DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER FAKULTÄT FÜR CHEMIE UND PHARMAZIE
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Development and Application of SILAC-PrEST:

An accurate method for absolute protein quantification

Marlis Zeiler

aus

Wien, Österreich

2014

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, am 21. August 2014

……..

Marlis Zeiler

Dissertation eingereicht am	11.02.2014
1. Gutachter	Prof. Dr. Matthias Mann
2. Gutachter	Prof. Thomas A. Neubert, PhD
Mündliche Prüfung am	12.03.2014

Table of contents

Summary

Mass spectrometry based proteomics has become a method of choice for identification and quantification of proteins in complex biological systems. It allows for absolute quantification – the determination of protein copy numbers and absolute concentrations by means of internal standards. This enables kinetic modeling, stoichiometry determination and facilitates comparison among different studies. Nonetheless, existing absolute quantification methods have limitations. For instance, in peptide-based absolute quantification, a single peptide is often spiked in at a late stage of the sample preparation workflow, compromising accuracy. Absolute quantification based on full-length standards, on the other hand, is time consuming and laborious and therefore low in throughput.

In this thesis, I established a new method for absolute protein quantification, which is capable of overcoming existing limitations and applied it to a global proteome study of cancer cells, platelets and stoichiometry determination of protein complexes. In the first project, I developed the novel absolute quantification strategy termed 'SILAC-PrEST', which uses SILAC-labeled protein epitope signature tags (PrESTs) as internal standards. These PrESTs were generated in the course of the Human Protein Atlas for polyclonal antibody production. They are designed to cover a 50-150 amino acids long, unique sequence of the protein of interest, and are fused with a purification and solubility tag. First, the solubility tag alone is highly purified and its concentration is determined by amino acid analysis. Subsequently, each PrEST is quantified against the solubility tag via the SILAC ratio (ratio of 'heavy' to 'light' labeled peptide). Next, the PrESTs are combined in a predefined proportion to generate a master mix which is spiked into cell lysate. The absolute copy number of the endogenous proteins is then deduced from the measured SILAC ratios. This quantification workflow successfully allowed concurrent quantification of 43 proteins ranging from 6,000 to 20 million copies in HeLa cells.

In a second, collaborative project, we combined the SILAC-PrEST method with an enrichment step using polyclonal antibodies generated by the Human Protein Atlas.

Epitope mapping revealed that the antibodies generated from the PrESTs recognize multiple linear epitopes and therefore allowed for enrichment of multiple tryptic peptides from the target. We termed this method 'immuno-SILAC' and in an application demonstrated the successful enrichment of endogenous and PrEST-derived peptides by a mixture of up to 50 immobilized antibodies. Immuno-SILAC enables efficient multiplexed absolute quantification of proteins from complex mixtures.

In a further project, I set out to characterize the entire murine platelet proteome using the SILAC-PrEST method to determine protein copy numbers. By analyzing proteomes of platelets of increasing purity we were able to identify and exclude contaminants originating from erythrocytes and plasma. I quantified thirteen platelets proteins directly by SILAC-PrEST and used these values as reference points to scale the copy number of each protein of the platelet proteome. Altogether we identified and quantified 4,400 proteins, constituting a close-to-complete reference map of the platelet proteome. Remarkably, proteins with estimated copy numbers of less than 10 per platelet were measured.

In the last and collaborative project, we applied the SILAC-PrEST method to determine the expression levels and stoichiometry of integrin beta 3 and its coactivators talin-1 and kindlin-3. In platelets, integrin αIIBβ3 is the major adhesion receptor. It binds to ligands upon activation via a process termed 'inside-out' signaling, whereby intracellular signaling and eventual binding of talin-1 and kindlin-3 to the cytoplasmic tail of the integrin, induces its conformational change. We discovered that the stoichiometry of the coactivators relative to integrin β3 was 2:1. Experiments on several mouse strains expressing different amounts of kindlin-3, revealed that high levels of kindlin-3 are required for activation of this integrin.

In summary, the projects presented in this thesis establish a novel, more accurate method for absolute protein quantification, which was successfully applied to diverse biological questions.

Introduction

Proteins are involved in virtually all processes and pathways in cellular systems. The entire set of proteins present in a cell is referred to as the proteome, a term first introduced in 1996 by Marc Wilkins as the entire 'protein complement expressed by a genome' [1]. Proteome research gained momentum by the availability of numerous genome maps of different organism in the 1990s. These gene maps serve as starting point for proteome research by predicting the sequence of the potential proteins in the system under investigation. The shift of emphasis from DNA to its protein products thus defined the start of the post-genome era for the proteomics community [2]. The human genome, for example, is believed to consist of about 20,000 protein coding genes. However, since one gene can give rise to multiple different proteins due to alternative splicing, sequence polymorphism and post-translational modifications, an enormous complexity is expected [3]. Furthermore, the abundance level of these polymorphic species varies highly between cell types, tissues and developmental stages, as well as responding to extra- and intracellular signals. This immense complexity and diversity can only be tackled by high-throughput technologies. In the beginning, proteomic studies relied on two-dimensional gel electrophoresis, but this technique proved unsuitable for in-depth characterization of full proteomes for a number of reasons [4]. In contrast, mass spectrometry (MS)-based proteomics has matured in the past 25 years to become a powerful and versatile technology that can sensitively and accurately identify and quantify proteins, their modifications and interactions [5, 6]. This technology is continuously developing thanks to advances in instrumentation, samples preparation strategies and computational analysis.

Mass-spectrometric measurements are not inherently quantitative, and various methods have been developed in response to this challenge. Many quantification methods are 'relative' in that they report ratios between two or more conditions. As such, the actual abundance remains unknown. In contrast, 'absolute' quantification allows determination of actual concentrations e.g. ng/ml of a clinical biomarker in plasma, or the copy number

of a protein in a cell. Therefore, in general only data from absolute quantification can be used for determination of kinetics and stoichiometries [7]. Moreover, absolute quantification data can easily be compared and exchanged among different studies and laboratories. Recently, absolute quantification became wide-spread in proteomics with the development of strategies employing isotope-labeled standards [8]. Apart from methods quantifying one or a small set of proteins, computational methods have also recently been developed that estimate the absolute abundance from the summed MS signals of all peptides identifying a protein [9-11].

In this thesis, I developed a method for absolute quantification termed 'SILAC-PrEST' based on a novel kind of stable isotope labeled standards. Below, besides the method development itself, applications of the absolute quantification strategy are described. The following introduction focuses on the basic workflow of MS-based proteomics, mass spectrometric instrumentation, proceeds to existing quantitative methods and concludes with a short introduction of the biological context in which this new SILAC-PrEST method was applied.

Mass spectrometry-based proteomics

MS-based proteomics is performed either 'bottom-up', where proteolytic peptide mixtures are analyzed [6] or 'top-down', where intact proteins are studied [12]. The former approach is more established, much more powerful and is therefore more widely used. A typical bottom-up MS-based proteomic experiment starts with extracting and optionally enriching proteins from tissue or cells. Next, the proteins are subjected to proteolytic digestion. Trypsin is commonly used as the protease to convert proteins into peptides since it specifically cleaves proteins on the carboxyl-terminal side of arginine and lysine residues, producing peptides with a size and charge range well-suited for MS analysis. To reduce complexity, samples may be fractionated either at the protein level based on the molecular weight using SDS-PAGE or gel filtration, or at the peptide level based on the peptide's physical and chemical properties, generally with strong anion exchange or isoelectric focusing.

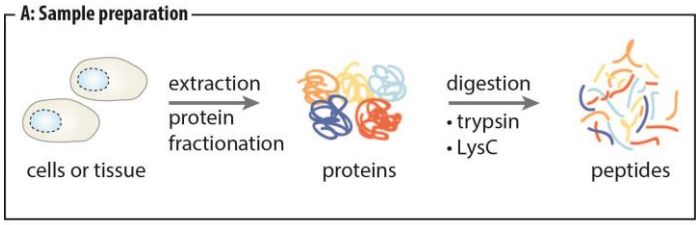


Figure 1: Simplified schematic outline of the sample preparation for a typical MS-based proteomics experiment

Next the peptides are separated by high-performance liquid chromatography (HPLC) and introduced 'online' into the mass spectrometer via electrospray. Mass spectrometers used in proteomics are capable of measuring the mass of biomolecules as well as of isolating and fragmenting them. They analyze the peptide mixture in a data-dependent fashion. First, the mass and intensity are measured in the so-called 'full scan' or 'MS1' scan and then the most abundant peptides are selected for fragmentation, generating 'tandem' or 'MS/MS' spectra. The peptides are identified by matching the fragmentation spectra to theoretical spectra generated from a sequence database (Figure 2). For protein assembly the identified peptides are computationally mapped to the corresponding protein.

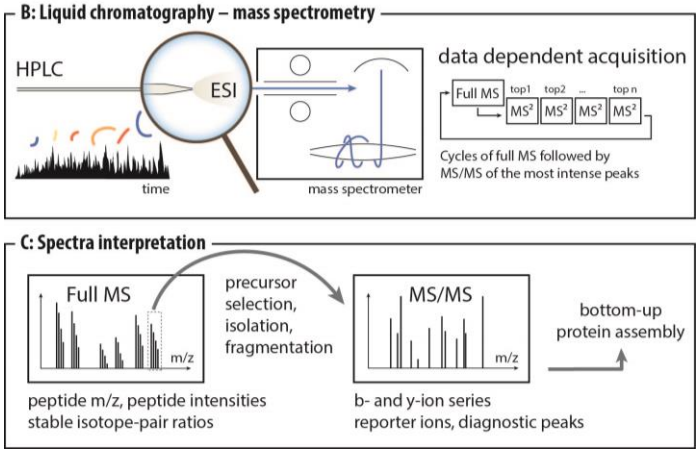


Figure 2: Schematic outline of a shotgun LC-MS/MS experiment divided into , (B) liquid chromatography and mass spectrometry, (C) spectra interpretation. Adapted from [13].

The aforementioned workflow is referred to as 'shotgun' proteomics [14] and is by nature a discovery based strategy, representing a powerful method for global characterization of proteomes (qualitatively and quantitatively). Using this shotgun proteomics approach, two recent studies identified a close to complete proteome of human cancer cell lines [15,

16]. One limitation of data dependent shotgun proteomics is the stochastic nature of the data acquisition, as the mass spectrometer does not always pick the same peptides for fragmentation. In particular, low-abundant peptides may only be stochastically sequenced and identified. However, scanning speed and sensitivity of the mass spectrometer has increased significantly in the recent years, and computational approaches for transferring identifications between samples have been developed to compensate for this potential limitation.

In addition to the shotgun approach, hypothesis-driven strategies are also available. Such 'targeted' methods by their nature monitor a number of predefined proteins. While the sample preparation is the same, the acquisition strategy differs. One implementation of targeted proteomics is multiple reaction monitoring (MRM), where predetermined precursor ions are selected for fragmentation and one corresponding product ion detected [17, 18]. One measurement of a mass-to-charge (*m/z)* pair is referred to as a 'transition' and several transitions are performed over the chromatographic peak, resulting in a trace of signal intensity versus retention time. To accurately quantify a protein, several peptides – each with at least two transitions – are generally targeted, resulting in no more than 50 proteins being concurrently monitored in a typical measurement. A potential challenge is the verification of the identity of the *m/z* pair measured in the MRM assay. This could in principle be overcome by measuring a pseudo tandem MS/MS spectrum, however at the expense of sensitivity and analysis speed. In practice, it is not trivial to select appropriate peptides for a robust, sensitive and specific assay. Various parameters such as MS properties, uniqueness, post-translational modifications, and cleavage sites have to be taken in consideration before choosing a peptide [18]. Nevertheless, once the identity of targeted species is established, the assay can be highly sensitive and accurate, especially if performed on high resolution instruments rather than the typically employed triple quadrupoles. MRM in principle allows absolute quantification of the targeted peptides if combined with heavy isotope-labeled peptide references. MRM can therefore be used to screen clinical samples for specific biomarkers in high-throughput. The targeted strategy

offers an attractive alternative to antibody based methods such as ELISA, however, in comparison to shotgun proteomics this comes at the expense of unbiased knowledge of the system.

Liquid chromatography mass spectrometry

Historically, mass spectrometry was restricted to the analysis of small molecules, owing to the lack of technique to gently ionize and vaporize biomolecules. The development of two ionization methods – electrospray ionization (ESI) [19] and matrix-assisted laser desorption ionization (MALDI) [20, 21] – allowed bringing biological macromolecules into gas-phase without destroying them, making peptides and proteins amenable to MS analysis. In electrospray, the peptides mixture is sprayed in a strong electrical field between the needle and the MS inlet, forming charged droplets that rapidly desolvate and eventually become 'lifted' into the gas phase (Figure 3) [19].

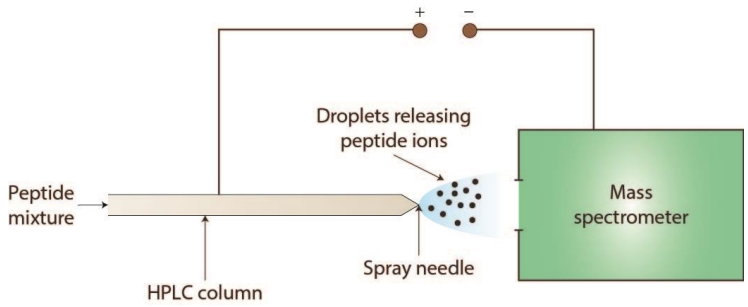


Figure 3: Electrospray ionization in combination with liquid chromatography. The peptides are separated on a typically reverse phase column. The peptides are ionized via electrospray and enter the mass spectrometer for analysis. Adapted from [22].

The development of these ionization methods – acknowledged with the Nobel Prize in Chemistry in 2002 – boosted the field of mass spectrometry based proteomics, particularly since ESI is highly compatible with preceding liquid separation technologies. Usually reverse phase chromatography is used to separate molecules by their hydrophobicity, so that only a limited diversity of peptides is introduced to the mass spectrometer at a given time. This consequently improves the accessibility of low-abundant peptides in a shotgun experiment and increases the depth of the resulting proteome [23].

Instrumentation

Regardless of the type of instrument, each mass spectrometer consists of three major components: an ion source, a mass analyzer and a detector. The heart of each mass spectrometer is the mass analyzer which separates ions according to their *m*/*z*. Commonly, time-of-flight (TOF), quadrupoles, Fourier transform ion cyclotron resonance (FT-ICR), linear ion trap (IT) and the recently developed Orbitrap analyzer are used in the field of proteomics as mass analyzers [24]. The performance of the individual mass analyzers can be characterized by mass range, analysis speed, sensitivity, resolution and mass accuracy. FT-ICR and Orbitrap analyzers, for example, are characterized by high mass resolving power and high mass accuracy while the ion trap – particularly the linear ion traps (LTQ) – features fast scan rates and high sensitivity [25]. Each mass analyzer has its merits and limitations and according to some, the ideal mass analyzer does not exist [26].

Hybrid mass spectrometers combining conceptually distinct mass analyzers and exploiting the benefits of each component increase the performance and versatility of the combined instruments [27]. Examples of such hybrid instruments include quadrupole TOFs, LTQ Orbitraps and quadrupole Orbitraps. The quadrupole TOF and quadrupole Orbitrap analyzer are broadly used and have proven invaluable in proteomics. In this thesis hybrid instruments (LTQ-Orbitrap, LTQ-Velos and Q Exactive) from Thermo Fisher Scientific have been used and therefore their mass analyzers (ion trap and Orbitrap) as well as their set-up are discussed in more detail below.

Linear ion traps

The ion trap is one of the most popular mass analyzers in MS-based proteomics. Ion traps are devices capable of storing, isolating and fragmenting ions and traditionally existed in a three dimensional configuration, which has more recently been largely superseded by an implementation of the ion trap based on a quadrupole (linear ion trap). They are characterized by their high scanning speed and high sensitivity, but yield only

moderately resolved spectra and relatively low mass accuracy. Therefore, ion traps are often combined with high resolution analyzers such as the Orbitrap or FT-ICR to form hybrid instruments. The quadrupole in the linear ion trap has four hyperbolic rods, each rod separated into three sections (Figure 4a). In the central section, one of the pairs of rods contain a slot for ion ejection and consequent detection by electron multipliers. An elaborate RF/AC/DC system is applied to the rods in all dimensions to achieve radial and axial trapping as well as isolation, excitation and ejection of ions (Figure 4b) [24, 25].

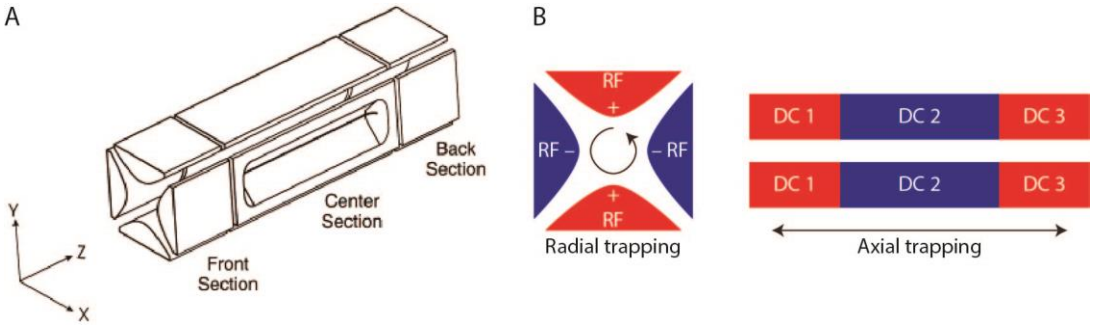


Figure 4: The linear ion trap: (a) Schematic view of the two dimensional ion trap (b) Different voltage are applied to achieve radial or axial trapping, respectively. Adapted from [25] and [28]

Orbitrap

In 2000, the Orbitrap analyzer was introduced by Makarov [29], although a first precursor was already described by Kingdon in 1923 [30]. The Orbitrap detector consists of an inner spindle-shaped electrode surrounded by an outer barrel-like electrode, which is split into two halves allowing radial ion injection. In order to efficiently inject ions into the Orbitrap, focused ion packages with narrow spatial and temporal distribution are required. This is achieved in the 'C-trap' upstream of the Orbitrap [31]. In this curved, C-shaped quadrupole, the ions are collisionally-cooled with nitrogen bath gas. In the following ion optics, the ions are accelerated and made to converge in order to be injected tangentially into the Orbitrap via a small aperture [32]. In the mass analyzer, the ions are trapped in an electrostatic field by orbiting around the inner electrode and simultaneously oscillating along the longitudinal axis.

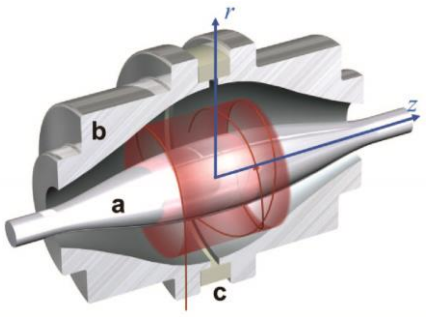


Figure 5: The geometry of the Orbitrap analyzer: (a) inner spindle shaped electrode, around which the ions move in addition to oscillating along the z-axis (b) the outer barrel-like electrode with a gap (c) ceramic ring electrically isolating the two halves from each other. Adapted from [32]

Initially in the development of the Orbitrap, the rotational frequency was used to deduce the *m*/*z* of the ions, but this yielded inaccurate results because these rotational frequencies also depended on initial properties of the injected ions. Therefore, the frequency of the harmonic oscillation along the z-axis is used instead. The electrical current resulting from the oscillating ions (the 'image current') is detected across the split of the outer electrodes as an electrical transient, and is Fourier-transformed (FT) to produce frequencies for each ion, which in turn are converted to accurate *m*/*z* values. The Orbitrap mass analyzer is capable of generating very high resolution and high mass accuracy measurements [33].

Tandem mass spectrometry

In the mass spectrometer, in the first analysis step the *m*/*z* value and intensity of an ion are determined (referred to as MS, MS^1, survey scan or full scan) and in the second analysis step they are fragmented and the fragmentation spectrum is recorded (termed 'tandem mass spectrometry', MS/MS or MS^2): Tandem MS involves activation of a selected and isolated precursor ion and then dissociation of the peptide backbone yielding various product ions [34]. To achieve fragmentation, electric fields are applied and consequently the ions collide with a low pressure bath gas. In collision induced dissociation (CID) and higher energy collisional dissociation (HCD), the cleavage typically occurs at the C-N peptide bond. Depending on where the charge is retained, the amino-terminal or carboxyl-terminal fragments can be observed. Following the Roepstorff-Fohlmann-

Biemann nomenclature, these fragments are referred to as b- or y-type ions, respectively (Figure 6) [35].

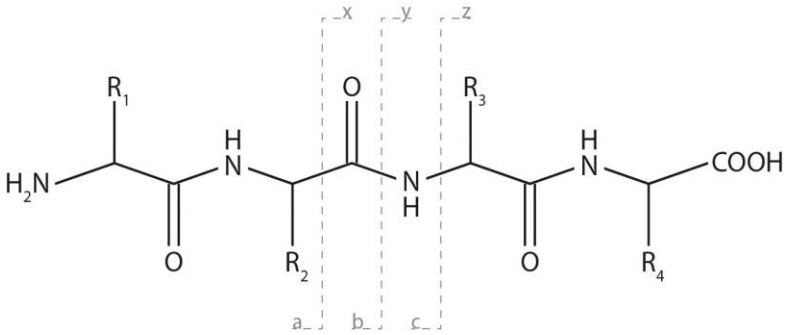


Figure 6: Peptide fragmentation: excitation of peptide ions induces cleavage of distinct bonds of the peptide backbone, yielding N-terminal and C-terminal ions. The fragments are named according to the Roepstorff–Fohlmann–Biemann nomenclature. Adapted from [22, 35].

Based on a complete series of fragmented ions, differing in mass by a single amino acid residue, the sequence can be determined either by manual interpretation ('*de-novo* sequencing') or computationally by comparing the fragmentation pattern to a sequence database (further described in the computational proteomics part).

Hybrid instrumentation

In case of the LTQ-Orbitrap, which was commercially introduced in 2005, the sensitivity of the ion trap was united with high resolving power of the Orbitrap analyzer [31, 36]. It quickly became a popular instrument for proteomics applications. Nevertheless, the set-up was constantly improved, and the next generation, termed LTQ-Orbitrap Velos, instruments were significantly faster and more sensitive. To achieve this boost in sensitivity, the ion optics were modified by replacing the tube lens with a S-lens, allowing better ion transmission. Furthermore, the linear ion trap was replaced by a dual-pressure ion trap that comprised two identical linear quadrupole cells. These cells are operated at different gas pressures, allowing capture and fragmentation to be uncoupled from scanning and detection. The first, high pressure ion trap efficiently captures the ions and generates fragments ions faster and more efficiently due to reduced activation time. The

second, low pressure ion trap allows for enhanced scanning speed while maintaining adequate resolution (Figure 7) [37].

In the LTQ-Orbitrap Velos, the introduction of highly efficient combination of the C-trap with the HCD collision cell enabled two fragmentation strategies. Before this, the full scan was measured at high resolution in the Orbitrap, while collision induced dissociation and fragmentation spectra acquisition were performed at low resolution in the highly sensitive ion trap in parallel. This mode of operation is referred to as 'high-low'. The new, highly efficient HCD fragmentation method allowed routine use of a 'high-high' mode, in which both the MS determination of peptides ('precursors') and of the fragments are performed in the Orbitrap analyzer. Therefore, in the new 'high-high' mode, both the survey spectrum and the fragmentation spectra are obtained at very high resolution and consequently with high mass accuracy. In addition, acquisition of the HCD fragment ions in the Orbitrap allows observation of low mass fragment ions, which are usually not observed in CID performed in ion traps, improving peptide characterization.

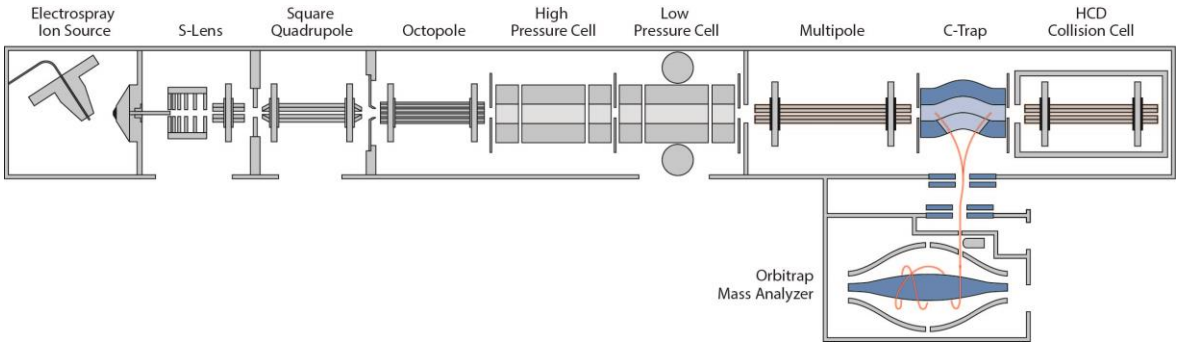


Figure 7: The hybrid mass spectrometer configuration of the LTQ Velos. Adapted from [37].

In the latest Orbitrap mass spectrometer, the bench top Q Exactive, the ion trap is replaced by a quadrupole operating as mass filter, leaving the Orbitrap as the only mass detector. In this setup, survey scans and the fragmentation scans are exclusively acquired at high resolution in the Orbitrap analyzer (Figure 8) [38].

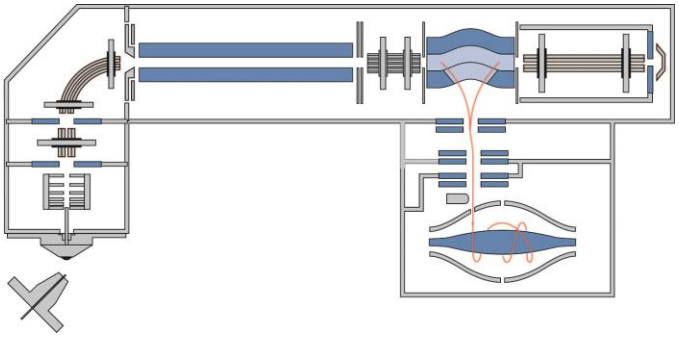


Figure 8: The Q Exactive set-up. Compared to previous models, the ion trap has been replaced by a quadrupole. Adapted from [38].

Computational proteomics

In a typical shotgun experiment, the complex peptide mixture is separated by liquid chromatography and analyzed by the mass spectrometer to generate a map of retention time and *m*/*z* values (Figure 9a). In brief, the mass spectrometer first acquires a survey scan, containing the precursor masses with intensity information. As explained above, five to twenty of the most abundant ions are then isolated and subjected to fragmentation. This cycle is repeated throughout the whole chromatographic gradient.

The large amount of data acquired prohibits manual interpretation, and rather requires sophisticated software. Various packages are available, most performing the following common steps: (i) peak detection (ii) identification of the corresponding MS/MS spectra (iii) statistical validation of the identifications (iv) protein assembly and (v) quantification. For this thesis the powerful MaxQuant software, including its own search engine 'Andromeda' was used and is therefore described in more detail [39, 40].

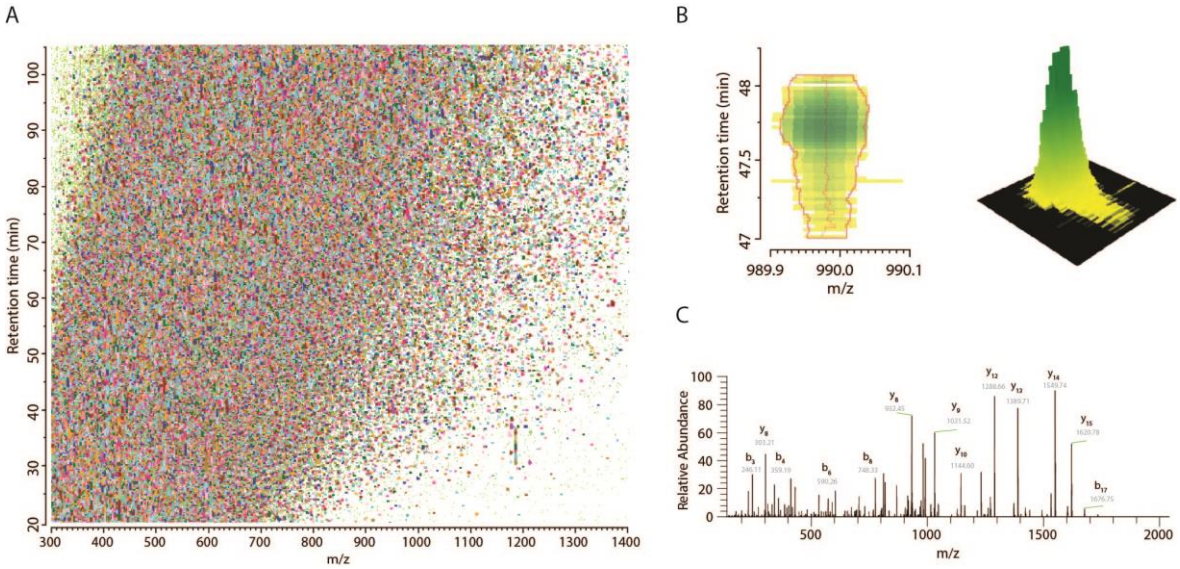


Figure 9: (a) Visualization of all peptides feature detected in a single LC-MS run. (b) Isotope cluster of one peptide, which is (c) targeted for fragmentation and successfully identified.

First, a series of MS scans spanning a peptide's elution profile is used to construct a three-dimensional peak, which is used to extract the accurate mass and the intensity (Figure 9b). The next and central step of the data analysis is the assignment of the corresponding

MS/MS spectrum to a peptide sequence, which is achieved by database searching. In this process, the observed spectra, retaining the top 12 peaks per 100 Da interval, are compared against theoretical fragment ion spectra. These theoretical spectra are obtained by an *in silico* digestion of all proteins in a reference protein sequence database. A successful match between the observed and theoretical spectrum requires the same *m*/*z* within an individualized mass tolerance not exceeding 20 ppm (Figure 9c). Clearly, the choice of database plays an important role since they vary in terms of completeness, degree of redundancy and quality of sequence annotation.

Currently, the standard database used in MS-based proteomics is UniProtKB, which consists of the manually-curated Swiss-Prot and the computationally generated TrEMBL (Translated EMBL Nucleotide Sequence Data Library) [41]. Inclusion of TrEMBL allows one to confirm existence of predicted proteins based on experimental data. Once the peptide sequence is established, the probability that the database match is a random event is determined and reported as score (*-10*log(probability)*), which can be interpreted as the confidence in the identification. Additionally, to identify false positive hits the 'target-decoy' strategy is employed: the MS/MS spectra are searched against a concatenated database of target (intended) but also decoy (reversed) sequences. Peptides identified in the decoy database are evidently incorrect and, assuming an equal probability of false identification from both databases, the score threshold can be used to control the false discovery rate (FDR) to a desired level (typically 1 %) [42].

Because peptides rather than proteins are identified in the bottom-up workflow, the actual proteins have to be reassembled computationally. However, mapping peptides to the corresponding proteins is not trivial, since peptides sequences can be shared between homologous proteins, protein families, multiple isoforms and even redundant entries in the reference database [43]. To obtain a protein list best describing the original sample, the concept of 'protein groups' was introduced.

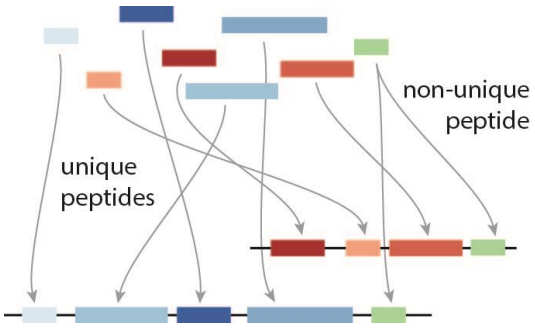


Figure 10: Protein grouping. Non unique peptides that might be assigned to two or more protein groups are placed into the protein group with most peptides. Adapted from [13].

In essence, proteins not distinguishable on the basis of identified peptides are collapsed into one group [44]. Peptides are considered 'unique' if they only occur in one protein group. 'Razor' peptides could also be assigned to another protein group, yet they are placed in the protein group with most peptides (Figure 10). After protein assembly the same basic principle of FDR is applied to receive a high-confidence protein list. Once the proteins in the sample are established, a measure of protein quantity is calculated. Various methods to achieve relative or absolute quantification are described in the next section.

To gain functional insights and to draw biological conclusions from the data set, further computational analysis needs to be performed. Downstream analyses can include hierarchical clustering, principal component analysis, tests for differential regulation, time series, pathway and ontology enrichment analysis. A myriad of software packages that offer suitable statistical tools and algorithms (mainly developed for microarray data) are publicly available, for example Bioconductor for R [45]. Here we use the Perseus program, which is very comprehensive and part of the MaxQuant software package [46].

Quantitative proteomics

Perturbation of a biological system usually results in quantitative change in protein expression levels, more so than qualitative change in the identity of the proteins being expressed. Previously, proteomic studies using mass spectrometry only yielded peptide or protein identification, i.e. they purely enumerated proteins without offering a quantitative estimate of the level of the proteins in a biological system [8, 47]. Mass spectrometry itself is not inherently a quantitative technique, since signal intensity of a peptide depends on numerous of its physicochemical features e.g. length, charge, hydrophobicity and ionization efficiency. Therefore, a wide variety of strategies for protein quantification have been developed.

Protein quantification can either measure relative changes between two experimental states, or absolute protein amounts in the sample. In relative quantification, two states are compared and fold changes are determined between them. In absolute quantification, protein concentrations or copy numbers per cell are assessed (note that absolute quantification in two states, also allows relative comparisons between them) [48]. In either case, most quantification techniques employ stable isotope labeling to introduce distinct mass differences that can be distinguished in the mass spectrometer. These mass differences are created by 'heavy' and 'light' stable isotope versions of the protein or peptide, and can be (i) incorporated metabolically, (ii) attached by chemical reactions, or (iii) introduced enzymatically during protein digestion. Quantification may also be achieved by spiking in stable isotope-labeled proteins or peptides as internal standards. As a complement to stable isotope labeling, 'label-free methods' that compare the peptide intensity or count the number of spectra, have emerged over the past decade and have been successfully established in the field (Figure 11). In the section below, the most common quantification methods are further described and practical aspects, applicability and pitfalls are discussed.

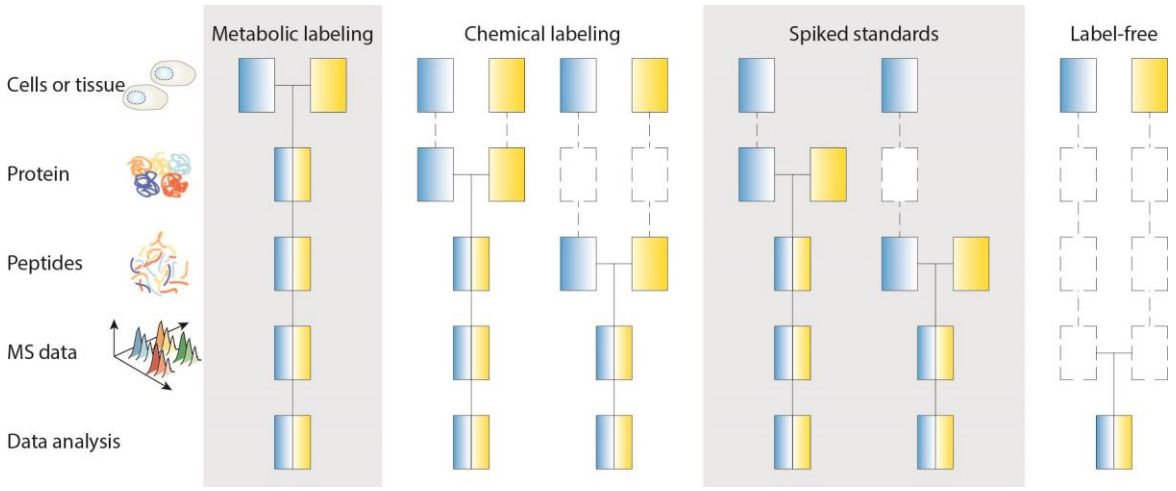


Figure 11: Quantitative proteomics workflow. The blue and yellow boxes symbolize the two experimental conditions to be compared. The horizontal line indicates the stage at which the samples are combined. The fewer the steps that are performed in separate, parallel processes, the higher the accuracy and the robustness of the quantification. Adapted from [48]).

Relative quantification using stable isotope-based methods

To introduce a mass shift, stable (non-radioactive) isotopes of nitrogen (^{15}N) and carbon (^{13}C) are widely used, however, oxygen (^{18}O) and hydrogen (^{2}H) have also been employed. The advantage of 'heavy' isotopes is that their incorporation into an amino acid does not change the physicochemical characteristics of the resulting proteins and their constituent peptides. This holds true during sample preparation, reverse phase chromatography separation and – importantly – ionization in the mass-spectrometer. This means that the MS signals derived from a peptide pair (heavy and light version) are separated by a defined mass shift, and that the MS intensity ratios reflect the quantitative abundance ratio of the peptide pair. However, this does not entirely hold true for deuterium ^{2}H, which is more polar and is therefore retained slightly longer on the reverse phase column than its ^{1}H counterpart. This leads to retention time shifts ('deuterium isotope effect'), which possibly introduces errors during quantification. Deuterium is thus less often used as label.

Metabolic Labeling

Metabolic labeling strategies introduce the isotope labels *in vivo* during cell growth and division. Therefore, different samples can be merged at an early step in the proteomic workflow, avoiding errors resulting from variation in parallel sample preparation and proteolytic digestion. Typically, cells are grown in media to which isotope-labeled compounds (salt or amino acids) are added. The labeled compounds are incorporated into the entire proteome of the sample [49]. Metabolic labeling was first performed in proteomics using ^{15}N by Chait and coworkers in 1999 to label yeast [50]. Since then it is mostly applied in prokaryotes and lower eukaryotes, in particular bacteria and yeast as well as plants. The advantage of metabolic labeling with ^{15}N is that every labeled peptide acquires a mass offset regardless of its amino acid composition. Thus, an ^{15}N -labeling experiment is uniquely unconstrained in the choice of protease used for digestion. However, the mass difference between labeled and unlabeled can only be calculated once the peptide sequence is known, which increases the complexity of the subsequent data analysis. Furthermore, care must be taken that only ^{15}N nitrogen sources are metabolically available to the organism.

A more commonly used metabolic labeling method is Stable Isotope Labeling of Amino Acids in Cell Culture (SILAC). This quantification technique introduced in 2002 [51] is based on *in vivo* incorporation of labeled amino acids that are essential for the organism. First, the quantification strategy was shown with deuterated leucine. However, owing to the above-mentioned retention time shifts of deuterium and the fact that not every tryptic peptide contains a leucine in the sequence, other amino acids are more frequently used today. In particular, $^{13}C_6{}^{15}N_4$ arginine and $^{13}C_6{}^{15}N_2$ lysine are utilized, which introduce a mass shift of 10 and 8 Dalton (Da), respectively. The advantage of labeling lysine and arginine is that in combination with trypsin, virtually all proteolytic peptides are amenable to quantification, since each tryptic peptide has at least one arginine or lysine (except the C-terminal peptide of the protein) [52]. The cell culture media is supplemented with the labeled amino acids and dialyzed serum is used to make the heavy amino acids

the exclusive source for protein synthesis. After 5 – 8 doublings virtually all proteins in the cells have incorporated the heavy amino acids, typically reaching an incorporation rate of 98 % or more. One possible pitfall in SILAC is the metabolic conversion of arginine to proline, leading to an additional peptide peak containing heavy proline. Although this happens to the same degree in light and heavy labeled cells, and therefore does not introduce an error, it makes the peptide mixtures more complex and is therefore undesirable. The Arg-Pro conversion can be addressed by titrating arginine concentration to eliminate excess arginine or by supplementing the media with proline [53]. After labeling, cells from the two states to be compared can be mixed and all sample preparation steps are then performed together. In the mass spectrometer, 'heavy' and 'light' isotope clusters appear for every peptide, constituting a SILAC pair, between which a ratio can be calculated (Figure 12). Introducing a third, 'medium' label, makes it possible to compare up to three conditions ('light', 'medium' and 'heavy') in a single experiment. For more complex comparisons, such as time courses, 'double-triple' SILAC can be employed, in which two experiments share a common conditions that is used as a reference point [54, 55].

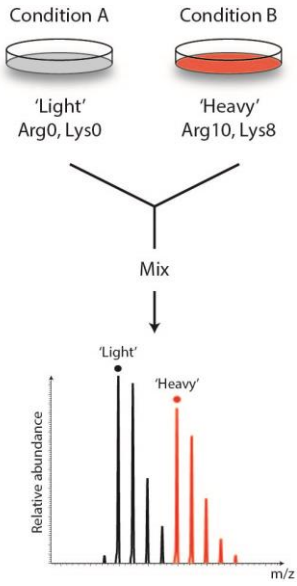


Figure 12: SILAC based quantification: Cells are grown in media containing heavy isotope labeled amino acids. The two conditions can be mixed at the protein level prior to optional sample fractionation and proteolytic digestion. In the mass spectrum a defined mass shift can be observed and peptide abundance can be directly compared. Adapted from [48].

The SILAC toolbox has been continuously extended and now also includes labeling of microorganisms such as bacteria [56, 57], yeast [58] as well as multicellular organisms such as rodents [59], nematodes [60], flies [61] and plants [62]. The SILAC method requires metabolic labeling of the studied system, and therefore cannot be used directly to analyze human tissue or clinical samples. To overcome this limitation, the super-SILAC approach [63] was introduced: A mixture of different SILAC-labeled cell lines are used as internal reference standard. The super-SILAC mix is ideally a good representative of the heterogeneous tissue and covers most proteins expressed in the complex tissue proteome. Differences between the tissue samples are determined on the basis of the ratio to the super-SILAC mix. This approach has successfully been applied to analyze human tumors from breast cancers [64] to lymphomas [65]. With the introduction of the super-SILAC approach, many more samples have become amenable to SILAC quantification. Yet another variation of SILAC is pulse-SILAC, in which the labeled amino acid is introduced for a given time period and which can be applied to study protein turnover [66].

Chemical Labeling

In addition to metabolic labeling, chemical labeling strategies have been established. In chemical labeling reactive side chains of amino acids or the peptide and protein termini are derivatized with an isotopic labeling agent (Figure 13). Therefore, virtually any sample, including clinical samples and body fluids can be quantified via chemical labels.

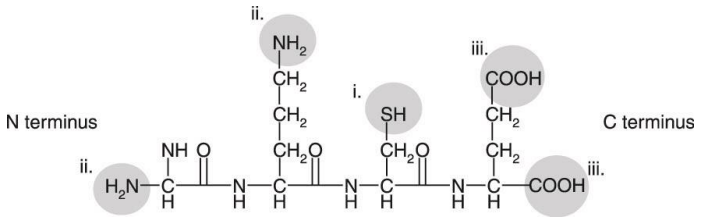


Figure 13: Chemical labeling. Any reactive amino acid side chain can in principle be used to attach an isotope label (i) Sulfhydryl group (cysteine), (ii) Amine-directed (amino terminus and ε-amino group of lysine), (iii) Carboxyl-directed (carboxyl terminus, aspartic and glutamic acid). Adapted from [48].

In 1999, Aebersold and coworkers introduced Isotope-Coded Affinity Tags (ICAT) [67] targeting sulfhydryl group of cysteines. This tag comprises a specific chemical reactive

group, an isotopically coded linker containing either zero or eight deuterium atoms and a biotin tag for enrichment of the labeled peptides. The advantage of the affinity tag is that the complexity of the sample is significantly reduced since only cysteine containing peptides, which are biotinylated through the tag are selected for quantification. However, because only cysteine-containing peptides contribute to quantification of a given protein, quantification accuracy is limited accordingly (and non-cysteine containing proteins are not quantified). Similarly, potentially interesting non cysteine-containing peptides carrying a post-translational modification (PTM) cannot be quantified.

