

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Sensitive and quantitative workflows
for the system-wide and targeted study of
post-translational modifications**

Fynn Michael Hansen

aus

Husum, Deutschland

2022

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Professor Dr. Matthias Mann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 24.03.2022

Fynn Michael Hansen

Dissertation eingereicht am

30.03.2022

1. Gutachter: Hon.-Prof. Dr. Matthias Mann

2. Gutachter: Prof. Dr. Brenda Schulman

Mündliche Prüfung am

27.05.2022

Abstract

The post-translational modification of proteins is a key mechanism to orchestrate cellular life. A wide variety of PTMs is involved in the regulation of diverse cellular processes including the cell cycle, protein synthesis and degradation. Through the fine-tuned modulation of protein activity, localization and interaction, PTMs dynamically regulate the proteome and cellular homeostasis. Conversely, their dysregulation is linked to a multitude of malignancies including various forms of cancer and neurodegenerative disorders. Consequently, there is a large body of research on the role of PTMs to better understand their regulation in the context of health and disease.

Mass spectrometry (MS) has developed into the gold-standard for system-wide as well as for targeted PTM analysis. Improvements of sample preparation procedures, MS instrumentation and data analysis tools continuously push the limits of MS-based proteomics towards faster and more comprehensive proteome acquisitions. However, due to the substoichiometric abundance of modified proteins and the resulting requirement of PTM enrichment procedures, MS-based PTM analysis remains particularly challenging. In this thesis, I aimed to develop and improve methods for the quantification and characterization of post-translational modifications, with a special focus on protein ubiquitination and phosphorylation.

A main focus of this thesis was the establishment of a powerful data independent acquisition (DIA)-based workflow for system-wide and in-depth ubiquitinome analysis. Compared to label-free data dependent acquisition (DDA), this workflow almost doubles the number of identified modified peptides, allowing for the detection of around 35,000 of them in a single liquid chromatograph (LC) MS run. At the same time, it markedly increases the quantification accuracy while reducing the number of missing peptide identifications across samples compared to DDA. We employed our improved high performance DIA workflow for ubiquitinome analysis in the context of the circadian rhythm. This identified hundreds of cycling ubiquitination sites, many of which occurred in close proximity to each other.

When the SARS-CoV-2 pandemic hit the world, we contributed to a multilevel proteomic analysis workflow of SARS-CoV-2 and SARS-CoV infected cells. Here, we adapted our DIA-based ubiquitinome analysis workflow to provide state-of-the-art proteome and phosphoproteome analysis of the cellular signaling response of viral infections and highlighted potential drug candidates for the treatment of SARS-CoV-2 infections.

In contrast to the unbiased, system-wide analysis of the ubiquitinome, the identification of modification sites on specific proteins of interest provides valuable insight into protein structure. I helped to identify ISGylation sites on TRAF2 that are relevant for its regulation and ubiquitination sites on RIPK2 upon NOD signaling. In further collaborations with the Schulman group, I mapped ubiquitination sites on *in vitro* ubiquitinated proteins. Especially the identification of ubiquitination sites on Fbp1 provided valuable insights into the structural elucidation of a supramolecular chelate E3-Fbp1 complex. In yet another collaboration with the Schulman group, I developed a targeted MS method for the quantification of ubiquitin chain linkages. Using this method, we evaluated the effects of different side chain architectures of acceptor lysines for ubiquitin chain linkages.

Another main focus was the characterization of mitochondria across different tissues on proteome and phosphoproteome level, which generated interesting biological insights along with a data resource that can be explored by the community to generate and test new hypothesis. To this end, I devised a web application that allows easy data access to explore the dataset.

A central, often neglected aspect of high-performance LC-MS experiments is the chromatography itself, in particular the chromatographic column. Here I contributed to the development of a multiplex column packing station allowing fast and low-cost manufacturing of high-performance columns.

Finally, data analysis of MS-based experiments generates extensive lists of identified modification sites, whose structural and functional annotation is still mostly done manually. With the development of the AlphaMap tool, we automated the annotation of PTM sites, contributing to an informative, streamlined PTM site analysis workflow. Moreover, the integration of AlphaFold protein structure predictions enabled us to evaluate PTMs on 3D protein structures on a global level.

Contents

Abstract	ii
1 Introduction	1
1.1 The proteome: more than the sum of its parts	1
1.1.1 Genome and transcriptome dynamics	1
1.1.2 Proteome dynamics	1
1.1.3 The proteoform explosion	3
1.2 Mass spectrometry - based proteomics.....	5
1.2.1 Sample preparation.....	6
1.2.2 Liquid chromatography – mass spectrometry (LC-MS)	8
1.2.2.1 Liquid chromatography	8
1.2.2.2 The mass spectrometer	9
1.2.2.3 Tandem mass spectrometry	13
1.2.2.4 Data acquisition	15
1.2.3 Peptide and protein identification	18
1.2.4 Protein quantification	19
1.3 Post-translational modifications	23
1.3.1 Ubiquitin – opening Pandora’s Box.....	23
1.3.1.1 The ubiquitin system.....	24
1.3.1.2 The ubiquitin code	26
1.3.1.3 MS-based ubiquitinome analysis.....	27
1.3.2 Phosphorylation.....	29
1.3.2.1 MS-based phosphoproteome analysis	30
1.3.2.2 Phosphorylation in mitochondria	32
2 Aims of this thesis.....	34
3 Publications	36
3.1 Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology.....	36
3.2 Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV.....	50
3.3 A regulatory region on RIPK2 is required for XIAP binding and NOD signaling activity	65
3.4 Identification of covalent modifications regulating immune signaling complex composition and phenotype.....	80
3.5 Interconversion between Anticipatory and Active GID E3 Ubiquitin Ligase Conformations via Metabolically Driven Substrate Receptor Assembly	102
3.6 GID E3 ligase supramolecular chelate assembly configures multipronged ubiquitin targeting of an oligomeric metabolic enzyme	126

3.7	Linkage-specific ubiquitin chain formation depends on a lysine hydrocarbon ruler.....	155
3.8	Mitochondrial phosphoproteomes are functionally specialized across tissues.....	167
3.9	A New Parallel High-Pressure Packing System Enables Rapid Multiplexed Production of Capillary Columns.....	215
3.10	AlphaMap: an open-source Python package for the visual annotation of proteomics data with sequence-specific knowledge.....	225
3.11	The structural context of PTMs at a proteome wide scale.....	230
4	<i>Discussion and Outlook</i>	262
5	<i>References</i>	266
	<i>Abbreviations</i>	280
	<i>Acknowledgements</i>	282

1 Introduction

1.1 The proteome: more than the sum of its parts

1.1.1 Genome and transcriptome dynamics

Life comes in many forms and shapes and every organism has its own building plan - the genome. Although this blueprint of life is shared throughout all cells of an organism, individual cells can display an astonishing heterogeneity. Johanssen first coined the terms 'genotype' and 'phenotype' in 1909 to distinguish between the genetic material and the observable traits of an organism [1]. Since then, tremendous efforts have been made to elucidate the intricate relationship between genotype and phenotype. With their seminal work on genetic mutations in *Neurospora crassa*, Beadle and Tatum introduced the 'one-gene – one protein' hypothesis in 1941 [2], implying that a single gene affects a single step in a metabolic pathway. Shortly after, Avery described deoxyribonucleic acid (DNA) as the container of the genetic information [3] and Watson and Crick solved the structure of DNA [4]. Discoveries such as alternative splicing [5] soon indicated that the 'one-gene – one protein' concept is oversimplified. In 2003, upon sequencing of the complete human genome, the International Human Genome Sequencing Consortium reported the existence of 20,000 genes [6], which was strikingly less than the initially anticipated 80,000-100,000 genes [7]. However, the seemingly low number of genes give rise to >150,000 protein coding transcripts [8], highlighting a greater variability on transcriptome and proteome than on genome level (Figure 1). In contrast to the genome, the transcriptome and the proteome are both dynamically changing in response to external or internal stimuli.

1.1.2 Proteome dynamics

The proteome describes the compendium of all expressed protein molecules in a biological system and can be subject to extensive and multifaceted regulation on a qualitative and quantitative level. Protein synthesis and degradation are finely regulated, counteracting processes and allow the proteome to dynamically adapt to intrinsic and external stimuli. An important factor contributing to proteome variability is the dynamically regulated pool of messenger ribonucleic acid (mRNA) sequences. Splice variants, as well as alternative open reading frame (ORF) translation events can lead to a multitude of functionally distinct proteins originating from the same gene. Regulation

of the abundance of individual mRNA species, as well as their translation efficiency, further define the proteome composition on a quantitative level. The lifespan of individual proteins largely depends on their stability, which is influenced by their three-dimensional structure, integration into stable protein complexes, as well as subcellular localization. Protein degradation is primarily facilitated by the ubiquitin-proteasome system (UPS), autophagy or lysosomal degradation. Furthermore, the subcellular protein compositions at e.g. membranes or in individual cellular compartments enable specific functionalities. Although protein synthesis directly depends on the transcriptome and an estimated 54%-84% of proteome variation at steady state has been estimated to be reflected by mRNA variance, the proteome and transcriptome correlate poorly during transition phases, after for example an external stimulus [9-11]. Reasons for the poor correlation are manifold and can include, for instance, different half-life of mRNA and proteins, delayed signal transduction and translational adaptation. In addition to the transcriptional and translational variation post-translational modifications (PTMs) strongly contribute to proteome diversity; in fact, some consider PTMS to be the main contributor [12].

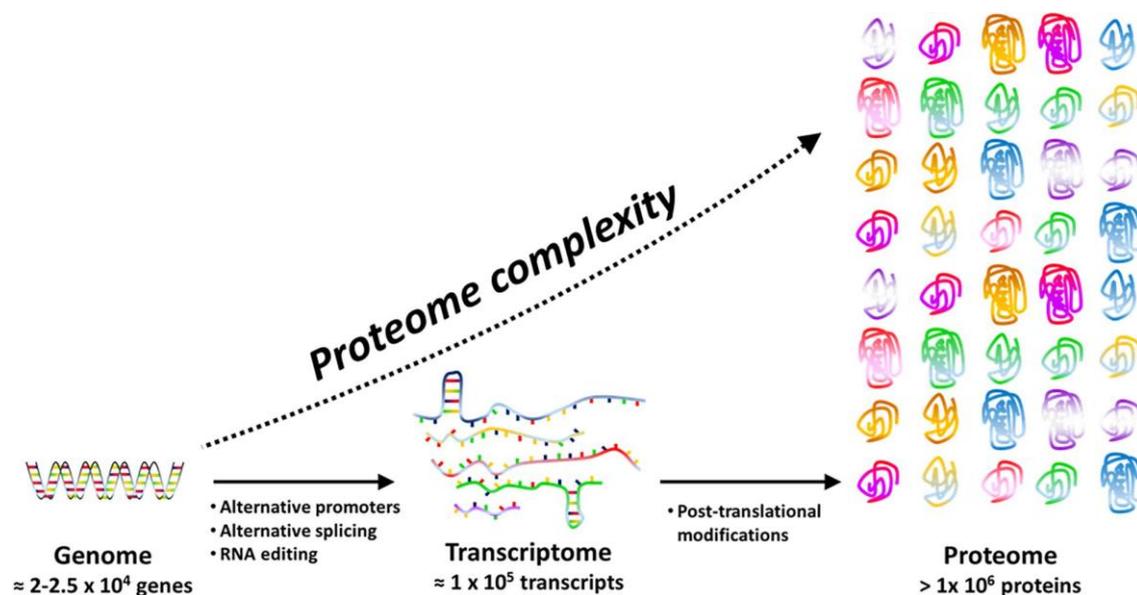


Figure 1 Proteome complexity – the proteoform explosion

The complexity of the proteome is based on transcriptional (e.g. alternative promoter usage), post-transcriptional (e.g. alternative splicing) and post-translational modifications (e.g. protein ubiquitination). Adopted from [13].

1.1.3 The proteoform explosion

As pointed out earlier, the ‘one-gene – one-protein’ concept did not withstand the test of time and should be rephrased as ‘one-gene – many proteins’. Individual members of proteins originating from the same genes are commonly referred to as protein isoforms and include alternative splicing or promotor usage variants. Post-translational modifications (PTMs), which by definition occur after protein synthesis, and genetic variations such as single-nucleotide polymorphisms are usually not included in this terminology and are instead referred to as proteoforms [14]. PTMs confer a tremendous variability to the proteome and can be classified into reversible and irreversible modifications (Figure 2) [15]. The latter include proteolytic cleavages, where for example specific signaling sequences are cleaved after protein transport to specific organelles such as the endoplasmic reticulum (ER) or the mitochondria and amino acid modifications such as deamidation, which alter protein stability. Reversible modifications, as the name implies, can be covalently attached and removed from proteins, which is mediated through enzymatic processes. For instance, the attachment and removal of phosphate or ubiquitin to and from substrate proteins is facilitated by kinases and phosphatases or E3 ligases and deubiquitinating enzymes (DUBs), respectively. Such reversible modifications can consist of the addition of complex molecules, small chemical groups and polypeptides for which glycosylation, phosphorylation and ubiquitination are prominent examples (Figure 2). To date, over 600 types of PTMs have been experimentally discovered (<http://www.uniprot.org/docs/ptmlist.txt>). Individual PTM types can display a huge variability themselves; for instance, while a phosphate group added to a protein of interest cannot be further modified, a ubiquitin molecule itself can be subject to multiple modifications [16]. The resulting set of possible PTMs combined with the potential co-occurrence of further PTMs on the same protein leads to an explosion of potential proteoforms (Figure 1). While the genome can be precisely determined and the size of the transcriptome can be estimated with some accuracy, the manifestation of the proteome with all its proteoforms remains difficult [17]. Nevertheless, unraveling the composition of proteoforms is a key for the understanding of intricate regulations of the proteome in health and disease. Mass spectrometry (MS) has experienced a tremendous development over the last decades and is the method of choice to decipher proteomes in a system-wide, unbiased way.

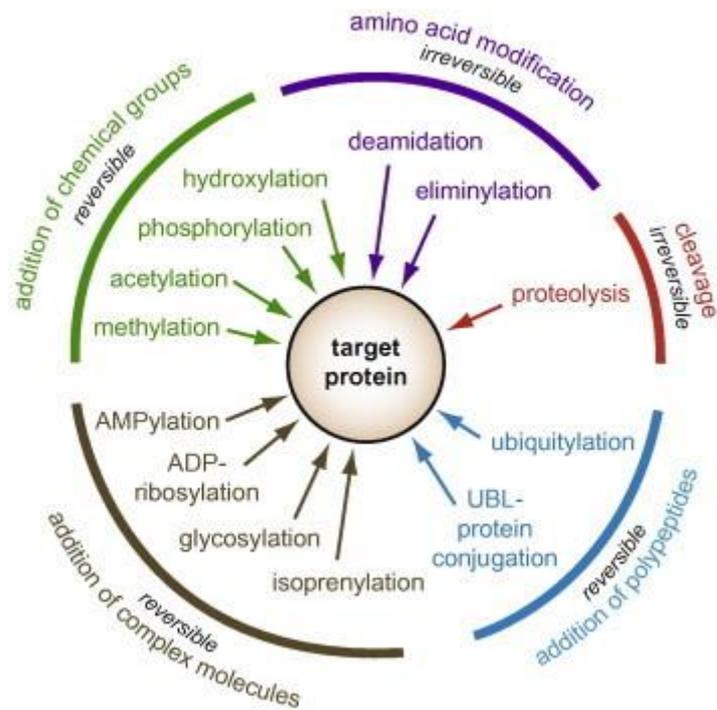


Figure 2 Compendium of PTM diversity

PTMs based on the character of the modification classify into two subgroupings: reversible or irreversible modifications, as well as the modification of the chemical structure of amino acid side chains such as the addition of chemical groups or complex molecules to specific amino acids, the covalent linkage of polypeptides, or the cleavage of the peptide bond between two amino acids. Adopted from [15].

1.2 Mass spectrometry - based proteomics

The underlying principle of mass-spectrometry is the detection of mass-to-charge ratios (m/z) of analyte ions and was already described at the end of the 19th century by Wilhelm Wien [18] and J. J. Thompson [19]. For a long time, the use of mass spectrometry was limited to the analysis of small molecules until almost a century later in the late 1980s “molecular elephants” – proteins – learned to fly. With the advent of matrix assisted laser desorption ionization (MALDI) [20] and electrospray ionization (ES) [21], it was possible to ionize large molecules such as proteins and analyze their m/z -values. Until this point, the sequence of polypeptides had commonly been determined by Edman degradation [22]. Thus, MS for protein analysis marked a turning point in modern protein sequence analysis [23]. Since then, improvements in MS instrumentation and measurement techniques have perpetually advanced the field of proteomics, making MS the gold standard for unbiased, system wide proteome analysis today.

MS-based experimental strategies for proteomics classically distinguish between bottom-up and top-down approaches. The latter describes the analysis of intact proteins and has the advantage of a very high sequence coverage and the ability to efficiently differentiate between proteoforms without the need to infer protein information from peptides [24]. Furthermore, the co-occurrence of multiple modifications on the same substrate protein can often only be analyzed with top-down approaches. However, due to poor ionization properties, complex ionization patterns and MS/MS spectra, the analysis of intact proteins from complex samples or even individual proteins is often not possible in practice. These constraints make top-down proteomics experimentally and computationally challenging and reduce the analytical depth compared to peptide-based approaches. Such peptide-based approaches are referred to as bottom-up or shotgun proteomics and imply that the protein information is inferred from these identified peptides. This approach is by far the most commonly applied technique for in-depth system-wide proteome analysis. Although computational analysis does not require expert knowledge, bottom-up approaches may be limited by low protein coverage and protein inference problems. It is for example not generally possible to distinguish if PTMs that were measured on different peptides derived from the same protein sequence were also present on the same or on different molecules.

Irrespective of the chosen experimental approach, MS-based proteomic workflows can be summarized into sample preparation, liquid chromatography, MS and data analysis.

The procedures discussed in this thesis are by no means exhaustive, but instead focus on the principles and instrumentation relevant to the work presented herein.

1.2.1 Sample preparation

Sample preparation is a crucial step in all MS-based experiments and it requires a high degree of robustness and reproducibility to provide a product of adequate quality for subsequent analytical steps. Although there is a huge variety of sample types and processing procedures, most proteomic sample preparation protocols share key features, such as protein extraction, reduction of disulfide bonds, alkylation of cysteines, enzymatic digestion and a final sample clean-up to remove MS-incompatible components and other contaminants (Figure 3A) [25].

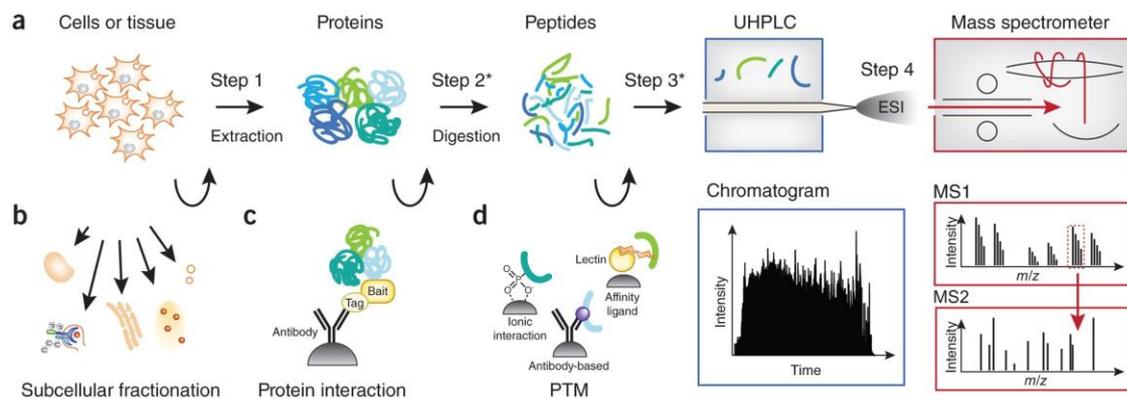


Figure 3 Bottom-up proteomic workflow.

A) Proteins are first extracted from biological material (step 1) and enzymatically digested to peptides (step 2). Peptide mixtures may be further fractionated or directly loaded onto ultrahigh pressure liquid chromatography systems (step 3). Peptides are ionized by electrospray (ES) ionization and analyzed via tandem mass spectrometry (step 4). B-D) Optional sample preparation, including the enrichment of subcellular compartments (B), specific proteins for e.g. interaction studies (C) and post-translationally modified peptides (D). Adopted from [25].

Protein extraction procedures are highly dependent on the sample type and may require mechanical (e.g. bead milling), physical (e.g. sonication), chemical (e.g. detergents) or biological (e.g. enzymes) procedures for efficient lysis [26]. Some biological questions may also involve the isolation of subcellular components such as the mitochondria and require more sophisticated extraction procedures (Figure 3B). To facilitate efficient proteolytic cleavage, all proteins of the resulting lysate or specifically enriched subsets (Figure 3C) are transferred into an unfolded state by first disrupting disulfide bonds with reducing agents like dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP). Reformation of these bonds is prevented by alkylation of cysteines with iodoacetamide

(IAA) or chloroacetamide (CAA). The ensuing protein digestion is most commonly performed with trypsin, which has a high cleavage specificity C-terminal to lysine and arginine residues [27]. This cleavage behavior renders the average peptide with two positive charges, one at the N-terminal amine and a second at the sidechain of lysine/arginine, which has positive effects for peptide ionization and detectability. However, generally not all produced peptides can be identified by MS, leading to an incomplete protein sequence coverage and loss of information. This is partially due to peptides that are either too short or too long for MS detection. The optimal length of peptides for MS detection ranges between 8-35 amino acids [28]. To improve sequence coverage, other enzymes like Glu-C, Lys-C, Asp-N, Arg-C, chymotrypsin, Lys-N and combinations of these can be employed [28].

Protein digestion is commonly performed in-solution at conditions that keep proteins in an unfolded denatured state, but at the same time keep the proteolytic enzyme in an active state [29]. However, classical biochemical protein analysis strategies may require a 1-D or 2-D gel electrophoresis step up-front protein digestion to for instance visualize a specific protein of interest. In such cases in-gel digestion procedures can be applied to digest proteins directly in the gel and extract peptides afterwards [30, 31]. These methods, however, usually have low throughput and are prone to experimental deviations reducing reproducibility.

Regardless of the digestion procedure, peptide mixtures require a sample clean-up step prior to MS measurement to remove contaminants and MS-incompatible components. These include for instance lipids, DNA and cell debris from the biological material or salts and detergents that were added for digestion. Sample clean-up is also often performed prior to the enrichment of modified peptides (Figure 3D). A simple, low-cost and broadly applicable solution for proteomics samples is to employ self-packed microcolumns termed STop And Go Extraction tips (StageTips) [32]. Stage-Tips are constructed of very small discs of an inert matrix containing separation material that are inserted into a pipet tip. Various stationary phase chemistries can be used for the sample clean-up procedure. While C18 is the most commonly used stationary phase chemistry, styrenedivinylbenzene - reversed phase sulfonate (SDB-RPS) can be used to efficiently remove detergents such as sodium deoxycholate (SDC), which is beneficial for optimal lysis and digestion conditions [33, 34]. In addition to the sample clean-up functionality, StageTips can further be used for sample concentration and fractionation.

Sample fractionation prior to LC-MS/MS is frequently used to increase proteome coverage and to build spectral libraries for data-independent acquisition (DIA) experiments [29] (see "Data acquisition"). Fractionation reduces the sample complexity and increases the total peptide amount analyzed across all fractions. Although

fractionation comes at the cost of increased analysis time, it increases the dynamic range and depth of analysis. In general, fractionation strategies can be divided in discrete and continuous approaches. StageTip fractionation is an example of the former, where different buffers with increasing elution strength are sequentially used to elute peptides. In contrast, continuous fractionation procedures make use of a binary buffer system with a continuous flow and sample collection into discrete fractions. The spider fractionator developed in our group is an example of such a continuous fractionation strategy [35]. A basic requirement for all off-line fractionation techniques prior to LC-MS/MS is a high orthogonality of both peptide separation techniques to grant an optimal isolation of peptide species in the retention time domain.

1.2.2 Liquid chromatography – mass spectrometry (LC-MS)

Mass spectrometry can be a stand-alone analysis technique without the need of additional sample separation up-front [36]. Flow injection analysis (FIA) and MALDI are examples of methods where no additional separation techniques are used [37-39]. These methods are popular for applications where a high throughput is required and low analytical depth can be accepted. However, for in-depth analysis of bottom-up proteomics additional separation that reduces the complexity of the ion mixture entering the mass spectrometer is required. Techniques based on the ion mobility rely on the separation of ions by their drift through a neutral gas in an electric field [38]. Although this technique is sufficient to separate ions, it is most commonly used in combination with liquid chromatographic separation. In fact, liquid chromatography – mass spectrometry (LC-MS) is by far the most widely used approach for proteomics experiments.

1.2.2.1 Liquid chromatography

The central motive for MS-coupled liquid chromatography is the reduction of the sample complexity before injection into the mass spectrometer for analysis. Chromatographic separation is based on the physicochemical properties of the analyte ions and their interaction with the mobile and stationary phase of an analytical column. Different stationary phases can be chosen for liquid chromatography, but reversed phase chromatography employing non-polar stationary phases is by far the most common. For the separation of peptides, C18 beads (long alkyl chains on silica particles) are used as stationary phase and a binary buffer system as mobile phase. Initially peptides bind to

the stationary phase, but then they are continuously eluted based on their hydrophobicity as the non-polar component in the mobile phase increases.

An elementary feature of an LC-system is its ability to efficiently separate different peptide species. The chromatographic performance can be summarized in the van Deemter equation (Height equivalent to a theoretical plate (HETP) = $A + B/u + Cu$), which is influenced by Eddy diffusion (A), random diffusion (B), mass transfer (C), and linear mobile-phase velocity (u) [40]. All factors can individually be adjusted to realize a favorable, small HETP value. A common chromatographic setup for bottom-up proteomics employs nano-flow high performance liquid chromatography (nano-HPLC) with flow rates between 100-400 nl/min in combination with small particle (< 3 μm diameter) packed capillaries with an inner diameter (ID) ranging from 50-150 μm . The weakest part in this setup is the packed capillary, which often has only weeks of peak performance before the performance drops. Although packed capillary columns are commercially available, many labs produce their own packed columns to save costs. To this end, we developed an improved column packing procedure, which combines the principles of high pressure and high-density slurry packing and allows capillary column packing times with <1.9 μm particles in 2 min [41].

The company Pharmafluidics recently introduced an alternative that avoids the intrinsic variability of packed bead beds and thus provides a very reproducible chromatography performance [42]. In their chip-based separation technique the flow-path is etched into a silica chip, creating a reproducible pillar architecture for the stationary phase. Although such commercial solutions are more expensive and not as flexible as custom-made columns, they provide the opportunity to establish a robust and reproducible chromatographic performance across labs.

1.2.2.2 The mass spectrometer

The mass spectrometer sits at the core of MS-based proteomics and advances in this field are closely linked to advances in instrumentation. It typically consists of three parts - an ionization source, a mass analyzer and a detector (Figure 4). Depending on the experimental needs, the user can choose between a variety of different techniques and instruments. Among the various ionization techniques, ESI is the one most widely used. Prominent mass analyzers include quadrupoles, ion traps (Orbitrap, quadrupole ion trap) and time-of-flight analyzers, which are often combined in modern MS instruments. The final part of a mass spectrometer is the detector, which is typically a variant of an

electron multiplier that amplifies the low signal originating from only few analyte ions or a detector that records an induced charge by oscillating ions in e.g. an Orbitrap.

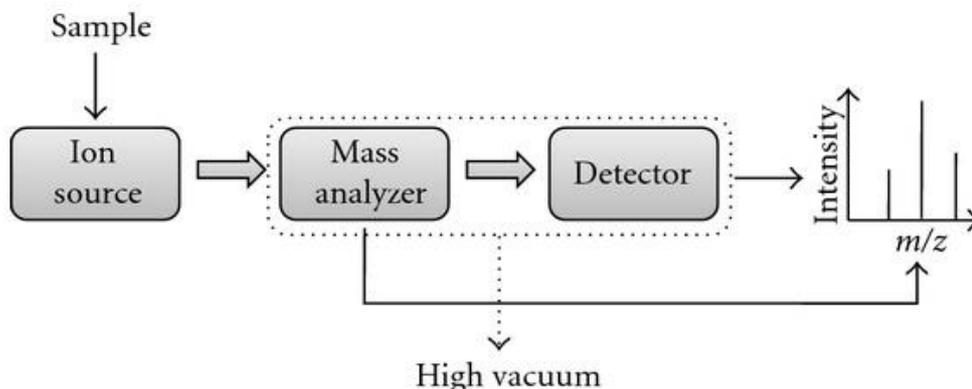


Figure 4 Components of a mass spectrometer.

A typical mass spectrometer consists of an ion source, a mass analyzer and a detector. Adopted from [43]

The most common source for ions is nano-HPLC in combination with ESI. Peptides eluting off a column first need to be ionized and transferred from liquid into the gas phase before they can enter the MS. This is achieved by ESI, a soft ionization technique for which John Fenn was awarded the Nobel Prize in 2002 [21]. By applying a kilovolt potential between the emitter, which is part of or connected to the column, and the transfer capillary, the entry of the mass spectrometer, a strong electric field is created (Figure 5) [44]. This electric field leads to the formation of a Taylor cone at the tip of the emitter. Small droplets leave the Taylor cone in a jet, which disperses into little droplets, the plume. As solvent evaporates, charged droplets become unstable until they reach the Rayleigh limit and further disperse into smaller droplets by coulomb fission. The formation of charged analyte ions in atmosphere is generally described by two models, which may also act together: the charge residue model (CRM) [45], where cycles of solvent evaporation and coulomb fission continue until only one analyte ion remains and the ion evaporation model (IEM) [46], where ions enter the gas phase through field desorption (Figure 5).

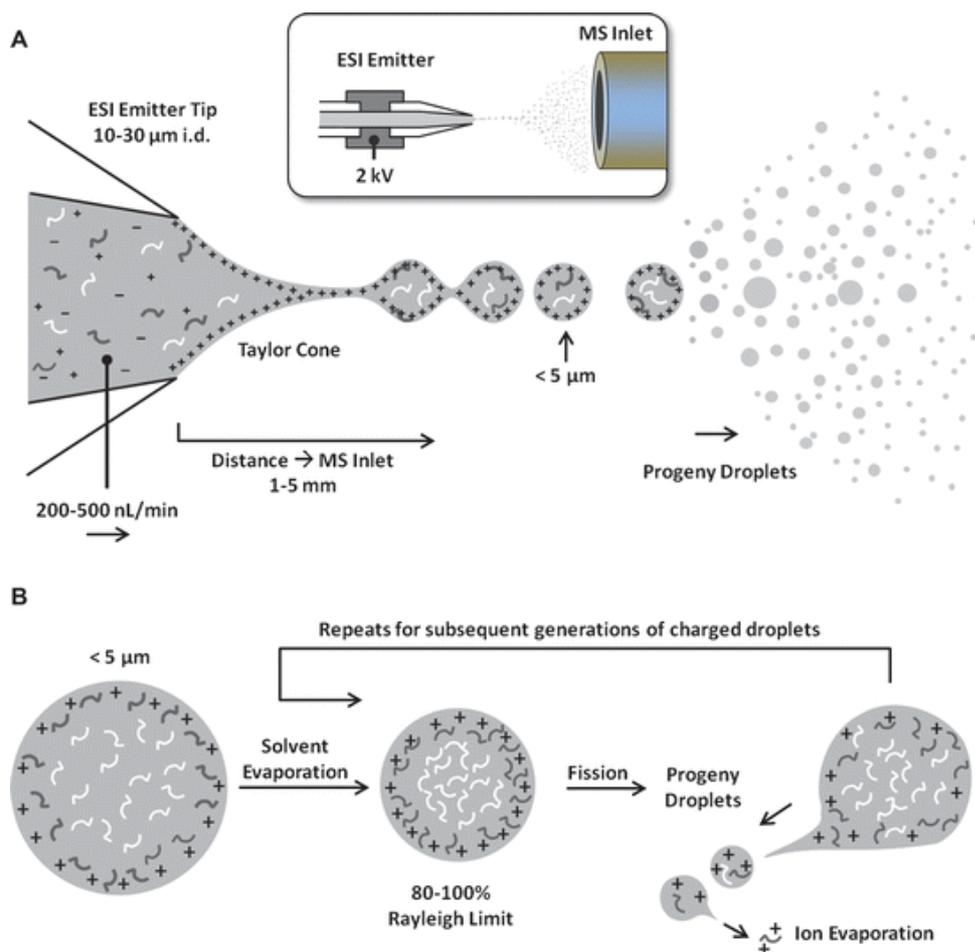


Figure 5 Electro spray ionization – elephants learn to fly

A. A strong electrical field is applied between an ESI emitter and the MS inlet. This electric field leads to the formation of a Taylor cone on the at the emitter tip, from where analytes dissolved in small droplets become airborne. Solvent evaporation shrinks droplets until the Rayleigh limit is reached and droplets disperse into smaller droplets. B. Repeated cycles, until all solvent is evaporated and only charged analytes remain, describe the charge residue model (CRM) while ejection of charged analyte ions form solvents describe the ion evaporation model (IEM). Adapted from [44].

Ions entering the mass spectrometer are analyzed in regard to their masses, or more precisely to their m/z -value. This is achieved through one or more mass analyzers. A popular one is the quadrupole, which consists of four cylindrical rods that are positioned equidistant from a center axis (Figure 6) [47]. A radio frequency (RF) voltage can be applied to opposing rods, which creates an oscillating electrical field. Depending on the applied RF only ion-trajectories of specific m/z -values are stabilized and can pass the quadrupole for detection. Quadrupoles are robust, have a high reproducibility and sensitivity, but suffer in speed and resolution. They are often used as mass filters in tandem with other mass analyzers like time-of-flight (TOF) or Orbitrap mass analyzers. TOF mass analyzers make use of the direct proportionality between the square root of the mass and the drift time of analyte ions that were accelerated with the same kinetic energy through a field free vacuum. Although these analyzers have a very high scan

speed and sensitivity, they used to suffer from a low resolving power [48]. Recently, through the combination of trapped ion mobility spectrometry (tims) with TOF mass spectrometers, TOF instruments have regained significant interest in the field of proteomics. The ability to separate precursor ions based on their ion mobility grants timsTOF instruments an additional dimension of precursor separation. In contrast, the Orbitrap is a comparably new mass analyzer that consists of an inner spindle and an outer barrel-shaped electrode, that is split into two parts [49] (Figure 6). Ions in the Orbitrap circle around the inner spindle, while moving back and forth the length of the spindle. The resulting harmonic axial oscillation is proportional to the m/z -values and is used as the readout for m/z -value detection [50]. Due to their high accuracy and resolution, Orbitraps are currently the most widely used mass analyzers in the field of proteomics.

Modern mass spectrometers allow in addition to the analysis of the m/z values also the filtering of analyte ions based on their ion mobility (IM). For instance, interfacing of the high-field asymmetric waveform ion mobility spectrometry (FAIMS) device with orbitrap instruments showed to increase proteomic depth [51]. An alternative strategy for IM separation uses trapped ion mobility spectrometry in combination with the parallel accumulation–serial fragmentation (PASEF) scan mode on TOF machines [52-54]. In general, IM provides an additional dimension for peptide separation, thus, increasing the resolving power for individual peptide analytes.

The actual detection of the signal imposed by analyte ions depends on the mass analyzer used upfront. The quadrupole and TOF mass analyzers employ electron multipliers that produce a measurable current upon analyte ion impact. Microchannel plate ion detectors, for instance, interfaced with a digitizer for signal recording are used in modern timsTOF instruments [52, 53].

In contrast, the Orbitrap records an induced charge difference between the two halves of the outer electrode. This image current is digitized and transferred via Fourier transformation into a mass spectrum.

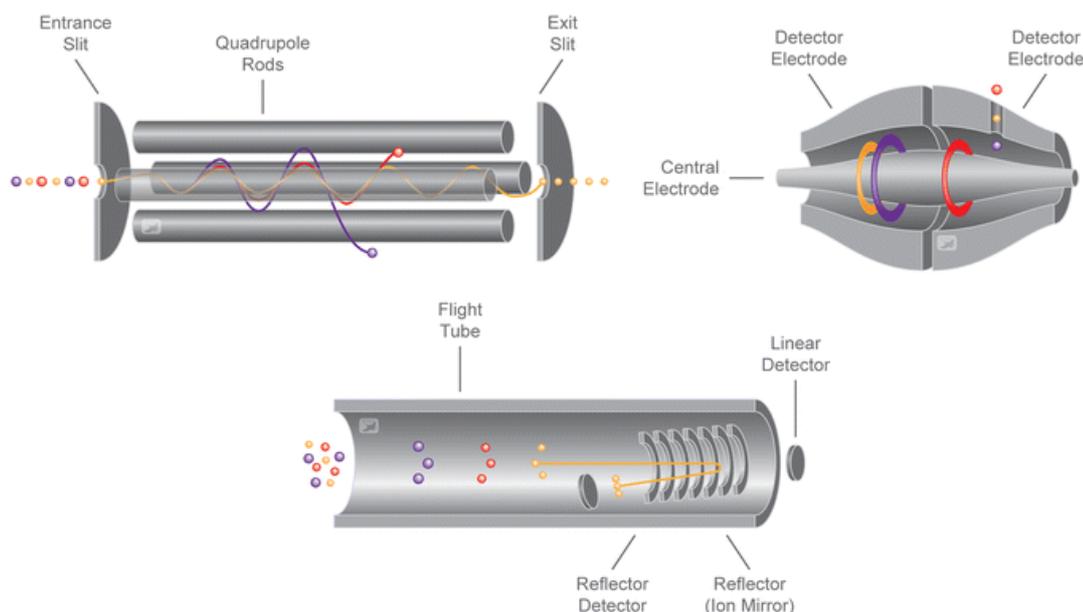


Figure 6 Mass analyzer

Three commonly used mass analyzers – upper left a quadrupole, upper right an Orbitrap and bottom a time of flight (TOF) mass analyzer. Adapted from [47]

1.2.2.3 Tandem mass spectrometry

For the complete characterization of peptides, their mass alone is not sufficient. Further sequence information is needed to more fully characterize a peptide and assign it to a protein. For this, tandem mass spectrometry is employed. The combination of a full MS scan (survey or MS1 scan) which records the m/z -values of the intact peptide and a tandem scan (MS/MS, MS2 scan) that records the m/z values of fragment ions together create a peptide fingerprint that is used for identification (Figure 7A).

Peptide precursor ions can be fragmented by various techniques such as collision-induced dissociation (CID), electron capture/transfer methods or photodissociation. Each method has different propensities to generate fragment ions (product ions), which are characterized by the peptide backbone bond that is broken (Figure 7B). In CID, peptides are collided in ion traps with an inert gas such as helium, which leads predominantly to the breakage of peptide bonds, producing b (N-terminal) or y (C-terminal) ions. The resonance excitation principle, however, leads to weak fragmentation yields and a low molecular mass cutoff in a three-dimensional ion trap [55]. In contrast, higher-energy collision dissociation (HCD), a special form of CID that was devised for the use with Orbitrap analyzers, allows consecutive fragmentations based on the non-resonance excitation principle [55, 56]. This leads to a broader

fragment ion range including also low mass fragments. Fragmentation in HCD is performed in a multipole collision cell, whereafter fragment ions are transferred into a C-trap, a bend version of a quadrupole, focused into small ion packages and injected into the Orbitrap. HCD on Orbitrap mass spectrometers generates high-quality MS spectra with high resolution and mass accuracy and is currently the most widely applied dissociation technique. Nevertheless, further fragmentation strategies can provide additional information. For instance, photodissociation methods like ultraviolet photodissociation (UVPD) [57-59] use high energy photons for dissociation, which in addition to b/y ions also produces a/x ions [60]. Furthermore, in electron transfer dissociation (ETD) [61], the transfer of an electron to a peptide leads to an N-C α bond breakage, generating c/z ions. ETD was also shown to be beneficial for PTM analysis, as labile modifications tend to be retained during fragmentation.

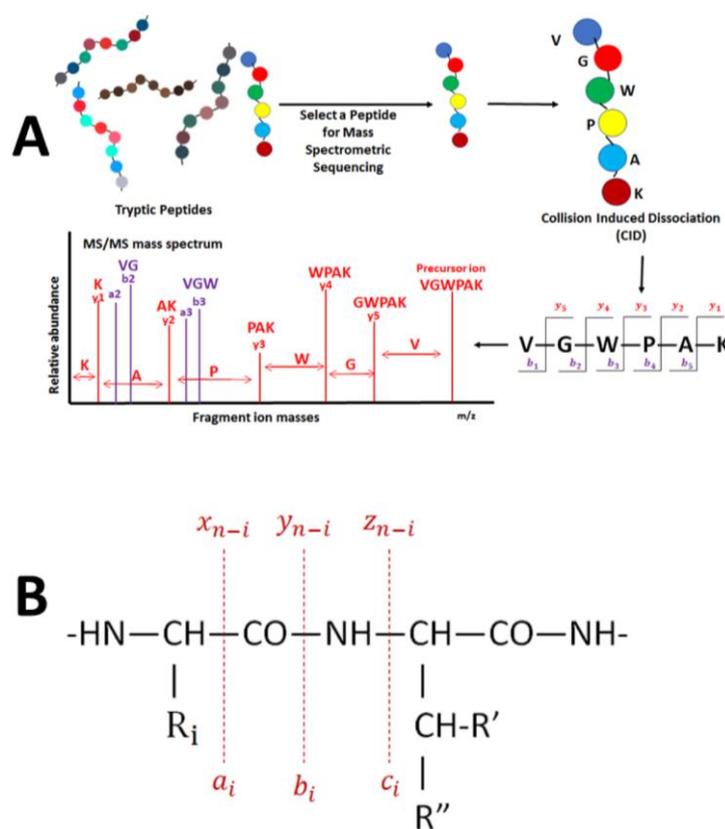


Figure 7 Peptide fragmentation

A) Generation of a tandem mass spectrum. A peptide is selected for fragmentation and resulting fragment ions are detected and visualized in a tandem mass spectrum. Mass differences between fragments allow the deduction of the peptide sequence. B) Peptide fragmentation can occur at three different bond types along the peptide backbone generating a/x, b/y and c/z fragment ions. Adapted from [62]

1.2.2.4 Data acquisition

Mass spectrometry offers multiple data acquisition strategies, which make different use of the individual components implemented in the mass spectrometer. Acquisition strategies can be separated into targeted and untargeted approaches. While the aim of the latter is the unbiased identification of as many peptides in a sample as possible, the former often aims at the reproducible and robust quantification of only few peptides [63].

Currently, the most widely used untargeted technique is data dependent acquisition (DDA). In a MS1 scan, the TopN most intense precursor ions are selected for fragmentation in consecutive MS2 acquisitions (Figure 8A-C). Precursor ions are isolated through a quadrupole operated as a mass filter for a small m/z -value range (typically 1.4 Thompson) around the desired m/z -value. The succession of one MS1 scan and multiple corresponding MS2 scans is called a duty cycle and depends on parameters such as mass resolution, ion collection time and the number of selected precursors for fragmentation. To avoid re-fragmentation of the same precursor ions over the course of their chromatographic elution, a dynamic exclusion time is applied, preventing further picking of these precursor ions. For sufficient chromatographic peak reconstruction, which is important for accurate quantification (see below), in practice at least 4-6 data points per chromatographic peak are required. Parameter adjustments aim to strike an optimal balance between analytical depth and accuracy. While DDA generates easy to interpret MS2 spectra, the intensity-based precursor selection is biased towards peptide identification of more abundant peptides. Furthermore, the semi-stochastic nature of precursor picking tends to produce missing values across multiple samples, which may impede accurate quantitative comparisons.

An alternative method is data-independent acquisition (DIA), which was already described two decades ago, but only gained traction in recent years [64-67]. In DIA, contrasting to DDA, the selection of m/z -ranges for fragmentation is not dependent on the MS1 scan. Instead, the whole m/z -range is split into mass windows, which each contain many peptides (Figure 8F-H). Fragmentation of all peptides in such a mass window leads to immensely complex MS2 spectra which need specialized software tools for analysis. However, this approach allows an unbiased sampling of all precursor ions and results in a higher dynamic range in trapping instruments such as the orbitrap and less missing values across multiple samples. Based on their data acquisition paradigm DIA strategies are classified as scanning methods, but data analysis can be performed in a targeted or untargeted fashion. Most implementations of DIA classically rely on a prerecorded spectral library, which is used to match peptide fragments in the actual DIA

measurement [68]. This library contains “peptide query parameters” (PQP) which include information about peptide sequence(s) for a given protein, peptide charge state distribution, fragment ions, fragmentation patterns and peptide retention times. This ‘peptide centric’ approach is often defined as a targeted analysis method. In recent years however, library-free DIA, which obtains all essential information from the acquired DIA measurements themselves [69] has become increasingly popular and can be regarded to as an untargeted analysis method.

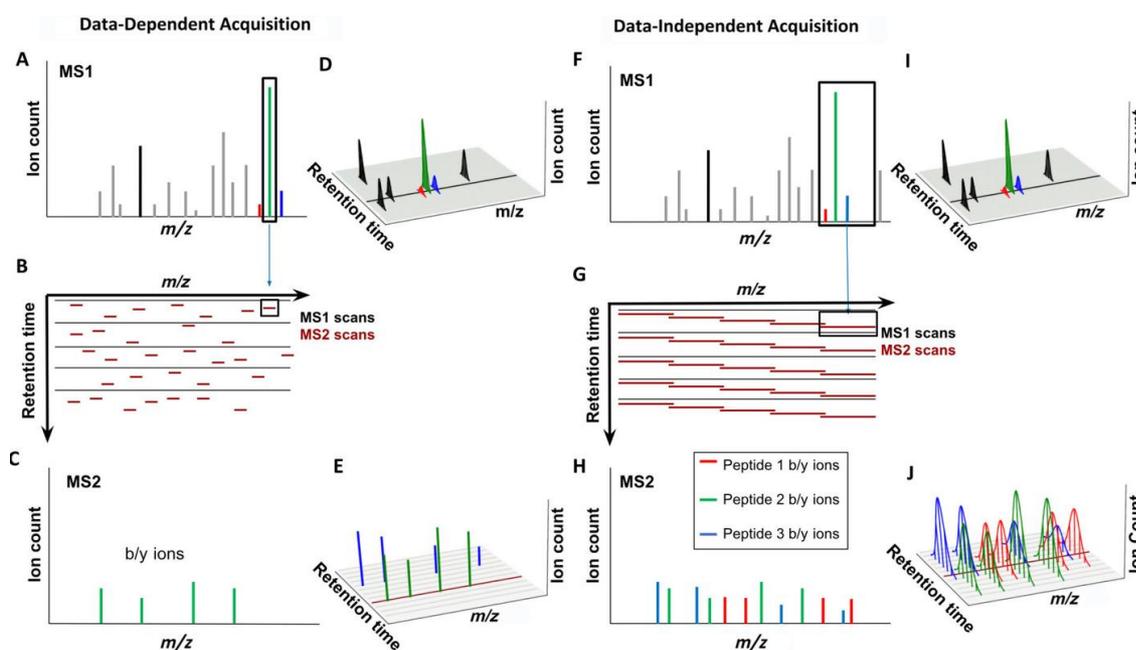


Figure 8 Conceptual comparison of data-dependent and data-independent acquisition.

In Data dependent-acquisition (DDA) (A-E) A) the most intense precursor ions, based on an MS1 survey scan are selected for isolation and fragmentation. B) The recording of a MS1 scan (long black line) is followed by multiple MS2 scans (short red lines) that semi-stochastically select the most intense precursor ions for fragmentation which is referred to as duty cycle. C) Fragmented precursor ions are recorded in a MS2 spectrum that ideally only contains fragments of the targeted precursor ion. D) MS1 signals are continuously recorded and used for precursor ion quantification. E) MS2 signals, in contrast are only semi-stochastically recorded for individually selected precursors and cannot be used for quantitation. Data-independent-acquisition (DIA) (F-J) F) selects all precursor ions of a predefined m/z window for fragmentation. G) After an MS1 scan the whole m/z range is sampled in predefined m/z windows for fragmentation. H) Resulting MS2 spectra display complex fragment ion mixtures of many precursor ions. I) Similar to DDA, the MS1 signal is continuously monitored and could be used for precursor ion quantification. J) In contrast to DDA, the DIA scheme continuously records fragment ions and allows the use of these fragment ions for precursor quantification. Adapted from [70]

While DDA and DIA are both hypothesis-free methods for the acquisition of whole proteomes, targeted methods like single ion monitoring (SIM) [71], selected or parallel reaction monitoring (SRM, PRM) [72] aim at the confident identification and quantification of single or few analyte ions (Figure 9). In SIM, a specific precursor ion is filtered by a quadrupole operated in filter mode, followed by the detection of the intact

peptide ion. Although only MS1 level information is used for quantification, additional MS2 scans can be acquired to confirm the peptide identity [73]. In SRM, MRM and PRM, in contrast to SIM, MS2 level information is used for quantitation. In SRM, the capabilities of triple quadrupole (QQQ) mass spectrometers are exploited [74]. A precursor ion is isolated by the first quadrupole, and then fragmented in the second quadrupole, which is followed by recording of a specific fragment ion in the third quadrupole. For more specificity, multiple fragments of the same precursor are detected sequentially which is then referred to as multiple reaction monitoring (MRM). In PRM, Orbitrap mass spectrometers instead of a QQQ mass spectrometer are used. This has the great advantage of much higher specificity and allows the parallel detection of multiple fragment ions of a selected precursor instead of the sequential detection in SRM/MRM. Although targeted analysis of individual peptides requires laborious method developments, they can offer excellent sensitivity and quantitative accuracy.

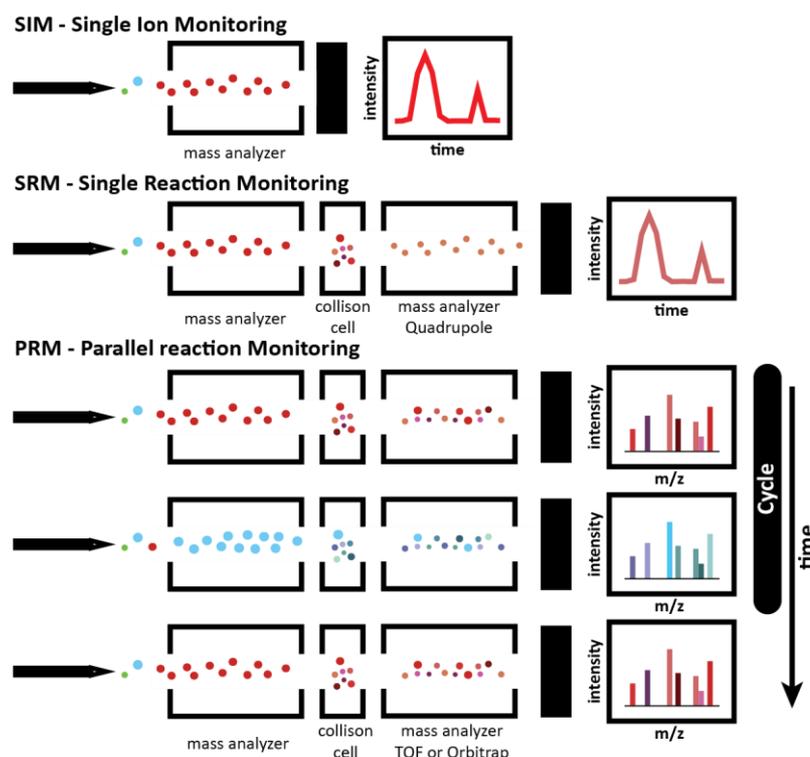


Figure 9 Overview of targeted acquisition methods

In Single Ion Monitoring (SIM) (top) precursor ions are filtered in a mass analyzer for direct detection via a mass detector. For Selected reaction monitoring (SRM) (middle) a precursor is selected by a mass analyzer operated in filter mode and fragmented in a collision cell. A second mass analyzer, which is typically a quadrupole for SRM, filters for a specific m/z that is subsequently detected. Similar to SRM, Parallel Reaction Monitoring (PRM) first selects and fragments a specific precursor ion, but uses all fragment ions for parallel detection, which is typically achieved with Orbitrap or time-of-flight mass analyzers and corresponding detection methods. [63]

1.2.3 Peptide and protein identification

The central motivation of classical MS experiments aiming at the characterization of whole proteomes is the identification and quantification of proteins. In bottom-up proteomics, this information needs to be inferred from the peptide level. Various data acquisition strategies can be employed to obtain experimental data and the selection of the most suitable method greatly depends on the scientific question. Regardless of the employed data acquisition strategy, the vast amount of data generated in MS-based experiments by far surpasses the capabilities of manual spectrum annotation, thus making software tools imperative. The list of software solutions is extensive and also depends on the type of MS-experiment. Currently, prominent tools are MaxQuant [75], MS-Fragger [76], Skyline [77], Spectronaut [78] and DIA-NN [79]. Two approaches for the analysis of acquired data and peptide identification can be distinguished – peptide-centric and spectrum-centric ones.

Spectrum-centric approaches employ a user specified sequence database, which is in-silico digested based on the experimentally used proteases. Hereafter, algorithms such as Andromeda for MaxQuant [80] search each experimentally acquired spectrum for the best matching peptide sequence and score the quality of established peptide-spectrum matches (PSM). These PSMs need to be statistically validated to ensure correct identification. A frequently used strategy to control the false-discovery rate (FDR) is the target-decoy strategy (reviewed in [81]). Here, spectra are not only searched against the proteome-based in-silico generated peptides (targets), but also for the same peptides with the reversed amino acid sequence (decoys). Based on the number of PSMs matching to the Decoy database, the PSM-score can be adjusted to only include an acceptable number of false identifications, which is usually 1% in discovery proteomics. DDA experiments, but also library-free DIA experiments, commonly employ spectrum-centric strategies [69].

Peptide-centric approaches use a priori knowledge about peptides and test whether these peptides are observed in a sample with a certain confidence. For library-based DIA experiments, PQPs are used to score detected peptide signals. Similar to the FDR control described for spectrum centric approaches, reversing the sequence of query peptides is a common strategy for FDR control in peptide-centric approaches [68, 82].

The detection and subsequent identification of peptides across multiple samples is crucial for robust quantification and strongly depends on the data acquisition strategy. DDA is still the most common approach for data acquisition and easy to implement with

relatively straightforward data analysis. However, the semi-stochastic precursor picking favors the more constant detection of abundant peptide species and can lead to missing peptide identifications across multiple samples. In contrast, the DIA paradigm is not biased to the analysis of high abundant peptides species and covers a higher dynamic range than DDA approaches, but data analysis requires specialized analysis tools for the deconvolution of complex MS2 spectra. Since the whole m/z range is systematically used for data acquisition, DIA generates considerably fewer missing values across samples. Such missing peptide values are especially detrimental for MS experiments that rely on peptide level information as final readout. The analysis of PTMs, for instance, typically reports modified peptides and in contrast to protein identification, that often profits from multiple peptides, a missing value cannot be compensated.

After search engines have confidently identified peptides, the correct set of proteins from which these peptides originate needs to be identified. Because of sequence similarities between proteins, e.g. isoforms, a clear one-to-one assignment of peptides to proteins is often not possible. The deduction of proteins from a complex peptide mixture containing peptides that can be explicitly assigned to one protein and those that cannot are collected into 'protein groups'. A prominent method addressing this problem is the parsimony principle, a version of the Occam's razor constraint that reports a minimal list of proteins explaining all peptide identifications [83].

1.2.4 Protein quantification

The confident identification of proteins is essential to uncover the qualitative proteome composition, but to gain a deeper understanding of molecular processes, changes in protein abundance need to be quantitatively evaluated. Protein quantification strategies can be classified into 'absolute' and 'relative' quantification methods. Relative quantification methods evaluate the difference between proteomes of different states, whereas absolute quantification aims to quantify the expression level or concentration of each protein species in a given sample. Absolute quantification usually relies on the spike-in of heavy labeled, synthetic proteins or peptides for quantification and is significantly more complex than relative quantification. Another classification of quantification strategies differentiates between 'label-free' and 'label-based' methods (Figure 10).

Label-free quantification (LFQ) is a straightforward relative quantification strategy that does not require any additional sample preparation steps. The easiest form of LFQ is performed by spectral counting, which uses the correlation of the number of tandem MS

spectra with the abundance of the respective protein [84]. A different approach uses the MS1 signal intensity of precursor ions over the course of the chromatographic peptide elution to create an extracted ion chromatogram (XIC). The area of this XIC is then used for the relative quantification of peptides across different samples (Figure 8D,E) [85]. LFQ strategies based on DDA often suffer from missing values and variability between individual MS measurements, which impedes reliable quantification across samples. However, feature detection algorithms such as 'Match between Runs' (MBR), that allow for peptide identification by using spectral information of other samples [86], and sophisticated sample normalization strategies, such as MaxLFQ that employs a non-linear optimization model for intensity normalization across samples [87], greatly improve the performance of standard LFQ experiments. The 'proteomic ruler', an absolute LFQ approach, uses the relationship between summed histone intensities, DNA amount and cell numbers to estimate protein copy numbers with surprising accuracy [88]. In contrast to DDA, LFQ based on DIA uses MS2 level information for quantification, which enables the recording of multiple quantitative values for a given precursor peptide and which is usually more robust. This is possible because the data acquisition scheme continuously covers the whole m/z range (Figure 8 I,J).

Label-based quantification strategies are classified into metabolic and chemical labeling methods, which both allow the multiplexing of samples. Multiplexing describes the combination of labeled samples for MS analysis, which reduces overall analysis time as well as experimental variability. Over the last decades, many chemical labeling strategies have been described that introduce various labels on protein or, more commonly, at the peptide level [70]. Common chemical labeling methods for bottom-up proteomics use isobaric tags, among which amine-reactive tags are most popular [89]. Especially, tandem mass tags (TMT), which consist of three groups that are readily cleaved by CID-based fragmentation, are widely used [90]. The reporter and mass normalizer group harbor a label characteristic distribution of isotopes and are connected to a NH_2 reactive group that enables the labeling of amines. Labeled peptides have the same chromatographic behavior and mass-to-charge ratio making them indistinguishable at the MS1 level. Upon precursor fragmentation for tandem mass spectrometry, however, sample specific reporter ions are generated and used for relative peptide quantification across samples. A major drawback of this technique is ratio compression that arises from co-fragmentation of not completely separated precursor ions and can lead to under estimation of true fold changes. This issue is often addressed by additional fragmentation steps of tag containing fragment ions (MS3) [91] or computational approaches to filter for precursor isolation purity [92, 93]. Chemical

labeling strategies that rely on the quantification of low m/z reporter ions are typically only possible with DDA strategies. Here, only individual precursor ions are selected for fragmentation and resulting reporter ions can easily be assigned to the precursor ions. This is not possible for DIA, since multiple precursor ions are simultaneously selected for fragmentation and reporter ion signals cannot be assigned to their respective precursors. EASI-tag is a chemical labeling strategy compatible with DIA, where labeled peptides retain a characteristic mass shift after fragmentation [94].

Metabolic labeling strategies, in contrast to chemical labeling, rely on the metabolic incorporation of stable 'heavy' isotopes (e.g. exchange of N^{14} for N^{15}), which allows early stage sample multiplexing. Similar to chemical labeling strategies, many metabolic labeling methods have been developed of which stable isotope labeling by amino acids in cell culture (SILAC) is the most popular [70, 95]. In SILAC 'light' amino acids (typically leucine, lysine or arginine) are replaced with their 'heavy' counterparts in a culture medium. Labeled lysine and arginine are commonly employed for standard bottom-up proteomic experiments using trypsin for digestion. While heavy labeled peptides, except for deuterium labeled peptides, have the same chromatographic behavior as light peptides, they can readily be distinguished by their mass-to-charge ratio at the MS1 level. A drawback of such quantification strategies is the increase in MS1 complexity, which usually results in a decrease of protein identification. Furthermore, these techniques are often limited to cell culture models, since the labeling of whole organisms, although possible, is time consuming and expensive. Super SILAC is an alternative approach to the labeling of whole organisms [96]. Here, a mixture of heavy labeled cells from different cell lines is spiked into unlabeled samples and used as a reference standard to enable relative quantification of multiple unlabeled samples against each other. The spike-in of just a few labeled proteins and peptides is a frequently used technique to determine absolute quantities of unlabeled proteins and peptides in a sample, respectively. The Protein Standard Absolute Quantification (PSAQ) strategy, for instance, uses heavy labeled full-length proteins of known concentration as spike-in to quantify a protein of interest [97]. Similarly, heavy synthetic peptides (AQUA peptides) can be used for absolute quantification on peptide level [98]. Although these spike-in methods in principle allow the absolute quantification of target proteins, they are limited to the analysis of only few proteins or peptides.

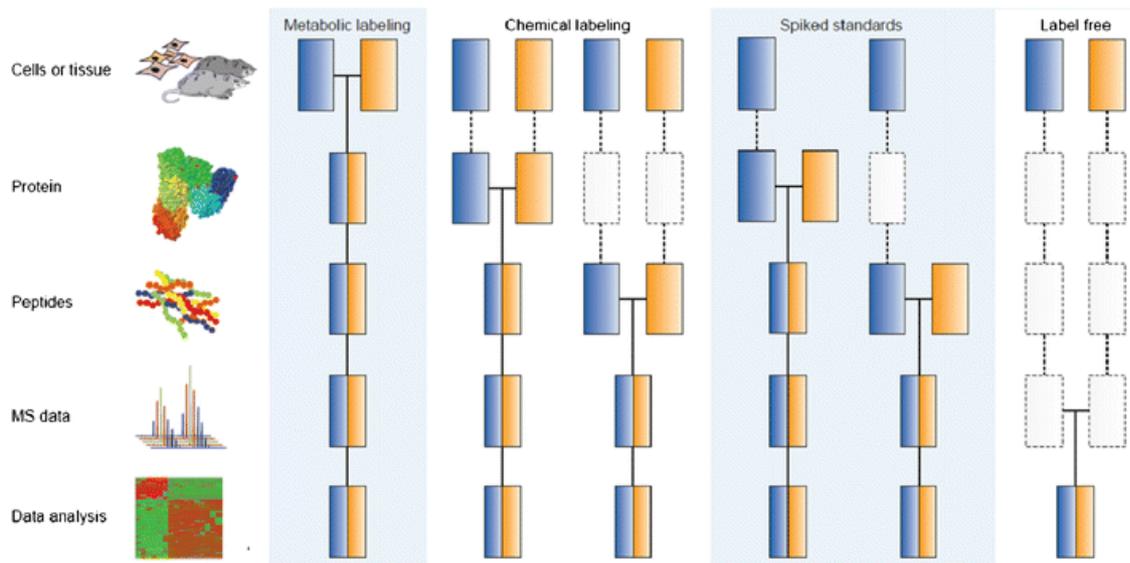


Figure 10 Common quantification strategies in proteomic experiments

Blue and yellow boxes represent different experimental conditions. Solid horizontal lines indicate sample combination. Dotted vertical lines and boxes highlight process steps where experimental variation could lead to quantification errors. Adopted from [99]

1.3 Post-translational modifications

The vast set of PTMs allows for a fine-tuned and dynamic adaption of the proteome to a changing environment. PTMs are involved in a plethora of cellular processes and dysregulation of these can have detrimental effects on biological systems. Understanding the intricate regulations and biological implications of PTMs is a challenging task and requires sophisticated experimental procedures. Mass spectrometry has proven to be a key technique for the study of PTMs and especially bottom-up strategies are irreplaceable for the system-wide analysis of PTM landscapes. However, the analysis of PTMs using bottom-up proteomics faces several challenges. Due to the notoriously low abundances of PTMs, additional enrichment strategies are required for their detection via mass spectrometry. In fact, efficient and robust enrichment strategies are a crucial aspect to elucidate the full extent of a modification. The impact of an appropriate enrichment strategy is obvious when looking at the tremendous gain of analytical depth brought by the development of a modification specific antibody for ubiquitinome analysis (see 'MS-based ubiquitinome analysis'). Furthermore, quantification of PTMs often only relies on the robust identification of a single modified peptide, whereas for protein identification and quantification multiple peptides can be used. Especially DDA methods with semi-stochastic precursor picking are prone to missing values across samples. To improve the reproducibility of peptide identification DDA strategies are often combined with metabolic or chemical labeling strategies. Additionally, in contrast to the mere identification of a peptide, the confident localization of a modification typically requires a higher peptide sequence coverage by fragment ions.

Depending on the PTM of interest, specialized workflows need to be established. For two of the most widely studied PTMs, phosphorylation and ubiquitination, continuous improvements of experimental workflows and instrumentation now allow routine identification of thousands of modification sites.

1.3.1 Ubiquitin – opening Pandora's Box

Ubiquitin was first described by Goldstein and colleagues in 1975 as "a universal constituent of living cells" [100]. In fact, the sequence of ubiquitin genes implied a strongly constrained evolution and the short 76 amino acid long protein sequence with 72 conserved amino acids between yeast, animals and plants is nearly identical [101-103]. Soon after the first reports of free ubiquitin, ubiquitin was also found to be covalently conjugated to other proteins [104, 105], which was subsequently proposed to

be a signal for downstream proteases [106]. Between 1981 and 1983 Hershko, Ciechanover and Rose described a multistep process for the tagging of ubiquitin to proteins featuring E1, E2 and E3 enzymes – the ubiquitin system [107-109]. This work was awarded with the Nobel Prize in 2004. In the following years, the proteasome was identified to be the ATP-dependent protease responsible for ubiquitin-dependent protein degradation [110, 111].

Research on ubiquitin has revealed an unexpected relevance of this small protein to a broad range of cellular processes that could not have been foreseen at its first discovery. The ubiquitin-proteasome-system (UPS) has revolutionized the view on protein degradation and opened up an immensely complex field of research [112]. It is now evident that ubiquitin and the UPS are involved in a plethora of intricate processes and actively regulate cellular homeostasis. Deregulation of this finely regulated system can lead to a multitude of diseases making the study of ubiquitin in all its facets a primary challenge of research.

1.3.1.1 The ubiquitin system

The attachment of ubiquitin to a substrate protein is a multistep mechanism sequentially mediated by members of three enzyme classes: ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s) and ubiquitin ligases (E3s) [113] (Figure 11). Today, two E1s, close to 30 E2s and over 600 E3s have been identified in the human genome. In the conserved first step of the ubiquitination cascade, an E1 activates ubiquitin in an ATP-dependent manner to form a thioester bond between the C-terminal carboxyl group of ubiquitin and a cysteine residue of the E1. Thereafter, ubiquitin is transferred onto an active site cysteine of an E2 via a transthioylation reaction. In the final step, an E3 interacts with the ubiquitin loaded E2 and a substrate protein to transfer the ubiquitin onto an amine group of a lysine residue in the substrate protein. Instead of a comparably weaker thioester bond seen for E1-UB and E2-UB conjugates, this last ubiquitin transfer results in the formation of an energetically more favorable iso-peptide bond.

E3s govern the efficiency and specificity of the ubiquitination process, and are commonly, depending on the presence of functional domains, differentiated into three main classes: RING (really interesting new gene), HECT (homologous to the E6AP carboxyl terminus) and RBR (Ring-between-Ring) ligases [114]. RING E3s are by far the largest group (>600) of ubiquitin ligases and characterized by their RING or U-box

fold catalytic domain. These ligases transfer ubiquitin in the final step of the ubiquitination cascade directly from the E2 to a substrate protein. Members of the HECT (>30) and RBR (>10) groups, in contrast, first transfer ubiquitin via a transthioylation reaction onto a catalytic cysteine residue on the E3 before passing the ubiquitin to the substrate protein [114].

Ubiquitin is most commonly conjugated to substrate proteins through the epsilon-amine group of a lysine sidechain. However, ubiquitin may also be conjugated to non-lysine residues. For instance, some E3 ligases target the thiol group of cysteines [115, 116] or alpha-amino groups of N-terminal residues [117] of substrate proteins. In addition, substrate proteins can be ubiquitinated through the formation of hydroxyester bonds on serine and threonine residues [118] and bacterial enzymes of the SidE family can link ubiquitin via Arg42 to serine and tyrosine through the formation of phosphoribosyl linkages [119, 120].

The process of ubiquitination is reversible which is mediated by a family of ubiquitin specific proteases, so called deubiquitinases (DUBs) [121]. DUBs are classified into two main classes – cysteine proteases and metalloproteases. The former comprise the four families of ubiquitin-specific proteases (USPs), ubiquitin C-terminal hydrolases (UCHs), Machado-Josephin domain proteases (MJDs) and ovarian tumor proteases (OTU), the latter the families of Jab1/Mov34/Mpr1 Pad1 N-terminal+ (MPN+) (JAMM) domain proteases [122]. The approximately 100 DUBs are essential to maintain the balance of protein ubiquitination and deubiquitination.

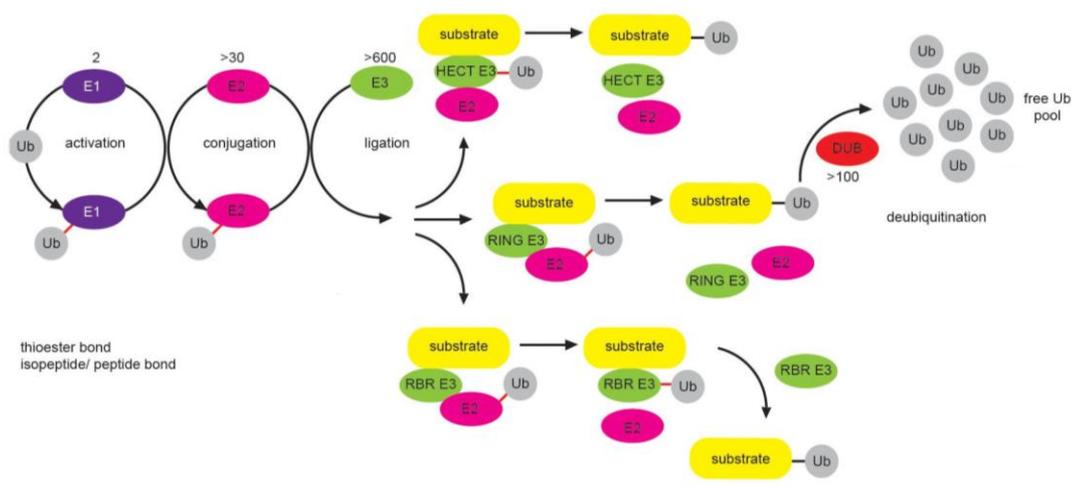


Figure 11 Ubiquitination system

A ubiquitination cascade transfers ubiquitin via ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2) and ubiquitin ligases (E3) onto a substrate protein. The three main groups of E3 ligases, HECT, RING and RBR and their mode of action are displayed. Ubiquitin in substrate proteins can be removed by deubiquitination enzymes (DUBs). Adopted from [113]

1.3.1.2 The ubiquitin code

Protein ubiquitination starts with the covalent addition of one ubiquitin molecule to a substrate protein, but by no means has to end here. On the contrary, monoubiquitinated proteins can be subject to further ubiquitination processes targeting additional lysine residues of the substrate protein, creating a multimonoubiquitinated protein (Figure 12). Moreover, an attached monoubiquitin can itself be ubiquitinated, which leads to the formation of ubiquitin chains. Such chains can have a multitude of structures, depending on the length of the ubiquitin chain and the linkage type(s). Generally, two ubiquitin molecules can be linked via 8 distinct linkage types, a linear link through an N-terminal alpha-amino group or a link through an epsilon-amino group of one of ubiquitin's seven lysine residues (K6, K11, K27, K29, K33, K48, K63). Polyubiquitin chains containing only one linkage type are referred to as homotypic. Otherwise, if different linkage types occur, they are termed heterotypic or mixed polyubiquitin chains. Moreover, a single ubiquitin in a chain can be modified multiple times resulting in the creation of branched chains. Thus, a myriad of different chain topologies can arise, which is further complicated by possible modifications of the ubiquitin molecule by other PTMs such as phosphorylation, acetylation or ubiquitin-like proteins such as Sumo2/3 or Nedd8 [123] (Figure 12).

It is well established that specific chain topologies and modifications confer specific functionalities. Homotypic K48-linked ubiquitin chains, for instance, mark proteins for proteasomal degradation [124] and phosphorylation of Ser65 on ubiquitin plays an essential role in mitophagy [125]. However, the informational content encrypted in the different chain formations – the ubiquitin code – is far from being fully understood. Unraveling of the ubiquitin code will help us to better understand the physiological implications of ubiquitination in health and disease. Therefore, many tools have been developed to study ubiquitination events. Among those, MS has had the most profound impact in deciphering the ubiquitin code [123].

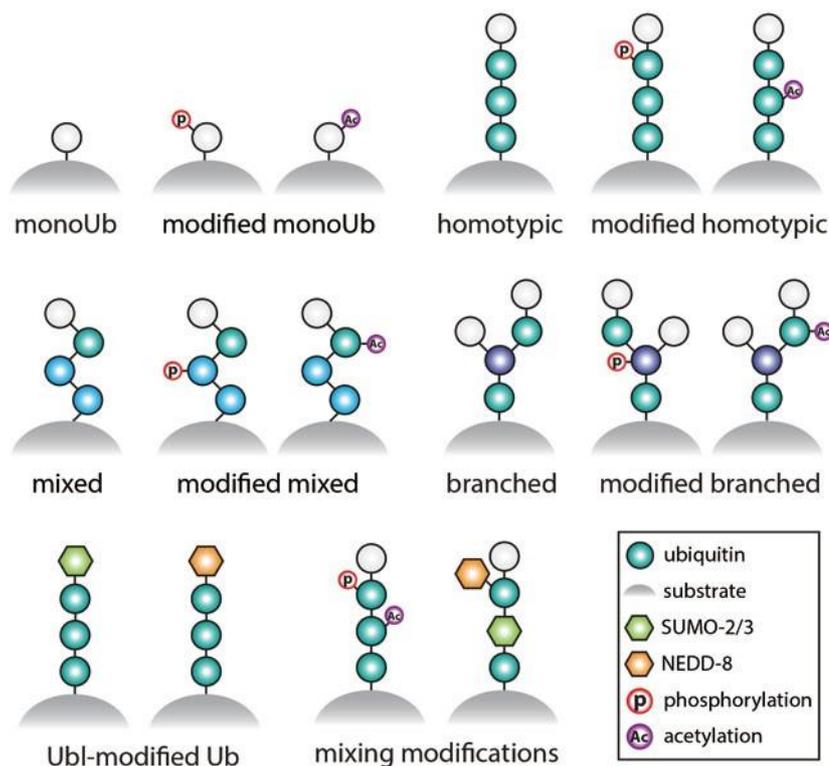


Figure 12 The ubiquitin code

A non-comprehensive representation of possible ubiquitin modifications, including ubiquitin-like and other chemical modifications. Adapted from [123]

1.3.1.3 MS-based ubiquitinome analysis

Since the discovery of ubiquitin and the first ubiquitinated protein, it soon became clear that protein ubiquitination is not an isolated phenomenon. The detection of ubiquitination events, however, is a challenging task and requires sophisticated analysis strategies. In general, ubiquitin modifications are of low abundance and require elegant enrichment strategies. Due to the dynamic and highly variable nature of ubiquitination events, allowing the formation of various chain topologies (see ‘The ubiquitin code’), enrichment strategies have to be carefully chosen to address different aspects of ubiquitin signaling. In this regard, MS in combination with various biochemical methods has become an extremely powerful tool for the analysis of ubiquitination events [126].

A central goal in the study of the ubiquitin system is the system-wide identification of ubiquitinated proteins, the ubiquitinome. Typically, bottom-up proteomic approaches are used for the analysis of the ubiquitinome. Here, the presence of a ubiquitin modification is indirectly observed through the detection of a characteristic diglycine (diGly) remnant on peptides. This diGly remnant, in fact, was already reported in 1977 upon tryptic

digestion of the first reported ubiquitinated protein [104]. However, the low abundance of ubiquitinated proteins make an upfront enrichment imperative. In 2003, the first MS-based large-scale study on protein ubiquitination reported 110 ubiquitination sites upon enrichment of 6xHis tagged ubiquitin [127]. In the following years, this approach and variations of it using different ubiquitin tags led to the identification of hundreds of ubiquitination sites [128-133]. Other studies forwent tagged ubiquitin strategies and instead used ubiquitin antibodies or other ubiquitin-binding entities to enrich directly for ubiquitinated proteins [134, 135].

The development of an antibody in 2010 that was raised against the diGly remnant on modified lysines marked a revolution in the MS-based analysis of the ubiquitinome [136]. Although this first study only resulted in the identification of a few hundred ubiquitination sites, it laid the foundation for further improvements in following years. The generation of more robust diGly motif-specific antibodies and sample preparation procedures resulted in the identification of tens of thousands of ubiquitination sites [137-141]. Although the employment of the diGly antibody has dramatically advanced the study of the ubiquitinome, this method has several limitations. For instance, different diGly antibodies display different sequence preferences for enrichment and they cannot enrich for N-terminal ubiquitination [139]. In addition, modifications of the ubiquitin-like (Ubl) proteins NEDD8 and ISG15 produce the same diGly remnant upon tryptic digestion as ubiquitination. Although the contribution of diGly peptides originating from these Ubl proteins to the total pool of diGly peptides is below 6%, an unambiguous assignment of the detected diGly site to the underlying modification is not possible [137, 142]. Akimov et al. recently generated an antibody that recognizes a ubiquitin distinctive 13 amino acid long remnant after LysC digestion, which is as well able to detect N-terminal ubiquitination [143].

An inherent problem to bottom-up proteomic strategies for ubiquitinome analysis is the loss of information concerning the ubiquitin chain topology upon enzymatic digestion. However, several biochemical methods can be combined with mass spectrometry to gain insight into the ubiquitin chain architecture. Early attempts enrich for ubiquitin employed recombinant proteins containing ubiquitin-binding domains (UBAs) and ubiquitin-interacting motives (UIMs) [128, 130]. Further developments and the combination of multiple UBAs and UIMs to tandem-repeated ubiquitin-binding entities (TUBEs) and tandem ubiquitin-interacting motifs (tUIMs), respectively, provide useful tools for the enrichment of polyubiquitin chains [144, 145]. Depending on the composition of tUIMs or TUBEs, specific chain types such as linear [146], K29-

linked [147], K48-linked [148] or K63-linked [149] polyubiquitin chains can be targeted. Other approaches use linkage-type specific antibodies or affimers for the enrichment of specific ubiquitin chain types [150, 151]. The aforementioned strategies have their strength in the classification of homotypic chain types, although only few branched chain types can be identified with these approaches [150, 151]. A more generic approach to identify branched ubiquitin chains is Ubiquitin Chain Enrichment Middle-down Mass Spectrometry (UbiChEM-MS) [152]. This technique uses ubiquitin binding domains for the enrichment of ubiquitinated proteins and combines minimal tryptic digestion and middle down mass spectrometry to identify multiple branch points on ubiquitin molecules. Multiple branch points on ubiquitin proteins can also be detected by a recently described method, termed UB-clipping [153]. This method uses the engineered viral protease Lb^{pro} to disassemble ubiquitin chains, while leaving the characteristic DiGly remnant on modified ubiquitin, which can then be identified via intact mass spectrometry.

System-wide ubiquitinome studies are frequently coupled with labeling approaches for quantification to reduce the impact of missing values across samples in DDA. In this regard, SILAC has often been favored, because the special enrichment procedure for diGly peptide enrichment complicates TMT labeling. Since the TMT reagent also labels the N-terminal amine of the DiGly remnant, the antibody-binding site for the diGly peptide enrichment will be obscured. Thus, workflows incorporating TMT labeling for quantification perform the labeling step after peptide enrichment, accepting more enrichment variation. To side-step the above issues, we developed a novel analysis strategy for system-wide ubiquitinome analysis using a DIA strategy (see 'Publications'). Amongst other benefits, DIA does not require labeling for consistent identifications across multiple samples, while providing excellent quantitative accuracy. Finally, there are also methods for the absolute quantification of the ubiquitin pool such as Ub-PSAQ [154].

1.3.2 Phosphorylation

The first characterization of reversible protein phosphorylation by Fischer and Krebs in 1955 [155] initiated a new field of research and was rewarded with the Noble Prize. Today, protein phosphorylation is one of the most widely studied PTMs. The process of phosphorylation, the addition of a phosphate group to a protein, and dephosphorylation, the removal of a phosphate group, are mediated by kinases and phosphatases, respectively. Over 500 kinases [156] and more than 180 phosphatases [157]. The vast

majority of reported phosphorylations is covalently linked to serine, threonine and tyrosine residues via a phosphate ester bond, but also non-canonical phosphorylation creating a phosphoramidate (histidine, lysine, arginine), phosphorothioate (cysteine) and phosphonate (aspartic acid, glutamic acid) have been reported [158].

Protein phosphorylation is clearly one of the most pervasive PTMs and involved in a plethora of cellular processes including protein binding, turnover, activity, localization, conformation and crosstalk [159] (Figure 13). Dysregulation of intricate phosphorylation processes is involved in a wide range of diseases such as various forms of cancer [160]. Tremendous research efforts have led to the development of various therapeutic agents successfully targeting components of the phosphorylation process, mostly kinases [161]. It is therefore essential to increase our understanding of the role of phosphorylation events in health and disease and MS-based phosphoproteomics is a powerful tool to study proteome wide phosphorylation events.

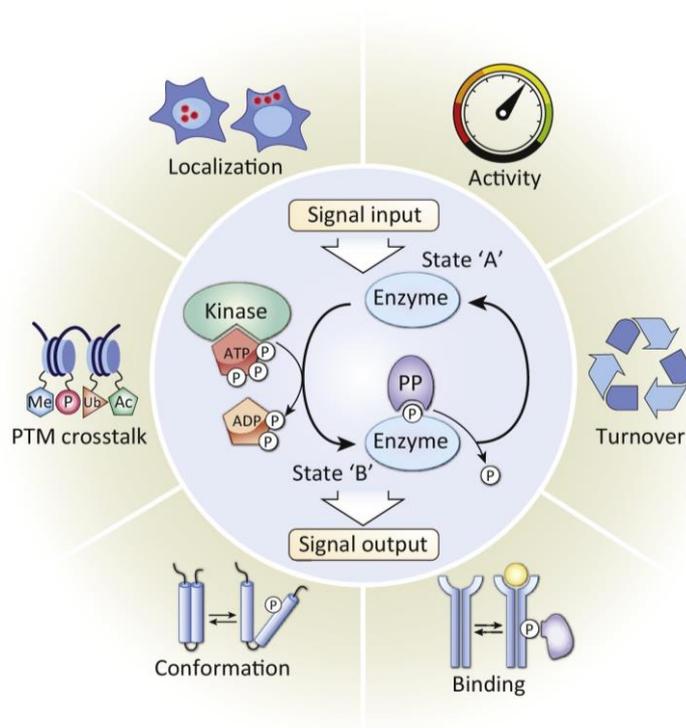


Figure 13 Reversible protein phosphorylation mediates various biological functions

Conceptual representation of reversible protein phosphorylation as a switch in signaling. Adapted from [159]

1.3.2.1 MS-based phosphoproteome analysis

Considering the diverse set of kinases, phosphatases and amino acids that can be phosphorylated along with the pervasive role taken by phosphorylation in the regulation

of cellular processes, it is not surprising that the phosphoproteome is highly dynamic and complex in nature. Depending on the number and distribution of phosphorylation sites, instances of a protein species can exist in various proteoforms at the same time. Additionally, the occupancy of individual phosphorylation sites may vary for an individual protein. Consequently, the large-scale analysis of the phosphoproteome is by no means trivial and powerful analysis strategies need to be devised to shed light on the intricate phosphorylation system.

State-of-the-art, system-wide phosphoproteome analysis typically combines bottom-up MS strategies with upfront phosphopeptide enrichment. Many enrichment strategies using different physicochemical peptide characteristics for the enrichment of phosphopeptides have been developed. Those employed most frequently in phosphopeptide enrichment approaches can be classified into (i) ion-exchange (IEX) and mixed-mode chromatography (MMC), (ii) antibody and protein domain-based enrichment and (iii) affinity-based chromatography [162]. Standard enrichment procedures will primarily lead to the identification of pSer, pThr and pTyr, due to their more stable linkage type compared to non-canonical phosphorylation.

Based on the negatively charged phosphate group, phosphorylated peptides display on average a lower charge state than unmodified peptides. This charge state difference can be exploited in IEX chromatography, for example strong anion exchange (SAX) or strong cation exchange (SCX) chromatography to enrich phosphorylated peptides [163, 164]. Mixed mode chromatography such as Electrostatic repulsion hydrophilic interaction chromatography (ERLIC) employ electrostatic effects superimposed on hydrophilic interaction to enrich phosphopeptides [165]. However, these methods are nowadays rather used for phosphopeptide fractionation in combination with other phosphopeptide enrichment strategies, such as affinity-based techniques. Popular affinity-based chromatography techniques include Immobilized Metal Ions Affinity Chromatography (IMAC) and ion Metal Oxide Affinity Chromatography (MOAC) that can either be used in combination with sample fractionation or as stand-alone methods. IMAC exploits the ability of metal ions, such as Fe^{3+} or Ga^{3+} , to interact with phosphate groups and separate them from unmodified peptides [166, 167]. In the most common setup, metal ions are immobilized on stationary material of a column and retain phosphopeptides in the mobile phase through electronic attraction and metal chelation. In contrast, metal oxides in MOAC form more stable bidentate bonds with phosphate, allowing for more acidic buffer conditions during enrichment [168]. A column format for phosphopeptide enrichment is not a prerequisite; in fact, high-performance methods

often use functionalized beads in a batch format for phosphopeptide enrichment. For instance, the EasyPhos method, developed in our group, employs TiO₂-beads for high-throughput enrichment of low input material [169]. Since the obtained proportion of pTyr (<1%) compared to pSer (90%) and pThr (10%) of these methods is generally very low, specialized antibodies for the specific enrichment of pTyr peptides are often deployed [170].

The generation of high-quality MS data, allowing efficient peptide identification as well as phosphosite localization, requires careful selection of the optimal fragmentation condition in the MS setup. During CID, the labile phospho ester bond tends to break first, leading to a loss of the modification. In contrast, HCD, ETD and variants of these provide more complete sets of fragment ions enabling a better localization of phosphorylation sites [57, 171].

Similar to diGly peptide quantification, several challenges need to be addressed for the confident quantification of phosphopeptides (see “MS-based ubiquitinome analysis”). As for other PTMs the quantification across multiple samples often relies on the identification of single peptides. Therefore, metabolic or chemical labeling strategies, such as SILAC or TMT greatly help to reduce missing values in DDA experiments. Recently, the use of DIA for phosphoproteome analysis was demonstrated to achieve excellent phosphopeptide quantification at an astonishing proteomic depth [172].

1.3.2.2 Phosphorylation in mitochondria

The occurrence of mitochondria is closely linked to a major event eukaryotic life. There are different versions of the endosymbiotic theory that try to explain the emergence of the small double membrane bound organelle [173]. Beyond any doubt, mitochondria are integral components of many pivotal cellular processes. Although they are best known for their function as “powerhouse” of the cell in bioenergetics pathways, they also have important functions in biosynthetic pathways, apoptosis and signaling [174]. Dysregulation of mitochondrial function can lead to a wide variety of mitochondrial diseases. Strikingly, many of these display tissue specificity, suggesting that the function or importance of mitochondria differ in a cell type or tissue specific context [175]. It is known that the abundance of mitochondria differs between cell types and early MS-based proteomics experiments showed that the mitochondrial make-up varies depending on the cellular context [176-178]. It is unlikely that these variances solely account for functional plasticity of mitochondria and there is mounting evidence that mitochondrial functions are frequently regulated by PTMs [179]. Mass spectrometry

potentially offers a powerful platform for the in-depth analysis of the mitochondrial proteome as well as its associated PTMs.

The MS-based study of mitochondria often requires special mitochondria isolation procedures to minimize the masking of mitochondrial proteome changes by the bulk proteome. Common isolation strategies deploy differential centrifugation (DC), DC in conjunction with ultracentrifugation on for e.g. Percoll gradients or magnetic bead-assisted methods (MACS) [180]. Novel approaches use tagged outer mitochondrial membrane proteins (MitoTags) for the specific enrichment of the organelle [181]. These methods strongly enrich mitochondria; however, they cannot completely exclude the enrichment of mitochondria associated non-mitochondrial proteins. Efforts in defining the mitochondrial proteome, using various biochemical and computational approaches, led to databases like MitoCarta2.0 [182] and IMPI (<http://impi.mrc-mbu.cam.ac.uk/>), both integrated in Mitominer4.0 [183]. Such efforts resulted in the annotation of more than 1000 proteins as mitochondrial of which only 13 are encoded by the circular mitochondrial DNA [184].

Similar efforts have been made to define the landscape of PTMs on mitochondrial proteins. Especially the analysis of the mitochondrial phosphoproteome is of great interest. Various studies have investigated the mitochondrial phosphoproteome and phosphosignaling in mitochondria generating mounting evidence that protein phosphorylation conveys important functionalities in mitochondrial biology [179]. For instance, the counterbalancing processes of mitochondrial fusion and fission, which are involved in organelle distribution, size balancing and maintenance of a healthy mitochondrial network, are regulated through phosphorylation events [185, 186]. Drp1, an essential protein in the initiation of mitochondria fission, translocates to the outer mitochondrial membrane (OMM) upon phosphorylation of serine 622 [187-189]. There, Drp1 is bound by Mff, which itself needs to be phosphorylated on serine 129 and 146 [190-192]. However, the investigation of such mitochondrial phosphorylation processes is usually restricted to individual cell types or tissues, making the evaluation of tissue- or cell type-specific regulations on phosphoproteome level difficult. Thus, we here addressed this issue by generating a resource for the community that allows the analysis of the mitochondrial proteome and phosphoproteome across 7 different tissues (submitted manuscript).

2 Aims of this thesis

The use of mass spectrometry for the identification of proteins has arguably created the field of proteomics. Continuous improvements in sample preparation procedures, instrumentation and analysis software nowadays allow the capture of whole proteomes within hours. These advances also unveiled the staggering complexity induced by PTMs on the proteome. However, the detection and analysis of the vast set of proteoforms remains challenging and of utmost importance for deciphering the profound implications of PTMs on the regulation of these biological processes. Thus, the overarching aim of this thesis is the advancement of MS-based proteomics for the analysis of PTMs with a special focus on protein ubiquitination and phosphorylation.

Since the introduction of the diGly remnant-specific antibodies, immunoprecipitation of diGly peptides for large-scale ubiquitinome studies has become the gold standard. Despite the great success of DIA methods for the analysis of proteomes and phosphoproteomes, only DDA strategies had been used for ubiquitinome studies. We reasoned that DIA should considerably improve currently employed DDA schemes and successfully set out to devise a DIA-based in-depth ubiquitinome analysis workflow (Publication 1). Compared to standard label-free DDA workflows, our DIA workflow for ubiquitinome analysis almost doubled the number of identified modification sites, while providing a more precise and accurate quantification. As the pandemic hit the world, we joined a collaborative effort, led by Prof. Andreas Pichlmair, to understand the host perturbations by SARS-CoV-2 and SARS-CoV by means of mass spectrometry. We swiftly adapted the DIA workflow for ubiquitinome analysis to the, at that time, newly available Orbitrap Exploris 480 mass spectrometer and used a method termed FAIMS to enable library generation with limited sample amounts. We further adopted this technique for phosphoproteome analysis. The resulting, well-recognized Nature publication remains to date the only study providing matching proteome, phosphoproteome and ubiquitinome data for SARS-CoV-2 infected cells (Publication 2). Apart from the unbiased, hypothesis free analysis of the ubiquitinome, mass spectrometry can also be used to answer specific questions regarding the modification state of individual proteins. For instance, to elucidate the functional role of RIPK2 in inflammatory signaling, we used a global ubiquitinome analysis strategy to identify modification sites on RIPK2 (Publication 3). Similarly, we also used an approach combining an immunoprecipitation step of a His-tagged TRAF2 and a diGly peptide enrichment step to identify modification sites on TRAF2 (Publication 4). The identification of specific modification sites on proteins offers profound insight in the

functionality of a protein. Our close collaboration in multiple projects with Prof. Schulman's group impressively shows how information on protein ubiquitination sites can be used to shed light on complex protein structures (Publications 5, 6). Here, *in vitro* ubiquitination assays, along with tailored protein digestion strategies were used to identify ubiquitination sites on proteins of interest. In particular, the identification of ubiquitination sites on Fbp1 helped to unveil a novel E3 assembly architecture for a GID E3 ligase (Publication 6). In yet another collaboration with the Schulman group, we investigated the influence of different lysine side chain length on the formation of ubiquitin chains. For this, we established a targeted MS method using stable isotope labeled peptides to quantify the linkage type of diubiquitin chains (Publication 7).

Another focus of this thesis, apart from protein ubiquitination, is protein phosphorylation, the most widely studied PTM. Although the pervasive implications of protein phosphorylation on cellular signaling are well accepted, its study in the context of mitochondria has only recently gained traction. In light of the diverse set of mitochondrial diseases, it appears apparent that mitochondrial function depends on the cellular environment. Compositional differences on the mitochondrial proteome across different tissues are well known, but no comparable data existed on the mitochondrial phosphoproteome level. To address this, we acquired matching mitochondrial proteome and phosphoproteome data of seven different tissues (Publication 8). This valuable resource shows many differences of the mitochondrial composition across tissues on mitochondrial proteome and phosphoproteome levels and we have made it readily accessible via a custom-made website.

A further aspect of this work is the construction of chromatography columns, which are key components of high-performance PTM analysis and proteomics in general. Here, I collaborated with a former PhD student of our group, Johannes Müller-Reif, to establish a novel multiplexed column packing procedure that allows the construction of high-performance columns in a minute timeframe (Publication 9).

Lastly, I also contributed to the AlphaMap analysis software tool, which was jointly devised by Eugenia Voytik and Dr. Isabel Bludau in our group. This tool greatly facilitated the computational analysis and interpretation of various PTMs on proteins (Publication 10). Dr. Isabel Bludau further extended this for 3D protein annotation based on AlphaFold2 structures. With this tool we characterized various PTMs and their 3D organization on a proteome-wide scale (Publication 11).

3 Publications

3.1 Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology

Fynn M. Hansen¹, Maria C. Tanzer¹, Franziska Brüning^{1,2}, Isabell Bludau¹, Che Stafford³, Brenda A. Schulman⁴, Maria S. Robles², Ozge Karayel¹ & Matthias Mann¹

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. ²Institute of Medical Psychology, Faculty of Medicine, LMU, Munich, Germany. ³Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. ⁴Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany.

Published in *Nature Communications* (2021)

Protein ubiquitination is an immensely complex PTM and involved in a plethora of cellular processes. Dysregulation of the intricate ubiquitination system can have detrimental effects on cellular homeostasis and is involved in severe diseases including cancers and neurodegenerative disorders. To better understand the regulation of the complex and widespread ubiquitination processes on a system-wide scale, it is essential to develop methods that allow an accurate, reproducible and in-depth analysis of the ubiquitinome. The enrichment of peptides carrying a diGly remnant after tryptic digestion with a diGly remnant-specific antibody marked a revolution of MS-based analysis of the ubiquitinome and is today the gold standard (see section 1.3.1). Nevertheless, the large-scale analysis of the ubiquitinome remains challenging. In recent years, DIA has excelled as a powerful alternative to DDA. Previous reports of high sensitivity and robust proteome as well as phosphoproteome studies suggested great potential of DIA for ubiquitinome analysis.

To this end, we devised a DIA-based workflow for ubiquitinome analysis. Compared to DDA we almost doubled diGly peptide identification numbers allowing the identification of approx. 35,000 diGly peptides in a single-shot format, while also improving quantitative precision and reproducibility. We applied the DIA-based ubiquitinome analysis to the investigation of TNF signaling and obtained markedly more significantly changing ubiquitination sites than with DDA. Moreover, we challenged our ubiquitinome analysis pipeline by investigating the ubiquitinome of the circadian rhythm. This identified hundreds of cycling ubiquitination sites many of which are organized in clusters. With this work we provide the community with a powerful workflow for large-scale unbiased ubiquitinome analysis.



ARTICLE

<https://doi.org/10.1038/s41467-020-20509-1>

OPEN

Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology

Fynn M. Hansen¹, Maria C. Tanzer¹, Franziska Brüning^{1,2}, Isabell Bludau¹, Che Stafford³, Brenda A. Schulman⁴, Maria S. Robles^{2✉}, Ozge Karayel^{1✉} & Matthias Mann^{1✉}

Protein ubiquitination is involved in virtually all cellular processes. Enrichment strategies employing antibodies targeting ubiquitin-derived diGly remnants combined with mass spectrometry (MS) have enabled investigations of ubiquitin signaling at a large scale. However, so far the power of data independent acquisition (DIA) with regards to sensitivity in single run analysis and data completeness have not yet been explored. Here, we develop a sensitive workflow combining diGly antibody-based enrichment and optimized Orbitrap-based DIA with comprehensive spectral libraries together containing more than 90,000 diGly peptides. This approach identifies 35,000 diGly peptides in single measurements of proteasome inhibitor-treated cells - double the number and quantitative accuracy of data dependent acquisition. Applied to TNF signaling, the workflow comprehensively captures known sites while adding many novel ones. An in-depth, systems-wide investigation of ubiquitination across the circadian cycle uncovers hundreds of cycling ubiquitination sites and dozens of cycling ubiquitin clusters within individual membrane protein receptors and transporters, highlighting new connections between metabolism and circadian regulation.

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. ²Institute of Medical Psychology, Faculty of Medicine, LMU, Munich, Germany. ³Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. ⁴Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany. ✉email: charo.robles@med.uni-muenchen.de; karayel@biochem.mpg.de; mmann@biochem.mpg.de

Ubiquitination is a reversible and highly versatile post-translational modification (PTM) involved in virtually all cellular processes. A ubiquitin conjugation cascade, involving ubiquitin activating (E1), conjugating (E2), and ligating (E3) enzymes, mediates the covalent attachment of the 76 amino acid long ubiquitin molecule to a ϵ -amine group of a lysine residue on a substrate protein. Its removal is mediated by an enzyme family called deubiquitinating enzymes (DUB). Ubiquitin itself can be ubiquitinated N-terminally or via one of its seven lysine residues, giving rise to a plethora of chain topologies, which encode a diverse and specific set of biological functions^{1,2}. Deregulation of this highly complex process has been linked to numerous diseases including neurodegenerative diseases^{3,4}, autoimmunity^{5,6}, and inflammatory disorders^{7–9}.

Protein ubiquitination is one of the most widely studied PTMs in the field of mass spectrometry (MS)-based proteomics. However, due to low stoichiometry of ubiquitination and varying ubiquitin-chain topologies, comprehensive profiling of endogenous ubiquitination is challenging and requires one or more enrichment steps prior to MS analysis¹⁰. Early reports to catalog ubiquitin conjugated proteins from yeast and human described various enrichment methods including the use of epitope-tagged ubiquitin or ubiquitin-associated domains (UBA)^{11–13}. After trypsinization previously ubiquitinated peptides bear a signature diGly remnant that can be targeted by a specific antibody¹⁴. Enrichment strategies employing such antibodies have enabled identification of thousands of ubiquitination sites by MS^{15–17}. A recently described antibody targets a longer remnant generated by LysC digestion to exclude ubiquitin-like modifications such as NEDD8 or ISG15¹⁸, however, the contribution of diGly sites derived from ubiquitin-like modifications is very low (<6%)¹⁵.

The commercialization of such antibodies has accelerated MS-based ubiquitinome analysis and enabled a variety of quantitative, systems-wide studies^{19–23}. However, large-scale analysis of ubiquitination events to study key signaling components remains challenging since in-depth diGly proteome coverage requires relatively large sample amounts and extensive peptide fractionation. These requirements, which largely stem from the low stoichiometry of the modification, come at the expense of throughput, robustness, and quantitative accuracy.

Thus far, ubiquitinome studies have employed data-dependent acquisition (DDA) methods combined with label-free or isotope-based quantification²⁴. Recently, data-independent acquisition (DIA) has become a compelling alternative to DDA for proteomics analysis enabling greater data completeness across samples^{25–28}. In contrast to intensity-based precursor picking of DDA, DIA fragments all co-eluting peptide ions within pre-defined mass-to-charge (m/z) windows and acquires them simultaneously²⁹. This leads to more precise and accurate quantification with fewer missing values across samples and higher identification rates over a larger dynamic range. DIA usually requires a comprehensive spectral library, from which the peptides are matched into single-run MS analyses. Recently, superior performance of DIA for sensitive and reproducible MS measurements has also been demonstrated for global protein phosphorylation analysis³⁰. Given the central importance of ubiquitination, we here set out to investigate the power of DIA for improving data completeness and sensitivity in a single-run analysis format.

For sensitive and reproducible analysis of the ubiquitin-modified proteome, we here devise a workflow combining diGly antibody-based enrichment with a DIA method tailored to the unique properties of the library peptides and to the linear quadrupole Orbitrap mass analyzer. We acquire extensive spectral libraries that altogether contained more than 90,000 diGly peptides allowing us to reproducibly analyze 35,000 distinct diGly

peptides in a single measurement of proteasome inhibitor-treated cells. The DIA-based diGly workflow markedly improves the number of identifications and quantitative accuracy compared to DDA. To investigate if our method would have advantages in the exploration of biological signaling systems, we first apply it to the well-studied TNF-signaling pathway, where it retrieves known ubiquitination events and uncovers novel ones. We then extend it to the analysis of circadian post-translational dynamics, so far poorly studied globally with regards to ubiquitination. This uncovers a remarkable extent and diversity of cycling ubiquitination events. These include closely spaced clusters with the same circadian phase, which are likely pointing to novel mechanisms. Together, our design and results establish a sensitive and accurate DIA-based workflow suitable for investigations of ubiquitin signaling at a systems-wide scale.

Results

DIA quantification enables in-depth diGly proteome coverage in single-shot experiments.

To obtain a comprehensive, in-depth spectral library for efficient extraction of diGly peptides in single-shot DIA analysis, we treated two human cell lines (HEK293 and U2OS) with a common proteasome inhibitor (10 μ M MG132, 4 h). After extraction and digestion of proteins, we separated peptides by basic reversed-phase (BRP) chromatography into 96 fractions, which were concatenated into 8 fractions (“Methods”, Supplementary Fig. 1a). Here, we isolated fractions containing the highly abundant K48-linked ubiquitin-chain derived diGly peptide (K48-peptide) and processed them separately to reduce excess amounts of K48-peptides in individual pools, which compete for antibody binding sites during enrichment and interfere with the detection of co-eluting peptides (Supplementary Fig. 1b). We found this to be a particular issue for MG132 treatment, as blockage of the proteasome activity further increases K48-peptide abundance in these samples. The resulting nine pooled fractions were enriched for diGly peptides, which were separately analyzed using a DDA method (PTMScan Ubiquitin Remnant Motif (K- ϵ -GG) Kit, CST) (Fig. 1a and Supplementary Fig. 1a-b). This identified more than 67,000 and 53,000 diGly peptides in MG132 treated HEK293 and U2OS cell lines, respectively (Fig. 1b). Furthermore, to fully cover diGly peptides of an unperturbed system, we also generated a third library using the same workflow but with untreated U2OS cells (used later for biological applications). This added a further 6000 distinct diGly peptides (Fig. 1b). In total, we obtained 89,650 diGly sites corresponding to 93,684 unique diGly peptides, 43,338 of which were detected in at least two libraries (Fig. 1c, see also source data at PRIDE: PXD019854). To our knowledge, this represents the deepest diGly proteome to date. According to the Phosphosite-Plus database³¹, 57% of the identified diGly sites were not reported before and 7.3% of them had previously been found to be acetylated or methylated, indicating that different PTMs can act on the same sites. Thus, the growing body of diGly sites can help to identify sites of potential PTM crosstalk, an important level of functional regulation of proteins³².

In possession of these large diGly spectral libraries, we evaluated DIA method settings for best performance in single-shot diGly experiments (Supplementary Data 1). Impeded C-terminal cleavage of modified lysine residues frequently generates longer peptides with higher charge states, resulting in diGly precursors with unique characteristics. Guided by the empirical precursor distributions, we first optimized DIA window widths—the transmission windows that together cover the desired precursor peptide range. This increased the number of identified diGly peptides by 6% (Supplementary Fig. 2a-b). Next, we tested different window numbers and fragment scan resolution settings,

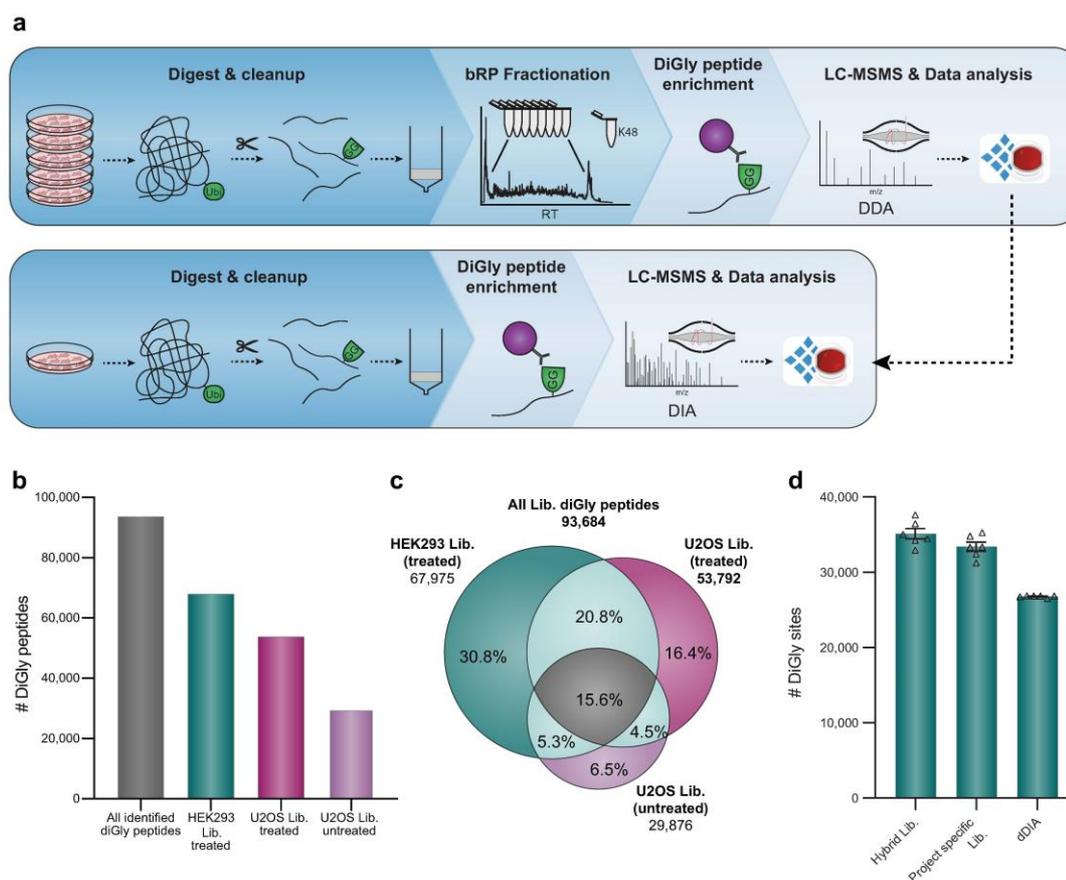
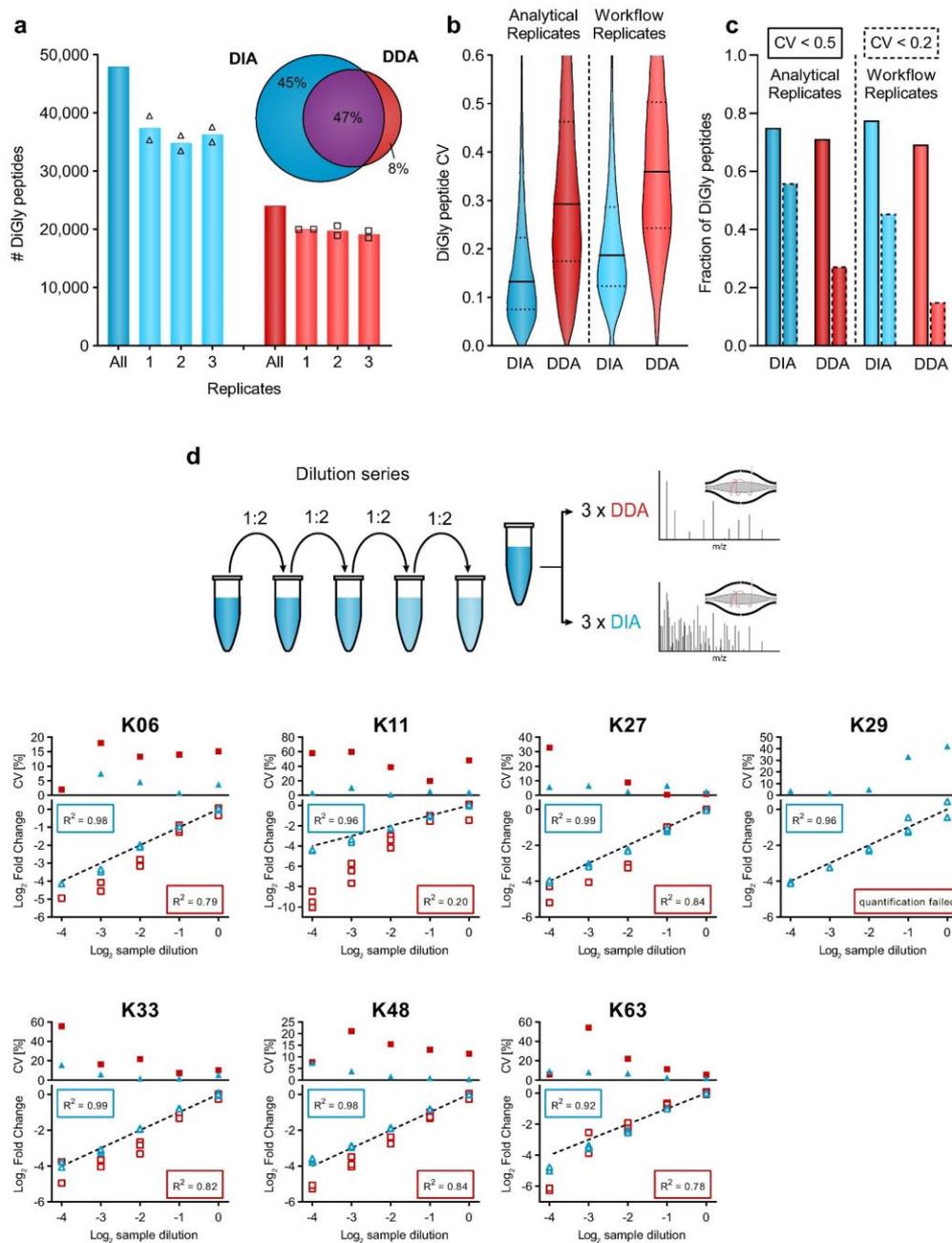


Fig. 1 In-depth diGly proteomics for DIA identification. **a** Experimental workflow for in-depth diGly peptide library construction (upper panel) and our single-run data-independent acquisition (DIA)-based workflow (lower panel). Protein digestion and peptide extraction are followed by basic reversed-phase (bRP) fractionation and diGly peptide enrichment. For library construction, samples were measured by data-dependent acquisition (DDA) and computationally processed (Spectronaut Pulsar). Individual samples are measured by our DIA workflow, including matching against a library for identification (Spectronaut software). **b** Number of identified diGly peptides in three different spectral libraries (MG132 treated HEK293 library—green, MG132 treated U2OS library—violet, U2OS library—light violet, all diGly peptides—gray). **c** Commonly and exclusively identified diGly peptides for different libraries (MG132 treated HEK293 library—green, MG132 treated U2OS library—violet, U2OS library—light violet). **d** Identified diGly sites (mean \pm SEM) of MG132 treated HEK293 cells using different DIA library search strategies ($n = 6$, three workflow replicates measured in analytical duplicates). Source data are provided as a Source data file.

to strike an optimal balance between data quality and a cycle time that sufficiently samples eluting chromatographic peaks. We found that a method with relatively high MS2 resolution of 30,000 and 46 precursor isolation windows performed best (13% improvement compared to the standard full proteome method that we started with) (Supplementary Fig. 2c). Furthermore, we determined the optimal antibody and peptide input combination to maximize peptide yield and depth of coverage in single DIA experiments. To mimic endogenous cellular levels, we used peptide input from cells not treated with MG132. From titration experiments, enrichment from 1 mg of peptide material using 1/8th of an anti-diGly antibody vial (31.25 μ g) turned out to be optimal (“Methods” and Supplementary Fig. 2d, e). With the improved sensitivity by DIA, only 25% of the total enriched material needed to be injected (Supplementary Fig. 2f).

Using our optimized DIA-based workflow, we identified a remarkable $33,409 \pm 605$ distinct diGly sites in single measurements of MG132 treated HEK293 samples. This implies that about half of the sites in the deep, cell line-specific spectral library was matched into the single runs. Interestingly, even without using any library, a search of six single runs identified $26,780 \pm 59$ diGly sites (direct DIA, “Methods”). Finally, employing a hybrid spectral library—generated by merging the DDA library with a direct DIA search—resulted in $35,111 \pm 682$ diGly sites in the same samples (Fig. 1d, Supplementary Data 2). Compared to recent reports in the literature²⁴, these numbers double diGly peptide identifications in a single-run format.

DIA improves diGly proteome quantification accuracy. To evaluate the reproducibility of the entire DIA-based diGly



workflow, we used MG132 treated HEK293 cells and performed three independent diGly peptide enrichments followed by DIA analysis in duplicates. This identified around 36,000 distinct diGly peptides in all replicates, 45% and 77% of which had coefficients of variations (CVs) below 20% and 50%, respectively (Fig. 2a–c, Supplementary Data 3). In contrast, a DDA method identified substantially fewer distinct diGly peptides and a smaller

percentage with good CVs (20,000 diGly peptides; 15% with CVs <20%; Fig. 2a–c). Overall, the six DIA experiments yielded almost 48,000 distinct diGly peptides, while the corresponding DDA experiments resulted in 24,000 diGly peptides. Furthermore, the improved reproducibility is apparent from the diGly site data matrix, which has considerably fewer missing values (Supplementary Fig. 2g).

Fig. 2 Accurate and reproducible diGly proteomics for DIA quantification. **a** Number of identified diGly peptides (mean, $n = 2$) for data-independent acquisition (DIA, blue, HEK293 hybrid library) and data-dependent acquisition (DDA, red) strategies ($n = 6$, three workflow replicates measured in analytical duplicates). Venn diagram depicts the proportion of shared and exclusively identified diGly sites between DIA and DDA approaches. **b** Coefficient of variation (CV) value distribution for DIA and DDA approaches. Solid and dotted lines denote median and 1st or 3rd quantile, respectively. **c** Fractions of CV values below 50% and 20% are shown with solid and dotted lines, respectively. **d** Dilution series of diGly enriched sample. Plots show individual ubiquitin-chain linkage type peptides measured via DIA (blue) or DDA (red) ($n = 3$). Top panels depict CV values of replicate measurements. Bottom panels show individual measurements compared to the expected dilution depicted as dotted line. R^2 values describe the goodness-of-fit of measured values to the expected dilution series (dotted line). Source data are provided as a Source data file.

To further investigate the quantitative precision and accuracy of our method, we turned to ubiquitin-chain linkage derived diGly peptides. These are the most abundant diGly peptides, all ranking in the top 20 by abundance and spanning three orders of magnitude in MS signal (Supplementary Fig. 2h). Diverse chain linkages confer various functions to proteins; hence, accurate quantification is important to decode the cellular roles of different ubiquitin linkage types. We performed a dilution series of a diGly sample and analyzed each dilution sample using both DIA and DDA methods in triplicates. Linear regression of measured vs. expected dilution factors, as a means to directly compare the performance of DIA against DDA, resulted in R^2 values higher than 0.92 for all seven chain peptides assessed, much higher than the corresponding values for DDA (R^2 0.20–0.84; Fig. 2d, Supplementary Data 3). Importantly for quantification purposes, the experimentally observed slope for DIA was much closer to 1 than for DDA.

Together, these analytical results establish that the DIA-based workflow substantially increased the number of diGly peptides identified while markedly improving the precision and accuracy of quantification compared to a DDA-based workflow.

In-depth ubiquitinome analysis of the TNF-signaling pathway.

The pro-inflammatory properties of TNF are heavily regulated by dynamic ubiquitination of its receptor-signaling complex (RSC)^{33,34} and global ubiquitinome changes upon TNF stimulation were described previously in a proteomics study³⁵. Encouraged by the technical capabilities of our DIA-based diGly workflow, we here aimed to test our DIA-based diGly workflow on this well-studied system, to demonstrate benefits of DIA over DDA based on accurate ubiquitination site quantification and, if possible, to extend the current knowledge of the TNF-regulated ubiquitinome (Fig. 3a). Applying both DIA- and DDA-based diGly workflows together quantified over 10,000 diGly sites in TNF-stimulated U2OS cells (Fig. 3b, Supplementary Fig. 3a, Supplementary Data 4). Both methods quantified a comparable number of ubiquitination sites (10,300 in DIA and 9500 in DDA experiment, Fig. 3b). However, the DIA experiment resulted in 248 significantly upregulated ubiquitination sites (5% FDR, median fold change 2.5), of which 37 mapped to 23 proteins known to be involved in TNF/NF κ B signaling (Fig. 3c). In stark contrast, the DDA approach identified only 38 significant upregulated ubiquitination sites (5% FDR and median fold change 4.1), of which 15 mapped to 7 TNF/NF κ B signaling proteins. In line with these numbers, gene ontology (GO) enrichment analysis had lower FDR values and larger group sizes for terms related to the TNF/NF κ B pathway in the DIA experiment compared to DDA (Fig. 3d). Similarly, there were more significantly downregulated ubiquitination sites (1260 in DIA vs. 517 in DDA, 5% FDR) and GOBP terms with lower FDR values in DIA than DDA experiments (Supplementary Fig. 3b and Supplementary Data 4). This large-scale downregulation of ubiquitination events may be due to the activation of deubiquitinating enzymes. In line with FDR threshold lines (Fig. 3c), power analysis exhibits lower fold-

change values (power of 0.8) for DIA compared to DDA, demonstrating increased reproducibility for DIA analysis (Supplementary Fig. 3c).

Several members of the TNF-signaling pathway have been implicated in viral infection and TNF-receptor blockage increases susceptibility to viral infection^{36,37}. The ‘viral processes’ term was significantly enriched in our DIA analysis, in line with literature reporting the involvement of TNF-mediated ubiquitination during viral infection. Underscoring the depth of the DIA analysis, the same term failed to reach significance in the DDA analysis (Fig. 3d, Supplementary Data. 4). In agreement with previous studies, both DIA and DDA analyses revealed increased ubiquitination of prominent members of the TNF-RSC, including TRAF2, RIPK1, and BIRC2^{38,39} (Fig. 3e). Increased protein ubiquitination was validated for TRAF2 and RIPK1 by western blot analysis (Supplementary Fig. 3d). DIA allowed the detection of further ubiquitination events associated with the TNF/NF κ B signaling (Fig. 3c). For instance, the death domain (DD) of RIPK1 mediates interaction with FADD and TRADD⁴⁰ and we found K642 in this domain to be ubiquitinated upon TNF stimulation. Furthermore, DIA but not DDA reveals regulated ubiquitination of all members—HOIP/RNF31, HOIL-1/RBCK1, and Sharpin—of the LUBAC complex, a critical E3 ligase complex in TNF signaling^{41,42} in agreement with a previous study that showed LUBAC auto-ubiquitination during inflammation⁴³ (Fig. 3e). p105/NFKB1, a precursor for p50 and inhibitor of NF κ B signaling⁴⁴ and we observed a striking 16-fold upregulation of K821 in its DD. Proteasome-mediated limited proteolysis of p105 during NF κ B signaling yields the active p50 subunit^{45–48} and the strong regulation of the K821 site suggests its involvement in this process.

DIA-based diGly analysis also uncovered TNF-regulated ubiquitination of numerous proteins known to be involved in other immune pathways. For instance, Peli2, an E3 ligase important for TLR and IL-1 signaling pathways⁴⁹ and its interaction partner TRAF6 were ubiquitinated upon TNF stimulation. We also found that STAT2, which mediates signaling by type I interferons⁵⁰, and USP13, which is involved in the antiviral response by deubiquitinating STING⁵¹, were ubiquitinated at K161 and K3218, respectively. Our results thus suggest further molecular mechanisms for crosstalk or cross-priming function of TNF to other immune pathways during viral and bacterial infections. In summary, our DIA-based ubiquitin workflow provides an in-depth view on the dynamic ubiquitination of core and peripheral members of TNF stimulation. Apart from validating the advantages of DIA over DDA, our results provide novel regulatory ubiquitination sites, conveying a more complete picture of the various aspects of TNF signaling.

Circadian rhythms are globally regulated by ubiquitination.

In mammals, circadian clocks are driven by interlocked transcription-translation feedback-loops. At the cellular and tissue level, they regulate oscillations of gene expression, protein abundance, and post-translational modifications^{52–56}. Ubiquitination plays a pivotal role in the core clock machinery (reviewed

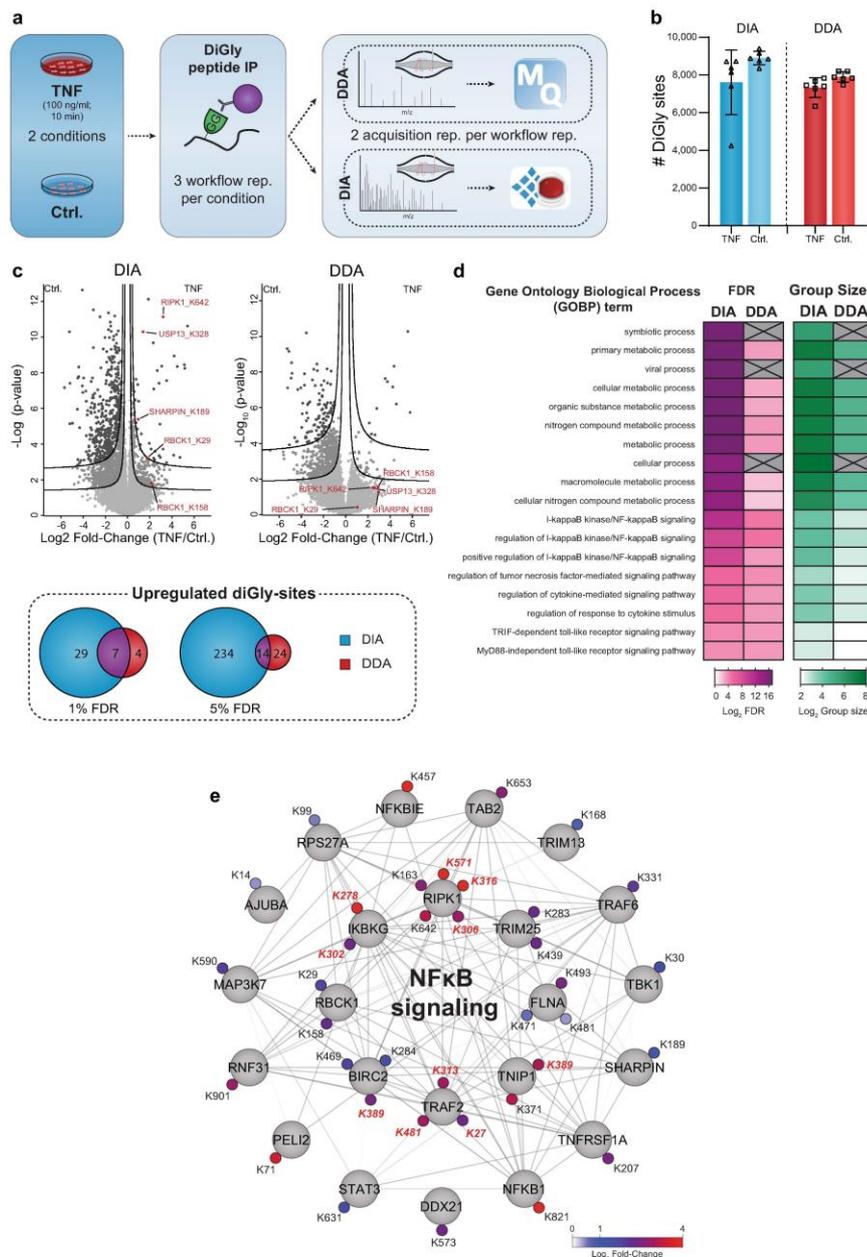


Fig. 3 DIA enables a detailed view of the TNF-regulated ubiquitinome. **a** Workflow for ubiquitinome analysis in tumor necrosis factor (TNF) signaling. **b** Identified diGly sites (\pm SD) for TNF treated (100 ng/ml for 10 min) and control U2OS cells in data-independent acquisition (DIA, blue) and data-dependent acquisition (DDA, red) experiments ($n = 6$, three workflow replicates measured in analytical duplicates). **c** Volcano plot of significantly regulated diGly sites at 5% false discovery rate (FDR) (FDR controlled, two-sided t -test, randomizations = 250, $s_0 = 0.1$) (lower line) and 1% (upper line) for DIA (blue) and DDA (red) and overlaps of significantly upregulated diGly sites for 1 and 5% FDR cutoffs (t -test, $s_0 = 0.1$). **d** Overrepresentation analysis of gene ontology biological process (GOBP) terms filtered for 5% corrected FDR (Fisher's Exact test). **e** Cytoscape network of proteins with significantly upregulated diGly sites in DIA that are associated with NfκB signaling (GO 0043122; GO 0051092; 5% FDR). Upregulated diGly sites also captured by DDA are marked in red (5% FDR). Source data are provided as a Source data file.

in ref.⁵⁷), exemplified by the ubiquitin-dependent spatiotemporal regulation of CRY proteins, the major negative clock regulators⁵⁸. Several studies have provided insights into ubiquitin-dependent events modulating core clock proteins and their effects^{59–61}. Given the unexpected degree of phosphorylation-mediated signaling temporally regulated in vivo⁵², we wondered if ubiquitination shows similar oscillations. With the high accuracy and reproducibility of our DIA-based diGly workflow, we reasoned that it would now be possible to obtain high coverage ubiquitinome quantification across a large time series sample set to answer this question.

To this end, we measured the proteome and ubiquitinome of synchronized U2OS cells—a well-established model to study the cell-autonomous circadian clock—collected every 4 h in biological quadruplicates across 32 h (Fig. 4a). Synchronization was validated by assessing the expression profile of core clock transcripts (*Bmal1* and *Per1*) and further confirmed by PER1 and CLOCK oscillations in our proteome data (Supplementary Fig. 5a–b). After filtering for ubiquitinated peptides present in at least half the samples, we obtained 10,886 ubiquitination sites mapping to 3238 proteins (Fig. 4b, Supplementary Data 5). Measurements were highly reproducible with median Pearson coefficients >0.95 for biological replicates (Supplementary Fig. 5b–c). A total of 7590 proteins were quantified in the proteome, of which at most 143 oscillated (q -value < 0.33). This small percentage of circadian regulation at the proteome level is in line with our previous proteomics results in tissues⁵³ and with transcriptomics results in this cellular system⁶². Next, we normalized the intensities of the diGly peptides encompassing each ubiquitination site to their corresponding protein abundance. The resulting quantitative values represent the occupancy of the ubiquitin sites irrespective of changes in protein abundance (“Methods”, Supplementary Fig. 5d).

Periodicity analysis showed that 8% of the ubiquitination sites on 18% of the proteins oscillated in a circadian manner (856 sites; 590 proteins, “Methods”, q -value < 0.1, Fig. 4c, Supplementary Fig. 5e). A large proportion of rhythmic sites peaked with phases clustered around 16–20 h after synchronization (Fig. 4c and Supplementary Fig. 5e). Remarkably, 59% of these were annotated to be membrane proteins, many more than expected by chance ($p < 10^{-172}$; Supplementary Data 5). Overrepresentation analysis revealed that these proteins are predominantly involved in transport of small molecules, such as ions, amines, and organic acids (Fig. 4d). These findings point to a potential metabolic function of circadian membrane protein ubiquitination.

A full quarter of rhythmic ubiquitinated proteins harbored more than one oscillating site (150 sites; Fig. 4e). To investigate the spatial arrangement of them, we developed a bioinformatic proximity analysis tool (available as part of our website for browsing and analyzing the cellular ubiquitinome <http://cyclingubi.biochem.mpg.de>). In 17% of these proteins, rhythmic ubiquitination sites were closer together than expected by chance ($p < 0.05$) and 73% were annotated as membrane proteins. Interestingly, we found several examples where these adjacent sites were mostly located in regions with potential regulatory function, such as N- and C-termini, cytosolic loops, and interaction domains (Table 1). For instance, K4, K30, and K37 of the sodium independent cystine-glutamate transporter (SLC7A11, 501 aa) are rhythmically ubiquitinated with similar phases (13.8; 13.3; 13.1 h, respectively, Fig. 4f). Likewise, the potassium chloride symporter NKCC1 (SLC12A2) has a cluster of eight rhythmically ubiquitinated sites in its C-terminal domain with similar phases (K948, K958, K966, K971, K976, K983, K991, K992; Supplementary Fig. 5f). This widely expressed solute carrier plays a key role in the regulation of ionic balance and cell volume⁶³. We also discovered novel oscillating ubiquitin

modifications in the MAGE domain of MAGED1, a protein that directly interacts with the core clock protein ROR α , to regulate *Bmal1*, *Rev-erba*, and *E4bp4* gene expression (Fig. 4g). Interestingly, despite these rhythmic outputs neither the *Maged1* transcript, protein expression nor its binding to ROR α oscillate⁶⁴. Our results now suggest that MAGED1 activity could instead be rhythmically controlled in a post-translational manner through the multiple ubiquitinations in its MAGE domain.

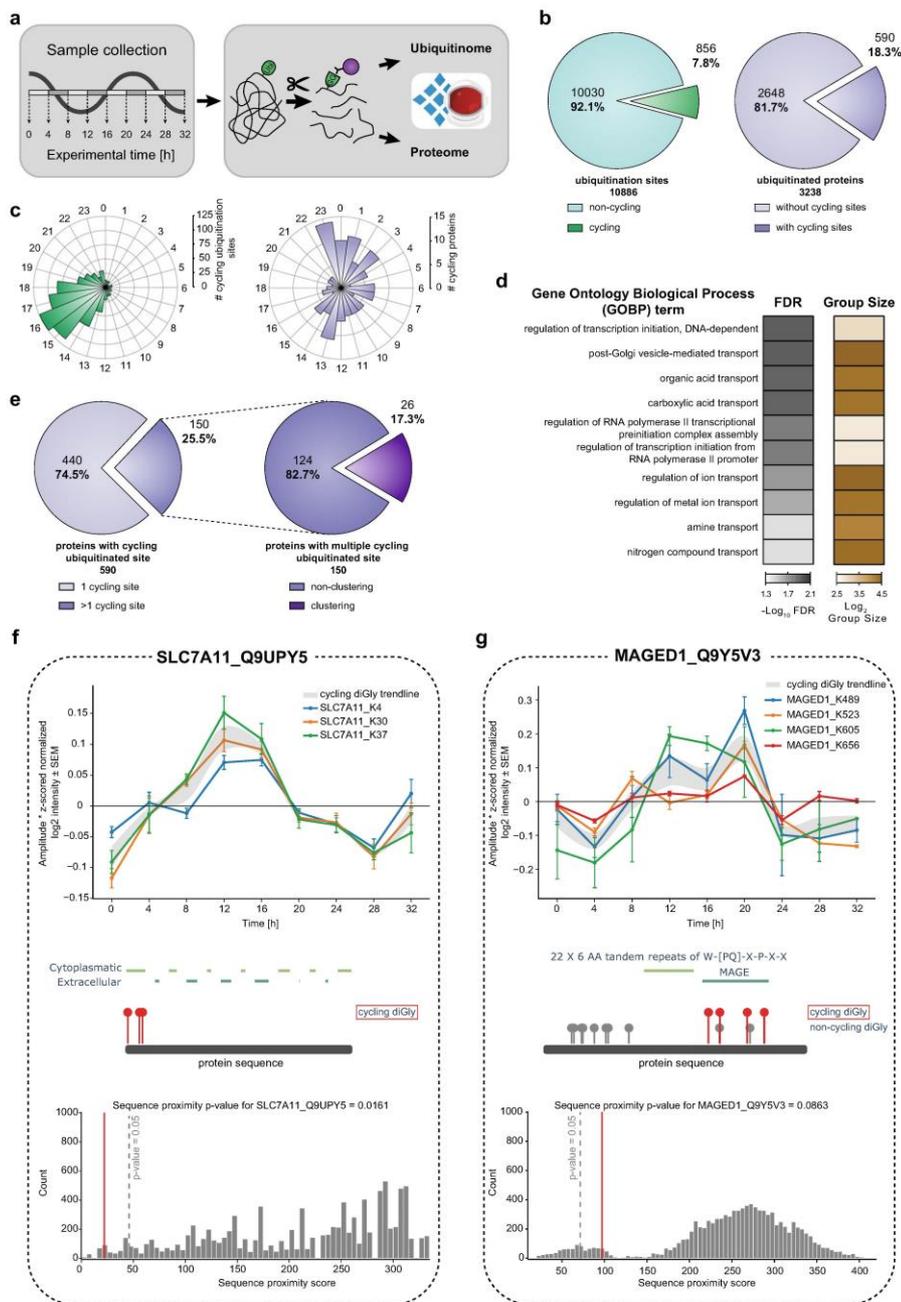
Together, this in-depth view of the circadian ubiquitinome, made possible by our DIA-based diGly workflow, reveals this PTM as a major regulatory mechanism driving rhythmic processes, which include essential cellular processes such as ion transport and osmotic balance.

Discussion

We here developed a sensitive and robust DIA-based workflow, capable of identifying 35,000 diGly peptides in single-run measurements. Both the depth of coverage and the quantitative accuracy are doubled compared to otherwise identical DDA experiments. Importantly the workflow requires no extra labeling step or offline fractionation, making it streamlined and easy to implement. Furthermore, it could be used for quantification of other PTMs relying on antibody-based enrichment such as lysine acetylation and tyrosine phosphorylation. A current limitation of the DIA method is that, like for any DIA-based analysis, including phosphoproteome analysis^{25,30}, the best coverage and quantification is obtained with custom-made, project-specific spectral libraries. Construction of such spectral libraries requires some effort, specialized equipment for fractionation and may not always be possible for samples with low amounts such as primary cells. Alternatively, gas phase⁶⁵ or ion mobility fractionation appear to be promising strategies to simplify the workflows for project-specific spectral library construction. Furthermore, library-free approaches may also greatly simplify DIA workflows in the future. Ongoing efforts to produce prediction tools for peptide MS/MS spectra and retention times will also greatly benefit PTM analysis^{66–69}.

While TMT-based workflows have the advantage of multiplexing compared to DIA workflows, they require peptide fractionation after labeling for in-depth analysis, limiting throughput. In contrast, the latest advances in nanoflow liquid chromatography now increasingly allow rapid, robust, and deep DIA-based proteome and phosphoproteome profiling, which is likely applicable to DIA-based ubiquitinome analysis as well. Furthermore, the LC-MS/MS analysis of our workflow requires only a few hundred μ g and it already enables the analysis of systems such as human primary cell culture models where protein material is limited. However, further sensitivity advances are limited by the initial antibody-based enrichment, which currently requires 0.5–1 mg of sample. If this step could be scaled down and the subsequent peptide purification eliminated altogether, sample amount requirements could become much smaller yet. A workflow without a peptide-clean-up step would also aid to further improve throughput and reproducibility, making the entire workflow more streamlined.

By converting from a DDA to a DIA workflow we demonstrate a dramatic increase in the number of ubiquitination sites that can consistently and significantly be quantified. Given the inherent sensitivity of our single-run approach allowing system-wide investigations of ubiquitination dynamics of biological processes, we applied it to TNF signaling. This provided an in-depth view on the ubiquitination dynamics of TNF signaling, covering core and peripheral signaling members, which a parallel DDA analysis failed to provide. Apart from validating the advantages of DIA over DDA, our results showed that like phosphorylation,



ubiquitination signaling events are rapidly induced after TNF stimulation. Unexpectedly, we still pinpointed novel TNF-regulated sites on proteins that were not previously described in this well-studied pathway. The rich resource provided here could be further explored to investigate the functions of these ubiquitination events in TNF signaling in health and disease.

System-wide circadian proteomics studies have so far been limited to the dynamic regulation of protein and phosphorylation levels—largely for technological reasons. Our in-depth quantitative diGly analysis of cell-autonomous circadian rhythms now extends those studies by providing a cell-intrinsic circadian map of ubiquitination dynamics. Quantifying more than 10,000

Fig. 4 Quantification of the rhythmic ubiquitinome. **a** Experimental workflow for rhythmic ubiquitinome analysis. **b** Proportion of oscillating ubiquitination sites (q -value < 0.1) quantified in >50% of all samples (left panel, green) and proteins with cycling ubiquitin sites (q -value < 0.1) (right panel, violet). **c** Rose plots indicate phase peaks for cycling ubiquitination sites (left panel, green) and proteins (right panel, violet). **d** Overrepresentation analysis of gene ontology biological processes (GOBP) filtered for top 10 significant terms. Significance is determined by 5% false discovery rate (FDR) (Fisher's Exact test). **e** Proportions of proteins with a single and multiple cycling ubiquitination sites (left panel) and those displaying cycling diGly site clusters (right panel). **f, g** Examples of proximity analysis of cycling ubiquitin clusters (<http://cyclingubi.biochem.mpg.de>). Cycling sites (q -value < 0.1, \pm SEM, $n = 4$ biologically independent experiments for each time point) (top) and their location in the protein sequence along with the domain annotation (middle) and proximity score (average distance, p -value < 0.1) (bottom) for **f** SLC7A11 (p -value = 0.0161) and **g** MAGED1 (p -value = 0.0863). Source data are provided as a Source data file.

Table 1 Ubiquitination clusters with potential regulatory circadian functions.						
Gene name	UniprotID	Proximity score (p -value)	Cycling ubiquitination sites	Membrane (GOCC)	Region/domain	
SLC7A5 (LAT1)	Q01650	0.0015	K19, K25, K30	x	N-terminus	
SLC16A1 (MCT1)	P53985	0.0036	K216, K223, K224	x	Cytosolic loop	
LAYN	Q6UX15	0.0064	K272, K273, K297, K311			
EPB41L5	Q9HCM4	0.0073	K508, K510	x		
MYH9	P35579	0.0079	K555, K651, K760, K821	x	Myosin motor, interaction with LIMCH1	
RTN4 (NOGO)	F8W914	0.0094	K327, K332, K336	x	Reticulon, C-terminus	
ABHD17B	Q5VST6	0.0098	K206, K207	x		
SLC3A2 (MDU1)	F5GZ56	0.0153	K114, K116	x		
SLC7A11 (xCT)	Q9UPY5	0.0161	K4, K30, K37	x	N-terminus	
PCNP	Q8WW12	0.0182	K94, K96			
SCRIB (LAP4)	A0A0G2JNZ2	0.0184	K53, K63	x	N-terminus	
PLXNB2	O15031	0.0189	K1743, K1757		C-terminus	
TOM1	O60784-2	0.0209	K443, K446	x	C-terminus	
VLDLR	P98155	0.0220	K828, K839	x	C-terminus	
H2AFY	O75367	0.0240	K292, K295		Macro	
SLC20A1 (GLVR1, PIT1)	Q8WUM9	0.0262	K286, K320, K394, K399, K456	x	Cytosolic loop	
KSR1	A0A0A0MQW1	0.0284	K92, K101	x		
HAS2	Q92819	0.0311	K73, K80			
SMARCD2	B9EGA3	0.0318	K200, K207			
TAX1BP1 (T6BP)	Q86VP1-2	0.0321	K561, K571			
SLC20A2 (GLVR2, PIT2)	Q08357	0.0394	K262, K272	x	Cytosolic loop	
HSP90A1 (HSP90B)	P08238	0.0411	K568, K577	x	Interaction with NR3C1	
SLC12A2 (NKCC1)	G3XAL9	0.0420	K237, K948, K958, K966, K971, K976, K983, K991, K992, K1125	x	SLC12	
PCDH5	Q9Y5E4	0.0426	K767, K784	x	C-terminus	
PPAP2B (LPP3)	O14495	0.0455	K8, K15	x	N-terminus	
SCAMP1	A0A087WXB0	0.0465	K63, K71	x	N-terminus	

Proteins with multiple cycling ubiquitination sites (q -value < 0.1) in close proximity to each other (p -value < 0.05). Membrane protein annotation by Gene Ontology Cellular Compartment (GOCC) term "membrane" and region/domain classifications are derived from UniProt and manual annotation.

unique ubiquitination sites in synchronized U2OS cells, a standard cellular model in chronobiology, revealed that 8% of them—located on 18% of the quantified ubiquitinated proteins—oscillated in abundance. Many of the cycling sites match into the DIA library of untreated, rather than the library of proteasome inhibited cells suggesting they could have regulatory, non-degradative functions.

