Aus dem Biomedizinischem Centrum München
Lehrstuhl Molekularbiologie
Klinik der Ludwig-Maximilians-Universität München
Vorstand: Prof. Dr. rer. nat. Peter B. Becker

Functional determinants of the intrinsically disordered region 2 in the general regulatory factor Abf1 from *Saccharomyces cerevisiae*

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Iris Langstein-Skora
aus
Bergisch Gladbach

2024

Mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuer(in):	Prof. Dr. rer. nat. Philipp Korber
Zweitgutachter(in):	Prof. Dr. rer. nat. Jörg Renkawitz
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	23. Juli 2024

Without people, you’re nothing.

Joe Strummer

This thesis is dedicated to my parents

Doris and Dr. Gerhard Langstein

and to my husband

Tobias Skora

TABLE OF CONTENTS

ZUSAMMENFASSUNG

Durch das, lange Zeit vorherrschende, Struktur-Funktions-Paradigma war das Interesse der Proteinforschung an intrinsisch ungeordneten Proteinen (*intrinsically disordered proteins*, IDPs) bzw. Proteinregionen (*intrinsically disordered regions*, IDRs) zunächst gering. Erst seit Ende der 1990er Jahre wurde zunehmend experimentell gezeigt, dass gerade das Fehlen einer definierten, dreidimensionalen Proteinstruktur unter physiologischen Bedingungen mit einer strukturellen und funktionellen Vielseitigkeit einhergeht. IDRs sind insbesondere an der Flüssig-Flüssig-Phasentrennung (LLPS) beteiligt und spielen auch bei neurodegenerativen Erkrankungen wie der amyotrophen Lateralsklerose (ALS), sowie der Alzheimer- und der Parkinson-Krankheit eine entscheidende Rolle.

Unser Ziel war es, am Beispiel einer essentiellen IDR die Bestandteile zu bestimmen, die für die Funktion der IDR entscheidend sind. Ein solches Verständnis ist notwendig, um zu verstehen, wie es IDRs gelingt, ihre vielfältigen zellulären Funktionen auszuüben, trotz geringer Konservierung der Sequenz und des Fehlens einer definierten dreidimensionalen Struktur. Abf1 ist ein sogenannter allgemeiner regulierender Faktor (GRF) in der Bäckerhefe *Saccharomyces cerevisiae*, der für das Überleben der Zelle essentiell, an der Chromatinorganisation und der Transkriptionsregulierung, sowie der Struktur von Telomeren beteiligt ist und möglicherweise eine Rolle bei der Replikation innehat. Er besitzt eine zweigeteilte, sequenzspezifische DNA-Bindungsdomäne (DBD) und zwei IDRs, IDR1 und IDR2. Wir untersuchten die funktionellen Bestandteile der IDRs in Abf1 mittels eines klassischen Plasmid-*Shuffling*-Tests. Zuerst konnten wir zeigen, dass ein N-terminaler Teil von IDR2 für die noch unbekannte, essentielle Funktion von Abf1 notwendig und hinreichend ist und Teil von IDR2 diente für alle weiteren Versuche als unsere Modell-IDR. Im evolutionären Vergleich von Hefespezies waren nur 5 von 18 IDRs aus Abf1-Orthologen mit konservierter Aminosäurezusammensetzung in der Lage, Lebensfähigkeit aufrechtzuerhalten. Das bedeutet, dass die Aminosäurezusammensetzung der IDR2 alleine für die essenzielle Funktion von Abf1 nicht ausreicht. In einem funktionellen Vergleich von IDRs von Proteinen, die funktionell ähnlicher oder verwandt zu Abf1 sind konnten nur einige davon die essentielle Funktion von Abf1 aufrechterhalten. Auf der Suche nach funktionell wichtigen Sequenzmotiven verwendeten wir blockweise Sequenzdurchmischungen, bei denen die gesamte Aminosäurezusammensetzung unverändert blieb, aber die lineare Sequenz innerhalb von Blöcken zufällig durchmischt wurde.

Auf diese Weise identifizierten wir ein 20 Aminosäurereste kurzes, lineares Motiv (*short linear motif*, SLiM), das "essenzielle Motiv" (*essential motif*, EM). Die Funktion des EM hing von seiner linearen Sequenz ab und war modular, da es Lebensfähigkeit verleihen konnte, wenn es in ansonsten nichtfunktionale IDR-Kontexte eingefügt wurde. Eine Unterregion von Abf1, zwei Aktivierungsdomänen von Gal4, eine transiente Helix des RNA-bindenden Proteins TDP–43, sowie fünf weitere Unterregionen aus menschlichen Transkriptionsfaktoren waren ebenfalls alle in der Lage, ansonsten nichtfunktionelle IDRs lebensfähig zu machen. Im Gegensatz zum EM von Abf1 war die Funktion all dieser anderen Sequenzen aber nicht von ihrer linearen Sequenz abhängig, da ihre Aminosäuren über den IDR-Kontext verteilt werden und immer noch Lebensfähigkeit vermittelten konnten. Die so erzeugten IDRs waren also nicht funktionell, weil sie ein entsprechendes SLiM (Sequenzspezifität), sondern lediglich, weil sie die benötigte globale Aminosäurezusammensetzung (chemische Spezifität) enthielten.

Letztere Ergebnis war überraschend, da die Funktion des Abf1-Wildtyps eben nicht nur von der chemischen Zusammensetzung des IDR-Kontexts, sondern auch vom linearen Sequenzmotiv EM abhing. Die Bedeutung der richtigen chemischen Zusammensetzung des IDR-Kontextes wurde durch Veränderungen der Azidität, der Hydrophobizität und der Aromatizität nachgewiesen: Alle drei Merkmale waren für einen lebensfähigen IDR-Kontext unerlässlich, unabhängig von der Art des SLiM oder funktionell ähnlichen Teilsequenzen. Wir fanden also, dass die essentielle Funktion einer IDR, die in der Wildtypform sowohl von der chemischen Zusammensetzung des IDR-Kontexts als auch von einem SLiM abhängt, vollständig durch einen geeigneten IDR-Kontext ohne SLiM ersetzt werden konnte.

Insgesamt konnte die Lebensfähigkeit der Abf1-IDR-Mutanten entweder durch eine Kombination aus Sequenzspezifität und chemischer Spezifität oder allein durch chemische Spezifität des IDR-Kontexts ohne SLiM vermittelt werden. Da beide Beiträge bekanntlich IDR-vermittelte Interaktionen begründen, vermuten wir, dass die essentielle Funktion von Abf1 auf solchen IDR-vermittelten Interaktionen beruht. Allgemeiner schlagen wir eine zweidimensionale Landschaft aus Sequenz- und chemischer Spezifität vor, um zu beschreiben, wie Kombinationen aus linearem Motiv (SLiM) und „Stärke“ des IDR-Kontextes die Bindung einer IDR an ihre(n) Partner ermöglichen.

Zusätzlich zur Mutantenstudie *in vivo* untersuchten wir mit unserem *in vitro* Chromatin-Rekonstitutions-System in genomweitem Maßstab, ob auch ein Transaktivator die bereits für GRF gezeigte Barriere-Funktion für die Nukleosomenpositionierung haben kann. Vorläufige Ergebnisse deuteten darauf hin, dass der Hefe-Transaktivator Pho4 eine schwächere Barriere für die Positionierung von Nukleosomen darstellt als Abf1.

SUMMARY

Due to the prevailing structure-function paradigm there was at first little interest of protein researchers in intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs). Since the end of the 1990s growing experimental evidence demonstrated that the absence of a defined three-dimensional protein structure under physiological conditions offers structural and therefore functional versatility. In particular, IDRs are centrally involved in liquid-liquid phase separation (LLPS) and play crucial roles in neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS), Alzheimer's as well as Parkinson's disease.

We aimed to decipher the functional determinants of an essential IDR, as they are key to understanding how IDRs carry out their diverse cellular functions despite poor sequence conservation and lack of a defined three-dimensional structure.

Abf1 is a so called general regulatory factor (GRF) in baker's yeast *Saccharomyces cerevisiae* that is essential for viability, involved in chromatin organization, the regulation of transcription and telomere structure and may have a role in replication. It has a bipartite sequence-specific DNA-binding domain (DBD) and two IDRs, IDR1 and IDR2. We examined the functional determinants of the IDRs in Abf1 via classical plasmid shuffling assay. First, we found that an N-terminal part of IDR2 is necessary and sufficient for the yet unknown Abf1 function that is essential for viability. The correspondingly C-terminally truncated IDR2 was our model IDR for all further studies. In an evolutionary comparison, only 5 out of 18 IDRs from Abf1 orthologs with conserved amino acid composition were able to confer viability. This means that amino acid composition of IDR2 alone is not sufficient to maintain the essential Abf1 function. In a functional comparison, we tested IDRs from proteins functionally similar or related to Abf1 and only in some cases their IDRs could substitute for the essential Abf1 function. In search for functionally important sequence motifs, we used sequential sequence shuffles that left the amino acid composition unaltered but shuffled blocks of linear sequence. This way, we identified a 20-residue short linear motif (SLiM) in IDR2, the "essential motif" (EM). EM function depended on its linear sequence and was modular as it conferred viability if inserted into otherwise inviable IDR contexts. An Abf1 subregion, two activation domains of Gal4, a transient helix from RNA-binding protein TDP-43 and five other subregions from human transcription factors were also all able to provide viability to otherwise inviable IDRs. In contrast to the EM of Abf1, the function of all these other sequences did not depend on their linear sequence as their amino acids could be distributed across the IDR context and still provide viability. Such IDRs were not functional due to a certain SLiM (sequence specificity),

but only due to necessary amino acid composition (chemical specificity). The latter finding was surprising, given that the function of wild type Abf1 depended not just on the chemical composition of its IDR context, but also on the linear sequence motif EM. We demonstrated the importance of proper IDR context chemistry by changes in acidity, hydrophobicity, and aromaticity: all three features were essential for a viable IDR context, independent of the kind of SLiM or functionally similar subsequences. Strikingly, we found that the essential function of an IDR, that depended in its wildtype form on both the proper chemical composition of the IDR context and on a SLiM, could be entirely replaced by a suitable IDR context without SLiM. Collectively, viability of *abf1*-IDR-mutants depended on either a combination of sequence-specificity and chemical specificity or on chemical specificity without SLiM alone. As both contributions are known to facilitate IDR-mediated interactions, we suggest that the essential function of Abf1 is based on IDR-mediated interactions. More generally, we propose a two-dimensional landscape of sequence- and chemical specificity to describe how combinations of linear motif (SLiM) and IDR context strength enable the binding of an IDR to its partner(s). In addition to the mutant studies in vivo, we employed our genome-wide *in vitro* chromatin reconstitution system to investigate if a transactivator could function as a nucleosome positioning barrier as shown for GRFs. Preliminary results suggested that the transactivator Pho4 is a weaker nucleosome positioning barrier than Abf1.

1 INTRODUCTION

1.1 Intrinsically disordered proteins and protein regions

Intrinsically disordered proteins and intrinsically disordered protein regions (from here on referred to as IDRs) are functional but lack the amino acid sequence-encoded ability to fold into well-defined three-dimensional structures on their own under physiological conditions. Nonetheless, their disordered state is considered thermodynamical stable[1] and they may also fold, at least partially, upon binding to interaction partners[2], although this feature is not a pre-requisite for function. IDRs are present in all three domains of life but are most prevalent in eukaryotes[3] and are classified as "long" if they comprise more than 30 to 40 residues[4,5]. IDR-mediated protein-protein interactions can occur with low binding affinity but high specificity (see 1.1.3) and they are present, e.g., when the function requires fast binding and regulation, for example in cell signaling (see 1.1.4).

1.1.1 IDRs in the focus of protein research: expanding the structure-function-paradigm

Until the late 1990ies, the general view in the field of protein research was that a well-defined, three-dimensional structure is a pre-requisite for an active, functional protein. This so-called structure-function paradigm assumed that the amino acid sequence determines the structure of a protein and the structure in turn determines the function of the protein. In other words: function follows form and hence function cannot be conducted without a well-defined folded structure. This view was shaped by the structures of proteins that were solved at that time, which were mostly well-folded, often globular proteins[6], such as myoglobin[7] (the first protein structure at all that was solved back in 1958), haemoglobin[8], chymotrypsin[9] or lactate dehydrogenase[10]. Nonetheless, it had also been reported that, for example, High-Mobility-Group Chromosomal Protein (HMG) 17 is disordered[11] and that monomeric glucagon adapts a flexible random coil formation in aqueous solution[12]. Another reason why IDRs were neglected in protein research until then was that IDPs behave differently from globular proteins in X-Ray crystallography and NMR spectroscopy: in the latter, disordered regions display a lack of amide hydrogen chemical shift dispersion[2], and in the former IDPs often prevent crystal formation. Even if they successfully crystallize, disordered regions hold no significant electron density[13,14] and disordered residues are therefore "invisible" in X-Ray crystallography and 2D NMR spectroscopy. Finally, the structure-function paradigm was questioned by a handful of

researchers[5,15]. This altered view on IDRs and the growing evidence that IDRs actually show great functional flexibility despite the lack of a defined structure led to an enormous increase in the interest in understanding how IDRs work. A schematic representation of the continuum from disordered to folded proteins is shown in Figure 1.

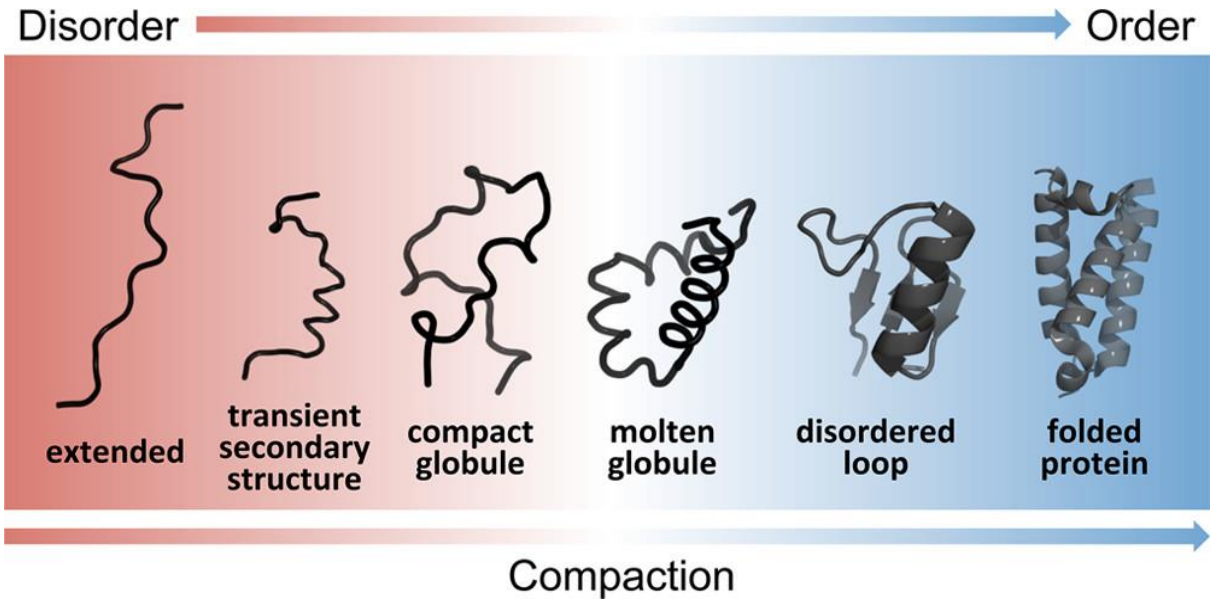

Figure 1: Schematic visualization of the diverse spectrum of (dis-)ordered protein structures. The color gradient from red (on the left) to blue (on the right) represents the structural variety from disordered proteins over compact and molten globules to folded proteins. Taken from[16] (https://pubs.acs.org/doi/10.1021/cr400525m) and reprinted with permission from the American Chemical Society (ACS). Further permissions related to the material excerpted must be directed to the ACS.

It is now established that IDRs are as important as folded proteins and play a crucial role in the biological context: from cell signaling and regulation[17,18] and their role as key players in liquid-liquid phase separation (LLPS)[19], to the emergence of diseases such as Alzheimer's disease[20], Parkinson's disease[21] and cancer[22]. Prominent examples are cyclin-dependent kinase inhibitor p21[23], tumor suppressor protein p53[24], RNA-binding proteins FUS[25] and TDP-43[26], yeast translation termination factor Sup35[27] and neuronal α-synuclein[28].

1.1.2 The correlation between amino acid sequence and conformation of IDRs

IDRs are characterized by their ability to function without adapting a distinct three-dimensional structure when in an unbound state. Instead, the conformations of IDRs are represented by an entire structural ensemble that depends on the properties of their amino acid sequence[29,30].
The compositions of amino acid sequences of IDRs are different from that of folded, globular protein domains. The sequences are of low complexity and show a compositional bias: they

are rich in polar and disorder-promoting amino acids (especially lysine, glutamic acid, proline and serine) and depleted in hydrophobic and order-promoting amino acids (i.e. cysteine, tryptophane, tyrosine, isoleucine and valine) and therefore have a high net charge and a low hydrophobicity[31–33]. IDRs are biopolymers and can be classified either as polar tracts, polyampholytes or polyelectrolytes based on the properties of their amino acid composition and can therefore also be described by the same parameters that are used to characterize the aforementioned. These parameters are the fraction of positively and negatively charged residues f_+ and f_-, the thereof derived fraction of charged residues (FCR) and net charge per residue (NCPR) and the patterning parameter κ. The FCR is the sum of f_+ and f_-. NCPR is the absolute value of the difference between f_+ and f_-[34]. The parameter κ quantifies how oppositely charged residues are distributed along a linear sequence. κ can take on a value between 0 and 1. A value of 0 corresponds to sequences where positively and negatively charged residues alternate evenly and a value of 1 to sequences where the oppositely charged residues each cluster together[30]. Sequences with higher κ-values exhibit higher binding affinities[35]. Based on f_+ and f_-, IDRs can be assigned to one of four conformational classes (Figure 2).

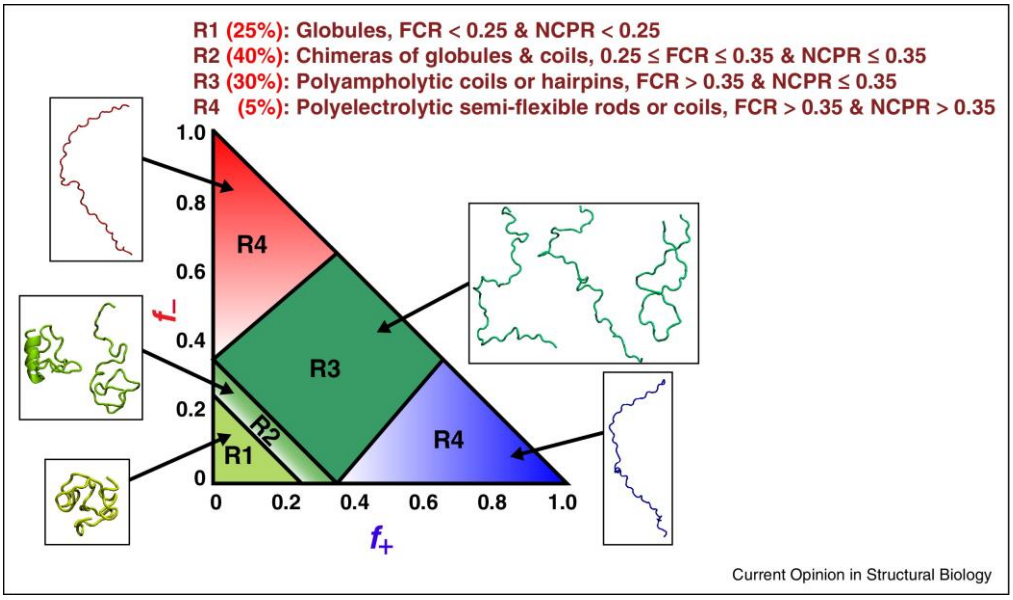

Figure 2: Diagram-of-states illustrating the correlation between IDR sequence and conformation. The number of positively and negatively charged residues in relation to the number of total residues in the IDR sequence (f_+ and f_-) defines the affiliation to one of the four conformational classes R1 to R4. Adapted and reprinted with permission from Elsevier[34].

The conformational ensemble in turn confers flexibility and enables IDRs to carry out a diverse range of biological functions with an equally versatile range of interaction partners[1].

1.1.3 IDRs as mediators of protein-protein interactions

Protein binding and interaction models were first generated for enzymes and their substrate. According to the traditional view, large, structured globular protein domains are required to mediate protein-protein interactions[36–38]. The lock and key model assumed that the interaction surface of two molecules, like the active site of an enzyme and its substrate, fit complementarily to each other like a lock and its key[38]. This model was replaced by the induced-fit model that assumed that the interacting structure of one of the interaction partners (the enzyme) is somewhat flexible and can adapt its structure to some extent in a way that it fits the respective interaction partner (the substrate)[39]. The current view includes disordered proteins as well: the entire range of all protein-protein interactions (the interactome) actually comprises a multitude of binding surfaces that exhibit an alternating degree of flexibility, ranging from stiff globular domains to flexible, disordered protein regions that lack native three-dimensional structures[36]. Moreover, the mechanism of coupled folding and binding was specifically proposed for IDRs[40]. According to this mechanism, the binding of an IDR to its interaction partner with at least a partial disorder-to-order transition can be achieved in two ways (Figure 3). In the first way, the conformational change in the IDR is induced by binding of the IDR to its interaction partner and completed upon binding. In the second way, a conformation that is similar to the one in the bound form is part of the conformational ensemble of the unbound IDR and is selected upon binding of the IDR and to its interaction partner[40].

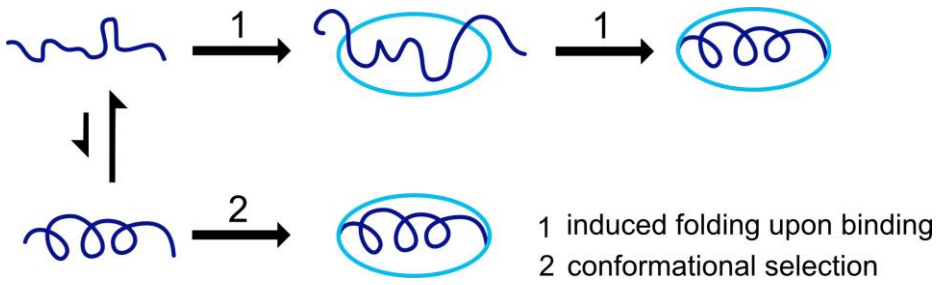

Figure 3: Schematic depiction of two possible ways for coupled folding and binding involving IDRs. For the first way shown, the fully disordered IDR meets its interaction partner, which in turn induces the IDR to fold. For the second way shown, the IDR adopts a conformation from its entire ensemble that resembles a conformation close to that in the bound form. The IDR undergoing the disorder-to-order transition is shown in dark blue and its interaction partner is depicted as a light blue oval. Based on[40].

In summary, although IDRs do not adopt rigid 3D-structures they can at least partially fold upon binding to their interaction partners[41,42]. IDRs are not inert, unstructured constructs but do engage in protein-protein interactions and are particularly well-suited for this due to their flexibility. Such interactions are mediated either by short linear motifs encoded in the linear IDR sequence (1.1.3.1) or by the surrounding IDR context (1.1.3.2).

1.1.3.1 Short linear motifs

Short linear motifs (SLiMs) are protein-protein interaction interfaces of 5-15 residues length that occur preferentially in IDRs[36,37]. Other designations for SLiMs are eukaryotic linear motif (ELM) or mini motif. SLiMs can, e.g., be present in calmodulin-binding sites[43], 14-3-3 protein recognition elements[42], phosphorylation[44] and deacetylation sites[45] or interaction motifs of SH2 and SH3 domains[46,47]. Furthermore, SLiMs are characterized by their modularity: within an IDR they function as molecular recognition modules for their binding and interaction partners[36,48–51]. The interaction surface between a SLiM and its binding partner is on average built by only three to four SLiM residues[52].

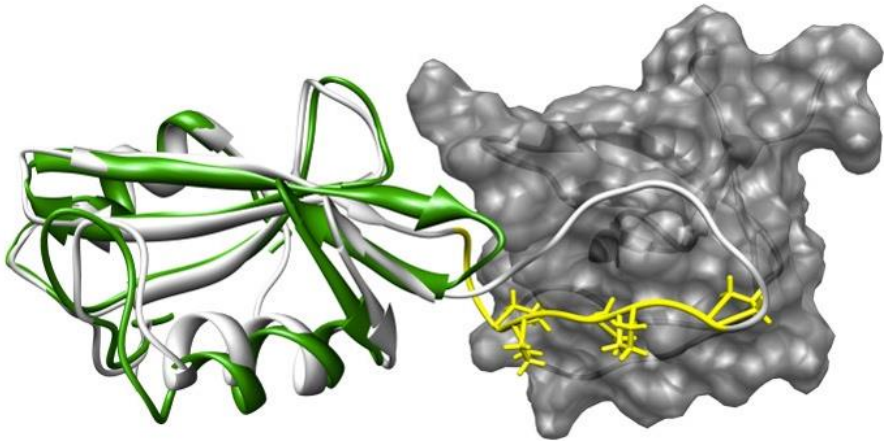

Figure 4: Structure of the SLiM of human proto-oncogene CRK binding to the SH3 domain of tyrosine kinase ABL1. The SH3 domain of ABL1 is shown as cartoon and overlayed surface in dark grey. Homologous SH2 domains of CRK (light gray, PDB 1JU5[53]) and Crk-like protein CRKL (green, PDB 2EO3) are overlapping. The SLiM (a PxxP motif) located in the loop of human CRK is depicted in yellow. Adapted[52] with permission from Kim Van Roey, Bora Uyar, Robert J. Weatheritt, Holger Dinkel, Markus Seiler, Aidan Budd, Toby J. Gibson, and Norman E. Davey, *Chemical Reviews* 2014 *114* (13), 6733-6778 , DOI: 10.1021/cr400585q. Copyright 2023 American Chemical Society.

Due to the limited number of residues that actually contribute to binding, the protein-protein interactions mediated by SLiMs occur with low micro molar affinity, typically in a range between one to 150 µM[51,54], which enables local flexibility and transient, reversible interactions[36,40]. Nevertheless, SLiM-mediated interactions display a high specificity, as it was demonstrated for the *S. cerevisiae* protein Pbs2: this protein contains a SLiM that interacts specifically with the SH3 domain of Sho1 (the Pbs2 interaction partner) as well as non-yeast SH3 domains, but no interactions occur between Pbs2 and the other 26 SH3 domains encoded in the *S. cerevisiae* genome[55]. Notably, it was observed that a SLiM can adapt an α-helical structure after binding to its interaction partner, thereby following the coupled folding and binding mechanism[56].

SLiMs exhibit a great evolutionary plasticity and the mechanism of an *ex nihilo* motif evolution has been proposed, in which a functional motif sequence is newly acquired "out of nothing" through randomly occurring mutations[57,58] (Figure 5). An example for evolutionary plasticity of SLiMs is a C-mannosylation site in human Interleukin-12 that is also found in other metazoans but abolished in rodents by the mutation of a single tryptophane to cysteine[57].