Another strategy is to derivatize amino or carboxyl groups, that are formed during proteolysis. Tryptic peptides, for example, contain a primary amine at their N-terminus and a basic amino acid at their C-terminus, and additionally an ε-amino group in the case of lysine. Here, the advantage is that all peptides have derivatizable groups independent of the amino acid sequence and of their reactive side chains [68]. Isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT), both very popular chemical labeling methods, target the amines with a tag that is built of a reporter group, a balancer group and an amine-specific reactive group [69, 70]. The total mass of the tag remains the same while the isotopic composition between the reporter group and balancer group varies in the different labels. The labeled peptides are indistinguishable in the survey scan. Upon fragmentation, however, the tag produces different MS/MS reporter ions in the low mass region of the tandem spectrum. Quantification values are determined from the ratios of these reporter ions. Multiplexing capacity is one of the strengths of isobaric labeling since the complexity of the chromatographic separation (labeled peptides comigrate) as well as the mass spectra does not increase as with SILAC labeling. Initially, iTRAQ and TMT were used in four-plex and six-plex experiments, respectively. More recently, multiplexing of up to 18 samples has been demonstrated by combining metabolic with isobaric labeling [71]. Although multiplexing saves measurement time, it also limits the dynamic range that can be observed in a proteomic measurement because less of each sample can be loaded on the LC column.

A weakness of isobaric labeling methods is that the peptide quantification is based on single MS/MS spectrum. Another caveat of method is that the reporter ions appear in the lower mass region (masses less than 150 *m*/*z*), therefore not all mass spectrometers are suitable for the analysis - e.g. ion traps hardly recover ions below 30 % of the precursor *m*/*z*. Another issue are ions of similar *m*/*z* that coelute and are therefore isolated and fragmented together with the selected ion. These 'contaminating' ions contribute to the reporter ion ratio and therefore interfere with peptide quantification [72]. Decreasing the isolation window of the peptide of interest partially address this issue, however, at the cost of decreased sensitivity. Two dedicated approaches to deal with interfering ions have recently been described: an MS^3-based method [73] and gas-phase purification [74]. In the former, the most abundant fragment ion from MS^2 is further isolated and fragmented to read out the reporter ion intensities. In gas-phase purification, the ion charge is reduced to increase the *m*/*z* difference between the selected and contaminating ion prior to fragmentation. Both strategies, however, suffer from reduction of sensitivity and acquisition speed.

Another established labeling technique in the quantitative proteomics field is dimethyl labeling, where primary amines of proteolytic peptides are targeted and converted into dimethylamines [75, 76]. First a Schiff base is generated with formaldehyde, which is then reduced by the addition of cyanoborohydride. Formaldehyde and cyanoborohydride containing varying number of heavy isotopes (^{13}C and ^{2}H) lead to mass shifts of 28, 32 and 36 Da for the 'light', 'intermediate' and 'heavy' label, respectively. The reaction is efficient and fast, and can also be performed online in reverse phase chromatography columns [77]. In contrast to iTRAQ or TMT, the reagents used for dimethyl labeling are inexpensive and therefore up-scaling up to milligrams of proteins is feasible. Because deuterium is part of the label, retention time shifts can occur.

Chemical labeling methods can be performed either on the protein or the peptide level. Clearly it is better that the samples are tagged and mixed at an early stage of the experiment. Errors introduced during the sample preparation would be eliminated if

tagging is performed at the protein level. However, chemical labeling methods targeting primary amines, in particular the ε-amine of lysine, cannot be performed prior to digestion as they interfere with trypsin and LysC digestion.

Absolute protein quantification using stable isotope-based methods

Accurate absolute quantification involves adding a known concentration of a stable isotope-labeled internal standard to the sample. The internal standard can be a peptide, a concatemer of peptides, a protein fragment or a full-length protein. The concept to spiking in an isotope-labeled internal standard was first described by Kai and coworkers in 1983, where they generated labeled peptides by digesting a protein in $H_2{}^{18}O$ water [78].

An example of the use of labeled peptides is AQUA (absolute quantification), in which synthetic peptides are added to the protein digest [79]. Commonly, peptides uniquely identifying a protein and likely to be observable in the mass spectrometer with unique transitions in MRM (the so-called 'proteotypic peptides') are used as standards [80]. To facilitate peptides synthesis, lengths below 15 aa are preferred and reactive side chains (cysteine, methionine) should be avoided. A major advantage of the AQUA technique is that post-translationally modified peptides can be synthesized, which can be useful to study signaling pathways [81]. However, one pitfall is that the standard is spiked in at the peptide level, and therefore does not corrected for errors introduced during sample processing and proteolytic digestion. This can lead to underestimation of the actual concentration in the sample. Moreover, incomplete digestion leading to missed cleaved peptides represents another source of error. Since AQUA peptides are commercially available, the method is broadly applied even though the synthetic, heavy peptides are relatively expensive. Owing to economic reasons targeted proteins are often quantified with only a single AQUA peptide, compromising accuracy.

To extend peptide-based absolute quantification methods to samples with very high dynamic range such as plasma, Anderson and coworkers in 2004 introduced a method

termed Stable Isotope Standards and Capture by Anti-Peptide Antibodies, short SISCAPA [82]. As the name implies, antibodies raised against the peptides of interest are used to enrich it along with its corresponding spike-in standard. Besides increasing the sensitivity by enrichment of low-abundant peptides, the MS analysis time is reduced owing to low complexity (because ideally only the peptide of interest needs to be analyzed). Both, AQUA and SISCAPA are typically used in combination with multiple reaction monitoring (MRM), where specific peptide and fragment ions are monitored (also see page 6).

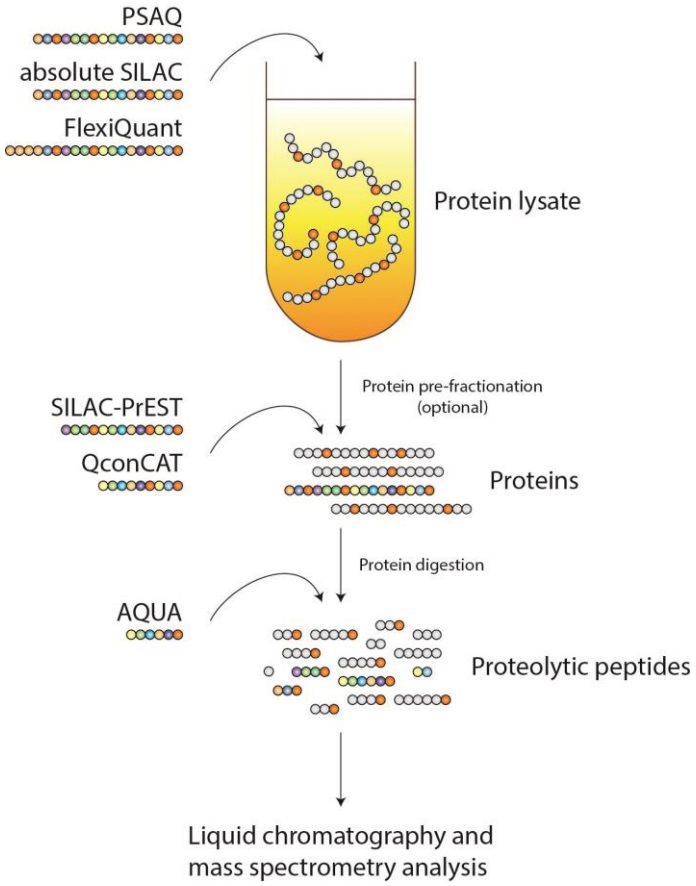


Figure 14: Sample preparation workflows in MS-based absolute quantification of proteins. Various methods are available, each introducing the internal standard at a different step in the workflow. Adapted from [57, 83]

For reliable quantification, it is highly desirable to quantify several peptides of the target protein in order to eliminate the effect of missed cleavages or possible unknown modifications on peptide quantification. The QconCAT (Quantification concatemer)

method is based on an artificial protein that is a concatenation of proteotypic peptides [84, 85]. This enables efficient synthesis of many labeled peptides, which can be used for simultaneous quantification of tens of proteins, each with multiple peptides, thus increasing the confidence in quantification. The peptides within the concatemer are equimolar, restricting quantification to peptides with similar concentrations. To overcome this issue, peptides with similar abundance are grouped together into one QconCAT constructs.

The QconCAT constitutes a synthetic gene, which can be recombinantly expressed and labeled in *Escherichia coli* (*E. coli*). The concatemer is purified with an affinity tag, quantified and then digested together with the sample to yield the internal standard peptides and endogenous peptides, therefore, potential errors introduced during proteolysis are corrected. Nevertheless, proteolysis efficiencies have to be monitored for each peptide because the protease digestion does not happen in the natural sequence context. Ito and coworkers further advanced the QconCAT methodology by surrounding the peptides on both sides with their natural flanking sequence, making cleavage efficiency more similar to the endogenous protein [86].

The gold standard for accurate and reliable quantification is isotope-labeled full-length proteins, since they are the complete equivalent of the target protein, having identical biochemical characteristics. Full-length protein standards can be added directly to cell extracts or body fluids, allowing for extensive sample pre-fractionation or sample depletion [83]. Such pre-fractionation can be especially valuable for samples such as plasma, where it may allow detection of low-abundant proteins. Three methods based on full- length labeled proteins recently established: Protein standard absolute quantification (PSAQ) [87], absolute SILAC [57] and FLEXIQuant [88]. The labeled proteins are typically produced recombinantly in cell-free systems or in *E. coli*. The full-length proteins are highly purified and precisely quantified, commonly using amino acid analysis. The FLEXIQuant method differs in this respect as it uses an *in vitro* synthesized FLEXIQuant standard that is tagged with a FLEX-peptide. Quantification of this standard is achieved

by adding a known amount of AQUA peptide (FLEX-peptide) to the extract only quantifying the standard, therefore extensive purification is not required. The production of full-length protein is laborious and expensive, and therefore only a small number of proteins of interest are targeted for quantification using these methods. Moreover, not all proteins can be produced recombinantly, e.g. membrane proteins represent a particularly difficult class. In summary, an ideal absolute quantification approach would use the native sequence context during digestion, use multiple peptides for quantification, allow multiplexed quantification of proteins and enable a streamlined and economic strategy for standard production.

Label-free quantification methods for relative and absolute quantification

In label-free quantification, as the name implies, no label is introduced to the sample. These strategies are therefore applicable to virtually any sample, including and in particular clinical samples [89]. Samples are prepared and analyzed separately and a measure of protein abundance is assigned to each protein in each sample. In relative, label-free quantification these are compared computationally, therefore allowing for comparison of an unlimited number of samples. Because sample processing, LC and MS analysis are carried out separately, the variation between samples is typically higher than in label-based methods (Figure 11). Consequently, label-free quantification methods require a robust and ideally automated sample handling workflow. Label-free quantification strategies can be divided into two classes: (i) spectral counting and (ii) intensity based methods (Figure 15).

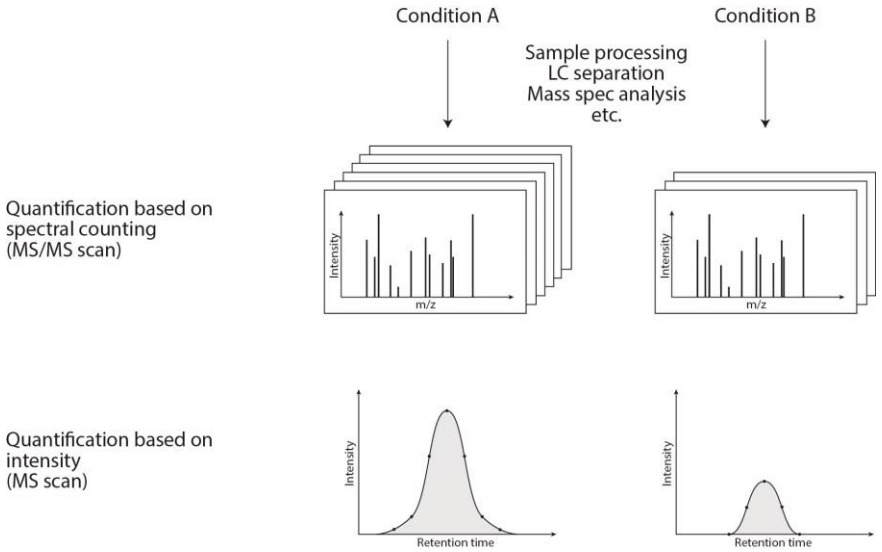


Figure 15: Label-free quantification: the first approach, spectral counting, compares the number of acquired spectra for each peptide, while the second, intensity based approach compares the chromatographic peak of the peptide. Adapted from [89].

In spectral counting, the number of fragmentation spectra identifying a protein are counted and used as a quantitative measure of the protein abundance. It is based on the empirical observation that the number of MS/MS spectra is directly related to the abundance of the peptide, therefore both absolute and relative quantification between different samples is possible [90, 91]. Spectral counting is a simple, intuitive method especially used in combination with low resolution mass spectrometric data. However, parameters such as the length of dynamic exclusion (time that a precursor mass is excluded from being sequenced again), chromatographic peak width and mass spectrometric scan speed can greatly influence the number of spectral counts and need to be tightly controlled. Therefore, to detect small changes in protein abundance to an acceptable precision, a relatively high number of MS/MS spectra is required [47]. Furthermore, protein size affects the number of proteolytic peptides detectable in the mass spectrometer and possibly introducing a bias. To account for the protein size, a 'spectral abundance factor' (SAF) was defined [92] and later further refined into a 'normalized spectral abundance factor' (NSAF) to make different runs comparable by normalizing with the total experiment SAF [93]. Another approach to normalize spectral counting data is termed 'Absolute Protein Expression' (APEX), which additionally

considers the physicochemical properties of the expected proteolytic peptides. This probability of observing a peptide in the mass spectrometer can be predicted using machine learning algorithms on large data sets [80]. Finally, to enable absolute protein quantification APEX relates total spectral counts to the total material loaded [94].

Besides counting spectra, the protein abundance can also be estimated by simple peptide counting. The protein abundance index (PAI) is defined as the number of observed peptides in the experiment divided by the number of theoretically observable peptides for each protein within the mass range of the MS [95]. The exponentially modified PAI (emPAI), an advancement of the PAI, is directly proportional to the protein content in a sample and can therefore be used to roughly estimate absolute protein quantities.

The second class of label-free strategies are based on peptide intensities. The integrated signal response of peptide ions in the mass spectrometer is compared between states, as it is well known that in the MS the peak intensity correlates linearly to the concentration of the peptide within a large concentration range (see for example [96, 97]). The method is applied to shotgun proteomics data, in which peptides are chromatographically separated according to particular physicochemical properties (typically hydrophobicity) and then their peptide ion signals are recorded by the mass spectrometer generating the intensity profile as a function of the retention time (Figure 15). Therefore the accuracy of this method depends on achieving high separation of eluting peaks in the retention time and m/z dimensions. To obtain reliable quantification, the data have to further be corrected for errors introduced during parallel handling, and as well as for run-to-run variations in performance of the chromatography, the mass spectrometer, and the amounts of injected samples. Common post-processing steps in advanced computational proteomics are feature detection, alignment of retention times, normalization of total MS intensities, and matching and transferring identification between measurements to compensate for stochastic peak picking. Within the MaxQuant software label-free quantification is performed via the MaxLFQ algorithm, an algorithm including advanced normalization that is fully compatible with any peptide or protein separation prior to LC-MS analysis

[98]. Even though the computational effort is much higher than in simple spectral counting, intensity-based label-free approaches provide much better precision than spectral counting, in particular for low-abundant proteins since the continuous intensity readout is more informative and less stochastic than single spectra counts.

Various strategies have been developed for absolute quantification of intensity-based label-free experiments. The 'Top3' approach proposed by Silva et al takes the median of the precursor intensities of the three best ionized (most intense) peptides of each protein as a measure of absolute protein abundance. One limitation of this method is that it requires at least three peptides of every protein to be quantified [99]. Another strategy to obtain absolute values, simply uses the sum of all peptides intensities of the protein as a measure for concentration. The summed intensities are divided by the total MS signal of all proteins, and the intensity values scaled either by the estimated total protein amount, or by using reference point derived from internal standards or natural standards such as histone protein content. One of these methods of scaling by estimated total protein amount is called intensity-based absolute quantification (iBAQ). In addition to normalizing the data by the observable peptides, it uses peptide intensities from, for example, commercially available protein standards (such as the Universal Protein Standard, UPS) to scale the data [11]. The newly developed 'proteomic ruler' concept, determines the total MS intensity of histones and makes use of its fixed relationship to the total DNA amount in the cell to deduce the cell number and thereby total protein content per cell [100]. The proteomics ruler method has the advantage that no spike-in is required for scaling, but the approach requires knowledge of parameters like the amount of DNA per cell.

Antibody-based proteomics

So far only mass spectrometry-based proteomics has been described. However, protein analysis in a global manner can also be performed by antibody based methods. Here, I describe the Swedish Human Protein Atlas (HPA) project (directed by Mathias Uhlén) because my thesis work is partly based on a collaboration with this venture. The HPA aims to systematically explore the human proteome using antibody-based proteomics. Within the project they generate two or more independent antibodies for each gene product and use these to image the abundance and distribution of all proteins in cells and tissues. The first version of the publicly available HPA website and database was released in 2005 and included 700 antibodies. After several revisions the database has accumulated 21 984 antibodies targeting 16 621 genes [101, 102]. The high-throughput antigen production, antibody generation and characterization are achieved in an elaborate workflow. In brief, the antigens, referred to as Protein Epitope Signature Tag (PrESTs), are automatically designed with following characteristics: (i) to have high specificity, and therefore they consist of 50 to 150 aa of an unique region representing the target protein (ii) to avoid 'troublesome' regions such as transmembrane regions and signal peptides. To achieve high-throughput production in *E. coli*, the PrESTs are fused to the Albumin Binding Protein (ABP), a solubility tag derived from the streptococcal protein G, which facilitates expression. Next, rabbits are immunized with the PrESTs and the antibodies are affinity purified from the polyclonal antisera using the target PrEST as ligand (Figure 16).

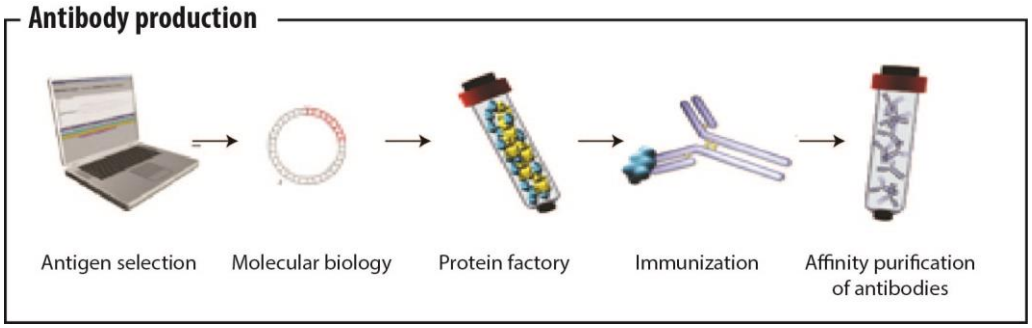


Figure 16: Workflow of antibody production: the antigens are selected computationally, cloned into an expression vector, and this is followed by recombinant protein expression and purification. Rabbits are

immunized with the antigens and eventually the antibody is purified from the antisera using the antigen as ligand.

The antibodies are validated by protein arrays and western blots. This allows exclusion of highly cross-reacting antibodies and ascertains that bands of appropriate molecular weights are recognized. Once quality controlled, the antibodies are used for protein profiling in a multitude of normal and cancerous tissues assembled in tissue microarrays, as well as in cell lines. The comprehensive HPA online database is divided into a tissue atlas, subcellular atlas and cell line atlas and also includes a large amount of high-resolution imaging data. Additionally the HPA performed deep sequencing of the mRNAs (RNA-seq) for some of the cell lines to obtain transcript expression levels.

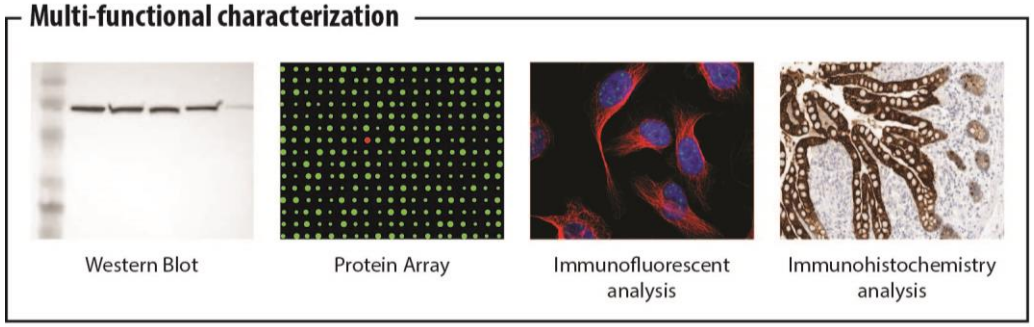


Figure 17: Multi-functional characterization: first the antibody is tested for specificity in a western blot as well as on a protein array, then the protein expression patterns are studies via immunofluorescence and immunohistochemistry analysis

The tissue atlas contains information of protein expression profiles based on immunohistochemistry of 46 normal tissues and 20 cancerous tissues from patients, and each image is annotated by pathologists. The subcellular atlas provides spatial information of protein expression from high-resolution, multicolor images of immunofluorescently stained cells. Besides U2OS cells, two additional cell lines based on the RNA sequencing data are selected for automated confocal microscopy analysis. The cell line atlas contains protein expression data based on immunohistochemistry for 46 distinct cell lines and 10 leukemia cell lines and PBMCs from patients (Figure 17). To summarize, the human protein atlas features millions of high-resolution images showing protein expression profiles in tissues and cell lines and spatial distribution of proteins. A large number of the antibodies have been made available commercially.

Platelets

The main biological application area of the SILAC-PrEST technology in my thesis were in platelet cells, which will be introduced in some detail now. Platelets are, besides erythrocytes and leukocytes, the smallest cells in the blood reaching a diameter of no more than 0.5 μm in mice or 1-2 μm in humans and one of their main functions is keeping the integrity of blood vessels. They are produced in the bone marrow by budding from their precursor cell: a giant, polyploidy cell called the megakaryocyte. Each megakaryocyte can give rise to thousands of platelets [103]. These cells circulate in the blood for 3-4 days in mice or 7-10 days in humans until they are degraded in the reticuloendothelial system of the spleen or liver. Platelets are devoid of a nucleus and DNA, however they share many features with conventional cells. For instance, they contain a cytoskeleton, ribosomes, mitochondria and secretory granules (dense, lysosomal and α-granules) (Figure 18). These granules are secreted upon platelet activation. While dense granules predominantly contain small molecules such as serotonin and ADP, α-granules enclose proteins such as the von Willebrand factor and fibrinogen. Additionally, platelets have an open canalicular system (OCS) that plays a major role in activation by changing the morphological shape from their resting discoid form into their active dendritic structure [104].

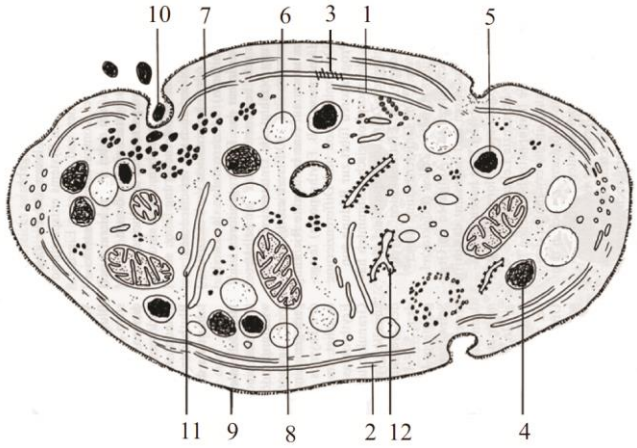


Figure 18: Ultrastructure of platelets: (1) microtubuli (2) actin (3) myosin (4) dense granules (5) α-granules (6) lysosome (7) glycogen (8) mitochondria (9) plasma membrane (10) membrane invagination (11) open canalicular system (12) dense tubular system. Adapted from [105].

Hemostasis

Hemostasis is the process of building a thrombus to stop bleeding. Temporally it is divided into three phase: (1) primary hemostasis is the phase of adhesion and aggregation of platelets, and the consequent formation of a thrombus sealing the lesion; (2) secondary hemostasis is the activation of the blood coagulation cascade leading to formation of a fibrin network that stabilizes the thrombus; and (3) fibrinolysis, where fibrin formation is inhibited and the clot is broken down after the vasculature has been repaired. Since platelets are mainly involved in primary hemostasis, the following section only explains this part of hemostasis.

In normal blood circulation within an intact vasculature, most platelets never adhere or aggregate during their entire life time. However, upon vascular injury subendothelial matrix, which includes proteins such as collagen, von Willebrand factor (VWF), fibronectin and laminin, is exposed. These adhesive proteins of the matrix serve as ligands for platelets in order to adhere and aggregate. In vessels with high shear rates such as small arteries and arterioles, the initial contact between platelets and damage vasculature occurs via VWF and the platelet surface receptor GPIbα. This interaction is not sufficient for stable adhesion, but fast on/off-rates enabling 'rolling' of the platelet on the vessel surface. This rolling then allows binding of the immunoglobulin superfamily receptor GPVI to collagen, which initiates activation. Molecularly, activation is characterized by intracellular signaling to shift integrins to a high-affinity state ('inside-out signaling') as well as by the release of ADP and thromboxane A_2 (TXA_2). In parallel to platelet recruitment, the extracellular matrix (ECM) releases tissue factor (thromboplastin) triggering thrombin formation. Together with ADP and TXA_2, thrombin stimulates G-protein coupled receptors (GPCR), which induce further signaling events and eventually lead to full platelet activation that include shape change and granula secretion [106]. Activated platelets recruit additional platelets by feedback amplification loops thereby inducing aggregation and formation of a hemostatic plug. The aggregation is primarily mediated by αIIbβ3 integrin, which bridges platelets to each other.

Clinical relevance

Normally, platelets smoothly circulate through the arteries and veins surveying the integrity of the vasculature or in case of vascular injuries they participate in physiological hemostasis. Unfortunately, however, they are also involved in pathological thrombus formation for example at sites of atherosclerotic plaque rupture, which can lead to vascular occlusion. Thrombotic occlusion of the coronary or cerebral artery can lead to reduced blood supply to the heart or the brain resulting in acute myocardial infarction or acute ischemic stroke, respectively. In the developed world these disease collectively are the most common cause of death. Therefore antiplatelet therapies have been developed to prevent pathological platelet activation by blocking critical platelets proteins involved in activation and aggregation. Prominent examples are aspirin (inhibition of TXA_2 production), Clopidogrel and Ticlopidine (both ADP receptor antagonist) as well as Abciximab (αIIbβ3 integrin antagonist) [107].

To develop novel antiplatelet therapies that on the one hand prevent thrombosis, but on the other hand avoid the risk of bleeding complications, requires deeper understanding of the molecular mechanism by which platelets become activated. Moreover, besides their main function in maintaining the integrity of the circulatory system by hemostasis, platelets are involved in many processes ranging from combating microbial infection, triggering inflammation and promoting angiogenesis and many others [108], which are far from completely understood. Therefore determination of the platelet proteome in a quantitative way should be very useful to biology and medicine and could reveal new molecular players as well as providing starting points for diverse functional studies.

Proteomics studies of platelets

Traditionally, proteomic studies of the platelet proteome have been performed by two-dimensional gel electrophoresis and together gel-based approaches have identified a few hundred proteins [109]. Concurrently, transcriptomics studies have been performed to predict the proteome [110], however because platelets lack a nucleus and no transcription occurs, the information that is gained by mRNA analysis is not easy to interpret. An initial

proteomics study of human platelets using modern MS instrumentation has already been performed and identified a few thousand proteins [111]. Yet a comprehensive quantitative platelet proteome map was not available for any mammalian species and remedying this situation was one of the goals of this thesis.

Results

Development of the SILAC-PrEST method

In this project I set out to develop an absolute quantification technique, which addresses the limitations of the present absolute quantification methods. To achieve this we joined forces with the Human Protein Atlas (HPA), which has produced a large resource of antigens, which have uniquely features making them a close-to ideal protein standard (Figure 19). First, antigens covering more than 75% of the human protein-coding genes have already been created by the HPA. Second, the 50 to 150 protein specific amino acids are selected to be unique regions of the target proteins with no or low homology to any other protein. They yield on average five MS-detectable tryptic peptides enabling robust and reliable quantification in principle. Third, each PrEST contains a common 120 aa albumin binding protein (ABP) solubility sequence enabling high-throughput quantification of the PrESTs. Importantly the PrEST standards do not need to be purified extensively since only the PrESTs and not the contaminations are quantified via the ABP. An inherent advantage of the ABP-PrEST fusion is that expression and purification of insoluble proteins are facilitated thereby enabling high-throughput production.

Figure 19: Protein epitope signature tag (PrEST) - protein fragment consisting of an affinity tag, solubility tag as well as protein specific region

As a baseline study we applied our quantification strategy to quantify more than 40 different proteins in HeLa cells. This was done with light standards in heavy labeled cell lines as well as heavy standards in non-labeled cells. Using this streamlined method we showed that accurate quantification of any of the targeted proteins was feasible – without many of the limitations of existing methods.

The study was published in Molecular and Cellular Proteomics in 2012:

'A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines'

Marlis Zeiler*, Werner L. Straube*, Emma Lundberg, Mathias Uhlén and Matthias Mann

Mol Cell Proteomics. 2012 Mar;11(3)

A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines*[S]

Marlis Zeiler‡, Werner L. Straube‡**, Emma Lundberg§, Mathias Uhlen§¶, and Matthias Mann‡‖**

Mass spectrometry-based proteomics increasingly relies on relative or absolute quantification. In relative quantification, stable isotope based methods often allow mixing at early stages of sample preparation, whereas for absolute quantification this has generally required recombinant expression of full length, labeled protein standards. Here we make use of a very large library of Protein Epitope Signature Tags (PrESTs) that has been developed in the course of the Human Protein Atlas Project. These PrESTs are expressed recombinantly in *E. coli* and they consist of a short and unique region of the protein of interest as well as purification and solubility tags. We first quantify a highly purified, stable isotope labeling of amino acids in cell culture (SILAC)-labeled version of the solubility tag and use it determine the precise amount of each PrEST by its SILAC ratios. The PrESTs are then spiked into cell lysates and the SILAC ratios of PrEST peptides to peptides from endogenous target proteins yield their cellular quantities. The procedure can readily be multiplexed, as we demonstrate by simultaneously determining the copy number of 40 proteins in HeLa cells. Among the proteins analyzed, the cytoskeletal protein vimentin was found to be most abundant with 20 million copies per cell, while the transcription factor and oncogene FOS only had 6000 copies. Direct quantification of the absolute amount of single proteins is possible via a SILAC experiment in which labeled cell lysate is mixed both with the heavy labeled solubility tag and with the corresponding PrEST. The SILAC-PrEST combination allows accurate and streamlined quantification of the absolute or relative amount of proteins of interest in a wide variety of applications. *Molecular & Cellular Proteomics 11: 10.1074/mcp.O111.009613, 1–13, 2012.*

MS-based proteomics has become a method of choice to study proteins in a global manner (1–3). Mass spectrometry is not inherently quantitative but many methods have been developed to overcome this limitation. Most of them are based on stable isotopes and introduce a mass shifted version of the peptides of interest, which are then quantified by their "heavy" to "light" ratio. Stable isotope labeling is either accomplished by chemical addition of labeled reagents, enzymatic isotope labeling, or metabolic labeling (4–6). Generally, these approaches are used to obtain relative quantitative information on proteome expression levels in a light and a heavy labeled sample. For example, stable isotope labeling by amino acids in cell culture (SILAC)[1] (7, 8) is performed by metabolic incorporation of light or heavy labeled amino acids into the proteome. Labeled proteomes can also be used as internal standards for determining protein levels of a cell or tissue proteome of interest, such as in the spike-in SILAC approach (9).

Absolute quantification is technically more challenging than relative quantification and can only be performed accurately for a single or a small number of proteins at a time (10). Typical applications of absolute quantifications are the determination of cellular copy numbers of proteins (important for systems biology) or the concentration of biomarkers in body fluids (important for medical applications). Furthermore, any precise method of absolute quantification, when performed in more than one sample, also yields the relative amounts of the protein between these samples.

Several methods for absolute quantification have emerged over the last years including absolute quantification (AQUA) (11), quantification concatamer (QConCAT) (12, 13), protein standard absolute quantification (PSAQ) (14), absolute SILAC (15), and FlexiQuant (16). They all quantify the endogenous protein of interest by the heavy to light ratios to a defined

From the ‡Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, D-82152 Martinsried, Germany; §Science for Life Laboratory, Royal Institute of Technology, SE-17165 Stockholm, Sweden; ¶AlbaNova University Center, Royal Institute of Technology, SE-10691 Stockholm, Sweden

⌘ *Author's Choice*—Final version full access.

Received March 15, 2011, and in revised form, September 22, 2011

Published, MCP Papers in Press, September 30, 2011, DOI 10.1074/mcp.O111.009613

[1] The abbreviations used are: SILAC, stable isotope labeling of amino acids in cell culture; AQUA, absolute quantification; PSAQ, protein standard absolute quantification; QconCAT, quantification concatamer; PrEST, protein epitope signature tag; AIF, all ion fragmentation; ABP, albumin binding protein.

amount of the labeled counterpart spiked into the sample and are chiefly distinguished by either spiking in heavy labeled peptides or heavy labeled full length proteins. The AQUA strategy is convenient and streamlined: proteotypic peptides (17) are chemically synthesized with heavy isotopes and spiked in after sample preparation. AQUA peptides are commercially available but currently relatively expensive, especially when many peptides or proteins need to be quantified. More fundamentally, the AQUA strategy suffers from quantification uncertainties that are introduced because of spiking in of the peptide standard after sample preparation and enzymatic proteolysis, which is a late stage in the workflow. Furthermore, any losses of the peptides—for example during storage—would directly influence quantification results. The QconCAT approach is based on artificial proteins that are concatamers of proteotypic peptides. This artificial protein is recombinantly expressed in *Escherichia coli* and spiked into the sample before proteolysis. QconCAT in principle allows efficient production of labeled peptides but does not automatically correct for protein fractionation effects or digestion efficiency in the native proteins *versus* the concatamers. The PSAQ, absolute SILAC and FlexiQuant approaches sidestep these limitations by metabolically labeling full length proteins by heavy versions of the amino acids arginine and lysine. PSAQ and FlexiQuant *in vitro* synthesize full-length proteins in wheat germ extracts or in bacterial cell extract, respectively, whereas absolute SILAC was described with recombinant protein expression in *E. coli*. The protein standard is added at an early stage, such as directly to cell lysate. Consequently, sample fractionation can be performed in parallel and the SILAC protein is digested together with the proteome under investigation. However, these advantages come at the cost of having to produce full length proteins, which limits throughput and generally restricts these methods to soluble proteins.

In this study we advance the absolute SILAC approach by making use of a highly scalable and already established system for protein standard production. We employ short Protein Epitope Signature Tags (PrESTs), which are produced in a high-throughput manner by the Human Protein Atlas project and subsequently used as antigens for antibody production (18–20). The ultimate goal of the Human Protein Atlas is to produce at least two specific antibodies to all human proteins and to use this resource to study the tissue distribution and the subcellular distribution of the human proteome (21–24). PrESTs incorporate a sequence of about 100 amino acids of the target protein chosen for minimal homology to other proteins. Other criteria include avoidance of signal peptide sequences and sequences from transmembrane spanning regions. These PrEST sequences are fused to a 6xHis tag for purification and to a solubility tag derived from the albumin binding domain of the streptococcal protein G (25). We reasoned that these attributes, combined with the fact that more than 30,000 PrESTs, representing 18,300 human genes have already been produced, would make the PrEST library an excellent resource for streamlined, absolute SILAC based quantification of human proteins.

EXPERIMENTAL PROCEDURES

Protein Epitope Signature Tags—The short protein fragments were produced in high-throughput by the Human Protein Atlas where they are used as antigens for antibody production (26, 27). In brief, suitable Protein Epitope Signature Tags (PrESTs) representing unique regions of each target protein were designed using the human genome sequence as template (EnsEMBL). Unique PrESTs with a size between 50 to 150 amino acids and low homology to other human proteins were selected, excluding epitope- and domain-sized similarities to other proteins, signal peptides, and transmembrane regions (26). The cloning, protein expression, and purification were performed as previously described (27, 28). Part of the quality control is that all PrESTs are evaluated and purity verified using SDS-PAGE and the molecular weight is determined by mass spectrometry before further use. This also excludes major "laddering" of the PrESTs. For optimal storage PrESTs were lyophilized and dissolved in 8 M urea and stored at −20 °C until further use. To ascertain that the PrESTs had an endogenous counterpart in HeLa cells, we selected 50 proteins spread over the abundance range of a HeLa proteome that we had measured at a depth of about 4000 proteins. Proteins were picked without regards to specific protein classes, cellular localizations, or functions. Of these 50 proteins, 43 were readily available from the Protein Atlas pipeline in recombinantly expressed form. For multiplexing experiments these 43 PrESTs were mixed together—each at the appropriate concentration. This "master mix" that was then spiked into cell lysates.

Cell Culture—For SILAC labeling, HeLa cells were cultured in Dulbecco's modified Eagle's medium (Invitrogen, Carlsbad, CA) containing 10% dialyzed fetal bovine serum (Invitrogen) and penicillin/streptomycin (Invitrogen). Heavy arginine (high purity Arg10, Cambridge Isotope Laboratories, Andover, MA) and heavy lysine (high purity Lys8, Cambridge Isotope Laboratory) were added to a final concentration of 33 μg/ml or 76 μg/ml, respectively. After six passages cells were fully labeled as assessed by mass spectrometry. Cells were counted using a Countess cell counter (Invitrogen) and aliquots of 10^6 cells were snap frozen and stored at −80 °C.

Protein Expression and Purification of ABP (Albumin Binding Protein)—The expression vector pAff8c (Human Protein Atlas) was modified via SLIC cloning (29) inserting a OneStrep affinity tag to the C terminus of the Albumin Binding Protein (ABP). To express heavy labeled ABP in *E. coli*, an expression strain auxotrophic for arginine and lysine was used (40). Cultures were grown in PA5052 minimal autoinduction media as previously described in (30) but with the addition of 18 normal (light) amino acids and heavy arginine and lysine. Cultures were grown overnight and harvested at an OD600 of about 5.7. *E. coli* cells were lysed in 100 mM Tris, 150 mM NaCl and Protease Inhibitor (Roche) using a Bioruptor (Diagenode, Denville, NY). Cell debris was removed by centrifugation and the supernatant was cleared by filtration through a 22-μm filter. The soluble ABP was purified by affinity chromatography on a StrepTap Hitrap column (GE Healthcare) coupled to an ÄKTA system. The protein was loaded in binding buffer (100 mM Tris, 150 mM NaCl, 1 mM dithiothreitol). After washing with 10 column volumes it was eluted with elution buffer (100 mM Tris, 150 mM NaCl, 1 mM dithiothreitol, 2.5 mM desthiobiotin) (31). The purity of the protein was evaluated by mass spectrometry via an in-solution digest followed by liquid chromatography tandem MS (LC MS/MS). Abundances of ABP and contaminants were estimated by adding the signal for their most intense peptides. ABP was dialyzed in phosphate-buffered saline (PBS), aliquoted, snap-frozen and stored at −80 °C. The concentration of purified ABP was measured by amino acid analysis (Genaxxon BioScience GmbH).

The heavy labeled PrESTs were produced as described above except that the PrEST was expressed and labeled using the auxotrophic *E. coli* strain in media containing Arg10 and Lys8.

Sample Preparation—HeLa cells were lysed in 100 mM Tris, 4% SDS, 100 mM dithiotreitol, incubated for 5 min at 95 °C and disrupted using a Bioruptor. The lysate was cleared by centrifugation through SpinX filters (22 μm, Corning, Corning, NY). The PrESTs were added at appropriate concentrations (see main text) to labeled HeLa cells and the samples were further processed by the FASP method (32). In brief, proteins were captured on a 30-kDa filter and SDS was exchanged with an urea containing buffer. Proteins were alkylated with iodoacetamide and trypsinized (Promega, Charbonnières, France). Further peptide separation was performed using pipette-based six fraction SAX as described (33).

The PrESTs and ABP were mixed and solubilized in denaturation buffer (6 M urea, 2 M thiourea in 10 mM HEPES, pH 8), reduced with dithiotreitol and subsequently alkylated with iodoacetamide. The protein mixture was digested with LysC (Wako) for 3 h, diluted with ammonium bicarbonate and further digested with trypsin overnight. The digestion was stopped by acidifying with trifluoroacetic acid and desalted on C_{18}-Empore disc StageTips (34).

Liquid Chromatography and Mass Spectrometry—Analysis of the PrESTs spiked into HeLa cells was performed on an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) coupled to an Easy nano-HPLC via a nanoelectrospray ion source (Proxeon Biosystems, now Thermo Fisher Scientific). The peptides were separated on a 20 cm fused silica emitter packed in-house with reversed phase material ReproSil-Pur 120 C18-AQ 1.8 μm resin (Dr. Maisch GmbH) and eluted with a 205-min gradient from 5–35% buffer B (80% acetonitrile, 0.5% acetic acid). The mass spectrometer was operated in a data dependent fashion to automatically measure MS and consecutive MS/MS. LTQ-Orbitrap full scan MS spectra (from 100 or 300 to 1650 *m/z*) were acquired with a resolution of 60,000 at *m/z* 400. The ten most abundant ions were sequentially isolated and fragmented using higher energy collisional dissociation (HCD) followed by analysis in the Orbitrap (35).

The ratios of the light PrEST *versus* heavy ABP peptides were analyzed online on the Exactive instrument with HCD option (Thermo Fisher Scientific) using the same nano-HPLC setup as described above. The peptides were eluted with a linear gradient with 5–30% buffer B over 40 min. The Exactive mass spectrometer identified peptides with all ion fragmentation (AIF) by performing alternating MS scans (300–1600 *m/z*) of the precursor ions and all ion fragmentation scans (100–1600 *m/z*) using stepped HCD fragmentation (36). Both scans were acquired at a resolution of 100,000 at *m/z* 200.

The heavy PrESTs *versus* light ABP peptides were analyzed using the TriVersa Nanomate (Advion Biosciences), a chip implementation of nanoelectrospray coupled to a LTQ Orbitrap XL. The samples were eluted in 50% methanol, 0.5% formic acid. A voltage of 1.6 kV and a nitrogen gas pressure of 0.35 psi was applied to spray the peptides into the mass spectrometer. Each sample is sprayed to a single nozzle on the electrospray ionization chip eliminating carryover. A standard data dependent top10 collision-induced dissociation fragmentation method was applied for 2 min acquiring ~40 full scans for quantification (37).

Data Analysis—Acquired data were analyzed with MaxQuant (38) (version 1.2.0.11) using the human IPI database (version 3.68; 87,083 entries). Common contaminants and the sequence of the ABP solubility tag were added to this database. For peptide identification we used Andromeda, a probabilistic search engine incorporated in to the MaxQuant framework (39). Carbamidomethylation of cysteine was included in the search as a fixed modification and methionine oxidation as well as N-terminal acetylation were included as variable modifications. We allowed two miscleavages and required a minimum of six amino acids per identified peptide. The initial mass tolerance for

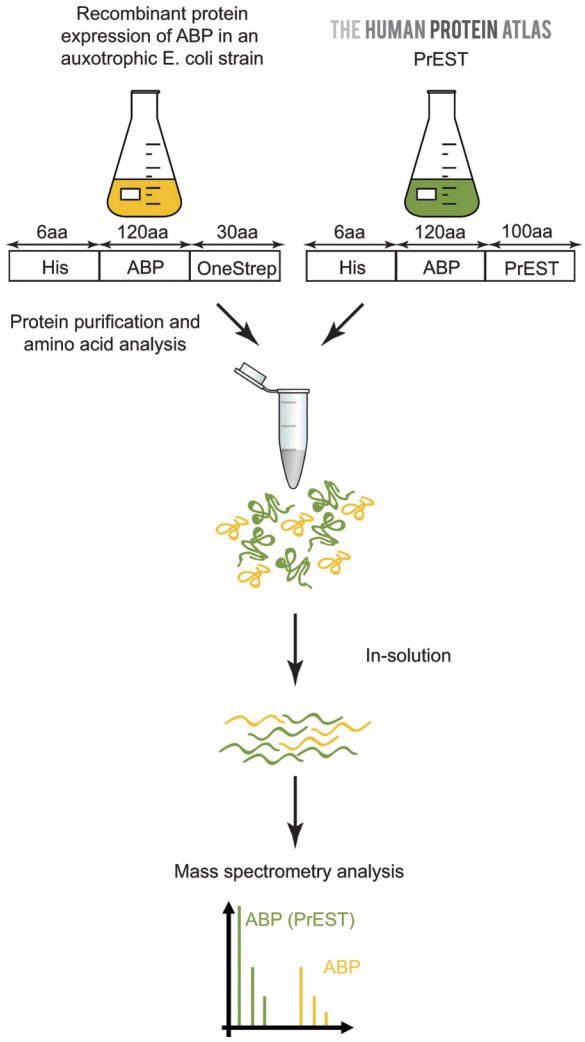


FIG. 1. **Schematic workflow for accurate determination of PrEST concentrations.** Heavy or light ABP is recombinantly expressed in an auxotrophic *E. coli* strain and purified using the C-terminal OneStrep tag. The heavy labeled ABP, whose concentration is measured separately by amino acid analysis, and the PrEST are mixed together and an in-solution digest is performed. Peptides are measured with a short LC MS/MS run on a benchtop mass spectrometer and the PrEST concentration is accurately determined by the SILAC ratio of the ABP peptides originating from the PrEST and the ABP.

precursor ions or fragment ions was set to 6 ppm and fragment masses were allowed to deviate by up to 0.5 Th. For statistical evaluation of the data obtained, the posterior error probability and false discovery rate (FDR) were used. The false discovery rate was determined by searching a reverse database and was set to 0.01 for peptide identification. Additional peptides were identified by the "match between run" option in MaxQuant, which matches precursor masses in a 2-min retention time window (after realignment of the runs) based on the accurate mass.