Our data reveal wide-spread rhythms of ubiquitination in membrane proteins, transporters, and receptors, all regulating major cellular processes such as cell volume, ion balance, and osmotic homeostasis. Intriguingly, often these cycling ubiquitination sites on membrane proteins are not randomly distributed over the protein sequence but rather cluster in certain regions such as the N- and C-terminus. Circadian rhythms in Mg^{2+} and K^+ cellular levels and their transport have been reported in a range of eukaryotic cell types suggesting an evolutionary conservation of this mechanism. Moreover, K^+ transport is a key

mechanism driving electrical excitability oscillations in the mammalian master clock and *Drosophila* neurons^{70,71} and in turn, plasma membrane potential feeds back to the cellular clock^{72,73}. Despite their fundamental cellular role, little is known about the regulatory mechanisms controlling rhythms of ion levels and size in cells^{74,75}. Our system-level data suggest that ubiquitination plays a major role in the rhythmic transport of ions and other compounds in the cell by temporally modulating the activity of membrane transporters. Such a mechanism would, for instance, explain the observation that red blood cells lose their daily electrophysiological rhythm after proteasome treatment⁷⁴.

We speculate that ubiquitin-dependent temporal regulation of transporter function for various substrates (e.g., sodium/phosphate/chloride—SLC20A1/SLC20A2, monocarboxylates—SLC16A1, sodium/potassium—ATP1A1, various amino acids—SLC3A2, SLC7A5, SLC7A11, and organic anions—ABCC3) and other receptors (e.g., TGFBR2 or PLXNB2) may serve as

temporal cellular switches to sense and respond to daily changes in nutrient availability. Interestingly, in our recent phosphoproteomics study of the synaptic compartment, we observed that many of the ubiquitination-related proteins had rhythmic phosphorylation sites⁷⁶. This suggests an interplay between post-translational modifications that together could fine-tune daily cycles of membrane-mediated processes essential for proper cellular and tissue metabolism. Given the central role of transporters in chronopharmacology^{77–79}, ubiquitin-dependent dynamic regulation of specific membrane transporters is an important functional aspect to consider for drug administration and patient health, both key goals of chronotherapy. The data of our rhythmic ubiquitinome analysis is accessible at <http://cyclingubi.biochem.mpg.de>, opening up new avenues for mechanistic investigations.

Methods

Cell culture, treatment, harvest, and lysis. HEK293 (human, DMSZ, ACC 635) and U2OS (human, American Type Culture Collection [ATCC], HTB-96) cells were cultivated in DMEM (Gibco, Invitrogen) supplemented with 10% fetal bovine serum (Gibco, Invitrogen), 100 U/ml penicillin (Gibco, Invitrogen), and 100 µg/ml streptomycin (Gibco, Invitrogen) at 37 °C in a humidified incubator with a 5% CO₂ atmosphere. For cell harvest, cells were washed twice with ice-cold PBS (Gibco, Invitrogen), centrifuged, snap-frozen in liquid nitrogen, and stored at –80 °C until lysis. Frozen cell pellets were lysed by adding lysis buffer (1% SDS in 100 mM Tris/HCl, pH 8.5) directly onto frozen cell pellets, followed by repeated aspiration and boiling for 5 min at 95 °C.

For proteasome inhibition, HEK293 or U2OS cells were treated with 10 µM MG132 (In vivo Gen) at ~80% confluence for 4 h and successively harvested. For circadian cycle experiments, cells were synchronized, when they reached at least 90% confluence, with dexamethasone (1 µM) for 1 h. Following this, U2OS were washed once with PBS and the medium was replaced. The first time point was collected after 24 h of synchronization continuing the collection every 4 h across 32 h for each of the 4 biological replicates. Collected cells were stored and lysed as described. For TNF stimulation of U2OS cells, confluent cultures were either stimulated with 100 ng/ml TNF for 10 min or left unstimulated. Cells were washed 3× with ice-cold PBS, directly lysed with lysis buffer, and boiled for 5 min at 95 °C.

Western blot analysis. To validate TNF-signaling induction U2OS cells were plated in 6-well plates and when confluent stimulated for 5, 10, 15, 30, and 60 min with 100 ng/ml TNF or left untreated. After stimulation, cells were washed in PBS and lysed in 4% SDS in 100 mM Tris/HCl, pH 8. Lysates were boiled, sonicated, and protein concentrations were estimated using BCA. SDS sample loading buffer (450 mM Tris-HCl, pH 8, 60% (v/v) glycerol, 12% (w/v) SDS, 0.02% (w/v) bromophenol blue, 600 mM DTT) was added to lysates before separation on 12% Novex Tris-glycine gels (Thermo Fisher Scientific, XP00120BOX). Separated proteins were transferred onto PVDF membranes (Merck Millipore, IPVH00010). Membranes were blocked in 5% BSA in PBST and antibodies diluted in 2% BSA in PBST.

For validating increased RIPK1 and TRAF2 ubiquitination upon TNF treatment U2OS cells were either left untreated or stimulated with TNF (100 ng/ml) for 10 min, washed in PBS, and lysed in 1 ml DISC buffer (150 mM NaCl, 30 mM Tris pH 7.5, 10% glycerol, 1% Triton X-100) with protease inhibitors (Roche) and 10 mM n-ethyl-maleimide (NEM). Samples were clarified by centrifugation at 16,000 × g for 15 min, equalized to 1 mg of protein, and added directly to 20 µl packed glutathione sepharose beads pre-bound with 100 µg GST-UBA (Ubiquillin-UBA x1)¹³. Beads were incubated on a rotating wheel at 4 °C overnight, washed five times with DISC buffer, and eluted with 2× SDS sample buffer. Proteins were separated on a 10% Novex Tris-glycine gel (Thermo Fisher Scientific, XP00105BOX) and transferred onto a nitrocellulose membrane (GE Healthcare Lifescience, 10600002).

Antibodies (diluted 1:1000) used for immunoblotting were as follows: anti-phospho p65 (CST, 3033 P), anti-p65 (CST, 4764 P), anti-IκBα (CST, 9242), anti-phospho p38 (CST, 9215), anti-p38 (CST, 9212), anti-β-actin (CST, 4970) for TNF-signaling validation and anti-RIP (BD Bioscience, 610458), anti-TRAF2 (CST, 4712), and anti-β-actin (Santa Cruz, sc-47778) for validation of increased RIPK1 and TRAF2 ubiquitination.

RNA isolation and qPCR. RNA was isolated from three biological replicates of each U2OS time point according to manufacture instruction using the RNeasy Plus Mini Kit (QIAGEN, #74134). Isolated RNA was reversely transcribed by using first-strand cDNA synthesis kit (Thermo Fisher Scientific, #K1612). qPCR was performed at the C1000 Thermal Cycler (Bio-Rad) with iQTM SYBR Green Supermix (Bio-Rad, #170-8862) with primers for *Bmal1* (forward: caggaaaa-taggccgaat; reverse: gcgatgaccttatctctg), *Per1* (forward: ggacactctgcagcag; reverse: gggagtgagggtggaagatcaa), and *Gapdh* (forward: ggacacatcctcagac;

reverse: gcccaatcagcaaaatcc). The in-build analysis tool of the CFX Manager Software (Version 3.1, Bio-Rad) was used to determine the normalized expression with the ΔΔC_q method of *Bmal1* and *Per1* compared to *Gapdh* in technical triplicates for all three biological replicates of each time point. The technical triplicates were further averaged and adjusted so that the highest value was set to 1. Following this, the average of all biological replicates and the SEM (standard error of the mean) was calculated for all the time points.

Protein digestion and peptide cleanup. Lysates were sonicated for 1 min (Branson Sonifier) and protein concentrations were estimated by tryptophan assay. After addition of CAA and TCEP to a final concentration of 10 and 40 mM, respectively, samples were incubated for 5 min at 45 °C for protein reduction and alkylation. Thereafter, samples were digested overnight at 37 °C using trypsin (1:100 w/w, Sigma-Aldrich) and LysC (1/100 w/w, Wako).

For proteome analysis, sample aliquots (~15 µg) were desalted in SDB-RPS StageTips (Empore). Briefly, samples were first diluted with 1% TFA in isopropanol to a final volume of 200 µl. Thereafter, samples were loaded onto StageTips and sequentially washed with 200 µl of 1% TFA in isopropanol and 200 µl 0.2% TFA/2% ACN. Peptides were eluted with 60 µl of 1.25% ammonium hydroxide (NH₄OH)/80% ACN and dried using a SpeedVac centrifuge (Eppendorf, Concentrator plus). Dried peptides were resuspended in buffer A* (2% ACN/0.1% TFA) supplemented with iRT peptides (1/30 v/v) (iRT Standard, Biognosys).

For diGly peptide enrichment, samples were four-fold diluted with 1% TFA in isopropanol and loaded onto SDB-RPS cartridges (Strata-X-C, 30 mg/3 ml or Strata-X-C, 200 mg/6 ml, Phenomenex Inc.). Before peptide loading, cartridges were equilibrated with 8 bed volumes (BV) of 30% MeOH/1% TFA and washed with 8 BV of 0.2% TFA. Samples were loaded by gravity flow and sequentially washed twice with 8 BV 1% TFA in isopropanol and once with 8 BV 0.2% TFA/2% ACN. Peptides were eluted twice with 4 BV 1.25% NH₄OH/80% ACN and diluted with ddH₂O to a final ACN concentration of 35% ACN. Thereafter, samples were snap-frozen in liquid nitrogen, lyophilized, and stored at 4 °C until diGly peptide enrichment.

DiGly peptide enrichment. Lyophilized peptides were resuspended in immunoaffinity purification buffer (IAP) (50 mM MOPS, pH 7.2, 10 mM Na₂HPO₄, 50 mM NaCl) and sonicated for 2.5 min (Bioruptor plus, Diagenode). Peptide concentration was estimated by tryptophan assay. DiGly remnant containing peptides were enriched using the PTMScan® Ubiquitin Remnant Motif (K-ε-GG) Kit (Cell Signaling Technology (CST)), which was kindly provided by CST. First, antibodies were cross-linking to beads. Following Udeshi et al.²² 1 vial of antibody coupled beads were first washed 3 times with 1 ml cold cross-linking wash buffer (100 mM sodium tetraborate decahydrate, pH 9.0), followed by 30 min incubation in 1 ml cross-linking buffer (20 mM dimethylpimipidate cross-linking wash buffer) for 30 min at room temperature and gentle agitation. The cross-linking reaction was quenched by two consecutive washes with 1 ml cold quenching buffer (200 mM ethanolamine, pH 8.0) and 2 h incubation in 1 ml quenching buffer at room temperature under gentle agitation. After quenching cross-linked beads were washed three times with 1 ml of cold IAP and used directly for diGly peptide enrichment or stored in 1 ml 0.02% sodiumazide in phosphate-buffered saline, pH 7.4. Unless otherwise stated, 1/8 of a vial of cross-linked antibody beads and 1 mg of peptide material were used for diGly peptide enrichments. For this, peptides were added to cross-linked antibody beads and the volume was adjusted to 1 ml with IAP buffer. After 1 h of incubation at 4 °C and gentle agitation, beads were washed twice with cold IAP and five times with cold ddH₂O. For this, beads were transferred into GF-filter StageTips and for each wash step, the according wash solution was added and passed through by centrifugal force. Thereafter, GF-StageTips were stacked onto SDB-RPS StageTips and peptides were directly eluted into SDB-RPS StageTips. For this, 50 µl 0.15% TFA were added twice onto the beads and passed through by centrifugation for 5 min at 100 × g. Thereafter, 100 µl of 0.2% TFA was added on top of peptide eluates, followed by sample loading onto the stationary material of SDB-RPS StageTips. Peptides were washed, eluted, and dried as described for proteomes samples, with the difference, that 0.2% TFA was used for the first wash step. Dried peptides were resuspended in 9 µl buffer A*, supplemented with iRT peptides (1/30 v/v) for LC/MS-MS analysis.

Basic reversed-phase fractionation. Basic reversed-phase (bRP) fractionation for diGly peptide and proteome spectral libraries were performed on an UFLC System (Shimadzu) and EASY-nLC 1000 (Thermo Fisher Scientific, Germany), respectively.

For diGly peptide separation, lyophilized samples were resuspended in Buffer A (5 mM NH₄HCO₃/2% ACN) and 5 mg peptide material (5 mg/ml) was loaded onto a reversed-phase column (ZORBAX 300Extend-C18, Agilent). Peptides were separated at a flow rate of 2 ml/min and a constant column temperature of 40 °C using a binary buffer system, consisting of buffer A and buffer B (5 mM NH₄HCO₃/90% ACN). An elution gradient at 0% B stepwise increased to 28 in 53 min and to 78 in 6 min was deployed. Eluting peptides were automatically collected into a 96-deepwell plate while well positions were switched in 40 s intervals.

For peptide fractionation on the EASY-nLC 1000 system, ~55 µg peptide material were loaded onto a 30 cm in-house packed, reversed-phase columns (250-

μm inner diameter, ReproSil-Pur C18-AQ 1.9 μm resin [Dr. Maisch GmbH]). Peptides were separated at a flow rate of 2 $\mu\text{l}/\text{min}$ using a binary buffer system of buffer A (PreOmics) and buffer B (PreOmics). An elution gradient at 3% B stepwise increased to 30% in 45 min, 60% in 17 min, and 95% in 5 min was used. Eluting peptides were concatenated into 24 fractions by switching the rotor valve of an automated concatenation system (Spider fractionator, PreOmics)⁸⁰ in 90 s intervals.

Library sample preparation. For individual deep diGly libraries, 2 \times 5 mg of peptide was fractionated by bRP fractionation. For K48-peptide containing fraction identification, 100 μl aliquots of fractions 46 to 54 were dried in a SpeedVac, resuspended in A*, and measured on an LTQ Orbitrap XL mass spectrometer. K48-peptide containing fractions of both plates were pooled in sample pool “K48” (Supplementary Fig. 1a). Remaining fractions of both plates were concatenated into P1–P8 (Supplementary Fig. 1a), snap-frozen, and lyophilized. Lyophilized peptides were resuspended in 1 ml IAP buffer and diGly peptides were enriched as described above. In case of HEK293 library generation, an optional second supernatant IP was conducted. For this, 500 μl of previous diGly peptide enrichment supernatants were pooled as indicated (Supplementary Fig. 1a) and used for sequential diGly peptide enrichment.

For the proteome library, aliquots of U2OS samples for proteome cycling analysis were used. Approximately 3 μg peptide material of individual time points of two biological replicates, after SDB-RPS cleanup, were pooled and fractionated via bRP fractionation as described above. Fractionated samples were dried using a SpeedVac and resuspended in A* supplemented with iRT peptides (1/30 v/v) for LC-MS/MS measurement and spectral library generation.

Nanoflow LC-MS/MS proteome measurements. Peptides were loaded onto a 50 cm, in-house packed, reversed-phase columns (75 μm inner diameter, ReproSil-Pur C18-AQ 1.9 μm resin [Dr. Maisch GmbH]). The column temperature was controlled at 60 $^{\circ}\text{C}$ using a homemade column oven and binary buffer system, consisting of buffer A (0.1% formic acid (FA)) and buffer B (0.1% FA in 80% ACN), was utilized for low pH peptide separation. An EASY-nLC 1200 system (Thermo Fisher Scientific), directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source, was employed for nanoflow liquid chromatography, at a flow rate of 300 nl/min. For individual measurements, 500 ng of peptide material was loaded and eluted with a gradient starting at 5% buffer B and stepwise increased to 30% in 95 min, 60% in 5 min, and 95% in 5 min.

The same general setup was used, for K48-peptide containing fraction identification, while the column and mass spectrometer were changed to a 20 cm column and an LTQ Orbitrap XL, respectively.

For DDA experiments, the Thermo Xcalibur (4.0.27.19) and LTQ Tune plus (2.5.5 SP2) software were used for Q Exactive HF-X and LTQ Orbitrap XL instruments, respectively. The Q Exactive HF-X was operated in Top12 mode with a full scan range of 300–1650 m/z at a resolution of 60,000. The automatic gain control (AGC) was set to 3e6 at a maximum injection time of 20 s. Precursor ion selection width was kept at 1.4 m/z and fragmentation was achieved by higher-energy collisional dissociation (HCD) (NCE 27%). Fragment ion scans were recorded at a resolution of 15,000, an AGC of 1e5 and a maximum fill time of 60 ms. Dynamic exclusion was enabled and set to 20 s. The LTQ Orbitrap XL was operated in Top10 mode with a full scan range of 300–1700 m/z at a resolution of 60,000. Precursor ion selection width was kept at 2.0 m/z and fragmentation was achieved by collision-induced dissociation (CID) (NCE 35%).

For DIA analysis, the MaxQuant Live software suite was utilized for data acquisition⁸¹. The full scan range was set to 300–1650 m/z at a resolution of 120,000. The AGC was set to 3e6 at a maximum injection time of 60 ms. HCD (NCE 27%) was used for precursor fragmentation and fragment ions were analyzed in 33 DIA windows at a resolution of 30,000, while the AGC was kept at 3e6.

Nanoflow LC-MS/MS diGly measurements. DiGly peptide enriched samples were measured on a Q Exactive HF-X using the same instrumental setup as for proteome analysis. For diGly single-run measurements one quarter (2 μl) and for diGly library preparation one-half (4 μl) of enriched samples were loaded for LC-MS/MS analysis, unless stated otherwise. Loaded peptides were eluted using a gradient starting at 3% buffer B and stepwise increased to 7% in 6 min, 20% in 49 min, 36% in 39 min, 45% in 10 min, and 95% in 4 min.

For DDA analysis, the MS was operated in Top12 mode with a full scan range of 300–1350 m/z at a resolution of 60,000. AGC was set to 3e6 at a maximum injection time of 20 s. Precursor ion selection width was kept at 1.4 m/z and fragmentation was achieved by HCD (NCE 28%). Fragment ion scans were recorded at a resolution of 30,000, an AGC of 1e5 and a maximum fill time of 110 ms. Dynamic exclusion was enabled and set to 30 s.

For DIA analysis, the MaxQuant Live software suite was employed for data acquisition⁸¹. The full scan range was set to 300–1350 m/z at a resolution of 120,000. The AGC was set to 3e6 at a maximum injection time of 60 ms. HCD (NCE 28%) was used for precursor fragmentation and resulting fragment ions were analyzed in 46 DIA windows at a resolution of 30,000 (unless otherwise

stated) and an AGC of 3e6. DIA window distribution parameters PdfMu and PdfSigma were set to 6.161865 and 0.348444, respectively, unless stated otherwise (Supplementary Data 6).

Raw data analysis. DDA raw data used for K48-peptide fraction identification and DIA and DDA comparisons were analyzed with MaxQuant (1.6.2.10) using default settings and enabled match between runs (MBR) functionality. Carbamidomethyl (C) was defined as fixed modification and Oxidation (M), Acetyl (Protein N-term), and DiGly (K) were set as variable modifications.

DDA raw data, used for spectral library construction, were processed with Spectronauts build in search engine pulsar (13.12.200217.43655)²⁸. Default settings were used for proteome spectral libraries. For diGly spectral libraries, the “Best N Fragments per peptides” maximum value was adjusted to 25. For hybrid library construction DIA raw files were processed together with DDA library raw files using the same search settings.

DIA raw files were processed using Spectronaut (13.12.200217.43655)²⁸. Proteome analysis was performed with default settings. For diGly analysis, diGly (K) was defined as an additional variable modification and PTM localization was enabled and set to 0. For dilution experiments, “XIC RT extraction window” was set to “static” with a window width of 10 min. Direct DIA searches used the same settings as described above.

Bioinformatics analysis. Data analysis was primarily performed in the Perseus software suite (1.6.7.0). For diGly site analysis, Spectronaut normal report output tables were aggregated to diGly sites using the peptide collapse plug-in tool for Perseus³⁰. DiGly sites were aggregated using the linear model-based approach and filtered for a localization probability >0.5. Data sets of both acquisition strategies, DIA and DDA, were filtered to contain >50% valid values in at least one experimental condition. Missing values were imputed based on a Gaussian normal distribution with a width of 0.3 and a downshift of 1.8. Student *t*-test statistics (FDR cutoff 1% or 5%; $\alpha = 0.1$) for TNF-stimulation experiments were performed in Perseus. Fisher’s Exact GOBP Term enrichment of upregulated diGly sites and cycling diGly sites was performed on the pantherdb website (<http://pantherdb.org/>) and in perseus, respectively, with Benjamin Hochberg FDR correction enabled and set to a 5% cutoff. Network representation of upregulated diGly sites was performed with the STRING app (1.5.1) in Cytoscape (3.7.2). The power analysis was performed in R (3.6.2), using the ‘pw’⁸² and ‘effectsiz’⁸³ packages. The Cohen’s distance was calculated based on a fixed power of 80%, a sample size of 6 per condition and a desired significance threshold of 1%. The test was set to a “two-sample” and “two-sided” *t*-test. A fold-change threshold was subsequently estimated by multiplying Cohen’s distance with the pooled standard deviation separately for each peptide. We only considered sites without missing values for this analysis.

For the cycling analysis of diGly sites, data were first filtered for diGly sites identified in at least 50% across all measurements. Proteins and diGly sites raw intensities were \log_2 transformed and normalized by median subtraction. For diGly site protein normalization the median values of biological quadruplicates were subtracted from normalized diGly sites. Missing values of protein data for subtraction were imputed based on a Gaussian normal distribution with a width of 0.3 and a downshift of 1.8. Cycling analysis of normalized protein and diGly site data was performed as previously described, but in this case with a period time of 24.8 h^{52,53}. A *q*-value cutoff of <0.1 and <0.33 was used to define cycling diGly sites and proteins, respectively.

Website tool. For profile plots individual *z*-scores for each protein abundance normalized diGly site and the median *z*-score and standard error of means (SEM) were subsequently determined for each time point. The resulting median *z*-scores and SEM values were multiplied with the cycling amplitude of each diGly site (Perseus periodicity analysis output). For sequence visualization and protein domain annotation each diGly site location was mapped to the first UniProt ID of its assigned protein group and was visualized based on its respective protein sequence stored in the fasta file that was used for MS/MS data analysis (human fasta, downloaded 2015). The protein sequences for visualization were obtained using the ‘fasta’ functions from pyteomics^{84,85}. Information about protein domains was obtained from UniProt (<https://www.uniprot.org/>, accessed 25.05.2020), including the following categories: “Topological domain”, “Motif”, “Region”, “Repeat”, “Zink finger”, and “Domain [FT]”.

To evaluate whether multiple observed cycling diGly sites are located in a specific region on the protein, we performed a proximity analysis. Three different metrics were evaluated: (1) the average distance (In amino acids) between all observed cycling diGly sites, (2) the minimum distance between any two observed cycling diGly sites, and (3) the maximum distance between any two observed cycling diGly sites. The observed distance metrics were compared to the distances expected from a random distribution of the diGly sites of a protein across all of its lysines. 10,000 random distributions were considered, and an empirical *p*-value was estimated based on the fraction of random samples with a smaller or equally small distance metric as the observed cycling diGly sites. For the main analysis, diGly sites with a *q*-value ≤ 0.1 were considered as cycling diGly sites.

Data preprocessing and visualization for the dashboard was performed using the python programming language. Following libraries were utilized for data processing: numpy (1.18.1), pandas (0.24.2), re, random, and pyteomics^{84,85} (4.2). Several libraries from the HoloViz (0.11.3) family of tools were used for data visualization and creation of the dashboard, including panel and holoviews (1.13.2), but also bokeh (2.0.1), plotly (4.6.0), and matplotlib (3.0.3).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All mass spectrometry data have been deposited on the ProteomeXchange Consortium via the PRIDE database with the dataset identifier PXD019854. A file linking mass spectrometry raw data in the ProteomeXchange folder to the associated experiments in the manuscript is available (Supplementary Data 7). The proximity analysis tool for the investigation of cycling diGly sites is available on <http://cyclingubi.biochem.mpg.de>. Information about protein domains was obtained from UniProt (<https://www.uniprot.org/>, accessed 25.05.2020). Source data are provided with this paper.

Code availability

Custom code for the proximity analysis, implemented on <http://cyclingubi.biochem.mpg.de> has been deposited on GitHub (<https://github.com/MannLabs/CyclingProximityAnalysis>).

Received: 20 July 2020; Accepted: 27 November 2020;

Published online: 11 January 2021

References

- Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).
- Yau, R. & Rape, M. The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **18**, 579–586 (2016).
- Karbowski, M. & Youle, R. J. Regulating mitochondrial outer membrane proteins by ubiquitination and proteasomal degradation. *Curr. Opin. Cell Biol.* **23**, 476–482 (2011).
- Spratt, D. E. et al. A molecular explanation for the recessive nature of parkin-linked Parkinson's disease. *Nat. Commun.* **4**, 1983 (2013).
- Seymour, R. E. et al. Spontaneous mutations in the mouse Sharpin gene result in multiorgan inflammation, immune system dysregulation and dermatitis. *Genes Immun.* **8**, 416–421 (2007).
- Boisson, B. et al. Immunodeficiency, autoinflammation and amylopectinosis in humans with inherited HOIL-1 and LUBAC deficiency. *Nat. Immunol.* **13**, 1178–1186 (2012).
- Babu, J. R., Geetha, T. & Wooten, M. W. Sequestosome 1/p62 shuttles polyubiquitinated tau for proteasomal degradation. *J. Neurochem.* **94**, 192–203 (2005).
- Varfolomeev, E. et al. c-IAP1 and c-IAP2 are critical mediators of tumor necrosis factor alpha (TNFalpha)-induced NF-kappaB activation. *J. Biol. Chem.* **283**, 24295–24299 (2008).
- Zucchelli, S. et al. Tumor necrosis factor receptor-associated factor 6 (TRAF6) associates with huntingtin protein and promotes its atypical ubiquitination to enhance aggregate formation. *J. Biol. Chem.* **286**, 25108–25117 (2011).
- Doll, S. & Burlingame, A. L. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem. Biol.* **10**, 63–71 (2015).
- Peng, J. et al. A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21**, 921–926 (2003).
- Mayor, T. & Deshaies, R. J. Two-step affinity purification of multiubiquitylated proteins from *Saccharomyces cerevisiae*. *Methods Enzymol.* **399**, 385–392 (2005).
- Hjerpe, R. et al. Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep.* **10**, 1250–1258 (2009).
- Xu, G., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunofluorescence profiling. *Nat. Biotechnol.* **28**, 868–873 (2010).
- Kim, W. et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).
- Wagner, S. A. et al. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell Proteom.* **10**, M111 013284 (2011).
- Bustos, D., Bakalarski, C. E., Yang, Y., Peng, J. & Kirkpatrick, D. S. Characterizing ubiquitination sites by peptide-based immunofluorescence enrichment. *Mol. Cell Proteom.* **11**, 1529–1540 (2012).
- Akimov, V. et al. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat. Struct. Mol. Biol.* **25**, 631–640 (2018).
- Wagner, S. A. et al. Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues. *Mol. Cell Proteom.* **11**, 1578–1585 (2012).
- Elia, A. E. et al. Quantitative proteomic atlas of ubiquitination and acetylation in the DNA damage response. *Mol. Cell* **59**, 867–881 (2015).
- Rose, C. M. et al. Highly multiplexed quantitative mass spectrometry analysis of ubiquitylomes. *Cell Syst.* **3**, 395–403 e394 (2016).
- Udeshi, N. D., Mertins, P., Svingkina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* **8**, 1950–1960 (2013).
- van der Wal, L. et al. Improvement of ubiquitylation site detection by Orbitrap mass spectrometry. *J. Proteom.* **172**, 49–56 (2018).
- Udeshi, N. D. et al. Rapid and deep-scale ubiquitylation profiling for biology and translational research. *Nat. Commun.* **11**, 359 (2020).
- Ludwig, C. et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, e8126 (2018).
- Gillet, L. C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Proteom.* **11**, O111 016717 (2012).
- Kelstrup, C. D. et al. Performance evaluation of the Q exactive HF-X for shotgun proteomics. *J. Proteome Res.* **17**, 727–738 (2018).
- Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell Proteom.* **14**, 1400–1410 (2015).
- Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).
- Bekker-Jensen, D. B. et al. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, 787 (2020).
- Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
- Aggarwal, S., Banerjee, S. K., Talukdar, N. C. & Yadav, A. K. Post-translational modification crosstalk and hotspots in sirtuin interactors implicated in cardiovascular diseases. *Front Genet.* **11**, 356 (2020).
- Harhaj, E. W. & Dixit, V. M. Deubiquitinases in the regulation of NF-kappaB signaling. *Cell Res.* **21**, 22–39 (2011).
- Silke, J. The regulation of TNF signalling: what a tangled web we weave. *Curr. Opin. Immunol.* **23**, 620–626 (2011).
- Wagner, S. A., Satpathy, S., Beli, P. & Choudhary, C. SPATA2 links CYLD to the TNF-alpha receptor signaling complex and modulates the receptor signaling outcomes. *EMBO J.* **35**, 1868–1884 (2016).
- Andersen, N. N. & Jess, T. Risk of infections associated with biological treatment in inflammatory bowel disease. *World J. Gastroenterol.* **20**, 16014–16019 (2014).
- Kim, S. Y. & Solomon, D. H. Tumor necrosis factor blockade and the risk of viral infection. *Nat. Rev. Rheumatol.* **6**, 165–174 (2010).
- Dziedzic, S. A. et al. ABIN-1 regulates RIPK1 activation by linking Met1 ubiquitylation with Lys63 deubiquitylation in TNF-RSC. *Nat. Cell Biol.* **20**, 58–68 (2018).
- Weinlich, R. & Green, D. R. The two faces of receptor interacting protein kinase-1. *Mol. Cell* **56**, 469–480 (2014).
- Hsu, H., Huang, J., Shu, H. B., Baichwal, V. & Goeddel, D. V. TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity* **4**, 387–396 (1996).
- Tokunaga, F. & Iwai, K. LUBAC, a novel ubiquitin ligase for linear ubiquitination, is crucial for inflammation and immune responses. *Microbes Infect.* **14**, 563–572 (2012).
- Ikeda, F. et al. SHARPIN forms a linear ubiquitin ligase complex regulating NF-kappaB activity and apoptosis. *Nature* **471**, 637–641 (2011).
- Heger, K. et al. OTULIN limits cell death and inflammation by deubiquitinating LUBAC. *Nature* **559**, 120–124 (2018).
- Savinova, O. V., Hoffmann, A. & Ghosh, G. The Nfkb1 and Nfkb2 proteins p105 and p100 function as the core of high-molecular-weight heterogeneous complexes. *Mol. Cell* **34**, 591–602 (2009).
- Collins, P. E., Mitxitorena, I. & Carmody, R. J. The ubiquitination of NF-kappaB subunits in the control of transcription. *Cells* **5**, 23 (2016).
- Kravtsova-Ivantsiv, Y. et al. KPC1-mediated ubiquitination and proteasomal processing of NF-kappaB1 p105 to p50 restricts tumor growth. *Cell* **161**, 333–347 (2015).
- Lin, L., DeMartino, G. N. & Greene, W. C. Cotranslational biogenesis of NF-kappaB p50 by the 26S proteasome. *Cell* **92**, 819–828 (1998).
- Ciechanover, A. et al. Mechanisms of ubiquitin-mediated, limited processing of the NF-kappaB1 precursor protein p105. *Biochimie* **83**, 341–349 (2001).
- Schauvliege, R., Janssens, S. & Beyaert, R. Pellino proteins: novel players in TLR and IL-1R signalling. *J. Cell Mol. Med.* **11**, 453–461 (2007).

50. Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nat. Rev. Immunol.* **14**, 36–49 (2014).
51. Sun, H. et al. USP13 negatively regulates antiviral responses by deubiquitinating STING. *Nat. Commun.* **8**, 15534 (2017).
52. Robles, M. S., Humphrey, S. J. & Mann, M. Phosphorylation is a central mechanism for circadian control of metabolism and physiology. *Cell Metab.* **25**, 118–127 (2017).
53. Robles, M. S., Cox, J. & Mann, M. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet.* **10**, e1004047 (2014).
54. Zhang, R., Lahens, N. F., Ballance, H. L., Hughes, M. E. & Hogenesch, J. B. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc. Natl Acad. Sci. USA* **111**, 16219–16224 (2014).
55. Szabo, A. et al. Ubiquitylation dynamics of the clock cell proteome and TIMELESS during a circadian cycle. *Cell Rep.* **23**, 2273–2282 (2018).
56. Wang, Y. et al. A proteomics landscape of circadian clock in mouse liver. *Nat. Commun.* **9**, 1553 (2018).
57. Stojkovic, K., Wing, S. S. & Cermakian, N. A central role for ubiquitination within a circadian clock protein modification code. *Front Mol. Neurosci.* **7**, 69 (2014).
58. Yoo, S. H. et al. Competing E3 ubiquitin ligases govern circadian periodicity by degradation of CRY in nucleus and cytoplasm. *Cell* **152**, 1091–1105 (2013).
59. Ullah, K. et al. The E3 ubiquitin ligase STUB1 attenuates cell senescence by promoting the ubiquitination and degradation of the core circadian regulator BMAL1. *J. Biol. Chem.* **295**, 4696–4708 (2020).
60. Correia, S. P. et al. The circadian E3 ligase complex SCF(FBXL3+CRY) targets TLK2. *Sci. Rep.* **9**, 198 (2019).
61. D'Alessandro, M. et al. Stability of wake-sleep cycles requires robust degradation of the PERIOD protein. *Curr. Biol.* **27**, 3454–3467 (2017).
62. Hughes, M. E. et al. Harmonics of circadian gene transcription in mammals. *PLoS Genet.* **5**, e1000442 (2009).
63. Demian, W. L. et al. The ion transporter NKCC1 links cell volume to cell mass regulation by suppressing mTORC1. *Cell Rep.* **27**, 1886–1896 (2019).
64. Wang, X. et al. Interaction of MAGED1 with nuclear receptors affects circadian clock function. *EMBO J.* **29**, 1389–1400 (2010).
65. Pino, L. K., Just, S. C., MacCoss, M. J. & Searle, B. C. Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries. *Mol. Cell Proteom.* **19**, 1088–1103 (2020).
66. Yang, Y. et al. In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. *Nat. Commun.* **11**, 146 (2020).
67. Searle, B. C. et al. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **11**, 1548 (2020).
68. Yang, Y., Horvatovich, P. & Qiao, L. Fragment mass spectrum prediction facilitates site localization of phosphorylation. *J. Proteome Res.* <https://doi.org/10.1021/acs.jproteome.0c00580> (2020).
69. Muller, J. B. et al. The proteome landscape of the kingdoms of life. *Nature* **582**, 592–596 (2020).
70. Kuhlman, S. J. & McMahon, D. G. Rhythmic regulation of membrane potential and potassium current persists in SCN neurons in the absence of environmental input. *Eur. J. Neurosci.* **20**, 1113–1117 (2004).
71. Flourakis, M. et al. A conserved bicycle model for circadian clock control of membrane excitability. *Cell* **162**, 836–848 (2015).
72. Nitabach, M. N., Sheeba, V., Vera, D. A., Blau, J. & Holmes, T. C. Membrane electrical excitability is necessary for the free-running larval *Drosophila* circadian clock. *J. Neurobiol.* **62**, 1–13 (2005).
73. Jones, J. R., Tackenberg, M. C. & McMahon, D. G. Manipulating circadian clock neuron firing rate resets molecular circadian rhythms and behavior. *Nat. Neurosci.* **18**, 373–375 (2015).
74. Henslee, E. A. et al. Rhythmic potassium transport regulates the circadian clock in human red blood cells. *Nat. Commun.* **8**, 1978 (2017).
75. Martins, B. M. C., Tooke, A. K., Thomas, P. & Locke, J. C. W. Cell size control driven by the circadian clock and environment in cyanobacteria. *Proc. Natl Acad. Sci. USA* **115**, E11415–E11424 (2018).
76. Bruning, F. et al. Sleep-wake cycles drive daily dynamics of synaptic phosphorylation. *Science* **366**, eaav3617 (2019).
77. Dallmann, R., Brown, S. A. & Gachon, F. Chronopharmacology: new insights and therapeutic implications. *Annu Rev. Pharm. Toxicol.* **54**, 339–361 (2014).
78. Tahara, Y. & Shibata, S. Chrono-biology, chrono-pharmacology, and chrono-nutrition. *J. Pharm. Sci.* **124**, 320–335 (2014).
79. Ozturk, N., Ozturk, D., Kavakli, I. H. & Okyar, A. Molecular aspects of circadian pharmacology and relevance for cancer chronotherapy. *Int. J. Mol. Sci.* **18**, 2168 (2017).
80. Kulak, N. A., Geyer, P. E. & Mann, M. Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol. Cell Proteom.* **16**, 694–705 (2017).
81. Wichmann, C. et al. MaxQuant. Live enables global targeting of more than 25,000 peptides. *Mol. Cell Proteom.* **18**, 982–994 (2019).
82. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn (L. Erlbaum Associates, 1988).
83. Ben-Shachar, M., Makowski, D. & Lüdtke, D. Compute and interpret indices of effect size. *CRAN* (2020).
84. Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V. & Gorshkov, M. V. Pyteomics—a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* **24**, 301–304 (2013).
85. Levitsky, L. I., Klein, J. A., Ivanov, M. V. & Gorshkov, M. V. Pyteomics 4.0: five years of development of a python proteomics framework. *J. Proteome Res.* **18**, 709–714 (2019).

Acknowledgements

F.B. and M.S.R. were supported by the Volkswagen Foundation (93071), M.S.R. also received funding from the German Research Foundation DFG (Project 329628492-SFB1321, INST 86/1800-1 FUGG, and project 428041612). We thank all the members of the Department of Signal Transduction and Proteomics, in particular, Igor Paron and Christian Deiml for MS technical assistance, Bianca Splettoesser for technical help with experimental work, and Johannes Müller, Alex Strasser, and Lisa Schweitzer for columns. We also thank Sarah Meszaros and Steven Dewitz for cell culture assistance. We are grateful to Arno F. Alpi, Jesper Olsen, Chuna Choudhary, André Michaelis, and Jakob Bader for constructive and insightful discussions. We specially thank Roberto Polakiewicz, Florian Gnad, Sean Landry, and Cell Signaling Technology (CST) for gifting of PTMScan® Ubiquitin Remnant Motif (K-ε-GG) Kits.

Author contributions

F.M.H., O.K., M.T., F.B., and M.S.R. designed experiments. TNF experiments were performed by F.M.H., M.T., and C.S. Circadian experiments were conducted by F.M.H. and F.B. Computational proximity analysis was performed by I.B. Data were analyzed by F.M.H., O.K., M.T., M.S.R., B.A.S., and M.M. wrote, reviewed, and edited the manuscript. All authors read and commented on the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-20509-1>.

Correspondence and requests for materials should be addressed to M.S.R., O.K. or M.M.

Peer review information *Nature Communications* thanks Ronald Hay and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

3.2 Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV

Alexey Stukalov^{1,15}, Virginie Girault^{1,15}, Vincent Grass^{1,15}, Ozge Karayel^{2,15}, Valter Bergant^{1,15}, Christian Urban^{1,15}, Darya A. Haas^{1,15}, Yiqi Huang^{1,15}, Lila Oubraham¹, Anqi Wang¹, M. Sabri Hamad¹, Antonio Piras¹, **Fynn M. Hansen**², Maria C. Tanzer², Igor Paron², Luca Zinzula³, Thomas Engleitner⁴, Maria Reinecke^{5,6,7}, Teresa M. Lavacca¹, Rosina Ehmann^{8,9}, Roman Wölfel^{8,9}, Jörg Jores¹⁰, Bernhard Kuster^{5,6,7}, Ulrike Protzer^{1,9}, Roland Rad⁴, John Ziebuhr¹¹, Volker Thiel^{12,13}, Pietro Scaturro^{1,14}, Matthias Mann² & Andreas Pichlmair^{1,9}

¹Technical University of Munich, School of Medicine, Institute of Virology, Munich, Germany. ²Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, Germany. ³Department of Molecular Structural Biology, Max-Planck Institute of Biochemistry, Martinsried, Germany. ⁴Institute of Molecular Oncology and Functional Genomics and Department of Medicine II, School of Medicine, Technical University of Munich, Munich, Germany. ⁵Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany. ⁶German Cancer Consortium (DKTK), Munich Partner Site, Munich, Germany. ⁷German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Bundeswehr Institute of Microbiology, Munich, Germany. ⁹German Center for Infection Research (DZIF), Munich Partner Site, Munich, Germany. ¹⁰Institute of Veterinary Bacteriology, Department of Infectious Diseases and Pathobiology, University of Bern, Bern, Switzerland. ¹¹Justus Liebig University Giessen, Institute of Medical Virology, Giessen, Germany. ¹²Institute of Virology and Immunology (IVI), Bern, Switzerland. ¹³Department of Infectious Diseases and Pathobiology, University of Bern, Bern, Switzerland. ¹⁴Leibniz Institute for Experimental Virology, Hamburg, Germany. ¹⁵These authors contributed equally: Alexey Stukalov, Virginie Girault, Vincent Grass, Ozge Karayel, Valter Bergant, Christian Urban, Darya A. Haas, Yiqi Huang.

Published in *Nature* (2021)

The Covid pandemic inflicted by the emergence of SARS-CoV-2 took the world by surprise and caused a tremendous research effort to elucidate the underlying principles of SARS-CoV-2 infections to find means for Covid treatment. MS-based proteomics was a core tool for such investigations among many other techniques. Multiple groups, led by Prof. Andreas Pichlmair, set out to reveal host perturbations by SARS-CoV-2 infections by multilevel proteomics and transcriptomics. The resulting study describes protein interactomes of infected cells and their influence on the transcriptome, phosphoproteome and ubiquitinome. Furthermore, host signaling differences between SARS-CoV-2 and SARS-CoV infection identified inhibitors of kinases and matrix metalloproteases as potential drugs for Covid treatment.

In this study we provided our expertise in PTM analysis. Especially my work on DIA-based ubiquitinome analysis helped to establish tailor-made DIA analysis strategies for this sample type. We expeditiously used the newly released FAIMS (High-Field Asymmetric Waveform Ion Mobility Spectrometry) device to develop an ion-mobility-based approach for spectral library construction that allowed us to record spectral libraries with low sample amount requirements.

Article

Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV

<https://doi.org/10.1038/s41586-021-03493-4>

Received: 12 June 2020

Accepted: 25 March 2021

Published online: 12 April 2021

 Check for updates

Alexey Stukalov^{1,15}, Virginie Girault^{1,15}, Vincent Grass^{1,15}, Ozge Karayel^{2,15}, Valter Bergant^{1,15}, Christian Urban^{1,15}, Darya A. Haas^{1,15}, Yiqi Huang^{1,15}, Lila Oubraham¹, Anqi Wang¹, M. Sabri Hamad¹, Antonio Piras¹, Fynn M. Hansen², Maria C. Tanzer², Igor Paron², Luca Zinzula³, Thomas Engleitner⁴, Maria Reinecke^{5,6,7}, Teresa M. Lavacca^{1,9}, Rosina Ehmann^{8,9}, Roman Wölfel^{8,9}, Jörg Jores¹⁰, Bernhard Kuster^{5,6,7}, Ulrike Protzer^{1,9}, Roland Rad⁴, John Ziebuhr¹¹, Volker Thiel^{12,13}, Pietro Scaturro^{1,14}, Matthias Mann² & Andreas Pichlmair^{1,9}✉

The emergence and global spread of SARS-CoV-2 has resulted in the urgent need for an in-depth understanding of molecular functions of viral proteins and their interactions with the host proteome. Several individual omics studies have extended our knowledge of COVID-19 pathophysiology^{1–10}. Integration of such datasets to obtain a holistic view of virus–host interactions and to define the pathogenic properties of SARS-CoV-2 is limited by the heterogeneity of the experimental systems. Here we report a concurrent multi-omics study of SARS-CoV-2 and SARS-CoV. Using state-of-the-art proteomics, we profiled the interactomes of both viruses, as well as their influence on the transcriptome, proteome, ubiquitinome and phosphoproteome of a lung-derived human cell line. Projecting these data onto the global network of cellular interactions revealed crosstalk between the perturbations taking place upon infection with SARS-CoV-2 and SARS-CoV at different levels and enabled identification of distinct and common molecular mechanisms of these closely related coronaviruses. The TGF- β pathway, known for its involvement in tissue fibrosis, was specifically dysregulated by SARS-CoV-2 ORF8 and autophagy was specifically dysregulated by SARS-CoV-2 ORF3. The extensive dataset (available at <https://covinet.innatelab.org>) highlights many hotspots that could be targeted by existing drugs and may be used to guide rational design of virus- and host-directed therapies, which we exemplify by identifying inhibitors of kinases and matrix metalloproteases with potent antiviral effects against SARS-CoV-2.

To identify protein–protein interactions of SARS-CoV-2 and SARS-CoV and cellular proteins, we transduced A549 lung carcinoma cells with lentiviruses expressing individual haemagglutinin-tagged viral proteins (Fig. 1a, Extended Data Fig. 1a, Supplementary Table 1). Statistical modelling of quantitative data from affinity purification followed by mass spectrometry (AP–MS) analysis identified 1,801 interactions between 1,086 cellular proteins and 24 SARS-CoV-2 and 27 SARS-CoV bait proteins (Fig. 1b, Extended Data Fig. 1b, Supplementary Table 2), substantially increasing the number of reported interactions of SARS-CoV-2 and SARS-CoV^{1,2,5,6,10,11} (Supplementary Table 10). The resulting virus–host interaction network revealed a wide range of cellular activities intercepted by SARS-CoV-2 and SARS-CoV (Fig. 1b, Extended Data Table 1, Supplementary Table 2). In particular, we observed that SARS-CoV-2 targets a number of key innate immunity regulators (ORF7b–MAVS and

ORF7b–UNC93B1), stress response components (N–HSPA1A) and DNA damage response mediators (ORF7a–ATM and ORF7a–ATR) (Fig. 1b, Extended Data Fig. 1c–e). Additionally, SARS-CoV-2 proteins interact with molecular complexes involved in intracellular trafficking (for example, endoplasmic reticulum–Golgi trafficking) and transport (for example, solute carriers and ion transport by ATPases) as well as cellular metabolism (for example, mitochondrial respiratory chain and glycolysis) (Fig. 1b, Extended Data Table 1, Supplementary Table 2). Comparing the AP–MS data of homologous SARS-CoV-2 and SARS-CoV proteins identified differences in the enrichment of individual host targets, highlighting potential virus-specific interactions (Fig. 1b (edge colour), c, Extended Data Figs. 1f, 2a, b, Supplementary Table 2). For instance, we recapitulated the known interactions between SARS-CoV NSP2 and prohibitins¹² (PHB and PHB2), but this interaction was not

¹Technical University of Munich, School of Medicine, Institute of Virology, Munich, Germany. ²Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, Germany. ³Department of Molecular Structural Biology, Max-Planck Institute of Biochemistry, Martinsried, Germany. ⁴Institute of Molecular Oncology and Functional Genomics and Department of Medicine II, School of Medicine, Technical University of Munich, Munich, Germany. ⁵Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany. ⁶German Cancer Consortium (DKTK), Munich Partner Site, Munich, Germany. ⁷German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Bundeswehr Institute of Microbiology, Munich, Germany. ⁹German Center for Infection Research (DZIF), Munich Partner Site, Munich, Germany. ¹⁰Institute of Veterinary Bacteriology, Department of Infectious Diseases and Pathobiology, University of Bern, Bern, Switzerland. ¹¹Justus Liebig University Giessen, Institute of Medical Virology, Giessen, Germany. ¹²Institute of Virology and Immunology (IVI), Bern, Switzerland. ¹³Department of Infectious Diseases and Pathobiology, University of Bern, Bern, Switzerland. ¹⁴Leibniz Institute for Experimental Virology, Hamburg, Germany. ¹⁵These authors contributed equally: Alexey Stukalov, Virginie Girault, Vincent Grass, Ozge Karayel, Valter Bergant, Christian Urban, Darya A. Haas, Yiqi Huang. ✉e-mail: andreas.pichlmair@tum.de

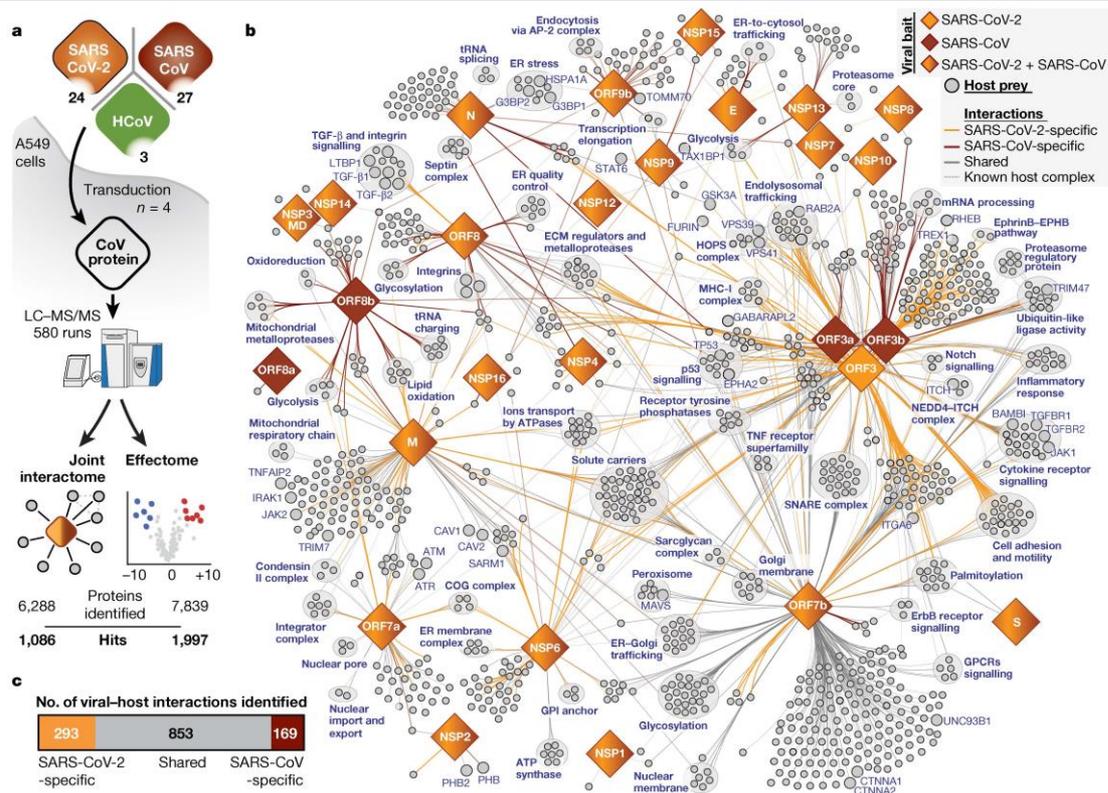


Fig. 1 | Joint analysis of SARS-CoV-2 and SARS-CoV virus-host protein-interactomes. **a**, Systematic comparison of interactomes and host proteome changes (effectomes) of the 24 SARS-CoV-2 and 27 SARS-CoV viral proteins, using 3 homologues from human coronaviruses (HCoV-NL63 and HCoV-229E) as reference for pan-coronavirus specificity. **b**, Combined virus-host protein-interaction network of SARS-CoV-2 and SARS-CoV measured by AP-MS. Homologous viral proteins are displayed as a single node. Shared and virus-specific interactions are denoted by the edge colour. The edge intensity

reflects the *P*-value for the interaction (with the smallest *P*-value represented by solid edges and the highest *P*-value (<0.001) represented by faded edges). ECM, extracellular matrix; ER, endoplasmic reticulum; GPCR, G-protein-coupled receptor; HOPS, homotypic fusion and protein-sorting; MHC, major histocompatibility complex; SNARE, soluble N-ethylmaleimide-sensitive factor attachment protein receptor; COG, conserved oligomeric Golgi. **c**, The numbers of unique and shared host interactions between the homologous proteins of SARS-CoV-2 and SARS-CoV.

conserved with SARS-CoV-2 NSP2, suggesting that the two viruses differ in their ability to modulate mitochondrial function and homeostasis through NSP2 (Extended Data Fig. 2a). The exclusive interaction of SARS-CoV-2 ORF8 with the TGF- β 1-LTBP1 complex is another interaction that potentially explains the differences in pathogenicity of the two viruses (Extended Data Figs. 1f, 2b). Notably, disbalanced TGF- β signalling has been linked to lung fibrosis and oedema, a common complication of severe pulmonary diseases including COVID-19¹³⁻¹⁶.

To map the virus-host interactions to the functions of viral proteins, we conducted a study of total proteomes of A549 cells expressing 54 individual viral proteins comprising the ‘effectome’ (Fig. 1a, Supplementary Table 3). This dataset provides clear links between changes in protein expression and virus-host interactions, as exemplified by ORF9b, which leads to a dysregulation of mitochondrial functions and binds to TOMM70, a known regulator of mitophagy^{2,17} (Fig. 1b, Supplementary Tables 2, 3). Global pathway-enrichment analysis of the effectome dataset confirmed that ORF9b of both viruses led to mitochondrial dysregulation^{2,18} (Extended Data Fig. 2c, Supplementary Table 3) and further highlighted virus-specific effects, as exemplified by the upregulation of proteins involved in cholesterol metabolism

(CYP51A1, DHCR7, IDI1 and SQLE) by SARS-CoV-2 NSP6 but not by SARS-CoV NSP6. Of note, cholesterol metabolism was recently shown to be implicated in SARS-CoV-2 replication and has been suggested as a promising target for drug development¹⁹⁻²¹. Besides perturbations at the pathway level, viral proteins also specifically modulated single host proteins, possibly explaining more specific molecular mechanisms involved in viral protein function. Focusing on the 180 most affected host proteins, we identified RCOR3, a putative transcriptional corepressor, as strongly upregulated by NSP4 of both viruses (Extended Data Figs. 2d, 3a). Notably, apolipoprotein B (APOB) was substantially regulated by ORF3 and NSP1 of SARS-CoV-2, suggesting that it has an important role in SARS-CoV-2 biology (Extended Data Fig. 3b).

Multi-omics profiling of virus infection

Although the interactome and the effectome provide in-depth information on the activity of individual viral proteins, we aimed to directly study their combined activities in the context of viral infection. To this end, we infected A549 cells expressing angiotensin-converting enzyme 2 (ACE2) (A549-ACE2 cells) (Extended Data Fig. 4a, b) with SARS-CoV-2

Article

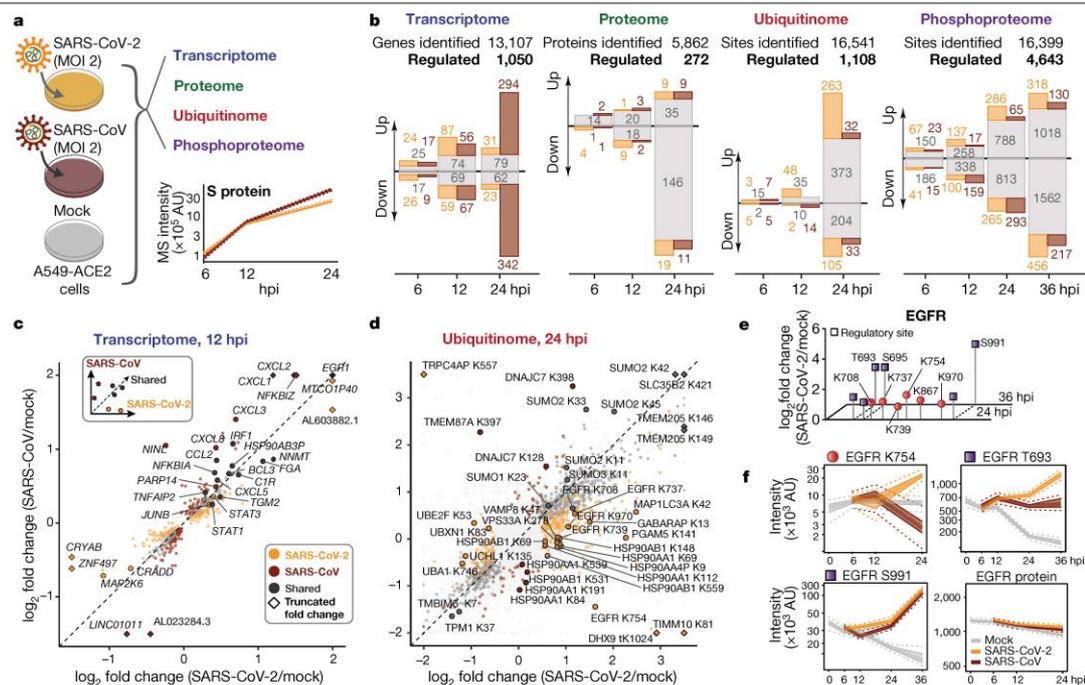


Fig. 2 | Multi-level profiling of SARS-CoV-2 and SARS-CoV infection.

a, Time-resolved profiling of parallel SARS-CoV-2 and SARS-CoV infection by multiple omics methods. The plot shows the mass spectrometry (MS) intensity estimates for spike proteins of SARS-CoV-2 and SARS-CoV over time ($n = 4$ independent experiments). MOI, multiplicity of infection. **b**, The numbers of distinct transcripts, proteins, ubiquitination and phosphorylation sites that are significantly up- or downregulated at given time points after infection (relative to mock infection at the same time point). Transcripts, proteins or sites that are regulated similarly by SARS-CoV-2 and SARS-CoV infection are shown in grey, those regulated specifically by SARS-CoV-2 are in orange and those regulated by SARS-CoV are in brown. **c**, **d**, Comparison of host transcriptome 12 h (c) and ubiquitinome 24 h (d) after infection (hpi) with SARS-CoV-2 (x-axis) or SARS-CoV (y-axis) (\log_2 fold change in comparison to the mock infection samples at the same time point). Significantly regulated transcripts by moderated *t*-test with false discovery rate-corrected two-sided

P-value ≤ 0.05 (c) and significantly regulated sites by Bayesian linear model-based unadjusted two-sided *P*-value $\leq 10^{-3}$, $|\log_2$ fold change| ≥ 0.5 (d) are coloured according to specificity as indicated. Diamonds indicate that the actual \log_2 fold change was truncated to fit into the plot. $n = 3$ independent experiments. **e**, Phosphorylation (purple squares) and ubiquitination (red circles) sites on EGFR that are regulated upon SARS-CoV-2 infection. The plot shows median \log_2 fold change of abundance compared with mock infection at 24 and 36 hpi. All identified phosphorylation sites have known regulatory function. **f**, Profile plots of time-resolved EGFR K754 ubiquitination, EGFR T693 and S991 phosphorylation, and total EGFR protein levels in A549-ACE2 cells infected with SARS-CoV-2 or SARS-CoV with indicated median (line), 50% (shaded region) and 95% (dotted line) confidence intervals. $n = 3$ (ubiquitination) or 4 (phosphorylation and total protein) independent experiments.

or SARS-CoV, and profiled the effects of viral infection on mRNA expression, protein abundance, ubiquitination and phosphorylation in a time-resolved manner (Fig. 2 a–b).

In line with previous reports²², we found that both SARS-CoV-2 and SARS-CoV can downregulate the type I interferon response and activate a pro-inflammatory signature at transcriptome and proteome levels (Fig. 2a–c, Extended Data Fig. 4c–f, i, Supplementary Tables 4, 8, Supplementary Discussion 1). However, SARS-CoV elicited a more pronounced activation of the NF- κ B pathway, correlating with its higher replication rate and potentially explaining the lower severity of pulmonary disease in cases of SARS-CoV-2 infection²³ (Supplementary Tables 4, 5). By contrast, SARS-CoV-2 infection led to higher expression of FN1 and SERPINE1, which may be linked to the specific recruitment of TGF- β factors (Fig. 1b), supporting regulation of TGF- β signalling by SARS-CoV-2.

To better understand the mechanisms underlying perturbation of cellular signalling, we performed comparative ubiquitination and phosphorylation profiling following infection with SARS-CoV-2 or SARS-CoV.

This analysis showed that 1,108 of 16,541 detected ubiquitination sites were differentially regulated by infection with SARS-CoV-2 or SARS-CoV (Fig. 2a, b, d, Extended Data Fig. 5a, Supplementary Table 6). More than half of the significant sites were regulated in a similar manner by both viruses. These included sites on SLC35 and SUMO family proteins, indicating possible regulation of sialic acid transport and the SUMO activity. SARS-CoV-2 specifically increased ubiquitination on autophagy-related factors (MAP1LC3A, GABARAP, VPS33A and VAMP8) as well as specific sites on EGFR (for example, K739, K754 and K970). In some cases, the two viruses targeted distinct sites on the same cellular protein, as exemplified by HSP90 family members (for example, K84, K191 and K539 on HSP90AA1) (Fig. 2d). Notably, a number of proteins (for example, ALCAM, ALDH3B1, CTNNA1, EDF1 and SLC12A2) exhibited concomitant ubiquitination and a decrease at the protein level after infection, pointing to ubiquitination-mediated protein degradation (Fig. 2d, Extended Data Figs. 4f, 5a, Supplementary Tables 5, 6). Among these downregulated proteins, EDF1 has a pivotal role in the maintenance of endothelial integrity and may be a link to endothelial dysfunctions

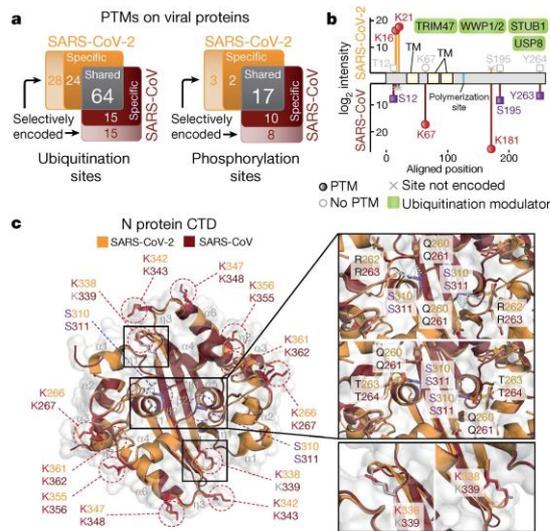


Fig. 3 | Phosphorylation and ubiquitination of SARS-CoV-2 and SARS-CoV viral proteins. **a**, Distribution of identified shared, differentially regulated and selectively encoded (sequence-specific) ubiquitination and phosphorylation sites on SARS-CoV-2 and SARS-CoV homologous proteins after infection of A549-ACE2 cells. PTM, post-translational modification. **b**, Mapping of the ubiquitination (red circles) and phosphorylation (purple squares) sites on an alignment of SARS-CoV-2 ORF3 and SARS-CoV ORF3a proteins, showing median log₂ intensities in virus-infected A549-ACE2 cells at 24 hpi. Functional (blue) and topological (yellow) domains are mapped on each sequence. Ubiquitin-modifying enzymes binding to ORF3 and ORF3a as identified in our AP-MS experiments (Extended Data Fig. 1b) are indicated (green). TM, transmembrane domain. **c**, Surface and ribbon representation of superimposed SARS-CoV (Protein Data Bank (PDB) ID: 2CJR, brown) and SARS-CoV-2 (PDB ID: 6YUN, orange) N protein CTD dimers (r.m.s.d. values of 0.492 Å for 108 matching Cα atoms). Secondary structures are numbered in grey (prefixed with α for α-helix, β for β-strand and η for non-structured regions). Side chains are colour coded depending on whether they are in ubiquitinated (red), phosphorylated (purple) or unmodified (grey) sites. The K338 ubiquitination site unique to SARS-CoV-2 is shown as a close-up for both monomers (bottom right). Close-ups of inter-chain residue interactions established by non-phosphorylated (top right) and phosphorylated (middle right) SARS-CoV-2 S310 or SARS-CoV S311.

described for COVID-19^{24,25}. Profound regulation of cellular signalling pathways was also observed at the phosphoproteomic level: among 16,399 total quantified phosphorylation sites, 4,643 showed significant changes after infection with SARS-CoV-2 or SARS-CoV (Extended Data Fig. 5b, c, Supplementary Table 7). Highly regulated sites were identified for the proteins of the MAPK pathways (for example, MAPKAPK2, MAP2K1, JUN and SRC), and proteins involved in autophagy signalling (for example, DEPTOR, RICTOR, OPTN, SQSTM1 and LAMTOR1) and viral entry (for example, ACE2 and RAB7A) (Extended Data Fig. 5b, d). Notably, RAB7A was recently shown to be an important host factor for SARS-CoV-2 infection that assists endosomal trafficking of ACE2 to the plasma membrane²⁶. We observed higher phosphorylation at S72 of RAB7A in SARS-CoV-2 infection compared with SARS-CoV or mock infection; this site is implicated in RAB7A intracellular localization and molecular association²⁷. The regulation of known phosphorylation sites suggests an involvement of central kinases (cyclin dependent kinases, AKT, MAPKs, ATM, and CHEK1) linked to cell survival, cell cycle progression, cell growth and motility, stress responses and the DNA

damage response; this was also supported by the analysis of enriched motifs (Extended Data Fig. 5e, f, Supplementary Tables 7, 8). Notably, SARS-CoV-2 infection, but not SARS-CoV infection, led to phosphorylation of the antiviral kinase EIF2AK2 (also known as PKR) at the critical regulatory residue S33²⁸. This differential activation of EIF2AK2 could contribute to the difference in the growth kinetics of the two viruses (Supplementary Table 4, 5).

Our data clearly point to an interplay of phosphorylation and ubiquitination patterns on individual host proteins. For instance, EGFR showed increased ubiquitination on 6 lysine residues at 24 h post-infection (hpi) accompanied by increased phosphorylation of T693, S695 and S991 at 24 and 36 hpi (Fig. 2e, f). Ubiquitination of all six lysine residues on EGFR was more pronounced following infection with SARS-CoV-2 than with SARS-CoV. Moreover, vimentin, a central co-factor for coronavirus entry²⁹ and pathogenicity^{30,31}, displayed distinct phosphorylation and ubiquitination patterns on several sites early (for example, S420) or late (for example, S56, S72 and K334) in infection (Extended Data Fig. 6a, b). These findings underscore the value of testing different post-translational modifications simultaneously and suggest a concerted engagement of regulatory machineries to modify target protein functions and abundance.

Post-translational modification of viral proteins

The majority of viral proteins were also post-translationally modified. Of the 27 detected SARS coronavirus proteins, 21 were ubiquitinated. Nucleocapsid (N), spike (S), NSP2 and NSP3 were the most heavily modified proteins in both viruses (Extended Data Fig. 6c, Supplementary Table 6). Many ubiquitination sites were common to both viruses. Around half of the sites that were exclusively ubiquitinated in either virus were conserved between SARS-CoV and SARS-CoV-2. The remaining specifically regulated ubiquitination sites were unique to each virus, indicating that these acquired adaptations can be post-translationally modified and may recruit cellular proteins with distinct functions (Fig. 3a). Our interactome data identified several host E3 ligases (for example, we identified interactions between SARS-CoV-2 ORF3 and TRIM47, WWP1, WWP2 and STUB1; and between SARS-CoV-2 M and TRIM7) and deubiquitinating enzymes (for example, interactions between SARS-CoV-2 ORF3 and USP8; SARS-CoV-2 ORF7a and USP34; and SARS-CoV N and USP9X), suggesting crosstalk between ubiquitination and viral protein functions (Fig. 1b, Extended Data Fig. 6d, Supplementary Table 2). Of particular interest are extensive ubiquitination events on the S protein of both viruses (K97, K528, K825, K835, K921 and K947), which are distributed on functional domains (N-terminal domain, C-terminal domain (CTD), fusion peptide and heptad repeat 1 domain), potentially indicating critical regulatory functions that are conserved between the two viruses (Extended Data Fig. 6e). We observed phosphorylation of 5 SARS-CoV-2 proteins (M, N, S, NSP3 and ORF9b) and 8 SARS-CoV proteins (M, N, S, NSP1, NSP2, NSP3, ORF3 and ORF9b) (Extended Data Fig. 6f, Supplementary Table 7), on sites corresponding to known recognition motifs. In particular, CAMK4 and MAPKAPK2 potentially phosphorylate sites on S and N, respectively. Phosphorylation of cellular proteins suggested that the activities of these kinases were enriched in cells infected with SARS-CoV-2 or SARS-CoV (Extended Data Figs. 5e, f, 6e, g). Moreover, N proteins of both SARS-CoV-2 and SARS-CoV recruit GSK3, which could potentially be linked to phosphorylation events on these viral proteins (Fig. 1b, Extended Data Fig. 6g, Supplementary Table 7). Notably, we identified novel post-translationally modified sites located at functional domains of viral proteins; we detected ubiquitination at SARS-CoV-2 N K338 and phosphorylation on SARS-CoV-2 and SARS-CoV N S310 and S311 (Extended Data Fig. 6g). Mapping these sites to the atomic structure of the CTD^{32,33} highlights critical positions for the function of the protein (Fig. 3c, Extended Data Fig. 6h, Supplementary Discussion 2). Collectively, while the identification of differentially regulated sites

Article

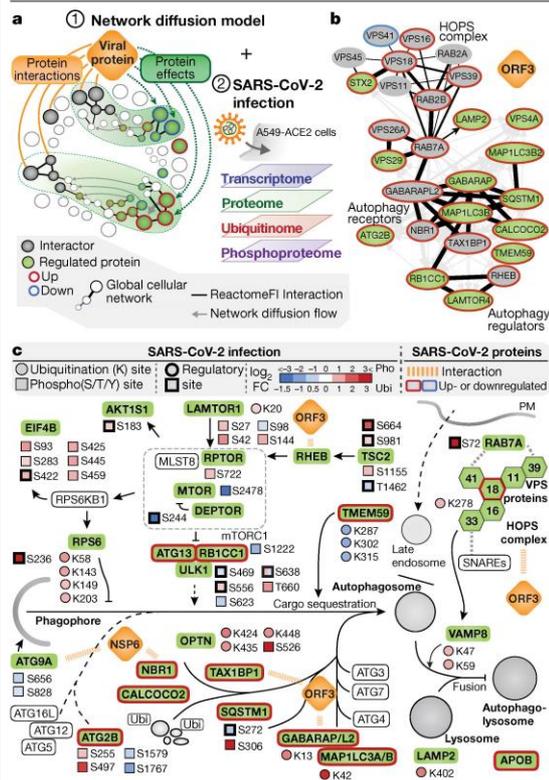


Fig. 4 | A network diffusion approach enables identification of molecular pathways linking protein-protein interactions with downstream changes in the host proteome. a, The network diffusion approach to identify functional connections between the host targets of a viral protein and downstream proteome changes. The results of network diffusion are integrated with omics datasets of SARS coronavirus infection to streamline the identification of affected host pathways. **b**, Subnetworks of the network diffusion linking host targets of SARS-CoV-2 ORF3 to factors involved in autophagy. The thickness of directed edges is proportional to the random-walk transition probability. Black edges denote connections present in ReactomeFl. **c**, Overview of perturbations to host-cell autophagy induced by SARS-CoV-2. The pathway regulation is derived from the network diffusion model for SARS-CoV-2 ORF3 and NSP6 and is overlaid with the changes in protein levels, ubiquitination (Ubi) and phosphorylation (Pho) induced by SARS-CoV-2 infection. FC, fold change; PM, plasma membrane.

(Supplementary Table 8). RHO GTPase activation, mRNA processing and the role of ABL in ROBO-SLIT signalling appeared to be regulated mostly through phosphorylation (Extended Data Fig. 7a). By contrast, processes connected to cellular integrity such as the formation of senescence-associated heterochromatin foci, apoptosis-induced DNA fragmentation and amino acid transport across the plasma membrane were modulated through concomitant phosphorylation and ubiquitination events, suggesting molecular links between these post-translational modifications. Ion transporters, especially the SLC12 family of cation-coupled chloride cotransporters—previously identified as cellular factors in pulmonary inflammation³⁴—were also regulated at multiple levels, evidenced by reduced protein abundance as well as differential post-translational modifications (Extended Data Fig. 7a).

The pathway-enrichment analysis provided a global and comprehensive picture of how SARS-CoV-2 and SARS-CoV affect the host. We next applied an automated approach to systematically explore the underlying molecular mechanisms contained in the viral interactome and effectome data. We mapped the measured interactions and effects of each viral protein onto the global network of cellular interactions³⁵ and applied a network diffusion approach³⁶ (Fig. 4a). This type of analysis uses known cellular protein-protein interactions, signalling and regulation events to identify connection points between cellular proteins that interact with viral proteins and the proteins affected by the expression of these viral proteins (Extended Data Figs. 1b, 2d, Supplementary Tables 2, 3). The connections inferred from the real data were significantly shorter than for randomized data, validating the relevance of the approach and the quality of the data (Extended Data Fig. 8a, b). The findings from this approach include the potential mechanisms by which ORF3 and NSP6 may regulate autophagy, the modulation of innate immunity by M, ORF3 and ORF7b, and the perturbation of integrin-TGF- β -EGFR-receptor tyrosine kinase signalling by ORF8 of SARS-CoV-2 (Fig. 4b, Extended Data Fig. 8c, d). Enriching these subnetworks with data on SARS-CoV-2 infection-dependent mRNA abundance, protein abundance, phosphorylation and ubiquitination (Fig. 4a) provided insights into the regulatory mechanisms activated by SARS-CoV-2. For instance, the analysis confirmed a role of NSP6 in both SARS-CoV-2- and SARS-CoV-induced autophagy³⁷ and revealed the SARS-CoV-2 specific inhibition of autophagic flux by ORF3 protein, leading to the accumulation of autophagy receptors (SQSTM1, GABARAPL2, NBR1, CALCOCO2, MAP1LC3A, MAP1LC3B and TAX1BP1), consistent with the accumulation of MAP1LC3B protein observed in cells infected with SARS-CoV-2 (Fig. 4c, Extended Data Fig. 8e, f). This inhibition may be a result of the interaction of the ORF3 protein with the HOPS complex (VPS11, VPS16, VPS18, VPS39 and VPS41), which is essential for autophagosome-lysosome fusion, as well as the differential phosphorylation of regulatory sites (for example, on TSC2, mTORC1 complex, ULK1, RPS6 and SQSTM1) and ubiquitination of key components (MAP1LC3A, GABARAPL2, VPS33A and VAMP8) (Fig. 4c, Extended Data Fig. 8g). This inhibition of autophagosome function may have direct consequences for protein degradation. The abundance of APOB, a protein that is degraded via autophagy³⁸, was selectively increased after SARS-CoV-2 infection or expression of SARS-CoV-2 ORF3 (Extended Data Fig. 3b, 8h). Accumulating APOB levels could increase the risk of arterial thrombosis³⁹, one of the main complications contributing to lung, heart and kidney failure in patients with COVID-19⁴⁰. The inhibition of the interferon (IFN)- α and IFN- β response observed at transcriptional and proteome levels was similarly explained by the network diffusion analysis (Extended Data Fig. 8i), which implicated multiple proteins of SARS-CoV-2 in the disruption of antiviral immunity. Further experiments functionally corroborated the inhibition of IFN- α and IFN- β induction or signalling by ORF3, ORF6, ORF7a, ORF7b and ORF9b (Extended Data Fig. 8j). Upon virus infection, we observed the regulation of TGF- β and EGFR pathways modulating cell survival, motility and innate immune responses (Extended Data Fig. 9a-d). Specifically, our network diffusion analysis revealed a connection between

may indicate pathogen-specific functions, insights from conserved post-translational modifications may also provide useful knowledge for the development of targeted pan-antiviral therapies.

Viral perturbation of key cellular pathways

Our unified experimental design in a syngeneic system enabled direct time-resolved comparison of SARS-CoV-2 and SARS-CoV infection across different levels. Integrative pathway-enrichment analysis demonstrated that both viruses largely perturb the same cellular processes at multiple levels, albeit with distinct temporal patterns (Extended Data Fig. 7a). For instance, transcriptional downregulation of proteins involved in tau protein kinase activity and Fe ion sequestration at 6 hpi was followed by a decrease in protein abundance after 12 hpi

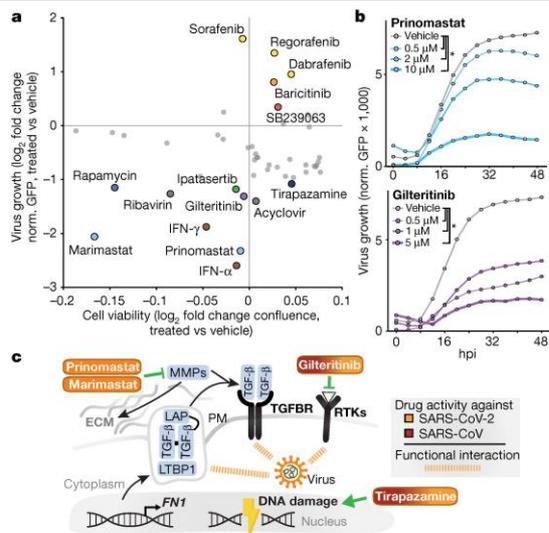


Fig. 5 | Identification of pathways targeted by SARS-CoV-2 using a multi-omics profiling approach enables systematic testing of candidate antiviral therapies. a, b, A549-ACE2 cells were treated with the indicated drugs 6 h before infection with SARS-CoV-2-GFP (MOI of 3). Changes in cell viability and virus growth (a) in drug-treated cells compared with untreated A549-ACE2 cells at 48 hpi. A confluence cut-off of $-0.2 \log_2$ fold change was applied to remove cytotoxic compounds. **b,** Time courses of virus replication after pre-treatment of cells with prinomastat or giliteritinib. $n = 4$ independent experiments; $^*P \leq 0.01$ compared with control treatment, unadjusted two-sided Wilcoxon test. Norm., normalized. **c,** Drugs potentially targeting pathways identified in our study. Colour indicates antiviral activity against SARS-CoV-2 or SARS-CoV (brown–orange gradient) or SARS-CoV-2 specifically (orange), as inferred from in vitro experiments.

the binding of the ORF8 and ORF3 proteins to TGF-β-associated factors (TGF-β1, TGF-β2, LTBP1, TGFBR2, FURIN and BAMBI), the differential expression of extracellular matrix regulators (FERMT2 and CDH1) and the virus-induced upregulation of fibrinogens (FGA, FGB), fibronectin (FNI) and SERPINE1⁴¹ (Extended Data Fig. 9a, b). The increased phosphorylation of proteins involved in MAPK signalling (for example, SHC1 (on S139), SOS1 (S1134/S1229), JUN (S63/S73), MAPKAPK2 (T334) and p38 (T180/Y182)) and receptor tyrosine kinase signalling (for example, phosphorylation of PI3K complex members PDPK1 (S241) and RPS6KA1 (S380)) as well as a higher expression of JUN, FOS and EGFR are further indications of regulation of TGF-β and EGFR pathways (Extended Data Fig. 9a, c, d). In turn, TGF-β and EGFR signalling are known to be potentiated by integrin signalling and activation of YAP-dependent transcription⁴², which we observed to be regulated in a time-dependent manner upon SARS-CoV-2 infection (Extended Data Fig. 9a). As well as promoting virus replication, activation of these pathways has been implicated in fibrosis^{13–15}, one of the hallmarks of COVID-19¹⁶.

Data-guided drug identification and testing

Together, the viral protein–host protein interactions and regulation of pathways observed at multiple levels identify potential points for targeting SARS-CoV and SARS-CoV-2 using well-characterized selective antiviral drugs. To test antiviral efficacy, we used time-lapse fluorescent microscopy of infection with a GFP reporter SARS-CoV-2⁴³. Inhibition of virus replication by treatment with IFN-α corroborated previous conclusions that efficient SARS-CoV-2 replication involves the inactivation

of this pathway at an early step^{9,44} and confirmed the reliability of this screening approach (Extended Data Fig. 10a). We tested a panel of 48 drugs that modulate the pathways perturbed by the virus for their effects on SARS-CoV-2 replication (Fig. 5a, Supplementary Table 9). Of note, inhibitors of B-RAF (sorafenib, regorafenib and dabrafenib), JAK1/2 (baricitinib) and MAPK (SB239063), which are commonly used to treat cancer and autoimmune diseases^{45–47}, significantly increased virus growth in an in vitro model of infection (Fig. 5a, Extended Data Fig. 10b, Supplementary Table 9). By contrast, inducers of DNA damage (tirapazamine and rabeprazole) or an mTOR inhibitor (rapamycin) suppressed virus growth. The highest antiviral activity was observed for giliteritinib (a designated inhibitor of FLT3 and AXL), ipatasertib (an AKT inhibitor), prinomastat and marimastat (matrix metalloprotease (MMP) inhibitors) (Fig. 5a, b, Extended Data Fig. 10c, Supplementary Table 9). These compounds profoundly inhibited replication of SARS-CoV-2 while having no effects or minor effects on cell growth (Extended Data Fig. 10b, Supplementary Table 9). Quantitative PCR analysis indicated antiviral activities for giliteritinib and tirapazamine against SARS-CoV-2 and SARS-CoV (Fig. 5c, Extended Data Fig. 10d, e). Notably, prinomastat and marimastat, specific inhibitors of MMP2 and MMP9, showed selective activity against SARS-CoV-2 but not against SARS-CoV (Fig. 5c, Extended Data Fig. 10f, g). Activities of MMPs have been linked to TGF-β activation and pleural effusions, alveolar damage and neuroinflammation (for example, Kawasaki disease), all of which are characteristic of COVID-19^{23,48–51}.

This drug screen demonstrates the value of our combined dataset, which profiles SARS-CoV-2 infection at multiple levels. We expect that further exploration of these rich data by the scientific community and additional studies of the interplay between different omics levels will substantially advance our molecular understanding of coronavirus biology, including the pathogenicity associated with specific human coronaviruses, such as SARS-CoV-2 and SARS-CoV. Moreover, this resource, together with complementary approaches^{26,52–54}, will streamline the search for antiviral compounds and serve as a base for rational design of combination therapies that target the virus from multiple synergistic angles, thus potentiating the effect of individual drugs while minimizing potential side effects on healthy tissues.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03493-4>.

- Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
- Gordon, D. E. et al. Comparative host–coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eabe9403 (2020).
- Bojkova, D. et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).
- Messner, C. B. et al. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst.* **11**, 11–24.e4 (2020).
- Samavarchi-Tehrani, P. et al. A SARS-CoV-2–host proximity interactome. Preprint at <https://doi.org/10.1101/2020.09.03.282103> (2020).
- Laurent, E. M. N. et al. Global BioID-based SARS-CoV-2 proteins proximal interactome unveils novel ties between viral polypeptides and host factors involved in multiple COVID-19-associated mechanisms. Preprint at <https://doi.org/10.1101/2020.08.28.272955> (2020).
- Klann, K. et al. Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication. *Mol. Cell* **80**, 164–174.e4 (2020).
- Huang, J. et al. SARS-CoV-2 infection of pluripotent stem cell-derived human lung alveolar type 2 cells elicits a rapid epithelial-intrinsic inflammatory response. *Cell Stem Cell* **27**, 962–973.e7 (2020).
- Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).
- Li, J. et al. Virus–host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med* **2**, 99–112.e7 (2021).
- von Brunn, A. et al. Analysis of intraviral protein–protein interactions of the SARS coronavirus ORFome. *PLoS ONE* **2**, e459 (2007).

Article

12. Cornillez-Ty, C. T., Liao, L., Yates, J. R., III, Kuhn, P. & Buchmeier, M. J. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J. Virol.* **83**, 10314–10318 (2009).
13. Andrianifahanana, M. et al. ERBB receptor activation is required for profibrotic responses to transforming growth factor β . *Cancer Res.* **70**, 7421–7430 (2010).
14. Pittet, J.-F. et al. TGF- β is a critical mediator of acute lung injury. *J. Clin. Invest.* **107**, 1537–1544 (2001).
15. George, P. M., Wells, A. U. & Jenkins, R. G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir. Med.* **8**, 807–815 (2020).
16. Mo, X. et al. Abnormal pulmonary function in COVID-19 patients at time of hospital discharge. *Eur. Respir. J.* **55**, 2001217 (2020).
17. Heo, J.-M. et al. Integrated proteogenetic analysis reveals the landscape of a mitochondrial-autophagosome synapse during PARK2-dependent mitophagy. *Sci. Adv.* **5**, eaay4624 (2019).
18. Shi, C.-S. et al. SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. *J. Immunol.* **193**, 3080–3089 (2014).
19. Hoagland, D. A. et al. Modulating the transcriptional landscape of SARS-CoV-2 as an effective method for developing antiviral compounds. Preprint at <https://doi.org/10.1101/2020.07.12.199687> (2020).
20. Castiglione, V., Chiriaco, M., Ermdin, M., Taddei, S. & Vergaro, G. Statin therapy in COVID-19 infection. *Eur. Heart J. Cardiovasc. Pharmacother.* **6**, 258–259 (2020).
21. Radenkovic, D., Chawla, S., Pirro, M., Sahebkar, A. & Banach, M. Cholesterol in relation to COVID-19: should we care about it? *J. Clin. Med.* **9**, 1909 (2020).
22. Chu, H. et al. Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an ex vivo study with implications for the pathogenesis of COVID-19. *Clin. Infect. Dis.* **71**, 1400–1409 (2020).
23. Zhu, Z. et al. From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* **21**, 224 (2020).
24. Cazzaniga, A., Locatelli, L., Castiglioni, S. & Maier, J. The contribution of EDF1 to PPAR γ transcriptional activation in VEGF-treated human endothelial cells. *Int. J. Mol. Sci.* **19**, 1830 (2018).
25. Gavrilaki, E. et al. Endothelial dysfunction in COVID-19: lessons learned from coronaviruses. *Curr. Hypertens. Rep.* **22**, 63 (2020).
26. Danilowski, Z. et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105 (2021).
27. Shinde, S. R. & Maddika, S. PTEN modulates EGFR late endocytic trafficking and degradation by dephosphorylating Rab7. *Nat. Commun.* **7**, 10689 (2016).
28. Wang, D. et al. Auto-phosphorylation represses protein kinase R activity. *Sci. Rep.* **7**, 44340 (2017).
29. Yu, Y. T.-C. et al. Surface vimentin is critical for the cell entry of SARS-CoV. *J. Biomed. Sci.* **23**, 14 (2016).
30. dos Santos, G. et al. Vimentin regulates activation of the NLRP3 inflammasome. *Nat. Commun.* **6**, 6574 (2015).
31. Ramos, I., Stamatakis, K., Oeste, C. L. & Pérez-Sala, D. Vimentin as a multifaceted player and potential therapeutic target in viral infections. *Int. J. Mol. Sci.* **21**, 4675 (2020).
32. Zinzula, L. et al. High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. *Biochem. Biophys. Res. Commun.* **538**, 54–62 (2021).
33. Chen, C.-Y. et al. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **368**, 1075–1086 (2007).
34. Perry, J. S. A. et al. Interpreting an apoptotic corpse as anti-inflammatory involves a chloride sensing pathway. *Nat. Cell Biol.* **21**, 1532–1543 (2019).
35. Wu, G., Dawson, E., Duong, A., Haw, R. & Stein, L. ReactomeFViz: a Cytoscape app for pathway and network-based data analysis. *F1000Res.* **3**, 146 (2014).
36. Reyna, M. A., Leiserson, M. D. M. & Raphael, B. J. Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* **34**, i972–i980 (2018).
37. Cottam, E. M., Whelband, M. C. & Wileman, T. Coronavirus NSP6 restricts autophagosome expansion. *Autophagy* **10**, 1426–1441 (2014).
38. Ohsaki, Y., Cheng, J., Fujita, A., Tokumoto, T. & Fujimoto, T. Cytoplasmic lipid droplets are sites of convergence of proteasomal and autophagic degradation of apolipoprotein B. *Mol. Biol. Cell* **17**, 2674–2683 (2006).
39. Khalil, M. F., Wagner, W. D. & Goldberg, I. J. Molecular interactions leading to lipoprotein retention and the initiation of atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **24**, 2211–2218 (2004).
40. Nicolai, L. et al. Immunothrombotic dysregulation in COVID-19 pneumonia is associated with respiratory failure and coagulopathy. *Circulation* **142**, 1176–1189 (2020).
41. Zavadi, J. et al. Genetic programs of epithelial cell plasticity directed by transforming growth factor- β . *Proc. Natl Acad. Sci. USA* **98**, 6686–6691 (2001).
42. Qin, Z., Xia, W., Fisher, G. J., Voorhees, J. J. & Quan, T. YAP/TAZ regulates TGF- β /Smad3 signaling by induction of Smad7 via AP-1 in human skin dermal fibroblasts. *Cell Commun. Signal.* **16**, 18 (2018).
43. Thi Nhu Thao, T. et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).
44. Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S. & Huang, C. Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Res.* **179**, 104811 (2020).
45. Seebacher, N. A., Stacy, A. E., Porter, G. M. & Merlot, A. M. Clinical development of targeted and immune based anti-cancer therapies. *J. Exp. Clin. Cancer Res.* **38**, 156 (2019).
46. O'Shea, J. J., Kontzias, A., Yamaoka, K., Tanaka, Y. & Laurence, A. Janus kinase inhibitors in autoimmune diseases. *Ann. Rheum. Dis.* **72** (Suppl 2), ii111–ii115 (2013).
47. Yong, H.-Y., Koh, M.-S. & Moon, A. The p38 MAPK inhibitors for the treatment of inflammatory diseases and cancer. *Expert Opin. Investig. Drugs* **18**, 1893–1905 (2009).
48. Hsieh, W.-Y. et al. ACE/ACE2 ratio and MMP-9 activity as potential biomarkers in tuberculous pleural effusions. *Int. J. Biol. Sci.* **8**, 1197–1205 (2012).
49. Ueland, T. et al. Distinct and early increase in circulating MMP-9 in COVID-19 patients with respiratory failure. *J. Infect.* **81**, e41–e43 (2020).
50. Villalta, P. C., Rocic, P. & Townsley, M. I. Role of MMP2 and MMP9 in TRPV4-induced lung injury. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **307**, L652–L659 (2014).
51. Marten, N. W. & Zhou, J. in *Experimental Models of Multiple Sclerosis* (eds Lavi, E. & Constantinescu, C. S.) 839–848 (Springer, 2005).
52. Zhang, J.-Y. et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat. Immunol.* **21**, 1107–1118 (2020).
53. Ellinghaus, D. et al. Genome-wide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
54. Ali, A. & Vijayan, R. Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Sci. Rep.* **10**, 14214 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and reagents

HEK293T, A549, Vero E6 and HEK293-R1 cells were authenticated by PCR-single-locus-technology and their respective culturing conditions were described previously⁵⁵. All cell lines were tested to be mycoplasma free. Expression constructs for C-terminal HA-tagged viral open reading frames were synthesized (Twist Bioscience and BioCat) and cloned into pWPI vector as described previously⁵⁶ with the following modifications: a starting ATG codon was added, internal canonical splicing sites were replaced with synonymous mutations and a C-terminal HA-tag, followed by an amber stop codon, was added to individual viral open reading frames. A C-terminally HA-tagged ACE2 sequence was amplified from an ACE2 expression vector (provided by S. Pöhlmann)⁵⁷ into the lentiviral vector pWPI-puro. A549 cells were transduced twice, and A549-ACE2 cells were selected with puromycin. Lentivirus production, transduction of cells and antibiotic selection were performed as described previously⁵². RNA isolation (Macherey–Nagel NucleoSpin RNA plus), reverse transcription (Takara Bio PrimeScript RT with gDNA eraser) and quantitative PCR with reverse transcription (RT–qPCR) (Thermo-Fisher Scientific PowerUp SYBR green) were performed as described previously⁵⁸. RNA isolation for next generation sequencing applications was performed according to the manufacturer's protocol (Qiagen RNeasy mini kit, RNase free DNase set). For detection of protein abundance by western blotting, HA–horseradish peroxidase (HRP) (Sigma-Aldrich; H6533; 1:2,500 dilution), ACTB–HRP (Santa Cruz; sc-47778; 1:5,000 dilution), MAP1LC3B (Cell Signaling; 3868; 1:1,000 dilution), MAVS (Cell Signaling; 3993; 1:1,000 dilution), HSPA1A (Cell Signaling; 4873; 1:1,000 dilution), TGF- β (Cell Signaling; 3711; 1:1,000 dilution), phospho-p38 (T180/Y182) (Cell Signaling; 4511; 1:1,000 dilution), p38 (Cell Signaling; 8690; 1:1,000 dilution) and SARS-CoV-2 or SARS-CoV N protein (Sino Biological; 40143-MM05; 1:1,000 dilution) antibodies were used. Secondary antibodies detecting mouse (Cell Signaling; 7076; 1:5,000 dilution; Jackson ImmunoResearch; 115-035-003; 1:5,000 dilution), rat (Invitrogen; 31470; 1:5,000 dilution), and rabbit IgG (Cell Signaling; 7074; 1:5,000 dilution) were coupled to HRP. For AP–MS and affinity purification–western blotting applications, HA beads (Sigma-Aldrich) and Thermo Fisher Scientific and Streptactin II beads (IBA Lifesciences) were used. Imaging of western blots was performed as described⁵⁸. Recombinant human IFN- α used for stimulation of cells in the reporter assay was a gift from P. Stäheli (Institute of Virology, University of Freiburg), recombinant human IFN- γ was purchased from PeproTech, and IVT4 was produced as described before⁵⁹. All compounds tested in the viral inhibitor assay are listed in Supplementary Table 9.

Virus strains, stock preparation, plaque assay and in vitro infection

SARS-CoV-Frankfurt-1, SARS-CoV-2-MUC-IMB-1 and SARS-CoV-2-GFP strains⁴³ were produced by infecting Vero E6 cells cultured in DMEM medium (10% FCS, 100 $\mu\text{g ml}^{-1}$ Streptomycin, 100 IU ml^{-1} penicillin) for 2 days (MOI of 0.01). Viral stock was collected and spun twice (1,000g for 10 min) before storage at -80°C . Titre of viral stock was determined by plaque assay. Confluent monolayers of Vero E6 cells were infected with serial fivefold dilutions of virus supernatants for 1 h at 37°C . The inoculum was removed and replaced with serum-free MEM (Gibco, Life Technologies) containing 0.5% carboxymethylcellulose (Sigma-Aldrich). Two days after infection, cells were fixed for 20 min at room temperature with formaldehyde added directly to the medium to a final concentration of 5%. Fixed cells were washed extensively with PBS before staining with water containing 1% crystal violet and 10% ethanol for 20 min. After rinsing with PBS, the number of plaques was counted and the virus titre was calculated.

A549-ACE2 cells were infected with either SARS-CoV-Frankfurt-1 or SARS-CoV-2-MUC-IMB-1 strains (MOI of 2) for the subsequent experiments. At each time point, the samples were washed once with $1\times$ TBS buffer and collected in sodium dodecyl sulfate (SDC) lysis buffer (100 mM Tris HCl pH 8.5; 4% SDC) for proteome-phosphoproteome-ubiquitinome analysis, sodium dodecyl sulfate (SDS) lysis buffer (62.5 mM Tris HCl pH 6.8; 2% SDS; 10% glycerol; 50 mM DTT; 0.01% bromophenol blue) for western blot, or buffer RLT (Qiagen) for transcriptome analysis. The samples were heat-inactivated and frozen at -80°C until further processing.

Affinity purification and mass spectrometric analyses of SARS-CoV-2, SARS-CoV, HCoV-229E and HCoV-NL63 proteins expressed in A549 cells

To determine the interactomes of SARS-CoV-2 and SARS-CoV and the interactomes of an accessory protein (encoded by ORF4 or ORF4a of HCoV-229E or ORF3 of HCoV-NL63) that presumably represents a homologue of the ORF3 and ORF3a proteins of SARS-CoV-2 and SARS-CoV, respectively, four replicate affinity purifications were performed for each HA-tagged viral protein. A549 cells (6×10^6 cells per 15-cm dish) were transduced with lentiviral vectors encoding HA-tagged SARS-CoV-2, SARS-CoV or HCoV-229E/NL63 proteins and protein lysates were prepared from cells collected 3 days after transduction. Cell pellets from two 15-cm dishes were lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1.5 mM MgCl_2 , 0.2% (v/v) NP-40, 5% (v/v) glycerol, cOmplete protease inhibitor cocktail (Roche), 0.5% (v/v) 750 U/ μl Sm DNase) and sonicated (5 min, 4°C , 30 s on, 30 s off, low settings; Bioruptor, Diagenode). Following normalization of protein concentrations of cleared lysates, virus protein-bound host proteins were enriched by adding 50 μl anti-HA-agarose slurry (Sigma-Aldrich, A2095) with constant agitation for 3 h at 4°C . Non-specifically bound proteins were removed by four subsequent washes with lysis buffer followed by three detergent-removal steps with washing buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1.5 mM MgCl_2 , 5% (v/v) glycerol). Enriched proteins were denatured, reduced, alkylated and digested by addition of 200 μl digestion buffer (0.6 M guanidinium chloride, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 4 mM chloroacetamide (CAA), 100 mM Tris-HCl pH 8, 0.5 μg LysC (WAKO Chemicals) and 0.5 μg trypsin (Promega) at 30°C overnight. Peptide purification on StageTips with three layers of C18 Empore filter discs (3M) and subsequent mass spectrometry analysis was performed as described previously^{55,56}. In brief, purified peptides were loaded onto a 20-cm reverse-phase analytical column (75 μm diameter; ReproSil-Pur C18-AQ 1.9 μm resin; Dr Maisch) and separated using an EASY-nLC 1200 system (Thermo Fisher Scientific). A binary buffer system consisting of buffer A (0.1% formic acid (FA) in H_2O) and buffer B (80% acetonitrile (ACN), 0.1% FA in H_2O) with a 90-min gradient (5–30% buffer B (65 min), 30–95% buffer B (10 min), wash out at 95% buffer B (5 min), decreased to 5% buffer B (5 min), and 5% buffer B (5 min)) was used at a flow rate of 300 nl per min. Eluting peptides were directly analysed on a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). Data-dependent acquisition included repeating cycles of one MS1 full scan (300–1650 m/z , $R = 60,000$ at 200 m/z) at an ion target of 3×10^6 , followed by 15 MS2 scans of the highest abundant isolated and higher-energy collisional dissociation (HCD) fragmented peptide precursors ($R = 15,000$ at 200 m/z). For MS2 scans, collection of isolated peptide precursors was limited by an ion target of 1×10^5 and a maximum injection time of 25 ms. Isolation and fragmentation of the same peptide precursor was eliminated by dynamic exclusion for 20 s. The isolation window of the quadrupole was set to 1.4 m/z and HCD was set to a normalized collision energy of 27%.

Proteome analyses of cells expressing SARS-CoV-2, SARS-CoV, HCoV-229E or HCoV-NL63 proteins

For the determination of proteome changes in A549 cells expressing SARS-CoV-2, SARS-CoV, HCoV-229E or HCoV-NL63 proteins, a fraction of

Article

1×10^6 lentivirus-transduced cells from the affinity purification samples were lysed in guanidinium chloride buffer (6 M guanidinium chloride, 10 mM TCEP, 40 mM CAA, 100 mM Tris-HCl pH 8), boiled at 95 °C for 8 min and sonicated (10 min, 4 °C, 30 s on, 30 s off, high settings). Protein concentrations of cleared lysates were normalized to 50 μ g, and proteins were pre-digested with 1 μ g LysC at 37 °C for 1 h followed by a 1:10 dilution (100 mM Tris-HCl pH 8) and overnight digestion with 1 μ g trypsin at 30 °C. Peptide purification on StageTips with three layers of C18 Empore filter discs (3M) and subsequent mass spectrometry analysis was performed as described previously^{55,56}. In brief, 300 ng of purified peptides were loaded onto a 50-cm reversed-phase column (75 μ m inner diameter, packed in house with ReproSil-Pur C18-AQ1.9 μ m resin (Dr Maisch)). The column temperature was maintained at 60 °C using a homemade column oven. A binary buffer system, consisting of buffer A (0.1% FA) and buffer B (80% ACN, 0.1% FA), was used for peptide separation, at a flow rate of 300 nl min⁻¹. An EASY-nLC 1200 system (Thermo Fisher Scientific), directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source, was employed for nano-flow liquid chromatography. Peptides were eluted by a linear 80 min gradient from 5% to 30% buffer B (0.1% v/v FA, 80% v/v ACN), followed by a 4 min increase to 60% B, a further 4 min increase to 95% B, a 4 min plateau phase at 95% B, a 4 min decrease to 5% B and a 4 min wash phase of 5% B. To acquire MS data, the data-independent acquisition (DIA) scan mode operated by the XCalibur software (Thermo Fisher) was used. DIA was performed with one full MS event followed by 33 MS/MS windows in one cycle resulting in a cycle time of 2.7 s. The full MS settings included an ion target value of 3×10^6 charges in the 300–1,650 m/z range with a maximum injection time of 60 ms and a resolution of 120,000 at m/z 200. DIA precursor windows ranged from 300.5 m/z (lower boundary of first window) to 1,649.5 m/z (upper boundary of 33rd window). MS/MS settings included an ion target value of 3×10^6 charges for the precursor window with an Xcalibur-automated maximum injection time and a resolution of 30,000 at m/z 200.

To generate the proteome library for DIA measurements purified peptides from the first and the fourth replicates of all samples were pooled separately and 25 μ g of peptides from each pool were fractionated into 24 fractions by high pH reversed-phase chromatography as described earlier⁶⁰. During each separation, fractions were concatenated automatically by shifting the collection tube every 120 s. In total 48 fractions were dried in a vacuum centrifuge, resuspended in buffer A* (0.2% trifluoroacetic acid (TFA), 2% ACN) and subsequently analysed by a top-12 data-dependent acquisition (DDA) scan mode using the same LC gradient and settings. The mass spectrometer was operated by the XCalibur software (Thermo Fisher). DDA scan settings on full MS level included an ion target value of 3×10^6 charges in the 300–1,650 m/z range with a maximum injection time of 20 ms and a resolution of 60,000 at m/z 200. At the MS/MS level the target value was 10^5 charges with a maximum injection time of 60 ms and a resolution of 15,000 at m/z 200. For MS/MS events only, precursor ions with 2–5 charges that were not on the 20-s dynamic exclusion list were isolated in a 1.4 m/z window. Fragmentation was performed by higher-energy C-trap dissociation with a normalized collision energy of 27 eV.

Infected time-course proteome–phosphoproteome–diGly proteome sample preparation

Frozen lysates of infected A549-ACE2 cells collected at 6, 12 and 24 hpi (and 36 hpi for the phosphoproteomics study) were thawed on ice, boiled for 5 min at 95 °C and sonicated for 15 min (Branson Sonifierer). Protein concentrations were estimated by tryptophan assay⁶¹. To reduce and alkylate proteins, samples were incubated for 5 min at 45 °C with TCEP (10 mM) and CAA (40 mM). Samples were digested overnight at 37 °C using trypsin (1:100 w/w, enzyme/protein, Sigma-Aldrich) and LysC (1:100 w/w, enzyme/protein, Wako).

For proteome analysis, 10 μ g of peptide material were desalted using SDB-RPS StageTips (Empore)⁶¹. In brief, samples were diluted with 1%

TFA in isopropanol to a final volume of 200 μ l and loaded onto StageTips, subsequently washed with 200 μ l of 1% TFA in isopropanol and 200 μ l 0.2% TFA/ 2% ACN. Peptides were eluted with 75 μ l of 1.25% ammonium hydroxide (NH₄OH) in 80% ACN and dried using a SpeedVac centrifuge (Eppendorf, Concentrator Plus). They were resuspended in buffer A* (0.2% TFA, 2% ACN) before LC–MS/MS analysis. Peptide concentrations were measured optically at 280 nm (Nanodrop 2000, Thermo Scientific) and subsequently equalized using buffer A*. One microgram of peptide was analysed by LC–MS/MS.

The rest of the samples were diluted fourfold with 1% TFA in isopropanol and loaded onto SDB-RPS cartridges (Strata-X-C, 30 mg per 3 ml, Phenomenex), pre-equilibrated with 4 ml 30% MeOH/1% TFA and washed with 4 ml 0.2% TFA. Samples were washed twice with 4 ml 1% TFA in isopropanol, once with 0.2% TFA/2% ACN and eluted twice with 2 ml 1.25% NH₄OH/80% ACN. Eluted peptides were diluted with ddH₂O to a final ACN concentration of 35%, snap frozen and lyophilized.

For phosphopeptide enrichment, lyophilized peptides were resuspended in 105 μ l of equilibration buffer (1% TFA/80% ACN) and the peptide concentration was measured optically at 280 nm (Nanodrop 2000, Thermo Scientific) and subsequently equalized using equilibration buffer. The AssayMAP Bravo robot (Agilent) performed the enrichment for phosphopeptides (150 μ g) by priming AssayMAP cartridges (packed with 5 μ l Fe³⁺-NTA) with 0.1% TFA in 99% ACN followed by equilibration in equilibration buffer and loading of peptides. Enriched phosphopeptides were eluted with 1% ammonium hydroxide, which was evaporated using a Speedvac for 20 min. Dried peptides were resuspended in 6 μ l buffer A* and 5 μ l was analysed by LC–MS/MS.

For diGly peptide enrichment, lyophilized peptides were reconstituted in IAP buffer (50 mM MOPS, pH 7.2, 10 mM Na₂HPO₄, 50 mM NaCl) and the peptide concentration was estimated by tryptophan assay. K- ϵ -GG remnant containing peptides were enriched using the PTMScan Ubiquitin Remnant Motif (K- ϵ -GG) Kit (Cell Signaling Technology). Cross-linking of antibodies to beads and subsequent immunopurification was performed with slight modifications as previously described⁶². In brief, two vials of cross-linked beads were combined and equally split into 16 tubes (~31 μ g of antibody per tube). Equal peptide amounts (600 μ g) were added to cross-linked beads, and the volume was adjusted with IAP buffer to 1 ml. After 1 h of incubation at 4 °C and gentle agitation, beads were washed twice with cold IAP and 5 times with cold ddH₂O. Thereafter, peptides were eluted twice with 50 μ l 0.15% TFA. Eluted peptides were desalted and dried as described for proteome analysis with the difference that 0.2% TFA instead of 1% TFA in isopropanol was used for the first wash. Eluted peptides were resuspended in 9 μ l buffer A* and 4 μ l was subjected to LC–MS/MS analysis.

DIA measurements

Samples were loaded onto a 50-cm reversed-phase column (75 μ m inner diameter, packed in house with ReproSil-Pur C18-AQ1.9 μ m resin (Dr Maisch)). The column temperature was maintained at 60 °C using a homemade column oven. A binary buffer system, consisting of buffer A (0.1% FA) and buffer B (80% ACN plus 0.1% FA) was used for peptide separation, at a flow rate of 300 nl min⁻¹. An EASY-nLC 1200 system (Thermo Fisher Scientific), directly coupled online with the mass spectrometer (Orbitrap Exploris 480, Thermo Fisher Scientific) via a nano-electrospray source, was employed for nano-flow liquid chromatography. The FAIMS device was placed between the nano-electrospray source and the mass spectrometer and was used for measurements of the proteome and the PTM-library samples. Spray voltage was set to 2,650 V, RF level to 40 and heated capillary temperature to 275 °C.

For proteome measurements we used a 100 min gradient starting at 5% buffer B followed by a stepwise increase to 30% in 80 min, 60% in 4 min and 95% in 4 min. The buffer B concentration stayed at 95% for 4 min, decreased to 5% in 4 min and stayed there for 4 min. The mass spectrometer was operated in data-independent mode (DIA) with a full scan range of 350–1,650 m/z at 120,000 resolution at 200 m/z ,

normalized automatic gain control (AGC) target of 300% and a maximum fill time of 28 ms. One full scan was followed by 22 windows with a resolution of 15,000, normalized AGC target of 1,000% and a maximum fill time of 25 ms in profile mode using positive polarity. Precursor ions were fragmented by HCD (NCE 30%). Each of the selected compensation voltage (CV) (-40, -55 and -70 V) was applied to sequential survey scans and MS/MS scans; the MS/MS CV was always paired with the appropriate CV from the corresponding survey scan.

For phosphopeptide samples, 5 μ l were loaded and eluted with a 70-min gradient starting at 3% buffer B followed by a stepwise increase to 19% in 40 min, 41% in 20 min, 90% in 5 min and 95% in 5 min. The mass spectrometer was operated in DIA mode with a full scan range of 300–1,400 m/z at 120,000 resolution at 200 m/z and a maximum fill time of 60 ms. One full scan was followed by 32 windows with a resolution of 30,000. Normalized AGC target and maximum fill time were set to 1,000% and 54 ms, respectively, in profile mode using positive polarity. Precursor ions were fragmented by HCD (NCE stepped 25–27.5–30%). For the library generation, we enriched A549 cell lysates for phosphopeptides and measured them with 7 different CV settings (-30, -40, -50, -60, -70, -80 or -90 V) using the same DIA method. The noted CVs were applied to the FAIMS electrodes throughout the analysis.

For the analysis of K- ϵ -GG peptide samples, half of the samples were loaded. We used a 120-min gradient starting at 3% buffer B followed by a stepwise increase to 7% in 6 min, 20% in 49 min, 36% in 39 min, 45% in 10 min and 95% in 4 min. The buffer B concentration stayed at 95% for 4 min, decreased to 5% in 4 min and stayed there for 4 min. The mass spectrometer was operated in DIA mode with a full scan range of 300–1,350 m/z at 120,000 resolution at m/z 200, normalized AGC target of 300% and a maximum fill time of 20 ms. One full scan was followed by 46 windows with a resolution of 30,000. Normalized AGC target and maximum fill time were set to 1,000% and 54 ms, respectively, in profile mode using positive polarity. Precursor ions were fragmented by HCD (NCE 28%). For K- ϵ -GG peptide library, we mixed the first replicate of each sample and measured them with eight different CV setting (-35, -40, -45, -50, -55, -60, -70 or -80 V) using the same DIA method.

Processing of raw MS data

AP-MS data. Raw MS data files of AP-MS experiments conducted in DDA mode were processed with MaxQuant (version 1.6.14) using the standard settings and label-free quantification (LFQ) enabled (LFQ min ratio count 1, normalization type none, stabilize large LFQ ratios disabled). Spectra were searched against forward and reverse sequences of the reviewed human proteome including isoforms (UniprotKB, release 2019.10) and C-terminally HA-tagged SARS-CoV-2, SARS-CoV and HCoV proteins by the built-in Andromeda search engine⁶³.

In-house Julia scripts (<https://doi.org/10.5281/zenodo.4541090>) were used to define alternative protein groups: only the peptides identified in AP-MS samples were considered for being protein group-specific, protein groups that differed by the single specific peptide or had less than 25% different specific peptides were merged to extend the set of peptides used for protein group quantitation and reduce the number of protein isoform-specific interactions.

Viral protein overexpression and DIA MS data. Spectronaut version 13 (Biognosys) with the default settings was used to generate the proteome libraries from DDA runs by combining files of respective fractionations using the human fasta file (Uniprot, 2019.10, 42 431 entries) and viral bait sequences. Proteome DIA files were analysed using the proteome library with the default settings and disabled cross run normalization.

SARS-CoV-2/SARS-CoV-infected proteome/PTM DIA MS data. Spectronaut version 14 (Biognosys)⁶⁴ was used to generate the libraries and analyse all DIA files using the human fasta file (UniprotKB, release 2019.10) and sequences of SARS-CoV-2/SARS-CoV proteins (Uniprot, release 2020.08). Orf1a polyprotein sequences were split into separate

protein chains according to the cleavage positions specified in the UniProt. For the generation of the PTM-specific libraries, the DIA single CV runs were combined with the actual DIA runs and either phosphorylation at serine, threonine or tyrosine, or GlyGly at lysine, was added as variable modification to default settings. The maximum number of fragment ions per peptide was increased to 25. The proteome DIA files were analysed using direct DIA approach with default settings and disabled cross run normalization. All post-translational modification DIA files were analysed using their respective hybrid library and either phosphorylation at Serine/Threonine/Tyrosine or GlyGly at Lysine was added as an additional variable modification to default settings with LOESS normalization and disabled PTM localization filter.

A collection of in-house Julia scripts (<https://doi.org/10.5281/zenodo.4541090>) were used to process the elution group (EG) -level Spectronaut reports, identify PTMs and assign EG-level measurements to PTMs. The PTM was considered if at least once it was detected with ≥ 0.75 localization probability in EG with q -value $\leq 10^{-3}$. For further analysis of given PTM, only the measurements with ≥ 0.5 localization probability and EG q -value $\leq 10^{-2}$ were used.

Bioinformatic analysis

Unless otherwise specified, the bioinformatic analysis was done in R (version 3.6), Julia (version 1.5) and Python (version 3.8) using a collection of in-house scripts (<https://doi.org/10.5281/zenodo.4541090> and <https://doi.org/10.5281/zenodo.4541082>).

Datasets. The following public datasets were used in the study: Gene Ontology and Reactome annotations (http://download.baderlab.org/EM_Genesets/April_01_2019/Human/UniProt/Human_GO_AllPathways_with_GO_jea_April_01_2019_UniProt.gmt); IntAct Protein Interactions (<https://www.ebi.ac.uk/intact/>, v2019.12); IntAct Protein Complexes (<https://www.ebi.ac.uk/complexportal/home>, v2019.12); CORUM Protein Complexes (<http://mips.helmholtz-muenchen.de/corum/download/allComplexes.xml.zip>, v2018.3); Reactome Functional Interactions (https://reactome.org/download/tools/Reactome-Fls/FlsInGene_020720_with_annotations.txt.zip); Human (v2019.10), Human-CoV, SARS-CoV-2 and SARS-CoV (v2020.08) protein sequences: <https://uniprot.org>.

Statistical analysis of MS data. MaxQuant and Spectronaut output files were imported into R using in-house `maxquantUtils` R package (<https://doi.org/10.5281/zenodo.4536603>). For all MS datasets, the Bayesian linear random effects models were used to define how the abundances of proteins change between the conditions. To specify and fit the models we used the `msglm` R package (<https://doi.org/10.5281/zenodo.4536605>), which uses the `rstan` package (version 2.19)⁶⁵ for inferring the posterior distribution of the model parameters. In all the models, the effects corresponding to the experimental conditions have regularized horseshoe+ priors⁶⁶, whereas the batch effects have normally distributed priors. Laplacian distribution was used to model the instrumental error of MS intensities. For each MS instrument used, the heteroscedastic intensities noise model was calibrated with the technical replicate MS data of the instrument. These data were also used to calibrate the logit-based model of missing MS data (the probability that the MS instrument will fail to identify the protein given its expected abundance in the sample). The model was fit using unnormalized MS intensities data. Instead of transforming the data by normalization, the inferred protein abundances were scaled by the normalization multiplier of each individual MS sample to match the expected MS intensity of that sample. This allows taking the signal-to-noise variation between the samples into account when fitting the model. Due to high computational intensity, the model was applied to each protein group separately. For all the models, 4,000 iterations (2,000 warmup + 2,000 sampling) of the no-U-turn Markov Chain Monte Carlo were performed in 7 or 8 independent chains, every 4th sample was collected

Article

for posterior distribution of the model parameters. For estimating the statistical significance of protein abundance changes between the two experimental conditions, the P -value was defined as the probability that a random sample from the posterior distribution of the first condition would be smaller (or larger) than a random sample drawn from the second condition. No-multiple hypothesis testing corrections were applied, since this is handled by the choice of the model priors.

Statistical analysis of AP-MS data and filtering for specific interactions. The statistical model was applied directly to the MS1 intensities of protein group-specific LC peaks (evidence.txt table of MaxQuant output). In R GLM formula language, the model could be specified as

$$\log(\text{Intensity}) \approx 1 + \text{APMS} + \text{Bait} + \text{Bait} : \text{Virus} + \text{MS1peak} + \text{MSbatch},$$

where the APMS effect models the average shift of intensities in AP-MS data in comparison to full proteome samples, Bait is the average enrichment of a protein in AP-MS experiments of homologous proteins of both SARS-CoV and SARS-CoV-2, and Bait:Virus corresponds to the virus-specific changes in protein enrichment. MS1peak is the log ratio between the intensity of a given peak and the total protein abundance (the peak is defined by its peptide sequence, PTMs and the charge; it is assumed that the peak ratios do not depend on experimental conditions⁶⁷), and MSbatch accounts for batch-specific variations of protein intensity. APMS, Bait and Bait:Virus effects were used to reconstruct the batch effect-free abundance of the protein in AP-MS samples.

The modelling provided the enrichment estimates for each protein in each AP experiment. Specific AP-MS interactions had to pass the two tests. In the first test, the enrichment of the candidate protein in a given bait AP was compared against the background, which was dynamically defined for each interaction to contain the data from all other baits, where the abundance of the candidate was within 50–90% percentile range (excluding top 10% baits from the background allowed the protein to be shared by a few baits in the resulting AP-MS network). The non-targeting control and Gaussian luciferase baits were always preserved in the background. Similarly, to filter out any potential side-effects of very high bait protein expression, the ORF3 homologues were always present in the background of M interactors and vice versa. To rule out the influence of the batch effects, the second test was applied. It was defined similarly to the first one, but the background was constrained to the baits of the same batch, and 40–80% percentile range was used. In both tests, the protein has to be fourfold enriched over the background (16 fold for highly expressed baits: ORF3, M, NSP13, NSP5, NSP6, ORF3a, ORF7b, ORF8b and HCoV-229E ORF4a) with P -value $\leq 10^{-3}$.

Additionally, we excluded the proteins that, in the viral protein expression data, have shown upregulation, and their enrichment in AP-MS data was less than 16 times stronger than observed upregulation effects. Finally, to exclude the carryover of material between the samples sequentially analysed by MS, we removed the putative interactors, which were also enriched at higher levels in the samples of the preceding bait, or the one before it.

For the analysis of interaction specificity between the homologous viral proteins, we estimated the significance of interaction enrichment difference (corrected by the average difference between the enrichment of the shared interactors to adjust for the bait expression variation). Specific interactions have to be fourfold enriched in comparison to the homologue with P -value $\leq 10^{-3}$.

Statistical analysis of DIA proteome effects upon viral protein over-expression. The statistical model of the viral protein overexpression dataset was similar to AP-MS data, except that protein-level intensities provided by Spectronaut were used. The PCA analysis of the protein intensities has identified that the second principal component is associated with the batch-dependent variations between the samples.

To exclude their influence, this principal component was added to the experimental design matrix as an additional batch effect.

As with AP-MS data, the two statistical tests were used to identify the significantly regulated proteins (column 'is_change' in Supplementary Table 3). First, the absolute value of median \log_2 -fold change of the protein abundance upon overexpression of a given viral protein in comparison to the background had to be above 1.0 with P -value $\leq 10^{-3}$. The background was individually defined for each analysed protein. It was composed of experiments, where the abundance of given protein was within the 20–80% percentile range of all measured samples. Second, the protein had to be significantly regulated (same median \log_2 -fold change and P -value thresholds applied) against the batch-specific background (defined similarly to the global background, but using only the samples of the same batch).

An additional stringent criterion was applied to select the most significant changes (column 'is_top_change' in Supplementary Table 3; Extended Data Fig. 1i).

For each protein we classified bait-induced changes as: 'high' when $|\text{median } \log_2 \text{ fold-change}| \geq 1$ and P -value $\leq 10^{-10}$ both in background and batch comparisons; 'medium' if $10^{-10} < P$ -value $\leq 10^{-4}$ with same fold-change requirement; and 'low' if $10^{-4} < P$ -value $\leq 10^{-2}$ with the same fold-change requirement. All other changes were considered non-significant.

We then required that 'shared' top-regulated proteins should have exactly one pair of SARS-CoV-2 and SARS-CoV high- or medium-significant homologous baits among the baits with either up- or down-regulated changes and no other baits with significant changes of the same type.

We further defined 'SARS-CoV-2-specific' or 'SARS-CoV-specific' top-regulated proteins to be the ones with exactly one high-significant change, and no other significant changes of the same sign. For 'specific' hits we additionally required that in the comparison of high-significant bait to its homologue $|\text{median } \log_2 \text{ fold-change}| \geq 1$ and P -value $\leq 10^{-3}$. When the homologous bait was missing (SARS-CoV-2 NSP1, SARS-CoV ORF8a and SARS-CoV ORF8b), we instead required that in the comparison of the high-significant change to the background $|\text{median } \log_2 \text{ fold-change}| \geq 1.5$.

The resulting network of most affected proteins was imported and prepared for publication in Cytoscape v.3.8.1⁶⁸.

Statistical analysis of DIA proteomic data of SARS-CoV-2 and SARS-CoV-infected A549-ACE2 cells. Similarly to the AP-MS DDA data, the linear Bayesian model was applied to the EG-level intensities. To model the protein intensity, the following linear model (in R notation) was used:

$$\log(\text{Intensity}(t)) \sim 1 + \sum_{t_i \leq t} (\text{after}(t_i) + (\text{infection} \pm \text{CoV2}) : \text{after}(t_i)) + \text{EG},$$

where the $\text{after}(t_i)$ effect corresponds to the protein abundance changes in mock-infected samples that happened between t_{i-1} and t_i after infection and it is applied to the modelled intensity at all time points starting from t_i ; $\text{infection} : \text{after}(t_i)$ ($t_i = 6, 12, 24$) is the common effect of SARS-CoV-2 and SARS-CoV infections occurring between t_{i-1} and t_i ; $\text{CoV2} : \text{after}(t_i)$ is the virus-specific effect within t_{i-1} and t_i hpi that is added to the log intensity for SARS-CoV-2-infected samples and subtracted from the intensity for SARS-CoV ones; EG is the elution group-specific shift in the measured log-intensities.

The absolute value of median \log_2 fold change between the conditions above 0.25 and the corresponding unadjusted P -value $\leq 10^{-3}$ were used to define the significant changes at a given time point in comparison to mock infection. We also required that the protein group is quantified in at least two replicates of at least one of the compared conditions. Additionally, if for one of the viruses (for example, SARS-CoV-2)

only the less stringent condition ($|\text{median log}_2 \text{ fold-change}| \geq 0.125$, $P\text{-value} \leq 10^{-3}$) was fulfilled, but the change was significant in the infection of the other virus (SARS-CoV), and the difference between the viruses was not significant, the observed changes were considered significant for both viruses.

Statistical analysis of DIA phosphoproteome and ubiquitinome data of SARS-CoV-2 and SARS-CoV infections. The data from single-double- and triple-modified peptides were analysed separately and, for a given PTM, the most significant result was reported.

The data were analysed with the same Bayesian linear model as proteome SARS-CoV and SARS-CoV-2 infection data. In addition to the intensities normalization, for each replicate sample the scale of the effects in the experimental design matrix was adjusted, so that on average the correlation between log fold changes of the replicates was 1:1. The same logic as for the proteome analysis, was applied to identify significant changes, but the median \log_2 fold change had to be larger than 0.5, or 0.25 for the less stringent test. We additionally required that the PTM peptides are quantified in at least two replicates of at least one of the compared conditions. To ignore the changes in PTM site intensities that are due to proteome-level regulation, we excluded PTM sites on significantly regulated proteins if the directions of protein and PTM site changes were the same and the difference between their median \log_2 fold changes was less than two. Phosphoproteomics data were further analysed with Ingenuity Pathway Analysis software (Qiagen; <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>)

Transcriptomic analysis of SARS-CoV-2 and SARS-CoV infected A549-ACE2 cells. For the analysis of the transcriptome data, Genecode gene annotations v28 and the human reference genome GRCh38 were derived from the Genecode homepage (EMBL-EBI). Viral genomes were derived from GenBank (SARS-CoV-2 - LR824570.1, and SARS-CoV - AY291315.1). Dropseq tool v1.12 was used for mapping raw sequencing data to the reference genome. The resulting UMI filtered count matrix was imported into R v3.4.4. CPM (counts per million) values were calculated for the raw data and genes having a mean cpm value less than 1 were removed from the dataset. A dummy variable combining the covariates infection status (mock, SARS-CoV, SARS-CoV-2) and time point was used for modelling the data within Limma (v3.46.0)⁶⁹.

Data were transformed with the Voom method⁶⁹ followed by quantile normalization. Differential testing was performed between infection states at individual time points by calculating moderated t -statistics and P -values for each host gene. A gene was considered to be significantly regulated if the false discovery rate-adjusted P -value was below 0.05.

Gene set enrichment analysis. We used Gene Ontology, Reactome and other EnrichmentMap gene sets of human proteins (version 2020.10)⁷⁰ as well as protein complexes annotations from IntAct Complex Portal (version 2019.11)⁷¹ and CORUM (version 2019)⁷². PhosphoSitePlus (version 2020.08) was used for known kinase-substrate and regulatory sites annotations, Perseus (version 1.6.14.0)⁷³ was used for annotation of known kinase motifs. For transcription factor enrichment analysis (Extended Data Fig. 2e) the significantly regulated transcripts were submitted to ChEA3 web-based application⁷⁴ and ENCODE data on transcription factor–target gene associations were used⁷⁵.

To find the non-redundant collection of annotations describing the unique and shared features of multiple experiments in a dataset (Fig. 1d, Extended Data Fig. 2l, m), we used in-house Julia package OptEnriched-SetCover.jl (<https://doi.org/10.5281/zenodo.4536596>), which employs evolutionary multi-objective optimization technique to find a collection of annotation terms that have both significant enrichments in the individual experiments and minimal pairwise overlaps.

The resulting set of terms was further filtered by requiring that the annotation term has to be significant with the specified unadjusted

Fisher's exact test P -value cut-off in at least one of the experiments or comparisons (the specific cut-off value is indicated in the figure legend of the corresponding enrichment analysis).

The generation of diagonally-split heat maps was done with the VegaLite.jl package (<https://github.com/queryverse/VegaLite.jl>).

Viral PTMs alignment. For matching the PTMs of SARS-CoV-2 and SARS-CoV the protein sequences were aligned using the BioAlignments.jl Julia package (v.2.0; <https://github.com/BioJulia/BioAlignments.jl>) with the Needleman–Wunsch algorithm using BLOSUM80 substitution matrix, and applying -5 and -3 penalties for the gap and extension, respectively.

For the cellular proteins, we required that the viral phosphorylation or ubiquitination site is observed with q -value $\leq 10^{-3}$ and localization probability ≥ 0.75 . For the PTMs with lower confidence (q -value $\leq 10^{-2}$ and localization probability ≥ 0.5) we required that the same site is observed with high confidence at the matching position of the orthologous protein of the other virus.

Network diffusion analysis. To systematically detect functional interactions, which may connect the cellular targets of each viral protein (interactome dataset) with the downstream changes it induces on proteome level (effectome dataset), we have used the network diffusion-based HierarchicalHotNet method³⁶ as implemented in Julia package HierarchicalHotNet.jl (<https://doi.org/10.5281/zenodo.4536590>). Specifically, for network diffusion with restart, we used the ReactomeFI network (version 2019)³⁵ of cellular functional interactions, reversing the direction of functional interaction (for example, replacing kinase→substrate interaction with substrate→kinase). The proteins with significant abundance changes upon bait over-expression ($|\text{median}(\log_2 \text{ fold change})| \geq 0.25$, $P \leq 10^{-2}$ both in the comparison against the controls and against the baits of the same batch) were used as the sources of signal diffusion with weights set to $w_i = \sqrt{|\text{median}(\log_2 \text{ fold change})| \cdot |\log_{10} P\text{-value}|}$, otherwise the node weight was set to zero. The weight of the edge $g_i \rightarrow g_j$ was set to $w_{ij} = 1 + w_j$. The restart probability was set to 0.4, as suggested in the original publication, so that the probability of the random walk to stay in the direct neighbourhood of the node is the same as the probability to visit more distant nodes. To find the optimal cutting threshold of the resulting hierarchical tree of strongly connected components (SCCs) of the weighted graph corresponding to the stationary distribution of signal diffusion and to confirm the relevance of predicted functional connections, the same procedure was applied to 1,000 random permutations of vertex weights as described in Reyna et al.³⁶ (vertex weights are randomly shuffled between the vertices with similar in and out degrees). Since cutting the tree of SCCs at any threshold t (keeping only the edges with weights above t) and collapsing each resulting SCC into a single node produces the directed acyclic graph of connections between SCCs, it allowed efficient enumeration of the paths from the 'source' nodes (proteins strongly perturbed by viral protein expression with vertex weight w , $w \geq 1.5$) to the 'sink' nodes (interactors of the viral protein). At each threshold t , the average inverse of the path length from source to sink nodes was calculated as:

$$L_{\text{avg}}^{-1}(t) = \frac{1}{N_{\text{source}} \cdot N_{\text{sink}}} \sum_p L_{\text{SCC}}^{-1}(p),$$

where N_{source} is the number of sources, N_{sink} is the number of sinks, $L_{\text{SCC}}(p)$ is the number of SCCs that the given path p from source to sink goes through, and the sum is for all paths from sources to sinks. The metric changes from 1 (all sources and sinks in the same SCC) to 0 (no or infinitely long paths between sources and sinks). For the generation of the diffusion networks we were using the t_{opt} threshold that maximized the difference between $L_{\text{avg}}^{-1}(t)$ for the real data and the third quartile of $L_{\text{avg}}^{-1}(t)$ for randomly shuffled data.

Article

In the generated SCC networks, the direction of the edges was reverted back, and the results were exported as GraphML files using in-house Julia scripts (<https://doi.org/10.5281/zenodo.4541090>). The catalogue of the networks for each viral bait is available as Supplementary Data 1.

To assess the significance of edges in the resulting network, we calculated the P -value of the edge $g_i \rightarrow g_j$ as the probability that the transition probability between the given pair of genes based on permuted data is higher than the transition probability based on the real data:

$$P(w_{\text{real}}(g_i, g_j) \leq w_{\text{perm}}(g_i, g_j)).$$

This P -value was stored as the 'prob_perm_walkweight_greater' edge attribute of GraphML output. The specific subnetworks predicted by the network diffusion (Fig. 4b–d) were filtered for edges with $P \leq 0.05$.

When the $g_i \rightarrow g_j$ connection was not present in the ReactomeFl network, to recover the potential short pathways connecting g_i and g_j , ReactomeFl was searched for intermediate g_k nodes, such that the edges $g_i \rightarrow g_k$ and $g_k \rightarrow g_j$ are present in ReactomeFl. The list of these short pathways is provided as the 'flowpaths' edge attribute in GraphML output.

The GraphML output of network diffusion was prepared for publication using yEd (v.3.20; <https://www.yworks.com>).

Intersection with other SARS coronavirus datasets. The intersection between the data generated by this study and other publicly available datasets was done using the information from respective supplementary tables. When multiple viruses were used in a study, only the comparisons with SARS-CoV and SARS-CoV-2 were included. For time-resolved data, all time points up to 24 hpi were considered. The dataset coverage was defined as the number of reported distinct protein groups for proteomic studies and genes for transcriptomic studies. Confident interactions or significant regulations were filtered according to the criteria specified in the original study. A hit was considered as 'confirmed' when it was significant both in the present study and the external data and showed the same trend.

qRT-PCR analysis

RNA isolation from SARS-CoV and SARS-CoV-2 infected A549-ACE2 cells was performed as described above (Qiagen). Five hundred nanograms total RNA was used for reverse transcription with PrimeScript RT with gDNA eraser (Takara). For relative transcript quantification PowerUp SYBR Green (Applied Biosystems) was used. Primer sequences can be provided upon request.

Co-immunoprecipitation and western blot analysis

HEK293T cells were transfected with pWPI plasmid encoding single HA-tagged viral proteins, alone or together with pTO-SII-HA expressing host factor of interest. 48 h after transfection, cells were washed in PBS, flash frozen in liquid nitrogen and kept at -80°C until further processing. Co-immunoprecipitation experiments were performed as described previously^{55,56}. In brief, cells were lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1.5 mM MgCl_2 , 0.2% (v/v) NP-40, 5% (v/v) glycerol, cComplete protease inhibitor cocktail (Roche), 0.5% (v/v) 750 $\text{U } \mu\text{l}^{-1}$ Sm DNase) and sonicated (5 min, 4°C , 30 s on, 30 s off, low settings; Bioruptor, Diagenode SA). HA or Streptactin beads were added to cleared lysates and samples were incubated for 3 h at 4°C under constant rotation. Beads were washed six times in the lysis buffer and resuspended in $1\times$ SDS sample buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 50 mM DTT, 0.01% bromophenol blue). After boiling for 5 min at 95°C , a fraction of the input lysate and elution were loaded on NuPAGE Novex 4–12% Bis-Tris (Invitrogen), and further submitted to western blotting using Amersham Protran nitrocellulose membranes. Imaging was performed by HRP luminescence (ECL, Perkin Elmer).

SARS-CoV-2 infected A549-ACE2 cell lysates were sonicated (10 min, 4°C , 30 s on, 30 s off, low settings; Bioruptor, Diagenode). Protein

concentration was adjusted based on Pierce660 assay supplemented with ionic detergent compatibility reagent. After boiling for 5 min at 95°C and brief centrifugation at maximum speed, the samples were loaded on NuPAGE Novex 4–12% Bis-Tris (Invitrogen), and blotted onto 0.22 μm Amersham Protran nitrocellulose membranes (Merck). Primary and secondary antibody stainings were performed according to the manufacturer's recommendations. Imaging was performed by HRP luminescence using Femto kit (ThermoFischer Scientific) or Western Lightning PlusECL kit (Perkin Elmer).

Mapping of post-translational modification sites on the NCTD structure

NCTD dimers of SARS-CoV-2 (PDB: 6YUN) and SARS-CoV (PDB: 2CJR) were superimposed by aligning the α -carbons backbone over 111 residues (from position 253/254 to position 364/365 following SARS-CoV-2/SARS-CoV numbering) by using the tool MatchMaker⁷⁶ as implemented in the Chimera software⁷⁷. Ubiquitination sites were visually inspected and mapped by using the PyMOL software (<https://pymol.org>). Phosphorylation on Ser310/311 was simulated in silico by using the PyTMs plugin as implemented in PyMOL⁷⁸. Inter-chain residue contacts, dimer interface area, free energy and complex stability were comparatively analysed between non-phosphorylated and phosphorylated SARS-CoV-2 and SARS-CoV NCTD by using the PDBEPIA server⁷⁹. Poisson–Boltzmann electrostatic surface potential of native and post-translationally modified NCTD was calculated by using the PBEQ Solver tool on the CHARMM-GUI server by preserving existing hydrogen bonds⁸⁰. Molecular graphics depictions were produced with the PyMOL software.

Reporter assay and IFN bioassay

The following reporter constructs were used in this study: pSRE-luc was purchased from Stratagene, EF1- α -ren was obtained from E. Gürlevik (Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Germany), pCAGGS-Flag-RIG-I was obtained from C. Basler (Department of Microbiology, Mount Sinai School of Medicine, USA), pIRF1-GAS-ff-luc, pWPI-SMN1-flag and pWPI-NSS (ZIKV)-HA was described previously^{56,81}.

For the reporter assay, HEK293-R1 cells were plated in 24-well plates 24 h before transfection. Firefly reporter and *Renilla* transfection control were transfected together with plasmids expressing viral proteins using polyethylenimine (PEI, Polysciences) for untreated and treated conditions. In 18 h cells were stimulated for 8 h with a corresponding inducer and collected in the passive lysis buffer (Promega). Luminescence of Firefly and *Renilla* luciferases was measured using dual-luciferase-reporter assay (Promega) according to the manufacturer's instructions in a microplate reader (Tecan).

Total amounts of IFN- α and IFN- β in cell supernatants were measured by using 293T cells stably expressing the firefly luciferase gene under the control of the mouse Mx1 promoter (Mx1-luc reporter cells)⁸². In brief, HEK293-R1 cells were seeded, transfected with pCAGGS-flag-RIG-I plus viral protein constructs and stimulated as described above. Cell supernatants were collected in 8 h. Mx1-luc reporter cells were seeded into 96-well plates in triplicates and were treated 24 h later with supernatants. At 16 h after incubation, cells were lysed in the passive lysis buffer (Promega), and luminescence was measured with a microplate reader (Tecan). The assay sensitivity was determined by a standard curve.

Viral inhibitor assay

A549-ACE2 cells were seeded into 96-well plates in DMEM medium (10% FCS, 100 $\mu\text{g } \text{ml}^{-1}$ streptomycin, 100 $\text{IU } \text{ml}^{-1}$ penicillin) one day before infection. Six hours before infection, or at the time of infection, the medium was replaced with 100 μl of DMEM medium containing either the compounds of interest or DMSO as a control. Infection was performed by adding 10 μl of SARS-CoV-2-GFP (MOI of 3) per well and plates were placed in the IncuCyte S3 Live-Cell Analysis System (Essen

Bioscience), where whole well real-time images of mock (phase channel) and infected (GFP and phase channel) cells were captured every 4 h for 48 h. Cell viability (mock) and virus growth (mock and infected) were assessed as the cell confluence per well (phase area) and GFP area normalized by cell confluence per well (GFP area/phase area) respectively using InCuCyte S3 Software (Essen Bioscience; version 2019B rev2).

For comparative analysis of antiviral treatment activity against SARS-CoV and SARS-CoV-2, A549-ACE2 cells were seeded in 24-well plates, as previously described. Treatment was performed for 6 h with 0.5 ml of DMEM medium containing either the compounds of interest or DMSO as a control, and infected with SARS-CoV-Frankfurt-1 or SARS-CoV-2-MUC-IMB-1 (MOI of 1) for 24 h. Total cellular RNA was collected and analysed by RT-qPCR, as previously described.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The raw sequencing data for this study have been deposited with the ENA at EMBL-EBI under accession number PRJEB38744. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁸³ partner repository with the dataset identifiers PXD022282, PXD020461 and PXD020222. Protein interactions identified in this study have been submitted to the IMEx (<https://www.imexconsortium.org>) consortium through IntAct⁸⁴ with the identifier IM-28109. The data and analysis results are accessible online via the interactive web interface at <https://covinet.innatelab.org>.

Code availability

In-house R and Julia packages and scripts used for the bioinformatics analysis of the data have been deposited to public GitHub repositories: <https://doi.org/10.5281/zenodo.4536605>, <https://doi.org/10.5281/zenodo.4536603>, <https://doi.org/10.5281/zenodo.4536590>, <https://doi.org/10.5281/zenodo.4536596>, <https://doi.org/10.5281/zenodo.4541090> and <https://doi.org/10.5281/zenodo.4541082>.