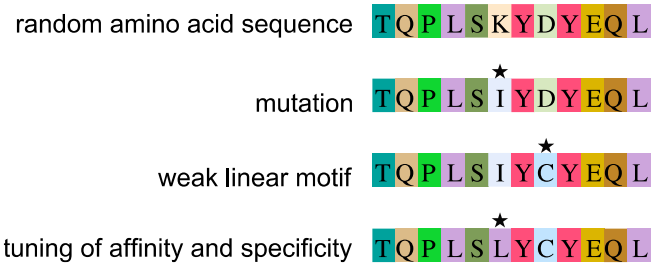

Figure 5: Example of an *ex nihilo* motif evolution of a short linear motif from a random amino acid sequence. The newly introduced mutation in comparison to the preceding sequence is indicated by an asterisk above the respective residue. Based on[58].

For the annotation of ELM classes within the ELM database[59], positions in the linear amino acid sequence of the SLiMs are distinguished between undefined and defined positions. The latter distinguishes between fixed positions where the exchange of one amino acid to another is not tolerated at any position and degenerated positions that allow the substitution of amino acids to a certain degree with amino acids that have similar physicochemical or structural properties. Undefined SLiM positions can be either wildcard positions, where any amino acid replacement works or prohibited positions, where some substitutions with the exception of a few amino acids are allowed[36]. In comparison to finding and identifying folded, globular domains, the prediction and identification of functional SLiMs, either via sequence alignment-based methods or via biological binding assays, was previously an especially challenging task due to their short length as well as their low affinity to their interaction targets[36,57]. To identify SLiMs, their evolution has been a useful aspect to take into. As SLiMs often evolve convergently[36] this feature has been employed to finding SLiMs. In addition, SLiMs are more conserved than their adjacent, non-functional residues[63] and their identification as "islands of conservation"[83–87] was used for their discovery before bioinformatic-based[64–66] and proteome-wide methods[48,50].

1.1.3.2 IDR context

In addition to short linear motifs, the IDR context itself can contribute to IDR-mediated protein-protein interactions. A broad view of the term "context" is of multidimensional nature and

comprises the context of the molecular interaction as well as the physicochemical, biomolecular and physiological environment. The physicochemical context includes, e.g., pH, salt concentration and viscosity, the biomolecular context describes interactors and location, whereas the physiological context encloses cell type, cell cycle and stress[67]. Here, the term IDR context is used to describe the IDR excluding any SLiMs. The context can confer specificity to protein interactions either by increasing the binding affinity or by averting non-native interactions[68]. The ability of IDRs to engage especially in multivalent interactions is encoded in the chemistry of the context without regard of the linear sequence arrangement of the amino acid residues[35,69–71].

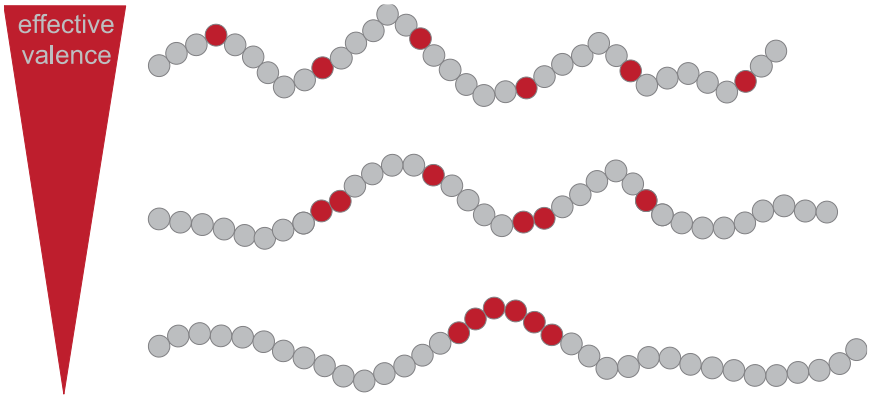

Figure 6: Schematic illustration of the effective valence of IDR context. The three different IDR contexts have the same apparent valence (the same number of residues that could participate in multivalent interactions) but a decrease in spatial distance is accompanied by a decrease in effective valence of the context. Inspired by[71].

An example for the role of sequence-encoded chemistry is the Sic1-Cdc4 system. Binding between the intrinsically disordered cyclin-dependent kinase inhibitor Sic1 and its interaction partner Cdc4, an adapter subunit of SCF ubiquitin ligases, is achieved through cumulative charge interactions: When Sic1 is phosphorylated at any six of its nine available phosphorylation sites, the binding of Cdc4 via its WD40 domain and subsequent destination of Sic1 for ubiquitin-dependent proteolysis is triggered[72–74]. In addition to such polyelectrostatic nature, sequence-encoded chemistry can also comprise multivalent polycationic-π interactions between tyrosine residues distributed along the IDR context and arginine residues at the binding surface of a respective interaction partner[75].

Multivalent interactions of IDR context can be stoichiometric[76–78] or support the phase separation behavior of biomolecular condensates[79–81].

Stoichiometric interactions can occur between two IDRs as is the case for the ultrahigh-affinity interaction of the predominantly unstructured, highly positively charged linker histone H1 and

its fully unstructured, highly negatively charged interaction partner nuclear protein prothymosin-α[76]. Moreover, an IDR context and a folded interaction partner can also engage in multivalent stoichiometric interactions. This was observed for the interaction between the disordered E-cadherin and the folded β-catenin in the cell adhesion complex, where a combination of many weak and few continuous context contacts results in a partial de- and re-attachment of the IDR to β-catenin, which gives regulatory enzymes access to the complex and at the same time enables interactions with high-affinity, specificity and flexibility[77]. Moreover, multivalent stoichiometric interactions between the IDR of a yeast TF are responsible for guiding the TF's promotor preferences[78]. Multivalent interactions involving the IDR context that influence the phase separation behavior of IDRs rely on the overall amino acid composition of the context instead of the linear sequence[79,80,82].

1.1.4 Functions of IDRs

The lack of a rigid, three-dimensional structure confers adaptability and flexibility to IDRs, that allow them to fulfill a broad range of important biological functions especially when fast and transient reactions, interactions and cellular responses are required. This comprises diverse cellular processes from cell signaling[17], cell division[83], cell cycle control[84] to posttranscriptional regulation[85], protein degradation[86] and apoptosis[87]. Moreover, IDRs hold an important role in transcription, liquid-liquid phase separation (LLPS) and chromatin organization (1.1.4.1) as well as in several diseases (1.1.4.2).

1.1.4.1 IDRs in transcription, chromatin organization and LLPS

Transcription regulation through a transcription factor (TF) requires its sequence-specific binding to DNA and the interaction with transcriptional coactivators or corepressors[88,89]. The domain architecture of a typical TF fits its function: besides at least one structured, sequence-specific DNA binding domain there is an often intrinsically disordered region which can either act as a transcription activation domain (TAD) or a transcription repression domain[89]. Prominent examples are Gal4, Gcn4 and Pho4 in *S. cerevisiae* and human tumor suppressor p53. The IDR of a TF can modulate affinity[90] as well as specificity[91] of the TF to DNA and stabilize their binding[92]. It has been shown in *S. cerevisiae* that amongst related TFs IDRs are necessary and sufficient to guide the DNA binding specificity independently from the DBD and also from the linear amino acid sequence[93]. Furthermore, IDRs can finetune transcription

by functioning as regulators of negative feedback loops as it was demonstrated for the TF HIF-1: during hypoxia its IDR subunit HIF-1α forms a complex with the TAZ1 domain of the transcriptional coactivators CBP/p300 and promotes the rapid activation of several genes among them *CITED2*[94,95]. The protein CITED2 can in turn form a ternary complex with the TAZ1 domain and HIF-1, which then causes the disassociation of HIF-1 from the complex and thereby brings the hypoxia response to a hold[94]. IDRs are preferred sites for post-translational modifications (PTMs) such as phosphorylation, acetylation, alkylation or glycosylation[96]. The phosphorylation state of the TF IDR can influence its interaction with DNA: in p53 the phosphorylation state of one specific threonine residue (T55) in the intrinsically disordered, N-terminal TAD works like a regulatory switch that either activates or terminates the transcription of genes involved in DNA damage and stress response T55[97,98].

Another prominent example for regulation of IDRs through PTMs are histones, proteins that are key to chromatin organization and form a nucleosome together with DNA. Nucleosomes are the building blocks of chromatin. One canonical nucleosome consists of a histone octamer that contains two copies each of the four core histones H3, H4, H2A and H2B, as well as 145-147 bp of DNA that are wrapped around the histone octamer and may be further stabilized by linker histone H1[99,100]. The C-terminal domain of linker histone H1 is also disordered in aqueous solution and becomes folded upon binding to DNA[101]. The N-terminal tails of the core histones are rich in positively charged amino acids, are largely intrinsically disordered in absence of DNA and enriched in sites for PTMs that allow for a very comprehensive regulation of chromatin structure and hence transcriptional regulation[102]. One example is acetylation. It is associated with gene activation: this PTM neutralizes the positive charge of lysine residues at the histone tail and thereby weakens the electrostatic interactions that stabilize the histone tail-DNA interactions[102]. This way, the deacetylation of nucleosomes behind an elongating RNA polymerase II can prevent the initiation of cryptic transcription[103]. Recent research has suggested that transcription, chromatin organization and LLPS influence each other. TFs can undergo LLPS, re-organize chromatin structure and regulate the transcription process[104,105]. LLPS is a non-stoichiometric de-mixing process that occurs when in a mixture of supersaturated macromolecules (e.g. proteins, nucleic acids, polysaccharides or a combination of all of them) two separate stably coexisting liquid phases are formed. This leads to the formation of liquid droplets or, in cells, biomolecular condensates such that one phase is enriched and the other phase depleted of certain macromolecules while the interchange of other molecules is still enabled[19,71,106,107]. This enables specific biochemical reactions in a

membraneless but compartmentalized space and is an advantage for the regulation of gene expression[108], DNA repair[109] and chromatin organization[110]. Here, IDRs are key players: LLPS promoting or opposing behavior of the IDR is mediated by the IDR context through multivalent interactions and mainly influenced by overall amino acid composition of the context rather than by linear sequence[79–82].

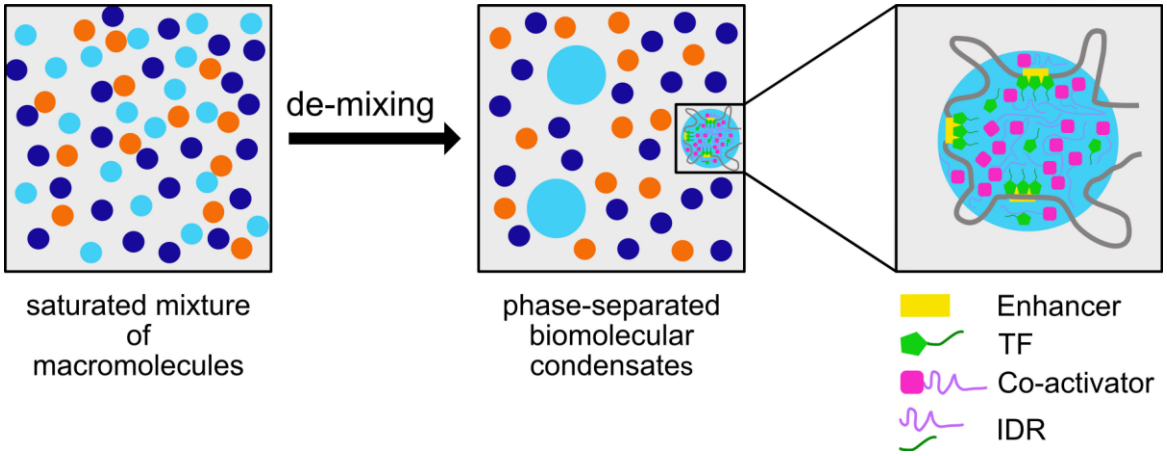

Figure 7: Schematic representation of liquid-liquid phase-separation leading to biomolecular condensates. The de-mixing of supersaturated macromolecules drives the formation of phase-separated membraneless biomolecular condensates. These enable the compartmentalization of biochemical reactions like transcription as depicted here and the enrichment of the necessary key factors. Inspired by[111,112].

Aromatic residues have been shown to support LLPS, especially if they are uniformly distributed along the IDR sequence[79,80]. It has been shown that the LLPS saturation concentration is modulated not only by any generic cationic-π interactions but by specific contributions of both tyrosine and arginine residues[81]. Moreover, LLPS can be driven by charge neutralization through the interaction of patterned charged sequence blocks that exhibit a considerable net charge with oppositely charged multivalent ligands[82].

PTMs within the IDR also govern its LLPS behavior. Depending on their phosphorylation state serine and tyrosine residues can either promote or oppose LLPS: the (non-phosphorylated) tyrosine residues present in RNA-binding protein FUS are required for the protein to undergo LLPS whereas the phosphorylation of its serine residues causes the disassembly of liquid droplets[79].

1.1.4.2 IDRs in diseases

The interest in IDR research is also linked to the discovery of their involvement in several neurodegenerative human diseases but also in cancers[110]. The sequence of IDRs often contains

regions with low complexity that hold several repeats of the same amino acid, like glutamine (so-called poly-Q tracts), which are associated with neurodegenerative diseases, like Huntington's disease, that involve the formation of misfolded or aggregated proteins[32,114]. Moreover, IDRs occur in other neuronal diseases: plaques of amyloid-β peptide and aggregates of the tau protein that occur in Alzheimer's disease[20], aggregates of α-synuclein in Parkinson's disease[28] or TDP-43 in amyotrophic lateral sclerosis (ALS)[113,115]. Furthermore, many misregulated TFs contribute to tumorgenesis like the prominent examples of the tumor-suppressor protein and "guardian of the genome" p53[24] and the proto-oncogene c-Myc[116,117].

1.2 The general regulatory factor Abf1

Autonomously replicating sequence (ARS) binding factor 1 (Abf1) belongs to a group of proteins in the budding yeast *Saccharomyces cerevisiae* known as general regulatory factors (GRFs). GRFs, such as Rap1, Reb1, Mcm1, Tbf1 and Cbf1, are multifunctional, abundant and usually essential DNA-binding proteins with specific binding sites that are key players in chromatin organization and transcriptional regulation[118–123]. Also Abf1 is encoded by an essential gene and is a sequence-specific DNA binding protein that recognizes the consensus sequence 5'- RTCRYNNNNNACG-3' in which R represents a purine, Y a pyrimidine and N any nucleo-base[124–129]. It was first described in the late 1980ies under different names by several groups independently from each other. First, it was termed SBF-B (silencing binding factor)[130]. Other names were OBF1[131] (origin binding factor 1), BAF1[132] (bidirectional acting factor), GF1[133,134], REB2[135] (ribosomal enhancer binding protein 2[136]) and EBF2[137] (enolase-binding factor). Abf1 consists of 731 amino acid residues and has a molecular weight of 81.7 kDa but migrates at 135 kDa when separated via SDS-PAGE[132,133]. It has a clear domain architecture: it consists of a bipartite DNA-binding domain (DBD) and two long IDRs: IDR1, which is inserted into the DBD, and the C-terminal IDR2 (Figure 8).

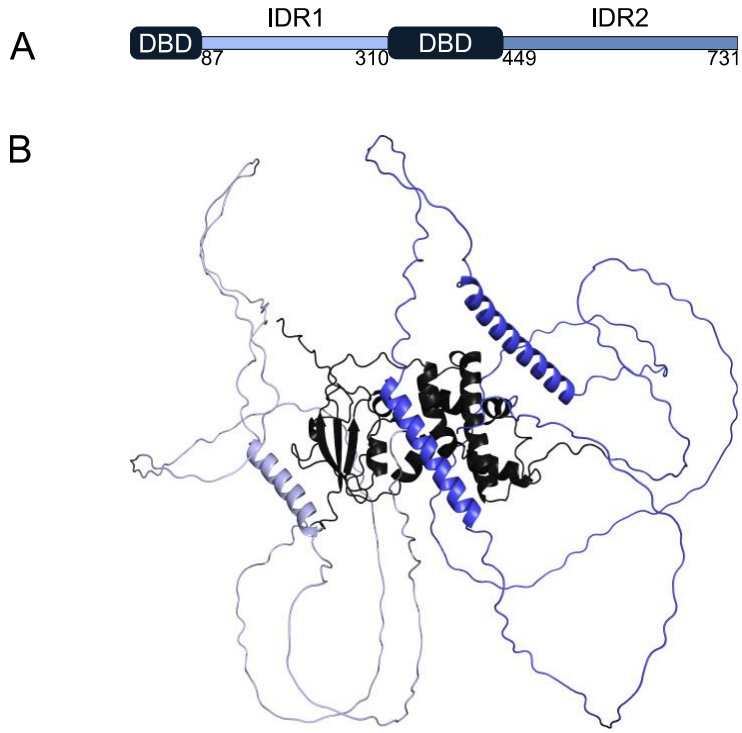

Figure 8: Architecture of ARS-binding factor 1. A) Schematic of the domain structure. Abf1 consists of a bipartite sequence-specific DNA-binding domain (DBD, shown in dark blue) and two long intrinsically disordered regions (IDRs), IDR1 (light blue), inserted into the DBDs, and the C-terminal IDR2 (blue). For simplicity, the bipartite DBD (residues 40-91 and 323-496) is not further distinguished from its surrounding structured domains. **B)** 3D structure (UniProt ID Abf1: P141641) as predicted by AlphaFold2[138,139] and visualized with PyMOL following the above mentioned color code but with the DBD depicted in black.

The Abf1 DBD is located at residues 40-91 and residues 323-496, displaying an atypical single zinc finger motif and a HTH domain, respectively[140,141]. Binding of Abf1 to DNA causes the DNA to bend[127] at an angle of 120°. The two IDRs in Abf1 are not required for DNA binding itself[132,140]. IDR2 contains the critical C-terminal sequence CS1, which comprises the endogenous Abf1 nuclear localization sequence[142] (NLS, residues 624-628), and CS2 (639-662). Both CS1 and CS2 are, in contrast to the very C-terminal residues 663 to 731, both required for cell growth[142,143]. Each IDR of Abf1 is more than 200 residues long and enriched in acidic residues. Acidic IDRs, usually shorter, were associated with transactivation domains. However, the C-terminal IDR2 of Abf1 does not function as one[144,145]. Although Abf1 is not a classical transactivator, it is involved in stimulating transcription activation by functioning as a nucleosome positioning barrier and chromatin re-organizing factor: through its binding at promoters Abf1 establishes a nucleosome depleted region (NDR) followed by a regular nucleosomal array downstream of the Abf1 binding site. Especially the NDR provides access to other activating *trans*-factors. Abf1 has been associated with mating-type locus silencing and telomer organization. Budding yeast has three mating-type loci, all located on

chromosome III: *MAT* and the silent loci *HML* and *HMR*[146,147]. Abf1 recognizes the B-element of *HMRE,* the sequence that silences HMR and comprises the A-, B- and E-element[148,149]. Abf1 also binds next to *ARS1* (autonomously replicating sequence 1)[130] as well as *ARS120*, which is located in a type X telomer[131]. Abf1 is also involved in nucleotide excision repair and important for ribosome biogenesis[150–152]. Several promoters of ribosomal genes contain an Abf1 binding site and, together with a T-rich promoter element, these sites are important for stimulated transcription at these promoters[118,153]. Abf1 likely does this in cooperation with ATP-dependent chromatin remodelers like ISW1a, ISW2 or RSC[118,120,121,143,154–156]. Therefore, Abf1 is ascribed to a group of TFs that are able to alter chromatin but cannot directly stimulate gene expression on their own, in contrast to factors that can do both simultaneously (e.g. Gal4), or factors that enhance expression but do not manage to open chromatin (such as Gcn4)[157,158]. It has also been shown that Abf1 can function as an insulator[119,159]. A C-terminally truncated version of Abf1 with residues 1-562 was sufficient for the chromatin-organizing function of the protein, whereas residues 604-662 are required for insulation[155,160]. Moreover, the effects of Abf1 loss were investigated via a temperature-sensitive (ts) yeast strain with the *abf1-1* ts allele. This allele disables growth at 37°C and caused a changed expression of 235 genes at this restrictive temperature[161]. At the promoter of the gene encoding ribosomal protein L45, the Abf1 binding site could be replaced with a binding site for the GRF Rap1 without loss of transcriptional activity[162]. Several experiments employing different assays were performed that investigated the functional similarity and interchangeability of the C-terminal regions of Abf1 and Rap1 but these experiments were inconclusive. Importantly, all previous studies that investigated truncations of the C-terminal IDR2, including various domain swaps, left the entire IDR1 of Abf1 untouched so that clear conclusion about the role of IDRs in Abf1 function were not possible[163,164]. To date, the function of Abf1 that makes it an essential protein is unknown.

1.3 The transactivator Pho4

Phosphate system positive regulatory protein (Pho4) is a yeast transactivator that belongs to the basic helix-loop-helix (bHLH) class of TFs[165]. Pho4 comprises 312 amino acids with a molecular weight of 34 kDa, has an acidic, intrinsically disordered N-terminal transactivation domain (TAD) (residues 75-99) including a predicted nine amino acids (9aa) TAD motif and binds as a homodimer to the consensus E-box sequence 5’-CACGTG-3’ in UASp elements

(upstream activating sequence phosphate regulated)[165–168]. Activity and cellular location of Pho4 are regulated through phosphorylation of serine-proline dipeptides at five sites (SP1, SP2, SP3, SP4 and SP6) in dependence of intracellular availability of inorganic phosphate[169]. At high phosphate abundance, Pho4 is phosphorylated by the Pho80-Pho85 complex and relocated into the cytoplasm. Upon phosphate starvation, unphosphorylated Pho4 remains in the nucleus and activates transcription of phosphate-responsive genes, e.g. *PHO5*, in collaboration with Pho2[170–172]. The *PHO5* promoter has been studied extensively as a model for promoter chromatin remodeling in the context of gene regulation[173,174]. Under repressive conditions, when inorganic phosphate is available in high concentrations, the Pho4 binding site UASp2 in the *PHO5* promoter is occupied by a nucleosome that prevents Pho4 from binding and triggering a comprehensive nucleosome remodeling transition[175].

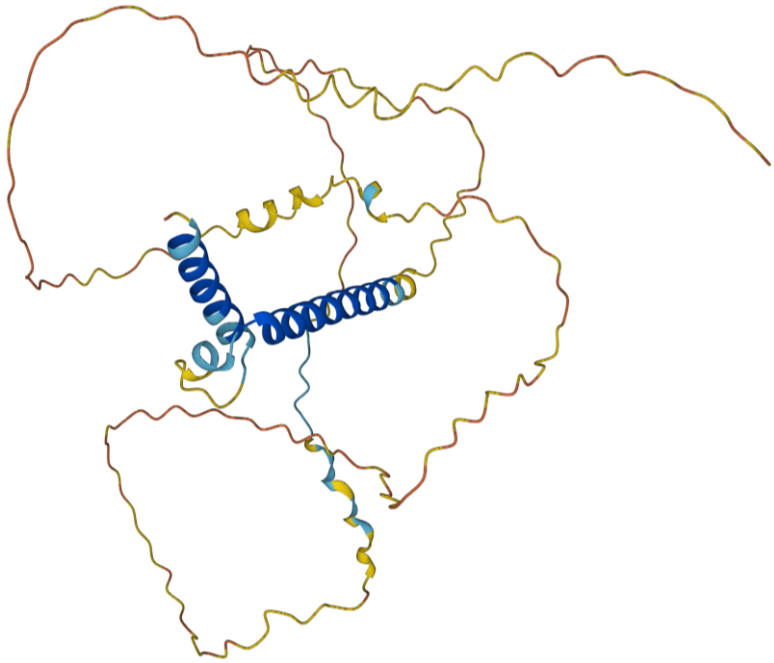

Figure 9: 3D structure of transactivator Pho4 as predicted by AlphaFold2[138,139]**.** Coloring according to the AlphaFold2 color code that distinguishes how confident regions were predicted: with very high confidence (dark blue), confident (light blue), low (yellow) and very low(orange). The bHLH-motif is represented by the two helices in blue that are orthogonal to each other. In this model the predicted 9aa TAD motif is represented as a short α-helical structure depicted in yellow and located above the left blue helix (Uniprot ID Pho4: P07270).

The ability of transactivators to activate transcription is facilitated by their effector domain, the TAD. It has been demonstrated for Pho4 that its TAD triggers both nucleosome remodelling and interacts with the transcriptional machinery[176]. However, the detailed mechanism regarding the mode of action of transactivators is an ongoing subject of research[89,177–179].

2 AIM OF THIS THESIS

Intrinsically disordered regions (IDRs) lack a well-defined three-dimensional structure, which was assumed to go hand in hand with a lack of function until the end of the 1990ies. The experimental evidence that IDRs are instead structurally and functionally versatile and moreover play a crucial role in neurodegenerative diseases such as ALS, Alzheimer's and Parkinson's disease has sparked great continuous interest in the fields of biochemical, biophysical, and medical research.

Despite the lack of tertiary structure, the determinants of IDR function are likely embedded in its amino acid sequence. Sequence alignments are only to a limited extent suitable to identify common features in IDRs as IDRs usually just share a compositional bias towards polar/charged residues and often contain tandem repeats of the same amino acid. Although several biophysical parameters, for example, net charge per residue (NCPR) or fraction of charged residues (FCR), can be used to describe IDR properties, the sequence-embedded determinants of IDR function remain largely unknown.

The GRF Abf1 is an essential protein in *S. cerevisiae* that has two long IDRs and a bipartite sequence-specific DBD. Abf1 functions as a nucleosome positioning barrier, acts as an insulator in transcription and is involved in ribosome biogenesis as well as telomer organization. Which function(s) is/are responsible for the essentiality of Abf1 remains to be elucidated.

The aim of this thesis is, to test if the essential function of Abf1 relies on one or both of its IDRs and if so to elucidate the functional determinants of the essential Abf1 IDR(s) *in vivo*. To this end, we chose a plasmid shuffling assay with viability of an investigated construct as read out: only if an IDR-mutant construct was able to substitute for the Abf1 wild type, i.e., contribute essential function of Abf1, the respective *S. cerevisiae* strain was viable. Furthermore, we intended to determine if the function of Abf1 was based on the function as a nucleosome-positioning barrier and if the functional difference between GRFs and transactivators can be found within their IDRs. To answer these questions, we employed our mechanistically well-defined genome-wide *in vitro* chromatin reconstitution assay.

3 MATERIAL AND METHODS

3.1 Material

3.1.1 Chemicals

General chemicals were purchased from AppliChem, Biozym, Carl Roth, Merck KGaA, neoFroxx, Serva or Sigma-Aldrich (now Merck KGaA). Chemicals with special applications are listed below.