The AIF data was processed as described above except that up to 50 peaks were analyzed per 100 *m/z* with a tolerance of 15 ppm. The precursor ion mass was matched with the possible fragment ion candidates on the basis of the cosine correlation value of at least 0.6 (36).

Enzyme-linked Immunosorbent Assay—Absolute amount measurements of proto-oncogene c-Fos and Stratifin (14–3-3 σ) was carried

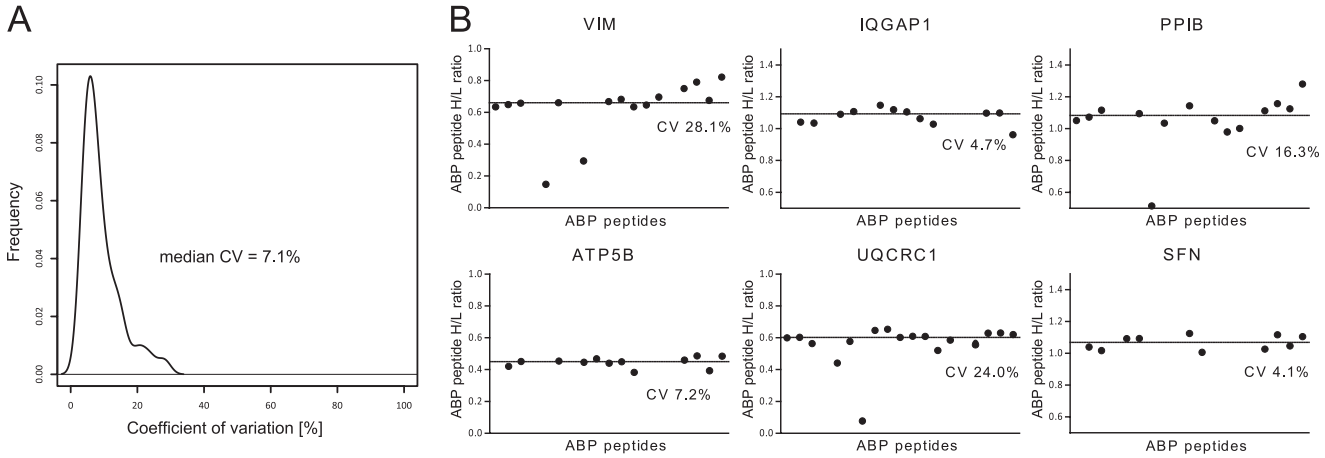


FIG. 2. **Accuracy of ABP quantification.** *A,* Density plot of the overall distribution of the 43 coefficients of variation (CVs) of the ABP peptides measured on a benchtop Exactive mass spectrometer. *B*, Representative example proteins showing the H/L peptide ratios of the ABP peptides deriving from the ABP standard and the ABP peptides in the PrESTs and their coefficients of variation (CVs).

out by ELISA. The kits were purchased from USCNK Life Science and performed according to the manufacturer's instructions. The HeLa cells were lysed in PBS, RIPA 1 (50 mM Tris pH 7.5, 150 mM NaCl, 1% Nonidet P-40) or RIPA 2 (50 mM Tris pH 7.5, 150 mM NaCl, 1% Nonidet P-40; 0.1% SDS) with protease inhibitors. The cells were disrupted by 3 freeze-thaw cycles and sonication using the Biorupter. For the ELISA the samples were diluted 1:10. Fluorescence activity was measured by a microplate reader (Tecan) and converted to actual concentration by a standard curve.

RESULTS

Unlike relative quantification, absolute quantification is a two-step process that requires measurement of firstly the absolute amount of the standard and secondly the relative amount of the standard compared with the analyte of interest. Determination and subsequent control of the level of standard is by no means trivial and can easily be the step that limits the overall accuracy of the approach. Below, we first describe a generic method to determine the absolute amount of each PrEST with high accuracy. Then we construct a "master mix" of different PrESTs and evaluate the ability of the SILAC - PrEST method to accurately quantify cellular proteins. We then apply the master mix to determine the copy numbers of 40 proteins in HeLa cells, a human cervical carcinoma cell line. Finally, we describe an alternative workflow for the quantification of single proteins of interest, in which the two steps are combined into one LC MS/MS analysis.

Accurate Measurement of PrEST Concentrations—Each PrEST is already fused to the ABP, a solubilization tag of 120 amino acids. *In silico* digest of ABP results in 40 tryptic peptides with a length between 6 and 30 amino acids (supplemental Table S1). We recombinantly expressed a heavy SILAC labeled version of the ABP protein tag. When necessary, we used a dual affinity approach based on an N-terminal His-tag and a C-terminal OneStrep tag to generate highly purified protein fragment and to ensure that only full length ABP was obtained. The absolute concentration of ABP protein fragment was determined by amino acid analysis, which is the most accurate method for protein quantification, but which is only applicable to highly purified proteins in relatively large amounts. Heavy SILAC incorporation into ABP was 99% and its purity was about 97% as judged by mass spectrometry (see Experimental Procedures). Because these two factors operate in a compensating direction and because of the small size of the effect, the measured concentration of ABP was not adjusted for them.

LC MS/MS of ABP indeed revealed many readily detectable tryptic peptides (see below). Each of the 43 PrESTs from the Protein Atlas Project was separately mixed with a known amount of labeled ABP as schematically outlined in Fig. 1 to allow for a SILAC LC-MS/MS experiment. As this experiment requires a separate LC MS/MS run for each PrEST it was likely to be rate-limiting for the overall project. We therefore decided to perform this analysis on an economical and robust benchtop Orbitrap instrument rather than on a Velos instrument. The Exactive instrument cannot isolate peptide precursors, therefore we identified the peptides by AIF (36) in 1-h runs.

Typically, at least eight labeled ABP peptides could be quantified against the corresponding ABP peptides from the PrESTs, leading to a median coefficient of variation (CV) of 7% for PrEST quantification (Fig. 2*A*).

To overcome the step of measuring the PrESTs concentration, which limits overall throughput, the heavy PrESTs were measured by static nanoelectrospray on an automated chip-based system (TriVersa Nanomate). This enabled higher throughput measurements of these simple mixtures of ABP peptides using low sample amounts. The peptide ratio showed a median coefficient of variation 5.5%, an improvement over the Exactive based measurement of 7%.

Importantly, a particular PrEST quantification can be repeated at this stage until a desired accuracy is achieved. Here, this was not done, because the accuracy of PrEST

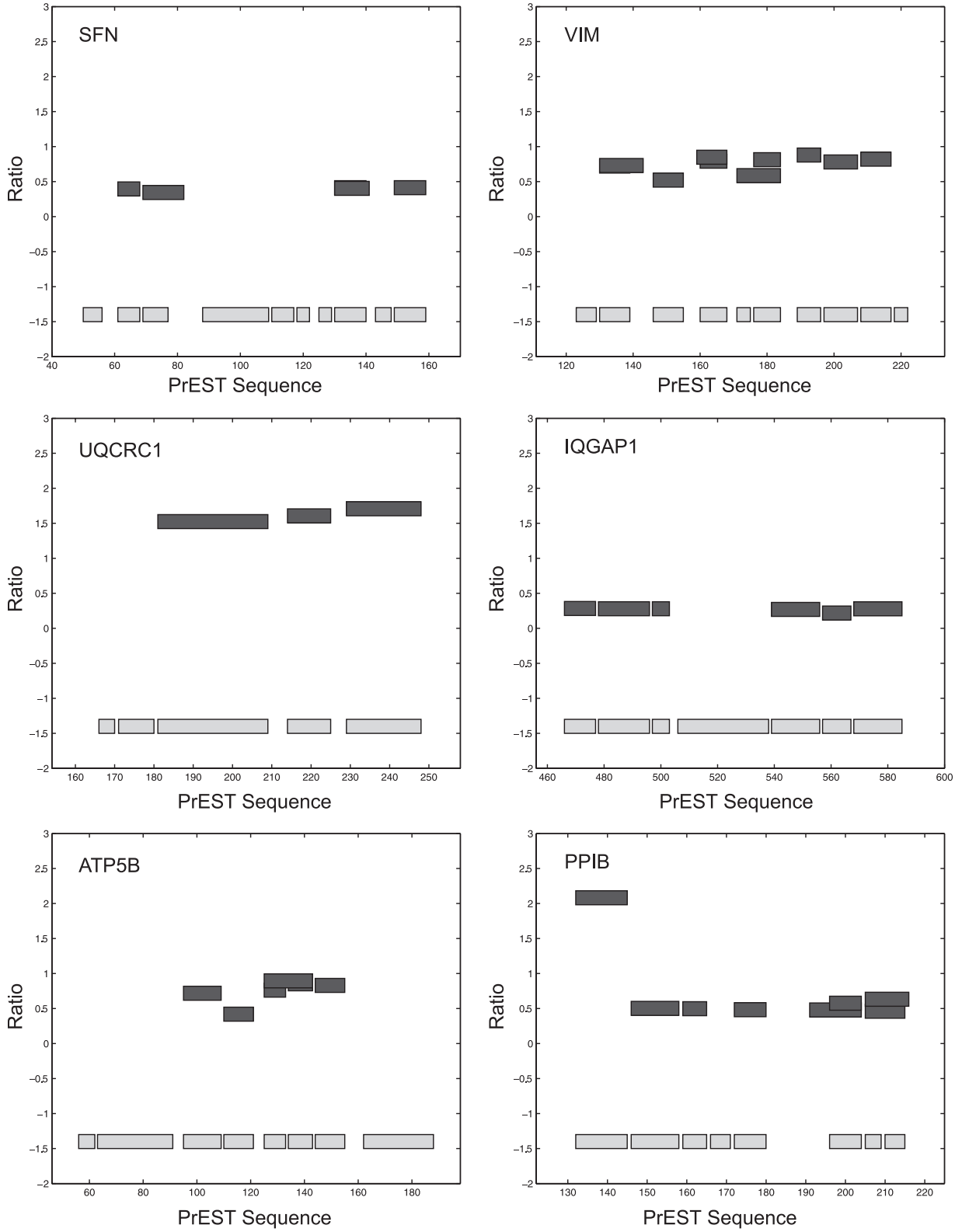


FIG. 3. **Peptide ratios along the PrESTs sequences.** The PrEST master mix was spiked into a lysate of HeLa cells and measured against the endogenous protein. The peptide ratios were extracted to quantify the proteins. The variation of the peptide ratios along the sequence is depicted. Overlapping peptides are because of missed cleavages. The gray bars correspond to the predicted limit tryptic peptides for the PrEST region.

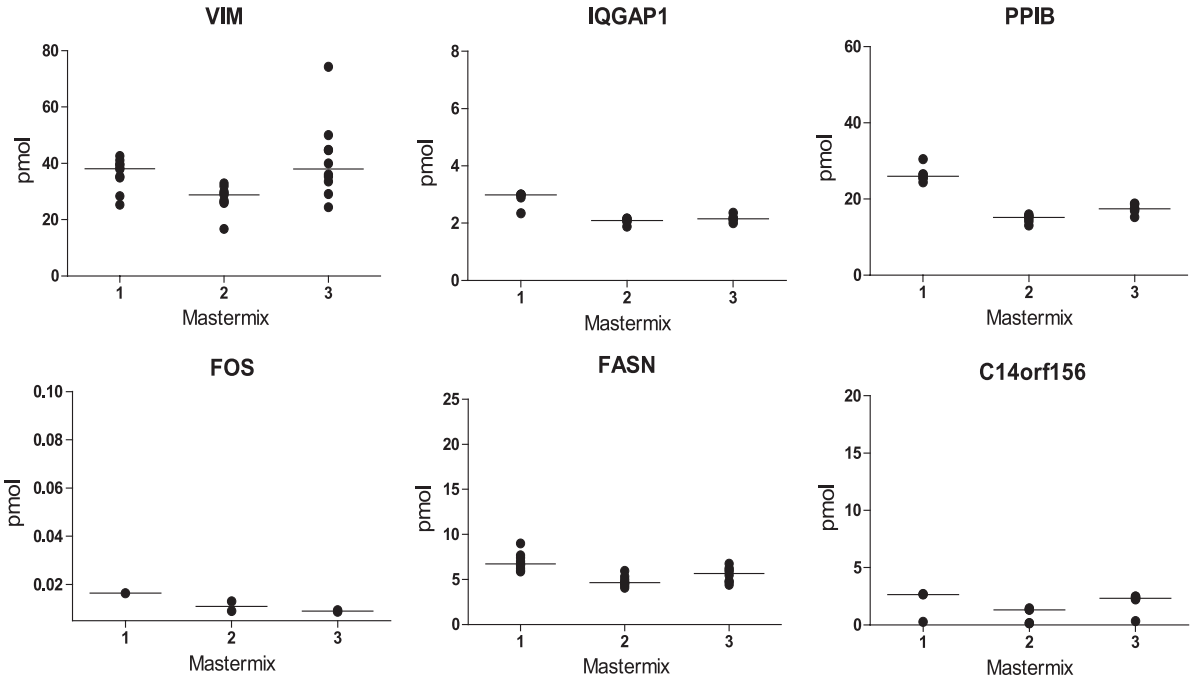


FIG. 4. **Reproducibility of the absolute quantification procedure.** Three independent quantification experiments for representative examples, in which the master mix preparation as well as the PrEST quantification were performed independently. The bars reflect the median of the peptide ratios for each protein.

quantification was estimated to be higher than that of the other steps in the workflow. A few typical examples of results from the PrEST quantification are shown in Fig. 2*B*. The median of the SILAC ABP ratios is used for robustness. This largely eliminates the contribution of outliers such as the ABP peptide (ISEATDGLSDFLKSQTPAEDTVK) in the PrESTs for proteins PPIB and UQCRC1, which had signals very close to noise levels. Note that the quantification accuracy does not depend on the cellular abundance or any other attributes of the target protein, because the same amounts of PrEST is used in each PrEST quantification experiment. Importantly, quantification accuracy in our workflow also does not depend on the purity of the PrEST because our method specifically measures the concentration of PrEST and not of total protein.

PrEST Master Mix and Endogenous Protein Quantification—Having quantified the PrEST amounts we proceeded to measuring protein expression levels in HeLa cells. For convenience we first used unlabeled PrESTs and quantified against heavy SILAC labeled HeLa cells. Because digested total cell lysates consist of hundreds of thousands of tryptic peptides, the addition of a single or even a large number of PrESTs does not change the overall complexity of the mixture. On the basis of the quantitative amounts established above, we here mixed 43 PrESTs together. In initial experiments we used equimolar mixtures of PrESTs, which were spiked into HeLa lysate in different amounts. The measured SILAC ratios established appropriate levels of each PrEST in the master mix, such that the SILAC ratios were within the most accurately quantifiable range, *i.e.* relatively close to one to one.

The master mix with appropriate levels of all the 43 PrESTs was spiked into the lysate of SILAC labeled cells. The mixture was digested according to the FASP protocol followed by SAX fractionation and resulting in six fractions that were separately measured with 4-h gradients on an LTQ Orbitrap Velos mass spectrometer. We were able to quantify 40 of the 43 proteins targeted by our PrEST master mix. Proteins were generally quantified with several PrEST derived peptides (average 3.98 and median 3), leading to an overall median CV of 12% (supplemental Table S2). As an example, the adhesion protein IQGAP1 was quantified with six peptides, which each gave nearly identical quantification results (CV 9.9%). Five of the six quantified tryptic peptides of ATP5B (mitochondrial ATP synthase subunit beta), had very close SILAC ratios, however, one peptide had a ratio that differed by 43% from the median. This peptide is clearly an outlier and its deviating value contributes substantially to the CV value, raising it from 8.7% to 23%. Note however, that we base protein quantification on the median of the peptide values; therefore the outlier peptide hardly contributes to the measured protein expression value and the CV value therefore underestimates the accuracy actually obtained in this experiment. For the same reason modifications of the endogenous proteins in the region covered by the PrEST could cause outlier peptide ratios, which would contribute little to the measured protein ratio (Fig. 3).

To independently assess the precision of this step of absolute protein quantification, we compared the ratios determined from "limit tryptic peptides" (those without internal arginine or lysine) to those determined from the longer ver-

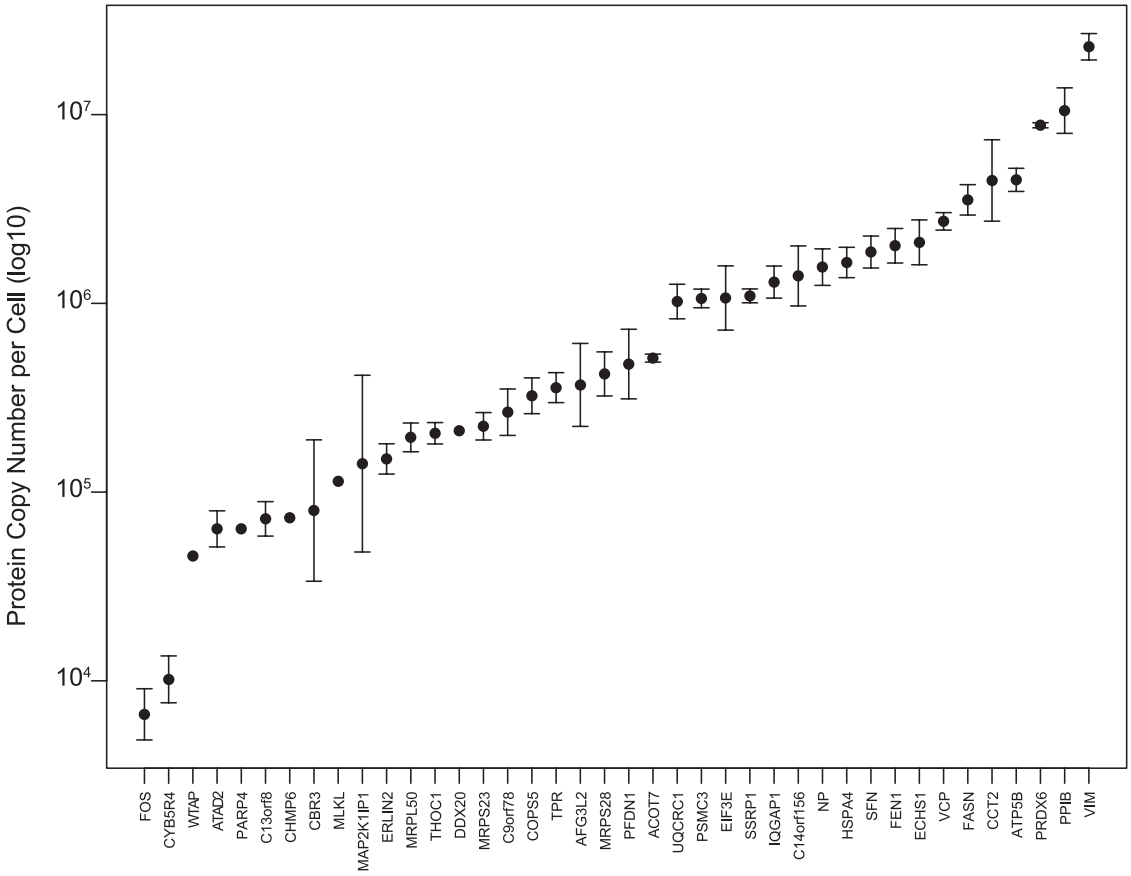


FIG. 5. **Protein copy numbers determined per HeLa cell.** The dot plot shows the protein copy numbers per cell measured in three independent experiments. The error bars correspond to the CVs. Proteins with copy numbers ranging from 6000 to 20,000,000 per cell were quantified (see also Table II).

sions of the peptide containing one or two missed tryptic cleavage sites. These peptides are very problematic for peptide standard based methods such as AQUA, but in our measurements very similar ratios were measured for such peptides. This shows that digestion proceeded identical for PrEST and endogenous protein (Table I). Thus, far from introducing uncertainty, in the SILAC-PrEST approach these peptides can provide additional quantification information.

To assess the degree of variability associated with both steps of the absolute quantification procedure, we repeated the entire workflow two more times, including PrEST quantification and master mix generation as well as measurement of cellular abundance of the target proteins. This analysis showed that the standard errors of the mean associated with all steps together are on average 20%. This value is excellent and to our knowledge the most accurate determination of cellular expression levels reported so far. Even more importantly, the errors of each of the steps in the workflow for each of the proteins are immediately apparent from the individual CVs. Thus all protein expression level measurements can be classified and accepted or discarded according to the confidence of measurements. Fig. 4 displays typical examples of protein expression determination from the triplicate measurements. Comparing the peptide ratio spreads to the variability of the mean protein values revealed that the preparation of the master mix contributed the largest variability whereas errors because of SILAC ratio determination were somewhat lower. Automated preparation of the master mix could therefore lead to further improvements in the future.

Protein Copy Number Determination in HeLa Cells—Next we used the absolute values for protein amounts in our HeLa cell lysate to calculate the corresponding copy numbers in cells. HeLa cells numbers were determined automatically in a cell counter (see Experimental Procedures). Given the known amount of each PrEST and their SILAC ratios with respect to the endogenous proteins we determined the cellular copy numbers of 40 different proteins. Very high accuracy of absolute quantification to within a standard error of 25% was achieved for 35 of 40 proteins (Table II).

Cellular copy numbers are only known for very few proteins and it is therefore interesting to relate these copy numbers to the known functions of the proteins (supplemental Table S3). The cytoskeletal protein vimentin forms intermediate filaments and was the most abundant protein with 20 million copies per cell. At the other extreme, the transcription factor and oncogene FOS is present in about 6000 copies in our HeLa cell

TABLE I

Comparison of limit tryptic peptides and peptides with missed tryptic cleavage sites. Peptides with one or two miscleavages as well as their ratios are depicted. The ratios of the two versions vary on average by 12%, which is in the normal range of variation of peptides derived from one protein

Gene name	Sequence	Missed cleavages	Ratio H/L mastermix (1)	CV (%)	Ratio H/L mastermix (2)	CV (%)	Ratio H/L mastermix (3)	CV (%)
ATP5B	IPVGPETLGR	0	0.85299	8.13	0.84263	16.92	1.1617	8.76
ATP5B	VLDSGAPIK	0	0.76219		0.67515		0.99543	
ATP5B	VLDSGAPIKIPVGPETLGR	1	0.89528		0.95159		1.1652	
CCT2	ILIANTGMDTDK	0	0.47498	39.26	0.37474	48.58	1.0965	–
CCT2	ILIANTGMDTDKIK	1	0.26858		0.1831		–	
CCT2	VAEIEHAEK	0	0.4707	6.37	0.45578	8.24	1.1598	10.43
CCT2	VAEIEHAEKEK	1	0.51512		0.51219		1.3444	
ECHS1	KLFYSTFATDDR	1	–	6.52	0.13372	23.26	–	0.01
ECHS1	LFYSTFATDDR	0	0.16792		0.15756		1.1803	
ECHS1	LFYSTFATDDRK	1	0.18416		0.20966		1.1801	
FASN	QQEQQVPILEK	0	0.73946	4.04	0.69517	2.02	0.8985	1.63
FASN	RQQEQQVPILEK	1	0.69837		0.67562		0.87805	
FEN1	LDPNKYPVPENWLHK	1	0.73377	1.04	0.68	2.58	1.2048	3.81
FEN1	YPVPENWLHK	0	0.72303		0.70528		1.1416	
HSPA4	EDQYDHLDAADMTK	0	0.2685	14.36	0.20351	5.45	0.79843	3.91
HSPA4	NKEDQYDHLDAADMTK	1	0.21899		0.21981		0.84382	
PPIB	DKPLKDVIIADCGK	2	0.47962	12.85	0.41764	14.10	1.0739	14.23
PPIB	DVIIADCGK	0	0.5755		0.51014		1.3142	
PRDX6	ELAILLGMLDPAEK	0	0.77082	14.35	0.61275	4.80	1.2449	1.05
PRDX6	ELAILLGMLDPAEKDEK	1	0.62879		0.65579		1.2636	
PRDX6	VVFVFGPDK	0	0.6815	12.56	0.74577	2.37	1.1822	15.71
PRDX6	VVFVFGPDKK	1	0.81434		0.72122		0.94586	
SFN	YLAEVATGDDK	0	0.41281	1.53	0.38665	5.75	0.98489	15.17
SFN	YLAEVATGDDKK	1	0.404		0.41942		0.79403	
TPR	LESALTELEQLR	0	0.1666	3.45	0.15943	8.07	1.1901	8.15
TPR	LESALTELEQLRK	1	0.17493		0.17872		1.3357	
VCP	DHFEEAMR	0	0.20879	1.21	0.20611	10.27	1.4096	2.44
VCP	RDHFEEAMR	1	0.20524		0.17819		1.459	
VCP	KYEMFAQTLQQSR	1	0.12676	33.99	NaN	–	0.47634	72.12
VCP	YEMFAQTLQQSR	0	0.20698		0.24338		1.4679	
VIM	QVDQLTNDK	0	0.79102	4.92	0.72609	2.57	1.3024	3.52
VIM	RQVDQLTNDK	1	0.84806		0.70013		1.2391	
VIM	EKLQEEMLQR	1	NaN	–	0.77968	7.15	NaN	–
VIM	LQEEMLQR	0	0.87912		0.8627		1.8464	
VIM	ILLAELEQLK	0	0.7214	0.83	0.71178	–	1.3289	49.04
VIM	ILLAELEQLKGQGK	1	0.7299		NaN		2.7399	
VIM	DNLAEDIMR	0	0.813	23.04	0.79552	1.43	1.6492	29.83
VIM	VEVERDNLAEDIMR	1	0.58525		0.77955		1.0746	

sample. As expected, proteins involved in cell signaling are generally expressed at lower values—as an example even the scaffolding factor mitogen-activated protein kinase scaffold protein 1 (MAP2K1IP1) is present at only 140,000 copies. However, ubiquitous signaling factors with a general chaperone-like role—such as 14-3-3 isoforms—are very highly expressed (14-3-3 sigma; 1.8 million copies). Two members of the mitochondrial ribosome have about 200,000 copies in this cell line (S23 and S28), whereas a third (L50) has about 400,000 (Note that not all ribosomal protein subunits have equal stoichiometry). The mitochondrial genome only encodes 13 genes therefore it is perhaps surprising that proteins involved in their translation are needed in such high copy numbers. A member of the respiratory chain, ATP5B, has about 4.5 million copies per HeLa cells—about fivefold higher than PSMC3, a regulatory component of the proteasome. The T-complex is a member of a chaperone system and as expected it has a very high copy number (about 4.5 million). Fatty acid synthase, a classical enzyme, is expressed at 3.5 million copies, whereas another enzyme acyl coenzyme A thioester hydrolase (ACOT7) is expressed about sevenfold lower (500,000 copies). Such expression numbers could be interesting for modeling metabolic pathways. These are anecdotal examples but they illustrate that knowledge of the absolute expression levels of cellular proteins can contribute to the understanding of their roles in the cell.

Absolute Quantification Using Heavy PrESTs—Above we used already expressed and purified PrESTs and quantified

TABLE II
Protein copy numbers per HeLa cell

Protein names	Gene name	Median	RSD (%)[a]	Mastermix 1	Mastermix 2	Mastermix 3
14-3-3 protein sigma	SFN	1,870,568	19.81	2,364,005	1,870,568	1,604,145
26S protease regulatory subunit 6A	PSMC3	1,062,048	11.37	1,062,048	950,200	1,192,875
28S ribosomal protein S23, mitochondrial	MRPS23	223,198	17.26	223,198	203,672	282,020
28S ribosomal protein S35, mitochondrial	MRPS28	422,825	24.80	473,409	284,783	422,825
39S ribosomal protein L50, mitochondrial	MRPL50	194,935	18.14	177,937	250,001	194,935
AFG3-like protein 2	AFG3L2	369,737	41.68	369,737	412,509	165,983
ATP synthase subunit beta, mitochondrial	ATP5B	4,511,967	14.68	5,672,473	4,376,424	4,511,967
ATPase family AAA domain-containing protein 2	ATAD2	63,835	23.40	63,835	61,373	91,846
Carbonyl reductase [NADPH] 3	CBR3	79,823	94.26	79,823	61,399	322,454
Charged multivesicular body protein 6	CHMP6	83,028	67.19	122,476	43,581	-
Coiled-coil domain-containing protein 55	CCDC55	–[b]	–	–	–	–
COP9 signalosome complex subunit 5	COPS5	323,791	22.62	323,791	284,218	435,937
Cytochrome b5 reductase 4	CYB5R4	10,180	30.80	16,205	10,180	9,515
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	1,022,450	19.50	1,022,450	713,318	1,025,854
Cytosolic acyl coenzyme A thioester hydrolase	ACOT7	512,746	4.79	512,746	472,208	514,556
Endoplasmic reticulum lipid raft-associated protein 2	ERLIN2	149,867	19.53	206,262	148,785	149,867
Enoyl-CoA hydratase, mitochondrial	ECHS1	2,105,336	28.10	2,965,394	1,723,133	2,105,336
Eukaryotic translation initiation factor 3 subunit 6	EIF3E	1,067,627	34.63	1,067,627	599,306	1,253,469
FACT complex subunit SSRP1	SSRP1	1,095,695	8.52	1,095,695	1,022,209	1,209,724
Fatty acid synthase	FASN	3,536,145	17.98	4,043,129	2,804,853	3,536,145
Flap endonuclease 1	FEN1	2,019,699	20.42	2,372,346	2,019,699	1,563,785
Heat shock 70 kDa protein 4	HSPA4	1,646,549	19.22	2,146,713	1,499,858	1,646,549
Hepatocellular carcinoma-associated antigen 59	C9orf78	265,003	25.76	289,516	171,397	265,003
Lysophosphatidylcholine acyltransferase 1	AYTL2	–	–	–	–	–
Mitogen-activated protein kinase scaffold protein 1	MAP2K1IP1	141,520	68.85	182,796	27,116	141,520
Mixed lineage kinase domain-like protein	MLKL	114,801	17.14	128,711	–	100,891
Nucleoprotein TPR	TPR	357,637	17.53	397,408	278,736	357,637
Peptidyl-prolyl cis-trans isomerase B	PPIB	10,502,199	29.14	15,610,836	9,112,850	10,502,199
Peroxiredoxin 6	PRDX6	8,781,079	3.07	8,881,373	8,377,838	8,781,079
Poly [ADP-ribose] polymerase 4	PARP4	63,971	7.07	60,775	67,168	–
Prefoldin subunit 1	PFDN1	476,849	36.22	476,849	523,643	243,332
Pre-mRNA-splicing regulator WTAP	WTAP	49,143	51.10	31,385	–	66,902
Probable ATP-dependent RNA helicase DDX20	DDX20	213,466	19.17	242,403	184,529	–
Proto-oncogene c-Fos	FOS	6,643	32.41	9,956	6,643	5,359
Purine nucleoside phosphorylase	NP	1,555,814	23.04	2,101,680	1,357,920	1,555,814
Ras GTPase-activating-like protein IQGAP1	IQGAP1	1,296,511	20.65	1,796,903	1,260,937	1,296,511
SRA stem-loop-interacting RNA-binding protein, mitochondrial	C14orf156	1,397,500	32.95	1,665,787	828,707	1,397,500
T-complex protein 1 subunit beta	CCT2	4,479,130	48.47	7,447,762	2,757,533	4,479,130
THO complex subunit 1	THOC1	204,962	13.16	239,173	184,576	204,962
Transitional endoplasmic reticulum ATPase	VCP	2,719,254	10.44	2,719,254	2,358,278	2,904,468
Uncharacterized protein C1orf65	C1orf65	–	–	–	–	–
Vimentin	VIM	22,886,339	15.22	22,974,646	17,376,010	22,886,339
Zinc finger protein 828	C13orf8	72,135	19.47	74,281	51,084	72,135

[a] Standard error of the mean (S.E.) for the three replicates in percent.
[b] No valid data obtained.

against heavy ABP protein and heavy SILAC-labeled cell lysate. Although convenient to determine copy numbers in cell lines, in other applications it would be more appropriate to express heavy labeled PrESTs, which can then be mixed into any proteome of choice—including tissue and clinical body fluid samples. To apply our absolute quantification approach to nonlabeled samples we expressed 28 of the PrESTs in heavy SILAC labeled *E. coli*, purified them and prepared a heavy master mix. To streamline quantification of PrEST levels, we developed an automated set up employing static nanoelectrospray (Advion NanoMate; see EXPERIMENTAL PROCEDURES). As expected, spiking the heavy master mix into normal, non-SILAC labeled cells allowed equally straightforward quantification of the targeted proteins, with good correlation to the previous experiment (Fig. 6).

Absolute Quantification in Single Experiments—We also wished to develop a variation on the SILAC-PrEST strategy to quantify single protein target. In this case, the two experimental steps involved in absolute protein quantification can be collapsed into one as outlined schematically in Fig. 7*A*. A

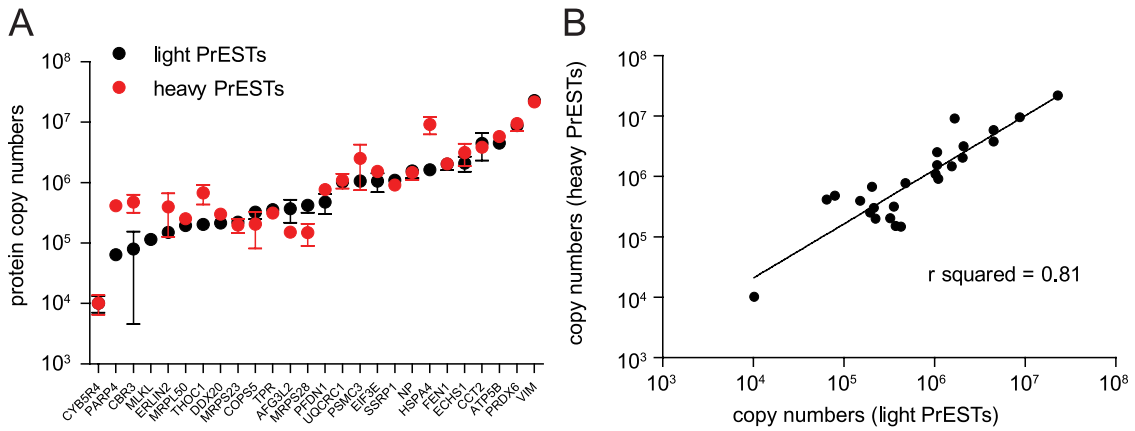


FIG. 6. **Absolute Quantification using heavy PrESTs.** *A*, Comparison of copy numbers obtained by quantifying light PrESTs against SILAC labeled heavy cell lysate (black symbols) *versus* quantifying heavy PrESTs against unlabeled cell lysate (red symbols). *B*, Values shown in *A* but plotted as a scatter graph.

precisely known amount of the ABP solubility tag is mixed into cell lysate together with the labeled PrEST. LC-MS/MS analysis of the sample then provides SILAC ratios of light ABP solubility tag to labeled PrEST ABP peptides. These ratios accurately quantify the amount of PrEST that was used. The same LC MS data also contain the ratios of labeled PrEST peptides to the unlabeled endogenous protein counterpart. Together, these ratios quantify the absolute amount of endogenous protein in a single experiment (Fig. 7*B*). Note that triple-SILAC labeling is not required in this approach because the ratios are determined against different regions of the PrEST construct, namely the common ABP solubility tag region (for quantifying the PrEST) and the protein specific PrEST region (for quantifying the endogenous protein).

This single-plex method for quantification was performed for three different HeLa proteins in which the SILAC-labeled cell lysate and SILAC-labeled ABP was quantified against unlabeled PrESTs. As shown in Fig. 7*C*, consistent values were obtained in these measurements based on triplicate experiments. The absolute levels generally agreed well with the copy numbers determined independently in the multiplexed PrEST-SILAC experiment described above (maximum difference between the means of 40%), validating both approaches.

Enzyme-linked Immunosorbent Assay—ELISA is a standard method in biochemical research to determine absolute amounts, or at least to reproducibly determine protein levels. We therefore compared the SILAC-PrEST method to this established technology. When performing the ELISA assay for Stratifin (14-3-3 σ) under typical conditions—filtered cell lysate and PBS as recommended by the manufacturer—the ELISA recorded less than 20% of the amount quantified by MS. (Note that there is no interference by 14-3-3 isoforms because these peptides are different.) The recommendation of the manufacturer to dissolve in PBS could not solubilize the pellet. The solubility was increased by adding the nonionic detergent Nonidet P-40, which was able to dissolve most of the sample pellet. Adding a low concentration of sodium dodecyl sulfate (SDS), an anionic detergent, further improvement significantly increased measured protein amount (Fig. 8*A*). Still the absolute amounts were underestimated twofold compared with mass spectrometry analysis, presumably because the FASP protocol enables complete solubilization by the use of 4% SDS.

We also investigated the levels of the transcription factor and proto-oncogene FOS by ELISA, the lowest abundance protein quantified in our mix. Here solubilization did not appear to be an issue and we received excellent agreement between quantitative values determined by MS and by ELISA using different buffer conditions (Fig. 8*B*).

DISCUSSION AND OUTLOOK

Here we have developed methods to determine the absolute levels of proteins in cells by taking advantage of the absolute SILAC concept as well as the availability of a large library of PrEST protein fragments. PrESTs already contain an ABP solubility tag, which facilitates recombinant expression of PrESTs against a wide variety of cellular targets. We found that this solubility tag is an excellent "quantification tag" because it generates a large number of readily quantifiable peptides upon tryptic digestion. After producing a highly purified and accurately quantified "gold standard" of the ABP tag alone, it can be used to quantify all PrESTs in turn. Importantly, the purity of the PrEST is not a concern because quantification is only performed against the PrEST component and not against possible *E. coli* or other contaminants.

PrESTs have already been produced against 80% of the human proteome and an "industrialized pipeline" for their production is in place. Although almost all PrESTs so far have been produced for human target proteins, they could in principle be made for any other species in exactly the same way. Furthermore, in many cases more than one PrEST has been made for the same protein to allow for the generation of paired antibodies and pair-wise validation of antibody staining patterns (23). Like-

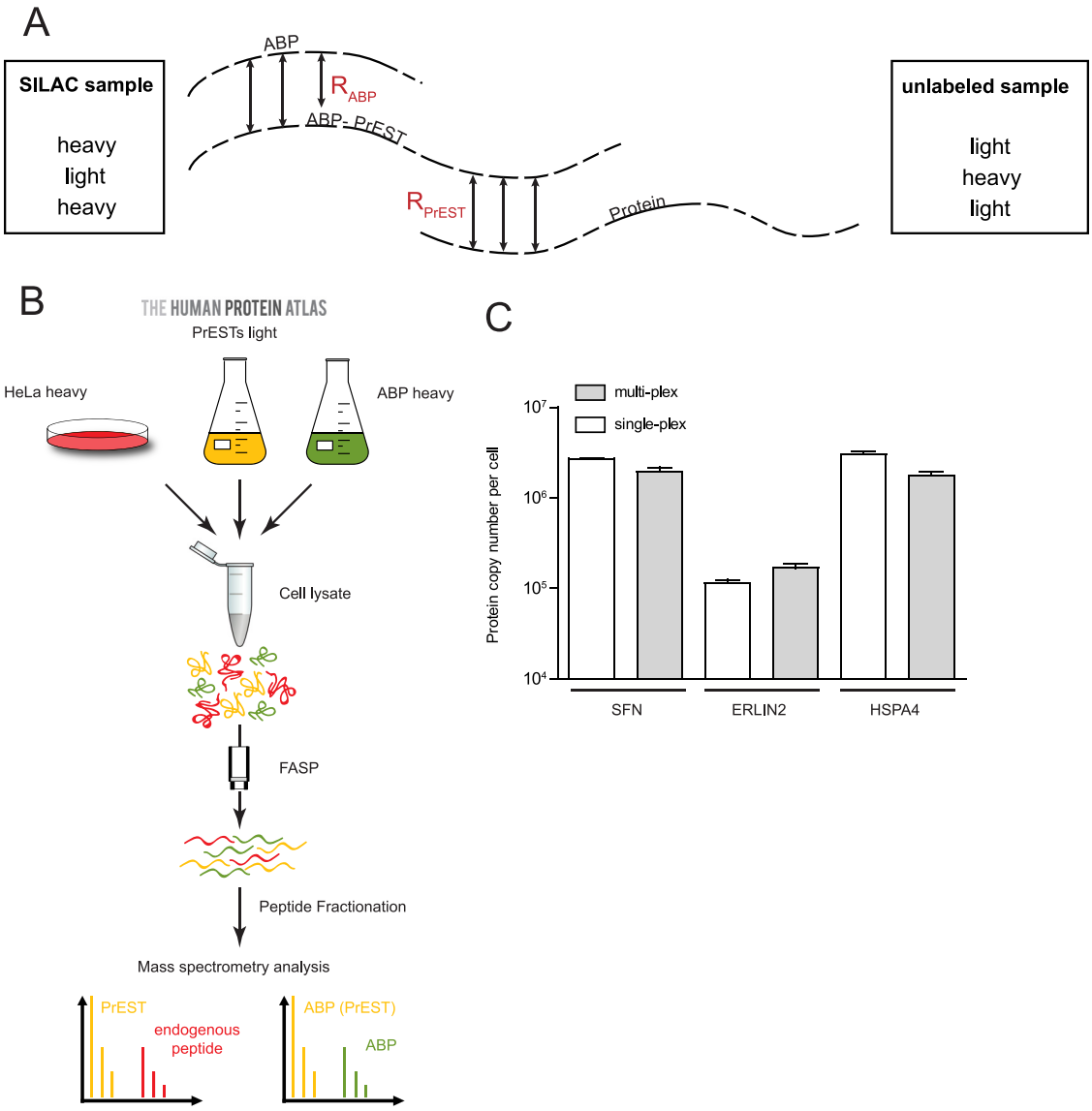


FIG. 7. **Direct quantification of a single protein in HeLa cell lysate.** *A*, Principle of the 'single-plex' strategy for the direct quantification of a single protein. In the same experiment, SILAC peptide ratios mapping to the ABP quantification tag determine the amount of PrEST whereas SILAC ratios mapping to the protein specific region of the PrEST construct determine the level of the endogenous proteins. The experiment can be performed with SILAC heavy labeled cells, unlabeled PrEST construct and heavy labeled ABP tag (*left* side) or *vice versa* (*right* side). *B*, Single-plex determination of absolute protein amount. In the workflow depicted here, an unlabeled PrEST construct as well as a heavy labeled ABP tag are both spiked into HeLa cell lysate before digestion. *C*, Comparison of copy numbers obtained from the "master mix" experiment with those from the single-plex experiments for three different proteins. Error bars are standard deviations of the mean from triplicate measurements.

wise, different PrESTs could be produced in cases where the current ones are not optimal for MS-based quantification.

For preparation of a SILAC-PrEST mixture we estimated the appropriate amount of PrESTs in a two-step procedure. In the future it may be simpler to estimate the amount of protein roughly from the peptide signals from in-depth proteome experiments. PrESTs spiked in at corresponding amounts will likely be in the easily quantifiable range in most cases. The overall accuracy of the SILAC-PrEST approach can be monitored for each step in the procedure and it currently appears to be limited by manual pipetting accuracy. It is likely that precision of all steps in the procedure can be improved significantly in the future.