55. Hubel, P. et al. A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape. *Nat. Immunol.* **20**, 493–502 (2019).
56. Scaturro, P. et al. An orthogonal proteomic survey uncovers novel Zika virus host factors. *Nature* **561**, 253–257 (2018).
57. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
58. Gebhardt, A. et al. The alternative cap-binding complex is required for antiviral defense in vivo. *PLoS Pathog.* **15**, e1008155 (2019).
59. Goldeck, M., Schlee, M., Hartmann, G. & Hornung, V. Enzymatic synthesis and purification of a defined RIG-I ligand. *Methods Mol. Biol.* **1169**, 15–25 (2014).
60. Kulak, N. A., Geyer, P. E. & Mann, M. Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol. Cell. Proteomics* **16**, 694–705 (2017).
61. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
62. Udeshi, N. D., Mertins, P., Svinikina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* **8**, 1950–1960 (2013).
63. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
64. Bober, M. & Miladinovic, S. General guidelines for validation of decoy models for HRM/DIA/SWATH as exemplified using Spectronaut. *F1000posters* <https://f1000research.com/posters/1097512> (2015).
65. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).
66. Bhadra, A., Datta, J., Polson, N. G. & Willard, B. The Horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal.* **12**, 1105–1131 (2017).
67. Goeminne, L. J. E., Gevaert, K. & Clement, L. Peptide-level robust ridge regression improves estimation, sensitivity, and specificity in data-dependent quantitative label-free shotgun proteomics. *Mol. Cell. Proteomics* **15**, 657–668 (2016).
68. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
69. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
70. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
71. Meldal, B. H. M. et al. Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.* **47** (D1), D550–D558 (2019).
72. Giurgiu, M. et al. CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **47** (D1), D559–D563 (2019).
73. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
74. Keenan, A. B. et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* **47** (W1), W212–W224 (2019).
75. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
76. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**, 339 (2006).
77. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
78. Warnecke, A., Sandalova, T., Achour, A. & Harris, R. A. PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. *BMC Bioinformatics* **15**, 370 (2014).
79. Paxman, J. J. & Heras, B. Bioinformatics tools and resources for analyzing protein structures. *Methods Mol. Biol.* **1549**, 209–220 (2017).
80. Jo, S., Vargyas, M., Vasko-Szedlar, J., Roux, B. & Im, W. PBEG-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.* **36**, W270–W275 (2008).
81. Vogt, C. et al. The interferon antagonist ML protein of thogoto virus targets general transcription factor IIB. *J. Virol.* **82**, 11446–11453 (2008).
82. Jorns, C. et al. Rapid and simple detection of IFN-neutralizing antibodies in chronic hepatitis C non-responsive to IFN-alpha. *J. Med. Virol.* **78**, 74–82 (2006).
83. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47** (D1), D442–D450 (2019).
84. Orchard, S. et al. The MintAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2014).

Acknowledgements We thank S. Pöhlmann for sharing ACE2 plasmids and R. Baier for technical assistance, M. Trojan for advice on drugs, J. Pancorbo and J. Albert-von der Gönna from the Leibniz Supercomputing Centre (<https://www.lrz.de>) for technical assistance. We further thank the Karl Max von Bauernfeind-Verein for support for the screening microscope. Work in the authors' laboratories was supported by an ERC consolidator grant (ERC-CoG ProDAP, 817798), the German Research Foundation (PI 1084/3, PI 1084/4, PI 1084/5, TRR179/TP11, TRR237/A07), the Bavarian State Ministry of Science and Arts (Bavarian Research Network FOR-COVID) and the German Federal Ministry of Education and Research (COVINET) to A. Pichlmaier. This work was also supported by the German Federal Ministry of Education and Research (CLINSPECT-M) to B.K. A.W. was supported by the China Scholarship Council (CSC). The work of J.Z. was supported by the German Research Foundation (SFB1021, A01 and B01; KFO309, P3), the State of Hessen through the LOEWE Program (DRUID, B02) and the German Ministry for Education and Research (COVINET, RAPID). The work of P.S. is supported by the Free and Hanseatic City of Hamburg and the German Research Foundation (SC 314/1-1). We thank R. Polakiewicz, F. Gnad and Cell Signaling Technology for the gift of the PTMScan Ubiquitin Remnant Motif (K-e-GG) Kits.

Author contributions Conceptualization: A.S., V. Girault, V. Grass, O.K., V.B., C.U., D.A.H., Y.H., J.Z., P.S., M.M. and A. Pichlmaier. Investigation: V. Girault, V. Grass, O.K., V.B., C.U., D.A.H., Y.H., L.O., A.W., A. Piras, F.M.H., M.C.T., I.P., T.M.L., R.E., J.J. and P.S. Data analysis: A.S., V. Girault, V. Grass, V.B., O.K., C.U., D.A.H., Y.H., M.S.H., F.M.H., M.C.T., L.Z., T.E. and M.R. Funding acquisition: R.W., B.K., U.P., R.R., J.Z., V.T., M.M. and A. Pichlmaier. Supervision: M.M., R.R. and A. Pichlmaier. Writing: A.S., V. Girault, V. Grass, O.K., V.B., C.U., D.A.H., Y.H., L.Z., M.M. and A. Pichlmaier.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03493-4>.

Correspondence and requests for materials should be addressed to A.P.

Peer review information Nature thanks Ivan Dikic, Trey Ideker, Michael Weekes, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

3.3 A regulatory region on RIPK2 is required for XIAP binding and NOD signaling activity

Valentin J Heim^{1,2}, Laura F Dagley^{1,2}, Che A Stafford³, **Fynn M Hansen**⁴, Elise Clayer^{1,2}, Aleksandra Bankovacki^{1,2}, Andrew I Webb^{1,2}, Isabelle S Lucet^{1,2}, John Silke^{1,2,†} & Ueli Nachbur^{1,2,†}

¹The Walter and Eliza Hall Institute of Medical Research, Parkville, Vic., Australia. ²Department of Medical Biology, University of Melbourne, Melbourne, Vic., Australia. ³Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany. ⁴Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. † These authors contributed equally to this work

Published in *EMBO reports* (2021)

The recognition of bacteria and the stimulation of the immune system through nucleotide-binding oligomerization domain (NOD) proteins NOD1 and NOD2 has a crucial role in the clearance of bacterial pathogens. Receptor-interacting serine/threonine-protein kinase 2 (RIPK2) is important in the NOD mediated signaling cascade and subject to modulation by phosphorylation and ubiquitination [193, 194]. Our study investigated the role of phosphorylation and ubiquitination on RIPK2 regulation and identified a new regulatory region of RIPK2 that is important for X-linked inhibitor of apoptosis protein (XIAP) engagement and NOD signaling.

In this study, I identified relevant ubiquitination sites induced through NOD signaling by MS-based large-scale ubiquitinome analysis. These results were essential to characterize the ubiquitination sites on RIPK2.



A regulatory region on RIPK2 is required for XIAP binding and NOD signaling activity

Valentin J Heim^{1,2}, Laura F Dagley^{1,2}, Che A Stafford³, Fynn M Hansen⁴, Elise Clayer^{1,2}, Aleksandra Bankovacki^{1,2,‡}, Andrew I Webb^{1,2}, Isabelle S Lucet^{1,2}, John Silke^{1,2,†,*}  & Ueli Nachbur^{1,2,†,**} 

Abstract

Signaling via the intracellular pathogen receptors nucleotide-binding oligomerization domain-containing proteins NOD1 and NOD2 requires receptor interacting kinase 2 (RIPK2), an adaptor kinase that can be targeted for the treatment of various inflammatory diseases. However, the molecular mechanisms of how RIPK2 contributes to NOD signaling are not completely understood. We generated FLAG-tagged RIPK2 knock-in mice using CRISPR/Cas9 technology to study NOD signaling mechanisms at the endogenous level. Using cells from these mice, we were able to generate a detailed map of post-translational modifications on RIPK2. Similar to other reports, we did not detect ubiquitination of RIPK2 lysine 209 during NOD2 signaling. However, using site-directed mutagenesis we identified a new regulatory region on RIPK2, which dictates the crucial interaction with the E3 ligase XIAP and downstream signaling outcomes.

Keywords inflammation; NOD signaling; RIPK2; ubiquitin; XIAP

Subject Categories Immunology; Post-translational Modifications & Proteolysis

DOI 10.15252/embr.202050400 | Received 11 March 2020 | Revised 30 July 2020 | Accepted 13 August 2020 | Published online 21 September 2020

EMBO Reports (2020) 21: e50400

Introduction

Nucleotide-binding oligomerization domain-containing (NOD) proteins NOD1 and NOD2 are intracellular pathogen recognition receptors (PRRs) that sense the bacterial peptidoglycan (PGN) fragments γ -D-Glu-m diaminopimelic acid (DAP) and muramyl dipeptide (MDP), respectively (Girardin *et al.*, 2003a,b). NOD1 and NOD2 play an important role in the clearance of bacterial pathogens, including *Mycobacterium tuberculosis* (Lee *et al.*, 2016), *Listeria monocytogenes* (Jeong *et al.*, 2014), and multiple *Chlamydiae*

species (Zou *et al.*, 2016). Aberrant NOD signaling has long been associated with a range of inflammatory disorders (Caruso *et al.*, 2014; Philpott *et al.*, 2014), and recent findings suggest that inhibition of the NOD pathways could be beneficial in the treatment of allergic asthma (Miller *et al.*, 2018) and type 2 diabetes mellitus (T2DM) (Amar *et al.*, 2011; Schertzer *et al.*, 2011; Denou *et al.*, 2015; Cavallari *et al.*, 2017).

Binding of the respective ligands to NOD1 and NOD2 leads to their self-oligomerization (Maharana *et al.*, 2015) and the recruitment of receptor-interacting serine/threonine-protein kinase 2 (RIPK2) via homotypic caspase recruitment domain (CARD)–CARD interactions (Inohara *et al.*, 2000). RIPK2 is the essential adaptor kinase in the NOD signaling pathway and drives nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) and mitogen-activated protein (MAP) kinase activation (Chin *et al.*, 2002; Park *et al.*, 2007). The kinase activity of RIPK2 was initially reported to be required for signal transduction and for critical autophosphorylation of RIPK2 on S176 in the activation loop of the kinase domain (Dorsch *et al.*, 2006) and Y474 in its CARD (Tigno-Aranjuez *et al.*, 2010). However, recent studies suggest that RIPK2 kinase activity is dispensable for NF- κ B activation and cytokine production (Goncharov *et al.*, 2018; Hrdinka *et al.*, 2018). Furthermore, it has been established that NOD signaling relies on ubiquitination of RIPK2 (Tigno-Aranjuez *et al.*, 2013). This process is coordinated by multiple ubiquitin E3 ligases, including X-linked inhibitor of apoptosis protein (XIAP) (Krieg *et al.*, 2009; Damgaard *et al.*, 2012; Heim *et al.*, 2019). XIAP binds to the kinase domain of RIPK2 via its baculovirus IAP repeat 2 (BIR2) domain (Krieg *et al.*, 2009; Bertrand *et al.*, 2011; Nachbur *et al.*, 2015) and generates K63-linked polyubiquitin chains on multiple lysine residues (Goncharov *et al.*, 2018). This leads to the recruitment of the linear ubiquitin chain assembly complex (LUBAC) (Damgaard *et al.*, 2012) and the generation of M1-linked polyubiquitin chains on RIPK2 that serve as binding platforms for I κ B kinase (IKK) and transforming growth factor beta-activated kinase 1 (TAK1) complexes.

1 The Walter and Eliza Hall Institute of Medical Research, Parkville, Vic., Australia

2 Department of Medical Biology, University of Melbourne, Melbourne, Vic., Australia

3 Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany

4 Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

*Corresponding author. Tel: +61 3 9345 2945; E-mail: silke@wehi.edu.au

**Corresponding author. Tel: +61 3 9345 2941; E-mail: nachbur@wehi.edu.au

†These authors contributed equally to this work

‡Present address: Translational Research, CSL Limited, Melbourne, Vic., Australia

While the importance of XIAP and LUBAC for immune responses mediated by NOD1 and NOD2 has been demonstrated *in vitro* and *in vivo*, not much is known about the function of individual ubiquitination sites on RIPK2. A putatively ubiquitinated lysine residue on RIPK2 (K209) was discovered more than 10 years ago in a systematic screening of lysine to arginine mutations (K/R) that disrupted NF- κ B activation in overexpression experiments (Hasegawa *et al*, 2008). A subsequent study showed that the K209R mutation blocked RIPK2 ubiquitination and signaling (Tigno-Aranjuez *et al*, 2013), and it was concluded that K209 is directly ubiquitinated and is indispensable for NOD2 responses. Nevertheless, ubiquitination of RIPK2 on K209 has not been demonstrated experimentally. Intriguingly, a recently published proteomics experiment reported multiple ubiquitination sites on RIPK2, but the authors did not identify K209 (Goncharov *et al*, 2018). Instead, they found ubiquitination sites on the C terminus of RIPK2 and generated a K410R/K538R double mutation that reduced MDP-dependent responses of THP-1 cells. Altogether, this highlights that our understanding of how post-translational modifications (PTMs) of RIPK2 regulate NOD signaling is incomplete.

Due to the association with inflammatory diseases, pharmaceutical companies have invested in the development of inhibitors for the NOD signaling pathway. RIPK2 has been established as a potential drug target, particularly in inflammatory bowel disease, and RIPK2-targeting kinase inhibitors have been developed (Damgaard *et al*, 2012; Tigno-Aranjuez *et al*, 2014; Nachbur *et al*, 2015). Recent studies showed that the inhibition of NOD signaling is not directly due to the inhibition of the kinase function of RIPK2, but rather by the disruption of the RIPK2-XIAP interaction (Nachbur *et al*, 2015; Goncharov *et al*, 2018; Hrdinka *et al*, 2018). This has led to the hypothesis that protein–protein interaction inhibitors could be used to treat NOD-driven diseases and highlights the need for a detailed understanding of post-translational modifications on RIPK2.

A significant issue that has hindered our understanding of such mechanisms of the NOD signaling pathway is the lack of specific biochemical tools. Most studies have been limited to overexpression experiments in cancer cell lines, which has many drawbacks including the formation of artefactual interactions or altered protein activities (von Mering *et al*, 2002; Aebersold & Mann, 2003). In the context of NOD signaling, it was shown that ectopic overexpression of NOD receptor complex components leads to pathway activation independent of PGN binding (McCarthy *et al*, 1998; Thome *et al*, 1998; Ogura *et al*, 2001). To study NOD2 signaling mechanisms and investigate the molecular determinants of RIPK2 activation, we established a new mouse strain with endogenously FLAG-tagged RIPK2. This allowed us to isolate RIPK2 from primary tissues and cells and to characterize the PTMs on RIPK2 that occur during MDP-induced signaling at endogenous levels. While we did not identify ubiquitination of the putative ubiquitination site K209, we identified a novel regulatory region that controls XIAP binding and is required for signal transduction.

Results and Discussion

FLAG-RIPK2 knock-in mice represent a novel tool to study endogenous NOD signaling mechanisms

To study NOD signaling at the endogenous level and to investigate how RIPK2 regulates signal transduction, we generated N-terminally

FLAG-tagged RIPK2 knock-in mice by microinjection of single guide RNAs (sgRNAs), recombinant Cas9 protein, and a dsDNA oligonucleotide encoding the FLAG-tagged sequence of RIPK2 with homologous arms upstream and downstream of the sgRNA targeted region into wild-type C57Bl/6 embryos (Fig EV1). Using this process, we generated mice harboring the desired FLAG-tagged version of RIPK2, as well as mice with defined insertions and deletions. After backcrossing to C57Bl/6 mice, we established a FLAG-RIPK2 knock-in mouse strain as well as a new RIPK2 knockout strain.

First, we explored the tissue distribution of RIPK2 by testing homogenates from organs of FLAG-RIPK2 and wild-type mice by Western blot. As the expression levels of RIPK2 in these organs were too low for detection using anti-FLAG antibodies, we subjected organ homogenates to anti-FLAG immunoprecipitation (Fig 1A) and probed the supernatants from boiled beads for the presence of RIPK2. We found high levels of FLAG-RIPK2 in homogenates from the lung and the spleen, and lower amounts in the brain, colon, small intestine, skin, and liver. We did not detect FLAG-RIPK2 in the kidney.

We then tested the functionality of FLAG-RIPK2 using IFN γ -primed bone marrow-derived macrophages (BMDMs). Priming with IFN γ is required for the BMDMs to take up and respond to MDP (Fekete *et al*, 2017). Upon MDP stimulation, cells generated from homozygous (KI/KI) or heterozygous (WT/KI) FLAG-RIPK2 mice induced NF- κ B and MAP kinase pathways equivalent to wild-type cells (Fig 1B). Western blot for RIPK2 also revealed even expression levels between wild-type and FLAG-tagged versions of RIPK2, which was particularly clear in the heterozygote samples (Fig 1B). After stimulation with MDP, ubiquitination of FLAG-RIPK2 in BMDMs from knock-in mice was comparable to ubiquitination of RIPK2 in cells from wild-type mice (Fig EV2). As expected, cells generated from our new strain of RIPK2 knockout mice were unresponsive to MDP and did not express detectable RIPK2. Cytokine production of BMDMs after MDP stimulation was then measured by ELISA (Fig 1C). In IFN γ -primed wild-type and FLAG-RIPK2 BMDMs, treatment of MDP induced the secretion of TNF, IL-6, and MCP-1 at equal levels, while RIPK2 KO BMDMs were unresponsive. To confirm that NOD2-dependent responses in FLAG-RIPK2 mice were indistinguishable from wild-type mice *in vivo*, we intraperitoneally (i.p.) injected MDP or PBS into wild-type, FLAG-RIPK2 and RIPK2 knockout mice and measured cytokine levels in the serum by ELISA (Fig 1D). MDP challenge caused a reproducible increase in the levels of IL-6 and TNF in the serum of wild-type and knock-in mice. In RIPK2 deficient mice, MDP injection did not result in an increase of cytokines. Altogether these data show that FLAG-RIPK2 mice and primary cells from these mice responded normally to NOD2 stimulation.

Post-translational modifications on RIPK2

Initially, it was thought that the kinase activity of RIPK2 was required for signaling down stream of NOD receptors (Nembrini *et al*, 2009). Recently, however, two groups independently showed that it is dispensable for NOD mediated NF- κ B activation and cytokine production (Goncharov *et al*, 2018; Hrdinka *et al*, 2018), although it is an open question whether the kinase activity is important in other RIPK2-regulated cellular processes, such as autophagy (Cooney *et al*, 2010; Homer *et al*, 2010; Anand *et al*, 2011; Lupfer

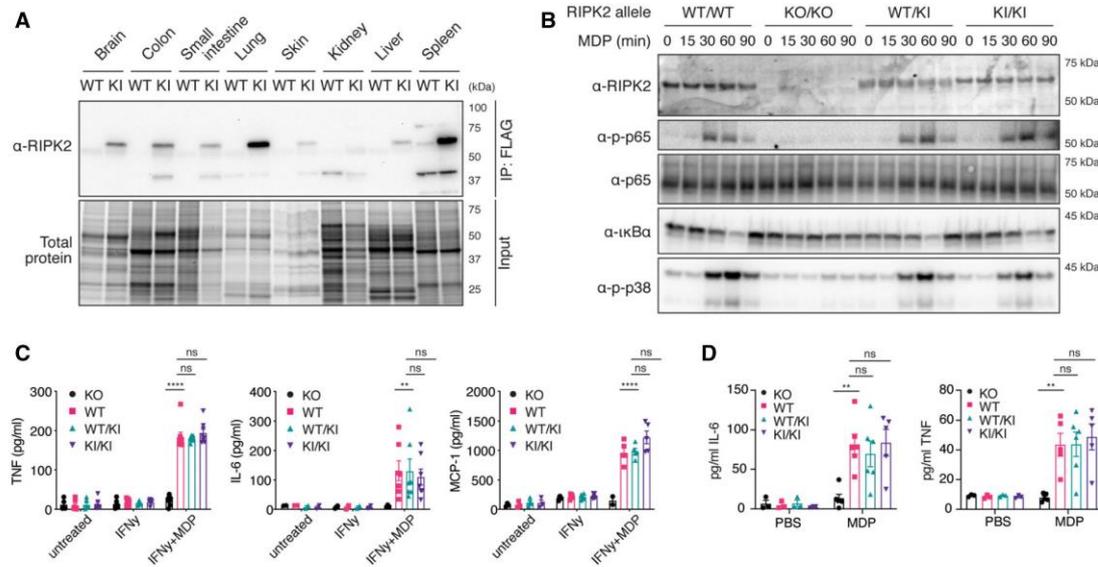


Figure 1. Flag-RIPK2 knock-in mice as a tool to study endogenous NOD signaling mechanisms.

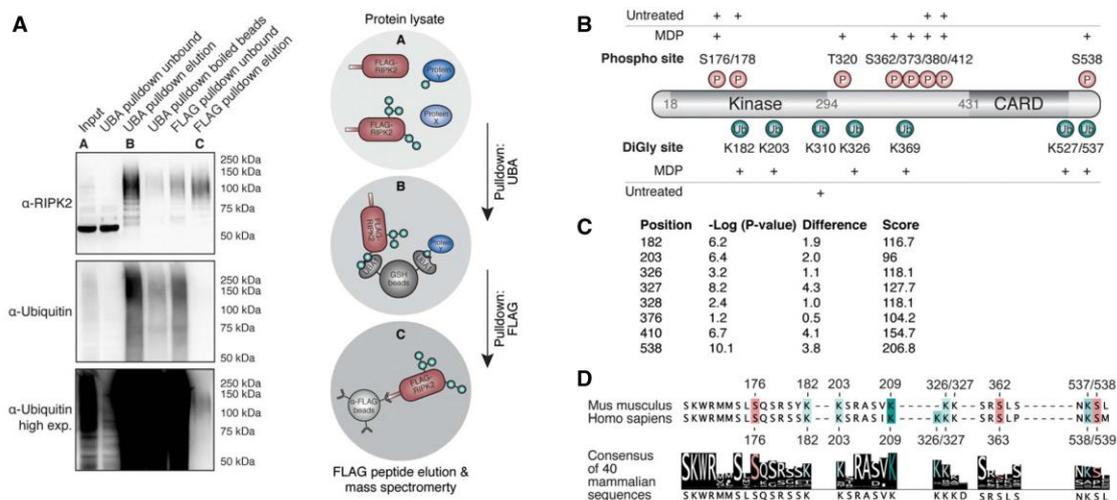
- A** Tissue distribution of RIPK2 determined by anti-FLAG immunoprecipitation. Organ homogenates from WT (wild-type) and KI (homozygous FLAG-RIPK2) mice were subjected to anti-FLAG immunoprecipitation and immunoblotting.
- B** Inflammatory signaling in wild-type (WT/WT), RIPK2 CRISPR KO (KO/KO), and FLAG-RIPK2 heterozygous (WT/KI) and homozygous (KI/KI) BMDMs. BMDMs were primed with IFN γ , stimulated with MDP for indicated times, and analyzed by immunoblotting.
- C** Cytokine production of RIPK2 CRISPR KO (KO), wild-type (WT), and FLAG-RIPK2 heterozygous (WT/KI) and homozygous (KI/KI) BMDMs in response to MDP. BMDMs were left untreated or treated with IFN γ alone or IFN γ and MDP overnight and cytokines were measured by ELISA. $N = 5-8$ mice. Shown is average \pm SEM. $^{ns}P > 0.05$; $^{**}P \leq 0.01$; $^{****}P \leq 0.0001$; two-way ANOVA.
- D** Serum cytokines in RIPK2 CRISPR KO (KO), wild-type (WT), and FLAG-RIPK2 heterozygous (WT/KI) or homozygous (KI/KI) mice after i.p. MDP administration. Mice were injected i.p. with PBS or MDP, sacrificed after 4 h and serum cytokines were measured by ELISA. $N = 3-6$ mice. Shown is average \pm SEM. $^{ns}P > 0.05$; $^{**}P \leq 0.01$; two-way ANOVA.

et al, 2013). These studies indicate that it is rather post-translational modification, particularly ubiquitination of the adaptor protein RIPK2, that is the critical mediator of NOD/RIPK2 signaling. Although it has been demonstrated that RIPK2 is post-translationally modified during signaling, we did not observe ubiquitination of RIPK2, as characterized by a high molecular weight smear on Western blots, in whole-cell lysates of MDP-stimulated BMDMs (Fig 1B). We also did not observe ubiquitination sites on RIPK2 when we performed anti-FLAG immunoprecipitation on MDP-stimulated FLAG-RIPK2 BMDMs followed by mass spectrometry analysis (Dataset EV1). These data indicate that the majority of cellular RIPK2 is not ubiquitinated and not part of the NOD signaling complex, even after stimulation, and suggests that an additional purification step is required to investigate RIPK2 in its activated state.

Therefore, we established a sequential pulldown protocol to enrich for the RIPK2 pool that participates in NOD2 signaling complexes. IFN γ -primed BMDMs were stimulated with MDP for 30 min and lysates were first enriched for ubiquitinated proteins using glutathione S-transferase (GST)-ubiquitin associated domain (UBA) bound to Sepharose beads (Hjerpe *et al*, 2009; Fiil *et al*, 2013). Bound proteins were then eluted with Glutathione, and eluates were subjected to anti-FLAG immunoprecipitation, followed by

elution with 3x-FLAG peptide (Fig 2A). The first pulldown (UBA, lane B) yielded a sample containing readily detectable levels of a ladder of RIPK2 species suggestive of ubiquitination, as well as many other ubiquitinated proteins (anti-ubiquitin, bottom panel). The sample obtained by sequential pulldown with anti-FLAG (lane C) also contained modified RIPK2; however, the background of ubiquitinated proteins was significantly reduced, further suggesting that this approach resulted in purification of activated RIPK2.

Tryptic digests of these samples were then generated and analyzed by mass spectrometry, revealing a substantial enrichment of ubiquitin and RIPK2 peptides. To specifically determine stimulation-dependent PTMs on RIPK2, this dataset was compared with datasets obtained from FLAG pulldowns of unstimulated BMDMs. In unstimulated BMDMs, only one single K-*e*-diglycine site (diGly, diglycine remnant on lysine after tryptic digestion of ubiquitinated proteins) on RIPK2 was observed at the C-terminal end of the kinase domain (K310). In contrast, MDP-stimulated BMDMs revealed multiple RIPK2 ubiquitination sites within the kinase domain (K182, K203), in the intermediate region (K326, K369) and in the CARD (K527, K537) (Fig 2B). The function of most of the detected sites is still uncharacterized; only one recent study showed that human THP-1 cells expressing a K410R/K538R double mutant (human K538



corresponds to K537 in mice) display reduced NF- κ B activation and cytokine production. We were not able to detect the previously described ubiquitination site K209 (Hasegawa *et al.*, 2008) using our stringent protocol.

Phosphopeptides on RIPK2 were detected in both unstimulated and MDP-stimulated BMDMs. Stimulation-dependent phosphorylation was detected in the intermediate region (T320, S362, S373) and the CARD (S539). The activation loop of RIPK2 was phosphorylated on two residues (S176, S178); however, these phosphopeptides were discovered in stimulated as well as in unstimulated cells. This was surprising since phosphorylation of the kinase activation loop is associated with RIPK2 activation (Dorsch *et al.*, 2006; Rahman *et al.*, 2014).

To further confirm the physiological importance of the RIPK2 ubiquitination sites, we examined the ubiquinome of the human monocytic cell line THP-1 employing diGly proteomics. In unstimulated cells, we did not detect any diGly sites on RIPK2 but we consistently observed several diGly marks after L18-MDP stimulation (Fig 2C, Dataset EV2). The sites identified in THP-1 cells reflected our results obtained using the sequential pulldown protocol in mouse cells, validating our initial approach. All stimulation-dependent diGly modifications that we detected on murine RIPK2 residues, which are also conserved in humans, were detected in L18-MDP-stimulated THP-1 cells (K182, K203, K326, K538). Additionally, two diGly-modified lysines (K376, K410) that are not conserved in mice were found. Again, with this second protocol using human cells, we were not able to detect ubiquitination of K209. As expected, we did

not detect any phosphorylation events on RIPK2 using this approach, indicating that there are no simultaneous ubiquitination and phosphorylation events on a single peptide after tryptic digest.

Most PTMs on RIPK2 are redundant for NOD2 signaling

While multiple ubiquitin E3 ligases and deubiquitinases (DUBs) have been suggested to regulate RIPK2 ubiquitination, the contribution of individual ubiquitination sites on RIPK2 has not been characterized in endogenous systems so far. Here, we demonstrated that RIPK2 is ubiquitinated on multiple lysine residues. This is not unusual as many proteins become ubiquitinated on multiple sites during signaling (Ball *et al.*, 2016; Wagner *et al.*, 2016). Typically, there appears to be flexibility in the lysine residues that can be ubiquitinated and often mutation of a single lysine has little impact on signaling. This is believed to be because E3 ligases are not usually restricted to a specific motif, in the way that for example kinases or caspases are, and can therefore be promiscuous in the lysine that they modify (Petroski & Deshaies, 2003; Wu *et al.*, 2003).

To assess the impact of individual RIPK2 PTMs on NOD signaling, we generated mutants of lysine and serine residues that we found to be modified upon NOD2 stimulation and are highly conserved among mammals (Fig 2D). Additionally, we included conserved sites that have previously been associated with RIPK2 activation (Fig 2D; Dorsch *et al.*, 2006; Hasegawa *et al.*, 2008; Tigno-Aranjuez *et al.*, 2010). These corresponded to K182, K203, K326, K327, S363, K538, S539, and K209 in humans.

We used these RIPK2 variants to test whether single ubiquitination and phosphorylation sites on RIPK2 are critical for its function in a close-to-endogenous setting. For this, we generated RIPK2-deficient THP-1 cells by transient transfection with Cas9 and RIPK2 sgRNA encoding plasmids and confirmed the knockout of RIPK2 by Western blot. These knockout cells were then transduced with doxycycline-inducible RIPK2 constructs to express RIPK2 at levels similar to wild-type THP-1. In contrast to overexpression studies, expression of RIPK2 alone did not autoactivate NF- κ B, but additional treatment with L18-MDP-induced transient phosphorylation of p65 and I κ B α and degradation of I κ B α (Fig 3A). Upon MDP stimulation, cells reconstituted with all mutant forms of RIPK2 activated NOD signaling normally, except the previously described K209R mutant.

After L18-MDP stimulation, we observed significantly reduced levels of IL-8 in RIPK2-deficient cells reconstituted with the K209R mutant, while all other mutants produced equal amounts of IL-8 compared with either wild-type THP-1 cells or RIPK2-deficient cells reconstituted with wild-type RIPK2 (Fig 3B).

Notably, RIPK2 K209R was the only form of RIPK2 which seemed to present in a second, higher molecular weight band after stimulation, similar to recently described Riposomes (Ellwanger *et al*, 2019). To test the impact of each individual modified site on the ubiquitination pattern on RIPK2 during NOD signaling, cells were subjected to UBA pulldowns and analyzed by Western blot. Stimulation with L18-MDP led to distinct RIPK2 polyubiquitination (Fig 3C) and the removal of single ubiquitination or phosphorylation sites, besides K209, did not affect RIPK2 ubiquitination.

It was previously reported that K209 is critical for RIPK2 ubiquitination and NOD signaling (Hasegawa *et al*, 2008). However, we observed that residual ubiquitination of RIPK2 K209R was still present. Since RIPK2 K209R failed to activate NF- κ B and to produce cytokines, and displayed reduced ubiquitination, we hypothesized that this mutation led to a loss of the critical K63- or M1-linked ubiquitin chains. To test this hypothesis, cells reconstituted with wild-type RIPK2 or K209R RIPK2 were stimulated with L18-MDP and subjected to either UBA pulldown or to pulldowns with K63- or M1-chain-specific antibodies (Newton *et al*, 2008; Matsumoto *et al*, 2012). Compared to wild-type RIPK2, the ubiquitination of K209R was reduced in all pulldowns; however, K63- and M1-linked chains were still detected, indicating that the K209R mutant is still able to be ubiquitinated on other lysine residues (Fig 3D). These reduced chains are, however, unable to induce downstream signaling which might explain the lack of global ubiquitination events after MDP stimulation in cells harboring the K209R mutation (Panda & Gekara, 2018).

Our results clearly indicate that most single phosphorylation and ubiquitination events on RIPK2 are redundant. While this is the case for many signaling proteins, there are some exceptions to this general rule. For example, mutation of K377 in RIPK1 has a profound effect on RIPK1 ubiquitination and TNFR1 induced activation of NF- κ B (O'Donnell *et al*, 2007). Our data as well as previous studies could lead to the assumption that K209 is another similarly special residue because the K209R mutation blocked overexpression-induced NF- κ B activation and RIPK2 ubiquitination (Hasegawa

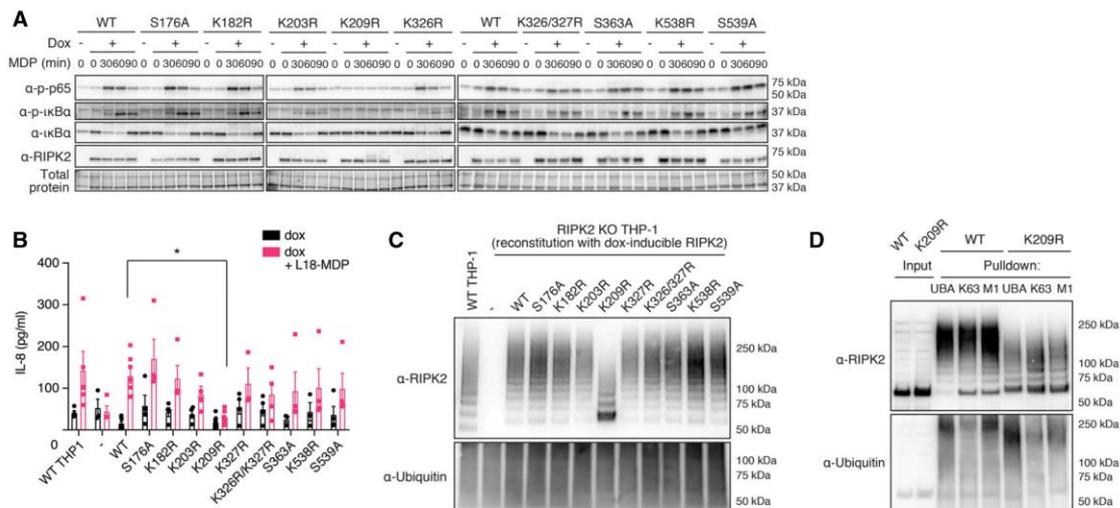


Figure 3. Characterization of RIPK2 diGly- and phosphosite mutations.

A Activation of the NF- κ B pathway by RIPK2 Lysine- and phosphosite mutants. RIPK2-deficient THP-1 cells reconstituted with wild-type RIPK2 or RIPK2 mutants were stimulated with L18-MDP, harvested at indicated time points and activation of the NF- κ B pathway was analyzed by immunoblotting.
B IL-8 production of wild-type THP-1 and RIPK2-deficient THP-1 cells reconstituted with wild-type RIPK2 or RIPK2 mutants and stimulated with L18-MDP was assessed by ELISA. $N = 4-8$ experiments. Shown is average \pm SEM. * $P \leq 0.05$; two-way ANOVA.
C RIPK2 ubiquitination determined by UBA pulldown in RIPK2-deficient cells reconstituted with wild-type or mutant RIPK2 after stimulation with L18-MDP.
D Detection of K63- and M1-linked ubiquitin chains by UBA pulldown or pulldown with ubiquitin chain type-specific antibodies in RIPK2-deficient cells reconstituted with wild-type or K209R RIPK2 after stimulation with L18-MDP.

et al., 2008) and the mutant has a loss of function phenotype when overexpressed together with ubiquitin in HEK 293T cells (Tigano-Aranjuez *et al.*, 2013). However, unlike K377 in RIPK1, K209 is on the C-lobe of the kinase domain of RIPK2, which is already indicative of a different function of this site.

K209 and I212 form a regulatory region that influences signal transduction

Since neither we nor others were able to detect ubiquitination of K209 by mass spectrometry analysis (Goncharov *et al.*, 2018), and mutation of K209 reduced, but did not completely abolish K63 and M1 linked ubiquitination of RIPK2 yet this same mutation stopped NF- κ B activation and cytokine production, we hypothesized that K209 is not a critical ubiquitination site but serves a different function. K209 is located at the N-terminal end of α E helix in the C-lobe of the kinase domain (Fig 4A; PDB 4C8B; (Canning *et al.*, 2015)). It is part of a hydrophobic pocket formed by amino acids of helix α E and amino acids in the loop between helix α E and α EF, which are suggestive of a regulatory interface. To test the hypothesis that disruption of this region prevents NOD2 signaling, we generated mutations of amino acids K209 and I208 that contribute to the making of

this pocket. While K209 is invariant among vertebrates, I208 is replaced by valine in most vertebrate RIPK2 sequences (Fig 4B). We also mutated I212, which is part of the α E helix with its side chain sitting deep within the pocket. I212 is highly conserved and only replaced by other hydrophobic amino acids, such as valine or methionine, in some vertebrate species.

We reconstituted RIPK2-deficient THP-1 cells with the new RIPK2 mutants we generated and tested them for NF- κ B activation, RIPK2 ubiquitination, and cytokine secretion. The expression levels of these RIPK2 mutants seemed equivalent (Fig 4C). Conservative I208V and I212M mutations did not prevent activation of NF- κ B or the production of IL-8 after MDP stimulation (Fig 4D). In contrast, cells that expressed the RIPK2 I212A mutant produced significantly more IL-8 than cells expressing wild-type RIPK2, and markers of NF- κ B activation were similar or enhanced (Fig 4C and E). We also observed enhanced ubiquitination after MDP stimulation of RIPK2 I212A compared with wild-type or other mutant forms of RIPK2 (Fig 4E). Finally, the substitution of isoleucine 212 with an aspartic acid (I212D) completely blocked RIPK2 ubiquitination and NF- κ B activation. These results suggest that the C-lobe pocket is critical in the regulation of RIPK2 activity and may act to recruit an E3 ligase, such as XIAP, to the NOD signaling complex.

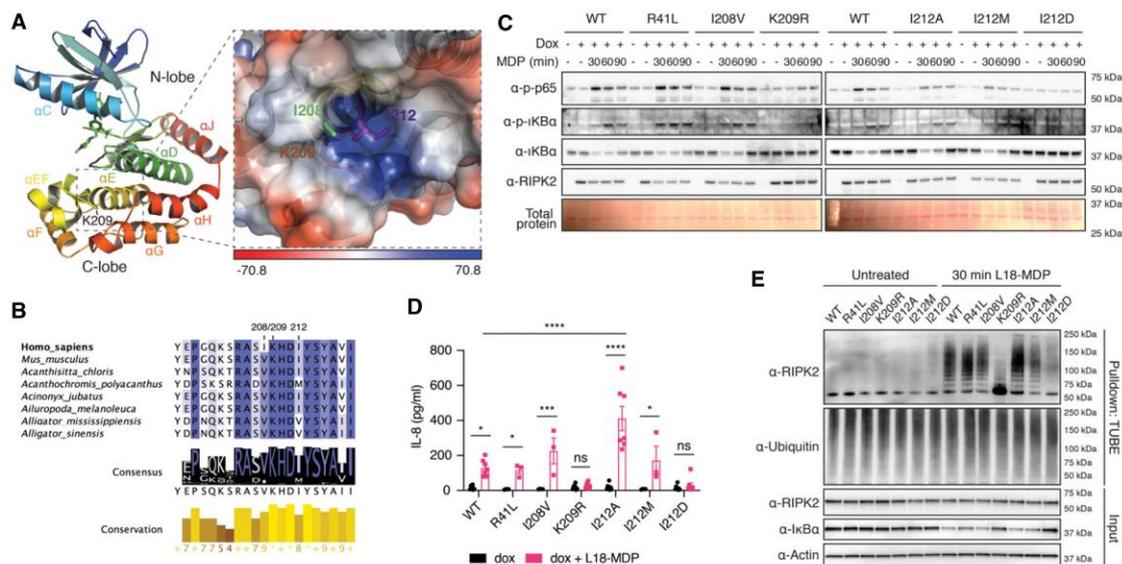


Figure 4. Functional studies of RIPK2 mutations introduced in close proximity to K209R.

- A Structural features of the RIPK2 kinase domain (left) and location of K209 within a hydrophobic pocket between helices α EF and α E (right). Shown is chain B of the RIPK2 kinase in complex with ponatinib from PDB:4C8B. The electrostatic interaction potential is shown as a blue to red gradient.
- B Conservation of amino acids creating a hydrophobic pocket between helices α EF and α E. Degree of conservation among 227 vertebrate species indicated by background color saturation.
- C Activation of the NF- κ B pathway by RIPK2 pocket mutants. RIPK2-deficient THP-1 cells were reconstituted with wild-type RIPK2 or RIPK2 mutants, stimulated with L18-MDP, harvested at indicated time points, and analyzed by immunoblotting.
- D IL-8 production of RIPK2-deficient THP-1 cells reconstituted with wild-type RIPK2 or RIPK2 mutants and stimulated with L18-MDP was assessed by ELISA. $N = 3-8$ experiments. Shown is average \pm SEM. $^{ns}P > 0.05$; $^{*}P \leq 0.05$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$; two-way ANOVA.
- E Ubiquitination of RIPK2 pocket mutants. RIPK2-deficient THP-1 cells were reconstituted with wild-type or mutant RIPK2, left unstimulated or stimulated with L18-MDP and subjected to UBA pull-down and immunoblotting.

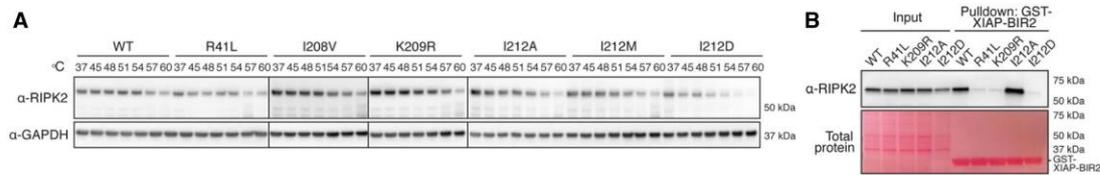


Figure 5. RIPK2 K209 and I212 mediate XIAP binding.

A Thermal stability assay using selected RIPK2 mutants. THP-1 cells expressing RIPK2 were subjected to heat treatment, and non-denatured fractions were analyzed by immunoblotting.
B Binding of RIPK2 to XIAP-BIR2. Lysates from RIPK2-deficient THP-1 cells reconstituted with WT or mutant RIPK2 were subjected to pull-down experiments with recombinantly expressed XIAP-BIR2 and analyzed by immunoblotting.

While these data confirm previous reports that K209 is indispensable for NOD mediated NF- κ B activation and cytokine production, and that RIPK2 ubiquitination is impaired when K209 is mutated to an arginine, the absence of direct ubiquitination of K209 indicated a different mechanism than previously reported. The reason for this is twofold: Firstly, we used two multi-replicate complementary experimental approaches to determine the phosphorylation- and ubiquitination signature on RIPK2 upon activation. The fact that we identified identical diGly sites on RIPK2 in mouse (BMDMs) and in human (THP-1) gives high confidence in our datasets. Despite the consistency of the PTM data across the two cell lines, we, like others in the field, failed to observe a diGly site corresponding to K209. Secondly, mutation of a residue in close proximity, but not directly affecting K209, resulted in an even more dramatic impact on RIPK2 ubiquitination than mutation of K209 itself. This suggests that structural integrity of this region is critical for signal transduction. To this end, we cannot exclude that our I212D mutation rendered K209 inaccessible for ubiquitination, but further studies will be required to examine this experimentally.

It should be noted that we not only failed to detect ubiquitinated K209, but we also did not detect unmodified K209 in our datasets on MDP-stimulated BMDMs or THP-1 cells. Nevertheless, when analyzing other datasets that identified RIPK2 in deep proteome and pan-kinome experiments (utilizing broad specificity kinase inhibitors to enrich for kinases), this region was readily identifiable as an unmodified tryptic peptide (Creixell *et al*, 2015; Ruprecht *et al*, 2015; Slany *et al*, 2016; Klaeger *et al*, 2017). This could be because we specifically enriched for activated RIPK2 in our protocols and we did not use kinase inhibitors in our approaches. We therefore cannot definitively refute the idea of K209 ubiquitination, although the current evidence suggests that there may be as-yet other unidentified post-translational modifications hindering its identification. Lastly, the fact that a third method did not lead to the detection of a diGly site on RIPK2 K209 (Goncharov *et al*, 2018) hints to the direction that this particular portion of RIPK2 is more complex than currently understood.

Mutation of K209 and I212 disrupts XIAP binding

An alternative explanation for the loss of the potential to activate NF- κ B and cytokine production of the K209R and I212D mutations is instability due to misfolding or impairment of the dimerization potential. The equivalent expression levels of the RIPK2 mutants

suggested that the structural integrity of RIPK2 was not overtly compromised. However, to exclude a trivial explanation for the complete lack of signaling of the I212D mutant in particular, and more rigorously show that the mutants retained their structural integrity, we devised an intracellular thermal stability test (Fig 5A) based on reports that kinase inhibitors can increase the thermal stability of their targeted kinases (Martinez Molina *et al*, 2013; Jafari *et al*, 2014; Alshareef *et al*, 2016; Martinez Molina & Nordlund, 2016; Seashore-Ludlow *et al*, 2018). We modified this assay to examine the effects of mutation on protein stability as a marker for structural integrity, and we think that it could be more widely used to determine whether a particular mutation affects structural integrity in a semi-quantitative manner.

At the physiological temperature of 37°C, wild-type RIPK2 and all RIPK2 mutants were expressed at equal levels. With increased temperature, all mutants displayed similar stability compared with wild-type RIPK2. In particular, all I212 mutants, I212A (activating), I212M (residue in Damsel fish), and I212D (inactivating) had an almost identical thermal stability profile, strongly suggesting that the mutations of this pocket had not caused major structural disruption. To estimate the effects of these mutations on the structural integrity of RIPK2, we employed molecular modeling using the DynaMut software (Rodrigues *et al*, 2018). This software predicted that all the mutants increase rigidity of the α G region, the activation loop, and a portion of the N-lobe α C helix with K209R (1.411 kcal/mol) > I212A (0.764 kcal/mol) > I212D (0.38 kcal/mol) (Fig EV3). Since this increased rigidity is predicted for both activating (I212A) and inactivating (K209R, I212D) mutations, it is unlikely to be the cause of the loss of function.

The ubiquitin E3 ligase XIAP is indispensable for NOD2 responses *in vitro* and *in vivo* (Bauler *et al*, 2008; Krieg *et al*, 2009; Damgaard *et al*, 2012; Stafford *et al*, 2018) and binds to the RIPK2 kinase domain via its Baculovirus IAP Repeat 2 (BIR2) domain (Krieg *et al*, 2009; Damgaard *et al*, 2012). The structural determinants underlying the interaction between the two proteins remain unclear. The XIAP-BIR2 domain contains a deep and distinctive hydrophobic cleft that typically mediates binding to proteins harboring an IAP-binding motif (IBM) as present in caspases or second mitochondria-derived activator of caspases (SMAC) (Wu *et al*, 2000; Verhagen *et al*, 2007). However, RIPK2 does not contain such a motif. It is possible that XIAP binds to RIPK2 in a non-canonical fashion, independently of an IBM. Such a mechanism has been shown with the caspase Dronc and the XIAP homolog DIAP1 in

Drosophila melanogaster. Binding of Dronc to DIAP1 is mediated via a short 12-residue peptide located between the CARD and the protease domain of Dronc that binds into the hydrophobic cleft of the DIAP in a similar fashion as observed in IBM-containing proteins. Therefore, one explanation could be that the RIPK2-XIAP interaction occurs in a similar, non-canonical fashion; however, the region around K209 and I212 has the shape of a pocket. The binding mode of RIPK2 and XIAP-BIR2, therefore, would be completely different from the previously observed interactions. Alternatively, XIAP could bind somewhere else, and mutation of K209 and I212 might lead to conformational changes that have an impact on the structure or orientation of the interaction interface. A region in the N-lobe of RIPK2, in particular residues R36 and R41, was identified as a critical interaction region for the BIR2 of XIAP (Hrdinka *et al*, 2018). However in all reported crystal structures, the RIPK2 kinase domain is organized in head-to-tail dimers (PDB: 4C8B, 5AR4, 6ES0, 6FU5, 5J79) and the regions around R36 and R41 are quite far apart from K209 (approximately 40 Å). According to these structures, it is unlikely that the BIR2 of XIAP binds both areas simultaneously, but the reason for reduced binding of the K209R and I212D mutant could be due to conformational impairment or that it may have influenced binding to XIAP indirectly.

To test whether mutation of K209 and I212 disrupted XIAP binding, a purified, recombinant, GST-coupled BIR2 domain of XIAP was used to precipitate RIPK2 from THP-1 lysates (Fig 5B). Wild-type RIPK2 was strongly enriched by XIAP-BIR2 pull-down, and I212A was even more abundant, correlating with the increased stimulus-dependent ubiquitination of this mutant and increased cytokine secretion compared with wild-type RIPK2. Binding of K209R and I212D to the BIR2 of XIAP was, however, drastically reduced and comparable to R41L, a RIPK2 mutant previously shown to have impaired binding to XIAP (Hrdinka *et al*, 2018).

In recent studies, higher order intracellular signaling platforms consisting of RIPK2 and NOD receptors were described (Gong *et al*, 2018; Pellegrini *et al*, 2018; Ellwanger *et al*, 2019). In particular, inhibition of XIAP by either siRNA or by SMAC mimetic compounds led to RIPK2-containing speck-like structures in cells, termed Ripsosomes (Ellwanger *et al*, 2019). Here, we described two mutations that perturbed the interaction between RIPK2 and XIAP. On close examination, RIPK2 K209R seems to accumulate in a higher order band in a Western blot after stimulation (Fig 3A), consistent with the hypothesis that in the absence of XIAP, RIPK2 is moved to Triton-insoluble Ripsosomes. However, the RIPK2 I212D or the RIPK2 R41L mutations, which reduce RIPK2-XIAP interaction to a similar level to the RIPK2 K209R mutation, did not lead to an equivalent band in Western blots after stimulation. Therefore, these new XIAP binding mutants will provide novel insight into the mechanism and function of Ripsosome formation and their relevance for NOD signaling.

Taken together, our work provides a detailed and systematic map of post-translational modifications on RIPK2 during NOD signaling, and we provide evidence that most single phosphorylation and ubiquitination events on RIPK2 are redundant in systems with close-to-endogenous levels of RIPK2. We identified a regulatory region on RIPK2 which influences the crucial interaction with XIAP. This region includes a pocket-shaped region around residues I212 and K209 on the C-lobe of RIPK2's kinase domain. Our findings give an explanation to the conundrum that has plagued the field to date:

why mutations of K209 reduce RIPK2 ubiquitination and block NOD signaling, yet ubiquitination of K209 has never been detected experimentally. As interfering with the RIPK2-XIAP interaction has emerged as a strategy to inhibit NOD signaling, it is tempting to speculate that this region could be targeted with small molecules for the treatment of diseases with increased NOD signaling. Although we cannot conclusively demonstrate that mutations in the region of this pocket do not disrupt RIPK2 in a manner which may be important for its function (kinase activity, autophagy, dimerization), our data collectively suggest that site-directed mutagenesis of either K209 or I212 blocks RIPK2 ubiquitination and inflammatory signaling by displacing XIAP.

Materials and Methods

Generation of FLAG-RIPK2 CRISPR knock-in mice

The FLAG-RIPK2 mice were generated by the MAGEC laboratory (WEHI) as previously described (Kueh *et al*, 2017) on a C57BL/6J background. To generate FLAG-RIPK2 mice, 20 ng/μl of Cas9 mRNA, 10 ng/μl of sgRNA (TGAACGGGGACGCCATCTGC;) and 40 ng/μl of oligo donor (GCCGCCCGGGACCTAGCCCGCGGCCAGGGTCCGGGGAGCGCCCGCGGACGCCGAGCCATGGACTACAAAGACGATGACGATAAAGGATCTACCAACGGGGACGCCATCTGCAGCGCGCTACCCCCCATCCCGTACCACAAGCTCGCCGACCTG) were injected into the cytoplasm of fertilized one-cell stage embryos. Twenty-four hours later, two-cell stage embryos were transferred into the oviducts of pseudo-pregnant female mice. Viable offspring were genotyped by next-generation sequencing.

Cell culture, generation of BMDMs, and stimulation protocols

Wild-type THP-1 cells and 293T cells were sourced from ATCC™. THP-1 cells were cultured in RPMI supplemented with 8% FBS and antibiotics (penicillin, streptomycin, GIBCO) at 37°C with 10% CO₂ in a humidified incubator. 293T cells were cultured in DMEM (GIBCO) with 8% FBS in the same conditions. BMDMs were generated from the femur and tibiae of mice and cultured for 6 days in DMEM (InvivoGen) supplemented with 8% FBS (GIBCO) and 20% L929 supernatant and antibiotics (penicillin, streptomycin). Cells were then detached using trypsin-EDTA and replated in 12- and 24-well tissue culture plates. Replated cultures of BMDMs were primed with murine interferon-γ (5 ng/ml, R&D Systems) overnight and again 2 h before stimulation with MDP (10 μg/ml, InvivoGen). THP-1 cells were stimulated with L18-MDP (200 ng/ml, Bachem).

Generation of RIPK2 CRISPR knockout THP-1 cells

RIPK2-deficient THP-1 cells were generated using a CRISPR/Cas9-based knockout workflow as previously described (Schmid-Burgk *et al*, 2014). Briefly, a sgRNA (GACCTGCGCTACCTGAGCCGGG) targeting RIPK2 was designed. THP-1 cells were nucleofected with one plasmid expressing sgRNA and one expressing mCherry-Cas9 (pLK0.1- gRNA-CMV-GFP, CMV-mCherry-Cas9) using the SG Cell Line 4D-Nucleofector™ X Kit S and a 4D-Nucleofector X unit. mCherry-positive cells were sorted and cloned by limiting dilution. After identifying clones, cells were replated and grown to identify

RIPK2 knockout clones by assessing RIPK2 expression on Western blot.

Intraperitoneal MDP injections

All *in vivo* experiments were performed according to the guidelines of the animal ethics committee of WEHI, ethics approval (2011.014, 2014.004 and 2017.004). Sex- and age-matched littermate controls were used within each experiment. For *in vivo* MDP challenge: wild-type, RIPK2 knockout, and FLAG-RIPK2 knock-in (hetero- and homozygous) mice were administered MDP (5 mg/kg, *i.p.* in 200 μ l PBS, Bachem) or PBS and sacrificed after 4 h. Peripheral blood was collected by cardiac puncture.

Western blotting

Following stimulation, cells were lysed in 2 \times SDS lysis buffer (126 mM Tris-HCl pH 8, 20% *v/v* glycerol, 4% *w/v* SDS, 0.02% *w/v* Bromophenolblue, 5% *v/v* 2-mercaptoethanol) and subjected to repeated freeze/boil cycles. Samples were separated using SDS-PAGE and transferred to polyvinylidene fluoride (PVDF) membranes. The following antibodies were used for probing: rabbit anti-RIPK2 (4142S, Cell Signaling Technology), rabbit anti-RIPK2 (SC 22763, Santa Cruz), mouse anti-FLAG (F1804, Sigma), anti- β actin (A-1978, Sigma), rabbit anti-p65 (631213, Upstate), rabbit anti-phospho p65 (3033, Cell Signaling Technology), rabbit anti-phospho p38 (9211, Cell Signaling Technology), mouse anti-phospho I κ B α (9246, Cell Signaling Technology), rabbit anti-I κ B α (9242, Cell Signaling Technology), mouse anti-ubiquitin (3936, Cell Signaling Technology), rabbit anti-GAPDH (2118, Cell Signaling Technology), human anti-K63-linked ubiquitin (Apu3.A8, Genentech), human anti-M1-linked ubiquitin (1F11, Genentech), anti-K27-linked ubiquitin (ab18153, abcam), goat anti-mouse Ig (1010-05), goat anti-rabbit Ig (4010-05) and goat anti-rat Ig HRP (horseradish peroxidase, 3010-05, Southern Biotech), and goat anti-human Ig (109-035-003, Jackson ImmunoResearch).

Cytokine measurement by ELISA

Cytokines from mouse serum or cell culture supernatant were measured by ELISA kits for IL-6, IL-8, TNF, and MCP-1, respectively (Invitrogen), according to the manufacturer's instructions. Sera and supernatants were diluted 1:10 for MCP-1 measurements.

Immunoprecipitation of RIPK2 from mouse tissues

Organs from 6 weeks old wild-type C57BL/6 mice and FLAG-RIPK2 knock-in mice were lysed in IP buffer (150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, 1% Triton X-100, 2 mM EDTA, all from SIGMA) supplemented with protease inhibitors (cOmplete protease inhibitor cocktail, Roche) using a TissueLysar II (Qiagen). Samples were clarified by centrifugation at 17,000 \times g for 30 min, the protein concentration was assessed using a BCA assay (Thermo Fisher) and 2 mg of protein per lysate were subjected to anti-FLAG immunoprecipitation using 15 μ l of packed magnetic anti-FLAG beads (M2, Sigma) for 4 h. Beads were washed three times in IP buffer, eluted with 2 \times SDS sample buffer, and subjected to immunoblotting.

Ubiquitin enrichments (UBA, TUBE and pulldowns with K63- and M1-specific ubiquitin antibodies)

20 \times 10⁶ THP-1 cells were treated with doxycycline (200 ng/ml) for 5 h and stimulated with L18-MDP (200 ng/ml, Invitrogen) for 30 min, washed in PBS, and lysed in 1–2 ml IP buffer (150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, 1% Triton X-100, 2 mM EDTA) with protease and phosphatase inhibitors and 5 mM n-ethylmaleimide (NEM). Samples were clarified by centrifugation at 17,000 \times g for 15 min and added directly to 20 μ l packed glutathione sepharose beads pre-bound with 100 μ g GST-TUBE (Ubiquillin-UBA4x) or GST-UBA (Ubiquillin-UBA1x) (Hjerpe *et al*, 2009). Beads were incubated on a rotating wheel at 4°C for at least 2 h, washed three times with IP buffer, and eluted with 2 \times SDS sample buffer.

To enrich for K63- and M1-linked ubiquitin species, 20 \times 10⁶ THP-1 cells were treated with doxycycline (200 ng/ml) for 5 h and stimulated with L18-MDP (200 ng/ml, Invitrogen) for 30 min, washed with PBS, and lysed in 1–2 ml IP buffer, as above, supplemented with 6 M Urea (for anti-K63-linked ubiquitin pulldowns) or 8 M Urea (for anti M1-linked ubiquitin pulldowns). Samples were clarified by centrifugation at 17,000 \times g for 15 min and 4 μ g of anti-K63 or anti-M1-linked ubiquitin antibodies [Genentech; (Matsumoto *et al*, 2012; Newton *et al*, 2008)] were added, followed by incubation on a rotating wheel at 4°C for at least 2 h. Antibodies were precipitated with 10 μ l of equilibrated protein G agarose (Thermo), washed three times in IP buffer without Urea, and eluted with 2 \times SDS sample buffer.

Two-step enrichment of modified RIPK2

Twelve dishes of confluent FLAG-RIPK2 BMDMs (equivalent of approximately 25 \times 10⁷ cells or 25 mg of total protein) were primed with IFN γ (5 ng/ml) overnight and fresh IFN γ was added the next morning for another 2 h before stimulation with MDP (10 μ g/ml) for 30 min. Cells were harvested, washed in PBS, and lysed in 2 ml IP buffer per dish (150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, 1% Triton X-100, 2 mM EDTA; supplemented with protease and phosphatase inhibitors and 5 mM n-ethylmaleimide (NEM)). Lysates were clarified by centrifugation at 17,000 \times g for 15 min, and supernatants were directly added to 100 μ l packed glutathione sepharose beads pre-bound with 1 mg GST-UBA. Beads were incubated on a rotating wheel at 4°C overnight, washed three times with IP buffer, and eluted twice with two volumes of IP buffer supplemented with 20 mM reduced glutathione (pHed to 7.5). Elutions were combined, diluted with an equal volume of IP buffer, and added to 50 μ l of packed magnetic anti-FLAG beads (M2, Sigma) and incubated on a rotating wheel at 4°C for 4 h. Beads were washed three times with IP buffer and eluted twice with two volumes of 3 \times -FLAG peptide (1 mg/ml) in TBS pH 7.5.

BMDM diGly proteomics

For the two-step protocol in BMDMs, eluted protein material from pulldowns of FLAG-RIPK2 expressing BMDMs was subjected to tryptic digestion using the FASP method as previously described (Wisniewski *et al*, 2009). Peptides were lyophilized using CentriVap (Labconco) prior to reconstituting in 80 μ l 0.1% FA/2% acetonitrile

(ACN). Peptide mixtures were analyzed by nanoflow reversed-phase liquid chromatography tandem mass spectrometry (LC-MS/MS) on an M-Class HPLC (Waters) coupled to a Q Exactive Orbitrap mass spectrometer (Thermo Fisher). Peptide mixtures were loaded in buffer A (0.1% formic acid, 2% acetonitrile, Milli-Q water) and separated by reverse-phase chromatography using C₁₈ fused silica column (packed emitter, I.D. 75 μ m, O.D. 360 μ m \times 25 cm length, IonOpticks, Australia) using flow rates and data-dependent methods as previously described (Delconte *et al*, 2016; Kedzierski *et al*, 2017).

THP-1 diGly proteomics

Frozen pellets from 50 \times 10⁶ WT THP-1 cells were lysed in 1% sodium deoxycholate (SDC), 100 mM Tris-HCl pH 8.5, immediately boiled for 5 min at 95°C and sonication for 30 s (Branson Sonifier). Protein concentrations were estimated by tryptophan assay. For protein reduction and alkylation, samples were incubated for 5 min at 45°C after addition of tris(2-carboxyethyl)phosphine (TCEP) and 2-chloroacetamide (CAA) to a final concentration of 10 mM and 40 mM, respectively. Samples were digested using trypsin (1:20 w/w, Sigma-Aldrich) in combination with LysC (1/100 w/w, Wako) at 37°C overnight. Protease activity was quenched by addition of four-sample volumes 1% trifluoroacetic acid (TFA) in isopropanol. Quenched samples were loaded onto SDB-RPS cartridges (StrataTM-X-C, 30 mg/3 ml, Phenomenex Inc), pre-equilibrated with 4 ml 30% methanol (MeOH)/1% TFA, and washed with 4 ml 0.2% TFA. After two washes with 4 ml 1% TFA in isopropanol and 1 wash with 0.2% TFA/2% acetonitrile (can), samples were eluted twice with 2 ml 1.25% ammonium hydroxide(NH₄OH)/80% ACN. Eluted samples were diluted with ddH₂O to a final ACN concentration of 35%, snap frozen, and dried by lyophilization.

Lyophilized peptides were reconstituted in IAP buffer (50 mM MOPS, pH 7.2, 10 mM Na₂HPO₄, 50 mM NaCl), and the peptide concentration was estimated by tryptophan assay. For proteome analysis, 10 μ g of peptide material was taken and desalted on SDB-RPS StageTips (Empore) (Kulak *et al*, 2014). Peptides were diluted to a final volume of 200 μ l with 0.2% TFA and loaded onto StageTips, followed by a wash with 200 μ l of 0.2% TFA and 200 μ l of 0.2% TFA/2% ACN, respectively. Captured peptides were eluted with 60 μ l of 1.25% Ammonium hydroxide(NH₄OH)/80% ACN and dried using a SpeedVac centrifuge (Eppendorf, Concentrator plus). Dried peptides were resuspended in buffer A* (2% ACN/0.1% TFA).

K- ϵ -GG remnant containing peptides were enriched using the PTMScan[®] Ubiquitin Remnant Motif (K- ϵ -GG) Kit (Cell Signaling Technology). Crosslinking of antibodies to beads and subsequent immunopurification was performed with slight modifications as previously described (Udeshi *et al*, 2013). Briefly, cross-linked beads were split equally into eight tubes (~31 μ g of antibody per tube), gently mixed with 1 mg peptide material (1 mg/ml) and incubated for 1 h at 4°C. Beads were washed twice with cold IAP and five times with cold ddH₂O, and peptides were eluted twice with 50 μ l 0.15% TFA. Eluted peptides were desalted and dried as described above and resuspended in 5 μ l buffer A* for LC/MS-MS analysis.

For the THP-1 diGly-enrichment analysis, samples were loaded onto a 50 cm reversed-phase column [75 μ m inner diameter, packed in house with ReproSil-Pur C18-AQ 1.9 μ m resin (Dr. Maisch

GmbH)]. The column temperature was maintained at 60°C using a homemade column oven. Peptides were separated with a binary buffer system of buffer A (0.1% formic acid (FA)) and buffer B (80% acetonitrile plus 0.1% FA), at a flow rate of 300 nl/min. Nano flow Liquid chromatography was performed with an EASY-nLC 1200 system (Thermo Fisher Scientific), which was directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source. For proteome measurements, 500 ng were loaded and eluted with a gradient starting at 5% buffer B and stepwise increased to 30% in 95 min, 60% in 5 min, and 95% in 5 min. The mass spectrometer was operated in Top15 data-dependent mode (DDA) with a full scan range of 300–1,650 m/z at 60,000 resolution with an automatic gain control (AGC) target of 3e6 and a maximum fill time of 20 ms. Precursor ions were isolated with a width of 1.4 m/z and fragmented by higher-energy collisional dissociation (HCD) (NCE 27%). Fragment scans were performed at a resolution of 15,000, an AGC of 1e5, and a maximum injection time of 28 ms. Dynamic exclusion was enabled and set to 30 s.

For K- ϵ -GG peptide samples, 2 μ l were loaded and eluted with a gradient starting at 3% buffer B and stepwise increased to 7% in 6 min, 20% in 49 min, 36% in 39 min, 45% in 10 min, and 95% in 4 min. The mass spectrometer was operated in Top12 data-dependent mode (DDA) with a full scan range of 250–1,350 m/z at 60,000 resolution with an automatic gain control (AGC) target of 3e6 and a maximum fill time of 20 ms. Precursor ions were isolated with a width of 1.4 m/z and fragmented by higher-energy collisional dissociation (HCD) (NCE 28%). Fragment scans were performed at a resolution of 30,000, an AGC of 1e5 and a maximum injection time of 110 ms. Dynamic exclusion was enabled and set to 15 s.

MS data processing

For BMDM data sets, raw files consisting of high-resolution MS/MS spectra were processed with MaxQuant (version 1.5.8.3) for feature detection and protein identification using the Andromeda search engine (Cox *et al*, 2011). Extracted peak lists were searched against the UniProtKB/Swiss-Prot *Mus musculus* database (October 2016) and a separate reverse decoy database to empirically assess the false discovery rate (FDR) using strict trypsin specificity allowing up to two missed cleavages. The minimum required peptide length was set to seven amino acids. The mass tolerance for precursor ions and fragment ions were 20 ppm and 0.5 Da, respectively. The search included variable modifications of oxidation (methionine), amino-terminal acetylation, carbamidomethyl (cysteine), GlyGly or ubiquitination (lysine), phosphorylation (serine, threonine, or tyrosine), and N-ethylmaleimide (cysteine). Raw MS data were also searched with PEAKS, version 8 (Bioinformatics Solutions) using a Swiss-Prot Human database and the same variable and fixed modifications as described above. A 0.1% and 1% FDR cutoff were applied at the PSM and peptide/protein levels, respectively.

Raw MS data from THP-1 cells were searched against the UniProt Human FASTA (21,051 sequences) using MaxQuant (version 1.6.2.10) with a 1% FDR at peptide and protein levels. The match and alignment time window for the match between run (MBR) algorithm were set to 0.7 min and 20 min, respectively. A ratio count of two was used for the MaxLFQ algorithm. Cysteine carbamidomethylation was defined as fixed, protein N-terminal acetylation and

methionine oxidation as variable modification. In case of K-ε-GG samples, “GlyGly (K)” was additionally selected as variable modifications. Enzyme specificity was set to trypsin and two missed cleavages were allowed, while permitting a maximum of five modifications per peptide.

Molecular modeling

The effects of K209R, I212D, and I212A mutations on the structural integrity were modeled using the DynaMut software (Rodrigues *et al*, 2018) using the kinase domain of RIPK2 (PDB: 4C8B) as the input.

Recombinant protein purification

pGEX-6 P-1 or pGEX-6 P-3 plasmids encoding XIAP-BIR2, Ubiquitin-UBA1x (UBA), or Ubiquitin-UBA4x (TUBE; Hjerpe *et al*, 2009) were transformed into BL21 (DE3) bacteria and grown in Super broth overnight at 37°C. Overnight culture was diluted 1:10 and grown until OD595 was 0.8. Isopropylthiogalactoside (IPTG) (0.3 mM) was added for 4 h at 30°C. Cells were pelleted and resuspended in Buffer A (50 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol, all from SIGMA) and sonicated. After centrifugation at 21,000 *g* for 30 min, the supernatant was incubated with glutathione sepharose4B (GE Healthcare) for 4 h, washed five times with Buffer A, and eluted 2 × 45 min with 10 mM reduced glutathione in Buffer A at 4°C.

Generation of doxycycline-inducible cell lines

Sequences of full length hsRIPK2 with an N-terminal 3x-Flag tag were synthesized by Genscript and cloned into doxycycline-inducible lentiviral expression vectors (pF_TRE3G_rTAAAd_puro (Takara Bio)). For lentiviral transfections, 2.5 µg of the plasmid of choice was transfected into HEK293T cells together with 1 µg pVSV-G and 1.5 µg pCMVΔR8.2 using an Effectene transfection kit (Qiagen). Twenty-four hours after transfection, the media was changed and virus was harvested after another 24 h. Media was filtered and supplemented with polybrene (4 µg/ml). Viral media was then applied to cell lines, centrifuged for 45 min at 1,000 *g* at 30°C. After 2 days of incubation, cells were selected using 2.5 µg/ml puromycin (Sigma).

Thermal shift assay

5 × 10⁶ THP-1 cells were treated with doxycycline (200 ng/ml, Sigma) for 5 h, washed in PBS, and resuspended in PBS supplemented with protease inhibitors. Cell suspensions were transferred into PCR tubes and incubated for 3 min at a temperature gradient (37–60°C) in a PCR machine. Samples were cooled to room temperature, lysed by repeated freeze-thawing, and centrifuged at 20,000 *g* for 20 min at 4°C. Supernatants were harvested and 2 × SDS sample buffer was added, before analysis by Western blot.

XIAP-BIR2 binding assay

20 × 10⁶ THP-1 cells were treated with doxycycline (200 ng/ml, Sigma) for 5 h, washed in PBS, and lysed in 1–2 ml IP buffer (150 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, 1% Triton X-100, 2 mM EDTA) with cOmplete protease inhibitor cocktail, Roche, and phosphatase

inhibitors (5 mM β-Glycerophosphate, 1 mM Sodium molybdate, 2 mM Sodium pyrophosphate, 10 mM Sodium fluoride) and 5 mM *n*-ethylmaleimide (NEM, Sigma). Lysates were clarified by centrifugation at 17,000 × *g* for 15 min and added directly to 20 µl packed glutathione sepharose beads pre-bound with 100 µg GST-XIAP-BIR2. Beads were incubated on a rotating wheel at 4°C for at least 2 h, washed three times with IP buffer, and eluted with 2 × SDS sample buffer.

Statistical analysis

The *P* values were calculated using two-way ANOVA using Prism v.8 (GraphPad). **P* ≤ 0.05, ***P* ≤ 0.01, ****P* ≤ 0.001, and *****P* ≤ 0.0001; *P* values > 0.05 are indicated as not significant (ns). Normal distribution was confirmed using the D’Agostino–Pearson test (Fig 3B), respectively, the Shapiro–Wilk test (Fig 4D) using GraphPad PRISM software.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol *et al*, 2019) with the dataset identifier PXD017741 (<http://www.ebi.ac.uk/pride/archive/projects/PXD017741>).

Expanded View for this article is available online.

Acknowledgements

We thank Prof. C. Day for the GST-BIR2 constructs and C. Gatt, K.McKenzie, T. Ballinger, and C.Epifanio for assistance with animal work and the MAGEC facility of WEHI for the generation of affinity-tagged RIPK2 mice. The work was supported by NHMRC grants 1046986, 1057888 and fellowships, 541901, 1058190 to JS and an ARC fellowship, FT130100166, to UN. This work was made possible through Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRISS (#9000220). This project has received funding from the European Union’s Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 754388 and from LMU Munich’s Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative (No. ZUK22).

Author contributions

VJH, JS, AIW, and UN developed the concept for this work. VJH, LFD, CAS, FMH, EC, AB, and ISL performed experimental work. AIW oversaw the mass spectrometry work and ISL the biophysical characterization of generated RIPK2 mutants. VJH, JS, and UN wrote and revised the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198–207
- Alshareef A, Zhang HF, Huang YH, Wu C, Zhang JD, Wang P, El-Sehemy A, Fares M, Lai R (2016) The use of cellular thermal shift assay (CETSA) to

- study Crizotinib resistance in ALK-expressing human cancers. *Sci Rep* 6: 33710
- Amar J, Chabo C, Waget A, Klopp P, Vachoux C, Bermudez-Humaran LG, Smirnova N, Berge M, Sulpice T, Lahtinen S et al (2011) Intestinal mucosal adherence and translocation of commensal bacteria at the early onset of type 2 diabetes: molecular mechanisms and probiotic treatment. *EMBO Mol Med* 3: 559–572
- Anand PK, Tait SW, Lamkanfi M, Amer AO, Nunez G, Pages G, Pouyssegur J, McCargill MA, Green DR, Kanneganti TD (2011) TLR2 and RIP2 pathways mediate autophagy of *Listeria monocytogenes* via extracellular signal-regulated kinase (ERK) activation. *J Biol Chem* 286: 42981–42991
- Ball KA, Johnson JR, Lewinski MK, Guatelli J, Verschuere E, Krogan NJ, Jacobson MP (2016) Non-degradative ubiquitination of protein kinases. *PLoS Comput Biol* 12: e1004898
- Bauler LD, Duckett CS, O’Riordan MX (2008) XIAP regulates cytosol-specific innate immunity to *Listeria* infection. *PLoS Pathog* 4: e1000142
- Bertrand MJM, Lippens S, Staes A, Gilbert B, Roelandt R, De Medts J, Gevaert K, Declercq W, Vandenabeele P (2011) cIAP1/2 are direct E3 ligases conjugating diverse types of ubiquitin chains to receptor interacting proteins kinases 1 to 4 (RIP1–4). *PLoS ONE* 6: e22356
- Canning P, Ruan Q, Schwerdt T, Hrdinka M, Maki JL, Saleh D, Suebsuwong C, Ray S, Brennan PE, Cuny GD et al (2015) Inflammatory signaling by NOD-RIPK2 is inhibited by clinically relevant type II kinase inhibitors. *Chem Biol* 22: 1174–1184
- Caruso R, Warner N, Inohara N, Nunez G (2014) NOD1 and NOD2: signaling, host defense, and inflammatory disease. *Immunity* 41: 898–908
- Cavallari JF, Fullerton MD, Duggan BM, Foley KP, Denou E, Smith BK, Desjardins EM, Henriksbo BD, Kim KJ, Tuinema BR et al (2017) Muramyl dipeptide-based postbiotics mitigate obesity-induced insulin resistance via Irf4. *Cell Metab* 25: 1063–1074 e1063
- Chin AI, Dempsey PW, Bruhn K, Miller JF, Xu Y, Cheng G (2002) Involvement of receptor-interacting protein 2 in innate and adaptive immune responses. *Nature* 416: 190–194
- Cooney R, Baker J, Brain O, Danis B, Pichulik T, Allan P, Ferguson DJ, Campbell BJ, Jewell D, Simmons A (2010) NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med* 16: 90–97
- Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10: 1794–1805
- Creixell P, Schoof EM, Simpson CD, Longden J, Miller CJ, Lou HJ, Perryman L, Cox TR, Zivanovic N, Palmeri A et al (2015) Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* 163: 202–217
- Damgaard RB, Nachbur U, Yabal M, Wong WW-L, Fiil BK, Kastirri M, Rieser E, Rickard JA, Bankovacki A, Peschel C et al (2012) The ubiquitin ligase XIAP recruits LUBAC for NOD2 signaling in inflammation and innate immunity. *Mol Cell* 46: 746–758
- Delconte RB, Kolesnik TB, Dagley LF, Rautela J, Shi W, Putz EM, Stannard K, Zhang JG, Teh C, Firth M et al (2016) CIS is a potent checkpoint in NK cell-mediated tumor immunity. *Nat Immunol* 17: 816–824
- Denou E, Lolmede K, Garidou L, Pomie C, Chabo C, Lau TC, Fullerton MD, Nigro G, Zakaroff-Girard A, Luche E et al (2015) Defective NOD2 peptidoglycan sensing promotes diet-induced inflammation, dysbiosis, and insulin resistance. *EMBO Mol Med* 7: 259–274
- Dorsch M, Wang A, Cheng H, Lu C, Bielecki A, Charron K, Clauser K, Ren H, Polakiewicz RD, Parsons T et al (2006) Identification of a regulatory autophosphorylation site in the serine–threonine kinase RIP2. *Cell Signal* 18: 2223–2229
- Ellwanger K, Briese S, Arnold C, Kienes I, Heim V, Nachbur U, Kufer TA (2019) XIAP controls RIPK2 signaling by preventing its deposition in speck-like structures. *Life Sci Alliance* 2: e201900346
- Fekete T, Koncz G, Szabo B, Gregus A, Rajnavolgyi E (2017) Interferon gamma boosts the nucleotide oligomerization domain 2-mediated signaling pathway in human dendritic cells in an X-linked inhibitor of apoptosis protein and mammalian target of rapamycin-dependent manner. *Cell Mol Immunol* 14: 380–391
- Fiil BK, Damgaard RB, Wagner SA, Keusekotten K, Fritsch M, Bekker-Jensen S, Mailand N, Choudhary C, Komander D, Gyrd-Hansen M (2013) OTULIN restricts Met1-linked ubiquitination to control innate immune signaling. *Mol Cell* 50: 818–830
- Girardin SE, Boneca IG, Carneiro LA, Antignac A, Jehanno M, Viala J, Tedin K, Taha MK, Labigne A, Zahringer U et al (2003a) Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* 300: 1584–1587
- Girardin SE, Travassos LH, Herve M, Blanot D, Boneca IG, Philpott DJ, Sansonetti PJ, Mengin-Lecreulx D (2003b) Peptidoglycan molecular requirements allowing detection by Nod1 and Nod2. *J Biol Chem* 278: 41702–41708
- Goncharov T, Hedayati S, Mulvihill MM, Izrael-Tomasevic A, Zobel K, Jeet S, Fedorova AV, Eidsencken C, deVoss J, Yu K et al (2018) Disruption of XIAP-RIP2 association blocks NOD2-mediated inflammatory signaling. *Mol Cell* 69: 551–565 e557
- Gong Q, Long Z, Zhong FL, Teo DET, Jin Y, Yin Z, Boo ZZ, Zhang Y, Zhang J, Yang R et al (2018) Structural basis of RIP2 activation and signaling. *Nat Commun* 9: 4993
- Hasegawa M, Fujimoto Y, Lucas PC, Nakano H, Fukase K, Núñez G, Inohara N (2008) A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF- κ B activation. *EMBO J* 27: 373–383
- Heim VJ, Stafford CA, Nachbur U (2019) NOD signaling and cell death. *Front Cell Dev Biol* 7: 208
- Hjerpe R, Aillet F, Lopitz-Otsoa F, Lang V, England P, Rodriguez MS (2009) Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep* 10: 1250–1258
- Home CR, Richmond AL, Rebert NA, Achkar JP, McDonald C (2010) ATG16L1 and NOD2 interact in an autophagy-dependent antibacterial pathway implicated in Crohn’s disease pathogenesis. *Gastroenterology* 139: 1630–1641, 1641 e1631–1632
- Hrdinka M, Schlicher L, Dai B, Pinkas DM, Bufton JC, Picaud S, Ward JA, Rogers C, Suebsuwong C, Nikhar S et al (2018) Small molecule inhibitors reveal an indispensable scaffolding role of RIPK2 in NOD2 signaling. *EMBO J* 37: e99372
- Inohara N, Koseki T, Lin J, del Peso L, Lucas PC, Chen FF, Ogura Y, Nunez G (2000) An induced proximity model for NF- κ B activation in the Nod1/RICK and RIP signaling pathways. *J Biol Chem* 275: 27823–27831
- Jafari R, Almqvist H, Axelsson H, Ignatushchenko M, Lundback T, Nordlund P, Martinez Molina D (2014) The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat Protoc* 9: 2100–2122
- Jeong YJ, Kang MJ, Lee SJ, Kim CH, Kim JC, Kim TH, Kim DJ, Kim D, Nunez G, Park JH (2014) Nod2 and Rip2 contribute to innate immune responses in mouse neutrophils. *Immunology* 143: 269–276
- Kedzierski L, Tate MD, Hsu AC, Kolesnik TB, Linossi EM, Dagley L, Dong Z, Freeman S, Infusini G, Starkey MR et al (2017) Suppressor of Cytokine Signaling (SOCS)5 ameliorates influenza infection via inhibition of EGFR signaling. *Elife* 6: e20444

- Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig PA, Reinecke M, Ruprecht B, Petzoldt S, Meng C et al (2017) The target landscape of clinical kinase drugs. *Science* 358: eaan4368
- Krieg A, Correa RG, Garrison JB, Le Negrato G, Welsh K, Huang Z, Knoefel WT, Reed JC (2009) XIAP mediates NOD signaling via interaction with RIP2. *Proc Natl Acad Sci USA* 106: 14524–14529
- Kueh AJ, Pal M, Tai L, Liao Y, Smyth GK, Shi W, Herold MJ (2017) An update on using CRISPR/Cas9 in the one-cell stage mouse embryo for generating complex mutant alleles. *Cell Death Differ* 24: 1821–1822
- Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* 11: 319–324
- Lee JY, Hwang EH, Kim DJ, Oh SM, Lee KB, Shin SJ, Park JH (2016) The role of nucleotide-binding oligomerization domain 1 during cytokine production by macrophages in response to *Mycobacterium tuberculosis* infection. *Immunobiology* 221: 70–75
- Lupfer C, Thomas PG, Anand PK, Vogel P, Milasta S, Martinez J, Huang G, Green M, Kundu M, Chi H et al (2013) Receptor interacting protein kinase 2-mediated mitophagy regulates inflammasome activation during virus infection. *Nat Immunol* 14: 480–488
- Maharana J, Sahoo BR, Bej A, Jena I, Parida A, Sahoo JR, Dehury B, Patra MC, Marthia SR, Balabantray S et al (2015) Structural models of zebrafish (*Danio rerio*) NOD1 and NOD2 NACHT domains suggest differential ATP binding orientations: insights from computational modeling, docking and molecular dynamics simulations. *PLoS ONE* 10: e0121415
- Martinez Molina D, Jafari R, Ignatushchenko M, Seki T, Larsson EA, Dan C, Sreekumar L, Cao Y, Nordlund P (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* 341: 84–87
- Martinez Molina D, Nordlund P (2016) The cellular thermal shift assay: a novel biophysical assay for *in situ* drug target engagement and mechanistic biomarker studies. *Annu Rev Pharmacol Toxicol* 56: 141–161
- Matsumoto ML, Dong KC, Yu C, Phu L, Gao X, Hannoush RN, Hymowitz SG, Kirkpatrick DS, Dixit VM, Kelley RF (2012) Engineering and structural characterization of a linear polyubiquitin-specific antibody. *J Mol Biol* 418: 134–144
- McCarthy JV, Ni J, Dixit VM (1998) RIP2 is a novel NF- κ B-activating and cell death-inducing kinase. *J Biol Chem* 273: 16968–16975
- von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417: 399–403
- Miller MH, Shehat MG, Alcedo KP, Spinel LP, Soualakova J, Tigno-Aranjuez JT (2018) Frontline Science: RIP2 promotes house dust mite-induced allergic airway inflammation. *J Leukoc Biol* 104: 447–459
- Nachbur U, Stafford CA, Bankovacki A, Zhan Y, Lindqvist LM, Fiil BK, Khakham Y, Ko H-J, Sandow JJ, Falk H et al (2015) A RIPK2 inhibitor delays NOD signalling events yet prevents inflammatory cytokine production. *Nat Commun* 6: 6442
- Nembrini C, Kisielow J, Shamshiev AT, Tortola L, Coyle AJ, Kopf M, Marsland BJ (2009) The kinase activity of Rip2 determines its stability and consequently Nod1- and Nod2-mediated immune responses. *J Biol Chem* 284: 19183–19188
- Newton K, Matsumoto ML, Wertz IE, Kirkpatrick DS, Lill JR, Tan J, Dugger D, Gordon N, Sidhu SS, Fellouse FA et al (2008) Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* 134: 668–678
- O'Donnell MA, Legarda-Addison D, Skountzou P, Yeh WC, Ting AT (2007) Ubiquitination of RIP1 regulates an NF- κ B-independent cell-death switch in TNF signaling. *Curr Biol* 17: 418–424
- Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Núñez G (2001) Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J Biol Chem* 276: 4812–4818
- Panda S, Gekara NO (2018) The deubiquitinase MYSM1 dampens NOD2-mediated inflammation and tissue damage by inactivating the RIP2 complex. *Nat Commun* 9: 4654
- Park J-H, Kim Y-G, McDonald C, Kanneganti T-D, Hasegawa M, Body-Malapel M, Inohara N, Núñez G (2007) RICK/RIP2 mediates innate immune responses induced through Nod1 and Nod2 but not TLRs. *J Immunol* 178: 2380–2386
- Pellegrini E, Desfosses A, Wallmann A, Schulze WM, Rehbein K, Mas P, Signor L, Gaudon S, Zenkeviciute G, Hons M et al (2018) RIP2 filament formation is required for NOD2 dependent NF- κ B signalling. *Nat Commun* 9: 4043
- Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M et al (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442–D450
- Petroski MD, Deshaies RJ (2003) Context of multiubiquitin chain attachment influences the rate of Sic1 degradation. *Mol Cell* 11: 1435–1444
- Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE (2014) NOD proteins: regulators of inflammation in health and disease. *Nat Rev Immunol* 14: 9–23
- Rahman MA, Sundaram K, Mitra S, Gavrillin MA, Wewers MD (2014) Receptor interacting protein-2 plays a critical role in human lung epithelial cells survival in response to Fas-induced cell-death. *PLoS ONE* 9: e92731
- Rodrigues CH, Pires DE, Ascher DB (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res* 46: W350–W355
- Ruprecht B, Zecha J, Heinzlmeir S, Medard G, Lemeer S, Kuster B (2015) Evaluation of kinase activity profiling using chemical proteomics. *ACS Chem Biol* 10: 2743–2752
- Schertzer JD, Tamrakar AK, Magalhaes JG, Pereira S, Bilan PJ, Fullerton MD, Liu Z, Steinberg GR, Giacca A, Philpott DJ et al (2011) NOD1 activators link innate immunity to insulin resistance. *Diabetes* 60: 2206–2215
- Schmid-Burgk JL, Schmidt T, Gaidt MM, Pelka K, Latz E, Ebert TS, Hornung V (2014) OutKnocker: a web tool for rapid and simple genotyping of designer nuclease edited cell lines. *Genome Res* 24: 1719–1723
- Seashore-Ludlow B, Axelsson H, Almqvist H, Dahlgren B, Jonsson M, Lundback T (2018) Quantitative interpretation of intracellular drug binding and kinetics using the cellular thermal shift assay. *Biochemistry* 57: 6715–6725
- Slany A, Bileck A, Kreutz D, Mayer RL, Muqaku B, Gerner C (2016) Contribution of human fibroblasts and endothelial cells to the hallmarks of inflammation as determined by proteome profiling. *Mol Cell Proteomics* 15: 1982–1997
- Stafford CA, Lawlor KE, Heim VJ, Bankovacki A, Bernardini JP, Silke J, Nachbur U (2018) IAPs Regulate distinct innate immune pathways to co-ordinate the response to bacterial peptidoglycans. *Cell Rep* 22: 1496–1508
- Thome M, Hofmann K, Burns K, Martinon F, Bodmer JL, Mattmann C, Tschopp J (1998) Identification of CARDIAC, a RIP-like kinase that associates with caspase-1. *Curr Biol* 8: 885–888
- Tigno-Aranjuez JT, Asara JM, Abbott DW (2010) Inhibition of RIP2's tyrosine kinase activity limits NOD2-driven cytokine responses. *Genes Dev* 24: 2666–2677

- Tigno-Aranjuez JT, Bai X, Abbott DW (2013) A discrete ubiquitin-mediated network regulates the strength of NOD2 signaling. *Mol Cell Biol* 33: 146–158
- Tigno-Aranjuez JT, Benderitter P, Rombouts F (2014) *In vivo* inhibition of RIPK2 kinase alleviates inflammatory disease. *J Biol Chem* 289: 29651–29664
- Udeshi ND, Mertins P, Svinkina T, Carr SA (2013) Large-scale identification of ubiquitination sites by mass spectrometry. *Nat Protoc* 8: 1950–1960
- Verhagen AM, Kratina TK, Hawkins CJ, Silke J, Ekert PG, Vaux DL (2007) Identification of mammalian mitochondrial proteins that interact with IAPs via N-terminal IAP binding motifs. *Cell Death Differ* 14: 348–357
- Wagner SA, Satpathy S, Beli P, Choudhary C (2016) SPATA2 links CYLD to the TNF-alpha receptor signaling complex and modulates the receptor signaling outcomes. *EMBO J* 35: 1868–1884
- Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6: 359–362
- Wu G, Chai J, Suber TL, Wu JW, Du C, Wang X, Shi Y (2000) Structural basis of IAP recognition by Smac/DIABLO. *Nature* 408: 1008–1012
- Wu G, Xu G, Schulman BA, Jeffrey PD, Harper JW, Pavletich NP (2003) Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. *Mol Cell* 11: 1445–1456
- Zou Y, Lei W, He Z, Li Z (2016) The role of NOD1 and NOD2 in host defense against chlamydial infection. *FEMS Microbiol Lett* 363: fnw170

3.4 Identification of covalent modifications regulating immune signaling complex composition and phenotype

Annika Frauenstein¹, Stefan Ebner¹, **Fynn M Hansen**², Ankit Sinha², Kshiti Phulphagar¹, Kirby Swatek³, Daniel Hornburg⁴, Matthias Mann² & Felix Meissner^{1,5}

¹Experimental Systems Immunology, Max Planck Institute of Biochemistry, Martinsried, Germany

²Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

³Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany

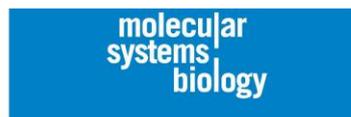
⁴Department of Genetics, School of Medicine, Stanford University, Stanford, CA, USA

⁵Institute of Innate Immunity, Department of Systems Immunology and Proteomics, Medical Faculty, University of Bonn, Bonn, Germany

Published in *Molecular Systems Biology* (2021)

Cellular functions rely on complex protein interaction networks that are governed by dynamic regulations through reversible interactions of distinct sets of proteins. PTMs play an important role in the orchestration of protein-protein interactions and the joint investigation of protein interactions, their modifications and phenotypes are important to decipher cellular signaling events. In this study, led by Dr. Annika Frauenstein, an analysis strategy devising modifications, interactions and phenotypes by affinity purification mass spectrometry (MIP-APMS) was developed to systematically dissect cellular signaling checkpoints. With the MIP-APMS pipeline we dissected the mode of action for MAPK14 inhibitors and the regulation of TRAF2 by dynamic phosphorylation and ISGylation.

In this collaboration, I aided in the identification of ISGylation sites on TRAF2. ISG15 is a ubiquitin-like protein that produces the same diGly remnants as ubiquitin after tryptic digestion, which we used here to our advantage. To detect low abundant ISGylation sites on TRAF2, we devised a workflow combining the MIP-APMS workflow with diGly peptide enrichment which revealed two modifications sites on TRAF2.



Identification of covalent modifications regulating immune signaling complex composition and phenotype

Annika Frauenstein¹, Stefan Ebner¹, Fynn M Hansen², Ankit Sinha², Kshiti Phulphagar¹, Kirby Swatek³, Daniel Hornburg⁴ , Matthias Mann² & Felix Meissner^{1,5,*}

Abstract

Cells signal through rearrangements of protein communities governed by covalent modifications and reversible interactions of distinct sets of proteins. A method that identifies those post-transcriptional modifications regulating signaling complex composition and functional phenotypes in one experimental setup would facilitate an efficient identification of novel molecular signaling checkpoints. Here, we devised modifications, interactions and phenotypes by affinity purification mass spectrometry (MIP-APMS), comprising the streamlined cloning and transduction of tagged proteins into functionalized reporter cells as well as affinity chromatography, followed by MS-based quantification. We report the time-resolved interplay of more than 50 previously undescribed modification and hundreds of protein–protein interactions of 19 immune protein complexes in monocytes. Validation of interdependencies between covalent, reversible, and functional protein complex regulations by knockout or site-specific mutation revealed ISGylation and phosphorylation of TRAF2 as well as ARHGEF18 interaction in Toll-like receptor 2 signaling. Moreover, we identify distinct mechanisms of action for small molecule inhibitors of p38 (MAPK14). Our method provides a fast and cost-effective pipeline for the molecular interrogation of protein communities in diverse biological systems and primary cells.

Keywords mass spectrometry; posttranslational modifications; protein interactions; proteomics; signaling networks

Subject Categories Immunology; Proteomics

DOI 10.15252/msb.202010125 | Received 16 November 2020 | Revised 8 July 2021 | Accepted 8 July 2021

Mol Syst Biol. (2021) 17: e10125

Introduction

Cellular functions rely on complex molecular networks that are mainly composed of proteins (Seet *et al.*, 2006; Pan *et al.*, 2012). Cell type- and context-specific functions require a tight orchestration of signaling, and their dysregulation is often associated with pathology (Arkin *et al.*, 2014). Experimental approaches that quantitatively capture the mechanisms of dynamic signaling networks are therefore highly valuable for establishing causal links to cellular phenotypes and the development of strategies for targeted interference.

Traditionally, the analysis of signal transduction mechanisms has focused on proteins with annotated functions in a given biological pathway. Pathway activation is probed with antibodies that determine the abundance of posttranslational modifications (PTMs) or interaction of selected proteins (protein–protein interactions, PPIs). Although valuable for testing pre-defined molecular states of selected proteins, the utility of this approach is limited by antibody availability, and prior knowledge of molecular and functional relationships. While employing antibodies would be applicable irrespective of the cell type, the discrimination of direct and indirect, as well as antibody-bound and bait-bound protein interactors, is often challenging because of limited antibody specificity (Marcon *et al.*, 2015). Conversely, while epitope tagging of selected proteins provides an alternative that guarantees specific enrichment with stable background binders—a defined set of proteins adhering to the affinity matrix—not all cell types are amenable to efficient genetic manipulations. An optimal strategy would therefore combine efficient and antibody-independent enrichment with universal applicability for eukaryotic cell types.

Mass spectrometry (MS)-based proteomics allows the detection of PTM and PPIs without prior knowledge. In recent years, MS-based proteomics has advanced tremendously and transitioned from identifying only a few proteins to comprehensively quantifying cellular proteomes and identifying modified proteins and protein interactions on a large scale (Larance & Lamond, 2015; Aebersold &

¹ Experimental Systems Immunology, Max Planck Institute of Biochemistry, Martinsried, Germany

² Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

³ Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany

⁴ Department of Genetics, School of Medicine, Stanford University, Stanford, CA, USA

⁵ Institute of Innate Immunity, Department of Systems Immunology and Proteomics, Medical Faculty, University of Bonn, Bonn, Germany

*Corresponding author (lead contact). Tel: +49 22828751210; E-mail: felix.meissner@uni-bonn.de

Mann, 2016). As such, it provides systems-wide views of cellular states with immense discovery potential, as indicated by large-scale efforts to map the entire interactomes in yeast (Gavin *et al.*, 2002; Ho *et al.*, 2002; Krogan *et al.*, 2006), drosophila (Guruharsha *et al.*, 2011), and human (Hein *et al.*, 2013; Hein *et al.*, 2015; Huttlin *et al.*, 2015), kinase and phosphatase interactomes (Gingras *et al.*, 2007; Couzens *et al.*, 2013; Yao *et al.*, 2017; Buljan *et al.*, 2020) as well as global views of specific PTMs (Choudhary *et al.*, 2009; Humphrey *et al.*, 2015; Lescarbeau *et al.*, 2016; Liu *et al.*, 2018).

Although it is well appreciated that the interplay of PTMs and PPIs determines how biological responses are regulated, MS-based technologies are almost always used to investigate PTMs and PPIs separately, and rely on distinct biochemical and analytical strategies. Hence, the analysis of PPIs is bait-centric, and selected proteins are affinity-enriched together with their interacting partners (Paul *et al.*, 2011). By contrast, PTM analysis generally focuses on a single modification type (e.g., phosphorylation), wherein modified peptides of all cellular proteins are affinity-enriched. Alternatively, in order to classify PTMs on specific proteins, affinity purification mass spectrometry (APMS) approaches with stringent washing and lysis conditions have been performed at the expense of PPI elucidation (Stutz *et al.*, 2017; Pankow *et al.*, 2019; Karayel *et al.*, 2020). Consequently, these two molecular modes of protein regulations are experimentally disconnected, hampering the discovery of the relationships between PTMs and PPIs in cellular signaling pathways. Furthermore, easy methods to simultaneously monitor different PTM types in a single sample are missing. Conventional enrichment strategies for distinct PTMs vary widely, and hence, mapping of multiple PTMs usually requires several sequential or parallel biochemical steps. This requires large amounts of starting material and results in low-sample throughput, while comprehensiveness is still limited as the enrichment strategies are tailored toward known biochemical properties of selected PTMs. A method that would capture in an unbiased manner all detectable PTMs in protein complexes of interest is therefore needed so as to comprehensively pinpoint molecular signaling checkpoints in complex biological systems.

The functional evaluation of emerging PTMs and PPIs is a common bottleneck in systems-wide discovery approaches. While initial screens are often performed in an experimental system that closely resembles cellular physiology, experimental validation of hits among all discovered candidates frequently relies on loss- or gain-of-function experiments in cell lines to achieve the necessary throughput. However, desirable would be an experimental setup that facilitates both discovery and validation in primary cells.

To develop a method for the systematic dissection of cellular signaling checkpoints by simultaneous PTM and PPI mapping in one experiment, we devised a streamlined pipeline—Modifications, Interactions and Phenotypes by APMS (MIP-APMS). We evaluated and technically optimized all steps of MIP-APMS, comprising (i) the epitope tagging of proteins of interest and mammalian cell transduction, (ii) affinity purification conditions for optimal interaction network and PPI enrichment, (iii) followed by MS-based PTM and PPI quantification and identification, and (iv) ultimate biochemical and phenotypic validation of interactors and PTMs in primary human immune cells. Integration of multiple MIP-APMS experiments generates dynamic signal transduction networks and pinpoints time-resolved co-regulations of PTMs and PPIs in

sequential signal transduction steps. We show the discovery potential of our pipeline by interrogating dynamically assembling protein communities in human monocyte immune signaling using Toll-like receptor (TLR) 2 activation and MAP kinase MAPK14 inhibition as paradigms. Our screen encompassing 19 protein complexes identified more than 50 previously undescribed PTMs, including phosphorylation, acetylation, methylation, ISGylation as well as other less well-described chemical modifications and elucidated an interaction network spanning more than 300 PPIs. We used the modular concept of MIP-APMS to test emerging data-driven hypotheses to validate PTMs and PPIs regulating immune signaling in reporter and primary cells. In this way, MIP-APMS enables the streamlined validation of crosstalk between different layers of protein regulation with broad applicability.

Results

Experimental and proteomics strategy for interrogating dynamic signal transduction networks

We devised MIP-APMS for the identification and perturbation of the functional checkpoints of cellular signaling pathways. MIP-APMS involves the following four stages with the indicated time frames (see Graphical Abstract, Figs 1 and EV1A):

- 1 Cloning of genes encoding epitope-tagged proteins and transduction of specialized cell types.
- 2 Streamlined quantification of various types of PTMs together with PPIs.
- 3 Implementation of an analytical strategy to pinpoint genetic or pharmacological signaling network perturbations.
- 4 Direct biochemical and functional evaluation of novel biological regulations in the same experimental system.

Universal cloning and transduction strategy

To enable interrogation of signaling cascades, we employed a cost-effective method for epitope tagging of proteins with a restriction enzyme-free approach, called restriction enzyme-free seamless ligation cloning extract (SLiCE) cloning (Zhang *et al.*, 2012). A modified weak phosphoglycerate kinase (PGK) promoter controls the expression of the GOIs, which are flanked by attL sites. Thereby, our vector system is compatible with commercial DNA assembly cloning strategies such as the NEBuilder platform or Gateway, which had been used before (Lambert *et al.*, 2014). As shown previously, employing lentiviral transduction for amphotrophic gene transfer extends the scope from readily transfectable cell lines, e.g., human embryonic kidney (HEK) cells, to non-dividing and terminally differentiated cells of primary origin (Huttlin *et al.*, 2015; Samavarchi-Tehrani *et al.*, 2018). In particular for application with primary immune cells, transduction is advantageous as other methods can activate innate immune signaling pathways and induce cell death (Fernandes-Alnemri *et al.*, 2009; Hornung *et al.*, 2009; Gaidt *et al.*, 2017). As a relevant and challenging experimental model system, we chose human monocytes, because these cells are not easily transfectable and execute a broad spectrum of cellular programs by the dynamic intracellular propagation of molecular signals downstream of cell surface receptors. For method development and

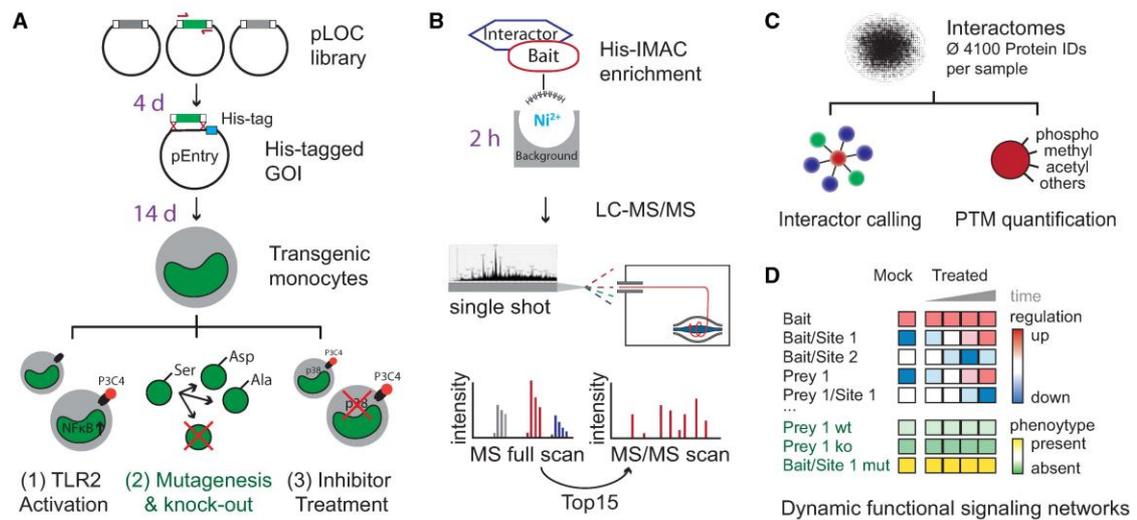


Figure 1. Experimental strategy for interrogating dynamic signal transduction networks with MIP-APMS.

A Generation of polyclonal transgenic cell lines by lentiviral transduction of genes encoding epitope-tagged wild-type or variant bait proteins. Analysis of PTMs and PPIs upon cellular activation (exemplified for the TLR1/2 activation by the agonist Pam3CSK4, P3C4), or pharmacological signal perturbation (exemplified by MAPK14 inhibitors). Time frames for the individual steps are indicated in violet.
 B Single-step His-IMAC affinity enrichment and single-run liquid chromatography–tandem mass spectrometry (LC-MS/MS).
 C Investigation of dynamic signal network topologies by simultaneous analysis of PPIs and multiple different PTMs. The numbers indicate analysis steps.
 D Schematic representation of PTM and PPI dynamics as a function of cellular phenotypes. Wt, wild type; ko, knockout; mut, mutation.
 Data information: See also Figs EV1–EV4.

phenotypic screening, we employed the monocytic cell line U937 and validated our results with primary cells. We achieved $92 (\pm 5)$ % cellular transduction efficiency after antibiotic marker selection (Fig EV1B). We further demonstrated the universality of our approach with primary human macrophages differentiated from peripheral monocytes (Fig EV1C and D, Table EV1) and primary human T cells (Fig EV1E, Table EV1).

We carefully characterized the functional properties of generated cell lines: The average copy numbers of the endogenous protein counterparts to the tagged proteins were 3.1 million per cell, increasing only slightly to an average of 4.3 million copies upon transduction (Fig EV2A). Importantly, global protein expression levels remained stable within cells upon expression of epitope-tagged bait proteins (Fig EV2B and C). We specifically confirmed that transduced cells exhibit no background immune activation by assessing expression levels of proteins involved in immune- and infection-associated pathways (Fig EV2D and E) and retain their full activation potential by assessing NF κ B activity with Luciferase reporter assays (Fig EV2F).

Simultaneous enrichment of PTMs and PPIs

Next, to study interdependency of PTMs and PPIs in signaling cascades, we evaluated biochemical enrichment strategies for epitope-tagged proteins with MS-based proteomics using high-performance liquid chromatography (HPLC) coupled to a linear quadrupole Orbitrap mass spectrometer (Q Exactive HF, Thermo)

operated in a data-dependent acquisition mode (Fig 1B) (Scheltema *et al.* 2014). We systematically compared typical short epitope tags: Flag-tag (Hopp *et al.* 1988), Strep-tag (Schmidt & Skerra, 2007), and polyhistidine tag (Hochuli *et al.* 1988). To quantitatively compare epitope tag-based enrichments, identification and label-free quantification (LFQ) were performed in the MaxQuant environment (Cox & Mann, 2008). While $> 1,000$ proteins were shared between all three enrichments (Fig EV3B), His-IMAC enrichment identified more background binding proteins. Exemplified for MAPK14, our results show high overlap of known interactors for Strep-tag and His-tag IPs with on-bead digestion, whereas Flag-tag and Strep-tag with elution yielded lower numbers of significant interactors (Fig EV3A, Table EV1). Notably, the highest median bait protein sequence coverage (Fig EV3C), highest intensity of MAPK14 (Fig EV3D), and highest number of significantly interacting proteins were achieved with His-IMAC.

Accordingly, we incorporated His-IMAC in the MIP-APMS protocol and further optimized the protocol for high bait enrichment and high-sequence coverage by titrating imidazole concentration in lysis and wash buffers, respectively (Fig EV3E and F). Following method optimization, the respective bait proteins were among the highest enriched proteins after MIP-APMS (Fig EV3G, Table EV1). We achieved a median sequence coverage of 70% for bait proteins (Fig EV3H), opening up the possibility of directly identifying and quantifying PTMs, such as phosphorylation, acetylation, or methylation as well as other less well-studied covalent protein modifications on

any bait protein (Fig 1C). Differently modified peptides were not analyzed separately, as in typical proteomics workflows, but instead the selected enriched proteins represented all present and detectable proteoforms. This made it possible to simultaneously quantify the differently modified and unmodified versions of peptides. MIP-APMS enables the efficient and cost-effective and robust analysis of PTMs and PPIs in a single experiment.

Dynamic signaling network analysis

To study how signaling networks rearrange upon cellular activation, we integrated quantitative PTM and PPI information from multiple MIP-APMS experiments. This enabled quantitative analysis of sequential steps in signal transduction, since it allowed for dynamic PTM and PPI crosstalk to be resolved providing a basis to identify molecular switches in signal transduction networks. We observed enrichments and de-enrichment of prey proteins in protein complex of interest and also dynamically regulated PTMs on both bait and prey proteins (Fig 1D, regulation up/down).

Biochemical and functional evaluation of novel biological regulations in the same experimental system

To validate our findings in follow-up studies, we employed the same experimental system used for discovery. We investigated the alterations in dynamic signaling networks of proteins mutated on single amino acid sites discovered in our study. Furthermore, by transforming our model system into NFκB reporter cells, we were able to reveal functional effects on NFκB activation of novel PTMs and PPIs by CRISPR-Cas9-mediated gene knockout and site-specific gene mutations, respectively (Fig 1D, phenotype, Fig 1A). As described below in more detail, we were able to derive functional molecular checkpoints in monocyte signal transduction networks.

Signaling networks of kinases, signaling adapters, and caspases in monocytes

We tested our MIP-APMS approach by interrogating the molecular composition of protein communities in mammalian cells *in situ*. Specifically, we investigated innate immune signaling complexes, assembled various protein classes, such as kinases, caspases, and tumor necrosis factor (TNF) receptor-associated factors (TRAFs) in human monocytes.

We generated 19 transgenic monocytic U937 cell lines and analyzed them with MIP-APMS, as described above. This identified and quantified an average of 4,106 proteins per measurement, including non-specifically binding proteins as expected for non-stringent APMS conditions (Trinkle-Mulcahy *et al*, 2008; Rees *et al*, 2011). We observed high median intra-bait and inter-bait Pearson correlations (> 0.9) between biological replicates (Fig EV3I) and between different cell lines (Fig EV3J). This highlights the overall reproducibility of the devised workflow. To discriminate specifically interacting proteins from background binders common to all baits, we compared enrichments from single vs. all other cell lines with a standard statistical test (two-sided *t*-test) at a stringent false discovery rate (FDR) of 1% to correct for multiple hypothesis testing (Hein *et al*, 2015; Keilhauer *et al*, 2015; Hubel *et al*, 2019). This resulted in a small fraction of significantly interacting proteins (378 proteins in total, with a median of 16 interactors per bait) compared to a large proportion of background binders (Table EV1, Fig EV4A). Notably, distinct protein intensity differences and *P*-values clearly distinguish specific bait and prey from unspecific background proteins (Fig EV4B and C). MIP-APMS prioritizes bait-specific preys, as proteins enriched in multiple experiments—including interconnected interactors—show lower enrichment differences and *P*-values (Fig EV4D) by unbiased statistical interactor calling (see Materials and Methods). We compared our LFQ intensity and *t*-test-based strategy to the results of the SAINT algorithm (spectral count based) exemplary for MAPK14 and identified largely similar interactors (Fig EV4E).

The identified interactors included previously described as well as novel proteins (Fig 2A, Table EV2). Unsupervised hierarchical clustering of label-free quantification (LFQ) intensity profiles of the significant interactors grouped specific interactors of corresponding bait proteins together (Fig EV4F). To determine the topology of the detected protein interaction network, we assembled proteins according to shared interactors. This enabled the identification of signaling hubs through common connections of bait and prey proteins that clustered together in the network (Fig 2B). The analysis recapitulated many known interactions, including the TRAF2–BIRC2 CORUM complex (Ruepp *et al*, 2010) involving the binary interaction of TRAF2 and BIRC2, supplemented by such players as TRAF1, TBK1, TANK, and IKKε (Wu *et al*, 2005). Some TRAF2 interactors, such as RIPK1, CASP8, and TNF (Hsu *et al*, 1996), were not detected in this experiment perhaps because they require distinct context-dependent cellular activation, e.g., through TNFR. These

Figure 2. Dissection of protein signaling networks in human monocytes using MIP-APMS.

- A Percentage of previously described interactors (green) and novel interactors (blue), and the count of significant interactors (FDR < 0.01, enrichment > 2) per bait protein (median interactor count: 16).
- B Protein–protein interaction network of clustered interaction data. Edges indicate interactions, with shared interactions connecting the individual MIP-APMS experiments. Red nodes correspond to bait proteins, green nodes to interactors reported in the literature, and blue nodes to novel interactors.
- C Numbers of acetylations, methylations, and phosphorylations identified on bait proteins and interactors.
- D Percentage of PTMs identified on bait proteins and interactors.
- E Numbers of PTMs on bait proteins/interactors of individual pull-downs.
- F Numbers of novel and described (Uniprot-annotated) acetylations, methylations, and phosphorylations.
- G Unsupervised clustering (Pearson correlation) of the z-scored intensity profiles of all PPIs (357) and PTMs (37) upon TLR2 activation, partitioned in seven clusters.
- H Dynamic profiles of co-regulated PTMs and PPIs with close network proximity, from the indicated clusters; median z-scored intensity of each time point (blue: median, gray: confidence interval = 0.95, method: loess); *n*, number of proteins in clusters 1–7. Selected proteins from each cluster are indicated, with the bait proteins in parentheses.

Data information: See also Fig EV3, Table EV2 for PPIs, and Table EV3 for PTMs.

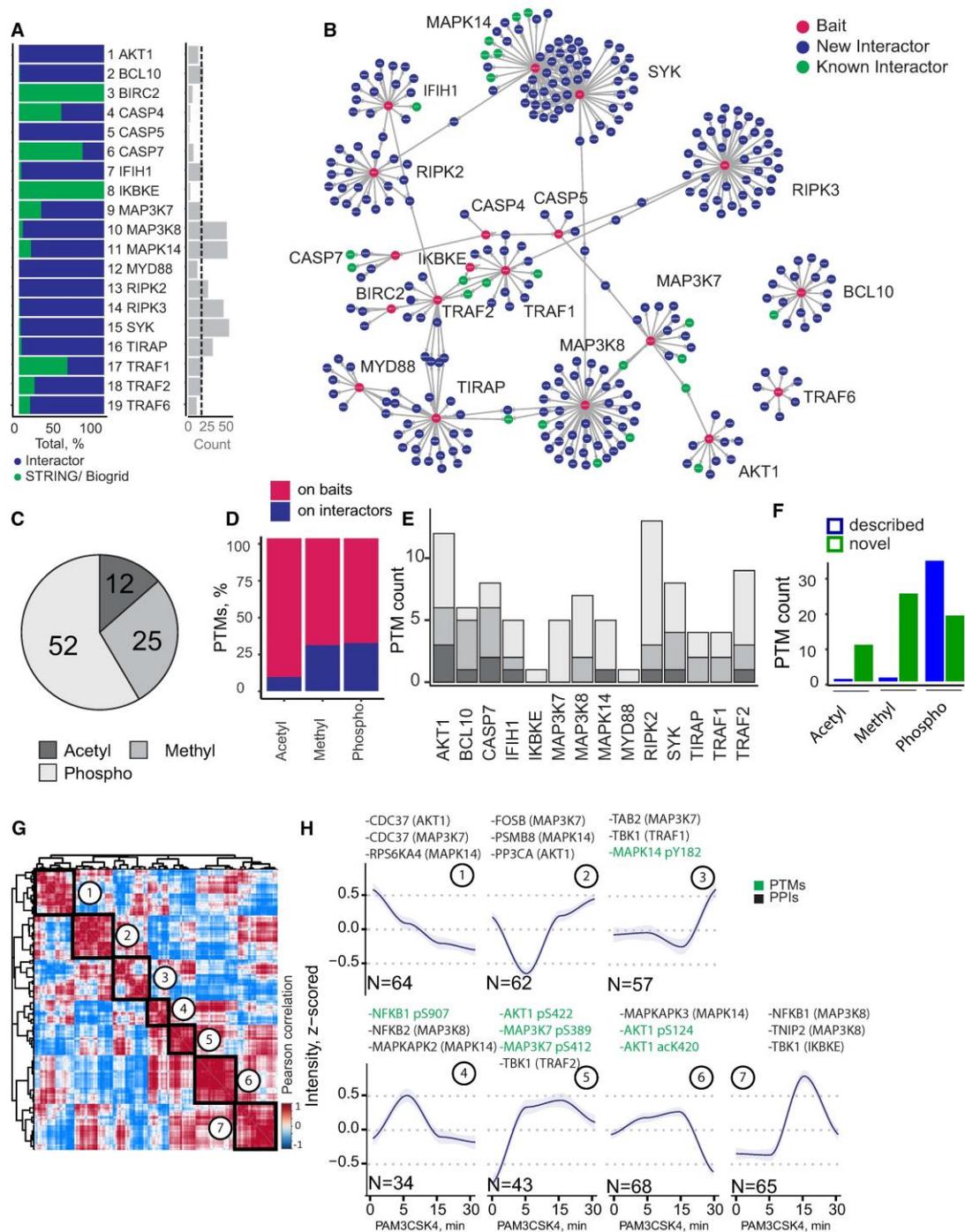


Figure 2.

observations validated the utility of MIP-APMS for the interrogation of intracellular signaling networks.

To identify and quantify PTMs in the same experimental setup, we re-analyzed our data with phosphorylation, acetylation, and methylation as variable posttranslational modifications. Even though we did not enrich for PTMs, we identified and quantified PTMs spanning phosphorylations, acetylations, and methylations on baits as well as prey proteins (88 PTMs on 19 bait proteins). Phosphorylation was the most abundant PTM in the dataset (52 sites), followed by methylation (25 sites) and acetylation (11 sites) (Fig 2C). While the majority of PTMs were detected on bait proteins, some (31 PTMs on 10 proteins) were also detected on prey proteins (26% of all known PTMs; Fig 2D). A remarkable 74% of the studied bait proteins or their respective interactors were posttranslationally modified, with some proteins, e.g., AKT1 and RIPK2, harboring more than 10 PTMs (Fig 2E). Notably, MIP-APMS identified 52 previously undescribed PTMs, in particular methylation and acetylation sites (Fig 2F). Furthermore, an unbiased analysis of covalent peptide modifications using the dependent peptide algorithm in MaxQuant, the string-based search algorithm Taggraph—based on a de novo search in PEAKS—and MS Fragger (Devabhaktuni et al, 2019)—revealed a series of less well-described covalent modifications on MAPK14 (Fig EV4G). Twenty-six modifications were shared between search engines (2.3% of all modifications for dependent peptides, 1.5% PEAKS/Taggraph, and 0.9% MS Fragger, Fig EV4H). Out of these 26 modifications, six were reproducibly identified and quantified in all replicates (Fig EV4I). To distinguish biologically regulated from other—for example—sample

preparation-introduced modifications, we quantified the identified modifications upon cell activation with specific searches in MaxQuant. Notably, only MAPK14 phosphorylation was differentially regulated between conditions. Moreover, acetylation, methylation, and phosphorylation detected on TRAF2, MAPK14, and MAP3K7 with specific searches were missed by open searches (dependent peptides of MaxQuant and PEAKS/Taggraph; Fig EV4J). This demonstrates that MIP-APMS can discover novel PTMs in signaling complexes; however, comparisons across search engines and confirmation with specific search strategies are advisable to increase confidence.

To capture the dynamics of cellular signal transduction, we next analyzed how the intracellular networks rearrange upon cellular activation via cell surface receptors. We stimulated cells via TLR2, as this pattern recognition receptor is prominently expressed in monocytes and induces a robust pro-inflammatory program that involves activation of the transcription nuclear factor (NF) κ B pathway (Oliveira-Nascimento et al, 2012; Rieckmann et al, 2017). We analyzed the dynamic signaling networks downstream of TLR2 using time course experiments in biological quadruplicates. Upon stimulation with the lipopeptide Pam3CysK (PAM3CSK4), cellular signaling was activated (Fig EV4K), and stable vs. dynamic PTMs and PPIs could be distinguished. Because of the short time frame of kinetic investigations (within 30-min post-cellular activation), we did not normalize protein levels to expression-induced protein abundance changes. On average, we detected two statistically significant dynamic PPIs and one dynamic PTMs per bait (Fig EV5A; Table EV2 and EV3). Our data suggest that phosphorylation is the most

Figure 3. N-Terminal phosphorylation of TRAF2 and ISG15 is dynamic functional regulators downstream of TLR2.

- A Volcano plot representing the interactome of TRAF2 (measured 15x in biological replicates) compared against all other pull-downs in the control group. The results of the t-tests are represented in volcano plots, which show the protein enrichment versus the significance of the enrichment. Numbers indicate enrichment ranks with the heatmap labels of (C) serving as the legend. Significant interactors of TRAF2 (two-tailed t-test, FDR < 0.01, enrichment > 4) are colored in blue (novel interactors) and green (known interactors).
- B Interactors of TRAF2 (blue: novel interactors, green: known interactors) with interconnecting proteins between different baits colored in gray.
- C Hierarchical clustering of significant interactors of TRAF2 upon activation with significant hits in at least one time point denoted with an asterisk. Cell activation was performed for 5, 15, and 30 min with the TLR2 ligand PAM3CSK4 (P3C4).
- D Intensity profile of the TRAF2 interactor TANK upon activation, normalized to TRAF2 bait LFQ intensity.
- E Hierarchical clustering of the TRAF2 PTMs (acetylation, methylation, and phosphorylation) upon activation, with significant hits (t-test) in at least one time point denoted with an asterisk.
- F Intensity profile of the phosphorylation of TRAF2 on Thr7 upon activation, normalized to TRAF2 bait intensity. Central band of the boxplot shows the median, boxes represent the IQR, 3 biological replicates were performed for UT, and 4 biological replicates were performed for additional time points. P-values were calculated by t-test. Asterisks indicate significant differences. *P-value < 0.05.
- G Intensity profile of TRAF2 interactors ISG15 and TRAF1 in different TRAF2 phospho-variants, normalized to TRAF2 wild-type intensities. Central band of the boxplot shows the median, boxes represent the IQR, and 4 biological replicates were performed for every condition. P-values were calculated by t-test. Asterisks indicate significant differences **P-value < 0.01, ***P-value < 0.001.
- H Induction of NF κ B determined based on luciferase luminescence in TLR2-activated U937 NF κ B reporter cells transfected with genes encoding different TRAF2 phospho-variants. Bar represents the median, error bars represent the standard deviation, and 4 biological replicates were performed for additional time points. P-values were calculated by t-test. Asterisks indicate significant differences. ***P-value < 0.001.
- I MS/MS Spectrum containing GlyGly modification K320 on TRAF2 after GlyGly enrichment on TRAF2 MIP-APMS.
- J Differences and P-values of ISG15 intensity in TRAF2 K->R mutants compared against TRAF WT
- K Intensity profile of TRAF2 interactors ISG15 and TANK in TRAF2 K->R mutants, normalized to TRAF2 wild-type intensities. Central band of the boxplot shows the median, boxes represent the IQR, and four biological replicates were performed for every condition. P-values were calculated by t-test. Asterisks indicate significant differences. ***P-value < 0.001.
- L Induction of NF κ B determined based on luciferase luminescence in TLR2-activated U937 NF κ B reporter cells transfected with genes encoding different TRAF2 K->R mutants (each bar represents a mean from three independent measurements; error bars represent the standard deviation; ***P-value < 0.001).
- M Intensity profile of TRAF2 interactors ISG15 and TANK in TRAF2-K389R and S11D mutants in human primary macrophages. Central band of the boxplot shows the median, boxes represent the IQR, and three biological replicates were performed for every condition. P-values were calculated by t-test. Asterisks indicate significant differences **P-value < 0.01, *P-value < 0.05, ***P-value < 0.001.
- Data information: Experiments in (A–L) were performed in U937 cell lines. Gray boxes indicate missing values. IQR stands for interquartile range and represents the 25th to 75th percentile. See also Table EV1–EV4.

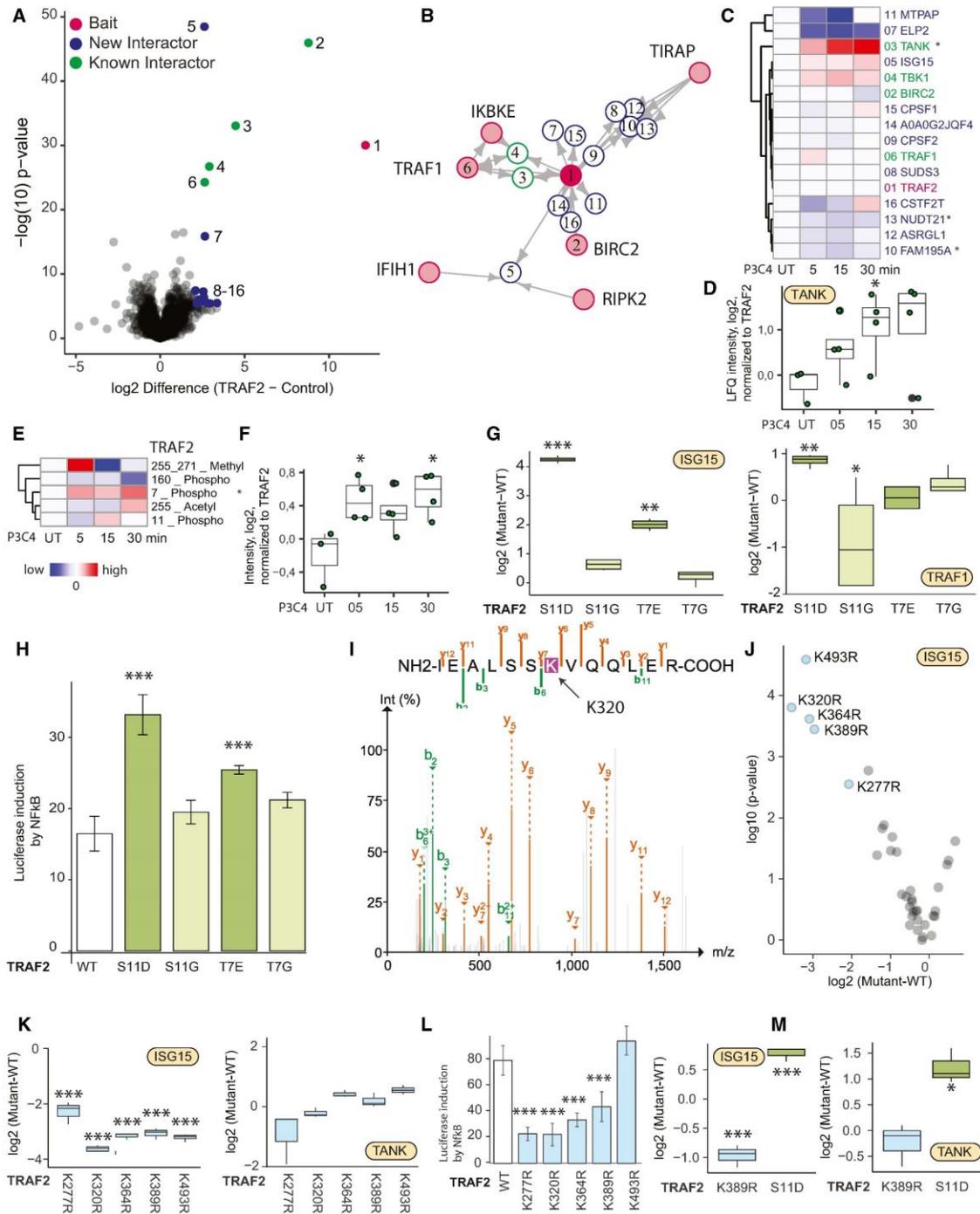


Figure 3.

dynamic PTM in the tested setting, followed by methylation (Fig EV5B).

Next, to study PTM and PPI interdependency, we correlated all PTM and PPI intensities and clustered them unbiasedly over the time course of TLR2 activation (Fig 2G, Table EV3). We detect the dynamic co-regulation on both molecular layers (PPIs and PTMs), identifying correlating and anti-correlating PTMs and PPIs during signaling pathway activation. We identified seven clusters with distinct kinetics, some peaking early (Fig 2H, Cluster: 4,5) and others late (Fig 2H, Cluster: 7) upon pathway activation, as well as up- (Fig 2H, Cluster: 4,5,6,1) vs. down-regulated (Fig 2H, Cluster: 1,2,3) PTMs and PPIs. Interestingly, interactors identified in more than one MIP-APMS experiment (e.g., CDC37: Cluster 1) were in close network proximity. Our approach facilitated an unbiased discovery of time-resolved molecular connections between dynamic PTMs and PPIs, exemplified by the correlated interaction of MAP3K8 interactors (NFKB1, NFKB2) and NFKB1 phosphorylation (Cluster 4,7), or the anti-correlated phosphorylation of the C-terminal kinase domain of AKT1 and the interaction with CDC37 (Clusters 1, 5). This demonstrates that the sensitivity and robustness of MIP-APMS enable the simultaneous determination of cellular signaling network rearrangements by PPIs, PTMs, and their interplay. We conclude that MIP-APMS is sufficiently sensitive and robust to capture dynamic signaling networks in mammalian cells *in situ*. It generates highly reproducible data that may be used for the discovery of novel dynamic PTMs in signal transduction cascades, and simultaneous evaluation of multiple PTMs and PPIs in signaling networks.

Dynamic phosphorylations and ISGylations regulate TRAF2 downstream of TLR2

We next evaluated MIP-APMS for the discovery of novel molecular checkpoints in intracellular immune signaling. We focused on significantly regulated PPIs (FDR < 0.01) and PTMs (*P*-value < 0.05) identified for TRAF2 and MAP3K7, and examined their biochemical and phenotypic relevance through network perturbations mediated by gain- and loss-of-function mutations.

TRAF2 is a central adaptor protein in TNF signaling and regulates pro-inflammatory cytokine production through NFκB and JNK signaling pathways (Borghi *et al*, 2016). As described above, the MIP-APMS analysis confirmed previously reported TRAF2 interactors, such as TNF receptor-associated factor TRAF1, baculoviral IAP repeat-containing protein BIRC2 (cIAP2), TRAF family member-associated NFκB activator TANK, and serine/threonine-protein kinase TBK1. In addition, we identified ELP2 and ISG15 as novel components of the TRAF2 complex (Fig 3A and B) and TANK, a negative regulator of TRAF2 (Cheng & Baltimore, 1996), as dynamically recruited to the TRAF2 complex. By contrast, the majority of other TRAF2 interactors remained unchanged upon activation (Fig 3C and D). While most other PTMs remained unchanged upon signal pathway activation, the analysis revealed dynamic N-terminal phosphorylations on Thr7 and Ser11 of TRAF2 (Fig 3E and F). Thus, the interactome and PTMs of TRAF2 are dynamically regulated upon NFκB activation via TLR2.

To test whether these dynamic N-terminal phosphorylations affected the composition and function of the TRAF2 protein complex, we used the MIP-APMS streamlined workflow to generate

protein phospho-variants, in which specific Ser or Thr residues were changed to Gly, or to Asp/Glu to mimic phosphorylation. We probed the resulting signaling network rearrangements using MIP-APMS and found a specific enrichment for ubiquitin-like protein ISG15 and TRAF1 by the phospho-mimetic TRAF2 variants compared to wild-type TRAF2 (Fig 3G; Table EV4). These data suggest that N-terminal phosphorylation of TRAF2 at both Thr7 and Ser11 stabilizes a protein complex with ISG15 and TRAF1.

To further assess the functional relevance of the N-terminal TRAF2 phosphorylation on cellular regulation, we introduced the phospho-mimetic and phospho-dead TRAF2 variants into NFκB reporter monocyte cell lines. TLR2-induced NFκB activation was elevated with TRAF2 N-terminal phospho-mimetics, whereas the phospho-dead variants showed activation comparable to that of wild-type TRAF2, indicating that N-terminal phosphorylation boosted downstream signal transduction (Fig 3H).

ISG15 is a ubiquitin-like protein that covalently modifies target proteins on lysine residues in a process called ISGylation (Loeb & Haas, 1992; Zhang & Zhang, 2011). After tryptic digest, isgylated peptides harbor GlyGly modifications on lysines that can be readily detected by LC-MS/MS. As we did not directly detect GlyGly-modified peptides, we combined MIP-APMS with GlyGly enrichment and indeed identified two GlyGly modification sites on TRAF2 (Positions K27, K320; Fig 3I). To deduce the impact of ISGylation on the TRAF2 interaction network, we performed site-directed mutagenesis of TRAF2 lysines and subjected the K→R mutant cell lines to MIP-APMS. Out of the total 32 K→R mutants, 5 showed strong (more than 4×) and significant depletion of ISG15 in the TRAF2 complex (Fig 3J, Table EV4). Interestingly, the most regulated site—K320—was also identified by our initial GlyGly enrichment, suggesting an ISGylation of TRAF2. Reduced ISG15 levels in the interactomes of certain K→R mutants further support this observation. In contrast, TANK levels—a TRAF2 complex member—remained unaltered in the different TRAF2 mutants, pointing toward a specific partial perturbation of the TRAF2 protein community by K→R site-directed mutagenesis (Fig 3K). We excluded potential clonal or TRAF2 mutant expression effects on ISG15 levels by comparing ISG15 levels of transgenic monocyte interactomes to full proteomes (Fig EV5D). Unchanged ISG15 intensities upon stringent MIP-APMS conditions (6 M GdmCl) in a TRAF2 MIP-APMS experiment as well as no evident interaction of recombinant ISG15 and TRAF2 in a size exclusion-based binding assay further support the covalent ISGylation of TRAF2 (Fig EV5E). Functional analysis of the K→R mutants revealed reduction in NFκB activation for K277R, K320R, K364R, and K389R mutants, suggesting that ISGylation of TRAF2 may act as a positive regulator downstream of TLR2 (Fig 3L). To expand our findings to primary human macrophages, we selected the novel phospho-mimetic TRAF2 mutant S11D and lysine-mutant K389R. These experiments confirm ISG15 enrichment in the TRAF2-S11D complex and depletion in the TRAF2-K389R complex (Fig 3M).

ARHGEF18 and FOSB are functional regulators downstream of TLR2

To further explore the utility of MIP-APMS for discovery of new interactors, we evaluated functional interactions of MAP3K7. MAP3K7 (TAK1) is a central kinase of the MAPK signaling pathway, with crucial roles in the activation of TRAF6 downstream of TLRs

and other receptors (B-cell receptor, TNF receptor) (Landström, 2010) and known as a major regulator of NF κ B signaling (Sato et al, 2005). The MIP-APMS analysis recapitulated the TNF α /NF κ B signaling complex 7 (CORUM) consisting of TAB1, TAB2, TAB3, and CDC37 (Fig 4A and B). Upon TLR2 activation, TAB1 and SNX17 were depleted from the MAP3K7 complex (Fig 4C), while phosphorylation of MAP3K7 on Ser389 increased, significantly (Fig 4D). This revealed dynamic regulation of both PTMs and PPIs during pathway execution.

From the nine previously unknown interactors, we selected the guanine nucleotide exchange factor ARHGEF18 and the transcriptional

regulators FOSB and FOXK2 for functional hypothesis testing. Because MAP3K7 is implicated in NF κ B activation, we used CRISPR to knock out the respective genes in monocytic NF κ B reporter cells and determined the pathway activity by luciferase induction that directly correlates with the activation of NF κ B (Fig 4E). Upon deletion of genes encoding TLR2 and MYD88 (the receptor and proximal adaptor of PAM3CSK4, respectively (Li et al, 2010)), we observed an almost complete inhibition of NF κ B activation. CRISPR knockout of MAP3K7, and the interactors ARHGEF18 and FOSB, led to a partial reduction of NF κ B activation, thereby linking this PPI to a functional downstream phenotype in the signaling cascade (all

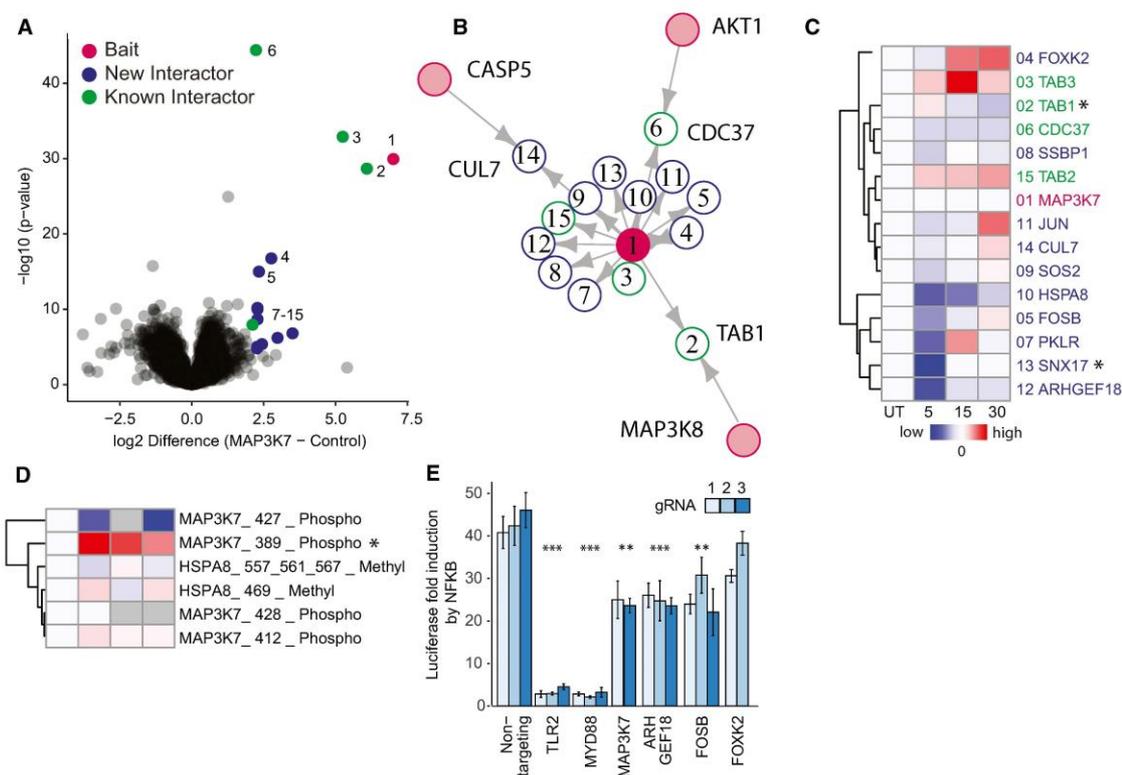


Figure 4. ARHGEF18 and FOSB are functional regulators downstream of TLR2.

- A** Volcano plot representing the interactome of MAP3K7 (measured 16x in biological replicates) compared against all other pull-downs in the control group. The results of the t -tests are represented in volcano plots, which show the protein enrichment versus the significance of the enrichment. Numbers indicate enrichment ranks with the heatmap labels of (C) serving as the legend. Significant interactors of MAP3K7 (two-tailed t -test, FDR < 0.01, enrichment > 4) are colored in blue (novel interactors) and green (known interactors).
- B** Interactors of MAP3K7 (blue: novel interactors, green: known interactors) with interconnecting proteins between different baits colored in gray.
- C** Heatmap of significant interactors of MAP3K7 upon activation, with significant hits in at least one time point (t -test, P -value < 0.05) denoted with an asterisk. Cell activation was performed for 5, 15, and 30 min with the TLR2 ligand PAM3CSK4 (P3C4).
- D** Heatmap of MAP3K7 PTMs (phosphorylation) upon activation, with significant hits (t -test, P -value < 0.05) in at least one time point denoted with an asterisk.
- E** Induction of NF κ B determined based on luciferase luminescence in U937 NF κ B reporter cells with CRISPR-Cas9 knockouts of the potential novel interactors of MAP3K7 upon TLR2 activation (each bar represents a mean of four independent measurements; error bars represent the standard deviation; P -values were calculated by t -test. Asterisks indicate significant differences. *** P -value < 0.001, ** P -value < 0.01).

Data information: Gray boxes indicate missing values. See also Appendix Figs S1–S14, Tables EV1 and EV2.

PTMs and PPIs of the characterized bait proteins are available in Appendix Figs S1–S14). We verified CRISPR-KO of ARHGEF18 and FOSB by Western blot analysis (Fig EV5F).

Hence, our MIP-APMS strategy can interrogate the functional relevance of individual molecular switches in a streamlined manner on the levels of PTMs as well as PPIs in signal transduction networks.

Dissecting drug mode of action for MAPK14 inhibitors with MIP-APMS

Small molecules are often used to interfere with specific cellular functions and are the mainstay of the drug industry. Definition of the target engagement of small molecules is a major challenge in drug discovery and novel proteomics approaches have been devised for this purpose (Schirle *et al.*, 2012). We reasoned that MIP-APMS could enable the identification of signaling network rearrangements induced by small molecules, providing a unique proteomic perspective on the mode of drug action. We selected previously described pharmacological inhibitors of p38 kinase (MAPK14) (JX-401 (Friedmann *et al.*, 2006), sorafenib (Edwards & Emens, 2010), and skepinone-L (Koeberle *et al.*, 2011)) and analyzed their mode of perturbation of the cellular signaling network assemblies involving MAPK14 (Fig 5A–C).

The obtained data indicated that skepinone-L and sorafenib interfered with the physiological intracellular signaling network of MAPK14 to a greater extent than JX-401 (Fig 5D; Table EV5). Further, interestingly, sorafenib and skepinone-L perturbed the interactions within the core complexes differently. While MAPKAPK5, a downstream substrate of MAPK14 (New *et al.*, 1998), was depleted in the MAPK14 protein complex upon treatment with both sorafenib and skepinone-L (Fig 5D), only sorafenib reduced the binding of RPS6KA4 (MSK1) and PTPN7 to MAPK14, and even more so upon cellular activation with TLR2 ligands (Fig 5E). Further, both sorafenib and skepinone-L induced hyper-phosphorylation of the MAPK14 phospho-loop on Tyr182, whereas an N-terminal phosphorylation site (Ser2) remained unaltered (Fig 5B and F). This indicated that PTMs and PPIs of MAPK14 are altered upon inhibitor treatment.