Chemical	Supplier	Identifier
5-Fluorotic acid (5-FOA)	Toronto Research Chemicals	F595000
	Diagnostic Chemicals Limited	1555
Kanamycin (G418)	Sigma-Aldrich	G8168
Gibco Bacto Peptone	Thermo Fisher Scientific	211820
Gibco Bacto Yeast Extract	Thermo Fisher Scientific	212720
Difco Yeast Nitrogen Base w/o Amino Acids	Becton, Dickinson and Company	291920
Bacto Agar	Becton, Dickinson and Company	214010
Universal Agarose	Bio&Sell	BS20.467.500
ME Agarose	Biozym	840014
Ultra Pure Salmon Sperm DNA solution	Thermo Fisher Scientific	15632011
Milk powder	Heirler Cenovis	3030

3.1.2 Consumables

Consumable	Supplier	Identifier
96-well micro-titer plate	Sarstedt	83.3924.005
Porablot NCL Nitrocellulose membrane	Macherey-Nagel	741280
SERVAGel TG PRiME, 4-20%	SERVA	43276.01
0.2 µM membrane Nalgene Rapid Flow	Thermo Fisher Scientific	291-4520
Slide-A-Lyzer MINI Dialysis Device, 3.5K MWCO	Thermo Fisher Scientific	69550
0.5 mm Zirconia glass beads	BioSpec Products	11079105
AMPure XP beads	Beckman Coulter	A63882

3.1.3 Standard yeast media

Standard media were prepared with demineralized water and autoclaved or filter sterilized through a 0.2 µM membrane as required. If not indicated, the recipes are stated for 1 L. For plates, 24 g Bacto Agar were added to the medium usually prior to autoclaving (5-FOA plates being an exception, see below). Specialized buffers and media are described in their corresponding section.

YPDAP	10 g Gibco Bacto Yeast Extract 20 g Gibco Bacto Peptone 20 g glucose 100 mg adenine 1 g KH_2PO_4
YNB w/o ura, w/o leu (pH 5.4)	6,7 g Difco Yeast Nitrogen Base w/o Amino Acids 1,6 g drop out-mix (w/o his, ura, leu, trp) 20 g glucose 84 mg histidine 84 mg tryptophan
YNB w/o ura (pH 5.4)	YNB w/o ura, w/o leu 84 mg leucine
YNB w/o leu (pH 5.4)	YNB w/o ura, w/o leu 84 mg uracil
5-FOA w/o leu plates	YNB w/o ura, w/o leu 50 mg uracil 1 g 5-FOA ad 500 ml with sterile H_2O, filter sterilize, stored at 60°C in incubator until combining with Bacto Agar 24 g Bacto Agar autoclaved in 500 ml H_2O

3.1.4 Amino acid composition of the drop out-mix

Nutrient	Supplier	Identifier	Amount (g)
adenine	Sigma-Aldrich	A-8626	3
L-alanine	Sigma-Aldrich	A-7627	2
L–arginine hydrochloride	Sigma-Aldrich	A-5131	2
L-asparagine anhydrous,	Sigma-Aldrich	A-0884	2
L-aspartic acid	Sigma-Aldrich	A-8949	2
L-cysteine	Sigma-Aldrich	C-1276	2
L-glutamic acid	Sigma-Aldrich	G-6904	2
L-glutamine	Sigma-Aldrich	G-3126	2
glycine	AppliChem	A1067	2
L-histidine	Merck	1.04351.0100	2
myo-inositol	Sigma-Aldrich	I-5125	2
L-isoleucine	Sigma-Aldrich	I-2752	2
L-leucine	Sigma-Aldrich	L-8000	4
L-lysine monohydrochloride	Sigma-Aldrich	L-5626	2
L-methionine	Sigma-Aldrich	M-9625	2
para-aminobenzoic acid	Merck	0059463	0.2
L-phenylalanine	Sigma-Aldrich	P-2126	2
L-proline	Sigma-Aldrich	P-0380	2
L-serine	Sigma-Aldrich	S-4500	2
L-threonine	Sigma-Aldrich	T-8625	2
L-tryptophan	Merck	1.08374.0100	2
L-tyrosine	Sigma-Aldrich	T-3754	2
uracil	Sigma-Aldrich	U-0750	2
L-valine	Sigma-Aldrich	V-0500	2

3.1.5 Bacterial strains

Bacterial strain	Genotype	Manufacturer
DH5α	Genotype: fhuA2 a(argF-lacZ)U169 phoA glnV44 a80a(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 F–, ara,Δ(lac-proAB) [Φ80d lacZΔM15]	NEB
Stellar	rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm	Clontech

3.1.6 Yeast strains

Name	Genotype	Source
Y24962	BY4743; MATa/MATα; *his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; LYS2/lys2Δ0; met15Δ0/MET15; ura3Δ0/ura3Δ0; YKL112w/YKL112w::kanMX4*	EUROSCARF
7-A2	MATa, *his3Δ1; leu2Δ0; ura3Δ0; YKL112w::kanMX4,* pRS416-*ABF1*	this work
7-C5	MATα, *his3Δ1; leu2Δ0; ura3Δ0; YKL112w::kanMX4*, pRS416-*ABF1*	this work
7-D4	MATa, *his3Δ1; leu2Δ0; ura3Δ0; YKL112w::kanMX4*, pRS416- *ABF1*	this work

In strains 7-A2, -C5 and -D4 genotype of *LYS2* and *MET15* was not determined.

3.1.7 Primers for Sanger Sequencing

Primers for Sanger sequencing were purchased from Sigma-Aldrich.

annealing region relative to *ABF1* gene	Part to be validated	Direction	Sequence (5' --> 3')
690 bp upstream of *ABF1*	O1	forward	GATTCAGCCGGGTACTAGTCTAGC
690 bp upstream of *ABF1*	O1, IDR1	forward	CCCATTAACGAAAGTCACCAAAG
IDR1	IDR1	forward	GCAAATAATAATACCAACCCTCCG
O1	IDR1	forward	CGAATTTCAGACGCGTTGCCC
O2	IDR1	reverse	CCAGTGACAAGTTATCTTCTAGTG
O2	IDR1	reverse	GCTTTATACGTCGTTATACGTCTG
IDR1	O2	forward	TCAATAACAACAATGTCGGTAGCC
O2	IDR2	forward	AGTTTCTCATTACGTGGAGGAGTC
IDR2	IDR2	forward	GTACCTCATCGAACACAACC
170 bp downstream of *ABF1*	Flag tag, SV40 NLS, IDR2	reverse	TATCAGAAGGTAAATGGCAAGACG

3.1.8 Antibodies

Primary antibody	Dilution	Company	Identifier
Anti-Flag M2 (mouse), also used for ChIP	1:5000	Sigma-Aldrich (Merck)	F1804
Anti-Histone H3 (rabbit)	1:5000	Abcam	ab1791
Secondary antibody			
IRDye 800CW goat anti-mouse IgG	1:5000	LI-COR Biosciences	926-32210
IRDye 680RD goat anti-rabbit IgG	1:5000	LI-COR Biosciences	926-68071

3.1.9 Kits, Enzymes, Marker

Name	Manufacturer	Identifier
1 kb Plus DNA Ladder	NEB	N3200L
1x Fast SYBR Green Master Mix	Thermo Fisher Scientific	4385612
Apyrase	NEB	M0398L
BamHI-HF	NEB	R0136L
Creatine kinase	Roche Applied Science	1012756600
dNTP mutagenesis Kit	Jena Bioscience	PP-101
DpnI	NEB	R0176L
EcoRI-HF	NEB	R3101L
EcoRV-HF	NEB	R3195L
HindIII-HF	NEB	R3104L
MNase	Sigma-Aldrich	N5386
NEBNext Multiplex Oligos for Illuminia	NEB	E7335L
NucleoBond Xtra Midi	Macherey-Nagel	740410.50
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	740609.250
NucleoSpin Plasmid No Lid	Macherey-Nagel	740499.250
Phusion Polymerase	NEB	M0530S/L
Qubit Assay Tubes	Thermo Fisher Scientific	Q32856
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
T4 DNA Polymerase	NEB	M0203S
T4 Ligase	NEB	M0202S
T4 PNK	NEB	M0201S
T5 exonuclease	NEB	M0363S
Taq DNA Ligase	NEB	M0208S
Taq DNA Polymerase	NEB	M0267S
Triple Color Protein Standard III	SERVA	39258.01
USER enzyme	NEB	M5505L
Zymolyase 100T	MP Biomedicals	08320931

3.1.10 Technical devices

Device	Manufacturer
BlueVertical PRiME mini slab gel unit	SERVA
Bioruptor Pico	Diagenode
ChemiDoc	BioRad
GENESYS 20 Visible Spectrophotometer	Thermo Fisher Scientific
LightCycler C 480 II	Roche
Mini trans-blot Cell	Biorad
Odyssey CLx	LI-COR
Percellys 24 tissue homogenizer (bead beater)	Bertin Technologies
PMQ II photometer	Zeiss
ProFlex PCR System	Thermo Fisher Scientific
Quantum ST4	Vilber
Qubit 3 Fluorometer	Thermo Fisher Scientific
Tetrad dissection microscope MSM400	Singer Instruments
Thermomixer C	Eppendorf
Vortex Genie 2	Scientific Industries
Zentrifuge Eppendorf 5424 R	Eppendorf
Zentrifuge Eppendorf 5810R	Eppendorf

3.1.11 Software

Software	Version	Application
SnapGene	4.3.11	visualization of DNA sequences and DNA sequence alignments
JalView	2.11.1.4	visualization of protein sequences and protein sequence alignments
PyMOL	2.5.2	visualization of three-dimensional protein structures
Affinity Designer	1.10.4	graphic figure design
LightCycler 480 SW	1.5.1	LightCycler qRT-PCR data handling
BioRad Image Lab	6.0	imaging of SDS gels recorded with ChemiDoc

3.2 Molecular cloning methods

All work involving *E. coli* cells was performed under sterile conditions as required.

3.2.1 PCR, Gibson assembly and DNA purification

For cloning PCRs, Phusion DNA polymerase (NEB) was used according to the manufacturer's protocol. Annealing temperatures were estimated/calculated using the NEB tm calculator online tool (https://tmcalculator.neb.com) that is tailored to the specific polymerase and buffers. Usually, 10 to 25 ng template DNA were used. To eliminate methylated template DNA afterwards, 50 µl PCR reactions were digested with 20 U DpnI at 37°C for 1 h. DNA was either subjected directly to the purification column after DpnI digest or after agarose gel electrophoresis and excision of the respective PCR fragment band (NucleoSpin Gel and PCR Clean-up). Insertion of a DNA fragment into a plasmid backbone was achieved by Gibson assembly[180] with a self-prepared mixture (Table 1).

Table 1: Self-prepared Gibson assembly mix

Component	Final concentration
5x Gibson assembly buffer (450 mM Tris-HCl, pH 7.5, 25% PEG 8000, 50 mM $MgCl_2$, 50 mM DTT, 1 mM dNTPs, 5 mM NAD^+)	1,3x
T5 exonuclease	0.005 U/µl
Phusion Polymerase	0.003 U/µl
Taq ligase	5.3 U/µl

A vector-to-insert ratio of 1:3 or 1:5 with 100-150 ng vector was used for the assembly. DNA concentration was measured with the Qubit dsDNA HS assay. The assembly reaction was incubated for 50 min at 50°C, chilled on ice for 2 min and transformed into competent *E. coli* cells prepared by Andrea Schmid as published[181]. In general, DNA was transformed as follows: chemically competent cells were thawed on ice for 10 min and incubated for another 20 min with at least 1 ng DNA or half of the total volume of the assembled Gibson mix. Then, 1 ml LB was added and mixed. Cells were heat shocked for 15 sec at 42°C and incubated on ice for 2 min immediately afterwards. To achieve evenly spaced colonies, 100 µl cells were plated directly onto LB plates containing the appropriate antibiotic. The remaining cells were pelleted for 30 sec at 15,000 rpm (Eppendorf 5424 R), the supernatant was decanted, cells were resuspended in the remaining media and also plated. Small scale DNA isolation was performed

(NucleoSpin Plasmid No Lid). To identify potentially correct clones, isolated DNA was digested with appropriate restriction enzymes, usually HindIII-HF and EcoRI-HF or EcoRV-HF (40 U each in a total reaction volume of 20 µl) for 20 min at 37°C and visualized in an 1% agarose gel in 1x TAE. The correct DNA sequence of selected clones was verified via Sanger sequencing (primers as in 3.1.7). If required, DNA of correct clones was isolated at larger scale using the NucleoBond Xtra Midi kit according to the manufacturer's recommendations.

3.2.2 Cloning of wildtype *ABF1* and derivative truncation constructs

Plasmids pRS315 and pRS416, bearing *LEU2* and *URA3* as auxotrophic marker, respectively, were linearized with BamHI-HF and either HindIII-HF (pRS315) or EcoRI-HF (pRS416)[182], respectively. Genomic DNA from *S. cerevisiae* strain BY4741 served as PCR template for amplification of the wildtype *ABF1* gene. To ensure the presence of both the endogenous promoter and terminator, 690 bp upstream and 170 bp downstream of the coding sequence were included. This *ABF1* fragment was inserted into the above plasmid backbones yielding the plasmids pRS315-*ABF1* and pRS416-*ABF1*. pRS315-*ABF1* served as master PCR template to generate the truncation construct derivates. Except for construct pRS315-*abf1*-IDR2$^{624-662}$, an SV40 nuclear localization sequence (NLS; amino acid sequence: PKKRKV) was added to the C-terminus of all constructs in pRS315 by including the respective DNA sequence in the PCR primer, as well as a triple Flag-tag C-terminal to the SV40 NLS to selected ones, using pAc-sgRNA-Cas9[183] as template for the triple Flag-tag. The SV40 NLS was included to ensure that the proteins are imported into the nucleus and the Flag-tag was included to enable the immunoprecipitation using an anti-Flag antibody (3.3.6). All these constructs were cloned as a combination of insert and backbone. Primer sequences used to generate the inserts for *ABF1* and truncation constructs are listed in Table 2 and the primers used to generate the backbones are listed in Table 3. After PCR, plasmid backbone and insert DNA were purified, ligated via Gibson assembly, transformed into *E. coli* and isolated and validated (3.2.1)

Table 2: Primer used to generate inserts for wildtype *ABF1* and truncation constructs

Construct	Insert primer forward (5'→ 3')	Insert primer reverse (5'→ 3')
pRS416-*ABF1*	CTCTAGAACTAGTGGATCCTTTCTTCCGGTGTCCGTTTC	GACGGTATCGATAAGCTTATACATGTGATTAATATCAGAAGGTAAATGGC
pRS315-*ABF1*	GATAAGCTTGATATCGAATTCTTTCTTCCGGTGTCCGTTTC	CTCTAGAACTAGTGGATCCATACATGTGATTAATATCAGAAGGTAAATGGC
pRS315-*ABF1*-NLS-Flag	AGGAAGGTGGACTATAAGGACCACGACGGAG	TGGGTATCTACTTATCGTCATCGTCTTTGTAATCAATATC
pRS315-*abf1*-ΔIDR1/2	GGTTCTGGTTCTGGTTCTACGTCTATGAATTTAGACGTCTTCAATTC	CTTCTCAACTGGGTATCTAGGAAGATGAAGAAGGTTGTAAATC
pRS315-*abf1*-ΔIDR1/2-NLS-Flag	CAACCTTCTTCATCTTCCCCCAAGAAGAAGAGGAAGGTGG	TGGGTATCTACTTATCGTCATCGTCTTTGTAATCAATATC
pRS315-*abf1*-ΔIDR1-NLS	GGTTCTGGTTCTGGTTCTACGTCTATGAATTTAGACGTCTTCAATTC	CTACACCTTCCTCTTCTTCTTGGGTTGACCTCTTAATTCTGGTTGAATG
pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$ NLS-Flag	GATGGTGAGCTTTCTGGCACGAACTTGAGAAGTAACTCTATCGAC	CTTCCTCTTCTTCTTGGGCTCTACCTCTTTCAAATGCGATG
pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$ -NLS-Flag	AGGAAGGTGGACTATAAGGACCACGACGGAG	TGGGTATCTACTTATCGTCATCGTCTTTGTAATCAATATC
pRS315-*abf1*-ΔIDR1-IDR2$^{624-662}$-NLS	GGTTCTGGTTCTGGTTCTTCCAAAAGACAGCATTTGTCAG	CTTCCTCTTCTTCTTGGGCTCTACCTCTTTCAAATGCGATG
pRS315-*abf1*-IDR2$^{624-662}$	GGTTCTGGTTCTGGTTCTTCCAAAAGACAGCATTTGTCAG	CTTCCTCTTCTTCTTGGGCTCTACCTCTTTCAAATGCGATG

Table 3: Primer used to generate plasmid backbones for wildtype *ABF1* and truncation constructs

Construct	Backbone primer forward (5'→ 3')	Backbone primer reverse (5'→ 3')
pRS315-*ABF1*-WT-NLS	CCCAAGAAGAAGAGGAAGGTGTAGATACCCAGTTGAGAAGACG	CTACACCTTCCTCTTCTTCTTGGGTTGACCTCTTAATTCTGGTTG
pRS315-*ABF1*-NLS-Flag	GACGATAAGTAGATACCCAGTTGAGAAGACGAGC	CCTTATAGTCCACCTTCCTCTTCTTCTTGGG
pRS315-*abf1*-ΔIDR1/2	ACAACCTTCTTCATCTTCCTAGATACCCAGTTGAGAAGACGAGC	AGAACCAGAACCAGAACCTGAGGATGCAGCATTGCC
pRS315-*abf1*-ΔIDR1/2-NLS-Flag	GACGATAAGTAGATACCCAGTTGAGAAGACGAGC	CTTCCTCTTCTTCTTGGGGGAAGATGAAGAAGGTTGTAAATCATTATACTCGC
pRS315-*abf1*-ΔIDR1-NLS	CCCAAGAAGAAGAGGAAGGTGTAGATACCCAGTTGAGAAGACGAGC	AGAACCAGAACCAGAACCTGAGGATGCAGCATTGCC
pRS315-*abf1*-ΔIDR2-NLS	CCCAAGAAGAAGAGGAAGGTGTAGATACCCAGTTGAGAAGACG	CACCTTCCTCTTCTTCTTGGGGGAAGATGAAGAAGGTTG
pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$-NLS	CGCATTTGAAAGAGGTAGAGCCCAAGAAGAAGAGGAAGGTGTAGATAC	GAGTTACTTCTCAAGTTCGTGCCAGAAAGCTCACCATCG
pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$-NLS-Flag	GACGATAAGTAGATACCCAGTTGAGAAGACGAGC	CCTTATAGTCCACCTTCCTCTTCTTCTTGGG

Construct	Backbone primer forward (5'→ 3')	Backbone primer reverse (5'→ 3')
pRS315-*abf1*-ΔIDR1-IDR2$^{449-623}$-NLS-Flag	CCCAAGAAGAAGAGGAAGGTG	CTTGTGAGCATCTTCTGAATC
pRS315-*abf1*-ΔIDR1-IDR2$^{\Delta 449-623}$-NLS	GGTTCTGGTTCTGGTTCTTCCAAAAGACAGC ATTTGTCAG	AGAACCAGAACCAGAACCGGAAGATGAAGAA GGTTGTAAATCATTATACTCG
pRS315-*abf1*-ΔIDR1-IDR2$^{624-662}$-NLS	CATTTGAAAGAGGTAGAGCCCAAGAAGAAGA GGAAGGTG	AGAACCAGAACCAGAACCGGAAGATGAAGAA GGTTGTAAATC
pRS315-*abf1*-IDR2$^{624-662}$	CATTTGAAAGAGGTAGAGCCCAAGAAGAAGA GGAAGGTG	AGAACCAGAACCAGAACCGGAAGATGAAGAA GGTTGTAAATC

3.2.3 Cloning of random mutagenesis constructs

Using plasmid pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$-NLS as PCR template, IDR2 residues 449 to 662 of Abf1 were randomly mutagenized using the dNTP-mutagenesis Kit (Jena Bioscience) according to manufacturer's recommendations. In short, in a first PCR, the region of interest was amplified in 30 cycles in presence of canonical dNTPs and the two mutagenic dNTP analogs 8-oxo-dGTP and dPTP. Then, 1 µl of this PCR reaction was subjected to a second PCR, also with 30 cycles but in absence of mutagenic dNTP analogues. To obtain mutants with a lower mutation rate, the first PCR was also carried out with 5, 10 and 15 cycles. After DpnI digest to remove the non-mutagenized original template, the library of randomly mutagenized constructs was purified and inserted into the template vector pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$-NLS via Gibson assembly, hereby substituting the wildtype sequence of IDR2 from residue 449 to 662. The Gibson assembly mix was transformed and various colonies were used to inoculate cultures for small-scale DNA isolation and processed as described (3.2.1, 3.3.1 and 3.3.4). To selected constructs, a C-terminal 3x Flag-tag was added via Gibson assembly using the same primers as for the generation of construct pRS315-*abf1*-ΔIDR1-IDR2$^{449-662}$-NLS-Flag (Table 2, Table 3). The viability of random mutagenesis constructs, to which the Flag-tag was added later on, was again verified on 5-FOA plates (3.3.4) and they were also transformed in diploid strain Y24962, sporulated and analyzed by tetrad dissection (3.3.2).

3.2.4 Cloning of rationally designed constructs

DNA sequences for rationally designed constructs were codon-optimized for *S. cerevisiae* and synthesized by Gene Art, Thermo Fisher Scientific. The ordered sequences of rational designs

were amplified from the standard vector they were provided in and inserted between ordered region 2 and the C-terminal SV40 sequence via Gibson assembly with pRS315-*abf1*-ΔIDR1-NLS-Flag serving as backbone. There were two exceptions: the first one is the EM-only construct, in which the sequence of interest is additionally flanked by two GSGSGS-linkers, and in the second one *S. cerevisiae ABF1* is fully replaced by its *K. lactis* homologue. The primers used for cloning of rationally designed constructs are listed in Table S1 and Table S2 in the appendix. The assembled constructs were further processed (3.2.1). Validated constructs were transformed into *S. cerevisiae* and their viability assessed in the 5-FOA plasmid shuffling assay (3.3.1 and 3.3.4). Expression and Abf1-site-specific DNA-binding of inviable constructs was verified via ChIP and qRT-PCR (3.3.6).

3.3 Yeast methods

All work involving *S. cerevisiae* was performed under sterile conditions as required.

3.3.1 Yeast transformation

Yeast transformation was performed according to CSH protocol[184], which achieves a highly efficient transformation by combining LiOAc treatment with PEG, single-stranded carrier DNA and DMSO[185,186]. In brief, 5 ml medium (YPDAP for Y24962, YNB w/o ura for 7-A2, -D4 and -C5, 3.1.6) were inoculated with a fresh single colony of the respective strain and incubated overnight (O/N) while shaking at 130 rpm and 30°C. Fifty ml YNB w/o ura were inoculated to an OD_{600} 0.1-0.2 (GENESYS 20 Visible Spectrophotometer) and incubated for 4-6 h while shaking at 130 rpm and 30°C. Cells were pelleted for 3 min (4000 rpm, Eppendorf 5810R) at room temperature (RT), washed with 10 ml (3 min centrifugation, 4000 rpm, RT) and resuspended in 0.5 to 1 ml 1x LiOAc buffer (0.1 M LiOAc, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, RT) depending on the number of transforming plasmids: each transformation requires 100 µL yeast cells meaning that, for example, for 8 transformation reactions the cell pellet was resuspended in 0.9 ml 1x LiOAc (RT).

For each transformation, 100 µl yeast cells, 1-5 µg plasmid DNA, 10 µl carrier DNA (10 mg/ml salmon sperm DNA, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, boiled at 95°C in a Thermomixer every 2-3 freeze and thaw cycles) and 280 µl PEG solution (50 % PEG 3500 (w/v), 1x TE and 100 mM LiOAc were combined, vortexed thoroughly and incubated for 45 min at RT. Forty-

three µl DMSO were added, the mixture was vortexed and incubated for 15 min at 42°C. The heat shock was stopped by incubation on ice for 2 min. Cells were washed three times (3 min centrifugation, 4000 rpm, RT, Eppendorf 5424R) with 500 µL sterile dH_2O. The pellet was resuspended at RT in 200 µl sterile 1x TE, of which 40 and 160 µl were plated on appropriate selection plates and incubated for 2-3 days at 30°C.

3.3.2 Liquid sporulation of strain Y24962 and tetrad dissection

Under nitrogen and carbon deprivation conditions, diploid *S. cerevisiae* cells are able to sporulate: one diploid cell undergoes a special meiosis that results in one tetrad, which harbors four haploid spores[187,188]. These spores can be separated from each other via tetrad dissection and used independently as haploid strains. To generate haploid strains 7-A2, 7-C5 and 7-C4, strain Y24962 (3.1.6) was purchased from EUROSCARF and transformed with pRS416-Abf1. Sporulation was conducted based on a protocol from CSHL, Yeast Genetics and Genomics Course 2017[184]: Two ml YNB w/o ura were inoculated with strain Y24962 and incubated while shaking O/N at 130 rpm and 23°C. Cells were pelleted for 3-5 min at 4000 rpm (Eppendorf 5810R) at RT, washed three times with sterile dH_2O and resuspended in 0.5 ml minimal SPO media (1% potassium acetate, 5 µg/ml histidine, 25 µg/ml leucine). With this resuspension, 3 ml minimal SPO media were inoculated to OD_{600} 0.1 to 0.3, resulting in a visibly cloudy culture. The culture was incubated under shaking at 23°C O/N and 130 rpm, then shifted to 30°C and incubated for 5-7 days. Liquid sporulation cultures were stored at 4°C.

Prior to tetrad dissection, the ascus wall was digested using freshly prepared zymolyase cocktail (82 mM K_2HPO_4, 166 mM KH_2PO_4, 1.2 M sorbitol, 10 µg/µl zymolyase): 200-600 µl sporulation culture were spun down (1 min, 15,000 rpm, Eppendorf 5424R) at RT until a pellet was clearly visible. Supernatant was removed, the pellet resuspended carefully in 100 µl zymolyase cocktail and incubated for 10 min at RT. Ascus wall digestion was stopped by addition of 300 µl 0.1 M KPO_4/Sorbitol solution (82 mM K_2HPO_4, 166 mM KH_2PO_4, 1.2 M sorbitol) and 10-20 µl of digested ascospores were transferred as a stripe onto one side of a YNB w/o ura plate. The tetrad dissection was carried out on a specialized microscope (MSM400, Singer instruments). There, the agar plate is held upside down and a glass needle, attached to a micromanipulator, allows to place the ascospores one by one on defined locations on the agar plate, which are arranged in a grid. Dissection plates were incubated for three days at 30°C. Spores from several tetrads were streaked out as patches on YNB w/o ura plates and

incubated for two days at 30°C. They were re-streaked as patches on YPDAP + kanamycin (0.2 mg/ml) selection plates and cells with genomic *ABF1* (systematic name: YKL112w) replaced by a kanamycin resistance cassette (YKL112w::kanMX4) were identified by growth on these plates, namely strains 7-A2, 7-C5 and 7-D4, which are biological replicates. To validate the (in-)viability of selected random mutagenesis constructs they were transformed into strain Y24962. These diploid strains were sporulated as described above but using YPDAP as pre-sporulation and minimal SPO media lacking leucine as sporulation media. Here, tetrads were dissected on YPDAP + 4% glucose plates.