Here we have demonstrated applications of SILAC-PrESTs for copy number determination in cell lines. However, the principle should be applicable in the same way to absolutely quantify proteins from any source and we plan to investigate this shortly.

Acknowledgments—We thank Shubin Ren for support with bioinformatic analysis, Christoph Schaab for providing Fig. 3, Sabine Suppmann for technical support and Tami Geiger for helpful discussions.

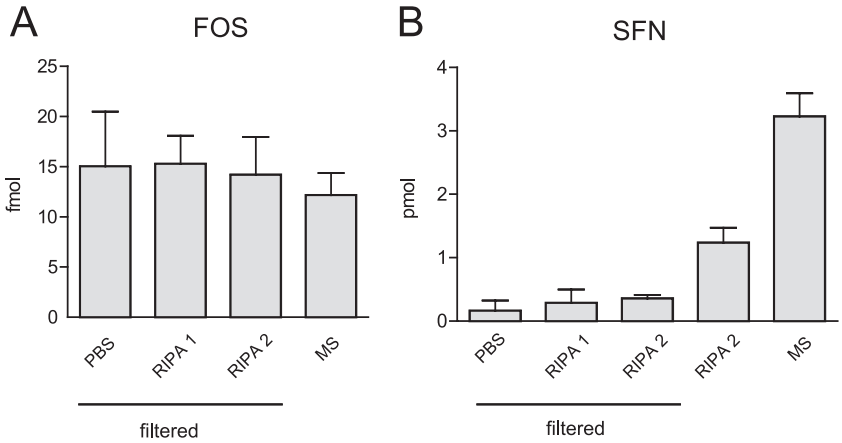


FIG. 8. **Comparison of SILAC-PrEST based quantification and ELISA.** Proto-oncogene c-Fos (*A*) and Stratifin (*B*) were quantified by ELISA to evaluate the SILAC-PrEST absolute quantification. Different ELISA compatible buffers and filtered *versus* unfiltered cell lysates were compared.

The *E. coli* strain auxotrophic for arginine and lysine was a kind gift from Ron Hay, University of Dundee.

Data availability - The acquired raw data was uploaded to Tranche (https://proteomecommons.org/tranche/) as "PrEST-SILAC absolute quantification" and is available via following hash code: HI8Dw7I jHuMUVLLXhNxbD/sJPWqKUuxUVCqa7ZncEXr/f30ksWIJPzy0/PKZ 2EyTgzRP6dU5E3BeeKy+1sUOI2ad0AAAAAAACP6Q==.

* This work was supported by the European Commission's 7th Framework Program PROSPECTS (HEALTH-F4-2008–201648, the Munich Center for Integrated Protein Science (CIPSM) and the Knut and Alice Wallenberg Foundation.

** Both authors made equal contributions to this work.

[S] This article contains supplemental Tables S1 to S4.

‖ To whom correspondence should be addressed: Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, D-82152 Martinsried, Germany. Tel.: + 49-89-8578-2557; E-mail: mmann@biochem.mpg.de.

REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
2. Cravatt, B. F., Simon, G. M., and Yates, J. R., 3rd (2007) The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991–1000
3. Gstaiger, M., and Aebersold, R. (2009) Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat. Rev. Genet.* **10,** 617–627
4. Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1,** 252–262
5. Bachi, A., and Bonaldi, T. (2008) Quantitative proteomics as a new piece of the systems biology puzzle. *J. Proteomics* **71,** 357–367
6. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389,** 1017–1031
7. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386
8. Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7,** 952–958
9. Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., and Mann, M. (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* **6,** 147–157
10. Brun, V., Masselon, C., Garin, J., and Dupuis, A. (2009) Isotope dilution strategies for absolute quantitative proteomics. *J. Proteomics* **72,** 740–749
11. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6940–6945
12. Beynon, R. J., Doherty, M. K., Pratt, J. M., and Gaskell, S. J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2,** 587–589
13. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., and Beynon, R. J. (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **1,** 1029–1043
14. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell. Proteomics* **6,** 2139–2149
15. Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. (2008) Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J. Proteome Res.* **7,** 1118–1130
16. Singh, S., Springer, M., Steen, J., Kirschner, M. W., and Steen, H. (2009) FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J. Proteome Res.* **8,** 2201–2210
17. Kuster, B., Schirle, M., Mallick, P., and Aebersold, R. (2005) Scoring proteomes with proteotypic peptide probes. *Nat. Rev. Mol. Cell. Biol.* **6,** 577–583
18. Uhlén, M., Björling, E., Agaton, C., Szigyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Sköllermo, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4,** 1920–1932
19. Nilsson, P., Paavilainen, L., Larsson, K., Odling, J., Sundberg, M., Andersson, A. C., Kampf, C., Persson, A., Al-Khalili Szigyarto, C., Ottosson, J., Bjorling, E., Hober, S., Wernérus, H., Wester, K., Pontén, F., and Uhlen, M. (2005) Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5,** 4327–4337
20. Berglund, L., Björling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szigyarto, C. A., Persson, A., Ottosson, J., Wernérus, H., Nilsson, P., Lundberg, E., Sivertsson, A., Navani, S., Wester, K., Kampf, C., Hober, S., Pontén, F., and Uhlén, M. (2008) A genecentric Human Protein Atlas for expression

profiles based on antibodies. *Mol. Cell. Proteomics* **7,** 2019–2027
21. Pontén, F., Gry, M., Fagerberg, L., Lundberg, E., Asplund, A., Berglund, L., Oksvold, P., Björling, E., Hober, S., Kampf, C., Navani, S., Nilsson, P., Ottosson, J., Persson, A., Wernérus, H., Wester, K., and Uhlén, M. (2009) A global view of protein expression in human cells, tissues, and organs. *Mol. Syst. Biol.* **5,** 337
22. Lundberg, E., and Uhlen, M. (2010) Creation of an antibody-based subcellular protein atlas. *Proteomics* **10,** 3984–3996
23. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Björling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28,** 1248–1250
24. Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., Algenäs, C., Lundeberg, J., Mann, M., and Uhlen, M. (2010) Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **6,** 450
25. Nygren, P. A., Eliasson, M., Abrahmsén, L., Uhlén, M., and Palmcrantz, E. (1988) Analysis and use of the serum albumin binding domains of streptococcal protein G. *J Mol Recognit* **1,** 69–74
26. Berglund, L., Björling, E., Jonasson, K., Rockberg, J., Fagerberg, L., Al-Khalili Szigyarto, C., Sivertsson, A., and Uhlén, M. (2008) A whole-genome bioinformatics approach to selection of antigens for systematic antibody generation. *Proteomics* **8,** 2832–2839
27. Larsson, M., Graslund, S., Yuan, L., Brundell, E., Uhlén, M., Höög, C., and Ståhl, S. (2000) High-throughput protein expression of cDNA products as a tool in functional genomics. *J. Biotechnol.* **80,** 143–157
28. Agaton, C., Galli, J., Höidén Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., Wurtz, D., Höög, C., Lundeberg, J., Ståhl, S., Pontén, F., and Uhlén, M. (2003) Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell. Proteomics* **2,** 405–414
29. Li, M. Z., and Elledge, S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4,** 251–256
30. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purification* **41,** 207–234
31. Skerra, A., and Schmidt, T. G. (2000) Use of the Strep-Tag and streptavidin for detection and purification of recombinant proteins. *Methods Enzymol.* **326,** 271–304
32. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
33. Wiśniewski, J. R., Zougman, A., and Mann, M. (2009) Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8,** 5674–5678
34. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670
35. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8,** 2759–2769
36. Geiger, T., Cox, J., and Mann, M. (2010) Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol. Cell. Proteomics* **9,** 2252–2261
37. Lu, A., Waanders, L. F., Almeida, R., Li, G., Allen, M., Cox, J., Olsen, J. V., Bonaldi, T., and Mann, M. (2007) Nanoelectrospray peptide mapping revisited: Composite survey spectra allow high dynamic range protein characterization without LCMS on an orbitrap mass spectrometer. *Int. J. Mass Spectrom.* **268,** 158–167
38. 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
39. 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
40. Matic, I., Jaffray, E. G., Oxenham, S. K., Groves, M. J., Barratt, C. L., Tauro, S., Stanley-Wall, N. R., and Hay, R. T. (2011) Absolute SILAC-Compatible Expression Strain Allows Sumo-2 Copy Number Determination in Clinical Samples. *Journal of proteome research* **10,** 4869–4875

Development of the immuno-SILAC method

Having PrESTs and corresponding antibodies in hand it was attractive to combine both to develop a very sensitive and fast quantification method for a number of proteins of interest by using the antibodies for enrichment of the peptides from complex samples. In particular, clinical samples such as plasma are highly complex and can have a very large concentration difference between abundant proteins and clinical relevant marker proteins. In some respects the method conceptually resembles other affinity enrichment methods such as SISCAPA [82, 112], however it also has major differences and advantages. Since polyclonal antibodies, raised against protein fragments (PrESTs) are used, multiple peptides can be enriched for the same PrEST target. In fact epitope mapping detected on average three epitopes per antibody. However, we also discovered that sometimes the epitopes contain a cleavage site making the corresponding peptides inaccessible for enrichment. The study clearly highlights one of the main advantages of using PrESTs compared to methods like SISCAPA: Because the standard is spiked in prior to digestion, differences in proteolytic efficiency between standard and sample and in particular missed cleavages do not affect quantification detrimentally. By multiplexing different antibodies for the enrichment step, many proteins can be quantified in parallel in a short time – demonstrated here with relatively short 15 min gradients. In summary, this study successfully established the immuno-SILAC concept in HeLa cells and compared it to the SILAC-PrEST method.

The study was published in Molecular and Cellular Proteomics in 2014:

'Immuno-proteomics using polyclonal antibodies and stable isotope labeled affinity-purified recombinant proteins'

Fredrik Edfors*, Tove Boström*, Björn Forsström, **Marlis Zeiler**, Henrik Johansson, Emma Lundberg, Sophia Hober, Janne Lehtiö, Matthias Mann and Mathias Uhlén

Mol Cell Proteomics. 2014 Jun;13(6)

Immunoproteomics Using Polyclonal Antibodies and Stable Isotope–labeled Affinity-purified Recombinant Proteins*[S]

Fredrik Edfors‡§, Tove Boström§¶, Björn Forsström‡, Marlis Zeiler‖, Henrik Johansson, Emma Lundberg‡, Sophia Hober¶, Janne Lehtiö**, Matthias Mann‖, and Mathias Uhlen‡¶‡‡**

The combination of immuno-based methods and mass spectrometry detection has great potential in the field of quantitative proteomics. Here, we describe a new method (immuno-SILAC) for the absolute quantification of proteins in complex samples based on polyclonal antibodies and stable isotope–labeled recombinant protein fragments to allow affinity enrichment prior to mass spectrometry analysis and accurate quantification. We took advantage of the antibody resources publicly available from the Human Protein Atlas project covering more than 80% of all human protein-coding genes. Epitope mapping revealed that a majority of the polyclonal antibodies recognized multiple linear epitopes, and based on these results, a semi-automated method was developed for peptide enrichment using polyclonal antibodies immobilized on protein A–coated magnetic beads. A protocol based on the simultaneous multiplex capture of more than 40 protein targets showed that approximately half of the antibodies enriched at least one functional peptide detected in the subsequent mass spectrometry analysis. The approach was further developed to also generate quantitative data via the addition of heavy isotope–labeled recombinant protein fragment standards prior to trypsin digestion. Here, we show that we were able to use small amounts of antibodies (50 ng per target) in this manner for efficient multiplex analysis of quantitative levels of proteins in a human HeLa cell lysate. The results suggest that polyclonal antibodies generated via immunization of recombinant protein fragments could be used for the enrichment of target peptides to allow for rapid mass spectrometry analysis taking advantage of a substantial reduction in sample complexity. The possibility of building up a proteome-wide resource for immuno-SILAC assays based on publicly available antibody resources is discussed. ***Molecular & Cellular Proteomics 13: 10.1074/mcp.M113.034140, 1611–1624, 2014.***

Mass spectrometry–based proteomics is fast developing in the direction of clinical applications. Therefore, reliable quantification methods for absolute protein concentration determination are indispensible tools for future applications. So far, enzyme-linked immunosorbent assays and similar antibody-based methods excel in the sensitive detection of low levels of proteins in complex matrices, whereas mass spectrometry enables unbiased approaches and can provide unsurpassed specificity. The fact that most proteomes have a very high dynamic range between high and low abundant proteins, in particular for clinical samples, such as plasma and serum, often makes it necessary to use protein depletion of the most abundant proteins (1, 2) and/or elaborate fractionations (3–5) before running the mass spectrometry analysis. This has prompted several investigators to introduce a protein or peptide capture step using specific antibodies to allow for immunoaffinity enrichment prior to the MS analysis. In this way, a "sandwich" assay is obtained, but instead of having a readout in the analysis step based on a second antibody, the analysis step is performed using MS. In such an approach, either the intact protein is captured using an anti-protein antibody (6) or a peptide derived from the protein is captured using an anti-peptide antibody that has been raised to the target peptide of interest (7–11). This is the principle behind stable isotope standards and capture by anti-peptide antibodies (SISCAPA),[1] developed by Anderson and co-workers (12–15). In immunoaffinity

From the ‡Science for Life Laboratory, KTH - Royal Institute of Technology, SE-171 21 Stockholm, Sweden; ¶Department of Proteomics, KTH - Royal Institute of Technology, SE-106 91 Stockholm, Sweden; ‖Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany; **Science for Life Laboratory, Department of Oncology-Pathology, Karolinska Institute, SE-171 21 Stockholm, Sweden

Received September 5, 2013, and in revised form, March 3, 2014
Published, MCP Papers in Press, April 10, 2014, DOI 10.1074/mcp.M113.034140

Author contributions: M.M. and M.U. designed research; F.E., T.B., and B.F. performed research; M.Z., H.J., and J.L. contributed new reagents or analytic tools; F.E., T.B., and B.F. analyzed data; F.E., T.B., B.F., M.M., and M.U. wrote the paper; E.L. and S.H. provided intellectual input.

[1] The abbreviations used are: SISCAPA, stable isotope standards and capture by anti-peptide antibodies; PrEST, protein epitope signature tag; ABP, albumin binding protein; SILAC, stable isotope labeling by amino acids in cell culture.

proteomics, it is preferable for the affinity of the anti-peptide capture antibody to be high, but the requirement for high selectivity is lower, because the mass spectrometer can readily distinguish and quantify the analyte peptide of interest despite the binding of other peptides in the digested sample.

A disadvantage with the immunoaffinity proteomics strategy is the limited availability of suitable antibodies that recognize peptides from the corresponding protein targets. The affinity enrichment of peptides usually requires the generation of custom antibodies for each target peptide, and this very time-consuming process makes high-throughput efforts very difficult to pursue. Most efforts so far have been aimed toward generating monoclonal antibodies against specific peptides selected as appropriate for mass spectrometric detection, which is a laborious and costly exercise. It would therefore be of great interest to explore whether antibodies generated toward native proteins or protein fragments could be used for the capture of peptides and in this way take advantage of the huge resource of already existing reagents for immunoproteomics.

Here, we investigated whether the publicly available resources on polyclonal antibodies could be used for immunoenrichment followed by quantitative proteomics. According to the Antibodypedia portal, there exist more than a million publicly available antibodies toward human protein targets, and more than 70% of these antibodies are polyclonal antibodies. These antibodies are of course interesting starting points as a resource for immunoproteomics, although this application was not intended at the time when the antibodies were generated. More specifically, we have investigated the use of polyclonal antibodies from the Human Protein Atlas project, covering more than 80% of all human protein-coding genes. These antibodies have been raised against human recombinant proteins called protein epitope signature tags (PrESTs), and we have therefore investigated the direct use of this resource for quantitative proteomics.

An attractive strategy for quantitative proteomics using immuno-enrichment is to use stable isotope approaches, including methods based on adding stable isotope–labeled peptides (16, 17), proteins (18, 19), or protein fragments (20). These methods are built on the detection of peptides generated by protease cleavage of the proteins in the sample, and the quantification is achieved by reading out the ratio between the endogenous peptide and the heavy-labeled spiked-in peptide. Because the endogenous protein and the labeled internal standard behave identically throughout the sample preparation including the immuno-enrichment, the relative ratio provides quantitative information, as the peptides can be distinguished by the mass spectrometer because of the shift in mass. We recently described (20) a method for protein quantification making use of the large library of PrESTs that has been developed in the course of the Human Protein Atlas (21) project. Heavy isotope–labeled PrESTs were quantified against an ultrapurified and accurately quantified protein standard using the albumin binding protein (ABP) tag. Thereafter, known amounts of heavy PrESTs were spiked into cell lysates, and the SILAC ratios were used to determine the cellular quantities of the endogenous proteins. That approach sidesteps the quantification-, storage-, and digestion-related causes of quantification error that are inherent to peptide-based methods. The PrEST-SILAC principle was used to simultaneously determine the copy numbers of 40 proteins in HeLa cells demonstrating quantitative measurements over a wide range of protein abundances, from the highly abundant cytoskeletal protein Vimentin, with 20 million copies, down to the low abundant transcription factor FOS, with only 6000 copies per cell.

Here, we combined the use of polyclonal antibodies for immunocapture with quantitative proteomics using heavy isotope–labeled proteins. A semi-automated immuno-SILAC method was developed for multiplex analysis of protein targets, taking advantage of the linear epitopes of the antibodies. A special effort was made to decrease the amounts of antibodies used in the assay. Based on the results, a new strategy for rapid mass spectrometry readout for target-specific proteomics is outlined in which antibodies are used for the multiplex immunocapture of peptides generated via trypsin digestion of cell extracts spiked with isotope-labeled recombinant protein fragments corresponding to the protein targets.

EXPERIMENTAL PROCEDURES

Generation of Antibodies—Antigens were designed using the software PRESTIGE (22). Gene fragments were amplified from a pool of RNA isolated from human tissues, cloned into a vector, and expressed in *Escherichia coli*. To generate polyclonal antibodies, purified and validated recombinant protein fragments were used for the immunization of New Zealand White rabbits, and the polyclonal rabbit sera were purified using their corresponding antigens as affinity ligands (23).

Epitope Mapping Using High-density Peptide Array—High-density peptide arrays were designed to contain 12-mer peptides with an overlap of 11 amino acid residues, in total covering all the antigen sequences. Parallel *in situ* peptide synthesis on microscope slides and removal of side chain protecting groups were performed by Roche NimbleGen Inc. (Madison, WI). Each slide containing 12 identical subarrays was covered with a PX12-mixer mask (Roche NimbleGen Inc.) according to the manufacturer's instructions. The polyclonal antibodies were combined into pools of 20 antibodies and diluted in binding buffer (10 mM Tris, 0.45% NaCl, pH 7.4, alkali soluble casein 0.5%) to a final concentration of 0.5 μg/ml for each antibody. The samples were added to the peptide arrays and incubated overnight at room temperature in a NimbleGen Hybridization Station (Roche NimbleGen Inc.). After primary incubation, the mixer masks were removed and the slides were washed in coplin jars twice with TBSTT (20 mM Tris, 0.9% NaCl, pH 7.4, 0.1% Tween 20, 0.4% Triton X-100) and twice with TBS, with each wash lasting 10 min. Secondary DyLight649 conjugated anti-rabbit antibodies (Jackson ImmunoResearch, West Grove, PA) were diluted to 0.15 μg/ml in binding buffer in LockMailer jars, and the slides were incubated for three hours on a shaking table. The slides were washed again twice with TBSTT and twice with TBS as described above, quickly rinsed three times in filtered de-ionized water, and dried using a microarray slide centrifuge. The slides were scanned at 2 μm resolution using a

NimbleGen MS200 scanner (Roche NimbleGen Inc.), and the median fluorescence intensities of the peptide features in the scanned images were analyzed using the NimbleScan software (Roche NimbleGen Inc.).

Preparation of PrEST Digest—PrESTs were mixed into three pools of 41, 42, and 44 targets, respectively. Samples containing 10 μg of each PrEST were first reduced with DTT and thereafter digested using the filter-aided sample preparation method (24). Briefly, the sample was added to a 30-kDa cutoff spin filter (Millipore, Bedford, MA), and the buffer was exchanged to denaturation buffer. The sample was alkylated with iodoacetamide and the buffer was changed to 50 mM NH_4HCO_3 before trypsin (Sigma-Aldrich, St. Louis, MO) was added, and the sample was incubated at 37 °C overnight.

Affinity Enrichment of Peptides Using Polyclonal Antibodies—Immunoaffinity enrichment of peptides from the trypsin-digested PrEST mixture was carried out with a subset of 127 polyclonal rabbit antibodies. A total of 250 ng of each antibody was pooled into 41-, 42-, and 44-plex pools, and the final volume was adjusted to 300 μl with PBS and Chaps detergent to yield a final concentration of 0.03% (w/v). In parallel, 5.3 mg of Protein A Dynabeads (Invitrogen, #10001D) was placed on a handheld magnet (Dynal, Oslo, Norway), and the storage buffer was removed before the beads were washed twice with wash buffer (1× PBS, 0.03% (w/v) CHAPS). Each subset of pooled antibodies was immobilized together with 150 μg of Protein A–coated beads per microgram of antibody and incubated for 30 min on a rotor mixer for 1 h at room temperature. A total of 200 ng of each trypsin-digested PrEST was diluted to 50 μl with PBS-supplemented CHAPS to a yield a final concentration of 1× PBS, 0.03% (w/v) CHAPS. All samples were prepared in duplicate and transferred to one 96-well standard microplate (ABgene, Hamburg, Germany) that was inserted into a Magnatrix 1200 (Magnetic Biosolutions AB, Stockholm, Sweden) automated bead processing system. Afterward, immobilized antibody–bead mixtures corresponding to 50 ng of antibody per target were transferred in triplicate to separate wells in a 96-well PCR plate (Thermo Scientific), and the plate was inserted into the Magnatrix 1200 system. The beads were washed twice with wash buffer and mixed with the peptide mixture from the trypsin-digested PrEST mixture using robotics. The 96-well standard microplate was manually covered with an opaque adhesive foil, and peptides were enriched for 16 h overnight at room temperature on a microtiter plate shaker at 1350 rpm. The following day, the plate was inserted into the Magnatrix 1200, and the beads were washed twice with wash buffer and then twice with 50 mM NH_4HCO_3. The enriched peptides were eluted with 10 μl of 0.1% formic acid (pH 2.5) for 2 min. All samples were heat treated at 96 °C for 5 min in order to denature antibodies that were eluted along with peptides from the solid bead support. Each sample was manually supplemented with 1 μl of 33% acetonitrile prior to storage at −20 °C until LC-MS analysis.

Production of Antigen Standards for Absolute Quantification—The expression vector pAff8c containing a fragment coding for the quantification standard HisABPOneStrep was transformed into *E. coli* Rosetta DE3 cells (Novagen, Merck, Darmstadt, Germany), and HisABPOneStrep was expressed according to the standard protocol used within the Human Protein Atlas project (25). After purification using immobilized metal ion affinity chromatography and buffer exchange to 1× PBS (10 mM NaP, 150 mM NaCl, pH 7.3) on a PD-10 desalting column (GE Healthcare, Uppsala, Sweden), a second purification step was performed using a StrepTrap™ HP column (GE Healthcare) on an ÄKTAexplorer system (GE Healthcare) according to the suggested protocol. The concentration of the purified protein was determined using amino acid analysis. Expression vectors containing PrEST fragments were transformed into an *E. coli* strain auxotrophic for lysine and arginine (26) for the production of heavy isotope–labeled PrESTs. A total of 41 targets were chosen for this purpose. Included PrESTs contained at least five theoretical tryptic peptides that had previously been detected in discovery proteomics experiments (data not presented here). These PrESTs were non-overlapping with the set used for the affinity enrichment screening. Cultivations of 10 ml were performed in 100-ml shake flasks using minimal autoinduction media as previously described (20, 27). Heavy isotope–labeled (^{13}C and ^{15}N) arginine and lysine (Cambridge Isotope Laboratories, Tewksbury, MA) and light versions of the remaining 18 amino acids (Sigma-Aldrich) were added to the medium to a final concentration of 200 μg/ml. After cultivation, the cells were lysed and the PrESTs were purified according to the standard Human Protein Atlas protocol (25).

Quantification of PrESTs—HisABPOneStrep and heavy isotope–labeled PrESTs were mixed in 50 mM NH_4HCO_3. The sample was reduced with DTT, alkylated with iodoacetamide, and digested with trypsin (Sigma-Aldrich) overnight. The samples were diluted in 5% acetonitrile, 0.1% formic acid, injected onto a 150 mm × 0.5 mm Zorbax 80SB-C18 column (Agilent, Santa Clara, CA), and separated using a 20-min gradient of 10%–40% acetonitrile with a flow rate of 20 μl/min on an Agilent 1200 capillary-LC system. The peptides were analyzed using an Agilent 6520 electrospray ionization quadrupole TOF mass spectrometer in a data-dependent manner; three precursor ions per cycle were chosen and fragmented via collision-induced dissociation. The isolation width was set at 4 *m/z.* Full-scan MS spectra were acquired between 300 and 2000 *m/z*, and product ion scans between 100 and 2000 *m/z*. Data analysis was performed using the software APP (unpublished), combining several MS analysis modules into one data analysis tool. The search engine X!Tandem (28) (version 2011.12.01.1) was used with the human UniProt database (70,555 entries, downloaded May 28, 2013) with the addition of the HisABPOneStrep sequence. Carbamidomethylation of cysteines was added as a fixed modification, and methionine oxidation was allowed as a variable modification. The minimum peptide length was five amino acids, and a maximum of two missed cleavages was allowed. Data from X!Tandem were further processed with PeptideProphet (29) and ProteinProphet (30), and SILAC ratios were determined with XPRESS software (31) (all from TPP v4.6 occupy rev 3). Three technical replicates were performed for each PrEST, and the median value was used when determining the protein concentration.

Preparation of HeLa Cell Lysates and Trypsin Digestion—HeLa cells (32) were cultivated in Dulbecco's modified Eagle's medium (Sigma-Aldrich) containing 10% fetal bovine serum (Sigma-Aldrich) and antibiotic and antimycotic solution (Sigma-Aldrich). Cells were released from the culture dish with a trypsin-EDTA solution (Sigma-Aldrich) and frozen at −80 °C in aliquots of 10 million cells per tube. Tubes were thawed on ice, and the cells were lysed with 1 ml of lysis buffer (0.1 M Tris-HCl, 4% SDS, 10 mM DTT, pH 7.6). Samples were incubated at 95 °C for 3 min and sonicated for 1 min. Aliquots of 100 μl corresponding to 1 million HeLa cells were used for tryptic digestion. For the first experiment, 1 pmol of each heavy isotope–labeled PrEST was spiked into the HeLa sample. Triplicate samples were then prepared in which the PrEST amounts had been adjusted to a ratio close to 1:1 relative to the corresponding endogenous protein. The correct amount of each PrEST was mixed and reduced with DTT before the PrEST mix was spiked into the HeLa lysate. The sample was diluted with denaturing buffer (8 M urea, 0.1 M Tris-HCl, pH 8.5) and centrifuged through a 0.65-μm spin filter (Millipore) to get rid of cell debris. Digestion was performed using the filter-aided sample preparation method (24) as described above.

Peptide Fractionation for PrEST-SILAC—Before MS analysis, 30 μg of the peptide mixture was divided into six fractions by means of strong anion exchange chromatography. This was done in a pipette-tip format as previously described (33). In brief, pipette tips were packed with strong anion exchange material (3M Bioanalytical Tech-

nologies, St. Paul, MN), and the peptide sample was loaded. Peptides were eluted according to isoelectric point with buffers of decreasing pH. After fractionation, eluted peptides were desalted using C18 StageTips.

Absolute Quantification of HeLa Cell Lysate—Immunoaffinity enrichment of peptides from trypsin-digested HeLa lysate, into which heavy PrESTs had been spiked prior to digestion, was carried out in the same way as described above for 41 polyclonal rabbit antibodies for which heavy-labeled PrESTs were available. Here, a total of 15 μg of trypsin-digested HeLa lysate with spiked-in heavy PrESTs was diluted to 50 μl with PBS-supplemented CHAPS to a yield a final concentration of 1× PBS, 0.03% (w/v) CHAPS. A total of 500 ng of each corresponding antibody was pooled (41-plex), and the final volume was adjusted to 500 μl with PBS and CHAPS detergent to yield a final concentration of 0.03% (w/v). All antibodies were immobilized onto 3.15 mg Protein A–coated magnetic beads. All samples were prepared in triplicate and processed in the same way as described above.

Liquid Chromatography and Mass Spectrometry—For PrEST-SILAC samples, 2 μg of peptides per fraction were analyzed, and for immuno-SILAC, only 50% of the sample was used. Peptides were first trapped on a Zorbax 300SB-C18 column (Agilent) and separated on an NTCC-360/100–5-153 (Nikkyo Technos Co., Ltd., Tokyo, Japan) column using a gradient of 6%–40% acetonitrile over 180 min (PrEST-SILAC) or 15 min (immuno-SILAC) with a flow rate of 0.4 μl/min on an Agilent 1200 nano-LC system. MS analysis was performed on a Q Exactive instrument (Thermo Fisher Scientific, San Jose, CA) operated in a data-dependent manner, with five precursors selected for fragmentation by higher energy collisional dissociation in each full MS scan. MS spectra were recorded between 300 and 1700 *m/z* at 70,000 resolution, and MS fragment ion spectra were recorded at 17,500 resolution.

Analysis of PrEST-SILAC and Immuno-SILAC MS Data—Data from Q Exactive MS runs were analyzed using MaxQuant software (34) (version 1.3.0.5) with the built-in search engine Andromeda (35). A human UniProt database (70,555 entries, downloaded May 28, 2013) was used in the search. The minimum peptide length was six amino acids, and two missed cleavages were allowed. Carbamidomethylation of cysteines was set as a fixed modification, and methionine oxidation and N-terminal acetylation were chosen as variable modifications. The initial MS mass tolerance for recalibration was 20 ppm, the initial mass deviation for the precursor ions was 4.5 ppm, and a maximum error of 20 ppm was allowed for MS/MS spectra. The false discovery rate was set at 0.01, and the match between runs option was used with a 2-min retention-time window. Identified peptides were grouped with their corresponding PrEST, and copy numbers were calculated for each peptide. The median peptide value was used as the copy number for the corresponding protein as well as the median value among the technical replicates.

RESULTS

Principle of the Targeted Immunoproteomics Method—The principle of immuno-SILAC is shown in Fig. 1. The method relies on the use of stable isotope–labeled recombinant protein fragment standards (PrESTs) produced in bacteria (Fig. 1*A*). In immuno-SILAC, a cell lysate is mixed with known amounts of accurately quantified heavy-labeled protein fragment standards, generated with heavy isotope–labeled versions (^{15}N and ^{13}C) of the amino acids arginine and lysine. The protein mixture is enzymatically digested, and the generated peptides are subsequently captured by antibodies immobilized onto Protein A–coated magnetic beads as illustrated in Fig. 1*C*. Following enrichment from the complex peptide mixture, target peptides are eluted with formic acid from the solid phase bead support. If the peptides are eluted in an MS-compatible buffer, only acetonitrile has to be added before LC-MS/MS analysis. The low complexity of the resulting sample enables very short analysis times. Here, a single 15-min HPLC gradient was sufficient for separation of a multiplexed sample in which up to 44 different antibodies were used for peptide enrichment. The ratios of light peptides originating from the endogenous digested proteins and peptides from the spiked-in heavy isotopic standards are compared, giving an absolute quantitative measurement of the studied proteins.

A schematic overview comparing the workflows for the related SISCAPA and immuno-SILAC methods is shown in Fig. 1*B*. In SISCAPA, known amounts of heavy-labeled synthetic peptides are spiked in after trypsin digestion, whereas in immuno-SILAC the protein fragments are added prior to trypsin digestion. The addition of protein fragments as standards before enzyme digestion has the advantage of compensating for miscleavages or otherwise incomplete digestion, as the heavy protein standards undergo the same processing as the endogenous proteins (36). In this way, possible losses during sample preparation do not introduce quantification errors.

Analysis of Linear Epitopes of Polyclonal Antibodies Using High-density Peptide Arrays—In order to analyze the number of linear epitopes for polyclonal antibodies generated in a standardized manner using protein fragments (PrESTs), a high-density array with 175,000 overlapping synthetic 12-mer peptides with a single-amino-acid lateral shift was designed covering 941 protein fragment sequences. The target proteins, a majority of all human kinases and a number of interesting biomarkers for cancer, were selected by the 7th European Union framework project Affinomics (37) for the ultimate goal of generating corresponding affinity reagents. Parallel *in situ* peptide synthesis of the arrays was achieved via repeated cycles of selective activation using a UV-light source and a micromirror device followed by the incorporation of amino acids with a photolabile protective group. Each of the synthesized peptide arrays was incubated with a pool containing 20 of the selected polyclonal antibodies, and a fluorophore-conjugated secondary anti-rabbit antibody was used to detect antibody–peptide interactions. The small shift of only a single amino acid between the overlapping peptides allowed very detailed mapping of the linear epitopes recognized by the antibodies. Previous results for polyclonal antibodies epitope-mapped together with two separate pools of 29 unrelated antibodies showed very similar binding profiles, indicating limited cross-reactivity of the unrelated antibodies to the antigen sequence peptides (data not shown). Two examples of epitope mapping are shown in Figs. 2*A* and 2*B*, where the bars represent the median fluorescence intensity of each of the overlapping peptides. The anti-AGAP2 antibody HPA023474 showed two distinct linear epitopes (Fig. 2*A*),

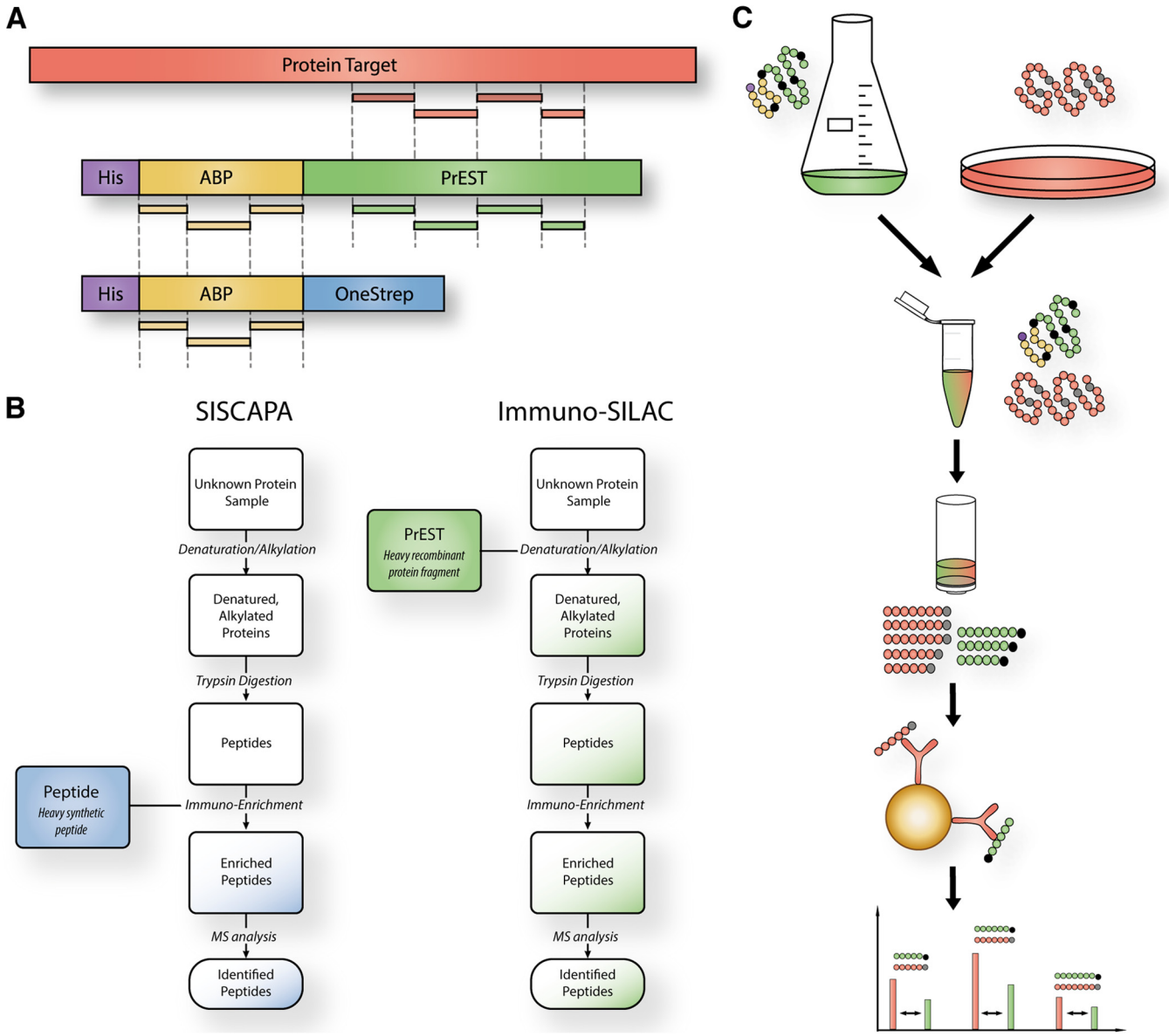


FIG. 1. **The principle of the immuno-SILAC method.** *A*, absolute protein quantification using PrESTs as the internal standard. Peptides originating from the albumin binding protein (ABP) tag (yellow) are used to quantify each PrEST against an ultrapurified ABP protein standard. The PrEST can thereafter be used as an internal standard in unknown samples to quantify the corresponding endogenous protein (red) in LC-MS. *B*, schematic representation of the immuno-SILAC workflow. Highly purified and accurately quantified isotopic heavy-labeled PrESTs are spiked into cell lysates prior to trypsin digestion, thereby minimizing the risk of differences arising between samples and standards during sample preparation. Antibodies coupled to magnetic solid phase support enrich target peptides from the digested sample, and the endogenous protein concentration is calculated from the ratio of heavy to light peptides detected via MS. *C*, comparison between the SISCAPA technology (7) using heavy-labeled peptides (blue) and immuno-SILAC using heavy-labeled protein fragments (green) for absolute protein quantification.

whereas the antibody HPA027341, targeting the protein fumarate hydratase, recognized four linear epitopes (Fig. 2*B*). The epitope mapping of the 941 antibodies showed that the number of linear epitopes varied for the different antibodies (Fig. 2*C*), but on average 2.9 linear epitopes were detected. The fact that most of the analyzed polyclonal antibodies recognized multiple linear epitopes suggests the possibility of using these polyclonal antibodies as capturing agents for the enrichment of peptides from trypsin-digested complex samples. It is noteworthy that ~40% of the identified epitopes contained a trypsin cleavage site (data not shown), suggesting that they might not be functional for immuno-enrichment of the corresponding peptides.

Immunocapture of Peptides for Mass Spectrometry Analysis—In order to investigate the performance of the epitope-mapped antibodies for immuno-enrichment of target protein

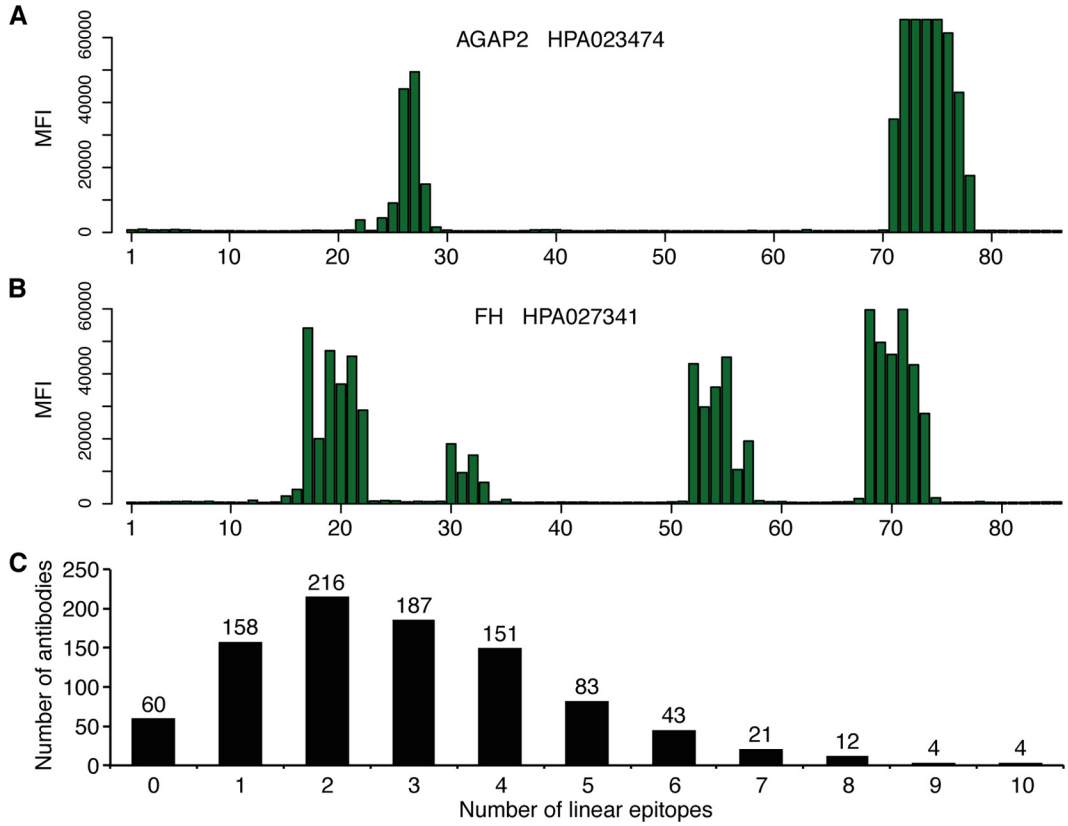


FIG. 2. **Epitope mapping of polyclonal antibodies on planar peptide arrays.** Linear epitopes of 941 polyclonal antibodies were analyzed using synthetic 12-mer peptides with an overlap of 11 amino acids, covering the antigen sequences. Two examples of the epitope mappings are shown in *A* (HPRA023474) and *B* (HPA027341), where each bar on the *x*-axis corresponds to one of the overlapping peptides and the height shows the median fluorescence intensity. *C*, distribution of the number of polyclonal antibodies recognizing 0 to 10 linear epitopes.

peptides from complex digested samples, we chose a random subset of 150 antibodies. For 127 of these, the corresponding recombinant PrESTs were available. The protein fragments had previously been expressed in *E. coli* and purified as described before (25). The 127 PrESTs were pooled together in equimolar amounts and used for trypsin cleavage using a standardized protocol (24). The peptide mixtures were transferred to a robotic work station for magnetic bead handling (38), and immuno-enrichment of target peptides was performed using a multiplex mixture of 40 to 45 of the corresponding polyclonal antibodies captured on Protein A–coated magnetic beads, using only 50 ng of each antibody. After overnight affinity capture, the beads were thoroughly washed and the enriched peptides were eluted with formic acid suitable for the subsequent mass spectrometry analysis. The MS analysis revealed that 57 out of the 127 target proteins were successfully identified by at least one tryptic peptide (supplemental Table S1). A representative subset of identified peptides and the corresponding mapped linear epitopes for the antibody used can be seen in Fig. 3. As shown by the examples, predicted epitopes as determined by epitope mapping using overlapping 12-mer peptides often correlated with the sequences of enriched peptides in immuno-SILAC. However, in some cases (*e.g.* FLT1 and CAMK4 (Fig. 3)), some peptides were not predicted from the high-density array mapping. It is tempting to speculate that in many of these cases, the epitopes seen by the antibody were not covered within the 12-mer peptides displayed on the high-density array, implying that these epitopes needed longer peptides to form than available on the peptide arrays. It is also apparent from the examples that many of the linear epitopes predicted by the epitope mapping were not observed after the MS analysis. Many of these epitopes contain lysines or arginines, and the corresponding peptides are thus not expected to be captured, as the epitope is cleaved by trypsin. It is also likely that many other peptides were not detected in the MS analysis because of technical issues (39). In summary, the results presented here, based on 127 antibodies, suggest that approximately half of the polyclonal antibodies enriched at least one peptide that could subsequently be detected via mass spectrometry.