Further, MIP-APMS also allowed testing of drug off-target effects (Fig EV5G–I). MAP3K7 phosphorylation on Ser367, Ser412, and Ser445 was significantly altered, and both JUN and TAB2 were depleted from the MAPK14 complex upon treatment with sorafenib. This suggests that the MAP3K7 protein complex, reported to be an upstream activator of MAPK14 (Martín-Blanco, 2000), is in part targeted by MAPK14-specific inhibitors. Enrichment of ELP2 (JX-401, Skepinone-L) and TBK1 (JX-401) was observed in the TRAF2 signaling complex. PTMs on TRAF2 were not affected by the inhibitor treatment. Hence, MIP-APMS can be used to dynamically resolve the interactome and PTM changes upon small molecule treatment and provides information on molecular relationships in signal transduction networks that facilitate understanding of drug mode of action.

Discussion

Cellular processes are orchestrated by signal transduction pathways that depend on PTMs and PPIs. However, how PTMs and PPIs

collaborate in structuring the dynamic signaling network topologies remains incompletely understood, in part because of the laborious experimental approaches involved in dissecting these interactions. Here, we describe MIP-APMS, a combined streamlined cell line generation and proteomics approach to interrogate functional signal transduction networks in intracellular signaling pathways. We quantified more than 370 PPIs and 80 PTMs across innate protein signaling cascades in human monocytes upon receptor activation or drug treatment. Among these are 50 previously undescribed PTMs, including those for which specific enrichment methods are less streamlined, such as ISGylation. Our approach revealed biochemical connections between PTMs and PPIs, as well as protein subnetworks that regulate cellular programs dependent on site-specific PTMs.

We employed MIP-APMS for streamlined and selected interference with protein subnetworks. Demonstrating this principle for the site-specific manipulation of protein phosphorylation as well as ISGylation on TRAF2 yielded differential interactomes of mutated proteins as well as altered cellular physiology. In this way, structural insights into interaction interfaces between protein complexes and crucial PTMs for stabilizing interacting proteins can be revealed. To our knowledge, this is the first description of protein ISGylation augmenting NF κ B activity. We disturbed protein interaction networks of the kinase MAPK14 with small molecules to shed additional light on the drug mode of action of kinase inhibitors. Both skepinone-L and sorafenib changed the protein interaction network by a different mode of action, whereas both inhibitors lead to phosphorylation of the MAPK14 phospho-loop.

Modifications, interactions and phenotypes-APMS experiments with temporal resolution further allow the elucidation of co-regulations at different biochemical layers—adding to our understanding of molecular connections along the sequential steps of signal transduction. By further increasing temporal resolution, it may become possible to resolve the causalities between regulation on the PTM and PPIs levels in even greater detail.

For epitope tagging, we employed constructs from the pLOC library (GE Healthcare); however, other cDNA libraries or gene synthesis can readily be employed with polymerase chain reaction (PCR) to obtain DNA fragments with respective homologous overhangs. We employed the cost-effective, non-commercial SLICE cloning strategy; however, commercial solutions using NEBuilder or Gateway are possible with our vector system. An advantage of the small peptide tag chosen for the enrichment strategy in the current study is that it results in little steric interference with physiological protein–protein interactions.

According to our evaluation, Strep-tag and His-tag-based enrichments resulted in high bait sequence coverage; however, His-tag captured known interaction partners most comprehensive. By optimizing a non-stringent lysis procedure with low detergent and salt concentrations and also low temperature in the MIP-APMS protocol, we aimed to capture PTMs together with stable as well as transient interactions. According to our analysis, on average 12.3% of the PTMs and 5.5% of the PPIs are dynamic; however, as biochemical procedures impact recovery of interactors and different thresholds for significance calling are employed, comparability of PTM and PPI dynamics across studies remains challenging. Incorporating chemical cross-linking approaches (Holding, 2015; Liu & Heck, 2015)

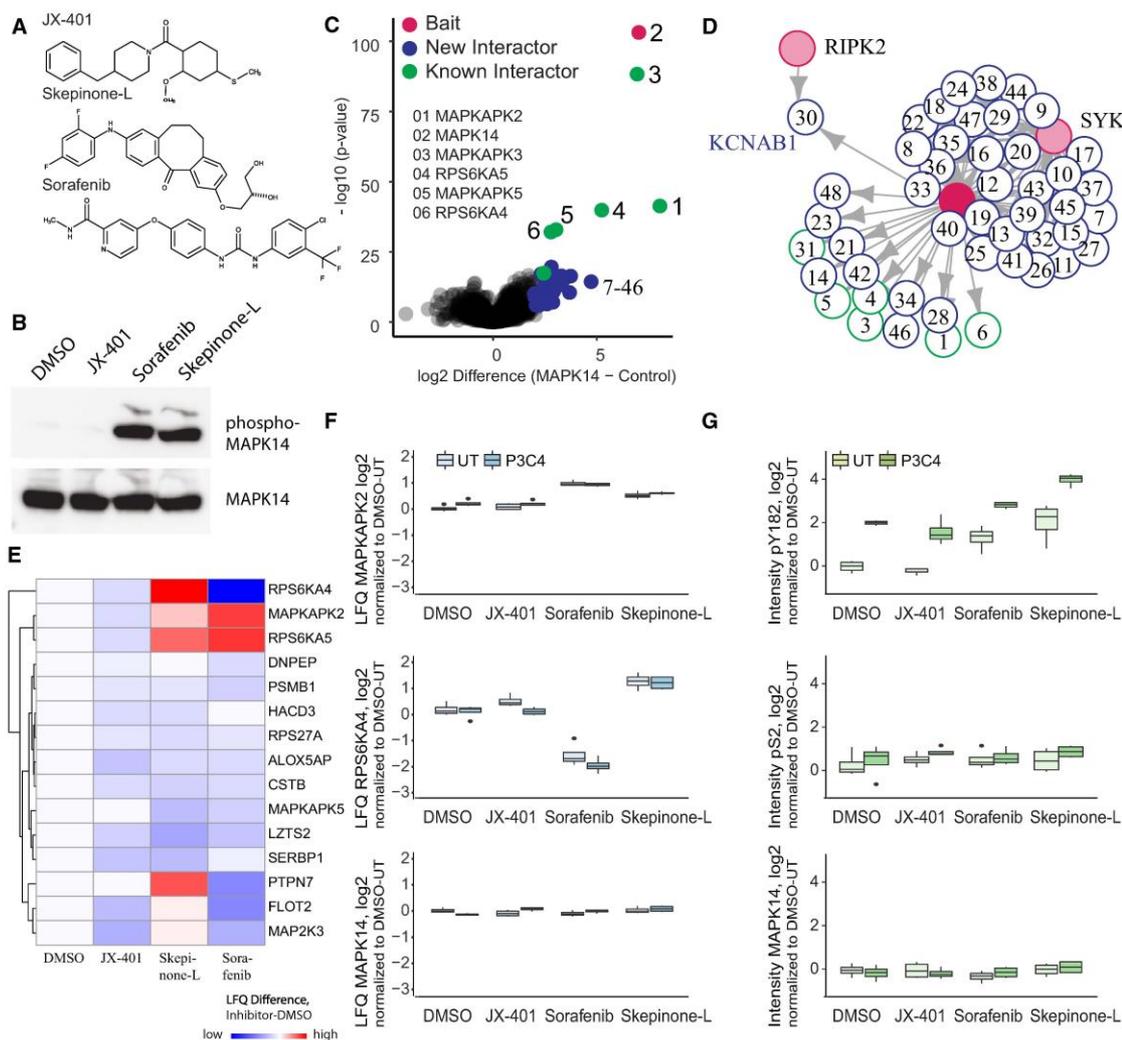


Figure 5. Dissecting drug mode of action for MAPK14 inhibitors with MIP-APMS.

- A Chemical structures of MAPK14 inhibitors JX-401, skepinone-L, and sorafenib.
- B Phosphorylation of MAPK14 in U937 WT after treatment with the inhibitors, analyzed by Western blotting using an alpha-phospho-MAPK14 antibody. Total MAPK14, detected by alpha-MAPK14 antibody, was used as a loading control.
- C The interactome of MAPK14 compared against all other pull-downs in the control group. The results of the t-tests are represented in volcano plots, which show the protein enrichment versus the significance of the enrichment. Numbers indicate enrichment ranks with the heatmap labels of (C) serving as the legend. Only the top interactors of MAPK14 are numbered. The complete list can be found in Table EV1.
- D Interactors of MAPK14 (blue: novel interactors, green: known interactors) with interconnecting proteins between different baits colored in gray.
- E Heatmap of MAPK14 interactors significantly altered upon treatment with the different MAPK14 inhibitors, with significant hits in at least one treatment (t-test, P -value < 0.05) denoted with an asterisk. Treatments were normalized to DMSO control. The complete list can be found in Table EV5.
- F LFQ intensity profiles of the MAPK14 interactors RPS6KA4 and MAPKAPK2 and MAPK14 after treatment with different MAPK14 inhibitors, normalized to MAPK14 bait intensity. Drug mode of action was analyzed in the presence (P3C4, 0.5 μ g/ml, 30 min) or absence of P3C4 after inhibitor treatment. Central band of the boxplot shows the median, boxes represent the IQR, and 4 biological replicates were performed for every condition.
- G Intensity profiles of MAPK14 phosphorylation on positions Ser2 and Tyr182 and MAPK14 protein intensity after treatment with different MAPK14 inhibitors, normalized to MAPK14 bait intensity. Drug mode of action was analyzed in the presence (P3C4, 0.5 μ g/ml, 30 min) or absence of P3C4 after inhibitor treatment. Central band of the boxplot shows the median, boxes represent the IQR, and 4 biological replicates were performed for every condition.
- Data information: Gray boxes indicate missing values. Bars represent median, error bars s.d. See also Fig EV5, Table EV5.

could further stabilize transient interactors. Combining MIP-APMS with structural information of the bait protein as well as its interactome can potentially reveal distinct interaction interfaces of protein complexes that are perturbed by site-specific covalent modification or drug action. It would be particularly interesting to integrate protein cross-linking with the PTM status, e.g., of wild-type vs. mutant variant proteins of interest, so that altered structural interaction interfaces can be resolved in addition to differential PTM-dependent PPIs.

We have already explored the strategy of disturbing protein interaction networks using small molecules to determine the effect of drugs on protein complexes. Whereas other proteomics approaches are aimed at identifying drug targets (Molina *et al.*, 2013), MIP-APMS elucidates changes in protein communities involving the selected targets. It may thus serve as an additional drug discovery tool to resolve target protein network properties or off-target effects.

Despite its advantages, MIP-APMS currently has some limitations, which can be addressed by developing the method further in the future. These include the possibility that bait protein levels are different from those of endogenous proteins and incomplete protein sequence coverage. MIP-APMS is based on epitope-tagged bait proteins, which are introduced into target cells by lentiviral cellular transduction. Although this strategy enables rapid experiments and functional interrogation with high bait throughput, protein production levels may deviate from endogenous levels with ectopic expression instead of genome editing (Lackner *et al.*, 2015). To address this and avoid excessive overproduction of the bait proteins (see Fig EV1B), we employed an engineered weak PGK promoter, as opposed to the commonly used strong cytomegalovirus (CMV) promoter (Qin *et al.*, 2010). In general, we recommend total proteome measurements as described in the current study to

evaluate whether normalization of changes caused by bait introduction is required.

Using MIP-APMS, we achieved 70% sequence coverage for bait proteins using a single-enzyme protein digestion strategy combined with data-dependent acquisition. To further increase sequence coverage and map PTMs on bait proteins even more comprehensively, additional proteases, e.g., chymotrypsin or GluC, could be used. As MIP-APMS does not include a second enrichment step, the method preferentially quantifies abundant PTMs on bait and prey proteins. Ubiquitylation, neddylation, and ISGylation are known as sub-stoichiometric PTMs and special biochemical enrichment or MS methods are commonly used for their detection (Kim *et al.*, 2011; Wagner *et al.*, 2011; Bustos *et al.*, 2012; Hansen *et al.*, 2021). We show that MIP-APMS combined with GlyGly enrichment facilitates the bait-centric identification of ubiquitin-like modification sites, exemplified for TRAF2. In the future, the total measuring time per sample at a comparable proteomics depth may be further reduced by using data-independent acquisition strategies and short LC gradients (Bruderer *et al.*, 2017; Bache *et al.*, 2018). Reproducibility, precision and accuracy of modified peptide quantification may be increased further by using isobaric labeling strategies as opposed to LFQ (Hogrebe *et al.*, 2018; Virreira Winter *et al.*, 2018).

In conclusion, MIP-APMS provides a versatile platform for parallel and time-resolved determination of PPIs and all PTMs of protein complexes in all transducible cells. It quantitatively resolves dynamic signaling network topologies and has broad applicability for the monitoring of virtually all coordinated intracellular programs. Owing to its conceptual design, emerging hypotheses on PTM and PPI involvement in selected signaling cascades are readily testable by protein mutation or loss of function impact on cellular phenotypes.

Materials and Methods

Reagents and Tools table

Reagent or Resource	Source	Identifier
Experimental Models: Cell Lines		
Human: HEK293T	ATCC	CRL-3216
Human: U937	ATCC	CRL-1593.2
Recombinant DNA		
CRISPR vector	Transomics	TELA1002
Gene synthesis	This paper	Appendix Table S6
pLOC vectors	GE Healthcare	Appendix Table S5
pMD2.G	Addgene	#12259
psPAX	Addgene	#12260
Antibodies		
Anti-rabbit IgG, HRP-linked antibody	Cell Signaling	7074
GAPDH (14C10) Rabbit mAb	Cell Signaling	2118
Phospho-p38 MAPK (Thr180/Tyr182) antibody	Cell Signaling	9211

Reagents and Tools table (continued)

Reagent or Resource	Source	Identifier
ARHGEF18	Sigma	HPA042689
MAP3K7	R&D	MAB5307
FOSB	R&D	AF2214
Bacterial and Virus Strains		
Signal Lenti NFκB Reporter (luc)	Qiagen	CLS-013L
XL1-Blue Competent Cells	Agilent Technologies	200249
Oligonucleotides		
PCR and cloning primers	This study	Appendix Table S1–S4
Chemicals, enzymes, and other reagents		
Blasticidin	Invivogen	ant-bl-1
cComplete™, Mini Protease Inhibitor Cocktail	Sigma	4693132001
DMEM	Life Technologies	31966047
JX-401	Santa Cruz Biotechnology	CAS 349087-34-9
LysC	Wako-Chemicals	129-02541
Ni-IDA Agarose	Jena Bioscience	AC-310-25
PAM3CSK4	Invivogen	tlrl-pms
Passive Lysis SX Buffer	Promega	E1941
Phosstop—20 TABLETS	Sigma	4906837001
Phusion® High-Fidelity DNA Polymerase	New England Biolabs	M0530S
Polybrene	Sigma	107689
Polyethylenimine, Linear, MW 25000, Transfection Grade (PEI 25K)	Polysciences	23966-1
Puromycin	Invivogen	ant-pr-5
RPMI-1640	Life Technologies	72400054
Skepinone-L	Merck	506174-5MG
Sorafenib	Santa Cruz Biotechnology	CAS 284461-73-0
Swal	New England Biolabs	R0604L
T4 DNA ligase reaction buffer	New England Biolabs	B0202S
Trypsin	Sigma	T6567-1mg
Critical commercial assays		
Dual-Luciferase® Reporter Assay System	Promega	E1910
QUIKChange II XL Site-Directed Mutagenesis Kit	Agilent	#200521
Oligonucleotides		
Oligos for pLOC cloning, site-directed mutagenesis, and gRNA cloning	This paper	Appendix Table S6
Software and Algorithms		
MaxQuant	(Cox & Mann, 2008)	http://www.biochem.mpg.de/5111795/maxquant
Perseus	(Tyanova et al, 2016)	http://www.biochem.mpg.de/5111810/perseus
R	NA	https://www.r-project.org/
Ggplot2	(Wickham, 2016)	https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf
Igraph	NA	http://igraph.org/r/
CHOPCHOP	(Labun et al, 2016)	http://chopchop.cbu.uib.no/index.php

Methods and Protocols

Experimental design

All experiments were performed in replicate. No aspect of the study was blinded. Sample size was not predetermined, and no outliers were excluded from analyses.

Molecular biology

Entry vector design

We made use of restriction enzyme-free seamless ligation cloning extract (SLiCE) cloning using universal primer pairs to insert coding sequences of genes from the Precision LentiORF Collection (pLOC) library (GE Healthcare) into target vectors under the control of a modified weak phosphoglycerate kinase (PGK) promoter, introducing C-terminal epitope tags into the encoded proteins (Zhang *et al.*, 2012). Our vector system is compatible with commercial DNA assembly cloning strategies such as the NEB Builder platform or Gateway due to *AttL* sites flanking the GOIs.

Entry vectors for SLiCE cloning were derived from the pLOC library (GE Healthcare). The vectors include a blasticidin resistance cassette for antibiotic-assisted cell-line selection and an IRES-GFP for FACS sorting. An efficient entry site for SLiCE cloning of the GOI was integrated by SLiCE cloning: The original pLOC vector was PCR-amplified using primers 1 and 2 containing overhangs with a *SwaI* restriction enzyme (New England Biolabs) cutting site for plasmid linearization, *attL1/2* sites for homologous recombination-based SLiCE cloning, and a His-GSG-Flag-tag for GOI epitope tagging. SLiCE cloning was performed as previously described (Zhang *et al.*, 2012). Briefly, 300 ng of the amplified pLOC vector, 1:10 (v/v) SLiCE extract (in-house), and 1:10 (v/v) T4 ligase buffer (New England Biolabs) was incubated for 1 h at 37°C. After incubation, the SLiCE mixture was used to transform XL1-blue bacteria (in-house) by heat shock. The transformants were selected on LB plates supplemented with 100 µg/ml ampicillin (LB-Amp plates) after overnight incubation at 37°C. Positive clones were identified by sequencing using primer 3.

The CMV promoter in the modified pLOC vector was exchanged for a weak PGK promoter by SLiCE cloning: the modified pLOC vector was PCR-amplified using primers 4 and 5 (see Appendix Table S1), the weak PGK promoter with homologous ends to the modified pLOC vector was *de novo* synthesized (see Appendix Table S6), and the two fragments were combined by SLiCE, as described before (Zhang *et al.*, 2012). Briefly, 300 ng of the amplified pLOC vector, 100 ng of the synthesized weak PGK promoter fragment, 1:10 (v/v) SLiCE extract, and 1:10 (v/v) T4 ligase buffer (New England Biolabs) were incubated for 1 h at 37°C. After incubation, the SLiCE mixture was used to transform XL1-blue bacteria by heat shock. The transformants were selected on LB-Amp plates after overnight incubation at 37°C. Positive clones were identified by sequencing using primer 6 (see Appendix Table S1), as above. The obtained vector was used in subsequent cloning steps as an entry vector, called pLOC entry vector (pLOC-PGKweak-C-HisGSGFlag-BLASTICIDIN).

Cloning for epitope tagging

Open-reading frame (ORF) clones were obtained from the Precision LentiORF Collection. GOI (see Appendix Table S5) were PCR-amplified from the pLOC library (GE Healthcare) using the universal

primers 7, 8, and 9. The CDS of MAP3K7 with *attL1/attL2* overhangs was obtained by gene synthesis (Thermo Fisher Scientific). The pLOC entry vector was digested with the restriction enzyme *SwaI* (New England Biolabs) according to the manufacturer's instructions. Then, 300 ng of linearized pLOC entry vector, 100 ng of amplified GOI, 1:10 (v/v) SLiCE extract, and 1:10 (v/v) T4 ligase buffer (New England Biolabs) were incubated for 1 h at 37°C. After incubation, the SLiCE mixture was used to transform XL1-blue bacteria by heat shock. The transformants were selected on LB-Amp plates supplemented with 10% (v/v) glucose after overnight incubation at 37°C. Positive clones were identified by sequencing using primers 4 and 10 (see Appendix Table S1).

Site-directed mutagenesis of selected phosphosites

For the site-directed mutagenesis of the N-terminal TRAF2 phosphosites and TRAF2 K→R mutants, the QUIKChange II XL site-directed mutagenesis kit (Agilent) was employed. The site-directed mutagenesis was performed by PCR amplification of pLOC-TRAF2 using specific primers (see Appendix Table S2), according to the manufacturer's instructions.

Molecular biology and protein purification for TRAF2-*ISG15* binding assays

ISG15, TRAF2, and influenza B virus NS1B were cloned into pCoofy vector as a N-terminal His-GST fusion. Plasmids were transformed into Rosetta (DE3) pLacI cells, grown in TB medium, and expression induced with 200 µM IPTG at OD₆₀₀ 0.4–0.8. After induction, cultures were grown for 16 h at 18°C. Cells were re-suspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol, protease inhibitor cocktail [Roche]) and lysed by sonication. Proteins were purified in tandem with His- and glutathione resin. Purified proteins were cleaved overnight with His-3C PreScission protease at 4°C. Following cleavage, the His-GST tag and His-3C protease were removed by a His pull-down. Proteins were either further purified by SEC (Superdex 75 10/300 GL, GE Life Sciences) or immediately buffer exchanged into storage buffer (50 mM Tris, 150 mM NaCl, 2 mM DTT). Proteins were concentrated and flash-frozen in liquid nitrogen.

Cell biology

Cell culture

U937 cells (CRL-1593.2) were purchased from the ATCC. The cells were cultured according to the manufacturer's instructions, in RPMI-1640 medium (Life Technologies) supplemented with 100 U/ml penicillin (GIBCO), 100 µg/ml streptomycin (GIBCO), and 10% (v/v) heat-inactivated fetal bovine serum (GIBCO; complete RPMI medium). The cells were incubated at 37°C under 5% CO₂.

HEK293T cells (CRL-3216) were purchased from ATCC. The cells were cultured according to the manufacturer's instructions, in DMEM (Life Technologies) supplemented with 100 U/ml penicillin (GIBCO), 100 µg/ml streptomycin (GIBCO), 1× Glutamax (GIBCO), and 10% heat-inactivated fetal bovine serum (complete DMEM medium). The cells were incubated at 37°C under 5% CO₂.

Primary human monocytes were obtained by culturing primary human monocytes enriched from buffy coats as described previously (Rieckmann *et al.*, 2017). Primary human macrophages were differentiated in RPMI-1640 medium (Life Technologies) supplemented with 100 U/ml penicillin (GIBCO), 100 µg/ml streptomycin (GIBCO),

10% (v/v) heat-inactivated fetal bovine serum (GIBCO), and 50 ng/ml M-Csf. The cells were incubated at 37°C under 5% CO₂.

NFκB reporter cell lines

U937 cell lines were transduced with Cignal Lenti NFκB-reporter constructs (Qiagen) according to the manufacturer's instructions. The transductants were selected in the presence of puromycin (5 μg/ml) for 14 days to establish stable cell lines.

Cell lines for epitope-tagged bait proteins

For lentivirus production, HEK293T cells (2×10^6 , one six-well) were transfected with sequence-validated pLOC-GOI vectors using polyethyleneimine (Polysciences) as a transfection reagent. Helper plasmids pMD2.G, psPAX, and the pLOC vector harboring the GOI were combined in a ratio of 1:1.5:2. After 4-h incubation in complete RPMI medium at 37°C under 5% CO₂, the transfection mix was removed and fresh complete RPMI medium was added. Lentiviral supernatant was collected after 48-h incubation at 37°C under 5% CO₂, centrifuged (500 g, 5 min), filtered (0.45 μm), and supplemented with 8 μg/ml polybrene (Sigma). Then, the virus (complete supernatant of one six-well) was added to 0.2 Mio U937 cells or U937-NFκB Reporter Cell lines, incubated for 4 h at 37°C under 5% CO₂, following which fresh medium was added. Selection pressure with blasticidin (10 μg/ml; Invivogen) was introduced after 48 h. The cells were cultured for 2 weeks under the selective pressure and then directly used in MIP-APMS experiments.

Transduction of primary human macrophages was performed as previously described (Berger *et al*, 2011). In short, 10 Mio macrophages were transduced with a mix of VPX-Vlps and pLOC lentivirus (v/v, 50%) in the presence of polybrene (8 μg/ml), incubated for 4 h at 37°C under 5% CO₂, following which fresh medium was added. Cells were harvested after 72h and then directly used in MIP-APMS experiments.

HEK293T (10 Mio) cells were transfected with pLOC-MAPK14-HisGSGFlag or pLOC-MAPK14-Strep using polyethyleneimine (Polysciences) as a transfection reagent. After 4-h incubation in complete DMEM medium at 37°C under 5% CO₂, the transfection mix was removed and fresh complete DMEM medium was added. Cells were harvested after 72h and then directly used in MIP-APMS, Flag-MS, and Strep-MS experiments.

TLR2 activation of U937 cells

Cells (5 Mio suspension) were seeded in deep-well 24-well plates, with one plate was used per cell line. TLR2 activation with PAM3CSK4 (0.5 μg/ml; Invivogen) was performed in a reverse time course and in quadruplicate, for 30, 15, 5, and 0 min at 37°C under 5% CO₂. The 0 min time point was not treated with PAM3CSK4. The cells were harvested by centrifugation and flash-frozen and stored at -80°C until MIP-APMS.

Drug mode of action on MAPK14

Cells (5 Mio) were seeded in deep-well 24-well plates. The cells were treated with MAPK14 inhibitors (sorafenib: 10 μM; skepinone-L: 80 nM; and JX-401: 10 μM) for 2 h at 37°C under 5% CO₂ in quadruplicate. Inhibitor-treated cells and controls either harvested directly or were activated with PAM3CSK3 (P3C4, 0.5 μg/ml; Invivogen) for 30 min at 37°C under 5% CO₂. Cells were harvested by centrifugation and frozen until MIP-APMS.

CRISPR/Cas9 Knockout

CRISPR knockout experiments were performed to identify potential novel interactors of MAP3K7 (see Appendix Table S3). For effective delivery of gRNA and Cas9, the transEDIT gRNA Plus Cas9 Expression vector with blasticidin was purchased from Transomics. For the experiment, gRNAs were designed using the web tool CHOP-CHOP (Labun *et al*, 2016) and cloned into the transEDIT vector according to the manufacturer's instructions (the primer list is provided in Appendix Table S3). Virus for each gRNA was produced as explained above (Cell Lines for Epitope-Tagged Bait Proteins). The U937-NFκB reporter cells were transduced and co-selected using puromycin (5 μg/ml) and blasticidin (10 μg/ml) at 37°C under 5% CO₂.

Luciferase reporter assay

U937-NFκB reporter cells (5×10^4) were seeded in quadruplicate on the day before the experiment. The cells were activated with PAM3CSK4 (0.5 μg/ml; Invivogen) for 6 h and harvested in passive lysis buffer (Promega). Luminescence of *Renilla* luciferase was determined in a dual-luciferase reporter assay (Promega), according to the manufacturer's instructions, using a microplate reader (Tecan).

Biochemistry

Western blots

One million U937 cells were stimulated, washed in PBS, and lysed in buffer (4% SDS, 40 mM HEPES [pH 7.4, 10 mM DTT] supplemented with protease inhibitors [Sigma-Aldrich, 4693159001]). Samples were centrifuged (16,000 g, 10 min), Li-LDS sample buffer was added to a final concentration of 1×, and the supernatant was incubated (5 min, 95°C). Proteins were separated on 12% Novex Tris-glycine gels (Thermo Fisher Scientific, XP00120BOX) and transferred onto PVDF membranes (Merck Millipore, IPVH00010) or Nitrocellulose membranes (Amersham, 10600002). Membranes were blocked in 5% BSA in PBST, and antibodies were diluted in 2% BSA in PBST. Antibodies used for immunoblotting were as follows (diluted 1:1,000): phospho-p38 MAPK (Thr180/Tyr182) antibody (Cell Signaling, 9211), GAPDH (14C10) rabbit mAb (Cell Signaling, 2118), p38 MAPK (R&D, AF8691), ARHGEF18 (Sigma, HPA042689), MAP3K7 (R&D, MAB5307), FOSB (R&D, AF2214) and anti-rabbit IgG, HRP-linked antibody (Cell Signaling, 7074).

His-IMAC enrichment

Frozen pellets of 19 cell lines with bait proteins containing 9x His-tags in deep-well 24-well plates were defrosted (5 min, 37°C). The cells were re-suspended in 800 μl of lysis buffer (10 mM HEPES [pH 7.5; Gibco], 50 mM NaCl [Sigma], 20 mM imidazole [Sigma], 0.05% NP-40 [Thermo Fisher], 1 mM MgCl₂ [Sigma], 50 U/ml benzonase [in-house], protease inhibitors [Roche, 1 tablet per 50 ml], and phosphatase inhibitors [Roche, 1 tablet per 50 ml]), incubated for 15 min on ice, and cleared by centrifugation (500 g, 5 min, 4°C). Supernatants were transferred to deep-well 96-well plates already containing equilibrated Ni-IDA beads (JenaBioScience GmbH, 50 μl slurry per well). The plates were incubated at 4°C for 1 h, shaking. The beads were washed three times (10 mM HEPES [pH 7.5], 50 mM NaCl, and 20 mM imidazole), and the supernatant was removed completely before proceeding.

Flag-enrichment

Frozen pellets of HEK293T-MAPK14-HisGSGFlag (1xFlag) and control cell lines were defrosted (5 min, 37°C). The cells were re-suspended in 800 µl of lysis buffer (10 mM HEPES [pH 7.5; Gibco], 50 mM NaCl [Sigma], 0.05% NP-40 [Thermo Fisher], 1 mM MgCl₂ [Sigma], 50 U/ml benzonase [in-house], protease inhibitors [Roche, 1 tablet per 50 ml], and phosphatase inhibitors [Roche, 1 tablet per 50 ml]), incubated for 15 min on ice, and cleared by centrifugation (500 g, 5 min, 4°C). Supernatants were transferred to deep-well 96-well plates already containing equilibrated anti-Flag M2 agarose gel (Sigma, 50 µl slurry per well). The plates were incubated at 4°C for 1 h, with shaking at over 1,500 rpm. The beads were washed three times (10 mM HEPES [pH 7.5], 50 mM NaCl) and the supernatant was removed completely before proceeding.

Strep-enrichment

Frozen pellets of HEK293T-MAPK14-Strep (1x Strep-tag II) and control cell lines were defrosted (5 min, 37°C). The cells were re-suspended in 800 µl of lysis buffer (10 mM HEPES [pH 7.5; Gibco], 50 mM NaCl [Sigma], 0.05% NP-40 [Thermo Fisher], 1 mM MgCl₂ [Sigma], 50 U/ml benzonase [in-house], protease inhibitors [Roche, 1 tablet per 50 ml], and phosphatase inhibitors [Roche, 1 tablet per 50 ml]), incubated for 15 min on ice, and cleared by centrifugation (500 g, 5 min, 4°C). Supernatants were transferred to deep-well 96-well plates already containing equilibrated MagStrep “type3” beads (iba, 50 µl slurry per well). The plates were incubated at 4°C for 1 h, with shaking at over 1,500 rpm. The beads were washed three times (10 mM HEPES [pH 7.5], 50 mM NaCl), and the supernatant was removed completely before proceeding with sample preparation for on-bead digestion. For elution, beads were incubated with 50 µl 1x buffer BXT (IBA Lifesciences) and purified proteins were eluted at room temperature for 30 min with constant shaking at 1,100 rpm on a ThermoMixer C incubator as described previously (Gordon et al, 2020). Proportional amounts of bead and elution were analyzed.

Combination of MIP-APMS with GlyGly enrichment

We used 500 Mio TRAF2-U937 cells and performed His-IMAC enrichment as described above adjusted for higher input. The sample was digested as explained below under sample preparation. Peptide desalting was performed on SepPack C18 columns as per the manufacturer's instruction. After elution, peptides were lyophilized overnight. The lyophilized sample was reconstituted in 900 µl cold immunoaffinity purification buffer (IAP; 50 mM MOPS, pH 7.2, 10 mM Na₂HPO₄, 50 mM NaCl). For the enrichment of diGly remnant containing peptides, antibodies of the PTMScan[®] Ubiquitin Remnant Motif (K-ε-GG) Kit (Cell Signaling Technology [CST] were first cross-linked to beads). For this, one vial of antibody-coupled beads was washed three times with 1 ml cold cross-linking buffer (100 mM sodium tetraborate decahydrate, pH 9.0), followed by 30-min incubation in 1 ml cross-linking buffer (20 mM dimethylpiperimidate in cross-linking wash buffer) for 30 min at room temperature and gentle agitation. After two consecutive washes with 1 ml cold quenching buffer (200 mM ethanolamine, pH 8.0) and 2-h incubation in 1 ml cold quenching, crosslinked beads were washed three times with 1 ml cold IAP buffer and 1/24 was immediately used for immunoaffinity purification. For this, peptides were added to crosslinked antibody beads and incubated for 2 h at 4°C under

gentle agitation. After incubation, beads were sequentially washed two times with cold IAP buffer and five times with cold ddH₂O in GF-StageTips. Thereafter, peptides were eluted twice with 50 µl 0.15% TFA into SDB-RPS StageTips. Eluted peptides were loaded onto stationary material and washed once with 200 µl 0.2% TFA and once with 200 µl 0.15% TFA/ 2% ACN. Peptides were eluted from SDB-RPS StageTips with 60 µl 1.25% ammonium hydroxide (NH₄OH)/80% ACN and dried using a SpeedVac centrifuge (Eppendorf, Concentrator plus). For mass spectrometry, dried peptides were re-suspended in 9 µl A* (2% ACN, 0.1% TFA).

Analytical size-exclusion chromatography binding assays

Binding assays were performed with ISG15 (1–157 aa) and TRAF2 variants (1–185 aa) on a Vanquish HPLC system (Thermo Fisher Scientific) using an AdvanceBio size-exclusion chromatography column (Agilent Technologies). As a positive control for ISG15 binding, the influenza B virus NS1B protein (1–103 aa) was used. Prior to analytical sizing, the column was pre-equilibrated with SEC buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT). ISG15 (30 µM) was mixed with TRAF2 variants or NS1B (25 µM) prior to injection on the column. Fractions were mixed with SDS sample buffer and resolved on a 4–20% gradient SDS/PAGE. Gels were visualized by Coomassie staining.

Quantitative proteomics analysis

MIP-APMS sample preparation

After His-IMAC, the beads were re-suspended in 50 µl of 8 M urea and 40 mM HEPES (pH 8.0). LysC digestion (Wako, 0.5 µg/µl, 1 µl) was performed for 3 h at room temperature - 25°C (with shaking, 1,500 rpm). Afterward, the samples were diluted (1:6) with water and digested with trypsin (Sigma; 0.5 µg/µl, 1 µl) for 16 h (room temperature, with shaking, 1,500 rpm). The digests were centrifuged (5 min, 500 g), and the supernatants were transferred to new 96-well plates. Cysteines were reduced by the addition of dithiothreitol (1 mM, room temperature, 1,500 rpm, 30 min), before proceeding to cysteine alkylation with iodoacetamide (55 mM, room temperature, 30 min, dark). Excess iodoacetamide was quenched by adding thiourea (100 mM, room temperature, 10 min) prior to acidification for peptide desalting with trifluoroacetic acid (TFA; final concentration: 1% v/v). Peptides were loaded onto C18 StageTips (EmporeTM, IVA-Analysentechnik). They were then eluted with 80% acetonitrile, dried using a SpeedVac, and re-suspended in a solution of 2% acetonitrile, 0.1% TFA, and 0.5% acetic acid.

Whole-proteome MS sample preparation

Cells were lysed in SDC-lysis buffer and digested with LysC and trypsin, as described previously (Kulak et al, 2017). Peptides were desalted on stacked poly(styrene-divinylbenzene) reversed-phase sulfonate plugs and eluted with a mixture of 80% acetonitrile, 19% ddH₂O, and 1% ammonia. MS measurements were performed in replicate ($n = 3$) using Q Exactive HF (Thermo Fisher Scientific).

LC-MS/MS

Peptides were separated using an EASY-nLC 1200 HPLC system (Thermo Fisher Scientific) coupled online to the Q Exactive HF and Q Exactive HF-X mass spectrometer via a nanoelectrospray source (Thermo Fisher Scientific), as described before (Scheltema et al,

2014; Kelstrup *et al*, 2018). Peptides were loaded in buffer A (0.5% formic acid) on in-house packed columns (75 μm inner diameter and 20 cm long; packed with 1.9- μm C18 particles from Dr. Maisch GmbH, Germany). Peptides were eluted using a nonlinear 95-min gradient of 5–60% buffer B (80% acetonitrile and 0.5% formic acid) at a flow rate of 300 nl/min and a column temperature of 55°C. The operational parameters were monitored in real-time by using the SprayQC software (in-house) (Scheltema & Mann, 2012). The Q Exactive HF and Q Exactive HF-X were operated in a data-dependent acquisition positive mode with a survey scan range of 300–1,650 m/z and a resolution of 60,000–120,000 at m/z 200. Up to 15 most abundant isotope patterns with a charge of > 1 were isolated using a 1.8 Thomson (Th) isolation window and subjected to high-energy collisional dissociation fragmentation at a normalized collision energy of 27. Fragmentation spectra were acquired with a resolution of 15,000 at m/z 200. Dynamic exclusion of sequenced peptides was set to 20 s to reduce repeated peptide sequencing. Thresholds for ion injection time and ion target values were set to 20 ms and 3E6 for the survey scans, and 55 ms and 1E5 for the MS/MS scans. Data were acquired using the Xcalibur software (Thermo Scientific).

Quantification and statistical analysis

Peptide identification and LC-MS/MS data analysis

MaxQuant software (version 1.5.3.16) was used to analyze MS raw files. MS/MS spectra were searched against the human Uniprot FASTA database (version July 2015, 91,645 entries) and a common contaminants database (247 entries) by the Andromeda search engine (Cox & Mann, 2008). Cysteine carbamidomethylation was set as a fixed modification, and N-terminal acetylation and methionine oxidation were set as variable modifications. To identify and quantify phosphorylation, acetylation, and methylation, variable modification search was consecutively performed. Enzyme specificity was set to trypsin, with a maximum of two missed cleavages and a minimum peptide length of seven amino acids. FDR of 1% was applied at the peptide and protein level. Peptide identification was performed with an allowed initial precursor mass deviation of up to 7 ppm and an allowed fragment mass deviation of 20 ppm. Nonlinear retention time alignment of all analyzed samples was performed using MaxQuant. Peptide identifications were matched across all samples within a time window of 1 min of the aligned retention times. Protein identification required at least one “razor peptide” in MaxQuant. A minimum ratio count of 1 was required for valid quantification events using the MaxQuant’s LFQ algorithm (MaxLFQ). Data were filtered for the presence of common contaminants and peptides only identified by site modification, and hits to the reverse database (Cox & Mann, 2008) were excluded from further analysis.

Dependent peptide in MaxQuant analysis was performed to analyze unbiased PTMs on MAPK14 with standard parameters (FDR < 0.01 , Mass bin size 0.0065 Da). For TagGraph analysis, sequence interpretations were first analyzed with the *de novo* search engine Peaks. Peaks analysis was performed with 10 ppm precursor mass tolerance and 0.01 Da fragment mass tolerance (Ma *et al*, 2003). TagGraph analysis was performed using human Uniprot FASTA database (version July 2015, 91,645 entries), with FDR cutoff of 0.1, and all other settings remained to unchanged as present in the software distribution.

Interactor calling

We integrate differences in intensity and abundance as described before (Keilhauer *et al*, 2015). We employ an AE-MS workflow with quantitative MS, which means that we use not only the information for protein identification but also for protein quantification for post-experiment interactor calling. To determine which proteins are substantially enriched (i.e., bait and prey proteins), AE-MS employs standard statistical testing (*t*-test) with a multiple hypothesis correction (FDR 0.01 for multiple hypothesis testing). In detail, each quantified protein had to be identified with more than one peptide and in more than 60% of replicates of at least one cell line to be considered valid. Protein LFQ intensities were log-transformed to the base of 2 and missing values imputed from a random normal distribution centered on the detection limit (width = 0.3, Down Shift:1.8). Samples were clustered by using Pearson correlation into different control groups in the Perseus environment leading to three separate groups (see Fig EV3K). To identify the interactors, a two-tailed Student’s *t*-test (permutation-based FDR < 0.01 with 250 randomizations, enrichment > 2) with a minimum of 10 valid values in the first group was performed in the Perseus environment, using all other cell lines in the respective control group (Tyanova *et al*, 2016). Here, the baits were loaded as first group and second group mode was selected as “complement”. Significant interactors were compared to the STRING and Biogrid databases (Szkarczyk *et al*, 2015; Chatr-Aryamontri *et al*, 2017) and overlaps were denoted in the Figs.

SAINT analysis via crapome

MAPK14 His IPs and controls (U937 transduced with His-Tag) were performed in triplicates and uploaded to the SAINT-based Crapome server (<https://reprint-apms.org>) (Mellacheruvu *et al*, 2013). As Experiment Type, we selected single-step epitope tag APMS and spectral counts as quantitation Type. As external controls, we selected PBMC (cell/tissue type), agarose (affinity support), and Q Exactive (Instrument type). The primary empirical fold change score (FC-A) was calculated by user controls using average for combining replicates (number of virtual controls = 10). The secondary fold change score was calculated by all controls (user + external controls) using geometric mean for combining replicates (number of virtual controls = 3). The probabilistic SAINT Score was calculated by user controls (combining replicates: average) and 10 virtual controls. Saint options were 2,000 n-burn, 4,000 n-iter, 0 LowMode, 1 MinFold, and 1 Normalize.

Analysis of dynamic PTMs and PPIs

Prior to the analysis of dynamic PPIs, LFQ intensities of significant interactors of each replicate were normalized to the LFQ intensities of the respective bait proteins to avoid loading artifacts.

$$\text{LFQ} - \text{intensity}(\text{prey} - \text{protein})_{\text{normalized}} = \frac{\text{LFQ intensity}(\text{prey} - \text{protein})}{\text{LFQ intensity}(\text{bait} - \text{protein})}$$

A two-tailed Student’s *t*-test (P -value < 0.05) was performed on the previously identified significant interactors comparing un-activated conditions versus activated conditions at different time points. Significant dynamic preys were reported with an asterisk in the heatmaps.

Conversely, intensities of modified peptides of each replicate were normalized to the intensity of the respective protein intensity to decrease the total coefficient of variation. PTMs that had valid values in at least 3 replicates of at least one time point were considered for the analysis. No imputation was performed.

$$\text{Intensity}(\text{modified peptide of protein } X)_{\text{normalized}} = \frac{\text{Intensity}(\text{modified peptide of protein } X)}{\text{Intensity}(\text{protein } X)}$$

A two-tailed Student's *t*-test (P -value < 0.05) was performed on the previously identified significant interactors and modified peptides, respectively, comparing un-activated conditions versus activated conditions at different time points. Significant dynamic preys/PTMs were reported with an asterisk in the heatmaps.

Unsupervised clustering

Intensities of dynamically regulated PPIs (357) and PTMs (178) upon TLR 2 activation were filtered for at least 70% valid values and normalized per time point (PPIs to bait protein intensity and PTMs to protein intensity of the modified protein as explained above). The median of each time point was calculated and then *Z*-scored. Pearson correlation was calculated between each of the PPIs and PTMs, and results were visualized by hierarchical clustering. The data were clustered and median *z*-scored intensities (confidence interval: 0.95) were plotted against the time course of TLR2 activation (method = loess, $y \sim x$). N shows the number of PPIs/PTMs corresponding to each cluster.

Analysis of whole-proteome data

Full proteomes were measured in triplicates as described under peptide identification and LC-MS/MS data analysis. Data were filtered for the presence of common contaminants and peptides only identified by site modification, and hits to the reverse database (Cox & Mann, 2008) were excluded from further analysis. As a requirement, each quantified protein had to be identified with more than one peptide and in more than 60% of replicates of at least one cell line to be considered valid. Protein LFQ intensities were log-transformed to the base of 2 and missing values imputed from a random normal distribution centered on the detection limit (width = 0.3, Down Shift: 1.8). To identify differentially expressed proteins between wildtype and transduced cell lines, a two-tailed Student's *t*-test (permutation-based FDR < 0.05 with 250 randomizations, enrichment > 2) with a minimum of two valid values in the first group was performed in the Perseus environment, using all other cell lines in the respective control group (Tyanova et al, 2016). Copy numbers were calculated with the Perseus Plugin Proteomic Ruler, which normalizes protein intensity to the molecular mass of each protein (Wićniewski et al, 2014).

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD010996. The datasets produced in this study are available in the following database: <https://www.ebi.ac.uk/pride/>

Project accession: PXD010996 (<http://www.ebi.ac.uk/pride/archive/projects/PXD010996>).

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Felix Meissner (felix.meissner@uni-bonn.de).

Expanded View for this article is available online.

Acknowledgements

We thank I. Paron, C. Deiml, J. Mueller, and K. Mayr for MS assistance; G. Sowa, J. Baum, V. Boeck S. Dewitz, and A. Ullrich for technical assistance; M. Oroshi for computer and database support; and M. Tanzer, J. Swietlik, J. Rieckmann, M. Steger, J. Bader, and C. Lubert for helpful discussions. This work was funded by the Max Planck Society for the Advancement of Science and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 165054336 (SFB 914), Project-ID 360372040 (SFB 1335), and Project-ID 408885537 (TRR274). Open Access funding enabled and organized by Projekt DEAL.

Author contributions

AF, SE, FMH, KP, and KS performed experiments. AF developed and implemented bioinformatics methods. AF and FM conceived the data analysis and interpreted the data. AS and DH assisted in data analysis. FM and MM conceived the study. FM supervised the experiments. AF and FM wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* 537: 347
- Arkin MR, Tang Y, Wells JA (2014) Small-molecule inhibitors of protein-protein interactions: progressing towards the reality. *Chem Biol* 21: 1102–1114
- Bache N, Geyer PE, Bekker-Jensen DB, Hoerning O, Falkenby L, Treit PV, Doll S, Paron I, Müller JB, Meier F et al (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol Cell Proteomics* 17: 2284–2296
- Berger G, Durand S, Goujon C, Nguyen X-N, Cordeil S, Darlix J-L, Cimarelli A (2011) A simple, versatile and efficient method to genetically modify human monocyte-derived dendritic cells with HIV-1-derived lentiviral vectors. *Nat Protoc* 6: 806–816
- Borghgi A, Verstrepen L, Beyaert R (2016) TRAF2 multitasking in TNF receptor-induced signaling to NF-κB, MAP kinases and cell death. *Biochem Pharmacol* 116: 1–10
- Bruderer R, Bernhardt OM, Gandhi T, Xuan Y, Sondermann J, Schmidt M, Gomez-Varela D, Reiter L (2017) Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol Cell Proteomics* 16: 2296–2309
- Buljan M, Ciuffa R, van Drogen A, Vichalkovski A, Mehnert M, Rosenberger G, Lee S, Varjosalo M, Pernas LE, Spegg V et al (2020) Kinase interaction network expands functional and disease roles of human kinases. *Mol Cell* 79: 504–520.e9
- Bustos D, Bakalarski CE, Yang Y, Peng J, Kirkpatrick DS (2012) Characterizing ubiquitination sites by peptide-based immunofluorescence enrichment. *Mol Cell Proteomics* 11: 1529–1540

- Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A et al (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 45(D1): D369–D379
- Cheng G, Baltimore D (1996) TANK, a co-inducer with TRAF2 of TNF- and CD 40L-mediated NF-kappaB activation. *Genes Dev* 10: 963–973
- Choudhary C, Kumar C, Gnaf F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325: 834
- Couzens A, Knight J, Kean M, Teo G, Weiss A, Dunham W, Lin Z, Bagshaw R, Sicheri F, Pawson T et al (2013) Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Sci Signal* 6: rs15
- Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* 26: 1367–1372
- Devabhaktuni A, Lin S, Zhang L, Swaminathan K, Gonzalez CG, Olsson N, Pearlman SM, Rawson K, Elias JE (2019) TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat Biotechnol* 37: 469–479
- Edwards JP, Emens LA (2010) The multikinase inhibitor sorafenib reverses the suppression of IL-12 and enhancement of IL-10 by PGE(2) in murine macrophages. *Int Immunopharmacol* 10: 1220–1228
- Fernandes-Alnemri T, Yu J-W, Datta P, Wu J, Alnemri ES (2009) AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* 458: 509–513
- Friedmann Y, Shriki A, Bennett ER, Golos S, Diskin R, Marbach I, Bengal E, Engelberg D (2006) JX401, A p38 α inhibitor containing a 4-benzylpiperidine motif, identified via a novel screening system in yeast. *Mol Pharmacol* 70: 1395
- Gaidt MM, Ebert TS, Chauhan D, Ramshorn K, Pinci F, Zuber S, O'Duill F, Schmid-Burgk JL, Hoss F, Buhmann R et al (2017) The DNA inflammasome in human myeloid cells is initiated by a STING-cell death program upstream of NLRP3. *Cell* 171: 1110–1124.e1118
- Gavin A-C, Böschke M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A-M, Cruciat C-M et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141
- Gingras A-C, Gstaiger M, Raught B, Aebersold R (2007) Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 8: 645–654
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL et al (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583: 459–468
- Guruharsha KG, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O et al (2011) A protein complex network of *Drosophila melanogaster*. *Cell* 147: 690–703
- Hansen FM, Tanzer MC, Brüning F, Bludau I, Stafford C, Schulman BA, Robles MS, Karayel O, Mann M (2021) Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat Commun* 12: 254
- Hein MY, Sharma K, Cox J, Mann M (2013) Proteomic analysis of cellular systems A2. In *Handbook of Systems Biology*, Marian Walhout AJ, Vidal M, Dekker J (eds), San Diego, CA: Academic Press: 3–25.
- Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F et al (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163: 712–723
- Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams S-L, Millar A, Taylor P, Bennett K, Boutilier K et al (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415: 180
- Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/Technology* 6: 1321–1325
- Hogrebe A, von Stechow L, Bekker-Jensen DB, Weinert BT, Kelstrup CD, Olsen JV (2018) Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nat Commun* 9: 1045
- Holding AN (2015) XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* 89: 54–63
- Hopp TP, Prickett KS, Price VL, Libby RT, March CJ, Pat Cerretti D, Urdal DL, Conlon PJ (1988) A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* 6: 1204–1210
- Hornung V, Ablasser A, Charrel-Dennis M, Bauernfeind F, Horvath G, Caffrey DR, Latz E, Fitzgerald KA (2009) AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* 458: 514–518
- Hsu H, Huang J, Shu H-B, Baichwal V, Goeddel DV (1996) TNF-dependent recruitment of the protein kinase RIP to the TNF receptor-1 signaling complex. *Immunity* 4: 387–396
- Hubel P, Urban C, Bergant V, Schneider WM, Knauer B, Stukalov A, Scaturro P, Mann A, Brunotte L, Hoffmann HH et al (2019) A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape. *Nat Immunol* 20: 493–502
- Humphrey SJ, Azimifar SB, Mann M (2015) High-throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nat Biotechnol* 33: 990
- Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K et al (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* 162: 425–440
- Karayel O, Tonelli F, Virreira Winter S, Geyer PE, Fan Y, Sammler EM, Alessi D, Steger M, Mann M (2020) Accurate MS-based Rab10 phosphorylation stoichiometry determination as readout for LRRK2 activity in Parkinson's disease. *Mol Cell Proteomics* 19: 1546–1560
- Keilhauer EC, Hein MY, Mann M (2015) Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol Cell Proteomics* 14: 120–135
- Kelstrup CD, Bekker-Jensen DB, Arrey TN, Hogrebe A, Harder A, Olsen JV (2018) Performance evaluation of the Q Exactive HF-X for shotgun proteomics. *J Proteome Res* 17: 727–738
- Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ et al (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* 44: 325–340
- Koeberle SC, Romir J, Fischer S, Koeberle A, Schattel V, Albrecht W, Grütter C, Werz O, Rauh D, Stehle T et al (2011) Skepinone-L is a selective p38 mitogen-activated protein kinase inhibitor. *Nat Chem Biol* 8: 141
- Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP et al (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440: 637
- Kulak NA, Geyer PE, Mann M (2017) Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol Cell Proteomics* 16: 694–705
- Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E (2016) CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* 44(W1): W272–W276
- Lackner DH, Carré A, Guzzardo PM, Banning C, Mangena R, Henley T, Oberndorfer S, Gapp BV, Nijman SMB, Brummelkamp TR et al (2015) A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat Commun* 6: 10237
- Lambert J-P, Tucholska M, Pawson T, Gingras A-C (2014) Incorporating DNA shearing in standard affinity purification allows simultaneous identification of both soluble and chromatin-bound interaction partners. *J Proteomics* 100: 55–59

- Landström M (2010) The TAK1–TRAF6 signalling pathway. *Int J Biochem Cell Biol* 42: 585–589
- Larance M, Lamond AI (2015) Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol* 16: 269
- Lescarbeau RS, Lei L, Bakken KK, Sims PA, Sarkaria JN, Canoll P, White FM (2016) Quantitative phosphoproteomics reveals Wee1 kinase as a therapeutic target in a model of proneural glioblastoma. *Mol Cancer Ther* 15: 1332–1343
- Li X, Jiang S, Tapping RI (2010) Toll-like receptor signaling in cell proliferation and survival. *Cytokine* 49: 1–9
- Liu F, Heck AJR (2015) Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr Opin Struct Biol* 35: 100–108
- Liu JJ, Sharma K, Zangrandi L, Chen C, Humphrey SJ, Chiu Y-T, Spetea M, Liu-Chen L-Y, Schwarzer C, Mann M (2018) *In vivo* brain GPCR signaling elucidated by phosphoproteomics. *Science* 360: eaao4927
- Loeb KR, Haas AL (1992) The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J Biol Chem* 267: 7806–7813
- Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 17: 2337–2342
- Marcon E, Jain H, Bhattacharya A, Guo H, Phanse S, Pu S, Byram G, Collins BC, Dowdell E, Fenner M et al (2015) Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat Methods* 12: 725
- Martín-Blanco E (2000) p38 MAPK signalling cascades: ancient roles and new functions. *BioEssays* 22: 637–645
- Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY et al (2013) The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nat Methods* 10: 730
- Molina DM, Jafari R, Ignatushchenko M, Seki T, Larsson EA, Dan C, Sreekumar L, Cao Y, Nordlund P (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* 341: 84
- New L, Jiang Y, Zhao M, Liu K, Zhu W, Flood LJ, Kato Y, Parry GC, Han J (1998) PRAK, a novel protein kinase regulated by the p38 MAP kinase. *EMBO J* 17: 3372–3384
- Oliveira-Nascimento L, Massari P, Wetzler LM (2012) The role of TLR2 in infection and immunity. *Front Immunol* 3: 79–79
- Pan CQ, Sudol M, Sheetz M, Low BC (2012) Modularity and functional plasticity of scaffold proteins as p(l)acemakers in cell signaling. *Cell Signal* 24: 2143–2165
- Pankow S, Bamberger C, Yates JR (2019) A posttranslational modification code for CFTR maturation is altered in cystic fibrosis. *Sci Signal* 12: eaan7984
- Paul FE, Hosp F, Selbach M (2011) Analyzing protein–protein interactions by quantitative mass spectrometry. *Methods* 54: 387–395
- Qin JY, Zhang L, Clift KL, Huler I, Xiang AP, Ren B-Z, Lahn BT (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* 5: e10611
- Rees JS, Lowe N, Armean IM, Roote J, Johnson G, Drummond E, Spriggs H, Ryder E, Russell S, St Johnston D et al (2011) *In vivo* analysis of proteomes and interactomes using Parallel Affinity Capture (iPAC) coupled to mass spectrometry. *Mol Cell Proteomics* 10: M110.002386
- Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, Jarrrossay D, Sallusto F, Shen-Orr SS, Lanzavecchia A, Mann M et al (2017) Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol* 18: 583
- Ruepp A, Waegle B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW (2010) CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res* 38 (suppl_1): D497–D501
- Samavarchi-Tehrani P, Abdouni H, Samson R, Gingras A-C (2018) A versatile lentiviral delivery toolkit for proximity-dependent biotinylation in diverse cell types *. *Mol Cell Proteomics* 17: 2256–2269
- Sato S, Sanjo H, Takeda K, Ninomiya-Tsuji J, Yamamoto M, Kawai T, Matsumoto K, Takeuchi O, Akira S (2005) Essential function for the kinase TAK1 in innate and adaptive immune responses. *Nat Immunol* 6: 1087–1095
- Scheltema RA, Mann M (2012) SprayQc: a real-time LC–MS/MS quality monitoring system to maximize uptime using off the shelf components. *J Proteome Res* 11: 3458–3466
- Scheltema RA, Hauschild J-P, Lange O, Hornburg D, Denisov E, Damoc E, Kuehn A, Makarov A, Mann M (2014) The Q exactive HF, a benchtop mass spectrometer with a pre-filter, high performance quadrupole and an ultra-high field orbitrap analyzer. *Mol Cell Proteomics* 13: 3698–3708
- Schirle M, Bantscheff M, Kuster B (2012) Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* 19: 72–84
- Schmidt TGM, Skerra A (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* 2: 1528–1535
- Seet BT, Dikic I, Zhou M-M, Pawson T (2006) Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 7: 473
- Stutz A, Kolbe C-C, Stahl R, Horvath GL, Franklin BS, van Ray O, Brinkschulte R, Geyer M, Meissner F, Latz E (2017) NLRP3 inflammasome assembly is regulated by phosphorylation of the pyrin domain. *J Exp Med* 214: 1725–1736
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP et al (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43(Database issue): D447–D452
- Trinkle-Mulcahy L, Boulton S, Lam YW, Urcia R, Boisvert FM, Vandermoere F, Morrice NA, Swift S, Rothbauer U, Leonhardt H et al (2008) Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J Cell Biol* 183: 223–239
- Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Meth* 13: 731–740
- Virreira Winter S, Meier F, Wichmann C, Cox J, Mann M, Meissner F (2018) EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nat Methods* 15: 527–530
- Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C (2011) A proteome-wide, quantitative survey of *in vivo* ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics* 10: M111.013284
- Wickham H (2016) *ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag. ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>
- Wińiewski JR, Hein MY, Cox J, Mann M (2014) A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics* 13: 3497
- Wu CJ, Conze DB, Li X, Ying SX, Hanover JA, Ashwell JD (2005) TNF- α induced c-IAP1/TRAF2 complex translocation to a Ubc6-containing compartment and TRAF2 ubiquitination. *EMBO J* 24: 1886
- Yao Z, Darowski K, St-Denis N, Wong V, Offensperger F, Villedieu A, Amin S, Maly R, Aoki H, Guo H et al (2017) A global analysis of the receptor tyrosine kinase-protein phosphatase interactome. *Mol Cell* 65: 347–360

Annika Frauenstein et al

Molecular Systems Biology

Zhang D, Zhang D-E (2011) Interferon-stimulated gene 15 and the protein ISGylation system. *J Interferon Cytokine Res* 31: 119–130
Zhang Y, Werling U, Edelmann W (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* 40: e55



License: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

3.5 Interconversion between Anticipatory and Active GID E3 Ubiquitin Ligase Conformations via Metabolically Driven Substrate Receptor Assembly

Shuai Qiao¹, Christine R. Langlois¹⁴, Jakub Chrustowicz¹⁴, Dawafuti Sherpa¹⁴, Ozge Karayel², **Fynn M. Hansen**², Viola Beier¹, Susanne von Gronau¹, Daniel Bollschweiler³, Tillman Schäfer³, Arno F. Alpi¹, Matthias Mann², J. Rajan Prabu¹

¹Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. ²Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. ³Cryo-EM Facility, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

Published in *Molecular Cell* (2020)

The post-translational modification of proteins is a vital cellular strategy to rapidly respond to environment perturbations. In *Saccharomyces cerevisiae*, for instance, carbon-stress induced gluconeogenesis is terminated upon glucose availability via the ubiquitination and subsequent degradation of gluconeogenic enzymes such as fructose-1,6-bisphosphatase (Fbp1), malate dehydrogenase (Mdh2), and isocitrate lyase (Icl1) [195]. The glucose-induced degradation-deficient (GID) E3 ligase complex holds a central role in the carbohydrate metabolism [196] and glucose-induced degradation of gluconeogenic enzymes. In this study, the Schulman group characterized state specific GID complex assemblies, as well as their mechanistic properties. Interestingly, they found an anticipatory GID complex that can readily assemble with different substrate receptors to enable specialized responses to various environmental stimuli.

I contributed to this study by mapping ubiquitination sites on Mdh2, a well-known substrate of the GID complex. For this, Mdh2 was ubiquitinated *in-vitro* and after optimized enzymatic digestion to increase protein sequence coverage, modification sites were mapped via MS. These sites may concurrently interact with Gid4 and Gid2 RING activated Ubc8~Ub intermediates. Furthermore, I analyzed the GID complex composition via MS after immunoprecipitation of tagged Gid8 and identified Gid7 to have a markedly lower abundance than other Gid components.

Interconversion between Anticipatory and Active GID E3 Ubiquitin Ligase Conformations via Metabolically Driven Substrate Receptor Assembly

Shuai Qiao,¹ Christine R. Langlois,^{1,4} Jakub Chrustowicz,^{1,4} Dawafuti Sherpa,^{1,4} Ozge Karayel,² Fynn M. Hansen,² Viola Beier,¹ Susanne von Gronau,¹ Daniel Bollschweiler,³ Tillman Schäfer,³ Arno F. Alpi,¹ Matthias Mann,² J. Rajan Prabu,¹ and Brenda A. Schulman^{1,5,*}

¹Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

²Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

³Cryo-EM Facility, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: schulman@biochem.mpg.de

<https://doi.org/10.1016/j.molcel.2019.10.009>

SUMMARY

Cells respond to environmental changes by toggling metabolic pathways, preparing for homeostasis, and anticipating future stresses. For example, in *Saccharomyces cerevisiae*, carbon stress-induced gluconeogenesis is terminated upon glucose availability, a process that involves the multiprotein E3 ligase GID^{SR4} recruiting N termini and catalyzing ubiquitylation of gluconeogenic enzymes. Here, genetics, biochemistry, and cryoelectron microscopy define molecular underpinnings of glucose-induced degradation. Unexpectedly, carbon stress induces an inactive anticipatory complex (GID^{Ant}), which awaits a glucose-induced substrate receptor to form the active GID^{SR4}. Meanwhile, other environmental perturbations elicit production of an alternative substrate receptor assembling into a related E3 ligase complex. The intricate structure of GID^{Ant} enables anticipating and ultimately binding various N-degron-targeting (i.e., “N-end rule”) substrate receptors, while the GID^{SR4} E3 forms a clamp-like structure juxtaposing substrate lysines with the ubiquitylation active site. The data reveal evolutionarily conserved GID complexes as a family of multisubunit E3 ubiquitin ligases responsive to extracellular stimuli.

INTRODUCTION

Eukaryotes use a plethora of mechanisms to cope with environmental perturbations. Much of our understanding of these processes comes from studies on the yeast *S. cerevisiae*, for example, chaperone induction to enable protein folding during heat stress, kinase activation to control osmolarity, and glycolytic or gluconeogenic enzyme expression to switch metabolism. An emerging concept is that cells also have “anticipatory” programs whereby an altered growth condition triggers not only

pathways rescuing cells from immediate dangers but also expression of proteins that could be required for subsequent shifts in conditions (Mitchell et al., 2009; Tagkopoulos et al., 2008). If the anticipated perturbation does occur, cells can more rapidly adapt to the new environment through expression of yet other genes. For example, chaperones are induced at temperatures below those causing global misfolding, thereby increasing proteostasis capacity should a more severe late stress further compromise cellular protein folding (Klaips et al., 2014). Determination of protein fate by ubiquitylation is another major mechanism orchestrating homeostasis (Ciechanover 2012; Varshavsky, 2012). Ubiquitylation depends on cellular signals directing E3 ligases to particular targets. Yet our understanding of E3-dependent responses to environmental change remains rudimentary. The questions of if and how E3 ligase structures play roles in cellular anticipation and responses to perturbations in the extracellular milieu are largely unexplored.

The ubiquitin (Ub) system has long been known to regulate yeast carbon catabolite repression (Zaman et al., 2008). Although yeast growing on non-fermentable carbon source (e.g., ethanol) require gluconeogenic production of glucose this energetically costly pathway is futile and therefore terminated when sugars are available. This not only involves multifaceted transcriptional responses but also glucose-induced degradation (Gid) of gluconeogenesis enzymes such as fructose-1,6-bisphosphatase (Fbp1), malate dehydrogenase (Mdh2), and isocitrate lyase (Icl1) (Chiang and Schekmar 1991; Chiang and Chiang, 1998; Gancedo, 1998; Hoffman and Chiang, 1996; Schork et al., 1994a, 1994b). The original Gid gene products defined by genetics and biochemistry include the E2 Ub-conjugating enzyme Gid3 (hereafter referred to as Ubc8), the deubiquitylating enzyme Ubp14 (Gid6), and a Gid complex loosely defined by physical interactions of Gid1, Gid2, Gid4, Gid5, Gid7, Gid8, and Gid9 (Braun et al., 2011; Francis et al., 2013; Menssen et al., 2012; Regelmann et al., 2003; Sant et al., 2008; Schüle et al., 2000). Although the Gid2 and Gid9 subunits each harbor RING (really interesting new gene) domains the other subunits lack sequences associated with ubiquitylation (Braun et al., 2011; Francis et al., 2013; Menssen et al., 2012; Schüle et al., 2000). A recent breakthrough in our understanding



of the GID E3 came from its assignment as an N-degron-targeting E3 (Chen et al., 2017).

N-degron (formerly termed “N-end rule”) and C-degron (collectively referred to as “terminal degron”) E3s recognize substrate N or C termini and regulate vast biology (Varshavsky, 2019). Nonetheless, beyond knowledge of pathways creating, exposing, or cloaking substrate N- or C-degrons, and structures showing their recognition by E3 ligases, there is limited structural information explaining regulation of terminal degron E3s (Brower et al., 2013; Choi et al., 2010; Dong et al., 2018; Hu et al., 2005; Koren et al., 2018; Lin et al., 2018; Matta-Camacho et al., 2010; Rao et al., 2001; Rusnac et al., 2018; Shemorry et al., 2013; Szoradi et al., 2018; Timms et al., 2019; Varshavsky, 2011; Wang et al., 2008). Moreover, Fbp1, Mdh2, and Icl1 each harbor natively exposed GID E3-targeting N-terminal prolines essential for their degradation (Hämmerle et al., 1998). The question of how their ubiquitylation could be regulated was answered by discovery that glucose availability determines expression of Gid4 (Menssen et al., 2018; Santt et al., 2008), which serves as a substrate receptor for the GID E3 by binding to an N-terminal proline (Chen et al., 2017; Dong et al., 2018). A crystal structure of peptide-bound human Gid4 showed the basis for N-terminal proline recognition (Dong et al., 2018). Although the mammalian GID E3 does not appear to regulate gluconeogenic enzymes (Lampert et al., 2018), and its N-degron substrates remain to be identified, numerous studies suggest that it may also act as a central component in cell fate determination essential for some developmental pathways (Han et al., 2016; Javan et al., 2018; Liu and Pfirrmann, 2019; Nguyen et al., 2017; Pfirrmann et al., 2015; Soni et al., 2006).

Here we reveal molecular mechanisms underlying assembly and activity of the largely mysterious GID E3 and provide general insight into ubiquitylation by the large cohort of terminal degron E3s and by those catalyzing ubiquitylation via heterodimeric RING-RING domains. Unexpectedly, our results also reveal mechanisms of stress anticipation and resolution through assembly of an E3 ligase, and that GID is not a singular complex. GID comprises a family of multisubunit E3s regulated through assembly with interchangeable N-degron-binding substrate receptors induced by distinct environmental perturbations.

RESULTS

Carbon-Source-Dependent Anticipatory versus Activated GID E3 Ligase Assemblies

As a prelude for developing and validating a recombinant system, we investigated properties of endogenous Gid proteins. The potential of Gid proteins to stably coassemble with one another *in vivo* was examined using a suite of yeast strains, each harboring a Gid gene tagged at its endogenous locus and validated for activity. Yeasts were grown in various carbon sources known to determine GID E3 ligase activity (Oh et al., 2017; Regelmann et al., 2003), and lysates were subjected to sucrose gradient fractionation (Figure 1A).

Migration of Gid subunits, and their relative levels in the four conditions (Figures 1A and S1), led to three major conclusions (Figure 1B). First, in carbon recovery conditions that prompt

degradation of gluconeogenesis enzymes, Gid1, Gid8, Gid5, Gid4, Gid9, and Gid2 comigrate, suggesting that these subunits form a minimal stable E3 ligase including the substrate receptor Gid4, which we term GID^{SR4}. Second, as expected, the relative level of Gid4 is highest during carbon recovery, in agreement with Gid4 expression being the glucose-regulated switch determining E3 activity (Menssen et al., 2018; Santt et al., 2008). Finally and unexpectedly, during carbon stress, the levels of all GID^{SR4} subunits except Gid4 increase, and they comigrate in the density gradients. This suggests that during energetically expensive growth on a non-fermentable carbon source, a seemingly unnecessary, inactive complex containing most Gid proteins is produced. This finding can be rationalized by the emerging concept of “anticipatory” programs preparing for a later shift in conditions. Thus, we term the complex containing Gid1, Gid2, Gid5, Gid8, and Gid9 “GID^{Ant},” surmising that when produced during carbon stress GID^{Ant} would be benign toward gluconeogenic enzymes but ready and primed for a potential later shift into glucose-containing media, which in turn would rapidly induce Gid4 expression and assembly of the active GID^{SR4} E3 ligase.

The distinct migration of Gid7 was corroborated by its substantially lower abundance relative to other Gid subunits observed by quantitative proteomic analysis of a tagged Gid8 immunoprecipitate. Also, Gid7 deletion did not appreciably affect GID assembly (Figures S1A and S1B). Although there may be settings when GID^{SR4} and GID^{Ant} further include Gid7 *in vivo*, at this point the role of Gid7 remains unknown. Gid7 may bind to a subset of GID complexes, additional factors may contribute to its binding, the interaction may be transient or low affinity, or Gid7 may play alternative roles in regulation.

To mechanistically define regulation, we generated recombinant GID^{Ant} and GID^{SR4} complexes (Figure 1C) that reconstitute known GID features. First, together with the E2 enzyme Ubc8, GID^{SR4}, but not GID^{Ant}, catalyzed robust polyubiquitylation of a recombinant gluconeogenic enzyme substrate, Mdh2 (Figure 1D). Second, in accordance with impaired degradation of a Gid substrate upon overexpressing a dominant-negative Ub K48R mutant *in vivo* (Schork et al., 1995), we found that in the context of otherwise lysineless Ub, only K48 supported substantial polyubiquitylation by our recombinant system (Figure S1D). Third, the N-terminal Pro of Mdh2 was required for its binding to Gid4 and ubiquitylation by GID^{SR4} (Figures 1E and 1F).

Three-dimensional (3D) reconstructions at 9 Å resolution obtained by cryoelectron microscopy (cryo-EM) further validated our recombinant system. Comparing the EM maps for recombinant GID^{Ant} and that purified from yeast cultured in carbon stress conditions revealed a common overall architecture (Figure 1G). Thus, it appears that the native GID^{Ant} purified from yeast, at least in terms of subunits overtly visible by cryo-EM at this resolution, is indeed a complex of Gid1, Gid2, Gid5, Gid8, and Gid9.

Prominent additional density correlating with the presence of the substrate receptor subunit Gid4 was readily visible in the map of recombinant GID^{SR4} (Figure 1G). The overall structure of the GID^{SR4} E3 resembles a clamp, with Gid4 corresponding to one jaw (Figure 1H). A high-resolution structure showed the substrate receptor linked via a scaffold to a catalytic module as described below.

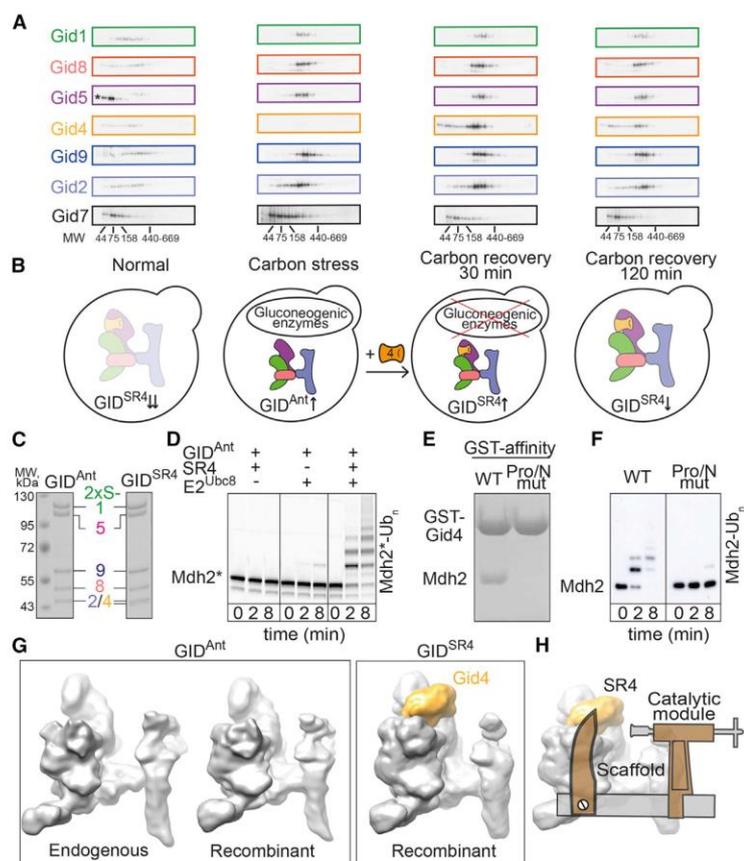


Figure 1. Regulation and Reconstitution of Yeast GID E3 Ligase Complexes

(A) Sucrose density gradient fractionation of *S. cerevisiae* lysates from cells harvested from four conditions: stationary growth in glucose-rich medium (normal), ethanol (carbon stress), and switch to glucose-rich medium for 30 and 120 min (carbon recovery 30 and 120 min, respectively). Gid subunits tagged at their endogenous loci were visualized by western blotting. Asterisk indicates non-specific anti-Flag interaction.

(B) Cartoons representing GID assemblies in different environmental conditions, on the basis of migration patterns of subunits in sucrose density gradients.

(C) Coomassie-stained SDS-PAGE of recombinant GID^{Ant} and GID^{SR4}.

(D) Fluorescent scan examining ubiquitylation of fluorescently labeled Pro/N-degron substrate Mdh2 (Mdh2^{*}). Assays test dependence on E (Ubc8) and substrate receptor (Gid4). Note GID^{Ant} + Gid4 = GID^{SR4}.

(E) Role of substrate N-terminal Pro, tested with wild-type (WT) Mdh2 or N-terminal Pro-to-Se mutant, in binding GST-tagged substrate receptor Gid4 *in vitro*.

(F) Ubiquitylation of WT or N-terminal Pro-to-Se mutant Mdh2-His₆ visualized by western blot with anti-His₆ tag antibody.

(G) Cryo-EM reconstructions at 9 Å resolution of endogenous and recombinant GID^{Ant} and recombinant GID^{SR4}. Density attributed to substrate receptor Gid4 is yellow.

(H) Clamp-like structure of GID^{SR4} assembled for substrate receptor Gid4 (SR4), scaffold, and catalytic modules.

See also Figure S1 and Tables S1 and S2.

Modular GID E3 Ligase Assembly

Refinement of the cryo-EM data for GID^{SR4} yielded a 3D reconstruction at 3.8 Å resolution (Tables 1, 2, S1, and S2; Figures S2–S6). Atomic coordinates for Gid4, Gid5, Gid8, and much of Gid1 and Gid9 were generated by a combinatorial approach involving cryo-EM maps of many variant complexes and automated and manual model building (Table 2; Figures S2–S6). Additional predicted domains from Gid1, Gid2, and Gid9 were approximately docked into lower resolution density (Figures S5B, S6A, and S6B). The multidomain nature of Gid proteins enabled structure validation through (1) testing effects of deleting specific subunits or domains on cryo-EM reconstructions; (2) strong correlations upon superimposing structures of human Gid4 substrate-binding and Gid1 Spla and Ryanodine receptor (SPRY) domains (1.0 and 0.73 root-mean-square deviation [RMSD], respectively) (Dong et al., 2018; Li et al., 2011); (3) visualizing predicted armadillo repeats in Gid5 and LisH-CTLH-CRA domains in Gid1, Gid8, and Gid9; and (4) concordance between the structure and effects of mutations observed in prior studies of GID E3 assembly *in vivo* (Braun et al., 2011; Menssen et al., 2012; Santt et al., 2008).

Overall, the EM data reveal that the GID E3 is organized around three structural and functional modules (Figure 2): the

scaffold—Gid1, Gid5, and Gid8, tightly interacting in a manner that outwardly projects protein interaction domains from each subunit; the substrate receptor—Gid4; and the catalytic module—the Gid2-Gid9 subcomplex. Details of this assembly and how it drives ubiquitylation of N-degron substrates are described below.

The Scaffold

The foundation of GID^{SR4} is an interdigitated assembly of Gid1, Gid8, and Gid5 (Figures 2A and S6C–S6F). Gid1 and Gid8 together form a heterodimeric trefoil-shaped structure. At the vertex, Gid1's LisH and C-terminal segment of the CRA domain (LisH-CRA^C), and adjacent elements, embrace paralogous regions from Gid8, rationalizing why Gid1 and Gid8 stabilize each other *in vivo* (Menssen et al., 2012). The three lobes of the trefoil are formed by (1) Gid1's SPRY domain, (2) Gid1's CTLH and N-terminal segment of the CRA domain (CRA^N), and (3) Gid8's CTLH-CRA^N domain and adjacent sequences (Figure S6C). The distal ends of the CTLH-CRA^N domains from both Gid1 and Gid8 radiate away from the core, while a continuous V-shaped surface between Gid1's SPRY and Gid8's CTLH domain engages an extended complementary surface

Table 1. Cryo-EM Data Collection, 3D Reconstruction, and Map Refinement

	GID ^{SR4}	GID ^{Ant}	Endogenous GID ^{Ant}	GID ^{SR4} ΔRINGs	GID ^{SR4}	GID ^{Ant}	GID ^{Scaffold} Plus SR ^{Gid4}	GID ^{Scaffold} Plus SR ^{Gid10}	GID ^{Scaffold}	GID ^{SR4} Minus Gid2/ ΔGid9 ^{RING}
EMDB code	N/A	N/A	10331	10332	10327	10326	10330	10329	10328	10333
Microscope/ detector	Arctica/ Falcon III	Arctica/ Falcon III	Arctica/ Falcon III	Arctica/ Falcon III	Krios/K2	Krios/ K2	Krios/K3	Krios/K3	Krios/K3	Krios/K3
Particles	387,982	388,646	192,972	378,602	615,139	994,904	1,106,310	2,132,595	2,132,595	1,645,121
Pixel size	1.612	1.997	1.612	1.612	1.06	1.06	1.09	1.09	1.09	1.09
Defocus range (μM)	1.5–3.5	1.5–3.5	1.5–3.5	1.5–3.5	1.1–3.2	1.1–3.2	1.1–3.2	1.1–3.2	1.1–3.2	1.1–3.2
Voltage (kV)	200	200	200	200	300	300	300	300	300	300
Electron dose (e/Å ² /s)	21.3	23.07	22.4	21.4	6.675	7.88	13.57	13.8	13.8	13.62
Exposure time (s)	3	3	3	3	8	7	4	4	4	4
Map resolution (Å) ^a	5.1	8.3	9.3	7.3	3.8	3.7	3.4	3.8	3.8	3.2
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	4.5–11.1	6.0–24.0	7.9–25.6	6.0–23.1	3.5–15.8	3.5– 11.9	3.1–7.2	3.5–6.5	3.5–7.6	3.0–8.3
Map sharpening B-factor (Å ²)	–246	–301	–618	–372	–92	–119	–114	–109	–99	–80

See also [Figures S2, S3, S4, and S6](#).

^aAccording to the Fourier shell correlation (FSC) cutoff criterion of 0.143 defined in [Rosenthal and Henderson \(2003\)](#).

from Gid5. Gid5's armadillo repeats stack in tandem in a continuous solenoid of roughly one and a half superhelical turns, with the N-terminal domain (NTD) filling the groove between Gid1 and Gid8, and a C-terminal domain (CTD) radiating outward ([Figures 2A, S6C, and S6E](#)). The scaffold is further buttressed by loops from all three proteins extending distances up to ≈ 70 Å to engage one another.

A protein interaction domain from Gid5 recruits the substrate receptor Gid4, and Gid8 binds the catalytic module Gid2-Gid9 ([Figures S6C and S6F](#)). Weak density corresponding to Gid1's CTLH domain also projects outward. Although the structural role of Gid1's CTLH domain is presently unknown, we speculate it binds Gid7 on the basis of its mutation specifically impairing this interaction *in vivo* ([Menssen et al., 2012](#)).

Scaffold Binding to Substrate Receptor Gid4 Generates GID^{SR4}

The substrate receptor, which recruits proteins for ubiquitylation, is an essential E3 ligase element. A prior structure showed that human Gid4's substrate-binding domain is largely a β -barrel, with a funnel-shaped opening at one end binding to short peptides via their N-terminal Pro ([Figure 3A](#)) ([Dong et al., 2018](#)). The structure of GID^{SR4} shows, in turn, how this substrate-binding domain is incorporated into an active E3 ligase ([Figures 2B and 3](#)). Gid4's C-terminal eight residues anchor the interaction, by extending into a channel in the concave surface of Gid5^{CTD} ([Figures 3B and 3C](#)). Here, successive Gid4 side chains protrude in opposite directions. Aromatic residues on one side fill pockets between Gid5 armadillo repeats. Those on the other side contribute to a composite Gid4/Gid5 interface with an aliphatic

stripe across Gid4's barrel. Indeed, mutation of key Gid4 and Gid5 contact residues impaired GID^{SR4}-catalyzed substrate ubiquitylation *in vitro* ([Figures 3D and 3E](#)). *In vivo*, the Gid5 mutations substantially impaired degradation of the gluconeogenic enzyme Fbp1, as did mutation of Gid4's C-terminal anchor. Although individual conservative amino acid substitutions in Gid4's substrate-binding domain did not have a measurable effect, introduction of a bulky residue or multiple Ala mutations caused substantially impaired glucose-induced degradation of Fbp1 ([Figures 3D and 3E](#)).

Additionally, weaker EM density showed Gid4 residues 91–116, upstream of the substrate-binding domain, meandering over 65 Å to loosely wrap around to the convex face of Gid5. Also a loop from Gid1's SPRY domain contacts a peripheral helical portion of Gid4's substrate-binding domain ([Figure 3F](#)). However, these residues are neither conserved nor essential for GID^{SR4} activity *in vitro* or *in vivo*, suggesting auxiliary roles ([Figure 3F](#)).

Dynamic Gid5 CTD in Anticipation of a Substrate Receptor

To understand the structure of the GID complex expressed during carbon stress, EM data for recombinant GID^{Ant} were refined to yield a 3D reconstruction at 3.7 Å resolution ([Figures 4, S2, and S4A](#); [Tables 1 and S1](#)). Comparison with the map of GID^{SR4} showed a striking difference in the density corresponding to Gid5's substrate-receptor-binding CTD, which is blurred in GID^{Ant}. The Gid5 armadillo repeats are visible, but poor density precluded refinement to high resolution ([Figure 4A](#)). Thus, we speculate that anticipation is manifested by conformational

Table 2. Model Refinement and Validation Statistics

Table 2. Model Refinement and Validation Statistics	
Map	GID ^{SR4} Minus Gid2/ Δ Gid9 ^{RING}
Refinement	
Model Composition	
Non-hydrogen atoms	16,071
Protein residues	2,031
Resolution	3.1
FSC map versus model@0.143 ^a	
RMS Deviations	
Bond lengths (Å)	0.005
Bond angles (Å)	0.960
Validation	
Molprobrity score/percentile	1.71
Clashscore/percentile	5.77
Rotamer outliers (%)	0.23
Ramachandran Plot	
% favored	94.15
% allowed	5.8
% outliers	0.05

See also [Figures S5A and S6](#).

^aAccording to the map-versus-model correlation coefficient definitions in [Afonine et al. \(2018\)](#).

dynamics of the Gid5 CTD armadillo repeats prior to capturing and curling around a substrate receptor subsequently available upon change in environmental conditions.

Because GID^{Ant} and GID^{SR4} are structurally similar beyond Gid5's CTD and its associated Gid4 ([Figure 4B](#)), we hypothesized that GID complexes may display intrinsic catalytic activity irrespective of ability to recruit substrate. To test this, we used an assay that monitors substrate-independent activation of E2~Ub intermediates ([Petroski and Deshaies, 2005; Wenzel et al., 2011](#)). First, the reactive Ubc8~Ub intermediate (the symbol ~ refers to thioester linkage) was generated enzymatically, and this reaction was quenched. Next, lysine was added simultaneously with various versions of GID E3s. Ub transfer from Ubc8, presumably to unanchored lysine, was monitored by both disappearance of Ubc8~Ub and appearance of free Ubc8 in SDS-PAGE. Although the Ubc8~Ub intermediate was relatively stable on its own over time, GID^{Ant} stimulated its rapid discharge with little effect of titrating a version of Gid4 suitable for substrate recruitment ([Figures 1D and 4C](#)). Thus, GID^{Ant} is intrinsically competent at activating Ub transfer even without a recruited N-degron substrate or its receptor.

A Family of Related GID E3s

The concept of a multiprotein E3 ligase that facultatively associates with a substrate receptor is conceptually reminiscent of cullin-RING and anaphase-promoting complex E3 families. However, these E3s use sets of interchangeable substrate receptors for distinct regulation ([Alfieri et al., 2017; Lydeard et al., 2013; Watson et al., 2019](#)). Thus, we hypothesized that other GID substrate receptors may exist, and we identified the ORF YGR066C as encoding a protein displaying homology to

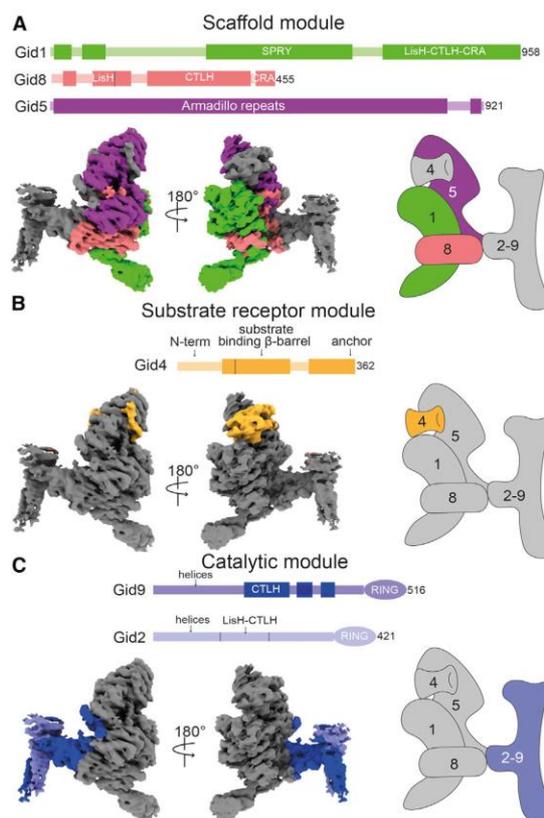


Figure 2. GID^{SR4} E3 Ligase Modular Architecture

Each panel shows a different module as a domain schematic (top), two view of cryo-EM density (bottom left), and a cartoon (bottom right). Subunits within module are color-coded, with others in gray. Darker boxes in the domain schematic represent regions of the density map into which an atomic model was built.

(A) Scaffold module comprising Gid1 (green), Gid8 (salmon), and Gid5 (purple) (B) Substrate receptor module consisting of Gid4 (orange). (C) Catalytic module composed of Gid2 (light blue) and Gid9 (dark blue). See also [Figures S2–S6](#) and [Tables S1 and S2](#).

Gid4, including the Gid5-binding hydrophobic stripe and C-terminal anchor ([Figure S4C](#)). While our manuscript was under consideration, YGR066C was published as a GID E3 substrate receptor and renamed “Gid10” ([Melnykov et al., 2019](#)). We have adopted this nomenclature and had already independently performed several experiments suggesting that Gid10 is a bona fide alternative substrate receptor for a GID E3. First, bacterial-expressed Gid10 binds our recombinant GID^{Ant} ([Figure 5A](#)). Second, Gid10 confers onto GID^{Ant} *in vitro* ubiquitylation activity toward an N-degron substrate, albeit with far lower efficiency than Gid4 ([Figure 5B](#)). Third, a 3.8-Å-resolution cryo-EM reconstruction of Gid10 bound to the scaffold module showed an overall similar structure to the Gid4-bound complex, including homologous placement of Gid10's C-terminal anchor and a β -barrel

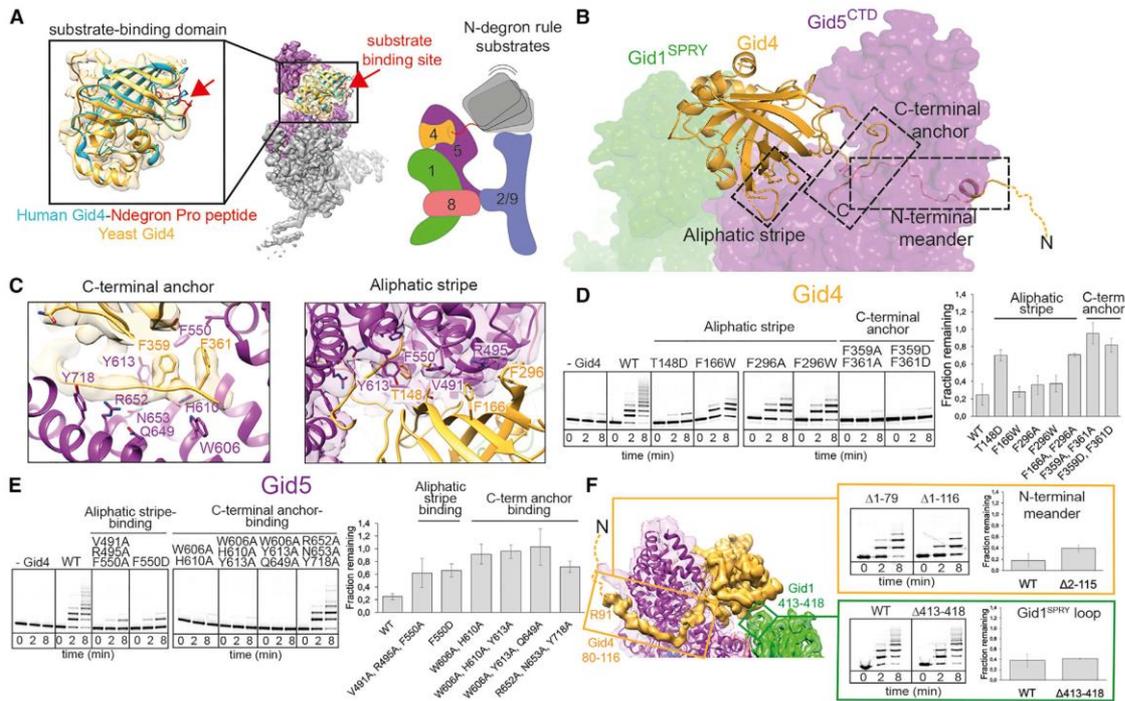


Figure 3. Formation of GID^{SR4} E3 Ligase through Incorporation of Substrate Receptor Gid4
 (A) Overlay of scaffold-bound *S. cerevisiae* Gid4 substrate-binding domain with crystal structure of human Gid4 bound to N-terminal Pro peptide (PDB: 6CDC) (left), showing potential substrate-binding site with a red arrow. Cartoon of N-degron substrate binding by GID^{SR4} (right).
 (B) Overview of Gid4 elements binding to GID scaffold. Gid4 (yellow cartoon) binds Gid5^{CTD} (purple surface) via a C-terminal anchor (C, C terminus), an aliphatic stripe and an N-terminal meander. Disordered residues connecting to N terminus (N) shown as dotted line.
 (C) Close-up of Gid4 (yellow) C-terminal anchor and aliphatic stripe interactions with Gid5 (purple). Residues mutated in (D) and (E) are represented as sticks.
 (D) Assays testing importance of Gid4 residues in aliphatic stripe and C-terminal anchor on *in vitro* Mdh2 ubiquitylation and *in vivo* Fbp1 degradation (quantified as fraction from time 0 remaining after switching from carbon stress to carbon recovery).
 (E) Assays testing importance of Gid5^{CTD} residues that interact with Gid4 aliphatic stripe and C-terminal anchor on *in vitro* Mdh2 ubiquitylation and *in vivo* Fbp1 degradation.
 (F) Left: structure and EM density map depicting auxiliary interactions between Gid5^{CTD} and Gid4 N-terminal meander (residues 80–116) and Gid1 SPRY domain loop (residues 413–418) with peripheral helical insertion in Gid4. Right: assays testing if these elements are not essential for *in vitro* ubiquitylation of Mdh2 and *in vivo* Fbp1 degradation.
 (D, E, and F) Error bars represent SD (n ≥ 3).
 See also Figures S5 and S6 and Table S2.

domain poised to bind N-degron substrates (Figures 5C, 5D, S2, and S4B; Tables 1 and S1). Indeed, deletion of Gid10's C-terminal anchor abrogates Gid10-dependent ubiquitylation of the recombinant substrate Mdh2 (Figure 5B).

Comparing EM maps with the two substrate receptors in detail shows a potential for the slightly different placement of Gid10 and Gid4 relative to the scaffold (Figures 5C and 5D). This raises the possibility that orientation of substrate-binding domains may underlie mechanisms regulating substrate degradation under different cellular conditions.

Although deletion of the Gid10 gene in yeast did not affect degradation of known Gid substrates after carbon source switching (data not shown and Melnykov et al., 2019), prior transcriptomics, along with our analyses of protein levels, do not

imply Gid10 expression under these conditions. Rather, Gid10 mRNA is expressed during various stresses, including high salinity and heat shock (Gasch et al., 2000; Wanichthanarak et al., 2014). Indeed, we observed Gid10 protein induction under these conditions, presumably leading to its incorporation into an alternative GID^{SR10} E3 complex (Figures 5E and 5F; Melnykov et al., 2019).

Embedding of a RING-RING Catalytic Module within Multisubunit E3 Ligase

Most E3 ligases depend on one or more RING domains binding to the E2 and the Ub in a thioester-linked E2~Ub intermediate, thereby stabilizing a “closed conformation” that activates discharge of Ub from the E2 active site (Dou et al., 2012;

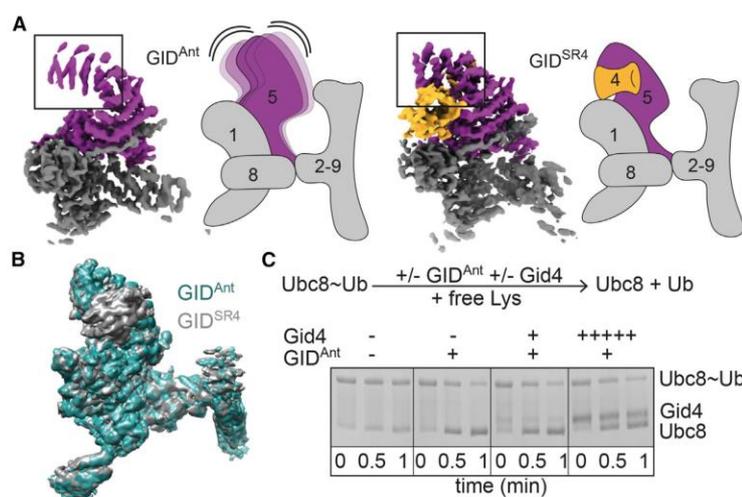


Figure 4. Structural Anticipation by GID^{Ant}

(A) Cryo-EM maps and cartoons showing GID^{Ant} and GID^{SR4} with Gid5 purple and Gid4 yellow. Black boxes highlight relatively weaker Gid5^{CT} density in GID^{Ant}, which we presume represent conformational flexibility in the absence of substrate receptor.

(B) Superposition of cryo-EM maps for GID^{SF} (gray) and GID^{Ant} (aqua) at low contour.

(C) Assay testing substrate-independent E3 activity. First, the Ubc8~Ub intermediate is generated enzymatically and this reaction is quenched. Next, free lysine was added. Reactivity probed by loss of Ubc8~Ub and appearance of Ub was tested without an E3, or with GID^{Ant} alone, or with addition of equimolar and 5× excess of Gid4. See also Figures S2 and S4A and Table S1.

Plechanovová et al., 2012; Pruneda et al., 2012). Thus, we sought to identify the structural locations and functional roles of the Gid2 and Gid9 RINGS. Having already placed the scaffold and substrate receptor modules, we attributed the remaining density to the catalytic module. This adopts a T-shaped structure, where the base of the T connects the catalytic domain to the scaffold (Figures 2, 6A, 6B, and S6). Here, Gid9's CTLH-CRA^N domain heterodimerizes with that from Gid8 in a manner resembling a pillar affixed to a base.

The top of the T appears to comprise multiple heterodimeric Gid2-Gid9 subdomains. The relatively poor resolution of this region may suggest mobility of the Gid2-Gid9 subdomains with respect to one another and relative to the scaffold. Although it was not possible to determine which elements derive from Gid2 or Gid9, the density was sufficiently visible at low contour to approximately localize predicted domains (Figures 6B, 6C, S5B, and S6B). One side of the top of the T is a four-stranded coiled coil, which we speculate corresponds to helices predicted at the N termini of Gid2 and Gid9 (Kelley et al., 2015).

Significantly, the structure of the catalytic core appears to place the Gid2 and Gid9 RING domains in a canonical RING-RING dimer assembly in the clamp-like structure of GID^{SR4}, forming the second “jaw” that faces Gid4 (Figures 1H and 6C). We arrived at this conclusion after considering that the remainder of the T-structure consists of two subdomains, and then roughly attributing the unassigned Gid2-Gid9 features. The subdomain at the extreme edge of the complex can be fitted with a homology model of the Gid2 and Gid9 RING domains superimposed on a canonical RING-RING dimer assembly found in many E3 ligases (Kelley et al., 2015). Notably, the notion that the RINGS heterodimerize is consistent with prior mutations of zinc ligands within either protein, which presumably lead to RING misfolding, decreasing Gid2-Gid9 interactions and eliminating glucose-induced substrate degradation *in vivo* (Braun et al., 2011; Regelmann et al., 2003).

To validate the locations of the RINGS we examined mutant versions of GID^{SR} lacking these domains by cryo-EM. Refinement of the data led to two major classes. One, indeed, showed selective elimination of the density we attribute to a Gid2-Gid9 RING-RING dimer, while the second class superimposed with the map obtained for a sample lacking the entire Gid2 subunit as well as the Gid9 RING domain (Table S1). This is consistent with heterodimeric assembly of the RING domains contributing to Gid2 incorporation into a GID E: *in vivo* (Braun et al., 2011). Last, we speculate that the remaining density at the T-junction corresponds to a heterodimeric assembly comprising the LisH-CRA^C domains from Gid2 and Gid1 and/or the ensuing CTLH domain from Gid2, which would match the size of this subdomain (Figure S6B). Moreover, this hypothesis is consistent with the relative orientation of Gid9's CTLH-CRA^N domain, which is inserted between the LisH and CRA^I elements in the sequence of Gid9.

Model of the Catalytic Center Suggests that the Heterodimeric RING Activates a Single Ubc8~Ub Facing Substrate

As a first step toward structurally modeling GID^{SR4}-catalyzed ubiquitylation, each RING domain docked into the EM density was superimposed with a prior structure of an isolated RING E2~Ub complex (Dou et al., 2012; Plechanovová et al., 2012; Pruneda et al., 2012), and then the docked E2 was replaced with Ubc8. Even with uncertain position of the Gid2-Gid1 RING-RING dimer, the structural modeling suggested that only one of the two RING domains would place Ubc8 to face the Gid5-bound substrate receptor.

To test if the Gid2 and/or Gid9 RING primarily binds Ubc8 or plays a supporting role in activating the Ubc8~Ub intermediate residues were selected for mutation on the basis of homology to three hallmark elements: (1) a hydrophobic surface that binds E2 loops conserved in Ubc8; (2) potential “linchpin” residues which can be located on either side of the domain, but irrespective of location insert between the E2 and its thioester-linked Ub to stabilize the noncovalent interface between them; and (3) a non-RING priming element flanking a RING sequence that function:

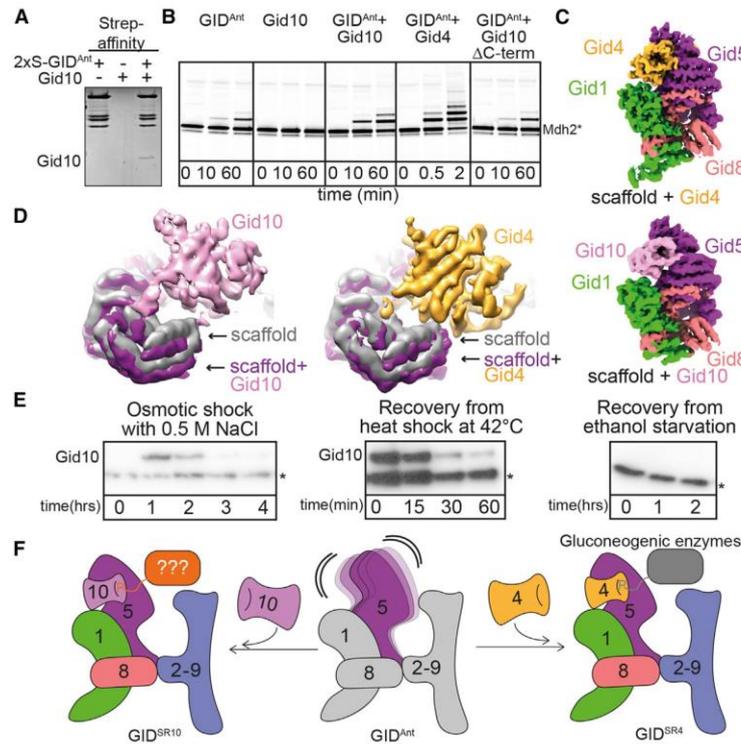


Figure 5. A Family of Multisubunit GID E3s with Swappable Substrate Receptors

(A) Streptactin pull-down of GID^{Ant} testing binding of Gid10. (B) Ubiquitylation assay testing potential of Gid10 to act as a substrate receptor for the GID^{SR4} substrate Mdh2. (C) Cryo-EM maps of Gid4- and Gid10-bound GID scaffold. (D) Close-up views of overlays of maps of GID scaffold alone and bound to Gid10 or Gid4. (E) Western blots showing expression of Gid10, tagged at the endogenous locus, under different environmental conditions. Asterisk indicates a protein interacting non-specifically with anti-Flag antibodies. (F) Model for family of GID E3s with interchangeable substrate receptors. See also Figures S2, S3B, S4B, and S4C and Table S1.

in trans to allosterically stabilize the closed conformation of the E2~Ub intermediate bound primarily to the opposite RING in a dimer (Figure 6D) (Brown et al., 2014; Dou et al., 2012; Kelley et al., 2015; Plechanovová et al., 2012; Pruneda et al., 2012; Scott et al., 2014; Zheng et al., 2000). Effects on GID^{SR4} E3 ligase activity *in vivo* were tested by introducing mutations into tagged versions of Gid2 and Gid9 expressed from their endogenous loci (Figure 6E). Effects of point mutations in predicted E2-binding and linchpin residues of Gid2 mirrored effects of wholesale deletion of Gid2's RING domain on glucose-induced degradation of Fbp1, while there was a relatively minimal effect of mutating Gid2's candidate non-RING priming element. The crucial role for the Gid2 RING's E2-binding site was also confirmed for GID^{SR4} E3 ligase activity *in vitro*. In contrast, the opposite pattern was observed for the Gid9 mutants, where only the candidate non-RING priming element significantly abrogated activity. The results suggest that Gid2's RING binds and activates the Ubc8~Ub intermediate, assisted by a non-RING priming element from Gid9, to face the substrate receptor (Figures 6E and 6F).

Model of GID^{SR4} Ubiquitylating an N-Degron Gluconeogenic Enzyme

The substrate-binding site on Gid4 is ≈50 Å away from the modeled catalytic center. Although the relatively weak EM density corresponding to the catalytic domain (Figure 6C) suggests flexibility, perhaps for conformational changes during catalysis, it is also possible that the large gap accommodates substrates

of GID^{SR4}, which are large oligomeric enzymes. To model ubiquitylation, the substrate Mdh2 was selected because (1) robust *in vitro* ubiquitylation of bacterially expressed Mdh2 demonstrated that post-translational modifications are not required for its N-degron-based substrate targeting (Figures 1D–1F), and (2) as an ≈80 kDa homodimer with 34 lysines, Mdh2 is the smallest and structurally most simplistic of known GID^{SR4} substrates (Figure 7). To place Mdh2, the N-terminal four residues were modeled on the basis of the prior structure of the human Gid4 substrate-binding domain bound to a 4-mer peptide. Next, an Mdh2 model was manually rotated while roughly constraining the location of the N-terminal-most ordered residue (L14) proximal to the substrate-binding site on Gid4 (Dong et al., 2018; Kelley et al., 2015).

Overall, the model suggests that some, but not all, Mdh2 lysines would preferentially access the ubiquitylation active site (Figure 7A). To test this, we used mass spectrometry to map sites that are ubiquitylated *in vitro* (Figure S7). Notably, the top sites include a cluster of K254, K256, and K259, as well as K330, and to a lesser extent K360 and K361, for which the 10-residues between Mdh2's N terminus and globular domain would easily accommodate the ≈10, ≈20, and ≈15 Å required, respectively, for Mdh2 to simultaneously engage Gid4 and approach the Gid2 RING activated Ubc8~Ub intermediate. Furthermore, substituting these residues with arginines severely impaired Mdh2 ubiquitylation (Figure 7B). Thus, although detailed knowledge awaits further structural studies, the EM data presented here enable the generation of a geometrically reasonable model for N-degron substrate ubiquitylation (Figure 7C).

DISCUSSION

The cryo-EM reconstructions reported here reveal E3 ligase assemblies that vary in response to extracellular stimuli

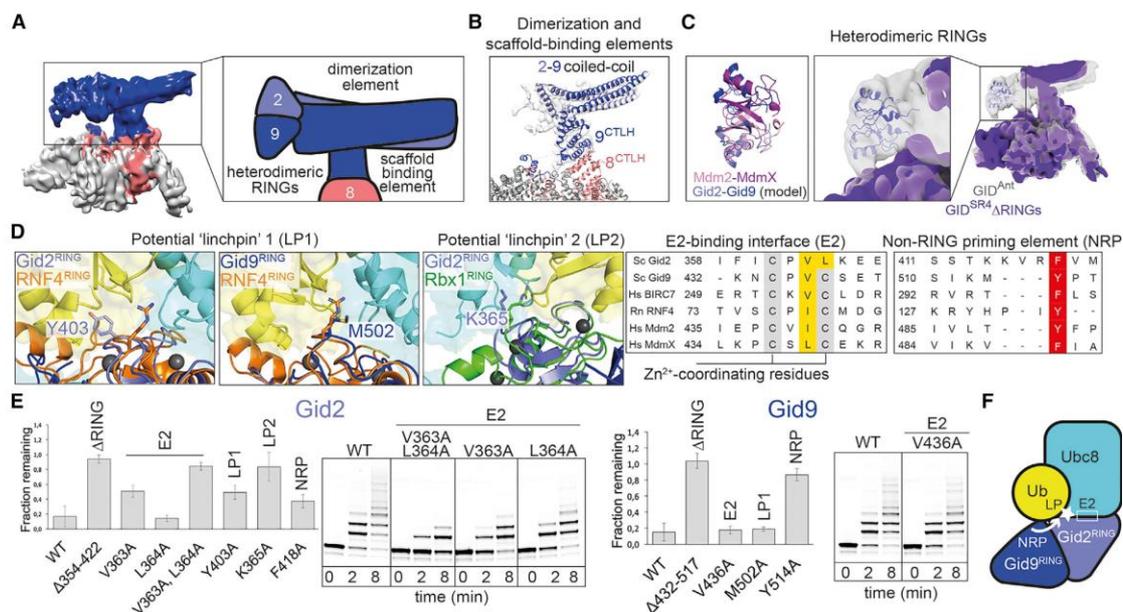


Figure 6. GID Catalytic Module

(A) Left: T-shaped Gid2-Gid9 catalytic module in low contour EM map of GID^{Ant}. Right: catalytic module elements shown in cartoon: scaffold-binding domain interacting with Gid8 (salmon), dimerization region, and heterodimeric RINGs.

(B) Homology models of catalytic module elements fitted into EM map generated by focused refinement and signal subtraction. The atomic models of Gid8 and Gid9 CTLH-CRA^N domains are shown, as is an approximated coiled-coil docked in additional density.

(C) The Gid2-Gid9 RING-RING domain was modeled in triangular density at the tip of the T-shaped catalytic module, as follows: (1) RING-RING domain was generated by superimposing homology models of Gid2 and Gid9 RINGs onto MDM2-MDMX structure (PDB: 2VJE) (left). (2) Model of Gid2-Gid9 RING-RING domain was docked into map of GID^{Ant} (center). (3) Density attributed to the Gid2-Gid9 RING-RING domain was not visible in EM map of GID^{SR4} with the RING deleted (solid violet map, right).

(D) Left: candidate Gid2 and Gid9 RING "linchpins" (LP) identified by superimposing their homology models with crystal structures of RNF4 (PDB: 4AP4) and RBX1 (PDB: 4P50) bearing linchpin residues R181 and R46, respectively. Corresponding Gid2 and Gid9 residues are shown as sticks. Right: sequence alignments of Gid2 and Gid9 with well-characterized RING domains identified potential E2-binding (E2, yellow) and non-RING priming element (NRP, red) residues.

(E) Assays testing effects of Gid2 and Gid9 mutations on GID E3 activity, Fbp1 degradation *in vivo*, and Mdh2 ubiquitylation *in vitro*. Error bars represent SI ($n \geq 3$).

(F) Cartoon summarizing model for Ubc8-Ub activation by Gid2-Gid9 RING-RING domains on the basis of mutational analysis shown in (E).

See also Figures S3C, S4A, S5, and S6 and Table S1.

(Figures 1, 4, 5, and S1), molecular mechanisms underlying their regulation (Figures 1, 2, 3, 4, 5, and 6), and a framework for GID E3 ligase-dependent ubiquitylation (Figure 7). The structural data also provide broad insights into major families of E3 ligases, namely, those recognizing terminal degrons and those displaying RING-RING catalytic domains. The modular multiprotein GID^{SR4} E3 assembly displays clamp-like properties, established by a central scaffold connecting the two jaws: a variable substrate receptor and the catalytic domain. The structure enables binding of a substrate's unfolded N terminus to Gid4, to direct lysines from a folded domain into the ubiquitylation active site. From the perspective of the other side of the complex, it seems that the RING-RING dimer is the culmination of an intricate heterodimeric Gid2-Gid9 assembly that activates the Ubc8~Ub intermediate facing the N-degron substrate. Some other E3s, for example,

BRCA1-BARD1 or HDM2-HDMX, that contain heterodimeric RING-RING domains, may likewise rely on complex, interconnected assemblies to couple a single active site with a substrate for ubiquitylation.

Although it has long been recognized that GID^{SR4} is part of the yeast response to environmental conditions via glucose-induced degradation of gluconeogenic enzymes (Sant et al., 2008), our data indicate that generation of a GID^{Ant} complex also occurs in response to an extracellular stimulus: carbon stress (Figure 1A). As GID^{Ant} would be inactive toward recruited substrate in the absence of a substrate receptor, we hypothesize that production of this complex allows cells to adapt more rapidly to potential later changes in the extracellular milieu. Our data raise the possibility that carbon stress may prepare cells for a potential return to nutrient-rich conditions. We also cannot rule out the possibility that GID^{Ant} could be coupled to a yet unknown substrate

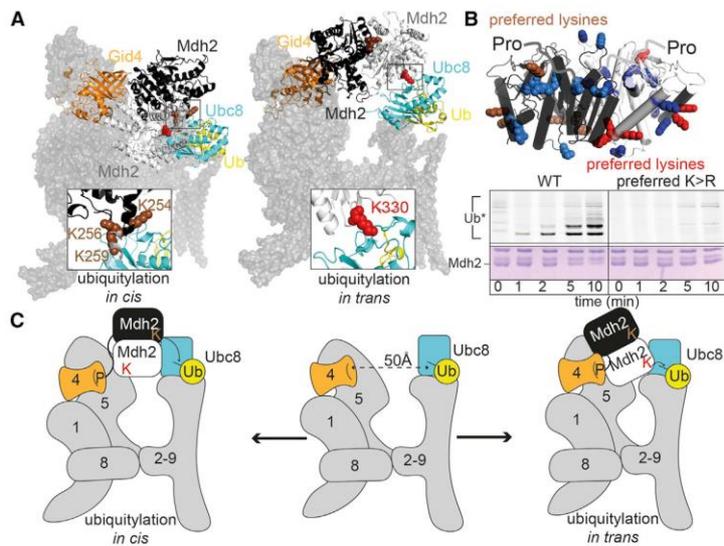


Figure 7. Model of GID^{SR4}-Catalyzed Ubiquitylation of N-Degron Substrate Mdh2

(A) Structural models for substrate ubiquitylation by GID^{SR4}, with a homology model for Mdh2 (protomers in black and gray) placed with its N-terminal Pro binding Gid4 and candidate lysine targets in the active site of a modeled Ubc8~Ub intermediate.

(B) Homology model of Mdh2 dimer showing preferred target lysines identified by mass spectrometry in brown and red and other lysines in blue. Assay testing effects of Arg replacement for preferred target lysines on Mdh2 ubiquitylation. Top: fluorescent scan detecting ubiquitin. Bottom: loading controls showing Mdh2, Gid2 (2), and Gid4 (4).

(C) Cartoon representing the structural models for GID^{SR4} ubiquitylation. The clamp-like structure enables multiple configurations for ubiquitylation of preferred lysines from a folded gluconeogenic enzyme substrate. See also Figure S7.

receptor to allow ubiquitylation of a distinct set of proteins during carbon stress.

We were puzzled by the apparently counterintuitive carbon stress-induced production of GID^{Ant} in anticipation of relief from starvation. It is conceptually appealing to envisage microbial anticipatory signaling cascades in terms of one stress serving as a signal for cells to cross-prepare for a looming new stress (Mitchell et al., 2009; Tagkopoulos et al., 2008). Our discovery that recombinant GID^{Ant} also binds the Gid4-like protein YGR066C/Gid10, which is induced under several distinct stress conditions (Figure 5; Melnykov et al., 2019), offers a potential mechanism for how GID^{Ant} could act as a multifaceted hub integrating responses to various extracellular stimuli. We speculate that carbon stress-induced production of GID^{Ant} may enable cells to prepare for ensuing osmotic stress or heat stress through the production of a Gid10-associated GID^{SR10} E3 ligase. It seems plausible that Gid10's substrates could be regulators of glycerol or salt intake, protein synthesis, or general stress responses. Mechanistically, it seems likely that substrate selectivity will be influenced not only by protein expression changes under different metabolic conditions but also by subtle differences in the β -barrel domains of Gid4 and Gid10 and their orientations relative to the scaffold (Figure 5). Future studies will be required to identify substrates of a GID^{SR10} E3, to visualize substrate ubiquitylation in action, and to understand cross-talk between GID^{Ant} assembly, association with multiple substrate receptors, and coupling responses to varying extracellular stimuli.

What does an E3 ligase in “anticipation” look like? Our cryo-EM reconstruction of GID^{Ant} suggested motion of Gid5's substrate receptor-binding CTD (Figure 4). Dynamic opening and closing of Gid5 could enable binding, release, and exchange of the substrate receptor. At this point, the molecular stimuli and structural mechanisms underlying substrate receptor dissociation or exchange remain unknown, although Gid4 turnover has

been shown to depend on GID E3-dependent ubiquitylation (Menssen et al., 2018).

We speculate that in humans, a GID^{Ant} complex will parallel its yeast counterpart and act as a hub integrating various signals, presumably through many additional binding partners, in turn leading to cell fate determination. Although binding to Gid4 likely generates a human GID^{SR4} E3 targeting substrates with N-terminal prolines (Dong et al., 2018), the functions of other partner proteins, including Gid7, remain elusive, and may regulate activity of GID^{SR4}, or perhaps form alternative assemblies with GID^{Ant} or Gid subunits. Indeed, two Gid7 homologs, along with many other proteins, have been shown to co-precipitate with human Gid subunits (Boldt et al., 2016; Huttlin et al., 2017; Lampert et al., 2018). In addition, the human GID E3 ligase has been reported to ubiquitylate a substrate that does not bear an N-terminal proline (Lampert et al., 2018), despite the apparent requirement of a proline to bind human Gid4 (Dong et al., 2018). This could potentially be reconciled on the basis of our discovery that the GID E3 ligase is not a singular complex but a family of E3 ligases with different substrate receptors (Figure 5). Additional human Gid subunits could substitute for Gid4, modulate substrate specificity, or localize the GID complex (Boldt et al., 2016; Lampert et al., 2018).

The concept of multiple GID E3 assemblies responding to different environmental stimuli is reminiscent of other multiprotein E3 ligases (e.g., cullin RING-ligases) and hubs such as mTOR that integrate signaling with various downstream functions required in certain cellular contexts (González and Hall, 2017; Lydeard et al., 2013; Saxton and Sabatini, 2017). Regulation of these assemblies through interchangeable receptors provides a framework for investigating the GID family. Are there cellular exchange factors that promote swapping Gid4 and Gid10 (Pierce et al., 2013), or inhibitory factors (Duda et al., 2012; Lyapina et al., 2001)? Is GID regulated by modifications or metabolites (González and Hall, 2017; Saxton and Sabatini,

2017)? Does substrate binding play a role in substrate-receptor selection (Emberley et al., 2012; Enchev et al., 2012)? And most curiously, are there other presently unknown substrate receptors? Although future studies will be required to unveil the molecular pathways and mechanisms underlying these complexities, the stunning structural intricacies of the seemingly simple yeast GID N-degron-targeting system, now revealed more than 25 years since the discovery of glucose-induced degradation, provide a blueprint for understanding this important family of multi-subunit E3 ligases.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **METHOD DETAILS**
 - Yeast strains and growth conditions
 - Fbp1 degradation assays
 - Affinity Enrichment Mass Spectrometry (AE-MS)
 - Protein digestion of *in vitro* ubiquitylation assays
 - LC-MS/MS sample preparation
 - LC-MS/MS Measurements
 - LC-MS/MS raw data processing
 - Density Fractionation by sucrose gradients
 - Plasmids preparation and Mutagenesis
 - Protein expression and purification for cryo-EM
 - Cryo EM sample preparation and Imaging
 - Data processing
 - Model building and refinement
 - Protein expression and purification for biochemical assays
 - Biochemical assays
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.molcel.2019.10.009>.

ACKNOWLEDGMENTS

We thank J. Frye and support from ALSAC/St. Jude for cloning Gid ORFs; A. Varshavsky for plasmids for promoter reference assays in yeast; J. Kellermann for assistance in all work in the Schulman lab; N. Nagaraj, V. Sanchez Caballero, N. Krombholz, and A. Wehner for intact mass spectrometry; I. Paron, C. Deiml, and J. Mueller for mass spectrometry maintenance; and M. Strauss, D. Scott, J.W. Harper, and members of the Schulman lab for discussions. This work was supported by the Max Planck Society.

AUTHOR CONTRIBUTIONS

S.Q. established recombinant GID and prepared cryo-EM samples. S.Q., D.B., T.S., and J.R.P. collected cryo-EM data. S.Q. and J.R.P. determined, refined, and validated atomic models. C.R.L. and V.B. generated yeast strains. C.R.L., V.B., and S.Q. performed yeast biochemistry. S.Q., J.C., and D.S. performed *in vitro* biochemistry. O.K. and F.M.H. performed mass spectrometry supervised by M.M. S.v.G. produced baculoviruses and insect cells expressing

GID complexes. S.Q., J.C., D.S., C.R.L., and B.A.S. wrote the paper. A.F. and B.A.S. coordinated and supervised.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 2, 2019

Revised: September 4, 2019

Accepted: October 8, 2019

Published: November 7, 2019

REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N. Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C. Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Bio* **74**, 531–544.
- Aliferi, C., Zhang, S., and Barford, D. (2017). Visualizing the complex function and mechanisms of the anaphase promoting complex/cyclosome (APC/C). *Open Biol.* **7**, 170204.
- Boldt, K., van Reeuwijk, J., Lu, Q., Koutroumpas, K., Nguyen, T.M., Texier, Y. van Beersum, S.E., Horn, N., Willer, J.R., Mans, D.A., et al.; UK10K Rare Diseases Group (2016). An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. *Nat. Commun.* **7**, 11491.
- Braun, B., Pfirrmann, T., Menssen, R., Hofmann, K., Scheel, H., and Wolf, D.-H. (2011). Gid9, a second RING finger protein contributes to the ubiquitin ligase activity of the Gid complex required for catabolite degradation. *FEBS Lett* **585**, 3856–3861.
- Brower, C.S., Platkov, K.I., and Varshavsky, A. (2013). Neurodegeneration associated protein fragments as short-lived substrates of the N-end rule pathway. *Mol. Cell* **50**, 161–171.
- Brown, N.G., Watson, E.R., Weissmann, F., Jarvis, M.A., VanderLinden, R. Grace, C.R.R., Frye, J.J., Qiao, R., Dube, P., Petzold, G., et al. (2014) Mechanism of polyubiquitination by human anaphase-promoting complex RING repurposing for ubiquitin chain assembly. *Mol. Cell* **56**, 246–260.
- Burnley, T., Palmer, C.M., and Winn, M. (2017). Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.* **73**, 469–477.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M. Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.
- Chen, S.J., Wu, X., Wadas, B., Oh, J.H., and Varshavsky, A. (2017). An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355**, eaal3655.
- Chiang, M.C., and Chiang, H.L. (1998). Vid24p, a novel protein localized to the fructose-1, 6-bisphosphatase-containing vesicles, regulates targeting of fructose-1,6-bisphosphatase from the vesicles to the vacuole for degradation. *J. Cell Biol.* **140**, 1347–1356.
- Chiang, H.L., and Schekman, R. (1991). Regulated import and degradation of cytosolic protein in the yeast vacuole. *Nature* **350**, 313–318.
- Choi, W.S., Jeong, B.C., Joo, Y.J., Lee, M.R., Kim, J., Eck, M.J., and Song H.K. (2010). Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. *Nat. Struct. Mol. Biol.* **17**, 1175–1181.
- Ciechanover, A. (2012). Intracellular protein degradation: from a vague idea through the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Biochim. Biophys. Acta* **1824**, 3–13.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011.

- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805.
- Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526.
- DiMaio, F., Echols, N., Headd, J.J., Terwilliger, T.C., Adams, P.D., and Baker, D. (2013). Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* 10, 1102–1104.
- Dong, C., Zhang, H., Li, L., Tempel, W., Loppnau, P., and Min, J. (2018). Molecular basis of GID4-mediated recognition of degrons for the Pro/N-end rule pathway. *Nat. Chem. Biol.* 14, 466–473.
- Dou, H., Buetow, L., Sibbet, G.J., Cameron, K., and Huang, D.T. (2012). BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* 19, 876–883.
- Duda, D.M., Olszewski, J.L., Tron, A.E., Hammel, M., Lambert, L.J., Waddell, M.B., Mittag, T., DeCaprio, J.A., and Schulman, B.A. (2012). Structure of a glomulin-RBX1-CUL1 complex: inhibition of a RING E3 ligase through masking of its E2-binding surface. *Mol. Cell* 47, 371–382.
- Emberley, E.D., Mosadeghi, R., and Deshaies, R.J. (2012). Deconjugation of Nedd8 from Cul1 is directly regulated by Skp1-F-box and substrate, and the COP9 signalosome inhibits deneddylated SCF by a noncatalytic mechanism. *J. Biol. Chem.* 287, 29679–29689.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
- Enchev, R.I., Scott, D.C., da Fonseca, P.C., Schreiber, A., Monda, J.K., Schulman, B.A., Peter, M., and Morris, E.P. (2012). Structural basis for a reciprocal regulation between SCF and CSN. *Cell Rep.* 2, 616–627.
- Fernandez-Leiro, R., and Scheres, S.H.W. (2017). A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D Struct. Biol.* 73, 496–502.
- Francis, O., Han, F., and Adams, J.C. (2013). Molecular phylogeny of a RING E3 ubiquitin ligase, conserved in eukaryotic cells and dominated by homologous components, the muskelin/RanBPM/CTLH complex. *PLoS ONE* 8, e75217.
- Gancedo, J.M. (1998). Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* 62, 334–361.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11, 4241–4257.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* 27, 14–25.
- González, A., and Hall, M.N. (2017). Nutrient sensing and TOR signaling in yeast and mammals. *EMBO J.* 36, 397–408.
- Guimaraes, C.P., Witte, M.D., Theile, C.S., Bozkurt, G., Kundrat, L., Blom, A.E.M., and Ploegh, H.L. (2013). Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* 8, 1787–1799.
- Hämmerle, M., Bauer, J., Rose, M., Szallies, A., Thumm, M., Dusterhus, S., Mecke, D., Entian, K.D., and Wolf, D.H. (1998). Proteins of newly isolated mutants and the amino-terminal proline are essential for ubiquitin-proteasome-catalyzed catabolite degradation of fructose-1,6-bisphosphatase of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273, 25000–25005.
- Han, T., Yang, C.S., Chang, K.Y., Zhang, D., Imam, F.B., and Rana, T.M. (2016). Identification of novel genes and networks governing hematopoietic stem cell development. *EMBO Rep.* 17, 1814–1828.
- Hoffman, M., and Chiang, H.L. (1996). Isolation of degradation-deficient mutants defective in the targeting of fructose-1,6-bisphosphatase into the vacuole for degradation in *Saccharomyces cerevisiae*. *Genetics* 143, 1555–1566.
- Hu, R.G., Sheng, J., Qi, X., Xu, Z., Takahashi, T.T., and Varshavsky, A. (2005). The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. *Nature* 437, 981–986.
- Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* 545, 505–509.
- Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* 21, 947–962.
- Javan, G.T., Salhotra, A., Finley, S.J., and Soni, S. (2018). Erythroblast macrophage protein (Emp): Past, present, and future. *Eur. J. Haematol.* 100, 3–9.
- Kaiser, S.E., Riley, B.E., Shaler, T.A., Trevino, R.S., Becker, C.H., Schulman, H., and Kopito, R.R. (2011). Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat. Methods* 8, 691–696.
- Keilhauer, E.C., Hein, M.Y., and Mann, M. (2015). Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol. Cell. Proteomics* 14, 120–135.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
- Klaips, C.L., Hochstrasser, M.L., Langlois, C.R., and Serio, T.R. (2014). Spatial quality control bypasses cell-based limitations on proteostasis to promote prion curing. *eLife* 3, e04288.
- Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* 15 (10B), 963–972.
- Koren, I., Timms, R.T., Kula, T., Xu, Q., Li, M.Z., and Elledge, S.J. (2018). The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* 173, 1622–1635.e14.
- Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* 11, 319–324.
- Lampert, F., Stafa, D., Goga, A., Soste, M.V., Gilberto, S., Olieric, N., Picotti, P., Stoffel, M., and Peter, M. (2018). The multi-subunit GID/CTLH E3 ubiquitin ligase promotes cell proliferation and targets the transcription factor Hbp1 for degradation. *eLife* 7, e35528.
- Li, J., Li, Z., Ruan, J., Xu, C., Tong, Y., Pan, P.W., Tempel, W., Crombet, L., Min, J., and Zang, J. (2011). Structural basis for specific binding of human MPP8 chromodomain to histone H3 methylated at lysine 9. *PLoS ONE* 6, e25104.
- Lin, H.C., Yeh, C.W., Chen, Y.F., Lee, T.T., Hsieh, P.Y., Rusnac, D.V., Lin, S.Y., Elledge, S.J., Zheng, N., and Yen, H.S. (2018). C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol. Cell* 70, 602–613.e3.
- Liu, H., and Pfirrmann, T. (2019). The Gid-complex: an emerging player in the ubiquitin ligase league. *Biol. Chem.* 400, 1429–1441.
- Lyapina, S., Cope, G., Shevchenko, A., Serino, G., Tsuge, T., Zhou, C., Wolf, D.A., Wei, N., Shevchenko, A., and Deshaies, R.J. (2001). Promotion of NEDD-CUL1 conjugate cleavage by COP9 signalosome. *Science* 292, 1382–1385.
- Lydeard, J.R., Schulman, B.A., and Harper, J.W. (2013). Building and remodeling Cullin-RING E3 ubiquitin ligases. *EMBO Rep.* 14, 1050–1061.

- Mastrorarde, D. (2003). SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183.
- Matta-Camacho, E., Kozlov, G., Li, F.F., and Gehring, K. (2010). Structural basis of substrate recognition and specificity in the N-end rule pathway. *Nat. Struct. Mol. Biol.* **17**, 1182–1187.
- Melnykov, A., Chen, S.J., and Varshavsky, A. (2019). Gid10 as an alternative N-recognin of the Pro/N-degron pathway. *Proc. Natl. Acad. Sci. U S A* **116**, 15914–15923.
- Menssen, R., Schweiggert, J., Schreiner, J., Kusevic, D., Reuther, J., Braun, B., and Wolf, D.H. (2012). Exploring the topology of the Gid complex, the E3 ubiquitin ligase involved in catabolite-induced degradation of gluconeogenic enzymes. *J. Biol. Chem.* **287**, 25602–25614.
- Menssen, R., Bui, K., and Wolf, D.H. (2018). Regulation of the Gid ubiquitin ligase recognition subunit Gid4. *FEBS Lett.* **592**, 3286–3294.
- Mitchell, A., Romano, G.H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., and Pilpel, Y. (2009). Adaptive prediction of environmental changes by microorganisms. *Nature* **460**, 220–224.
- Nguyen, A.T., Prado, M.A., Schmidt, P.J., Sendamarai, A.K., Wilson-Grady, J.T., Min, M., Campagna, D.R., Tian, G., Shi, Y., Dederer, V., et al. (2017). UBE2O remodels the proteome during terminal erythroid differentiation. *Science* **357**, eaan0218.
- Oh, J.H., Chen, S.J., and Varshavsky, A. (2017). A reference-based protein degradation assay without global translation inhibitors. *J. Biol. Chem.* **292**, 21457–21465.
- Petroski, M.D., and Deshaies, R.J. (2005). Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* **123**, 1107–1120.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
- Pfirrmann, T., Villavicencio-Lorini, P., Subudhi, A.K., Menssen, R., Wolf, D.H., and Hollemann, T. (2015). RMND5 from *Xenopus laevis* is an E3 ubiquitin-ligase and functions in early embryonic forebrain development. *PLoS ONE* **10**, e0120342.
- Pierce, N.W., Lee, J.E., Liu, X., Sweredoski, M.J., Graham, R.L., Larimore, E.A., Rome, M., Zheng, N., Clurman, B.E., Hess, S., et al. (2013). Cnd1 promotes assembly of new SCF complexes through dynamic exchange of F box proteins. *Cell* **153**, 206–215.
- Plechanovová, A., Jaffray, E.G., Tatham, M.H., Naismith, J.H., and Hay, R.T. (2012). Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120.
- Pruneda, J.N., Littlefield, P.J., Soss, S.E., Nordquist, K.A., Chazin, W.J., Brzovic, P.S., and Klevit, R.E. (2012). Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942.
- Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. (2001). Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. *Nature* **410**, 955–959.
- Regelmann, J., Schüle, T., Josupeit, F.S., Horak, J., Rose, M., Entian, K.D., Thumm, M., and Wolf, D.H. (2003). Catabolite degradation of fructose-1,6-bisphosphatase in the yeast *Saccharomyces cerevisiae*: a genome-wide screen identifies eight novel GID genes and indicates the existence of two degradation pathways. *Mol. Biol. Cell* **14**, 1652–1663.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745.
- Rusnac, D.V., Lin, H.C., Canzani, D., Tien, K.X., Hinds, T.R., Tsue, A.F., Bush, M.F., Yen, H.S., and Zheng, N. (2018). Recognition of the diglycine C-end degron by CRL2(KLHDC2) ubiquitin ligase. *Mol. Cell* **72**, 813–822.e4.
- Santt, O., Pfirrmann, T., Braun, B., Juretschke, J., Kimmig, P., Scheel, H., Hofmann, K., Thumm, M., and Wolf, D.H. (2008). The yeast GID complex, a novel ubiquitin ligase (E3) involved in the regulation of carbohydrate metabolism. *Mol. Biol. Cell* **19**, 3323–3333.
- Saxton, R.A., and Sabatini, D.M. (2017). mTOR signaling in growth, metabolism, and disease. *Cell* **168**, 960–976.
- Scheres, S.H. (2012). A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418.
- Schork, S.M., Bee, G., Thumm, M., and Wolf, D.H. (1994a). Catabolite inactivation of fructose-1,6-bisphosphatase in yeast is mediated by the proteasome. *FEBS Lett.* **349**, 270–274.
- Schork, S.M., Bee, G., Thumm, M., and Wolf, D.H. (1994b). Site of catabolite inactivation. *Nature* **369**, 283–284.
- Schork, S.M., Thumm, M., and Wolf, D.H. (1995). Catabolite inactivation of fructose-1,6-bisphosphatase of *Saccharomyces cerevisiae*. Degradation occurs via the ubiquitin pathway. *J. Biol. Chem.* **270**, 26446–26450.
- Schüle, T., Rose, M., Entian, K.D., Thumm, M., and Wolf, D.H. (2000). Ubc8 functions in catabolite degradation of fructose-1,6-bisphosphatase in yeast. *EMBO J.* **19**, 2161–2167.
- Scott, D.C., Sviderskiy, V.O., Monda, J.K., Lydeard, J.R., Cho, S.E., Harper, J.W., and Schulman, B.A. (2014). Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. *Cell* **157**, 1671–1684.
- Shemorry, A., Hwang, C.S., and Varshavsky, A. (2013). Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* **50**, 540–551.
- Soni, S., Bala, S., Gwynn, B., Sahr, K.E., Peters, L.L., and Hanspal, M. (2006). Absence of erythroblast macrophage protein (Emp) leads to failure of erythroblast nuclear extrusion. *J. Biol. Chem.* **281**, 20181–20189.
- Storici, F., and Resnick, M.A. (2006). The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329–345.
- Szoradi, T., Schaeff, K., Garcia-Rivera, E.M., Itzhak, D.N., Schmidt, R.M., Bircham, P.W., Leiss, K., Diaz-Miyar, J., Chen, V.K., Muzzey, D., et al. (2018). SHRED is a regulatory cascade that reprograms Ubr1 substrate specificity for enhanced protein quality control during stress. *Mol. Cell* **70**, 1025–1037.e5.
- Tagkopoulos, I., Liu, Y.C., and Tavazoie, S. (2008). Predictive behavior within microbial genetic networks. *Science* **320**, 1313–1317.
- Timms, R.T., Zhang, Z., Rhee, D.Y., Harper, J.W., Koren, I., and Elledge, S.J. (2019). A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* **365**, eaaw4912.
- Tyanova, S., Temu, T., Carlson, A., Sinitcyn, P., Mann, M., and Cox, J. (2015). Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453–1456.
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740.
- Varshavsky, A. (2011). The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20**, 1298–1345.
- Varshavsky, A. (2012). The ubiquitin system, an immense realm. *Annu. Rev. Biochem.* **81**, 167–176.
- Varshavsky, A. (2019). N-degron and C-degron pathways of protein degradation. *Proc. Natl. Acad. Sci. U S A* **116**, 358–366.
- Wang, K.H., Roman-Hernandez, G., Grant, R.A., Sauer, R.T., and Baker, T.F. (2008). The molecular basis of N-end rule recognition. *Mol. Cell* **32**, 406–414.
- Wanichthanarak, K., Nookaew, I., and Petranovic, D. (2014). yStreX: yeast stress expression database. *Database (Oxford)* **2014**, bau068.
- Watson, E.R., Brown, N.G., Peters, J.M., Stark, H., and Schulman, B.A. (2019). Posing the APC/C E3 Ubiquitin Ligase to Orchestrate Cell Division. *Trends Cell Biol.* **29**, 117–134.
- Weissmann, F., Petzold, G., VanderLinden, R., Huis In 't Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A., and Peters, J.M. (2016). biGBac enables rapid gene assembly for the expression of large

- multisubunit protein complexes. *Proc. Natl. Acad. Sci. U S A* *113*, E2564–E2569.
- Wenzel, D.M., Lissounov, A., Brzovic, P.S., and Klevit, R.E. (2011). UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* *474*, 105–108.
- Zaman, S., Lippman, S.I., Zhao, X., and Broach, J.R. (2008). How *Saccharomyces* responds to nutrients. *Annu. Rev. Genet.* *42*, 27–81.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* *193*, 1–12.
- Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* *102*, 533–539.
- Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* *14*, 331–332.
- Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* *7*, e42166.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal ANTI-FLAG M2 antibody	Sigma	Cat#F1804; RRID: AB_262044
Anti-HA antibody produced in rabbit	Sigma	Cat#H6908; RRID: AB_260070
Goat anti-rabbit IgG Dylight488 conjugated	Invitrogen	Cat#35552; RRID: AB_844398
Goat anti-mouse IgG Dylight633 conjugated	Invitrogen	Cat#35512; RRID: AB_1307538
Anti-Rabbit peroxidase antibody produced in goat	Sigma	Cat#A9169; RRID: AB_258434
Anti-Mouse IgG Peroxidase antibody produced in goat	Sigma	Cat#A4416; RRID: AB_258167
Anti-His antibody produced in mouse	Cell Signaling Technology	Cat#9991; RRID: AB_2797714
Bacterial and Virus strains		
<i>E. coli</i> BL21 RIL (DE3)	Stratagene	Cat#230245
Chemicals, Peptides, and Recombinant proteins		
complete EDTA free protease inhibitor cocktail	Roche	Cat#05056489001
Aprotinin from bovine lung	Sigma	Cat#A1153-10MG
Leupeptin	Sigma	Cat#L2884-250MG
Benzamidine	Sigma	Cat#B6506-25G
Peptide for C-terminal sorting (H-GGGGGFYVK-FAM-NH ₂)	MPIB	N/A
Fluorescein-5-maleimide	Anaspec	Cat#AS-81405
NAP-5 desalting column	GE Healthcare	Cat#17-0853-01
PD-10 desalting column	GE Healthcare	Cat#52130800
3x FLAG Peptide	MPIB	Cat#L1033
Anti-DYKDDDDK G1 Affinity Resin	Genscripts	Cat#L00432
MG132 Proteasome inhibitor	InvivoGen	Cat#tlrl-mg132; CAS: 133407-82-06
Deposited Data		
GID ^{SR4}	This study	EMDB: 10327
GID ^{SR4} minus Gid2/ΔGid9 ^{RING}	This study	EMDB: 10333; PDB: 6SWY
GID ^{Scaffold} plus SR ^{Gid4}	This study	EMDB: 10330
GID ^{Scaffold}	This study	EMDB: 10328
GID ^{Scaffold} plus SR ^{Gid10}	This study	EMDB: 10329
GID ^{Ant}	This study	EMDB: 10326
GID ^{SR4} ΔRINGs	This study	EMDB: 10332
Endogenous GID ^{Ant}	This study	EMDB: 10331
Unprocessed image data	This study	https://doi.org/10.17632/nd5zc59vfg.1
Proteomics data	This study	PRIDE database: PXD015396
Experimental Models: Cell lines		
Sf9 Insect cells	Thermo Fisher	Cat#11496015
High Five Insect cells	Thermo Fisher	Cat#B85502
Experimental Models: Organisms/Strains		
<i>Saccharomyces cerevisiae</i> : Strain S288C: BY4741; MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Euroscarf	Cat#Y00000
CRLY39; BY4741 GID1-3xFLAG::KANMX	This study	N/A
CRLY42; BY4741 GID5-3xFLAG::KANMX	This study	N/A
CRLY44; BY4741 GID7-3xFLAG::KANMX	This study	N/A
CRLY45; BY4741 GID8-3xFLAG::KANMX	This study	N/A
CRLY50; BY4741 gid2::3xFLAG-GID2	This study	N/A

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
CRLY52; BY4741 gid9::3xFLAG-GID9	This study	N/A
CRLY68; BY4741 gid4::3xFLAG-GID4	This study	N/A
PHH2; BY4741 GID8-3xFLAG::KANMX, gid7-3xHA::HPHNT1	This study	N/A
CRLY131; BY4741 gid2::3xFLAG-GID2K365A	This study	N/A
CRLY132; BY4741 gid2::3xFLAG-GID2Y403A	This study	N/A
CRLY133; BY4741 gid2::3xFLAG-GID2F418A	This study	N/A
CRLY135; BY4741 gid2::3xFLAG-GID2Δ354-422	This study	N/A
VBY73; BY4741 gid2::3xFLAG-GID2L364A	This study	N/A
VBY74; BY4741 gid2::3xFLAG-GID2V363A	This study	N/A
VBY75; BY4741 gid2::3xFLAG-GID2L364A, V363A	This study	N/A
VBY76; BY4741 gid9::3xFLAG-GID9V436A	This study	N/A
CRLY110; BY4741 gid9::3xFLAG-GID9M502A	This study	N/A
CRLY113; BY4741 gid9::3xFLAG-GID9Y514A	This study	N/A
CRLY114; BY4741 gid9::3xFLAG-GID9Δ432-517	This study	N/A
VBY60; BY4741 gid5::GID5R652A, N653A, Y718A-3xFLAG::HPHNT1	This study	N/A
VBY61; BY4741 gid5::GID5V491A, R495A, F550A-3xFLAG::HPHNT1	This study	N/A
VBY62; BY4741 gid5::GID5W606A, Y613A, Q649A-3xFLAG::HPHNT1	This study	N/A
VBY63; BY4741 gid5::GID5W606A, H610A, Y613A-3xFLAG::HPHNT1	This study	N/A
VBY64; BY4741 gid5::GID5W606A, H610A-3xFLAG::HPHNT1	This study	N/A
VBY65; BY4741 gid5::GID5F550D-3xFLAG::HPHNT1	This study	N/A
VBY59; BY4741 gid1::GID1Δ411-420-3xFLAG::HPHNT1	This study	N/A
VBY66; BY4741 gid1::GID1Δ413-418-3xFLAG::HPHNT1	This study	N/A
VBY49; BY4741 gid4::GID4F359A, F361A	This study	N/A
VBY50; BY4741 gid4::GID4F359D, F361D	This study	N/A
VBY81; BY4741 gid4::GID4T148D	This study	N/A
VBY82; BY4741 gid4::GID4F166W	This study	N/A
VBY83; BY4741 gid4::GID4F296W	This study	N/A
VBY84; BY4741 gid4::GID4F296A	This study	N/A
CRLY74; BY4741 gid10::3xFLAG-GID10	This study	N/A
Recombinant DNA		
pCSJ95	Chen et al., 2017	N/A
pFLN2 Gid1	This study	N/A
pFLN2 Gid1 (Δ413-418)	This study	N/A
pFLN2 2xStrep-3c-Gid1	This study	N/A
pFLN2 Gid2	This study	N/A
pFLN2 Gid2 (Δ354-421)	This study	N/A
pFLN2 Gid2 (V363A)	This study	N/A
pFLN2 Gid2 (L364A)	This study	N/A
pFLN2 Gid2 (V363A, L364A)	This study	N/A
pFLN2 Gid4	This study	N/A
pFLN2 Gid10	This study	N/A
pFLN2 Gid5	This study	N/A
pFLN2 Gid5 (V491A, R495A, F550A)	This study	N/A
pFLN2 Gid5 (F550D)	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pFLN2 Gid5 (W606A, H610A)	This study	N/A
pFLN2 Gid5 (W606A, H610A, Y613A)	This study	N/A
pFLN2 Gid5 (W606A, Y613A, Q649A)	This study	N/A
pFLN2 Gid5 (R652A, N653A, Y718A)	This study	N/A
pFLN2 Gid8	This study	N/A
pFLN2 Gid8-TEV-2xStrep	This study	N/A
pFLN2 Gid9	This study	N/A
pFLN2 Gid9 (Δ 432-516)	This study	N/A
pFLN2 Gid9 (V436A)	This study	N/A
pBIG1a Gid1:Gid5	This study	N/A
pBIG1a Gid1:Gid5:Gid2	This study	N/A
pBIG1a 2xStrep-3c-Gid1:Gid5	This study	N/A
pBIG1a 2xStrep-3c-Gid1:Gid5:Gid2	This study	N/A
pBIG1b Gid8:Gid9	This study	N/A
pBIG1b Gid8-TEV-2xStrep:Gid9	This study	N/A
pBIG1c Gid2:Gid4	This study	N/A
pBIG2a-d Gid1:Gid8-TEV-2xS:Gid2:Gid9:Gid5	This study	N/A
pBIG2a-d Gid1:Gid8-TEV-2xS:Gid2:Gid9:Gid5:Gid4	This study	N/A
pBIG1e Gid1:Gid8-TEV-2xS:Gid9:Gid5	This study	N/A
pBIG1e Gid1:Gid8-TEV-2xS:Gid2:Gid5	This study	N/A
pBIG2a-d Gid8-TEV-2xS:Gid2:Gid9:Gid5	This study	N/A
pBIG2a-d Gid1:Gid8-TEV-2xS:Gid2:Gid9	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-79)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116; T148D)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116; F166W)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116; F296A)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116; F296W)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116; F359A, F361A)	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116; F359D, F361D)	This study	N/A
pGEX GST-TEV-Gid10 (Δ 1-56)	This study	N/A
pGEX GST-TEV-Gid10 (Δ 1-56; Δ 289-292)	This study	N/A
pGEX GST-3c-Ub	This study	N/A
pGEX GST-3c-Ub K0 (all K > R)	This study	N/A
pGEX GST-3c-Ub K6 (all K > R; R6K)	This study	N/A
pGEX GST-3c-Ub K11 (all K > R; R11K)	This study	N/A
pGEX GST-3c-Ub K27 (all K > R; R27K)	This study	N/A
pGEX GST-3c-Ub K29 (all K > R; R29K)	This study	N/A
pGEX GST-3c-Ub K33 (all K > R; R33K)	This study	N/A
pGEX GST-3c-Ub K48 (all K > R; R48K)	This study	N/A
pGEX GST-3c-Ub K63 (all K > R; R63K)	This study	N/A
pGEX GST-TEV-Cys-Ub	This study	N/A
pET3b Ub	This study	N/A
pRSFduet Ubc8-6xHis	This study	N/A
pRSFduet Mdh2-6xHis	This study	N/A
pRSFduet Mdh2-6xHis (P2S)	This study	N/A
pRSFduet Mdh2-6xHis (K254R, K256R, K259R, K330R, K360R, K361R)	This study	N/A

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
pRSFduet Mdh2-GGGGS-sortag-6xHis	This study	N/A
pFastbac GST-TEV-Uba1	This study	N/A
Software and Algorithms		
SerialEM	Mastronarde, 2003	http://bio3d.colorado.edu/SerialEM/
MOTIONCOR2	Zheng et al., 2017	
Gctf	Zhang, 2016	https://www.mrc-lmb.cam.ac.uk/kzhang/Gctf/
Gautomatch	Kai Zhang	https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/
Relion3.0	Fernandez-Leiro and Scheres, 2017; Scheres, 2012; Zivanov et al., 2018	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
Phyre ²	Kelley et al., 2015	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX	Goddard et al., 2018	https://www.rbvi.ucsf.edu/chimeraX/
Pymol	Schrodinger	https://pymol.org/2/
CCP-EM	Burnley et al., 2017	http://www.ccpem.ac.uk/download.php
Buccaneer	Cowtan, 2006	https://www.ccp4.ac.uk/newsletters/newsletter44/articles/buccaneer.html
Coot	Emsley and Cowtan, 2004; Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix	Adams et al., 2010; Afonine et al., 2018; DiMaio et al., 2013	https://www.phenix-online.org/
Molprobit	Chen et al., 2010	http://molprobit.biochem.duke.edu/
Image Studio	LI-COR Biosciences	https://www.licor.com/bio/image-studio/?gclid=EAlaQobChMInO7thc224wIVSxbTCh3irwQdEAAAYASAAEgLj8fD_BwE
Other		
QUANTIFOIL® R1.2/1.3, 100 Holey Carbon Films, Grids: Cu 200 mesh	Quantifoil Micro Tools GmbH	https://www.quantifoil.com

LEAD CONTACT AND MATERIALS AVAILABILITY

Information and requests for resources and reagents should be directed to the Lead Contact, Brenda Schulman (schulman@biochem.mpg.de).