3.3.3 Colony PCR to determine mating types of strains 7-A2, 7-C5 and 7-D4

The mating type of haploid strains 7-A2, 7-C5 and 7-D4 was identified via colony PCR using primers as published[189]. A small piece of yeast colony was picked-up from the plate with an inoculation loop and resuspended in 10 µl of zymolyase solution (2.5 mg/ml zymolyase 1.2 M sorbitol, 0.1 M $NaPO_4$). The mixture was first incubated for 10 min at 37°C followed by a 5 min incubation at 95°C and next vortexed vigorously for at least 15 seconds. The cell debris was pelleted for 1 min at 15,000 rpm (Eppendorf 5424 R). The supernatant was transferred to a new 1.5 ml Eppendorf tube and either directly used for PCR or stored at -20°C. PCRs were conducted using Taq Polymerase with ThermoPol buffer (NEB), following manufacturer's protocol and recommended thermocycling conditions. Mating type specific PCR products (404 bp for MATα and 544 bp for MATa[189]) were identified via separation in an 1% agarose gel in 1x TAE.

3.3.4 5-FOA plasmid shuffling assay

5-FOA is a classic reagent used to select for *ura3*$^-$ cells[190]. This plasmid shuffling assay takes advantage of 5-FOA toxicity to cells expressing the *URA3* gene. Haploid strains (7-A2, 7-C5, 7-D4) with wildtype *ABF1* on pRS416, a plasmid bearing a *URA3* marker, were transformed with an *abf1*-IDR-mutant construct on pRS315, a plasmid containing a *LEU2* marker. The loss of the *URA3* bearing plasmid pRS416-*ABF1* due to 5-FOA's toxicity results in cell death, since *ABF1* is an essential gene in *S. cerevisiae*. Therefore, the 5-FOA plasmid shuffling assay can be used to determine, if a certain *abf1*-IDR-mutant construct, provided on the *LEU2* plasmid, is functional and hence supports viability. Transformed cells were plated on YNB w/o ura,

w/o leu plates and incubated for 3 days at 30°C. Three to six single colonies were re-streaked for single colonies on YNB w/o ura, w/o leu plates and incubated at 30°C for 2-3 days. Next, single colonies were streaked out as patches on YNB w/o ura, w/o leu plates and incubated for at 2-3 days. Finally, cells from these patches were streaked again as patches on 5-FOA w/o leu plates and incubated at 30°C for 2-3 days. Cells harboring a functional Abf1 version, that is able to compensate for the loss of the wildtype version, grow on 5-FOA plates. These colonies were streaked out on YNB w/o ura plates to verify, that these cells cannot grow without supplemented uracil, i.e., that pRS416-Abf1 indeed got lost and that cell growth on the 5-FOA plates actually arose from the *abf1*-IDR-mutant version.

3.3.5 Spotting assay

To assess the (in-)viability of selected strains in a more nuanced way, a spotting assay was conducted. Twenty ml of appropriate media (YNB w/o leu for strains viable on 5-FOA and after selection on 5-FOA, YNB w/o ura, w/o leu for strains inviable on 5-FOA) were inoculated with cells from a freshly streaked plate and incubated O/N at 30°C under shaking at 130 rpm. Cells equivalent to 1 OD_{600} were spun down and resuspended in 1 ml sterile dH_2O. Ten-fold serial dilutions with sterile dH_2O were prepared in a 96-well micro-titer plate ranging from row 1 to 6 (with row 1 containing undiluted cells with OD_{600} = 1 to row 6 containing only sterile water). After mixing the dilutions thoroughly, 10 µl of each strain and dilution were spotted on 5-FOA plates and incubated for 3 days at 30°C.

3.3.6 Chromatin immunoprecipitation

To ensure that *abf1*-IDR-mutant constructs are expressed, imported into the nucleus and capable to specifically bind DNA at Abf1 sites, chromatin immunoprecipitation (ChIP) was performed and assessed via quantitative RealTime-PCR (qRT-PCR). In brief, 100 OD_{600} of logarithmic yeast cells (Zeiss PMQ II photometer) were crosslinked at RT with 1% formaldehyde (final concentration) under gentle shaking and crosslinking was quenched with 250 mM glycine (final concentration). Cell lysis was conducted with Zirconia beads in a bead beater. Chromatin was sheared through sonication (Bioruptor) at 4.5°C and high intensity, 30 cycles, 30 sec on/off. The anti-Flag antibody M2 (Sigma) was used for immunoprecipitation O/N at 4°C and under gentle rolling. Incubation O/N at 65°C under shaking reversed the

crosslinking and DNA was purified with AMPure XP beads. Primers used for qRT-PCR are listed in Table 4.

Table 4: qPCR primers

Primer	Forward primer (5' --> 3')	Reverse primer (5' --> 3')
Abf1-1-1	CAGGCACGTTCGTGTAAACAT	AACGAAATCTTCGAACGGCTTTT
Abf1-2-1	ACCGGGACTGTAGTTTATGAGG	AGCCTTTCTTGGTGGCGAAA
Abf1-3-1	ATTCACTACGACGCCACGAA	TCTCAACAATGCAACTCGTATCC
PGK1-2	AACGGTCCACCAGGTGTTTT	CAGCAGCAGAGCTCTTGACA
PDH1-1	ATTCCATGCGCAAACTGCTG	TCATTGCCGCCTCTTGAGTT
ADH1-1	GTGCTCACGGTGTCATCAAC	GCATACCGACCAAAACGGTG

The qRT-PCR was performed with LightCycler 480 II (Roche) using 1x Fast SYBR Green Master Mix (Thermo Fisher Scientific). All ChIP assays were done by Andrea Schmid.

3.4 Biochemical methods

If not indicated further, specifications of chemicals, enzymes and kits used here can be found in the methods section of the referenced papers.

3.4.1 Salt gradient dialysis

Chromatin assembly via salt gradient dialysis was performed as described in detail in [191,192]. Histone octamers were purified from *D. melanogaster* embryos as described[191,193] and *S. cerevisiae* plasmid library pGP546 was also prepared as previously published[194,195]. Ten µg *S. cerevisiae* plasmid library pGP546 and 3-8 µg *D. melanogaster* histone octamers, 20 µg BSA and 0.05% IGEPAL CA630 were combined in 2x high salt buffer to 100 µl and transferred to dialysis cups (Slide-A-Lyzer MINI Dialysis Device). The mixture was dialyzed O/N at 30°C starting in 300 ml high salt buffer (10 mM Tris-HCl, pH 7.6, 2 M NaCl, 1 mM EDTA, 0.05% IGEPAL CA630, 14.3 mM β–mercaptoethanol), against 3 L low salt buffer (10 mM Tris-HCl, pH 7.6, 50 mM NaCl, 1 mM EDTA, 0.05% IGEPAL CA630, 1.4 mM β–mercaptoethanol). Next day, the buffer was exchanged with 1 liter of fresh low salt buffer and dialyzed for another hour. Chromatin was transferred to a fresh Eppendorf tube, spun down briefly for 1 min at 10,000 rpm (Eppendorf 5424 R) and DNA concentration measured photometrically (DS-11

Spectrophotometer) against 1x low salt buffer. To assess the successful SGD assembly, an MNase control digest was conducted with 10 µl SGD chromatin as in 3.4.3 but without proteinase K treatment.

3.4.2 Genome-wide *in vitro* remodeling of SGD-chromatin with purified factors

Genome-wide remodeling was performed as published previously[120,195]: SGD-chromatin was incubated with different combinations of purified factors. INO80 complex was kindly provided by the Hopfner group, Gene Center Munich, and used in the remodeling reaction at a concentration of 5-10 nM[195]. Abf1 containing a N-terminal His_6-tag was purified as previously described[120] and usually used at 45 nM. Pho4 was purified as published and used at 200 nM[196]. Pho4ΔAD was used at 100 nM. Dr. Philipp Korber kindly provided purified Pho4ΔAD and Dr. Nils Krietenstein provided purified Abf1, Pho4 and Pho4ΔAD. Remodeling was performed at 30°C for 1-2 h in 100 µl in 26.6 mM HEPES-NaOH pH 7.5, 1 mM Tris-HCl, pH 7.6, 85.5 mM NaCl, 8 mM KCl, 10mM $(NH_4)_2SO_4$, 10 mM creatine phosphate (Sigma-Aldrich), 3 mM $MgCl_2$, 2.5 mM ATP, 0.1 mM EDTA, 0.6 mM EGTA, 1 mM DTT, 14% glycerol, 20 ng/µl creatine kinase. The remodeling reaction was started by the addition of SGD-chromatin corresponding to 0.5-1 µg DNA to the reaction and stopped by a 30 min incubation with 0.2-0.4 U apyrase (NEB).

3.4.3 MNase digestion, library generation and Illumina sequencing

Remodeled samples were treated with MNase to generate predominantly mononucleosomes as previously published[191,195]. In short, 104 µl remodeling reaction (0.5-1 µg DNA as SGD chromatin) were digested with 100 Units MNase (Sigma) for 5 min at 30°C. MNase digestion was stopped by addition of 10 mM EDTA and 0.5% SDS (final concentrations). Proteins were digested by a 30-60 min proteinase K treatment at 37°C. DNA was precipitated with ethanol and separated via gel electrophoresis. In general, gel electrophoresis was performed in 1.5% agarose gels in 1x TAE (ME agarose, Biozym) at 120 V for 1.5 hours. The mononucleosomal band was excised, isolated (PureLink Quick gel extraction kit) and 10-50 ng mononucleosomal DNA were used for preparing the sequencing library. To that end, the mononucleosomal DNA end polishing was achieved by incubation with 1.25 U Taq Polymerase, 3.75 U T4 DNA polymerase, 12.5 U T4 PNK for 15 min at 12°C, 15 min at 37 °C and 20 min at 72°C in a total

volume of 25 µl. Adapters were ligated (0.75 µM NEB Adapters with 800 U T4 Ligase) through a 15 min incubation at 25°C and opened in a subsequent 15 min incubation with 2 U of USER enzyme at 37°C. Adapter ligated mononucleosomal DNA was purified using AMPure XP beads according to the manufacturer's protocol and amplified and indexed via PCR with NEBNext Multiplex Oligos for Illumina using the thermocycling conditions listed in Table 5.

Table 5: Thermocycling conditions to amplify mononucleosomal DNA with multiplex oligos for Illumina sequencing

Temperature	Time	Cycles
98°C	30 sec	1
98°C	10 sec	
65°C	30 sec	8-10x
72°C	30 sec	
72°C	5 min	1
4°C	∞	1

PCR products were separated via gel electrophoresis and the band of mononucleosomal DNA fused with adapters and multiplex oligos was excised, purified and sequenced at the Laboratory of Functional Genome Analysis (LaFuGa, Gene Center, LMU Munich) in a 50 bp single end run on Illumina HiSeq 1500.

3.4.4 Protein extraction from *S. cerevisiae*, SDS-PAGE and Western-blotting

Selected pRS315-*abf1*-NLS-Flag constructs were separated via sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)[197] and transferred onto a nitrocellulose membrane. Five ml of appropriate yeast selection media were inoculated and incubated O/N at 30°C while shaking at 130 rpm. The cells were pelleted (21.000 rpm, RT, 2 min, Eppendorf 5424 R) and the supernatant was discarded. The cell pellet was washed with 1 ml dH_2O (21,000 rpm, RT, 1 min, Eppendorf 5424 R) and resuspended in 500 µl breaking solution (0.2 M NaOH, 0.2% β-mercaptoethanol). After incubating on ice for 10 min, 25 µl trichloroacetic acid were added to the samples and mixed, followed by another 10 min incubation on ice. Samples were centrifuged (21,000 rpm, 4°C, 5 min, Eppendorf 5424 R), resuspended in 100 µl SDS-loading buffer (50 mM Tris-HCl, pH 6.8, 8 M urea, 0.1 M DTT, 2% SDS, 6% glycerol, 0.004% bromphenol blue) and incubated at 95°C for 5 min (Thermomixer C, Eppendorf). Ten µl were subjected to SDS-PAGE using pre-cast 4-12%

gradient gels (SERVA). SDS-PAGE was run at 200 V for 1,5 h in 1x TG buffer. Proteins were transferred onto a nitrocellulose membrane via a wet blotting procedure for 70 minutes on ice at 115 V in transfer buffer (25 mM Tris-HCl, 192 mM glycine, in 20% MeOH (v/v)). The membrane was washed with dH_2O for 5 min at RT and blocked for 40 min at RT in blocking solution (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% (w/v) milk powder). The respective primary antibody (1:5000 in blocking solution) was incubated O/N at 4°C. The membrane was washed three times for 5 min at RT with 1x TBS (50 mM Tris-HCl, pH 7.5, 150 mM NaCl) and incubated with the secondary antibody (1:5000 in blocking solution) for 1 hour at RT. After three washing steps (5 min at RT with 1x TBS) the membrane was dried and expression of Flag-tag bearing pRS315-*abf1* constructs was visualized (LI-COR Biosciences).

4 RESULTS

4.1 Deciphering IDR features essential for *S. cerevisiae* Abf1 function

This part of the thesis is the result of a collaboration with Prof. Dr. Alex Holehouse and his group and it is available as non-peer reviewed preprint on BioRxiv[198]. All biophysical and bioinformatic analyses that were applied to generate the following results were conducted by him or members of his group and figures generated by them are used with permission as indicated in the figure legends.

4.1.1 The 5-FOA plasmid shuffling assay determines the viability of an *abf1*-IDR-mutant

Abf1 is an essential protein in *S. cerevisiae* consisting of a bipartite, sequence-specific DNA binding domain and two long, acidic, intrinsically disordered regions[125,140]. To ascertain the functional determinants of Abf1's IDRs, we employed the approach of a classical plasmid shuffling assay (Figure 10).

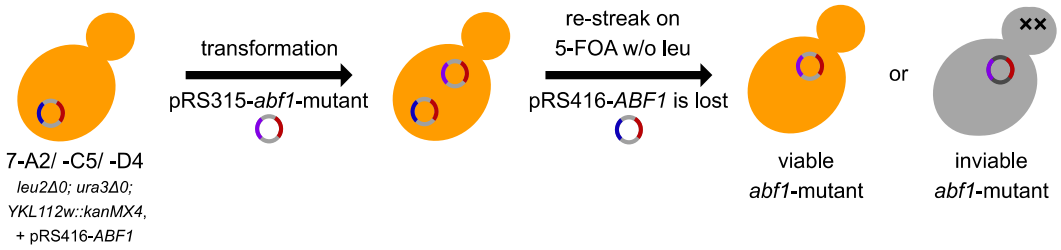

Figure 10: Schematic overview of the 5-FOA plasmid shuffling assay. Three independent biological replicates (7-A2, 7-C5 and 7-D4) of the haploid yeast strain, where *ABF1* was deleted in the chromosome, were used for the viability assay. The only wildtype *ABF1* version (systematic *ABF1* name: YKL112w) in the cell is provided on a plasmid with a *URA3* marker (pRS416-*ABF1*). The *LEU2* marker plasmid pRS315 bearing the *abf1*-mutation, is transformed into a biological replicate of the haploid strain and transformants re-streaked on 5-FOA w/o leu plates. Only cells that lost the plasmid pRS416-*ABF1* are retained in this step, as 5-FOA is toxic to cells expressing *URA3*. The protein encoded by the *abf1*-IDR-mutant gene is the only Abf1 version remaining in the cell. This allows to assess, if the respective mutant is able to provide the essential Abf1 function.

Starting from the diploid *S. cerevisiae* strain BY4743 with heterozygous deletion of the *ABF1* gene (MATa/MATα; *his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; LYS2/lys2Δ0; met15Δ0/MET15; ura3Δ0/ura3Δ0; YKL112w/YKL112w::kanMX4)*, we generated three independent haploid strains (biological replicates 7-A2, -C5 and -D4) via tetrad dissection. In these strains, the wildtype *ABF1* gene, with the systematic name YKL112w, was deleted in the chromosome but instead provided on a *URA3* marker plasmid pRS416-*ABF1*. The *LEU2* marker plasmid pRS315 with the *abf1*-

IDR-mutant gene was transformed into this strain. Plasmid pRS416-*ABF1* was lost through counterselection on 5-FOA, taking advantage of the agent's toxicity to cells expressing *URA3*: 5-FOA is converted to 5-F-dUMP, which irreversibly inhibits thymidylate synthase and eventually blocks DNA synthesis due to lack of dTTP[199,200]. The only Abf1 protein version remaining in the cell was the one encoded by the *abf1*-IDR-mutant gene on the *LEU2*-marker plasmid pRS315- *abf1*-IDR-mutant. Therefore, only the mutants still able to fulfill the essential Abf1 function for viability resulted in living *S. cerevisiae* growing as patches on 5-FOA w/o leu plates. Strains were scored as viable, if they grew as dense patches while strains, that exhibited a smeary appearance, which mainly reflected the original streak out, or only sparse individual colonies were scored as inviable (Figure 11).

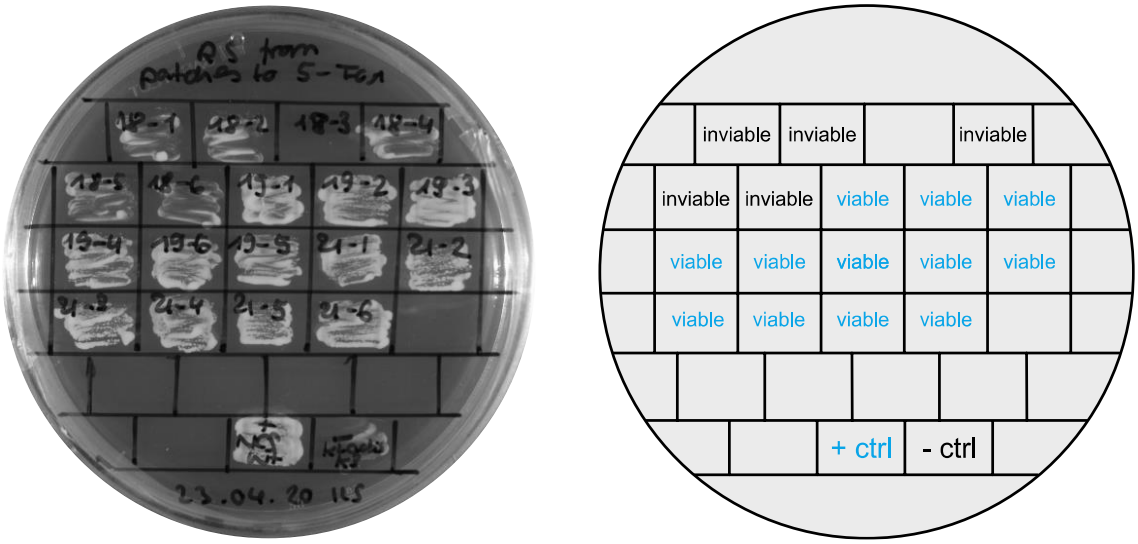

Figure 11: Example of a 5-FOA w/o leu plate used to determine the viability of *abf1*-IDR-mutants. On the left, a 5-FOA w/o leu plate is shown, on which strains transformed with different pRS315-*abf1*-IDR-mutant plasmids were re-streaked as patches from YNB w/o ura, w/o leu plates. Viable strains grew as more or less dense patches while inviable strains looked like a smear. In the schematic on the right, strains of this plate are labeled viable (blue) and inviable (black). Positive (+ ctrl) and negative (- ctrl) controls are indicated. As a positive control the following strain was used here: plasmid pRS315-*abf1*-ΔIDR1-IDR2$^{449\text{-}662}$ was transformed in one of the three haploid strains (7-D4, -C5- or -A2). As a negative control here served the equivalent strain, in which the plasmid that contained the corresponding partial *K. lactis* Abf1-IDR2 (pRS315-*abf1*-ΔIDR1-*K. lactis*-IDR2), which was shown before by us to be inviable, was transformed.

4.1.2 The intrinsically disordered region 2 of Abf1 is essential for cell viability

S. cerevisiae Abf1 has a very distinct domain structure: it holds a bipartite sequence-specific DBD and two IDRs, IDR1 and IDR2 (Figure 12). To decipher if the latter are essential for the protein's essential function, we investigated the viability of different truncation constructs affecting IDR1, IDR2 and combinations of both. The DBD-only construct without any IDR did

not support viability. IDR1 could be deleted without influencing cell viability while deletion of IDR2 caused inviability, thus IDR2 is the IDR essential for the essential Abf1 function. Several truncation constructs of IDR2 were generated based on the published observations that two critical regions, C-terminal sequence 1 (CS1, 624-628), which harbors the endogenous Abf1 NLS, and CS2 (639-662), but not residues 662-731, were required for cell growth[142,143]. The last 100 C-terminal residues of Abf1 are partially interchangeable with the C-terminus (582-827) of general regulatory factor (GRF) Rap1[164]. In line with and even going beyond these observations, we showed that a C-terminal truncation of IDR2 up to residue 663 supported viability, even in the absence of IDR1. Even more, further C-terminal truncations of IDR2 residues up to 624, thereby eliminating CS1 and CS2 were still sufficient for viability in the employed 5-FOA plasmid shuffling assay if a heterologous NLS was included. In contrast, CS1/CS2 (624-662) alone without IDR1 did not support viability. Nonetheless, the construct, in which just IDR1 and CS1/2 both are present, provided viability.

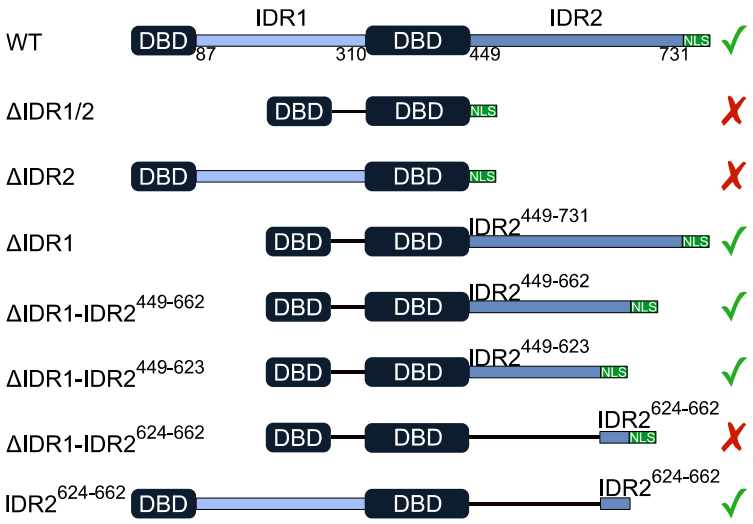

Figure 12: Domain structure of *S. cerevisiae's* WT Abf1 and truncation constructs. DBDs are marked in dark blue, IDR1 and IDR2 in light blue color. The C-Terminal SV40 NLS is depicted in green. The linker regions (three glycine serine repeats $(GlySer)_3$), shown as a black line serve as flexible connections within the DBD or between DBD and a truncated IDR. The viability of the respective construct is depicted by a green check mark, inviability is shown with a red cross.

Most constructs investigated in this work hold a C-terminal SV40 NLS to ensure that expressed mutant proteins were imported into the nucleus also in the absence of the endogenous Abf1 NLS (CS1). In this way, the inviability of constructs due to failed nuclear import of the proteins was prevented. In previously published studies the NLS was only added to constructs to specifically verify that CS1 was indeed the endogenous Abf1 NLS[142]. Moreover, we added a C-terminal triple Flag-tag to almost all constructs to enable immunoprecipitation via an anti-

Flag antibody: ChIP coupled to qRT-PCR was performed for all inviable constructs to control that the proteins are expressed, imported into the nucleus and sequence-specifically bound to Abf1 binding sites. Some of the truncation constructs (Table 2 and Table 3) and selected random mutagenesis constructs (4.1.4) also have a C-terminal triple Flag-Tag.

Taken together, Abf1 essentially requires a certain length and kind of IDR for its essential function and this requires from the wild type Abf1 at least a part of IDR2. The maximally truncated IDR2 construct that was still viable in absence of IDR1 and presence of a heterologous NLS, comprised IDR2$^{449\text{-}623}$. We chose this disordered region plus residues of CS1/2 as a model IDR for the following work, namely construct pRS315-*abf1*-ΔIDR1-IDR2$^{449\text{-}662}$ -NLS-Flag.

4.1.3 Compositional conservation is not sufficient for the essential Abf1 function

We analyzed all *S. cerevisiae* IDRs to evaluate the compositional conservation among their orthologs across 19 yeast species, in other words, how conserved charged, hydrophobic and polar residues are in these IDRs. This analysis revealed that IDRs can be quite conserved in terms of composition despite poor conservation of their linear sequence. In IDR2$^{449\text{-}662}$ of Abf1, the composition of hydrophobic and charged residues was similarly or more conserved as for the majority of *S. cerevisiae* IDRs, but the composition of polar residues was less conserved (Figure 13). Positively charged amino acids and hydrophobic residues in IDR2$^{449\text{-}662}$ were more conserved than in 93% of all IDRs, negatively charged amino acids more than 64%. Polar residues were more conserved in IDR2$^{449\text{-}662}$ than in 48% of all IDRs.

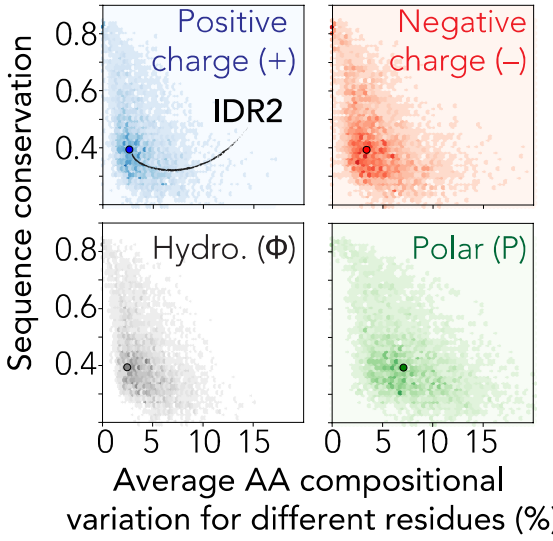

Figure 13: Conservation of sequence and composition for all *S. cerevisiae* IDRs. Average sequence conservation assessed by linear alignment in comparison to per-residue compositional conservation. Composition is on average more conserved than the linear sequence among *S. cerevisiae* IDRs. Position of Abf1's IDR2$^{449\text{-}662}$ is marked by an outlined dot. Charged and hydrophobic residues are compositionally similarly or more conserved in Abf1's IDR2$^{449\text{-}662}$ as for the majority of *S. cerevisiae* IDRs, but composition of polar residues is less conserved. For details of this analysis see[198]. *Used with permission from Dr. Alex Holehouse.*

Based on the observation that IDR2$^{449\text{-}662}$ is compositionally conserved, we investigated if compositionally similar IDRs from Abf1 orthologs were able to preserve viability. To this end, we substituted IDR2$^{449\text{-}662}$ in construct pRS315-*abf1*-ΔIDR1-IDR2$^{449\text{-}662}$-NLS-Flag with sequences corresponding to Abf1's IDR2$^{449\text{-}662}$ from 18 Abf1 orthologs (Figure 14) and assessed their viability.