Comparative Analysis of PrEST-SILAC and Immuno-SILAC—In order to investigate the success rate of finding functional immunoproteomics pairs using corresponding recombinant PrESTs as internal standards, we analyzed a new set of 41 protein targets known to be present in HeLa cells based on RNA sequencing data (40). The targets were chosen based on sequences of the corresponding PrESTs, with a minimum of

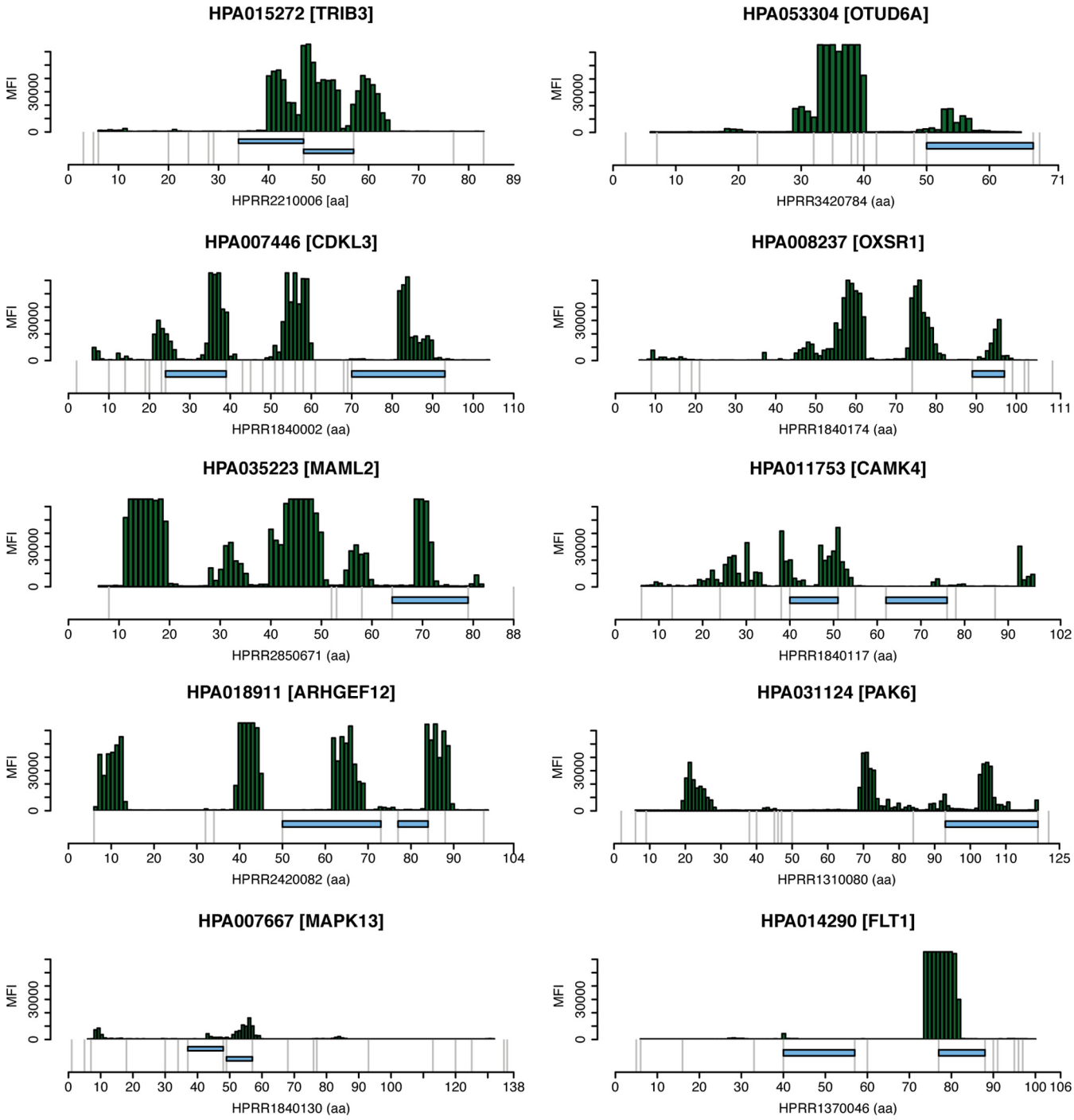


FIG. 3. **Comparison of mapped epitopes and peptides identified in immuno-SILAC screening of polyclonal antibodies against trypsin-digested PrESTs.** The upper part of each panel shows the binding intensities in median fluorescence intensity to overlapping 12-mer peptides in the epitope mapping, where epitopes are shown as clusters of bound peptides. The blue horizontal bars in the lower part show the locations of peptides identified in immuno-SILAC screening on their corresponding antigen, and gray vertical lines indicate trypsin cleavage sites.

five tryptic peptides required. A HeLa cell lysate was analyzed using both the PrEST-SILAC method described earlier (20) and the immuno-SILAC protocol in parallel. Isotope-labeled protein fragments corresponding to 41 human protein targets were spiked into a cell lysate sample as schematically shown in Fig. 1*C*, and the mixture of heavy standards and light peptides was captured using the immobilized multiplex antibodies ($n = 41$). Of the 41 antibodies used in the multiplex analysis, 22 managed to capture peptide(s) corresponding to the correct target protein. Two targets were successfully quantified using two different antibodies, resulting in 20 quantified proteins. The targets are shown in Table I, and the cell

TABLE I

Quantified protein copy number per HeLa cell for immuno-SILAC and PrEST-SILAC. Copy numbers are shown for each replicate, along with the median copy number, relative standard deviation (RSD), and number of peptides used for quantification

Gene name	Copy number replicate 1	Copy number replicate 2	Copy number replicate 3	Median copy number	RSD (%)	Number of peptides
Immuno-SILAC						
ACOT7	1 283 838	998 876	498 230	998 876	42.9	2
ANXA1	6 664 952	9 833 375	6 362 332	6 664 952	25.2	4
ANXA3	2 219 240	2 298 319	4 650 459	2 298 319	45.2	1
BLVRB	2 148 052	-	3 002 937	2 575 495	23.5	1
CANT1	-	70 703	42 667	56 685	35.0	1
CAPG	3 078 134	2 579 694	-	2 828 914	12.5	1
CLPP	1 208 794	1 087 915	777 082	1 087 915	21.7	1
DAP3	1 498 865	1 016 070	900 236	1 016 070	27.9	6
DECR1	-	-	588 225	588 225	-	1
DIMT1	-	-	235 127	235 127	-	1
ERLIN1	529 872	923 508	1 013 719	923 508	31.3	2
P4HA1	277 258	307 834	431 686	307 834	24.1	1
PDIA5	94 844	166 129	186 790	166 129	32.3	3
PRPF4	306 890	316 067	342 904	316 067	5.8	1
PTPN1	763 589	1 035 203	686 727	763 589	22.1	2
SERPINB6	885 677	2 539 203	1 477 373	1 477 373	51.3	2
SIL1	-	117 997	-	117 997	-	1
SLC25A24	2 317 150	2 099 723	1 239 592	2 099 723	30.2	1
STUB1	573 646	474 906	490 305	490 305	10.4	1
UGDH	193 277	501 192	605 305	501 192	49.5	1
PrEST-SILAC						
ACOT7	771 360	1 170 086	955 768	955 768	20.7	4
ANXA1	12 516 639	10 077 946	11 725 516	11 725 516	10.9	14
ANXA3	1 899 959	2 428 880	4 058 745	2 428 880	40.2	12
BLVRB	779 755	836 484	624 563	779 755	14.7	4
CANT1	-	67 847	-	67 847	-	2
CAPG	2 238 047	2 535 800	2 762 024	2 535 800	10.5	4
CLPP	1 017 569	705 756	639 815	705 756	25.6	3
DAP3	329 798	-	638 890	484 344	45.1	2
DECR1	-	-	-	-	-	-
DIMT1	337 769	408 358	342 916	342 916	10.8	6
ERLIN1	559 997	963 844	803 794	803 794	26.2	5
P4HA1	1 114 521	627 467	733 443	733 443	31.0	11
PDIA5	134 089	253 862	152 319	152 319	25.5	4
PRPF4	411 509	274 967	566 398	411 509	34.9	6
PTPN1	520 007	542 602	597 039	542 602	7.2	5
SERPINB6	576 320	974 111	1 269 089	974 111	37.0	9
SIL1	96 363	155 829	111 533	111 533	25.5	1
SLC25A24	1 077 749	1 015 007	925 753	1 015 007	7.6	4
STUB1	719 533	604 201	692 742	692 742	9.0	4
UGDH	1 000 743	888 750	746 425	888 750	14.5	5

Number of peptides: the total number of different peptides used for the quantification.

lysate was an extract from the HeLa cell cultivation. The concentrations of the endogenous proteins were calculated based on the ratio between the light peptides from the sample and the corresponding heavy peptide from the protein standard. In parallel, the same protein targets were analyzed using the PrEST-SILAC protocol. For immuno-SILAC, 15 μg of a digested HeLa sample was used in the peptide enrichment and prior to MS analysis; a 15-min HPLC gradient was needed for sufficient peptide separation. For PrEST-SILAC, 30 μg of peptides was divided into six fractions, and then each fraction was further separated on an HPLC column using a 3-h gradient. The difference in sample complexity between the two methods is illustrated in Fig. 4, where example chromatograms from one immuno-SILAC run and one of the six fractions from a corresponding PrEST-SILAC run are shown. An example MS spectrum (at the retention time indicated by the arrow) showing a peptide from SERPINB6 is shown for both methods. The peptide intensities are similar in the two spectra, but the absence of interfering peaks in the immuno-SILAC spectra indicates a better separation of peptides along the HPLC gradient. To further demonstrate the difference in sample complexity between the two methods, intensities for peptides corresponding to the target proteins were compared with the total intensity of all identified endogenous peptides

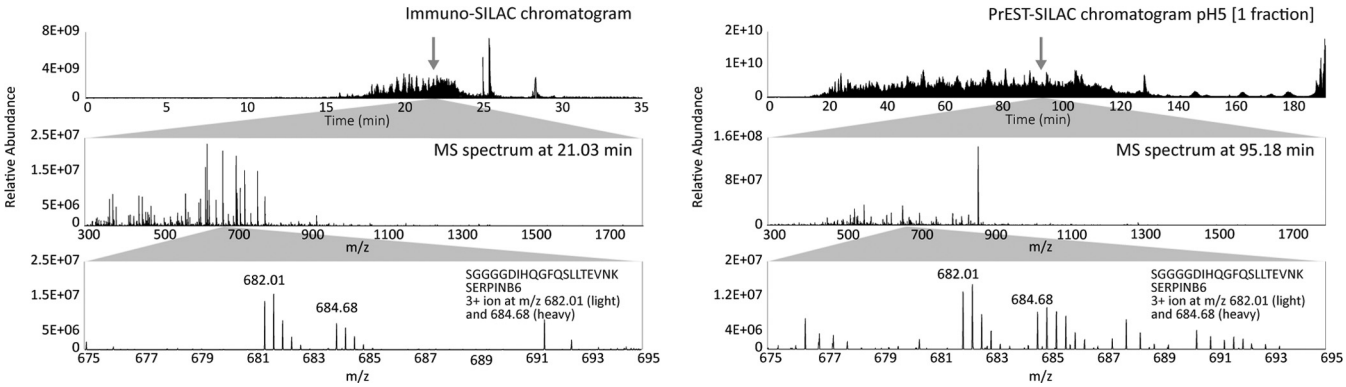


FIG. 4. **Comparison of HPLC chromatograms and MS spectra from one immuno-SILAC and one PrEST-SILAC sample.** The upper panels show HPLC chromatograms for one immuno-SILAC sample (left) and one PrEST-SILAC fraction (right). The peptide SGGGGDIHQGFQSLLTEVNK from SERPINB6 was identified in both experiments as a triple-charged ion of *m/z* 682.01 (light version). The middle panels show extracted full MS spectra at the respective retention times, and the lower panels show the same MS spectra at *m/z* between 675 and 695.

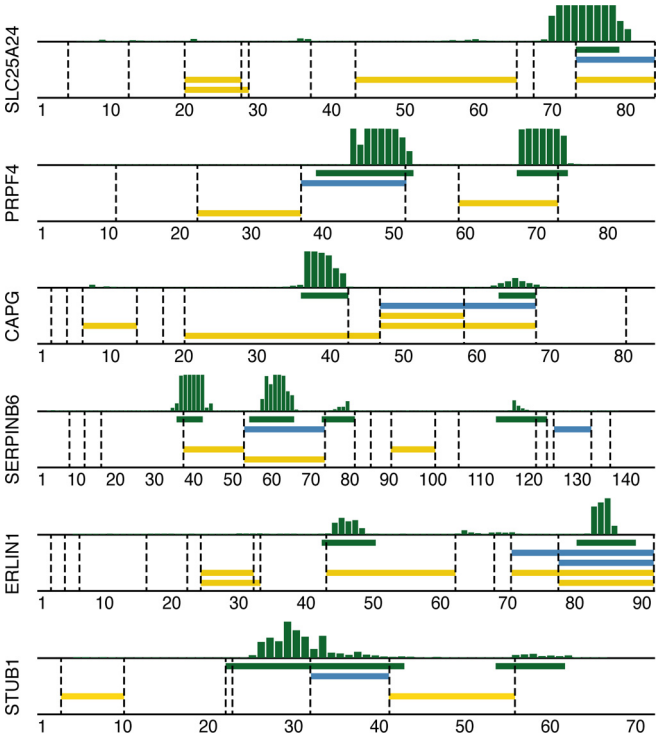


FIG. 5. **Comparison of epitope mapping, immuno-SILAC, and PrEST-SILAC peptides.** Horizontal bars show the locations of linear epitopes (green) and peptides identified in immuno-SILAC (blue) and PrEST-SILAC (yellow) on their corresponding antigen; dashed vertical lines indicate trypsin cleavage sites. Above each plot, the raw epitope mapping binding intensities to overlapping 12-mer peptides are shown as vertical bars.

(common contaminants excluded). The proportion of target peptides in immuno-SILAC was 83%, as compared with 0.5% for PrEST-SILAC (data not shown).

In Fig. 5, a comparison of identified peptides from targets quantified via both immuno-SILAC (blue) and PrEST-SILAC (yellow) is shown for 6 of the 20 proteins together with the epitopes mapped on peptide arrays (green). In general, PrEST-SILAC identified, as expected, more unique peptides than immuno-SILAC, as no specific peptides were enriched. However, the peptide enrichment step reduces the complexity of the input material, consequently decreasing the analysis time drastically. For SLC25A24, four peptides were found using PrEST-SILAC, and one peptide was found when using the corresponding antibody for peptide enrichment, as this peptide contained the consensus epitope determined on the peptide array. The same peptide was also identified in PrEST-SILAC, indicating that it was present at sufficient levels for MS analysis even without peptide enrichment. This was the case for most of the analyzed proteins, as the target proteins were all moderate to highly abundant proteins in HeLa cells. Interestingly, some peptides were identified only in immuno-SILAC (see PRPF4 and STUB1). In the case of PRPF4, two epitope regions were detected during mapping, but the most C-terminal epitope contains a trypsin cleavage site (indicated by dotted vertical lines) and, as expected, the antibody failed to bind any of the two resulting tryptic peptides. The N-terminal epitope region probably consists of two overlapping epitopes, of which at least one is still intact after digestion and can be enriched. For SERPINB6, two different peptides can be identified.

Quantitative Analysis of 20 Targets in HeLa Cell Lysates Using Immuno-SILAC and PrEST-SILAC—A quantitative analysis of 20 protein targets in a human HeLa cell lysate was performed. The same targets were analyzed using three technical replicates for both immuno-SILAC and PrEST-SILAC methods. Given the amount of HeLa cells used in the assay, absolute quantification of endogenous proteins as copy numbers per HeLa cell was determined from the ratio of heavy to light peptides. Heavy standard amounts were spiked into the lysate close to the level of endogenous protein in order to generate SILAC ratios close to 1:1 and hence more reliable quantitative data. Relative standard deviations ranged between 10% and 40% for most targets, with somewhat lower numbers for PrEST-SILAC (Table I).

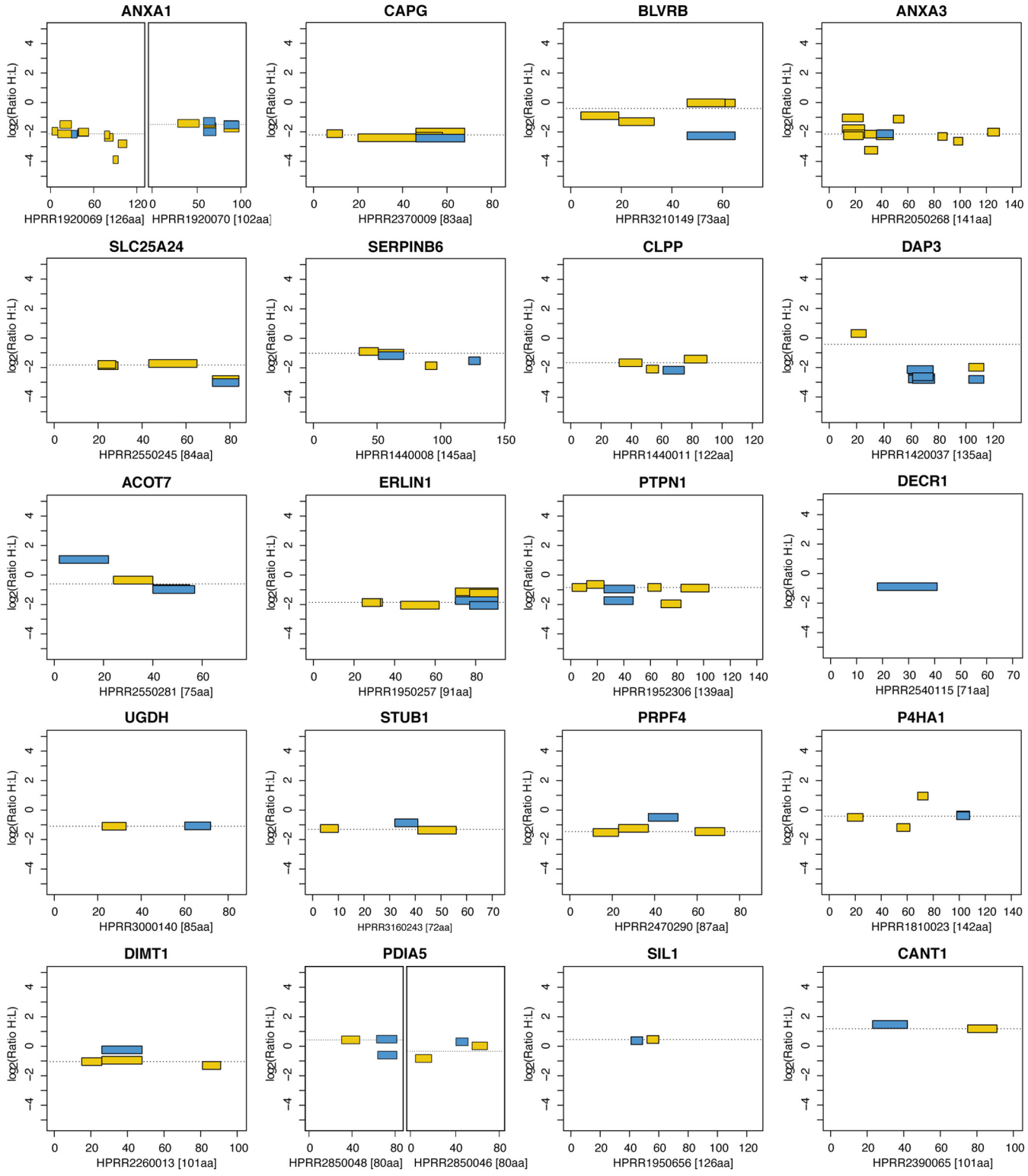


FIG. 6. **Comparison of peptide ratios along the PrEST sequences for PrEST-SILAC (yellow) and immuno-SILAC (blue).** Identified peptides from the two methods are plotted along the PrEST-sequence (*x*-axis) showing ratios (*y*-axis) between heavy and light peptides used for quantification of endogenous protein in HeLa cell lysate. The dotted line represents the median ratio for PrEST-SILAC. Overlapping peptide sequences come from missed cleavages. Two different heavy PrESTs were used for annexin 1 (ANXA1) and PDIA5.

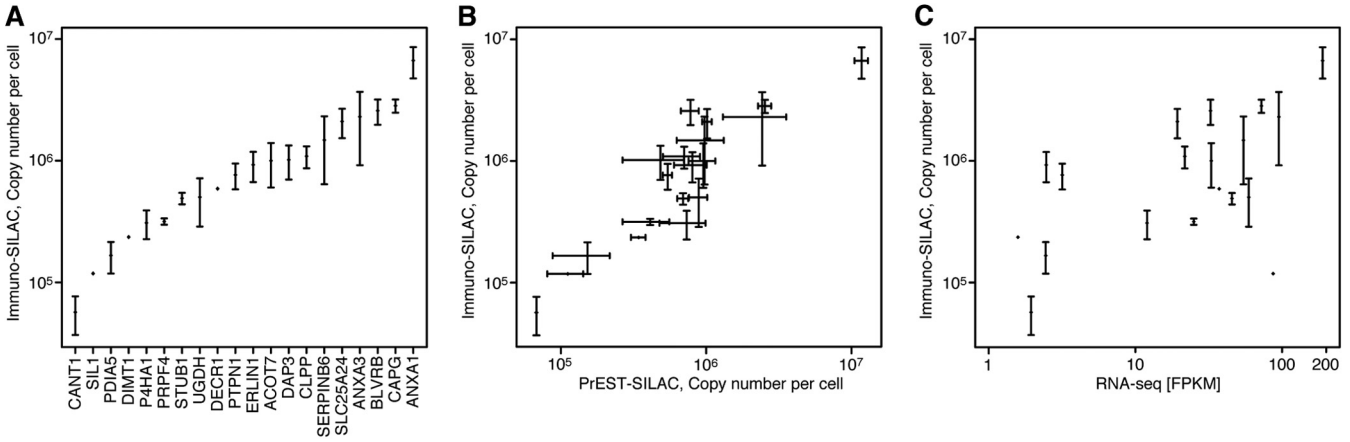


FIG. 7. **Immuno-MS results from antibodies toward 20 different target proteins in HeLa cell lysates.** *A*, immuno-SILAC quantification of 20 target proteins as copy number per cell. The analysis was performed in triplicate, and median values were plotted with error bars showing the standard deviation. *B*, comparison of copy numbers per cell obtained using immuno-SILAC (*y*-axis) and PrEST-SILAC (*x*-axis). *C*, comparison of protein copy number (*y*-axis) and transcript abundance from RNA sequencing (*x*-axis).

Ratios obtained from the MS analysis for all peptides used for quantification in immuno-SILAC in comparison with PrEST-SILAC are shown in Fig. 6. A good correlation between the PrEST-SILAC and the immuno-SILAC can be observed for all 20 targets. Note that miscleaved peptides also can be used for quantitative analysis. Here, miscleaved peptides refer to peptides that still contained one or more potential cleavage sites after trypsin digestion. These miscleaved peptides become an issue in methods relying on spiked-in peptide standards for protein quantification (41), such as AQUA (17), in which peptides are spiked in after the enzymatic digestion. However, it is noteworthy that information from these miscleaved peptides can be used in the quantitative analysis in immuno-SILAC; here we took advantage of heavy-labeled protein fragments enzymatically digested along with the protein target. The observed ratios suggest that the enzymatic digestion was equal for both protein fragment standards and endogenous proteins, and these peptides add extra and valuable information rather than introducing ambiguity to the subsequent quantitative analysis.

The absolute quantifications for all 20 targets as determined via immuno-SILAC are shown in Fig. 7*A*. Among the 20 identified proteins, copy numbers ranged from 6.6 million copies per cell for annexin 1 down to 57,000 copies per cell determined for calcium activated nucleotidase, which represents moderately to highly abundant proteins. The absolute quantitative data given by immuno-SILAC was compared with that obtained via the previously described PrEST-SILAC method (20). Cell copy numbers determined via each respective absolute quantitative method (Table I) were plotted against each other as illustrated in Fig. 7*B*. In total, 17 and 18 proteins were quantified in at least two out of three replicates using the immuno-SILAC and PrEST-SILAC methods, respectively. Out of these, 59% of the targets (10 out of 17 proteins) were detected with only a single peptide using immuno-SILAC, whereas only 6% of the targets (1 out of 18) were detected with only one peptide using PrEST-SILAC. It should be noted that the data used for quantification with the two methods differed, as in general PrEST-SILAC identified more peptides per protein than immuno-SILAC. However, even though many proteins were quantified with only one proteotypic peptide in immuno-SILAC, good correlation could be observed between the two methods. To investigate whether the determined protein cell copy numbers showed any correlation to mRNA levels of the corresponding genes, we determined the transcript levels of all genes in the HeLa cell culture using deep sequencing of mRNA molecules using RNA sequencing (42). Fig. 7*C* shows a comparison between protein cell copy numbers as determined from the immuno-SILAC experiments and the absolute transcript levels for the 20 targets analyzed here. Even though a completely linear relationship between protein and RNA levels cannot be assumed and it is especially difficult to compare protein and RNA using data from only a single cell line, the analysis still showed a trend between RNA and protein levels. It is clear from Fig. 7*C* that the RNA levels to some extent indicated protein levels.

DISCUSSION

Here, we show that polyclonal antibodies raised against recombinant protein fragments can be used for immunoproteomics analysis. The immuno-SILAC method takes advantage of the fact that the majority of the epitopes of the generated antibodies are directed to relatively short linear epitopes and are therefore useful for immunocapture of peptides generated after trypsin cleavage. In addition, the method takes advantage of the fact that protein fragments used for generating the antibodies can be efficiently expressed in bacteria (*E. coli*) using cultivation in minimal media with isotope-labeled arginines and lysines followed by a standardized protocol for affinity purification using an affinity tag. Thus, it is

easy to generate protein standards, and these can be quantified in an exact manner using the PrEST-SILAC method previously described by us (20).

The method resembles other immunoproteomics methods, such as SISCAPA (7) and the use of AQUA peptides (17), but this new method has several important differences. First, the use of protein fragments often generates multiple peptides, and it is in many cases possible to obtain quantitative data from several independent peptides using the same antibody and protein standard. In addition, the fact that the protein fragments are added prior to the trypsin cleavage ensures that uncleaved endogenous peptides will not affect the quantification as long as the digestion efficiency of the protein standard is the same as that of the endogenous protein target. Furthermore, miscleaved peptides can be used as additional standards, and the problem of miscleaved peptides is therefore transformed into an opportunity to generate more data that can be used in the analysis. Finally, the generation of the protein standards does not involve peptide synthesis, and the quantification of the protein standards can be easily performed using the quantification tag included as a fusion on every protein fragment (20).

The publicly available antibodies from the Human Protein Atlas project constitute a huge antibody resource for the immuno-SILAC method described here. These polyclonal antibodies are generated in a standardized manner involving immunizations of animals using recombinant protein fragments selected for their low sequence identity to other human proteins. In addition, more than a million antibodies toward human targets are listed in the antibody portal Antibodypedia (43), and ~70% of these antibodies are polyclonal antibodies. However, the fact that the method described here relies on polyclonal antibodies makes it important to decrease the consumption of antibody reagent in each assay. Although it is possible to renew polyclonal antibodies through re-immunization of the same antigen in more animals, batch-to-batch variations exist, and one can never be sure of an unlimited resource of renewable antibodies, as is the case with monoclonal antibodies. We have investigated various protocols to lower the consumption of the antibody reagent used in each assay. The protocol described here requires only 50 ng of antibody for each target, and these antibodies do not need to be chemically immobilized to a solid support and can be simply mixed with Protein A–coated magnetic beads in a multiplex manner. Although the amount of antibodies varies greatly for publicly available polyclonal antibodies, the mean concentration of the ~18,000 polyclonal antibodies generated within the framework of the Human Protein Atlas program is ~140 μg/ml (unpublished); thus an aliquot of 100 μl can be used for more than 250 individual assays, and the whole batch with ~10 ml of affinity-purified antibodies would last for 25,000 assays. Furthermore, it is likely that the need for antibodies can be reduced even further through the use of smaller magnetic beads and altered capture procedures, such as the use of microfluidics. Thus, the publicly available polyclonal antibodies covering more than 90% of the human protein-coding genes listed in Antibodypedia are an attractive resource for efforts to develop new immunoproteomics assays.

Here, we have screened polyclonal antibodies raised against human protein fragments to investigate their functionality for immunoproteomics. In total, the immuno-enrichment experiments generated 79 new immunoproteomics pairs, each with an antibody suitable for immuno-SILAC. Here, approximately half of the analyzed antibodies (57/127) yielded functional antibodies for peptide immuno-enrichment in the initial screening, and the subsequent quantitative HeLa analysis had a similar success rate (22/41). It is noteworthy that the high-density microarray assay identified approximately three linear epitopes per polyclonal antibody on average, but the results presented here suggest that only a fraction of these epitopes are suitable for immuno-SILAC.

The rapid turn-around times coupled with the ease of automating all the unit operations of sample preparation makes the immuno-SILAC method ideal for both research applications and future clinical diagnostic assays. In total, the PrEST-SILAC method described here required an MS analysis time of around 24 h for the six fractions of the sample. To compare, the MS analysis time for the 40-plex immuno-SILAC experiments took ~40 min. It is noteworthy that more than 80% of the total intensities from the mass spectrometry analysis corresponded to the target peptides for the immuno-SILAC experiment, suggesting that the analysis time can be reduced even further. The method allows for multiplex analysis, and we have used it for the simultaneous analysis of 20 protein targets using both PrEST-SILAC and immuno-SILAC. As a large number of antibodies are publicly available, it might not be unrealistic to increase the multiplexing to several hundreds of protein targets. Particularly for samples with a huge dynamic range between abundant proteins and proteins to be diagnosed, such as plasma and serum analysis, the immunocapture step prior to the mass spectrometry analysis is an attractive way forward. In summary, we have shown that small amounts of polyclonal antibodies can be used for efficient multiplex analysis of quantitative levels of proteins, opening up the possibility of building up a proteome-wide resource for immuno-SILAC reagents based on already available public antibody resources.

Acknowledgments—We acknowledge the entire staff of the Human Protein Atlas program and the Science for Life Laboratory for valuable contributions.

* Funding was provided by the Knut and Alice Wallenberg Foundation, the ProNova center through a grant from VINNOVA, and PROSPECTS and Affinomics 7th Framework grants from the European Directorate.

[S] This article contains supplemental material.

‡‡ To whom correspondence should be addressed.

§ These authors contributed to this work equally.

REFERENCES

1. Bjorhall, K., Miliotis, T., and Davidsson, P. (2005) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* **5,** 307–317
2. Steel, L. F., Trotter, M. G., Nakajima, P. B., Mattu, T. S., Gonye, G., and Block, T. (2003) Efficient and specific removal of albumin from human serum samples. *Mol. Cell. Proteomics* **2,** 262–270
3. Jin, W. H., Dai, J., Li, S. J., Xia, Q. C., Zou, H. F., and Zeng, R. (2005) Human plasma proteome analysis by multidimensional chromatography prefractionation and linear ion trap mass spectrometry identification. *J. Proteome Res.* **4,** 613–619
4. Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, S. S., Su, Q., McGrath, A. M., Estock, M. A., Parmar, P. P., Zhao, M., Huang, S. T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N. L., Taylor, J., and Steiner, S. (2003) The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **3,** 1345–1364
5. Pernemalm, M., Lewensohn, R., and Lehtio, J. (2009) Affinity prefractionation for MS-based plasma proteomics. *Proteomics* **9,** 1420–1427
6. Berna, M., Ott, L., Engle, S., Watson, D., Solter, P., and Ackermann, B. (2008) Quantification of NTproBNP in rat serum using immunoprecipitation and LC/MS/MS: a biomarker of drug-induced cardiac hypertrophy. *Anal. Chem.* **80,** 561–566
7. Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W., and Pearson, T. W. (2004) Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* **3,** 235–244
8. Whiteaker, J. R., Zhao, L., Abbatiello, S. E., Burgess, M., Kuhn, E., Lin, C., Pope, M. E., Razavi, M., Anderson, N. L., Pearson, T. W., Carr, S. A., and Paulovich, A. G. (2011) Evaluation of large scale quantitative proteomic assay development using peptide affinity-based mass spectrometry. *Mol. Cell. Proteomics* **10,** M110.005645
9. Neubert, H., Gale, J., and Muirhead, D. (2010) Online high-flow peptide immunoaffinity enrichment and nanoflow LC-MS/MS: assay development for total salivary pepsin/pepsinogen. *Clin. Chem.* **56,** 1413–1423
10. Hoofnagle, A. N., Becker, J. O., Wener, M. H., and Heinecke, J. W. (2008) Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. *Clin. Chem.* **54,** 1796–1804
11. Ahn, Y. H., Lee, J. Y., Lee, J. Y., Kim, Y. S., Ko, J. H., and Yoo, J. S. (2009) Quantitative analysis of an aberrant glycoform of TIMP1 from colon cancer serum by L-PHA-enrichment and SISCAPA with MRM mass spectrometry. *J. Proteome Res.* **8,** 4216–4224
12. Razavi, M., Frick, L. E., LaMarr, W. A., Pope, M. E., Miller, C. A., Anderson, N. L., and Pearson, T. W. (2012) High-throughput SISCAPA quantitation of peptides from human plasma digests by ultrafast, liquid chromatography-free mass spectrometry. *J. Proteome Res.* **11,** 5642–5649
13. Anderson, N. L., Jackson, A., Smith, D., Hardie, D., Borchers, C., and Pearson, T. W. (2009) SISCAPA peptide enrichment on magnetic beads using an in-line bead trap device. *Mol. Cell. Proteomics* **8,** 995–1005
14. Whiteaker, J. R., Zhao, L., Zhang, H. Y., Feng, L. C., Piening, B. D., Anderson, L., and Paulovich, A. G. (2007) Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal. Biochem.* **362,** 44–54
15. Whiteaker, J. R., Zhao, L., Anderson, L., and Paulovich, A. G. (2010) An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. *Mol. Cell. Proteomics* **9,** 184–196
16. Barr, J. R., Maggio, V. L., Patterson, D. G., Jr., Cooper, G. R., Henderson, L. O., Turner, W. E., Smith, S. J., Hannon, W. H., Needham, L. L., and Sampson, E. J. (1996) Isotope dilution–mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.* **42,** 1676–1682
17. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 6940–6945
18. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell. Proteomics* **6,** 2139–2149
19. Singh, S., Springer, M., Steen, J., Kirschner, M. W., and Steen, H. (2009) FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *J. Proteome Res.* **8,** 2201–2210
20. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M., and Mann, M. (2012) A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol. Cell. Proteomics* **11,** O111.009613
21. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28,** 1248–1250
22. Berglund, L., Bjorling, E., Jonasson, K., Rockberg, J., Fagerberg, L., Al-Khalili Szigyarto, C., Sivertsson, A., and Uhlen, M. (2008) A whole-genome bioinformatics approach to selection of antigens for systematic antibody generation. *Proteomics* **8,** 2832–2839
23. Agaton, C., Falk, R., Hoiden Guthenberg, I., Gostring, L., Uhlen, M., and Hober, S. (2004) Selective enrichment of monospecific polyclonal antibodies for antibody-based proteomics efforts. *J. Chromatogr. A* **1043,** 33–40
24. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
25. Tegel, H., Steen, J., Konrad, A., Nikdin, H., Pettersson, K., Stenvall, M., Tourle, S., Wrethagen, U., Xu, L., Yderland, L., Uhlen, M., Hober, S., and Ottosson, J. (2009) High-throughput protein production—lessons from scaling up from 10 to 288 recombinant proteins per week. *Biotechnol. J.* **4,** 51–57
26. Matic, I., Jaffray, E. G., Oxenham, S. K., Groves, M. J., Barratt, C. L., Tauro, S., Stanley-Wall, N. R., and Hay, R. T. (2011) Absolute SILAC-compatible expression strain allows Sumo-2 copy number determination in clinical samples. *J. Proteome Res.* **10,** 4869–4875
27. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234
28. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467
29. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392
30. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75,** 4646–4658
31. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19,** 946–951
32. Scherer, W. F., Syverton, J. T., and Gey, G. O. (1953) Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J. Exp. Med.* **97,** 695–710
33. Wisniewski, J. R., Zougman, A., and Mann, M. (2009) Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **8,** 5674–5678
34. 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
35. 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
36. Kuhn, E., Whiteaker, J. R., Mani, D. R., Jackson, A. M., Zhao, L., Pope, M. E., Smith, D., Rivera, K. D., Anderson, N. L., Skates, S. J., Pearson, T. W., Paulovich, A. G., and Carr, S. A. (2012) Interlaboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Mol. Cell. Proteomics* **11,** M111.013854
37. Stoevesandt, O., and Taussig, M. J. (2012) European and international collaboration in affinity proteomics. *Nat. Biotechnol.* **29,** 511–514
38. Uhlen, M., Hultman, T., Wahlberg, J., Lundeberg, J., Bergh, S., Pettersson, B., Holmberg, A., Stahl, S., and Moks, T. (1992) Semi-automated solid-phase DNA sequencing. *Trends Biotechnol.* **10,** 52–55
39. Mirzaei, H., and Regnier, F. (2006) Enhancing electrospray ionization effi-

ciency of peptides by derivatization. *Anal. Chem.* **78,** 4175–4183
40. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7,** 548
41. Glatter, T., Ludwig, C., Ahrne, E., Aebersold, R., Heck, A. J., and Schmidt, A. (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11,** 5145–5156
42. Danielsson, F., Wiking, M., Mahdessian, D., Skogs, M., Ait Blal, H., Hjelmare, M., Stadler, C., Uhlen, M., and Lundberg, E. (2013) RNA deep sequencing as a tool for selection of cell lines for systematic subcellular localization of all human proteins. *J. Proteome Res.* **12,** 299–307
43. Bjorling, E., and Uhlen, M. (2008) Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol. Cell. Proteomics* **7,** 2028–2037
44. Malm *et al.* (2014), unpublished

Copy number analysis of the murine platelet proteome

Platelets are small cells circulating in the vasculature mediating hemostasis. In case of injuries of the vessels they seal the lesion by building a clot. Their physiological function to maintain hemostasis is crucial since abnormalities might lead to excessive bleeding during and after injury. Deep understanding of platelet activation and aggregation is only possible if all proteins involved are known. We decided to measure the murine proteome since mice represent a valuable model to study disease, in particular in knock-out mice. Using MS-based proteomics in combination with SILAC-PrESTs we identified nearly all platelet proteins and their estimated their copy numbers, paving the way for further functional studies.

Since this study was performed in mice, we established the PrEST production independently of the Human Protein Atlas project, which is specific to the human system. This also gave us the liberty to design the PrEST sequences to be optimal for mass spectrometric analysis, specifically by selecting unique regions having many tryptic peptides.

The study is currently in revision at Molecular and Cellular Proteomics:

'Copy number analysis of the murine platelet proteome spanning the complete abundance range'

Marlis Zeiler, Markus Moser and Matthias Mann

Copy number analysis of the murine platelet proteome spanning the complete abundance range

Marlis Zeiler[1], Markus Moser[2] and Matthias Mann[1*]

[1]Max-Planck-Institute of Biochemistry, Department of Proteomics and Signal Transduction, Am Klopferspitz 18, D-82152 Martinsried, Germany

[2]Max-Planck-Institute of Biochemistry, Department of Molecular Medicine, Am Klopferspitz 18, D-82152 Martinsried, Germany

*To whom correspondence may be addressed: Matthias Mann, Ph.: +49-89-8578-2557; Fax: +49-89-8578-2219; E-Mail: mmann@biochem.mpg.de

Running title: Copy number analysis of the murine platelet proteome

Abbreviations:

ABP – albumin binding protein, DTS – dense tubular system, ER - endoplasmic reticulum, FASP – filter aided sample preparation, GO – Gene ontology, GP – Glycoprotein, GPCR – G protein coupled receptor, iBAQ - intensity-based absolute quantification method, NSAF - normalized spectral abundance factor, OCS – open canalicular system, PCP – protein correlation profiling, PIP2 - phosphatidylinositol 4,5-bisphosphate, PLC – phospholipase C, PrEST – protein epitope signature tag, PRP – platelet rich plasma, SAX – strong anion exchange, SILAC – Stable isotope labeling by amino acids in cell culture, vWF - von Willebrand factor

Knowledge of the identity and quantity of expressed proteins of a cell type is a prerequisite for a complete understanding of its molecular functions. Mass spectrometry (MS)-based proteomics has allowed the identification of the entire protein complement of yeast and the close-to-complete set of proteins expressed in mammalian cell lines. Using recent technological advances we here characterize the proteome of murine platelets, key actors in mediating hemostasis and thrombosis. We accurately measured the absolute protein concentrations of thirteen platelet proteins by SILAC-Protein Epitope Signature Tags (PrESTs) and used them as reference points to estimate the copy number of all proteins of the platelet proteome. To distinguish contaminants such as plasma or erythrocyte proteins from true platelet proteins, we monitored protein abundance profiles across multiple purification steps. In total, we absolutely quantified 4,400 platelet proteins, with estimated copy numbers ranging from less than ten to about a million per cell. Stoichiometries derived from our data correspond well with previous studies. Our study provides a close-to-complete reference map of platelet proteins, which will be useful to the community, for instance for interpreting mouse models of human platelets diseases.

Platelets are cells derived from the cytoplasm of megakaryocytes, which are found in bone marrow and constantly produce and release platelets into the blood circulation. In the blood they circulate and survey the integrity of the vasculature. Upon injury of the endothelium platelets prevent hemorrhages and uncontrolled blood loss by sealing the vascular lesions. The ability to form aggregates is important for their hemostatic function, however, pathological platelet activation, for example during rupture of an atherosclerotic plaque, may reduce blood supply to the heart or brain during vascular occlusion and thereby induce cardiac infarction or stroke. It is therefore important to understand the molecular processes that control platelet activation and aggregation and to develop new therapeutic strategies to block critical platelet proteins involved in these processes [1]. A deeper, quantitative understanding of the platelet proteome will facilitate the identification of new drug targets and therefore the development of novel anti-platelet therapies.

With a diameter of only 0.5 -1 μm in mouse and 2-5 μm in human, platelets are the smallest blood cell type and have a very short life span of 3-4 days (human 7-10 days). Platelets lack a nucleus and therefore there is no transcription that could replenish their

residual megakaryocyte-derived mRNA. As a consequence mRNA levels are very low. Nevertheless, platelets translate mRNA into protein upon activation, however, whether this is important for platelet function is not clear [2]. The low mRNA levels make transcriptomics challenging because even a minimal contamination of the platelet sample by nucleated cells could make a substantial contribution to the measured transcriptome. In addition, functional interpretation of the measured transcript levels is complicated by the fact that they may reflect the parental megakaryocyte transcriptome rather than platelet specific processes [3]. Despite these difficulties, several studies have measured mouse and human platelet transcriptomes leading to the identification of approximately 6,500 and 9,500 transcripts, respectively [4]. In contrast to transcriptomics, proteomics approaches are intrinsically better suited to understand the cellular functions of platelets, because proteins are the biochemical functional units. Furthermore, they are the drug targets in antithrombotic or antiplatelet therapy [1]. Historically, studies of the platelet proteome have been performed by two-dimensional gel electrophoresis and typically have quantified up to several dozens of proteins [5]. This approach has now been superseded by mass spectrometry (MS)-based proteomics with much higher resolution and mass accuracy. The high peptide sequencing speed of modern instrumentation, combined with other technological advances enables the mapping of close-to-complete proteomes with high confidence, despite the broad dynamic range of protein quantities expressed [6, 7]. Recently, Burkhart et al. employed modern mass spectrometric instrumentation to confidently identify the to date deepest proteome of about 4,000 human platelet proteins [8]. Based on the tendency of the shotgun proteomics workflow to identify peptides from more abundant proteins more frequently (spectral counting), the authors were able to derive a quantitative measure of the majority of the identified proteome. These values were then scaled to copies per cell through a literature review of absolute copy number measurements from diverse sources, such as quantitative Western blotting.

We have recently developed a method for absolute protein quantification, in which we produced isotope labeled recombinant protein fragments (PrESTs) in *E. coli* and combined them with stable isotope labeling in cell culture (SILAC) [9, 10]. In this SILAC-PrEST method, protein fragments are expressed as fusion proteins with the Albumin Binding Protein (ABP) as a solubility tag. Upon purification their absolute concentrations are determined in relation to a common sample of ultra-pure ABP, whose concentration has previously been measured by amino acid analysis. The heavy PrESTs with known concentration are then spiked into cell lysates and the SILAC ratios of several peptides enabled calculation of the cellular concentrations and copy numbers of their endogenous protein counterparts. Using this approach, we previous were able to quantify HeLa cell proteins with copy numbers ranging from thousands to several millions per cell [10].