All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

METHOD DETAILS

Yeast strains and growth conditions

All yeast strains used in this study are listed in Key Resources Table and are derivatives of BY4741. For strain construction, standard genetic techniques were employed (Janke et al., 2004; Knop et al., 1999; Storici and Resnick, 2006). All yeast strains were verified by DNA sequencing, western blotting for protein expression, and were shown to be competent for Fbp1 degradation (see below).

Unless otherwise specified, for assays described here, yeast strains were grown to OD₆₀₀ of 1.0 in synthetic complete (SC) medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose, plus a mixture of amino acids). If strains were carrying a plasmid, the appropriate amino acids were omitted. Cells were then centrifuged at 1900xg for 3 minutes, washed once with SE medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% ethanol, plus a mixture of amino acids), and then resuspended in fresh, pre-warmed SE media to an OD₆₀₀ of 1.0. Cells were grown at 30°C for 19 hours, at which point they were harvested by centrifugation at 1,900xg for 3 minutes, and resuspended to an OD₆₀₀ of 1.0 in fresh SC medium. At the indicated time points, cells were harvested by centrifugation at 11,200xg for 2 minutes, and flash frozen in liquid nitrogen for later analysis.

For growth under heat shock conditions, cells were grown in YPD at 30°C to mid-log phase and then shifted to 42°C for 30 minutes before returning the cultures to 30°C growth. For growth under high salinity conditions, cells were grown to mid-log phase in YPD then pelleted and resuspended in fresh YPD + 0.5 M NaCl and allow to grow at 30°C. At the indicated time points, an aliquot of cell was harvested by centrifugation.

Fbp1 degradation assays

Fbp1 degradation assays were carried out using the promoter reference technique as previously described (Oh et al., 2017). Briefly cells were first transformed with a plasmid co-expressing Fbp1 and a control protein (DHFR) from identical promoters containing an element that once transcribed binds tetracycline to inhibit translation. After growth for 19 hours in medium containing 2% ethanol cells were resuspended to an OD₆₀₀ of 1.0 in SD medium lacking the appropriate amino acids and containing 2% glucose and 0.5 mM tetracycline. At the indicated time points, 1 OD₆₀₀ equivalent of cells were harvested. Cells were lysed by resuspension in 0.2 M NaOH followed by incubation on ice for 20 minutes, and then pelleted by centrifugation at 11,200xg for 2 minutes. The supernatant was removed and the cell pellet was resuspended in HU buffer containing 1X complete protease inhibitor tablets (Roche), heated a 70°C for 10 minutes, and then the resulting lysate was precleared by centrifugation for 5 minutes at 11,200xg. Samples were loaded on a 12% SDS-PAGE gel, followed by analysis by western blotting. Blots were imaged on a Typhoon scanner (GE Healthcare) and bands were quantified using ImageStudio software (Li-Cor). At each time point, the amount of Fbp1 was normalized to the DHFR control protein.

Affinity Enrichment Mass Spectrometry (AE-MS)

Experiments were carried out in triplicates and as previously described (Keilhauer et al., 2015). 50 OD₆₀₀ of yeast were resuspended in buffer containing 150 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM MgCl₂, Complete protease inhibitor tablets (Roche), PhosStop tablets (Roche), 5% Glycerol, 1% NP-40, and 1% Benzonase. Cells were lysed by glass bead lysis in a FastPrep-24 instrument (MP Bio medicals) during three rounds of 20 s each at a speed of 4.0 M/s. Lysates were then pre-cleared by centrifugation at 4,000xg for 10 minutes. FLAG-tagged protein was then pulled-down using anti-DYKDDDDK magnetic beads (Miltenyi Biotec). After applying the beads to the magnetic columns, beads were washed three times with wash buffer containing 150 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.05% NP-40, and 5% Glycerol, followed by two washes in buffer containing 150 mM NaCl, 50 mM Tris-HCl pH 7.5, and 5% Glycerol. Bound proteins were then partially digested by addition of 25 µL of elution buffer I containing 5 ng/µL Trypsin, 2 M Urea, 50 mM Tris-HCl pH 7.5, and 1 mM DTT. After incubation for 30 minutes at room temperature, the proteins were eluted from the column by elution with 2x50 µL of buffer containing 2 M Urea, 50 mM Tris-HCl pH 7.5, and 5 mM CAA. Samples were then incubated overnight at room temperature to ensure complete tryptic digestion. Next day, protease activity was quenched by acidification with Trifluoroacetic acid to a final concentration of 1%. Samples were processed as described in 'LC-MS/MS sample preparation'

Protein digestion of *in vitro* ubiquitylation assays

Protein concentrations of the samples from *in vitro* ubiquitylation assays were measured by Bradford assay (BioRad). 4 µg of sample were 4-fold diluted in digestion buffer (1 M Urea in 50 mM Ammonium Bicarbonate, pH 8.0) followed by addition of TCEP and CAA to a final concentration of 10 mM and 40 mM, respectively, for reduction and alkylation for 5 min at 45°C. The samples were either digested using Trypsin (1:20 w/w, Sigma-Aldrich) alone, Trypsin (1:40 w/w)/GluC (1:40 w/w, BioLab) or Trypsin (1:40 w/w)/AspI (1:40 w/w, Promega) at 37°C overnight. In all cases, protease activity was quenched by acidification with Trifluoroacetic acid to a final concentration of 1%.

LC-MS/MS sample preparation

Acidified samples were loaded onto SDB-RPS StageTips, pre-equilibrated with 30% Methanol /1% Trifluoroacetic acid and washed with 0.2% Trifluoroacetic acid. StageTips were prepared by inserting two layers of SDB-RPS matrix (Empore) into a 200 µL pipette tip using an in-house prepared syringe device as described previously (Kulak et al., 2014). The StageTips were centrifuged at 1000xg. Loaded samples were sequentially washed with 0.2% Trifluoroacetic acid and 2% Acetonitrile/0.2% Trifluoroacetic acid, followed by elution with 1.25% NH₄OH/80% Acetonitrile. Eluates were dried using a SpeedVac centrifuge (Eppendorf, Concentrator plus). Peptides were resuspended in buffer A* (2% Acetonitrile /0.1% Trifluoroacetic acid) and briefly sonicated (Branson Ultrasonics) before LC MS-MS analysis.

LC-MS/MS Measurements

For mapping ubiquitylation sites, peptide concentration was estimated by UV spectrometry and approximately 200 ng was loaded on a 50 cm reversed phase column (75 µm inner diameter, packed in house with ReproSil-Pur C18-AQ 1.9 µm resin [Dr. Maisch GmbH]). Column temperature was maintained at 60°C using a homemade column oven. Peptides were separated with a binary buffer system of buffer A (0.1% Formic acid (FA)) and buffer B (80% Acetonitrile plus 0.1% FA), at a flow rate of 300 nl/min. We used an EASY-nLC 1200 system (Thermo Fisher Scientific), which was directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source. Peptides were eluted with a gradient starting at 3% buffer B and stepwise increase to 8% in 8 min, 36% in 32 min, 45% in 4 min and 95% in 4 min. The mass spectrometer was operated in Top12 data-dependent mode (DDA) with a full scan range of 250-1350 m/z at 60,000 resolution with an automatic gain control (AGC) target of 3e6 and a maximum

fill time of 20ms. Precursor ions were isolated with a width of 1.4 m/z and fragmented by higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 28%. Fragment scans were performed at a resolution of 30,000, an AGC of 1e5 and a maximum injection time of 110ms. Dynamic exclusion was enabled and set to 15 s.

For AE-MS samples, peptides were loaded on a 50 cm reversed phase column (75 μ m inner diameter, packed in house with ReproSil-Pur C18-AQ 1.9 μ m resin [Dr. Maisch GmbH]). Column temperature was maintained at 60°C using a homemade column oven. Peptides were separated with a binary buffer system of buffer A (0.1% Formic acid (FA)) and buffer B (80% Acetonitrile plus 0.1% FA), at a flow rate of 300 nL/min. We used an EASY-nLC 1200 system (Thermo Fisher Scientific), which was directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source. Peptides were eluted with a gradient starting at 5% buffer B and stepwise increased to 30% in 40 min, 60% in 4 min and 95% in 4 min. The mass spectrometer was operated in Top12 data-dependent mode (DDA) with a full scan range of 300-1650 m/z at 60,000 resolution with an automatic gain control (AGC) target of 3e6 and a maximum fill time of 20ms. Precursor ions were isolated with a width of 1.4 m/z and fragmented by higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 27%. Fragment scans were performed at a resolution of 15,000, an AGC of 1e5 and a maximum injection time of 60ms. Dynamic exclusion was enabled and set to 30 s.

LC-MS/MS raw data processing

For mapping ubiquitylation sites, raw MS data were searched against the UniProt Yeast FASTA using MaxQuant (version 1.6.2.10) (Cox and Mann, 2008; Cox et al., 2011) with a 1% FDR at peptide and protein level. Cysteine carbamidomethylation was set as fixed, protein N-Terminal acetylation, methionine oxidation and lysine diGly as variable modifications. The minimum peptide length was set to 7 amino acids, enzyme specificity was set to trypsin and two missed cleavages were allowed, permitting a maximum of 5 modifications per peptide. MS/MS spectra identifying ubiquitinated peptides of interest were obtained and exported using MaxQuant Viewer (Tyanova et al., 2015).

For AE-MS runs, raw MS data were searched against the UniProt Yeast FASTA using MaxQuant (version 1.6.0.15) (Cox and Mann, 2008; Cox et al., 2011) with a 1% FDR at peptide and protein level. Cysteine carbamidomethylation was set as fixed, protein N-Terminal acetylation, methionine oxidation. The minimum peptide length was set to 7 amino acids, enzyme specificity was set to trypsin and two missed cleavages were allowed, permitting a maximum of 5 modifications per peptide. Quantification was performed by label free quantification (MaxLFQ), with a minimum ratio count of 2. 'Match between runs' was enabled, with a matching time window of 0.7 min (Cox et al., 2014). Bioinformatic analyses were performed with Perseus (www.perseus-framework.org) (Tyanova et al., 2016). Missing data points were replaced by data imputation (width: 0.3 and shift: 1.8) after filtering for valid values (100% data completeness in at least one experimental group). Significance was assessed using t test (5% FDR), for which replicates were grouped.

Density Fractionation by sucrose gradients

At the indicated time points, 100 OD₆₀₀ equivalents were harvested by centrifugation at 1900xg for 3 minutes. Cells were resuspended in lysis buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM MgCl₂, 1% NP-40, and protease inhibitors (Roche), and lysed by glass bead lysis using a FastPrep-24 (MP Biomedicals). The resulting lysate was pre-cleared by centrifugation at 4,000xg for 10 minutes, and the supernatant was normalized by Bradford for total protein content, loaded onto a 5%–40% sucrose gradient, and centrifuged at 34,300 rpm for 16 hours at 4°C. Each gradient was harvested into fourteen equal fractions, and run on SDS-PAGE, followed by analysis by western blotting with the appropriate antibody. Approximate molecular weights for fractions were determined using the protein standards provided with Gel Filtration Calibration Kit HMW (GE Healthcare). Briefly, 2 mg of each protein standard were resuspended in lysis buffer and run on a 5%–40% gradient as described above.

Plasmids preparation and Mutagenesis

Genes encoding GID subunits and Mdh2 substrate were originally amplified using *S. cerevisiae* BY4742 genomic DNA as a template. Upon initiating functional studies, we noticed that all sequences except *Gid5* match those in the *Saccharomyces* Genome Database (SGD), where the *Gid5* sequence corresponds to accession number NP_012247. The sequence of *Gid5* used in the structure corresponds to accession number AJR43361 from the strain YJM1133. The difference is a single *Gid5* Y758N residue substitution. We tested the functionality of the sequence used in the structure in initial assays probing GID^{SR4} activity toward Mdh2, and we confirmed by cryo EM that the overall structures of GID^{SR4} are similar with the two versions of *Gid5*. We converted to the SGD sequence for functional studies.

The genes encoding GID subunits were combined into one baculoviral expression vector with the biGBac method (Weissmann et al., 2016). All the plasmids used in this study are listed in the Key Resources Table.

The constructs for recombinant protein expression were generated by Gibson assembly method (Gibson et al., 2009) with a home-made Gibson reaction mix. To generate all the mutant versions of the constructs, QuickChange (Stratagene) protocol was applied. All coding sequences used for protein expression were entirely verified by sequencing.

Protein expression and purification for cryo-EM

GID complexes and all the subcomplexes used for the single particle cryo EM analysis (GID^{SR4}, GID^{Ant}, GID^{Scaffold} plus substrate receptor *Gid4* (SR^{Gid4}), GID^{Scaffold} plus substrate receptor *Gid10* (SR^{Gid10}), GID^{SR4} minus *Gid2*/ Δ *Gid9*^{RING}, GID^{SR4} Δ RINGS), were

expressed in insect cell. For protein expression, Hi5 insect cells were transfected with recombinant baculovirus variants carrying the respective protein coding sequences and grown for 60 hours to 72 hours in EX-CELL 420 Serum-Free Medium at 27°C.

Cell pellets were resuspended in a lysis buffer containing 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM DTT, 10 µg/ml leupeptin 20 µg/ml aprotinin, 2 mM benzamidine, EDTA-free protease inhibitor tablet (1 tablet per 50 mL of the buffer) and 1 mM PMSF. The tagged complexes were purified from cell lysates by Strep-Tactin affinity chromatography by pulling on the Twin-Strep tag fused to the Gid1 N terminus. Elutions were further purified by anion exchange chromatography and size exclusion chromatography in 25 mM MES pH 6.5, 500 mM NaCl and 1 mM DTT.

For endogenous GID^{Ant} purification, yeast strain CRL45 harboring 3X Flag tag at the C terminus of Gid8 was grown in the YPI medium at 30°C and 130 rpm to an OD₆₀₀ of 1. Yeast cells were spun down and rinsed with YP medium to remove the remaining glucose. The pellet was resuspended with YPE medium and grown at 30°C and 130 rpm for 16-19 hours. Next, cell pellets were passed through a 50 mL syringe, to get thin noodle-like pellets and flash-frozen in liquid nitrogen. Frozen yeast noodles were cryo milled using Retsch ZM200 Ultra Centrifugal Mill. Powder was dissolved in the lysis buffer described above. GID complex was purified from the cell lysate by Flag affinity chromatography (Anti-DYKDDDDK G1 resin, GenScript). The resin bound GID complex was washed with 25 mM MES pH 6.5, 500 mM NaCl, 1 mM DTT, and protein was eluted with 150 µg/ml Flag peptide. The elutions were directly used to make the cryo EM grids.

Cryo EM sample preparation and imaging

To prepare cryo EM grids, Vitrobot Mark IV (Thermo Fisher Scientific) was used. 3.5 - 4 µl of freshly purified protein at 0.25 mg/ml was applied to glow discharged Quantifoil holey carbon grids (R1.2/1.3 200 mesh) and incubated for 30 s at 4°C and 100% humidity. Grids were immediately blotted with Whatman no.1 filter paper (blot time 10 s, blot force 10) and vitrified by plunging into liquid ethane.

For GID^{SR4}, GID^{Ant}, GID^{Scaffold} plus SR^{Gid4}, GID^{Scaffold} plus SR^{Gid10}, GID^{SR4} minus Gid2/ΔGid9^{RING}, GID^{SR4} ΔRINGS and the endogenous GID^{Ant}, cryo EM data were collected on a Talos Arctica transmission electron microscope operated at 200 kV and equipped with a Falcon III direct detector. Automated data collection was carried out using EPU software at a nominal magnification of 92,000x which corresponds to 1.612 Å/pixel at the specimen level, with a total exposure of 63 e⁻/Å² and the target defocus range between 1.5-3.5 µm.

For GID^{SR4} and GID^{Ant} data were collected on a FEI Titan Krios microscope operated at 300 kV, equipped with a post-column GID and a K2 Summit direct detector operating in a counting mode. SerialEM software was used to automate data collection (Mastrorade, 2003). Images were recorded at a nominal magnification of 130,000x (1.06 Å/pixel) with target defocus range between 1.1 and 3.2 µm and approximate total exposure of 54 e⁻/Å².

For GID^{Scaffold} plus SR^{Gid4}, GID^{Scaffold} plus SR^{Gid10}, GID^{SR4} minus Gid2/ΔGid9^{RING}, images were acquired as described above except using a K3 direct electron detector instead of K2 and at a nominal magnification of 81,000x corresponding to 1.09 Å/pixel at the specimen level. A SerialEM multi-record mode was used to collect data.

Data processing

Movie frames were motion-corrected and dose-weighted using the Motioncorr2 (Zheng et al., 2017) program. Contrast transfer function parameters were estimated from dose-weighted, aligned micrographs using Gctf (Zhang, 2016). Particles were automatically picked by Gautomatch (<https://www.mrc-lmb.cam.ac.uk/kzhang/>). Further processing was carried out using Relion (Fernandez Leiro and Scheres, 2017; Scheres, 2012; Zivanov et al., 2018). Poor quality images were discarded by manual inspection and only particles in the high-quality images were extracted. Iterative rounds of 2D classifications were done to clean up the data. 3f classifications were done using the initial model generated and clean set of particles from 2D classification. For large datasets the particles were split into smaller groups for which 3D classifications were carried out separately, and another round of classification was done if necessary. The resulting 3D classes were manually inspected, and those with complete features were selected for further processing. Particles selected from 3D classification were finally re-extracted, re-centered and subjected to auto-refinement (with and without a mask).

In addition to generating reconstructions for entire complexes, maps with improved quality over specific regions were obtained as follows. A map encompassing the majority of the catalytic module was obtained by multibody refinement of data from GID^{Ant}, treating the scaffold module (Gid1-Gid8-Gid5, aka GID^{Scaffold}) and the catalytic module (the Gid2-Gid9 subcomplex) as two separate entities. The resultant map over the catalytic module reached 5.1 Å resolution, and enabled visualizing the 4-stranded coiled coil subdomain. Meanwhile, the highest quality map for the CTLH-CRAN domain of Gid9 was obtained using the GID^{SR4} minus Gid2/ΔGid9^{RING} data set, by focused auto-refinement using a mask over this domain.

Finally, automatic B-factor weighting as well as high resolution noise substitution were done using post-processing in Relion. Local resolution estimates were done as implemented in Relion (Fernandez-Leiro and Scheres, 2017; Scheres, 2012; Zivanov et al., 2018). All the reported resolutions are based on the gold-standard Fourier Shell Correlation (FSC) at 0.143 criterion. Processing details for each dataset are provided in Figures S2, S3, S4A, and S4B. Maps generated in this study are summarized in Table S1.

Model building and refinement

The scheme for model building is shown in Figures S5A and S5B, and described here for each module. The scaffold module comprising Gid1, Gid8 and Gid5, was built using the 3.4 Å resolution reconstruction of GID^{Scaffold} plus SR^{Gid4}. Most of the Gid5

Gid8 as well as SPRY and LisH domains of Gid1 could be built automatically using Buccaneer (Cowtan, 2006), as implemented in CCP-EM software suite (Burnley et al., 2017), with some portions built manually using Coot (Emsley and Cowtan, 2004; Emsley et al., 2010).

The same map was used to build the substrate receptor module, Gid4, manually in Coot (Emsley and Cowtan, 2004; Emsley et al., 2010). The building of Gid4 was guided by a crystal structure of human Gid4 (PDB ID: 6CCR) and sequence alignment of ScGid4 and HsGid4, secondary structure prediction generated by the Phyre² server (Kelley et al., 2015), and the positions of side-chain features (e.g., aromatic residues) as markers.

Segments of Gid9 from the catalytic module were guided by differences in EM reconstructions lacking portions of Gid9. The CTLH-CRAN portion of Gid9 was best visualized and built manually with the 3.5 Å resolution map of Gid9^{CTLH-CRA} generated by focused refinement using data obtained from GID^{SR4} minus Gid2/ΔGid9^{RING}. A loop from Gid9 (Gid9^{loop}, residues 291-323) was built using the final 3.2 Å resolution map of GID^{SR4} minus Gid2/ΔGid9^{RING}.

Atomic model refinement was performed using 'phenix.real_space_refine' available in PHENIX software suite (Adams et al., 2010; Afonine et al., 2018; DiMaio et al., 2013) and the model was validated using Molprobit (Chen et al., 2010). The entire model was checked manually, and regions that lack the sequence registers due to weak/unclear density were modeled as polyalanine. Representative EM density is shown in Figure S5C, and the residues in the final model are summarized in Table S2. Data collection, 3D reconstruction, model refinement and validation details are given in Tables 1 and 2. Figures of maps and models were prepared with Chimera (Pettersen et al., 2004), ChimeraX (Goddard et al., 2018) and PyMol-v 1.8.2.

Protein expression and purification for biochemical assays

Insect cell expression as well as cell pellet resuspension for the WT and all the mutant versions of GID^{Ant} and GID^{SR4} used for biochemical assays followed the procedure described in the section 'Protein expression and purification for cryo EM'. Proteins were purified from insect cell lysates using Strep-Tactin affinity chromatography by pulling on a Twin-Strep tag fused to Gid8 C terminus. The eluted proteins were further purified by size exclusion chromatography in 25 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT (Buffer B).

To ensure that all assays contained equal concentrations of WT and mutant versions of Gid4 and Gid10 irrespective of their ability to bind GID^{Ant}, these proteins that were added exogenously to the *in vitro* assays were expressed as GST-TEV fusions in *E. coli* BL21 (DE3) RIL cells in a Terrific Broth (TB) medium overnight at 20°C. Cell pellets were resuspended in the lysis buffer containing 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM DTT and 1 mM PMSF (Buffer A). Proteins were purified from bacterial lysates with glutathione affinity chromatography and digested overnight at 4°C with tobacco etch virus (TEV) protease to liberate the GST tag. For further purification, they were subjected to size exclusion chromatography in Buffer B. Remaining free GST as well as uncleaved GST-fusion protein was removed by pass-back over a glutathione affinity resin.

Ubc8, Mdh2 and Mdh2 P2S mutant were expressed in *E. coli* BL21 (DE3) RIL cells in a Terrific Broth (TB) medium overnight at 20°C. Cell pellets were resuspended in Buffer A and proteins were purified from the bacterial lysates by Nickel-Affinity chromatography with Ni-NTA Sepharose resin by pulling on the 6xHis tag fused to proteins C terminus. The elutions were further purified by anion exchange chromatography and size exclusion chromatography in Buffer B.

Untagged WT ubiquitin used for the multi-turnover assays was expressed in *E. coli* BL21 (DE3) RIL cells and purified from bacterial lysates with a glacial acetic acid method (Kaiser et al., 2011). It was further purified by gravity S column ion exchange chromatography and size exclusion chromatography in Buffer B. No-lysine and single-lysine Ub variants as well as WT Ub used for the Ub chain type determination assay were expressed as GST-HRV 3C fusions in *E. coli* BL21 (DE3) RIL cells in a Terrific Broth (TB) medium overnight at 20°C and purified by glutathione affinity chromatography. To liberate the GST tag, elutions were incubated with HRV13 3C protease for 3 hours at room temperature. Further purification was done with size exclusion chromatography in Buffer B that separated Ub from the free GST and uncleaved GST-fusion proteins. Cys-ubiquitin used for fluorescent labeling was expressed as a GST-TEV fusion in *E. coli* BL21 (DE3) RIL cells in a Terrific Broth (TB) medium overnight at 23°C and purified by glutathione affinity chromatography. After incubating GST resin with the bacterial lysate, TEV protease was added to the beads to liberate ubiquitin from the GST tag. Further purification was done with anion exchange chromatography and size exclusion chromatography in 50 mM HEPES pH 7.0 and 150 mM NaCl.

To generate fluorescent Mdh2 (Mdh2-FAM) for ubiquitylation assays, fluorescein was attached to its C terminus using a sortase A-mediated reaction (Guimaraes et al., 2013). For the reaction, 50 μM Mdh2 fused to a C-terminal sortag (LPETGG) and a 6xHis tag was mixed with 250 μM of a fluorescent peptide (GGGGG-FAM) and 50 μM of sortase A. Reaction was carried out for 30 minutes on ice in a buffer consisting of 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 10 mM CaCl₂. To get rid of unreacted Mdh2, the reaction mixture was supplemented with 20 mM imidazole and passed through Ni-NTA Sepharose resin. Labeled Mdh2 was purified with size exclusion chromatography in Buffer B.

For fluorescent labeling of ubiquitin, a version of Ub with a cysteine introduced upstream its N-terminal methionine (Cys-Ub) was used. To reduce the cysteine before mixing it with fluorescein-5-maleimide, Cys-Ub was supplemented with 20 mM DTT. After 10 minutes of incubation, the protein was desalted into 50 mM HEPES pH 7.0, 150 mM NaCl with a NAP-5 column. Desalted Cys-Ub was mixed with fluorescein-5-maleimide (dissolved in DMSO) at 1:5 molar ratio and incubated at 4°C overnight. The reaction was quenched by adding 5 mM DTT and the reaction mixture was desalted into Buffer B with a PD-10 column. The final purification was performed by size exclusion chromatography in Buffer B.

Biochemical assays

Unless otherwise stated, *in vitro* ubiquitylation monitored a fluorescently-labeled substrate Mdh2-FAM. All assays were performed at room temperature in a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl, 5 mM ATP, 10 mM MgCl₂ and 0.25 mg/mL BSA. At each time point, a part of the reaction mixture was quenched by mixing it with SDS-PAGE loading dye. To check the activity of Gid4, Gid5 and Gid1 mutants, as well as to show that Mdh2 ubiquitylation is Gid4 and E2 dependent, the reaction involved mixing of 0.2 μM Uba1, 2 μM Ubc8-6xHis, 1 μM GID^{Ant} (containing either WT or indicated mutants of Gid5 or Gid1), 1 μM Gid4 (Δ1-116; WT or an indicated mutant), 1 μM Mdh2-FAM and 100 μM Ub. For the assay testing importance of the N-terminal meander of Gid4, the exogenously added Gid4 started with the residue at position 80 or 117.

The assay validating dependence of GID^{SR4} activity on the N-terminal proline of its substrate was performed at the same condition; but western blotting with anti-6xHis antibodies was used to visualize ubiquitylation of unlabeled WT and P2S mutant of Mdh2-6xHis (note that complete cleavage of the N-terminal Met residue was confirmed by mass spectrometry).

For testing the mutations in Gid2 and Gid9 RING domains, the assay contained 0.2 μM Uba1, 2 μM Ubc8-6xHis, 1 μM GID^{SR4} (containing either WT or indicated mutants of Gid2 or Gid9), 1 μM Mdh2-FAM and 100 μM Ub. To test the activity of an alternative substrate receptor Gid10, the reaction was composed of 0.2 μM Uba1, 2 μM Ubc8-6xHis, 0 or 1 μM GID^{Ant}, 0 or 1 μM Gid10 (either Δ1-51 or Δ1-56 and Δ289-292 version) or 1 μM Gid4 (Δ1-116), 1 μM Mdh2-FAM and 100 μM Ub. Determination of the type of Ub chain formed by GID^{SR4} was done by using a panel of single-Lys Ub variants, with all other lysines mutated to arginines. The reaction mixture was composed of 0.2 μM Uba1, 2 μM Ubc8-6xHis, 1 μM GID^{SR4}, 1 μM Mdh2-FAM and 20 μM Ub (WT, lysineless K0 Ub or any of the single Lys Ub variants).

In order to validate the preferred ubiquitylation sites on Mdh2 mapped with mass-spectrometry, we have compared ubiquitylation of WT Mdh2 with its version, in which all the preferred target lysines were mutated to arginines (K254R, K256R, K259R, K330F, K360R, K361R). The assay comprised 0.2 μM Uba1, 2 μM Ubc8-6xHis, 0.5 μM GID^{SR4}, 0.5 μM Mdh2-6xHis (WT or the preferred K > R mutant) and 100 μM fluorescently labeled FAM-Cys-Ub. Progress of the reaction was visualized by monitoring the fluorescently-labeled Cys-Ub.

To analyze if addition of the substrate receptor Gid4 to the GID^{Ant} has any impact on its intrinsic E3 ligase activity, a substrate-independent discharge assay was employed. To separate an effect of E2~Ub discharge from its E1-dependent loading, this assay was performed in a pulse-chase format. In the pulse reaction, loading of E2 was performed by mixing 0.5 μM Uba1, 10 μM Ubc8-6xHis and 30 μM Ub in a buffer containing 25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM ATP and 2.5 mM MgCl₂. After 15 minutes incubation of the pulse mixture at room temperature, E2 loading was stopped by addition of 50 mM EDTA. For the chase reaction, the quencher pulse mixture was mixed with an equal volume of the chase initiating mixture containing 1 μM GID^{Ant}, 0, 1 or 5 μM Gid4 (Δ1-116) and 40 μM lysine pH 8.0 in 25 mM HEPES pH 7.5 and 100 mM NaCl, and incubated at room temperature. The discharge was quenched at each of the time points by mixing the discharge reaction with SDS-PAGE loading dye without any reducing agent and visualized with a non-reduced SDS-PAGE stained with Coomassie.

To test if our recombinant Mdh2 binds to Gid4 according to the Pro/N-degron pathway, purified GST-tagged Gid4 (Δ1-116) was mixed with two-fold molar excess of Mdh2-6xHis (WT or the P2S mutant) in a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT. After incubating the proteins for 30 minutes on ice, 20 μL of GST resin was added to the mixture and further incubated for 1 hour. As a negative control, Mdh2-6xHis was mixed with GST resin in absence of Gid4. GST beads were then thoroughly washed and proteins were eluted. The elution fractions were analyzed with SDS-PAGE to check for the presence or absence of an Mdh2 band. A similar binding test was applied to check if an alternative substrate receptor Gid10 interacts with GID^{Ant}. Here, GID^{Ant}, which was Twin Strep-tagged on Gid8 C terminus, was mixed with a two-fold molar excess of Gid10 (Δ1-56) and Strep Tactin pull-down was performed. As a negative control, Gid10 was mixed with Strep Tactin resin in the absence of GID^{Ant}.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the *in vivo* Fbp1 degradation assay, experiments were performed in at least biological triplicate. Fbp1 degradation pattern was visualized by western-blot and the bands were quantified. Bars on graphs represent average ($n > 3$) and error bars represent standard deviation.

DATA AND CODE AVAILABILITY

The PDB and EM maps are available from the RCSB and EMDB as follows: GID^{SR4} minus Gid2/ΔGid9^{RING}, EMDB: 10333, PDB ID 6SWY; GID^{Scaffold}, EMDB:10328; GID^{Scaffold} plus SR^{Gid4}, EMDB:10330; GID^{Scaffold} plus SR^{Gid10}, EMDB:10329; GID^{SR4}, EMDB: 10327; GID^{Ant}, EMDB: 10326; Endogenous GID^{Ant}, EMDB: 10331; GID^{SR4} ΔRINGS, EMDB: 10332. Proteomics data can be accessed on Proteome Xchange via the Pride database with the dataset identifier PXD015396.

The unprocessed image data are available at: <https://doi.org/10.17632/nd5zc59vfg.1>.

3.6 GID E3 ligase supramolecular chelate assembly configures

multi-pronged ubiquitin targeting of an oligomeric metabolic enzyme

Dawafuti Sherpa^{1,3}, Jakub Chrustowicz^{1,3}, Shuai Qiao¹, Christine R. Langlois¹, Laura A. Hehl¹, Karthik Varma Gottemukkala¹, **Fynn M. Hansen**², Ozge Karayel², Susanne von Gronau¹, J. Rajan Prabu¹, Matthias Mann², Arno F. Alpi¹, Brenda A. Schulman^{1,4}

¹Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried 82152, Germany. ²Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried 82152, Germany

Published in *Molecular Cell* (2021)

It has long been known that ubiquitin is added to target proteins through a ubiquitination cascade involving various combinations of E1s, E2s and E3s. Depending on the E3 ligase, the mode of action how ubiquitin is transferred from an E2 to a target protein can show substantial differences (see also section 1.3.1). In this study, the group of Prof. Brenda Schulman reconstituted a minimal GID E3 ligase that is active towards tetrameric Fbp1. Strikingly, the E3 ligase assembly resembles a behemoth organometallic supramolecular chelate.

I contributed to this study by mapping ubiquitination sites on Fbp1. To this end, I performed a multistage digestion strategy tailored to the Fbp1 protein sequence to optimize the sequence coverage and the number of observable modification sites. This greatly helped in the identification of ubiquitination sites on Fbp1 that are crucial for the composition of the supramolecular chelate E3-Fbp1 complex.

Article

GID E3 ligase supramolecular chelate assembly configures multipronged ubiquitin targeting of an oligomeric metabolic enzyme

Dawafuti Sherpa,^{1,3} Jakub Chrustowicz,^{1,3} Shuai Qiao,¹ Christine R. Langlois,¹ Laura A. Hehl,¹ Karthik Varma Gottemukkala,¹ Fynn M. Hansen,² Ozge Karayel,² Susanne von Gronau,¹ J. Rajan Prabu,¹ Matthias Mann,² Arno F. Alpi,¹ and Brenda A. Schulman^{1,4,*}

¹Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried 82152, Germany

²Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried 82152, Germany

³These authors contributed equally

⁴Lead contact

*Correspondence: schulman@biochem.mpg.de

<https://doi.org/10.1016/j.molcel.2021.03.025>

SUMMARY

How are E3 ubiquitin ligases configured to match substrate quaternary structures? Here, by studying the yeast GID complex (mutation of which causes deficiency in glucose-induced degradation of gluconeogenic enzymes), we discover supramolecular chelate assembly as an E3 ligase strategy for targeting an oligomeric substrate. Cryoelectron microscopy (cryo-EM) structures show that, to bind the tetrameric substrate fructose-1,6-bisphosphatase (Fbp1), two minimally functional GID E3s assemble into the 20-protein Chelator-GID^{SR4}, which resembles an organometallic supramolecular chelate. The Chelator-GID^{SR4} assembly avidly binds multiple Fbp1 degrons so that multiple Fbp1 protomers are simultaneously ubiquitylated at lysines near the allosteric and substrate binding sites. Importantly, key structural and biochemical features, including capacity for supramolecular assembly, are preserved in the human ortholog, the CTLH E3. Based on our integrative structural, biochemical, and cell biological data, we propose that higher-order E3 ligase assembly generally enables multipronged targeting, capable of simultaneously incapacitating multiple protomers and functionalities of oligomeric substrates.

INTRODUCTION

Cells rapidly adapt their metabolic pathways in response to nutrient availability (Tu and McKnight, 2006; Zaman et al., 2008; Zhu and Thompson, 2019). Shifts in metabolic enzyme activities are achieved by regulation at every conceivable level. Metabolite-responsive transcriptional programs activate pathways that maximally use available nutrients and repress those rendered unnecessary or counterproductive. For oligomeric enzymes, catalytic activities are subject to metabolite-mediated allosteric control (Koshland, 1963a, 1963b; Monod et al., 1963). In eukaryotes, undesired metabolic activities are often terminated by ubiquitin-mediated degradation (Nakatsukasa et al., 2015).

Degradation is typically controlled by recognition of proteins as substrates of E3 ubiquitin (Ub) ligases. However, little is known about whether or how E3 ligases are specifically tailored for oligomeric assemblies of metabolic enzymes. One of the first identified targets of nutrient-dependent degradation, budding yeast fructose-1,6-bisphosphatase (Fbp1), is an oligomer (Chiang and Schekman, 1991). Fbp1 is a gluconeogenic enzyme essential for yeast growth on non-fermentable carbon sources. A shift from gluconeogenic to glycolytic conditions renders gluconeogenesis

superfluous. Accordingly, Fbp1 activity and expression are curtailed (Gancedo, 1971; Schork et al., 1994a, 1994b, 1995). The switch to glycolytic conditions induces Ub-mediated degradation of Fbp1 and other gluconeogenic enzymes, including malate dehydrogenase (Mdh2) and phosphoenolpyruvate carboxylase (Pck1), mediated by the multiprotein E3 ligase termed "GID"; yeast mutants of Gid subunits are glucose-induced-degradation deficient (Braun et al., 2011; Chiang and Schekman, 1991; Hämmerle et al., 1998; Menssen et al., 2012; Regelmann et al., 2003; Santt et al., 2008; Schork et al., 1994b, 1995). Although the GID E3 is conserved across eukaryotes and regulates important physiology (Lampert et al., 2018; Liu et al., 2020; Liu and Pfirrmann, 2019; Salemi et al., 2017), its regulation and targets are best characterized in budding yeast.

Much like well-studied multiprotein E3 ligases, such as anaphase-promoting complex/cyclosome (APC/C) or cullin-RING ligases, GID is not a singular complex—a core catalytic and scaffolding assembly is modulated by other subunits (Barford, 2020; Karayel et al., 2020; Liu and Pfirrmann, 2019; Melnykov et al., 2019; Qiao et al., 2020; Rusnac and Zheng, 2020; Watson et al., 2019). The constituents of various GID assemblies and how they achieve regulation are beginning to emerge. Previous structural studies have elucidated the core assembly and



recapitulated some GID regulation (Qiao et al., 2020). Briefly, a core inactive complex, GID^{Ant}, contains the heterodimeric E3 ligase RING and RING-like subunits (Gid2 and Gid9) and scaffold subunits (Gid1, Gid5, and Gid8). Coexpression of these subunits in insect cells enables purification of recombinant GID^{Ant} and systematic interrogation of GID functions. Within the GID^{Ant} scaffold, Gid5 can bind the structurally homologous, interchangeable substrate-binding receptors Gid4 and Gid10 (Karayel et al., 2020; Melnykov et al., 2019; Qiao et al., 2020). Of these, the molecular basis of substrate binding by Gid4 is well understood: glucose-induced incorporation of Gid4 into the GID E3 enables recognition of substrate “Pro/N-degron” motifs depending on an N-terminal proline (Chen et al., 2017; Dong et al., 2018; Hämmerle et al., 1998; Regelmann et al., 2003; Santt et al., 2008). Indeed, *in vitro*, adding Gid4 transforms GID^{Ant} into an active GID^{SR4} complex that collaborates with the cognate E2, Ubc8 (also known as Gid3) to ubiquitylate Mdh2, as explained by a structure of GID^{SR4} (Qiao et al., 2020). Mutations probing the GID^{SR4} structure also showed that this assembly is required for glucose-induced Fbp1 degradation *in vivo* (Qiao et al., 2020).

Perplexingly, despite the crucial role of Fbp1 in regulating gluconeogenesis, its ubiquitylation has not been reconstituted *in vitro* using defined GID E3 ligase components. *In vivo*, Fbp1 degradation depends on another protein, Gid7, which associates with other Gid subunits (Menssen et al., 2012; Regelmann et al., 2003; Santt et al., 2008). Gid7 is evolutionarily conserved across eukaryotes. Mammals even have two orthologs, WDR26 and MKLN1, which are subunits of the “CTLH” complex that corresponds to the yeast GID E3 (Boldt et al., 2016; Francis et al., 2013; Kobayashi et al., 2007; Lampert et al., 2018; Liu and Pfirrmann, 2019; Salemi et al., 2017). The CTLH E3, named for the preponderance of CTLH domains (in Gid1, Gid2, Gid7, Gid8, and Gid9 and their orthologs), has intrinsic E3 ligase activity, although Pro/N-degron substrates have not yet been identified despite human Gid4 binding this motif (Cao et al., 2020; Dong et al., 2018; Lampert et al., 2018; Liu et al., 2020; Liu and Pfirrmann, 2019; Maitland et al., 2019; Zavortink et al., 2020).

Here we reconstitute a minimal GID E3 ligase active toward Fbp1 by combining GID^{SR4} and Gid7. Cryoelectron microscopy (cryo-EM) reveals its structure as a 20-protein supramolecular chelate E3 ligase assembly specifically tailored for Fbp1’s quaternary structure. Structural and biochemical data suggest that the human Gid7 orthologs likewise transform a GID^{SR4}-like E3 ligase core into higher-order assemblies. Our data reveal supramolecular chelate assembly of a pre-existing, functionally competent E3 ligase complex as a structural and functional principle to achieve multipronged Ub targeting tailored to an oligomeric substrate.

RESULTS

Reconstitution of Fbp1 ubiquitylation

Considering that the Gid7 protein, not visualized previously, is required for glucose-induced Fbp1 degradation *in vivo* (Regelmann et al., 2003), we tested its effect *in vitro*. Our assay setup probes modulation of the core recombinant GID^{Ant} assembly upon adding other purified components individually or in combination. First, adding Gid4 marginally stimulated Fbp1 ubiquityla-

tion despite substantially potentiating ubiquitylation of Mdh2, another canonical Pro/N-degron substrate, and Pck1, whose recognition by the GID E3 remains elusive (Figure 1A). However, adding Gid7 together with Gid4 substantially increased Fbp1 ubiquitylation. Comparing reactions with wild-type (WT) Ub or a Ub version lacking lysines (K0Ub) that cannot form polyUb chains indicated that adding Gid7 increases substrate consumption, the number of modified Fbp1 sites, and the number of Ubs in polyUb chains (Figures 1A and 1B). Second, the remarkable activation upon addition of Gid7 was specific to Fbp1; effects on Pck1 were negligible, and effects on Mdh2 were nuanced in increasing polyUb chain length while attenuating the amount of Mdh2 molecules consumed in the assay (Figure 1A). Third, adding Gid7 actually suppressed intrinsic GID E3 ligase activity, as shown by effects on Ub transfer from a pre-formed Ubc8~Ub intermediate to free lysine in solution (Figure S1A). Binding of Fbp1’s degron per se is insufficient to overcome this inhibition because Gid7 likewise subdued ubiquitylation of a model peptide substrate in which Fbp1’s degron sequence, PTLV, is connected to a lysine acceptor through an intervening flexible linker (Figure S1B).

To gain mechanistic insights, we quantified effects of including Gid7 in a chromatographically purified version of the E3 by performing enzyme kinetics. Compared with GID^{SR4}, a version of the E3 complex fully incorporating Gid7 displayed a relatively 10-fold lower Michaelis-Menten constant, K_m , for Fbp1 ubiquitylation and 10-fold increase in the reaction turnover number k_{cat} (Figures 1C, 1D, S1C, and S1D). Adding purified Gid7 to GID^{SR4} had similar effects (Figures 1C and S1C).

Consistent with the biochemical data, glucose-induced ubiquitylation of Fbp1 *in vivo* is impaired by Gid7 deletion (Figure 1E). To examine effects on degradation, we employed a promoter reference technique that monitors degradation of exogenously expressed proteins (here, C-terminally FLAG-tagged Fbp1, Mdh2, or Pck1) while normalizing for effects on expression (Chen et al., 2017; Oh et al., 2017). Our assay agreed with prior studies showing that glucose-induced degradation of Fbp1, Mdh2, and Pck1 depends on Gid4 (Chen et al., 2017; Qiao et al., 2020; Regelmann et al., 2003; Santt et al., 2008). However, Gid7 deletion substantially stabilized only Fbp1 (Figure 1F). This deficit in Fbp1 degradation upon Gid7 deletion was not rescued by Gid4 overexpression (Figure S1E). Also, quantitative mass spectrometry analyses of the yeast proteome confirmed that, of known gluconeogenic GID E3 substrates, Fbp1 was most affected by Gid7 deletion (Figure S1F).

A supramolecular Chelator-GID^{SR4} E3 assembly encapsulates the tetrameric Fbp1 substrate

To understand the mechanism of Fbp1 recognition by the GID E3, we purified an Fbp1-active recombinant complex and analyzed its structure by cryo-EM (Figures S2A and S4; Table S1). A 13-Å-resolution map of the assembly even without the substrate showed a remarkable GID E3 structure: an exterior oval supporting several inward-pointing globular domains. Strikingly, the longest exterior dimension of 305 Å is roughly comparable with that of a singly capped 26S proteasome, 1.3 times that of the multiprotein Fanconi anemia E3 ligase complex and 1.5 times that of APC/C (Figure 2A)

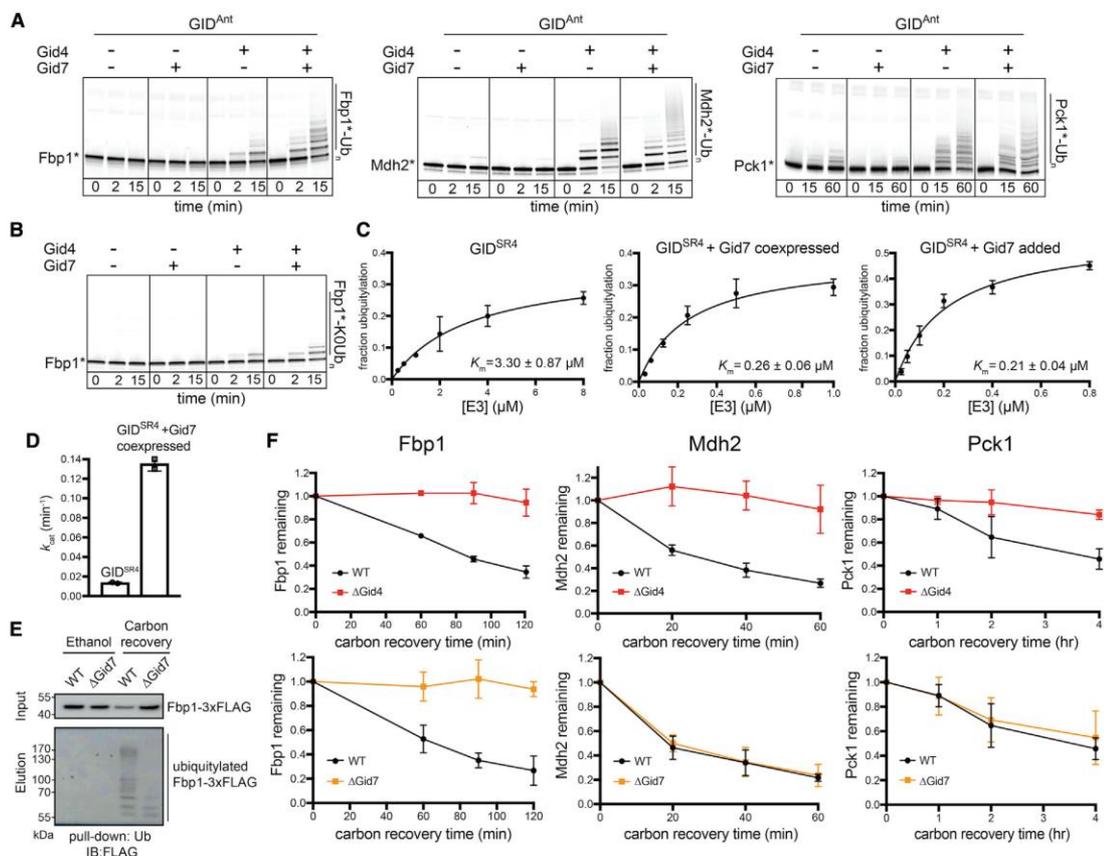


Figure 1. Fbp1 ubiquitylation and degradation require a distinct Gid7-containing GID E3 ligase

(A) Fluorescence scans of SDS-PAGE gels showing *in vitro* ubiquitylation assays. These assays test the roles of Gid4 and Gid7 in ubiquitylation of C-terminally fluorescently labeled Fbp1 (left), Mdh2 (center), and Pck1 (right). GID^{Ant} contains 2 protomers each of Gid1 and Gid8 and 1 of Gid2, Gid5, and Gid9. An asterisk indicates that substrates are fluorescently labeled.

(B) *In vitro* ubiquitylation assay as in (A) but performed with lysine-less Ub (K0Ub) to determine the number of Fbp1 ubiquitylation sites.

(C) Plots showing fraction of Fbp1 ubiquitylation as a function of concentration of GID^{SR4} (left) or its complex with Gid7 (center and right). K_m values were determined by fitting to the Michaelis-Menten equation. Error bars, SD ($n = 2$).

(D) Comparison of k_{cat} between GID^{SR4} and its complex with Gid7, determined from plots in Figure S1D. Error bars, SD ($n = 2$).

(E) Assessing *in vivo* ubiquitylation of Fbp1 (C-terminally 3×FLAG-tagged at the endogenous locus) under carbon starvation (ethanol) and after 2 h of carbon recovery in WT and Δ Gid7 yeast strains. Following capture of ubiquitylated proteins with TUBEs (tandem ubiquitin binding entities), Fbp1-3×FLAG was visualized by anti-FLAG immunoblotting.

(F) Glucose-induced degradation *in vivo* of exogenously expressed substrates Fbp1 (left), Mdh2 (center), and Pck1 (right), quantified using the promoter reference technique. Substrate levels were quantified as the ratio of substrate detected relative to the level after switching from carbon starvation to carbon recovery conditions in WT, Δ Gid4 (top panels), and Δ Gid7 (bottom panels) strains. Points represent mean, and error bars represent SD ($n \geq 3$).

See also Figure S1.

(Brown et al., 2016; Chen et al., 2016; Haselbach et al., 2017; Lander et al., 2012; Schweitzer et al., 2016; Shakeel et al., 2019; Wehmer et al., 2017). Unlike these compact assemblies, however, this GID complex displays a behemoth hollow center with interior edges of 270 and 130 Å in the longest and shortest dimensions, respectively—larger than a cullin-RING ligase ubiquitylating a substrate (Baek et al., 2020).

The organization of the oval GID assembly was gleaned from comparison with cryo-EM maps of subcomplexes (Figure 2B; Table S1). Two copies of the previously defined GID^{SR4} structure (Qiao et al., 2020) fit in the large assembly. An additional Gid1-Gid8 subcomplex can be observed bound to GID^{SR4}. These duplicated Gid1 and Gid8 protomers are components of recombinant GID^{Ant} used for biochemical assays (Qiao et al., 2020) but

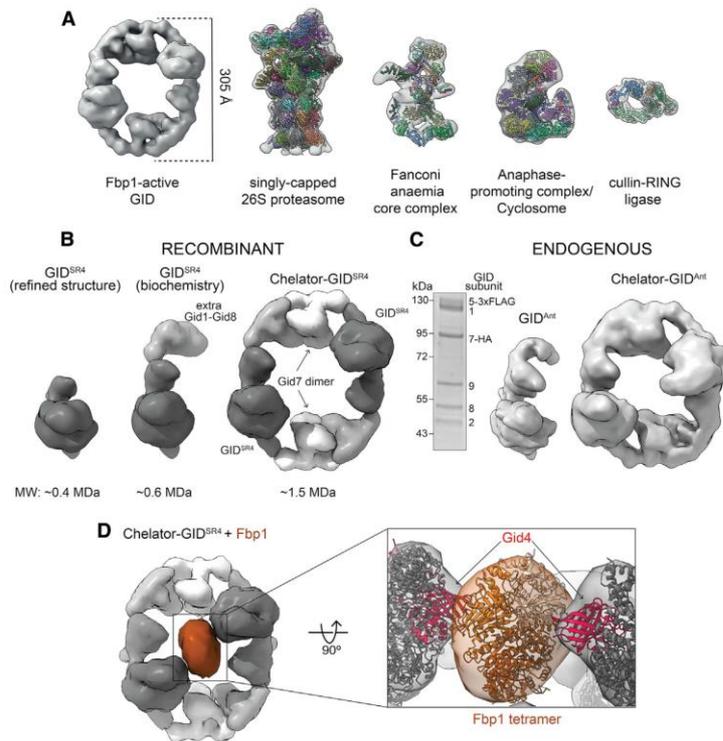


Figure 2. Multidentate capture of the Fbp1 tetramer by the Chelator-GID^{SR4} assembly

(A) Cryo-EM map of GID E3 active toward Fbp1 compared for scale with low-pass-filtered maps of the singly capped 26S proteasome (EMDB: EMD-3536; PDB: 5MPB), Fanconi anaemia core complex (EMDB: EMD-10290; PDB: 6SRI), APC/C (EMDB: EMD-3433; PDB: 5L9T), and cullin-RING E3 ubiquitylation complex (EMDB: EMD-10585; PDB: 6TTU).

(B) Cryo-EM maps and molecular weights of recombinant GID assemblies. Structurally determined GID^{SR4} (left, low-pass-filtered, dark gray, EMDB: EMD 10327; PDB: 6SWY) is a stoichiometric complex of Gid1, Gid8, Gid5, Gid4, Gid2, and Gid9. The purification conditions used here include an additional Gid1-Gid8 subcomplex (gray) bound to GID^{SR4} center, taken for the biochemical assays). The oval higher-order Chelator-GID^{SR4} assembly includes Gid7 dimers (right, white).

(C) Coomassie-stained SDS-PAGE (left) and cryo-EM maps of endogenous yeast GID^{Ant} (center) and Chelator-GID^{Ant} (right) assemblies (prepared by anti-FLAG immunoprecipitation of lysates from yeast with Gid5 3×FLAG tagged and Gid7 hemagglutinin (HA) tagged at their endogenous loci and grown under conditions when Gid4 is not induced). (D) Cryo-EM map of Chelator-GID^{SR4} (gray) bound to the Fbp1 tetramer (brown). The close up shows 2 red Gid4 protomers (modeled from PDB: 6SWY) simultaneously contacting the docked Fbp1 crystal structure.

See also Figure S2 and Tables S1 and S2.

are not visible upon map refinement to high resolution. We interpreted the remaining density in the large oval GID assembly as Gid7 dimers, one at each vertex, given size exclusion chromatography-multi angle light scattering (SEC-MALS) data indicating that purified Gid7 dimerizes (Figure S2B). The data reveal a 1.5-MDa eicosameric GID assembly composed of 4 Gid1: 2 Gid2: 2 Gid4: 2 Gid5: 4 Gid7: 4 Gid8: 2 Gid9 protomers.

We sought to determine whether this GID assembly might be formed *in vivo*. Prior studies did (Santt et al., 2008) or did not (Qiao et al., 2020) observe Gid7 cosedimenting with other GID proteins in density gradients. This raised the possibility that, like the equally giant 26S proteasome, some subunits or regulatory partners may be prone to dissociation; for example, based on lysis conditions (Leggett et al., 2002). Thus, we examined sedimentation of a core subunit, Gid8 tagged at the endogenous locus, as a marker for a GID assembly because it cosediments with all other GID^{SR4} subunits even under relatively harsh lysis conditions (Qiao et al., 2020). Yeast lysates prepared by cryomilling were subjected to sucrose density gradient fractionation. Anti-FLAG immunoblotting showed Gid8 migrating at a lower molecular weight in a Gid7 deletion compared with the WT, irrespective of whether yeast was grown under carbon starvation or recovery in glucose, which induces GID E3 ligase activity (Figure S2C). Moreover, cryo-EM data of endogenous GID purified from yeast grown under carbon starvation yielded 3D recon-

structions corresponding to the recombinant assemblies with and without Gid7 at 14.2- and 9.5-Å resolution, respectively (Figures 2C and S2D).

Why is the minimum E3 ligase for Fbp1 so gigantic and hollow? Given the substantial effect on K_m in our enzyme kinetics analyses, we hypothesized that such an assembly would form to accommodate the substrate. To characterize the substrate, we determined the crystal structure of yeast Fbp1, which confirmed its tetrameric assembly (Figures 2D and S2B; Table S2). We next resolved a cryo-EM structure with Fbp1 bound to the GID E3, which led to several conclusions (Figure 2D; Table S1). First, Fbp1 was readily docked in the center of the large GID E3 oval. Second, two Fbp1 edges approach the substrate binding Gid4 subunits within each GID^{SR4} on opposite sides of the oval. Third, the density attributed to Gid7 does not directly contact Fbp1 but connects two Fbp1-binding GID^{SR4} complexes. Thus, Gid7 activates GID E3 activity toward Fbp1 indirectly by driving supramolecular assembly.

The resultant GID assembly resembles an organometallic supramolecular chelate in which multiple giant organic molecules capture a much smaller ligand through multiple discrete points of contact. Thus, we call the giant oval complex “Chelator-GID^{SR4}” based on its supramolecular assembly in which two GID^{SR4} complexes simultaneously capture degrons displayed from two protomers in the tetrameric Fbp1 substrate.

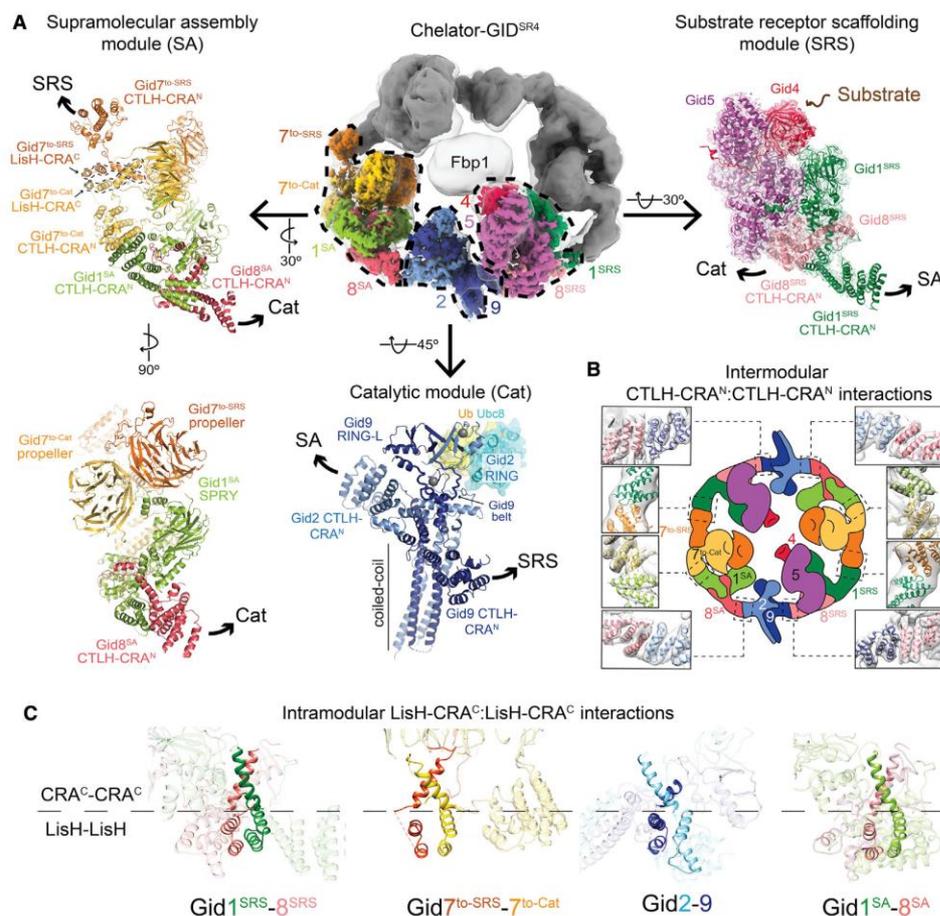


Figure 3. High-resolution details of Chelator-GID^{SR4} modular assembly

(A) Focused refined maps of the substrate receptor scaffolding (SRS), catalytic (Cat), and supramolecular assembly (SA) modules, colored according to subunit identity, fit in half of the overall map of Fbp1-bound Chelator-GID^{SR4} (top center). The GID^{SR4} structure (PDB: 6SWY) fits the SRS module (Gid1^{SRS}, dark green; Gid8^{SRS}, salmon; Gid5, purple; Gid4, red). A brown arrow points to Gid4's substrate binding site (top right). The Cat module comprises Gid2 (sky blue) and Gid9 (navy). Zinc ions are shown as gray spheres. Ubc8-Ub was modeled by aligning Gid2 RING with an E2-Ub-bound RING structure (PDB: 5H7S). The SA module comprises Gid1^{SA} (green), Gid8^{SA} (pink) and 2 Gid7 protomers, Gid7^{to-Cat} (yellow), and Gid7^{to-SRS} (orange) facing the Cat or SRS module, respectively. Superscript text refers to a module for a given Gid1 or Gid8 protomer. Arrows point to connected modules.

(B) Cartoon of Chelator-GID^{SR4} with close ups of intermodule CTLH-CRA^N:CTLH-CRA^N interactions fit into the map of Chelator-GID^{SR4} (gray).

(C) Intramodule LisH-CRA^C:LisH-CRA^C (solid ribbon) interactions in Chelator-GID^{SR4}.

See also Figures S3 and S4 and Table S1.

High-resolution structures of modules in Chelator-GID^{SR4}

A series of focused refinements enabled building atomic models of the three functionally distinct modules comprising Chelator-GID^{SR4} (Figures 3A, S2E, S3A, and S4; Table S1): (1) the substrate receptor scaffolding (SRS) module contained in GID^{SR4}, responsible for bridging the substrate receptor to the other E3 ligase subunits; (2) the catalytic (Cat) module, also present in GID^{SR4}, which binds and activates the Ubc8~Ub inter-

mediate; and (3) a previously undescribed supramolecular assembly (SA) module.

A 3.4-Å map of the Chelator-GID^{SR4} SRS module fit the prior coordinates for this region (PDB: 6SWY) (Figures 3A and S4B). As described previously, the globular substrate-binding domain of Gid4 fits snugly in a complementary concave surface of the scaffold subunit Gid5. This arrangement is supported by a base from Gid1^{SRS} and Gid8^{SRS}, which form an intricate heterodimer involving their LisH-CTLH-CRA domains.

Focused refinement over the Cat module yielded a 3.8-Å-resolution reconstruction (Figures 3A and S4C). The map quality permitted *de novo* building and refinement of atomic coordinates for the majority of Gid2 and Gid9 (Figure S3A). The catalytic function is mediated by a region of Gid2 that adopts an E3 ligase RING domain fold (albeit stabilized by a single zinc in the E2~Ub binding site) together with a portion of Gid9 that adopts a unique RING-like (RING-L) structure (Figure S3B; Braun et al., 2011; Qiao et al., 2020; Regelmann et al., 2003). Folding of the Gid2 RING depends on its incorporation into the intricately configured Gid2-Gid9 heterodimer. The Gid2 RING is embedded in an unprecedented intermolecular heart-shaped domain, stabilized by Gid9 elements, including an intermolecular zinc-binding domain; a belt that encases roughly three quarters of the base of Gid2's RING; the RING-L domain, which packs against the remaining side of Gid2's RING; and the extreme C terminus, which contributes to Gid2's RING in a manner analogous to canonical RING dimers (Budhidarmo et al., 2012). Gid2 and Gid9 are further intertwined by their N termini co-assembling in an ~70-Å-long 4-helix coiled coil (Figures 3A and S3A).

Within Chelator-GID^{SR4}, the two Gid2-Gid9 E3 ligase domains face the two degron-binding Gid4 subunits. A model of the Gid2 RING-Ubc8~Ub intermediate based on published isolated RING E3-E2~Ub complexes shows the Gid2 RING domain recruiting Ubc8, whereas its linked Ub would be activated by Gid2 and Gid9 in the canonically activated conformation (Figures 3A and S3B; Dou et al., 2012; Plechanovová et al., 2012; Pruneda et al., 2012). The model explains the previously reported effects of Gid2 and Gid9 point mutations on Fbp1 degradation (Qiao et al., 2020).

A 3.6-Å resolution map of the SA module within Chelator-GID^{SR4} enabled building of an atomic model (Figures 3A and S4D). The two Gid7 protomers form an asymmetric dimer on one side of the module. Gid1^{SA} and Gid8^{SA} form an interdigitated scaffold that connects the Gid7 dimer to the Cat module.

Each Gid7 protomer consists of an N-terminal LisH-CTLH-CRA motif and an atypical β-propeller. The LisH-CTLH-CRA motifs form elongated helical double-sided dimerization domains (Figure S3C). The LisH and CTLH helices initially progress in one direction. The distal end is capped by the first two CRA helices. The remaining CRA helices reverse and traverse the length of the domain, pack against CTLH helices along the way, and terminate adjacent to the LisH helices. We refer to one side of the LisH-CTLH-CRA structure as "LisH-CRA^C" because it contains the LisH and C-terminal CRA helices. Accordingly, the other side is called "CTLH-CRA^N." The Gid7 LisH-CRA^C motifs mediate homodimerization, much like LisH-CRA^C motifs mediate heterodimerization between Gid1^{SRS} and Gid8^{SRS} and between Gid2 and Gid9 (Qiao et al., 2020).

β-Propellers are protein interaction domains formed by toroidally arranged β sheet "blades" (Chen et al., 2011a). The 7-bladed propellers from the two Gid7 protomers ensue from the LisH-CTLH-CRA motifs at different relative angles and interact with each other. The resultant asymmetric double-propeller domain binds part of Gid1^{SA}. The SA module is further stabilized by distinctive interactions between the CTLH-CRA^N domains from Gid1^{SA}, a loop from Gid8^{SA}, and the CTLH-CRA^N

domain from a Gid7 protomer we call Gid7^{to-Cat} because it points toward the Cat module (Figure S3D). The remainder of the Gid1^{SA} and Gid8^{SA} subcomplex superimposes on corresponding regions of Gid1^{SRS} and Gid8^{SRS}. At the two edges of the SA module, the CTLH-CRA^N domains from the SRS-facing Gid7 protomer (Gid7^{to-SRS}) and Gid8^{SA} connect to the SRS and Cat modules, respectively.

Supramolecular chelate assembly is supported by inter- and intramodule LisH-CTLH-CRA domain interactions

The relative arrangement of E3 ligase elements—the Gid4 substrate receptor and the Gid2-Gid9 RING-RING-L complex—in Chelator-GID^{SR4} depends on the exterior oval band. The oval is established by two types of intersubunit interactions—within the modules and mediating intermodule connections—in a daisy chain-like arrangement of LisH-CTLH-CRA domains (Figures 3B and 3C).

In Chelator-GID^{SR4}, the modules are connected to each other by outward-facing heterotypic dimerization of CTLH-CRA^N domains at the edges of each module (Figure 3B). The CTLH-CRA^N domains connect modules in a side-by-side manner. In the GID^{SR4} assembly, the SRS and Cat modules are adjoined by interactions between the CTLH-CRA^N domains of Gid8^{SRS} and Gid9. The Cat and SA modules are bridged by interactions between the CTLH-CRA^N domains of Gid2 and Gid8^{SA}. Notably, Gid2's CTLH-CRA^N domain also packs against Gid9's RING-L domain, which may explain how formation of the Chelator-GID^{SR4} assembly affects intrinsic Ub transferase activity (Figures 3A, S1A, and S3B). The oval structure also depends on adjoining the SRS and SA modules through interactions between the CTLH-CRA^N domains of Gid1^{SRS} and Gid7^{to-SRS}. Despite the similarity of intermodule interactions at a secondary structural level, specificity is dictated by contacts between domains, ensuring formation of the Chelator-GID^{SR4} assembly.

Chelator-GID^{SR4} assembly mediates avid recruitment of the tetrameric substrate Fbp1

Comparing the major classes of Chelator-GID^{SR4} alone or bound to Fbp1 showed relative repositioning of the SRS module toward the center of the oval to bind the substrate, resembling a Venus flytrap capturing its prey (Figure 4A). An individual Fbp1 Pro/N-degron was visualized bound to Gid4 in a locally refined map of SRS (Figures 4B and S4B). Fbp1's N-terminal proline and two subsequent residues are recruited much like short peptides binding human Gid4 (Chen et al., 2017; Dong et al., 2018; Hämmerle et al., 1998). Comparing the substrate-bound Chelator-GID^{SR4} structure with the substrate-free GID^{SR4} (Qiao et al., 2020) shows remodeling of several Gid4 loops to embrace the N-terminal residues PTL of the Fbp1 substrate (Figure 4B).

Notably, the Pro/N-degrons and several subsequent residues are not visible in the Fbp1 crystal structure, suggesting that they are intrinsically disordered (Figure 4C). These elements could emanate from opposite sides of the disk-like Fbp1 catalytic domain. In the complex with Chelator-GID^{SR4}, degrons from both sides appear to simultaneously ensnare Gid4 substrate receptors. Such avid binding would rationalize the 10-fold lower K_m in Fbp1 ubiquitylation assays (Figure 1C). To further test

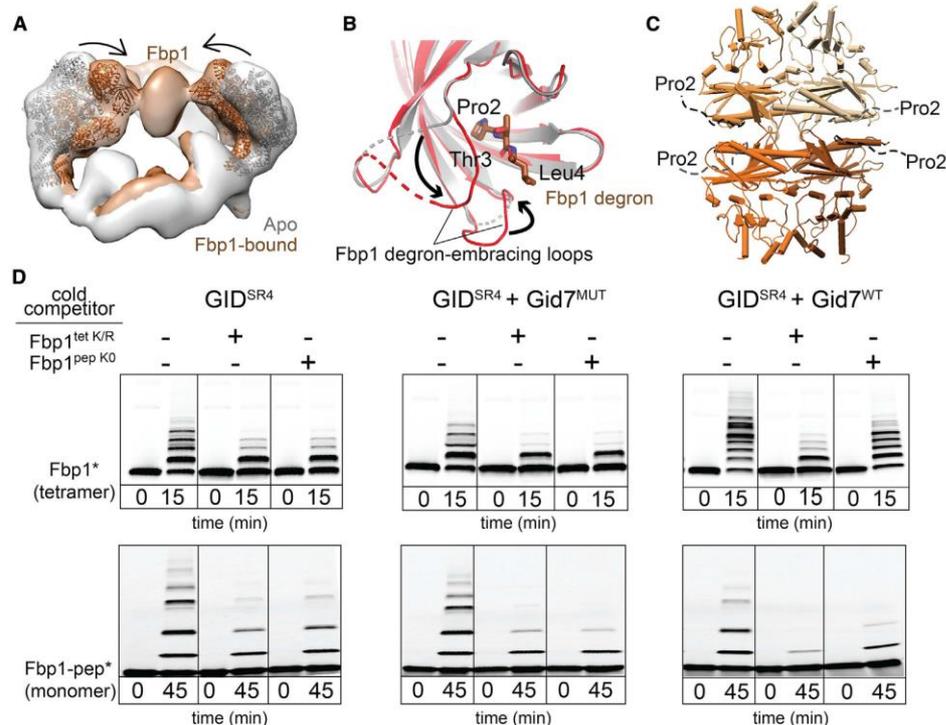


Figure 4. Chelator-GID^{SR4} assembly specifies multivalent binding for the tetrameric Fbp1 substrate

(A) Superimposed maps of substrate-free (gray) and Fbp1-bound Chelator-GID^{SR4} (brown) show relative inward movement of SRS modules (ribbon) upon substrate recruitment.

(B) Conformational differences between Gid4 in GID^{SR4} (PDB: 6SWY, gray) and Fbp1-bound Chelator-GID^{SR4} (red). The first three residues of Fbp1 (the Pro/N-degron) bound to Gid4 are shown as sticks.

(C) Crystal structure of the Fbp1 tetramer, with the N-terminal region (residues 2–19), including the degron not visible in the electron density, depicted as dotted lines. Fbp1 protomers are shown in various brown shades.

(D) Competitive *in vitro* ubiquitylation assays probing multivalent E3-substrate interactions. Chelator-GID^{SR4} has two substrate binding sites and two catalytic centers, whereas two other E3 assemblies (GID^{SR4} or GID^{SR4} + Gid7^{MUT} lacking the Lish-CTLH-CRA motif, Δ1–285) have only one substrate binding site and one catalytic center. Substrates are oligomeric (tetrameric Fbp1) or monomeric (a peptide harboring a single acceptor Lys, Fbp1-pep) and fluorescently labeled at the C terminus (denoted by an asterisk). Competitors are oligomeric (tetrameric Fbp1^{tet K/R}, with preferred target lysines mutated to arginines) or monomeric (lysine-less peptide, Fbp1^{pep K0}).

See also Figure S4 and Tables S1 and S2.

the possibility of avid substrate capture, we performed competitive qualitative ubiquitylation assays. Unlabeled monomeric and tetrameric Fbp1 competitors had a comparable inhibitory effect on ubiquitylation of fluorescent Fbp1 by GID^{SR4} or GID^{SR4} mixed with a Gid7 mutant that does not support supramolecular assembly (Figure 4D). However, compared with an unlabeled monomeric inhibitor, the unlabeled Fbp1 tetramer was strikingly more effective at impeding Chelator-GID^{SR4} ubiquitylation of fluorescent Fbp1. The same inhibitory trends were observed for ubiquitylation of a fluorescent monomeric peptide substrate, confirming that the Fbp1 tetramer complements the Chelator assembly. The data are consistent with avid Fbp1 recruitment to Chelator-GID^{SR4} depending on supramolecular assembly of the E3 ligase and its substrate.

Chelator-GID^{SR4} assembly establishes dual site-specific Ub targeting

We next mapped regions of Fbp1 engaging the ubiquitylation active sites. Locating di-Gly sites by mass spectrometry identified Chelator-GID^{SR4}-mediated ubiquitylation of two pairs of neighboring lysines, K32/K35 and K280/K281, preferentially from 18 potential target lysines on the surface of Fbp1 (Figure S5). The importance of these lysines was confirmed mutationally (Figures 5A and 5B). Use of K0 Ub had shown modification of up to two sites in an Fbp1 protomer during the time course of the experiment (Figure 1B). Eliminating either lysine pair reduced this to monoubiquitylation, with a slightly greater effect on the K32/K35 mutant (Figure 5A). The results suggest that either region can be ubiquitylated independent of the other but that, for

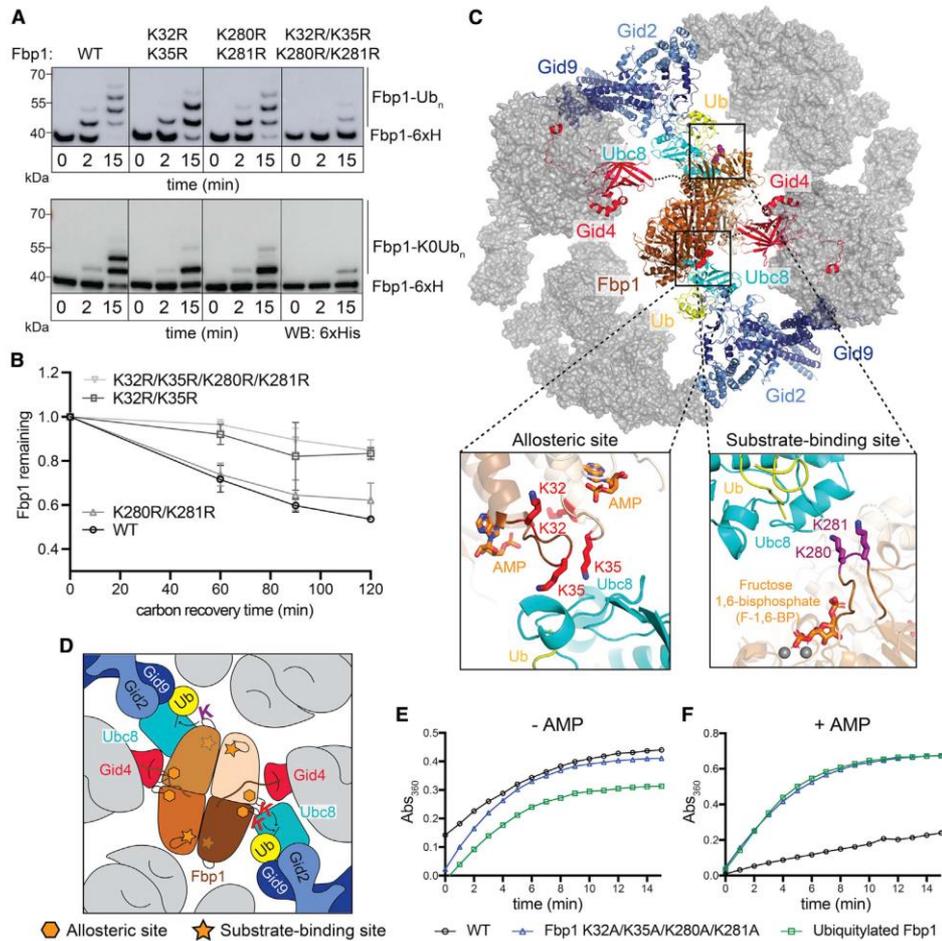


Figure 5. Chelator-GID^{SR4} configures simultaneous targeting of specific lysine clusters in metabolic regulatory regions of the Fbp1 tetramer
 (A) *In vitro* ubiquitylation of Fbp1-6xHis, detected by anti-His immunoblotting, with WT (top) or K0 (bottom) Ub, testing the effects of mutating the major Fbp1 Ub-targeted lysines identified by mass spectrometry.
 (B) Glucose-induced degradation *in vivo* of exogenously expressed WT or lysine mutant versions of Fbp1. Substrate levels were quantified as the ratio of substrate detected relative to the level after switching from carbon starvation to carbon recovery conditions. Points represent mean, and error bars represent SD (n = 3).
 (C) Structural model of Chelator-GID^{SR4}-mediated ubiquitylation of Fbp1. Ubc8-Ub was modeled by aligning a RING-E2-Ub structure (PDB: 5H7S) on Gid2 RING. Dotted lines indicate disordered Fbp1 N termini. Close ups show major Fbp1 ubiquitylation sites near substrate (Fructose-1,6-bisphosphate, F-1,6-BP) and allosteric AMP binding sites modeled from structures with human Fbp1 (PDB: 5ZWK and 5ET6).
 (D) Structure-based cartoon of Fbp1 ubiquitylation as shown in (C). Stars and hexagons represent substrate-binding and allosteric sites in Fbp1, respectively.
 (E) *In vitro* Fbpase activity of purified WT, polyubiquitylated, and mutant Fbp1 (K32A/K35A/K280A/K281A).
 (F) Fbpase activity assay as in (E), testing the responses of purified WT, polyubiquitylated, and mutant Fbp1 (K32A/K35A/K280A/K281A) to the allosteric inhibitor AMP.
 See also Figure S5.

a given protomer, ubiquitylation is restricted to one lysine within a pair. Testing the effects of the mutations on Fbp1 degradation confirmed the importance of these lysines *in vivo*, with substantial stabilization even upon mutating only the K32/K35 lysine pair (Figure 5B).

To understand how the Chelator-GID^{SR4} supramolecular assembly determines regulation, we generated a structural model of ubiquitylation (Figures 5C and 5D). Fbp1 was first anchored via two degrons, one from each side binding a Gid4. Ubc8~Ub was modeled on the Gid2-Gid9 RING-RING-L domains based

on homology to another RING-E2~Ub assembly (Nayak and Sivaraman, 2018). Fbp1 was subjected to constrained rotation to localize the K32 and K35 region of one protomer adjacent to one active site. This led to two striking observations. First, the K32 and K35 regions of two pairs of protomers are adjacent to each other. Second, and unexpectedly, when a K32 and K35 region is aligned with one active site, the K280 and K281 region of a different Fbp1 protomer is simultaneously situated in the other Chelator-GID^{SR4} active site. Thus, the Chelator-GID^{SR4} supramolecular assembly complements the tetrameric structure of Fbp1 by enabling simultaneous capture of two Pro/N degrons and simultaneous ubiquitylation of multiple protomers within the Fbp1 tetramer.

Given that Fbp1 is allosterically regulated in response to metabolite binding (Ke et al., 1990a, 1990b), we inspected the structure for potential functional importance of the ubiquitylation sites (Figures 5C and 5D). Intriguingly, the K32 and K35 residues reside in a loop abutting the allosteric site that regulates Fbp1 activity by binding the non-competitive inhibitor AMP (Ke et al., 1990b). K280 and K281 are located adjacent to another interprotomer interface, relatively near the substrate binding site (Ke et al., 1990a). We thus examined the effects of Chelator-GID^{SR4} ubiquitylation on Fbp1 activity. A K32A/K35A/K280A/K281A mutant and a ubiquitylated version of Fbp1 show Fbpase activity in our assay. However, allosteric modulation by AMP was substantially impaired in both cases (Figures 5E and 5F). Thus, Chelator-GID^{SR4} targets sites related to Fbp1's metabolic function.

Structural and mechanistic parallels in human CTLH E3

To determine whether structural principles governing activity of the yeast GID E3 are conserved in higher eukaryotes, we studied the human CTLH complex, whose subunits mirror those of Chelator-GID^{SR4} (Figure 6A).

We first reconstituted a recombinant complex that we call "CTLH^{SR4}," which parallels yeast GID^{SR4}. A low-resolution cryo-EM envelope showed that the corresponding human subunits form SRS (hGid4-ARMC8-RANBP9-TWA1) and Cat (RMND5A-MAEA) modules (Figure S6A). As for yeast GID^{SR4} (Qiao et al., 2020), the CTLH^{SR4} Cat module is relatively poorly resolved, but the coordinates for the yeast Gid2-Gid9 subcomplex derived from Chelator-GID^{SR4} readily fit in the density. A 3.2-Å-resolution map obtained by focused refinement enabled building of atomic coordinates for the human SRS module, which superimposes on its yeast counterpart (Figures 6B, S6B and S7; Table S1).

We tested whether the structural conservation extended to the enzymatic mechanism. Because the Pro/N-end degron targets of the CTLH E3 remain unknown, we generated a model peptide substrate: an N-terminal PGLW sequence reported previously to optimally bind hGid4 (Dong et al., 2018, 2020), connected via a flexible linker to a C-terminal target lysine. With this peptide substrate, we tested the effects of structure-based point mutations on ubiquitylation. The hGid4 residues mediating its incorporation into CTLH^{SR4} and RMND5A and MAEA residues that activate UBE2H~Ub are crucial for peptide substrate ubiquitylation (Figures S6C–S6H). Moreover, as with GID^{SR4} (Qiao et al., 2020), only K48 of all Ub lysines was sufficient to support polyUb chain

formation by CTLH^{SR4}, albeit to a substantially lesser degree than WT Ub (Figure S6I). Thus, it seems that the human CTLH core module parallels that in yeast GID assemblies.

We examined by cryo-EM whether the human Gid7 orthologs WDR26 and MKLN1 have capacity for supramolecular assembly. We obtained reconstructions for two subcomplexes containing WDR26. Coexpressing WDR26 with scaffolding and catalytic subunits (ARMC8-RANBP9-TWA1-RMND5A-MAEA) yielded a complex broadly resembling Chelator-GID^{SR4} in that it forms a hollow oval of similar dimensions (Figures 6A and 6C). Docking structures of human and yeast subcomplexes into the density showed that a WDR26 dimer is the SA module. However, WDR26 binds directly to RANBP9-TWA1 in the scaffold, without duplicates of these subunits corresponding to yeast Gid1^{SA}-Gid8^{SA}. The distinct WDR26-dependent supramolecular assembly places four—not two—ARMC8 subunits poised to each bind a hGid4 to capture substrate degrons in the CTLH oval.

The distinctive arrangement of SA and SRS modules was preserved in a 6-Å resolution map of WDR26, RANBP9, TWA1, ARMC8, hGid4, and the poorly understood CTLH subunit YPEL5 (Figure 6C; Table S1). Interestingly, YPEL5 binds at the junction of the two protomers in the WDR26 double-propeller domain.

A low-resolution map showed yet another SA for another human Gid7 ortholog, MKLN1, bound to the CTLH SRS module (Figure 6D; Table S1). Like WDR26, MKLN1 binds directly to RANBP9-TWA1 in the scaffold without intervening duplicates of these subunits. However, in accordance with previous studies (Deltó et al., 2015; Kim et al., 2014), MKLN1 forms a tetramer. Four MKLN1 protomers bind between two CTLH SRS modules, demonstrating potential for even higher-order CTLH complex assemblies.

We confirmed roles of WDR26 and MKLN1 in human CTLH complex assembly by sedimentation analyses of lysates from K562 cells or lines in which the human Gid7 orthologs were deleted. Immunoblotting of fractions from sucrose density gradients of parental K562 cell lysates showed comigration of CTLH subunits, corresponding to a complex with a molecular weight greater than that predicted for a uniformly stoichiometric assembly (600–800 kDa according to standards) (Figure 6E). However, probing migration of the core subunit RANBP9 as a marker for the CTLH complex showed that the assembly changes markedly, toward fractions of 150–350 kDa, in CRISPR-Cas9 genome-edited lines lacking WDR26, MKLN1, or both or the Cat module subunit MAEA (Figures 6F and S6J). Interestingly, migration of WDR26 and MKLN1 in higher-molecular-weight fractions is not interdependent (Figure 6G), possibly indicating that each Gid7 ortholog can reside in distinct CTLH assemblies. Much of the total CTLH population shifted to lower-molecular-weight fractions upon deletion of WDR26, with a lesser effect of deleting MKLN1. This may suggest that a greater proportion of the CTLH complex in these cells depends on WDR26 for supramolecular assembly, perhaps because of a higher relative concentration of WDR26 or factors differentially regulating WDR26 or MKLN1 assembly into CTLH complexes.

Overall, the results suggest that CTLH E3 assemblies contain SRS, Cat, and SA modules with features resembling those of Chelator-GID^{SR4}. Moreover, differences in structural configuration of complexes containing MKLN1 or WDR26 offer prospects

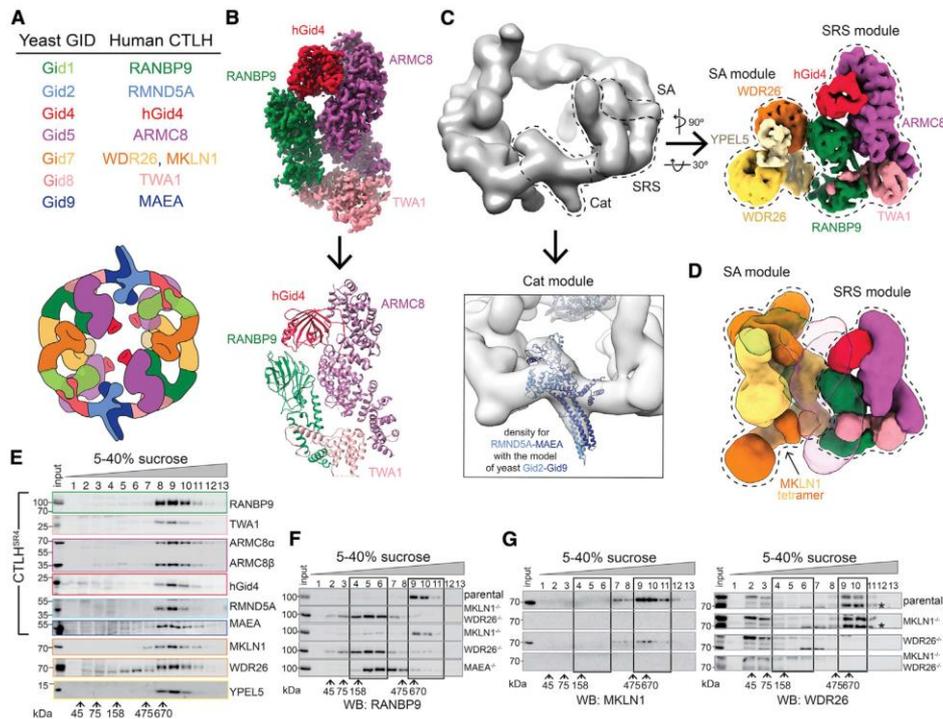


Figure 6. Higher-order assemblies of human CTLH E3

(A) Color-coded guide to yeast GID subunits and their human orthologs in the CTLH complex (top). Two colors indicate multiple protomers of a subunit. Cartoon colored as in the top, representing CTLH oval assembly where the SA module is the WDR26-YPEL5 dimer (bottom).

(B) 3.2-Å-resolution segmented map of CTLH SRS module (RANBP9-TWA1-ARMC8-hGid4) obtained by focused refinement of CTLH^{SR4} (top) and its corresponding model (bottom).

(C) Cryo-EM maps of CTLH assemblies containing the Cat (RMND5A-MAEA), SRS (RANBP9-TWA1-ARMC8 alone or bound to hGid4), and/or supramolecular assembly (WDR26 with or without YPEL5) modules, as indicated. Subunits are colored according to the guide in (A). Top left: low-resolution map of WDR26-mediated SA of CTLH (RANBP9-TWA1-ARMC8-MAEA-RMND5A-WDR26). Right: 6.5-Å-resolution map of the human CTLH SRS module (RANBP9-TWA1-ARMC8-hGid4) subcomplex with an SA module comprising WDR26-YPEL5. Bottom panel: the yeast Gid2-Gid9 structure in the corresponding CTLH Cat module.

(D) 10.4-Å-resolution map of the human CTLH SRS module with MKLN1 as the SA module. The second copy of the SRS module in the subcomplex is transparent.

(E) Immunoblots of fractions from sucrose gradients of K562 cell lysates, probed with the indicated antibodies.

(F) Immunoblots probing for the core CTLH subunit (RANBP9) in fractions from sucrose gradients of lysates from parental K562 and WDR26^{-/-}/MKLN1^{-/-}, MKLN1^{-/-}, WDR26^{-/-}, and MAEA^{-/-} knockout cells. Black boxes delineate high- and low-molecular weight (MW) peak fractions.

(G) As in (F) but probed as indicated with anti-MKLN1 or -WDR26 antibodies. *, WDR26 band.

See also Figures S6 and S7 and Table S1.

that CTLH may adopt a variety of supramolecular E3 assemblies that could impart distinct functionalities.

DISCUSSION

Here we discovered multipronged substrate targeting by an E3 ligase chelate supramolecular assembly tailored to the oligomeric quaternary structure of its metabolic enzyme substrate. In the absence of chelate assembly, GID^{SR4} is a competent E3 ligase that can bind a substrate degron, activate the intrinsic reactivity of its E2 partner (the Ubc8~Ub intermediate), and promote Ub transfer from Ubc8 to a recruited substrate (Qiao et al., 2020). GID^{SR4} is also competent *in vivo* insofar as Gid7

is not required for glucose- and GID-dependent degradation of several substrates (Figure 1). Instead of binding directly to its specified substrate Fbp1, Gid7 alters the GID assembly (Figures 2 and 3).

Although other E3s have been reported to self-assemble (Balaji and Hoppe, 2020), this is typically achieved by catalytic or substrate receptor subunits; for example, the dimeric RING domains of single-subunit E3s or dimeric F-box and BTB substrate receptors in multisubunit cullin-RING ligases (Dou et al., 2012; McMahon et al., 2006; Ogura et al., 2010; Plechanová et al., 2012; Welcker et al., 2013; Zhuang et al., 2009). Substrate-bound multivalent E3s can undergo liquid-liquid phase-separation (Bouchard et al., 2018). However, the transformation into

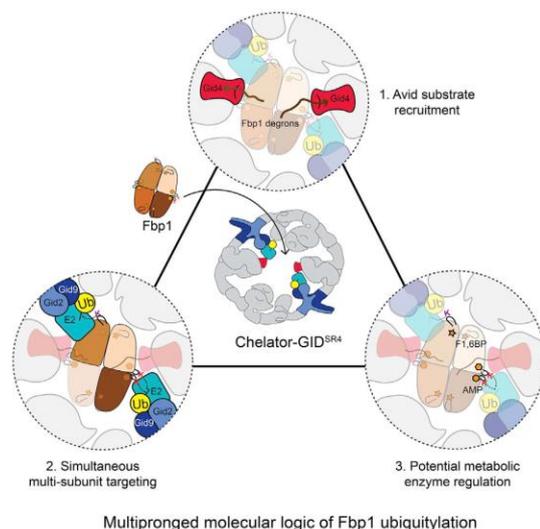


Figure 7. Molecular logic of multipronged Ub targeting of Fbp1 by Chelator-GID^{SR4}

Supramolecular chelate assembly specifies oligomeric metabolic enzyme targeting. (1) Opposing Gid4 subunits avidly bind multiple degrons of tetrameric Fbp1. (2) Opposing RING-E2-Ub active sites simultaneously target specific lysine clusters. (3) Targeted lysines map to metabolically important regions of oligomeric substrate.

Chelator-GID^{SR4} is a distinctive, extreme, and specific adjustment of E3 ligase architecture (Figures 2 and 3).

Resembling an organometallic chelate interacting with its central ligand, Chelator-GID^{SR4}'s multiple distinct points of contact with Fbp1 not only include the degron-binding sites from two opposing Gid4 substrate receptors but also the ubiquitylation active sites from Ubc8~Ub intermediates activated by two opposing Gid2-Gid9 catalytic domains (Figures 4, 5, and 7). Relative to the monodentate GID^{SR4}, the Chelator-GID^{SR4} assembly enables more molecules within the Fbp1 tetramer to be ubiquitylated simultaneously, increasing Ub density on a given Fbp1 tetramer (Figures 1A and 1B). Interestingly, there is not a 1:1 correspondence between the number of degron binding sites in Chelator-GID^{SR4} and the number of degrons in Fbp1. The Fbp1 tetramer has four exposed potential degrons, two on each side, both seemingly poised to capture one central-facing Gid4 in Chelator-GID^{SR4} (Figure 4C). An excess number of degrons is reminiscent of substrates recruited to the cullin-RING ligase receptor Cdc4, whose single binding site can continually and dynamically sample multiple degrons (Mittag et al., 2008). For Chelator-GID^{SR4}-bound Fbp1, we speculate that the arrangement of degrons allows their rapid interchange. This could potentially mediate switching between the protomers positioned adjacent to the active sites.

The human CTLH E3 complex displays striking parallels to Chelator-GID^{SR4}, albeit with interesting twists. In particular, the different Gid7 orthologs form distinct supramolecular assem-

blies (Figure 6). We speculate that the unique assemblies define distinct functions, as implied by varying phenotypic alterations upon their individual mutation (Bauer et al., 2018; Nassan et al., 2017; Skraban et al., 2017; Zhen et al., 2020).

Taken together with previous data (Lampert et al., 2018; Qiao et al., 2020), it is now clear that there is not a single yeast GID or human CTLH complex. Rather, GID and CTLH are examples of responsive systems of multiprotein assemblies with an active E3 core that can be elaborated by supramolecular assembly. Although the function of one such assembly is shown here, the variations revealed by human Gid7 orthologs suggest that they, and presumably other subunits, also co-configure substrate binding and ubiquitylation active sites in accordance with the molecular organization and quaternary structure of particular substrates. The Chelator model presented here demonstrates how GID (and presumably CTLH) utilizes an elegant molecular logic: the response to a signal such as glucose availability converges on numerous aspects of its substrate's structure and function to achieve precise physiological regulation (Figure 7).

Limitations

Chelator-GID^{SR4} is remarkably specific in ubiquitylating particular Fbp1 lysines in metabolic regulatory regions. However, the physiological roles of Fbp1 ubiquitylation impairing allosteric regulation and metabolic function are unknown. Future studies will be required to determine how metabolic flux is coupled with GID-dependent ubiquitylation during termination of gluconeogenesis.

Although Chelator-GID^{SR4} is active toward Mdh2 and Pck1, it is unclear why these oligomeric substrates are less dependent than Fbp1 on Gid7-mediated supramolecular assembly. One speculative possibility could be that any potential advantage of avid binding is offset by accessibility of numerous ubiquitylation sites to GID^{SR4}. Future studies will be required to understand how Pck1 and other GID E3 substrates, including the Gid4 substrate receptor itself, are recognized and ubiquitylated (Hämmerle et al., 1998; Karayel et al., 2020; Menssen et al., 2018).

Finally, although discovery of the Chelator configuration provides a basis for understanding higher-order GID assembly, what other assemblies or sub-assemblies may form and their functions remain unknown. Clearly, other arrangements are observed for human CTLH complexes with WDR26. MKLN1 forms an even higher-order assembly with the human SRS module. Some yeast GID assemblies migrate in the void volume, as seen by size-exclusion chromatography (Figure S2A). Moreover, the mechanistic roles of additional subunits, including YPEL5 (Figure 6), or regulatory partners, such as Cdc48/p97, remain unknown (Barbin et al., 2010; Lampert et al., 2018). We await future studies revealing functions of other variations of GID and CTLH assemblies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability

- **METHOD DETAILS**

- Yeast strain construction and growth conditions
- *In vivo* yeast substrate degradation assays
- Purification of endogenous yeast GID for cryo EM
- Sucrose gradient fractionation of yeast lysates (Figure S2C)
- *In vivo* Fbp1 ubiquitylation assay (Figure 1E)
- Plasmid preparation and Mutagenesis
- Insect cell expression and purification of GID/CTLH complexes
- Bacterial expression and purification
- *In vitro* biochemical assays
- Biochemical assays with yeast GID
- Determination of kinetic parameters of Fbp1 ubiquitylation by GID E3
- Biochemical assays with human CTLH^{SR4}
- SEC for initial characterization of GID supramolecular assembly
- SEC-MALS
- Fbp1 enzyme activity assay
- Analysis of global proteome of WT versus ΔGid7 yeast (Karayel et al., 2020)
- Determination of preferentially targeted lysines in Fbp1 by LC-MS/MS (Figure S5)
- Cell culture and generation of CRISPR-Cas9 knock out cell lines
- Human cell lysate fractionation by sucrose density gradient
- Cryo EM sample preparation and Imaging
- Cryo EM data processing
- Model building and refinement
- Fbp1 crystallization and data processing

- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2021.03.025>.

ACKNOWLEDGMENTS

We thank A. Varshavsky for promoter reference plasmids; S. Übel and S. Petera for peptide synthesis; D. Bollschweiler, T. Schäfer, J. Basquin, K. Valer-Saldana, and S. Pleyer for assistance with cryo-EM and crystallography; Swiss Light Source, Villigen, Switzerland for Fbp1 crystal data collection; M. Strauss for assistance with initial negative-stain data collection on the yeast Chelator-GID complex; M. Yamaguchi for Gid7 constructs; G. Kleiger for guidance regarding kinetics; I. Paron for technical assistance with mass spectrometry; and K. Baek, B. Bräuning, and the Schulman lab for advice and support. The B.A.S. lab is funded by the Leibniz Prize from Deutsche Forschungsgemeinschaft (SCHU 3196/1-1). B.A.S. and M.M. are supported by the Max Planck Society.

AUTHOR CONTRIBUTIONS

Initial conceptualization, D.S., J.C., and B.A.S.; methodology, D.S., J.C., J.R.P., C.R.L., O.K., A.F.A., and B.A.S.; investigation, D.S., J.C., S.Q., C.R.L., L.A.H., K.V.G., F.M.H., O.K., J.R.P., and A.F.A.; resources, D.S.,

J.C., K.V.G., C.R.L., S.v.G., and A.F.A.; writing – original draft, D.S., J.C., and B.A.S.; writing – review & editing, D.S., J.C., C.R.L., A.F.A., and B.A.S.; supervision, M.M., A.F.A., and B.A.S.; funding acquisition, M.M. and B.A.S.

DECLARATION OF INTERESTS

B.A.S. is an honorary professor at Technical University of Munich, Germany and adjunct faculty at St. Jude Children's Research Hospital, Memphis, TN, USA and is on the scientific advisory board of Interline Therapeutics.

Received: October 15, 2020

Revised: February 17, 2021

Accepted: March 17, 2021

Published: April 26, 2021

REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* 74, 814–840.
- Baek, K., Krist, D.T., Prabu, J.R., Hill, S., Klügel, M., Neumaier, L.M., von Gronau, S., Kleiger, G., and Schulman, B.A. (2020). NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly. *Nature* 578, 461–466.
- Balaji, V., and Hoppe, T. (2020). Regulation of E3 ubiquitin ligases by homotypic and heterotypic assembly. *F1000Res.* 9, F1000 Faculty Rev-88.
- Barbin, L., Eisele, F., Santt, O., and Wolf, D.H. (2010). The Cdc48-Ufd1-Npl4 complex is central in ubiquitin-proteasome triggered catabolite degradation of fructose-1,6-bisphosphatase. *Biochem. Biophys. Res. Commun.* 394, 335–341.
- Barford, D. (2020). Structural interconversions of the anaphase-promoting complex/cyclosome (APC/C) regulate cell cycle transitions. *Curr. Opin. Struct. Biol.* 61, 86–97.
- Bauer, A., Jagannathan, V., Högl, S., Richter, B., McEwan, N.A., Thomas, A., Cadieu, E., André, C., Hytönen, M.K., Lohi, H., et al. (2018). MKNL1 splicing defect in dogs with lethal acrodermatitis. *PLoS Genet.* 14, e1007264.
- Biyani, N., Righetto, R.D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K.N., and Stahlberg, H. (2017). Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol.* 198, 124–133.
- Boldt, K., van Reeuwijk, J., Lu, Q., Koutroumpas, K., Nguyen, T.M., Texier, Y., van Beersum, S.E., Horn, N., Willer, J.R., Mans, D.A., et al.; UK10K Rare Diseases Group (2016). An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. *Nat. Commun.* 7, 11491.
- Bouchard, J.J., Otero, J.H., Scott, D.C., Szulc, E., Martin, E.W., Sabri, N., Granata, D., Marzahn, M.R., Lindorff-Larsen, K., Salvatella, X., et al. (2018). Cancer Mutations of the Tumor Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. *Mol. Cell* 72, 19–36.e8.
- Braun, B., Pfirrmann, T., Menssen, R., Hofmann, K., Scheel, H., and Wolf, D.H. (2011). Gid9, a second RING finger protein contributes to the ubiquitin ligase activity of the Gid complex required for catabolite degradation. *FEBS Lett.* 585, 3856–3861.
- Brown, N.G., VanderLinden, R., Watson, E.R., Weissmann, F., Oudureau, A., Wu, K.P., Zhang, W., Yu, S., Mercredi, P.Y., Harrison, J.S., et al. (2016). Dual RING E3 Architectures Regulate Multiubiquitination and Ubiquitin Chain Elongation by APC/C. *Cell* 165, 1440–1453.
- Budhidarmo, R., Nakatani, Y., and Day, C.L. (2012). RINGS hold the key to ubiquitin transfer. *Trends Biochem. Sci.* 37, 58–65.
- Burnley, T., Palmer, C.M., and Winn, M. (2017). Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.* 73, 469–477.

- Cao, W.X., Kabelitz, S., Gupta, M., Yeung, E., Lin, S., Rammelt, C., Ihling, C., Pekovic, F., Low, T.C.H., Siddiqui, N.U., et al. (2020). Precise Temporal Regulation of Post-transcriptional Repressors Is Required for an Orderly *Drosophila* Maternal-to-Zygotic Transition. *Cell Rep.* **31**, 107783.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.
- Chen, C.K., Chan, N.L., and Wang, A.H. (2011a). The many blades of the β -propeller proteins: conserved but versatile. *Trends Biochem. Sci.* **36**, 553–561.
- Chen, I., Dorr, B.M., and Liu, D.R. (2011b). A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. USA* **108**, 11399–11404.
- Chen, S., Wu, J., Lu, Y., Ma, Y.B., Lee, B.H., Yu, Z., Ouyang, Q., Finley, D.J., Kirschner, M.W., and Mao, Y. (2016). Structural basis for dynamic regulation of the human 26S proteasome. *Proc. Natl. Acad. Sci. USA* **113**, 12991–12996.
- Chen, S.J., Wu, X., Wadas, B., Oh, J.H., and Varshavsky, A. (2017). An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355**, eaal3655.
- Chiang, H.L., and Schekman, R. (1991). Regulated import and degradation of a cytosolic protein in the yeast vacuole. *Nature* **350**, 313–318.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011.
- Delto, C.F., Heisler, F.F., Kuper, J., Sander, B., Kneussel, M., and Schindelin, H. (2015). The LisH motif of muskelin is crucial for oligomerization and governs intracellular localization. *Structure* **23**, 364–373.
- DiMaio, F., Echols, N., Headd, J.J., Terwilliger, T.C., Adams, P.D., and Baker, D. (2013). Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **10**, 1102–1104.
- Dong, C., Zhang, H., Li, L., Tempel, W., Loppnau, P., and Min, J. (2018). Molecular basis of GID4-mediated recognition of degrons for the Pro/N-end rule pathway. *Nat. Chem. Biol.* **14**, 466–473.
- Dong, C., Chen, S.J., Melnykov, A., Weirich, S., Sun, K., Jeltsch, A., Varshavsky, A., and Min, J. (2020). Recognition of nonproline N-terminal residues by the Pro/N-degron pathway. *Proc. Natl. Acad. Sci. USA* **117**, 14158–14167.
- Dou, H., Buetow, L., Sibbet, G.J., Cameron, K., and Huang, D.T. (2012). BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* **19**, 876–883.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
- Fernandez-Leiro, R., and Scheres, S.H.W. (2017). A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. D Struct. Biol.* **73**, 496–502.
- Francis, O., Han, F., and Adams, J.C. (2013). Molecular phylogeny of a RING E3 ubiquitin ligase, conserved in eukaryotic cells and dominated by homologous components, the muskelin/RanBPM/CTLH complex. *PLoS ONE* **8**, e75217.
- Gancedo, C. (1971). Inactivation of fructose-1,6-diphosphatase by glucose in yeast. *J. Bacteriol.* **107**, 401–405.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25.
- Hämmerle, M., Bauer, J., Rose, M., Szallies, A., Thumm, M., Dusterhus, S., Mecke, D., Entian, K.D., and Wolf, D.H. (1998). Proteins of newly isolated mutants and the amino-terminal proline are essential for ubiquitin-proteasome-catalyzed catabolite degradation of fructose-1,6-bisphosphatase of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 25000–25005.
- Haselbach, D., Schrader, J., Lambrecht, F., Henneberg, F., Chari, A., and Stark, H. (2017). Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat. Commun.* **8**, 15578.
- Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**(11).
- Kaiser, S.E., Riley, B.E., Shaler, T.A., Trevino, R.S., Becker, C.H., Schulman, H., and Kopito, R.R. (2011). Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat. Methods* **8**, 691–696.
- Karayel, O., Michaelis, A.C., Mann, M., Schulman, B.A., and Langlois, C.R. (2020). DIA-based systems biology approach unveils E3 ubiquitin ligase-dependent responses to a metabolic shift. *Proc. Natl. Acad. Sci. USA* **117**, 32806–32815.
- Ke, H.M., Thorpe, C.M., Seaton, B., Lipscomb, W.N., and Marcus, F. (1990a). Structure refinement of fructose-1,6-bisphosphatase and its fructose 2,6-bisphosphate complex at 2.8 Å resolution. *J. Mol. Biol.* **212**, 513–539.
- Ke, H.M., Zhang, Y.P., and Lipscomb, W.N. (1990b). Crystal structure of fructose-1,6-bisphosphatase complexed with fructose 6-phosphate, AMP, and magnesium. *Proc. Natl. Acad. Sci. USA* **87**, 5243–5247.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858.
- Kim, K.H., Hong, S.K., Hwang, K.Y., and Kim, E.E. (2014). Structure of mouse muskelin discoidin domain and biochemical characterization of its self-association. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2863–2874.
- Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**(10B).
- Kobayashi, N., Yang, J., Ueda, A., Suzuki, T., Tomaru, K., Takeno, M., Okuda, K., and Ishigatsubo, Y. (2007). RanBPM, Muskelein, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* **396**, 236–247.
- Koshland, D.E., Jr. (1963a). Correlation of Structure and Function in Enzyme Action. *Science* **142**, 1533–1541.
- Koshland, D.E., Jr. (1963b). Properties of the active site of enzymes. *Ann. N.Y. Acad. Sci.* **103**, 630–642.
- Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324.
- Lampert, F., Stafa, D., Goga, A., Soste, M.V., Gilberto, S., Olieric, N., Picotti, P., Stoffel, M., and Peter, M. (2018). The multi-subunit GID/CTLH E3 ubiquitin ligase promotes cell proliferation and targets the transcription factor Hbp1 for degradation. *eLife* **7**, e35528.
- Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**, 186–191.
- Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H., and Finley, D. (2002). Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**, 495–507.
- Liu, H., and Pfirrmann, T. (2019). The Gid-complex: an emerging player in the ubiquitin ligase league. *Biol. Chem.* **400**, 1429–1441.
- Liu, H., Ding, J., Köhnlein, K., Urban, N., Ori, A., Villavicencio-Lorini, P., Walentek, P., Klotz, L.O., Hollemann, T., and Pfirrmann, T. (2020). The GID ubiquitin ligase complex is a regulator of AMPK activity and organismal lifespan. *Autophagy* **16**, 1618–1634.

- Maitland, M.E.R., Onea, G., Chiasson, C.A., Wang, X., Ma, J., Moor, S.E., Barber, K.R., Lajoie, G.A., Shaw, G.S., and Schild-Poulter, C. (2019). The mammalian CTLH complex is an E3 ubiquitin ligase that targets its subunit muskulin for degradation. *Sci. Rep.* **9**, 9864.
- Mastrorade, D. (2003). SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microsc. Microanal.* **9**, 1182–1183.
- McMahon, M., Thomas, N., Itoh, K., Yamamoto, M., and Hayes, J.D. (2006). Dimerization of substrate adaptors can facilitate cullin-mediated ubiquitylation of proteins by a “tethering” mechanism: a two-site interaction model for the Nrf2-Keap1 complex. *J. Biol. Chem.* **281**, 24756–24768.
- Melnykov, A., Chen, S.J., and Varshavsky, A. (2019). Gid10 as an alternative N-recognin of the Pro/N-degron pathway. *Proc. Natl. Acad. Sci. USA* **116**, 15914–15923.
- Menssen, R., Schweiggert, J., Schreiner, J., Kusevic, D., Reuther, J., Braun, B., and Wolf, D.H. (2012). Exploring the topology of the Gid complex, the E3 ubiquitin ligase involved in catabolite-induced degradation of gluconeogenic enzymes. *J. Biol. Chem.* **287**, 25602–25614.
- Menssen, R., Bui, K., and Wolf, D.H. (2018). Regulation of the Gid ubiquitin ligase recognition subunit Gid4. *FEBS Lett.* **592**, 3286–3294.
- Mittag, T., Orlicky, S., Choy, W.Y., Tang, X., Lin, H., Sicheri, F., Kay, L.E., Tyers, M., and Forman-Kay, J.D. (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. USA* **105**, 17772–17777.
- Monod, J., Changeux, J.P., and Jacob, F. (1963). Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329.
- Nakatsukasa, K., Okumura, F., and Kamura, T. (2015). Proteolytic regulation of metabolic enzymes by E3 ubiquitin ligase complexes: lessons from yeast. *Crit. Rev. Biochem. Mol. Biol.* **50**, 489–502.
- Nassan, M., Li, Q., Croarkin, P.E., Chen, W., Colby, C.L., Veldic, M., McElroy, S.L., Jenkins, G.D., Ryu, E., Cunningham, J.M., et al. (2017). A genome wide association study suggests the association of muskulin with early onset bipolar disorder: Implications for a GABAergic epileptogenic neurogenesis model. *J. Affect. Disord.* **208**, 120–129.
- Nayak, D., and Sivaraman, J. (2018). Structure of LNX1:Ubc13~Ubiquitin Complex Reveals the Role of Additional Motifs for the E3 Ligase Activity of LNX1. *J. Mol. Biol.* **430**, 1173–1188.
- Ogura, T., Tong, K.I., Mio, K., Maruyama, Y., Kurokawa, H., Sato, C., and Yamamoto, M. (2010). Keap1 is a forked-stem dimer structure with two large spheres enclosing the intervening, double glycine repeat, and C-terminal domains. *Proc. Natl. Acad. Sci. USA* **107**, 2842–2847.
- Oh, J.H., Chen, S.J., and Varshavsky, A. (2017). A reference-based protein degradation assay without global translation inhibitors. *J. Biol. Chem.* **292**, 21457–21465.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
- Plechanovová, A., Jaffray, E.G., Tatham, M.H., Naismith, J.H., and Hay, R.T. (2012). Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120.
- Pruneda, J.N., Littlefield, P.J., Soss, S.E., Nordquist, K.A., Chazin, W.J., Brzovic, P.S., and Klevit, R.E. (2012). Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942.
- Qiao, S., Langlois, C.R., Chrustowicz, J., Sherpa, D., Karayel, O., Hansen, F.M., Beier, V., von Gronau, S., Bollschweiler, D., Schäfer, T., et al. (2020). Interconversion between Anticipatory and Active GID E3 Ubiquitin Ligase Conformations via Metabolically Driven Substrate Receptor Assembly. *Mol. Cell* **77**, 150–163.e9.
- Regelmann, J., Schüle, T., Josupeit, F.S., Horak, J., Rose, M., Entian, K.D., Thumm, M., and Wolf, D.H. (2003). Catabolite degradation of fructose-1,6-bisphosphatase in the yeast *Saccharomyces cerevisiae*: a genome-wide screen identifies eight novel GID genes and indicates the existence of two degradation pathways. *Mol. Biol. Cell* **14**, 1652–1663.
- Rusnac, D.V., and Zheng, N. (2020). Structural Biology of CRL Ubiquitin Ligases. *Adv. Exp. Med. Biol.* **1217**, 9–31.
- Salemi, L.M., Maitland, M.E.R., McTavish, C.J., and Schild-Poulter, C. (2017). Cell signalling pathway regulation by RanBPM: molecular insights and disease implications. *Open Biol.* **7**, 170081.
- Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J., Sorzano, C., and Vargas, J. (2020). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv*. <https://doi.org/10.1101/2020.06.12.148296>.
- Santt, O., Pfirrmann, T., Braun, B., Juretschke, J., Kimmig, P., Scheel, H., Hofmann, K., Thumm, M., and Wolf, D.H. (2008). The yeast GID complex, a novel ubiquitin ligase (E3) involved in the regulation of carbohydrate metabolism. *Mol. Biol. Cell* **19**, 3323–3333.
- Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.
- Schor, S.M., Bee, G., Thumm, M., and Wolf, D.H. (1994a). Catabolite inactivation of fructose-1,6-bisphosphatase in yeast is mediated by the proteasome. *FEBS Lett.* **349**, 270–274.
- Schor, S.M., Bee, G., Thumm, M., and Wolf, D.H. (1994b). Site of catabolite inactivation. *Nature* **369**, 283–284.
- Schor, S.M., Thumm, M., and Wolf, D.H. (1995). Catabolite inactivation of fructose-1,6-bisphosphatase of *Saccharomyces cerevisiae*. Degradation occurs via the ubiquitin pathway. *J. Biol. Chem.* **270**, 26446–26450.
- Schweitzer, A., Aufderheide, A., Rudack, T., Beck, F., Pfeifer, G., Piltzko, J.M., Sakata, E., Schulten, K., Förster, F., and Baumeister, W. (2016). Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc. Natl. Acad. Sci. USA* **113**, 7816–7821.
- Shakeel, S., Rajendra, E., Alcón, P., O'Reilly, F., Chorev, D.S., Maslen, S., Degliesposti, G., Russo, C.J., He, S., Hill, C.H., et al. (2019). Structure of the Fanconi anaemia monoubiquitin ligase complex. *Nature* **575**, 234–237.
- Skraban, C.M., Wells, C.F., Markose, P., Cho, M.T., Nesbitt, A.I., Au, P.Y.B., Begtrup, A., Bernat, J.A., Bird, L.M., Cao, K., et al. (2017). WDR26 Haploinsufficiency Causes a Recognizable Syndrome of Intellectual Disability, Seizures, Abnormal Gait, and Distinctive Facial Features. *Am. J. Hum. Genet.* **101**, 139–148.
- Storici, F., and Resnick, M.A. (2006). The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymology* **409**.
- Tu, B.P., and McKnight, S.L. (2006). Metabolic cycles as an underlying basis of biological oscillations. *Nat. Rev. Mol. Cell Biol.* **7**, 696–701.
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46** (W1), W296–W303.
- Watson, E.R., Brown, N.G., Peters, J.M., Stark, H., and Schulman, B.A. (2019). Posing the APC/C E3 Ubiquitin Ligase to Orchestrate Cell Division. *Trends Cell Biol.* **29**, 117–134.
- Wehmer, M., Rudack, T., Beck, F., Aufderheide, A., Pfeifer, G., Piltzko, J.M., Förster, F., Schulten, K., Baumeister, W., and Sakata, E. (2017). Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc. Natl. Acad. Sci. USA* **114**, 1305–1310.
- Weissmann, F., Petzold, G., VanderLinden, R., Huis In 't Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A., and Peters, J.M. (2016). biGBac enables rapid gene assembly for the expression of

Molecular Cell
Article

large multisubunit protein complexes. *Proc. Natl. Acad. Sci. USA* **113**, E2564–E2569.

Welcker, M., Larimore, E.A., Swanger, J., Bengoechea-Alonso, M.T., Grim, J.E., Ericsson, J., Zheng, N., and Clurman, B.E. (2013). Fbw7 dimerization determines the specificity and robustness of substrate degradation. *Genes Dev.* **27**, 2531–2536.

Zaman, S., Lippman, S.I., Zhao, X., and Broach, J.R. (2008). How *Saccharomyces* responds to nutrients. *Annu. Rev. Genet.* **42**, 27–81.

Zavortink, M., Rutt, L.N., Dzitoyeva, S., Henriksen, J.C., Barrington, C., Bilodeau, D.Y., Wang, M., Chen, X.X.L., and Rissland, O.S. (2020). The E2 Marie Kondo and the CTLH E3 ligase clear deposited RNA binding proteins during the maternal-to-zygotic transition. *eLife* **9**, e53889.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12.

Zhen, R., Moo, C., Zhao, Z., Chen, M., Feng, H., Zheng, X., Zhang, L., Shi, J., and Chen, C. (2020). Wdr26 regulates nuclear condensation in developing erythroblasts. *Blood* **135**, 208–219.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332.

Zhu, J., and Thompson, C.B. (2019). Metabolic regulation of cell growth and proliferation. *Nat. Rev. Mol. Cell Biol.* **20**, 436–450.

Zhuang, M., Calabrese, M.F., Liu, J., Waddell, M.B., Nourse, A., Hammel, M., Miller, D.J., Walden, H., Duda, D.M., Seyedin, S.N., et al. (2009). Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases. *Mol. Cell* **36**, 39–50.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal ANTI-FLAG M2 antibody	Sigma Aldrich	Cat# F1804; RRID: AB_262044
Anti-HA antibody produced in rabbit	Sigma Aldrich	Cat# H6908; RRID: AB_260070
Goat anti-rabbit IgG Dylight488 conjugated	Invitrogen	Cat# 35552; RRID: AB_844398
Goat anti-mouse IgG Dylight633 conjugated	Invitrogen	Cat# 35512; RRID: AB_1307538
Anti-rabbit peroxidase antibody produced in goat	Sigma Aldrich	Cat# A9169; RRID: AB_258434
Anti-mouse IgG Peroxidase antibody produced in goat	Sigma Aldrich	Cat# A4416; RRID: AB_258167
Anti-His antibody produced in mouse	Cell Signaling Technology	Cat# 9991; RRID: AB_2797714
Goat polyclonal anti-RMND5A antibody	Santa Cruz	Cat# sc-161202; RRID: AB_2181510
Sheep polyclonal anti-MAEA antibody	R&D Systems	Cat# AF7288-SP; RRID: AB_10971438
Rabbit polyclonal anti-RANBP9 antibody	Abnova	Cat# PAB16671; RRID: AB_10677213
Rabbit polyclonal anti-TWA1 antibody	Novus	Cat# NBP1-32596; RRID: AB_2274921
Mouse monoclonal anti-ARMC8 antibody	Santa Cruz	Cat# sc-365307; RRID: AB_10850172
Mouse monoclonal anti-MKLN1 antibody	Santa Cruz	Cat# sc-398956; RRID: AB_2737249
Rabbit polyclonal anti-WDR26 antibody	Bethyl Laboratories	Cat# A302-245A; RRID: AB_1730876
Rabbit polyclonal anti-YPEL5 antibody	Thermo Fisher	Cat# PA5-26957; RRID: AB_2544457
Sheep polyclonal anti-hGid4	This study	N/A
HaloLink Resin	Promega	Cat# G1912
ANTI-FLAG M2 affinity gel	Sigma Aldrich	Cat# A2220
His-Select Nickel affinity gel	Sigma Aldrich	Cat# P6611
Glutathione Sepharose 4B	GE Healthcare	Cat# 17075605
StrepTactin Sepharose High Performance resin	cytiva	Cat# 28935599
His-Halo UBA ^{UBOLN1}	This study	N/A
Critical commercial assays		
EnzChek Phosphate Assay Kit	ThermoFisher Scientific	Cat# E6646
Bacterial and virus strains		
<i>E. coli</i> BL21 RIL (DE3)	MPIB	N/A
<i>E. coli</i> DH5 α	MPIB	N/A
Chemicals, peptides, and recombinant proteins		
complete EDTA free	Roche	Cat# 05056489001
Aprotinin from bovine lung	Sigma	A1153-10MG
Leupeptin	Sigma	L2884-250MG
Benzamidine	Sigma	B6506-25G
GGGGGFYVK-FAM	MPIB	N/A
PTLVNGWPR	MPIB	N/A
PTLVNGPRRDSTEGFTGRGWSGRGWS	MPIB	N/A
KGGK-FAM	MPIB	N/A
PGLWRSPRRDSTEGFTGRGWSGRG	MPIB	N/A
WSKGGK-FAM	MPIB	N/A
3xFLAG peptide	MPIB	N/A

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Continued		
Deposited data		
Apo Chelator-GID ^{SR4}	This study	EMDB: EMD-12541
Chelator-GID ^{SR4} + Fbp1	This study	EMDB: EMD-12557
GID ^{SR4}	This study	EMDB: EMD-12548
SRS module	This study	EMDB: EMD-12559; PDB: 7NS3
Cat module	This study	EMDB: EMD-12560; PDB: 7NS4
SA module	This study	EMDB: EMD-12563; PDB: 7NSB
Endogenous GID ^{Ant}	This study	EMDB: EMD-12538
Endogenous Chelator-GID ^{Ant}	This study	EMDB: EMD-12540
CTLH ^{SR4}	This study	EMDB: EMD-12537
CTLH ^{SR4} SRS module	This study	EMDB: EMD-12564; PDB: 7NSC
CTLH-WDR26 SA and SRS modules	This study	EMDB: EMD-12545
CTLH-WDR26 supramolecular assembly	This study	EMDB: EMD-12542
CTLH-MKLN1 SA and SRS modules	This study	EMDB: EMD-12547
Fbp1 (crystal structure)	This study	PDB: 7NS5
Proteomics data	This study	PXD024462
Raw image data	This study	http://dx.doi.org/10.17632/rfpsg6939c.1
Experimental models: Cell lines		
Sf9 Insect cells	Thermo Fisher	Cat# 11496015
High Five Insect cells	Thermo Fisher	Cat# B85502
K562 human cells	ATCC	ATCC#CCL-243; RRID: CVCL_00004
Experimental models: Organisms/strains		
<i>Saccharomyces cerevisiae</i> : Strain S288C: BY4741; MATa his3Δ1leu2Δ0 met15Δ0 ura3Δ0	Euroscarf	Cat# Y00000
CRLY12; BY4741, Gid4::KANMX	This study	N/A
CRLY14; BY4741, Gid7::KANMX	This study	N/A
CRLY45; BY4741, Gid8::Gid8-3xFLAG-KANMX	Qiao et al., 2020	N/A
CRLY131; BY4741, Gid2::3xFLAG-Gid2 (K365A)	Qiao et al., 2020	N/A
CRLY241; BY4741, Gid7::KANMX, Gid8::Gid8-3xFLAG-HPHNT1	This study	N/A
CRLY267; BY4741, Gid7::Gid7-3xHA-HPHNT1, Gid5::Gid5-3xFLAG-KANMX	This study	N/A
CRLY498; BY4741, Fbp1::Fbp1-3xFLAG-HPHNT1, Pdr5::NATNT2	This study	N/A
CRLY504; BY4741, Fbp1::Fbp1-3xFLAG-HPHNT1, Pdr5::NATNT2, Gid7::KANMX	This study	N/A
Recombinant DNA		
pCSJ95	Chen et al., 2017	N/A
pCSJ125	Chen et al., 2017	N/A
VBP6; pRS313-pGPD-Pck1-3xFLAG-CYC-pGPD-DHFR-HA-CYC	This study	N/A
DSJC1; pRS313-pGPD-Fbp1-K32R/K35R-3xFLAG-CYC-pGPD-DHFR-HA-CYC	This study	N/A
DSJC2; pRS313-pGPD-Fbp1-K280R/K281R-3xFLAG-CYC-pGPD-DHFR-HA-CYC	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DSJC3; pRS313-pGPD-Fbp1-K32R/K35R/K280R/K281R-3xFLAG-CYC-pGPD-DHFR-HA-CYC	This study	N/A
pRS415-pTEF-CYC	This study	N/A
pRS415-pTEF-GFP-ScGid4-CYC	This study	N/A
pLIB Gid1	This study	N/A
pLIB Gid2	This study	N/A
pLIB Gid4	This study	N/A
pLIB Gid5	This study	N/A
pLIB Gid7	This study	N/A
pLIB Gid8-TEV-2xStrep	This study	N/A
pLIB Gid8	This study	N/A
pLIB Gid9	This study	N/A
pLIB RANBP9	This study	N/A
pLIB RMND5A	This study	N/A
pLIB hGid4	This study	N/A
pLIB ARMC8	This study	N/A
pLIB 2xStrep-3C-ARMC8	This study	N/A
pLIB MAEA	This study	N/A
pLIB WDR26	This study	N/A
pLIB GST-TEV-WDR26	This study	N/A
pLIB MKLN1	This study	N/A
pLIB YPEL5	This study	N/A
pLIB MAEA Y394A	This study	N/A
pLIB RMND5A R340A	This study	N/A
pLIB RMND5A I338A/L339A	This study	N/A
pLIB GST-TEV-Uba1	This study	N/A
pBIG2 Gid1:Gid8-TEV-2xS:Gid5:Gid4:Gid2:Gid9	This study	N/A
pBIG2 Gid1:Gid8-TEV-2xS:Gid5:Gid2:Gid9	This study	N/A
pBIG2 Gid1:Gid8-TEV-2xS:Gid5:Gid2:Gid9:Gid7	This study	N/A
pBIG2 Gid1:Gid8-TEV-2xS:Gid5:Gid4:Gid2:Gid9:Gid7	This study	N/A
pBIG1 RANBP9:TWA1-TEV-2xS:ARMC8	This study	N/A
pBIG1 RANBP9:TWA1:2xS-3C-ARMC8	This study	N/A
pBIG2 RANBP9:TWA1-TEV-2xS:ARMC8:RMND5A:MAEA	This study	N/A
pBIG2 RANBP9:TWA1-TEV-2xS:ARMC8:RMND5A	This study	N/A
pBIG2 RANBP9:TWA1-TEV-2xS:ARMC8:MAEA	This study	N/A
pGEX GST-TEV-Gid4 (Δ 1-116)	This study	N/A
pGEX GST-TEV-Gid7	This study	N/A
pGEX GST-TEV-Gid7 (Δ 1-285)	This study	N/A
pGEX GST-TEV-hGid4 (Δ 1-99)	This study	N/A
pGEX GST-TEV-hGid4 (R189A)	This study	N/A
pGEX GST-TEV-hGid4 (Y154A)	This study	N/A
pGEX GST-TEV-hGid4 (F174A)	This study	N/A
pGEX GST-TEV-hGid4 (F239A)	This study	N/A
pGEX GST-TEV-hGid4 (C156D)	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pGEX GST-TEV-hGid4 (E298A)	This study	N/A
pGEX GST-TEV-hGid4 (H147D)	This study	N/A
pGEX GST-TEV-hGid4 (Y158A/F174A)	This study	N/A
pGEX GST-TEV-hGid4 (Y158A/F239A)	This study	N/A
pGEX GST-TEV-hGid4 (F174A/F239A)	This study	N/A
pGEX GST-TEV-hGid4 (Y297A/F229A)	This study	N/A
pGEX GST-TEV-hGid4 (Δ 297-300)	This study	N/A
pRSF Ubc8-6xHis	This study	N/A
pRSF Fbp1-6xHis	This study	N/A
pRSF Fbp1 (K32R/K35R)-6xHis	This study	N/A
pRSF Fbp1 (K280R/K281R)-6xHis	This study	N/A
pRSF Fbp1 (K32R/K35R/K280R/K281R)-6xHis	This study	N/A
pRSF Fbp1-GGGGS-sortag-6xHis	This study	N/A
pRSF Mdh2-GGGGS-sortag-6xHis	This study	N/A
pRSF Pck1-GGGGS-sortag-6xHis	This study	N/A
pQlink Fbp1-TEV-V5-2xS	This study	N/A
pRSF Ube2H-6xHis	This study	N/A
pGEX GST-3C-Ub	This study	N/A
pGEX GST-3C-Ub K0 (all K > R)	This study	N/A
pGEX GST-3C-Ub K6 (all K > R; R6K)	This study	N/A
pGEX GST-3C-Ub K11 (all K > R; R11K)	This study	N/A
pGEX GST-3C-Ub K27 (all K > R; R27K)	This study	N/A
pGEX GST-3C-Ub K29 (all K > R; R29K)	This study	N/A
pGEX GST-3C-Ub K33 (all K > R; R33K)	This study	N/A
pGEX GST-3C-Ub K48 (all K > R; R48K)	This study	N/A
pGEX GST-3C-Ub K63 (all K > R; R63K)	This study	N/A
pET3b Ub	This study	N/A
pET29 sortase A	Chen et al., 2011b	N/A
Software and algorithms		
FOCUS	Biyani et al., 2017	https://focus.c-cina.unibas.ch/documentation.php
SerialEM	Mastronarde, 2003	https://bio3d.colorado.edu/SerialEM/
MOTIONCOR2	Zheng et al., 2017	https://emcore.ucsf.edu/ucsf-software
Gctf	Zhang, 2016	https://www2.mrc-lmb.cam.ac.uk/download/gctf/
Gautomatch	Kai Zhang	https://www2.mrc-lmb.cam.ac.uk/download/gautomatch-053/
Relion3.0/3.1	Fernandez-Leiro and Scheres, 2017; Scheres, 2012; Zivanov et al., 2018	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
Phyre ²	Kelley et al., 2015	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
SWISS-MODEL	Waterhouse et al., 2018	https://swissmodel.expasy.org
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX	Goddard et al., 2018	https://www.rbvi.ucsf.edu/chimerax/
PyMOL v2.1	Schrödinger	https://pymol.org/2/
CCP-EM	Burnley et al., 2017	https://www.ccpem.ac.uk/download.php
Buccaneer	Cowtan, 2006	http://www.yesbl.york.ac.uk/~cowtan/buccaneer/buccaneer.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Coot	Emsley and Cowtan, 2004; Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix	Adams et al., 2010; Afonine et al., 2018; DiMaio et al., 2013	https://www.phenix-online.org/
Molprobrity	Chen et al., 2010	http://molprobrity.biochem.duke.edu/
Image Studio	LI-COR Biosciences	https://www.licor.com/bio/image-studio/
Fiji/ImageJ	Schindelin et al., 2012	https://imagej.net/Welcome
GraphPad Prism version 8.0	GraphPad Software	http://www.graphpad.com:443/
ImageQuant TL Toolbox version 8.2	Cytiva (formerly GE Healthcare)	https://www.cytivalifesciences.com
DeepEMhancer	Sanchez-Garcia et al., 2020	http://www.biorxiv.org
Other		
QUANTIFOIL® R1.2/1.3, 100 Holey Carbon Films, Grids: Cu 200 mesh	Quantifoil Micro Tools GmbH	https://www.quantifoil.com
IMEM	Thermo Fisher	Cat# 12440-053

RESOURCE AVAILABILITY**Lead contact**

Information and requests for resources and reagents should be directed to the Lead Contact, Prof. Dr. Brenda Schulman (schulman@biochem.mpg.de).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

The accession codes for the PDB models and EM maps are available in RCSB and EMDB, respectively, as follows: Apo Chelator-GID^{SR4}, EMDB: EMD-12541; Chelator-GID^{SR4} + Fbp1, EMDB: EMD-12557; GID^{SR4}, EMDB: EMD-12548; SRS module, EMDB: EMD-12559; PDB: 7NS3; Cat module, EMDB: EMD-12560; PDB: 7NS4; SA module, EMDB: EMD-12563; PDB: 7NSB; Endogenous GID^{Ant}, EMDB: EMD-12538; Endogenous Chelator-GID^{Ant}, EMDB: EMD-12540; CTLH^{SR4}, EMDB: EMD-12537; CTLH^{SR4} SRS module, EMDB: EMD-12564; PDB: 7NSC; CTLH-WDR26 supramolecular assembly, EMDB: EMD-12542; CTLH-WDR26 SA and SRS modules, EMDB: EMD-12545; CTLH-MKLN1 SA and SRS modules, EMDB: EMD-12547; Fbp1 crystal structure, PDB: 7NS5.