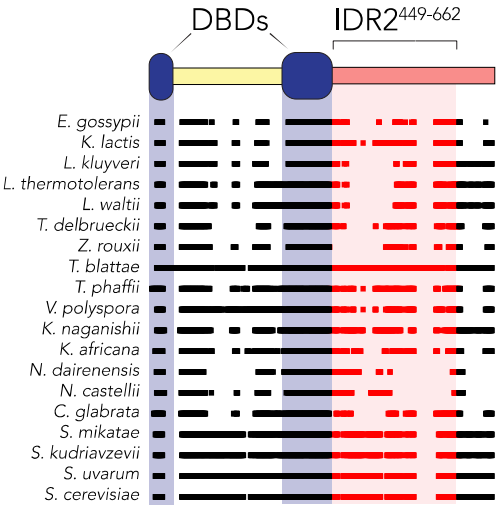

Figure 14: Schematics of the domain structure of *S. cerevisiae* Abf1 and the sequence alignments of orthologs. DBDs are highlighted in blue, subregions corresponding to Abf1's IDR2$^{449\text{-}662}$ in red. *Used with permission from Dr. Alex Holehouse.*

Only five out of 18 tested orthologous sequences supported viability, three of them belonging to the closely related *Saccharomyces sensu stricto* strains (Figure 15).

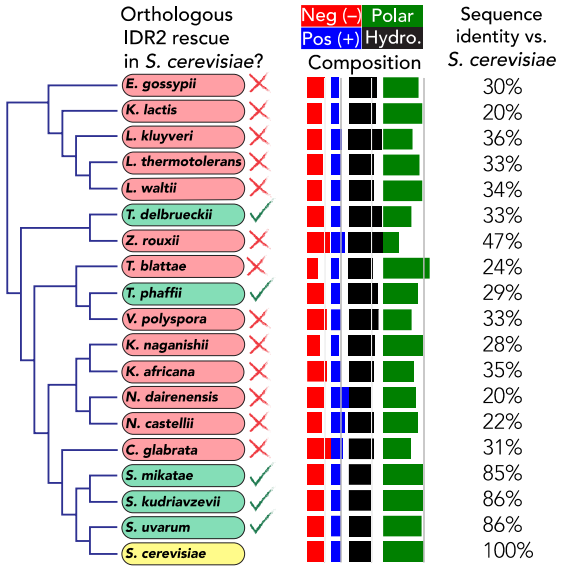

Figure 15: Phylogenetic tree, sequence composition and identity of *S. cerevisiae* Abf1 IDR2$^{449\text{-}662}$ and orthologs. Orthologs that are able to substitute Abf1's IDR2$^{449\text{-}662}$ in the plasmid shuffling assay are marked in green and indicated with a green checkmark. Orthologous sequences that cannot substitute and result in inviable strains are shown in red and indicated with a red cross. *Used with permission from Dr. Alex Holehouse.*

The Abf1 IDR2 orthologs differ in sequence identity to *S. cerevisiae* Abf1 from 20-86% and show different lengths, but are compositionally similar. Nonetheless, composition was not indicative here to distinguish viable from inviable sequences, therefore, amino acid composition alone, at least not as provided by these evolved sequences, is not sufficient to maintain the essential Abf1 function.

4.1.4 IDR2 is both robust and sensitive to randomly introduced mutations

We altered the linear sequence of IDR2$^{449\text{-}662}$ through random mutagenesis of this region in construct pRS315-*abf1*-ΔIDR1-IDR2$^{449–662}$-NLS by error-prone PCR. Random mutagenesis allowed here to test for different IDR features simultaneously in an unbiased way and features that are important for the protein's essential function might be revealed through such mutagenesis constructs. The aim was to examine if viable and inviable ones could be distinguished by certain parameters such as disorder, NCPR, hydrophobicity or amino acid composition, meaning the frequency of each aa in the sequence. To this end, we investigated 48 mutagenized sequences, of which 11 were viable and 37 inviable (Figure 9A).

Computational analyses predicted all of them to remain disordered but no parameter qualified to distinguish viable from inviable sequences (data not shown). Moreover, the results regarding viability obtained through the 5-FOA plasmid shuffling assay were verified by tetrad dissection for all inviable and selected viable constructs (data not shown). Nonetheless, spotting assays on 5-FOA plates done later by Andrea Schmid, showed that one of the constructs, NCS506, that was initially scored as inviable, did support growth albeit very slowly (data not shown). Intuitively, one would expect sequences with a lower mutation rate to be more likely to support viability than those with a higher mutational burden. This has not proven to be correct for IDR2$^{449\text{-}662}$: one viable sequence held 55 mutations in the amino acid sequence while another inviable one exhibited only 21. There was no significant link between mutational burden and viability (Figure 16B). Therefore, IDR2 is both robust and sensitive to the introduction of point mutations.

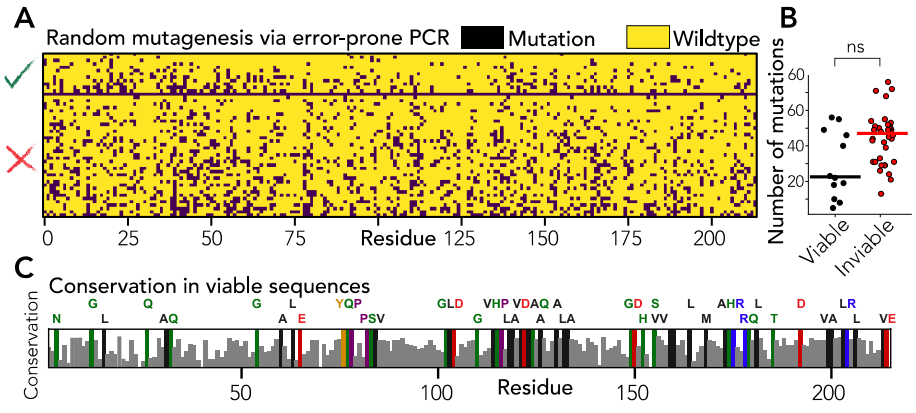

Figure 16: Impact of random mutagenesis on viability of Abf1 IDR2$^{449\text{-}662}$. A) Randomly induced mutations in IDR2$^{449\text{-}662}$. Viable sequences are shown on top indicated with a green check-mark, inviable sequences are on the bottom and indicated with a red cross. Each line represents one mutagenized sequence and each column displays one amino acid residue. Shown in yellow are residues consistent with the WT sequence, mutated amino acids are indicated by black rectangles. **B)** Comparison of extent of mutagenesis between viable (black) and inviable (red) sequences. ns: not significant **C)** Degree of conservation of residues in viable IDR2$^{449\text{-}662}$ sequences. Residues that were never mutated in viable sequences are highlighted above the plot. Shown in black are aliphatic hydrophobic residues. Acidic aa are depicted in blue, basic aa in red. Polar residues are shown in green, aromatic residues in orange and prolines are colored magenta. *Used with permission from Dr. Alex Holehouse.*

IDR2$^{449\text{-}662}$ was also randomly mutagenized to ascertain, if the essential Abf1 function relies on the presence of SLiMs, 5-15 residues comprising versatile binding modules for low-affinity protein-interactions in IDRs[36]. If SLiM(s) are required for the essential function, they might be revealed through disruption by mutagenesis as their functionality relies on their linear sequence[36]. Several positions were conserved in viable sequences (Figure 16 C) although there were not enough conserved residues in one consecutive stretch to identify a SLiM and the

conservation pattern not significantly different from a random pattern (statistical analysis done by Dr. Alex Holehouse, not shown).

Random mutagenesis alters the linear sequence which then results in changes in composition of charged, hydrophobic and polar residues. We treated all viable and inviable sequences, by themselves, as a mini group of "orthologs": the linear sequence conservation was plotted versus the compositional variation within both groups as it was done for all other yeast IDRs (Figure 17). This showed that not only the viable but also the inviable random mutagenesis sequences showed less prominent differences in sequence conservation and composition variation than the IDR2$^{449\text{-}662}$ orthologs. This explains the result gained through the investigation of compositionally similar Abf1 IDR2$^{449\text{-}662}$ orthologs: the degree of compositional conservation among the orthologs was lower than for the inviable mutagenesis sequences and therefore was unlikely to maintain Abf1 function. Even a rather high degree of compositional conservation as seen for the inviable random mutagenesis sequences was not sufficient for a viable Abf1 IDR2$^{449\text{-}662}$-variant.

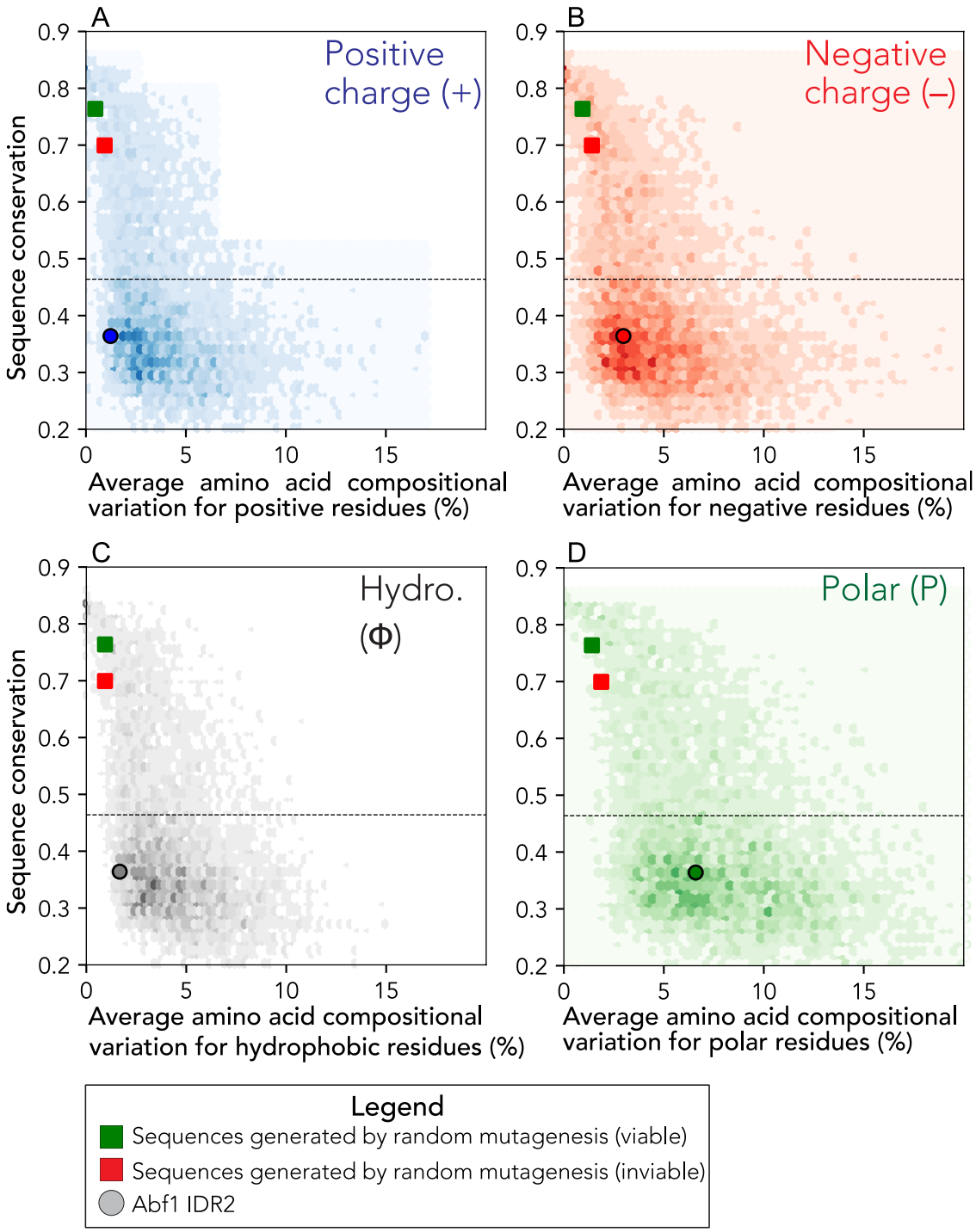

Figure 17: Conservation of sequence and composition of viable and inviable sequences generated via random mutagenesis and across *S. cerevisiae* IDRs. Average sequence conservation assessed by linear alignment in comparison to per-residue compositional conservation. As Figure 13, but sequences generated by random mutagenesis were included as if they represented "evolutionary orthologs" of Abf1 IDR2$^{449\text{-}662}$. Composition variation of viable and inviable random mutagenesis sequences differed less among them than the composition variation of WT Abf1 IDR2$^{449\text{-}662}$ among its yeast orthologs. *Adapted with permission from Dr. Alex Holehouse.*

4.1.5 IDRs of functionally related proteins can only in some cases substitute IDR2 function

We investigated if IDRs of proteins that are functionally similar or related to Abf1 can confer viability: Abf1's IDR1, IDRs of GRFs (Rap1, Mcm1, Reb1), yeast transactivators (Gcn4, Gal4, Pho4) and human insulator CTCF (Figure 18 A and B).

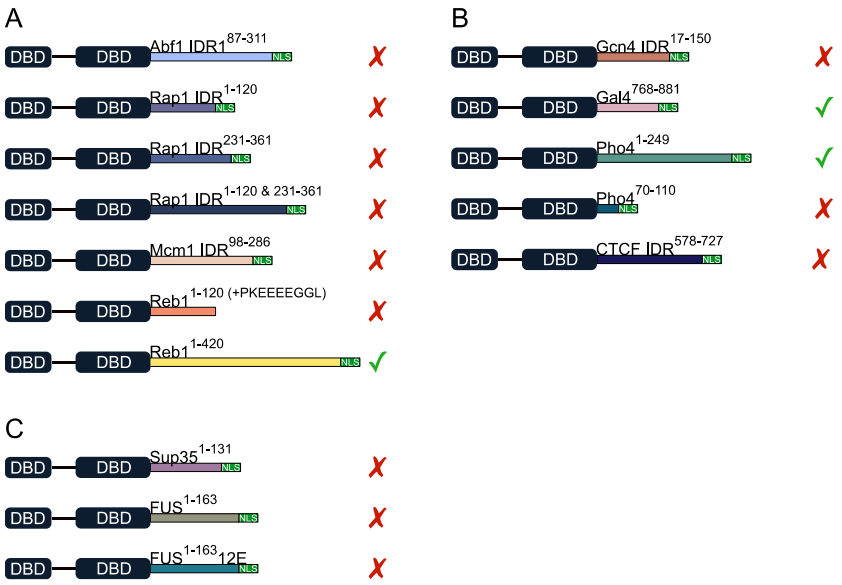

Figure 18: IDRs of functionally related and unrelated proteins mostly could not substitute Abf1's IDR2. IDR2$^{449-662}$ was replaced with **A)** Abf1-IDR1 and other GRF-IDRs, **B)** IDRs of yeast transactivators Gcn4, Gal4, Pho4 and human insulator CTCF, and **C)** yeast translation termination factor Sup35, low-complexity domain of human RNA-binding protein FUS as well as its phosphomimetic version FUS$^{1-163}$12E. Substitutions that resulted in viable strains in the 5-FOA plasmid shuffling assay are indicated with a green checkmark, constructs that resulted in inviable strains are marked with a red cross.

IDR1 of Abf1 was not sufficient to confer viability when located at its endogenous position between the bipartite DBD (Figure 5). As substitute for IDR2$^{449-662}$, IDR1 also resulted in an inviable construct which indicated that the IDR position relative to the DBDs did not matter here. Mcm1's only IDR as well as three constructs of Rap1's IDRs' residues, 1-120, 231-361 and a combination of both, resulted in inviability. For Reb1 a short (1-120) and a long IDR version (1-420) were tested of which only the latter resulted in a viable construct, so did Gal4's extended transactivation domain and all of Pho4 without just its DBD (1-249). In contrast, neither Gcn4's nor Pho4's transactivation domain (Pho4$^{70-110}$) or CTCF's IDR conferred viability. IDR-length was not a critical parameter also here, as the 250 residues comprising combined Rap1 IDR was inviable and Gal4's 140 residue extended transactivation domain was viable.

Furthermore, we inspected IDRs of functionally unrelated proteins like yeast translation termination factor Sup35 (residues 1-131) or the low-complexity domain of human RNA-binding protein FUS (residues 1-163) as well as its phosphomimetic version (FUS^{1-163}12E) (Figure 18 C). The latter was chosen, as IDRs, and especially the FUS IDR, are regulated by posttranslational modifications (PTMs) like phosphorylation[96,201]. The phosphomimic of FUS with the additional 12 negative charges in the strongly hydrophobic FUS background may have mimicked somewhat the compositional properties of $IDR2^{449-662}$. Nonetheless, neither of these three were able to confer viability as a replacement of Abf1's $IDR2^{449-662}$.

Taken together, these results suggest that the essential function of $IDR2^{449–662}$ requires very specific molecular recognition pattern not solely based on IDR length, functional relationship or compositional similarity, as it cannot easily be mimicked by IDRs similar in these features to Abf1's $IDR2^{449-662}$.

4.1.6 Global and local sequence shuffles reveal an essential motif in $IDR2^{449-662}$

As there was so far no clear compositional signature and as the results from the random mutagenesis regarding linear sequence motifs were not conclusive, we continued to ask, if there was a linear sequence necessary for the function of IDR2. Now we employed global and local sequence shuffles. In both cases, the amino acid composition remained unaltered, but the linear sequence of either the entire $IDR2^{449-662}$ (global shuffle) or in short, non-overlapping windows (local shuffle) was randomly shuffled (Figure 19 A). We assumed that destroying putative SLiMs through the disruption of their linear sequence through sequence shuffles provided a more systematic way to uncover these motifs than their potential disruption by random mutagenesis.

We tested three different globally shuffled sequences (Figure 19 B). In these shuffles, the positions conserved in viable sequences obtained through random mutagenesis, were not changed. All three global shuffles were inviable and hence, those 15 conserved positions as well as composition alone were not sufficient to provide the essential Abf1 function. Instead, this suggested that a SLiM in $IDR2^{449-662}$ was required for function.

To narrow down this SLiM, we applied a new approach of sequential local sequence shuffles: First, the sequence of $IDR2^{449-662}$ was shuffled (Figure 19 C) in 30-residue windows (constructs LS 1-7). This identified a 60-residue window (shuffled in LS 4 and 5) that could not be shuffled without losing viability, while keeping this window intact we could shuffle all the rest (LS 8).

We then further divided and shuffled this window into smaller 10-residue sub-windows (LS 9-14). The inviability of LS 11 and 12 pinpointed the SLiM location. We kept the linear sequence of these 20 residues unaltered in LS15, and shuffled the entire remaining sequence instead. Viability of LS15 (and with lower resolution also of LS 8) confirmed the location of the SLiM and excluded that another SLiM was necessary in addition. We named this 20-residue window the "essential motif" (EM).

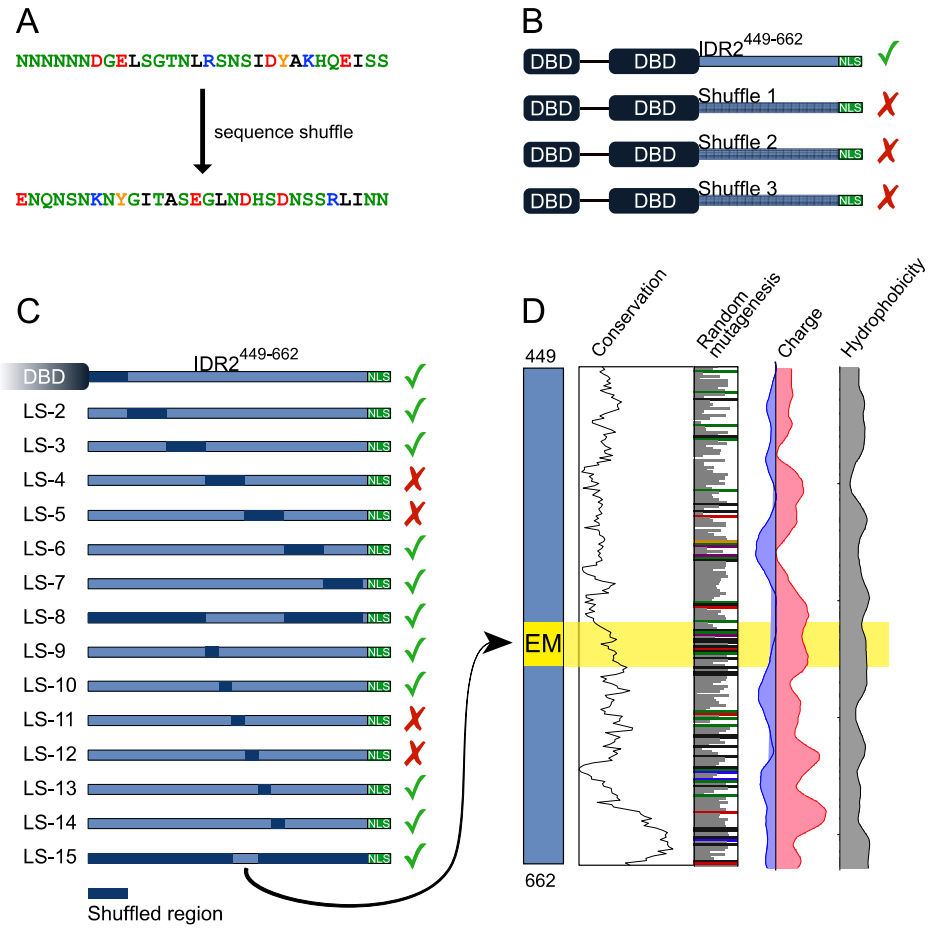

Figure 19: Sequence shuffles reveal an essential motif in Abf1 IDR2$^{449-662}$. A) Illustration of an exemplary sequence shuffle. **B)** Conserved positions from random mutagenesis were held fixed in global shuffles. **C)** Sequential local sequence shuffles along IDR2$^{449-662}$ revealed one essential motif (EM) via inviability of constructs where the EM region was shuffled. (In)viability is indicated by red crosses and green check marks, respectively. **D)** EM features within IDR2$^{449-662}$ with respect to conservation between orthologs, conservation in random mutagenesis, charge (positive blue, negative red) and hydrophobicity. *Adapted with permission from Dr. Alex Holehouse.*

This EM is neither especially conserved between orthologs investigated in 4.1.3, nor does it exhibit very pronounced characteristics with respect to charge or hydrophobicity. However, the EM overlaps with a region that was more conserved in viable sequences of random mutagenesis constructs (Figure 19 D). AlphaFold2[138,139] predicted this motif to partially form a

hydrophobic-faced transient helix (Figure 20), a feature employed in IDR-mediated interactions[202,203].

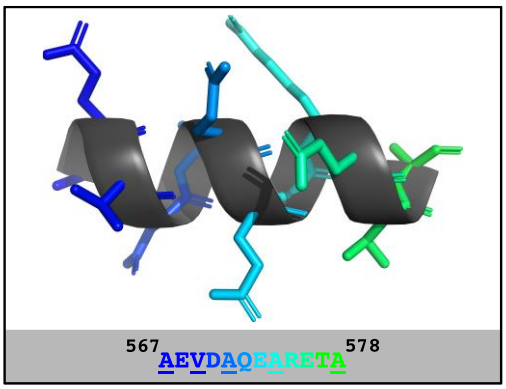

Figure 20: Three-dimensional structure prediction of the transient helix formed by the EM in Abf1 IDR2. Structure (UniProt ID: P141641) as predicted by AlphaFold2[138,139] and visualized with PyMOL. Protein backbone is shown as a cartoon structure and colored in dark grey, amino acid side chains are shown as sticks in shades of blue and green. The corresponding linear sequence of the essential motif, which is predicted to form a transient helix with a hydrophobic face, is shown below and colored accordingly. Hydrophobic residues are underlined.

Taken together, Abf1 IDR2 contains a 20 aa SLiM, the EM, that is predicted to form a transient helix, cannot be shuffled and is essential for the essential Abf1 function in the context of Abf1 IDR2$^{449\text{-}662}$.

4.1.7 The essential motif of Abf1 and its derivates confer viability to inviable IDRs

A characteristic of SLiMs is their modularity: within an IDR context they can function as low-affinity molecular recognition modules for binding and interaction partners[36,48–51]. Therefore, if Abf1's essential motif acts as a modular SLiM, it should be able to confer function as a module independent from its surrounding IDR context. "Context" describes the IDR without a SLiM. To ascertain if the EM fulfilled this modularity criterion, we placed it into the FUS$^{1\text{-}163}$12E context, which (lacking the EM) was inviable if fused to Abf1's DBDs (Figure 18 C). The phosphomimetic FUS$^{1\text{-}163}$12E low-complexity domain was especially suited, as it is a functionally non-related IDR, did not provide viability, is not predicted to form secondary structures[138], has similar sequence properties as Abf1's IDR2$^{449\text{-}662}$ and holds evenly spaced hydrophobic and acidic residues. The insertion of the EM into FUS$^{1\text{-}163}$12E context resulted in a viable construct (Figure 21). This shows that Abf1's EM acts like a module and can provide viability to an otherwise inviable IDR.

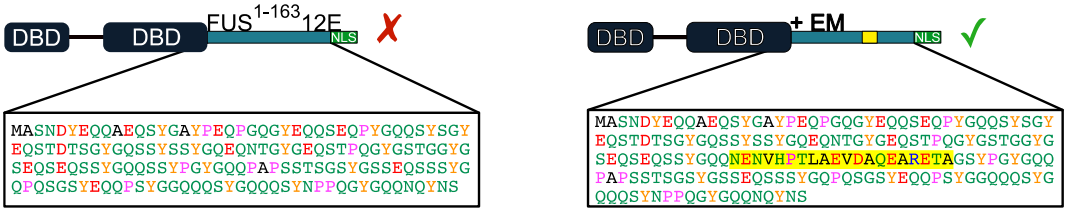

Figure 21: The essential motif of Abf1 is a modular SLiM that provides viability to an inviable IDR. Schematics and amino acid sequences of FUS$^{1-163}$ phosphomimetic low-complexity domain without (left) and with (right) insertion of Abf1's EM (highlighted in yellow). Viability is indicated with a green checkmark, inviability with a red cross. Amino acids are colored as follows: aliphatic, hydrophobic residues in black, acidic in red, glycine and polar residues in green, tyrosine in orange and prolines in magenta.

We assessed, if another IDR, that could substitute Abf1 IDR2 function (Figure 18 A and B), also contained a modular SLiM that we could find by homology search relative to Abf1 IDR2. Global sequence alignment between Abf1 and transactivator Gal4 using EMBOSS Needle[204] showed overall quite poor homologies but revealed one slightly homologous subregion, which we named "Abf1$^{G4\text{-like}}$" in Abf1 and subregion "Gal4^{G4}" in Gal4 (Figure 22). Gal4^{G4} comprises 17 residues. Abf1$^{G4\text{-like}}$ is 19 residues long, two of which overlap with the C-terminal part of Abf1's essential motif.

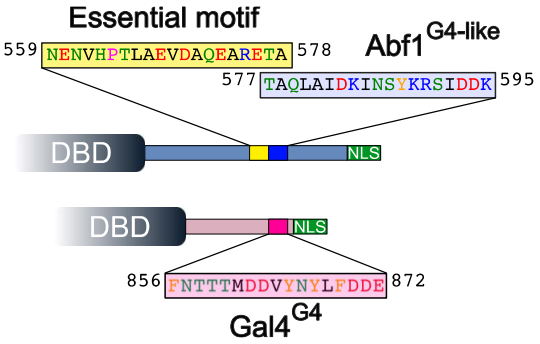

Figure 22: Schematic representation of Abf1's essential motif, and the Abf1$^{G4\text{-like}}$ and Gal4^{G4} low homology subregions. Abf1's IDR2$^{449-662}$ is depicted in light blue, Gal4's disordered transactivation domain is shown in light pink. Sequences showing slight homology between both IDRs are highlighted in blue (Abf1$^{G4\text{-like}}$) and pink (Gal4^{G4}). The essential motif of Abf1 is highlighted in yellow. The sequences were aligned using EMBOSS Needle[204] and alignments were performed by Dr. Alex Holehouse.