In this study we set out to analyze the murine platelet proteome by the high resolution, quantitative methods developed in our laboratory [11, 12]. Using a quadrupole Orbitrap mass spectrometer [13], we obtained label-free quantification values for more than 4,400 proteins. These values were converted to copy numbers per platelet using PrESTs of 13 proteins for calibration. Furthermore, to distinguish true platelet proteins from contaminants experimentally, we followed their decreasing intensity profile through successive stages of purification. Our accurate and quantitative picture of a mammalian platelet proteome shows that it is much larger than might have been expected from its specialized functions.

EXPERIMENTAL PROCEDURES

Platelet preparation - Mice (strain C57BL/6) were bled under anesthesia from the retro-orbital plexus and approximately 1ml blood was collected using heparin (20 U/ml in TBS) as the anticoagulation reagent. Blood was then centrifuged at 100 g for 7 min to obtain the platelet rich plasma (PRP), which we termed 'crude fraction'. The PRP was centrifuged at 700 g to concentrate the platelets in the top layer (termed 'purified fraction'). This procedure was repeated once to obtain the 'highly purified fraction'. Eventually the platelet pellet was resuspended in 1 ml of Tyrode's buffer containing PGI_2 and apyrase, followed by a centrifugation step (Fig. 1a) (termed 'ultra-purified fraction'). For protein correlation profiling since higher sample amounts are required we mixed blood from different mice and then performed the extensive purification described above. For each sample 30 x 10^6 platelets each were resuspended in lysis buffer (2 % SDS, 100 mM Tris pH 7.5, 100 mM DTT), boiled at 95 °C and further processed using the FASP method [11]. In brief, SDS was exchanged to urea on a 30 kDa filter. Peptides were eluted after digestion with trypsin and subjected to a StageTip-based [14] Strong Anion exchange (SAX) fractionation [12]. Platelet counts were determined using a Hemavet950 analyzer (Drew Scientific).

Absolute Quantification - For absolute protein quantification using the SILAC-PrEST method [10], the synthetic genes of the protein standards were fused to Albumin Binding Protein (ABP). The murine PrESTs were designed to be optimal for mass spectrometric analysis, specifically we selected unique regions having many tryptic peptides also allowing us to distinguish isoforms. SILAC standards were produced using an auxotrophic *E. coli* strain [15] in the presence of heavy arginine ($^{13}C_6{}^{15}N_4$) and heavy lysine ($^{13}C_6{}^{15}N_2$). The recombinant His-tag containing proteins were purified via Ni-NTA columns and quantified with aliquots of a light ABP preparation, on which amino acid analysis had been performed. Next, the thirteen protein standards were mixed and spiked into the lysed platelets at approximately endogenous concentration. The sample was further processed using the FASP method [11].

Mass spectrometry - The samples were eluted from the stage tip, resuspended in buffer A* (2 % acetonitrile (ACN), 0.1 % trifluoroacetic acid (TFA)) and loaded onto a fresh 50 cm C_{18} column packed with Reprosil-Pur 1.9 μm resin (Dr. Maisch GmbH). The samples were separated on a UPLC system using a 180 min gradient ranging from 5 to 30 % buffer B (80 % ACN, 0.1 % TFA) at a constant flow rate of 250 nl/min and injected via a nanoelectrospray ion source into the mass spectrometer. We used a quadrupole Orbitrap mass spectrometer (Q Exactive [13], Thermo Fisher Scientific) in a data dependent fashion, acquiring a full scan (300 – 1750 m/z, 70,000 resolution at m/z 200, target value 3e6 ions, maximum fill time 20 ms) and up to ten subsequent MS/MS scans (17,500 resolution, target value 1e5 ions, maximum fill time 120 ms) using higher energy collision fragmentation (HCD) for peptide identification [13]. Xcalibur software (Thermo Fisher Scientific) was used to acquire data.

Data Analysis - The raw data was analyzed using MaxQuant version 1.4.1.4 [16] with the integrated search engine Andromeda [17]. For peptide identification, the fragmentation spectra were searched against the UniProt mouse database (downloaded in May 2013) containing 50,829 entries to which 247 common contaminants were added. A 'software lock mass' was used to recalibrate and improve the mass accuracy of the precursor masses [18]. During the main search the maximum allowed initial mass deviation of the precursor ions was set to 4.5 ppm, while the maximum mass deviation of the fragmentation ions was set to 20 ppm. Cysteine carbamidomethylation was set as 'fixed modification' and N-terminal acetylation and methionine oxidation as 'variable modification'. The 'Enzyme' parameter was set as Trypsin, for which N-terminal cleavage to proline and two miss-cleavages were allowed. A minimum of 7 amino acids were required for valid peptide identification. In addition to the standard peptide search, the 'second peptide' identification and the 'match between run' options were enabled in Andromeda. For statistical evaluation of the data, a posterior error probability and a false discovery rate cut off (determined by target-decoy searching) of 0.01 were used for peptides and proteins. For SILAC quantification the MaxQuant standard settings were applied

requiring at least two ratio counts between SILAC peptides pairs. Bioinformatics analysis was performed with the Perseus tool (version 1.4.1.4) available with the MaxQuant environment. For clustering and subsequent identification of contaminating proteins from other cell types and plasma we used label-free intensities [19]. In this case we replaced missing values using data imputation which is based on the assumption that missing values are caused by the detection limit of MS measurement [20]. First, we determined the Gaussian distribution of the logarithmized data and next we used a normal distribution with adjusted mean and standard deviation in order to simulate signals of low abundant proteins. We chose parameters (width = 0.3, downshift = 1.8) where the distribution of the imputed values was placed at the lower end of the distribution of measured values. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org) via the PRIDE partner repository [21] with the dataset identifier PXD000747.

RESULTS AND DISCUSSION

Removal of contaminant proteins via protein correlation profiling - Platelets are highly abundant in blood and can be prepared in high yields by differential centrifugation, but proteins from lysed erythrocytes and plasma are inevitably present in platelet fractions. To distinguish true platelet proteins from such contaminants, we adapted the protein correlation profiling approach (PCP) [22, 23]. Originally, PCP was developed to define organelle proteomes by quantifying the distribution of various organelle marker proteins across different subcellular fractions, and subsequent matching of proteins with profiles similar to the marker. Here, we instead followed protein abundance profiles across purification fractions to identify the contamination profile. Mouse platelets were separated from erythrocytes and plasma via multiple centrifugation and washing steps (Fig. 1a). Aliquots were taken at each step (crude, purified, highly purified, ultra-purified fractions, see EXPERIMENTAL PROCEDURES) to perform protein correlation profiling. All samples were measured on a Q Exactive mass spectrometer after FASP sample

preparation and SAX fractionation [12]. To ensure meaningful profiles, we required at least three quantification values either from the three different mice and the highly pure platelet preparation or from the three unpurified fractions. After this stringent filtering, 4,585 protein groups remained.

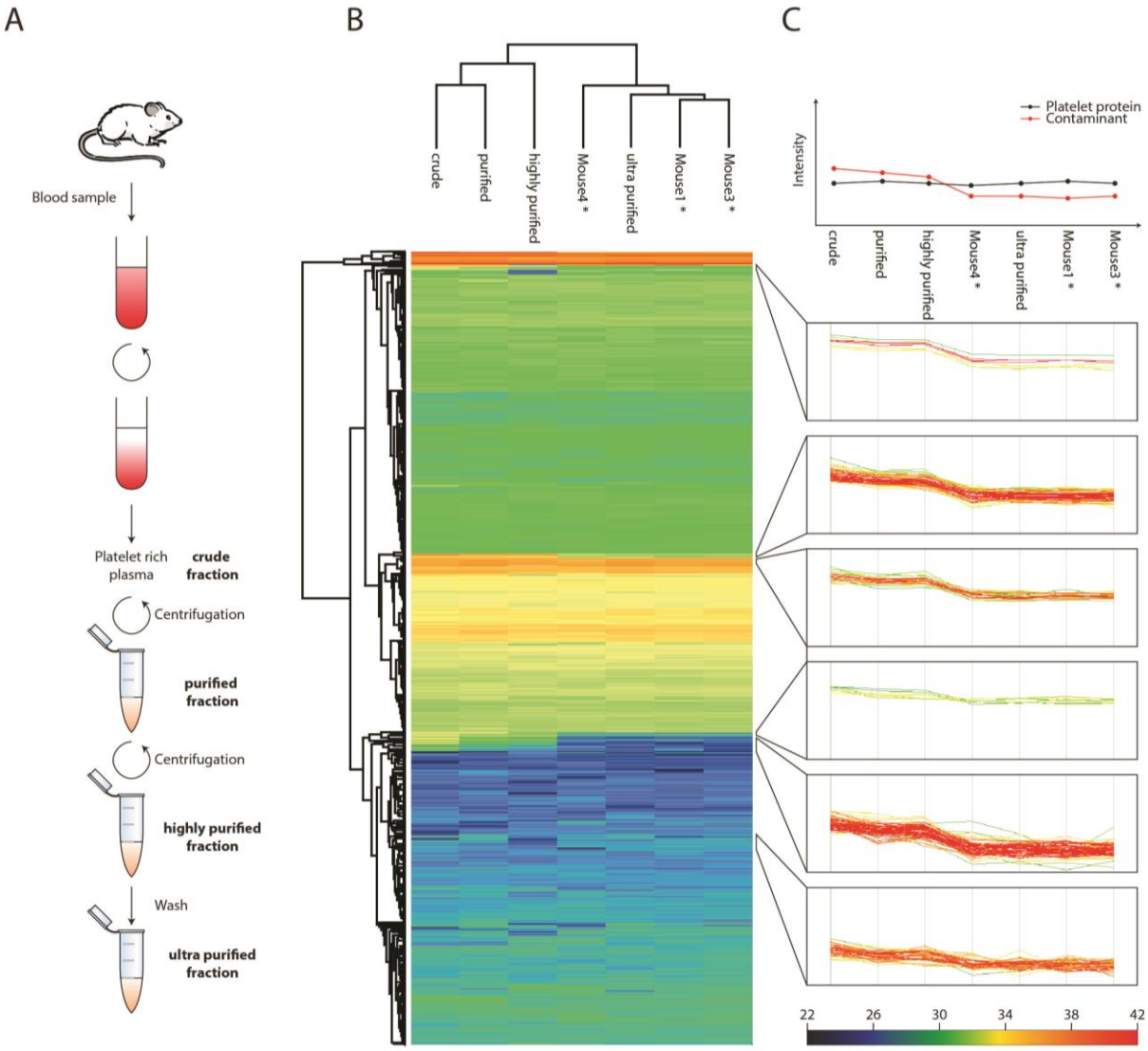


Figure 1: Platelet preparation and protein correlation profiling (A) The platelet preparation workflow: platelets were purified by the depicted centrifugation and washing steps. For the Protein Correlation Profiling several mice were pooled and at each purification step a sample was taken. For the 'mouse' samples only at the ultra-purified fraction a sample was collected (* = ultra-purified) (B) Unsupervised hierarchical clustering of protein abundance profiles of the different stages of purity (C) Prototypical abundance profile of contaminating proteins (upper panel) and actual profiles of contaminating proteins at every abundance level.

To identify proteins that have similar abundance profiles across different purification steps, we first performed unsupervised hierarchical clustering (Fig. 1b) of the label-free intensities (Fig. 1c). From the hierarchical clustering, we observed two major branches of the dendrogram grouped based on the different levels of purity— all the 'ultra-purified' mouse replicates grouped together, separate from the cruder samples (Fig. 1b). Intriguingly, groups of proteins at each abundance level had similar profiles over the different samples (six clusters in total in six abundance ranges). By cluster analysis and by comparing protein profiles to known plasma proteins, we identified 191 contaminant proteins. Since there was a gradual decline of protein intensities of the contaminations between the highly purified and ultra-purified, we additionally perform a Welch's t-test and detected 191 significant outliers at an FDR of 0.05 (Fig. 2a). Interestingly, these outliers covered 90 % of the 191 proteins identified via cluster analysis (18 unique proteins each); together amounting to 209 contaminants, of which 55 % (115) are 'secreted' according to the Uniprot keyword annotation (Fig. 2b). These include apolipoproteins, serine protease inhibitors, antibodies and complement factors, confirming that plasma was the main source of contaminant proteins. However, contaminants also included known erythrocytes markers such as erythrocytic spectrin, erythrocyte membrane protein band 4.2 and carbonic anhydrase 1 and 2, indicating that erythrocytes were another source of contamination besides plasma (Supplementary Table 1).

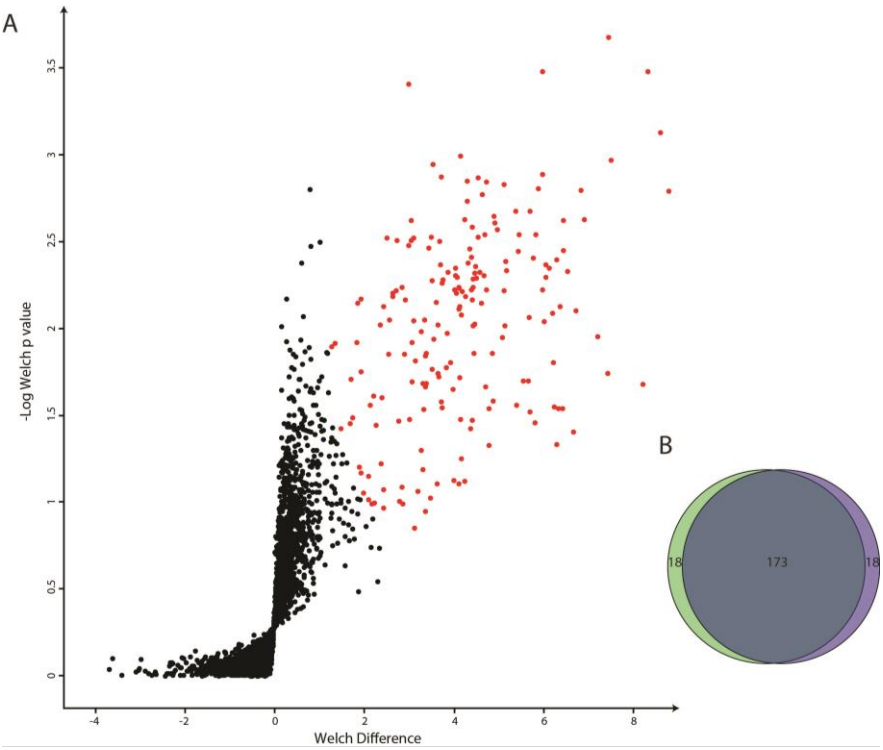


Figure 2: Identification of the contaminations via Welch's t-test. (A) Volcano plot of one-sided Welch's t-test, statistically significant (FDR of 0.05) hits are marked in red. (B) Venn diagram depicting the overlap between hierarchical clustering and Welch's t-test.

However, we note that protein correlation profiling does not solve the problem of contaminating proteins completely. Due to the platelet's canalicular system and 'sponge-like' surface as well as their uptake of plasma through vesicles [24], some of these proteins remain bound to or are even taken up by platelets making them difficult to identify as contaminants.

The platelet proteome with accurate absolute protein quantification - To determine a reference map of the platelet proteome we performed a second quantitative analysis. In these experiments we performed the same purification strategy but exclusively measured quantitative data on the 'ultra-purified' platelets from three different mice. The different mice provided the range of variation. The final mean copy number values were determined as a combination of these values and the ultra-purified fraction of the combined blood sample to obtain the best possible accuracy. To evaluate and visualize the

quality of our measurements, we plotted label-free protein abundance values of different replicates against each other (Figure 3). This resulted in a Pearson correlation coefficient of R=0.98, indicating excellent performance of our workflow and of MaxQuant's label-free algorithm as well as consistency between the inbred mice. Plotting protein abundances of the ultra-purified versus the crude fractions highlighted contaminants previously identified through protein correlation profiling and Welch's t-test as outliers from the trend line (depicted in red in Figure 3). Once we eliminated the contaminant proteins, a final platelet proteome of 4,376 protein groups was obtained (Supplementary Table 2).

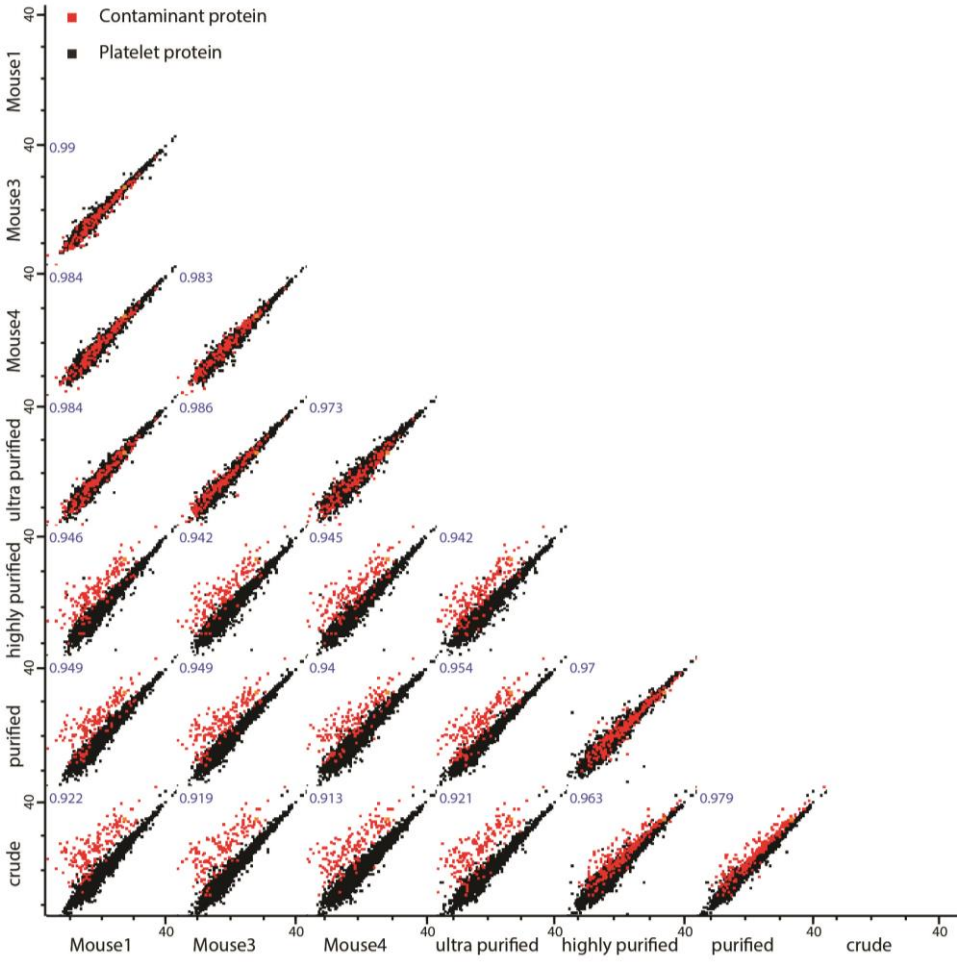


Figure 3: Reproducibility of the proteomic measurements. Scatter plot of the proteins comparing replicates showing reproducibility as indicated by the Pearson correlation coefficients. Contaminant proteins are offset in the semi-purified samples because they are progressively de-enriched.

Next, we wished to determine the copy numbers of each platelet protein. To this end, we first selected thirteen platelet proteins (Table 1), which cover a wide abundance range of the platelet proteome. Absolute protein quantification was achieved using the recently developed SILAC-PrEST quantification method [10]. Briefly, from each of the thirteen proteins we selected approximately 150 amino acids of unique sequence, containing multiple tryptic peptides. We recombinantly expressed these PrESTs with a purification and a solubility tag and quantified these standards in a SILAC experiment (see EXPERIMENTAL PROCEDURES). Next, we mixed the PrESTs with known concentration in appropriate ratios to obtain a master mix [10]. This heavy mix was combined with the platelet lysate allowing multiplexed concentration determination of the corresponding endogenous proteins via their SILAC ratios (Fig. 4b). To quantify all other platelet proteins their intensities were then normalized and scaled by the number of theoretically observable peptides through the intensity-based absolute quantification (iBAQ) algorithm incorporated in the MaxQuant software, using the PrESTs as the iBAQ standards [25]. The copy numbers reported are the mean of three different ultra-purified mice and the ultra-purified fraction from the protein correlation profiling. The coefficient of variation (in %) for each protein copy number was calculated from the four measurements and is reported in Supplementary table 2. The overall median coefficient of variation of the all copy numbers between the different mouse replicates is 20.6 %.

To assess the accuracy of iBAQ-derived protein copy numbers we plotted them against the values measured via SILAC-PrEST. The estimates from iBAQ quantification agreed well with the measured values using the PrESTs, with estimated copy numbers generally located within an average 1.8-fold range, depending on protein abundance (Figure 4d). The data from the two different methods concurred well with a high correlation coefficient of 0.94, indicating the reliability of the copy numbers determined by interpolation.

Table 1: Selected platelet proteins for reference points

Gene Name	Protein Name	Copy Number	Coefficient of Variation (%)
Akt1	RAC-alpha serine/threonine-protein kinase	500	35.6
Akt2	RAC-beta serine/threonine-protein kinase	700	40.9
Akt3	RAC-gamma serine/threonine-protein kinase	1,600	109.0
Fermt3	Fermitin family homolog 3	286,700	3.5
Fyn	Tyrosine-protein kinase Fyn	3,400	26.6
Gp6	Platelet glycoprotein VI	19,700	33.9
Itgb3	Integrin beta-3	7,300	7.9
Itgb2	Integrin beta-2	136,300	4.8
P2rx1	P2X purinoceptor 1	5,500	24.7
Prkca	Protein kinase C	24,900	16.6
Prkg1	cGMP-dependent protein kinase 1	2,100	29.5
Selp	P-selectin	42,200	31.3
Tln1	Talin-1	267,000	4.1

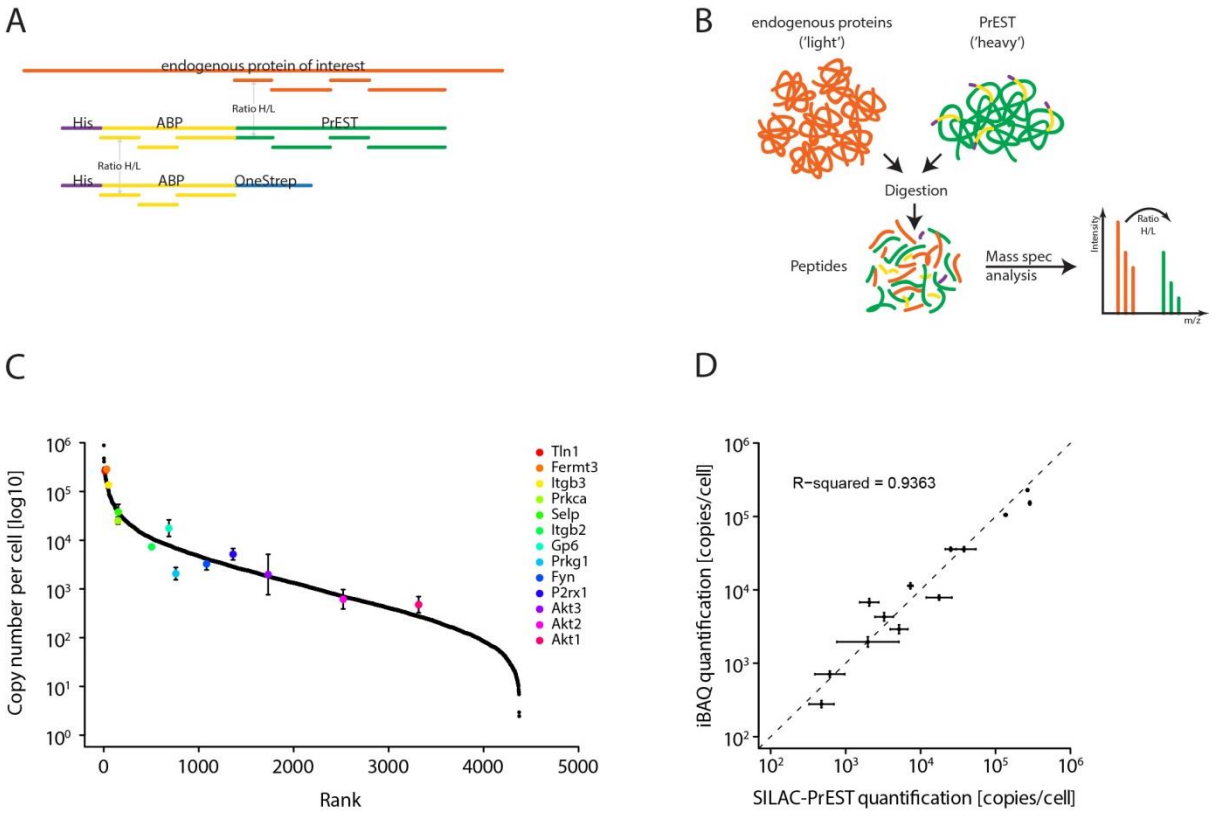


Figure 4: Absolute protein quantification (A) Quantification of the PrEST. Highly purified ABP are quantified by amino acid analysis. The concentration of the SILAC-labeled PrESTs is determined using the common ABP derived tryptic peptides. (B) In a separate experiment, previously quantified PrESTs are multiplexed and spiked into the platelets lysate. The sample is processed via filter-aided sample

preparation (FASP) and the concentration of the endogenous proteins is read out using the SILAC ratio determined in the mass spectrometer. (C) Thirteen different PrESTs were used to quantify proteins of interest over the whole abundance range. (D) Comparison of protein copy numbers of thirteen selected proteins obtained using SILAC-PrEST to those calculated using iBAQ method.

Adhesion and signaling receptors – Having protein copy numbers by hand we set out to examine important adhesion and signaling receptors. The primary function of platelets is to control hemostasis by scanning the blood system for vascular lesions. Most platelets never adhere or aggregate during their entire lifetime and are degraded in the reticuloendothelial system of the spleen or liver. In the case of a vessel lesion, however, different platelets receptors mediate adhesion, aggregation and signaling [26]. Figure 5 depicts the major adhesion and signaling receptors of platelets.

The GPIb-IX-V complex consists of four proteins (GPIbα:GPIbβ:GPIX:GPV) and represents the second most important platelet receptor after αIIbβ3 integrin. Its absence or deficiency in human leads to Bernard- Soulier syndrome, the second most common bleeding disorder [27]. It is responsible for initial platelet recruitment and platelet activation and is the major receptor for von Willebrand Factor (VWF), a protein binding to subendothelial collagen whose absence is the most common cause for coagulation deficiency. The GPIb-IX-V complex associates in a certain stoichiometry, which was thought to be 2:2:2:1 (GPIbα:GPIbβ:GPIX:GPV). However, Li and co-workers [28] recently reported that GPIbβ is bound via two disulfide bridges, which would lead to two GPIbβ molecules for one GPIbα protein. Indeed, the stoichiometry of 1.3:4.4:1.8:1 found in our study agrees with this conclusion. Given this stoichiometry, we find 36,000 ±5000 GPIb-IX-V complexes to be present in each platelet on average.

Integrins represent the major class of adhesion and signaling receptors and are expressed in almost all cells. They consist of α and β subunits and only correctly folded and associated heterodimers are transported to the plasma membrane. Upon activation integrins shift towards an active conformation with high affinity for their ligands [29]. Platelets express six different integrins, αIIbβ3, αvβ3, α2β1, α5β1, α6β1 and αLβ2, each

binding to different ligands. Mutations in the αIIb or β3 genes cause Glanzmann thrombostenia, the most common bleeding diathesis in patients. As expected from its prominent cellular and disease roles, the αIIbβ3 integrin was measured as the most abundant adhesion receptor with 110,000 ±7,900 to 130,000 ±10,100 copies per platelet estimated from our dataset. The α2β1 integrin is the second most important integrin, which mediates adhesion to collagen. We find it expressed at 18,000 ±1,300 copies per platelet. The laminin receptor α6β1 integrin is involved in binding to laminins of the basement membrane, which is not thought to be of equal importance to the binding of platelets to collagen. Therefore it was surprising to find that the laminin receptor was expressed at similar levels (20,000 ±1,500 copies). This indicates that it may have a so far underestimated role in platelet adhesion to subendothelial lesions, which expose laminins of the basement membrane. The expression levels of the other integrins are listed in Table 2. Our dataset reflects the 1:1 stoichiometry of the various integrin complexes within the expected accuracy, further supporting the validity of our copy number estimates.

Table 2: Integrin Stoichiometry

Integrin subunit	Protein Copy Number	Collagen receptor	Fibrinogen receptor	Vitronectin receptor	Fibronectin receptor	Laminin receptor	unknown role
Itga2b	106,600		106,600				
Itgß3	131,700		~131,000*	~600			
Itga5	1,900				1,900		
Itga2	17,600	17,600					
Itgß1	30,900	~13,530			~1,460	~15,910	
Itga6	20,700					20,700	
Itgav	500			500			
ItgaL	5,800						5,800
Itgß2	11,000						11,000

*Integrins are heterodimers. Copy numbers of each of these (Collagen receptor, Fibrinogen receptor etc.) were estimated from the measured copy numbers of their alpha subunits. The proportion of beta subunits in each integrin receptor was allocated to this ratio (indicated by ~ symbols).

Glycoprotein VI (GPVI) represents an Ig superfamily member and is the other major collagen receptor. We directly quantified it by a SILAC-PrEST leading to a copy number of 19,000 ± 6,700 per platelet. Interestingly, both collagen receptors are expressed at similar levels. GPVI forms a complex with the common FcRg chain (75,000 ±15,400 copies), which is critical for the signaling process. Although GPVI may also have a role in platelet adhesion to collagen, its major role is platelet activation via phospholipase Cγ2 (11,000 ±1,300 copies), which induces Ca^{2+} release and the activation of α2β1 integrins. Collagen-bound GPVI and α2β1-derived signals are both required for full platelet and αIIbβ3 integrin activation [30]. The knowledge of copy numbers of the different collagen receptors might help to understand their roles in collagen adhesion and subsequent signaling.

Beside these adhesion receptors, platelets contain a number of other receptor types. As many of them are thought to be expressed at very low levels, it is remarkable that they were identified and quantified in our measurements without any special enrichment step (Fig. 5). G-protein-coupled receptors belong to the seven transmembrane receptors and represent the major agonist receptor family in platelets. Protease-activated receptors Par4 (5,200 ±1,500) and Par3 (3,900 ±1,400), which reside on mouse platelets (in contrast to human platelets, where PAR-1 and PAR-4 are expressed), are activated by thrombin, one of the most potent platelet agonist. ADP is another important platelet agonist, which is secreted from platelet dense granules and acts as an autocrine agonist. It is bound by the two purinergic receptors, $P2Y_1$ (1,000 ±100) and $P2Y_{12}$ (2,700 ±500). Many FDA-approved antiplatelet drugs such as Ticopidine, Clopidogrel, Prasugrel target $P2Y_{12}$ and cause significant inhibition of ADP-mediated platelet activation [1]. Other GPCRs important in transmitting signals from thromboxane, prostacyclin (PGI_2), and epinephrine are thromboxane A2 receptor (1,900 ±800), prostanoid IP receptor (2,200 ±800), and alpha-2A adrenergic receptor (140 ±60), respectively. The serotonin receptor 5HT2A is also known to be expressed on platelets; however, because we only quantified it in two of three

replicates (with average copy number 175 ±20), it did not pass our stringent threshold for copy number assignment.

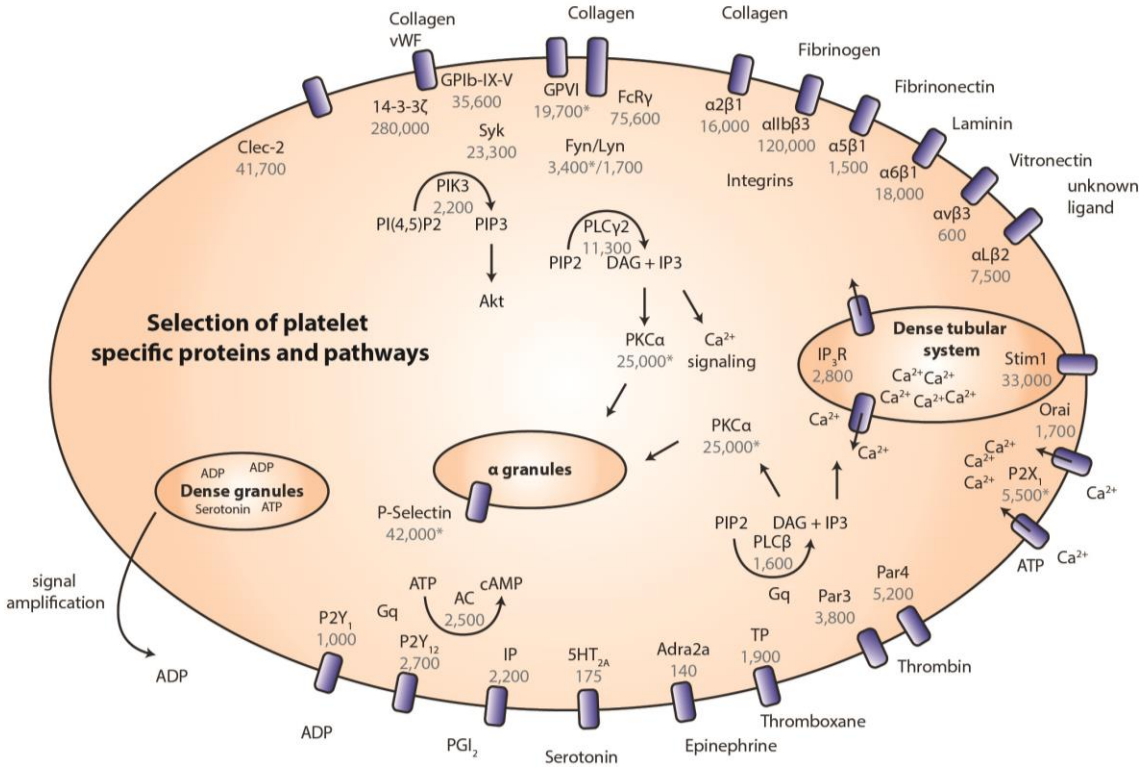


Figure 5: Selection of platelet specific proteins and pathways. Overview of major adhesion and signaling receptors of platelets as well as downstream signaling pathways. Copy numbers derived in this study are depicted in grey (Copy numbers with an asterisk were measured with SILAC-PrEST quantification).

One central step in platelet activation triggered by the above mentioned adhesion and signaling receptors is an increase in intracellular calcium levels. This is mediated by phospholipase C hydrolyzing PIP_2 to IP_3, which in turn induces the release of Ca^{2+} from the dense tubular system (DTS). Depletion of Ca^{2+} from the DTS activates Stim1 (33,000 ±1,900), causing it to relocate to the plasma membrane to open the ion channel Orai1 (1,700 ±300) resulting in Ca^{2+} influx from plasma. Another important Ca^{2+} channel is the P2X purinoceptor, which is activated independently by ATP. We measured 5,500 ±1,400 copies of this receptor via direct PrEST quantification. Increased calcium levels are for instance crucial for αIIbβ3 integrin activation, leading to platelet adhesion and

aggregation. Knowledge of the copy numbers of these proteins may be informative for pharmacological approaches to target calcium signaling.

Determination and quantification of isoforms - Many protein families include different isoforms, which may have distinct functions. Quantification of isoforms by shotgun proteomics is complicated by the fact that peptides may be shared between these isoforms. Here, we restricted our analysis to cases in which the majority of peptides were sequence unique peptides for the isoform in question.

In platelets, two predominant phospholipase C isoforms can be found: Beta (β1 – 220 ±30 copies per cell, β3 – 1,600 ±200, β4 – 1,400 ±100) and gamma (γ1 – 60±20 , γ2 – 11,300 ±1,300). Depending on the upstream receptor a different PLC isoform is activated. In case of the gamma isoform, we can confirm that in platelets γ2 is by far the predominant isoform [31, 32]. Downstream of PLC, protein kinase C (PKC) is activated by diacylglycerol (DAG) and increase Ca^{2+}. We detected five different PKC isoforms: α (24,900 ±2,200), β (30 ±10), δ (4,000 ±200), θ (8,000 ±500), ε (1,600 ±300). By different knock-out studies (review in [33]) it was shown that the different isoforms play distinct roles in the platelet activation process. PKC α for example phosphorylates, among other substrates, proteins involved in the secretion of the α-granules and dense granules. Mice deficient in Prkca exhibit reduced secretion of ATP [34]. Other examples include the Akt kinases of which all three mammalian isoforms have been shown to occur in platelets [35]. Using SILAC-PrESTs we quantified all isoforms Akt1 (500 ±200), Akt2 (700 ±300) and Akt3 (1,600 ±1,700), and confirm that Akt3 is the most abundant isoform as reported recently [35]. It is known that they have different roles in platelet activation [36, 37]. Even though the Akt isoforms have 75% sequence homology (but not sequence identity), we were able to distinguish and quantify the Akt isoforms with multiple unique peptides. These examples indicate that shotgun proteomic data with sufficiently high coverage enables the determination of expression level differences of distinct isoforms.

Global view of protein copy numbers - Altogether using our shotgun proteomics workflow, we identified and measured 4,376 platelet protein groups with average sequence coverage of 38.5%. Estimated abundances from our data span a very wide full dynamic range - six orders of magnitude - from two copies up to 900,000 copies per platelet (Fig. 6a). As expected, the most abundant platelet protein is the cytoskeletal protein actin with close to one million copies per platelet. Notably, one third of the identified platelet proteome (1,500 proteins) has copy numbers below 500. Interestingly, for several proteins we even estimated copy numbers less than 10 per platelet, however, measurements in this range are likely to be less accurate. The lowest copy number of a protein known to have an important function in the proteome, was the alpha-2A adrenergic receptor (A2AR) with 140±60 copies per platelet. Conversely, some proteins that we would have expected in our platelet proteome, such as nitric oxidase have not been identified.

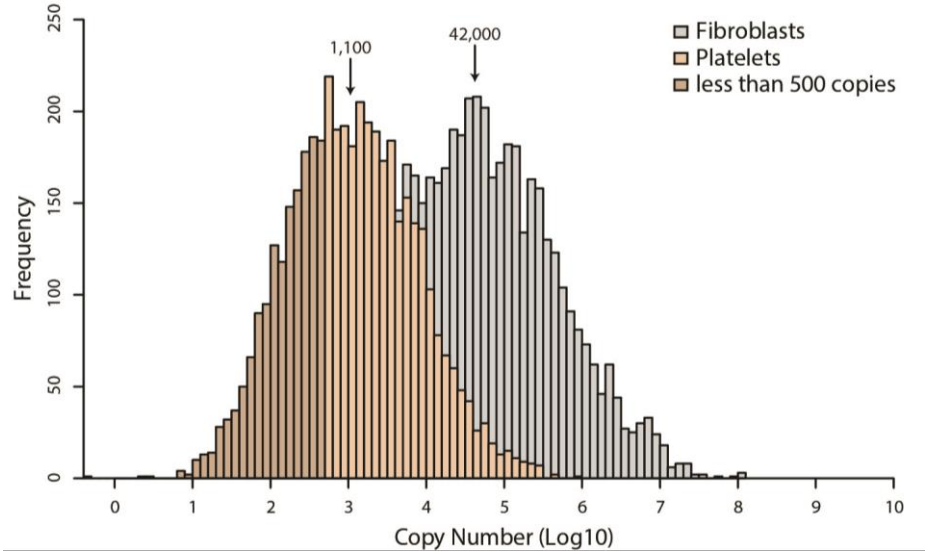


Figure 6: Histogram of all copy numbers. Comparison between the protein abundance distribution of fibroblast and platelet copy numbers. Proteins below 500 are depicted in dark brown.

We were intrigued by the high proportion of proteins with low copy numbers and asked if they were likely to be functional. To investigate this, we did a bioinformatic enrichment

analysis of groups or pathways that were statistically significantly overrepresented. To this end we performed a so called 1D enrichment analysis (with abundance being the one dimension) on the basis of annotation terms such as Uniprot keywords or Gene Ontology (GO) in the Perseus software, which is part of MaxQuant [38] (Supplementary Table 3). It employs a two-sample Welch's t-test for difference of means of the category versus all remaining proteins and yields categories with statistically significance and therefore does not require arbitrary cutoffs for abundance categories. We then did the same analysis for fibroblasts, for which absolute copy numbers have been reported recently (note that fibroblasts are much larger and therefore copy numbers are in general much higher [25]). First, we could confirm that protein categories that were overrepresented in the high abundance region were platelet specific protein groups since these annotation terms comprised 'fibrinogen complex' (mean copy number value 220,000), 'platelet alpha granule' (140,000) and 'platelet activation' (120,000). In contrast, diverse categories such as 'ribosome assembly' or 'ribosomal subunits' (mean copy numbers 6,000,000 to 10,000,000) were enriched in the high abundant region of fibroblasts. Another abundant hit in fibroblasts was 'nucleosome' with a mean copy number of approximately 4,000,000. This unbiased analysis directly confirms the different functional processes in platelets vs. fibroblasts – one cell type with a specialization on cell adhesion and associated signaling but little translation in contrast to a dividing cell line allocating cellular resources to protein synthesis.

In the low abundance region of platelet distribution, chromatin binding (mean copy number of 600), tRNA processing (500), RNA-mediated gene silencing (300), Zinc finger proteins (400), nucleic acid binding (600) and different categories of proteins involved in the ubiquitination pathway were detected. Fibroblasts, in contrast did not show any of these enrichments in the low abundance region. Enrichment in the low copy number region of the platelet proteome is not surprising as platelets lack a nucleus and are therefore not expected to have a functional requirement for proteins involved in transcription, DNA repair or replication. Indeed, the fact that low amounts of these

proteins are nevertheless clearly quantifiable, suggests that they are vestiges of the budding process from the megakaryocytes. The presence of only very low copy numbers of proteins in the ubiquitylation process may reflect a relative absence of protein degradation – the mirror image of low or absent translation.

CONCLUSIONS AND OUTLOOK

In this study, we performed a deep and quantitative proteome analysis of murine platelets. Platelets are purified from blood therefore it is crucial to discriminate between false positive proteins and contaminations from erythrocytes and plasma. To differentiate platelet proteins from common contaminations deriving from sample preparation, we successfully applied a variant of the protein correlation profiling approach based on the fact that true platelet proteins should be enriched, whereas contaminating proteins disappear during purification. After subtracting these, 4,400 platelet proteins were confidently identified. Notably, this proteome depth was achieved without laborious fractionation, enrichment for membrane proteins or post-translationally modified peptides.

Absolute quantification using the recently published SILAC-PrEST method allowed us to accurately quantify a number of platelet proteins in an absolute manner, which then enabled us to estimate the copy numbers of all identified platelet proteins per cell. These copy numbers enabled us to examine complex stoichiometry. For instance we correctly determined the composition of the GPib-IX-V complex and of most integrins. We are confident that the stoichiometry of many complexes can be deduced from our data.

We noted that some of our copy numbers are higher than previously published, for example in the case of integrin αIIbβ3 which was reported to have 80,000 copies per unstimulated platelet. However, inspection of the literature revealed that only the receptors in the outer plasma membrane had been labeled and quantified [39]. It is known that a significant proportion (30%) of the receptor is stored in the open canalicular system

(OCS) as well as in platelet α granules [40], leading to an agreement with our SILAC-PrEST measurement of 120,000 copies per platelet. In general, many receptors are stored on alpha and dense granules and only incorporated into the plasma membrane upon activation. Since we measure the total protein content, some copy numbers might not agree with those numbers, which have been generated from surface measurements. Furthermore, to the extent that copy numbers are known, they have often been determined in human platelets, making them not directly comparable.

Due to the fact that most murine platelet proteins fulfill identical functions as their human counterparts and mice represent valuable model organism to study human pathology, which is also demonstrated by numerous genetically modified mice mimicking human platelet diseases, our mouse platelet proteome study may further help in understanding the role of individual proteins in platelet physiology. At the same time significant differences between platelets from mice and men exist, such as platelet size, counts and expression of certain critical proteins (e.g. the above-mentioned PAR1 in human vs. PAR3 in mice). Thus, our data can also serve to further study and follow up differences in protein expression and isoforms.