All proteomics data have been deposited on ProteomeXchange with the dataset identifier PRIDE: PXD024462.

All the unprocessed image data have been deposited to Mendeley Data : <http://dx.doi.org/10.17632/rfpg6939c.1>

METHOD DETAILS**Yeast strain construction and growth conditions**

The yeast strains used in this study are specified in the Key Resources Table. They were constructed as derivatives of BY4741 using standard genetic techniques (Janke et al., 2004; Knop et al., 1999; Storici and Resnick, 2006) and were verified using PCR, DNA sequencing and immunoblotting (to confirm protein expression). Unless stated otherwise, yeast strains were grown to OD₆₀₀ of 1.0 in synthetic dropout (SD-glucose; 0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose, amino acid mix) or yeast peptone-based medium (YPD; 1% yeast extract, 2% peptone, 2% glucose) as indicated in the respective assays.

In vivo yeast substrate degradation assays

Degradation assays were performed to test the dependency of Fbp1, Mdh2, and Pck1 degradation on Gid4 and Gid7 (Figure 1F) using the promoter reference technique adapted from Oh et al. (2017). The respective strains were transformed with a plasmid harboring the open reading frame of either Fbp1-3xFLAG, Mdh2-3xFLAG or Pck1-3xFLAG and the control protein DHFR-HA, both expressed from identical promoters. Cells were grown in SD-glucose medium to OD₆₀₀ of 1.0 before being starved in SE medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% ethanol, amino acid mix) for 19 hours. Subsequently, an equivalent of 1 OD₆₀₀ was transferred to SD-glucose medium containing 0.5 mM tetracycline that inhibits translation of the respective substrate and DHFR by binding to specific RNA-regions within their ORFs. At the indicated time points, 1 mL or 1 OD₆₀₀ of cells was harvested. Cell lysis was performed by resuspending the pellets in 800 μ L 0.2 M NaOH and incubating them on ice for 20 minutes with subsequent centrifugation at 11,200xg for 1 minute at 4°C. The pellets were aspirated and resuspended in 50 μ L HU buffer (8 M Urea, 5%

SDS, 1 mM EDTA, 100 mM DTT, 200 mM Tris-HCl, pH 6.8, protease inhibitor, bromphenol blue), heated at 70°C for 10 minutes and then centrifuged again for 5 minutes at 11,200xg at 4°C. Protein levels of the substrates and a control protein DHFR were visualized by immunoblotting with anti-FLAG and anti-HA antibodies, respectively, and imaged using a Typhoon scanner (GE Healthcare). The bands were quantified using the ImageStudioLite software (LI-COR) and the substrate signal was normalized relative to the DHFR signal for every sample. At least three biological replicates were considered for all *in vivo* assays and the standard deviation was presented using error bars.

To validate the major ubiquitylation sites in Fbp1 *in vivo* (Figure 5B), the above-described PRT degradation assays were carried out in a similar manner with Fbp1-3xFLAG mutants, in which the lysines targeted by Chelator-GID^{SR4} (K32, K35, K280 and K281) were mutated to arginine.

To test if overexpression of Gid4 affects degradation of Fbp1 in Δ Gid7 yeast (Figure S1E), the GFP-Gid4 overexpression plasmid was transformed together with the Fbp1-3xFLAG PRT plasmid into different yeast strains (WT, Δ Gid7 and a Gid2^{K365A} catalytically inactive mutant). The cells were grown in SD medium lacking histidine and leucine, which served as selection markers for the Gid4 overexpression plasmid. After 8 h growth in SD-glucose media, samples of 1 OD₆₀₀ were harvested and analyzed as described above.

Purification of endogenous yeast GID for cryo EM

To purify endogenous GID complex, 3 l of a yeast strain with Gid7 and Gid5 C-terminally tagged at their endogenous loci with an HA and 3xFLAG tag, respectively, were grown in YPD medium for 8 hours. Subsequently, the cells were washed and resuspended to OD₆₀₀ of 1.0 in YPE medium (1% yeast extract, 2% peptone, 2% ethanol). Cells were harvested at OD₆₀₀ of 18.0. The pellet was resuspended in the lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM CaCl₂, 0.2 M sorbitol, complete protease inhibitor tablets) and frozen in liquid nitrogen in the form of small beads. For lysis, the frozen yeast pellets were subjected to cryogenic grinding using a cryo-mill (SPEX Sample Prep-6875 Freezer/Mill). The obtained yeast powder was thawed and centrifuged at 35,000 rpm for 10 minutes, and the resultant supernatant was incubated with ANTI-FLAG M2 affinity resin for an hour. After thorough washing, the protein was eluted using 3xFLAG peptide and visualized by Coomassie-stained SDS-PAGE. The eluted complex was concentrated to 1 mg/ml and analyzed by cryo EM.

Sucrose gradient fractionation of yeast lysates (Figure S2C)

Yeast strains with Gid8 C-terminally tagged at its endogenous locus with a 3xFLAG tag, with or without Gid7 deleted were grown in YPD media for 8 hours. Subsequently, they were switched to YPE medium and grown for 19 to 24 hours. One part of both cultures was harvested, while the other was switched to YPD medium for glucose recovery and harvested after 2 hours. The pellets were resuspended and lysed using a cryo-mill (as described above). To perform sucrose gradient fractionation of yeast lysates, roughly 300–500 mg of yeast powder was resuspended in the lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM CaCl₂, 0.2 M sorbitol, complete protease inhibitor tablets). To aid in resolubilization, lysates were incubated for 15 minutes at 4°C with gentle agitation, and then pre-cleared by centrifugation at 17,000xg for 10 minutes. Protein concentration was normalized by Bradford assay, lysates were loaded onto a 5%–40% sucrose gradient, and centrifuged at 34,300 rpm for 16 hours at 4°C. Gradients were then fractionated into fourteen equal fractions and loaded onto a 12% SDS-PAGE gel. Proteins were visualized by immunoblotting and imaged with Amersham Typhoon imager (GE Healthcare).

In vivo Fbp1 ubiquitylation assay (Figure 1E)

Yeast strains with Fbp1 tagged at its endogenous locus with 3xFLAG were grown to OD₆₀₀ of 1.0–1.5 in YPD, pelleted by centrifugation at 3,000 rpm for 3 min, washed with pre-warmed YPE, resuspended to an OD₆₀₀ = 1 in fresh, pre-warmed YPE, and grown at 30°C for 18 hours. Cultures for the ethanol condition were then diluted to an OD = 1 in fresh, pre-warmed YPE containing 1% DMSO. For the recovery condition, cells were pelleted by centrifugation at 3,000 rpm for 3 minutes, and resuspended in fresh pre-warmed YPD containing 1% DMSO. After two hours of growth at 30°C, 50 ODs of cells were pelleted by centrifugation at 3,000 rpm for 3 minutes, and flash frozen in liquid nitrogen.

Samples were resuspended in 1 mL lysis buffer (50 mM Tris-HCl, pH7.5, 150 mM NaCl, 2 mM EDTA, 50 mM NaF, 0.1% SDS, 1% NP-40, 0.5% Na-deoxycholate, 1% glycerol, 20 mM NEM, and complete protease inhibitor tablets), and lysed by 3 rounds of 20 s in a FastPrep-24 instrument, resting 5 minutes on ice between each round. Lysates were then pre-cleared by centrifugation at 4,000xg for 10 minutes, and the supernatant was added to pre-equilibrated His-Halo-UBA^{UBQLN1}-conjugated agarose beads, and incubated for 2 hours at 4°C with gentle rotation. Beads were separated by centrifugation at 800xg for 1 minute, washed once with lysis buffer and four times with wash buffer (50 mM Tris-HCl, pH7.5, 150 mM NaCl, 2 mM EDTA, 50 mM NaF, 0.1% SDS, 1% NP-40, 0.5% Na-deoxycholate, 1% glycerol). Proteins were eluted by addition of sample buffer, and heating at 95°C for 5 minutes. Samples were then loaded on a 12% SDS-PAGE gel and visualized by immunoblotting.

Plasmid preparation and Mutagenesis

All the genes encoding yeast GID subunits and the substrates Fbp1, Mdh2 and Pck1 were originally amplified from *S. cerevisiae* BY4741 genomic DNA. The genes coding for subunits of human CTLH were obtained from human cDNA library (Max Planck Institute of Biochemistry), except for hGid4, which was codon-optimized for bacterial expression system and synthesized by GeneArt gene

synthesis service (Thermo Fisher Scientific). The sequences of all the CTLH genes correspond to the canonical UniProt sequences, besides ARMC8, for which isoform 2 (missing the residues 2-15 of the canonical sequence) was used based on the prior literature (Kobayashi et al., 2007).

The constructs for recombinant protein expression were generated by Gibson assembly method (Gibson et al., 2009), whereas the mutant versions of the genes were prepared by the QuickChange protocol (Stratagene). All the coding sequences used for protein expression were verified by DNA sequencing. To express GID/CTLH subunits from a single baculoviral expression vector, the genes were combined by the biGBac method (Weissmann et al., 2016). All the plasmids used in this study are listed in the [Key resources table](#).

Insect cell expression and purification of GID/CTLH complexes

Both yeast GID and human CTLH complexes used for the biochemical assays and cryo EM were expressed in insect cells. For protein expression, Hi5 insect cells were transfected with recombinant baculovirus variants carrying the respective protein-coding sequences and grown for 60 to 72 hours in EX-CELL 420 Serum-Free Medium at 27°C. After harvesting, insect cell pellets were resuspended in a lysis buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT, 10 µg/ml leupeptin, 20 µg/ml aprotinin, 2 mM benzamide, EDTA-free Complete protease inhibitor tablet (Roche, 1 tablet per 50 mL of buffer) and 1 mM PMSF.

All recombinant yeast GID complexes were purified from insect cell lysates by StrepTactin affinity chromatography by pulling on a twin-Strep tag fused at the Gid8 C terminus. Further purification was performed by anion exchange chromatography and size exclusion chromatography (SEC) in the final buffer containing 25 mM HEPES pH 7.5, 200 mM NaCl and 5 mM (Buffer A) or 1 mM DTT (Buffer B) for cryo EM and biochemical assays, respectively. To ensure a stoichiometric level of the substrate receptor Gid4 in all cryo EM samples, all GID complexes were expressed without Gid4, and a bacterially-expressed truncated version of Gid4 (Δ 1-116) was added at a 2-fold molar excess to Gid^{Ant} (Gid1-Gid8-Gid2-Gid9-Gid5) before final SEC. To assemble Chelator-GID^{SR4}, both Gid4 (Δ 1-116) and Gid7 were added to Gid^{Ant} at a 2-fold molar excess before final SEC. For the sample of Chelator-GID^{SR4} with Fbp1 bound, 2-fold molar excess of the substrate was added to a purified and concentrated complex just before cryo EM grids preparation. A list of yeast GID complexes analyzed by cryo EM along with strategies for their expression and purification is shown below:

1. Chelator-GID^{SR4}: Gid1, Gid2, Gid5, Gid8-2xS, Gid9 coexpressed in Hi5 insect cells; bacterially expressed Gid4 (Δ 1-116) and Gid7 added before final SEC; purified by StrepTactin affinity, IEX and SEC
2. Fbp1-bound Chelator-GID^{SR4}: Gid1, Gid2, Gid5, Gid8-2xS, Gid9 coexpressed in Hi5 insect cells; bacterially expressed Gid4 (Δ 1-116) and Gid7 added before final SEC; purified by StrepTactin affinity, IEX and SEC; Fbp1-6xHis added directly before cryo EM grids preparation
3. GID^{SR4}: Gid1, Gid2, Gid5, Gid8-2xS, Gid9 coexpressed in Hi5 insect cells; bacterially expressed Gid4 (Δ 1-116) added before final SEC; purified by StrepTactin affinity, IEX and SEC

CTLH^{SR4} and CTLH-MKLN1 subcomplex comprising SA and SRS modules were purified from insect cell lysates by StrepTactin affinity chromatography by pulling on a twin-Strep tag fused at the TWA1 C terminus, whereas the CTLH-WDR26 subcomplex comprising SA and SRS modules was pulled on a twin-Strep tag at the ARMC8 N terminus. Further purification was performed by anion exchange chromatography and size exclusion chromatography in Buffer A or Buffer B. As for yeast GID, the CTLH subcomplexes used for cryo EM were saturated with hGid4 by mixing them with the bacterially-expressed truncated version of hGid4 (Δ 1-99) and running SEC. CTLH-WDR26 supramolecular assembly was purified from lysates by a tandem affinity chromatography, by first pulling on TWA1-2xS and then GST-WDR26. The pull-down fractions were run on SEC in Buffer A. A list of human CTLH complexes analyzed by cryo EM along with strategies for their expression and purification is shown below:

1. CTLH-WDR26 supramolecular assembly: RANBP9, TWA1-2xS, ARMC8, RMND5A, MAEA, GST-WDR26 coexpressed in Hi5 insect cells; purified by tandem StrepTactin and GST affinity and SEC
2. CTLH-WDR26 SA and SRS modules: RANBP9, TWA1, 2xS-ARMC8, WDR26, YPEL5 coexpressed in Hi5 insect cells; bacterially expressed hGid4 (Δ 1-99) added before final SEC; purified by StrepTactin affinity, IEX and SEC
3. CTLH-MKLN1 SA and SRS modules: RANBP9, TWA1-2xS, ARMC8, MKLN1 coexpressed in Hi5 insect cells; bacterially expressed hGid4 (Δ 1-99) added before final SEC; purified by StrepTactin affinity, IEX and SEC
4. CTLH^{SR4}: RANBP9, TWA1-2xS, ARMC8, RMND5A, MAEA coexpressed in Hi5 insect cells; bacterially expressed hGid4 (Δ 1-99) added before final SEC; purified by StrepTactin affinity, IEX and SEC

Bacterial expression and purification

All bacterial expressions were performed in *E. coli* BL21 (DE3) RIL cells in a Terrific Broth (TB) medium overnight at 18°C.

All the mutant and WT versions of Gid4 (both yeast and human ortholog) and Gid7 were expressed as GST-TEV fusions. After harvesting, cell pellets were resuspended in the lysis buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT and 1 mM PMSF and purified from bacterial lysates by glutathione affinity chromatography, followed by overnight digestion at 4°C with tobacco etch virus (TEV) protease to liberate the GST tag. Further purification was carried out with size exclusion chromatography in Buffer B. Additionally, a pass-back over glutathione affinity resin was performed to get rid of the remaining uncleaved GST-fusion protein and free

GST. Ubc8, Ubc2H, Ub (for generating ubiquitylated Fbp1), Fbp1 (WT and mutants), Mdh2 and Pck1 were expressed as their C-terminally 6xHis-tagged versions. After harvesting, cell pellets were resuspended in the lysis buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM β -mercaptoethanol, 10 mM imidazole and 1 mM PMSF, and purified from bacterial lysates by nickel affinity chromatography, followed by anion exchange and size exclusion chromatography in Buffer A or Buffer B (for structural studies and activity assays, respectively). Fbp1-V5-2xS (for Fbpase activity assays) was purified by StrepTactin affinity chromatography and SEC in Buffer B.

Untagged WT ubiquitin used for *in vitro* assays was purified via glacial acetic acid method (Kaiser et al., 2011), followed by gravity S column ion exchange chromatography and size exclusion chromatography in Buffer B. Different Ub variants as well as WT Ub used for the ubiquitin chain type determination assay were expressed as GST-3C fusions and purified by glutathione affinity chromatography, followed by incubation with HRV-3C protease for 3 hours at room temperature. Further purification was done with size exclusion chromatography in Buffer B.

Fluorescent tagging of the GID substrates Fbp1, Mdh2 and Pck1 used for all the biochemical assays was performed with a sortase A-mediated reaction, which catalyzed fusion of fluorescein to the C terminus of the substrate. The reaction mix contained 50 μ M of the substrate, which was C-terminally tagged with a sortag (LPETGG) and a 6xHis tag, 250 μ M of a fluorescent peptide (GGGGFYVK-FAM) and 50 μ M of sortase A (Chen et al., 2011b). The labeling reaction was carried out for 30 minutes at room temperature in a buffer comprising 50 mM Tris-HCl pH 8, 150 mM NaCl and 10 mM CaCl_2 . The reaction mixture was consecutively passed-back through the Ni-NTA Sepharose resin to get rid of unreacted Fbp1. Further purification was done with size exclusion chromatography in Buffer B.

All the labeled and unlabeled peptides used in the biochemical assays were synthesized in the MPIB Biochemistry Core Facility.

In vitro biochemical assays

All *in vitro* activity assays were performed at room temperature in a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl, 5 mM ATP and 10 mM MgCl_2 . To ensure that all the reaction mixtures contained equal concentrations of WT and mutant versions of Gid4 and Gid7, these proteins were added exogenously for all assays besides kinetics. To analyze kinetics of Fbp1 ubiquitylation, the assays were performed with purified GID^{SR4} and Chelator-GID^{SR4} obtained by co-expressing all of their subunits, as well as GID^{SR4} mixed with Gid7 before starting the reaction. All the reactions were quenched at indicated time points by mixing an aliquot of the total reaction mix with SDS-PAGE loading buffer. Ubiquitylation of fluorescent substrates was visualized by a fluorescent scan of SDS-PAGE gel using the Amersham Typhoon imager (GE Healthcare).

Biochemical assays with yeast GID

The influence of Gid7 and Gid4 on ubiquitylation of the full-length (Figures 1A and 1B) and peptide versions (Figure S1B) of Fbp1 was tested in a multiturnover assay format using 0.2 μ M Uba1, 1 μ M Ubc8-6xHis, 0.5 μ M GID^{Ant}, 0 or 1 μ M Gid4, 0 or 2 μ M Gid7, 1 μ M full-length Fbp1-FAM or a fluorescently labeled model peptide substrate and 20 μ M Ub (WT or all K > R (K0) version). The model peptide substrate was designed with the N-terminal Fbp1 sequence (aa 2-16) and a single lysine placed at position 27 (to span the distance between the substrate receptor Gid4 and the catalytic center measured in the structure of Chelator-GID^{SR4}). Similarly, the influence of Gid7 and Gid4 on ubiquitylation of other gluconeogenic substrates, Mdh2 and Pck1, was tested in a multiturnover assay using their fluorescently labeled versions and carried out under identical conditions (Figure 1A). The same assay format and conditions were employed to qualitatively compare Fbp1 ubiquitylation activity of GID^{SR4} exogenously mixed with Gid7 to that of the SEC-purified Chelator-GID^{SR4} (containing co-expressed Gid7) (Figure S1C). All the assays were performed in at least duplicates and some of them were quantified using image analysis software ImageQuant (GE healthcare; version 8.2).

To test the influence of Gid7 on intrinsic activity of GID E3, a substrate-independent pulse-chase assay monitoring discharge of Ubc8~Ub to free lysine in solution was employed (Figure S1A). In the pulse reaction, loading of Ubc8 was performed by mixing 0.5 μ M Uba1, 10 μ M Ubc8-6xHis, 30 μ M Ub, 2.5 mM MgCl_2 and 1 mM ATP. After 15 minutes at room temperature, Ubc8 loading was stopped by incubation of the pulse mixture with 50 mM EDTA on ice for 5 minutes. For the chase reaction, the quenched pulse mixture was mixed with an equal volume of the chase-initiating mixture containing 1 μ M GID^{SR4} complex, 0 or 2 μ M Gid7 (WT or Δ 1-285 mutant) and 25 mM lysine pH 8.0. The discharge was carried out at room temperature, quenched at different time points and visualized by non-reducing SDS-PAGE stained with Coomassie.

Avid binding of Fbp1 to Chelator-GID^{SR4} was verified by performing a competition ubiquitylation assay in a multiturnover format (Figure 4D). The reactions were initiated by mixing 0.2 μ M Uba1, 1 μ M Ubc8-6xHis, 0.5 μ M E3 GID^{SR4}, 0 or 2 μ M Gid7 (WT or its Δ 1-284 mutant), 0.5 μ M of fluorescently labeled tetrameric Fbp1 or a monomeric model peptide substrate containing Fbp1 degron (as described above), 20 μ M of an unlabeled competitor (full-length Fbp1-6xHis with major target lysines K32, K35, K280, K281 mutated to R or a lysine-less 9-residue peptide containing Fbp1 N-terminal sequence) and 20 μ M Ub. Before starting the reaction, GID^{SR4} was incubated with Gid7 for 3 minutes.

To validate the preferred ubiquitylation sites in Fbp1 determined by proteomics, multi-turnover ubiquitylation assays were performed using mutants of Fbp1, in which the pairs of major target lysines were mutated to arginine separately or together (Figure 5A). The reaction mixtures contained 0.2 μ M Uba1, 1 μ M Ubc8-6xHis, 0.5 μ M GID^{Ant}, 1 μ M Gid4, 2 μ M Gid7, 1 μ M Fbp1-6xHis (WT or target K > R mutants) and 20 μ M Ub (WT or its all K > R (K0) version). Ubiquitylation of the substrates was visualized by immunoblotting with anti-His antibody.

Determination of kinetic parameters of Fbp1 ubiquitylation by GID E3

To examine the effect of Gid7 on the Michaelis-Menten constant K_m for Fbp1 ubiquitylation by GID E3 (Figure 1C), mult turnover assays were performed by titrating the E3 concentration and with substrate levels that were below K_m . Assays were performed with GID^{SR4}, GID^{SR4} mixed with Gid7 as well as a purified Chelator-GID^{SR4} (GID^{SR4} coexpressed with Gid7). Reactions were quenched at time points such that the initial velocities of all reactions were well within the linear range (determined by running time courses for reactions that contained the highest E3 concentration from the titrations). Reactions with GID^{SR4} comprised 0.2 μ M Uba1, 1 μ M Ubc8, 0.25–8 μ M GID^{SR4}, 0.5 μ M Fbp1-FAM and 20 μ M Ub, and were quenched after 8 minutes. For GID^{SR4} exogenously mixed with Gid7, reactions comprised 0.2 μ M Uba1, 1 μ M Ubc8, 0.025–0.8 μ M GID^{SR4} mixed with a 2-fold excess of Gid7, 0.1 μ M Fbp1-FAM and 20 μ M Ub, and were quenched after 3 minutes. In the case of Chelator-GID^{SR4}, the reaction mixes contained 0.2 μ M Uba1, 1 μ M Ubc8, 0.03–1 μ M Chelator-GID^{SR4}, 0.1 μ M Fbp1-FAM and 20 μ M Ub, and the reactions were quenched after 2 minutes. Reaction substrate and products were resolved by SDS-PAGE and quantified using ImageQuant (GE healthcare; version 8.2). Fraction of Fbp1 that had been modified by one or more ubiquitins was then plotted as a function of E3 concentration in GraphPad Prism and fit to the Michaelis-Menten equation using non-linear curve fitting. All reactions were performed in duplicate.

Since the method described in the previous paragraph involved titration of E3 levels rather than that of the substrate, k_{cat} was estimated using the following protocol. Initial velocities were measured for both GID^{SR4} and Chelator-GID^{SR4} by performing a time course where the ratios of both E3 to K_m and substrate to K_m were the same for each E3 complex (2.7 and 0.4, respectively). The fraction of ubiquitylated Fbp1 was plotted in GraphPad Prism as a function of time (Figure S1D) and the rate of the reaction was estimated by linear regression. Having calculated the rate, initial velocities V_0 were calculated using the following equation: $V_0 = rate \cdot [S]$. V_{max} was then estimated using a modified form of the Michaelis-Menten equation: $V_{max} = \frac{V_0 \cdot (K_m + [S])}{[S]}$, where $[S] = \frac{K_m}{2.5}$ because the substrate concentration was 2.5 times lower than K_m . To obtain k_{cat} values, V_{max} was divided by the E3 concentration: $k_{cat} = \frac{V_{max}}{[E3]}$.

Biochemical assays with human CTLH^{SR4}

All *in vitro* ubiquitylation assays with CTLH^{SR4} were performed using a 30-residue fluorescent model peptide substrate harboring an N-terminal hGid4-interacting sequence PGLW and a single lysine placed at position 27, which is an optimal distance between the catalytic module and hGid4 based on the cryo EM structure.

To probe the residues of hGid4 that mediate its incorporation into CTLH^{SR4}, structure-based hGid4 mutants (corresponding to homologous mutations in yeast Gid4 (Qiao et al., 2020)) were tested in a binding test (Figure S6D) and ubiquitylation assays (Figure S6E). For the binding test, 10-fold molar excess of the purified WT and mutant hGid4 (Δ 1–99) was mixed with 20 μ g of RANBP9-TWA1-ARMC8-RMND5A-MAEA complex (tagged with a twin-Strep tag at TWA1 C terminus) in a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT. After incubating the proteins for 30 minutes on ice, 40 μ L of the StrepTactin resin was added to the mixture and further incubated for 1 hour. As a control, RANBP9-TWA1-ARMC8-RMND5A-MAEA complex and hGid4 were mixed with StrepTactin alone. After throughout wash of the resin, elution fractions were collected and analyzed with SDS-PAGE stained with Coomassie. Ubiquitylation reactions were performed in a mult turnover format by mixing 0.2 μ M Uba1, 2 μ M Ube2H-6xHis, 1 μ M RANBP9-TWA1-ARMC8-RMND5A-MAEA complex, 1 μ M hGid4 (Δ 1–99, WT or an indicated mutant), 0.5 μ M fluorescent model peptide substrate and 20 μ M Ub.

The catalytic mechanism of CTLH was examined by testing mutants of RMND5A and MAEA in substrate-independent discharge reactions (Figure S6G) and ubiquitylation assays (Figure S6H). The substrate-independent reactions monitored the discharge of Ube2H~Ub to free lysine in solution in a pulse-chase format, applying the conditions as described for the assay with yeast GID. For the ubiquitylation mult turnover assays, the reactions contained 0.2 μ M Uba1, 2 μ M Ube2H-6xHis, 1 μ M RANBP9-TWA1-ARMC8-RMND5A-MAEA complex (containing either WT or indicated mutants of RMND5A or MAEA), 1 μ M hGid4 (Δ 1–99), 0.5 μ M fluorescent model peptide substrate and 20 μ M Ub.

For characterizing the ubiquitin chain type formed by CTLH^{SR4} in conjunction with Ube2H, a mult turnover assay was performed (Figure S6I). The reaction mix contained 0.2 μ M Uba1, 2 μ M Ube2H-6xHis, 1 μ M RANBP9-TWA1-ARMC8-RMND5A-MAEA complex, 1 μ M hGid4 (Δ 1–99), 0.5 μ M fluorescent model peptide substrate and 20 μ M Ub (WT, lysine-less (all K > R) or one of its single-lysine variants (with all but one lysine mutated to arginine)).

SEC for initial characterization of GID supramolecular assembly

For initial test of how Gid7 affects GID complex assembly (Figure S2A), 200 μ L of 10 μ M Gid7 and GID^{SR4} alone or together (mixed in 1:1 ratio) were loaded onto a Superose 6 column (GE Healthcare) equilibrated with 25 mM HEPES 7.5, 150 mM NaCl and 5 mM DTT. SEC fractions were analyzed with Coomassie-stained SDS-PAGE.

SEC-MALS

To determine the oligomeric state of Fbp1 and Gid7 (Figure S2B), the proteins were subjected to SEC-MALS analysis. For each run, 100 μ L of samples at 1 mg/mL were loaded onto Superdex 200 column equilibrated with a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl and 5 mM DTT. SEC-MALS was conducted in the MPIB Biochemistry Core Facility.

Fbp1 enzyme activity assay

To test the effect of Fbp1 ubiquitylation on its activity (Figure 5E) and sensitivity to allosteric regulation by AMP (Figure 5F), EnzChek Phosphate Assay Kit (ThermoFisher Scientific) was employed. This assay quantifies inorganic phosphate (P_i) released from fructose-1,6-bisphosphate by Fbp1 through enzymatic conversion of 2-amino-6-mercapto-7-methyl-purine riboside (MESG) to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine by purine nucleoside phosphorylase (PNP). This leads to a shift in maximum absorbance from 330 nm for MESG to 360 nm for the final reaction product (2-amino-6-mercapto-7-methylpurine).

To obtain fully ubiquitylated Fbp1, it was subjected to an overnight multiturnover ubiquitylation reaction at room temperature consisting of 0.2 μM Uba1, 1 μM Ubc8, 0.5 μM GID^{Ant}, 1 μM Gid4, 2 μM Gid7, 10 μM Fbp1-V5-2xS and 100 μM 6xHis-3c-Ub. The reaction mix was run on SEC (using Superose 6 column) to separate different components of the assay. Fractions corresponding to the ubiquitylated Fbp1 were pooled and incubated with Ni-NTA resin for 30 minutes. After throughout wash, the bound protein was eluted and visualized by SDS-PAGE.

Fbp1 activity assays were performed according to the manufacturer's instructions at room temperature. First, all the reagents provided in the assay kit (MESG, PNP and 20x reaction buffer), 0.5 mM fructose-1,6-bisphosphate substrate and 0.6 mM AMP (only for the Fbp1 inhibition assay) were pre-mixed and incubated for 5 min. Then, the reaction was initiated by addition of 53 nM of the WT, target lysine mutant (K32A/K35A/K280A/K281A) or fully ubiquitylated Fbp1, and the reaction progress was followed by measuring a time-course of absorbance at 360 nm (A₃₆₀, absorbance of the final reaction product) using CLARIOStar Plus microplate reader (BMG LABTECH) in a UV-transparent 96-well plate. The values of A₃₆₀ obtained for the buffer-only control were subtracted from all the experimental measurements, which were then plotted in GraphPad Prism.

Analysis of global proteome of WT versus ΔGid7 yeast (Karayel et al., 2020)

To test which proteins are dependent on Gid7 for their *in vivo* degradation, we compared the global proteome of WT with that of the ΔGid7 yeast (Figure S1F). Cells were grown in SD media to an OD of 1-1.5, pelleted by centrifugation, washed in pre-warmed SE media, and resuspended to an OD of 1 in fresh, pre-warmed SE media. Cultures were then allowed to grow at 30°C for 18 hours, after which cells were again pelleted by centrifugation, and resuspended in fresh, pre-warmed SD media to an OD of 1. Following growth at 30°C for 2 hours, 50 ODs of cells were pelleted by centrifugation, flash frozen and stored at -80°C until lysis. The frozen pellets were mixed with SDC lysis buffer (1% SDC and 100 mM Tris pH 8.5) and immediately heat-treated for 5 minutes at 95°C. Lysates were homogenized by sonication at 4°C using a Bioruptor and then diluted to achieve equal protein concentrations in a 96-well plate. Samples were next incubated for 5 minutes at 45°C with 40 mM CAA and 10 mM TCEP for reduction and alkylation and digested overnight at 37°C using trypsin (1:100 w/w, Sigma-Aldrich) and LysC (1/100 w/w, Wako). Next day, peptide material was desalted using SDB-RPS StageTips (Empore) (Kulak et al., 2014) and resuspended in buffer A (0.2% TFA/2% ACN). Peptide concentrations were measured by absorbance at 280 nm (Nanodrop 2000, Thermo Scientific) and equalized using buffer A*. 300 ng peptides were subjected to LC-MS/MS analysis.

Samples were loaded onto a 20 cm reversed phase column (75 μm inner diameter, packed in house with ReproSil-Pur C18-AQ 1.9 μm resin (Dr. Maisch GmbH)). The column temperature was maintained at 60°C using a homemade column oven. A binary buffer system, consisting of buffer I (0.1% formic acid (FA) and buffer II (80% ACN plus 0.1% FA), was used for peptides separation, at a flow rate of 450 nl/min. An EASY-nLC 1200 system (Thermo Fisher Scientific), directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source, was employed for nano-flow liquid chromatography. We used a gradient starting at 5% buffer B, increased to 35% in 18 and a half minute, 95% in a minute and stayed at 95% for three and a half min. The mass spectrometer was operated in DIA mode. Full MS resolution was set to 120,000 with a full scan range of 300-1650 m/z, a maximum fill time of 60 ms and an automatic gain control (AGC) target of 3e6. One full scan was followed by 12 windows with a resolution of 30,000 in profile mode. Precursor ions were fragmented by stepped higher-energy collisional dissociation (HCD) (NCE 25.5, 27,30%).

Spectronaut version 13 (Biognosys) was used to analyze DIA raw files using the yeast FASTA file (Swissprot, 2018) and the proteome library previously published (Karayel et al., 2020) with default settings and enabled cross run normalization. The Perseus software package version 1.6.0.7 was used for the data analysis (Tyanova et al., 2016). Protein intensities were log₂-transformed and filtered to make sure that identified proteins showed expression in all biological triplicates of at least one condition. The missing values were subsequently replaced by random numbers that were drawn from a normal distribution (width = 0.3 and down shift = 1.8). For volcano plots, we used permutation-based FDR, which was set to 0.05 in conjunction with an S0-parameter of 0.1 to determine the significance.

Determination of preferentially targeted lysines in Fbp1 by LC-MS/MS (Figure S5)

To determine the preferentially targeted lysines in Fbp1, it was ubiquitylated by Chelator-GID^{SR4} and subjected to proteomic analysis. To capture the initial ubiquitylation events, the assay was performed in a single-turnover pulse-chase format, wherein the concentration of the substrate was significantly exceeding that of E2~Ub. In the pulse, 10 μM Ubc8 was loaded with 30 μM lysine-less ubiquitin mutant (all K > R) and 0.5 μM Uba1 for 15 minutes at room temperature and quenched with 50 mM EDTA. To start the chase, the pulse reaction was mixed with an equal volume of the chase-initiating mixture containing 1 μM GID^{Ant}, 2 μM Gid7, 2 μM Gid4 and 4 μM Fbp1-6xHis and incubated at room temperature. After 1 minute, the reaction was quenched by adding 10 mM DTT, which was then removed by desalting before proteomic analysis.

Proteins were digested and prepared for LC-MS/MS measurements as previously described (Qiao et al., 2020). Briefly, samples were diluted in digestion buffer (1 M urea in 50 mM ammonium bicarbonate, pH 8.0), followed by addition of TCEP and CAA to a final concentration of 10 mM and 40 mM, respectively. After reduction and alkylation for 5 minutes at 45°C, samples were enzymatically digested using either trypsin (1:20 w/w, Sigma-Aldrich) alone, trypsin (1:40 w/w)/GluC (1:40 w/w, BioLab) or trypsin (1:40 w/w)/AspN (1:40 w/w, Promega) at 37°C overnight. Thereafter, protease activity was quenched and peptides were loaded and cleaned on SDP-RPS StageTips. Peptides were subsequently eluted with 1.25% ammonium hydroxide/80% ACN, dried using a SpeedVac centrifuge (Eppendorf, Concentrator plus) and resuspended in buffer A (2% ACN/0.1% TFA) for LC/MS-MS analysis.

Peptide concentration was estimated by UV spectrometry and approximately 200 ng were loaded on a 50 cm reversed phase column (75 µm inner diameter, packed in-house with ReproSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch GmbH)). Column temperature was maintained at 60°C using a homemade column oven. Peptides were separated with a binary buffer system of buffer A (0.1% formic acid (FA)) and buffer B (80% acetonitrile plus 0.1% FA), at a flow rate of 300 nl/min. We used an EASY-nLC 1200 system (Thermo Fisher Scientific), which was directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source. Peptides were eluted with a gradient starting at 3% buffer B and stepwise increased to 8% in 8 min, 36% in 32 min, 45% in 4 minutes and 95% in 4 min. The mass spectrometer was operated in Top12 data-dependent mode (DDA) with a full scan range of 250–1350 m/z at 60,000 resolution with an automatic gain control (AGC) target of 3e6 and a maximum fill time of 20 ms. Precursor ions were isolated with a width of 1.4 m/z and fragmented by higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 28%. Fragment scans were performed at a resolution of 30,000, an AGC of 1e5 and a maximum injection time of 110 ms. Dynamic exclusion was enabled and set to 15 s.

Raw MS data were searched against UniProt Yeast FASTA using MaxQuant (version 1.6.2.10) with a 1% FDR at peptide and protein level. Cysteine carbamidomethylation was set as fixed, protein N-terminal acetylation, methionine oxidation and lysine diGly as variable modifications. The minimum peptide length was set to 7 amino acids, enzyme specificity was set to trypsin and two missed cleavages were allowed, permitting a maximum of 5 modifications per peptide. MS/MS spectra identifying ubiquitylated peptides of interest were obtained and exported using MaxQuant Viewer.

Cell culture and generation of CRISPR-Cas9 knock out cell lines

K562 erythroleukemia cell line was obtained from ATCC (CCL-243TM) and cultured in IMDM completed with 10% (v/v) FBS (GIBCO) and antibiotics (100 units/ml penicillin, 0.1 mg/ml streptomycin, GIBCO). Cell densities were kept between $0.1-1 \times 10^6$ cells/mL, and cultures were regularly checked for the absence of mycoplasma contamination. For CRISPR-Cas9-(D10A) nickase-mediated functional knockouts of MAEA, MKLN1 and WDR26, paired sense and antisense guide RNAs (gRNA) were designed to target MAEA in exon 2, exon 5 in MKLN1 and exon 1 in WDR26 genetic locus. Sense and antisense gRNA were cloned into pBABED-U6-Puromycin plasmid (gift from Thomas Macartney, University of Dundee, UK) and pX335-Cas9(D10A) (Addgene) (Cong et al., 2013), respectively. K562 cells were co-transfected with vectors encoding the pair of gRNAs using Lipofectamine LTX reagent (Invitrogen) following manufacturer's instructions. Twenty-four hours after transfection, cells were selected in puromycin (2 µg/ml) for 2 days, followed by expansion, and single-cell dilution to obtain cell clones. Successful knockout clones were confirmed by immunoblotting and genomic sequencing of targeted loci (Figure S6J).

Human cell lysate fractionation by sucrose density gradient

1×10^7 cells were harvested by centrifugation at 360 x g, washed once with ice-cold PBS, and resuspended in lysis buffer (40 mM HEPES pH 7.5, 120 mM NaCl, 1 mM EGTA, 0.5% NP40, 1 mM DTT, and Complete protease inhibitor mix (Roche)). Cells were homogenized by pushing them 10 times through a 23G syringe. The obtained lysate was cleared by centrifugation at 23,000 x g for 30 minutes at 4°C. 3 mg of total protein were loaded on top of a 5%–40% sucrose gradient (weight/volume, in lysis buffer) and centrifuged in a SW60 rotor at 34,300 rpm for 16 hours at 4°C. Fractions were collected from top of the gradient and separated by SDS-PAGE, followed by immunoblotting using the following antibodies: RMND5A (Santa Cruz), MAEA (R&D systems), RANBP9 (Novus Biologicals), TWA1 (Thermo Fisher), ARMC8 (Santa Cruz), WDR26 (Bethyl Laboratories), MKLN1 (Santa Cruz) and YPEL5 (Thermo Fisher). Antibodies that recognize hGid4 were generated by immunizing sheep with bacterially expressed GST-hGid4 ($\Delta 1-99$). Western blots were developed using Clarity Western ECL Substrate (BioRad) and imaged using Amersham Imager 600 (GE Lifesciences).

Cryo EM sample preparation and imaging

Cryo EM grids were prepared using Vitrobot Mark IV (Thermo Fisher Scientific) operated at 4°C and 100% humidity. 3.5 µl of freshly purified proteins at 0.3–0.5 mg/ml were applied to glow-discharged Quantifoil holey carbon grids (R1.2/1.3 200 mesh). Grids were immediately blotted with Whatman no. 1 filter paper (blot time: 3 s, blot force: 3) and vitrified by plunging into liquid ethane.

Cryo EM data were first screened and collected on a Talos Arctica or Glacios transmission electron microscope (Thermo Fisher Scientific) operated at 200 kV, equipped with a Falcon III (Thermo Fisher Scientific) or K2 (Gatan) direct electron detector, respectively. Automated data collection was carried out using EPU software (Thermo Fisher Scientific) or SerialEM (Mastronarde, 2003). High-resolution datasets were collected on a Titan Krios (Thermo Fisher Scientific) microscope operated at 300 kV, equipped with a post-column GIF and a K3 Summit direct electron detector (Gatan) operating in a counting mode. SerialEM was used to automate data collection (Mastronarde, 2003). Details of cryo EM data collection and map refinement are listed in Table S1.

Cryo EM data processing

Frames were motion-corrected with dose weighting using MotionCorr2 (Zheng et al., 2017) and subjected to estimation of contrast transfer function parameters with Gctf (Zhang, 2016). Auto-picking of particles was performed with Gautomatch (<https://www.mrc-lmb.cam.ac.uk/kzhang/>) and for most datasets, it was aided by provision of a template obtained from previous low-resolution datasets. For Titan Krios datasets, movies were being pre-processed on-the-fly during data collection with Focus (Biyani et al., 2017), which also automatically discarded poor quality images. All the subsequent stages of data processing were carried out with Relion (Fernandez-Leiro and Scheres, 2017; Scheres, 2012; Zivanov et al., 2018). To clean up the data, extracted particles were subjected to either several rounds of 2D classification, followed by a 3D classification or submitted directly to a masked 3D classification. The chosen subset of particles was subjected to auto-refinement without and with a mask. To improve the quality of maps obtained after consensus refinement, a 3D classification without particle alignment was performed and a class having the most complete features was selected.

High-resolution maps of yeast substrate receptor scaffolding (SRS), catalytic (Cat) and supramolecular assembly (SA) modules were obtained from the Chelator-GID^{SR4} dataset with its substrate Fbp1 bound. For the SRS module, a more resolved half of the Chelator-GID^{SR4} was first auto-refined and a focused 3D classification without particle alignment was performed with a mask over GID^{SR4}. Then, focused refinement was performed, wherein the Cat module was masked out. For Cat and SA modules, the number of particles used for alignment was doubled by taking advantage of the Chelator-GID^{SR4} having C2 symmetry. First, the map of the entire complex was auto-refined with C2 symmetry imposed and masks were created for each of its halves. Then, the signal for each half was separately subtracted and the resulting semi-elliptical particles were aligned by auto-refinement. After masking out the SRS module, a focused 3D classification without particle alignment was performed separately for Cat and SA modules. After one more round of 3D classification without particle alignment with a higher T-value, the particles were subjected to CTF refinement and final auto-refinement.

For high-resolution CTLH^{SR4} dataset, the density corresponding to the catalytic module (RMND5A-MAEA) was masked out due to its mobility relative to the substrate receptor scaffolding module (RANBP9-TWA1-ARMC8-hGid4). For visualization of less resolved parts of the map, such as RANBP9^{LisH-CRAC}-TWA1 as well as the N- and C-termini of ARMC8, subsequent rounds of focused 3D classifications with masks over these regions were carried out. The final auto-refinement was preceded by a CTF refinement.

All maps were post-processed by automatic B-factor weighting and high-resolution noise substitution in Relion. In addition, to aid in building atomic models, the refined maps of the Chelator-GID^{SR4} SA module and CTLH^{SR4} SRS module were post-processed with DeepEMhancer (Sanchez-Garcia et al., 2020) and are deposited as additional maps in EMDb. The estimated resolutions of all reconstructions are based on the gold-standard Fourier Shell Correlation (FSC) at 0.143 criterion. Simplified schematic of processing for both Titan Krios datasets are presented in Figures S4 and S7 (for Chelator-GID^{SR4} and CTLH^{SR4}, respectively).

Model building and refinement

Manual building of all models was performed with Coot (Emsley and Cowtan, 2004; Emsley et al., 2010), whereas structure visualization and analysis was carried out with Chimera (Pettersen et al., 2004), ChimeraX (Goddard et al., 2018) and Pymol-v2.1 (<https://pymol.org/2/>).

The atomic model of CTLH substrate receptor scaffolding module was prepared as follows. Most of ARMC8 was built automatically with Buccaneer (Cowtan, 2006) and refined manually with Coot. The model of the substrate receptor hGid4 was generated by docking its crystal structure (PDB: 6CDC) into the EM map and manual building of its N- and C-termini. The crystal structure of RANBP9 SPRY domain (PDB: 5JI7) was fitted into the electron density map and served as a starting point for manual building of its downstream region. Manual building of TWA1 was guided by fitting parts of its homology model into the map (generated by SWISS-MODEL (Waterhouse et al., 2018), based on the structure of yeast Gid8 in GID^{SR4}, PDB: 6SWY) and secondary structure prediction obtained from Phyre² server (Kelley et al., 2015).

The structure of the SRS module in Chelator-GID^{SR4} was generated by fitting the atomic coordinates of the corresponding part of GID^{SR4} (PDB: 6SWY) and manual refinement. The loops of Gid4 surrounding its substrate-binding cavity as well as Fbp1 degron were built manually. Coordinates of most of Gid8 and Gid1 in the SA module were fitted from the structure of GID^{SR4} and the missing or differing parts, such as Gid1 CTLH-CRA^N, were built manually. The LisH-CRA^C as well as CTLH-CRA^N domains of Gid7 were built manually, guided by secondary structure predictions. Manual building of Gid7 β -propellers was aided by their homology model from Phyre² (Kelley et al., 2015). All of the Cat module was built manually and the geometry of the zinc binding site was constrained to account for tetrahedral coordination of zinc ions.

All the models were subjected to iterative rounds of manual building with Coot and real space refinement in PHENIX (Adams et al., 2010; Afonine et al., 2018; DiMaio et al., 2013) until a satisfactory model quality, in terms of its geometry and agreement with the map, was obtained.

Fbp1 crystallization and data processing

Crystallization trials of Fbp1-6xHis were performed in the MPIB Crystallization Facility. Before setting up the crystallization trays, the purified Fbp1 was concentrated to 12 mg/mL and combined with 0.5 mM of its allosteric inhibitor AMP and the substrate fructose-1,6-bisphosphate. Crystals used for X-ray data collection were obtained at 4°C in the buffer containing 16% PEG 3350, 0.2 M MgCl₂

and 0.1 M Bis-Tris pH 6 using a vapor diffusion method performed in a sitting-drop format. Crystals were cryoprotected using 20% ethylene glycol and stored by flash freezing in liquid nitrogen until data collection.

Diffraction dataset was recorded at X10SA beam line, Swiss Light Source (SLS) in Villigen, Switzerland. Data were recorded at 0.5 degree rotation intervals using Dectris Pilatus 2M-F detector. Data were indexed, integrated, and scaled using XDS package to a resolution limit of 1.95 Å. Phasing was performed through molecular replacement using a structure of human Fbp1 (PDB: 1FTA) with PHASER integrated into the PHENIX software suite (Adams et al., 2010; Afonine et al., 2018; DiMaio et al., 2013). Model building was done using Coot (Emsley and Cowtan, 2004; Emsley et al., 2010), whereas refinement was carried out with phenix.refine. Details of X-ray diffraction data collection and refinement statistics are listed in Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the assays described in the section “*In vivo* yeast substrate degradation assays,” protein bands visualized by western blots were quantified using ImageStudioLite software (Li-Cor). For statistical analysis, at least three biological replicates were considered and the standard deviation of the replicates was presented using error bars.

Fluorescently labeled proteins in *in vitro* ubiquitylation reactions were quantified in ImageQuant (GE Healthcare) and the calculated fractions of ubiquitylated substrates were plotted in GraphPad Prism. All *in vitro* assays were performed in at least duplicates and the standard deviation represented by error bars are shown wherever necessary. For determination of K_m for Fbp1 ubiquitylation by GID E3 with kinetics, the fraction of ubiquitylated Fbp1 was fit to the Michaelis-Menten equation in GraphPad Prism. k_{cat} was calculated based on a slope of a linear phase of Fbp1 ubiquitylation reaction fitted in GraphPad Prism.

3.7 Linkage-specific ubiquitin chain formation depends on a lysine hydrocarbon ruler

Joanna Liwocha^{1,9}, David T. Krist^{1,8,9}, Gerbrand J. van der Heden van Noort^{2,9}, **Fynn M. Hansen**³, Vinh H. Truong⁴, Ozge Karayel³, Nicholas Purser⁵, Daniel Houston⁵, Nicole Burton⁵, Mark J. Bostock^{6,7}, Michael Sattler^{6,7}, Matthias Mann³, Joseph S. Harrison⁴, Gary Kleiger⁵, Huib Ovaa^{2,10} and Brenda A. Schulman¹

¹Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany. ²Oncode Institute and Department of Cell and Chemical Biology, Chemical Immunology, Leiden University Medical Centre, Leiden, the Netherlands. ³Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. ⁴Department of Chemistry, University of the Pacific, Stockton, CA, USA. ⁵Department of Chemistry and Biochemistry, University of Nevada, Las Vegas, Las Vegas, NV, USA. ⁶Biomolecular NMR and Center for Integrated Protein Science Munich at Department Chemie, Technical University of Munich, Garching, Germany. ⁷Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany. ⁸Present address: Carle Illinois College of Medicine, Champaign, IL, USA. ⁹These authors contributed equally: Joanna Liwocha, David T. Krist, Gerbrand J. van der Heden van Noort. ¹⁰Deceased: Huib Ovaa.

Published in *Nature Chemical Biology* (2021)

Protein ubiquitination is an immensely complex PTM that is not limited to the mere addition of single ubiquitin molecules to a substrate protein. Substrate-bound ubiquitin itself can be subject to further ubiquitination resulting in the formation of ubiquitin chains. Two ubiquitin molecules may be linked through one of the seven epsilon amines of the lysine sidechains or the N-terminal alpha amine. The 'ubiquitin code' describes the various architectures of resulting polyubiquitin chains and their different functionalities (see also section 1.3.1.2). In this study, the Schulman group evaluated how the reactive amine geometry affects the ubiquitin code. By generating synthetic ubiquitin molecules with defined, non-natural lysine sidechain length, they demonstrated that the aliphatic side chain is a determinant of the ubiquitin code.

For this study, I developed a targeted SIM assay for the absolute quantification of diubiquitin chain linkages. Specific diubiquitin linkage types can be identified by characteristic peptides that display a missed cleavage site after modified diGly remnant carrying lysins. The absolute quantification of these characteristic peptides was enabled by spiking in synthetic stable isotope labeled versions of these linkage type-specific peptides.



Linkage-specific ubiquitin chain formation depends on a lysine hydrocarbon ruler

Joanna Liwocha^{1,9}, David T. Krist^{1,8,9}, Gerbrand J. van der Heden van Noort^{2,9}, Fynn M. Hansen³, Vinh H. Truong⁴, Ozge Karayel³, Nicholas Purser⁵, Daniel Houston⁵, Nicole Burton⁵, Mark J. Bostock^{6,7}, Michael Sattler^{6,7}, Matthias Mann^{6,3}, Joseph S. Harrison⁴, Gary Kleiger⁵✉, Huib Ovaa^{2,10}✉ and Brenda A. Schulman¹✉

Virtually all aspects of cell biology are regulated by a ubiquitin code where distinct ubiquitin chain architectures guide the binding events and itineraries of modified substrates. Various combinations of E2 and E3 enzymes accomplish chain formation by forging isopeptide bonds between the C terminus of their transiently linked donor ubiquitin and a specific nucleophilic amino acid on the acceptor ubiquitin, yet it is unknown whether the fundamental feature of most acceptors—the lysine side chain—affects catalysis. Here, use of synthetic ubiquitins with non-natural acceptor site replacements reveals that the aliphatic side chain specifying reactive amine geometry is a determinant of the ubiquitin code, through unanticipated and complex reliance of many distinct ubiquitin-carrying enzymes on a canonical acceptor lysine.

Ubiquitin (UB) chains are a major post-translational modification controlling protein function in eukaryotic cells. Eight distinct chain types are formed from linkage of the C terminus of one UB to an amino group acceptor (seven lysines and the N terminus) on another UB. The different UB chains form a ‘UB code’ that is read by cognate binding domains, which control the fates of modified proteins^{1–4}. Studies of endogenous and recombinant proteins have shed light on this code, showing that K48-linked chains often direct proteasomal degradation, while K63-linked chains mediate diverse regulation by modulation of multi-subunit complex assembly^{1–4}. Structural studies have shown how specificity is determined by the distinct spacing between hydrophobic patches presented by UB molecules linked in various chain types^{1–4}. In some cases, the actual isopeptide linkages between the UBs, and the surrounding residues, also dictate recognition of specific UB chains.

Recently, chemical biology approaches have elucidated the principles governing important aspects of UB biology⁵. Indeed, synthetic UB chains with defined linkages and chemically unique properties have highlighted the mechanisms underlying protein degradation by the proteasome^{6,7} and revealed the potential of hundreds of UB-binding domains to partner with their cognate chain types⁸.

Despite this progress in deciphering how the code is ‘read’ by the downstream machineries that recognize UB chains, the mechanisms underlying the generation of specific UB chain linkages remain incompletely understood. Chains are forged by combinations of UB-conjugating enzymes (E2s) and UB ligases (E3s). In humans, various pairings among ~30 E2s and ~600 E3s mediate UB ligation to selected target proteins and determine the generation of UB chains with specific linkages. Different E2s and E3s employ distinct enzymatic mechanisms to achieve polyubiquitylation^{9,10}. Some

E2 enzymes can generate chains themselves, whereby, after enzymatic linkage of the C terminus of a UB to an E2 catalytic cysteine, UB is transferred from the resultant E2~UB intermediate (where ~ refers to a thioester bond) to a lysine on an ‘acceptor UB’. The preferred UB acceptor lysine may be intrinsic to an E2 and/or may be influenced by an E2 partner protein¹¹. In some cases, UB transfer from the E2 is stimulated by the hallmark ‘Really Interesting New Gene’ (RING) domain in many E3s. An E2 may also transfer UB to an active-site cysteine of some E3s, as in a ‘Homologous to E6AP C-terminus’ (HECT) catalytic domain, from which the donor UB is linked to an acceptor UB to generate a chain.

Previous studies have identified E2 or E3 residues that are critical for catalysis and present an acceptor UB to the active site, as well as the roles of acceptor UB residues surrounding the targeted lysine^{11–17}. However, whether features of a UB’s target lysine beyond its nucleophilic primary amino group—such as the distance between the primary amine and the UB polypeptide backbone—influence UB chain formation remains unknown. Within classes of UB-carrying enzymes (for example, E2 or HECT E3), catalytic domains adopt similar structures that have the capacity to catalyze covalent bond formation between the donor UB and assorted free amino-acid acceptors (lysine, cysteine, serine and threonine)^{9,10,18,19}. Because the substrates of the UB system are often degraded even with mutation of the preferred lysines, it seems that targeting by some E2 and E3 enzymes is relatively lax. This contrasts with protein interaction domains or histone-modifying enzymes, which strictly depend on lysines for specific salt-bridge geometries or substrate targeting^{20,21}.

To investigate whether acceptor lysine side-chain features beyond the primary amino group influence UB chain formation,

¹Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, Martinsried, Germany. ²Oncode Institute and Department of Cell and Chemical Biology, Chemical Immunology, Leiden University Medical Centre, Leiden, the Netherlands. ³Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. ⁴Department of Chemistry, University of the Pacific, Stockton, CA, USA. ⁵Department of Chemistry and Biochemistry, University of Nevada, Las Vegas, Las Vegas, NV, USA. ⁶Biomolecular NMR and Center for Integrated Protein Science Munich at Department Chemie, Technical University of Munich, Garching, Germany. ⁷Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany. ⁸Present address: Carle Illinois College of Medicine, Champaign, IL, USA. ⁹These authors contributed equally: Joanna Liwocha, David T. Krist, Gerbrand J. van der Heden van Noort. ¹⁰Deceased: Huib Ovaa. ✉e-mail: gary.kleiger@unlv.edu; h.ovaa@lumc.nl; schulman@biochem.mpg.de

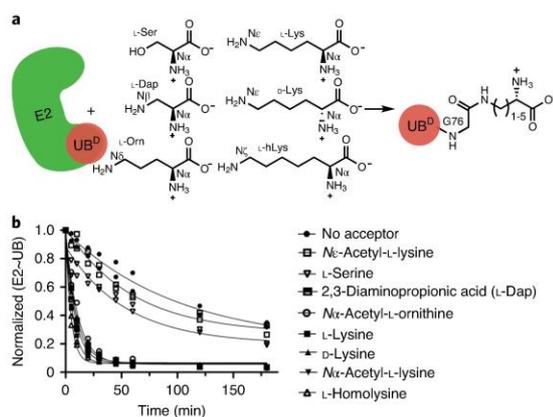


Fig. 1 | UBE2N-UB/UBE2V1/RNF4 RING E3 complex reacts preferentially with free amino acids harboring amine acceptors and various side-chain hydrocarbon linkers. **a**, Cartoon of the experimental scheme, monitoring the reactivity of E2-UB (D refers to the ‘donor’ fluorescent UB to be transferred from E2) towards various free amino acids. **b**, Time-course of fluorescent UB discharge from UBE2N-UB/UBE2V1/RNF4 RING E3 to the indicated amino acids, normalized to the starting signal of fluorescent UB thioester-bonded to UBE2N. $N = 2$ independent experiments. For samples derived from the same experiment, gels were processed in parallel.

we employed a suite of synthetic UBs harboring replacements for K11, K48 or K63 with shorter or longer aliphatic side chains, and tested their reactivities with a broad set of ubiquitylating enzymes. Our results demonstrate that the geometry between the polypeptide backbone and primary amine strongly influences chain formation for diverse polyubiquitylating enzymes. Thus, the lysine side chain itself helps establish the UB code.

Results

Acceptor UB lysine geometry required by K63-specific E2. The simplest activity of an E2 involves UB transfer to a nucleophilic amino acid, free in solution. For some E2s, such discharge onto an isolated amino-acid acceptor (for example, lysine, cysteine and threonine) correlates with a preferred residue type modified in the context of a protein target^{18,19}. We examined the reactivity of the well-characterized K63-linked UB chain forming enzyme, the human heterodimeric E2 UBE2N/UBE2V1 complex, which uniquely partners a canonical E2 subunit (UBE2N) with the dedicated catalytically inactive E2-like UBE2V1^{16,22,23}. UBE2V1 guides K63 (on an acceptor UB) towards the thioester linkage between the active site cysteine of UBE2N and the C terminus of the donor UB^{13,22}. The rate of formation of this K63-linked UB chain is accelerated by the RING domain of RNF4 E3 (hereafter referred to as RNF4). RNF4 stabilizes the active conformation of the donor UB thioester-bonded to the UBE2N active site²⁴. Moreover, coupling with UBE2V1 and RNF4 stimulates the intrinsic reactivity of the UBE2N-UB intermediate as monitored by UB discharge to free lysine²³, albeit less efficiently than to an acceptor UB’s K63.

We examined transfer of the donor UB from RNF4-UBE2V1-activated UBE2N to various free amino acids using a pulse-chase assay (Fig. 1a). UBE2N was charged with fluorescent donor UB in the pulse reaction using E1 enzyme. After quenching this reaction, the resultant UBE2N-UB intermediate was incubated with RNF4, UBE2V1 and an amino acid. We initially tested L-lysine (four methylene units in the side chain, referred to here

as C₄) and two controls: L-serine, not known to accept UB from RNF4-UBE2V1-UBE2N, and Nε-acetyl-L-lysine, with a blocked ε amino group. As expected, L-lysine had high reactivity compared to controls (Fig. 1b). Reactivity of Nε-acetyl-L-lysine, with a blocked α amino but available ε amino group, verified lysine’s Nε-amine as the preferred acceptor. With this established, we tested lysine analogs differing in side-chain length. C₁, C₃ and C₅ analogs (L-2,3-diaminopropionic acid, Nα-acetyl-L-ornithine and L-homolysine, respectively) demonstrated robust reactivity (Fig. 1b and Extended Data Fig. 1a,b), indicating a lack of an absolute requirement for aliphatic chain length between the backbone and nucleophilic amino group of lysine analogs free in solution.

We next wondered how the L-lysine architecture within the context of an acceptor UB would affect UBE2N/UBE2V1 reactivity (Fig. 2a). Solid-phase peptide synthesis was used to generate five UBs with K63 analogs differing by the number of methylene groups—one, two, three, four or five—between the α carbon and the side-chain amino group: L-2,3-diaminopropionic acid (Dap, referred to here as ^{K63}UB_{C1} for one methylene group in the analog replacing native K63), L-2,4-diaminobutyric acid (Dab, referred to here as ^{K63}UB_{C2} for two methylene groups in the analog replacing K63), L-ornithine (Orn, referred to here as ^{K63}UB_{C3} for three methylene groups in the analog replacing K63), L-lysine (Lys, referred to here as ^{K63}UB_{C4} for four methylene groups in the native acceptor) and L-homolysine (hLys, referred to here as ^{K63}UB_{C5} for five methylene groups in the analog replacing K63) (Fig. 2b).

UBE2N/UBE2V1 activity was again measured using a pulse-chase assay, with the acceptor now being UB and the product a di-UB chain. Remarkably, unlike in the discharge to free amino acids, removal or addition of only a single methylene from or onto a canonical K63 side chain greatly reduced di-UB chain formation. The striking preference for the native lysine persisted in reactions accelerated by the RNF4 E3 (Fig. 2c).

Lysine geometry impacts many di-UB forming E2s and E3s.

Because acceptor UB placement for UBE2N is unique in depending on a partner (UBE2V1)²³, UBE2V1’s grip may limit the ability of the reactive amine to reposition in the active site upon addition or removal of a methylene. We thus wondered how changes to the lysine architecture affect other E2s that are reliant on their own surfaces to orient an acceptor UB. Accordingly, we assayed the activities of two K48 linkage-specific E2s, UBE2G1 and UBE2R2, towards a ^{K48}UB_{C1-C5} suite²⁵⁻²⁷. Significant di-UB product was only observed with the ^{K48}UB_{C4} acceptor—for the E2s alone and for UBE2R2 and UBE2G1 reactions stimulated by cullin-RING ligase E3s CRL1 or CRL4, respectively²⁵⁻²⁹, and for substrate-linked acceptors (Fig. 2d). These latter assays depended on CRL receptors recruiting specific substrate degron motifs. The CRL1 receptor FBW7, a tumor suppressor protein, recruits phosphopeptide motifs in targets including the cell cycle regulator cyclin E³⁰. For the CRL4 receptor CRBN, the chemotherapeutic agent Pomalidomide induces recognition of zinc finger motifs in neosubstrates including Ikaros family transcription factors^{31,32}. CRL1^{FBW7} and CRL4^{CRBN} substrates were generated by sortase-mediated transpeptidation of degron peptides (cyclin E phosphopeptide and IKZF1 zinc finger, respectively) with synthetic UBs. Only native lysine supported substantial UB-chain elongation onto CRL-bound substrates (Fig. 2d).

To determine if the preference for native lysine is preserved for HECT E3 ligases—where UB is transferred from E2 to the HECT catalytic cysteine and then onto the substrate lysine—we assayed the NEDD4 HECT domain³³ and a version of its yeast ortholog Rsp5p harboring substrate-binding WW and catalytic domains³⁴. Both forge K63-linked chains³³⁻³⁵. Again, robust di-UB formation was only observed with native lysine acceptor ^{K63}UB_{C4}. Di-UB formation was greatly reduced with ^{K63}UB_{C1-C3} or ^{K63}UB_{C5}, including for a substrate³⁶ recruited to Rsp5p (Fig. 3a,c).

ARTICLES NATURE CHEMICAL BIOLOGY

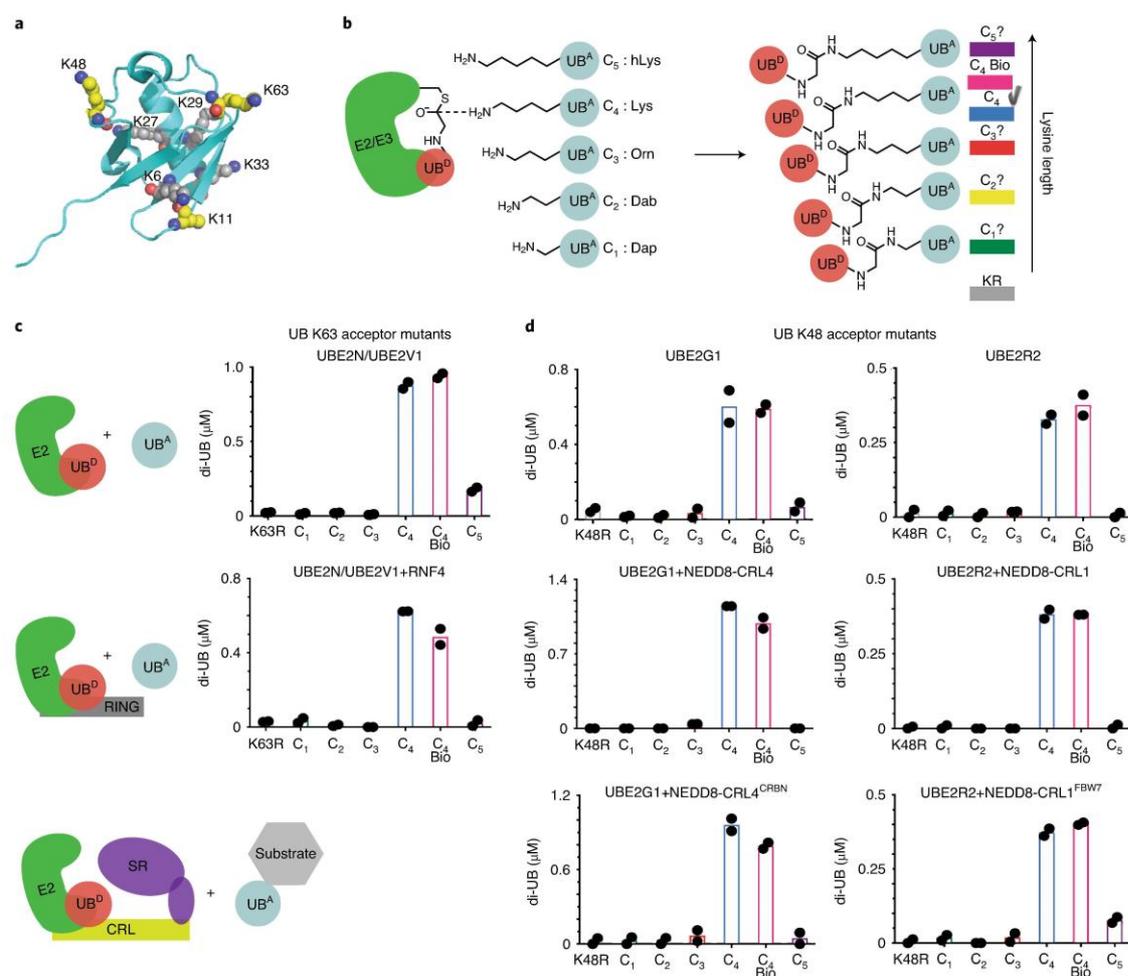


Fig. 2 | K48 and K63 chain-forming E2s show strong preferences for a native lysine acceptor on UB. **a**, Structure of UB (PDB: 3CMM), showing lysines as spheres and carbon (yellow) and nitrogen (blue) atoms highlighted for K11, K48 and K63 where analogs were installed for this study. **b**, Cartoon of the experimental scheme, monitoring the reactivity of E2-UB (D refers to the fluorescent donor UB to be transferred from E2) and formation of di-UB with various versions of acceptor UB (UB^A). Color coding is used for acceptor lysine analogs denoted as C₁ to C₅ based on having one to five side-chain methylene groups, respectively. KR represents lysine mutation to arginine. **c**, Plots showing the amount of di-UB chain produced by UBE2N/UBE2V1 with UBs harboring the indicated acceptor side chain at position 63 in the absence (top) or presence (bottom) of the E3 RING domain from RNF4. **d**, Plots showing the amount of di-UB chain produced by E2s UBE2G1 and UBE2R2 in the absence (top) or presence (middle) of cognate E3s—NEDD8-CRL4 or NEDD8-CRL1—that activate di-UB synthesis. E2-dependent di-UB forming activity toward E3-bound substrates was tested with substrates (sortase-mediated ^{K48}UB_{C1-C5} linked to phospho-cyclin E peptide or IKZF1 ZF 2-3) of neddylated CRL1^{FBW7} or CRL4^{CRBN} (bottom). For all plots, di-UB levels (μM) represent the final time points from the reactions (Source Data Fig. 2), *N* = 2 independent experiments. For samples derived from the same experiment, gels were processed in parallel.

As controls, the ^{K63}UB_{C1-C5} analogs served as acceptors with the K48-specific E2 UBE2G1, demonstrating proper folding for the synthetic UBs harboring K63 substitutions (Extended Data Fig. 2a). Similarly, UBE2N/UBE2V1, NEDD4 and Rsp5p also produce nearly wild-type amounts of di-UB chains with UBs harboring lysine analogs on the non-acceptor position 48 (^{K48}UB_{C1-C5}, Fig. 3b and Extended Data Fig. 2b). Moreover, proton NMR spectra for recombinant UB ('C₄ Bio'), synthetic UB (aka 'C₄') and ^{K48}UB_{C5} showed good dispersion and were superimposable except for a

few resonances, presumably reflecting the overall minor impact of sequence differences between them (Met1 in C₄ Bio substituted with NorLeu in synthetic UBs, and Lys48 versus the C₅ side chain, Extended Data Fig. 3).

Taken together, the data show that K63- and K48-specific E2 and E3 enzymes utilizing distinct modes of acceptor UB recruitment display exquisite specificity for the attacking lysine architecture in the context of an acceptor UB. Notably, E2~UB and HECT E3~UB active sites are structurally distinct. Thus, the demand for native

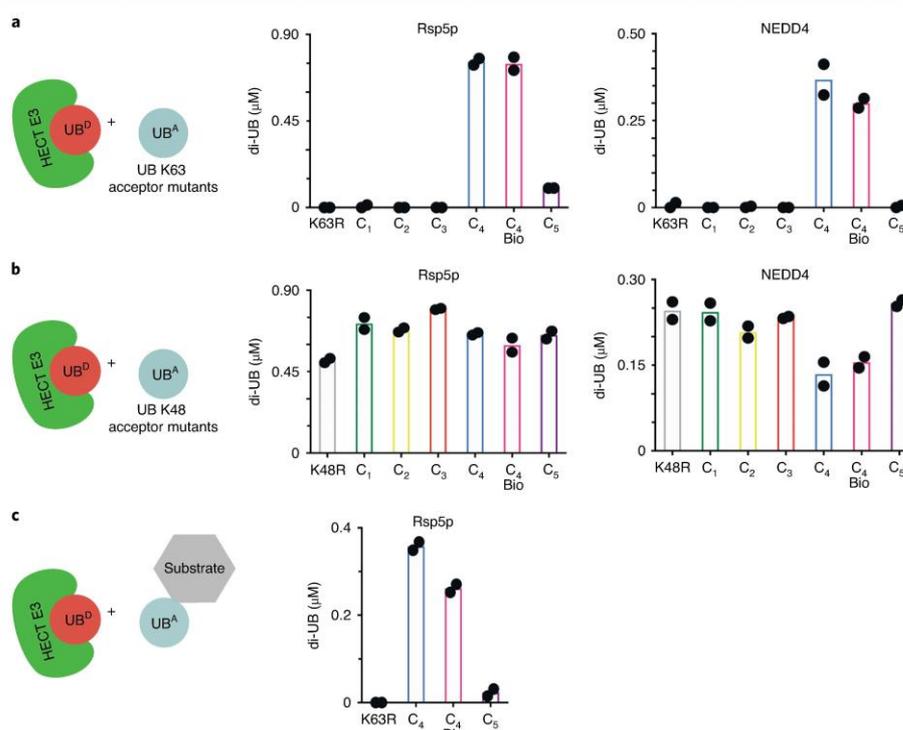


Fig. 3 | K63 chain-forming HECT E3 ligases show strong preferences for a native lysine acceptor on UB. **a**, Cartoon of the experimental scheme (left) and reactivity of the yeast HECT E3 Rsp5p (middle) or human HECT E3 NEDD4 (right) and formation of di-UB chains with $K^{63}UB_{C1-C5}$ acceptors (UB^A). **b**, Same as in **a**, except with $K^{48}UB_{C1-C5}$ acceptors. **c**, HECT E3 ligase-dependent di-UB forming activity in the context of an Rsp5p-bound substrate (sortase-mediated UB, UB K63R or $K^{63}UB_{C5}$ linkage to the WW-domain-binding PPPY degron motif of the substrate Sna4p). For all plots, di-UB levels (μM) represent the final time points from the reactions (Source Data Fig. 3), $N=2$ independent experiments. For samples derived from the same experiment, gels were processed in parallel.

lysine acceptor geometry for chain building seems to be a general property that could extend across many of the hundreds of E2/E3 ligation systems.

The K48 side chain impacts the multifunctional E2 UBE2D3. We pondered whether there may be exceptions to linkage-specific ubiquitylation relying on native lysine (C_4) acceptors. The E2 UBE2S, which generates K11 di-UB linkages, is an intriguing candidate, as UBE2S relies on acceptor UB-assisted catalysis¹⁴. UBE2S displays weak di-UB chain synthesis activity on its own due to the high K_m for the acceptor¹⁴. This is overcome by fusing a UB-binding domain to UBE2S, or with the anaphase-promoting complex/cyclosome (APC/C) E3, whose RING domain recruits the acceptor $UB^{37,38}$. Results from our qualitative assays suggest that UBE2S is less sensitive to lysine side-chain length, as di-UB formation occurred with $K^{11}UB_{C2-C5}$ acceptors, although $K^{11}UB_{C1}$ was relatively inactive (Extended Data Fig. 4). This is not due to an overt folding defect, as all $K^{11}UB_{C1-C5}$ analogs are acceptors for UBE2N/UBE2V1-dependent K63-linked di-UB formation.

We also examined the relatively promiscuous UB chain-forming enzyme UBE2D3 (also known as UBCH5C): UBE2D3 collaborates with numerous E3s, transfers UB to E3 catalytic cysteines and substrate lysines, generates several UB chain linkages, and forms branched UB chains in multiple turnover polyubiquitylation reactions^{39,40}. Pulse-chase assays examining di-UB products

of UBE2D3~UB revealed preferential targeting to K11 and K63, according to absolute quantitation by MS (Fig. 4a). However, with a $K^{48}UB_{C5}$ acceptor added to UBE2D3~UB, the SDS-PAGE mobilities of di-UB products differed from those formed with a native UB acceptor. Because different UB chain linkages could impact electrophoretic migration, the result hints at distinct products (Fig. 4b).

We developed a targeted MS strategy to quantify the distributions of UB chain linkages formed with native lysines. Although the method does not detect chains linked to the unnatural amino acid, it quantifies relative UB linkages to the remaining lysines in reactions with $K^{11}UB_{C5}$, $K^{48}UB_{C5}$ or $K^{63}UB_{C5}$ compared to reactions with the C_4 acceptor UB. With $K^{11}UB_{C5}$ or $K^{63}UB_{C5}$ acceptors, UBE2D3~UB generates di-UBs with linkage-type distributions similar to reactions with UB_{C4} (Fig. 4c,d and Extended Data Fig. 5). However, adding an extra methylene group to the side chain at K48 alters the distribution of di-UB species formed. The change between preferred acceptors could be accounted for by two observations. First, there is a relative redistribution from K63 to K11 linkages. Second, although the di-UBs linked via K27, K29 and K33 remain a minor proportion of the total, utilization of these non-preferred acceptors increased compared to UB_{C4} . Thus, the location on UB is a determinant of the requirement for a UB_{C4} by a multifunctional ubiquitylating enzyme.

Impact of side-chain architecture revealed by MD. The potential structural effects of adding a methylene group to the acceptor side

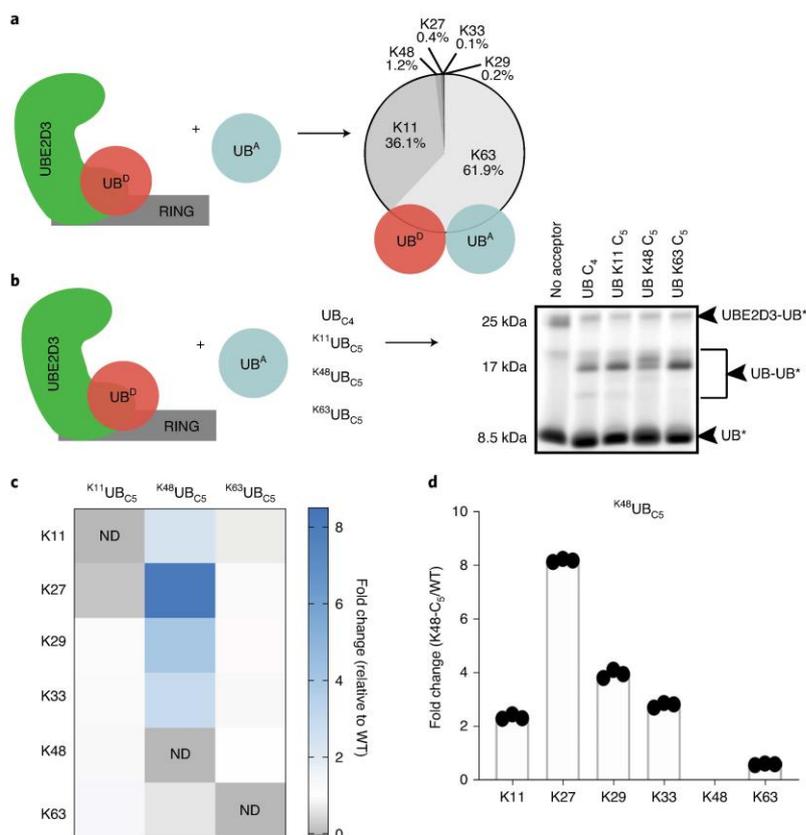


Fig. 4 | The location of lysine analogs on acceptor UB impacts the distribution of di-UB chain linkage types generated by the E2 enzyme UBE2D3.

a, Cartoon of the experimental scheme and reactivity of UBE2D3-UB (D refers to the donor UB to be transferred from E2) in the presence of the E3 RING domain from RNF4 and formation of di-UB with UB_{C_4} acceptor (UB^A). The distribution of the generated di-UB linkage types is shown. **b**, Same as in **a**, except with fluorescent donor UB and various acceptor UBs (UB^A). Notice that di-UB products with distinct electrophoretic mobilities were observed for each acceptor, indicating that the lysine analogs probably affect the di-UB chain linkage identity ($N=2$ independent experiments). **c**, Relative fold changes of di-UB linkage types for reactions with UBE2D3 and the E3 RING domain from RNF4 comparing $K^{11}UB_{C_5}$, $K^{48}UB_{C_5}$ or $K^{63}UB_{C_5}$ acceptors with UB_{C_4} . ND, not defined. **d**, Plot showing the relative changes in UBE2D3/RNF4-generated di-UB chain linkages when comparing products containing $K^{48}UB_{C_5}$ or UB_{C_4} acceptors. For **c** and **d**, $N=3$ technical replicates.

chain were revealed by MD simulations on native UB_{C_4} or UB_{C_5} at position 11, 48 or 63. In two independent 50-ns simulations for UBs with native lysine or C_5 at positions 11, 48 and 63, the overall UB globular fold (residues 1–70) was preserved, with 1.618, 1.271, 1.209 and 1.494 Å average C α root mean square deviation (RMSD) across the simulations, respectively. Nonetheless, the relative differences for C_5 —at all three sites—include (1) an increased potential range of distances between the α carbon and side-chain amine for C_5 , with limited overlap in the distribution of relative side-chain amine position (Fig. 5a)—this would effectively impart a greater radius to the C_5 side chain when considering the backbone as the axis of rotation; (2) an expanded number of potential rotamers from 81 to 273, with more accessed by the C_5 side chain in every simulation ($C_4:C_5$ rotamer ratios for residues 11, 48 and 63 of 56:82, 65:117 and 43:96, respectively); (3) different dynamics for χ angles, particularly χ_4 , which oscillated more frequently between the three rotamer bins for C_5 —this would cause more rapid fluctuation of relative side-chain amine positions (Fig. 5b and Extended Data Fig. 6a).

With C_5 at positions 48 and 63, there were also subtle but reproducible increases in fluctuations in the ϕ and ψ angles (Fig. 5c,d), and an increased number of allowable ϕ/ψ combinations ($C_4:C_5$ ratios for residues 11, 48 and 63 of 185:175, 138:169 and 73:90, respectively). Collectively, between the backbone and rotamer combinations, we typically observed more states accessible to the C_5 residue, with a noticeable increase of over 1,000 additional states at positions 48 and 63 ($C_4:C_5$ ratios at positions 11, 48 and 63 of 2,942:3,016, 2,942:4,261 and 1,188:2,561, respectively).

We wished to further probe the potential effects of the C_5 side chain as an acceptor in di-UB chain formation. The only structurally characterized reaction is a donor UB–UBE2N/UBE2V1/acceptor UB complex, where the acceptor UB's K63 points towards, but is 12.5 Å from, the donor UB's carbonyl to which it becomes linked during di-UB synthesis²³. Nonetheless, we adapted an intermediate of the acceptor K63 based on modeling and on constraints from enzymology and the crystal structures of wild-type UBE2N/UBE2V1^{41,42}. Three independent MD simulations (25 ns each)

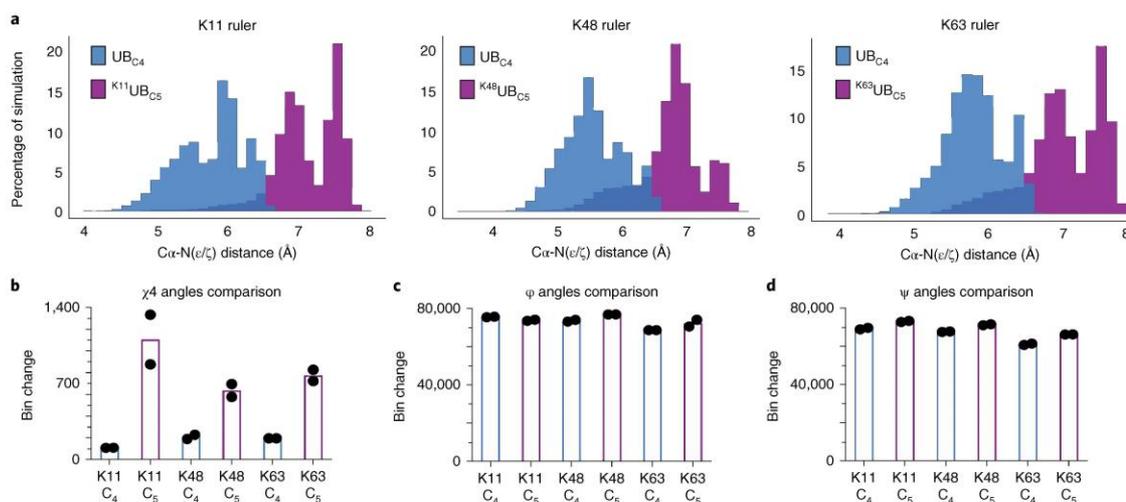


Fig. 5 | Molecular dynamics (MD) simulations reveal pleiotropic structural effects on UBs harboring lysine analogs. a, Distribution of the distances between lysine acceptor amine and C α atoms for UB_{C4} versus UB_{C5} in MD simulations performed for UB. **b**, Plot showing the dynamics of γ^4 side-chain rotomers for various UB_{C5} acceptors compared with UB_{C4}. Bins are divided by 120° intervals. **c**, Same as in **b**, except for the ψ main-chain torsion angle. Bins are divided by 10° intervals. **d**, Same as in **b**, except for the ϕ main-chain torsion angle. For all plots, $N=2$ independent experiments.

showed the C₅ acceptor side chain preferentially adopting extended conformers, and more frequently fluctuating between rotamers, as in the simulations of UB alone. Although both C₄ and C₅ side-chain amines maintained a similar distance to the UBE2N~UB active site (Extended Data Fig. 6b–d), closer inspection revealed two appreciable differences between the simulations: (1) lysine occupied a favorable trajectory towards the active site for a greater proportion of the simulations, whereas C₅ more frequently rotated between rotamer bins and approached the active site from different angles (Extended Data Fig. 6e,f); (2) there was a greater deviation in the conformation of UBE2N's so-called 'active site gate loop' (residues 115–120). Interestingly, this gate loop is important for stabilizing noncovalent interactions between the donor UB tail and UBE2N, configuring catalytic residues and positioning the acceptor lysine relative to the thioester bond for catalysis^{41,43}. Distortion of the gate loop conformation, as observed with the C₅ side chain, could reduce the probability of adopting a structure favoring ligation (Extended Data Fig. 6g,h).

Impact of UB acceptor lysine geometry on kinetic parameters. To illuminate mechanistic roles for lysine, quantitative biochemical experiments were performed. Substantially increasing reaction times and protein levels under steady-state conditions enabled quantification for C₅ as acceptor for di-UB formation by the E2s UBE2N/UBE2V1 (with or without RNF4 E3), UBE2R2 and by the HECT E3 Rsp5p.

The reactions with both E2s showed similar profiles overall: k_{cat} values were lower with C₅ replacements for acceptor lysines—16-fold and 14-fold, respectively (Table 1)—consistent with the striking results from the pulse-chase assays (Fig. 2). Although defects in enzyme activity can manifest themselves through various perturbations, failure to activate the acceptor lysine amine or decreasing affinity of the acceptor UB for the E2 are quite common^{13,14,44}. A pioneering investigation of the related modification SUMOylation suggested that E2s catalyze ubiquitylation at least in part through the active site complementing the acceptor lysine to

achieve pK_a suppression⁴⁴. Despite being unable to estimate apparent pK_a for reactions with E3s due to a loss of enzyme activity at high pH, we were able to determine apparent pK_a values in the reactions with E2s.

UBE2N/UBE2V1 activity (with a K92R mutation to decrease auto-ubiquitylation at high pH²³) was measured in the presence of ^{K63}UB_{C4} or ^{K63}UB_{C5} across varying pHs (Table 1 and Extended Data Fig. 7a,b). Although caution should be taken when interpreting apparent pK_a values, because both k_{cat} and K_{M} may display pH dependencies of their own, the data fit best to a model where a single ionizing species is responsible for the pH dependency of k_{obs} (Extended Data Fig. 7b). Surprisingly, pK_a values were similar in reactions with ^{K63}UB_{C4} or ^{K63}UB_{C5} (8.9 and 9.0, respectively; Table 1). Parallel experiments with UBE2R2 showed apparent pK_a values of 6.6 and 7.3 for ^{K48}UB_{C4} and ^{K48}UB_{C5}, respectively (Table 1 and Extended Data Fig. 7e). For both E2s, differences in pK_a values are insufficient to account for those between rates of di-UB formation with acceptor lysine or C₅ side chains in reactions at elevated pH (nearly 100-fold for UBE2R2 at pH 9.7, Table 1). The estimated K_{M} values of ^{K63}UB_{C4} or ^{K63}UB_{C5} for UBE2N/UBE2V1 were within two-fold, and those of ^{K48}UB_{C4} or ^{K48}UB_{C5} for UBE2R2 within four-fold, suggesting similar affinities for lysine- and C₅-bearing acceptor UBs and their respective E2s (Table 1 and Extended Data Fig. 7c,f). Thus, defective catalysis seemingly arises from other effects of the additional methylene in the acceptor UB side chain.

An E3 may affect mechanisms underlying acceptor UB lysine specificity. Although the RNF4 RING domain greatly impacted UBE2N/UBE2V1-catalyzed di-UB formation (lowering the K_{M} of acceptor UB for E2 and increasing k_{cat} by ~17-fold and 11-fold, respectively (Table 1 and Extended Data Fig. 7d)), only modest effects were observed for K_{M} (~2.5 fold) as well as k_{cat} (~4 fold) in the presence of ^{K63}UB_{C5}. In combination, these effects are not greater than those observed without E3.

By contrast, kinetic experiments performed on the HECT E3 Rsp5p showed a remarkable 16-fold lower K_{M} for the acceptor ^{K63}UB_{C4} compared with ^{K63}UB_{C5}, with only a ~2.5-fold difference in

Table 1 | k_{cat} , $\text{p}K_{\text{a}}^{\text{app}}$ and K_{M} for UB-carrying enzymes with native versus homolysine acceptor UBs

E2/E3	UB	Lys	$\text{p}K_{\text{a}}^{\text{app}}$	$k_{\text{obs}} (\text{h}^{-1})$, top pH	K_{M} (10^{-6} M)	k_{cat} (h^{-1})
UBE2N/V1	C ₄ Bio	K63			190	6.1
UBE2N/V1	C ₄	K63	8.9	15.8	398	3.4
UBE2N/V1	C ₅	K63	9.0	0.58	284	0.21
UBE2N/ V1 + RNF4	C ₄	K63			23	39.1
UBE2N/ V1 + RNF4	C ₅	K63			58	9.3
UBE2R2	C ₄	K48	6.6	2.67	528	15.8
UBE2R2	C ₅	K48	7.3	0.028	1,940	1.1
Rsp5p	C ₄	K63			21	1.11
Rsp5p	C ₅	K63			335	0.44

Kinetic parameters for several UB-carrying enzymes, including the apparent $\text{p}K_{\text{a}}$ ($\text{p}K_{\text{a}}^{\text{app}}$), the rate of di-UB formation (k_{obs}) for UBE2N/UBE2V1 and UBE2R2 at pH10.1 or 9.7, respectively (k_{obs} , (h^{-1}) top pH), and the K_{M} and k_{cat} of UB_{C4} or UB_{C5} acceptors for E2 or HECT E3. Each value represents the mean of duplicate data points (Source data).

k_{cat} (Table 1 and Extended Data Fig. 7g). Overall, the kinetic results unveil a diverse spectrum of effects of the lysine side-chain ruler on UB-carrying enzyme activities.

Discussion

Our data show that many different UB chain-forming enzymes are strikingly sensitive to the lysine side chain hydrocarbon linker at the ångström length scale, as determined by a single methylene. Biochemical assays show that UB_{C5} can affect K_{M} , k_{cat} and $\text{p}K_{\text{a}}$ (Table 1). Meanwhile, MD simulations unveiled pleiotropic structural effects of C₅, including additional degrees of freedom, more side-chain flexibility and more dynamics in the backbone in UB itself (Fig. 5). It might stand to reason that side chains that are too short simply cannot span the distance between the acceptor UB backbone and UB-carrying enzyme active site. However, the fact that the UB_{C5} analogs impacted most tested enzymes indicates further roles of the acceptor side chain.

For both UBE2N/UBE2V1 and UBE2R2, the mild effects on apparent $\text{p}K_{\text{a}}$ and/or K_{M} are insufficient to explain the defects in k_{cat} observed upon acceptor lysine substitution with C₅ (Table 1). The MD simulations pointed towards several possible features of the lysine side-chain length that may be optimal for E2-catalyzed UB chain formation. For example, for enzymes where substrate binding and/or lysine positioning are rate-limiting, it seems that the increased entropy afforded by an extra methylene in the acceptor side chain could decrease the frequency of catalytic encounter (Fig. 5). Interestingly, this mechanism would differ from that of another E2, UBE2W, for which a confluence of disorder between a flexible substrate N terminus and a non-canonical E2 C terminus guides ubiquitylation to a substrate's N-terminal amine¹⁵. Rather than demanding disorder, the systems tested herein appear to favor a calibrated reach by the nucleophile that also must have restrained degrees of freedom.

In addition to entropic effects on the side chain, the hydrocarbon linker length would also affect catalysis. For example, as shown for UBE2N (without UBE2V1-RING E3 partners), computational studies support a model where there is a precise 'hole' fitting the lysine amine, and attack on the thioester carbonyl is rate-limiting¹⁶. Our data may suggest that the acceptor UB lysine itself is optimal not only for accessing the amine hole, but also for the chemistry of ubiquitylation. Indeed, the MD simulations of the UB-UBE2N/

UBE2V1/UB complex point to multiple ways the acceptor lysine side-chain length could impact catalysis, including through an optimal geometric approach to the active site, and through conformationally toggling the active-site gate loop in the UB-E2 intermediate. Moreover, in agreement with previous studies suggesting this loop in UBE2N essentially closes around the acceptor K63 to promote formation of the transition state⁴³, our MD simulations showed distortion of the active-site gate loop with the suboptimal C₅ side chain. This would be consistent with UB discharge to free side-chain amine acceptors irrespective of hydrocarbon length, and a dramatic impact on k_{cat} in the context of acceptor UB presented from UBE2V1. One would also predict little impact on K_{M} in such a case, although lack of an effect on K_{M} may also reflect that the additional methylene does not impact acceptor UB recruitment to this auxiliary UB-binding domain.

The impact of acceptor side-chain length on the HECT E3 Rsp5p represents the opposite extreme. The predominant effect on K_{M} implies a role of the acceptor lysine itself in productive binding to the E3. It is possible that local interactions—awaiting elucidation by future structural studies—dominate acceptor UB recruitment¹⁷. It is also possible that placement of the acceptor lysine in the active site allosterically stabilizes the enzyme-UB conformation that binds the acceptor¹⁸.

Although our study relied on installing side-chain chemical variants, it seems likely that, in the cellular milieu, many natural factors—including linkage within a chain and binding to protein partners—could influence presentation of acceptor lysines resulting in specificity with E2 and E3 enzymes. Strong preferences for the lysine side chain itself may contribute to robust ubiquitylation sufficient to elicit proteasomal degradation, even when preferred targeting sites are unavailable. Such features may also influence the successes or failures of targeted protein degradation strategies that rely on small molecules to direct proteins of therapeutic interest to ubiquitylating enzymes^{49,50}.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-020-00696-0>.

Received: 2 April 2020; Accepted: 14 October 2020;

Published online: 7 December 2020

References

- Dikic, I., Wakatsuki, S. & Walters, K. J. Ubiquitin-binding domains—from structures to functions. *Nat. Rev. Mol. Cell Biol.* **10**, 659–671 (2009).
- Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).
- Yau, R. & Rape, M. The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **18**, 579–586 (2016).
- Kwon, Y. T. & Ciechanover, A. The ubiquitin code in the ubiquitin-proteasome system and autophagy. *Trends Biochem. Sci.* **42**, 873–886 (2017).
- Mulder, M. P. C., Witting, K. F. & Ovaia, H. Cracking the ubiquitin code: the ubiquitin toolbox. *Curr. Issues Mol. Biol.* **37**, 1–20 (2020).
- Singh, S. K. et al. Synthetic uncleavable ubiquitinated proteins dissect proteasome deubiquitination and degradation, and highlight distinctive fate of tetraubiquitin. *J. Am. Chem. Soc.* **138**, 16004–16015 (2016).
- Sun, H. et al. Diverse fate of ubiquitin chain moieties: the proximal is degraded with the target, and the distal protects the proximal from removal and recycles. *Proc. Natl Acad. Sci. USA* **116**, 7805–7812 (2019).
- Zhang, X. et al. An interaction landscape of ubiquitin signaling. *Mol. Cell* **65**, 941–955 (2017).
- Buetow, L. & Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **17**, 626–642 (2016).
- Zheng, N. & Shabek, N. Ubiquitin ligases: structure, function and regulation. *Annu. Rev. Biochem.* **86**, 129–157 (2017).

11. Mattioli, F. & Sixma, T. K. Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nat. Struct. Mol. Biol.* **21**, 308–316 (2014).
12. Bernier-Villamor, V., Sampson, D. A., Matunis, M. J. & Lima, C. D. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* **108**, 345–356 (2002).
13. Eddins, M. J., Carlile, C. M., Gomez, K. M., Pickart, C. M. & Wolberger, C. Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nat. Struct. Mol. Biol.* **13**, 915–920 (2006).
14. Wickliffe, K. E., Lorenz, S., Wemmer, D. E., Kuriyan, J. & Rape, M. The mechanism of linkage-specific ubiquitin chain elongation by a single-subunit E2. *Cell* **144**, 769–781 (2011).
15. Smit, J. J. et al. The E3 ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension. *EMBO J.* **31**, 3833–3844 (2012).
16. Berndsen, C. E., Wiener, R., Yu, I. W., Ringel, A. E. & Wolberger, C. A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nat. Chem. Biol.* **9**, 154–156 (2013).
17. Stieglitz, B. et al. Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature* **503**, 422–426 (2013).
18. Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
19. Pao, K. C. et al. Activity-based E3 ligase profiling uncovers an E3 ligase with esterification activity. *Nature* **556**, 381–385 (2018).
20. Virdee, S., Macmillan, D. & Waksman, G. Semisynthetic Src SH2 domains demonstrate altered phosphopeptide specificity induced by incorporation of unnatural lysine derivatives. *Chem. Biol.* **17**, 274–284 (2010).
21. Temimi, A. H. K. A. et al. Lysine possesses the optimal chain length for histone lysine methyltransferase catalysis. *Sci. Rep.* **7**, 16148 (2017).
22. McKenna, S. et al. Noncovalent interaction between ubiquitin and the human DNA repair protein Mms2 is required for Ubc13-mediated polyubiquitination. *J. Biol. Chem.* **276**, 40120–40126 (2001).
23. Branigan, E., Plechanovová, A., Jaffray, E. G., Naismith, J. H. & Hay, R. T. Structural basis for the RING-catalyzed synthesis of K63-linked ubiquitin chains. *Nat. Struct. Mol. Biol.* **22**, 597–602 (2015).
24. Branigan, E., Carlos Penedo, J. & Hay, R. T. Ubiquitin transfer by a RING E3 ligase occurs from a closed E2-ubiquitin conformation. *Nat. Commun.* **11**, 2846 (2020).
25. Petroski, M. D. & Deshaies, R. J. Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* **123**, 1107–1120 (2005).
26. Choi, Y.-S. et al. Differential ubiquitin binding by the acidic loops of Ubc2g1 and Ubc2r1 enzymes distinguishes their Lys-48-ubiquitylation activities. *J. Biol. Chem.* **290**, 2251–2263 (2015).
27. Hill, S., Harrison, J. S., Lewis, S. M., Kuhlman, B. & Kleiger, G. Mechanism of Lysine 48 selectivity during polyubiquitin chain formation by the Ubc2R1/2 ubiquitin-conjugating enzyme. *Mol. Cell. Biol.* **36**, 1720–1732 (2016).
28. Lu, G. et al. UBE2G1 governs the destruction of cereblon neomorphic substrates. *eLife* **7**, e40958 (2018).
29. Hill, S. et al. Robust cullin-RING ligase function is established by a multiplicity of poly-ubiquitylation pathways. *eLife* **8**, e51163 (2019).
30. Skaar, J. R., Pagan, J. K. & Pagano, M. Mechanisms and function of substrate recruitment by F-box proteins. *Nat. Rev. Mol. Cell Biol.* **14**, 369–381 (2013).
31. Kronke, J. et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
32. Lu, G. et al. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
33. Maspero, E. et al. Structure of a ubiquitin-loaded HECT ligase reveals the molecular basis for catalytic priming. *Nat. Struct. Mol. Biol.* **20**, 696–701 (2013).
34. Kamadurai, H. B. et al. Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *eLife* **2**, e00828 (2013).
35. Kim, H. C. & Huibregtse, J. M. Polyubiquitination by HECT E3s and the determinants of chain type specificity. *Mol. Cell. Biol.* **29**, 3307–3318 (2009).
36. Gupta, R. et al. Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast. *Mol. Syst. Biol.* **3**, 116 (2007).
37. Bremm, A., Freund, S. M. V. & Komander, D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nat. Struct. Mol. Biol.* **17**, 939–947 (2010).
38. Brown, N. G. et al. Dual RING E3 architectures regulate multiubiquitination and ubiquitin chain elongation by APC/C. *Cell* **165**, 1440–1453 (2016).
39. Brzovic, P. S. & Klevit, R. E. Ubiquitin transfer from the E2 perspective: why is UbcH5 so promiscuous? *Cell Cycle* **5**, 2867–2873 (2006).
40. Swatek, K. N. et al. Insights into ubiquitin chain architecture using Ub-clipping. *Nature* **572**, 533–537 (2019).
41. Lee, B. L., Singh, A., Mark Glover, J. N., Hendzel, M. J. & Spyropoulos, L. Molecular basis for K63-linked ubiquitination processes in double-strand DNA break repair: a focus on kinetics and dynamics. *J. Mol. Biol.* **429**, 3409–3429 (2017).
42. Garg, P. et al. Structural and functional analysis of ubiquitin-based inhibitors that target the backsides of E2 enzymes. *J. Mol. Biol.* **432**, 952–966 (2020).
43. Rout, M. K. et al. Stochastic gate dynamics regulate the catalytic activity of ubiquitination enzymes. *J. Am. Chem. Soc.* **136**, 17446–17458 (2014).
44. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nat. Struct. Mol. Biol.* **13**, 491–499 (2006).
45. Vittal, V. et al. Intrinsic disorder drives N-terminal ubiquitination by Ube2w. *Nat. Chem. Biol.* **11**, 83–89 (2015).
46. Jones, W. M., Davis, A. G., Wilson, R. H., Elliott, K. L. & Sumner, I. A conserved asparagine in a ubiquitin-conjugating enzyme positions the substrate for nucleophilic attack. *J. Comput. Chem.* **40**, 1969–1977 (2019).
47. Wang, M., Cheng, D., Peng, J. & Pickart, C. M. Molecular determinants of polyubiquitin linkage selection by an HECT ubiquitin ligase. *EMBO J.* **25**, 1710–1719 (2006).
48. Scott, D. C. et al. Structure of a RING E3 trapped in action reveals ligation mechanism for the ubiquitin-like protein NEDD8. *Cell* **157**, 1671–1684 (2014).
49. Burslem, G. M. & Crews, C. M. Proteolysis-targeting chimeras as therapeutics and tools for biological discovery. *Cell* **181**, 102–114 (2020).
50. Verma, R., Mohl, D. & Deshaies, R. J. Harnessing the power of proteolysis for targeted protein inactivation. *Mol. Cell* **77**, 446–460 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Constructs, protein expression and purification. All expression constructs were prepared using standard molecular biology methods. Modifications to protein amino-acid sequences were accomplished using PCR and the Quikchange mutagenesis kit (Agilent). The human E2 constructs used in this study are Glutathione S-transferase (GST)-Tobacco Etch Virus Protease (TEV)-UBE2R2, GST-TEV-UBE2N, GST-TEV-UBE2N harboring a K92R mutation, GST-TEV-UBE2D2, GST-TEV-UBE2D3, His-GST-Ps3C-UBE2V1, UBE2G1-TEV-His and GST-TEV-UBE2S (1–196) fused with the human USP5/IsoT (residues 173–289) domain—here called UBE2S_IsoT⁷⁰. Human HECT E3 ligase NEDD4 was expressed as a GST-TEV-NEDD4 construct, and yeast HECT E3 Rsp5p containing residues 383–C (with WW-domain-binding PPPY degen motif of the substrate Sna4p) was expressed as a GST-TEV-Rsp5p construct³¹. All E2s, both HECT E3s and His-sortase A were expressed in BL21-Gold(DE3) bacterial cells. Proteins were purified by either GST or nickel affinity chromatography and cleaved on beads overnight with TEV or 3C protease. Cleaved protein solutions were then subjected to ion exchange chromatography followed by size exclusion chromatography in 25 mM HEPES pH7.5, 150 mM NaCl and 1 mM DTT buffer. Human SKP1-FBW7 complex, NEDD8, APPBP1-UBA3 (the E1 to activate NEDD8), UBE2M (a NEDD8 E2-conjugating enzyme) and fluorescently labeled wild-type, K11R, K48R or K63R UB were generated as previously described^{18,21,52}. APC/C and CDH1 were expressed and purified as previously described³¹. The RING-RING fusion version of RNF4 was expressed and purified as previously described³¹. Coding regions for CUL1, CUL4A (38-C), RBX1 (5-C) His-TEV-DDB1, CRBN, GST-TEV-IKZF1 (encompassing zinc fingers 2–3 containing amino acids 141–243^{Δ197–246} referred to as IKZF1 ZF 2–3)³⁸ and UBA1 were all subcloned into pLIB vectors³⁸. Baculoviruses for CUL1, GST-TEV-RBX1 5-C, CUL4A 38-C, HIS-TEV-DDB1 and CRBN were first prepared and isolated from Sf9 cells, followed by CUL1 and GST-TEV-RBX1 5-C, CUL4A 38-C and GST-TEV-RBX1 5-C, HIS-TEV-DDB1 and CRBN co-infection of Hi5 cells for co-expression^{51,52}. Proteins were purified by His or GST affinity chromatography followed by overnight TEV cleavage. Cleaved protein solutions were then subjected to ion exchange chromatography followed by size exclusion chromatography in 25 mM HEPES pH7.5, 150 mM NaCl and 1 mM DTT buffer. The covalent modification of CUL1-RBX1 (CRL1) and CUL4A-RBX1 (CRL4) with the CRL activator protein NEDD8 (termed neddylation) was performed as previously described^{51,52,57}. All variants of UB used in this study were generated as previously described⁵³.

A plasmid for the bacterial expression of K63R human UB was prepared by using a previous construct for a GST fusion³⁸ to wild-type human UB containing a conserved N-terminal TEV cleavage site (ENLYFQG) and protein kinase A consensus sequence (RRASVG) for radiolabeling. Mutation of K63 to Arg was accomplished by the Quikchange method, using DNA oligo sequences 5'-GATTACAACATTGAGAGGGAGTCCACCTTACATC-3' for the forward primer and 5'-GATGTAAGGTGGACTCCCTCTGAATGTTGTAATC-3' for the reverse one. The construct DNA sequence was validated by Sanger sequencing. The plasmid was transformed into chemically competent BL21(DE3) *Escherichia coli* bacteria for expression at 37 °C. Protein purification was accomplished using standard approaches³⁸, with the final step being gel filtration chromatography into a buffer containing 30 mM Tris, pH8.0, 100 mM NaCl, 1 mM DTT and 10% glycerol. Purified K63R UB was concentrated to ~250 μM based on an extinction coefficient of 1,280 M⁻¹ cm⁻¹ and flash frozen in liquid nitrogen before storage at -80 °C. K63R UB (50 or 100 μM) was radiolabeled in the presence of 5 kU of cyclic adenosine monophosphate-dependent protein kinase (New England Biolabs) and [³²P]-adenosine triphosphate (ATP) for 1 h at 30 °C.

All UB-conjugating enzymes (E2s) and their associated E3s employed in this study are listed in Supplementary Table 1.

Donor UB discharge assay (pulse-chase) to free amino acids. UBE2N (20 μM) was loaded with 20 μM fluorescent donor UB (UB*) in the presence of 0.3 μM UBA1 in buffer containing 50 mM Hepes, pH7.5, 100 mM NaCl, 2.5 mM MgCl₂, 1.5 mM ATP and 0.05 mg ml⁻¹ BSA. Loading reactions were incubated for 0.5 h and quenched by adding EDTA to a final concentration of 30 mM. The reaction was then initiated by adding UBE2N-UB* (0.5 μM final) to a substrate mix containing 0.5 μM UBE2V1, 0.5 μM RNF4 RING domain and 35 mM amino-acid acceptors (N_ε-acetyl-L-lysine, L-serine, L-Dap, N_ε-acetyl-L-ornithine, L-lysine, D-lysine, N_α-acetyl-L-lysine or L-homolysine). Reactions were quenched with either non-reducing or reducing SDS-PAGE sample buffer after 0, 5, 10, 20, 30, 45, 60, 120 or 180 min, and substrates and products were separated by SDS-PAGE. Gels were scanned on an Amersham Typhoon system (GE Healthcare) and the intensities of all fluorescent bands were quantified using ImageQuant TL (GE Healthcare). The E2-UB* band intensities were divided by the total fluorescent intensity in each lane and normalized to the 0 time point. Data were plotted in GraphPad Prism 8 (GraphPad Software) and fitted to an exponential decay function using nonlinear regression (Fig. 1 and Extended Data Fig. 1). All reactions were performed in duplicate. Source Data Fig. 1 and Source Data Extended Data Fig. 1 contain all gels obtained from this experiment.

Transpeptidation reactions. Sortase-mediated transpeptidation was utilized to link the C terminus of various acceptor UBs to the N terminus

of a cyclin E phosphopeptide (Nterm-GGGGLPSGLL(pT)PPQ(pS)GKKQSSDYKDDDDK-Cterm), IKZF1 ZF 2–3 or Sna4p peptide (Nterm-GGGGQSLVESPPYPENLYFQGDYKDDDDK-Cterm). UBs were synthesized or expressed recombinantly that contained a G76S mutation followed by the GSGSLPETGG sortase recognition sequence. Briefly, 50 μM UB was mixed with 100 μM substrate and 10 μM His-sortase on ice in a buffer containing 50 mM Tris, pH8.0, 150 mM NaCl, 10 mM CaCl₂ for 1 h. Next, the reaction mixture was exposed to nickel-agarose beads to remove His-sortase. Final products were purified by size exclusion chromatography in 25 mM HEPES pH7.5, 150 mM NaCl and 1 mM DTT buffer. For the UB-Sna4p fusions, an additional overnight incubation with TEV was included to remove a FLAG tag from Sna4p peptide, followed by size exclusion chromatography.

Donor UB discharge assay (pulse-chase) to UB analogs. E2s (20 μM) were loaded with 20 μM fluorescent donor UB (UB*) to form the E2-UB* complex in the presence of 0.3 μM UBA1 in a buffer containing 50 mM Hepes, pH7.5, 100 mM NaCl, 2.5 mM MgCl₂, 1.5 mM ATP and 0.05 mg ml⁻¹ BSA. Loading reactions were quenched with EDTA (30 mM final) after a 0.5-h incubation period at room temperature. Reactions were initiated by the addition of various UB acceptors, and in some cases E3s (Supplementary Tables 2–4 report the final concentrations of these reagents for all pulse-chase reactions) in a buffer containing 25 mM Hepes, pH7.5, 150 mM NaCl together with E2-UB* (~0.5 μM final). All reactions were performed at room temperature for the indicated times and quenched with non-reducing SDS-PAGE sample buffer. Substrates and products were separated by SDS-PAGE and scanned on an Amersham Typhoon system (GE Healthcare). The intensities of all fluorescent bands were quantified using ImageQuant TL (GE Healthcare). The amount of di-UB chain was calculated by first dividing the di-UB* band intensity by the total UB* intensity in each lane of the gel. The fraction of di-UB* product was then multiplied by the total amount of UB* (μM) used in the reactions. All reactions were performed in duplicate.

For UBE2S, donor UB can be transferred to a lysine on the E2 surface (termed autoubiquitination). To minimize this, E1-UB* was prepared and added to UBE2S protein immediately before initiation of the reaction. Briefly, 10 μM UBA1 was first loaded with 20 μM UB* at room temperature for 0.5 h. E1-UB* was desalted twice, using a Zeba desalting column, to quench loading into a buffer containing 25 mM HEPES pH7.5 and 150 mM NaCl. Reactions were initiated as described above by the addition of UB acceptors and UBE2S with APC/C and its coactivator CDH1 to achieve final concentrations of ~5 μM E1-UB* and 0.2 μM E2. Reactions were processed as described above.

For reactions containing CRL1-bound substrate, the SKP1/FBW7 substrate receptor was utilized to reconstitute the full CRL (CRL1^{FBW7}), which binds to phosphorylated cyclin E peptide. For reactions containing CRL4-bound substrate, the substrate adapter CRBN was utilized to reconstitute the full CRL (CRL4^{CRBN}). Here, the small molecule pomalidomide (2 μM final) facilitates complex formation between the sortase IKZF1 ZF 2–3-UB fusion and CRL4^{CRBN}. All reactions for CRL-bound substrates had an approximate final concentration of 0.5 μM E2-UB that had been generated in the pulse step. All CRL-dependent reactions were processed as described above. Source Data Figs. 2 and 3 and Source Data Extended Data Figs. 2 and 4 contain all gels obtained from this experiment.

In-gel digestion protocol for liquid chromatography-tandem mass spectrometry. UBE2D3 pulse-chase reactions were run as described above (see biochemical assay section). Briefly, UBE2D3 was loaded with either fluorescently labeled UB (UB*) or GST-UB. Note that GST-tagged donor UB was used to separate di-UB³¹, which is formed during the pulse reaction as a side product, from the desired di-UB product between donor and acceptor UBs. Chase reactions contained 1 μM RNF4 and 100 μM UB_{c4}^{K11}, UB_{c3}^{K48} or UB_{c3}^{K63}. After a 15-min incubation for UB_{c4} or 1 h for UB_{c3}, reactions were quenched with SDS-PAGE sample buffer. Reactants and products were separated by SDS-PAGE. Gels with samples containing UBE2D3-UB* were scanned on an Amersham Typhoon system (GE Healthcare) and used to generate the image in Fig. 4b. Gels with samples containing UBE2D3-GST-UB were first stained with Coomassie brilliant blue to identify and excise the desired GST-UB³¹-UB^A product band. After staining, the gel was subsequently destained by soaking for several hours in 10% acetic acid, 40% methanol and 60% de-ionized water with at least two changes of the solvent to achieve a clear background. The gel band corresponding to GST-UB³¹-UB^A was excised and chopped into smaller pieces (~1 × 1 mm). Gel pieces were washed twice with 50% 50 mM ammonium bicarbonate, pH8.0 (ABC)/50% EtOH and then completely dehydrated by incubation in absolute EtOH. The gel pieces were then dried in a Speed-vac system (Eppendorf, Concentrator plus), rehydrated in 200 μl of 1% (wt/vol) Sodium deoxycholate (SDC) buffer (10 mM Tris-(2-carboxyethyl)phosphine (TCEP), 40 mM 2-chloroacetamide (CAA), 0.5 μg trypsin, 0.5 μg LysC in 100 mM Tris-HCl pH8.5) and incubated at 37 °C overnight. The next day, peptides were extracted from gel pieces by two consecutive rounds of adding isopropanol buffer (1% 2,2,2-trifluoroacetic acid (TFA) in isopropanol) to the samples and subsequent collection of the liquid phase. At this step, stable isotope-labeled (SIL) analogs of chain specific native di-Gly peptide standards were added to the samples, which provided chromatographic orientation for the detection of endogenous (light)

counterparts. For absolute quantification of different di-Gly peptides in UB_{C4} samples, SIL analog spike-in amounts were adjusted to yield peptide quantification ratios between 0.1 and 10 (20 and 2 fmol per injection for K11_GG, K48_GG, K63_GG and K27, K29, K33, respectively).

Liquid chromatography–tandem mass spectrometry sample preparation. Stage Tips were prepared by inserting three layers of an Styrenedivinylbenzene-reverse phase sulfonate (SDB-RPS) matrix (Empore) into a pipette tip using an in-house prepared syringe device as described previously^{29,30}. The peptides, mixed with isopropanol buffer, were loaded onto the StageTips. The tips were washed with isopropanol buffer and subsequently with 2% ACN/0.2% TFA. Elution was performed using 80% ACN/1.25% NH₄OH. Eluates were collected in PCR tubes and dried using a Speed-vac centrifuge. Peptides were resuspended in buffer A* (2% ACN/0.2% TFA) and briefly sonicated (Branson Ultrasonics) before LC/MS-MS analysis.

Liquid chromatography–tandem mass spectrometry measurements. Peptides were loaded on a 50-cm reversed-phase column (75 μm inner diameter, packed in house with ReproSil-Pur C18-AQ 1.9-μm resin (Dr Maisch)). The column temperature was maintained at 60 °C using a homemade column oven. An EASY-nLC 1200 system (Thermo Fisher Scientific) was directly coupled online with the mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) via a nano-electrospray source, and peptides were separated with a binary buffer system of buffer A (0.1% formic acid (FA)) and buffer B (80% acetonitrile, 0.1% FA), at a flow rate of 300 nL min⁻¹. Peptides were eluted with a gradient starting at 7% buffer B (0.1% (vol/vol) FA, 80% (vol/vol) ACN) and stepwise increased to 14% in 4 min and 26% in 22 min. After each gradient, buffer B concentration was increased to 95% in 2 min and maintained at this concentration for 6 min.

The mass spectrometer was programmed to acquire in targeted scan mode in which every full scan, with resolution of 60,000 at 200 *m/z* (3 × 10⁶ ions accumulated with a maximum injection time of 20 ms), was followed by 20 multiplexed selected ion monitoring (SIM) scans employing a multiplexing degree of four. Light (endogenous) and heavy counterpart peptides were always simultaneously recorded in the same scan. Each SIM scan covered a range of *m/z* of 150–2,000 with resolution of 120,000 (5 × 10⁴ ions accumulated with a maximum injection time of 65 ms, 1.4-*m/z* isolation window and 0.4-*m/z* isolation offset). The targeted peptides with *m/z* values are listed in Supplementary Table 5.

Data analysis. Raw MS data were processed using Skyline, an open-source software project^{37,38}. Graphical displays of chromatographic traces were manually inspected for proper peak picking of MS1 filtered endogenous peptides based on co-eluting SIL peptides. All quantification was done on the precursor ion level, based on area. Only the most abundant peak of the isotope cluster was used for quantitation.

Bioinformatics analyses in this study were performed with Microsoft Excel and data visualized using GraphPad Prism (GraphPad Software). The background signal detected in the sample lacking acceptor UB was subtracted from the corresponding signals of samples containing acceptor UBs. Next, each chain peptide was normalized to the first tryptic peptide of the N-terminally modified UB^A (M1Nle) sequence: NleQIFVK. Because this peptide was used for normalization, K6-linked di-UB was not measured in our protocol. Finally, fold changes of each chain peptide (relative to WT) are calculated using the equations in the next section and shown in Fig. 4c,d and Extended Data Fig. 5a,b. All gels used in this experiment are shown in Source Data Fig. 4.

Equations. Correction for background signal by subtraction of signal detected in 'no acceptor' reactions:

$$AUC_{rep,i}^{UB_{C4}^{Acceptor,BGcorr}} = AUC_{rep,i}^{UB_{C4}^{Acceptor}} - \overline{AUC}^{UB_{C4}^{NoAcceptor}} \quad \{i|1 \leq i \leq 3\} \quad (1)$$

Normalization of GlyGly peptides to NleQIFVK:

$$GG-Pep. AUC_{rep,i}^{UB_{C4}^{Acceptor, Norm}} = \frac{GG-Pep. AUC_{rep,i}^{UB_{C4}^{Acceptor,BGcorr}}}{NleQIFVK AUC_{rep,i}^{UB_{C4}^{Acceptor,BGcorr}}} \quad \{i|1 \leq i \leq 3\} \quad (2)$$

Foldchange calculation of UB_{C5} to UB_{C4}:

$$GG-Pep. FC_{rep,i}^{UB_{C5}^{Acceptor, Norm}} = \frac{GG-Pep. AUC_{rep,i}^{UB_{C5}^{Acceptor, Norm}}}{GG-Pep. AUC_{rep,i}^{UB_{C4}^{Acceptor, Norm}}} \quad \{i|1 \leq i \leq 3\} \quad (3)$$

Estimating pK_d^{app} values for ubiquitylation reactions. For UBE2N/UBE2V1, the pK_d^{app} values for synthetic UBs (⁸⁶³UB_{C4} or ⁸⁶³UB_{C5}) were measured by a steady-state kinetics assay that detects isopeptide bond formation between radiolabeled donor UB and unlabeled acceptor. First, a titration series was created using Bis-Tris propane buffer with pH values of 5.7, 6.1, 6.5, 6.9, 7.3, 7.7, 8.1, 8.5, 8.9, 9.3, 9.7 and 10.1. Stocks of Bis-Tris propane buffer, 10× reaction buffer (20 mM ATP, 10 mM DTT, 50 mM MgCl₂ and 500 mM NaCl), radiolabeled K63R donor UB, UBE2V1

and UBE2N (WT or K92R) proteins were added in the above order to autoclaved individual Eppendorf tubes. The final concentrations in the ubiquitylation reactions were 50 mM Bis-Tris Propane, 1× reaction buffer, 0.25 μM human UBA1, 5 μM radiolabeled K63R donor UB and 2 μM UBE2V1/K92R UBE2N complex. Following a 1-min incubation period, either synthetic ⁸⁶³UB_{C4} or ⁸⁶³UB_{C5} was added to initiate the reaction (100 μM final concentration). Reactions with ⁸⁶³UB_{C4} were quenched after 2 min and 45 s and reactions with synthetic ⁸⁶³UB_{C5} were quenched after 15 min in either non-reducing or reducing 2× SDS-PAGE buffer (100 mM Tris-HCl, pH 6.8, 20% glycerol, 30 mM EDTA, 4% SDS and 0.02% bromophenol blue). The reaction products and substrate were resolved by SDS-PAGE on 18% Tris-glycine polyacrylamide gels, followed by autoradiography and detection on a Typhoon 9410 Imager. Quantification of substrate and products was performed using Image Quant (GE Healthcare). The fraction of di-UB product for each reaction was measured by normalizing the signal for product over the total signal in the lane. These fractions were then multiplied by the donor UB concentration and divided by both the UBE2N/UBE2V1 complex concentration and the time of incubation. The velocities were plotted as a function of the pH of the reaction and fit to a sigmoidal four-parameter logistic curve with the Hill slope constrained to 1 (GraphPad Prism software, version 8.3). Note that this model assumes that the reaction velocities are dependent on a single ionizing species that becomes activated at high pH. A similar procedure was followed for UBE2R2 and ⁸⁴⁸UB_{C4} or ⁸⁴⁸UB_{C5}, except for the following modifications. The final concentrations in the ubiquitylation reactions were 0.5 μM human UBA1, 15 μM radiolabeled K63R donor UB and 10 μM UBE2R2 protein. Reactions with ⁸⁶³UB_{C4} were quenched after 5 min and reactions with ⁸⁶³UB_{C5} were quenched after 60 min. The times of incubation were selected to ensure that all reaction velocities were within the linear range and that donor UB consumption was not sufficient to result in a lower concentration than E2. All reactions were performed in duplicate. All gels used in this experiment are shown in Source Data Table 1 and Source Data Extended Data Fig. 7.

Estimating the K_d for UBE2N/UBE2V1 complex. The K_d values of acceptor UBs were measured by a steady-state kinetics assay that detects isopeptide bond formation between radiolabeled donor UB and unlabeled acceptor. A 10× reaction buffer was prepared with 500 mM Bis-Tris propane, pH 7.3, 20 mM ATP, 10 mM DTT, 50 mM MgCl₂ and 500 mM NaCl. First, a twofold dilution series was established for acceptor UB proteins that had first been dialyzed into a buffer containing 30 mM Bis-Tris propane, pH 7.3. The starting concentrations of the dilution series were 1.4 mM for bacterial ⁸⁶³UB_{C4}, 1.6 mM for synthetic ⁸⁶³UB_{C4} and 1.3 mM ⁸⁶³UB_{C5} (note that initiation of the reaction results in a further twofold dilution of each acceptor UB). Next, the following reagents were added from stock solutions to an Eppendorf tube to achieve final concentrations in each reaction of 1× reaction buffer, 0.25 μM human UBA1, 5 μM K63R donor UB and 2 μM UBE2N/UBE2V1 for bacterial or synthetic UB_{C4}, or 15 μM K63R donor UB and 10 μM UBE2N/UBE2V1 for ⁸⁶³UB_{C5}. After a 2-min incubation period, aliquots of the master mix were evenly dispensed to clean Eppendorf tubes. Ubiquitylation reactions were then initiated by adding an equal volume of acceptor UB to the Eppendorf tubes containing the master mix. Reactions were incubated for either 15 or 30 min (⁸⁶³UB_{C4} or ⁸⁶³UB_{C5}, respectively) before quenching in either non-reducing or reducing 2× SDS-PAGE buffer containing 100 mM Tris-HCl, pH 6.8, 20% glycerol, 30 mM EDTA, 4% SDS and 0.02% bromophenol blue. The processing of reactions and estimation of velocities were performed as described in the previous section. The reaction velocities were fit to the Michaelis–Menten equation to estimate K_d (GraphPad Prism software, version 8.3). Reactions containing the RING domain of RNF4 were performed similarly with the following changes. All reactions contained 0.5 μM human UBA1, 15 μM K63R donor UB, 10 μM UBE2N/UBE2V1 and 1 μM RNF4. The starting concentrations of the dilution series were 1.3 mM for UB_{C4} and 1.25 mM for K63 UB_{C5}. Reactions were incubated for either 0.5 or 2.5 min (UB_{C4} or ⁸⁶³UB_{C5}, respectively) before quenching. The times of incubation were selected to ensure that all reaction velocities were within the linear range and that donor UB consumption was not sufficient to result in a lower concentration than E2. Reactions were performed in duplicate. All gels used in this experiment are shown in Source Data Table 1.

Estimating the K_d for UBE2R2 and Rsp5p. The K_d values of acceptor UBs ⁸⁴⁸UB_{C4} and ⁸⁴⁸UB_{C5} for UBE2R2 and Rsp5p were measured similarly to the protocol described in the previous section with the following modifications. For UBE2R2, a twofold dilution series was established for acceptor UB proteins that had first been dialyzed into a buffer containing 30 mM Bis-Tris propane, pH 7.3 with starting concentrations of 8.4 mM for UB_{C4} and 12.5 mM for ⁸⁴⁸UB_{C5}. The final concentrations in each reaction contained 0.5 μM human UBA1, 15 μM K48R donor UB and 10 μM UBE2R2 protein. Reactions were incubated for either 1 or 2.5 min for each replicate for the UB_{C4} titration series and for either 15 or 16 min for each replicate of the ⁸⁴⁸UB_{C5} titration series before quenching. For Rsp5p, the starting concentration of the acceptor UB dilution series was 1.6 mM for both UB_{C4} and ⁸⁶³UB_{C5}. The final concentrations in each reaction contained 0.5 μM human UBA1, 7 μM K63R donor UB and 5 μM UBE2D2 and Rsp5p proteins. Reactions were incubated for either 5 or 30 min (UB_{C4} or ⁸⁶³UB_{C5}, respectively) before quenching. The times of incubation were selected to ensure that all reaction velocities were within the linear range and that donor UB consumption

was not sufficient to result in a lower concentration than E2 or E3. Reactions were performed in duplicate. All gels used in this experiment are shown in Source Data Table 1.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

All raw gels are included in source data files. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD021286. Source data are provided with this paper.

Code availability

ROSETTA software can be downloaded from www.rosettacommons.org and is available free to academic users.

References

- Scott, D. C. et al. Two distinct types of E3 ligases work in unison to regulate substrate ubiquitylation. *Cell* **166**, 1198–1214 (2016).
- Baek, K. et al. NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly. *Nature* **578**, 461–466 (2020).
- Brown, N. G. et al. Mechanism of polyubiquitination by human anaphase-promoting complex: RING repurposing for ubiquitin chain assembly. *Mol. Cell* **56**, 246–260 (2014).
- Plechanovova, A. et al. Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nat. Struct. Mol. Biol.* **18**, 1052–1059 (2011).
- Sievers, Q. L. et al. Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science* **362**, eaat0572 (2018).
- Weissmann, F. et al. biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc. Natl Acad. Sci. USA* **113**, E2564–E2569 (2016).
- Duda, D. M. et al. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995–1006 (2008).
- Ziemba, A. et al. Multimodal mechanism of action for the Cdc34 acidic loop: a case study for why ubiquitin-conjugating enzymes have loops and tails. *J. Biol. Chem.* **288**, 34882–34896 (2013).
- Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nano-electrospray and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
- Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).

Acknowledgements

This study is dedicated the memory of our inspiring mentor, colleague and beloved friend, Huib Ovaa, whom we miss dearly. We thank J.R. Prabu, J. Kellermann, S. von Gronau, D. Scott, S. Uebel, S. Pettera, V. Sanchez, K. Baek, D. Horn-Ghetko and S. Kostrhon for assistance, reagents and helpful discussions. We also thank C. Talavera-Ormeño and P. Hekking for assistance with peptide synthesis. B.A.S. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 789016-NEDD8Activate), and from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation—SCHU 3196/1-1). B.A.S. and M.M. are supported by the Max Planck Society. Also, N.P., D.H., N.B. and G.K. were supported by a grant from the National Institutes of Health (R15GM117555-02). G.J.v.d.H.v.N. was supported by grants from NWO (VIDI and Off-Road). H.O. was supported by a VICI grant from the Netherlands Foundation for Scientific Research (NWO). Work by M.S. and M.J.B. was performed within the framework of SFB 1035 (German Research Foundation DFG, Sonderforschungsbereich 1035, no. 201302640, project Z01).

Author contributions

Syntheses of UB analogs were designed and executed by G.J.v.d.H.v.N. J.L. performed all biochemical assays. Kinetics experiments were carried out by G.K., N.P., D.H. and N.B. MS experiments were designed and conducted by M.M., E.M.H. and O.K. MD simulations were performed by V.H.T. and J.S.H. NMR was carried out by M.J.B. and M.S. The manuscript was prepared by J.L., D.T.K., G.J.v.d.H.v.N., G.K., H.O. and B.A.S., with input from all authors. The project was supervised by B.A.S., G.K., H.O. and D.T.K.

Competing interests

H.O. was a shareholder of UbiqBio. All other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41589-020-00696-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41589-020-00696-0>.

Correspondence and requests for materials should be addressed to G.K., H.O. or B.A.S.

Reprints and permissions information is available at www.nature.com/reprints.

3.8 Mitochondrial phosphoproteomes are functionally specialized across tissues

Fynn M. Hansen¹, Laura S. Kremer², Ozge Karayel¹, Inge Kühl³, Isabell Bludau¹, Nils-Göran Larsson², Matthias Mann¹

¹ Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, 82152, Germany

² Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden.

³ Department of Cell Biology, Institute of Integrative Biology of the Cell, UMR9198, CEA, CNRS, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

Available on *BioRxiv* and submitted for publication

Mitochondria are essential organelles involved in critical biological processes such as energy metabolism and cell survival. Their dysfunction is linked to numerous human pathologies that often manifest in a tissue-specific manner. Thus, understanding the differences between mitochondria in various cellular environments is crucial to understand their functional plasticity. Our group as well as others have investigated the mitochondrial proteome across different tissues [176-178], but a large-scale study of matching mitochondrial proteomes and phosphoproteomes was missing.

To this end, we isolated mitochondria of 7 different mouse tissues and analyzed their mitochondrial proteome and phosphoproteomes via LC-MS. I transformed the acquired data into a rich resource that shows literature-known tissue-specific differences between mitochondria, but also many interesting and novel observations such as phosphorylation clusters on MIGA2 which can be linked to mitochondrial fusion. Thus, this resource provides a valuable dataset for the community to test and make new hypothesis. Furthermore, the dataset is readily available for researchers through a web application.

Mitochondrial phosphoproteomes are functionally specialized across tissues

Fynn M. Hansen¹, Laura S. Kremer², Ozge Karayel¹, Inge Kühl³, Isabell Bludau¹, Nils-Göran Larsson², Matthias Mann¹

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, 82152, Germany

²Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden.

³Department of Cell Biology, Institute of Integrative Biology of the Cell, UMR9198, CEA, CNRS, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

Lead contact: mmann@biochem.mpg.de

Abstract

Mitochondria are essential organelles involved in critical biological processes such as energy metabolism and cell survival. Their dysfunction is linked to numerous human pathologies that often manifest in a tissue-specific manner. Accordingly, mitochondrial fitness depends on versatile proteomes specialized to meet diverse tissue-specific requirements. Furthermore, increasing evidence suggests that phosphorylation may also play an important role in regulating tissue-specific mitochondrial functions and pathophysiology. We hypothesized that recent advances in mass spectrometry (MS)-based proteomics would now enable in-depth measurement to quantitatively profile mitochondrial proteomes along with their matching phosphoproteomes across tissues. We isolated mitochondria from mouse heart, skeletal muscle, brown adipose tissue, kidney, liver, brain, and spleen by differential centrifugation followed by separation on Percoll gradients and high-resolution MS analysis of the proteomes and phosphoproteomes. This in-depth map substantially quantifies known and predicted mitochondrial proteins and provides a resource of core and tissue

modulated mitochondrial proteins (mitophos.de). We also uncover tissue-specific repertoires of dozens of kinases and phosphatases. Predicting kinase substrate associations for different mitochondrial compartments indicates tissue-specific regulation at the phosphoproteome level. Illustrating the functional value of our resource, we reproduce mitochondrial phosphorylation events on DRP1 responsible for its mitochondrial recruitment and fission initiation and describe phosphorylation clusters on MIGA2 linked to mitochondrial fusion.

Introduction

Mitochondria are double-membrane-bound organelles with an essential role in homeostasis of eukaryotic cells. They are often referred to as the “powerhouse of the cell” due to their prominent function in bioenergetics. Among many other processes, they are also involved in several biosynthetic processes such as balancing redox systems, the regulation of metabolic by-products like reactive oxygen species (ROS) (Spinelli and Haigis, 2018) and hold a central role in cell death (Bock and Tait, 2020). The function and stability of mitochondria depend on their intrinsic bioenergetics regulation and finely orchestrated interaction with the cellular microenvironment. Energy conversion via the oxidative phosphorylation system (OXPHOS) plays an essential role in harvesting energy from ingested nutrients. Moreover, the morphology of mitochondria within an eukaryotic cell is actively regulated by fusion and fission events which dynamically modulate their number, size, and localization (Liesa et al., 2009). Regulation of mitochondrial dynamics also affects the interplay of mitochondria with other cellular structures, such as the cytoskeleton for active regulation of their localization (Moore and Holzbaur, 2018), and organelles like the endoplasmic reticulum (ER) and lipid droplets to regulate many physiological processes such as energy metabolism and ion buffering. The mitochondria-associated membrane (MAM), which is the contact site of the outer mitochondrial membrane with the ER, comprises a unique set of proteins mediating this interaction and fine-tune mitochondrial functions with the cellular microenvironment (Kwak et al., 2020; Nunnari and Suomalainen, 2012). Dysregulation of any of these intricate processes can lead to severe mitochondrial dysfunctions and diseases, including neurodegenerative diseases, cardiovascular disorders, myopathies, obesity, and cancers, which can manifest in a cell type- and tissue-specific manner (Suomalainen and Battersby, 2018).

The fitness of mitochondria depends on the production and maintenance of functional as well as versatile proteomes specialized to carry out a variety of functions within the eukaryotic metabolism and meet

diverse cellular and tissue-specific requirements (Kuznetsov et al., 2009). The mitochondrial proteome includes over a thousand proteins (see below), but only a small fraction of 13 proteins are encoded on the circular mitochondrial DNA molecule (Anderson et al., 1981). Thus, the majority of mitochondrial proteins are encoded by the nuclear genome, synthesized outside of mitochondria and subsequently imported into the organelle, implying that mechanisms controlling mitochondrial protein quality (e.g., correct protein folding and import) are essential for health and integrity of mitochondria (Jadiya and Tomar, 2020). Furthermore, investigations of mitochondrial dynamics and functional plasticity have revealed regulatory roles for post-translational modifications (PTMs), including phosphorylation (Niemi and Pagliarini, 2021). Studies have shown that phosphorylation of several mitochondrial proteins is involved in the regulation of central processes such as metabolic function, for instance through phosphorylation of the E1 α subunit of PDH (Patel et al., 2014), mitophagy (Kolitsida et al., 2019) and fission (Cribbs and Strack, 2007; Ducommun et al., 2015; Lewis et al., 2018; Taguchi et al., 2007; Toyama et al., 2016). Thus, deregulation of protein phosphorylation might be an important underlying feature of mitochondrial physiology and pathophysiology. Currently, there is a significant knowledge gap of the mitochondrial variable proteomic composition and to what extent it is phosphorylated in a tissue-specific manner and how post-translational regulation influences organelle function. A detailed understanding of the functional specialization of mitochondria at the protein and phosphorylation levels is needed to elucidate the contribution of mitochondria to health and disease.

Large-scale mass spectrometry (MS)-based quantitative proteomics studies from our and other groups have already shed light on the proteomic composition of mitochondria of various mammalian tissues and cell types, mostly highlighting that the majority of proteins are shared between mitochondria of different tissues (Forner et al., 2006; Mootha et al., 2003; Pagliarini et al., 2008). The breadth and depth of such studies has been largely driven by technological advances in the field in combination with improvements of mitochondria isolation procedures, such as differential centrifugation (DC), DC in conjunction with

ultracentrifugation on e.g. Percoll gradients, magnetic bead-assisted methods (MACS) (Kappler et al., 2016) or MitoTags (Bayraktar et al., 2019). Efforts in defining the mitochondrial proteome lead to databases like MitoCarta2.0 and IMPI (<http://impi.mrc-mbu.cam.ac.uk/>), both integrated in Mitominer4.0 (Smith and Robinson, 2019). A recent quantitative and high confidence proteome of human mitochondria identified 1134 different proteins that vary over six orders in magnitude in abundance (Morgenstern et al., 2021). Similarly, efforts have been undertaken to map the mitochondrial phosphoproteome, and identified dozens to hundreds of phosphorylation sites on mitochondrial proteins (listed in Supplementary Table 1). However, there has been a dramatic improvement in the technology of phosphoproteomics workflows during the last years, leading to the routine identification and quantification of thousands of phosphorylation sites in cell culture and in vivo systems (Bekker-Jensen et al., 2020; Humphrey et al., 2018) which had not been available in earlier studies. Furthermore, comparisons between mitochondrial phosphoproteomes have been difficult because only a single or a few tissues were analyzed. This complicates the combination and comparison of data sets across studies to obtain a clear view of mitochondrial diversity on proteome and phosphoproteome levels. Thus, a concerted effort is needed to systematically and quantitatively profile mitochondrial proteomes together with their matching phosphoproteomes from the same biological source. This would further help to investigate the dynamic composition of mitochondria and help identify the tissue-specific repertoire of mitochondria-resident kinases and phosphatases and their substrate associations.

Here we performed a systematic analysis of the mitochondrial composition at the level of proteins, major functional entities, and phosphorylation in seven mouse tissues – brain, brown adipose tissue (BAT), heart, kidney, liver, skeletal muscle (SKM), and spleen. Our study employs state of the art MS-based proteomics technology to systematically map divergent composition and phosphorylation of mitochondria between tissues and provides functionally valuable insights into their proteome and post-translational regulations. This study contributes to our understanding of tissue-specific mitochondrial processes controlled by

protein abundance and phosphorylation and helps to manipulate these in health and disease. Our mitochondrial (phospho)proteomes are composed into an extensive resource and made freely accessible via mitophos.de.

Results

Comprehensive mitochondria proteome coverage across various mouse tissues

To advance our understanding of tissue-specific functional specializations of mitochondria at the protein level, we set out to characterize proteomes of mitochondria collected from various mouse tissues by LC-MS/MS analysis. To this end, we first isolated mitochondria from seven tissues – brain, brown adipose tissue (BAT), heart, kidney, liver, skeletal muscle (SKM), spleen - from six 18-21 weeks old C57BL/6N mice (three females and three males). Mouse tissues were homogenized with a Dounce homogenizer, and crude mitochondria were isolated via differential centrifugation and subsequently purified on a Percoll density gradient to obtain ultra-pure mitochondria isolates (Kuhl et al., 2017). Importantly, this procedure was shown to efficiently exclude contaminations from other cellular compartments (Wieckowski et al., 2009). Proteomes of these ultra-pure mitochondrial samples were acquired by a state-of-the-art proteomics workflow (Figure 1A), allowing the robust identification and quantification (coefficient of variation (CV) < 20%) of proteins that covered a dynamic range of more than five orders of magnitude (Supplementary File 1).

In total, we identified over 7000 proteins of which 1620 were annotated as mitochondrial proteins using MitoCarta3.0 (Rath et al., 2021) and the IMPI (<http://impi.mrc-mbu.cam.ac.uk/>) database. This essentially covers (92%) the mitochondrial proteome by the measure of MitoCarta3.0 and even in the IMPI database, which also includes predicted mitochondrial proteins, we still identified 62% (Figure 1B). For further analysis, we filtered for proteins identified in more than half of the biological replicates in at least one tissue, which resulted in 1548 mitochondrial proteins. This still represents over 90% of MitoCarta3.0 and 59% of the IMPI databases and highlights the deep and reproducible mitochondrial proteome coverage of this study (Figure 1C). Interestingly, half of these mitochondrial proteins were identified across all tissues while only 9% were exclusively detected in one specific tissue (Figure 1C), confirming previous

reports on mitochondrial proteomes by us and others (Calvo and Mootha, 2010; Forner *et al.*, 2006; Johnson *et al.*, 2007; Mootha *et al.*, 2003). Of these, almost half (65 proteins) were both reproducibly identified and not in the lowest 20% of ranked abundances (Supplementary File 2), making them clear candidates for tissue specific mitochondrial proteins. A similar proportion of the mitochondrial proteome was also exclusive to two or more tissues by the same criterion.