Interestingly, AlphaFold2 also predicted all residues of subregion Abf1$^{G4\text{-like}}$ and residues ^{861}MDDVYNYL^{868} of Gal4^{G4} to form a transient helix (Figure 23).

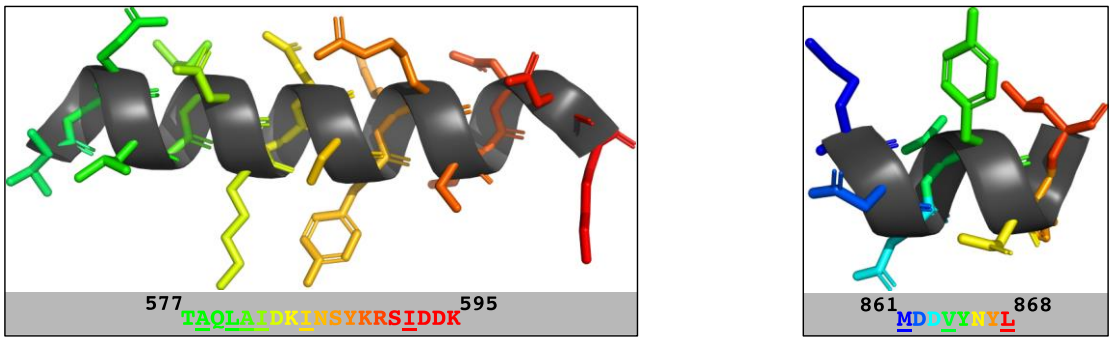

Figure 23: Three-dimensional structure prediction of the transient helices formed by the Abf1$^{G4\text{-}like}$ subregion in Abf1 IDR2 and a part of the subregion Gal4^{G4} in Gal4. Abf1$^{G4\text{-}like}$ is shown on the left, Gal4^{G4} is shown on the right. Structures (UniProt ID Abf1: P141641, UniProt ID Gal4: P04386) as predicted by AlphaFold2[138,139] and visualized with PyMOL. Protein backbones are shown as a cartoon structure and colored in dark grey, amino acid side chains are shown as sticks in rainbow colors. The corresponding linear sequences of the subregions, which are predicted to form a transient helix with a hydrophobic face, are shown below and colored accordingly. Hydrophobic residues are underlined.

We inserted the two subregions Abf1^{G4}-like and Gal4^{G4} into the IDRs of FUS$^{1\text{-}163}$12E and Sup35$^{1\text{-}131}$. Both subregions provided viability to the otherwise inviable low-complexity context. Moreover, non-essential Abf1 IDR1 and Rap1 IDR 231-361 became viable through the insertion of the Gal4^{G4} subregion (Figure 24).

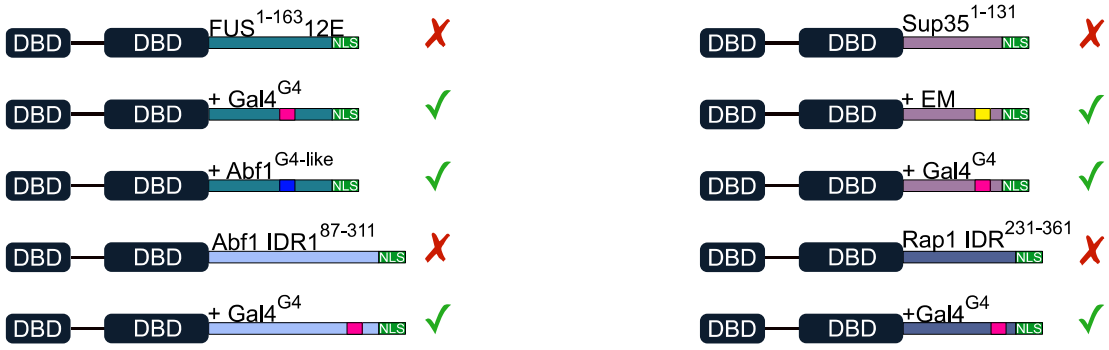

Figure 24: Modularity of subregions provides viability to inviable IDR context and may indicate putative SLiMs. The Gal4^{G4} and/or Abf1$^{G4\text{-}like}$ subregions provided viability to FUS$^{1\text{-}163}$12E, Sup35$^{1\text{-}131}$, Abf1's IDR1 and Rap1 IDR$^{231\text{-}361}$. The EM of Abf1 is also able to provide viability to the Sup35$^{1\text{-}131}$ IDR.

To specifically ask if another (transient) hydrophobic-faced helix could provide viability to the FUS$^{1\text{-}163}$12E context, we chose a well-described 24-residues (residues 321-345) transient helix-forming region from human RNA-binding protein TDP-43, a protein associated with ALS and Alzheimer's disease[115,205,206](Figure 25).

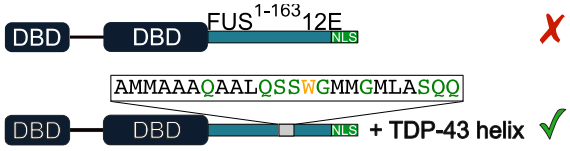

Figure 25: TDP-43 helix provides viability to an otherwise inviable FUS$^{1-163}$12E context. Schematics of insertion of the TDP-43 region into the FUS$^{1-163}$12E context with symbols and color coding of the amino acids as in Figure 21.

The hydrophobic TDP-43 helix provided viability to the otherwise inviable FUS$^{1-163}$12E IDR. In summary: the essential motif of Abf1-IDR2 is a modular SLiM that depends on its linear sequence and hence cannot be shuffled. The EM conferred viability to several inviable IDRs. Two slightly homologous regions in Abf1 and Gal4 both also independently provided viability, so did the hydrophobic-faced helix of TDP-43.

4.1.8 Motif distribution is required to validate *bona fide* SLiMs

We wished to validate if subregions Abf1$^{G4–like}$ and Gal4^{G4} were genuine SLiMs. As negative control, we selected sequences from transactivation domains of five TFs with similar length and composition as Gal4^{G4} and hypothesized that they could not provide viability and would therefore result in inviable constructs (Figure 26).

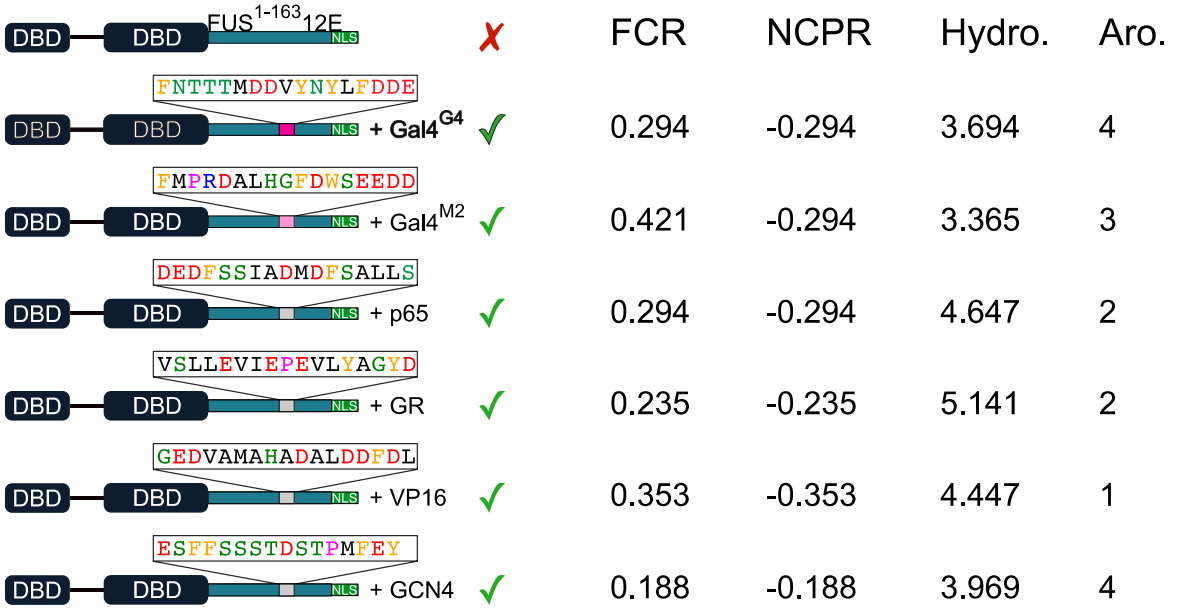

Figure 26: Compositionally matched subsequences of transactivation domains from various TFs all provide viability to FUS $^{1-163}$12E context. Subsequences and their respective biophysical parameters: Fraction of charged residues (FCR), net charge per residue (NCPR), hydrophobicity (Hydro.) and number of aromatic residues present in the subsequence (Aro.). Symbols and amino acid coloring as in Figure 21.

These sequences were taken from TFs from different species/virus: *S. cerevisiae* ($Gal4^{G4}$, $Gal4^{M2}$, Gcn4), *H. sapiens* (p65, glucocorticoid receptor (GR)) and Human herpes simplex virus 2 (VP16). Surprisingly, all five subsequences were able to provide viability to the FUS^{1-163}12E context.

This result made us re-assess our understanding of SLiMs that are required for the essential function in Abf1's $IDR2^{449-662}$. Given the current definition of a SLiM, it is characterized by its modularity and linear sequence. In other words: if the modular SLiM is inserted into an otherwise non-viable IDR context, it provides viability to this context and shuffling the sequence turns the viable IDR context back into an inviable one. This approach revealed that the essential motif of Abf1's $IDR2^{449-662}$ is a *bona fide* SLiM. We showed that the $Gal4^{G4}$ subsequence also fulfills the modularity criteria. To ascertain, if the function conferred by this subsequence relies on its linear sequence or just on its composition, we inserted a shuffled $Gal4^{G4}$ subsequence into the FUS^{1-163}12E context. Moreover, we went one step further and first shuffled, then distributed the amino acids of the subsequence randomly across the entire FUS^{1-163}12E IDR (Figure 27).

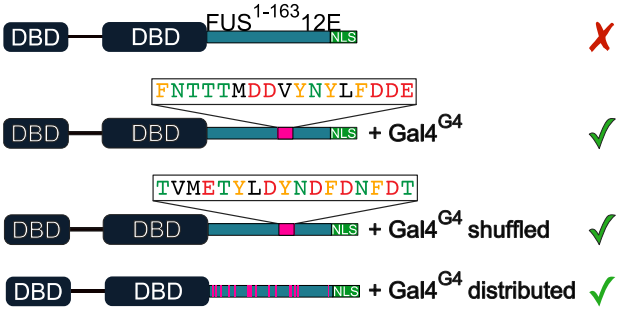

Figure 27: The $Gal4^{G4}$ subsequence still provides viability if shuffled or distributed across the FUS^{1-163}12E context. Symbols and amino acid coloring as in Figure 21.

Strikingly, the $Gal4^{G4}$ subsequence still provided viability in both cases. Hence, the composition alone was sufficient for the subsequence to mimic the effect of Abf1's EM. Even more surprisingly: just the mere presence of the respective amino acids even if distributed across the IDR was sufficient to confer viability to the FUS^{1-163}12E context.

To validate this surprising finding by independent cases, we examined if the EM of Abf1 and subsequences from TFs p65 and GR also tolerated a distribution across IDR context (Figure 28). The EM did not provide viability if distributed, neither in the FUS^{1-163}12E nor in the $Sup35^{1-131}$ context. The subsequences taken from p65 and GR were both still able to provide viability in both IDR contexts.

This verifies the EM as a genuine SLiM, as it loses its viability upon shuffling and distribution. In contrast, subsequences from p65 and GR were not true SLiMs, as they only met the criterion of modularity but their viability did not depend on their linear sequence. Therefore, sequence distribution of a (putative) motif across an IDR represents another, more stringent test that must be employed to validate *bona fide* SLiMs.

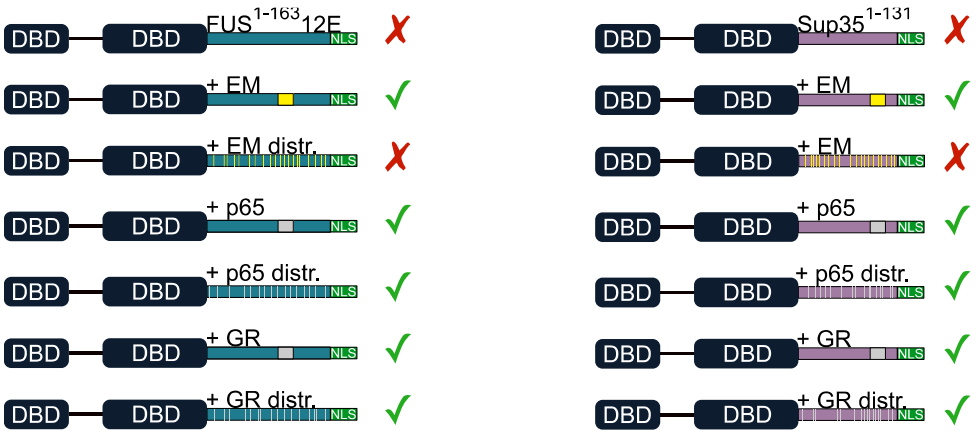

Figure 28: Other subregions that seemed to act like SLiMs turned out to just contribute to the composition of the IDR context as they still conferred function if distributed. The essential motif found in IDR2$^{449-662}$ of Abf1 did not support viability anymore if distributed across low-complexity domains of FUS$^{1-163}$12E and Sup35$^{1-131}$. Subsequences from TFs p65 and GR provided viability even if distributed across either IDR.

Given that the EM was a true SLiM, we also tested if the EM alone even without IDR context was sufficient for viability. For this we used an EM-only construct, in which the EM was just flanked by $(GlySer)_3$ linkers but no IDR context was present (schematic not shown). This construct did not provide viability, hence underscoring the essential requirement of the IDR context.

In conclusion: the EM is a *bona fide* SLiM that fulfills the modularity criterion and depends for its function on its linear sequence. It is not sufficient to provide viability without or if distributed across IDR context. The three examined subsequences from the other TFs Gal4, p65 and GR were not *bona fide* SLiMs, as their function did not depend on their linear sequence, but their composition alone was sufficient.

4.1.9 Viable IDR2$^{449-662}$ context relies on its proper chemical specificity

A subsequence that is distributed across an IDR contributes to the chemistry of the general IDR context but does not function as a “classic” modular SLiM anymore. So far, we found that one determinant of IDR viability is the presence of a SLiM or a compositionally similar

subsequence. Moreover, we saw that the presence of an IDR context is essential: the construct with the EM alone lacking a surrounding IDR was not viable.

These findings motivated us to ascertain if a viable IDR context required specific chemical features. First, we investigated the importance of aromaticity: we chose the construct with the FUS$^{1-163}$12E IDR context and the Gal4^{G4} motif present and changed all aromatic residues (here: 24 tyrosine residues) in the context to either serine or leucine (Figure 29).

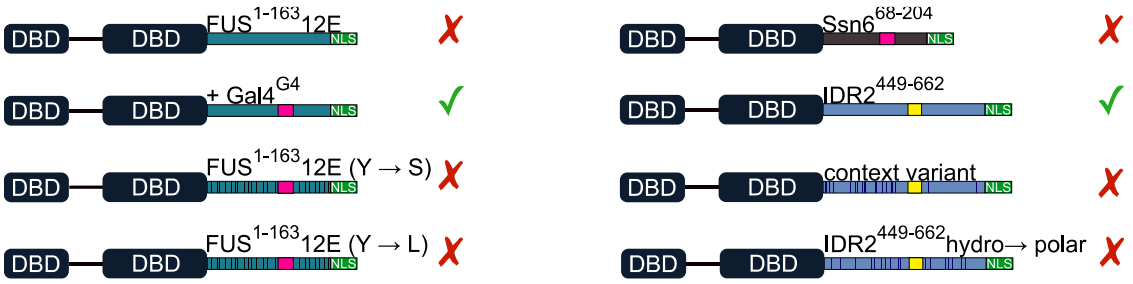

Figure 29: Aromaticity and hydrophobicity are essential features of a viable IDR-context. Exchange of 24 tyrosine residues in the FUS$^{1-163}$12E Gal4^{G4} context to serine or leucine resulted in inviable constructs. The Gal4^{G4} subsequence (magenta) cannot provide viability to the Glutamine-rich IDR of *S. pombe* transcription co-repressor Ssn6. Reduced hydrophobicity in *S. cerevisiae* IDR2$^{449-662}$ of Abf1 resulted in an inviable IDR-context. A context variant created by random mutagenesis with an altered IDR context, but unimpaired essential motif (yellow) was also inviable.

The constructs were designed based on the published observations that the tyrosine residues present in FUS are the basis for multivalent interactions required for LLPS [79,207].

Both constructs resulted in inviability, which underscores the importance of aromaticity for a viable IDR-context. Additionally, we investigated the viability of a Glutamine-rich IDR context using residues 68-206 of the *S. pombe* transcription co-repressor Ssn6 where the Gal4^{G4} subsequence was inserted. This construct was also inviable. Furthermore, we investigated the role of the composition of the IDR context in IDR2$^{449-662}$ of Abf1 (Figure 29). In this IDR, we showed that global shuffling of the context was tolerated as long as the linear sequence of the EM was not changed (Figure 19). One construct generated through random mutagenesis (4.1.4) already implicated that the context was important: the EM was unaltered, but the context was changed through mutagenesis and the construct was inviable. We then examined the importance of hydrophobicity in the context of IDR2$^{449-662}$ of Abf1 by means of a construct with an unaltered EM but with reduced hydrophobicity in the context. This construct was inviable.

We then used a previously described construct (4.1.4) which was generated by random mutagenesis that held 55 point mutations in the amino acid sequence (schematic not shown).

Here, we selected residues that were polar in the WT sequence but altered to hydrophobic residues in the construct generated through random mutagenesis. These hydrophobic residues were then mutated back to polar residues. This made the resulting sequence more similar to the WT but also transformed the construct from a viable to an inviable one. Spotting assays on 5-FOA plates without leu conducted by Andrea Schmid later showed that this construct was viable but with a very slow growth rate (data not shown). The inviability of these two IDR2$^{449-662}$-based constructs demonstrated that there were also features present in this IDR context that determined its viability and one of them was indeed hydrophobicity. These findings further emphasized the essentiality of the IDR-context for viability.

Next, we asked if the context features also mattered for other (sub-)sequences. To test this, we used the FUS$^{1-163}$12E context that was complemented with the hydrophobic TDP-43 helix and changed all aromatic residues in the context to serine (Figure 30).

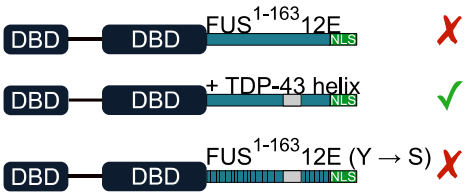

Figure 30: Aromaticity in IDR-context is an essential determinant also in combination with the TDP-43 helix subsequence. Changing all aromatic residues in the FUS$^{1-163}$12E context complemented with the TDP-43 helix altered the construct from a viable to an inviable one.

These changes altered the construct to an inviable one. Therefore, aromaticity is here again an essential feature for IDR context and is independent from the inserted subsequence that initially provided viability to the disordered region.

Moreover, we examined the relevance of acidity in the IDR context. To this end, we depleted all acidic residues in the constructs FUS$^{1-163}$12E Gal4^{G4} and Sup35$^{1-131}$ Gal4^{G4}. The acidic residues, both in the IDR context and in the Gal4^{G4} subsequences, were substituted by either serine, glycine, or glutamine residues (Figure 31).

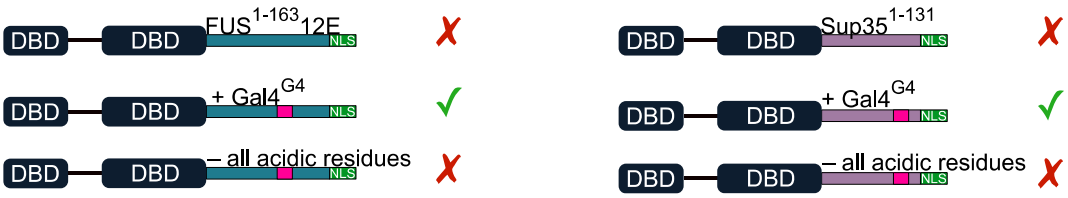

Figure 31: Acidity is another essential feature of a viable IDR. The removal of all acidic residues (in the context as well as in the subsequence) in the FUS$^{1-163}$12E Gal4^{G4} and in the Sup35 Gal4^{G4} constructs resulted in their inviability.

The depletion of all acidic residues led to inviability in both constructs, FUS$^{1\text{-}163}$12E Gal4^{G4} and Sup35$^{1\text{-}131}$ Gal4^{G4}. Hence, sufficient acidity in context (and subsequence) is another feature required for a viable IDR.

Taken together, we investigated the role of IDR context and the importance of chemical specificity for the IDR context in terms of acidity, hydrophobicity and aromaticity. We found that all three features were essential for a viable IDR context, independent from the kind of SLiM or functionally similar subsequences. Moreover, an enrichment in glutamine residues is not a feature required for a viable IDR context.

4.1.10 IDR mediated interactions can be described via a two-dimensional binding landscape

Our findings so far showed that the viability of Abf1's IDR2$^{449\text{-}662}$ was determined by the presence of a SLiM (the EM) as well as chemical features of the IDR context. Viable constructs could be designed by the combination of subsequences with similar composition as the EM embedded in certain contexts or even, surprisingly, by context-only designs where these subsequences were distributed across the context. Both aspects, SLiMs and contexts, are known as facilitators of IDR-mediated protein-protein interactions: SLiMs consisting of just a few amino acid residues enable low- to medium-affinity stoichiometric 1:1 binding whereas the sequence context is mostly discussed in connection with multivalent interactions involving LLPS[52,79–81,208,209], although context can also mediate stoichiometric interactions[76,77,79–81].

We conceptualized in a more general way that IDR viability provided through IDR-mediated interactions can be described via a two-dimensional binding landscape (Figure 32).

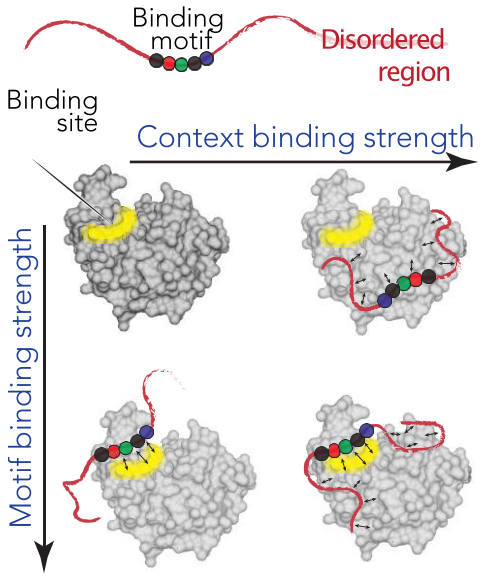

Figure 32: Motif and context binding strength contribute to a two-dimensional binding landscape that describes IDR-mediated interactions. The linear binding motif (colored circles) is located within the disordered region (red). The interaction partner is shown as a grey surface with a binding site in yellow. IDR binding to its interaction partner can be achieved through context binding strength (top right), motif binding strength (bottom left) or both (bottom right). *Adapted with permission from Dr. Alex Holehouse.*

Here, binding of an IDR to its interaction partner can be achieved through motif binding strength, context binding strength or combinations of both. As such, binding is enabled through sequence-specificity and/or chemical-specificity. Sequence-specificity is provided by linear binding motifs/SLiMs. Chemical specificity is provided by sequence-encoded chemical features of the context required for viability: aromaticity, hydrophobicity and acidity. Both, sequence- and chemical specificity can act orthogonally and interoperably.

Course-grained simulations of a 1:1 binding of an IDR to its interaction partner were performed to simulate the two-dimensional binding landscape (Figure 33). The transition from an unbound state to a bound state can be accomplished either by an increase in context binding strength (transition from 1 to 2; Figure 33 A) or in motif binding strength (transition from 1 to 3). Some examples for the approximated location of different rationally designed constructs in the two-dimensional binding landscape and their transitions from unbound and inviable to bound and viable or *vice versa* due to changes in motif- or context binding strength are shown in Figure 33 B.

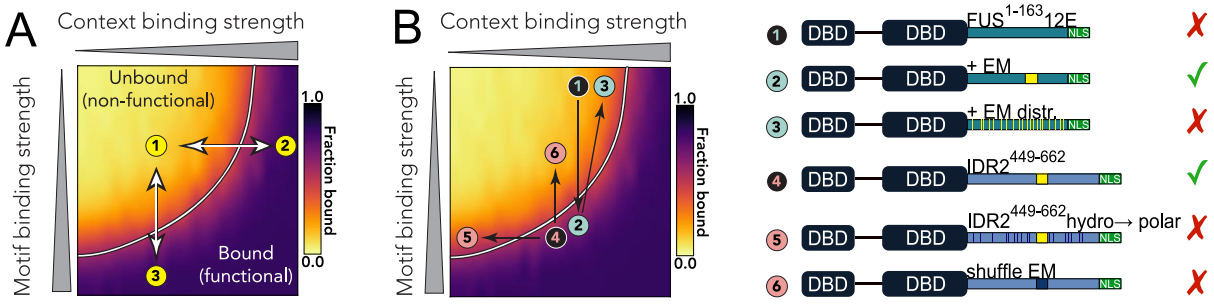

Figure 33: The two-dimensional binding landscape based on course-grained simulations. A) The binding of the IDR to its respective binding partner can be achieved through changes in motif binding strength, context binding strength or both. The unbound fraction is depicted in yellow, the border between unbound and bound is illustrated as a white line and the bound fraction is visualized in purple. **B)** Classification of exemplary constructs in the two-dimensional binding landscape. For details of the simulation see [198]. *Adapted with permission from Dr. Alex Holehouse.*

The context binding strength of the inviable construct FUS$^{1\text{-}163}$12E is increased through the addition of essential motif of Abf1 and results in a viable construct (transition from 1 to 2). The distribution of the essential motif across the FUS IDR causes an extensive decrease of motif binding strength and only a small increase in context binding strength (transition from 2 to 3). This does not suffice for a bound state as FUS$^{1\text{-}163}$12E with a distributed EM is inviable.

IDR2$^{449\text{-}662}$ of Abf1 is assumed to be close to the border between the bound and unbound state (see also below Figure 35). The decrease of hydrophobic residues results in the reduction of context binding strength (transition from 4 to 5). Shuffling the linear sequence of the essential motif of Abf1 causes a decrease in motif binding strength (transition from 4 to 6). In both cases binding is abolished and results in inviability of the respective construct.

We showed that an EM-only construct that lacks IDR context cannot provide viability (4.1.8). In contrast, the constructs with distributed subsequences showed that IDR-only constructs could provide viability.