In conclusion, we here took advantage of recent advances in MS-based proteomics and quantification techniques to achieve a very large and high confidence dataset of platelet proteins. Estimated absolute quantification to a level down to single copy per cell should provide a basis for further characterization of platelets and their functionality.

ACKNOWLEDGEMENTS

We thank the Department for Signal Transduction and Proteomics for help and fruitful discussions, in particular Jeffrey Liu and Georg Borner for careful reading of the manuscript as well as Katharina Zettl for excellent technical assistance. This work was supported by the European Commission's 7th Framework Program PROSPECTS (grant

agreement HEALTH-F4-2008-201648) and the 'Deutsche Forschungsgemeinschaft' SFB914 (project A1).

REFERENCES

1. Michelson, A.D., *Antiplatelet therapies for the treatment of cardiovascular disease.* Nat Rev Drug Discov, 2010. 9(2): p. 154-69.
2. Weyrich, A.S., et al., *Protein synthesis by platelets: historical and new perspectives.* J Thromb Haemost, 2009. 7(2): p. 241-6.
3. McRedmond, J.P., et al., *Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes.* Mol Cell Proteomics, 2004. 3(2): p. 133-44.
4. Rowley, J.W., et al., *Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes.* Blood, 2011. 118(14): p. e101-11.
5. Zufferey, A., et al., *Platelet proteomics.* Mass Spectrom Rev, 2012. 31(2): p. 331-51.
6. Altelaar, A.F., J. Munoz, and A.J. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics.* Nat Rev Genet, 2013. 14(1): p. 35-48.
7. 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.
8. Burkhart, J.M., et al., *The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways.* Blood, 2012. 120(15): p. e73-82.
9. 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.
10. Zeiler, M., et al., *A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines.* Mol Cell Proteomics, 2012. 11(3): p. O111 009613.
11. Wisniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nature methods, 2009. 6(5): p. 359-62.
12. Wisniewski, J.R., A. Zougman, and M. Mann, *Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome.* J Proteome Res, 2009. 8(12): p. 5674-8.
13. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.* Mol Cell Proteomics, 2011. 10(9): p. M111 011015.

14. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. 2(8): p. 1896-906.

15. Matic, I., et al., *Absolute SILAC-compatible expression strain allows Sumo-2 copy number determination in clinical samples.* J Proteome Res, 2011. 10(10): p. 4869-75.

16. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nature biotechnology, 2008. 26(12): p. 1367-72.

17. Nagaraj, N., et al., *Deep proteome and transcriptome mapping of a human cancer cell line.* Mol Syst Biol, 2011. 7: p. 548.

18. Cox, J., A. Michalski, and M. Mann, *Software lock mass by two-dimensional minimization of peptide mass errors.* Journal of the American Society for Mass Spectrometry, 2011. 22(8): p. 1373-80.

19. Cox, J., et al., *MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction.* Mol Cell Proteomics, 2014.

20. Deeb, S.J., et al., *Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles.* Mol Cell Proteomics, 2012. 11(5): p. 77-89.

21. Vizcaino, J.A., et al., *The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013.* Nucleic Acids Res, 2013. 41(Database issue): p. D1063-9.

22. Foster, L.J., et al., *A mammalian organelle map by protein correlation profiling.* Cell, 2006. 125(1): p. 187-99.

23. Andersen, J.S., et al., *Proteomic characterization of the human centrosome by protein correlation profiling.* Nature, 2003. 426(6966): p. 570-4.

24. Berger, G., J.M. Masse, and E.M. Cramer, *Alpha-granule membrane mirrors the platelet plasma membrane and contains the glycoproteins Ib, IX, and V.* Blood, 1996. 87(4): p. 1385-95.

25. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. 473(7347): p. 337-42.

26. Michelson, A.D., *Platelets*. 2012: Elsevier Science & Technology Books.

27. Lopez, J.A., et al., *Bernard-Soulier syndrome.* Blood, 1998. 91(12): p. 4397-418.

28. Luo, S.Z., et al., *Glycoprotein Ibalpha forms disulfide bonds with 2 glycoprotein Ibbeta subunits in the resting platelet.* Blood, 2007. 109(2): p. 603-9.

29. Legate, K.R., S.A. Wickstrom, and R. Fassler, *Genetic and cell biological analysis of integrin outside-in signaling.* Genes Dev, 2009. 23(4): p. 397-418.

30. Petzold, T., et al., *beta1 integrin-mediated signals are required for platelet granule secretion and hemostasis in mouse.* Blood, 2013. 122(15): p. 2723-31.

31. Tate, B.F. and S.E. Rittenhouse, *Thrombin activation of human platelets causes tyrosine phosphorylation of PLC-gamma 2.* Biochim Biophys Acta, 1993. 1178(3): p. 281-5.

32. Blake, R.A., G.L. Schieven, and S.P. Watson, *Collagen stimulates tyrosine phosphorylation of phospholipase C-gamma 2 but not phospholipase C-gamma 1 in human platelets.* FEBS Lett, 1994. 353(2): p. 212-6.

33. Harper, M.T. and A.W. Poole, *Isoform-specific functions of protein kinase C: the platelet paradigm.* Biochem Soc Trans, 2007. 35(Pt 5): p. 1005-8.

34. Konopatskaya, O., et al., *PKCalpha regulates platelet granule secretion and thrombus formation in mice.* J Clin Invest, 2009. 119(2): p. 399-407.

35. O'Brien, K.A., et al., *An important role for Akt3 in platelet activation and thrombosis.* Blood, 2011. 118(15): p. 4215-23.

36. Woulfe, D., et al., *Defects in secretion, aggregation, and thrombus formation in platelets from mice lacking Akt2.* J Clin Invest, 2004. 113(3): p. 441-50.

37. Chen, J., et al., *Impaired platelet responses to thrombin and collagen in AKT-1-deficient mice.* Blood, 2004. 104(6): p. 1703-10.

38. Cox, J. and M. Mann, *1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data.* BMC Bioinformatics, 2012. 13 Suppl 16: p. S12.

39. Wagner, C.L., et al., *Analysis of GPIIb/IIIa receptor number by quantification of 7E3 binding to human platelets.* Blood, 1996. 88(3): p. 907-14.

40. Schmitt, A., et al., *Of mice and men: comparison of the ultrastructure of megakaryocytes and platelets.* Exp Hematol, 2001. 29(11): p. 1295-302.

SILAC-PrEST applied to protein complex stoichiometry

One of the major adhesion receptors on platelets that is crucial for activation and aggregation is the integrin αIIBβ3 complex. Integrins have the ability to change their conformation from an inactive state with low affinity to ligands to an active state. The activation of integrins ('inside-out' signaling) is mediated by two protein families: talins and kindlins. In platelets talin-1 and kindlin-3 directly bind to the tail of the integrin αIIBβ3 inducing the conformational change. Previously it was thought that talin alone plays a role in integrin activation, but recently it was shown that kindlin-3 is equally essential [113, 114]. Here, we investigated the interplay of these two proteins by generating mouse strains expressing different amounts of kindlin-3. Furthermore we used SILAC-PrEST quantification to reveal the expression levels and stoichiometry of integrin 3, kindlin-3 and talin-1.

The study was performed in mice, therefore PrESTs were produced independently of the Human Protein Atlas project. We designed the PrEST sequences to be optimal for mass spectrometric analysis, specifically unique regions having many tryptic peptides, as well as considering uniqueness in case of isoforms. To be as accurate as possible two independent PrESTs were designed and produced to target each of the proteins of interest.

The manuscript for this study is currently in preparation:

'Platelets and neutrophils require different Kindlin-3 copy numbers to control integrin-mediated functions in vivo'

Sarah Klapproth*, Federico Moretti*, **Marlis Zeiler**, Raphael Ruppert, Matthias Mann, Markus Sperandio, Reinhard Fässler and Markus Moser

Platelets and neutrophils require different Kindlin-3 copy numbers to control integrin-mediated functions *in vivo*

Sarah Klapproth[1,2], Federico Moretti[1], Marlis Zeiler[3], Raphael Ruppert[1], Matthias Mann[3], Markus Sperandio[2], Reinhard Fässler[1], Markus Moser[1#]

[1]Max-Planck-Institute of Biochemistry, Department of Molecular Medicine, Am Klopferspitz 18, D-82152 Martinsried, Germany

[2]Walter Brendel Center for Experimental Medicine, Ludwig-Maximilians-University Munich, D-81377 Munich, Germany

[3]Max-Planck-Institute of Biochemistry, Department of Proteomics and Signaltransduction, Am Klopferspitz 18, D-82152 Martinsried, Germany

Corresponding author: Markus Moser (moser@biochem.mpg.de)

Keywords: Kindlin-3, talin-1, integrin, integrin activation, LAD-III

Many of the diverse functions of blood cells depend on integrins and the cytoplasmic proteins kindlin-3 and talin-1. To relate specific functions of platelets and polymorphonuclear neutrophils (PMNs) to specific kindlin-3 levels we generated mouse strains expressing 50%, 10% and 5% of kindlin-3. We report that in contrast to Kindlin-3 null mice, which die of severe bleeding and leucocyte adhesion deficiency, mice expressing as little as 5-10% of kindlin-3 were viable, although only a limited number of integrins could be activated on platelets and neutrophils. Interestingly, tail wounding led to a severely prolonged bleeding tendency due to reduced platelet adhesion and aggregation but leukocyte adhesion to endothelial cells in the inflamed ear and the cremaster muscle was rather mildly impaired. Absolute quantification of protein copy numbers revealed similar kindlin-3 and talin-1 copy numbers in platelets and neutrophils. Interestingly, the ratio of the copy number of kindlin-3 and talin-1 relative to β3 integrin in platelets was 2:1, whereas in neutrophils the ratio relative to β2 integrins was 1:2. These data indicate that platelet-mediated hemostasis requires high kindlin-3 levels to ensure full-fledged integrin activation, whereas leukocyte adhesion and extravasation require lower levels of active integrins.

Integrins are cell adhesion receptors that consist of α and β subunits. They anchor cells to the extracellular matrix and assemble large signaling hubs, with which they regulate essential cellular processes including cell adhesion, migration, proliferation, survival and differentiation. A hallmark of integrins is their ability to reversibly switch between an active and an inactive conformation. In their inactive conformation they have low affinity for ligands, while upon activation (inside-out signaling) the affinity increases, which leads to ligand binding, integrin clustering and finally signaling (outside-in signaling). Two principal protein families are known to regulate integrin affinity, the talins (talin-1 and -2) and kindlins (kindlin-1, -2, -3). They both bind directly β integrin tails and thereby induce the allosteric change in the integrin ectodomain (Moser et al., 2009b).

A fast allosteric change and rapid integrin-ligand interaction is particularly important for blood cells such as platelets and leukocytes, whose surface integrins continuously

encounter ligands in plasma or on the vascular endothelium, respectively. Therefore, it is of fundamental importance for these cells to maintain integrins in an inactive state and rapidly activate them on demand, for example during bleeding when platelet aggregation is needed to seal vascular injuries, or during infection when leukocytes need to extravasate and kill microbial invaders. Blood cells express talin-1 and kindlin-3 whose binding to β integrin tails is tightly controlled by mechanism(s) that are largely unknown. Furthermore, it is also unclear whether a specific stoichiometry is critical for the integrin activation process.

In the present study, we engineered several mouse strains expressing different kindlin-3 protein levels. This allowed testing how different talin-1/kindlin-3 ratios influence integrin-mediated functions. We found that 5% of kindlin-3 in blood cells is sufficient for embryonic and postnatal development. However, upon exposure to stress the low levels of kindlin-3 impair adhesive functions of platelets and polymorphonuclear neutrophil (PMNs). Furthermore, platelets possess more kindlin-3 and talin-1 relative to their integrin levels than PMNs suggesting that rapid and full-fledged integrin activation is required for platelet functions, while only a limited number of integrins are activated during leukocyte extravasation.

RESULTS AND DISCUSSION

Generation of conditional and hypomorphic kindlin-3 mouse mutants. Kindlin-3 deficient mice die shortly after birth of anemia caused by bleeding and erythrocyte defects (Kruger et al., 2008; Moser et al., 2009b). To enable the disruption of the kindlin-3 gene at later stages or in specific blood cells, we generated a conditional kindlin-3 mouse by flanking exon 3-6 with *loxP* sites (suppl. Fig. 1). An intercross of these kindlin-3 floxed (K3fl/fl) mice with a deleter Cre strain produced offspring lacking exon 3-6 and the *frt*-flanked neomycin cassette, and as expected, with the same defects observed in constitutive kindlin-3 null mice (Moser et al., 2009a; Moser et al., 2009b) (suppl. Fig. 1D,E).

We noticed that the presence of the neomycin cassette in intron 6 of the floxed kindlin-3 gene (suppl. Fig. 1A; K3+/n mice) significantly reduced kindlin-3 protein levels (Figure 1A,B). Intercrossing these mice to homozygosity (K3n/n) or with mice carrying a kindlin-3 null allele (K3n/-) further reduced kindlin-3 protein levels in spleen and thymus to around 10% and 5%, respectively (Figure 1A,B). Sequence analysis of kindlin-3 mRNA revealed aberrant splicing into the neomycin cassette as the major cause for reduced protein levels (data not shown). Thus, the introduction of a neomycin cassette into intron 6 of the kindlin-3 gene produced a hypomorphic kindlin-3 allele and by intercrossing these mice with wild type or kindlin-3 null mice we could differently reduce kindlin-3 protein levels.

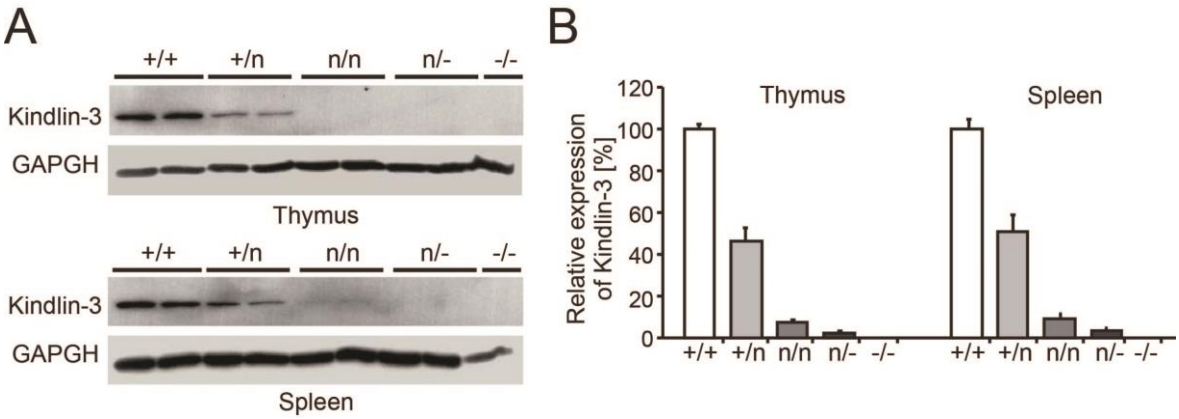


Figure 1: Kindlin-3 protein levels in spleen and thymus. (A) Western blot analyses of kindlin-3 in spleen and thymus lysates from K3+/+, K3n/n, K3n/- mice. GAPDH served as loading control. (B) Densitometric quantification of (A).

Surprisingly, the different hypomorphic mouse strains including the K3n/- mice developed and aged normally, were fertile and showed normal cellularity in spleen and thymus (suppl. Fig. 2A-C). Furthermore, their erythroid counts in blood were normal indicating that they did not suffer from microbleedings or erythrocyte defects. However, the total number of white blood cells (WBC), in particular PMNs and lymphocytes, was significantly increased in K3n/n and K3n/- mice suggesting that low kindlin-3 levels compromise leukocyte extravasation and tissue patrolling (supplementary table 1).

Kindlin-3 hypomorphic mice show a bleeding tendency. Kindlin-3 deficiency causes fatal bleedings in mice and man (Kuijpers et al., 2009; Malinin et al., 2009; Moser et al.,

2008; Svensson et al., 2009). Although platelets from K3n/n and K3n/- mice expressed around 10% and 5% of kindlin-3 levels, respectively, compared to wild type controls (Fig. 2A), they did not suffer from spontaneous hemorrhages during development or postnatal life (Fig. 2B, and data not shown). However, bleeding assays revealed that wild type and K3+/n mice stopped bleeding within 5 min, while K3n/n mice suffered from a similar bleeding tendency as kindlin-3 null mice (Fig. 2C). Interestingly, thrombin treatment of wild type and K3+/n platelets robustly induced binding of fibrinogen by platelet αIIbβ3, while fibrinogen binding was reduced to less than 20% in K3n/n and K3n/- platelets and almost entirely lost in K3-/- platelets (Fig. 2D). Consistently, thrombin-induced activation of αIIbβ3 integrins measured by flow cytometry using the JON/A-PE antibody decreased concomitantly with lowering levels of kindlin-3 in platelets (Fig. 2E). Adhesion and spreading of platelets from K3n/n and K3n/- mice on a fibrinogen-coated surface was severely impaired (Fig. 2F). Since treatment of platelets with manganese, which can bypass integrin activation (Mould et al., 2002) rescued fibrinogen binding (not shown) and platelet adhesion but not platelet spreading (Fig. 2F), we conclude that the diminished kindlin-3 levels in K3n/n and K3n/- platelets impaired both integrin activation and outside-in signaling leading to actin reorganization. These findings were corroborated with *in vitro* aggregation assays, in which platelets from K3n/n and K3n/- mice produced much smaller aggregates (Fig. 2G) and displayed a biphasic aggregation response (Fig. 2H); the platelet shape change was followed by a rapid initial aggregation, while the second phase of aggregation, which depends on secreted auto- and paracrine factors, was significantly reduced in K3n/n and K3n/- platelets. Addition of fibrinogen normalized the second aggregation phase suggesting that indeed integrin-triggered platelet degranulation and release of secondary agonists is impaired and/or delayed in kindlin-3 hypomorphic platelets (Fig. 2G). Furthermore, thrombin-induced platelet clot retraction was delayed with platelets from K3n/n and K3n/- mice in platelet-rich plasma; however, addition of manganese normalized the clot retraction times indicating that acto-myosin induced pulling forces are induced by active integrins independent of kindlin-3

(Fig. 2I), while actin reorganization and platelet spreading require active integrins and depend on kindlin-3 (Fig. 2F).

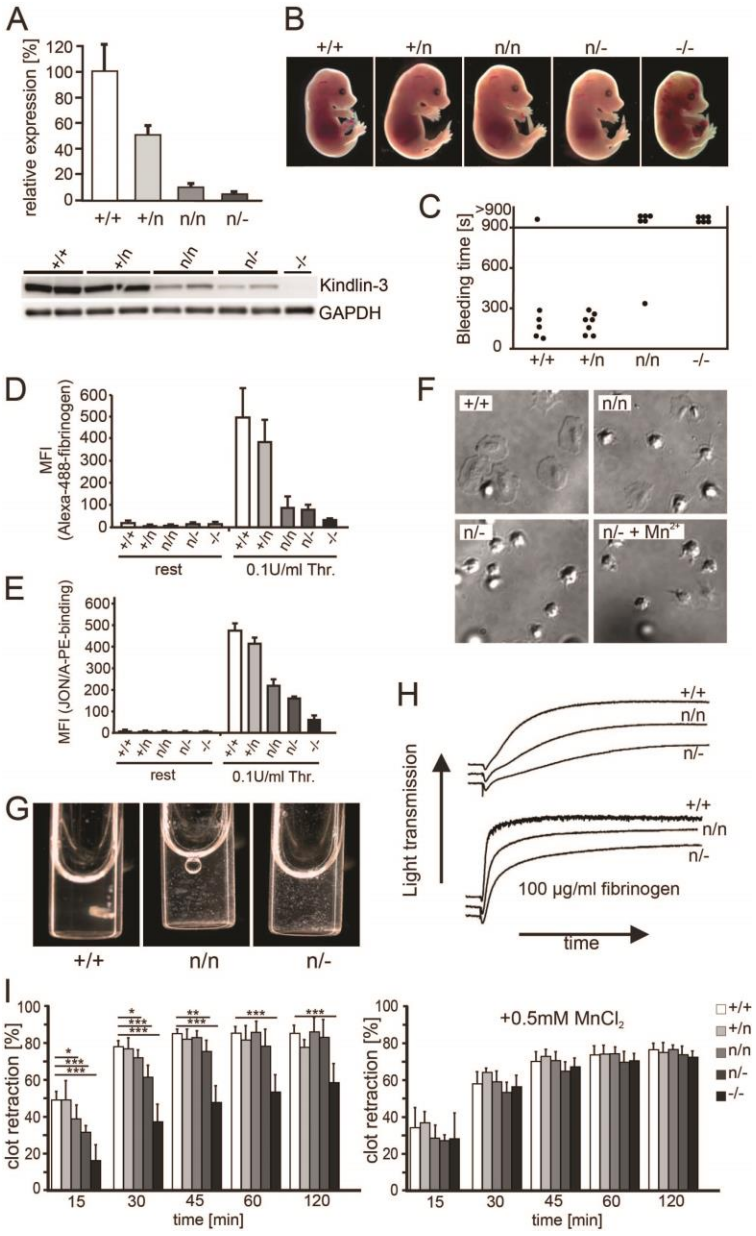


Figure 2: Function of platelet integrins is kindlin-3 dose dependent. (A) Densitometric quantification of kindlin-3 levels in K3+/+, K3n/n, K3n/- and K3-/- platelets (N=5;5;4;4). GAPDH served as loading control. (B) Kindlin-3 K3+/+, K3n/n, K3n/- E14.5 embryos are free bleedings. (C) Tail bleeding times of K3+/+, K3n/n, K3n/- and K3-/- chimeric mice. (D) Binding of fibrinogen to platelets after treatment with 0.1 U/ml thrombin. Resting platelets were used as control. N=5;5;5;5;3. (E) Activation of αIIbβ3 integrin on platelets upon stimulation with 0.1 U/ml thrombin. N=5;5;5;5;3. (F) Washed platelets were stimulated with 0.01 U/ml thrombin and allowed to adhere to coated fibrinogen for 10 min in the presence or absence of 3 mM Mn2+. (G) Aggregates of K3+/+, K3n/n and K3n/- platelets in response to 0.1 U/ml thrombin. (H) Platelet aggregation assays with K3+/+, K3n/n, K3n/- platelets in response to thrombin in the absence and presence

of 100 μg/ml fibrinogen. (I) Clot retraction of platelets from K3+/+, K3n/n, K3n/- and K3-/- mice in the absence or presence of 0.5 mM Mn2+. N=7;4;6;6;7. Values are given as mean ± SD. Significant differences are indicated by asterisk (*: p<0.05; **: p<0.01; ***p<0.001).

Altogether, these data indicate that lowering kindlin-3 in platelets to 5-10% of wild type levels allows them to activate a limited number of integrins, which is sufficient to fulfill basal functions but insufficient to scope with stress conditions.

K3n/n and K3n/- mice suffer from leukocyte adhesion defects. Absence of Kindlin-3 also impairs β2 integrin functions on leukocytes resulting in leukocyte adhesion deficiency in mouse and man (Moser et al., 2009a). PMNs from the bone marrow of wild type, K3n/+, K3n/n and K3n/- mice displayed reduced kindlin-3 levels ranging from 50% in K3n/+, to 10% in K3n/n and 5% in K3n/- mice (Fig. 3A). With these different mouse strains we investigated in several experiments to which extent leukocyte adhesion, extravasation and phagocytosis become compromised when kindlin-3 drops below a certain threshold. First, stimulation of K3n/n and K3n/- PMNs with phorbol-12-myristate-13-acetate (PMA) showed diminished adhesion to ICAM-1, although adhesion levels were still significantly higher than for K3-/- PMNs (Fig. 3B). Next, we treated earflaps with phorbol ester (crotoin oil) for 1 h and observed a delay in the extravasation of K3n/n and K3n/- PMNs compared to wild type and K3+/n PMNs. Notably, 4 h after treatment K3n/n and K3n/- PMNs have crossed the vascular wall and distributed in the interstitium like wild type PMNs, while K3-/- PMNs failed to extravasate (Fig. 3C). In order to study leukocyte adhesion and extravasation more quantitatively we analysed TNFα-stimulated cremaster muscle venules *in vivo* using intravital microscopy. The experiments revealed that leukocyte adhesion was normal in wild type and K3n/+ mice, decreased by about 20% in K3n/n mice and by about 50% in K3n/- mice (Fig. 3D). Concomitantly with the reduced *in vivo* adhesion of leukocytes, the numbers of rolling leukocytes were increased in K3n/- mice (Fig 3E). Similarly, β2 integrin-mediated phagocytosis of serum-opsonized bacteria was reduced in K3n/n and K3n/- PMNs and abolished in K3-/- PMNs (Fig. 3F). Altogether, these findings indicate that the low kindlin-3 levels in K3n/n and K3n/- PMNs impairs

phagocytosis and adhesion. However, the compromised adhesion still allows a rather efficient leukocyte extravasation in inflammatory models.

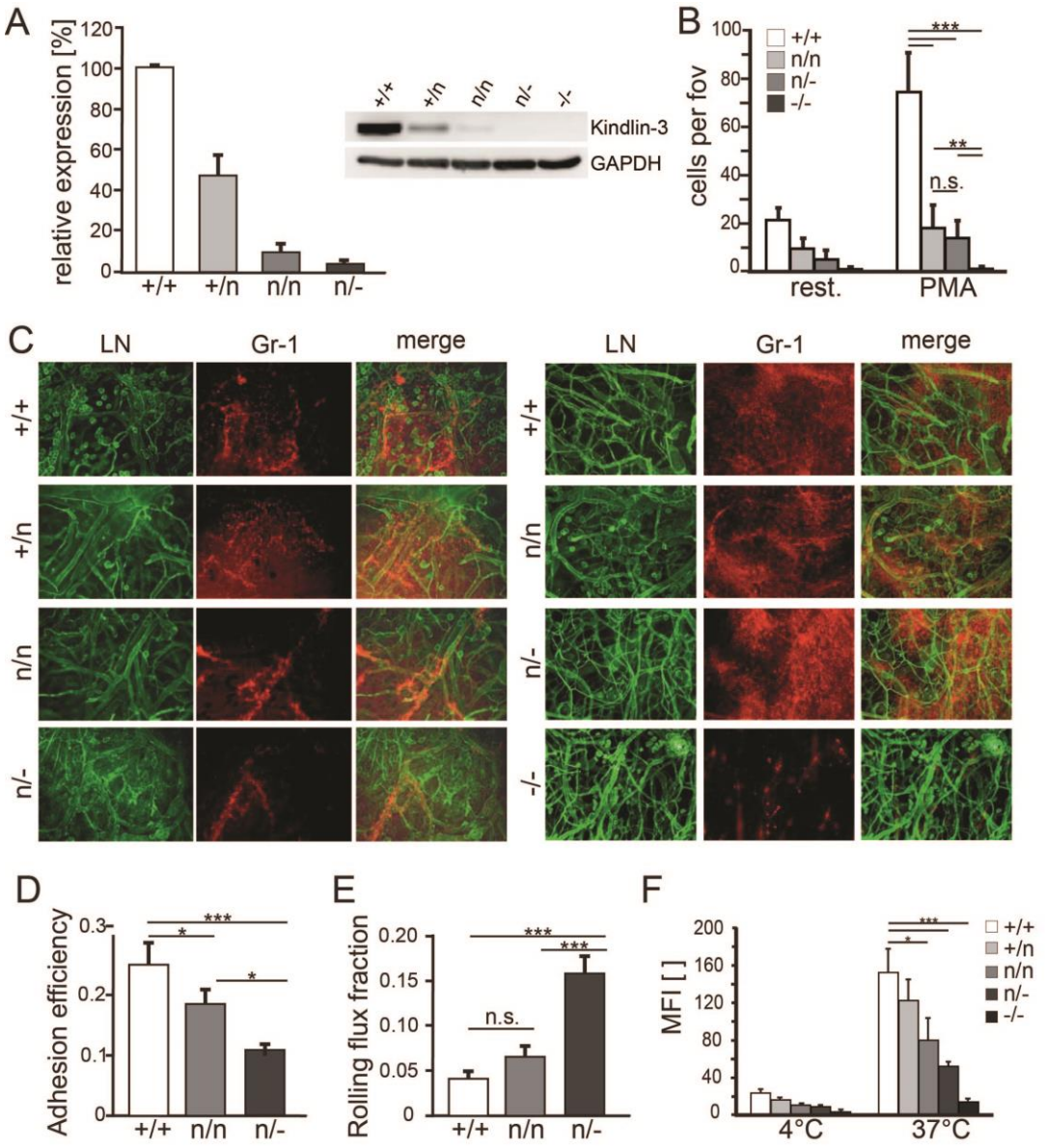


Figure 3: Leukocyte adhesion is Kindlin-3 modulated. (A) Densitometric quantification of kindlin-3 levels in K3+/+, K3n/n, K3n/- and K3-/- PMNs (N=3). GAPDH served as loading control. (B) Static adhesion of PMNs to ICAM-1 upon PMA stimulation. N=4. (C) Whole mounts of ears treated with phorbol ester for 1 (left) and 4 h (right) stained with a pan-laminin antibody (LN; green) to visualize endothelial basement membranes and a Gr-1 antibody (red) to visualize PMNs. Scale bar: 200 μm. (D-E) Leukocyte adhesion and rolling in TNFα-stimulated cremaster muscle venules assessed in 12 venules of 4 K3+/+ mice, in 19 venules of 4 K3n/n mice and in 17 venules of 3 K3n/- mice. (D) Leukocyte adhesion efficiency determined as number of adherent leukocytes per mm2 vascular surface area divided by the systemic leukocyte count. (E) Leukocyte rolling flux fraction determined as rolling leukocytes passing an imaginary perpendicular line over the vessel corrected by the total number of passing leukocytes. Values are mean ± SEM. (F)

Phagocytosis of fluorescently labeled E.coli particles by K3+/+, K3n/n, K3n/- and K3-/- PMNs. N=3;5;4;4;3. Significant differences are indicated by asterisk (*: p<0.05; **: p<0.01; ***p<0.001).

Kindlin-3, talin-1 and integrin copy numbers in platelets and PMNs. The basal functions of platelets and PMNs from K3n/n and K3n/- mice are efficiently executed *in vivo*. However, upon stress exposure the functions of platelets and PMNs become compromised. Interestingly, stress situations affect K3n/n and K3n/- derived platelets more severely than leukocytes. To test whether this difference could be due to different copy number ratios of kindlin-3 and talin-1 proteins in platelets versus leukocytes we performed Western blot analysis with purified blood cell populations. In comparison with various white blood cells platelets contain high levels of kindlin-3 and talin-1 (Fig. 4A). Analysis of the platelet proteome by high resolution mass spectrometry (MS) (Cox and Mann, 2011) confirmed that kindlin-3, talin-1 and β3 integrins are among the most abundant platelet proteins (Fig. 4B).

To determine the stoichiometry of kindlin-3, talin-1, β2 and β3 integrins in platelets and PMNs, we used a recently developed MS method to quantify absolute protein numbers by expressing parts of the target protein sequence in a heavy labeled form (Zeiler et al.). Specifically, short protein fragments of approximately 150 amino acids from murine kindlin-3, talin-1, β3 and β2 integrins were fused to the albumin-binding protein (ABP) and expressed in the presence of $^{13}C_6{}^{15}N_2$-lysine (Lys^8) and $^{13}C_6{}^{15}N_4$-arginine (Arg^{10}) in an auxotrophic *Escheria coli* strain, which labeled them with heavy isotopes. A precisely quantified cocktail of different stable isotope labeled protein standards from kindlin-3, talin-1, β2 and β3 integrins was then mixed with platelet and PMN lysates and their absolute copy numbers per cell were calculated after determining the heavy to light ratio by MS (Fig 4C; for details see material and methods and Zeiler et al., 2012). Kindlin-3 expression in platelet lysates from wild type, K3+/n, K3n/n and K3n/- mice revealed a stepwise decrease, while talin-1 and β3 integrin levels were similar in the different platelet populations (Fig. 4D). The measurements also revealed that a wild type platelet contained between 260,000 and 290,000 copies of talin-1 and kindlin-3, respectively, and about 130,000 copies of β3 integrin molecules (Fig. 4D). The PMNs also contained a

stoichiometric number of kindlin-3 and talin-1 molecules of about 420,000 molecules per cell each. Furthermore, the total numbers of β2 integrins were approximately 800,000 molecules per cell, and hence was twice the number of talin-1 and kindlin-3 (Fig. 4E). These data indicate that platelets have an excess of talin-1 and kindlin-3 over integrins, whereas PMNs have an excess of integrins over talin-1 and kindlin-3.

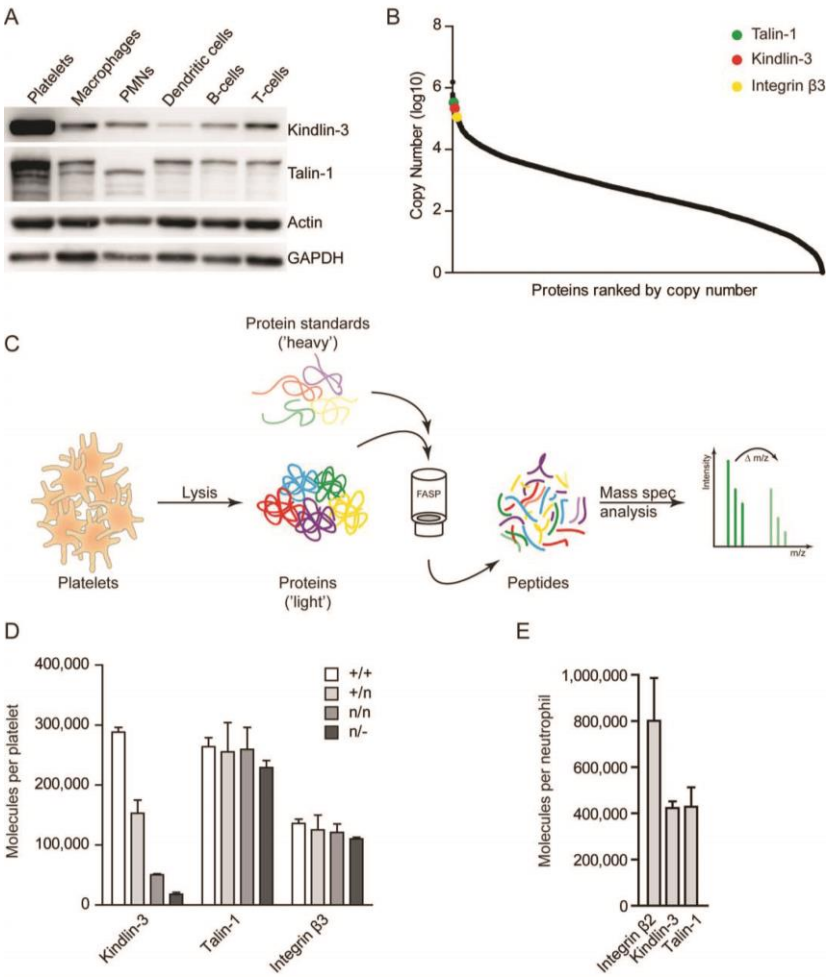


Figure 4: Quantification of kindlin-3, talin-1 and integrin protein levels in platelets and PMNs. (A) Western blot of kindlin-3 and talin-1 from different blood cell lysates. GAPDH and actin served as loading controls. (B) Copy number counts of platelet proteins were estimated with intensity based absolute quantification. (C) Schematic representation of the workflow for quantification of protein copy numbers in platelet lysates. Heavy labeled protein standards with known concentration were spiked into platelet lysates and peptide ratios were determined by mass spectrometry. (D) Mass spectrometry-based absolute quantification of the copy numbers of kindlin-3, talin-1, β3 integrin and β2 integrin in K3+/+, K3n/n and K3n/- platelets. (wt N=3; +/n, n/n and n/- N=2) . Values are given as median ± range. (E) Mass spectrometry-based absolute quantification of the copy numbers of kindlin-3, talin-1 and β2 integrin in wild type PMNs. N=3. Values are given as median ± range.

Integrin-dependent functions of blood cells are strictly controlled by kindlin-3. In this study kindlin-3 hypomorphic mice were generated to address which levels of kindlin-3 are required for basal and stress induced platelet and PMNs functions. The results allow drawing interesting conclusions; first, expression of 5 to 10% of kindlin-3 in blood cells is sufficient for survival. Second, the steady state functions of platelets and PMNs with 5-10% of normal kindlin-3 levels are maintained, although the elevated WBC counts indicate reduced tissue patrolling. Third, low kindlin-3 expression allows activation of only a reduced number of integrins leading to decreased ligand binding and adhesion. This shows that kindlin-3 does not act in a catalytic manner. Fourth, under stress conditions such as bleeding or inflammation the low levels of kindlin-3 limit platelet functions more than PMN functions indicating that platelets require a higher threshold of kindlin-3 to cope with their demands. Fifth, although kindlin-3 is required for integrin-dependent intracellular signalling controlling platelet spreading and degranulation, it is not required for acto-myosin pulling and platelet contraction, which is in contrast to talin-1 (Haling et al., 2011).

Why are decreased kindlin-3 levels affecting platelets more than PMNs? The integrin-mediated adhesion machinery is highly expressed in platelets indicating that it is of paramount importance for the organism. The binding of plasma fibrinogen might require more functional integrins than the binding of PMNs to insoluble ICAM-1, which is upregulated and clustered on inflamed endothelial cells allowing a fast rebinding of broken integrin-ligand bonds. We determined the absolute numbers of kindlin-3, talin-1, β2 and β3 integrins in platelets and PMNs and found stoichiometric expression of both integrin activators, which further corroborates their cooperativeness during integrin regulation. In addition, surface density of β3 integrins on platelets (approximately 20,000 copies per μm^2 platelet surface area based on a total surface area of 8 μm^2) and the corresponding concentration of talin-1 and kindlin-3 in platelets (40,000 copies each per fL platelet volume; total platelet volume appr. 7fL (von Hundelshausen et al., 2007)) are much higher than the corresponding values in PMNs (3,000 copies of β2 integrins per μm^2

neutrophil surface area and 1,500 copies of talin-1 and kindlin-3 per fL PMN volume; based on 250 μm^2 surface area and 300fL volume for PMNs (Ting-Beall et al., 1993)). These values may very well reflect differences in mechanical stress, platelets and neutrophils are exerted to (platelet adhesion in arteries versus neutrophil adhesion in venules). Furthermore, the twofold excess of talin-1 and kindlin-3 over β3 integrins in platelets ensuring rapid and full-fledged integrin activation critical for primary hemostasis. On the contrary, PMNs express only half as much integrin activators than β2 integrins suggesting that only a limited number of β2 integrins is activated. Remarkably, even neutrophils from kindlin-3 hypomporphic mice can efficiently adhere and leave the inflamed vasculature. Probably because a highly dynamic and strongly restricted number of active β2 integrins on leukocytes is sufficient to control the dynamic interaction with the inflamed endothelium and subendothelial tissue. In this regard, studies on leukocyte adhesion deficiency type I patients also showed that the severity of clinical infectious complications is directly related to the degree of β2 integrin deficiency. Severely immune deficient patients have essentially undetectable β2 integrin expression on leukocytes, whereas moderate phenotypes are found in patients expressing 2.5 to 6% of β2 integrins on PMNs (Anderson et al., 1985; Anderson and Springer, 1987). The clinical data mentioned above, together with our results, indicate that integrin activation and function is regulated in a large part by the precise control of the stoichiometry of integrin activators relative to integrin levels, thus enabling different blood cells to achieve their diverse biological functions.

MATERIAL AND METHODS

Animals. Kindlin-3$^{-/-}$ mice have been described previously (Moser et al., 2008). To generate conditional kindlin-3 mouse mutants (kindlin-3 fl/fl) a targeting vector was cloned, in which exons three to six of the kindlin-3 gene were flanked by loxP sites. This vector was electroporated into 129 ES cells. A Neo-cassette flanked by frt sites allowed G418 selection. Several homologous recombinant clones were injected into C57BL/6 blastocysts and resulted in germ line chimeras. The presence of the Neo-cassette within

the kindlin-3 genomic locus resulted in strongly reduced kindlin-3 expression giving rise to a hypomorphic kindlin-3 allele. Mating with mice carrying a wild type, hypomorphic or kindlin-3 null (K3-/-) allele resulted in offspring with different levels of kindlin-3. In parallel, the Neo-cassette was removed from the kindlin-3 genomic locus by mating kindlin-3 hypomorphic mice with a strain expressing a deleter flipase resulting in conditional kindlin-3 mice. Finally, conditional kindlin-3 mice were crossed with a deleter Cre mouse strain (Betz et al., 1996) to obtain K3-/- mice.

All mouse experiments were performed with the approval of the District Government of Bavaria.

Generation of fetal liver cell chimeras. Fetal liver cells were obtained from embryonic day 15 K3+/+ or K3-/- embryos by pushing liver tissue through a cell strainer (Falcon). 4 x 10^6 cells were then injected into the tail vein of lethally irradiated (twice 6.5 Gy) recipient C57BL/6 mice. Mice were used not before 6 weeks after transfer.

Antibodies. The following antibodies were used for immunostaining of cells: mouse anti-vinculin antibody, (from Sigma-Aldrich); Cy3-labeled secondary antibody (Jackson ImmunoResearch Laboratories, Inc.). Phalloidin-488 dye was obtained from Invitrogen.

The following antibodies were used for flow cytometry: hamster IgG anti-integrin β1, isotype control hamster IgG, and isotype control mouse IgG1 (all from Biolegend); rat IgG2a anti-integrin β2, rat IgG2a anti-integrin β7, rat IgG2a anti-integrin α4, rat IgG2a anti-integrin α5, rat IgG2a anti-integrin αL, rat IgG1 anti-integrin αv, isotype control rat IgG1, and isotype control rat IgG2a (all from BD); hamster IgG anti-integrin β3, mouse IgG1 anti-integrin β5, rat IgG2b anti-integrin αM and isotype control rat IgG2b (all from eBioscience).

The following antibodies were used for Western blotting: mouse anti-glyceraldehyde 3-phosphate dehydrogenase (anti-GAPDH; Merck); rabbit anti-actin and mouse anti-talin (both from Sigma-Aldrich); rabbit anti-P42/44 MAPK, rabbit anti-pP42/44 MAPK

Thr202/Tyr204, rabbit anti-Akt, and rabbit anti-pAkt Ser473 (all from Cell Signaling Technology); rabbit anti-kindlin-3 antibody is homemade (Ussar et al., 2006).

Peripheral blood cell counts. Mice were bled from the retro-orbital sinus into EDTA-coated tubes. Blood cell counts were determined using a Hemavet 950 analyzer (Drew Scientific Inc., Oxford, CT).

Cell preparation. CD4-positive T-cells were isolated from spleen using the CD4+ T cell Isolation Kit II from Miltenyi Biotech (Bergisch Gladbach, Germany). PMNs (polymorphonuclear neutrophils) were FACS sorted from whole bone marrow for high expression of Gr1 using a FITC-labeled antibody (eBioscience). B220-positive B-cells were FACS-sorted from whole spleen. Macrophages and dendritic cells (DCs) were derived from whole bone marrow by culturing in medium containing M-CSF and GM-CSF.

Platelet analyses. Integrin activation studies were performed with blood samples washed twice with modified Tyrode´s HEPES buffer (134 mM NaCl, 0.34 mM Na_2HPO_4, 2.9 mM KCl, 12 mM $NaHCO_3$, 20 mM HEPES, pH 7.0) containing 5 mM glucose, 0.35% bovine serum albumin (BSA) and 1 mM $CaCl_2$, which then were activated with 0.1 U/ml thrombin (Sigma Aldrich) and either stained with fluorophore-labeled anti-activated αIIbβ3 (JON/A, Emfret Analytics) or fibrinogen (Molecular Probes).

Platelet adhesion and spreading was analysed on fibrinogen-coated (1mg/ml; Sigma Aldrich) glass bottom dishes blocked with 3% BSA. Washed platelets were re-suspended at a concentration of $0.5x\ 10^6$ platelets/ml and then further diluted 1:10 in Tyrode's-HEPES buffer. Shortly before plating, platelets were activated with 0.01 U/ml thrombin. Platelets were allowed to spread for 10 min and imaged by differential interference contrast (DIC) microscopy. Additionally, washed platelets were allowed to adhere to fibrinogen in the presence of 3 mM Mn^{2+} without thrombin stimulation.

Platelet aggregation was measured with 2×10^8/ml washed platelets stimulated with 0.1 U/ml Thrombin in the absence or presence of 100 ug/ml human fibrinogen. Light

transmission was recorded on a ChronoLog aggregometer over 15 min and was expressed as arbitrary units with the transmission through buffer defined as 100% transmission.