Mitochondrial enrichment efficiency can be determined by the proportion of summed signal intensity for mitochondrial proteins in relation to all identified proteins in measured samples (Williams *et al.*, 2018). Applying this strategy, we determined the proportions to be very high (>95%) for liver, kidney, SKM, heart and BAT, but lower for brain (65%) and spleen (45%) (Figure 2A). These trends are consistent with the established literature (Fecher *et al.*, 2019; McLaughlin *et al.*, 2020; Roberts *et al.*, 2021; Williams *et al.*, 2018) and can likely be explained by the high tissue heterogeneity of brain (Fecher *et al.*, 2019; Menacho and Prigione, 2020) and spleen. Indeed, spleen tissue consists of various types of immune cells (Lewis *et al.*, 2019), which might impede high purity enrichment of its organelles. For brain, contaminations by synaptosomes, which themselves contain neuronal mitochondria, have frequently been observed (Mootha *et al.*, 2003). Yet, correlation of both mitochondrial and all identified proteins between biological replicates yielded Pearson correlation coefficients higher than 95% in all tissues (Figure 2B and Figure 2-figure supplement 1A). Principle component analysis (PCA) further shows that biological replicates cluster together and underlines functional similarities between tissues such as heart and SKM as well as tissue related diversity of mitochondria proteomes (Figure 2C and Figure 2-figure supplement 1B).

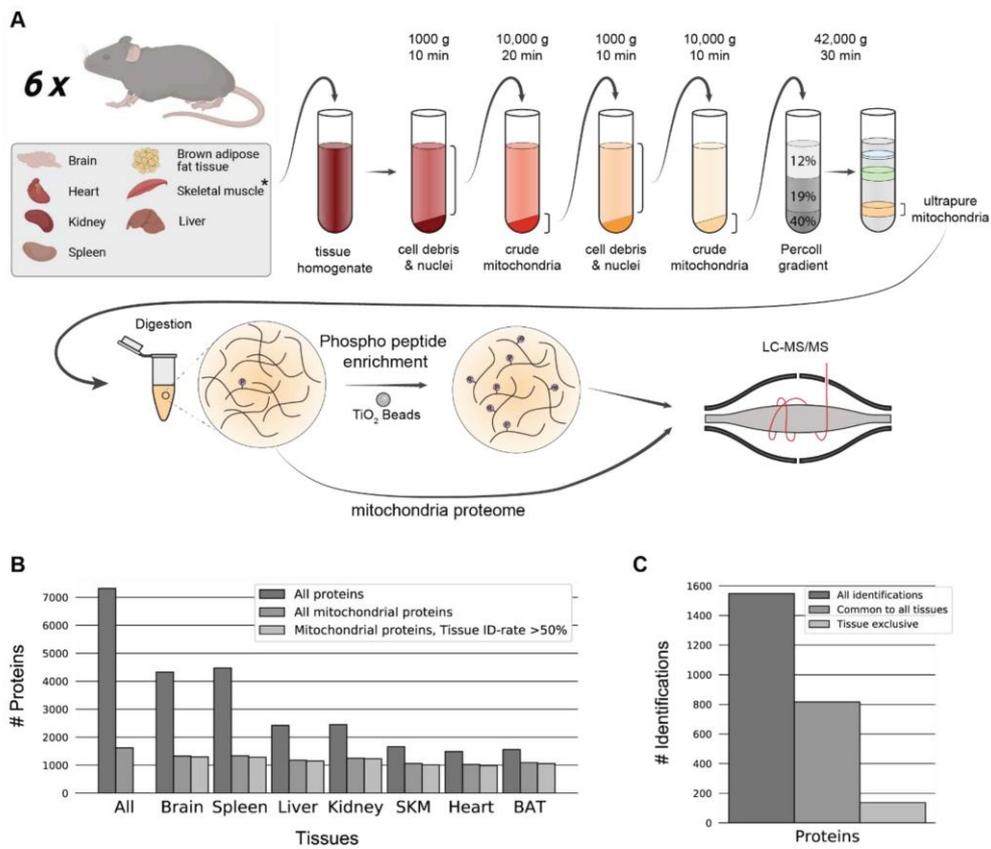


Figure 1 Mitochondrial proteome and phosphoproteome preparation

(A) Workflow of tissue preparations for mitochondrial proteome and phosphoproteome enrichment, and LC-MS/MS analysis (n=6). Tissues were first homogenized (* for skeletal muscle, see Methods), crude mitochondria were isolated and ultrapure mitochondria were obtained using a Percoll gradient. Proteins were digested and prepared for phosphoproteome analysis via TiO₂ enrichment or subjected to LC-MS/MS. (B) Distinct protein identification across biological replicates (n=6). Annotation of proteins as mitochondrial is based on MitoCarta3.0 and the IMPI database. (C) Mitochondrial protein numbers after filtering for mitochondrial proteins identified in at least 50% of biological replicates (n=6) in at least one tissue. Skeletal muscle (SKM), brown adipose tissue (BAT). See also Figure 1 – Source Data 1.

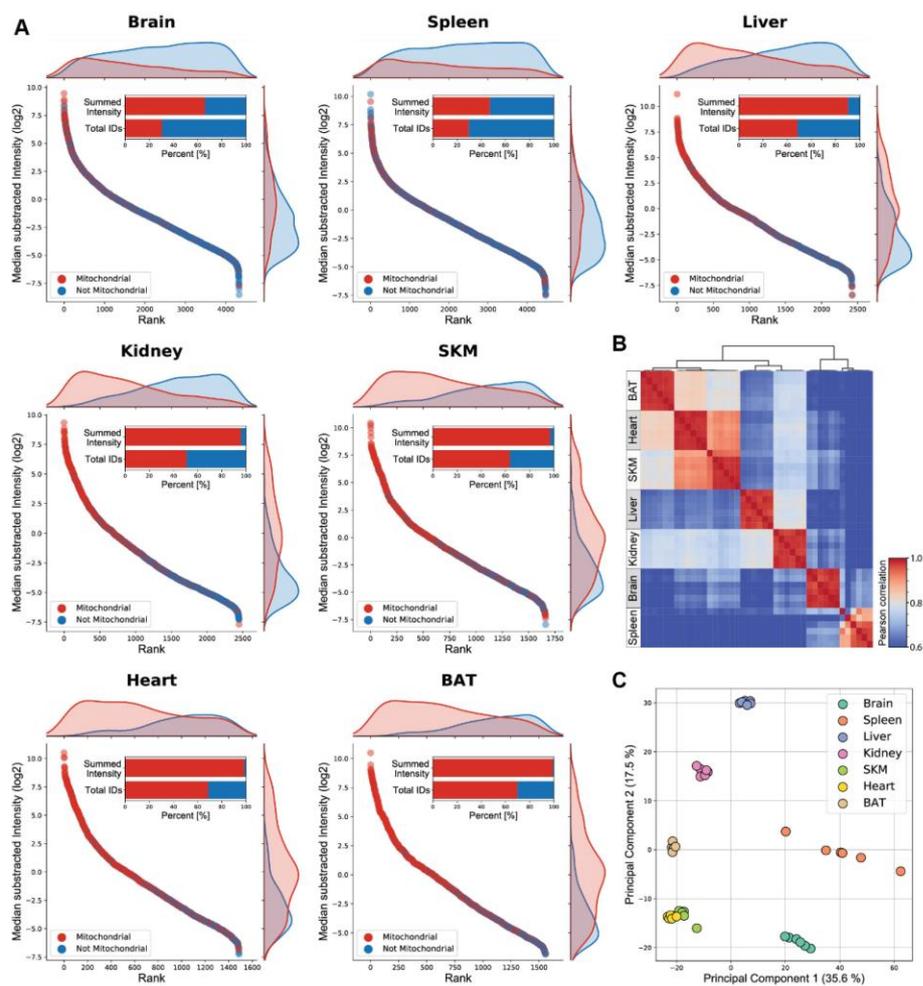


Figure 2 Mitochondria enriched samples show tissue-specific clustering

(A) Mitochondrial (red) and not mitochondrial (blue) proteins identified, based on the MitoCarta3.0 and IMPI database, are ranked by their intensity for each individual tissue. Histograms on the top and right display the distribution of mitochondrial (red) and not mitochondrial (blue) proteins along the rank and the Intensity axis, respectively. The percentage of all identified mitochondrial (red) and not mitochondrial (blue) proteins and their summed intensities are displayed in bar graphs. (B) Heatmap showing Pearson correlation coefficient for biological replicates ($n=6$) for mitochondrial proteins of all mitochondrial enriched samples. (C) Principal component analysis of mitochondrial proteins of all acquired biological replicates ($n=6$). Skeletal muscle (SKM), brown adipose tissue (BAT). See also Figure 2 – Source Data 1.

Mitochondrial proteome composition reveals tissue-specific functions

To gain further insights into tissue-specific functional differences in mitochondria, we investigated differences in the composition of mitochondrial proteomes across tissues. First, we focused on the oxidative phosphorylation (OXPHOS) system, which is essential for production of the energy-rich metabolite ATP and other processes like free radical generation and apoptosis (Huttemann et al., 2007). We found that proteins of the electron transport chain (Complex I – Complex IV) and ATP-synthase (Complex V) displayed high abundances in heart and SKM tissues, supporting the physiological requirement of high ATP levels in both muscular tissues to sustain processes like muscle contraction (Ferreira et al., 2010; Ventura-Clapier et al., 2011) (Figure 3A). Conversely, levels of Complex V proteins were substantially lower in mitochondria from BAT compared to all other measured tissues (Figure 3A), which agrees with the specialized function of BAT in non-shivering thermogenesis (Jastroch et al., 2010; Kajimura and Saito, 2014; Oelkrug et al., 2015). This was further supported by the high abundance of the uncoupling protein 1 (UCP1) in our proteome measurements, the key mediator of the heat-generating proton leak in the mitochondria of BAT (Figure 3B).

In contrast to the well described UCP1, several other members of the SLC25 family remain uncharacterized (Ruprecht and Kunji, 2020). In our dataset, we identified a total of 47 members of the SLC25 family (Kunji et al., 2020). We found that levels of SLC25A23 and SLC25A25, two ATP-Mg²⁺/P_i carrier paralogues, (del Arco and Satrustegui, 2004; Fiermonte et al., 2004), displayed high abundance in mitochondria of brain tissue (Figure 3C and Figure 3D). Interestingly, knockout of SLC25A23 was shown to increase neuronal vulnerability (Rueda et al., 2015), which corroborates our observation and suggests an important role of SLC25A25 in this tissue type. We also quantified a third paralogue, SLC25A24, which showed a higher abundance in spleen tissue. Notably, the spleen harbors a large pool of B-cells and reduced SLC25A24 levels were previously linked to B-cell malignancies (Sandhu et al., 2013) (Figure 3E). Although these three ATP-Mg²⁺/P_i carriers were not detected in mitochondria from heart tissue, we identified another class of

ATP carriers, including SLC25A4 or SLC25A31, which both showed increased abundance in heart compared to all other measured tissues. Such differences between mitochondrial proteome compositions are easily retrieved from our dataset, facilitating a better understanding of mitochondrial plasticity across tissues.

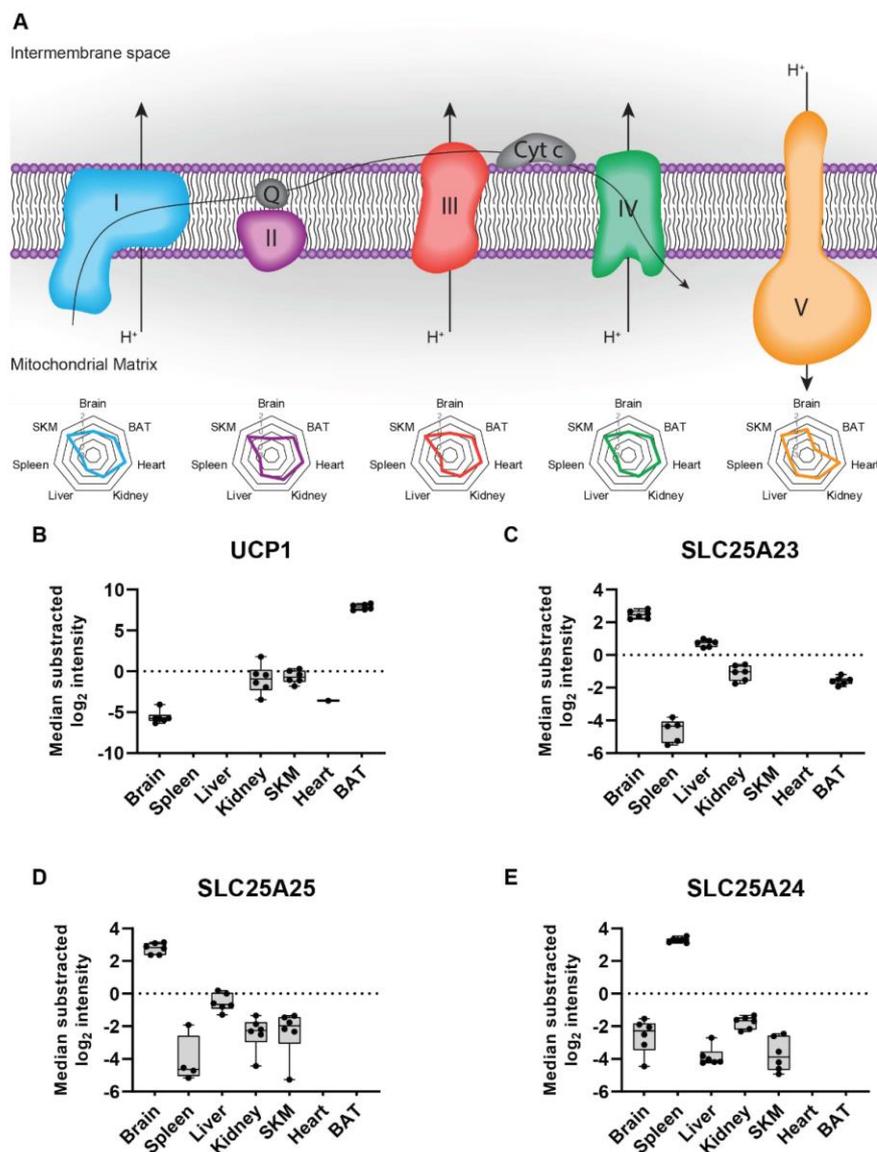


Figure 3 Tissue-specific protein contribution of SLC25 proteins the mitochondrial proteome

(A) Representation of the oxidative phosphorylation (OXPHOS) system including from left to right the electron transport chain (Complex I (blue), Complex II (violet), Complex III (red), Complex IV (green)) and ATP synthase (Complex V (yellow)). Radar plots show the relative contribution of the corresponding complex across the analyzed tissues to the overall mitochondrial composition (see methods for detailed description). (B) Normalized intensity (median of all \log_2 transformed mitochondrial proteins of a sample was subtracted from all \log_2 protein intensities of that sample) of the uncoupling protein 1 (UCP1), (C) SLC25A23, (D) SLC25A25, and (E) SLC25A24 across all analyzed tissues (black dots indicate individual identifications). Protein abundance differences in relation to the reference tissue (BAT in B, Brain in C and D and spleen in E) were significant (p -value <0.0001) by one-way ANOVA analysis (Figure 3 - Source Data 1). Data in this figure is based on the analysis of six replicates ($n=6$) for each tissue. Skeletal muscle (SKM), brown adipose tissue (BAT). See also Figure 3 – Source Data 1.

Tissue-specific function of mitochondria-associated proteins

The identification of key proteins mediating the crosstalk of mitochondria with their cellular environment is crucial to better understand their tissue-specific regulation and this concept has already attracted considerable interest in recent years (Montes de Oca Balderas, 2021). Although the characterization of local proteomes of organellar contact sites usually require special centrifugation-based isolation methods, we anticipated that a considerable fraction of mitochondria-associated proteins would also enrich along with mitochondria in our samples. We first performed an annotation term enrichment analysis of all identified proteins. While we observed significant enrichment of several mitochondria related terms in all tissues as expected, the non-mitochondrial protein pool was largely enriched for terms related to the tissue of origin (Figure 4 – supplement 1, Figure 4 - Source Data 1). For instance, terms like ‘epoxygenase P450 pathway’ in liver, ‘positive regulation of B cell activation’ in spleen or ‘positive regulation of excitatory postsynaptic potential’ in brain tissue highlight tissue-specific functions.

Next, we performed a network analysis to evaluate the nature and quality of co-enriched proteins, more specifically whether non-mitochondrial proteins identified in these samples are associating proteins with functional roles or biological contaminations that are likely tissue-specific and high abundant. This analysis

revealed several clusters of known mitochondrial complexes such as the Tim23 complex or processes like the ubiquinone biosynthetic process, whose members were robustly identified in all tissues, and several tissue-specific clusters including both mitochondrial and non-mitochondrial proteins (Figure 4). For instance, in line with terms enriched for co-sedimenting proteins in liver tissue, we identified a cluster of cytochrome P450 superfamily members exclusively in liver tissue (Figure 4). Intriguingly, CYP2E1 and CYP1A2, two specific members of this family, were previously shown to be targeted to the mitochondria (Avadhani et al., 2011; Genter et al., 2006; Robin et al., 2001), and we now find evidence of the enrichment of many more family members in this tissue. Another identified cluster consisted of G proteins, some of which were identified throughout all tissues and previously annotated as mitochondrial (e.g. GNB1, GNB2 and GNG5) or shown to localize to mitochondria (GNAI2) (Beninca et al., 2014). Interestingly, like the majority of the proteins in the G protein cluster, many G protein coupled receptors (GPCR) were exclusively identified in brain tissue. A prominent member of these brain specific GPCRs is CNR1, which was reported to localize to mitochondria where it plays an important role in the regulation of memory processes through the modulation of the mitochondrial energy metabolism (Hebert-Chatelain et al., 2016). These examples highlight the value of our dataset for uncovering novel mitochondrial and mitochondria-associated proteins in an unbiased way.

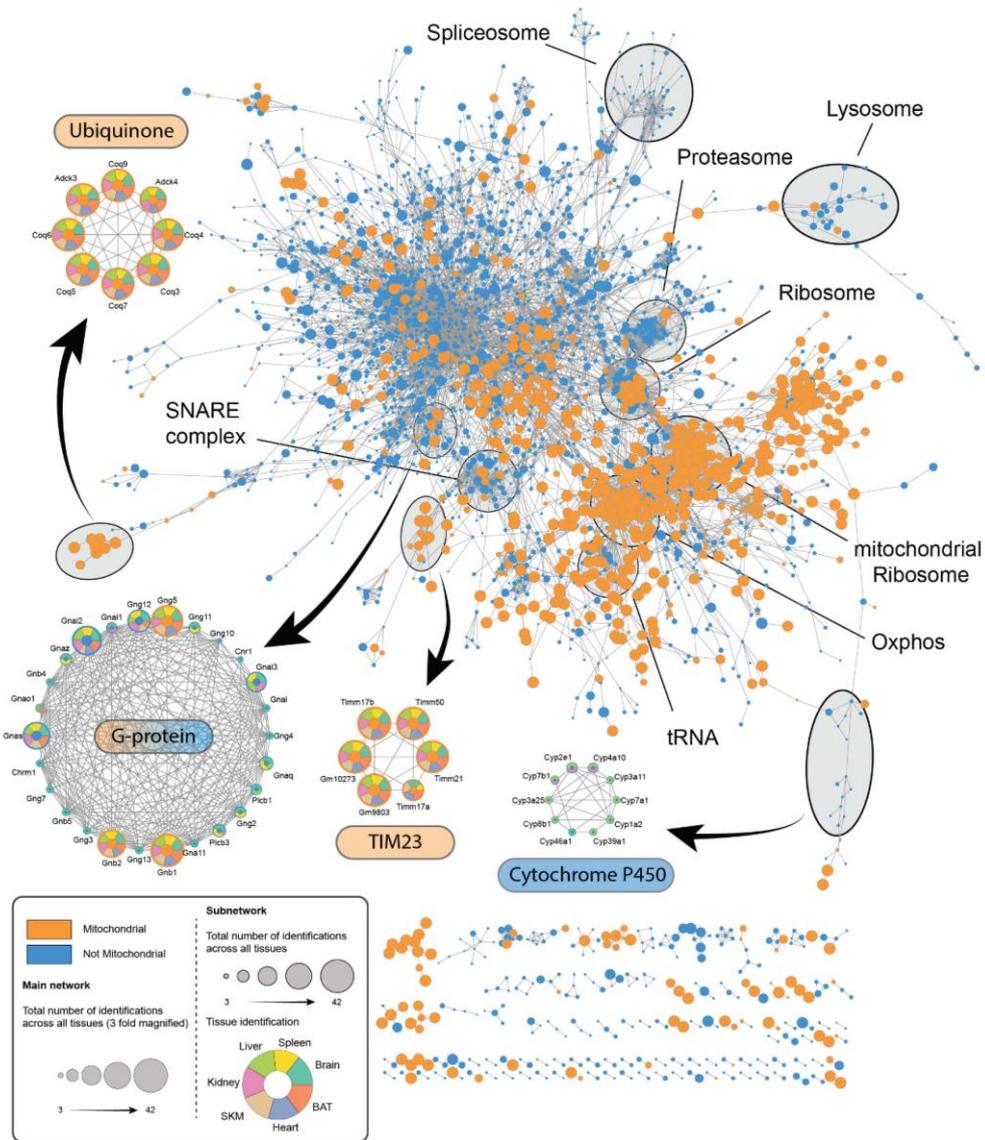


Figure 4 Proteome of mitochondria enriched samples displays tissue-specific complexes

Cytoscape network analysis of reproducibly (> 50% identification rate in at least one tissue, six replicates were analyzed (n=6)) identified proteins of all mitochondria enriched samples. The main network depicts mitochondrial (orange) and non-mitochondrial (blue) proteins with at least one edge (String score >0.95). The size of individual nodes represents the number of identifications ranging from 3 (small circle) to 42 (big circle). Subnetworks display mitochondrial (orange) and non-mitochondrial

(blue) proteins and tissues in which they were identified (dark green – brain; yellow – spleen; light green – liver; pink – kidney; ochre – SKM; blue – heart; orange – BAT). Skeletal muscle (SKM), brown adipose tissue (BAT). See also Figure 4 – Source Data 1.

Mitochondrial kinases and phosphatases show tissue specificity

Post-translational modification of proteins, specifically phosphorylation, plays a crucial role in the orchestration of mitochondrial protein function (Niemi and Pagliarini, 2021). However, almost no mitochondrial kinases with mitochondrial targeting sequences have been consistently reported and most kinases shown to associate with mitochondria have been found on or interact with the outer membrane (Kotrasova et al., 2021). Given our deep mitochondrial proteomes, we investigated relative abundances of kinases and phosphatases, which are annotated as or suggested to be mitochondrial, across mouse tissues.

Firstly, we observed clear differences in abundances of identified mitochondrial kinases and phosphatases, including well described matrix kinases and phosphatases, between tissues (Figure 5A). For instance, PDK1, PDK2, and PDK4 contributed preferentially to the composition of heart, SKM and BAT mitochondria, while the PDK3 was more abundant in brain, spleen, and kidney mitochondria. These tissue-related differences are in line with earlier reports and suggest a specialized function of PDK3, which may originate in its insensitivity to pyruvate inhibition (Klyuyeva et al., 2019; Sadana et al., 2018). Similarly, levels of the heterodimeric pyruvate dehydrogenase phosphatase consisting of PDP1 and PDPr were elevated in brain, SKM, and heart compared to the remaining tissues, whereas PDP2 contributes more to the composition of liver, kidney, and BAT mitochondria. While our study confirmed previous reports on the differential expressions of these proteins in a tissue-specific manner (Huang et al., 1998; Huang et al., 2003), it also provided quantitative data to assess the magnitude of these differences. Together, our results imply a tailored regulation of kinase and phosphatase abundances across tissues to modulate the mitochondrial phosphoproteome.

To investigate if and how kinase and phosphatase levels translate into protein phosphorylation, we analyzed the mitochondrial phosphoproteomes of the same samples collected from all seven tissues (Figure 1A). This analysis resulted in the identification of 1263 phosphorylation sites on 626 mitochondrial proteins (Figure 5B, Figure 5 - Source Data 1). After stringent filtering of the data for more than four identifications across six biological replicates in at least one tissue and a site localization score higher than 75%, we obtained a dataset of 758 phosphorylation sites on 423 mitochondrial proteins (Figure 5C, Figure 5 - Source Data 1). Of these high confidence sites, 16% have previously not been reported in mouse according to the PhosphoSitePlus (PSP) database (Hornbeck et al., 2012). Strikingly, in contrast to the mitochondrial proteomes, 37% of the phosphosites identified on mitochondrial proteins were exclusive to one tissue, and only 8% were identified in all the tissues measured (Figure 5C). This might indicate that mitochondrial diversity is more strongly pronounced at the phosphorylation than the protein level.

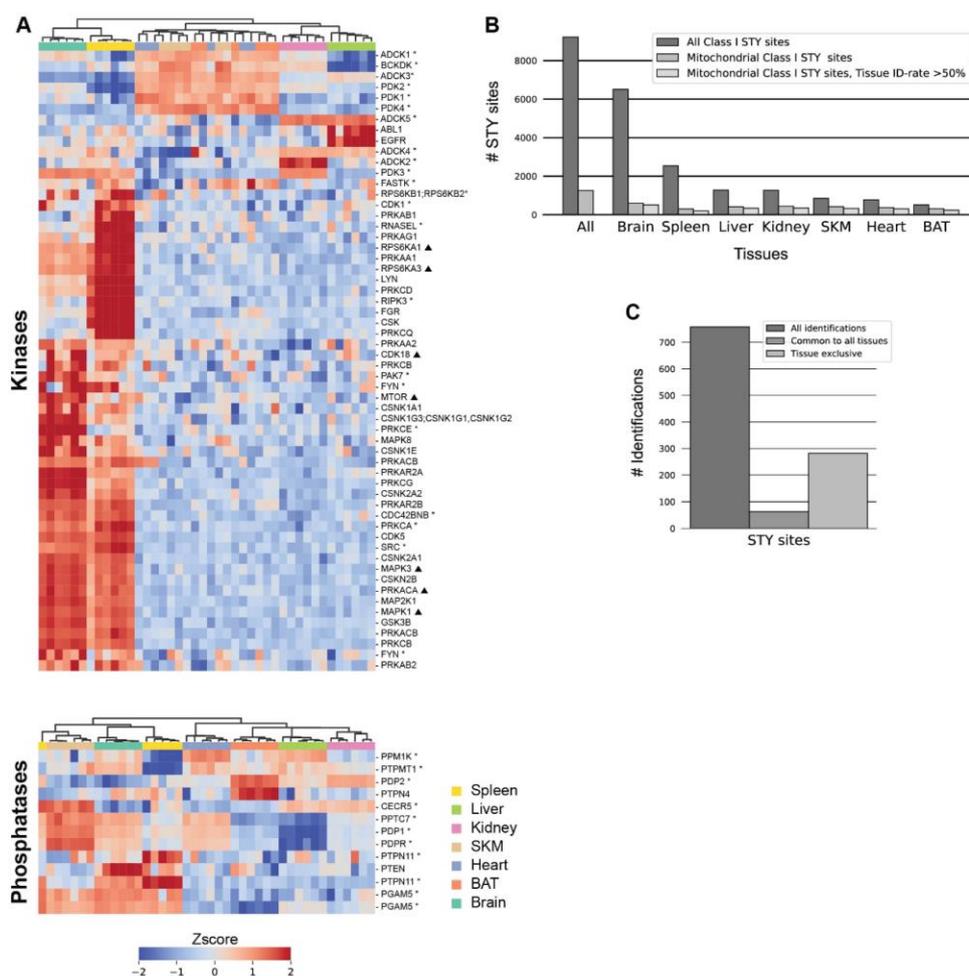


Figure 5 Tissue specificity of mitochondrial kinases and phosphatases

(A) Z-scored protein abundances for predicted (triangle), known (star) (based on MitoCarta3.0 and IMPI database), and manually curated (Figure 5 - Source Data 1) mitochondrial kinases (top) and phosphatases (bottom) across analyzed tissues. Skeletal muscle (SKM), brown adipose tissue (BAT). (B) STY site identification numbers. (C) Mitochondrial STY site numbers after filtering for mitochondrial STY sites identified in at least 5 out of 6 biological replicates in one tissue. Identification numbers for all identified STY sites (left), STY sites common to all tissues (middle), and STY sites exclusive to one tissue (right) are shown. Data in this figure is based on the analysis of six replicates (n=6) for each tissue. Skeletal muscle (SKM), brown adipose tissue (BAT). See also Figure 5 – Source Data 1.

Mitochondrial phosphoproteomes exhibit extensive intra-mitochondrial phosphorylation

To understand the distribution of mitochondrial phosphoproteins across mitochondrial compartments we examined the sub-mitochondrial localization based on the curated MitoCarta3.0 annotation. Mitochondria are typically divided into four main compartments, i.e. mitochondrial outer membrane (OMM), intermembrane space (IMS), inner mitochondrial membrane (IMM) and matrix, although the complex organization of the IMM possibly may define additional compartments (Colina-Tenorio et al., 2020). In line with the high proportion of shared mitochondrial proteomes across all tissues (Figure 1C), the overall localization of proteins was not different between tissues and closely resembled the distribution of all annotated mitochondrial proteins in the database (Figure 6A). However, when performing the same analysis using the phosphorylated mitochondrial proteins, we observed a significant shift towards a localization to the OMM in all tissues (adj. p-values $<6.4 \times 10^{-10}$) (Figure 6A, Figure 6 – supplement 1). Surprisingly, our data also showed that depending on the tissue type, more than 60% of phosphorylated mitochondrial proteins had an intra-mitochondrial annotation - IMM, IMS or matrix localization - (Figure 6A, Figure 6 – supplement 1). Interestingly, 36-54% of OMM, but only 7-22% of intra-mitochondrial proteins were phosphorylated. Here, especially brain (10%) and spleen (7%) tissues showed low intramitochondrial phosphorylation rates.

This prompted us to further investigate the localization of specific kinase-substrate associations (KSA) across sub-mitochondrial localizations using NetworKin3.0 (Horn et al., 2014). Prominently, more than 40% of the predicted KSA in the IMM were linked to the PKC kinase family (Figure 6B). Studies have already reported the localization of PKC kinase family members to mitochondria, as well as an increased phosphorylation of the IMM protein COX IV after PKC ϵ activation (Baines et al., 2002; Jaburek et al., 2006; Majumder et al., 2000; Ogbi and Johnson, 2006; Ping et al., 2002). Moreover, the MAPK group appeared to act on proteins localized to the OMM and matrix, while the PDHK family was specifically associated with the matrix proteins (Figure 6B). The members of the latter kinase family are known to localize to the

mitochondria matrix (Hitosugi et al., 2011), further supporting the validity of identified KSA. However, molecular studies are needed to investigate such KSA, whether phosphorylation of intra-mitochondrial proteins occurs *in situ* or outside mitochondria before being imported into mitochondria, how and which kinases/phosphatases translocate to or into mitochondria and whether these phosphorylation events are functionally relevant.

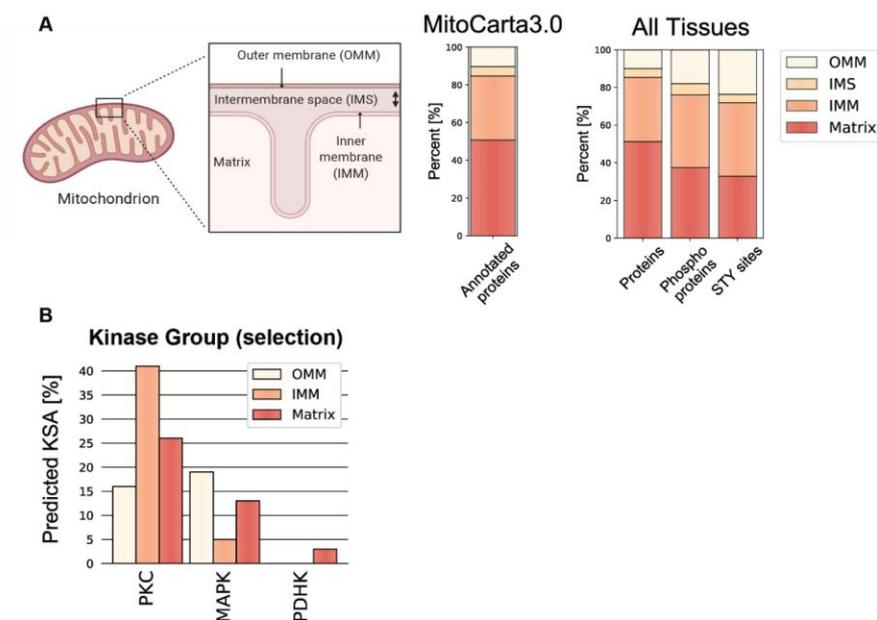


Figure 6 Localization distribution of mitochondrial phosphoproteome diverges from mitochondrial proteome

(A) Simplified scheme of a mitochondrion with four different mitochondrial localizations – outer mitochondrial membrane (OMM), intermembrane space (IMS), inner mitochondrial membrane (IMM), matrix – and the distribution of mitochondrial proteins contained in and classified by the MitoCarta3.0 database. Bar graphs show the precentral distribution of mitochondrial proteins (left), phosphoproteins (middle) and STY sites (right) across different mitochondrial localizations. (B) Predicted Kinase substrate associations (KSA) by the NetworKin3.0 tool for selected kinase families. Data in this figure is based on the analysis of six replicates for 7 different tissues. See also Figure 6 – Source Data 1.

Mitochondrial phosphoproteome reveals tissue-specific modulation of fusion and fission events

The tissue-specific phosphorylation of mitochondrial proteins suggested functional differences in mitochondria and prompted us to investigate the influence of phosphorylation on mitochondrial dynamics. We focused on proteins involved in mitochondrial fusion and fission, two important counteracting events involved in organelle distribution, size balancing and maintenance of a healthy mitochondrial network (Liu et al., 2020; Silva Ramos et al., 2019). Especially proteins involved in the fission process are regulated by a range of protein modifications, including phosphorylation (Cribbs and Strack, 2007; Taguchi *et al.*, 2007; van der Bliek et al., 2013).

Throughout all tissues, MIGA1 (FAM73A) and MIGA2 (FAM73B), two homologues regulating mitochondrial fusion by functioning downstream of the mitofusins, showed different abundances (Figure 6 – supplement 2). This is especially interesting since MIGA1 and MIGA2 can form hetero and homodimers, highlighting a different regulation of fusion in different tissues (Zhang et al., 2016). Moreover, we identified multiple phosphorylation site clusters on MIGA2, while none were identified on MIGA1 (Figure 6 – supplement 2). Intriguingly, similar phosphorylation clusters were identified on Miga in *Drosophila melanogaster* (Xu et al., 2020). Interestingly, two phosphorylation sites on Miga, S246 and S249, were reported to be essential for Vap33 interaction and the establishment of endoplasmic reticulum–mitochondria contact site (ERMCS), suggesting that phosphorylation on MIGA2 has similar functions in mammals (Xu *et al.*, 2020).

Moreover, we observed that GTPase dynamin-related protein 1 (DRP1), a crucial player initiating mitochondrial fission (Bleazard et al., 1999; Cereghetti et al., 2008), displayed higher abundance in brain compared to other tissues (Figure 7A). This observation supports the importance of mitochondrial fission in neurons, where mitochondria switch to a fragmented morphology to enter and travel through axons

(Lewis *et al.*, 2018). Additionally, we also found elevated levels of serine 622 phosphorylation on DRP1 in brain tissue. This site has been shown to regulate DRP1 translocation to mitochondria (Cereghetti *et al.*, 2008; Cribbs and Strack, 2007; Taguchi *et al.*, 2007), indicating that it is actively localized to mitochondria, likely to regulate constant fission events in the brain tissue (Figure 7B).

In mammals, four DRP1 receptor proteins, that are all integral membrane proteins of the OMM, have been reported: mitochondrial fission protein 1 (FIS1), mitochondrial fission factor (MFF), and mitochondrial dynamics protein MiD49 (MIEF1) and MiD51 (MIEF2) (Loson *et al.*, 2013). MFF and FIS1, the fission promoting receptors, were robustly quantified in all tissues and displayed higher abundances in brain and spleen compared to other tissues (Figure 7A). However, MIEF1/2, which counteract DRP1-mediated fission (Dikov and Reichert, 2011; Liu *et al.*, 2013), were generally too low to be robustly quantified in the measured tissues. Additionally, we detected higher levels of MFF phosphorylation at the serine 129 and 146 residues in brain tissue compared to all other tissues in which they were detected (Figure 7C and 7D). Both sites are essential for the recruitment of DRP1 and initiation of fission (Ducommun *et al.*, 2015; Lewis *et al.*, 2018; Toyama *et al.*, 2016).

Elevated DRP1 levels have been shown to increase ROS levels (Watanabe *et al.*, 2014). Intriguingly, we identified oxidation resistance 1 (OXR1), exclusively in the mitochondria of brain tissue, where it plays an important role in the protection of neuronal cells from oxidative stress (Volkert and Crowley, 2020). This likely indicates a protective function of OXR1 in brain tissue as a response to the prevalent fragmented organellar morphology induced by DRP1. In addition, we found OXR1 to be hyperphosphorylated and three out of 12 high confidence sites were novel (Figure 7E). Tissue specificity and the lack of functional annotation of phosphorylation sites that are identified in our study make OXR1 an exciting candidate to be investigated in the future.

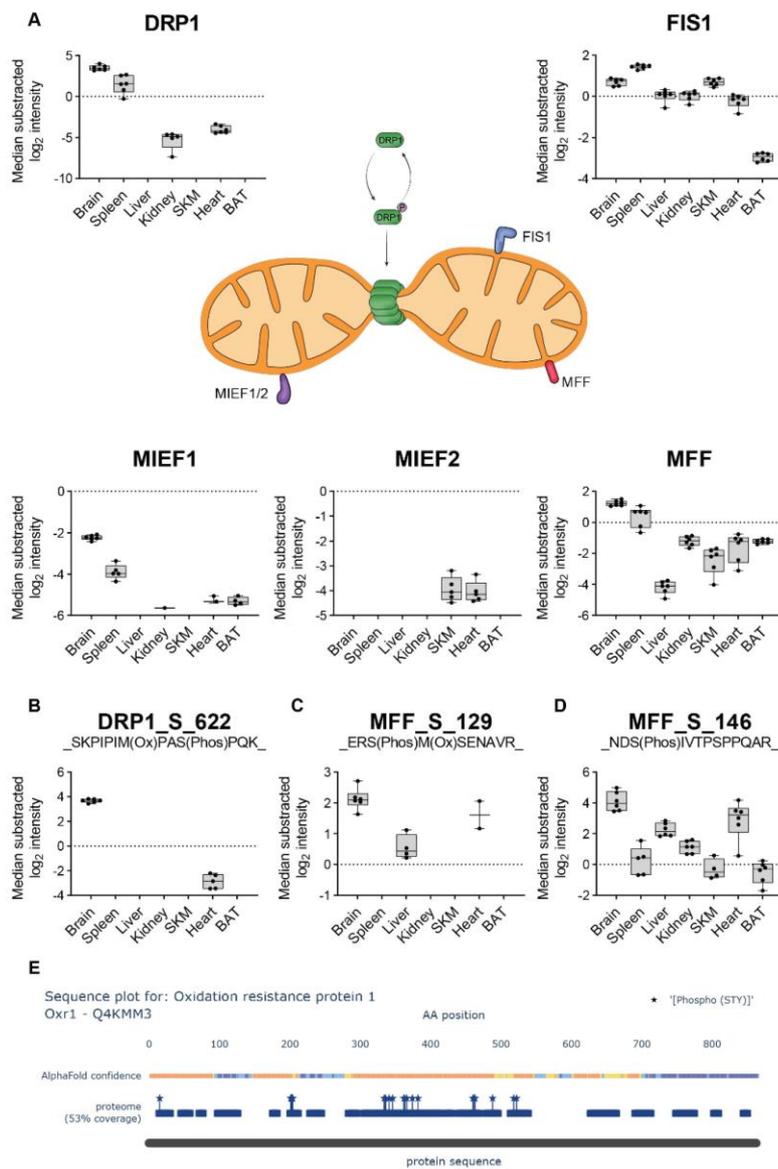


Figure 7 Phosphoproteome reveals tissue-specific functionality for mitochondrial fission

(A) Scheme displays the reversible phosphorylation and the connected localization change to mitochondria of DRP1 (green). Normalized intensities (median of all log₂ transformed mitochondrial proteins of a sample was subtracted from all log₂ protein intensities of that sample) of DRP1 across all analyzed tissues (black dots indicate individual identifications) are displayed in the upper left. Mitochondrial DRP1 receptors – FIS1 (blue), MIEF1/2 (violet), MFF (red) – and corresponding box plots are shown next

to their receptors. (B) Normalized intensities (median of all \log_2 transformed mitochondrial phosphopeptide of a sample was subtracted from all \log_2 peptide intensities of that sample) of the phosphopeptide showing S622 phosphorylation on DRP1 (black dots indicate individual identifications). (C) Same as (B) showing S129 phosphorylation on MFF. (D) Same as (B) showing S146 phosphorylation on MFF. Significance of protein abundance differences in relation to the reference tissue (brain for DRP1 panel, brain and spleen in FIS1 and MFF panel) were estimated (p-values <0.001, except for SKM in FIS1 panel) by one-way ANOVA analysis (Figure 7 - Source Data 1). Significance of phosphosites abundance differences in (B) and (C, only brain and liver) were estimated by a two-side t-test (Figure 7 - Source Data 1). Significance of MFF_S_146 abundance differences in relation to the reference tissue (brain) were estimated (p-value<0.05) by one-way ANOVA analysis (Figure 7 - Source Data 1). Data in this figure is based on the analysis of six replicates (n=6). Skeletal muscle (SKM), brown adipose tissue (BAT). See also Figure 7 – Source Data 1.

Web application makes Mouse Mitochondria Atlas data readily accessible

As indicated by the above examples, this study presents a rich resource to explore the mitochondrial proteomes and phosphoproteomes across mouse tissues. Preceding examples show the potential of this resource for investigation of tissue-specific mitochondrial regulations on the proteome and phosphoproteome level, ultimately permitting the generation and analysis of new hypotheses. However, utilization of such resources largely depends on the ease of data access for exploration.

To this end, we created a web application mitophos.de offering the end user an interface to easily explore datasets, including MitoCarta3.0 networks, abundance comparisons across tissues and sequence analysis (Figure 8A). As an example, Figure 8 illustrates the MICOS complex, which has a central role in mitochondria (Khosravi and Harner, 2020). In the network view one can see (I) members of the complex as well as the phosphorylation sites identified on these proteins and (II) in which tissues and how reproducibly they are identified in our dataset (Figure 8B). Moreover, the user can inspect individual abundance distributions of all identified proteins/STY sites across all measured tissues. For instance, MINOS1, a core component of the MICOS complex (Bohnert et al., 2015), displayed high abundance in

mitochondria isolated from heart, SKM and BAT tissues (Figure 8C). Moreover, in the sequence view the AlphaMap tool (Voytik et al., 2021) is integrated to map all identified peptides, including phosphorylated peptides, onto their respective protein sequence along with structural information such as topological domains and transmembrane regions (Figure 8D) and to visualize phosphorylation sites in their 3D structures (unpublished data) as predicted by AlphaFold (Jumper et al., 2021).

Together, this data-rich and comprehensive tool is an entry point to investigate the herein presented resource and will assist in future efforts to functionally characterize mitochondrial proteins and their respective phosphorylation sites.

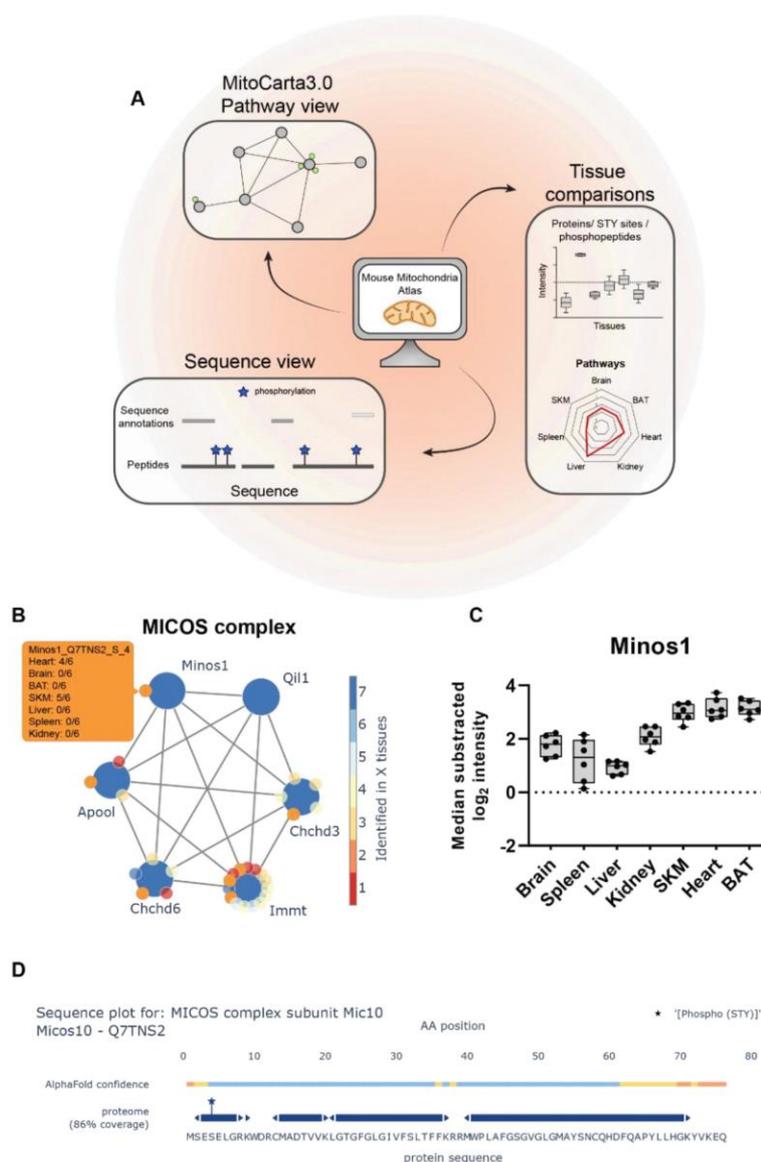


Figure 8 Web application readily enables easy data access

(A) Scheme of Mouse Mitochondria Atlas application features (mitophos.de). (B) MICOS complex view based on MitoCarta3.0 annotation. Large and small nodes represent proteins and class I STY sites, respectively. Edges represent on String interaction scores >0.4 and color of nodes indicate the number of tissues in which proteins/STY sites were identified. (C) Normalized intensities (median of all \log_2 transformed mitochondrial proteins of a sample was subtracted from all \log_2 protein intensities of

that sample) for Minos1 across all analyzed tissues (black dots indicate individual identifications, box and error bar). (D) Sequence plot of Micos1 shows structural information (based on UniProt annotations) and protein coverage based on identified peptides. All identified STY sites are marked with a star. Data in this figure is based on the analysis of six replicates (n=6). Skeletal muscle (SKM), brown adipose tissue (BAT).

Discussion

Here we present a tissue-specific atlas of mouse mitochondrial proteomes and phosphoproteomes, an in-depth resource towards a better understanding of the composition and function of this vital organelle in a tissue-specific manner. Previous MS-based studies combining mitochondrial phosphoproteome and proteome measurements typically focused on a single or few tissues and were generally shallower than our study. In addition, differences in study designs such as utilization of various organisms or mitochondria/phosphopeptide enrichment protocols, analysis pipelines, and mitochondrial protein annotation databases, complicate the integration of such datasets to understand tissue specificity. We now globally and precisely quantified different protein expression and phosphorylation patterns at subcellular level across seven mouse tissues, providing a detailed view on mitochondrial diversity. The breadth and depth of coverage achieved by the integration of tissue-specific mitochondrial proteomes and phosphoproteomes provide unbiased insights into the mitochondrial composition and function. This allows the generation and assessment of novel hypotheses related to mitochondrial biology, which cannot be generated with focused studies alone. Integrated mitochondrial proteomes and phosphoproteomes that are diverse between tissues can readily be explored at mitophos.de.

Our data revealed that the functional diversity of mitochondria is defined by protein abundance rather than compositional differences since more than half of the mitochondrial proteome was shared by all analyzed tissues and 90% by at least two tissues. For instance, the electron transport chain is an integral part of the mitochondrial composition and its components are found across all tissues, however, their abundance shows substantial differences to meet tissue-specific energy demands. Thus, dysregulation of individual proteins can strongly affect mitochondria in one tissue, leading to severe diseases, while mitochondria in a different tissue remain largely unaffected. Moreover, we found that 9% of the proteome displays tissue specificity, further contributing to our understanding of tissue-specific effects of

mitochondrial protein dysregulation (Russell et al., 2020). This is an important concept, which can now be studied in an unbiased manner using our resource data, providing opportunities for development of targeted treatments for mitochondrial diseases. For instance, members of the SLC25 family are linked to metabolic diseases in distinct tissues (Palmieri and Monne, 2016) and various cancers (Rochette et al., 2020). However, biological functions of a large repertoire of mitochondrial SLCs are still unknown. For example, inactivation of SLC25A25 was assessed in mouse skeletal muscle tissue where it caused a reduced metabolic efficiency (Anunciado-Koza et al., 2011). Given its high abundance in mitochondria of brain tissue and in glioma cells (Traba et al., 2012), it will also be interesting to investigate its role in this tissue, particularly whether SLC25A25 deficiency influences neuronal fitness.

Mitochondria are essential cellular entities that are involved in a wide variety of cellular processes (McBride et al., 2006) through dynamic interaction and constant communication with other organelles such as the (ER), nucleus and peroxisomes via membrane contact sites (Desai et al., 2020; Perrone et al., 2020; Shai et al., 2018) or protein complexes, such as the ribosome (Lashkevich and Dmitriev, 2021). We suggest that non-mitochondrial proteins identified in the samples might present signatures that could convey important biological information regarding mitochondria-associated structures. Given the high level of mitochondrial enrichment combined with highly reproducible LC-MS/MS measurements, such protein signatures are unlikely to be solely based on unspecific enrichment of abundant proteins. For example, the proteasome is robustly identified in most of the mitochondria enriched samples, which is in line with its involvement in the degradation of misfolded mitochondrial proteins (Basch et al., 2020; Kodron et al., 2021). We also prominently observed that several members of the cytochrome P450 superfamily are enriched in liver mitochondria, demonstrating the versatile interaction of mitochondria with their cellular environment. Furthermore, our data identified non-mitochondrial ribosomal proteins in all tissues, which could be explained by the local translation of nuclear-encoded mitochondrial mRNAs (Lashkevich and Dmitriev, 2021). It was recently shown that RNA-bearing late endosomes associate with

mitochondria and ribosomes forming hotspots of local protein synthesis in axons (Cioni et al., 2019) and that mitochondria fuel such local translation in neurons, enabling synaptic plasticity (Rangaraju et al., 2019).

There is mounting evidence that phosphorylation of mitochondrial proteins fulfills important functions to maintain cellular health as exemplified by the fission process in this study. Deregulation of mitochondrial protein phosphorylation can lead to diseases such as cancer, diabetes, heart and neurological disorders. It was recently reported that 91% of mitochondrial proteins on MitoCarta3.0 have at least one phosphorylation site reported on the PSP database (Niemi and Pagliarini, 2021). However, this analysis also includes proteins that do not always localize to mitochondria (Ben-Menachem et al., 2011) and phosphorylation sites that can specifically be captured upon perturbations and stimuli. Our study revealed that around half of the mitochondria-localized proteins were phosphorylated in tissues at steady state - 39% in brain, 15% in spleen, 29% in liver, 28% in kidney, 32% in SKM, 31% in heart and 22% in BAT. This suggests that the mitochondrial proteome and phosphoproteome compositions are dynamically modulated in response to environmental changes. Furthermore, we identified over 60 kinases and 10 phosphatases that either are localized to mitochondria or associate with mitochondria, providing a global view on important modulators of mitochondrial protein phosphorylation. Mapping tissue-specific mitochondrial kinases and phosphatases is an important step towards understanding their role in the regulation of the mitochondrial phosphoproteome in different tissues and hence developing therapeutics for mitochondrial diseases. For example, a mitochondrial phosphatase, phosphoglycerate mutase family member 5 (PGAM5), has recently emerged as an important regulator of mitochondrial homeostasis. Deletion of PGAM5 has been shown to result in Parkinson's-like movement disorder in mice (Lu et al., 2014) and T cell dysfunction in primary cells (Panda et al., 2016). While its diverse roles largely remain to be uncovered (Liang et al., 2021), a novel PGAM5 inhibitor was recently suggested as a potential therapeutic for brain ischemic stroke (Gao et al., 2021). We observed that PGAM5 displays elevated levels

in the mitochondria isolated from brain, SKM and spleen, possibly explaining the tissue-specific phenotypes induced by its absence and where in the body molecules targeting its phosphatase activity would exert their effects.

The data in this study will contribute to our understanding of the tissue-specific composition and function of mitochondria and serve as a gateway for investigation of specific questions related to mitochondrial biology. Future biochemical and more focused investigations are needed to validate our findings and test the hypotheses arising from our study. For instance, it is pivotal to experimentally validate kinase-substrate associations of previously unknown phosphorylations on mitochondrial proteins and their functional implications in the cell. Additionally, the impact of those phosphorylations on the localization of target proteins and, more specifically, the question of whether mitochondrial proteins are phosphorylated before or after entering the mitochondria remain to be investigated. Accessibility, for instance, of previously undescribed phosphorylation sites on mitochondrial proteins can be assessed using advanced structural tools (Jumper *et al.*, 2021) to determine if they are likely targeted by a mitochondrial kinase or a cytoplasmic kinase before being imported (Schober *et al.*, 2021). Moreover, future developments towards better enrichment strategies for the isolation of mitochondria from different tissues and advances in the MS technology will aid to further improve the depth and quality of the mitochondrial proteomes and phosphoproteome.

Materials and Methods

Experimental model and subject details

Six C57BL/6N mice (3male, 3 female) were housed in a 12-hours light/dark cycle in standard ventilated cages under specific-pathogen-free conditions with constant temperature (21°C) and humidity (50 - 60%) and fed ad libitum with a standard mouse diet. At the age of 18-21 weeks mice were sacrificed by cervical dislocation. The study was approved by the by the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein–Westfalen, Germany, and performed in accordance with European law.

Tissue preparation and isolation of ultra-pure mitochondria

Mice were sacrificed by cervical dislocation and the 7 tissues - heart, skeletal muscle (SKM), brown adipose tissue (BAT), spleen, kidney, liver, and brain – were rapidly removed. Heart, spleen, and kidney, tissues were homogenized in mitochondrial isolation buffer containing 320 mM sucrose, 1 mM EDTA, and 10 mM Tris-HCl, pH 7.4, supplemented with EDTA-free complete protease inhibitor cocktail and PhosSTOP tablets (Roche). For isolation of mitochondria from BAT, liver, and brain the mitochondrial isolation buffer was additionally supplemented with 0.2% bovine serum albumin (Sigma-Aldrich). Subsequently, crude mitochondria were isolated from the homogenates by two rounds of differential centrifugation (see Figure 1a). Isolation of crude mitochondria from SKM was performed as previously described (Frezza et al., 2007)). Crude mitochondrial pellets from all tissues were further purified on a Percoll density gradient as described recently (Kuhl *et al.*, 2017). Briefly, mitochondrial pellets were washed once in 1xM buffer (220 mM mannitol, 70mM sucrose, 5mM HEPES pH 7.4, 1 mM EGTA pH 7.4; pH was adjusted with potassium hydroxide; supplemented with EDTA-free complete protease inhibitor cocktail and PhosSTOP tablets (Roche)) and subsequently purified on a Percoll (GE healthcare) density gradient of 12%:19%:40% via centrifugation in a SW41 rotor at 42, 000 g at 4°C for 30 min in a Beckman Coulter Optima L- 100 XP ultracentrifuge using 14 mm × 89 mm Ultra-Clear Centrifuge Tubes (Beckman Instruments Inc.). Ultra-

pure mitochondria were harvested at the interphase between 19% and 40% and washed three times with 1xM buffer. Dry mitochondrial pellets were snap-frozen in liquid nitrogen and stored at -80°C until further use.

Mitochondrial (phospho)proteome sample preparation

Frozen ultra-pure mitochondria pellets were resuspended in lysis buffer (4%SDC, 100mM Tris/HCl, pH8.5), boiled for 5 min at 95°C and sonicated in 30 s intervals for 15 min (Bioruptor). Protein concentration was estimated via Tryptophan assay (Kulak et al., 2014) and was adjusted with lysis buffer to a total volume of 270 µl containing 400 µg of protein for brain, SKM, liver, heart, kidney, and 140 µg of protein for BAT samples. Proteins were reduced and alkylated by adding 30 µl of 10x reduction/alkylation solution (100 mM Tris (2-carboxyethyl)phosphine hydrochloride (TCEP) and 400 mM 2-chloroacetamide (CAA), followed by 5min incubation at 45°C. Subsequently, 1:100 Trypsin and LysC were added for overnight protein digestion at 37°C. For proteome analysis 10 µl (brain, SKM, liver, heart, kidney) and 20 µl (BAT, spleen) aliquots were taken and loaded on SDB-RPS StageTips. Peptides were washed with 200µl wash buffer (0.2% TFA/2% ACN (vol/vol)) and then eluted with SDB-RPS elution buffer (1.25% NH₄OH, 80% ACN (vol/vol)) and dried in a SpeedVac. Dried peptides were resuspended in A* buffer (2% ACN/0.1% TFA).

The remaining samples were processed following the EasyPhos protocol for phosphopeptide enrichment (Humphrey *et al.*, 2018). In brief, samples were first mixed with isopropanol and EP buffer (48% TFA, 8 mM KH₂PO₄), followed by phosphopeptide enrichment with 5mg TiO₂ beads per sample (GL Sciences). For this, samples were mixed with TiO₂ beads in loading buffer (6% TFA/80% ACN (vol/vol)) at a concentration of 1 mg/µl and incubated for 5 min at 40°C by shaking at 1200rpm. Subsequently beads were washed four times with 1 ml of wash buffer (5% TFA,60% isopropanol (vol/vol)) and phosphopeptides were eluted from beads using 60 µl of elution buffer (40% ACN, 5% NH₄OH) and concentrated in a SpeedVac for 30 min at 45°C. Samples were immediately diluted with 100 µl of SDBRPS loading buffer (99% isopropanol, 1% TFA

(vol/vol)) and loaded on SDB-RPS StageTips. Thereafter, phosphopeptides were washed and eluted as described above and resuspended in 6 μ l A*.

LC-MS/MS

For all measurements peptides were loaded onto a 50cm, in-house packed, reversed-phase column (75 μ m inner diameter, 1. diameter, ReproSil-Pur C18-AQ 1.9 μ m resin [Dr. Maisch GmbH]) and separated with and binary buffer system consisting of buffer A (0.1% formic acid (FA)) and buffer B (0.1% FA in 80% ACN). The column temperature was controlled by a homemade column oven and maintained at 60°C. For nanoflow liquid chromatography an EASY-nLC 1200 system (Thermo Fisher Scientific), directly coupled online with a Q Exactive HF-X (Thermo Fisher Scientific) via a nano-electrospray source, was operated at a flow rate of 300 nl/min and 350 nl/min for mitochondrial proteome and phosphoproteome measurements, respectively.

For mitochondrial proteome measurements 500 μ g of peptides were loaded and separated using a gradient starting at 5% buffer B, increasing to 30% buffer B in 80 min, 60% buffer B in 4 min and 95% buffer B in 4 min. The MS was operated in DDA mode (Top12) with a full scan range of 300-1650 m/z and a MS1 and MS2 resolution of 60,000 and 15,000, respectively. The automatic gain control (AGC) was set to 3e6 and 1e5 for MS1 and MS2, while the maximum injection time was set to 20 ms and 60 ms, respectively. Precursor ion selection width was kept at 1.4 m/z and fragmentation was achieved by higher-energy collisional dissociation (HCD) (NCE 27%). Dynamic exclusion was enabled and set to 20 s.

For mitochondrial phosphoproteome measurements, 5 μ l as loaded and separated using a gradient starting at 3% buffer B, increasing to 19% buffer B in 40 min, 41% buffer B in 20 min and 90% buffer B in 5 min. The MS was operated in DDA mode (Top10) with a full scan range of 300-1600 and a MS1 and MS2 resolution of 60,000 and 15,000, respectively. The automatic gain control (AGC) was set to 3e6 and 1e5 for MS1 and MS2, while the maximum injection time was set to 120 ms and 60 ms, respectively. Precursor

ion selection width was kept at 1.6 m/z and fragmentation was achieved by higher-energy collisional dissociation (HCD) (NCE 27%). Dynamic exclusion was enabled and set to 30 s.

Raw data analysis

DDA raw data were analyzed with MaxQuant (1.6.14.0) against the mouse fasta file (downloaded 19. October 2020) using default settings. PSM and protein data discovery rate were controlled at 1% FDR. The match between runs (MBR) functionality was enabled and set in a way that only biological replicates belonging to the same tissue type were allowed to match each other. This eliminates the possibility of potentially false MBR identification transfer between tissues. Carbamidomethyl (C) was selected as fixed modification and Acetyl (Protein N-term) and oxidation (M) were defined as variable modifications. For mitochondrial phosphoproteome analysis, STY site phosphorylation was additionally selected as variable modification.

Bioinformatics analysis

Data analysis was performed using the python programming language using python (3.8.12) and the following packages: alphamap, matplotlib (3.5.0), mygene (3.2.2), numpy (1.19.2), pandas (1.1.3), pyteomics (4.3.3), requests (2.26.0), scipy (1.7.2), seaborn (0.11.2), sklearn (0.0), upsetplot (0.6.0). All notebooks used for data analysis are available at GitHub (<https://github.com/MannLabs>). Identified proteins were filtered for at least 3 valid values in at least one tissue. Similarly, phosphorylation sites were filtered for at least 5 valid values in at least one tissue and a localization probability >75%. The UniProt API was used to map 'ACC+ID' protein group identifiers provided by the MQ analysis to 'ENSEMBL_ID', 'P_ENTREZGENEID' and 'STRING_ID' for further analysis. 'ENSEMBL_ID' and 'P_ENTREZGENEID' identifier were used for mitochondrial protein annotation based on the IMPI (IMPI_2020_Q3, downloaded 27. October 2020) and MitoCarta3.0 (downloaded 1. January 2021) databases, respectively. Network analysis and visualization were performed with the StringApp (1.6.0) in Cytoscape (3.8.2). Kinase annotations are

based on manual annotations (Figure 6 - Source Data 1) and 'pkinfam' (downloaded 9. April 2021). Networkin3.0 was used for kinase substrate association (KSA) predictions, while the Networkin score was set to 1. Gene Ontology (GO) annotations for the GOBP term enrichment analysis were retrieved from UniProt (accessed 8. March 2021). Enrichment analysis was performed in Perseus (1.6.7.0) against the set of identified proteins in the corresponding tissue and the results were filtered for an intersection size >10. Missing values were only imputed for PCA and heatmap analysis. For this, a Gaussian normal distribution with a width of 0.3 relative to the standard deviation of measured values and a downshift of 1.8 standard deviations were used. For data normalization, intensity values were \log_2 transformed and then filtered for known and predicted mitochondrial proteins. The median value of these mitochondrial proteins was subtracted from all \log_2 transformed values. Significance testing for individual proteins and phosphopeptides as shown in Figure 3 and Figure 7 was performed with the ordinary one-way ANOVA method or by two-sided t-tests in GraphPad Prism (9.3.1) (Figure 3 - Source Data 1, Figure 7 - Source Data 1). Significance testing for differences in mitochondrial protein, phosphoprotein and phosphosite localization was performed in RStudio (1.3.1093) (Figure 6 - Source Data 1). OMM proportions were used for fitting a beta-regression model using the betareg R package with default settings (Cribari-Neto and Zeileis, 2010; Grun et al., 2012). P-values were estimated with the lrtest function of the lrttest package (Zeileis and Hothorn, 2015) and p-values were adjusted with the "fdr" method of the p.adjust function of the stats base package (Benjamini and Hochberg, 1995)

Website tool

The website tool is structured into four sections. The first three 'Pathway view', 'Sequence view', and 'Tissue comparison' are for displaying data, while the fourth section provides explanations for each individual section. In the 'Pathway view' and 'Tissue comparison', proteome and phosphoproteome data filtered for at least three and 5 identifications in at least one tissue, respectively. Intensity values were normalized as described above and used for data representation in the 'Tissue comparison' tab or z-scored

across tissues and used for the 'Pathway view tab'. Here, median z-score values of the six biological replicates per tissue are displayed in the data table. The polar plot represents the median z-score of all pathway/complex members of a given tissue. Network/complex annotations were retrieved from MitoCarta3.0 and protein interactions are based on STRING interaction scores. These STRING interaction scores were retrieved from STRING (17. November 2021) using the 'STRING_ID' and the STRING API. For the 'Sequence view', the 'evidence.txt' of the MaxQuant output files was directly used as input to annotated sequences.

The python programming language was used for data processing and visualization for the Dashboard. The following libraries were used for data processing: numpy (1.19.2), pandas (1.19.2), re, sys, os, and pyteomics (4.3.3). Several libraries from the HoloViz family of tools were used for data visualization and creation of the dashboard, including panel (1.14.6), holoviews (1.14.6), bokeh (2.2.2), plotly (4.12.0), and param (1.10.0). Network visualization was achieved with the NetworkX package (Hagberg et al., 2008). The Alphamap tool (Voytik *et al.*, 2021) was integrated to display linear protein sequence annotations as well as to visualize 3D protein structures.

Resource availability

Data and code availability

Datasets generated in this study have been deposited at ProteomeXchange and are publicly available as of the date of publication. The accession number is listed in the key resource table. (Identifier: PXD030062).

All original code has been deposited on GitHub (<https://github.com/MannLabs>) and is available as of the date of publication.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgments

This work was supported by grants to N.G.L from the Swedish Research Council (2015-00418), Swedish Cancer Foundation (2021.1409), the Knut and Alice Wallenberg foundation, European Research Council (ERC Advanced Grant 2016-741366), grants from the Swedish state under the agreement between the Swedish government and the county councils (SLL2018.0471). L.S.K was supported by an EMBO long-term fellowship (ALTF 570-2019). I.K. was supported by the French Muscular Dystrophy Association (AFM-Téléthon #23294) and Agence nationale de la recherche (ANR-20-CE12-0011). We thank all members of the Department of Signal Transduction and Proteomics, in particular, Isabel Bludau for helpful discussions regarding the website tool, Igor Paron for MS technical assistance, Mario Oroshi for technical assistance during website setup, Marta Murgia for helpful discussions on mitochondrial biology and Johannes Müller-Reif, Patricia Skowronek and Sophia Steigerwald for columns. We are grateful to Jesper Olsen and Brenda Schulman for constructive and insightful discussions. Figure 1A and Figure 6A were partly created with BioRender.com.

Author contributions

F.M.H., L.S.K. designed experiments. Mouse work was performed by L.S.K. and I.K.. Proteomic experiments were conducted by F.M.H. and O.K.. Website was constructed by F.M.H. and I.B.. Data were analyzed by F.M.H.. All authors contributed to writing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

References

- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* *290*, 457-465. 10.1038/290457a0.
- Anunciado-Koza, R.P., Zhang, J., Ukropec, J., Bajpeyi, S., Koza, R.A., Rogers, R.C., Cefalu, W.T., Mynatt, R.L., and Kozak, L.P. (2011). Inactivation of the mitochondrial carrier SLC25A25 (ATP-Mg²⁺/Pi transporter) reduces physical endurance and metabolic efficiency in mice. *J Biol Chem* *286*, 11659-11671. 10.1074/jbc.M110.203000.
- Aponte, A.M., Phillips, D., Hopper, R.K., Johnson, D.T., Harris, R.A., Blinova, K., Boja, E.S., French, S., and Balaban, R.S. (2009). Use of (32)P to study dynamics of the mitochondrial phosphoproteome. *J Proteome Res* *8*, 2679-2695. 10.1021/pr800913j.
- Avadhani, N.G., Sangar, M.C., Bansal, S., and Bajpai, P. (2011). Bimodal targeting of cytochrome P450s to endoplasmic reticulum and mitochondria: the concept of chimeric signals. *FEBS J* *278*, 4218-4229. 10.1111/j.1742-4658.2011.08356.x.
- Baines, C.P., Zhang, J., Wang, G.W., Zheng, Y.T., Xiu, J.X., Cardwell, E.M., Bolli, R., and Ping, P. (2002). Mitochondrial PKCepsilon and MAPK form signaling modules in the murine heart: enhanced mitochondrial PKCepsilon-MAPK interactions and differential MAPK activation in PKCepsilon-induced cardioprotection. *Circ Res* *90*, 390-397. 10.1161/01.res.0000012702.90501.8d.
- Bak, S., Leon, I.R., Jensen, O.N., and Hojlund, K. (2013). Tissue specific phosphorylation of mitochondrial proteins isolated from rat liver, heart muscle, and skeletal muscle. *J Proteome Res* *12*, 4327-4339. 10.1021/pr400281r.
- Basch, M., Wagner, M., Rolland, S., Carbonell, A., Zeng, R., Khosravi, S., Schmidt, A., Aftab, W., Imhof, A., Wagener, J., et al. (2020). Msp1 cooperates with the proteasome for extraction of arrested mitochondrial import intermediates. *Mol Biol Cell* *31*, 753-767. 10.1091/mbc.E19-06-0329.
- Bayraktar, E.C., Baudrier, L., Ozerdem, C., Lewis, C.A., Chan, S.H., Kunchok, T., Abu-Remaileh, M., Cangelosi, A.L., Sabatini, D.M., Birsoy, K., and Chen, W.W. (2019). MITO-Tag Mice enable rapid isolation and multimodal profiling of mitochondria from specific cell types in vivo. *Proc Natl Acad Sci U S A* *116*, 303-312. 10.1073/pnas.1816656115.
- Bekker-Jensen, D.B., Bernhardt, O.M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C.D., Reiter, L., and Olsen, J.V. (2020). Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat Commun* *11*, 787. 10.1038/s41467-020-14609-1.
- Ben-Menachem, R., Tal, M., Shadur, T., and Pines, O. (2011). A third of the yeast mitochondrial proteome is dual localized: a question of evolution. *Proteomics* *11*, 4468-4476. 10.1002/pmic.201100199.
- Beninca, C., Planaguma, J., de Freitas Shuck, A., Acin-Perez, R., Munoz, J.P., de Almeida, M.M., Brown, J.H., Murphy, A.N., Zorzano, A., Enriquez, J.A., and Aragay, A.M. (2014). A new non-canonical pathway of Galpha(q) protein regulating mitochondrial dynamics and bioenergetics. *Cell Signal* *26*, 1135-1146. 10.1016/j.cellsig.2014.01.009.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* *57*, 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bleazard, W., McCaffery, J.M., King, E.J., Bale, S., Mozdy, A., Tieu, Q., Nunnari, J., and Shaw, J.M. (1999). The dynamin-related GTPase Dnm1 regulates mitochondrial fission in yeast. *Nat Cell Biol* *1*, 298-304. 10.1038/13014.

- Bock, F.J., and Tait, S.W.G. (2020). Mitochondria as multifaceted regulators of cell death. *Nat Rev Mol Cell Biol* 21, 85-100. 10.1038/s41580-019-0173-8.
- Bohnert, M., Zerbes, R.M., Davies, K.M., Muhleip, A.W., Rampelt, H., Horvath, S.E., Boenke, T., Kram, A., Perschil, I., Veenhuis, M., et al. (2015). Central role of Mic10 in the mitochondrial contact site and cristae organizing system. *Cell Metab* 21, 747-755. 10.1016/j.cmet.2015.04.007.
- Boja, E.S., Phillips, D., French, S.A., Harris, R.A., and Balaban, R.S. (2009). Quantitative mitochondrial phosphoproteomics using iTRAQ on an LTQ-Orbitrap with high energy collision dissociation. *J Proteome Res* 8, 4665-4675. 10.1021/pr900387b.
- Calvo, S.E., and Mootha, V.K. (2010). The mitochondrial proteome and human disease. *Annu Rev Genomics Hum Genet* 11, 25-44. 10.1146/annurev-genom-082509-141720.
- Cereghetti, G.M., Stangherlin, A., Martins de Brito, O., Chang, C.R., Blackstone, C., Bernardi, P., and Scorrano, L. (2008). Dephosphorylation by calcineurin regulates translocation of Drp1 to mitochondria. *Proc Natl Acad Sci U S A* 105, 15803-15808. 10.1073/pnas.0808249105.
- Cioni, J.M., Lin, J.Q., Holtermann, A.V., Koppers, M., Jakobs, M.A.H., Azizi, A., Turner-Bridger, B., Shigeoka, T., Franze, K., Harris, W.A., and Holt, C.E. (2019). Late Endosomes Act as mRNA Translation Platforms and Sustain Mitochondria in Axons. *Cell* 176, 56-72 e15. 10.1016/j.cell.2018.11.030.
- Colina-Tenorio, L., Horten, P., Pfanner, N., and Rampelt, H. (2020). Shaping the mitochondrial inner membrane in health and disease. *J Intern Med* 287, 645-664. 10.1111/joim.13031.
- Cribari-Neto, F., and Zeileis, A. (2010). Beta Regression in R. *J Stat Softw* 34, 1-24.
- Cribbs, J.T., and Strack, S. (2007). Reversible phosphorylation of Drp1 by cyclic AMP-dependent protein kinase and calcineurin regulates mitochondrial fission and cell death. *EMBO Rep* 8, 939-944. 10.1038/sj.embor.7401062.
- Cui, Z., Hou, J., Chen, X., Li, J., Xie, Z., Xue, P., Cai, T., Wu, P., Xu, T., and Yang, F. (2010). The profile of mitochondrial proteins and their phosphorylation signaling network in INS-1 beta cells. *J Proteome Res* 9, 2898-2908. 10.1021/pr100139z.
- del Arco, A., and Satrustegui, J. (2004). Identification of a novel human subfamily of mitochondrial carriers with calcium-binding domains. *J Biol Chem* 279, 24701-24713. 10.1074/jbc.M401417200.
- Deng, N., Zhang, J., Zong, C., Wang, Y., Lu, H., Yang, P., Wang, W., Young, G.W., Wang, Y., Korge, P., et al. (2011). Phosphoproteome analysis reveals regulatory sites in major pathways of cardiac mitochondria. *Mol Cell Proteomics* 10, M110 000117. 10.1074/mcp.M110.000117.
- Deng, W.J., Nie, S., Dai, J., Wu, J.R., and Zeng, R. (2010). Proteome, phosphoproteome, and hydroxyproteome of liver mitochondria in diabetic rats at early pathogenic stages. *Mol Cell Proteomics* 9, 100-116. 10.1074/mcp.M900020-MCP200.
- Desai, R., East, D.A., Hardy, L., Faccenda, D., Rigon, M., Crosby, J., Alvarez, M.S., Singh, A., Mainenti, M., Hussey, L.K., et al. (2020). Mitochondria form contact sites with the nucleus to couple prosurvival retrograde response. *Sci Adv* 6. 10.1126/sciadv.abc9955.
- Dikov, D., and Reichert, A.S. (2011). How to split up: lessons from mitochondria. *EMBO J* 30, 2751-2753. 10.1038/emboj.2011.219.
- Ducommun, S., Deak, M., Sumpton, D., Ford, R.J., Nunez Galindo, A., Kussmann, M., Viollet, B., Steinberg, G.R., Foretz, M., Dayon, L., et al. (2015). Motif affinity and mass spectrometry proteomic approach for the discovery of cellular AMPK targets: identification of mitochondrial fission factor as a new AMPK substrate. *Cell Signal* 27, 978-988. 10.1016/j.cellsig.2015.02.008.
- Fecher, C., Trovo, L., Muller, S.A., Snaidero, N., Wettmarshausen, J., Heink, S., Ortiz, O., Wagner, I., Kuhn, R., Hartmann, J., et al. (2019). Cell-type-specific profiling of brain mitochondria reveals functional and molecular diversity. *Nat Neurosci* 22, 1731-1742. 10.1038/s41593-019-0479-z.
- Feng, J., Zhu, M., Schaub, M.C., Gehrig, P., Roschitzki, B., Lucchinetti, E., and Zaugg, M. (2008). Phosphoproteome analysis of isoflurane-protected heart mitochondria: phosphorylation of adenine

- nucleotide translocator-1 on Tyr194 regulates mitochondrial function. *Cardiovasc Res* *80*, 20-29. 10.1093/cvr/cvn161.
- Ferreira, R., Vitorino, R., Alves, R.M., Appell, H.J., Powers, S.K., Duarte, J.A., and Amado, F. (2010). Subsarcolemmal and intermyofibrillar mitochondria proteome differences disclose functional specializations in skeletal muscle. *Proteomics* *10*, 3142-3154. 10.1002/pmic.201000173.
- Fiermonte, G., De Leonadis, F., Todisco, S., Palmieri, L., Lasorsa, F.M., and Palmieri, F. (2004). Identification of the mitochondrial ATP-Mg/Pi transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J Biol Chem* *279*, 30722-30730. 10.1074/jbc.M400445200.
- Forner, F., Foster, L.J., Campanaro, S., Valle, G., and Mann, M. (2006). Quantitative proteomic comparison of rat mitochondria from muscle, heart, and liver. *Mol Cell Proteomics* *5*, 608-619. 10.1074/mcp.M500298-MCP200.
- Frezza, C., Cipolat, S., and Scorrano, L. (2007). Organelle isolation: functional mitochondria from mouse liver, muscle and cultured fibroblasts. *Nat Protoc* *2*, 287-295. 10.1038/nprot.2006.478.
- Gao, C., Xu, Y., Liang, Z., Wang, Y., Shang, Q., Zhang, S., Wang, C., Ni, M., Wu, D., Huang, Z., and Pang, T. (2021). A novel PGAM5 inhibitor LFHP-1c protects blood-brain barrier integrity in ischemic stroke. *Acta Pharm Sin B* *11*, 1867-1884. 10.1016/j.apsb.2021.01.008.
- Genter, M.B., Clay, C.D., Dalton, T.P., Dong, H., Nebert, D.W., and Shertzer, H.G. (2006). Comparison of mouse hepatic mitochondrial versus microsomal cytochromes P450 following TCDD treatment. *Biochem Biophys Res Commun* *342*, 1375-1381. 10.1016/j.bbrc.2006.02.121.
- Grimrud, P.A., Carson, J.J., Hebert, A.S., Hubler, S.L., Niemi, N.M., Bailey, D.J., Jochem, A., Stapleton, D.S., Keller, M.P., Westphall, M.S., et al. (2012). A quantitative map of the liver mitochondrial phosphoproteome reveals posttranslational control of ketogenesis. *Cell Metab* *16*, 672-683. 10.1016/j.cmet.2012.10.004.
- Grun, B., Kosmidis, I., and Zeileis, A. (2012). Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. *J Stat Softw* *48*, 1-25.
- Guedouari, H., Savoie, M.C., Jean, S., Djeungoue-Petga, M.A., Pichaud, N., and Hebert-Chatelain, E. (2020). Multi-omics Reveal that c-Src Modulates the Mitochondrial Phosphotyrosine Proteome and Metabolism According to Nutrient Availability. *Cell Physiol Biochem* *54*, 517-537. 10.33594/000000237.
- Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
- Hebert-Chatelain, E., Desprez, T., Serrat, R., Bellocchio, L., Soria-Gomez, E., Busquets-Garcia, A., Pagano Zottola, A.C., Delamarre, A., Cannich, A., Vincent, P., et al. (2016). A cannabinoid link between mitochondria and memory. *Nature* *539*, 555-559. 10.1038/nature20127.
- Hitosugi, T., Fan, J., Chung, T.W., Lythgoe, K., Wang, X., Xie, J., Ge, Q., Gu, T.L., Polakiewicz, R.D., Roesel, J.L., et al. (2011). Tyrosine phosphorylation of mitochondrial pyruvate dehydrogenase kinase 1 is important for cancer metabolism. *Mol Cell* *44*, 864-877. 10.1016/j.molcel.2011.10.015.
- Hopper, R.K., Carroll, S., Aponte, A.M., Johnson, D.T., French, S., Shen, R.F., Witzmann, F.A., Harris, R.A., and Balaban, R.S. (2006). Mitochondrial matrix phosphoproteome: effect of extra mitochondrial calcium. *Biochemistry* *45*, 2524-2536. 10.1021/bi052475e.
- Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J., and Linding, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods* *11*, 603-604. 10.1038/nmeth.2968.
- Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* *40*, D261-270. 10.1093/nar/gkr1122.

- Huang, B., Gudi, R., Wu, P., Harris, R.A., Hamilton, J., and Popov, K.M. (1998). Isoenzymes of pyruvate dehydrogenase phosphatase. DNA-derived amino acid sequences, expression, and regulation. *J Biol Chem* 273, 17680-17688. 10.1074/jbc.273.28.17680.
- Huang, B., Wu, P., Popov, K.M., and Harris, R.A. (2003). Starvation and diabetes reduce the amount of pyruvate dehydrogenase phosphatase in rat heart and kidney. *Diabetes* 52, 1371-1376. 10.2337/diabetes.52.6.1371.
- Humphrey, S.J., Karayel, O., James, D.E., and Mann, M. (2018). High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. *Nat Protoc* 13, 1897-1916. 10.1038/s41596-018-0014-9.
- Huttemann, M., Lee, I., Samavati, L., Yu, H., and Doan, J.W. (2007). Regulation of mitochondrial oxidative phosphorylation through cell signaling. *Biochim Biophys Acta* 1773, 1701-1720. 10.1016/j.bbamcr.2007.10.001.
- Jaburek, M., Costa, A.D., Burton, J.R., Costa, C.L., and Garlid, K.D. (2006). Mitochondrial PKC epsilon and mitochondrial ATP-sensitive K⁺ channel copurify and coreconstitute to form a functioning signaling module in proteoliposomes. *Circ Res* 99, 878-883. 10.1161/01.RES.0000245106.80628.d3.
- Jadiya, P., and Tomar, D. (2020). Mitochondrial Protein Quality Control Mechanisms. *Genes (Basel)* 11. 10.3390/genes11050563.
- Jastroch, M., Divakaruni, A.S., Mookerjee, S., Treberg, J.R., and Brand, M.D. (2010). Mitochondrial proton and electron leaks. *Essays Biochem* 47, 53-67. 10.1042/bse0470053.
- Johnson, D.T., Harris, R.A., French, S., Blair, P.V., You, J., Bemis, K.G., Wang, M., and Balaban, R.S. (2007). Tissue heterogeneity of the mammalian mitochondrial proteome. *Am J Physiol Cell Physiol* 292, C689-697. 10.1152/ajpcell.00108.2006.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589. 10.1038/s41586-021-03819-2.
- Kajimura, S., and Saito, M. (2014). A new era in brown adipose tissue biology: molecular control of brown fat development and energy homeostasis. *Annu Rev Physiol* 76, 225-249. 10.1146/annurev-physiol-021113-170252.
- Kappler, L., Li, J., Haring, H.U., Weigert, C., Lehmann, R., Xu, G., and Hoene, M. (2016). Purity matters: A workflow for the valid high-resolution lipid profiling of mitochondria from cell culture samples. *Sci Rep* 6, 21107. 10.1038/srep21107.
- Khosravi, S., and Harner, M.E. (2020). The MICOS complex, a structural element of mitochondria with versatile functions. *Biol Chem* 401, 765-778. 10.1515/hsz-2020-0103.
- Klyuyeva, A., Tuganova, A., Kedishvili, N., and Popov, K.M. (2019). Tissue-specific kinase expression and activity regulate flux through the pyruvate dehydrogenase complex. *J Biol Chem* 294, 838-851. 10.1074/jbc.RA118.006433.
- Kodron, A., Mussulini, B.H., Pilecka, I., and Chacinska, A. (2021). The ubiquitin-proteasome system and its crosstalk with mitochondria as therapeutic targets in medicine. *Pharmacol Res* 163, 105248. 10.1016/j.phrs.2020.105248.
- Kolitsida, P., Zhou, J., Rackiewicz, M., Nolic, V., Dengjel, J., and Abeliovich, H. (2019). Phosphorylation of mitochondrial matrix proteins regulates their selective mitophagic degradation. *Proc Natl Acad Sci U S A* 116, 20517-20527. 10.1073/pnas.1901759116.
- Kotrasova, V., Keresztesova, B., Ondrovicova, G., Bauer, J.A., Havalova, H., Pevala, V., Kutejova, E., and Kunova, N. (2021). Mitochondrial Kinases and the Role of Mitochondrial Protein Phosphorylation in Health and Disease. *Life (Basel)* 11. 10.3390/life11020082.
- Kruse, R., and Hojlund, K. (2017). Mitochondrial phosphoproteomics of mammalian tissues. *Mitochondrion* 33, 45-57. 10.1016/j.mito.2016.08.004.

- Kuhl, I., Miranda, M., Atanassov, I., Kuznetsova, I., Hinze, Y., Mourier, A., Filipovska, A., and Larsson, N.G. (2017). Transcriptomic and proteomic landscape of mitochondrial dysfunction reveals secondary coenzyme Q deficiency in mammals. *Elife* 6. 10.7554/eLife.30952.
- Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* 11, 319-324. 10.1038/nmeth.2834.
- Kunji, E.R.S., King, M.S., Ruprecht, J.J., and Thangaratnarajah, C. (2020). The SLC25 Carrier Family: Important Transport Proteins in Mitochondrial Physiology and Pathology. *Physiology (Bethesda)* 35, 302-327. 10.1152/physiol.00009.2020.
- Kuznetsov, A.V., Hermann, M., Saks, V., Hengster, P., and Margreiter, R. (2009). The cell-type specificity of mitochondrial dynamics. *Int J Biochem Cell Biol* 41, 1928-1939. 10.1016/j.biocel.2009.03.007.
- Kwak, C., Shin, S., Park, J.S., Jung, M., Nhung, T.T.M., Kang, M.G., Lee, C., Kwon, T.H., Park, S.K., Mun, J.Y., et al. (2020). Contact-ID, a tool for profiling organelle contact sites, reveals regulatory proteins of mitochondrial-associated membrane formation. *Proc Natl Acad Sci U S A* 117, 12109-12120. 10.1073/pnas.1916584117.
- Lashkevich, K.A., and Dmitriev, S.E. (2021). mRNA Targeting, Transport and Local Translation in Eukaryotic Cells: From the Classical View to a Diversity of New Concepts. *Mol Biol*, 1-31. 10.1134/S0026893321030080.
- Lee, J., Xu, Y., Chen, Y., Sprung, R., Kim, S.C., Xie, S., and Zhao, Y. (2007). Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. *Mol Cell Proteomics* 6, 669-676. 10.1074/mcp.M600218-MCP200.
- Lewandrowski, U., Sickmann, A., Cesaro, L., Brunati, A.M., Toninello, A., and Salvi, M. (2008). Identification of new tyrosine phosphorylated proteins in rat brain mitochondria. *FEBS Lett* 582, 1104-1110. 10.1016/j.febslet.2008.02.077.
- Lewis, S.M., Williams, A., and Eisenbarth, S.C. (2019). Structure and function of the immune system in the spleen. *Sci Immunol* 4. 10.1126/sciimmunol.aau6085.
- Lewis, T.L., Jr., Kwon, S.K., Lee, A., Shaw, R., and Polleux, F. (2018). MFF-dependent mitochondrial fission regulates presynaptic release and axon branching by limiting axonal mitochondria size. *Nat Commun* 9, 5008. 10.1038/s41467-018-07416-2.
- Li, N., Qian, S., Li, B., and Zhan, X. (2019). Quantitative analysis of the human ovarian carcinoma mitochondrial phosphoproteome. *Aging (Albany NY)* 11, 6449-6468. 10.18632/aging.102199.
- Liang, M.Z., Ke, T.L., and Chen, L. (2021). Mitochondrial Protein PGAM5 Emerges as a New Regulator in Neurological Diseases. *Front Mol Neurosci* 14, 730604. 10.3389/fnmol.2021.730604.
- Liesa, M., Palacin, M., and Zorzano, A. (2009). Mitochondrial dynamics in mammalian health and disease. *Physiol Rev* 89, 799-845. 10.1152/physrev.00030.2008.
- Liu, T., Yu, R., Jin, S.B., Han, L., Lendahl, U., Zhao, J., and Nister, M. (2013). The mitochondrial elongation factors MIEF1 and MIEF2 exert partially distinct functions in mitochondrial dynamics. *Exp Cell Res* 319, 2893-2904. 10.1016/j.yexcr.2013.07.010.
- Liu, Y.J., McIntyre, R.L., Janssens, G.E., and Houtkooper, R.H. (2020). Mitochondrial fission and fusion: A dynamic role in aging and potential target for age-related disease. *Mech Ageing Dev* 186, 111212. 10.1016/j.mad.2020.111212.
- Loson, O.C., Song, Z., Chen, H., and Chan, D.C. (2013). Fis1, Mff, MiD49, and MiD51 mediate Drp1 recruitment in mitochondrial fission. *Mol Biol Cell* 24, 659-667. 10.1091/mbc.E12-10-0721.
- Lu, W., Karuppagounder, S.S., Springer, D.A., Allen, M.D., Zheng, L., Chao, B., Zhang, Y., Dawson, V.L., Dawson, T.M., and Lenardo, M. (2014). Genetic deficiency of the mitochondrial protein PGAM5 causes a Parkinson's-like movement disorder. *Nat Commun* 5, 4930. 10.1038/ncomms5930.

- Majumder, P.K., Pandey, P., Sun, X., Cheng, K., Datta, R., Saxena, S., Kharbanda, S., and Kufe, D. (2000). Mitochondrial translocation of protein kinase C delta in phorbol ester-induced cytochrome c release and apoptosis. *J Biol Chem* 275, 21793-21796. 10.1074/jbc.C000048200.
- McBride, H.M., Neuspiel, M., and Wasiak, S. (2006). Mitochondria: more than just a powerhouse. *Curr Biol* 16, R551-560. 10.1016/j.cub.2006.06.054.
- McLaughlin, K.L., Hagen, J.T., Coalson, H.S., Nelson, M.A.M., Kew, K.A., Wooten, A.R., and Fisher-Wellman, K.H. (2020). Novel approach to quantify mitochondrial content and intrinsic bioenergetic efficiency across organs. *Sci Rep* 10, 17599. 10.1038/s41598-020-74718-1.
- Menacho, C., and Prigione, A. (2020). Tackling mitochondrial diversity in brain function: from animal models to human brain organoids. *Int J Biochem Cell Biol* 123, 105760. 10.1016/j.biocel.2020.105760.
- Montes de Oca Balderas, P. (2021). Mitochondria-plasma membrane interactions and communication. *J Biol Chem* 297, 101164. 10.1016/j.jbc.2021.101164.
- Moore, A.S., and Holzbaur, E.L.F. (2018). Mitochondrial-cytoskeletal interactions: dynamic associations that facilitate network function and remodeling. *Curr Opin Physiol* 3, 94-100. 10.1016/j.cophys.2018.03.003.
- Mootha, V.K., Bunkenborg, J., Olsen, J.V., Hjerrild, M., Wisniewski, J.R., Stahl, E., Bolouri, M.S., Ray, H.N., Sihag, S., Kamal, M., et al. (2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629-640. 10.1016/s0092-8674(03)00926-7.
- Morgenstern, M., Peikert, C.D., Lubbert, P., Suppanz, I., Klemm, C., Alka, O., Steiert, C., Naumenko, N., Schendzielorz, A., Melchionda, L., et al. (2021). Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context. *Cell Metab*. 10.1016/j.cmet.2021.11.001.
- Niemi, N.M., and Pagliarini, D.J. (2021). The extensive and functionally uncharacterized mitochondrial phosphoproteome. *J Biol Chem*, 100880. 10.1016/j.jbc.2021.100880.
- Nunnari, J., and Suomalainen, A. (2012). Mitochondria: in sickness and in health. *Cell* 148, 1145-1159. 10.1016/j.cell.2012.02.035.
- Oelkrug, R., Polymeropoulos, E.T., and Jastroch, M. (2015). Brown adipose tissue: physiological function and evolutionary significance. *J Comp Physiol B* 185, 587-606. 10.1007/s00360-015-0907-7.
- Ogbi, M., and Johnson, J.A. (2006). Protein kinase Cepsilon interacts with cytochrome c oxidase subunit IV and enhances cytochrome c oxidase activity in neonatal cardiac myocyte preconditioning. *Biochem J* 393, 191-199. 10.1042/BJ20050757.
- Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., et al. (2008). A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134, 112-123. 10.1016/j.cell.2008.06.016.
- Palmieri, F., and Monne, M. (2016). Discoveries, metabolic roles and diseases of mitochondrial carriers: A review. *Biochim Biophys Acta* 1863, 2362-2378. 10.1016/j.bbamcr.2016.03.007.
- Panda, S., Srivastava, S., Li, Z., Vaeth, M., Fuhs, S.R., Hunter, T., and Skolnik, E.Y. (2016). Identification of PGAM5 as a Mammalian Protein Histidine Phosphatase that Plays a Central Role to Negatively Regulate CD4(+) T Cells. *Mol Cell* 63, 457-469. 10.1016/j.molcel.2016.06.021.
- Patel, M.S., Nemeria, N.S., Furey, W., and Jordan, F. (2014). The pyruvate dehydrogenase complexes: structure-based function and regulation. *J Biol Chem* 289, 16615-16623. 10.1074/jbc.R114.563148.
- Perrone, M., Caroccia, N., Genovese, I., Missiroli, S., Modesti, L., Pedriali, G., Vezzani, B., Vitto, V.A.M., Antenori, M., Lebedzinska-Arciszewska, M., et al. (2020). The role of mitochondria-associated membranes in cellular homeostasis and diseases. *Int Rev Cell Mol Biol* 350, 119-196. 10.1016/bs.ircmb.2019.11.002.
- Ping, P., Song, C., Zhang, J., Guo, Y., Cao, X., Li, R.C., Wu, W., Vondriska, T.M., Pass, J.M., Tang, X.L., et al. (2002). Formation of protein kinase C(epsilon)-Lck signaling modules confers cardioprotection. *J Clin Invest* 109, 499-507. 10.1172/JCI13200.

- Rangaraju, V., Lauterbach, M., and Schuman, E.M. (2019). Spatially Stable Mitochondrial Compartments Fuel Local Translation during Plasticity. *Cell* 176, 73-84 e15. 10.1016/j.cell.2018.12.013.
- Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res* 49, D1541-D1547. 10.1093/nar/gkaa1011.
- Roberts, R.F., Bayne, A.N., Goiran, T., Levesque, D., Boisvert, F.M., Trempe, J.F., and Fon, E.A. (2021). Proteomic Profiling of Mitochondrial-Derived Vesicles in Brain Reveals Enrichment of Respiratory Complex Sub-assemblies and Small TIM Chaperones. *J Proteome Res* 20, 506-517. 10.1021/acs.jproteome.0c00506.
- Robin, M.A., Anandatheerthavarada, H.K., Fang, J.K., Cudic, M., Otvos, L., and Avadhani, N.G. (2001). Mitochondrial targeted cytochrome P450 2E1 (P450 MT5) contains an intact N terminus and requires mitochondrial specific electron transfer proteins for activity. *J Biol Chem* 276, 24680-24689. 10.1074/jbc.M100363200.
- Rochette, L., Meloux, A., Zeller, M., Malka, G., Cottin, Y., and Vergely, C. (2020). Mitochondrial SLC25 Carriers: Novel Targets for Cancer Therapy. *Molecules* 25. 10.3390/molecules25102417.
- Rueda, C.B., Traba, J., Amigo, I., Llorente-Folch, I., Gonzalez-Sanchez, P., Pardo, B., Esteban, J.A., del Arco, A., and Satrústegui, J. (2015). Mitochondrial ATP-Mg/Pi carrier S_{Ca}MC-3/Slc25a23 counteracts PARP-1-dependent fall in mitochondrial ATP caused by excitotoxic insults in neurons. *J Neurosci* 35, 3566-3581. 10.1523/JNEUROSCI.2702-14.2015.
- Ruprecht, J.J., and Kunji, E.R.S. (2020). The SLC25 Mitochondrial Carrier Family: Structure and Mechanism. *Trends Biochem Sci* 45, 244-258. 10.1016/j.tibs.2019.11.001.
- Russell, O.M., Gorman, G.S., Lightowers, R.N., and Turnbull, D.M. (2020). Mitochondrial Diseases: Hope for the Future. *Cell* 181, 168-188. 10.1016/j.cell.2020.02.051.
- Sadana, P., Geyer, R., Pezoldt, J., Helmsing, S., Huehn, J., Hust, M., Dersch, P., and Scrima, A. (2018). The invasins D protein from *Yersinia pseudotuberculosis* selectively binds the Fab region of host antibodies and affects colonization of the intestine. *J Biol Chem* 293, 8672-8690. 10.1074/jbc.RA117.001068.
- Sandhu, S.K., Fassan, M., Volinia, S., Lovat, F., Balatti, V., Pekarsky, Y., and Croce, C.M. (2013). B-cell malignancies in microRNA Emu-miR-17~92 transgenic mice. *Proc Natl Acad Sci U S A* 110, 18208-18213. 10.1073/pnas.1315365110.
- Sathe, G., Deepha, S., Gayathri, N., Nagappa, M., Parayil Sankaran, B., Taly, A.B., Khanna, T., Pandey, A., and Govindaraj, P. (2021). Ethylmalonic encephalopathy ETHE1 p. D165H mutation alters the mitochondrial function in human skeletal muscle proteome. *Mitochondrion* 58, 64-71. 10.1016/j.mito.2021.02.011.
- Schober, F.A., Atanassov, I., Moore, D., Calvo-Garrido, J., Moedas, M.F., Wedell, A., Freyer, C., and Wredenber, A. (2021). Stable Isotope Labeling of Amino Acids in Flies (SILAF) Reveals Differential Phosphorylation of Mitochondrial Proteins Upon Loss of OXPHOS Subunits. *Mol Cell Proteomics* 20, 100065. 10.1016/j.mcpro.2021.100065.
- Schulenberg, B., Aggeler, R., Beechem, J.M., Capaldi, R.A., and Patton, W.F. (2003). Analysis of steady-state protein phosphorylation in mitochondria using a novel fluorescent phosphosensor dye. *J Biol Chem* 278, 27251-27255. 10.1074/jbc.C300189200.
- Shai, N., Yifrach, E., van Roermund, C.W.T., Cohen, N., Bibi, C., L, I.J., Cavellini, L., Meurisse, J., Schuster, R., Zada, L., et al. (2018). Systematic mapping of contact sites reveals tethers and a function for the peroxisome-mitochondria contact. *Nat Commun* 9, 1761. 10.1038/s41467-018-03957-8.
- Silva Ramos, E., Motori, E., Bruser, C., Kuhl, I., Yeroslaviz, A., Ruzzenente, B., Kauppila, J.H.K., Busch, J.D., Hultenby, K., Habermann, B.H., et al. (2019). Mitochondrial fusion is required for regulation of mitochondrial DNA replication. *PLoS Genet* 15, e1008085. 10.1371/journal.pgen.1008085.

- Smith, A.C., and Robinson, A.J. (2019). MitoMiner v4.0: an updated database of mitochondrial localization evidence, phenotypes and diseases. *Nucleic Acids Res* *47*, D1225-D1228. 10.1093/nar/gky1072.
- Spinelli, J.B., and Haigis, M.C. (2018). The multifaceted contributions of mitochondria to cellular metabolism. *Nat Cell Biol* *20*, 745-754. 10.1038/s41556-018-0124-1.
- Suomalainen, A., and Battersby, B.J. (2018). Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat Rev Mol Cell Biol* *19*, 77-92. 10.1038/nrm.2017.66.
- Taguchi, N., Ishihara, N., Jofuku, A., Oka, T., and Mihara, K. (2007). Mitotic phosphorylation of dynamin-related GTPase Drp1 participates in mitochondrial fission. *J Biol Chem* *282*, 11521-11529. 10.1074/jbc.M607279200.
- Toyama, E.Q., Herzig, S., Courchet, J., Lewis, T.L., Jr., Loson, O.C., Hellberg, K., Young, N.P., Chen, H., Polleux, F., Chan, D.C., and Shaw, R.J. (2016). Metabolism. AMP-activated protein kinase mediates mitochondrial fission in response to energy stress. *Science* *351*, 275-281. 10.1126/science.aab4138.
- Traba, J., Del Arco, A., Duchen, M.R., Szabadkai, G., and Satrustegui, J. (2012). SCaMC-1 promotes cancer cell survival by desensitizing mitochondrial permeability transition via ATP/ADP-mediated matrix Ca(2+) buffering. *Cell Death Differ* *19*, 650-660. 10.1038/cdd.2011.139.
- van der Bliek, A.M., Shen, Q., and Kawajiri, S. (2013). Mechanisms of mitochondrial fission and fusion. *Cold Spring Harb Perspect Biol* *5*. 10.1101/cshperspect.a011072.
- Ventura-Clapier, R., Garnier, A., Veksler, V., and Joubert, F. (2011). Bioenergetics of the failing heart. *Biochim Biophys Acta* *1813*, 1360-1372. 10.1016/j.bbamcr.2010.09.006.
- Volkert, M.R., and Crowley, D.J. (2020). Preventing Neurodegeneration by Controlling Oxidative Stress: The Role of OXR1. *Front Neurosci* *14*, 611904. 10.3389/fnins.2020.611904.
- Voytik, E., Bludau, I., Willems, S., Hansen, F.M., Brunner, A.D., Strauss, M.T., and Mann, M. (2021). AlphaMap: an open-source python package for the visual annotation of proteomics data with sequence specific knowledge. *Bioinformatics*. 10.1093/bioinformatics/btab674.
- Watanabe, T., Saotome, M., Nobuhara, M., Sakamoto, A., Urushida, T., Katoh, H., Satoh, H., Funaki, M., and Hayashi, H. (2014). Roles of mitochondrial fragmentation and reactive oxygen species in mitochondrial dysfunction and myocardial insulin resistance. *Exp Cell Res* *323*, 314-325. 10.1016/j.yexcr.2014.02.027.
- Wieckowski, M.R., Giorgi, C., Lebedzinska, M., Duszynski, J., and Pinton, P. (2009). Isolation of mitochondria-associated membranes and mitochondria from animal tissues and cells. *Nature Protocols* *4*, 1582-1590. 10.1038/nprot.2009.151.
- Williams, E.G., Wu, Y., Wolski, W., Kim, J.Y., Lan, J., Hasan, M., Halter, C., Jha, P., Ryu, D., Auwerx, J., and Aebersold, R. (2018). Quantifying and Localizing the Mitochondrial Proteome Across Five Tissues in A Mouse Population. *Mol Cell Proteomics* *17*, 1766-1777. 10.1074/mcp.RA118.000554.
- Xu, L., Wang, X., Zhou, J., Qiu, Y., Shang, W., Liu, J.P., Wang, L., and Tong, C. (2020). Miga-mediated endoplasmic reticulum-mitochondria contact sites regulate neuronal homeostasis. *Elife* *9*. 10.7554/eLife.56584.
- Zeileis, A., and Hothorn, T. (2015). Diagnostic Checking in Regression Relationships.
- Zhang, Y., Liu, X., Bai, J., Tian, X., Zhao, X., Liu, W., Duan, X., Shang, W., Fan, H.Y., and Tong, C. (2016). Mitoguardin Regulates Mitochondrial Fusion through MitoPLD and Is Required for Neuronal Homeostasis. *Mol Cell* *61*, 111-124. 10.1016/j.molcel.2015.11.017.
- Zhao, X., Bak, S., Pedersen, A.J., Jensen, O.N., and Hojlund, K. (2014). Insulin increases phosphorylation of mitochondrial proteins in human skeletal muscle in vivo. *J Proteome Res* *13*, 2359-2369. 10.1021/pr401163t.
- Zhao, X., Leon, I.R., Bak, S., Mogensen, M., Wrzesinski, K., Hojlund, K., and Jensen, O.N. (2011). Phosphoproteome analysis of functional mitochondria isolated from resting human muscle reveals

extensive phosphorylation of inner membrane protein complexes and enzymes. *Mol Cell Proteomics* **10**, M110 000299. 10.1074/mcp.M110.000299.

3.9 A New Parallel High-Pressure Packing System Enables Rapid Multiplexed Production of Capillary Columns

Johannes B. Müller-Reif¹, **Fynn M. Hansen**¹, Lisa Schweizer¹, Peter V. Treit¹, Philipp E. Geyer^{1,2}, and Matthias Mann^{1,2}

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany; and ²NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

Published in *Molecular & Cellular Proteomics* (2021)

Mass spectrometers are continuously being improved, producing machines with increasing sensitivity, resolution and data acquisition speeds. An often neglected, but crucial aspect of LC-MS setups is the high-performance liquid chromatography. Especially the robust performance of chromatographic columns is essential for the chromatographic separation of peptides. Commercial columns are often very expensive; thus, many laboratories produce in-house packed columns for LC-MS. However, packing of high-performance chromatography columns was time-intensive and required trained personnel. In this study, Dr. Müller-Reif devised a column packing station, that allows the multiplexed packing of high-performance chromatographic columns in minutes compared to hours.

In this work, I helped in the conceptualization of the experimental outline and acquired data for the evaluation of column performances.



A New Parallel High-Pressure Packing System Enables Rapid Multiplexed Production of Capillary Columns

Johannes B. Müller-Reif¹, Fynn M. Hansen¹, Lisa Schweizer¹, Peter V. Treit¹, Philipp E. Geyer^{1,2}, and Matthias Mann^{1,2,*}

Reversed-phase HPLC is the most commonly applied peptide-separation technique in MS-based proteomics. Particle-packed capillary columns are predominantly used in nanoflow HPLC systems. Despite being the broadly applied standard for many years, capillary columns are still expensive and suffer from short lifetimes, particularly in combination with ultra-high-pressure chromatography systems. For this reason, and to achieve maximum performance, many laboratories produce their own in-house packed columns. This typically requires a considerable amount of time and trained personnel. Here, we present a new packing system for capillary columns enabling rapid, multiplexed column packing with pressures reaching up to 3000 bar. Requiring only a conventional gas pressure supply and methanol as the driving fluid, our system replaces the traditional setup of helium-pressured packing bombs. By using 10× multiplexing, we have reduced the production time to just under 2 min for several 50 cm columns with 1.9- μm particle size, speeding up the process of column production 40 to 800 times. We compare capillary columns with various inner diameters and lengths packed under different pressure conditions with our newly designed, broadly accessible high-pressure packing station.

State-of-the-art MS-based proteomic pipelines typically consist of a sample preparation workflow to digest proteins and harvest pure peptides, an LC system for peptide separation, a mass spectrometer, and a sophisticated bioinformatics pipeline for raw data interpretation and subsequent statistical analysis (1, 2). The LC system plays a central role by partially separating the complex mixture of tens of thousands of peptides in a time-resolved manner according to their physicochemical properties, making them ultimately manageable for the MS system over the course of a gradient (3, 4). The most widely applied technique for high-performance applications is reversed-phase separation, originally introduced in the 1970s (5). In essence,

chromatographic systems are made of programmable pumps with the ability to form a gradient of a mixture of different agents. In the case of reversed-phase LC, the stationary phase is nonpolar, separating analytes by hydrophobicity over the course of a gradient of an increasing nonpolar mobile phase. The LC system is coupled to the mass spectrometer by electrospray (ES) ionization via an emitter (6). Glass or steel needles are commonly connected to the column. Particle-packed capillaries for chromatography can also be used for ES without being coupled to an additional emitter (7–9). These basic attributes are shared by most LC-MS systems, and differences are mainly defined by operational flow. Nanoflow LC operates at flow rates of several hundred nanoliters per minute and is the standard in proteomics because of the high sensitivity obtainable.

High flow rates in the μl to ml range, applied to columns with large inner diameters (IDs), are typically used in high-throughput or industrial-scale analysis and analytical MS application areas. Although these microflow and analytical-flow systems limit sensitivity, recent work has demonstrated robust and reproducible performance (10, 11). Reproducibility and stability of those systems are high, but drawbacks are lowered sensitivity and a need for high sample amounts. Compared with developments in sample preparation, MS instrumentation, scan modes, and software, the LC apparatus has been largely unchanged in cutting-edge MS-based proteomics. Although identifications in proteomics experiments have doubled in single-shot experiments, this can mainly be traced to improvement on the MS instrumentation and software (12–17). Current trends in LC developments aim rather toward systems for higher throughput and increasing robustness required for clinical applications (18), whereas the race for better separation in single-shot high performance runs with increasingly higher pump pressures has been comparatively abandoned. Consequently, a typically used setup for maximum sensitivity and performance for most experiments

From the ¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany; and ²NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

*For correspondence: Matthias Mann, mmann@biochem.mpg.de.

Present address for Johannes B. Müller-Reif and Philipp E. Geyer: OmicEra Diagnostics GmbH, Planegg, Germany.



Mol Cell Proteomics (2021) 20 100082 1

© 2021 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1016/j.mcpro.2021.100082>

Multiplexed High-Pressure Column Packing

still consists of columns around 75- μm ID with a length of 20 to 50 cm, packed with sub-2- μm particles. Although, better performance could be reached by longer columns or smaller particles, both conditions would result in higher operational pressures that tend to make the LC systems unstable (4, 19). For example, very high pressures can lead to leaks in the LC flow paths, resulting in poor reproducibility and subsequently a loss of measurement time.

Commercially available capillary columns in the aforementioned dimensions are expensive, especially considering how frequently they must be replaced (e.g., in our laboratories, a 50 cm column with 75- μm ID has an average turnaround time from 10 to 14 days). Therefore, many high-throughput laboratories produce packed capillaries in-house. Empty glass capillaries, ready to be packed and used, can be either purchased or produced from cheap polyimide-coated capillaries using a laser puller. Typically, a gas pressure system is deployed to pack such columns with particles in the low μm range, and instructions on the manufacturing process can be found online with open access (https://proteomicsresource.washington.edu/docs/protocols05/Packing_Capillary_Columns.pdf). However, this process is inherently slow, and interesting methods have recently been established with the aim of speeding up the packing process with high pressure (20) or dense bead slurry, as in the FlashPack method (21).

Combining these principles, we here present a high-pressure packing system for capillary columns using a high-concentration bead slurry that has previously been described as beneficial for column performance (22). These high slurry concentrations and packing pressures of 1000 to 2000 bar allow us to achieve packing times for 50 cm columns in the minute range with our system, compared with hours for traditional procedures. Deploying a manifold system and a pump capable of high flow rates further multiplexes packing to up to ten columns simultaneously and makes column production 40 to 800 times more time efficient than in previous systems. We observe consistently good column performance for packing pressures at over 1000 bar with no adverse effects on the column backpressure and lifetime, while packing times continued to decrease at higher pressures. We provide a detailed blueprint of the system so it can readily be set up in interested laboratories (supplemental Table S1).

EXPERIMENTAL PROCEDURES

Preparation of Fused Silica

Fused silica from Polymicro (TSP075365 for 75- μm ID, TSP100365 for 100- μm ID, or TSP150365 for 150- μm ID) was cut to 140 cm. Polyimide coating was removed by a Bunsen burner and the silica surface was polished with an ethanol-soaked tissue in the middle of the cut capillary at a width of 2 cm. An ES emitter tip was pulled with a laser puller (Sutter P2000) at the polished part of the capillary resulting in two empty capillary columns ready to be packed.

Sample Preparation: Protein Digestion and in-Stage Tip Purification

HeLa cells were cultured in high-glucose Dulbecco's modified Eagle's medium with 10% fetal bovine serum and 1% penicillin/streptomycin (all from Life Technologies, Inc). Cells were counted using a countess cell counter (Invitrogen), and aliquots of 1×10^6 cells were washed twice with PBS (Life Technologies, Inc), snap-frozen, and stored at -80°C . Sample preparation was carried out with the PreOmics iST kit (www.preomics.de). We used one HeLa pellet with one million cells per cartridge, determined the peptide concentration after peptide cleanup via NanoDrop, and adjusted the peptide concentration to 0.2 mg/ml.

Ultra-High-Pressure LC and MS

Samples were measured using LC-MS instrumentation consisting of an EASY-nLC 1200 ultra-high-pressure system (Thermo Fisher Scientific), coupled to an Orbitrap Exploris 480 instrument (Thermo Fisher Scientific) using a nano-ES ion source (Thermo Fisher Scientific). Purified peptides were separated on high-pressure packed columns as described in the Results and Discussion section. For each LC-MS/MS analysis with 75- μm ID columns, 500 ng peptides were used. For 100- μm ID columns, 888 ng peptides were used, and for 150- μm ID columns, 2000 ng peptides were used to adjust for the higher column volume. Peptides were loaded in buffer A* (2% acetonitrile (v/v), 0.1% trifluoroacetic acid (v/v)) and eluted with a linear 105 min gradient of 5 to 30% of buffer B (0.1% formic acid, 80% (v/v) acetonitrile), followed by a 10 min increase to 95% of buffer B and a 5 min wash of 95% buffer B. For the 75- μm ID columns, the flow rate was 300 nL/min, 535 nL/min for 100- μm ID columns, and 1200 nL/min for 150- μm ID columns to adjust for linear flow velocity. The column temperature was kept at 60°C by an in-house developed oven containing a Peltier element, and parameters were monitored in real time by the SprayQC software. MS data were acquired with a Top15 data-dependent MS/MS scan method. MS1 automatic gain control target was set to 300% in the 300 to 1650 m/z range with a maximum injection time of 25 ms and a resolution of 60,000 at m/z 200. Fragmentation of precursor ions was performed by higher-energy C-trap dissociation with a normalized collision energy of 30 eV. MS/MS scans were performed at a resolution of 15,000 at m/z 200 with an automatic gain control target of 100% and a maximum injection time of 28 ms. Dynamic exclusion was set to 30 s to avoid repeated sequencing of identical peptides.

Each column was equilibrated with two 120 min HeLa runs before the representative run for column cross-comparison.

Data Analysis

MS raw files were analyzed by MaxQuant software, version 1.6.11.0, and peptide lists were searched against the human UniProt FASTA database (release 2019_01, 188441 entries). A contaminant database generated by the Andromeda search engine was configured with cysteine carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate to 0.01 for protein and peptide levels with a minimum length of seven amino acids for peptides, and the false discovery rate was determined by searching a reverse database. Enzyme specificity was set as C-terminal to arginine and lysine as expected using trypsin and LysC as proteases. A maximum of two missed cleavages were allowed. Peptide identification was performed with an initial precursor mass deviation up to 7 ppm and a fragment mass deviation of 20 ppm. All proteins and peptides matching to the reversed database were filtered out.

Bioinformatics Analysis

Bioinformatics analyses were performed in Python (version 3.6.4.) using NumPy (1.19.2), Pandas (1.1.4), Matplotlib (3.3.2), Seaborn (0.11.0), and SciPy (1.5.2) packages.

Experimental Design and Statistical Rationale

The overall experimental design was focused on making different capillary columns for proteomics experiments as comparable as possible. To achieve this, statistical analysis was performed from triplicate experiments for the packing time and pressure performance experiments. Experimental conditions for column cross-comparisons were chosen to eliminate outer influences, including measurements on similar LC and MS systems and equilibration procedures.

RESULTS AND DISCUSSION

A High-Pressure Packing Chamber for High-Density Bead Slurries

A central challenge of nano-flow chromatography in proteomics laboratories is the constant demand for new capillary columns. Owing to their costs, commercial columns cannot be treated as a disposable item. However, in our hands, we frequently observe peak performance only for a short life span for ultra-high-performance applications. Therefore, to reach the needed quantity and cost requirements, we and many other laboratories produce own capillary columns. However, the throughput of production is limited, especially for columns with a small ID and extended length such as the 50 cm 75- μm ID columns used in most applications in our laboratories. We produce pulled or fritted capillaries and pack them with solid phase material, typically sub-2- μm C18 beads. A skilled person can pull hundreds of empty columns within a day, and fritted columns are also easy to produce. However, the packing process is inherently low-throughput and error-prone, which makes high-performance columns prized items in MS laboratories. In particular, the use of longer column lengths is—in our experience—a precondition for ultra-high-performance.

We hypothesized that high-throughput packing of capillary columns could be achieved by highly concentrated bead slurries (21) in combination with very-high-pressure packing (>1000 bar) (20). However, an increased packing pressure and bead slurry concentration can lead to column blocking, slowing down and eventually halting the packing procedure. Chloroform as a bead solvent was reported as an approach to avoid this issue because it can solvate higher bead concentrations. However, in combination with our bead particles, we observed poor chromatographic performance during proteomic experiments. Instead, we combined elevated packing pressure with the FlashPack system (21), which prohibited bead aggregation at the column entrance *via* stirring.

To test our concept, we constructed a custom-made chamber for high-pressure packing, where the pressure derives from a conventional HPLC system (EASY-LC 1000 in our case). The device consists of a central chamber, containing

the bead slurry and magnetic stirring bar, and has three openings. A large-bore access allows filling the chamber with the bead slurry, a microbore fitting holds the capillary entrance into the chamber, and a nanoviper connection is used as an inlet for the pressure from the HPLC system (supplemental Fig. S1). The slurry applied to pack columns in this system can be highly concentrated. To prepare the slurry, we mixed about 100- μl of bead particles with 500- μl of methanol. After brief vortexing and 1 minute of sonication in a sonication bath, we let the slurry settle for 5 min, whereupon we loaded 200- μl of slurry into the chamber with a 500- μl Hamilton pipette. The prototype packing chamber enabled us to fill single capillaries within minutes using the HPLC high-pressure pumps (950 bar). However, this system was not suited for high-throughput column production, and moreover, the low pump volume of the HPLC system resulted in noncontinuous packing as the pump had to be refilled several times until a column was filled with beads.

Encouraged by aspects of our newly devised packing system, we set out to further streamline column production. We replaced the small-volume HPLC pump with a Maximator HD-pump (Experimental Procedures). This high-flow continuous system converts driving gas from a standard laboratory gas supply line at a pressure ratio of 1:660 to a fluid outlet with a maximal pressure rating of 4000 bar and maximal flow capacity of 140 ml/min (Fig. 1). To use the FlashPack principle, we used methanol as the packing medium, which settles C18 beads at the chamber bottom (supplemental Fig. S2). The high flow capacity allowed us to implement multiple pump outlets for multiplex packing of up to ten columns with our station. We redesigned the original packing chamber to fit high-pressure connections (supplemental Fig. S3). For optimal stirring, we further created a rack system with magnets mounted on electric motors *via* 3D printed components to fit directly underneath the packing stations (detailed in Experimental Procedures and supplemental Fig. S4). Moreover, we connected a high-pressure range manometer to monitor packing pressure and added a pressure relief valve for efficient and controlled depressurization of the system, a notoriously time-consuming process. Although the system is typically running at 1500 bar in our laboratory, the relief of pressure takes only 60 s, without flowback from the running beads from the capillary. In addition, the system is secured from capacity exceeding driving gas pressure by a control valve, which prevents the pump to be exposed to a higher input than 6 bar. As with conventional packing systems, the weakest connection is the sealing of the capillary to the high-pressure chamber. We used a standard polyether-ether-ketone ferrule used in HPLC applications in combination with a newly designed, reinforced polyether-ether-ketone screw cap (supplemental Fig. S3D) to pin the column under very high pressure. Nevertheless, if the system pressure exceeds the durability of the material, the column is ejected. Owing to the low compression capabilities of methanol, this is dangerous if

Multiplexed High-Pressure Column Packing

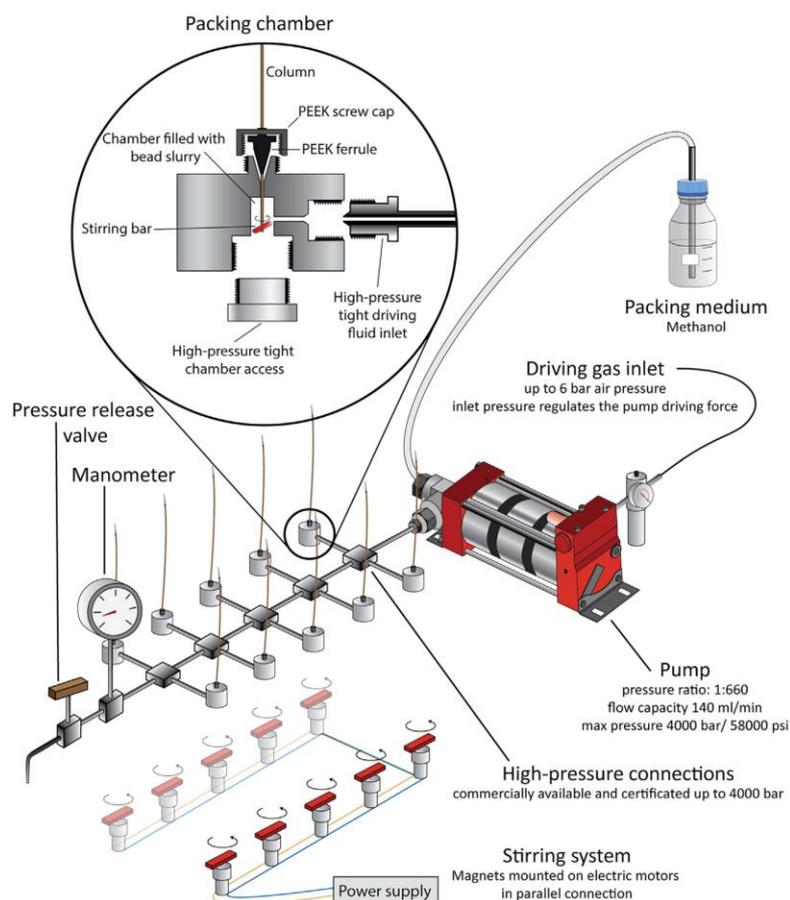


FIG. 1. **High-pressure packing station.** The scheme of the high-pressure packing station with detailed description of the crucial parts. The high-pressure pump is powered by a driving gas inlet and increases the pressure of a packing medium that is provided in a large volume flask by 660-fold. The compressed packing medium is channeled to ten packing chambers and placed on top of a magnetic stirring rack. A manometer is installed to monitor the system pressure and a pressure-release valve to facilitate time-efficient system depressurization. The inset depicts a packing chamber in detail, including high-pressure fittings, a stirring bar, and a capillary column.

one has body parts directly above the fitting when a rupture occurs and hence this must be prevented. Compared with gas, which can compress much more than liquid, no explosion risk should arise from our new packing station. To pass health and safety standards, we set up the packing system in a chemical hood with air circulation to pump off any methanol or bead particle aerosols and minimize the possibility for physical contact.

Ultra-Fast Column Packing

The time required to fill a capillary column with beads depends on two variables, the bead concentration of the packing slurry and the flow rate through the capillary. Empty capillaries with a pulled ES emitter have high flow rates in the $\mu\text{l}/\text{min}$

range even for conventional gas-based packing bombs with lower pressure (<100 bar). However, as the bead bed grows, the flow rate through the column decreases drastically. Hence, the high-density bead slurry of FlashPack enables short packing times especially for shorter columns (21). We anticipated that combining this principle with the potentially high flow rates of our extremely high-pressure system would significantly reduce packing times.

To quantify the production throughput of our system, we consecutively packed 50-cm capillaries with 75- μm ID at different pressures (1000–2500 bar) and measured the time required. With a freshly filled bead reservoir, packing at the lowest tested pressure took on average 4.7 min. Increasing pressure to 2000 bar results in packing times just over a

Multiplexed High-Pressure Column Packing

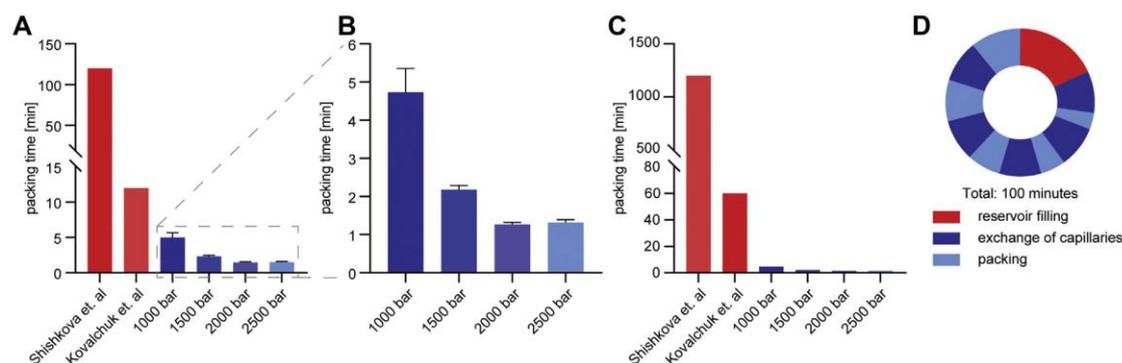


FIG. 2. **Comparison of packing times.** A, packing times of single columns as described in previous efforts and for different packing pressures (data collected in triplicates, displayed with SD) with a detailed view of the tested pressure conditions (B). C, production time for ten columns considering multiplexing (2 \times multiplexing for the system of Kovalchuk *et al.* and 10 \times for the system presented here) (20, 21). D, times of a packing cycle of 10 \times 5 columns, taking a total of 100 min with filling of the reservoir and changing of capillaries between the actual packing steps.

minute. Even higher pressure did not result in faster packing. Overall, our system decreased the time for making a single column 10- to 100-fold compared with previous packing procedures (20, 21) (Fig. 2, A and B). Of note, the total production throughput is even higher due to multiplex packing and the option to quickly exchange capillaries and bead slurries. This results in a speed-up factor of 40 to 800 (Fig. 2, C). Once filled with bead slurry and mounted on the high-pressure system, the packing chambers can be used to pack several columns consecutively. This merely requires depressurizing the system *via* the pressure relief valve and exchanging the filled columns with empty capillaries. Consecutive packing of several columns from the same reservoir will decrease the packing speed because of the removal of beads from the reservoir. To fully restore packing speed, the bead chamber has to be opened and refilled, which takes about 10 min for all ten chambers together. Typically, we refilled the reservoir after five capillary exchanges. The average turn-around cycle for producing ten columns is thus 20 min, allowing the production of hundreds of columns in a working day (Fig. 2, D). An additional advantage of the high-throughput system is that it allows us to discard improperly packed columns, which occur in approximately 10% of cases.

The high-pressure system faces the same two main challenges as usual packing stations, which are particle clogging within the capillary and bead aggregation at the column entrance. Particle clogging can only be avoided by clean working conditions. This means dust-free storage and clean cutting of fused silica and the use of filtered fluids and dust-free particles for bead slurry preparation. Bead aggregation from dense slurry can be circumvented by optimized stirring conditions according to the FlashPack principle (21).

Influence of Packing Pressure on Column Performance

To evaluate the effect of packing pressure on column performance on realistic samples, we analyzed three of our laboratory standard HeLa digests on each column. Across all packing conditions, we observed no significant variation in the number of identified peptides and protein groups (Fig. 3, A/B). Moreover, the median peak widths of identified peptides were comparable for all conditions (Fig. 3, C). Correlation between the noncorrected retention times of peptides analyzed using columns produced at varying pressures was remarkably high (Pearson correlation coefficient >0.996) and not significantly altered from replicates packed with similar pressure conditions (Fig. 3, D).

Another factor often used to characterize column performance is the tailing factor that can be calculated as depicted in Figure 3, E (23). Usually, the peak width at 5% peak height is used for peak width calculation but in proteomics experiments where tens of thousands of peaks are investigated, the base-to-base peak width is typically calculated, although full width at half maximum is also often given. We decided to calculate the peak tailing at baseline as a metric. In general, the distribution of peak shapes was wider than what would be expected from an analysis run of few analytes, but the median typically centered around the optimum of 1. The median of the peak tailing at baseline was below 1.0 for the lower and shifts above 1.0 for higher packing pressures up to a median of 1.2 (Fig. 3, F). In the literature, tailing factors in the range between 1 and 1.2 are often described (24). The shift towards this range with the higher packing pressures could result from denser compressed bead beds. As described above, the general performance was not altered for the proteomics metrics, which leads us to the conclusion that the minor change in peak tailing at baseline with higher packing pressures is not

Multiplexed High-Pressure Column Packing

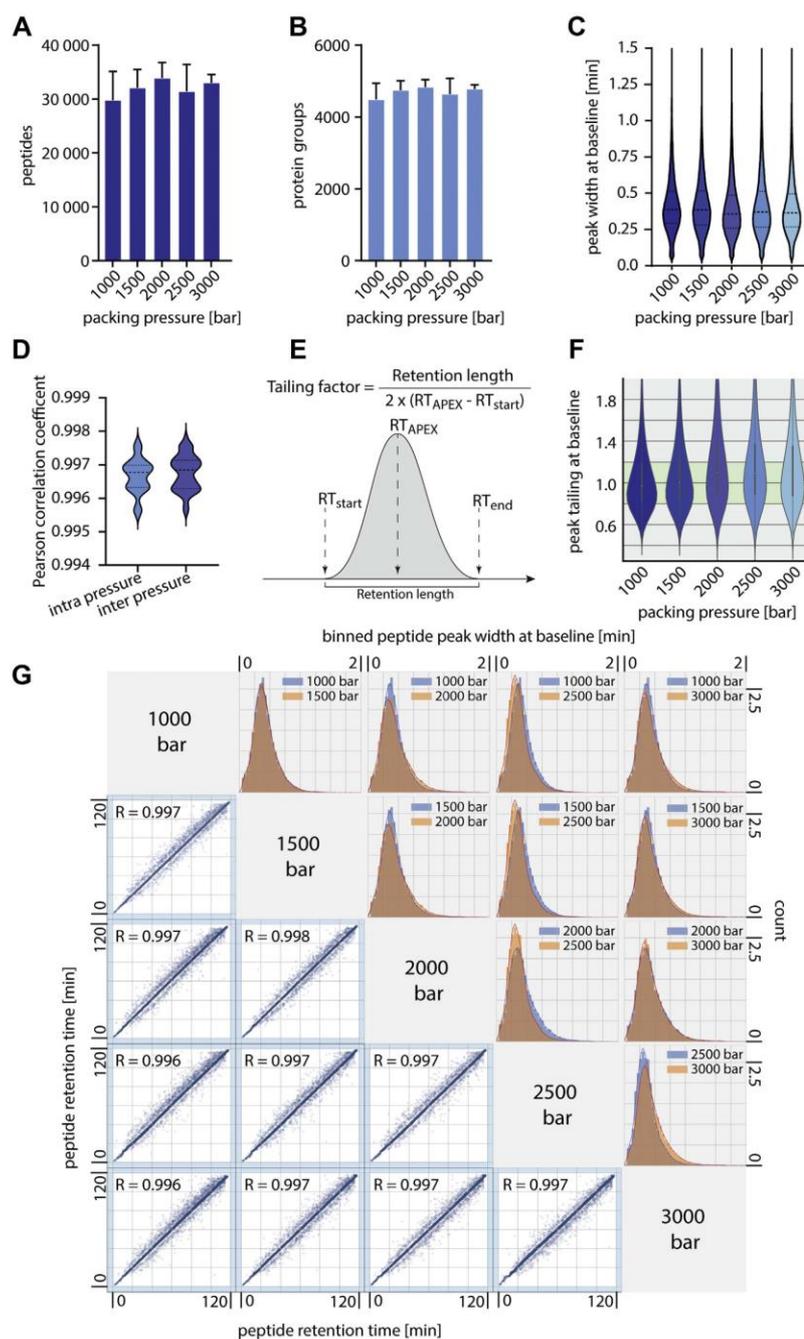


FIG. 3. **Comparison of capillary columns packed at different pressures.** A, numbers of identified peptides of triplicate measurements of 500 ng HeLa digests on columns filled at the indicated packing pressures. Peptides were separated on 50 cm and 75- μm ID columns packed with 1.9- μm Reprosil AQ Beads (Dr Maisch) with a 2 h gradient. B, numbers of identified protein groups of the same conditions as in panel A. Error bars indicate the standard deviation from triplicate measurements. C, median peak widths at baseline of identified peptides. D, distribution

changing the LC-MS performance. This manifests in an only slightly altered distribution of peak widths between representative experiments of columns packed at different pressures (Fig. 3, G). From the correlation of peptide retention times, it is visible that for all representative comparisons, the peptides elute in a narrow and reproducible time window that is not influenced by the applied packing pressure. This retention time stability is accompanied by similar separation properties of the different columns, which can be visualized directly by the peak width at baseline of analyzed molecules. Figure 3, G shows bulk analysis of all identified peptides with nearly overlapping peak width at baseline distributions, whereas the minor differences do not constitute a significant trend toward a better performance for lower or higher packing pressures of capillary columns. We did not observe a significant change in column backpressure from the different packing conditions. Based on these results, it seems that the packing pressure has no or only minimal effect on the column performance.

LC-MS Performance of Columns With Different Lengths and IDs

The length and ID of capillary columns allow their adaptation to a plethora of sample materials and LC systems, specifically regarding separation power and backpressure. In MS-based proteomics, 75- μm ID columns in combination with flow rates in the range of 200 to 400 nL per minute are typical. Hence, we packed such capillary columns with different lengths (20, 30, 50 cm) with our high-pressure system and compared their performance. Packing time for the shorter columns was even faster and in the range of 30 s. The longest columns produced the smallest peak widths and subsequently resulted in the highest numbers of identified peptides and proteins (Fig. 4, A and B). Interestingly, the distribution of peptide intensities did not change significantly, and the peak tailing at baseline also remained unaffected (Fig. 4, C and D).

Over the last years, the demand for high-throughput analysis has become apparent for the analysis of clinical samples, especially blood plasma as we have described before (25). This has been addressed by a novel HPLC principle with preformed gradients and slightly higher flow rates (18) and by higher-flow systems operating in the high microliter per minute range (10, 26). As these strategies require columns with a higher ID to maintain acceptable pressure during analysis, we produced columns with 75- μm , 100- μm , and 150- μm ID and tested their performance.

When comparing column IDs, the experimental setup has to be adapted to the conditions. To enable direct comparison of

capillaries between different IDs, we scaled the flow rates to reach the same linear velocities and the amount of input material to the column volume (Experimental Procedures). For the 100- μm ID columns, this results in a flow rate of 535 nL/min and 888 ng of peptides for loading, whereas for the 150- μm ID column, 1200 nL/min and 2 μg of peptide material was loaded to be comparable to the 300 nL/min and 500 ng used for the 75- μm ID columns. This requirement of higher sample amount already limits the applicability of larger column diameters for samples with limited accessibility. The 1400 μL of pump volume from the Easy-LC 1200 used for the experiment was sufficient to run a 2 h gradient with the 150- μm ID column, but longer gradients or higher flow rates would exceed the capabilities of the LC system and require lower flow rates. The higher column IDs led to slightly broader peak widths, but peptide and protein identifications were not affected. Owing to the correction of the sample input amount, we did not see a difference in the peptide intensity distributions, and the peak tailing at baseline was also not affected by the column ID (Fig. 4, E-H).

CONCLUSION

Here, we aimed to increase the throughput and to streamline the production of capillary columns for MS-based proteomics. We provide a detailed list for the commercial parts and blueprints describing the construction of our high-pressure packing station. The setup can be built at relatively low costs (<\$10,000), compared with the cumulative expenses for high-performing commercial columns. We designed this new station to fill multiple columns simultaneously within a few minutes, which accelerates the packing process of capillary columns more than a 100-fold compared with traditional gas pressure-driven stations. In this way, we hope our system helps researchers by streamlining the often work-intensive and fragile column production process. In addition, the extreme high pressures enable the packing of long, high-performing columns (>50 cm). The ability to produce high-performing columns at high-throughput allows for the possibility of only using capillary columns at the peak of their performance, replacing them as soon as peak broadening or decreased ionization is observed. Reassuring in terms of robustness of the packing process itself and the stability achieved at exceedingly high pressures, we have not observed variation in the performance characteristics over a wide range of packing pressure from 1000 to 3000 bar. We hope the technology described here will enable laboratories of any

of Pearson correlation coefficients calculated on peptide retention times between columns packed at the same pressure and columns packed at different pressures (p -value of unpaired t -test for difference: 0.6). E, visualization of the tailing factor calculation, this is typically done at 5% peak height. F, Peak tailing at baseline for all identified peptides from runs with 75- μm ID columns and different packing pressures. G, correlation of peptide retention times across packing conditions. The density of peptides is color-coded. The histograms show the peak widths at baseline distribution of five representative runs. ID, inner diameter.

Multiplexed High-Pressure Column Packing

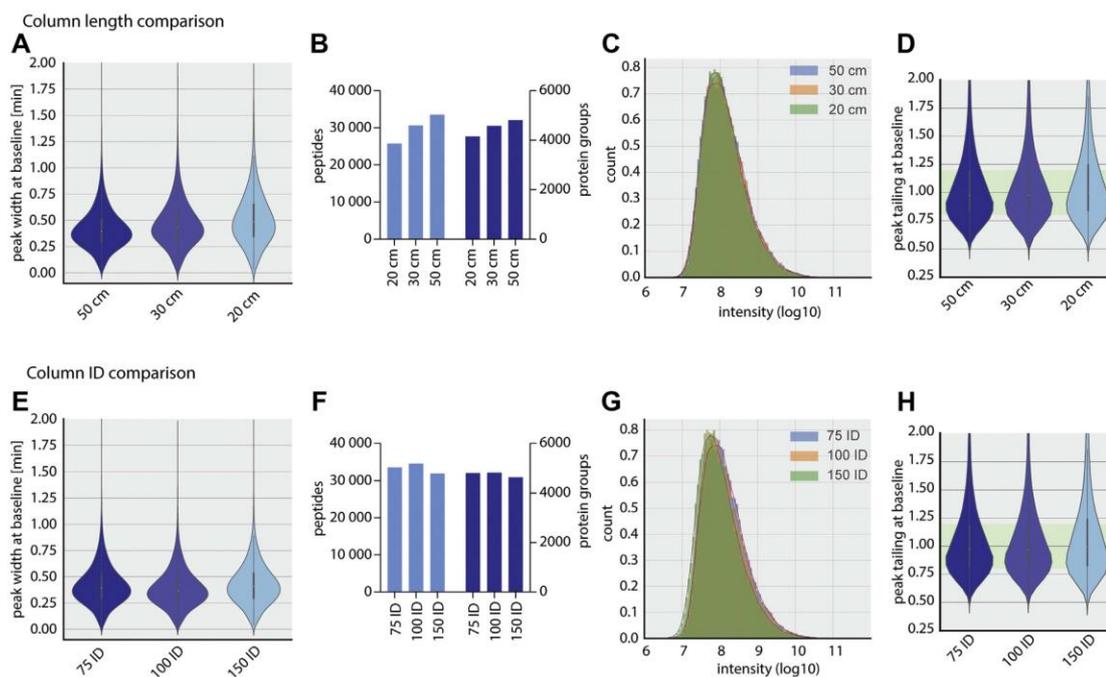


FIG. 4. **Length and inner diameter comparison.** All columns were packed with 1000 bar packing pressure. A, peak width distribution from HeLa runs with different column length with the respective number of peptide and protein identifications (B), peptide intensity distribution (log10) (C), and peak tailing at baseline distribution (D). E, peak width distribution from HeLa runs with different column IDs with the respective number of peptide and protein identifications (F), peptide intensity distribution (log10) (G), and peak tailing at baseline distribution (H). ID, inner diameter.

size to mass-produce high-performance long capillary columns.

DATA AVAILABILITY

The MS-based proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository and are available via ProteomeXchange with identifier PXD024296.

Supplemental data—This article contains supplemental data (21).

Acknowledgments—We thank all members of the Proteomics and Signal Transduction Group and the Clinical Proteomics Group for help and discussions and in particular Igor Paron and Christian Deiml for technical assistance and Mario Orosi for help with the online resource. We thank the mechanical workshop and the educational workshop of the Max Planck Institute of Biochemistry especially Martin Wied, Andreas Kucher and Harry Spangenberg for the fabrication and iterative optimization of all self-constructed parts.

The work carried out in this project was partially supported by the Max Planck Society for the Advancement of Science.

Author contributions—J. B. M.-R. designed and assembled the packing station parts and carried out the bioinformatics analyses. J. B. M.-R., L. S., F. M. H., P. E. G., and P. V. T. designed the experiments, performed and interpreted the MS-based proteomics analyses, generated text and figures, and wrote the manuscript. M. M. supervised and guided the project, designed the experiments, and interpreted MS-based proteomics data.

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: ES, electrospray; ID, inner diameter.

Received February 26, 2021, and in revised form, April 1, 2021
Published, MCPRO Papers in Press, April 20, 2021, <https://doi.org/10.1016/j.mcpro.2021.100082>

REFERENCES

- Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
- Aebersold, R., and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355
- Michalski, A., Cox, J., and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority

- is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10**, 1785–1793
- Shishkova, E., Hebert, A. S., and Coon, J. J. (2016) Now, more than ever, proteomics needs better chromatography. *Cell Syst.* **3**, 321–324
 - Horváth, C., Melander, W., and Molnár, I. (1976) Solvophobic interactions in liquid chromatography with nonpolar stationary phases. *J. Chromatogr. A.* **125**, 129–156
 - Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71
 - Kennedy, R. T., and Jorgenson, J. W. (1989) Preparation and evaluation of packed capillary liquid chromatography columns with inner diameters from 20 to 50 micrometers. *Anal. Chem.* **61**, 1128–1135
 - Emmett, M. R., and Caprioli, R. M. (1994) Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* **5**, 605–613
 - Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Micro-columns with self-assembled particle frits for proteomics. *J. Chromatogr. A.* **979**, 233–239
 - Bian, Y., Zheng, R., Bayer, F. P., Wong, C., Chang, Y. C., Meng, C., Zolg, D. P., Reinecke, M., Zecha, J., Wiechmann, S., Heinzlmeir, S., Scherr, J., Hemmer, B., Baynham, M., Gingras, A. C., et al. (2020) Robust, reproducible and quantitative analysis of thousands of proteomes by microflow LC-MS/MS. *Nat. Commun.* **11**, 157
 - Bian, Y., Bayer, F. P., Chang, Y. C., Meng, C., Hoefler, S., Deng, N., Zheng, R., Boychenko, O., and Kuster, B. (2021) Robust microflow LC-MS/MS for proteome analysis: 38000 runs and counting. *Anal. Chem.* **93**, 3686–3690
 - Bernhardt, O., Selevsek, N., Gillet, L., Rinner, O., Picotti, P., Aebersold, R., and Reiter, L. (2014) Spectronaut: a fast and efficient algorithm for MRM-like processing of data independent acquisition (SWATH-MS) data. *F1000Res.* **5**, 1092
 - Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
 - Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005) The Orbitrap: A new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443
 - Kelstrup, C. D., Bekker-Jensen, D. B., Arrey, T. N., Högberg, A., Harder, A., and Olsen, J. V. (2018) Performance evaluation of the Q Exactive HF-X for shotgun proteomics. *J. Proteome Res.* **17**, 727–738
 - Meier, F., Beck, S., Grassl, N., Lubeck, M., Park, M. A., Raether, O., and Mann, M. (2015) Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387
 - Meier, F., Brunner, A. D., Frank, M., Ha, A., Bludau, I., Voytik, E., Kaspar-Schoenefeld, S., Lubeck, M., Raether, O., Bache, N., Aebersold, R., Collins, B. C., Röst, H. L., and Mann, M. (2020) diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236
 - Bache, N., Geyer, P. E., Bekker-Jensen, D. B., Hoerning, O., Falkenby, L., Treit, P. V., Doll, S., Paron, I., Müller, J. B., Meier, F., Olsen, J. V., Vorm, O., and Mann, M. (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol. Cell Proteomics* **17**, 2284–2296
 - Richards, A. L., Merrill, A. E., and Coon, J. J. (2015) Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* **24**, 11–17
 - Shishkova, E., Hebert, A. S., Westphall, M. S., and Coon, J. J. (2018) Ultra-high pressure (>30,000 psi) packing of capillary columns enhancing depth of shotgun proteomic analyses. *Anal. Chem.* **90**, 11503–11508
 - Kovalchuk, S. I., Jensen, O. N., and Rogowska-Wrzesinska, A. (2019) FlashPack: Fast and simple preparation of ultrahigh-performance capillary columns for LC-MS. *Mol. Cell Proteomics* **18**, 383–390
 - Bruns, S., Franklin, E. G., Grinias, J. P., Godinho, J. M., Jorgenson, J. W., and Tallarek, U. (2013) Slurry concentration effects on the bed morphology and separation efficiency of capillaries packed with sub-2 μm particles. *J. Chromatogr. A.* **1318**, 189–197
 - Anderson, J. L., Berthod, A., Pino, V., and Stalcup, A. (2015) *Analytical Separation Science*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany
 - Birdsall, R. E., Kellett, J., Yu, Y. Q., and Chen, W. (2019) Application of mobile phase additives to reduce metal-ion mediated adsorption of non-phosphorylated peptides in RPLC/MS-based assays. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **1126–1127**, 121773
 - Geyer, P. E., Holdt, L. M., Teupser, D., and Mann, M. (2017) Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13**, 942
 - [preprint] Messner, C., Demichev, V., Bloomfield, N., Ivosev, G., Wasim, F., Zelezniak, A., Lilley, K., Tate, S., and Ralser, M. (2019) ScanningSWATH enables ultra-fast proteomics using high-flow chromatography and minute-scale gradients. *bioRxiv*. <https://doi.org/10.1101/656793>

3.10 AlphaMap: an open-source Python package for the visual annotation of proteomics data with sequence-specific knowledge

Eugenia Voytik^{1,†}, Isabell Bludau^{1,†}, Sander Willems¹, **Fynn M. Hansen**¹, Andreas-David Brunner¹, Maximilian T. Strauss¹ and Matthias Mann^{1,2}

¹ Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany and ² Department of Clinical Proteomics, NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
† The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Published in *Bioinformatics* (2021)

MS-based search engines allow the routine identification of different PTMs across a wide variety of data acquisition schemes leading to vast sets of identified modification sites. The manual curation of such sites is time consuming and in practice often limited to only a subset of identified sites. In addition, the visualization of modification sites on proteins is not easy, making the interpretation of the spatial distribution of modification sites on proteins cumbersome. In this work, Eugenia Voytik and Dr. Isabel Bludau of the Mann group set out to devise a software tool that allows the generic annotation of identified modification sites and their localization on the corresponding protein sequence. This tool bridges a long-standing gap between large scale data identification and the previously cumbersome manual data interpretation and provides a great access point to explore PTM-related experiments.

Based on my expertise in the experimental work with PTMs, I helped with the conceptualization of the AlphaMap tool providing valuable feedback from the user point of view. Furthermore, I was involved in the rigorous testing of the tool.

Sequence analysis

AlphaMap: an open-source Python package for the visual annotation of proteomics data with sequence-specific knowledge

Eugenia Voytik^{1,†}, Isabell Bludau^{1,†}, Sander Willems¹, Fynn M. Hansen¹,
Andreas-David Brunner¹, Maximilian T. Strauss¹ and Matthias Mann^{1,2,*}

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany and

²Department of Clinical Proteomics, NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Olga Vitek

Received on July 30, 2021; revised on September 2, 2021; editorial decision on September 13, 2021; accepted on September 22, 2021

Abstract

Summary: Integrating experimental information across proteomic datasets with the wealth of publicly available sequence annotations is a crucial part in many proteomic studies that currently lacks an automated analysis platform. Here, we present AlphaMap, a Python package that facilitates the visual exploration of peptide-level proteomics data. Identified peptides and post-translational modifications in proteomic datasets are mapped to their corresponding protein sequence and visualized together with prior knowledge from UniProt and with expected proteolytic cleavage sites. The functionality of AlphaMap can be accessed via an intuitive graphical user interface or—more flexibly—as a Python package that allows its integration into common analysis workflows for data visualization. AlphaMap produces publication-quality illustrations and can easily be customized to address a given research question.

Availability and implementation: AlphaMap is implemented in Python and released under an Apache license. The source code and one-click installers are freely available at <https://github.com/MannLabs/alphamap>.

Contact: mmann@biochem.mpg.de

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Bottom-up mass spectrometry (MS) has become the leading technology for identifying and quantifying proteomes (Aebersold and Mann, 2003, 2016; Müller *et al.*, 2020). Since peptides rather than intact proteins are measured, visualizing identified peptides and post-translational modifications (PTMs) together with known protein sequence information is an important aspect of downstream MS data exploration. However, the ability to easily integrate and visualize experimental data together with already known sequence annotations is an unmet need in the proteomics community. Although established visualization platforms provide manual visualization of a single experimental sample or dataset at a time (Omasits *et al.*, 2014), there is a lack of tools that support state-of-the-art data analysis software frameworks and that can visualize experimental sequence coverage across multiple samples or datasets in combination with available sequence annotations mined from UniProt, the standard knowledgebase for protein information (Bateman, 2019). To make this wealth of information easily accessible to proteomics

researchers, we developed AlphaMap, a Python package that facilitates the visual exploration of peptide-level proteomics data.

2 The AlphaMap computational framework

In line with other recently developed software tools from our lab (Strauss *et al.*, 2021; Willems *et al.*, 2021), we implemented AlphaMap in pure Python because of its clear, easy to understand syntax and the availability of excellent supporting scientific libraries. To read fasta files, we leverage the Pyteomics Python package (Goloborodko *et al.*, 2013; Levitsky *et al.*, 2019). Plotly is a well-established plotting library that we use for generating AlphaMap's sequence visualization (Plotly Technologies Inc., 2015), allowing flexible customization and great user interactivity. To enable easy access to the AlphaMap functionality with a low barrier of entry, a stand-alone graphical user interface (GUI) was implemented using the Panel library (Rudiger *et al.*, 2021). AlphaMap can be launched either as a browser-based GUI after simple local installation or as a

standard Python module installed via PyPI (Python Software Foundation, n.d.) or directly from its GitHub repository.

In line with the AlphaPept ecosystem (Strauss *et al.*, 2021), we make the AlphaMap code openly available on GitHub, using its many supporting features for unit and system testing via GitHub actions. For code development, we adopted the concept of ‘literate programming’ (Knuth, 1984), which combines the algorithmic code with readable documentation and testing. Using the nbdev package, the codebase can directly be inspected in well documented Jupyter Notebooks, from which the code is automatically extracted (Kluyver *et al.*, 2016). We envision that these design principles will encourage

the broader community to integrate AlphaMap in their own data analysis and visualization workflows with the possibility to easily adopt the code according to specific needs.

3 Overview of the AlphaMap workflow

AlphaMap uses peptide-level proteomics data as input. It currently supports the direct import of data processed by MaxQuant (Cox and Mann, 2008), Spectronaut (Bruderer *et al.*, 2015), DIA-NN (Demichev *et al.*, 2020), FragPipe (Kong *et al.*, 2017) and our

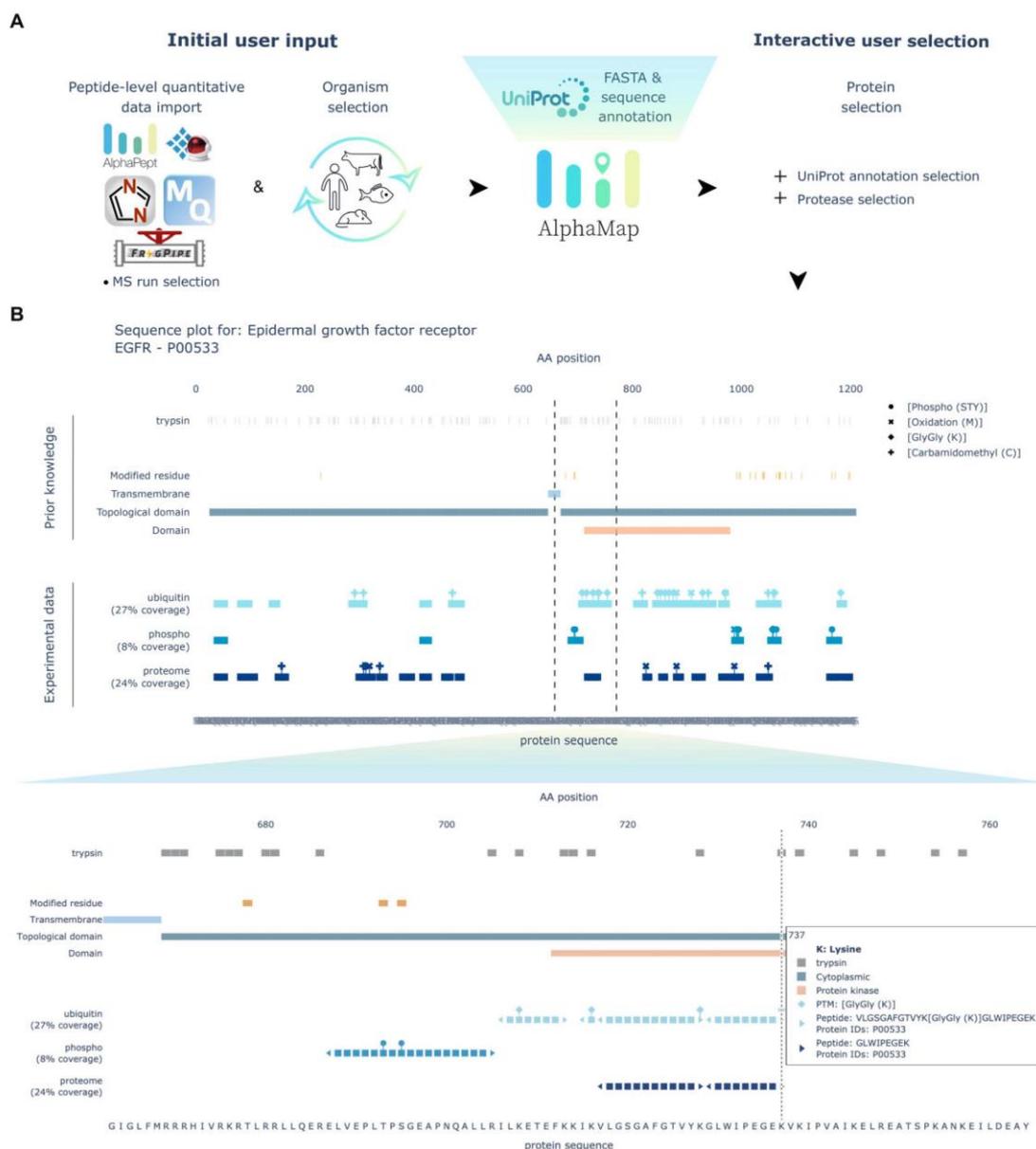


Fig. 1. (A) Overview of the AlphaMap workflow from MS data upload to the interactive sequence visualization. (B) Exemplary sequence visualization for epidermal growth factor receptor (EGFR). A zoom-in on a selected sequence region, indicated by dashed lines, is provided at the lower part of the panel

recently introduced AlphaPept framework (Strauss *et al.*, 2021). In contrast to Protter (Omasits *et al.*, 2014), users can select multiple independent datasets for co-visualization. These could either have been processed by the same or with different MS analysis tools. It is also possible to select only a single sample, or a subset of samples of a given input file for individual sequence visualization. In addition to the peptide-level data generated from LC-MS analysis, AlphaMap leverages a plethora of manually curated sequence-specific protein level information available from UniProt. Fasta files and UniProt sequence annotations are readily accessible in AlphaMap for the 13 most popular UniProt organisms as well as for SARS-CoV and SARS-CoV-2. Functionality to enable the integration of additional organisms is further available as part of our Python package. Finally, the user can select the different layers of information that should be displayed in the interactive sequence representation, including selected protease cleavage sites and UniProt sequence annotations. Figure 1A shows a schematic overview of the AlphaMap workflow. Detailed instructions for its installation and usage are further provided in the [supplementary user guide](#). In addition to interactive sequence visualization of a user-selected protein, AlphaMap provides individual links to external databases and tools for further sequence evaluation in UniProt (Bateman, 2019), PhosphoSitePlus (Hornbeck *et al.*, 2015), Protter (Omasits *et al.*, 2014), PDB (Berman *et al.*, 2000) and Peptide Atlas (Desiere *et al.*, 2006).

4 Application of AlphaMap to investigate full proteome and PTM data

Figure 1B shows the sequence visualization of the peptides and PTMs identified for the epidermal growth factor receptor (EGFR) in human A549-ACE2 cells that were infected with SARS-CoV-2 or SARS-CoV (an exemplary viral protein detected in this dataset is visualized in the [Supplementary Material](#)) (Stukalov *et al.*, 2021). We show three independent experimental traces: one for full proteome data, one for phospho-enriched peptides and one for ubiquitin-enriched peptides. The proteome data indicates a homogeneous coverage across the entire protein sequence. As expected, phosphorylation and ubiquitination are limited to the C-terminal region of the protein, which is annotated to be exposed to the cytosol. In addition, the kinase domain of EGFR is highly ubiquitinated in our dataset, whereas the surrounding cytosolic regions are phosphorylated. Interestingly, AlphaMap reports that most of our observed phosphorylation sites have been previously identified, whereas none of the identified ubiquitination sites are annotated in UniProt. Please note that unmodified peptides are also observed in both the phospho- and ubiquitin-enriched samples due to the imperfect selectivity of enrichment protocols.

Beyond the uses highlighted here, we envision AlphaMap to facilitate data analysis and interpretation for a variety of different applications:

- Candidate validation: AlphaMap can be used to assess the sequence coverage of identified biomarker candidates (or other proteins of interest) to evaluate possible sequence variations or unexpected anomalies on the basis of readily available sequence information.
- Preparation of panels for publication: Sequence visualizations from AlphaMap can directly highlight the precise MS derived information about proteins of interest in biological or clinical projects.
- Technical comparisons: AlphaMap can be used to evaluate sequence coverage between different data acquisition strategies such as data-dependent and data-independent acquisition, alternative instrument platforms or software tools.
- Optimization of sample processing: Visualization of protein cleavage sites for different proteases can help to optimize sample

processing with the goal to achieve a more complete sequence coverage.

5 Conclusion

AlphaMap offers an interactive GUI and a Python package for visualizing peptide-level bottom-up proteomics data on the basis of individual protein sequences, including information of curated UniProt sequence annotations and expected proteolytic cleavage sites. We expect that future developments by us and the community will extend the variety of available annotations in AlphaMap, for example by including prior knowledge of sequence conservation or predicted functional domains. In addition, we will integrate quantitative information and differential analysis results into the AlphaMap sequence representations. We envision that AlphaMap will assist MS-based proteomics researchers in inspecting peptide- and PTM-level data, thereby providing valuable information in the process of candidate validation in biological and clinical context.

Author contributions

I.B. conceptualized the project and together with E.V. and M.M. wrote the manuscript with contributions from all authors. I.B. and E.V. implemented the core AlphaMap functions. E.V. implemented the GUI. S.W. provided important help with the AlphaMap installers. F.M.H. and A.-D.B. provided valuable ideas for the concept and visualization in AlphaMap and F.M.H. further contributed by rigorous testing. M.T.S. designed the general AlphaPept ecosystem and assisted with the nbdev environment. M.M. supervised the study and provided critical feedback on all aspects of the presented work.

Funding

This study was supported by The Max-Planck Society for Advancement of Science and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de, G64b-A1070-2018/131-2 DMB-1805-0008). I.B. acknowledges funding support from her Postdoc.Mobility fellowship granted by the Swiss National Science Foundation [P400PB_191046].

Conflict of Interest: none declared.

Acknowledgements

The authors thank Julia Schessner, Barbara Steigenberger, Jakob Bader and Sophia Mädler for testing and providing critical feedback on AlphaMap. They are grateful to Özge Karayel and Maria C. Tanzer for valuable discussions and for providing experimental data.

References

- Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.
- Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature*, **537**, 347–355.
- Bateman, A.; UniProt Consortium. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Berman, H.M. *et al.* (2000) The protein data bank. In *Nucleic Acids Res.*, **28**, 235–242.
- Bruderer, R. *et al.* (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics*, **14**, 1400–1410.
- Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
- Demichev, V. *et al.* (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods*, **17**, 41–44.

- Desiere, F. et al. (2006) The PeptideAtlas project. *Nucleic Acids Res.*, **34**, D655–D658.
- Goloborodko, A.A. et al. (2013) Pyteomics – a python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrometry*, **24**, 301–304.
- Hornbeck, P.V. et al. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
- Kluyver, T. et al. (2016) Jupyter Notebooks—a publishing format for reproducible computational workflows. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas – Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, Göttingen, Germany, pp. 87–90.
- Knuth, D.E. (1984) Literate programming. *Comput. J.*, **27**, 97–111.
- Kong, A.T. et al. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods*, **14**, 513–520.
- Levitsky, L.I. et al. (2019) Pyteomics 4.0: five years of development of a Python proteomics framework. *J. Proteome Res.*, **18**, 709–714.
- Müller, J.B. et al. (2020) The proteome landscape of the kingdoms of life. *Nature*, **582**, 592–596.
- Omasits, U. et al. (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, **30**, 884–886.
- Plotly Technologies Inc. (2015) *plotly*. Montréal, QC. <https://plot.ly> (27 September 2021, date last accessed).
- Python Software Foundation. (n.d.) Python Package Index – PyPI. <https://pypi.org/> (27 September 2021, date last accessed).
- Rudiger, P. et al. (2021) *holoviz/panel: Version 0.11.3*. doi: 10.5281/ZENODO.4692827.
- Strauss, M.T. et al. (2021) AlphaPept, a modern and open framework for MS-based proteomics. *BioRxiv*, 2021.07.23.453379. doi:10.1101/2021.07.23.453379.
- Stukalov, A. et al. (2021) Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature*, **594**, 246–252.
- Willems, S. et al. (2021) AlphaTims: indexing trapped ion mobility spectrometry – time of flight data for fast and easy accession and visualization. *Mol. Cell. Proteomics*, 100149.

3.11 The structural context of PTMs at a proteome wide scale

Isabell Bludau¹, Sander Willems¹, Wen-Feng Zeng¹, Maximilian T. Strauss², **Fynn M. Hansen**¹, Maria C. Tanzer¹, Özge Karayel¹, Brenda A. Schulman³, Matthias Mann^{1,2}

¹ Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany ² Department of Clinical Proteomics, NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark ³ Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

Available on *BioRxiv* and submitted for publication

Over the last decades, MS-based proteomics has unveiled a tremendous number of PTMs, many of them with unknown function. Although, knowledgebases such as PhosphoSitePlus [197] condense and annotate the vast set of reported PTMs, the structural context of PTMs has not been addressed on a global level. AlphaFold has made an enormous impact in the field of protein structure prediction [198] and now allows the structural mapping of PTMs. In this study, Dr. Isabell Bludau in our group uncovered global patterns of PTMs across structured and intrinsically disordered regions, highlighting sites of spatial co-regulation and sites of potential PTM crosstalk. With the extension of the AlphaMap tool (see 3.10), users can readily visualize PTMs of their own experiments on predicted 3D protein structures.

Based on my expertise in the field of PTMs, I helped with hypothesis generation and result interpretation. Furthermore, I helped with the incorporation of key datasets of my previous projects (Publication 1 and Publication 2).

The structural context of PTMs at a proteome wide scale

Isabell Bludau¹, Sander Willems¹, Wen-Feng Zeng¹, Maximilian T. Strauss², Fynn M. Hansen¹, Maria C. Tanzer¹, Ozge Karayel¹, Brenda A. Schulman³, Matthias Mann^{1,2}

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

²Department of Clinical Proteomics, NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

³Department of Molecular Machines and Signaling, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

Abstract

The recent revolution in computational protein structure prediction provides folding models for entire proteomes, which can now be integrated with large-scale experimental data. Mass spectrometry (MS)-based proteomics has identified and quantified tens of thousands of post-translational modifications (PTMs), most of them of uncertain functional relevance. In this study, we determine the structural context of these PTMs and investigate how this information can be leveraged to pinpoint potential regulatory sites. Our analysis uncovers global patterns of PTM occurrence across folded and intrinsically disordered regions. We found that this information can help to distinguish regulatory PTMs from those marking improperly folded proteins. Interestingly, the human proteome contains thousands of proteins that have large folded domains linked by short, unstructured regions that are strongly enriched in regulatory phosphosites. These include well-known kinase activation loops that induce protein conformational changes upon phosphorylation. This regulatory mechanism appears to be widespread in kinases but also occurs in other protein families such as solute carriers. It is not limited to phosphorylation but includes ubiquitination and acetylation sites as well. Furthermore, we performed three-dimensional proximity analysis which revealed examples of spatial co-regulation of different PTM types and potential PTM crosstalk. To enable the community to build upon these first analyses, we provide tools for 3D visualization of proteomics data and PTMs as well as python libraries for data accession and processing.

Introduction

Post translational modifications (PTMs) are an important mechanism to regulate the activity and function of proteins. Mass spectrometry (MS)-based proteomics has become the method of choice to not only identify and quantify proteomes (Aebersold and Mann, 2016, 2003), but also to investigate PTMs on a proteome-wide scale (Bekker-Jensen et al., 2020; Hansen et al., 2021; Sharma et al., 2014). Despite impressive technological progress, a key challenge in the PTM and signaling fields remains to distinguish PTMs that are of direct functional relevance from the tens of thousands that are routinely measured. This is necessary to match the proteomics data to dedicated, low throughput biochemical follow-up studies that characterize the biological functions of candidate PTMs.

To assess functional relevance of PTMs on a more global scale, Beltrao and co-workers recently presented a machine learning model that uses information from different features indicative of proteomic, structural, regulatory or evolutionary relevance to derive a functional score for a large catalog of phosphosites (Ochoa et al., 2019). Another recent study directly evaluated the functional relevance of phosphorylations purely based on available structural information (Kamacioglu et al., 2021). Based on these and many previous studies, we know that phosphorylations are predominantly observed on spatially accessible amino acids and particularly in intrinsically disordered regions (IDRs) (Iakoucheva et al., 2004; Kamacioglu et al., 2021; Tyanova et al., 2013). Furthermore, it stands to reason that phosphorylations in flexible regions within folded domains and on binding interfaces are more likely to be functional compared to those that are buried or less accessible in rigidly folded regions (Kamacioglu et al., 2021). Although these studies impressively highlight the value of integrating structural information into the analysis of PTMs, they have been limited to phosphorylation and to the set of available experimentally derived structures deposited in PDB, which furthermore inherently favor stable regions of proteins (Burley et al., 2021).

Recently, there has been a key breakthrough in computational protein structure prediction from just the amino acid sequence of a protein. The novel deep learning models in AlphaFold2 (henceforth referred to as AlphaFold) (Jumper et al., 2021), rapidly followed by RoseTTAFold (Baek et al., 2021), were shown to regularly achieve high accuracy in predicting protein structures that are largely comparable to those determined by experimental methods. By providing structural information for almost the complete human proteome as well as the proteomes of over 20 model organisms, the AlphaFold protein structure database (AlphaFold DB, <https://alphafold.ebi.ac.uk>) now enables structural investigations on a proteome-wide scale, thus promising to accelerate our understanding

of the structure-to-function relationship of proteins (Tunyasuvunakool et al., 2021; Varadi et al., 2021).

Here we set out to combine this wealth of structural information with proteomics data, especially large-scale PTM information, with the goal of shedding new light on the long-standing question of functional relevance of PTMs. We present a first systematic assessment of how PTM data can be integrated with deep-learning predicted structures on a proteome-wide scale. We then explore key features in the structure function domain by combining predictions of functional relevance with domain features and discover a multitude of sites with potential regulatory roles. To enable the community to further explore the numerous related biological questions, we provide a Python package called StructureMap, that allows to easily and quickly access and integrate structural data from AlphaFold DB with proteomics data and information on PTMs. Finally, we provide an extended version of our previously published AlphaMap tool for sequence visualization (Voytik et al., 2021), which now enables the mapping of peptides and PTMs to three-dimensional protein structures.

Results

Estimation of side chain exposure and intrinsically disordered regions from predicted protein structures

To make the information provided by predicted structures accessible for systematic analyses, we first wanted to extract it into tractable metrics such as amino acid side chain exposure or the categorization of amino acids into structured and intrinsically disordered regions (IDRs). At this point, it is important to make a clear distinction between predicted and experimentally derived structures. Experimental structures are often incomplete and may only cover a specific sequence region. In contrast, the predicted structures in AlphaFold DB in principle cover the entire protein sequence from N- to C-terminus. Importantly, each amino acid in the predicted structure is associated with a specific prediction confidence (pLDDT) derived from the deep learning models (Jumper et al., 2021). Additionally, the relative position of amino acids to each other is annotated with an expected distance error in Ångströms (predicted aligned error, PAE). Although experimental structures are available for a large set of proteins, we decided to base all our analyses on predicted structures only. This prevents ambiguities in the integration of multiple conflicting structures and allows us to leverage the complete sequence information, confidence metrics and PAE estimates.

We found that it is crucial to take AlphaFold's confidence and PAE metrics into account for best accuracy, in accordance with others (Akdal et al., 2021). Therefore, we developed a prediction-aware metric. To evaluate amino acid side chain exposure, we built upon the previously introduced half-sphere exposure (HSE) (Hamelryck, 2005; Heffernan et al., 2016). This method essentially calculates the half-sphere of a given amino acid in the direction of its side chain at a defined radius, and counts the number of alpha carbon atoms from other residues within it, with a larger number reflecting less exposure and vice-versa (Supplementary Figure 1A). We adjusted the HSE to take prediction uncertainties into account, meaning that an alpha carbon atom is only considered as a neighbor if it still lies within the defined radius after addition of its PAE for this alpha carbon pair. We further introduce an angle parameter that determines whether to consider the full-sphere, half-sphere, or any other angle in direction of the amino acid side chain. Accordingly, we termed our metric prediction-aware part-sphere exposure (pPSE). To illustrate, Figure 1A shows the AlphaFold predicted structure of Mitogen-activated protein kinase 3 (MAPK3) colored by pLDDT and by the pPSE using a radius of 12 Å and an angle of 70°.

The higher the pPSE, the more other amino acids are in close proximity to the amino acid being evaluated, and hence the more structured its environment. In this respect pPSE offers a similar metric as the commonly used solvent accessible surface area (SASA), or relative SASA (RSA). However, the pPSE directly considers side chain orientation and, more importantly, it takes the prediction error of AlphaFold into account. Estimating the pPSEs for all amino acids in the 20,053 predicted human protein structures on AlphaFold only takes minutes on a laptop computer with our implementation. This makes the tool especially useful for system-wide studies where tens of thousands of proteins are evaluated for particular properties. In a community effort, it was recently shown that a smoothed AlphaFold confidence metric (pLDDT) or RSA metric based on the predicted structures can confidently determine IDRs (Akdal et al., 2021), improving on IUPred2, a state-of-the-art tool for IDR prediction (Mészáros et al., 2018). We found that our smoothed pPSE metric even obtains slightly better results when using a radius of 24 Å and a full sphere (the true positive rate improves from 83% for RSA to 86% for pPSE, see Figure 1A, 1B, 1C and Supplementary Figure 1C and 1D). Importantly, considering the positional uncertainty during pPSE estimation considerably improves IDR prediction compared to neglecting it (TPR increase from 79% to 86%).

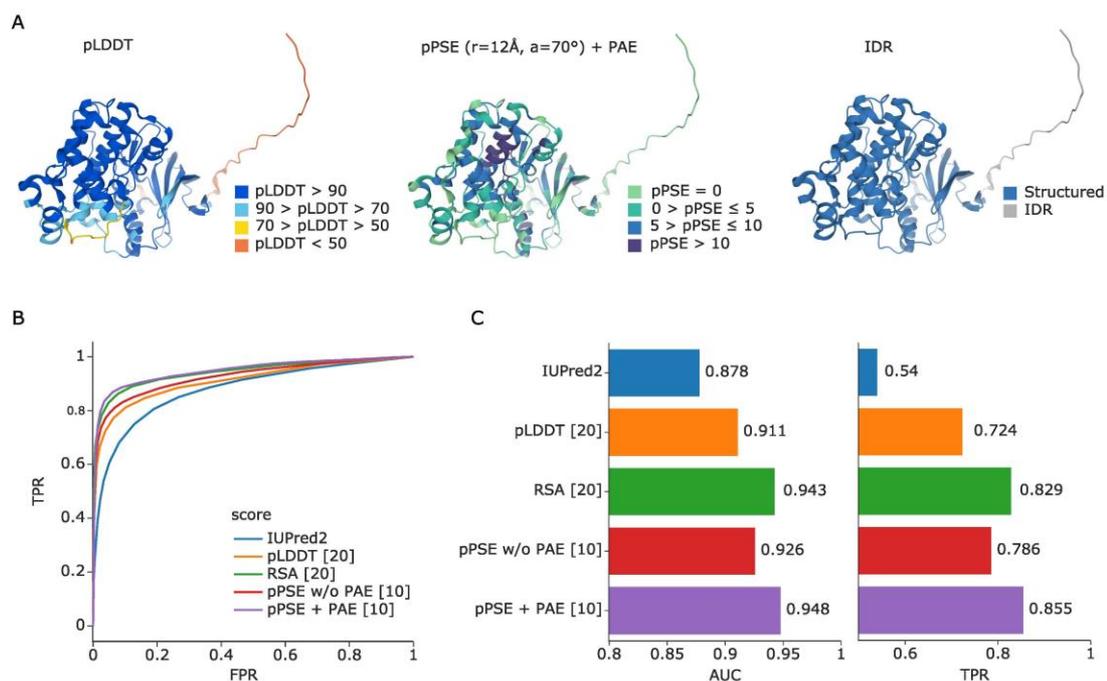


Figure 1. Estimation of amino acid side chain exposure and intrinsically disordered regions (IDRs). (A) AlphaFold predicted structure of Mitogen-activated protein kinase 3 (MAPK3) colored by prediction confidence (pLDDT, left), colored by our prediction-aware part-sphere exposure (pPSE) metric using a radius of 12 Å and an angle of 70° (center), and colored by our prediction of structured regions and IDRs (right). (B) Receiver operating characteristic (ROC) curve for predicting IDRs based on IUPred2 in comparison to the smoothed pLDDT confidence scores from AlphaFold, the smoothed relative solvent accessible surface areas (RSA), and the pPSE with (+) and without (w/o) considering the predicted aligned error (PAE) (radius = 24 Å, angle = 70°). (C) Corresponding area under the curve (AUC) values and the true positive rates (TPRs) at a 5% false positive rate (FPR). The numbers in square brackets behind each metric indicate the smoothing windows that were used, see Supplementary Figure 1 for a comprehensive parameter screen.

Most PTMs are enriched in intrinsically disordered regions, whereas ubiquitinations accumulate in structured domains

Having the proteome-wide information on IDRs at hand, we next performed an enrichment analysis of different PTMs located within those regions across the entire human proteome. We first used all PTMs annotated in the PhosphoSitePlus database that overlapped with structural data, comprising a total of 334,529 sites, including phosphorylations (p), ubiquitinations (ub), sumoylations

(sm), acetylations (ac), methylations (m) and the glycosylations O-GalNAc (gl) and O-GlcNAc (ga) (Hornbeck et al., 2015). In agreement with previous observations of phosphorylations (Iakoucheva et al., 2004; Tyanova et al., 2013), most PTMs were indeed significantly enriched in IDRs (Figure 2A). In contrast, our analysis revealed that ubiquitinations and, to a lesser extent, acetylations were significantly underrepresented in IDRs (ubiquitination: odds-ratio = 0.6, adj. p-value \approx 0, number of sites = 91,388; acetylation: odds-ratio 0.9, adj. p-value = 6e-08, number of sites = 21,202). However, if only PTM sites with a known regulatory function are considered, this effect disappeared for ubiquitination and was even reversed for acetylation (ubiquitination: odds-ratio = 1.0, adj. p-value = 0.7, number of sites = 451; acetylation: odds-ratio 2.0, adj. p-value = 1e-18, number of sites = 631).

A possible explanation for non-regulatory ubiquitination sites in structured regions is the tagging of misfolded proteins for degradation by the proteasome. Importantly, most datasets that contribute ubiquitination sites to PhosphoSitePlus are from samples treated with proteasome inhibitors (Figure 2B). This leads to the accumulation of misfolded proteins that presumably expose normally inaccessible lysine residues. Furthermore, ubiquitin might also be specifically attached to structured regions to destabilize the protein fold, creating new short IDRs that are required for proteasome binding and subsequent degradation (Carroll et al., 2020).

To directly test our hypothesis, we contrasted ubiquitination sites from proteasome inhibitor-treated and untreated samples in the same experiment (Hansen et al., 2021). Interestingly, ubiquitination sites unique to the proteasome inhibition condition confirm the overrepresentation of ubiquitination in structured regions (Figure 2A right, odds-ratio = 0.4, adj. p-value = 0, number of sites = 19,517). The same effect can still be observed for the sites shared between both datasets (odds-ratio = 0.6, adj. p-value = 1e-193, number of sites = 11,741), whereas the ubiquitination sites unique to the untreated condition are enriched in IDRs (odds-ratio = 1.4, adj. p-value = 1e-32, number of sites = 6321), similar to most other PTMs. Overall, 78 % of ubiquitinated lysines unique to the inhibitor treatment condition and 71 % of shared ubiquitin sites were in structured regions. In notable contrast, in the uninhibited condition, only 50 % of observed ubiquitinations are in structured regions.

To further pursue these findings, we disregarded all amino acids in predicted IDRs and asked if PTMs are enriched in amino acids with side chains of high versus low exposure within structured regions. To this end, we calculated the pPSE for each amino acid at a radius of 12 Å and an angle of 70°. We considered amino acids with a pPSE \leq 5 to have a high exposure and those with a pPSE $>$ 5 as low exposure (see Supplementary Figure 1B and the methods section for details on the cutoff selection). Due to much lower numbers of annotated PTMs in structured regions, statistical

significance decreases, but phosphorylations were still enriched in amino acids with high side chain exposure, whereas ubiquitinations were enriched in those with low side chain exposure (Supplementary Figure 2). This indicates that ubiquitinations are located on lysines that are buried within the structure of a properly folded protein rather than on outwards facing amino acids of a helix or beta-sheet at the protein's surface.

In addition to these global analyses, we further explored modified proteins individually to test if PTMs were specifically enriched in certain structural elements. For ubiquitination, this revealed 71 proteins with a significant enrichment in structured regions (odds-ratio < 1, adj. p-value ≤ 0.05). Interestingly 80% of them were DNA or RNA binding proteins. A striking example is Ras GTPase-activating protein-binding protein 2 (G3BP2), an RNA-binding protein that plays an essential role in cytoplasmic stress granule formation (Matsuki et al., 2013). Here all 9 ubiquitination sites are in structured regions (Figure 2C).

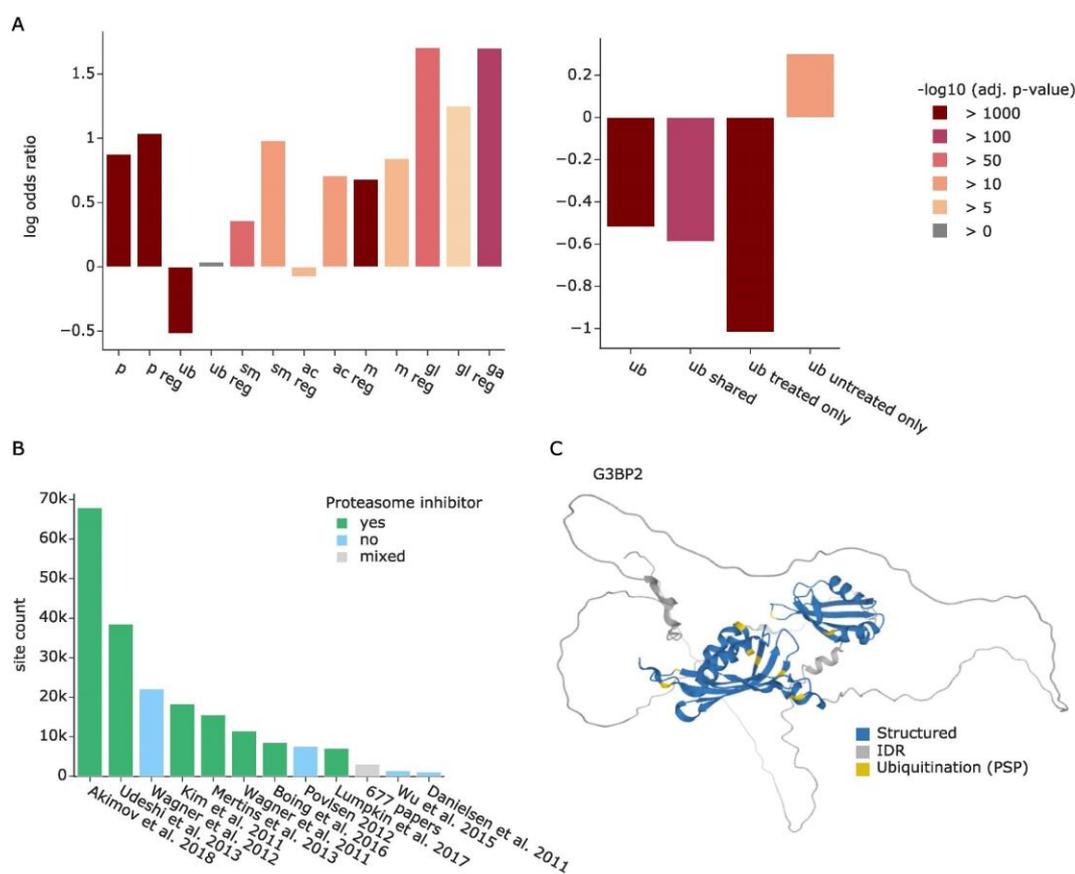


Figure 2. Enrichment analysis of PTMs in IDRs. (A) Enrichment of different PTMs annotated in the PhosphoSitePlus database in IDRs (left). Enrichment of ubiquitinated lysines annotated in PhosphoSitePlus versus ubiquitinations detected in a dataset treated with proteasome inhibitor or untreated (right). PTMs are abbreviated as follows: phosphorylations (p), ubiquitinations (ub), sumoylations (sm), acetylations (ac), methylations (m) and the glycosylations O-GalNAc (gl) and O-GlcNAc (ga). **(B)** Overview of datasets that contribute ubiquitination sites to PhosphoSitePlus and their use of proteasome inhibitors. **(C)** AlphaFold predicted structure of Ras GTPase-activating protein-binding protein 2 (G3BP2) colored by structured regions (blue) and predicted IDRs (grey) as well as ubiquitination sites annotated in PhosphoSitePlus (yellow).

Improving sequence motif analysis through structural context

Post-translational modifications are commonly introduced by dedicated enzymes such as kinases for phosphorylation, E3-ligases for ubiquitination and proteases for proteolytic cleavage, which generally recognize specific sequence motifs. Given that most PTMs have a preference for exposed amino acids, we reasoned that sequence motifs in accessible protein regions should be preferred compared to inaccessible ones, adding another layer of selectivity. To explore this hypothesis, we first selected a curated list of kinase phosphorylation motifs available in Perseus (Tyanova et al., 2016). Based on phosphosites in both PhosphoSitePlus and a recent in-depth, COVID-related phospho study (Stukalov et al., 2021), we first confirmed that phosphorylations are generally enriched in kinase phosphorylation motifs compared to all possible STY sites. This effect is even more pronounced for regulatory sites and sites from the Stukalov study (Figure 3A). Confirming our hypothesis, motifs in IDRs and motifs harboring exposed STY sites were indeed preferentially modified (Figure 3B and Supplementary Figure 3A).

These results highlight that proteome-wide structural information can provide valuable insights for motif analysis and help interpret experiments determining enzyme-substrate relationships. We illustrate this with a large in vitro kinase substrate screen by Ishihama and colleagues (Sugiyama et al., 2019). These authors dephosphorylated HeLa cell lysates with phosphatases, which they then deactivated by heat. The resulting – partially denatured – sample then served as a substrate pool to which 385 different recombinant human protein kinases were individually added to investigate which kinases phosphorylate which specific amino acid sites. Their study resulted in an unprecedented set of 20,669 phosphosites, 175,574 proposed kinase-substrate relationships and 1,427 kinase phosphorylation motifs.

Based on the denaturation of the proteome, and as already indicated by the authors, we speculated that many of the previously inaccessible sites would now be amenable to phosphorylation,

providing an ideal test case for structure-based interpretation. Indeed, our analysis revealed that the identified phosphosites were not enriched in IDRs, in contrast to the above examples. However, the sites that overlapped with other studies did show the expected enrichment in IDRs (Figure 3C). These results suggest that structural information can be used to refine the list of reported phosphosites to a set that better represents the sites expected to occur on endogenously folded proteins.

To further test the effect of 3D exposure filtering on sequence motif analysis, we selected kinases and performed a motif analysis separately for sites of high surface exposure ($pPSE \leq 5$) and sites of low exposure ($pPSE > 5$) using the PSSMSearch tool (Krystkowiak et al., 2018; O'Shea et al., 2013). As can clearly be seen in Figure 3D, kinases showed striking differences in sequence motifs between sites of high and low exposure. While motifs for sites of high exposure are mostly in agreement with the reference set provided by Perseus (Tyanova et al., 2016), this was not the case for the sites with low exposure. To account for the fact that there are fewer sites in structured regions, we also selected random, equally sized subsets of the high-exposure sites and repeated the motif analysis. This resulted in similar patterns as for the full set of sites, but with lower enrichment scores (Supplementary Figure 3B).

In the case of the RAC-alpha serine/threonine-protein kinase (ACT1), the phosphosites of high exposure clearly display the R-x-R-x-x-pS-F motif (Figure 3D, top left panel). In contrast, phosphosites of low exposure only provide a noisy motif (Figure 3D, top right panel). For other kinases, such as Stress-activated protein kinase JNK1, a serine/threonine-specific protein kinase, the phosphosites of low exposure even have an unexpected enrichment for a phosphorylated central tyrosine residue and the proline at the +1 position is hardly enriched (Figure 3D, right panel).

Together, these results establish that the structural information from AlphaFold and the tools presented herein can guide determining potential regulatory PTM sites found by in vitro screens, increasing the confidence of measured kinase-substrate pairs by filtering out a subset that are less likely to be true in-vivo substrates. As we have shown, this can improve kinase phosphorylation motif predictions and help to interpret individual sites of interest. Here we focused on phosphorylations but we expect similar benefits for the analysis of any other types of motifs, including enzyme recognition or general protein binding.

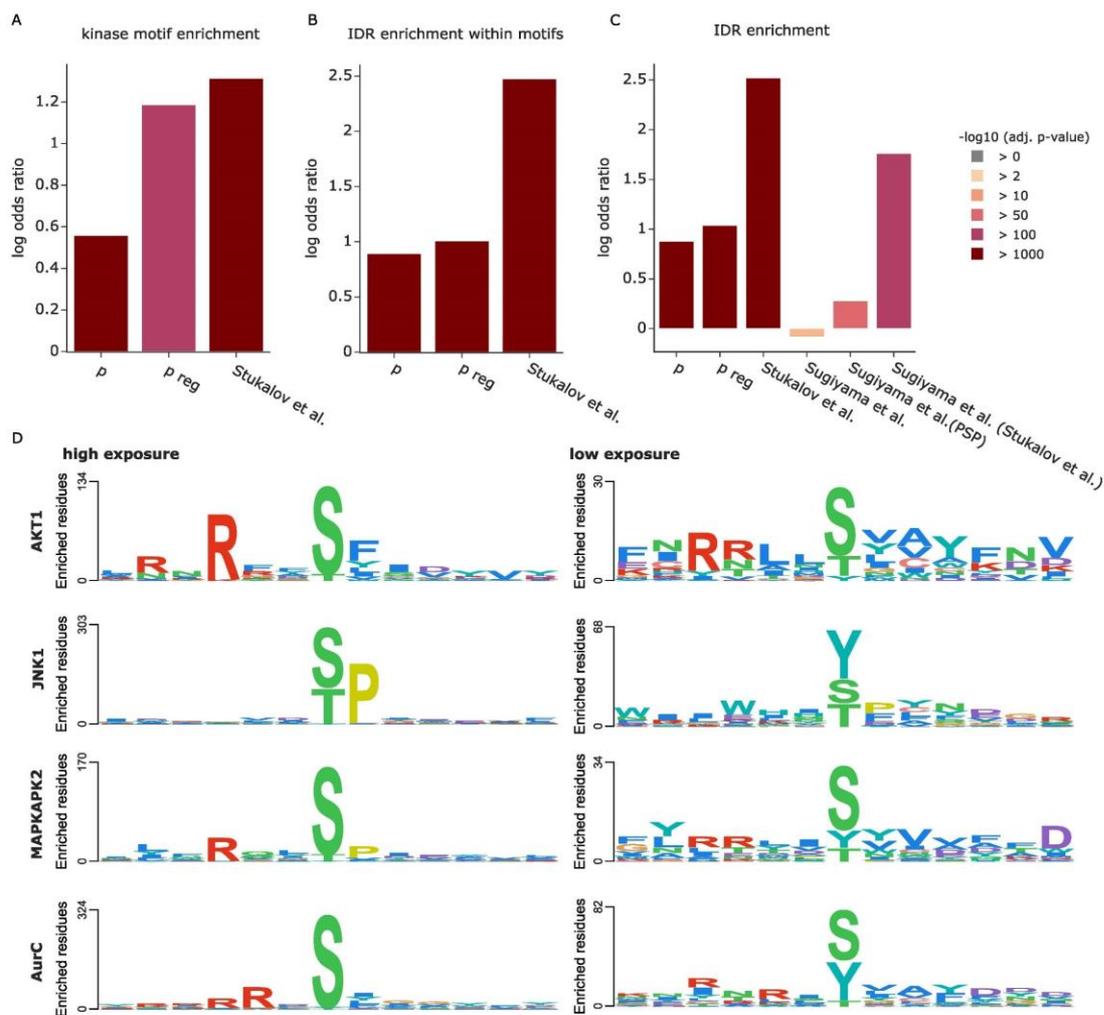


Figure 3. Exploiting the 3D context of kinase phosphorylation motifs. (A) Enrichment of phosphorylation events in kinase motifs compared to all possible STY sites. **(B)** Enrichment of phosphorylations in kinase motifs within IDRs compared to all possible kinase motif occurrences. **(C)** Enrichment of phosphorylations in IDRs compared to all possible STY sites. The phosphosites reported by Sugiyama et al. (Sugiyama et al., 2019) were filtered for sites also reported in PhosphoSitePlus and by Stukalov et al. (Stukalov et al., 2021). **(D)** Sequence logos for different kinases. The PSSMSearch tool (Krystkowiak et al., 2018) was used with a log odds scoring method (O’Shea et al., 2013).

Functionally relevant PTMs are enriched in short IDRs within large structured domains exemplified by kinase activation loops

Having established the enrichment of most PTMs in unstructured regions in the dataset, we further explored the structural context of this effect. We found that these PTMs are often located in short IDRs that are embedded within larger structured domains. To investigate if this was a random effect or whether these short IDRs could be of biological relevance, we extracted all proteins with short IDRs of maximally 20 amino acids length between two flanking structured regions of at least 80 amino acids. Among the 20,053 human proteins in AlphaFold DB, 2,454 have such a pattern. Notably, enrichment analysis of these proteins revealed a significant overrepresentation of GO molecular functions related to ATP binding, protein kinase activity, ATPase activity, transmembrane transporter activity and motor activity (Figure 5A).

To further evaluate the relevance of the short IDRs, we analyzed their occupancy with functional phosphosites. To this end, we first considered regulatory phosphosites in PhosphoSitePlus (Hornbeck et al., 2015). Our analysis revealed that these regulatory sites, as compared to phosphosites in general, are significantly enriched in short IDRs versus all IDRs (odds-ratio = 1.57, adj. p-value = 0.001, Figure 2B). We also extracted a second set of phosphosites from the above-mentioned study of Beltrao and co-workers, where sites were given a functional score between zero and one, with scores equal to or above 0.5 considered functional (Ochoa et al., 2019). We indeed observed that the higher the score, the stronger the enrichment of functional phosphosites in short IDRs, ranging from an odds-ratio of 1.43 at a cut-off of 0.5 to an odds-ratio of 7.7 at 0.9 (Figure 4B). This raises the exciting possibility that at least a subset of these phosphorylation sites may play important roles in structural and functional rearrangements of their neighboring domains.

Due to the strong enrichment of kinases among proteins with short IDRs, we next investigated whether they overlap with any known kinase substructures annotated in KinaseMD (Hu et al., 2021). That database contains substructure annotations for 388 kinases, 365 of which also have predicted structures in AlphaFold DB. We found a large overlap of 72 short IDRs with the 309 annotated activation loops, but none with the 171 G-loops or 230 C α -helix positions in these kinases. If the five amino acids flanking a short IDR are also considered, this number increases to 79 (also see methods section). These results are particularly interesting, because the activation loops of many kinases undergo structural rearrangements upon phosphorylation (Nolen et al., 2004). Strikingly, 55 of the 79 kinases (70 %) with an overlap of the extended short IDR and the annotated activation loop have an annotated regulatory phosphosite or a functional score higher than 0.5, for a total of 99 different

phosphosites. To illustrate, receptor-interacting serine/threonine-protein kinase 2 (RIPK2) and Mitogen-activated protein kinase kinase kinase kinase 1 (MAP4K1) both show an overlap between our predicted short IDR and the annotated activation loop with known regulatory phosphosites (Figure 4C, D and E).

Next, we evaluated short IDRs outside of annotated kinase activation loops. An interesting example of these is the serine/threonine-protein kinase CHK2 (CHEK2). Although the annotated activation loop (amino acids 367 to 389) was not detected as a short IDR, our data contained an alternative short IDR region (amino acids 262 and 263). Directly flanking this IDR is a phosphorylation site with a functional score of 0.44 (S260). Notably, CHEK2 was reported to be autophosphorylated at residue S260, which is important for triggering a conformational change in CHEK2 that favors a dissociation of dimers into fully active monomers (Zannini et al., 2014).

We also observed short IDRs in many proteins apart from kinases. One example is Band 3 anion transport protein (SLC4A1) where we found that the short IDR from residue 354 to 369 contains a known regulatory phosphosite on Y359 (Flatt et al., 2020). In addition to phosphorylations, other PTMs might also be biologically relevant in the short IDRs. Indeed, one of three regulatory ubiquitination sites in another solute carrier protein - SLC22A6 – is located directly in a short IDR, whereas the other two are in close proximity (Figure 4F). These three ubiquitination sites have previously been shown to play an important role for the internalization of this protein (Li et al., 2013).

Compared to phosphorylation, the percentage of other PTM sites with known functions is even smaller. Our findings suggest that selecting candidates from PTM sites within or in close proximity to short IDRs is a promising strategy to discover functional relevance. We found that our predicted short IDRs extended by five amino acids contain a wealth of PTM sites that are not yet annotated as regulatory in PhosphoSitePlus (1,437 phosphosites, 898 ubiquitination sites, 118 acetylations, 43 sumoylations, 53 methylations, 33 GalNAc and 1 GlcNAc) (Figure 4G, Supplementary dataset 1). We further provide a list of all human short IDRs for researchers to explore their favorite proteins, enabling the integration of own experimental data from PTM studies or other types of studies, such as mutational screens (Supplementary dataset 2).

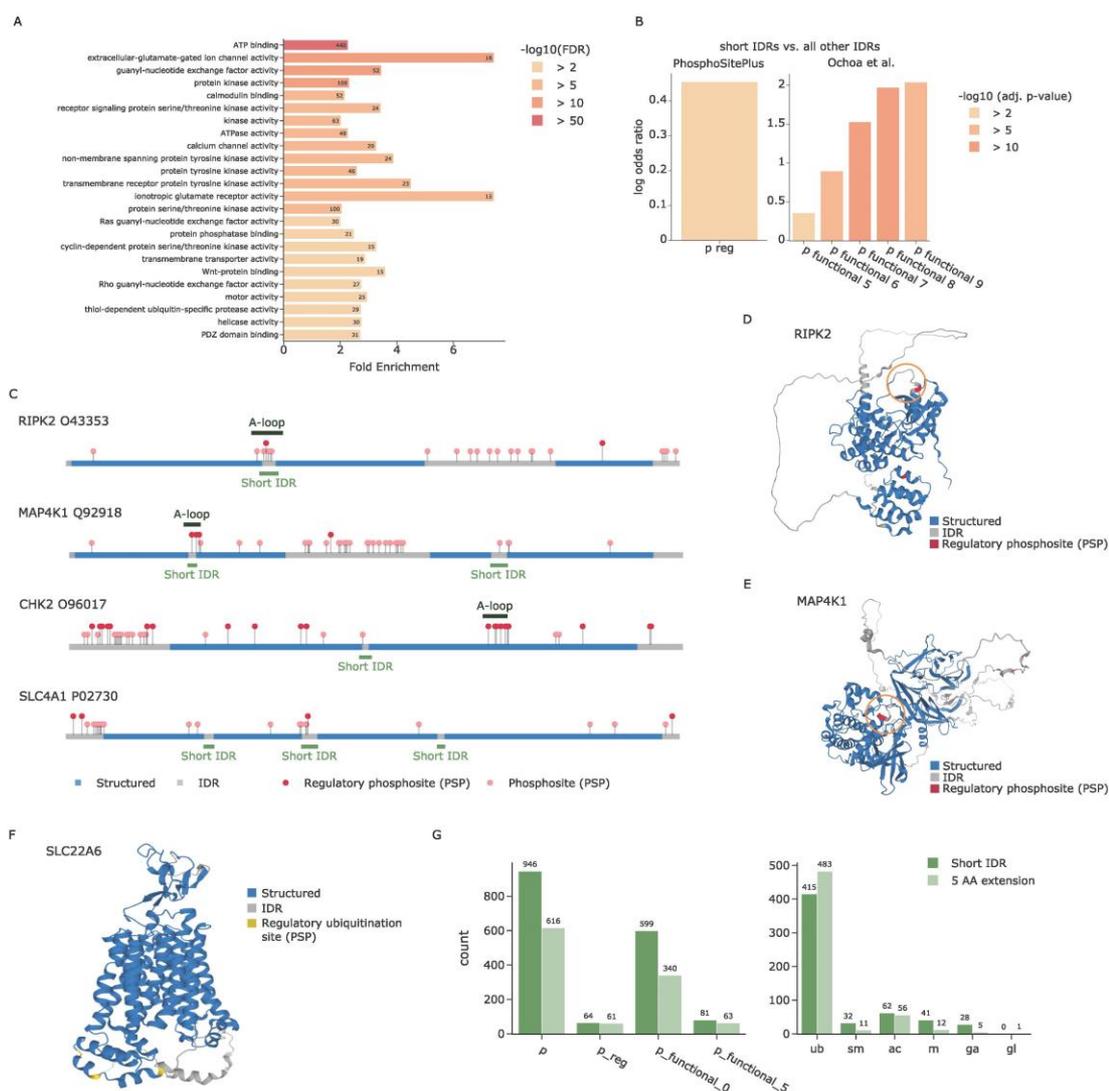


Figure 4. Regulatory PTMs accumulate in short IDRs. (A) Enrichment analysis of proteins with short IDRs. **(B)** Enrichment of regulatory phosphosites from PhosphoSitePlus in short IDRs compared to all other IDRs. **(C)** Sequence plot showing the N- to C-terminus of different proteins colored by whether the amino acid is part of a structured region (blue) or an IDR (grey). All phosphosites annotated in PhosphoSitePlus are indicated by circles. Regulatory sites are colored in dark red and stand out higher than non-regulatory sites (salmon). Regions of short IDRs including a five amino acid extension are indicated in light green below the sequence. Annotated kinase activation loops (A-loops) from KinaseMD are indicated in dark green above the sequence. The predicted structures of RIPK2 **(D)** and MAP4K1 **(E)** are colored by structured regions (blue) and predicted IDRs (grey) as well as known regulatory phosphosites annotated in PhosphoSitePlus (dark red). Specific regions of interest are highlighted by an orange circle. **(F)** The predicted structure of SLC22A6 is colored by structured regions (blue) and predicted IDRs (grey) as well as known regulatory ubiquitination sites annotated in PhosphoSitePlus

(yellow). **(G)** Overview of phosphorylations (left) and other PTMs (right) that lie within short unstructured regions or their flanking five amino acids (AA).

PTMs on proteins preferentially occur in three-dimensional clusters

It is well known that many proteins have hotspots of modifications. For example, multisite phosphorylation in specific sequence regions is critical in regulating the activity of many enzymes and their binding properties (Cohen, 2000). Furthermore, in the context of circadian biology, we recently showed that regulated ubiquitinations frequently occur in sequence proximity (Hansen et al., 2021). Beyond co-localization of the same PTM types, phospho-acceptors near PTM-modified lysines were shown to be preferentially phosphorylated in comparison to more distant residues (Beltrao et al., 2012). Those prior findings were obtained on the basis of linear sequence analysis. Now, with the spatial coordinates of each PTM acceptor residue provided by AlphaFold, we set out to evaluate PTM proximity in three-dimensional space.

First, we investigated whether PTM acceptors near a modified amino acid residue are more frequently observed to also be modified compared to more distant residues or to random expectation. For this, we extended the strategy of Krogan and co-workers (Beltrao et al., 2012) to evaluate distance in 3D space and to assess both individual PTM types and PTM co-occurrence (see methods section). Importantly, our metric considers the predicted positional uncertainty between any two PTM sites as a factor in the analysis. Furthermore, we only take structured regions and short IDRs into account for the proximity analysis (that is, we removed all IDRs of more than 20 amino acids). This ensures that proximity results are not influenced by regions of high structural uncertainty, as is the case for IDRs. It further avoids any biases that arise from the fact that many PTMs are enriched in IDRs and tend to cluster there in linear sequence space.

Our analysis revealed that the observed sites of PTM types annotated in PhosphoSitePlus indeed form 3D modification hotspots. Phosphorylations, ubiquitinations, sumoylations, acetylations and methylations each form tight clusters in 3D space, where proximal amino acids are preferentially modified compared to more distant residues or an equivalent random selection (Figure 5A). Due to an overall lower number of O-GlcNAc and O-GalNAc modified sites, the results for these modifications are less conclusive but also show a similar trend (Supplementary Figure 4). In addition to evaluating PTM types by themselves, we further investigated co-localization of different PTM types. This confirmed that phospho-acceptors near modified lysines (including ubiquitination, sumoylation and acetylation) are more frequently phosphorylated compared to random expectation (Beltrao et al.,

2012) (Figure 5B). This is also true for methylated lysines and arginines. Conversely, investigating ubiquitination sites near other PTMs revealed that they also preferentially occur close to phosphosites (Figure 5C). Other lysine modifications, however, often compete for the same or directly neighboring residues, but they do not generally favor proximity. Overall, our analysis reveals that many proteins have specific 3D regions and folds that are particularly prone to being modified by the same or different PTMs. This structurally supports the notion of PTM cross-talk.

Following these global analyses, we next explored 3D PTM clusters of all individual proteins. For this, we calculated all pairwise distances between modified amino acids and compared their average against a distribution of random PTM sites (see methods section). We again only considered structured protein regions and included the positional uncertainty between any two PTM sites in the distance calculation. Clustering analysis of phosphorylation and ubiquitination sites in PhosphoSitePlus revealed many proteins with significant PTM clusters, showing a strong enrichment for transmembrane proteins (Figure 5D). On those proteins, we detected these PTMs on the cytosolic domains and in 3D proximity, nicely confirming that the proximity analysis worked as intended.

To enable a more fine-grained inspection of PTM clusters, we explored an in-house phospho-dataset (Stukalov et al., 2021). Of 47 phosphoproteins with three or more sites in structured regions, 3D proximity analysis yielded six significant ones (adj. p-value ≤ 0.05 and ≥ 3 phosphosites, Supplementary Dataset 3). As an example, the mitochondrial Pyruvate dehydrogenase E1 component subunit alpha (PDHA1) had an adjusted proximity p-value of 0.005 and all six detected phosphosites are located within one protein pocket (Figure 5E). Three of these (S232, S293 and S300) have previously been reported to be substrates of PDK family kinases. Any single one of these is sufficient to inactivate PDHA1 and dephosphorylation of all sites is required for reactivation (Kato et al., 2008). Our 3D analysis suggests that the other three phosphorylation sites would have the same effect. Another interesting example is Aldo-keto reductase family 1 member B1 (AKR1B1) (adj. p-value = 0.03). In linear sequence space, S211 and S215 are close together, whereas S23 and Y49 are far apart. However, in the folded protein structure, all four are in the same pocket (Figure 5F). Interestingly, it also contains the annotated NADP binding site of AKR1B1, which consists of two distinct sequence stretches from amino acids 10 to 19 and 211 to 273 (based on UniProt annotation).

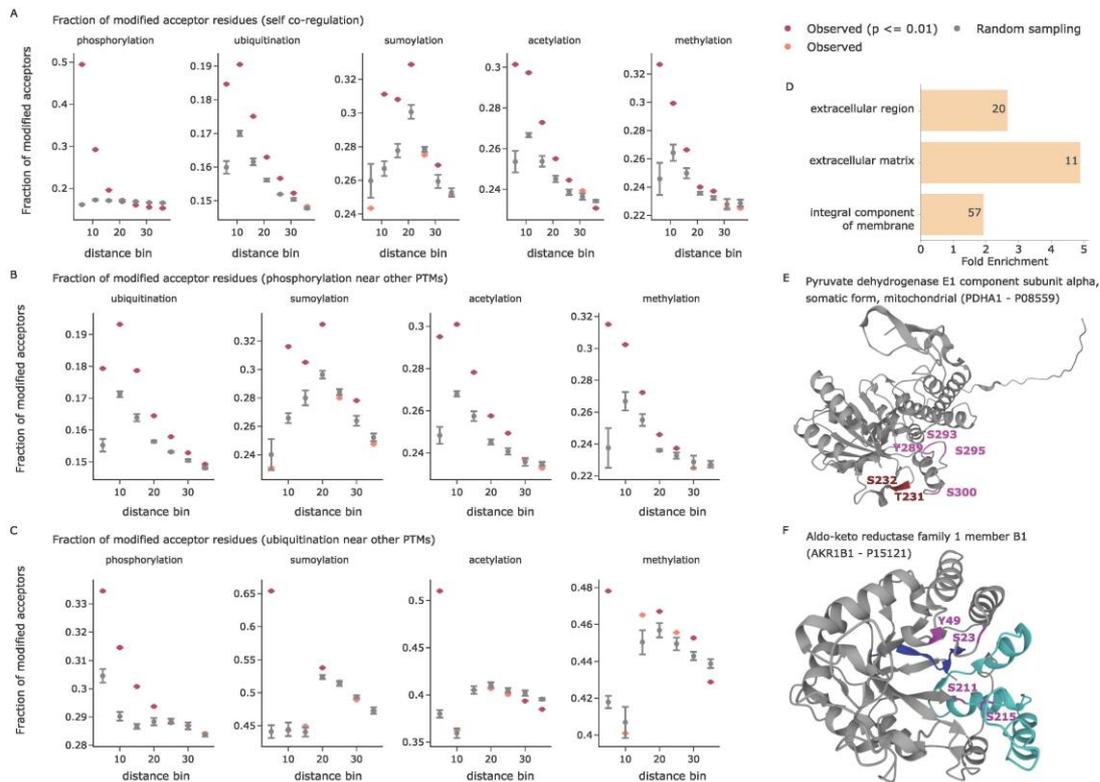


Figure 5. PTM proximity analysis in 3D. (A) The fraction of modified PTM acceptor residues is shown as a function of the 3D distance to a given modified amino acid in Å. Observed values (indicated in red when statistically significant and colored in salmon otherwise) are compared to the mean of five random samples including the same number of modified PTM sites (grey). Error bars indicate one standard deviation. The x-axes are divided in distance bins ranging from each previous bin to the indicated cutoff in Å. (B) The fraction of modified phospho-acceptor residues is shown as function of the 3D distance to a given modified amino acid in Å. (C) The fraction of ubiquitinated lysines is shown as function of the 3D distance to a given modified amino acid in Å. The smallest bin shows competition for the same central lysine residue. (D) Enrichment analysis of proteins with 3D phospho- and/or ubiquitination clusters (FDR ≤ 0.01). (E) The predicted structure of Pyruvate dehydrogenase E1 component subunit alpha (PDHA1). Phosphorylations on the phospho-loop A are indicated in dark red (T231 and S232). The phosphorylations on phospho-loop B are indicated in magenta (Y289, S293, S295 and S300) (Kato et al., 2008). (F) The predicted structure of Aldo-keto reductase family 1 member B1 (AKR1B1). Residues annotated as NADP binding sites are highlighted in blue (amino acids 10 to 19) and turquoise (amino acids 211 to 273). Phosphorylations are indicated in magenta.

Community resources for enabling the systematic integration of PTM data with structure predictions from AlphaFold

This study only scratches the surface of biological insights that can be gained from combining PTM data with structural information. To enable the community to further investigate various research questions of interest to them in their own data or in public repositories, we created a toolset that facilitates systematic data exploration and integration. First, we provide StructureMap, an open-source Python package for processing predicted structures from AlphaFold DB and for integrating the data with PTM information. Its functionalities include (1) accession of predicted structures from AlphaFold DB and extraction of essential information into an internal data format, (2) calculation of the pPSEs of individual amino acids as well as estimation of IDRs, (3) extraction of short IDRs for PTM site prioritization, (4) import and formatting of PTM datasets, (5) enrichment analyses of PTMs in different structural regions and IDRs, (6) motif analysis in 3D context and filtering based on side chain exposure, and (7) multi-dimensional proximity analysis of PTMs. To further enable easy visualization of PTMs on the three-dimensional structure of proteins, we also extended our previously published AlphaMap tool, which is available as Python library as well as a stand-alone application with graphical user interface (Voytik et al., 2021). The source code of both tools is openly available with an Apache license on the MannLabs GitHub page and includes extensive documentation to readily enable researchers to understand and further adopt code to any specific needs (<https://github.com/MannLabs/structuremap> and <https://github.com/MannLabs/alphamap>).

Discussion and Outlook

PTMs provide essential mechanisms to regulate the activity and function of proteins. Although mass-spectrometry based proteomics routinely enables the identification and quantification of thousands of PTMs, systematic assessment of their functional relevance remains a persisting challenge. While previous work already demonstrated the merits of structural information for PTM analyses (Beltrao et al., 2012; Iakoucheva et al., 2004; Kamacioglu et al., 2021; Ochoa et al., 2019; Tyanova et al., 2013), the recent revolution in computational protein structure prediction (Baek et al., 2021; Jumper et al., 2021) only now enables the proteome-wide integration of structural information with PTM data. In this study, we provide a first overview of how the comprehensive structural context of all detected PTMs can provide global and protein-specific insights into biological mechanisms, to filter in vitro datasets for physiological relevance and to identify promising candidates for biochemical follow-up studies.

Key to most of our analyses was that we used whole proteome structural information to determine the exposure of each individual amino acid side chain, thus providing a measure of how amenable that residue is for harboring a modification (Figure 1A). In contrast to experimentally derived structures, the *in silico* structures come with prediction errors and positional uncertainties, which turned out to be crucial for assessing amino acid exposure (Akdal et al., 2021) (Figure 1B/C and Supplementary Figure 1). Our analyses confirmed that most PTMs are strongly enriched on exposed amino acids compared to residues that are buried within the protein fold (Supplementary Figure 2). This effect is even more evident when comparing IDRs and structured regions (Figure 2A) and for specific kinase phosphorylation motifs (Figure 3B and Supplementary Figure 3A).

In contrast to the other analyzed PTMs, ubiquitinations were strikingly enriched on structured regions and on amino acids that are expected to be inaccessible (Figure 2A and Supplementary Figure 2). We showed that this effect is triggered by proteasome inhibition, supporting the idea that the ubiquitinations on structured regions are predominantly placed on misfolded proteins, tagged for degradation. Thus, our toolkit can help distinguish ubiquitination associated with protein quality control from that mediating site-specific regulation. Interestingly, a majority of the proteins with ubiquitin modifications enriched in structured regions are DNA or RNA binders, many of which are known to be ubiquitinated in cellular stress response (Maxwell et al., 2021). As an example, the RNA-binding protein G3BP2, which is essential for cytoplasmic stress granule formation, has all of its nine ubiquitination sites in structured regions (Matsuki et al., 2013). SET8 is ubiquitinated during DNA damage response, causing its degradation followed by chromatin rearrangements (Jørgensen et al., 2011), and our analysis placed 9 of 10 ubiquitination sites in normally structured regions (adj. p-value = 0.01). Together, these results raise the possibility that ubiquitination of DNA and RNA binders and their direct regulators provides an effective regulatory mechanism for cellular stress responses. Further investigations could combine the above analyses with information about different ubiquitin side chain architectures to elucidate possible 3D topologies and associated functionalities.

As PTMs on properly folded proteins are expected to reside on exposed amino acid side chains, we reasoned that our 3D analysis could help to prioritize sites from experiments performed under less than physiological conditions. Here, we exemplified such a case by an *in vitro* kinase substrate screen (Sugiyama et al., 2019). While the screen allowed defining sequence preferences for kinase phosphorylation motifs, determining which sites could mediate *bona fide* regulation remained a challenge. Modified sites that are observed on inaccessible amino acid residues can be filtered out to reduce the target list to more physiologically relevant sites. Moreover, in our analysis, the motifs of best retained sites were strikingly better defined than those based on the discarded ones (Figure 3D).

We strongly suggest to employ such analyses in future in vitro PTM studies. Apart from phosphorylations, we expect similar benefits for the analysis of other types of PTMs or even for linear motifs involved in specific molecular binding events.

More generally, 3D analysis can highlight sites with a high potential to be functionally relevant. Our unbiased, global analysis revealed that regulatory phosphosites are strongly enriched in short unstructured regions between large folded domains (Figure 4B). Many of these short IDRs correspond to annotated kinase activation loops, which are known to undergo structural rearrangements upon phosphorylation (Nolen et al., 2004). Our analysis systematically reveals such functionally highly relevant sequence regions on a proteome-wide scale, opening up multiple interesting routes for further investigations: First, short IDRs in regions without known functional relevance could specifically be investigated. Second, PTMs lying in or adjacent to such regions could be prioritized in the selection of candidate PTMs for biochemical follow-up studies, given their potential to cause structural rearrangements with functional consequences. We provide the community with a resource of such short IDRs in the human proteome and also with a set of promising PTM candidates from PhosphoSitePlus, which lie within or directly adjacent to these short IDRs. Beyond phosphorylation, these candidate PTMs contain hundreds of ubiquitinations as well as tens of sumoylations, acetylations and methylations (Figure 4G). It would be exciting if future work in protein structure prediction enables the direct prediction of protein folds including PTMs.

Stepping back, the structural context of PTMs can not only be evaluated by means of spatial metrics, such as the exposure and folding state of individual amino acids, but also by using the three-dimensional fold directly. To this end, we evaluated PTM proximity in 3D space, revealing that not only phosphorylations, but also most other PTM types tend to cluster in protein folds (Figure 5A). Beyond this co-localization of PTMs of the same type, we further showed that phosphorylations and lysine modifications preferentially occur in 3D proximity (Figure 5B and 5C), extending previous findings in linear sequence space (Beltrao et al., 2012). Interestingly, lysine modifications, including ubiquitination, sumoylation, acetylation and methylation often compete for the same or directly neighboring residues, but they do not generally favor proximity (Figure 5C). A possible explanation for this effect could be the high reactivity of specific lysine residues, which will be an interesting follow-up to our study (Abbasov et al., 2021). We further investigated PTM proximity on individual proteins in a proteome-wide fashion. This systematic analysis resulted in the identification of interesting proteins with 3D clusters of phosphorylations in specific protein pockets. For instance, in PDHA1, the relevant phosphorylation sites may all be important for enzyme activation and inactivation (Kato et al., 2008). In AKR1B1, the relevant phosphorylations surround the known NADP binding site. We only

observed functionally relevant PTM hotspots in protein pockets for phosphorylations here, but we expect that future improvements in the coverage of other PTM types will also enable the identification of 3D clusters for them. Recent work extends the prediction of structures of individual proteins to the prediction of structures for entire protein complexes (Evans et al., 2021; Humphreys et al., 2021). The integration of proteome-wide estimates of binding interfaces may make studies of PTM co-regulation across proteins, such as the synergistic sumoylation of multiple proteins during DNA repair (Psakhye and Jentsch, 2012), feasible on a global scale.

In summary, our study is the first systematic assessment of the structural context of PTMs and how to leverage this information to gain novel biological insights and to augment proteomic studies. By providing the open-source Python packages StructureMap and AlphaMap, we equip the scientific community with a toolset to easily integrate information from AlphaFold predicted structures into any kind of proteomics study and to visualize proteomics data on the predicted protein fold. We envision that this will enable scientists to directly assess observed PTMs in their structural context and to evaluate their physiological feasibility, to prioritize promising candidates for functional follow-up studies, and to find three-dimensional hotspots of PTM regulation. Future work could explore 3D motifs for enzyme substrate recognition as well as the integration of protein binding information. Beyond our own work presented here, we are convinced that the systematic integration of structural information with proteomics data will open up new avenues to (re-)evaluate both old and new research questions.

Contributions

IB and MM conceived the study, IB performed all analyses; IB and SW implemented the StructureMap package; IB, WFZ and MTS integrated the 3D visualization into AlphaMap; IB, FMH, MT, MM and OK formulated different biological hypotheses and interpreted the results; BAS provided critical feedback to the paper; MM coordinated and supervised the study; IB and MM wrote the paper with input from all authors.

Acknowledgements

This study was supported by The Max-Planck Society for Advancement of Science and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de). IB acknowledges funding support from her Postdoc.Mobility fellowship granted by the Swiss National Science Foundation [P400PB_191046].

We thank Julia P. Schessner for valuable discussions about the implementation of the pPSE exposure metric. We also thank J. Rajan Prabu for sharing his knowledge in structural bioinformatics.

Conflict of interest

The authors declare that they have no conflict of interest.

Methods

Prediction of intrinsically disordered regions (IDRs) and benchmarking against IUPred2

To benchmark IDR prediction, we first downloaded ground truth datasets for both disordered and structured protein regions as defined by Beltrao and colleagues (Akdal et al., 2021). To avoid inconsistencies, we removed any amino acids that were labeled as both disordered and structured. We downloaded and formatted the AlphaFold structure predictions for 3062 of the remaining 3080 proteins. The other 18 proteins did not have a complete structure deposited in AlphaFold DB (Tunyasuvunakool et al., 2021; Varadi et al., 2021). As a reference dataset for our benchmark, we obtained direct IDR estimates for all proteins from IUPred2A (<https://iupred2a.elte.hu/>) (Mészáros et al., 2018).

RSA values for all amino acids of the benchmarking proteins were calculated using DSSP from Bio.PDB (version 1.79) using default parameters (Hamelryck and Manderick, 2003). Similar to Beltrao and colleagues, we applied a local smoothing of both the pLDDT and RSA values by averaging along the sequence with a window size of 5, 10, 15, 20 and 25 amino acids (Akdal et al., 2021).

To evaluate the ability of our pPSE metric to predict IDRs, we calculated the pPSE values of all amino acids with different parameter settings. Given the goal of finding IDRs, we chose a constant angle term of 180° . This corresponds to the full sphere exposure, which is independent of side chain orientation. For the distance, we screened over multiple parameters including 12 Å, 16 Å, 20 Å, 24 Å and 28 Å. pPSE values were calculated both with and without considering the PAE. Similar to the pLDDT and RSA values, pPSE values were also smoothed by averaging along the sequence with a window size of 5, 10, 15, 20 or 25 amino acids.

The performance of IUPred2, pLDDT, RSA and pPSE in predicting IDRs was evaluated based on the AUC and TPR at 5% FPR, when screening over different scoring thresholds. The best results in this analysis were obtained by the pPSE metric when considering PAEs and a distance threshold of 24 Å.

Here, a smoothed pPSE ≤ 34.27 obtained a TPR of 85% at a FPR of 5%. IDR prediction for our proteome wide analysis was therefore performed based on these parameters for pPSE calculation.

For evaluating the benchmark analysis, it is important to keep in mind that the annotations of structured regions and IDRs in the benchmark dataset were unfortunately not associated with a specified amino acid sequence, but only with sequence positions. While this should be correct for most proteins, there might be some differences between the original sequences for which the IDRs were predicted and the sequences used to predict structures by AlphaFold, thus potentially leading to slight inconsistencies.

All data and code for the IDR benchmark analysis will be available as a Jupyter notebook.

Definition of short IDRs

The goal of our 'short IDR' detection strategy was to find flexible regions that are embedded into large folded domains. We defined short IDRs as amino acid stretches that we predicted to be IDRs and that comprise maximally 20 residues sandwiched between two flanking structured regions of at least 80 amino acids. Based on our strategy to identify IDRs by a sliding average of the pPSE exposure, some short IDRs only contain very few amino acids that reach the threshold to be classified as IDRs. To account for this effect, we introduced the extended regions of five amino acids on either side of short IDRs. We further expect the extension of short IDRs to be very beneficial in identifying potentially interesting PTMs, given that modifications on residues directly next to a flexible region could also affect its 3D conformation.

pPSE parameters for estimating amino acid side chain exposure

The pPSEs for side chain exposure estimation were calculated based on a distance threshold of 12 Å and an angle of 70°. These values were chosen based on considerations about the average size of amino acids of approximately 3.5 Å and side chain flexibility around the direction of the beta carbon. Visualization of the resulting pPSE values showed expected patterns, as can be seen in Figure 1A (middle):

- Amino acids at the core of the protein fold have the highest pPSE values.
- Beta sheets have an alternating pattern of high and low exposure sites.
- Outwards-facing amino acids of alpha-helices have a higher exposure than inwards-facing sites.

- Amino acids in badly predicted sequence regions (e.g. the tailing sequence) have a pPSE value close to zero.

To categorize amino acids into low versus high exposure, we selected a pPSE threshold of 5. This means that amino acids with a $pPSE \leq 5$ were considered to have a high exposure. Supplementary Figure 1B illustrates the distribution of pPSE values across all amino acids in structured protein regions (non-IDRs).

PTM enrichment analyses

All enrichment analyses were performed using the two-sided Fischer's exact test available in `scipy.stats` version 1.7.1 (Virtanen et al., 2020). P-values were subsequently adjusted for multiple testing (Benjamini and Hochberg, 1995) using `statsmodels.stats.multitest.multipletests` version 0.13.0 (Seabold and Perktold, 2010).

Motif analysis with PSSMMatch

The kinase specific phosphosites were extracted from Supplementary Table S2 from Sugiyama et al. (Sugiyama et al., 2019). UniProt IDs were matched to UniProt accessions and the sequence windows corresponding to ± 6 amino acids around each phosphosite were extracted from the current human protein fasta file (version: 2021_03, downloaded on 02.08.2021). Any sites not matching the expected phosphoacceptor residue were removed from the analysis. For each selected kinase, targeted sequence windows were filtered as follows:

- 1) All reported sequence windows with a phosphosite $pPSE \leq 5$. The number of resulting sequence windows is defined as N_{exposed} .
- 2) All reported sequence windows with a phosphosite $pPSE > 5$. The number of resulting sequence windows is defined as $N_{\text{not-exposed}}$. Usually, $N_{\text{exposed}} > N_{\text{not-exposed}}$.
- 3) A randomly selected subset of size $N_{\text{not-exposed}}$ of the reported sequence windows with a phosphosite $pPSE \leq 5$.

Motif analysis was subsequently performed using the PSSMSearch tool for each of the three sets of sequence windows by selecting 'log odds' as scoring method (Krystkowiak et al., 2018; O'Shea et al., 2013).

Gene ontology enrichment analysis with DAVID

All gene ontology enrichment analyses presented herein were performed on the DAVID platform version 6.8 (Huang et al., 2009a, 2009b).

For the enrichment analysis of proteins with short IDRs, we compared the 2454 proteins with short IDRs against the background of all 20053 human proteins with available structural information. Direct Gene Ontology terms of the 'Molecular function' category were considered. Filtering was set to an FDR threshold of 0.01, a minimum fold enrichment of 2 and a minimum count of 10.

PTM import from PhosphoSitePlus

Lists of PTMs were downloaded from PhosphoSitePlus on 01.08.2021 (Hornbeck et al., 2015). PTM types include phosphorylations (p), ubiquitinations (ub), sumoylations (sm), acetylations (ac), methylations (m) and the glycosylations O-GalNAc (gl) and O-GlcNAc (ga). Additionally, the set of regulatory PTMs was downloaded and filtered for the modifications mentioned above. For each PTM type, sites were filtered for a selected set of the most common acceptor residues as follows:

- p: S, T, Y
- ub, sm and ac: K
- m: K, R (Note: all types of methylations were grouped together)
- ga, gl: S, T

In rare cases where regulatory sites were reported for a specific PTM type, but no matching entry was found in the according dedicated PTM list, these sites were added there.

The data processing of PhosphoSitePlus data will be available as a Jupyter notebook.

Global PTM proximity analysis

In our global PTM proximity analysis we investigated whether PTM acceptors near a modified amino acid residue are more frequently observed to also be modified compared to more distant residues or random expectation, a concept previously introduced by Beltrao et al. (Beltrao et al., 2012). Here we extended the concept to evaluate PTM proximity in 3D space and to assess both individual PTM types and PTM co-occurrence. Starting from a set of observed modifications, e.g. phosphorylations, we calculate the 3D distance to all observed modifications of either the same type, e.g. also phosphorylations, or of a different type, e.g. ubiquitinations. Importantly, we consider the positional uncertainty of AlphaFold predictions by adding the PAE to each distance. We also generate

5 random permutations, where the same number of modifications are distributed randomly across all possible acceptor residues. Finally, we select distance bins and count the number of modified residues in each of these bins for both the real observations and the randomized background. To ensure that unstructured regions do not bias the results, only structured regions and short IDRs were considered in this analysis (i.e. we removed all IDRs stretching more than 20 amino acids).

For PTM self-proximity we started distance bins at 1 Å ranging up to a maximum of 35 Å in step sizes of 5 Å. For PTM co-localization analysis, we started distance bins already at 0 Å to enable the evaluation of competition for the same site, especially between different lysine modifications such as ubiquitination, sumoylation and acetylation.

Per protein PTM cluster analysis

To find proteins with a significant co-localization of PTMs, we calculated all pairwise distances between the alpha carbons of modified acceptor residues. The matching PAE provided by AlphaFold was added to each calculated distance to account for positional uncertainties. The average distance was subsequently compared to the average distances of 10,000 random permutations among the modified acceptor residues, thus resulting in empirical p-values. These p-values were adjusted for multiple testing (Benjamini and Hochberg, 1995) using statsmodels.stats.multitest.multipletests version 0.13.0 (Seabold and Perktold, 2010).

Protein structure visualization with AlphaMap

The 3D structure visualization in AlphaMap was implemented by integrating a Mol* Viewer (Sehnal et al., 2021). Code was adapted from <https://github.com/molstar/pdbe-molstar>.

References

Abbasov ME, Kavanagh ME, Ichu TA, Lazear MR, Tao Y, Crowley VM, am Ende CW, Hacker SM, Ho J, Dix MM, Suciu R, Hayward MM, Kiessling LL, Cravatt BF. A proteome-wide atlas of lysine-reactive chemistry. *Nat Chem* 2021 1311 2021;13:1081–92. <https://doi.org/10.1038/s41557-021-00765-4>.

Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature* 2016;537:347.

Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198.

Akdel M, Pires DE V, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, Shenoy A, Zhu W, Kundrotas P, Serra VR, Rodrigues CHM, Dunham AS, Burke D, Borkakoti N, Velankar S, Frost A, Lindorff-Larsen K, Valencia A, Ovchinnikov S, Durairaj J, Ascher DB, Thornton JM, Davey NE, Stein A, Elofsson A, Croll TI, Beltrao P. A structural biology community assessment of AlphaFold 2 applications. *BioRxiv* 2021:2021.09.26.461876.
<https://doi.org/10.1101/2021.09.26.461876>.

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Dustin Schaeffer R, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues A V., Van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Christopher Garcia K, Grishin N V., Adams PD, Read RJ, Baker D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (80-) 2021;373:871–6.
https://doi.org/10.1126/SCIENCE.ABJ8754/SUPPL_FILE/ABJ8754_MДАР_REPRODUCIBILITY_CHECKLIST.PDF.

Bekker-Jensen DB, Bernhardt OM, Hoglebe A, Martinez-Val A, Verbeke L, Gandhi T, Kelstrup CD, Reiter L, Olsen J V. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat Commun* 2020;11.
<https://doi.org/10.1038/S41467-020-14609-1>.

Beltrao P, Albanèse V, Kenner LR, Swaney DL, Burlingame A, Villén J, Lim WA, Fraser JS, Frydman J, Krogan NJ. Systematic functional prioritization of protein posttranslational modifications. *Cell* 2012;150:413–25. <https://doi.org/10.1016/J.CELL.2012.05.036/ATTACHMENT/C23EB872-F0EF-436E-8146-15572DBC02C8/MMC1.PDF>.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 1995;57:289–300.

Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow G V., Christie CH, Dalenberg K, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranovic V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao YP, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of

biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res* 2021;49:D437–51. <https://doi.org/10.1093/NAR/GKAA1038>.

Carroll EC, Greene ER, Martin A, Marqusee S. Site-specific ubiquitination affects protein energetics and proteasomal degradation. *Nat Chem Biol* 2020 168 2020;16:866–75. <https://doi.org/10.1038/s41589-020-0556-3>.

Cohen P. The regulation of protein function by multisite phosphorylation – a 25 year update. *Trends Biochem Sci* 2000;25:596–601. [https://doi.org/10.1016/S0968-0004\(00\)01712-6](https://doi.org/10.1016/S0968-0004(00)01712-6).

Evans R, O’neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. Protein complex prediction with AlphaFold-Multimer 2021. <https://doi.org/10.1101/2021.10.04.463034>.

Flatt JF, Stevens-Hernandez CJ, Cogan NM, Eggleston DJ, Haines NM, Heesom KJ, Picard V, Thomas C, Bruce LJ. Expression of South East Asian Ovalocytic Band 3 Disrupts Erythroblast Cytokinesis and Reticulocyte Maturation. *Front Physiol* 2020;11:357. <https://doi.org/10.3389/FPHYS.2020.00357/BIBTEX>.

Hamelryck T. An amino acid has two sides: A new 2D measure provides a different view of solvent exposure. *Proteins Struct Funct Bioinforma* 2005;59:38–48. <https://doi.org/10.1002/PROT.20379>.

Hamelryck T, Manderick B. PDB file parser and structure class implemented in Python. *Bioinforma Appl NOTE* 2003;19:2308–10. <https://doi.org/10.1093/bioinformatics/btg299>.

Hansen FM, Tanzer MC, Brüning F, Bludau I, Stafford C, Schulman BA, Robles MS, Karayel O, Mann M. Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat Commun* 2021;12. <https://doi.org/10.1038/s41467-020-20509-1>.

Heffernan R, Dehzangi A, Lyons J, Paliwal K, Sharma A, Wang J, Sattar A, Zhou Y, Yang Y. Highly accurate sequence-based prediction of half-sphere exposures of amino acid residues in proteins. *Bioinformatics* 2016;32:843–9. <https://doi.org/10.1093/BIOINFORMATICS/BTV665>.

Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014:

Mutations, PTMs and recalibrations. *Nucleic Acids Res* 2015;43:D512–20.

<https://doi.org/10.1093/nar/gku1267>.

Hu R, Xu H, Jia P, Zhao Z. KinaseMD: kinase mutations and drug response database. *Nucleic Acids Res* 2021;49:D552–61. <https://doi.org/10.1093/NAR/GKAA945>.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009a;4:44–57.

<https://doi.org/10.1038/nprot.2008.211>.

Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009b;37:1–13.

<https://doi.org/10.1093/nar/gkn923>.

Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde S, Stancheva VG, Li X-H, Liu K, Zheng Z, Barrero DJ, Roy U, Fernández IS, Szakal B, Branzei D, Greene EC, Biggins S, Keeney S, Miller EA, Fromme JC, Hendrickson TL, Cong Q, Baker D. Structures of core eukaryotic protein complexes. *BioRxiv* 2021:2021.09.30.462231.

<https://doi.org/10.1101/2021.09.30.462231>.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 2004;32:1037–49.

<https://doi.org/10.1093/NAR/GKH253>.

Jørgensen S, Eskildsen M, Fugger K, Hansen L, Larsen MSY, Kousholt AN, Syljuåsen RG, Trelle MB, Jensen ON, Helin K, Sørensen CS. SET8 is degraded via PCNA-coupled CRL4(CDT2) ubiquitylation in S phase and after UV irradiation. *J Cell Biol* 2011;192:43–54. <https://doi.org/10.1083/JCB.201009076>.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nat* 2021 5967873 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>.

Kamacioğlu A, Tuncbag N, Ozlu N. Structural analysis of mammalian protein phosphorylation at a proteome level. *Structure* 2021;29:1219-1229.e3. <https://doi.org/10.1016/J.STR.2021.06.008>.

Kato M, Wynn RM, Chuang JL, Tso SC, Machius M, Li J, Chuang DT. Structural Basis for Inactivation of the Human Pyruvate Dehydrogenase Complex by Phosphorylation: Role of Disordered Phosphorylation Loops. *Structure* 2008;16:1849–59.

<https://doi.org/10.1016/J.STR.2008.10.010/ATTACHMENT/6F54FB3C-2DD6-47C2-B4F3-BD4B2928EAA3/MMC1.PDF>.

Krystkowiak I, Manguy J, Davey NE. PSSMSearch: a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic Acids Res* 2018;46:W235–41. <https://doi.org/10.1093/NAR/GKY426>.

Li S, Zhang Q, You G. Three Ubiquitination Sites of Organic Anion Transporter-1 Synergistically Mediate Protein Kinase C–Dependent Endocytosis of the Transporter. *Mol Pharmacol* 2013;84:139–46. <https://doi.org/10.1124/MOL.113.086769>.

Matsuki H, Takahashi M, Higuchi M, Makokha GN, Oie M, Fujii M. Both G3BP1 and G3BP2 contribute to stress granule formation. *Genes Cells* 2013;18:135–46. <https://doi.org/10.1111/GTC.12023>.

Maxwell BA, Gwon Y, Mishra A, Peng J, Nakamura H, Zhang K, Kim HJ, Taylor JP. Ubiquitination is essential for recovery of cellular activities after heat shock. *Science* 2021;372. <https://doi.org/10.1126/SCIENCE.ABC3593>.

Mészáros B, Erdős G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* 2018;46:W329–37. <https://doi.org/10.1093/NAR/GKY384>.

Nolen B, Taylor S, Ghosh G. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. *Mol Cell* 2004;15:661–75. <https://doi.org/10.1016/J.MOLCEL.2004.08.024>.

O’Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods* 2013 1012 2013;10:1211–2. <https://doi.org/10.1038/nmeth.2646>.

Ochoa D, Jarnuczak AF, Viéitez C, Gehre M, Soucheray M, Mateus A, Kleefeldt AA, Hill A, Garcia-Alonso L, Stein F, Krogan NJ, Savitski MM, Swaney DL, Vizcaíno JA, Noh KM, Beltrao P. The functional landscape of the human phosphoproteome. *Nat Biotechnol* 2019 383 2019;38:365–73. <https://doi.org/10.1038/s41587-019-0344-3>.

Psakhye I, Jentsch S. Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* 2012;151:807–20. <https://doi.org/10.1016/J.CELL.2012.10.021>.

Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. PROC 9th PYTHON Sci CONF 2010.

Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* 2021;49:W431–7. <https://doi.org/10.1093/NAR/GKAB314>.

Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski JR, Cox J, Mann M. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* 2014;8:1583–94. <https://doi.org/10.1016/J.CELREP.2014.07.036>.

Stukalov A, Girault V, Grass V, Karayel O, Bergant V, Urban C, Haas DA, Huang Y, Oubraham L, Wang A, Hamad MS, Piras A, Hansen FM, Tanzer MC, Paron I, Zinzula L, Engleitner T, Reinecke M, Lavacca TM, Ehmann R, Wölfel R, Jores J, Kuster B, Protzer U, Rad R, Ziebuhr J, Thiel V, Scaturro P, Mann M, Pichlmair A. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* 2021;594:246–52. <https://doi.org/10.1038/S41586-021-03493-4>.

Sugiyama N, Imamura H, Ishihama Y. Large-scale Discovery of Substrates of the Human Kinome. *Sci Reports* 2019 91 2019;9:1–12. <https://doi.org/10.1038/s41598-019-46385-4>.

Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the human proteome. *Nat* 2021 5967873 2021;596:590–6. <https://doi.org/10.1038/s41586-021-03828-1>.

Tyanova S, Cox J, Olsen J, Mann M, Frishman D. Phosphorylation Variation during the Cell Cycle Scales with Structural Propensities of Proteins. *PLOS Comput Biol* 2013;9:e1002842. <https://doi.org/10.1371/JOURNAL.PCBI.1002842>.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016;13:731–40. <https://doi.org/10.1038/NMETH.3901>.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2021. <https://doi.org/10.1093/NAR/GKAB1061>.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Vázquez-Baeza Y, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020 17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.

Voytik E, Bludau I, Willems S, Hansen FM, Brunner A-D, Strauss MT, Mann M. AlphaMap: an open-source Python package for the visual annotation of proteomics data with sequence-specific knowledge. *Bioinformatics* 2021. <https://doi.org/10.1093/BIOINFORMATICS/BTAB674>.

Zannini L, Delia D, Buscemi G. CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol* 2014;6:442. <https://doi.org/10.1093/JMCB/MJU045>.

4 Discussion and Outlook

Mass spectrometry is the gold standard for the unbiased and system-wide interrogation of whole proteomes. Ongoing developments and improvements of sample preparation procedures, instrumentation and analysis software enable the analysis of proteomes at increased depth and speed. Here, I generally use “proteome” as an umbrella term describing groups of canonical proteins rather than the complex composition of their many proteoforms. However, state-of-the-art mass spectrometry has the power to report individual proteoforms, dramatically expanding the informational content of MS-based proteomics [199]. An increasing awareness of the biological implications evoked by different proteoforms of a protein has led to a growing interest in the analysis of proteomes resolved on proteoform level.

In this thesis, I developed and applied various methods and strategies to identify proteoforms characterized by PTMs on the global and individual protein levels. My major focus was the establishment of a robust DIA strategy for ubiquitinome analysis. The development of a sophisticated DIA-based workflow enabled an unprecedented depth of ubiquitinome analysis allowing the identification of 35,000 diGly peptides in a single measurement. We found that compared to standard label-free DDA, DIA almost doubles the number of identified diGly peptides, while improving quantitative accuracy and data completeness. Robust identification and quantification also allowed the investigation of ubiquitination throughout the circadian rhythm unveiling hundreds of cycling ubiquitination sites. DIA for ubiquitinome analysis clearly is a powerful alternative to DDA approaches and the workflow presented here can readily be adapted for other antibody-based enrichment procedures. Further improvements of various aspects of the workflow hold great promise to enable an even more powerful pipeline for high-throughput automated, high sensitivity ubiquitinome analysis. For instance, development of alternatives to the classically used diGly remnant-specific antibody such as nanobodies or aptamers could increase enrichment efficiencies. Recently, the diGly-specific antibody has also become commercially available on magnetic beads, which promises lower peptide background binding and furthermore facilitates workflow automation on robotic platforms. In fact, the DDA-based UbiFast protocol uses these magnetic beads and demonstrates automatic diGly peptide enrichment in combination with TMT labeling [200]. However, that protocol relies on cleaned up peptides and thus still requires manual sample processing and peptide cleanup steps upfront. The removal of these manual steps would make the workflow more robust, streamlined and would reduce sample loss. Sample multiplexing through chemical labeling strategies such as TMT

could reduce measurement time, thus increasing sample throughput. DIA, however, is not compatible with labeling strategies that rely on reporter ions in the low m/z mass range. Here, the EASI-tag labeling strategy is an appealing alternative to the aforementioned chemical labeling approach, and would also allow the multiplexing of samples with our DIA workflow for ubiquitinome analysis [94]. Furthermore, the LC-MS setup can also be improved. Novel LC platforms, such as the Evosep One reduce overhead time between individual measurements and work with pre-formed and highly reproducible gradients [201]. The combination of the Evosep system with our in-house packed columns produces high quality chromatographic peak separation. However, the bead bed between packed columns is not entirely consistent and may also change during extended column use, leading to somewhat variable peak separation and a reduced reproducibility. The use of μ Pac columns with their standardized column architecture is an interesting alternative to reduce chromatographic variance between measurements [42]. In addition, the integration of ion mobility (IM) for peptide separation enhances peptide resolving power. Especially diGly peptides, which are on average longer and have a higher charge state, are prone to show a different behavior in the IM space than unmodified tryptic peptides. We successfully deployed IM separation with the FAIMS device to construct extensive spectral libraries for DIA measurements. In my opinion, IM separation on TimsTOF instruments operated in the DIA-PASEF mode [54] are highly interesting for diGly peptide measurements. In contrast to the FAIMS device, the ion loss during measurements can be reduced to a minimum by the parallel accumulation on the PASEF mode. Furthermore, due to the distinct features of diGly peptides in the IM space these peptides can preferentially be targeted for data acquisition, reducing the sampling of unmodified peptides. DIA approaches previously used to require extensive spectral libraries for high performance, which is often not feasible when biological sample amounts are limited. Recent advances in analysis software, however, greatly improved library free DIA analysis [79], and in my recent results library-based and library-free analysis for ubiquitinome analysis perform almost at par. In view of these extensive possibilities for further workflow improvements, the study presented in this thesis only marks the beginning of further exciting developments for DIA-based ubiquitinome analysis.

To understand the implications of protein modification, it is often essential to identify and quantify the specific site of modification. Although this can be achieved through system-wide approaches, targeted MS assays or MS in combination with *in vitro* modification assays often provide more information. The design and establishment of such experimental approaches requires optimization, because peptides greatly vary in their chromatographic and ionization behavior and detectability. More robust chromatography

as well as peptide 'flyability' prediction would greatly facilitate the establishment of such methods [202, 203]. In contrast to the mere identification of modification sites that are provided by standard MS analysis tools, further tools highlighting the location of modifications on the linear or 3D protein structure would be helpful to interpret observed modification sites in regard to the protein structure. To this end, the StructureMap tool from our lab, which uses predicted protein structures of AlphaFold, greatly helps to obtain such structural context. Future integration of information on PTMs for the prediction of the structure of modified proteins will definitely help to understand the impact of PTMs on protein regulation.

Post-translational modification of proteins not only allows the regulation of individual protein function, but also the fine tuning of cellular or organellar signaling networks. To help to understand how mitochondria regulation is orchestrated throughout different cellular environments, we assessed the mitochondria composition of 7 different mouse tissues on proteome and phosphoproteome level. This resource portrays known mitochondrial differences as well as novel unknown regulations and broadens our knowledge of differential mitochondria phosphorylation, although the functional relevance of individual phosphorylation sites are yet to be determined. Such studies on mitochondrial proteomes and phosphoproteomes rely on efficient enrichment of mitochondria, for which density centrifugation and MitoTags have successfully been deployed [181]. We found that the fraction of non-mitochondrial proteins after enrichment strongly depends on the tissue of origin and it will be interesting to see if these different proportions are enrichment artefacts or based on more diverse mitochondrial interactions with their cellular environment. Of note, the reported mitochondrial differences are based on tissue pieces without further discrimination of different cell types contained within these samples. Future studies investigating distinct cell types within a tissue for their mitochondrial phosphoproteome will shed new light on the differential regulation of mitochondria through protein phosphorylation.

Post-translational modifications are crucial for the regulation of cellular homeostasis and their dysregulation can cause a multitude of diseases. Intervening in dysregulated PTMs has already proven to be an essential strategy for the treatment of cancers. Kinase inhibitors such as Imatinib and Gefitinib are first-line treatments for Chronic myelogenous leukemia (CML) and non-small cell lung carcinoma (NSCLC), respectively [204]. Similarly, compounds such as PROTACs or molecular glues that use E3 ligases for the degradation of protein targets are being investigated in clinical trials for the treatment of malignancies [205]. To understand the full effect of drugs targeting the regulation of PTMs, it is essential to investigate the corresponding effects on the PTM landscape on a large as well as small scale to fully appreciate the biological implications

of these drugs. This could prevent the occurrence of adverse effects, and may also unveil opportunities for the treatment of further malignancies. Moreover, PTM sites might be used as diagnostic biomarkers for the early identification of disorders – as was shown by Özge Karayel in our group for phosphorylated Rab10 in the context of stratification and treatment of Parkinson's disease carrier [206]. In summary, MS-based proteomics is at the heart of such investigations and I hope that the work provided in this thesis will help to deepen our understanding of post-translational modifications in health and disease.

5 References

1. Johannsen, W., *Elemente der exakten Erblchkeitslehre*. Zeitschrift für Induktive Abstammungs- und Vererbungslehre, 1909. **2**(1): p. 136-137.
2. Beadle, G.W. and E.L. Tatum, *Genetic Control of Biochemical Reactions in Neurospora*. Proc Natl Acad Sci U S A, 1941. **27**(11): p. 499-506.
3. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii*. J Exp Med, 1944. **79**(2): p. 137-58.
4. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
5. Chow, L.T., et al., *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA*. Cell, 1977. **12**(1): p. 1-8.
6. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome*. Nature, 2004. **431**(7011): p. 931-45.
7. Pennisi, E., *Human Genome Project. And the gene number is...?* Science, 2000. **288**(5469): p. 1146-7.
8. Frankish, A., et al., *Gencode 2021*. Nucleic Acids Res, 2021. **49**(D1): p. D916-D923.
9. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control*. Nature, 2011. **473**(7347): p. 337-42.
10. Li, J.J., P.J. Bickel, and M.D. Biggin, *System wide analyses have underestimated protein abundances and the importance of transcription in mammals*. PeerJ, 2014. **2**: p. e270.
11. Jovanovic, M., et al., *Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens*. Science, 2015. **347**(6226): p. 1259038.
12. Beyer, A., et al., *Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale*. Mol Cell Proteomics, 2004. **3**(11): p. 1083-92.
13. Pessione, E., *The Russian Doll Model: How Bacteria Shape Successful and Sustainable Inter-Kingdom Relationships*. Front Microbiol, 2020. **11**: p. 573759.
14. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity*. Nat Methods, 2013. **10**(3): p. 186-7.
15. Ribet, D. and P. Cossart, *Post-translational modifications in host cells during bacterial infection*. FEBS Lett, 2010. **584**(13): p. 2748-58.

16. Yau, R. and M. Rape, *The increasing complexity of the ubiquitin code*. Nat Cell Biol, 2016. **18**(6): p. 579-86.
17. Aebersold, R., et al., *How many human proteoforms are there?* Nat Chem Biol, 2018. **14**(3): p. 206-214.
18. Wien, W., *Untersuchungen über die elektrische Entladung in verdünnten Gasen*. Annalen der Physik, 1898. **301**(6): p. 440-452.
19. Thomson, J.J., *XL. Cathode Rays*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1897. **44**(269): p. 293-316.
20. Karas, M., et al., *Matrix-assisted ultraviolet laser desorption of non-volatile compounds*. International Journal of Mass Spectrometry and Ion Processes, 1987. **78**: p. 53-68.
21. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules*. Science, 1989. **246**(4926): p. 64-71.
22. Edman, P., *A method for the determination of amino acid sequence in peptides*. Arch Biochem, 1949. **22**(3): p. 475.
23. Mann, M., *The Rise of Mass Spectrometry and the Fall of Edman Degradation*. Clin Chem, 2016. **62**(1): p. 293-4.
24. Fornelli, L., et al., *Top-down proteomics: Where we are, where we are going?* J Proteomics, 2018. **175**: p. 3-4.
25. Meissner, F. and M. Mann, *Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology*. Nat Immunol, 2014. **15**(2): p. 112-7.
26. Islam, M.S., A. Aryasomayajula, and P.R. Selvaganapathy, *A Review on Macroscale and Microscale Cell Lysis Methods*. Micromachines, 2017. **8**(3).
27. Vandermarliere, E., M. Mueller, and L. Martens, *Getting intimate with trypsin, the leading protease in proteomics*. Mass Spectrom Rev, 2013. **32**(6): p. 453-65.
28. Swaney, D.L., C.D. Wenger, and J.J. Coon, *Value of using multiple proteases for large-scale mass spectrometry-based proteomics*. J Proteome Res, 2010. **9**(3): p. 1323-9.
29. Ye, X.T., et al., *Integrated proteomics sample preparation and fractionation: Method development and applications*. Trac-Trends in Analytical Chemistry, 2019. **120**.
30. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels*. Anal Chem, 1996. **68**(5): p. 850-8.
31. Rabilloud, T., et al., *Power and limitations of electrophoretic separations in proteomics strategies*. Mass Spectrom Rev, 2009. **28**(5): p. 816-43.
32. Rappsilber, J., Y. Ishihama, and M. Mann, *Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics*. Anal Chem, 2003. **75**(3): p. 663-70.

33. Leon, I.R., et al., *Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis*. Mol Cell Proteomics, 2013. **12**(10): p. 2992-3005.
34. Kulak, N.A., et al., *Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells*. Nat Methods, 2014. **11**(3): p. 319-24.
35. Kulak, N.A., P.E. Geyer, and M. Mann, *Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics*. Mol Cell Proteomics, 2017. **16**(4): p. 694-705.
36. Meyer, J.G., et al., *Quantitative shotgun proteome analysis by direct infusion*. Nat Methods, 2020. **17**(12): p. 1222-1228.
37. Chekmeneva, E., et al., *Optimization and Application of Direct Infusion Nanoelectrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology*. J Proteome Res, 2017. **16**(4): p. 1646-1658.
38. Gachumi, G., et al., *Fast Quantification Without Conventional Chromatography, The Growing Power of Mass Spectrometry*. Anal Chem, 2020. **92**(13): p. 8628-8637.
39. Nanita, S.C. and L.G. Kaldon, *Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis*. Anal Bioanal Chem, 2016. **408**(1): p. 23-33.
40. VanDeemter, J.J., F.J. Zuiderweg, and A. Klinkenberg, *Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography (Reprinted from Chem Engng Sci, vol 5, pg 271-289, 1956)*. Chemical Engineering Science, 1995. **50**(24): p. 3869-3882.
41. Muller-Reif, J.B., et al., *A New Parallel High-Pressure Packing System Enables Rapid Multiplexed Production of Capillary Columns*. Mol Cell Proteomics, 2021. **20**: p. 100082.
42. De Beeck, J.O., et al., *A well-ordered nanoflow LC-MS/MS approach for proteome profiling using 200 cm long micro pillar array columns*. bioRxiv, 2019: p. 472134.
43. Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte*. Int J Anal Chem, 2012. **2012**: p. 282574.
44. Eyers, C., S.J. Gaskell, and C. Royal Society of, *Quantitative proteomics. New Developments in Mass Spectrometry 1*. 2014, Cambridge, England: The Royal Society of Chemistry.
45. Dole, M., L.L. Mack, and R.L. Hines, *Molecular Beams of Macroions*. Journal of Chemical Physics, 1968. **49**(5): p. 2240-&.
46. Iribarne, J.V. and B.A. Thomson, *On the evaporation of small ions from charged droplets*. Journal of Chemical Physics, 1976. **64**(6): p. 2287-2294.

47. Pol, J., et al., *Molecular mass spectrometry imaging in biomedical and life science research*. Histochem Cell Biol, 2010. **134**(5): p. 423-43.
48. Beck, S., et al., *The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics*. Mol Cell Proteomics, 2015. **14**(7): p. 2014-29.
49. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. Anal Chem, 2000. **72**(6): p. 1156-62.
50. Haag, A.M., *Mass Analyzers and Mass Spectrometers*. Adv Exp Med Biol, 2016. **919**: p. 157-169.
51. Hebert, A.S., et al., *Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer*. Anal Chem, 2018. **90**(15): p. 9529-9537.
52. Meier, F., et al., *Parallel Accumulation-Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device*. J Proteome Res, 2015. **14**(12): p. 5378-87.
53. Meier, F., et al., *Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer*. Mol Cell Proteomics, 2018. **17**(12): p. 2534-2545.
54. Meier, F., et al., *diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition*. Nat Methods, 2020. **17**(12): p. 1229-1236.
55. Ichou, F., et al., *Comparison of the activation time effects and the internal energy distributions for the CID, PQD and HCD excitation modes*. J Mass Spectrom, 2014. **49**(6): p. 498-508.
56. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis*. Nat Methods, 2007. **4**(9): p. 709-12.
57. Reilly, J.P., *Ultraviolet photofragmentation of biomolecular ions*. Mass Spectrom Rev, 2009. **28**(3): p. 425-47.
58. Antoine, R., J. Lemoine, and P. Dugourd, *Electron photodetachment dissociation for structural characterization of synthetic and bio-polymer anions*. Mass Spectrom Rev, 2014. **33**(6): p. 501-22.
59. Brodbelt, J.S., *Photodissociation mass spectrometry: new tools for characterization of biological molecules*. Chem Soc Rev, 2014. **43**(8): p. 2757-83.
60. Halim, M.A., et al., *Combined Infrared Multiphoton Dissociation with Ultraviolet Photodissociation for Ubiquitin Characterization*. J Am Soc Mass Spectrom, 2016. **27**(9): p. 1435-42.
61. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.

62. Dupree, E.J., et al., *A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field*. Proteomes, 2020. **8**(3).
63. Borrás, E. and E. Sabido, *What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry*. Proteomics, 2017. **17**(17-18).
64. Masselon, C., et al., *Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures*. Anal Chem, 2000. **72**(8): p. 1918-24.
65. Bateman, R.H., et al., *A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation*. Journal of the American Society for Mass Spectrometry, 2002. **13**(7): p. 792-803.
66. Purvine, S., et al., *Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer*. Proteomics, 2003. **3**(6): p. 847-50.
67. Venable, J.D., et al., *Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra*. Nat Methods, 2004. **1**(1): p. 39-45.
68. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis*. Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.
69. Tsou, C.C., et al., *DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics*. Nat Methods, 2015. **12**(3): p. 258-64, 7 p following 264.
70. Pappireddi, N., L. Martin, and M. Wuhr, *A Review on Quantitative Multiplexed Proteomics*. Chembiochem, 2019. **20**(10): p. 1210-1224.
71. Gatlin, C.L., et al., *Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry*. Anal Biochem, 1998. **263**(1): p. 93-101.
72. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics*. Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
73. Thomas, A., et al., *Expanded test method for peptides >2 kDa employing immunoaffinity purification and LC-HRMS/MS*. Drug Test Anal, 2015. **7**(11-12): p. 990-8.
74. Picotti, P. and R. Aebersold, *Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions*. Nat Methods, 2012. **9**(6): p. 555-66.
75. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nat Biotechnol, 2008. **26**(12): p. 1367-72.

76. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics*. Nat Methods, 2017. **14**(5): p. 513-520.
77. Egertson, J.D., et al., *Multiplexed peptide analysis using data-independent acquisition and Skyline*. Nat Protoc, 2015. **10**(6): p. 887-903.
78. Bruderer, R., et al., *Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues*. Mol Cell Proteomics, 2015. **14**(5): p. 1400-10.
79. Demichev, V., et al., *DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput*. Nat Methods, 2020. **17**(1): p. 41-44.
80. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment*. J Proteome Res, 2011. **10**(4): p. 1794-805.
81. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry*. Nat Methods, 2007. **4**(3): p. 207-14.
82. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial*. Mol Syst Biol, 2018. **14**(8): p. e8126.
83. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of shotgun proteomic data: the protein inference problem*. Mol Cell Proteomics, 2005. **4**(10): p. 1419-40.
84. Rappsilber, J., et al., *Large-scale proteomic analysis of the human spliceosome*. Genome Res, 2002. **12**(8): p. 1231-45.
85. Bondarenko, P.V., D. Chelius, and T.A. Shaler, *Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry*. Anal Chem, 2002. **74**(18): p. 4741-9.
86. Geiger, T., et al., *Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins*. Mol Cell Proteomics, 2012. **11**(3): p. M111 014050.
87. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ*. Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.
88. Wisniewski, J.R., et al., *A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards*. Mol Cell Proteomics, 2014. **13**(12): p. 3497-506.
89. Rauniyar, N. and J.R. Yates, 3rd, *Isobaric labeling-based relative quantification in shotgun proteomics*. J Proteome Res, 2014. **13**(12): p. 5293-309.
90. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS*. Anal Chem, 2003. **75**(8): p. 1895-904.

91. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics*. Nat Methods, 2011. **8**(11): p. 937-40.
92. O'Connell, J.D., et al., *Proteome-Wide Evaluation of Two Common Protein Quantification Methods*. J Proteome Res, 2018. **17**(5): p. 1934-1942.
93. Myers, S.A., et al., *Evaluation of Advanced Precursor Determination for Tandem Mass Tag (TMT)-Based Quantitative Proteomics across Instrument Platforms*. J Proteome Res, 2019. **18**(1): p. 542-547.
94. Virreira Winter, S., et al., *EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification*. Nat Methods, 2018. **15**(7): p. 527-530.
95. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics*. Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
96. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue*. Nat Methods, 2010. **7**(5): p. 383-5.
97. Brun, V., et al., *Isotope-labeled protein standards: toward absolute quantitative proteomics*. Mol Cell Proteomics, 2007. **6**(12): p. 2139-49.
98. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
99. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present*. Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
100. Goldstein, G., et al., *Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells*. Proc Natl Acad Sci U S A, 1975. **72**(1): p. 11-5.
101. Sharp, P.M. and W.H. Li, *Ubiquitin genes as a paradigm of concerted evolution of tandem repeats*. J Mol Evol, 1987. **25**(1): p. 58-64.
102. Graham, R.W., D. Jones, and E.P. Candido, *UbiA, the major polyubiquitin locus in Caenorhabditis elegans, has unusual structural features and is constitutively expressed*. Mol Cell Biol, 1989. **9**(1): p. 268-77.
103. Catic, A. and H.L. Ploegh, *Ubiquitin--conserved protein or selfish gene?* Trends Biochem Sci, 2005. **30**(11): p. 600-4.
104. Goldknopf, I.L. and H. Busch, *Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24*. Proc Natl Acad Sci U S A, 1977. **74**(3): p. 864-8.
105. Ciehanover, A., Y. Hod, and A. Hershko, *A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes*. Biochem Biophys Res Commun, 1978. **81**(4): p. 1100-5.

106. Hershko, A., et al., *Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis*. Proc Natl Acad Sci U S A, 1980. **77**(4): p. 1783-6.
107. Hershko, A., A. Ciechanover, and I.A. Rose, *Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown*. J Biol Chem, 1981. **256**(4): p. 1525-8.
108. Ciechanover, A., et al., *"Covalent affinity" purification of ubiquitin-activating enzyme*. J Biol Chem, 1982. **257**(5): p. 2537-42.
109. Hershko, A., et al., *Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown*. J Biol Chem, 1983. **258**(13): p. 8206-14.
110. Hough, R., G. Pratt, and M. Rechsteiner, *Purification of two high molecular weight proteases from rabbit reticulocyte lysate*. J Biol Chem, 1987. **262**(17): p. 8303-13.
111. Waxman, L., J.M. Fagan, and A.L. Goldberg, *Demonstration of two distinct high molecular weight proteases in rabbit reticulocytes, one of which degrades ubiquitin conjugates*. J Biol Chem, 1987. **262**(6): p. 2451-7.
112. Glickman, M.H. and A. Ciechanover, *The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction*. Physiol Rev, 2002. **82**(2): p. 373-428.
113. Kliza, K. and K. Husnjak, *Resolving the Complexity of Ubiquitin Networks*. Front Mol Biosci, 2020. **7**: p. 21.
114. Zheng, N. and N. Shabek, *Ubiquitin Ligases: Structure, Function, and Regulation*. Annu Rev Biochem, 2017. **86**: p. 129-157.
115. Cadwell, K. and L. Coscoy, *Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase*. Science, 2005. **309**(5731): p. 127-30.
116. Williams, C., et al., *A conserved cysteine is essential for Pex4p-dependent ubiquitination of the peroxisomal import receptor Pex5p*. J Biol Chem, 2007. **282**(31): p. 22534-43.
117. Ciechanover, A. and R. Ben-Saadon, *N-terminal ubiquitination: more protein substrates join in*. Trends Cell Biol, 2004. **14**(3): p. 103-6.
118. McDowell, G.S. and A. Philpott, *New Insights Into the Role of Ubiquitylation of Proteins*. Int Rev Cell Mol Biol, 2016. **325**: p. 35-88.
119. Kalayil, S., et al., *Insights into catalysis and function of phosphoribosyl-linked serine ubiquitination*. Nature, 2018. **557**(7707): p. 734-738.
120. Zhang, M., et al., *Members of the Legionella pneumophila Sde family target tyrosine residues for phosphoribosyl-linked ubiquitination*. RSC Chem Biol, 2021. **2**(5): p. 1509-1519.
121. Komander, D., M.J. Clague, and S. Urbe, *Breaking the chains: structure and function of the deubiquitinases*. Nat Rev Mol Cell Biol, 2009. **10**(8): p. 550-63.

122. Reyes-Turcu, F.E., K.H. Ventii, and K.D. Wilkinson, *Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes*. *Annu Rev Biochem*, 2009. **78**: p. 363-97.
123. Swatek, K.N. and D. Komander, *Ubiquitin modifications*. *Cell Res*, 2016. **26**(4): p. 399-422.
124. Finley, D., et al., *Inhibition of proteolysis and cell cycle progression in a multiubiquitination-deficient yeast mutant*. *Mol Cell Biol*, 1994. **14**(8): p. 5501-9.
125. Wauer, T., et al., *Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis*. *EMBO J*, 2015. **34**(3): p. 307-25.
126. Rose, A. and T. Mayor, *Exploring the Rampant Expansion of Ubiquitin Proteomics*. *Methods Mol Biol*, 2018. **1844**: p. 345-362.
127. Peng, J., et al., *A proteomics approach to understanding protein ubiquitination*. *Nat Biotechnol*, 2003. **21**(8): p. 921-6.
128. Mayor, T., et al., *Analysis of polyubiquitin conjugates reveals that the Rpn10 substrate receptor contributes to the turnover of multiple proteasome targets*. *Mol Cell Proteomics*, 2005. **4**(6): p. 741-51.
129. Tagwerker, C., et al., *A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking*. *Mol Cell Proteomics*, 2006. **5**(4): p. 737-48.
130. Mayor, T., et al., *Quantitative profiling of ubiquitylated proteins reveals proteasome substrates and the substrate repertoire influenced by the Rpn10 receptor pathway*. *Mol Cell Proteomics*, 2007. **6**(11): p. 1885-95.
131. Meierhofer, D., et al., *Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry*. *J Proteome Res*, 2008. **7**(10): p. 4566-76.
132. Kirkpatrick, D.S., et al., *Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin*. *Proteomics*, 2005. **5**(8): p. 2104-11.
133. Danielsen, J.M., et al., *Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level*. *Mol Cell Proteomics*, 2011. **10**(3): p. M110 003590.
134. Vasilescu, J., et al., *Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry*. *J Proteome Res*, 2005. **4**(6): p. 2192-200.
135. Tan, F., et al., *Proteomic analysis of ubiquitinated proteins in normal hepatocyte cell line Chang liver cells*. *Proteomics*, 2008. **8**(14): p. 2885-96.
136. Xu, G., J.S. Paige, and S.R. Jaffrey, *Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling*. *Nat Biotechnol*, 2010. **28**(8): p. 868-73.

137. Kim, W., et al., *Systematic and quantitative assessment of the ubiquitin-modified proteome*. Mol Cell, 2011. **44**(2): p. 325-40.
138. Wagner, S.A., et al., *A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles*. Mol Cell Proteomics, 2011. **10**(10): p. M111 013284.
139. Wagner, S.A., et al., *Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues*. Mol Cell Proteomics, 2012. **11**(12): p. 1578-85.
140. Udeshi, N.D., et al., *Large-scale identification of ubiquitination sites by mass spectrometry*. Nat Protoc, 2013. **8**(10): p. 1950-60.
141. Lee, K.A., et al., *Ubiquitin ligase substrate identification through quantitative proteomics at both the protein and peptide levels*. J Biol Chem, 2011. **286**(48): p. 41530-41538.
142. Na, C.H., et al., *Synaptic protein ubiquitination in rat brain revealed by antibody-based ubiquitome analysis*. J Proteome Res, 2012. **11**(9): p. 4722-32.
143. Akimov, V., et al., *UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites*. Nat Struct Mol Biol, 2018. **25**(7): p. 631-640.
144. Hjerpe, R., et al., *Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities*. EMBO Rep, 2009. **10**(11): p. 1250-8.
145. Sims, J.J., et al., *Polyubiquitin-sensor proteins reveal localization and linkage-type dependence of cellular ubiquitin signaling*. Nat Methods, 2012. **9**(3): p. 303-9.
146. Rahighi, S., et al., *Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation*. Cell, 2009. **136**(6): p. 1098-109.
147. Kristariyanto, Y.A., et al., *K29-selective ubiquitin binding domain reveals structural basis of specificity and heterotypic nature of k29 polyubiquitin*. Mol Cell, 2015. **58**(1): p. 83-94.
148. Kristariyanto, Y.A., et al., *A single MIU motif of MINDY-1 recognizes K48-linked polyubiquitin chains*. EMBO Rep, 2017. **18**(3): p. 392-402.
149. Kulathu, Y., et al., *Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain*. Nat Struct Mol Biol, 2009. **16**(12): p. 1328-30.
150. Michel, M.A., et al., *Ubiquitin Linkage-Specific Affimers Reveal Insights into K6-Linked Ubiquitin Signaling*. Mol Cell, 2017. **68**(1): p. 233-246 e5.
151. Yau, R.G., et al., *Assembly and Function of Heterotypic Ubiquitin Chains in Cell-Cycle and Protein Quality Control*. Cell, 2017. **171**(4): p. 918-933 e20.
152. Crowe, S.O., et al., *Ubiquitin Chain Enrichment Middle-Down Mass Spectrometry Enables Characterization of Branched Ubiquitin Chains in Cellulo*. Anal Chem, 2017. **89**(8): p. 4428-4434.
153. Swatek, K.N., et al., *Insights into ubiquitin chain architecture using Ub-clipping*. Nature, 2019. **572**(7770): p. 533-537.

154. Kaiser, S.E., et al., *Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools*. Nat Methods, 2011. **8**(8): p. 691-6.
155. Fischer, E.H. and E.G. Krebs, *Conversion of phosphorylase b to phosphorylase a in muscle extracts*. J Biol Chem, 1955. **216**(1): p. 121-32.
156. Manning, G., et al., *The protein kinase complement of the human genome*. Science, 2002. **298**(5600): p. 1912-34.
157. Chen, M.J., J.E. Dixon, and G. Manning, *Genomics and evolution of protein phosphatases*. Sci Signal, 2017. **10**(474).
158. Hunter, T., *Why nature chose phosphate to modify proteins*. Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1602): p. 2513-6.
159. Humphrey, S.J., D.E. James, and M. Mann, *Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation*. Trends Endocrinol Metab, 2015. **26**(12): p. 676-687.
160. Singh, V., et al., *Phosphorylation: Implications in Cancer*. Protein J, 2017. **36**(1): p. 1-6.
161. Ardito, F., et al., *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review)*. Int J Mol Med, 2017. **40**(2): p. 271-280.
162. Low, T.Y., et al., *Widening the Bottleneck of Phosphoproteomics: Evolving Strategies for Phosphopeptide Enrichment*. Mass Spectrom Rev, 2021. **40**(4): p. 309-333.
163. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins*. Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12130-5.
164. Zhang, K., *From purification of large amounts of phospho-compounds (nucleotides) to enrichment of phospho-peptides using anion-exchanging resin*. Anal Biochem, 2006. **357**(2): p. 225-31.
165. Alpert, A.J., *Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides*. Anal Chem, 2008. **80**(1): p. 62-76.
166. Neville, D.C., et al., *Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry*. Protein Sci, 1997. **6**(11): p. 2436-45.
167. Posewitz, M.C. and P. Tempst, *Immobilized gallium(III) affinity chromatography of phosphopeptides*. Anal Chem, 1999. **71**(14): p. 2883-92.
168. Leitner, A., *Phosphopeptide enrichment using metal oxide affinity chromatography*. Trac-Trends in Analytical Chemistry, 2010. **29**(2): p. 177-185.
169. Humphrey, S.J., et al., *High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform*. Nat Protoc, 2018. **13**(9): p. 1897-1916.
170. Rush, J., et al., *Immunoaffinity profiling of tyrosine phosphorylation in cancer cells*. Nat Biotechnol, 2005. **23**(1): p. 94-101.

171. Nagaraj, N., et al., *Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation*. J Proteome Res, 2010. **9**(12): p. 6786-94.
172. Bekker-Jensen, D.B., et al., *Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries*. Nat Commun, 2020. **11**(1): p. 787.
173. Martin, W.F., S. Garg, and V. Zimorski, *Endosymbiotic theories for eukaryote origin*. Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1678): p. 20140330.
174. Herst, P.M., et al., *Functional Mitochondria in Health and Disease*. Front Endocrinol (Lausanne), 2017. **8**: p. 296.
175. Suomalainen, A. and B.J. Battersby, *Mitochondrial diseases: the contribution of organelle stress responses to pathology*. Nat Rev Mol Cell Biol, 2018. **19**(2): p. 77-92.
176. Mootha, V.K., et al., *Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria*. Cell, 2003. **115**(5): p. 629-40.
177. Forner, F., et al., *Quantitative proteomic comparison of rat mitochondria from muscle, heart, and liver*. Mol Cell Proteomics, 2006. **5**(4): p. 608-19.
178. Pagliarini, D.J., et al., *A mitochondrial protein compendium elucidates complex I disease biology*. Cell, 2008. **134**(1): p. 112-23.
179. Niemi, N.M. and D.J. Pagliarini, *The extensive and functionally uncharacterized mitochondrial phosphoproteome*. J Biol Chem, 2021. **297**(1): p. 100880.
180. Kappler, L., et al., *Purity matters: A workflow for the valid high-resolution lipid profiling of mitochondria from cell culture samples*. Sci Rep, 2016. **6**: p. 21107.
181. Bayraktar, E.C., et al., *MITO-Tag Mice enable rapid isolation and multimodal profiling of mitochondria from specific cell types in vivo*. Proc Natl Acad Sci U S A, 2019. **116**(1): p. 303-312.
182. Calvo, S.E., K.R. Clauser, and V.K. Mootha, *MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins*. Nucleic Acids Res, 2016. **44**(D1): p. D1251-7.
183. Smith, A.C. and A.J. Robinson, *MitoMiner v4.0: an updated database of mitochondrial localization evidence, phenotypes and diseases*. Nucleic Acids Res, 2019. **47**(D1): p. D1225-D1228.
184. Anderson, S., et al., *Sequence and organization of the human mitochondrial genome*. Nature, 1981. **290**(5806): p. 457-65.
185. Silva Ramos, E., et al., *Mitochondrial fusion is required for regulation of mitochondrial DNA replication*. PLoS Genet, 2019. **15**(6): p. e1008085.
186. Liu, Y.J., et al., *Mitochondrial fission and fusion: A dynamic role in aging and potential target for age-related disease*. Mech Ageing Dev, 2020. **186**: p. 111212.

187. Cribbs, J.T. and S. Strack, *Reversible phosphorylation of Drp1 by cyclic AMP-dependent protein kinase and calcineurin regulates mitochondrial fission and cell death*. EMBO Rep, 2007. **8**(10): p. 939-44.
188. Taguchi, N., et al., *Mitotic phosphorylation of dynamin-related GTPase Drp1 participates in mitochondrial fission*. J Biol Chem, 2007. **282**(15): p. 11521-9.
189. Cereghetti, G.M., et al., *Dephosphorylation by calcineurin regulates translocation of Drp1 to mitochondria*. Proc Natl Acad Sci U S A, 2008. **105**(41): p. 15803-8.
190. Ducommun, S., et al., *Motif affinity and mass spectrometry proteomic approach for the discovery of cellular AMPK targets: identification of mitochondrial fission factor as a new AMPK substrate*. Cell Signal, 2015. **27**(5): p. 978-88.
191. Toyama, E.Q., et al., *Metabolism. AMP-activated protein kinase mediates mitochondrial fission in response to energy stress*. Science, 2016. **351**(6270): p. 275-281.
192. Lewis, T.L., Jr., et al., *MFF-dependent mitochondrial fission regulates presynaptic release and axon branching by limiting axonal mitochondria size*. Nat Commun, 2018. **9**(1): p. 5008.
193. Tigno-Aranjuez, J.T., J.M. Asara, and D.W. Abbott, *Inhibition of RIP2's tyrosine kinase activity limits NOD2-driven cytokine responses*. Genes Dev, 2010. **24**(23): p. 2666-77.
194. Tigno-Aranjuez, J.T., X. Bai, and D.W. Abbott, *A discrete ubiquitin-mediated network regulates the strength of NOD2 signaling*. Mol Cell Biol, 2013. **33**(1): p. 146-58.
195. Chen, S.J., et al., *An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes*. Science, 2017. **355**(6323).
196. Santt, O., et al., *The yeast GID complex, a novel ubiquitin ligase (E3) involved in the regulation of carbohydrate metabolism*. Mol Biol Cell, 2008. **19**(8): p. 3323-33.
197. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations*. Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
198. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold*. Nature, 2021. **596**(7873): p. 583-589.
199. Carbonara, K., M. Andonovski, and J.R. Coorsen, *Proteomes Are of Proteoforms: Embracing the Complexity*. Proteomes, 2021. **9**(3).
200. Rivera, K.D., et al., *Automating UbiFast for High-throughput and Multiplexed Ubiquitin Enrichment*. Mol Cell Proteomics, 2021. **20**: p. 100154.
201. Bache, N., et al., *A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics*. Mol Cell Proteomics, 2018. **17**(11): p. 2284-2296.

202. Serrano, G., E. Guruceaga, and V. Segura, *DeepMSPeptide: peptide detectability prediction using deep learning*. *Bioinformatics*, 2020. **36**(4): p. 1279-1280.
203. Cheng, H., et al., *PepFormer: End-to-End Transformer-Based Siamese Network to Predict and Enhance Peptide Detectability Based on Sequence Only*. *Anal Chem*, 2021. **93**(16): p. 6481-6490.
204. Cohen, P., D. Cross, and P.A. Janne, *Kinase drug discovery 20 years after imatinib: progress and future directions*. *Nat Rev Drug Discov*, 2021. **20**(7): p. 551-569.
205. Maneiro, M., et al., *PROTACs, molecular glues and bifunctionals from bench to bedside: Unlocking the clinical potential of catalytic drugs*. *Prog Med Chem*, 2021. **60**: p. 67-190.
206. Karayel, O., et al., *Accurate MS-based Rab10 Phosphorylation Stoichiometry Determination as Readout for LRRK2 Activity in Parkinson's Disease*. *Mol Cell Proteomics*, 2020. **19**(9): p. 1546-1560.

Abbreviations

DNA	described deoxyribonucleic acid
mRNA	messenger ribonucleic acid
ORF	open reading frame
UPS	ubiquitin-proteasome system
PTM	post-translational modification
ER	endoplasmic reticulum
MS	Mass spectrometry
m/z	mass-to-charge ratios
ES	electrospray ionization
MALDI	matrix assisted laser desorption ionization
DTT	dithiothreitol
TCEP	tris(2-carboxyethyl)phosphine
IAA	iodoacetamide
CAA	chloroacetamide
StageTips	STop And Go Extraction tips
SDB-RPS	styrenedivinylbenzene - reversed phase sulfonate
SDC	sodium deoxycholate
DIA	data-independent acquisition
FIA	Flow injection analysis
HETP	Height equivalent to a theoretical plate
LC-MS	liquid chromatography – mass spectrometry
CRM	charge residue model
IEM	ion evaporation model
ID	inner diameter
RF	radio frequency
TOF	time-of-flight
Tims	trapped ion mobility spectrometry
IM	Ion mobility
FAIMS	high-field asymmetric waveform ion mobility spectrometry
PASEF	parallel accumulation–serial fragmentation
CID	collision-induced dissociation
HCD	higher-energy collision dissociation
UVPD	ultraviolet photodissociation
ETD	electron transfer dissociation
DDA	data dependent acquisition
PQP	peptide query parameters”
SIM	selected ion monitoring
SRM	single reaction monitoring
PRM	parallel reaction monitoring
QQQ	triple quadrupole
MRM	multiple reaction monitoring
PSM	peptide-spectrum matches
FDR	false-discovery rate

LFQ	Label-free quantification
XIC	extracted ion chromatogram
TMT	tandem mass tags
SILAC	stable isotope labeling by amino acids in cell culture
PSAQ	Protein Standard Absolute Quantification
RING	really interesting new gene
HECT	homologous to the E6AP carboxyl terminus
RBR	Ring-between-Ring
DUB	deubiquitinase
USP	ubiquitin-specific protease
UCH	ubiquitin C-terminal hydrolase
MJD	Machado-Josephin domain protease
OTU	ovarian tumor protease
JAMM	Jab1/Mov34/Mpr1 Pad1 N-terminal+ (MPN+)
diGly	diglycine
Ubl	ubiquitin-like
UBA	ubiquitin-binding domains
UIM	ubiquitin-interacting motives
TUBE	tandem-repeated ubiquitin-binding entities
tUIM	tandem ubiquitin-interacting motifs
UbiChEM-MS	Ubiquitin Chain Enrichment Middle-down Mass Spectrometry
IEX	ion-exchange
MMC	mixed-mode chromatography
SAX	strong anion exchange
SCX	strong cation exchange
ERLIC	Electrostatic repulsion hydrophilic interaction chromatography
IMAC	Immobilized Metal Ions Affinity Chromatography
MOAC	Metal Oxide Affinity Chromatography
DC	differential centrifugation
MACS	magnetic bead-assisted methods
OMM	outer mitochondrial membrane
CML	Chronic myelogenous leukemia
NSCLC	non-small cell lung carcinoma
PROTAC	proteolysis targeting chimera

Acknowledgements

I would like to express my deepest gratitude to all those who have guided me on this exciting journey.

My first and foremost thanks are devoted to Prof. Dr. Matthias Mann, who gave me the great opportunity to pursue my research interests. Driven by the knowledge of your support, guidance and trust you made my work possible. Furthermore, you enabled my close collaboration with Prof. Dr. Brenda Schulman, who's enthusiasm for science and constant firework of ideas has been a great source of inspiration for me. I am very thankful for being a part of both your teams.

I also would like to thank my examiners board, consisting of Prof. Dr. Julian Stingele, Prof. Dr. Lucas Jae, Prof. Dr. Johanna Klughammer and Prof. Dr. Veit Hornung for critical reading of my thesis. A special thanks goes to Prof. Dr. Jesper Olsen, who was a valuable member of my TAC committee and great source of advice.

In the framework of this thesis, I was very fortunate to collaborate with the amazing groups of Prof. Dr. Charo Robles and Prof. Dr. Nils-Göran Larsson. Here, I especially thank Franziska Brüning, Laura Kremer and Inge Kühl who made troubleshooting fun.

In general, I thank the research group of Matthias Mann with all it's current and past members for creating such a vibrant research atmosphere, where scientific advice can be found around each corner.

For the excellent administrative support, I want to thank Alison Dalfovo, Theresa Schneider, Ute Sackers, Medini Steger and Mario Oroshi. For cell culture support I want to thank Bianca Splettstoesser. Furthermore, I want to highlight the outstanding technical support of Igor Paron, Sophia Steigerwald, Johannes Müller-Reif, Lisa Schweizer and Patricia Skowronek, who enable with mass spectrometer maintenance and the production of high-performance chromatographic columns cutting edge LC-MS experiments.

Lively discussions and the entertaining atmosphere in the office would not have been possible without Matteo Pecoraro, Florian Schober, André Michaelis, Jakob Bader and Tung Vu. It's a shame that the SARS-COV-2 pandemic took a toll on our office life.

Isabell Bludau, your enthusiasm for biology and PTMs, paired with your bioinformatics wizardry broaden my scientific horizon, but more importantly you are a great friend.

Özge Karayel, it is difficult to express how thankful I am for having you as my mentor. Your commitment is a constant source of inspiration on a scientific and personal level and I cannot imagine how my time in the lab would have been without you.

Finally, I would like to all my friends and family. Mama and Papa, without you nothing of all this would have been possible (quite literally). You equipped me with the tools to become who I am and there is no way to express how grateful I am for this. Fenna, you accompanied me for many formative years of my life, of which I don't want to miss a single moment. My greatest thanks are devoted to Leonie, you are the great constant by my side and I cannot possibly show how happy I am to have you in my life.