Platelet clot retraction was measured in platelet-rich plasma (PRP) obtained from blood drawn into 1/10 volume of 3.8% sodium citrate. Platelets were diluted to 1x 10^8/ml with platelet-poor plasma obtained from wild type mice. PRP was added to siliconized glass cuvettes containing a paperclip, and clot retraction was initiated after addition of 2 mM $CaCl_2$ and 1 U/ml thrombin (Sigma Aldrich). In parallel, samples were also treated with 0.5 mM $MnCl_2$. Samples were incubated at 25°C for 2 hrs and clot retraction was analysed by weighing the residual serum to calculate the percentage of serum extruded from the clot at the indicated time points.

Skin inflammation model. Croton oil (Sigma-Aldrich) was diluted to 1% in acetone and topically applied to the ventral and the dorsal sides (20µl each) of ears from K3+/+, K3n/n, K3n/- and K3-/- chimeric mice. After 1 and 4 h, mice were sacrificed and the earflaps were split into two halves by carefully separating the dorsal and ventral skin. For histological analysis, dorsal and ventral ear halves were subjected to whole-mount immunostaining. After fixation in paraformaldehyde (PFA), ear halves were blocked with 1% BSA (PAA Laboratories) in PBS for 1h at room temperature, probed with biotin-labeled anti-Gr-1 (RB6-8C5; BD Biosciences-Pharmingen) to identify PMNs and anti-pan-laminin (L9393; Sigma-Aldrich) to visualize vessels, diluted in 1% BSA in PBS overnight at 4°C (wile shaking), and finally washed with 1% BSA in PBS. Antibodies were detected with a repeated cycle of staining with anti-rabbit Alexa Fluor 488 (Invitrogen) and anti-rat Cy3 (Dianova) before tissue was embedded in elvanol and representative images taken with a Zeiss Axio Imager equipped with an ApoTome (Zeiss).

Intravital microscopy. rmTNF-α (R&D Systems) was injected intrascrotally at a dose of 500 ng per mouse in a volume of 0.15 ml sterile physiological NaCl solution. Two hours later, the mice were anesthesized (Frommhold et al., 2008) and placed onto a heating pad to maintain body temperature at 37 °C. After intubation and carotid artery cannulation for blood sampling and application of antibodies and chemokines, the cremaster muscle was

surgically prepared for intravital microscopy as previously described (Sperandio et al., 2006). Intravital microscopy was performed on an upright microscope (Olympus BX51) with a saline immersion objective (40x and 0.8 numerical aperture). Experiments were recorded via CCD camera (model CF8/1, Kappa) on a Panasonic S-VHS recorder and with a digital camera (LaVision Biotech), and stored on a computer with Imspector software package (LaVision Biotech). The digital recordings were used offline to generate movie clips with ImageJ software (US National Institutes of Health). During the experiment, systemic blood samples were obtained (10 µl into 90 µl Türck's solution, Merck) to assess systemic white blood cell counts. Rolling flux was defined as the number of rolling cells per min and rolling PMN flux fraction as the percentage of rolling PMNs in all PMNs passing the same vessel in 1 min (Sperandio et al., 2006). PMNs that did not move for more than 30 s were considered to be adherent. The number of adherent leukocytes was assessed as the number of adherent cells per mm^2 vessel surface area (Frommhold et al., 2008).

Adhesion assays. For PMN adhesion to ICAM-1, 96-well plates were coated with 4 µg/ml recombinant human ICAM-1 (R&D Systems) in coating buffer (150 mM NaCl, 20 mM Tris-HCl, and 2 mM $MgCl_2$, pH 9) ON at 4°C and blocked with 3% BSA in PBS for 1 h at RT. PMNs sorted from bone marrow were either left untreated, or treated either with 33 ng/ml PMA (Phorbol-12-Myristate-13-Acetate; Calbiochem) or 1 mM $MnCl_2$. The adhesion assay was performed with 50.000 cells per well for 30 min. Adherent cells were fixed with 4% PFA after washing. The number of adherent cells was determined by acquisition of 3 phase contrast pictures from each well.

Phagocytosis assay. Phagocytosis was measured using the pHrodoTM Red *E.coli* BioParticle Phagocytosis Kit for Flow cytometry (invitrogen) according to the manufacturer´s instructions. Blood was collected from the retro-orbital sinus and each blood sample was measured in triplicates. Granulocytes were identified by their position in a forward vs. side scatter plot.

Absolut quantification of proteins in platelets and neutrophils. We produced two protein standards spanning over different regions of talin-1, b3 and b2 integrins, and 3 standards for kindlin-3. The protein standards were designed with an approximate length of 150 amino acids and with as many unique tryptic cleavage sites as possible. They covered in kindlin-3 the aminoacids 114-276, 304-477 and 522-605; in talin-1 1752-1948 and 2021-2188; in β3 integrin 34-162 and 589-714; and in β2 integrin 176-325 and 334-480. Subsequently they were fused to the albumin binding protein (ABP), whose cDNA was was inserted into the expression vector pET28a(+) then fused with the gene sequences of interest using HindIII and XhoI. In a next step the proteins fused to ABP were expressed in an auxotrophic E. coli strain and labeled with 'heavy' isotopes ($^{13}C_6{}^{15}N_2$-lysine and $^{13}C_6{}^{15}N_4$-arginine). The labeled proteins were purified via His_6-Tag using spin columns from Qiagen (Hilden, Germany).

The proteins of interest were quantified as previously described (Zeiler et al.). In brief, unlabeled ABP was quantified beforehand with amino acid analysis. These measurements were used to quantify the 'heavy' protein standards via SILAC ratios. The different protein standards with known concentration were then mixed. The platelets or neutrophils were lysed and the 'protein standard mix' was spiked in at approximately endogenous level. The mixture was further processed using the filter-aided sample preparation (FASP) method (Wisniewski et al., 2009), in which proteins were captured on a 30 kDa filter and SDS was removed with an urea-containing buffer. Proteins were alkylated with iodoacetamide and typsinized. The peptides were measured by MS, and peptide ratios between endogenous and labeled peptides were extracted.

For the platelet proteome FASP peptides were further fractionated with strong anion exchange chromatography. Six fractions of pH 11, 8, 6, 5, 4 and 3 were obtained.

All samples were measured using the LTQ-Orbitrap Elite or Q Exactive proteomic pipeline (Michalski et al., 2012). Raw mass spectrometric data were analyzed using the MaxQuant software (Cox and Mann, 2008). Detailed description of the MS analysis as well as the data analysis can be found in the supplementary material.

For calculation of the absolute copy number per cell we used the peptide ratios and converted the ratios to pmol. We took the median of all peptides amounts (derived from different protein standards) to calculate the copy number. We required at least 3 peptides for the quantification of each protein.

Statistical analysis. All data are shown as mean ± SD or mean ± SEM as indicated in the figure legends. To test significance level an unpaired student t-Test was performed.

ACKNOWLEDGEMENTS

We thank Michal Grzejszczyk for excellent technical assistance, Kevin Flynn, Karim Dib, Roy Zent and Ambra Pozzi for critically reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 914 TPs A1, A5 and B1) and the Max Planck Society.

The authors declare that they have no competing financial interests.

REFERENCES

Anderson, D.C., F.C. Schmalsteig, M.J. Finegold, B.J. Hughes, R. Rothlein, L.J. Miller, S. Kohl, M.F. Tosi, R.L. Jacobs, T.C. Waldrop, and et al. 1985. The severe and moderate phenotypes of heritable Mac-1, LFA-1 deficiency: their quantitative definition and relation to leukocyte dysfunction and clinical features. *The Journal of infectious diseases* 152:668-689.

Anderson, D.C., and T.A. Springer. 1987. Leukocyte adhesion deficiency: an inherited defect in the Mac-1, LFA-1, and p150,95 glycoproteins. *Annual review of medicine* 38:175-194.

Betz, U.A., C.A. Vosshenrich, K. Rajewsky, and W. Muller. 1996. Bypass of lethality with mosaic mice generated by Cre-loxP-mediated recombination. *Current biology : CB* 6:1307-1316.

Cox, J., and M. Mann. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26:1367-1372.

Cox, J., and M. Mann. 2011. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 80:273-299.

Cox, J., N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, and M. Mann. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10:1794-1805.

Frommhold, D., A. Ludwig, M.G. Bixel, A. Zarbock, I. Babushkina, M. Weissinger, S. Cauwenberghs, L.G. Ellies, J.D. Marth, A.G. Beck-Sickinger, M. Sixt, B. Lange-Sperandio, A. Zernecke, E. Brandt, C. Weber, D. Vestweber, K. Ley, and M. Sperandio. 2008. Sialyltransferase ST3Gal-IV controls CXCR2-mediated firm leukocyte arrest during inflammation. *The Journal of experimental medicine* 205:1435-1446.

Kruger, M., M. Moser, S. Ussar, I. Thievessen, C.A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fassler, and M. Mann. 2008. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* 134:353-364.

Kuijpers, T.W., E. van de Vijver, M.A. Weterman, M. de Boer, A.T. Tool, T.K. van den Berg, M. Moser, M.E. Jakobs, K. Seeger, O. Sanal, S. Unal, M. Cetin, D. Roos, A.J. Verhoeven, and F. Baas. 2009. LAD-1/variant syndrome is caused by mutations in FERMT3. *Blood* 113:4740-4746.

Malinin, N.L., L. Zhang, J. Choi, A. Ciocea, O. Razorenova, Y.Q. Ma, E.A. Podrez, M. Tosi, D.P. Lennon, A.I. Caplan, S.B. Shurin, E.F. Plow, and T.V. Byzova. 2009. A point mutation in KINDLIN3 ablates activation of three integrin subfamilies in humans. *Nature medicine* 15:313-318.

Michalski, A., E. Damoc, J.P. Hauschild, O. Lange, A. Wieghaus, A. Makarov, N. Nagaraj, J. Cox, M. Mann, and S. Horning. 2011. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP* 10:M111 011015.

Michalski, A., E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Muller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.J. Dunyach, J. Cox, S. Horning, M. Mann, and A. Makarov. 2012. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* 11:O111 013698.

Moser, M., M. Bauer, S. Schmid, R. Ruppert, S. Schmidt, M. Sixt, H.V. Wang, M. Sperandio, and R. Fassler. 2009a. Kindlin-3 is required for beta2 integrin-mediated leukocyte adhesion to endothelial cells. *Nature medicine* 15:300-305.

Moser, M., K.R. Legate, R. Zent, and R. Fassler. 2009b. The tail of integrins, talin, and kindlins. *Science* 324:895-899.

Moser, M., B. Nieswandt, S. Ussar, M. Pozgajova, and R. Fassler. 2008. Kindlin-3 is essential for integrin activation and platelet aggregation. *Nature medicine* 14:325-330.

Mould, A.P., J.A. Askari, S. Barton, A.D. Kline, P.A. McEwan, S.E. Craig, and M.J. Humphries. 2002. Integrin activation involves a conformational change in the alpha 1 helix of the beta subunit A-domain. *The Journal of biological chemistry* 277:19800-19805.

Sperandio, M., J. Pickard, S. Unnikrishnan, S.T. Acton, and K. Ley. 2006. Analysis of leukocyte rolling in vivo and in vitro. *Methods in enzymology* 416:346-371.

Svensson, L., K. Howarth, A. McDowall, I. Patzak, R. Evans, S. Ussar, M. Moser, A. Metin, M. Fried, I. Tomlinson, and N. Hogg. 2009. Leukocyte adhesion deficiency-III is

caused by mutations in KINDLIN3 affecting integrin activation. *Nature medicine* 15:306-312.

Ussar, S., H.V. Wang, S. Linder, R. Fassler, and M. Moser. 2006. The Kindlins: subcellular localization and expression during murine development. *Experimental cell research* 312:3142-3151.

Wisniewski, J.R., A. Zougman, N. Nagaraj, and M. Mann. 2009. Universal sample preparation method for proteome analysis. *Nat Methods* 6:359-362.

Zeiler, M., W.L. Straube, E. Lundberg, M. Uhlen, and M. Mann. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol Cell Proteomics* 11:O111 009613.

SUPPLEMENTARY METHODS

Liquid chromatography and mass spectrometry analysis. The peptides were separated on a 20 cm column packed with Reprosil-Pur 1.9 μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) using an UPLC system (Thermo Fisher Scientific). A linear gradient 180 min 5–25% buffer B (80% acetonitrile, 0,5 % acetic acid) at a constant flow rate of 300nl/min was applied. The UPLC was coupled via nanoelectrospray ion source (Proxeon Biosystems) to an LTQ-Orbitrap Elite mass spectrometer (Michalski et al., 2012). The data was acquired with a data-dependent top10 method, automatically selecting the most abundant precursor ions for fragmentation. Full scans, 300-1750m/z, were measured with a target value 1e6 at a resolution of 60,000 followed by 10 fragmentation spectra using higher energy collisional dissociation also measured in the Orbitrap using 5e4 ions at a resolution of 15,000. The injection time for MS spectra was restricted to 100 ms and for MS/MS spectra to 150 ms. The platelet proteome was measured on a QExactive instrument (Michalski et al., 2011)Thermo Scientific) using the same liquid chromatography setup as described above. A linear gradient 180 min 5–35% buffer B at a constant flow rate of 300 nl/min was applied. Full scans, 300-1650m/z, were measured with a target value 3e6 at a resolution of 70,000 followed by 20 fragmentation spectra using higher energy collisional dissociation and measured in the Orbitrap using

1e6 ions at a resolution of 17,500. The injection time for MS spectra was restricted to 20 ms and for MS/MS spectra to 60 ms.

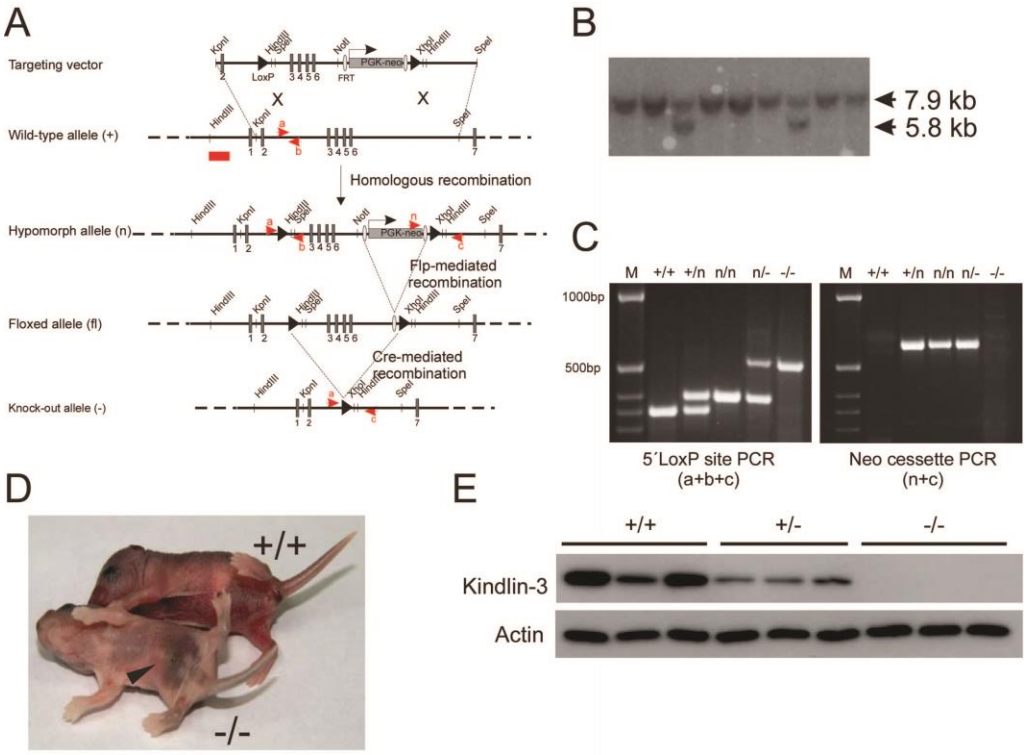
Data analysis. The raw files were analyzed using MaxQuant (version 1.3.7.4; (Cox and Mann, 2008). For peptide identification Andromeda search engine incorporated into MaxQuant (Cox et al., 2011) was used to search fragmentation spectra against the UniProtKB mouse database (downloaded on 16 June 2012), which contains 54,232 entries. 247 common contaminants were added to this database. A time-dependent mass recalibration algorithm was used for recalibration to improve the mass accuracy of precursor ions. Peptide identification was based on a search with an initial mass deviation of the precursor ion of up to 6 ppm and the allowed fragment mass deviation was set to 20 ppm. The search included cysteine carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. Enzyme specificity was set to trypsin allowing N-terminal cleavage to proline. Two mis-cleavages were allowed and a minimum of seven amino acids per identified peptide was required. The second peptide identification option in Andromeda was enabled. For statistical evaluation of the data obtained, the posterior error probability and false discovery rate (FDR) were used. The false discovery rate was determined by searching a reverse database and was set to 0.01 for peptide and protein identification. Additional peptides were identified by the "match between run" option in MaxQuant, which matches precursor masses in a 2 min retention time window (after realignment of the runs) based on the accurate mass

SUPPLEMANTARY TABLE

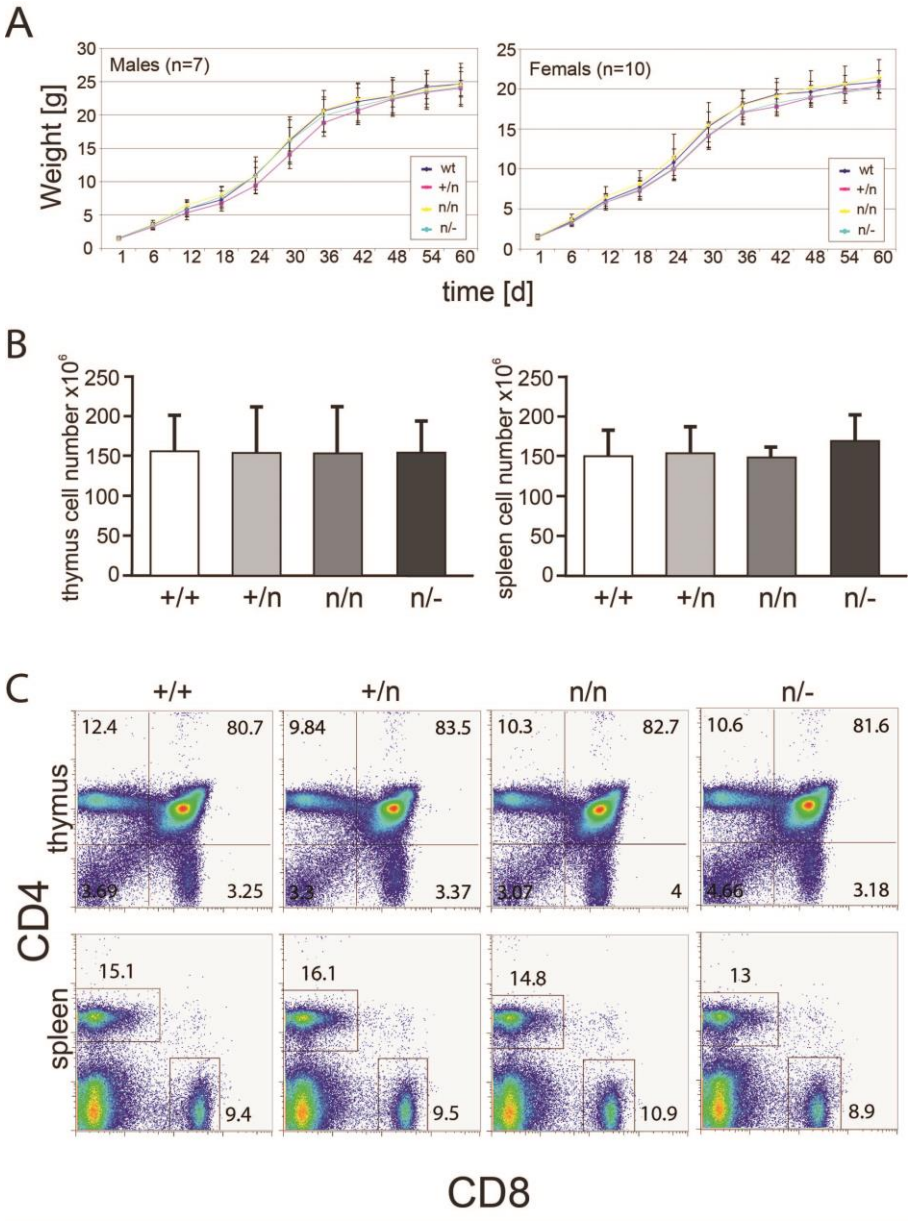
Blood counts of K3+/+, K3+/n, K3n/n, K3n/- and K3-/- chimeras. Values are shown as mean ± SD. WBC, white blood cell; NE, neutrophil; LY, lymphocyte; RBC, red blood cell; Hb, hemoglobin; PLT, platelet.

		+/+ [n=9]		+/n [n=8]		n/n [n=9]		-/n [n=9]		-/-chimera [n=4]	
		mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
[10^3/µl]	WBC	5,70	(± 2,45)	6,38	2,05	13,30	2,58	9,76	4,01	18,23	2,97
[10^3/µl]	NE	1,01	(± 0,49)	1,21	0,46	4,15	1,48	2,53	2,38	13,97	2,99
[10^3/µl]	LY	4,32	(± 2,09)	4,72	1,49	8,37	2,27	6,62	3,68	3,33	0,55
[10^6/µl]	RBC	9,44	(± 1,27)	9,19	0,82	8,35	1,21	8,35	0,95	2,44	0,47
[g/dL]	Hb	13,93	(± 2,29)	13,56	1,19	12,56	2,12	12,71	1,53	5,78	0,77
[10^6/µl]	PLT	603,44	(± 138,5)	452,0	61,8	593,2	112,5	695,3	216,4	648,0	167,1

SUPPLEMENTARY FIGURES



Supplementary Figure 1: Generation of mutant Kindlin-3 mice. (A) Targeting strategy of kindlin-3. Partial map of the kindlin-3 gene and the targeting vector carrying a loxP sites flanking exons 3 to 6 of the kindlin-3 gene. A frt flanked neomycin resistance cassette was inserted into intron 6. Exons are indicated as rectangles and restriction enzymes by name. Probe used for Southern blotting is shown as red rectangle. (B) Southern blot on DNA samples from R1 ES cells electroporated with the targeting construct. The genomic DNA was digested with HindIII and hybridized with the probe indicated in A. (C) Genomic PCR performed with primers indicated as red arrows in A to identify mice carrying the K3+/+, K3n/n, K3n/- and K3-/- alleles. (D) Deletion of the kindlin-3 gene by crossing conditional kindlin-3 mice with deletor Cre mice. (E) Western blots of thymus lysates from K3+/+ and K3-/- mice.



Supplementary Figure 1: Normal development of K3+/n, K3n/n, K3n/- mice. (A) Weight gains. (B) Cellularity of the thymus and spleen. (C) Differentiation of precursor cells into CD4- and CD8-positive T-cells.

Conclusion and future outlook

In this thesis, I have established a method for accurate absolute protein quantification termed 'SILAC-PrEST'. To date, quantitative western blotting is widely applied to detect and quantify proteins in research settings. However, only a limited number of proteins and samples can be quantified using this labor-intensive technique. Thus, absolute protein quantification methods based on MS represent an attractive alternative thanks to multiplexing capabilities and high selectivity [115]. Numerous strategies have been developed over the last years to absolutely quantify proteins using MS-based proteomics, and it could be argued that absolute quantification instead of relative quantification methods should dominate the field of quantitative proteomics. However, there are a number of limitations in the currently available absolute quantification methods that would first have to be tackled. This was the main aim of this thesis.

In the SILAC-PrEST quantification developed here, the standard is spiked in at an early stage in sample preparation, therefore both the standard and the sample are subjected to identical treatments e.g. proteolytic digestion. Indeed, we observed that the cleavage efficiency of PrEST and endogenous protein is virtually the same as peptides with missed cleavage sites have the same SILAC ratio. Furthermore, since the PrESTs typically cover 150 aa, multiple peptides can generally be used for quantification which significantly improves accuracy and robustness. Lastly, each PrEST includes a 120 aa ABP domain in addition to the protein specific sequence, making all protein classes (including membrane proteins) amenable to recombinant expression and purification and enabling a high-throughput production pipeline. Although the method developed here is clearly powerful and superior to existing ones, it can still be further advanced by for instance combining it with a targeted acquisition mode such as multiple reaction monitoring or multiplexed single ion monitoring on the latest high resolution instruments such as the Q Exactive [116]. Given this, application in a clinical setting appears feasible.

To further increase sensitivity of the approach the immuno-SILAC workflow was established. It combines antibody-based enrichment with absolute protein quantification,

exploiting the high enrichment factor of antibodies with the quantification accuracy of the SILAC-PrEST method. We utilize the polyclonal antibodies from the Human Protein Atlas project, which are raised against the PrESTs, to enrich multiple peptides. Several proteins can be quantified by multiplexing the MS-readout. Additionally, measuring time is significantly reduced due to the lowered sample complexity. Therefore, one could envision the immuno-SILAC method to be part of biomarker development pipeline and applied to clinical samples. In particular, body fluids such as plasma are highly complex and have a very large dynamic range between abundant and clinically relevant marker proteins. An additional enrichment step would enable quantification of low-abundant proteins such as cytokines. Currently, enzyme-linked immunosorbent assay (ELISA) are considered the gold standard assay to routinely quantify biomarkers in clinical settings. However in ELISA, assay development is time-consuming and various factors can influence the quality of the assay, including specificity, avidity, concentration of antibodies as well as incubation time, temperature, sample volume, dilution, and pH. At present numerous candidate biomarkers are discovered by the proteomics field and therefore verification and validation is much needed [117]. This could be achieved with an accurate, sensitive method such as immuno-SILAC.

The SILAC-PrEST technology can easily be combined with label-free quantification methods, where the PrESTs are used as reference points for scaling the intensities. This enables absolute quantification of all proteins detected in the measurement. However, it should be kept in mind that these absolute quantification values are estimates, expected to be accurate generally within a factor of two but also less accurate on a case by case basis. Nevertheless, knowing the identity of all proteins in a cell and having an estimate of their quantities provides an important basis for studying complex biological processes. In case of the platelet study described here it opens up more possibilities for investigation of platelet activation and aggregation in health and disease.

Another powerful application of the SILAC-PrEST method is the determination of protein stoichiometry in mixtures. Proteins rarely operate in isolation but work together or

assemble into a complex to fulfill their functional roles. Although MS-based methods allow identification of protein-protein interactions in high-throughput (see for example reference [118]) so far protein stoichiometries have usually relied on methods such as size-exclusion chromatography, analytical ultracentrifugation and gel electrophoresis. However, since MS-based methods for absolute protein quantification have been established, protein stoichiometries can also be deduced [119, 120]. In the case of platelets, integrin activation is achieved by mutual binding of two proteins talin-1 and kindlin-3. We set out to estimate their abundance and stoichiometry, and showed that they are expressed at the same levels, suggesting that they directly cooperate.

Next to the developments in absolute protein quantification using spike in standards label-free methods have matured independently. Although these label-free methods have existed for many years, they have become mainstream only in the last few years owing to progress in the resolution and robustness of LC-MS systems and advances in algorithms. It is a valid question to ask if label-free quantification methods will eventually supersede absolute protein quantification based on labeled spike-in standards. Indeed, from an economic and convenience perspective, label-free experiments are attractive. However, the accuracy of label-free methods currently cannot match that of quantification approaches based on spike in standards, largely because individual peptides can display unpredictable behavior in mass spectrometry, which currently cannot be accounted for in the label-free algorithm. Nevertheless, to obtain a global, quantitative view of a proteome, label-free methods are very feasible especially if combined with spike-in references that allow for correct absolute scaling, they may be able to answer most of the systems biology questions.

In conclusion, the novel SILAC-PrEST method described in my thesis allows for accurate and robust quantification and shows great promise for further applications in biological research and clinical medicine.

Abbreviations

aa	Amino acid(s)
ABP	Albumin binding protein
ADP	Adenosine diphosphate
APEX	Absolute protein expression
AQUA	Absolute quantification
CID	Collision induced dissociation
Da	Dalton
E. coli	*Escherichia coli*
ECM	Extracellular matrix
ELISA	Enzyme-linked immunosorbent assay
emPAI	Exponentially modified protein abundance index
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FDR	False discovery rate
FT	Fourier transformation
GP	Glycoprotein
GPCR	G-protein couples receptor
HCD	Higher energy collisional dissociation
HPA	Human Protein Atlas
HPLC	High-performance liquid chromatography
iBAQ	Intensity-based absolute quantification
ICAT	Isotope coded affinity tags
ICR	Ion cyclotron resonance
IT	Ion trap
iTRAQ	Isobaric tags for relative and absolute quantification
LMCO	Low mass cut off
LTQ	Linear trap quadrupole
m/*z*	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrum
NHS	N-hydroxy succinimide
NSAF	Normalized spectral abundance factor
OCS	Open canalicular system
PAI	Protein abundance index
PBMC	Peripheral blood mononucleated cell
PSAQ	Protein standard absolute quantification

PTM	Post translational modification
QconCAT	Quantification concatemer
QQQ	Triple quadrupole
RP	Reverse phase
SAF	Spectral abundance factor
SILAC	Stable isotope labeling of amino acids in cell culture
SIM	Single ion monitoring
SISCAPA	Stable isotope standards and capture by anti-peptide antibodies
TOF	Time-of-flight
TrEMBL	Translated EMBL Nucleotide Sequence Data Library
TXA_2	Thromboxane A2
UPS	Universal protein standard
VWF	Von Willebrand factor
XIC	Extracted ion chromatogram

References

1. Wilkins, M.R., et al., *From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.* Biotechnology (N Y), 1996. **14**(1): p. 61-5.
2. James, P., *Protein identification in the post-genome era: the rapid rise of proteomics.* Q Rev Biophys, 1997. **30**(4): p. 279-331.
3. Lewin, B., et al., *Lewin's GENES X*. 2011: Jones & Bartlett Learning.
4. Gygi, S.P., et al., *Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.* Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9390-5.
5. Mann, M., et al., *The coming age of complete, accurate, and ubiquitous proteomes.* Mol Cell, 2013. **49**(4): p. 583-90.
6. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
7. Kito, K. and T. Ito, *Mass spectrometry-based approaches toward absolute quantitative proteomics.* Curr Genomics, 2008. **9**(4): p. 263-74.
8. 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.
9. de Godoy, L.M., et al., *Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast.* Nature, 2008. **455**(7217): p. 1251-4.
10. Malmstrom, J., et al., *Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans.* Nature, 2009. **460**(7256): p. 762-5.
11. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.
12. Tran, J.C., et al., *Mapping intact protein isoforms in discovery mode using top-down proteomics.* Nature, 2011. **480**(7376): p. 254-8.
13. Hein, M.Y., et al., *Chapter 1 - Proteomic Analysis of Cellular Systems*, in *Handbook of Systems Biology*. 2013, Academic Press: San Diego. p. 3-25.
14. Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, *An automated multidimensional protein identification technology for shotgun proteomics.* Anal Chem, 2001. **73**(23): p. 5683-90.
15. Beck, M., et al., *The quantitative proteome of a human cell line.* Mol Syst Biol, 2011. **7**: p. 549.
16. Nagaraj, N., et al., *Deep proteome and transcriptome mapping of a human cancer cell line.* Mol Syst Biol, 2011. **7**: p. 548.
17. Wolf-Yadlin, A., et al., *Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks.* Proc Natl Acad Sci U S A, 2007. **104**(14): p. 5860-5.
18. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial.* Mol Syst Biol, 2008. **4**: p. 222.
19. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

20. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.
21. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.
22. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.
23. Michalski, A., J. Cox, and M. Mann, *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, 2011. **10**(4): p. 1785-93.
24. Hoffmann, E.d. and V. Stroobant, *Mass spectrometry : principles and applications*. 3rd ed. 2007, Chichester, West Sussex, England ; Hoboken, NJ: J. Wiley. xii, 489 p.
25. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* J Am Soc Mass Spectrom, 2002. **13**(6): p. 659-69.
26. Brunnee, C., *The Ideal Mass Analyzer - Fact or Fiction.* International Journal of Mass Spectrometry and Ion Processes, 1987. **76**(2): p. 125-237.
27. Glish, G.L. and D.J. Burinsky, *Hybrid mass spectrometers for tandem mass spectrometry.* J Am Soc Mass Spectrom, 2008. **19**(2): p. 161-72.
28. Scientific, T.F., *LTQ Orbitrap Biotech Operation*.
29. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Anal Chem, 2000. **72**(6): p. 1156-62.
30. Kingdon, K.H., *A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures.* Physical Review, 1923. **21**(4): p. 408-418.
31. Makarov, A., et al., *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer.* Anal Chem, 2006. **78**(7): p. 2113-20.
32. Scigelova, M. and A. Makarov, *Orbitrap mass analyzer--overview and applications in proteomics.* Proteomics, 2006. **6 Suppl 2**: p. 16-21.
33. Perry, R.H., R.G. Cooks, and R.J. Noll, *Orbitrap mass spectrometry: instrumentation, ion motion and applications.* Mass Spectrom Rev, 2008. **27**(6): p. 661-99.
34. Sleno, L. and D.A. Volmer, *Ion activation methods for tandem mass spectrometry.* J Mass Spectrom, 2004. **39**(10): p. 1091-112.
35. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.
36. Makarov, A., et al., *Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer.* J Am Soc Mass Spectrom, 2006. **17**(7): p. 977-82.
37. Olsen, J.V., et al., *A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed.* Mol Cell Proteomics, 2009. **8**(12): p. 2759-69.
38. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.* Mol Cell Proteomics, 2011. **10**(9): p. M111 011015.
39. 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.

40. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
41. Apweiler, R., et al., *Update on activities at the Universal Protein Resource (UniProt) in 2013.* Nucleic Acids Research, 2013. **41**(D1): p. D43-D47.
42. 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.
43. Deutsch, E.W., H. Lam, and R. Aebersold, *Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics.* Physiol Genomics, 2008. **33**(1): p. 18-25.
44. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.
45. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.
46. Cox, J. and M. Mann, *Quantitative, high-resolution proteomics for data-driven systems biology.* Annu Rev Biochem, 2011. **80**: p. 273-99.
47. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.
48. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol, 2005. **1**(5): p. 252-62.
49. Gouw, J.W., J. Krijgsveld, and A.J. Heck, *Quantitative proteomics by metabolic labeling of model organisms.* Mol Cell Proteomics, 2010. **9**(1): p. 11-24.
50. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6591-6.
51. 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.
52. Ong, S.E., I. Kratchmarova, and M. Mann, *Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC).* J Proteome Res, 2003. **2**(2): p. 173-81.
53. Bendall, S.C., et al., *Prevention of amino acid conversion in SILAC experiments with embryonic stem cells.* Mol Cell Proteomics, 2008. **7**(9): p. 1587-97.
54. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.
55. Blagoev, B., et al., *Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.* Nat Biotechnol, 2004. **22**(9): p. 1139-45.
56. Soufi, B., et al., *Stable isotope labeling by amino acids in cell culture (SILAC) applied to quantitative proteomics of Bacillus subtilis.* J Proteome Res, 2010. **9**(7): p. 3638-46.
57. Hanke, S., et al., *Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level.* J Proteome Res, 2008. **7**(3): p. 1118-30.
58. de Godoy, L.M., et al., *Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system.* Genome Biol, 2006. **7**(6): p. R50.
59. Kruger, M., et al., *SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function.* Cell, 2008. **134**(2): p. 353-64.

60. Larance, M., et al., *Stable-isotope labeling with amino acids in nematodes.* Nat Methods, 2011. **8**(10): p. 849-51.
61. Sury, M.D., J.X. Chen, and M. Selbach, *The SILAC fly allows for accurate protein quantification in vivo.* Mol Cell Proteomics, 2010. **9**(10): p. 2173-83.
62. Gruhler, A., et al., *Stable isotope labeling of Arabidopsis thaliana cells and quantitative proteomics by mass spectrometry.* Mol Cell Proteomics, 2005. **4**(11): p. 1697-709.
63. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue.* Nat Methods, 2010. **7**(5): p. 383-5.
64. Geiger, T., et al., *Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics.* Nat Protoc, 2011. **6**(2): p. 147-57.
65. Deeb, S.J., et al., *Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles.* Mol Cell Proteomics, 2012. **11**(5): p. 77-89.
66. Schwanhausser, B., et al., *Global analysis of cellular protein translation by pulsed SILAC.* Proteomics, 2009. **9**(1): p. 205-9.
67. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
68. Regnier, F.E. and S. Julka, *Primary amine coding as a path to comparative proteomics.* Proteomics, 2006. **6**(14): p. 3968-79.
69. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
70. 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.
71. Dephoure, N. and S.P. Gygi, *Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast.* Sci Signal, 2012. **5**(217): p. rs2.
72. Bantscheff, M., et al., *Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer.* Mol Cell Proteomics, 2008. **7**(9): p. 1702-13.
73. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.
74. Wenger, C.D., et al., *Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging.* Nat Methods, 2011. **8**(11): p. 933-5.
75. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. **75**(24): p. 6843-52.
76. Boersema, P.J., et al., *Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates.* Proteomics, 2008. **8**(22): p. 4624-32.
77. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nat Protoc, 2009. **4**(4): p. 484-94.
78. Desiderio, D.M. and M. Kai, *Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue.* Biomed Mass Spectrom, 1983. **10**(8): p. 471-9.
79. 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.

80. Mallick, P., et al., *Computational prediction of proteotypic peptides for quantitative proteomics.* Nat Biotechnol, 2007. **25**(1): p. 125-31.
81. Kettenbach, A.N., J. Rush, and S.A. Gerber, *Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides.* Nat Protoc, 2011. **6**(2): p. 175-86.
82. Anderson, N.L., et al., *Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA).* J Proteome Res, 2004. **3**(2): p. 235-44.
83. Brun, V., et al., *Isotope dilution strategies for absolute quantitative proteomics.* J Proteomics, 2009. **72**(5): p. 740-9.
84. Pratt, J.M., et al., *Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes.* Nat Protoc, 2006. **1**(2): p. 1029-43.
85. Beynon, R.J., et al., *Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides.* Nat Methods, 2005. **2**(8): p. 587-9.
86. Kito, K., et al., *A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes.* J Proteome Res, 2007. **6**(2): p. 792-800.
87. Dupuis, A., et al., *Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks.* Proteomics, 2008. **8**(22): p. 4633-6.
88. Singh, S., et al., *FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides.* J Proteome Res, 2009. **8**(5): p. 2201-10.
89. Megger, D.A., et al., *Label-free quantification in clinical proteomics.* Biochim Biophys Acta, 2013. **1834**(8): p. 1581-90.
90. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.* Anal Chem, 2004. **76**(14): p. 4193-201.
91. Gilchrist, A., et al., *Quantitative proteomics analysis of the secretory pathway.* Cell, 2006. **127**(6): p. 1265-81.
92. Powell, D.W., et al., *Cluster analysis of mass spectrometry data reveals a novel component of SAGA.* Mol Cell Biol, 2004. **24**(16): p. 7249-59.
93. Zybailov, B., et al., *Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae.* J Proteome Res, 2006. **5**(9): p. 2339-47.
94. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.* Nat Biotechnol, 2007. **25**(1): p. 117-24.
95. Rappsilber, J., et al., *Large-scale proteomic analysis of the human spliceosome.* Genome Res, 2002. **12**(8): p. 1231-45.
96. 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.

97. Chelius, D. and P.V. Bondarenko, *Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry.* J Proteome Res, 2002. **1**(4): p. 317-23.
98. Cox, J., et al., *MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction.* Mol Cell Proteomics, 2014.
99. Silva, J.C., et al., *Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.* Mol Cell Proteomics, 2006. **5**(1): p. 144-56.
100. Wiśniewski, J.R., et al., *A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards.* Mol Cell Proteomics, 2014.
101. *The Human Protein Atlas Version 12*. 2014 24th January 2014].
102. Ponten, F., K. Jirstrom, and M. Uhlen, *The Human Protein Atlas--a tool for pathology.* J Pathol, 2008. **216**(4): p. 387-93.
103. Patel, S.R., J.H. Hartwig, and J.E. Italiano, Jr., *The biogenesis of platelets from megakaryocyte proplatelets.* J Clin Invest, 2005. **115**(12): p. 3348-54.
104. Michelson, A.D., *Platelets*. 2012: Elsevier Science & Technology Books.
105. Hees, H. and F. Sinowatz, *Histologie: Kurzlehrbuch der Zytologie und mikroskopischen Anatomie ; [mit 12 Schemata und 15 Tabellen]*. 2000: Deutscher Aerzte-Verlag.
106. Varga-Szabo, D., I. Pleines, and B. Nieswandt, *Cell adhesion mechanisms in platelets.* Arterioscler Thromb Vasc Biol, 2008. **28**(3): p. 403-12.
107. Michelson, A.D., *Antiplatelet therapies for the treatment of cardiovascular disease.* Nat Rev Drug Discov, 2010. **9**(2): p. 154-69.
108. Leslie, M., *Cell biology. Beyond clotting: the powers of platelets.* Science, 2010. **328**(5978): p. 562-4.
109. Garcia, A., et al., *Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry.* Proteomics, 2004. **4**(3): p. 656-68.
110. Rowley, J.W., et al., *Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes.* Blood, 2011. **118**(14): p. e101-11.
111. Burkhart, J.M., et al., *The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways.* Blood, 2012. **120**(15): p. e73-82.
112. Whiteaker, J.R., et al., *Evaluation of large scale quantitative proteomic assay development using peptide affinity-based mass spectrometry.* Mol Cell Proteomics, 2011. **10**(4): p. M110 005645.
113. Moser, M., et al., *Kindlin-3 is essential for integrin activation and platelet aggregation.* Nat Med, 2008. **14**(3): p. 325-30.
114. Moser, M., et al., *The tail of integrins, talin, and kindlins.* Science, 2009. **324**(5929): p. 895-9.
115. Aebersold, R., A.L. Burlingame, and R.A. Bradshaw, *Western blots versus selected reaction monitoring assays: time to turn the tables?* Mol Cell Proteomics, 2013. **12**(9): p. 2381-2.
116. Gallien, S., et al., *Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer.* Mol Cell Proteomics, 2012. **11**(12): p. 1709-23.

117. Rifai, N., M.A. Gillette, and S.A. Carr, *Protein biomarker discovery and validation: the long and uncertain path to clinical utility.* Nat Biotechnol, 2006. **24**(8): p. 971-83.
118. Hubner, N.C., et al., *Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions.* J Cell Biol, 2010. **189**(4): p. 739-54.
119. Schmidt, C., et al., *Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring.* Anal Chem, 2010. **82**(7): p. 2784-96.
120. Nanavati, D., et al., *Stoichiometry and absolute quantification of proteins with mass spectrometry using fluorescent and isotope-labeled concatenated peptide standards.* Mol Cell Proteomics, 2008. **7**(2): p. 442-7.

Acknowledgments

First of all I would like to express my gratitude to Matthias Mann for letting me join his unique department, giving me the opportunity to learn top-notch state of the art mass spectrometry and to work on challenging projects and collaborations and for being such an inspiring and visionary PhD supervisor.

I am particularly grateful for wonderful and fruitful collaborations with Mathias Uhlén and Emma Lundberg from the Human Protein Atlas as well as Markus Moser from the Fässler department.

I would like to thank Korbinian Mayr, Richard Scheltema and Igor Paron for having the patience to answer countless mass spec related questions and for the great 24-7 mass spec support. Likewise thanks goes to Jürgen Cox for taking the time to answer countless bioinformatics and statistics questions. For their support at the beginning of my PhD I would like to thank Werner Straube and Sabine Suppmann.

Thanks goes to all the office mates that passed through E102: Tami Geiger, Sara Zanivan, Marco Hein, Atul Deshmukh, Paul Boersema, Herbert Schiller, Michael Wierer, Fabian Coscia and Marta Murgia. I very much enjoyed everyone's company and fruitful scientific and nonscientific discussions and I am grateful for the friendships that developed out of it. I am thankful to Tar Viturawong for lots of bioinformatics support and taking the challenge to successfully teach us R. Altogether, I would like to thank all members of the department for generating such a great working environment.

Special thanks goes to Gabi Stöhr and Sally Deeb for mastering the PhD together with all the up and downs involved and your invaluable friendship.

Last but not least, I would like to thank my parents for their continuous support and encouragement. It is great to know that they will always be there for me no matter what.