To confirm this via a *de novo* design according to this principle, we chose hydrophobicity as the critical feature to provide chemical specificity and enable IDR-mediated binding through sufficient context binding strength. We designed a construct in which we distributed the same seven hydrophobic residues contained in the Gal4^{G4} subsequence across the FUS$^{1\text{-}163}$12E IDR: three aliphatic (methionine) and four aromatic (tyrosine) residues (Figure 34).

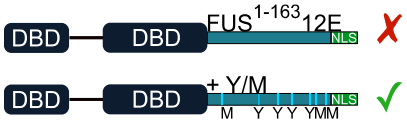

Figure 34: *De novo* design of a viable IDR only through increase in context binding strength.
Seven hydrophobic residues distributed across the FUS$^{1-163}$12E context were necessary and sufficient to provide viability in absence of a motif. Viability was achieved through increase in context binding strength.

The addition of the hydrophobic residues was necessary and sufficient to provide viability to that construct. Hence, we were able to design a viable artificial IDR *de novo* by increasing the context binding strength and presumably enabling IDR-mediated interactions through seven additional hydrophobic residues. As IDR2 of Abf1 usually required a SLiM to be viable, the viability of this construct confirms that sequence-specificity can be replaced by chemical specificity, i.e., that both types indeed act interoperably.

We previously delineated that a viable IDR context requires hydrophobicity (including aromaticity) and acidity. This prompted us to ask, if we can find a way to quantify our findings and to classify inviable and viable sequences if taken as context-only IDRs. To this end, we developed a composition-based metric that successfully distinguished 88 sequences based on a charge and binding score (Figure 35).

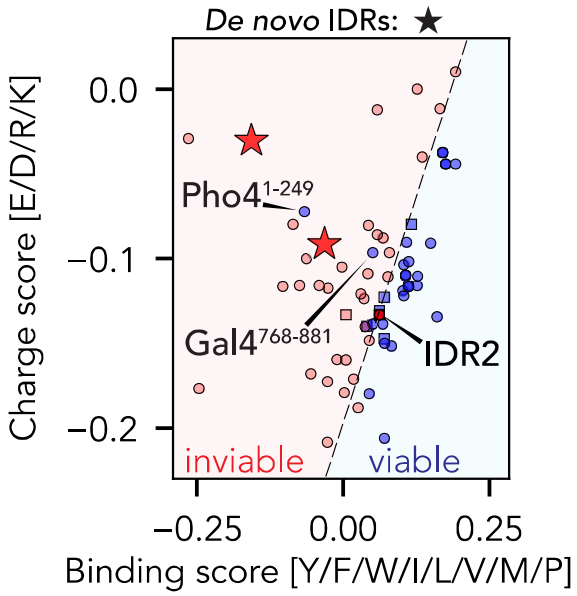

Figure 35: Classification of inviable and viable sequences based on charge and binding score. The scores are based on the weighted sequence composition with regard to the residues given in squared brackets. Inviable sequences are shown in red, viable sequences are depicted in blue. Squares/circles represent sequences that do/do not contain the essential motif, respectively, or the TDP-43 helix. Fully synthetic *de novo* designed sequences which were used to titrate the space for inviable sequences are illustrated as red stars. IDR2$^{449-662}$ of Abf1 (here indicated as IDR2) is located on the dashed line and hence on the brink of (in)viability. *Adapted with permission from Dr. Alex Holehouse.*

As assumed above (Figure 33), IDR2$^{449-662}$ of Abf1 is close to the border between the bound and unbound state in the two-dimensional landscape diagram. Constructs Pho4 $^{1-249}$ and Gal4 $^{768-881}$ are two examples for sequences that are in the space of inviable constructs but are actually viable. This suggests that they contain SLiMs that confer viability to an otherwise inviable context. This prediction would be that shuffling of these sequences will abolish viability, which will be tested in future experiments.

To verify the quantifications, we tested two fully synthetic, artificial constructs, which are shown as stars in Figure 35. As basis for these designs, we created a 164-residues comprising IDR based on the FUS$^{1-163}$12E context in which all acidic and aromatic residue were depleted. To this context we added the composition of one and three equivalents of the Gal4^{G4} motif and distributed them across the IDR (Figure 36).

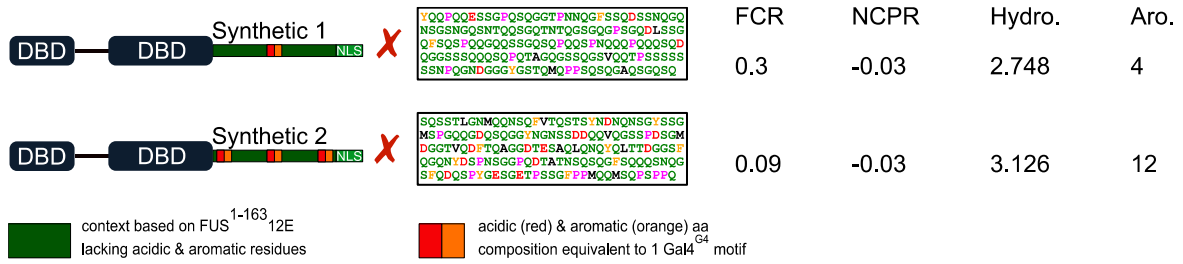

Figure 36: Fully synthetic *de novo* designed sequences used to titrate the sequence space of inviable sequences as in Figure 35. Constructs synthetic 1 and 2 are shown as schematics as well as their respective amino acid sequence. The corresponding biophysical parameters of synthetic 1 and 2 are listed on the right (FCR, NCPR, hydrophobicity (Hydro.) and number of aromatic residues of the entire sequence (Aro.)). Symbols and amino acid coloring as in Figure 21.

The compositional equivalents of one and three Gal4^{G4} motifs were both not sufficient for viability.

In conclusion, our results allowed us to describe IDR-mediated interactions via a two-dimensional binding landscape, in which the binding of an IDR to its binding partner is achieved through combinations of motif binding strength and context binding strength. We demonstrated that context binding strength relies on hydrophobic and charged residues. Furthermore, we quantified our findings with a composition-based metric that so far classified inviable context-sequences based on the charge and hydrophobicity.

4.1.11 The essential Abf1 function most likely does not depend on LLPS

The importance of LLPS in chromatin organization has previously been shown and IDRs have an important role in LLPS[105,210,211]. As FUS only engages in phase separation in presence of an adequate binding partner (as for example RNA) we asked, if viable constructs can display features that are not compatible with LLPS[201,207]. To this aim, we designed three LLPS-breaking constructs based on the FUS$^{1-163}$12E context (Figure 37).

Figure 37: Viability of LLPS-breaking constructs suggest that the essential Abf1 function does not depend on a mechanism involving phase separation. The left constructs altered valence 1 and 2 were based on different numbers of 25-residue repetitions taken from the FUS$^{1-163}$12E sequence with addition of the Gal4^{G4} subsequence and thereby varied IDR-length were viable. The construct with all tyrosine residues organized as aromatic clusters across the FUS$^{1-163}$12E Gal4^{G4} IDR, which should favor aggregation but not LLPS, also supported viability. *Adapted with permission from Dr. Alex Holehouse.*

IDRs behave like flexible polymers and LLPS supporting properties of polymers are length-dependent: the shorter the polymers are, the less they favor LLPS[80,81,212]. Based on this knowledge, we designed two constructs with altered valence of IDR subregion repeats. The first 167 residues comprising construct was built of six repetitive 25-residue units based on the FUS$^{1-163}$12E sequence to which one Gal4^{G4} subsequence was added (altered valence 1). The second construct was built similarly but consisted of only two repeats and 67 residues (altered valence 2).

As an orthogonal biophysical property known to affect LLPS-forming properties of IDRs, we exploited constructs with clustered in contrast to distributed aromatic residues. This design was based on the published observation that aromatic clusters favor aggregation and suppress LLPS[80,213]. We started again from the FUS$^{1-163}$12E context in combination with the Gal4^{G4} subsequence but re-organized all tyrosine residues in six aromatic clusters across the IDR.

Even though two of these three constructs were predicted to break LLPS-forming features, they were able to support viability. This suggested that the essential function of Abf1 is probably not based on a mechanism involving LLPS-based IDR-mediated interactions. Nevertheless, we cannot exclude that maybe a non-essential Abf1 function involves LLPS.

4.2 Investigation of transactivator Pho4 as nucleosome positioning barrier *in vitro*

Our *in vivo* approach (4.1) demonstrated that an IDR with certain properties was necessary for the essential function of Abf1. It is not resolved so far, which function of Abf1 is essentially required for viability. Abf1 is a GRF known to modulate chromatin structure *in vivo* as an nucleosome-depleted region (NDR) generating factor[121,129,156,214] (see 1.2) and was demonstrated to function as a nucleosome positioning barrier in cooperation with ATP-dependent remodelers like INO80[120]. We wondered if this barrier function was the basis for the essential function of Abf1 and started to investigate with our mechanistically well-defined *in vitro* reconstitution system which features make a DNA binding factor a nucleosome positioning barrier. As a first step, we wanted to ascertain if also a transactivator, not so far known as a nucleosome positioning barrier, can act as a nucleosome positioning barrier in our genome-wide *in vitro* reconstitution system. We chose Pho4, as its protein sequences besides its DNA binding domain (construct Pho4$^{1-249}$ in Figure 18 B) was able to replace Abf1 *in vivo*. SGD-assembled genome-wide chromatin incubated with the ATP-dependent chromatin remodeler INO80[215] and purified full-length Pho4 was able to generate an NDR over its binding sites and to position flanking nucleosomes (Figure 38 B).

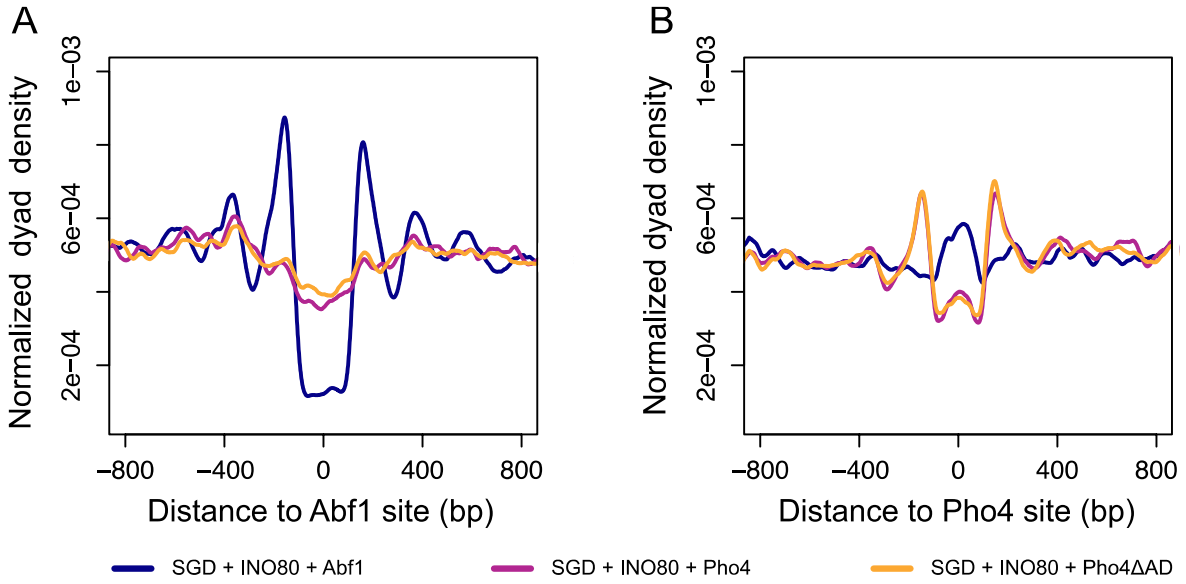

Figure 38: The transactivator Pho4 is a weaker nucleosome positioning barrier than the GRF Abf1. *In vitro* reconstituted SGD chromatin incubated with the ATP-dependent remodeler INO80 and the GRF Abf1 (blue line) or transactivator Pho4 (magenta line) or Pho4ΔAD (orange line). INO80 and Abf1 were used in a 1:2 molar ratio, INO80 to Pho4 ratio was 1:20 and INO80 and Pho4ΔAD in a 1:10 ratio. Plots were aligned to the respective protein's binding sites (by PWM): **A)** 1029 Abf1 binding sites[216], **B)** 1198 Pho4 binding sites[129]. Bioinformatic analyses were performed and plots were generated by Dr. Tamas Schauer.

This was even true for a truncated Pho4 version that contained mainly the DNA binding domain, but lacked the transactivation domain (Pho4ΔAD^{166,217}, residues 163-312). Therefore, the transactivation domain of Pho4 is not involved in nucleosome positioning at Pho4 binding sites. However, in both cases, the flanking nucleosome arrays were less pronounced than seen with the positive control of the combination of INO80 with Abf1 (Figure 31, panel B vs. A). Hence, Pho4 is a weaker nucleosome positioning barrier than Abf1. These results are not conclusive yet, as the concentrations of Abf1 versus Pho4 or Pho4ΔAD were not matched, but the beginning of an in-depth mechanistic analysis.

5 DISCUSSION

The lack of well-defined structure has led the field of protein research to see IDRs for almost half a century as unstructured blobs without relevant biological functions. This changed with growing evidence that IDRs are actually very functionally versatile due to this lack of rigid structure. During the last two decades, IDRs became popular objects of protein research that again experienced a great boost when their involvement in LLPS and crucial role in several human diseases was uncovered. Yet, the modus operandi of IDRs is still incompletely understood and subject of current research. This thesis aimed to ascertain if the essential function of Abf1 relies on one or both of its two IDRs and if this was the case, we wanted to elucidate the functional determinants of the essential Abf1 IDR(s) *in vivo*. Moreover, we wanted to verify if the essential function of Abf1 was based on its ability to act as a barrier for nucleosome positioning and if the discriminating feature of function between GRFs and transactivators lies within their IDRs.

We showed that the essential function of Abf1 relies on a part of IDR2, namely IDR2$^{449\text{-}623}$. Here, the regions CS1 and CS2 (IDR2$^{624\text{-}662}$), which were previously described to be responsible for Abf1 to activate transcription and to participate in chromatin remodleing[143], were not necessary as long as an artificial NLS was provided[142]. It was shown before that CS1/2 provide the NLS to the wt Abf1. Nonetheless, we included CS1/2 in the construct IDR2$^{449\text{-}662}$ that we used as our model IDR and examined its functional determinants by a combined approach of orthologs, sequence shuffles, domain swaps, random mutagenesis and rational designs. We uncovered that the functional determinants of the investigated IDR2 of Abf1 are defined on two levels: chemical specificity of the IDR context and sequence-specificity conferred by a true SLiM. Against the conventional view that approached IDR context and SLiMs separately also with regards to the type of interactions they both engage in, they actually work interoperably and synergistically. Both will be discussed in more detail below. In retrospective, the random mutagenesis approach was not sufficient to uncover the functional determinants of IDR2 as it probably affected sequence-specificity and context chemistry simultaneously.

The approach of global shuffles has been used by others to test the effect of a randomized, scrambled Sup35 sequences on the prion-forming ability of the protein[69]. The global shuffles of IDR2 revealed, that the chemical specificity of the IDR context alone was not sufficient to

confer viability. In addition, the compositional conservation within orthologs was not sufficient to identify the essential function in an IDR like Abf1-IDR2.

The approach of local shuffles was first employed by us and enabled us to successfully identify an essential SLiM in IDR2$^{449\text{-}662}$. Interestingly, the essential motif we found as well as subregions Abf1$^{G4\text{-}like}$ and partially Gal4^{G4} are each predicted by Alpha2Fold to form a transient helix. We hypothesized that the transient helicity is a common theme among SLiMs and could be used to identify them. Therefore, we intentionally chose the TDP-43 subsequence because it has been demonstrated to form a transient α-helical structure. Transient helicity is a feature associated with IDR-mediated interactions[202,203] and the transient α-helical structure of TDP-43 engages in helix–helix interactions and thereby enhances the function of TDP-43 in phase separation[218]. This suggests that the essential function of Abf1 could involve an interaction that is mediated by its transient, α-helical EM. The interaction could occur between Abf1 and nucleosomes or between Abf1 and RNA polymerase II, as it has been shown that Abf1 functions as an insulator by roadblocking this polymerase and terminating transcription[202]. The interaction could involve the recruitment or the interception of other TFs to active transcription sites.

Furthermore, we demonstrated that the EM alone is not sufficient but that a functional IDR2$^{449\text{-}662}$ relies also on proper IDR context, which includes the right kind of chemical specificity, such as acidity, aromaticity and hydrophobicity. It has been observed by others that hydrophobic residues of an IDR direct the binding of a TF towards its promoter[78] and that the strength of a TAD depends on its acidic, hydrophobic and solvent-exposed aromatic residues[177,178]. Moreover, it has been observed that the multivalent interactions of the IDR context rely on sequence composition rather than sequence order[75,79,80,82]. Most surprisingly, we showed that an IDR context with the proper chemical specificity can sustain Abf1 function even without the presence of a SLiM: While all tested putative motifs/subsequences ticked the box for the modularity criterion, only the EM lost its ability to confer viability when the linear sequence of the motif was distributed. Therefore, the EM proved to be a *bona fide* SLiM while the other "motifs" just contributed to the IDR context. It was quite unexpected that an IDR that evolved a combination of SLiM and context, could also work just via context.

The common view saw IDR-mediated interactions either through the perspective of the IDR context or through the perspective of SLiMs[67,76,209,219]. Our *de novo* designs demonstrated that

the contributions of these two interaction modes, at least for Abf1, can vary in both a compensatory and synergistical way. We propose that the binding of an IDR to its interaction partner can be described via a two-dimensional binding landscape where binding is accomplished through motif binding strength (sequence-specificity), context binding strength (chemical specificity) or combinations of both. In other words, a strong context binding strength can compensate for a weak motif binding strength and *vice versa* but binding can also be achieved if both context and motif binding strength combined are just sufficient enough. The contribution of the motif was demonstrated by turning an inviable/non-functional IDR into a viable/functional one through the insertion of a motif/subsequence. The contribution of the context binding strength was shown by turning an inviable IDR into a viable one simply through distributing hydrophobic residues along the IDR context. We demonstrated the variable degrees of motif vs. context contributions for Abf1 but suggest that this is a general feature of IDRs. In addition to the direct contribution of IDR context to interactions, we speculate IDR context is also required to keep the SLiM accessible for protein-protein interactions. To date we do not know if there is a SLiM that would be functional without IDR context.

Moreover, we suggest testing motif distribution as a new standard to properly identify SLiMs. The here provided distribution of a linear sequence along an IDR context is a new approach to truly test for the characteristic of linearity: a *bona fide* SLiM must not confer functionality when being distributed. The importance of the accurate identification of a *bona fide* SLiM is underlined by the discovery that viruses, including SARS-Cov-2, successfully use SLiMs as vulnerable spots of their host cells: by mimicking host cell SLiMs, viruses tamper with regulatory cellular processes, which in turn makes SLiM-binding pockets promising targets for broad-spectrum antiviral inhibitors[220–222].

For SLiMs, a mechanism of motif evolution, in which a functional motif sequence is newly acquired through randomly occurring mutations *ex nihilo*, “out of nothing”, has been proposed[57,58]. Our model of the two-dimensional binding landscape suggests an even higher evolutionary plasticity as it allows for motifs to arise during evolution if the context strength decreases simultaneously and *vice versa*. The auxiliary SLiMs embedded in the IDR context can buffer and compensate for the addition or the loss of binding strength and can even develop into genuine SLiMs through just a few mutations.

Now that we uncovered that IDR-mediated interactions can be described via a two-dimensional binding landscape, the next aim is the development of a model that accurately predicts for any IDR where it is located in this landscape based on its sequence and encoded biophysical parameters. Our model IDR itself is probably on the border of viability, meaning close to the border of unbound/non-functional fractions in the two-dimensional binding landscape. We speculate that this simplifies the tunability of the molecular function of IDR2 for example by PTMs in comparison to an IDR2 that is deeply rooted in the bound/functional fraction and might require more drastic modifications to cross the bound/unbound border.

The dysfunctionality of most of the Abf1 orthologs can have different causes. Firstly, the IDR portion of the ortholog that corresponded to the *S. cerevisiae* IDR2 could have been too short, thereby lacking functionally important features that were not included in our *in vivo* approach: Research by others has demonstrated that the *K. lactis* homolog of Abf1 is functional in *S. cerevisiae*[163] and was confirmed by us using plasmid shuffling assay. This was shown for the full-length protein while our tested chimera protein, the portion of *K. lactis* Abf1-IDR corresponding to S. cerevisiae Abf1-IDR2$^{449\text{-}662}$ fused to the *S. cerevisiae* Abf1-DBD was not functional with regards to executing the essential function of Abf1. The reason for this might be that full-length of *K. lactis* Abf1 contained more IDR than we included in the chimera protein and that the omitted parts were important for function. Secondly, the interaction partners of Abf1 might have evolved differently in the other species investigated than in *S. cerevisiae* and in a way that the interactions with the Abf1 of the respective species can occur but not the interaction between the ortholog and the interaction partner in *S. cerevisiae*.

In summary, this work underlines the versatility of IDRs and demonstrates that one IDR is not like the other.

6 OUTLOOK

After uncovering the functional determinants of Abf1 IDR2 the next step is the investigation of the interactome of Abf1 as we presume that the essential function of Abf1 is coupled to the binding of at least one, if not several interaction partners. The interactome can for example be identified by surface plasmon resonance (SPR)[223,224] or by enzyme-catalyzed proximity labeling[225–227], both followed by mass spectrometry (MS) analysis.

Although TFs can undergo LLPS, which leads to re-organization of chromatin structure and regulation of transcription[104,105], we did not find strong evidence that the essential function of Abf1 includes LLPS based on IDR-mediated interactions : the *de novo* designed constructs that displayed features incompatible with LLPS (decreased polymer length or aromatic clusters present) were functional.

To decipher the essential function of Abf1 further experiments are required. Using the anchor-away approach[228], the *in vivo* ablation of wildtype Abf1 in the presence of inviable abf1-IDR-mutants and the possible impact on chromatin structure and/or transcription can be examined. By testing inviable *abf1*-IDR-mutants that have been validated by ChIP, it can be ensured that the mutants are expressed, present in the nucleus and binding to the respective binding site. Strand-specific RNA-sequencing can be used to examine if the insulator function of Abf1 is retained in *abf1*-IDR-mutants and *de novo* designs. Selected designs will be tested in the genome-wide *in vitro* reconstitution assay to ascertain if the function of Abf1 as a nucleosome positioning barrier is coupled to the essential function of IDR2 *in vivo*. It still needs to be elucidated if inviability is caused by the inability of Abf1 to function as a nucleosome positioning barrier, e.g. to generate phased nucleosomal arrays at promoter regions, or by some other function of Abf1.

Another open question is what differentiates GRFs from transactivators. We assume that the difference could lie within the IDR-regulated interactomes. Although our experiments are not conclusive yet the results so far showed, that Pho4 is a weaker nucleosome positioning barrier than Abf1 and that its transactivation domain is not required for this residual barrier function *in vitro*.

7 REFERENCES

1. Forman-Kay, J. D. & Mittag, T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* **21**, 1492–1499 (2013).
2. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T. & Dahlquist, F. W. The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, σ28. *Nat. Struct. Biol.* **4**, 285–291 (1997).
3. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337**, 635–645 (2004).
4. Obradovic, Z. *et al.* Predicting intrinsic disorder from amino acid sequence. *Proteins* **53 Suppl 6**, 566–72 (2003).
5. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–31 (1999).
6. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **37**, 509–516 (2012).
7. Kendrew, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662–666 (1958).
8. Bolton, W. & Perutz, M. F. Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 å resolution. *Nature* **228**, 551–552 (1970).
9. Wright, C. S., Hess, G. P. & Blow, D. M. Structure of crystalline methyl-chymotrypsin. *J. Mol. Biol.* **63**, 295–303 (1972).
10. Abad-Zapatero, C., Griffith, J. P., Sussman, J. L. & Rossmann, M. G. Refined crystal structure of dogfish M4 apo-lactate dehydrogenase. *J. Mol. Biol.* **198**, 445–67 (1987).
11. ABERCROMBIE, B. D. *et al.* Studies on the Conformational Properties of the High-Mobility-Group Chromosomal Protein HMG 17 and Its Interaction with DNA. *Eur. J. Biochem.* **84**, 173–177 (1978).
12. Boesch, C., Bundi, A., Oppliger, M. & Wüthrich, K. 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur. J. Biochem.* **91**, 209–14 (1978).
13. Huber, R. Conformational flexibility in protein molecules. *Nature* **280**, 538–539 (1979).
14. Chouard, T. Structural biology: Breaking the protein rules. *Nature* **471**, 151–153 (2011).
15. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).

16. van der Lee, R. *et al.* Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
17. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
18. Holehouse, A. S. & Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell Biol.* (2023). doi:10.1038/s41580-023-00673-0
19. Posey, A. E., Holehouse, A. S. & Pappu, R. V. Phase Separation of Intrinsically Disordered Proteins. in *Methods in enzymology* **611**, 1–30 (Elsevier Inc., 2018).
20. Gadhave, K. *et al.* The dark side of Alzheimer's disease: unstructured biology of proteins from the amyloid cascade signaling pathway. *Cell. Mol. Life Sci.* **77**, 4163–4208 (2020).
21. Ahmed, R. & Forman-Kay, J. D. Aberrant phase separation: linking IDR mutations to disease. *Cell Res.* **33**, 583–584 (2023).
22. Chen, Y. *et al.* Mechanisms governing target search and binding dynamics of hypoxia-inducible factors. *Elife* **11**, 1–32 (2022).
23. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11504–9 (1996).
24. Toufektchan, E. & Toledo, F. The guardian of the genome revisited: P53 downregulates genes required for telomere maintenance, DNA repair, and centromere structure. *Cancers (Basel).* **10**, (2018).
25. Chen, C., Ding, X., Akram, N., Xue, S. & Luo, S. Z. Fused in sarcoma: Properties, self-assembly and correlation with neurodegenerative diseases. *Molecules* **24**, 1–17 (2019).
26. Carey, J. L. & Guo, L. Liquid-Liquid Phase Separation of TDP-43 and FUS in Physiology and Pathology of Neurodegenerative Diseases. *Front. Mol. Biosci.* **9**, 1–20 (2022).
27. Balbirnie, M., Grothe, R. & Eisenberg, D. S. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β-sheet structure for amyloid. *Proc. Natl. Acad. Sci.* **98**, 2375–2380 (2001).
28. Stefanis, L. α-Synuclein in Parkinson's Disease. *Cold Spring Harb. Perspect. Med.* **2**, a009399–a009399 (2012).
29. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc.*

Natl. Acad. Sci. U. S. A. **107**, 8183–8 (2010).

30. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13392–7 (2013).
31. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **42**, 38–48 (2001).
32. Lise, S. & Jones, D. T. Sequence patterns associated with disordered regions in proteins. *Proteins* **58**, 144–50 (2005).
33. Mittal, A., Holehouse, A. S., Cohan, M. C. & Pappu, R. V. Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. *J. Mol. Biol.* **430**, 2403–2421 (2018).
34. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–12 (2015).
35. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).
36. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. BioSyst.* **8**, 268–281 (2012).
37. Fuxreiter, M., Tompa, P. & Simon, I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950–956 (2007).
38. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der Dtsch. Chem. Gesellschaft* **27**, 2985–2993 (1894).
39. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.* **44**, 98–104 (1958).
40. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38 (2009).
41. Parker, D. *et al.* Role of Secondary Structure in Discrimination between Constitutive and Inducible Activators. *Mol. Cell. Biol.* **19**, 5601–5607 (1999).
42. Bustos, D. M. & Iglesias, A. A. Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins* **63**, 35–42 (2006).
43. Radivojac, P. *et al.* Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. *Proteins* **63**, 398–410 (2006).
44. Iakoucheva, L. M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–49 (2004).
45. Khan, A. N. & Lewis, P. N. Unstructured conformations are a substrate requirement for

the Sir2 family of NAD-dependent protein deacetylases. *J. Biol. Chem.* **280**, 36073–8 (2005).

46. Ren, S., Uversky, V. N., Chen, Z., Dunker, A. K. & Obradovic, Z. Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. *BMC Genomics* **9 Suppl 2**, S26 (2008).
47. Beltrao, P. & Serrano, L. Comparative Genomics and Disorder Prediction Identify Biologically Relevant SH3 Protein Interactions. *PLoS Comput. Biol.* **1**, e26 (2005).
48. Wigington, C. P. *et al.* Systematic Discovery of Short Linear Motifs Decodes Calcineurin Phosphatase Signaling. *Mol. Cell* **79**, 342-358.e12 (2020).
49. Davey, N. E. *et al.* Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *FEBS J.* **284**, 485–498 (2017).
50. Benz, C. *et al.* Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Mol. Syst. Biol.* **18**, 1–27 (2022).
51. Neduva, V. *et al.* Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.* **3**, e405 (2005).
52. Van Roey, K. *et al.* Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **114**, 6733–78 (2014).
53. Donaldson, L. W., Gish, G., Pawson, T., Kay, L. E. & Forman-Kay, J. D. Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. *Proc. Natl. Acad. Sci.* **99**, 14053–14058 (2002).
54. Seet, B. T. & Pawson, T. MAPK signaling: Sho business. *Curr. Biol.* **14**, R708-10 (2004).
55. Zarrinpar, A., Park, S.-H. & Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676–680 (2003).
56. O'Shea, C. *et al.* Structures and Short Linear Motif of Disordered Transcription Factor Regions Provide Clues to the Interactome of the Cellular Hub Protein Radical-induced Cell Death1. *J. Biol. Chem.* **292**, 512–527 (2017).
57. Neduva, V. & Russell, R. B. Linear motifs: evolutionary interaction switches. *FEBS Lett.* **579**, 3342–5 (2005).
58. Davey, N. E., Cyert, M. S. & Moses, A. M. Short linear motifs - ex nihilo evolution of protein regulation. *Cell Commun. Signal.* **13**, 43 (2015).
59. Gould, C. M. *et al.* ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.* **38**, D167–D180 (2010).

60. Edwards, A. S. & Newton, A. C. Phosphorylation at Conserved Carboxyl-terminal Hydrophobic Motif Regulates the Catalytic and Regulatory Domains of Protein Kinase C. *J. Biol. Chem.* **272**, 18382–18390 (1997).
61. Bu, J. Y., Shaw, A. S. & Chan, A. C. Analysis of the interaction of ZAP-70 and syk protein-tyrosine kinases with the T-cell antigen receptor by plasmon resonance. *Proc. Natl. Acad. Sci.* **92**, 5106–5110 (1995).
62. Pidoux, A. L. & Armstrong, J. Analysis of the BiP gene and identification of an ER retention signal in Schizosaccharomyces pombe. *EMBO J.* **11**, 1583–1591 (1992).
63. Glotzer, M., Murray, A. W. & Kirschner, M. W. Cyclin is degraded by the ubiquitin pathway. *Nature* **349**, 132–138 (1991).
64. Neduva, V. & Russell, R. B. DILIMOT: discovery of linear motifs in proteins. *Nucleic Acids Res.* **34**, W350–W355 (2006).
65. Edwards, R. J., Davey, N. E. & Shields, D. C. SLiMFinder: A Probabilistic Method for Identifying Over-Represented, Convergently Evolved, Short Linear Motifs in Proteins. *PLoS One* **2**, e967 (2007).
66. Davey, N. E., Shields, D. C. & Edwards, R. J. SLiMDisc: Short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res.* **34**, 3546–3554 (2006).
67. Bugge, K. *et al.* Interactions by Disorder – A Matter of Context. *Front. Mol. Biosci.* **7**, 1–16 (2020).
68. Stein, A. & Aloy, P. Contextual specificity in peptide-mediated protein interactions. *PLoS One* **3**, e2524 (2008).
69. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12825–30 (2005).
70. Amin, A. N., Lin, Y.-H., Das, S. & Chan, H. S. Analytical Theory for Sequence-Specific Binary Fuzzy Complexes of Charged Intrinsically Disordered Proteins. *J. Phys. Chem. B* **124**, 6709–6720 (2020).
71. Choi, J. M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
72. Borg, M. *et al.* Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitive. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9650–9655 (2007).
73. Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–21 (2001).
74. Orlicky, S., Tang, X., Willems, A., Tyers, M. & Sicheri, F. Structural basis for

phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* **112**, 243–56 (2003).

75. Song, J., Ng, S. C., Tompa, P., Lee, K. A. W. & Chan, H. S. Polycation-π Interactions Are a Driving Force for Molecular Recognition by an Intrinsically Disordered Oncoprotein Family. *PLoS Comput. Biol.* **9**, (2013).
76. Borgia, A. *et al.* Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).
77. Wiggers, F. *et al.* Diffusion of a disordered protein on its folded ligand. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–9 (2021).
78. Jonas, F. *et al.* The molecular grammar of protein disorder guiding genome-binding locations. *Nucleic Acids Res.* **51**, 4831–4844 (2023).
79. Lin, Y., Currie, S. L. & Rosen, M. K. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* **292**, 19110–19120 (2017).
80. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
81. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).
82. Pak, C. W. *et al.* Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* **63**, 72–85 (2016).
83. Buske, P. J. & Levin, P. A. A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo. *Mol. Microbiol.* **89**, 249–263 (2013).
84. Wang, Y. *et al.* Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nat. Chem. Biol.* **7**, 214–221 (2011).
85. Buljan, M. *et al.* Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. *Mol. Cell* **46**, 871–883 (2012).
86. van der Lee, R. *et al.* Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution. *Cell Rep.* **8**, 1832–1844 (2014).
87. Kussie, P., Gorina, S. & Marechal, V. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science (80-.).* **274**, 948–53 (1996).
88. Liu, J. *et al.* Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–88 (2006).
89. Staller, M. V. Transcription factors perform a 2-step search of the nucleus. *Genetics* (2022). doi:10.1093/genetics/iyac111

90. Liu, Y., Matthews, K. S. & Bondos, S. E. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila Hox protein ultrabithorax. *J. Biol. Chem.* **283**, 20874–20887 (2008).
91. Zaret, K. S. Pioneer Transcription Factors Initiating Gene Network Changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
92. Laptenko, O. *et al.* The p53 C Terminus Controls Site-Specific DNA Binding and Promotes Structural Changes within the Central DNA Binding Domain. *Mol. Cell* **57**, 1034–1046 (2015).
93. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol. Cell* **79**, 459-471.e4 (2020).
94. Berlow, R. B., Dyson, H. J. & Wright, P. E. Hypersensitive termination of the hypoxic response by a disordered protein switch. *Nature* **543**, 447–451 (2017).
95. Bhattacharya, S. *et al.* Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1. *Genes Dev.* **13**, 64–75 (1999).
96. Bah, A. & Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* **291**, 6696–705 (2016).
97. Wu, Y. *et al.* Phosphorylation of p53 by TAF1 inactivates p53-dependent transcription in the DNA damage response. *Mol. Cell* **53**, 63–74 (2014).
98. Sun, X., Dyson, H. J. & Wright, P. E. A phosphorylation-dependent switch in the disordered p53 transactivation domain regulates DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
99. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
100. Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. *Nat. Struct. Mol. Biol.* **26**, 3–13 (2019).
101. Roque, A., Iloro, I., Ponte, I., Arrondo, J. L. R. & Suau, P. DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. *J. Biol. Chem.* **280**, 32141–7 (2005).
102. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
103. Li, B. *et al.* Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* **316**, 1050–4 (2007).
104. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation

Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).

105. Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e21 (2019).
106. Walter, H. & Brooks, D. E. Phase separation in cytoplasm, due to macromolecular crowding, is the basis for microcompartmentation. *FEBS Lett.* **361**, 135–9 (1995).
107. Alberti, S. The wisdom of crowds: Regulating cell function through condensed states of living matter. *J. Cell Sci.* **130**, 2789–2796 (2017).
108. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-.).* **361**, (2018).
109. Kilic, S. *et al.* Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments. *EMBO J.* **38**, e101379 (2019).
110. Sanulli, S. *et al.* HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* **575**, 390–394 (2019).
111. Ren, J., Zhang, Z., Zong, Z., Zhang, L. & Zhou, F. Emerging Implications of Phase Separation in Cancer. *Adv. Sci.* **9**, 1–23 (2022).
112. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).
113. Huang, A. & Stultz, C. M. Finding order within disorder: Elucidating the structure of proteins associated with neurodegenerative disease. *Future Med. Chem.* **1**, 467–482 (2009).
114. Faux, N. G. *et al.* Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* **15**, 537–551 (2005).
115. Hasegawa, M. *et al.* Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann. Neurol.* **64**, 60–70 (2008).
116. Littlewood, T. D., Kreuzaler, P. & Evan, G. I. All Things to All People. *Cell* **151**, 11–13 (2012).
117. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
118. Lascaris, R. F., Groot, E., Hoen, P. B., Mager, W. H. & Planta, R. J. Different roles for abf1p and a T-rich promoter element in nucleosome organization of the yeast RPS28A gene. *Nucleic Acids Res.* **28**, 1390–6 (2000).
119. Fourel, G., Miyake, T., Defossez, P. A., Li, R. & Gilson, É. General Regulatory Factors (GRFs) as genome partitioners. *J. Biol. Chem.* **277**, 41736–41743 (2002).
120. Krietenstein, N. *et al.* Genomic Nucleosome Organization Reconstituted with Pure

Proteins. *Cell* **167**, 709-721.e12 (2016).

121. Yan, C., Chen, H. & Bai, L. Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. *Mol. Cell* **71**, 294-305.e4 (2018).
122. Bosio, M. C., Fermi, B. & Dieci, G. Transcriptional control of yeast ribosome biogenesis: A multifaceted role for general regulatory factors. *Transcription* **8**, 254–260 (2017).
123. Rossi, M. J., Lai, W. K. M. & Pugh, B. F. Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. *Genome Res.* **28**, 497–508 (2018).
124. Buchman, a R., Kimmerly, W. J., Rine, J. & Kornberg, R. D. Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **8**, 210–25 (1988).
125. Rhode, P. R., Sweder, K. S., Oegema, K. F. & Campbell, J. L. The gene encoding ARS-binding factor I is essential for the viability of yeast. *Genes Dev.* **3**, 1926–39 (1989).
126. Beinoravičiūtė-Kellner, R., Lipps, G. & Krauss, G. In vitro selection of DNA binding sites for ABF1 protein from Saccharomyces cerevisiae. *FEBS Lett.* **579**, 4535–4540 (2005).
127. McBroom, L. D. B. & Sadowski, P. D. DNA bending by Saccharomyces cerevisiae ABF1 and its proteolytic fragments. *J. Biol. Chem.* **269**, 16461–8 (1994).
128. Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* **36**, 1331–1339 (2004).
129. Badis, G. *et al.* A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* **32**, 878–87 (2008).
130. Shore, D., Stillman, D. J., Brand, A. H. & Nasmyth, K. A. Identification of silencer binding proteins from yeast: possible roles in SIR control and DNA replication. *EMBO J.* **6**, 461–7 (1987).
131. Eisenberg, S., Civalier, C. & Tye, B. K. Specific interaction between a Saccharomyces cerevisiae protein and a DNA element associated with certain autonomously replicating sequences. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 743–6 (1988).
132. Halfter, H., Kavety, B., Vandekerckhove, J., Kiefer, F. & Gallwitz, D. Sequence, expression and mutational analysis of BAF1, a transcriptional activator and ARS1-binding protein of the yeast Saccharomyces cerevisiae. *EMBO J.* **8**, 4265–72 (1989).
133. Diffley, J. F. X. & Stillman, B. Purification of a yeast protein that binds to origins of

DNA replication and a transcriptional silencer. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2120–4 (1988).

134. Dorsman, J. C., Gozdzicka-Jozefiak, A., van Heeswijk, W. C. & Grivell, L. A. Multifunctional DNA proteins in yeast: the factors GFI and GFII are identical to the ARS-binding factor ABFI and the centromere-binding factor CPF1 respectively. *Yeast* **7**, 401–12 (1991).
135. Morrow, B. E., Johnson, S. P. & Warner, J. R. Proteins that bind to the yeast rDNA enhancer. *J. Biol. Chem.* **264**, 9061–8 (1989).
136. Morrow, B. E., Ju, Q. & Warner, J. R. Purification and characterization of the yeast rDNA binding protein REB1. *J. Biol. Chem.* **265**, 20778–83 (1990).
137. Brindle, P. K., Holland, J. P., Willett, C. E., Innis, M. A. & Holland, M. J. Multiple factors bind the upstream activation sites of the yeast enolase genes ENO1 and ENO2: ABFI protein, like repressor activator protein RAP1, binds cis-acting sequences which modulate repression or activation of transcription. *Mol. Cell. Biol.* **10**, 4872–85 (1990).
138. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
139. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
140. Cho, G., Kim, J., Mo Rho, H. & Jung, G. Structure-function analysis of the DNA binding domain of Saccharomyces cerevisiae ABF1. *Nucleic Acids Res.* **23**, 2980–2987 (1995).
141. Diffley, J. F. & Stillman, B. Similarity between the transcriptional silencer binding proteins ABF1 and RAP1. *Science* **246**, 1034–8 (1989).
142. Loch, C. M., Mosammaparast, N., Miyake, T., Pemberton, L. F. & Li, R. Functional and Physical Interactions between Autonomously Replicating Sequence-Binding Factor 1 and the Nuclear Transport Machinery. *Traffic* **5**, 925–935 (2004).
143. Miyake, T., Loch, C. M. & Li, R. Identification of a multifunctional domain in autonomously replicating sequence-binding factor 1 required for transcriptional activation, DNA replication, and gene silencing. *Mol. Cell. Biol.* **22**, 505–16 (2002).
144. Ma, J. & Ptashne, M. A new class of yeast transcriptional activators. *Cell* **51**, 113–9 (1987).
145. Rossi, M. J., Lai, W. K. M. & Pugh, B. F. Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. *Genome Res.* **28**, 497–508 (2018).
146. Abraham, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S. & Hicks, J. B. Regulation

of mating-type information in yeast. Negative control requiring sequences both 5' and 3' to the regulated region. *J. Mol. Biol.* **176**, 307–31 (1984).

147. Haber, J. E. Mating-type gene switching in Saccharomyces cerevisiae. *Annu. Rev. Genet.* **32**, 561–99 (1998).
148. Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. Characterization of a 'silencer' in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**, 41–8 (1985).
149. Brand, A. H., Micklem, G. & Nasmyth, K. A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. *Cell* **51**, 709–19 (1987).
150. Reed, S. H., Akiyama, M., Stillman, B. & Friedberg, E. C. Yeast autonomously replicating sequence binding factor is involved in nucleotide excision repair. *Genes Dev.* **13**, 3052–8 (1999).
151. Yu, S. *et al.* ABF1-binding sites promote efficient global genome nucleotide excision repair. *J. Biol. Chem.* **284**, 966–73 (2009).
152. Bosio, M. C. *et al.* Abf1 and other general regulatory factors control ribosome biogenesis gene expression in budding yeast. *Nucleic Acids Res.* **45**, 4493–4506 (2017).
153. Gonçalves, P. M. *et al.* Transcription activation of yeast ribosomal protein genes requires additional elements apart from binding sites for Abf1p or Rap1p. *Nucleic Acids Res.* **23**, 1475–80 (1995).
154. Venditti, P., Costanzo, G., Negri, R. & Camilloni, G. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. *Biochim. Biophys. Acta* **1219**, 677–89 (1994).
155. Yarragudi, A., Miyake, T., Li, R. & Morse, R. H. Comparison of ABF1 and RAP1 in Chromatin Opening and Transactivator Potentiation in the Budding Yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.* **24**, 9152–9164 (2004).
156. Bai, L., Ondracka, A. & Cross, F. R. Multiple Sequence-Specific Factors Generate the Nucleosome-Depleted Region on CLN2 Promoter. *Mol. Cell* **42**, 465–476 (2011).
157. Levo, M. *et al.* Systematic Investigation of Transcription Factor Activity in the Context of Chromatin Using Massively Parallel Binding and Expression Assays. *Mol. Cell* **65**, 604-617.e6 (2017).
158. Yarragudi, A., Miyake, T., Li, R. & Morse, R. H. Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.* **24**, 9152–64 (2004).

159. Jonge, W. J. De, Brok, M., Lijnzaad, P., Kemmeren, P. & Holstege, F. C. P. Genome-wide off-rates reveal how DNA binding dynamics shape transcription factor function. 1–16 (2020).
160. Fourel, G., Miyake, T., Defossez, P.-A., Li, R. & Gilson, E. General regulatory factors (GRFs) as genome partitioners. *J. Biol. Chem.* **277**, 41736–43 (2002).
161. Yarragudi, A., Parfrey, L. W. & Morse, R. H. Genome-wide analysis of transcriptional dependence and probable target sites for Abf1 and Rap1 in Saccharomyces cerevisiae. *Nucleic Acids Res.* **35**, 193–202 (2007).
162. Kraakman, L. S., Mager, W. H., Grootjans, J. J. & Planta, R. J. Functional analysis of the promoter of the gene encoding the acidic ribosomal protein L45 in yeast. *Biochim. Biophys. Acta* **1090**, 204–10 (1991).
163. Gonçalves, P. M., Maurer, K., Mager, W. H. & Planta, R. J. Kluyveromyces contains a functional ABF1-homologue. *Nucleic Acids Res.* **20**, 2211–5 (1992).
164. Gonçalves, P. M. *et al.* C-terminal domains of general regulatory factors Abf1p and Rap1p in Saccharomyces cerevisiae display functional similarity. *Mol. Microbiol.* **19**, 535–43 (1996).
165. Shao, D., CL, C. & LW, B. A cysteine residue in helixII of the bHLH domain is essential for homodimerization of the yeast transcription factor Pho4p. *Nucleic Acids Res.* **26**, 710–714 (1998).
166. Ogawa, N. & Oshima, Y. Functional Domains of a Positive Regulatory Protein, PH04, for Transcriptional Control of the Phosphatase Regulon in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **10**, 2224–2236 (1990).
167. Shimizu, T. *et al.* Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J.* **16**, 4689–4697 (1997).
168. Piskacek, S. *et al.* Nine-amino-acid transactivation domain: Establishment and prediction utilities. *Genomics* **89**, 756–768 (2007).
169. Komeili, A. & O'Shea, E. K. Roles of Phosphorylation Sites in Regulating Activity of the Transcription Factor Pho4. *Science (80-.).* **284**, 977–980 (1999).
170. Fascher, K. D., Schmitz, J. & Hörz, W. Role of trans-activating proteins in the generation of active chromatin at the PHO5 promoter in S. cerevisiae. *EMBO J.* **9**, 2523–2528 (1990).
171. Komeili, A. & O'Shea, E. K. Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. *Science* **284**, 977–80 (1999).
172. Kaffman, A., Herskowitz, I., Tjian, R. & O'Shea, E. K. Phosphorylation of the

Transcription Factor PHO4 by a Cyclin-CDK Complex, PHO80-PHO85. *Science (80-.).* **263**, 1153–1156 (1994).

173. Wolff, M. R., Schmid, A., Korber, P. & Gerland, U. Effective dynamics of nucleosome configurations at the yeast PHO5 promoter. *Elife* **10**, 1–28 (2021).
174. Korber, P. & Barbaric, S. The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Res.* **42**, 10888–10902 (2014).
175. Musladin, S., Krietenstein, N., Korber, P. & Barbaric, S. The RSC chromatin remodeling complex has a crucial role in the complete remodeler set for yeast PHO5 promoter opening. *Nucleic Acids Res.* **42**, 4270–4282 (2014).
176. Svaren, J., Schmitz, J. & Hörz, W. The transactivation domain of Pho4 is required for nucleosome disruption at the PHO5 promoter. *EMBO J.* **13**, 4856–62 (1994).
177. Staller, M. V. *et al.* A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst.* **6**, 444-455.e6 (2018).
178. Staller, M. V *et al.* Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Syst.* **13**, 334-345.e5 (2022).
179. Erkina, T. Y. & Erkine, A. M. Nucleosome distortion as a possible mechanism of transcription activation domain function. *Epigenetics Chromatin* **9**, 40 (2016).
180. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).
181. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–80 (1983).
182. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**, 19–27 (1989).
183. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Mutagenesis and homologous recombination in Drosophila cell lines using CRISPR/Cas9. *Biol. Open* **3**, 42–49 (2014).
184. Dunham, M., Brown, G. W. & Ünal, E. Yeast Genetics and Genomics 2017 Course Manual. 317 (2017).
185. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168 (1983).
186. Soni, R., Carmichael, J. P. & Murray, J. A. H. Parameters affecting lithium acetate-mediated transformation of Saccharomyces cerevisiae and development of a rapid and

simplified procedure. *Curr. Genet.* **24**, 455–459 (1993).

187. Neiman, A. M. Sporulation in the Budding Yeast Saccharomyces cerevisiae. *Genetics* **189**, 737–765 (2011).
188. Freese, E. B., Chu, M. I. & Freese, E. Initiation of yeast sporulation of partial carbon, nitrogen, or phosphate deprivation. *J. Bacteriol.* **149**, 840–51 (1982).
189. Huxley, C., Green, E. D. & Dunham, I. Rapid assessment of S. cerevisiae mating type by PCR. *Trends Genet.* **6**, 236 (1990).
190. Boeke, J. D., Trueheart, J., Natsoulis, G. & Fink, G. R. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. in *Methods in enzymology* **154**, 164–175 (1987).
191. Oberbeckmann, E. *et al.* Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. *Nat. Commun.* **12**, 1–17 (2021).
192. Krietenstein, N., Wippo, C. J., Lieleg, C. & Korber, P. Genome-wide in vitro reconstitution of yeast chromatin with in vivo-like nucleosome positioning. *Methods Enzymol.* **513**, 205–32 (2012).
193. Simon, R. H. & Felsenfeld, G. A new procedure for purifying histone pairs H2A + H2B and H3 + H4 from chromatin using hydroxylapatite. *Nucleic Acids Res.* **6**, 689–96 (1979).
194. Jones, G. M. *et al.* A systematic library for comprehensive overexpression screens in Saccharomyces cerevisiae. *Nat. Methods* **5**, 239–41 (2008).
195. Oberbeckmann, E. *et al.* Genome information processing by the INO80 chromatin remodeler positions nucleosomes. *Nat. Commun.* **12**, (2021).
196. Ertel, F. *et al.* In vitro reconstitution of PHO5 promoter chromatin remodeling points to a role for activator-nucleosome competition in vivo. *Mol. Cell. Biol.* **30**, 4060–76 (2010).
197. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–5 (1970).
198. Langstein-Skora, I. *et al.* Sequence- and chemical specificity define the functional landscape of intrinsically disordered regions. *bioRxiv* (2022). doi:10.1101/2022.02.10.480018
199. Hoskins, J. & Scott Butler, J. Evidence for distinct DNA- and RNA-based mechanisms of 5-fluorouracil cytotoxicity inSaccharomyces cerevisiae. *Yeast* **24**, 861–870 (2007).
200. Parker, W. B. & Cheng, Y. C. Metabolism and mechanism of action of 5-fluorouracil. *Pharmacol. Ther.* **48**, 381–395 (1990).
201. Monahan, Z. *et al.* Phosphorylation of the FUS low-complexity domain disrupts phase

separation, aggregation, and toxicity. *EMBO J.* **36**, 2951–2967 (2017).

202. Oldfield, C. J. *et al.* Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454–70 (2005).
203. Chi, S.-W. *et al.* Structural details on mdm2-p53 interaction. *J. Biol. Chem.* **280**, 38795–802 (2005).
204. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
205. Amador-Ortiz, C. *et al.* TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann. Neurol.* **61**, 435–45 (2007).
206. Uryu, K. *et al.* Concomitant TAR-DNA-Binding Protein 43 Pathology Is Present in Alzheimer Disease and Corticobasal Degeneration but Not in Other Tauopathies. *J. Neuropathol. Exp. Neurol.* **67**, 555–564 (2008).
207. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **60**, 231–241 (2015).
208. Benz, C. *et al.* Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Mol. Syst. Biol.* **18**, 1–27 (2022).
209. Davey, N. E. The functional importance of structure in unstructured protein regions. *Curr. Opin. Struct. Biol.* **56**, 155–163 (2019).
210. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
211. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
212. Rubinstein, M. & Colby, R. H. *Polymer Physics*. (2003).
213. Schmidt, H. B., Barreau, A. & Rohatgi, R. Phase separation-deficient TDP43 remains functional in splicing. *Nat. Commun.* **10**, 4890 (2019).
214. Ganapathi, M. *et al.* Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. *Nucleic Acids Res.* **39**, 2032–44 (2011).
215. Shen, X., Mizuguchi, G., Hamiche, A. & Wu, C. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* **406**, 541–4 (2000).
216. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
217. McAndrew, P. C., Svaren, J., Martin, S. R., Hörz, W. & Goding, C. R. Requirements for

Chromatin Modulation and Transcription Activation by the Pho4 Acidic Activation Domain. *Mol. Cell. Biol.* **18**, 5818–5827 (2015).

218. Conicella, A. E. *et al.* TDP-43 α-helical structure tunes liquid-liquid phase separation and function. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5883–5894 (2020).
219. Van Roey, K., Gibson, T. J. & Davey, N. E. Motif switches: Decision-making in cell regulation. *Curr. Opin. Struct. Biol.* **22**, 378–385 (2012).
220. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **36**, 159–169 (2011).
221. Simonetti, L., Nilsson, J., McInerney, G., Ivarsson, Y. & Davey, N. E. SLiM-binding pockets: an attractive target for broad-spectrum antivirals. *Trends Biochem. Sci.* **48**, 420–427 (2023).
222. Mészáros, B. *et al.* Short linear motif candidates in the cell entry system used by SARS-CoV-2 and their potential therapeutic implications. *Sci. Signal.* **14**, 1–26 (2021).
223. Cytiva. Biacore™ sensor surface Handbook. 71 (2022).
224. Xue, J., Bai, Y. & Liu, H. Hybrid methods of surface plasmon resonance coupled to mass spectrometry for biomolecular interaction analysis. *Anal. Bioanal. Chem.* **411**, 3721–3729 (2019).
225. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
226. Hwang, J. & Espenshade, P. J. Proximity-dependent biotin labelling in yeast using the engineered ascorbate peroxidase APEX2. *Biochem. J.* **473**, 2463–2469 (2016).
227. Hollstein, L. S., Schmitt, K., Valerius, O., Stahlhut, G. & Pöggeler, S. Establishment of in vivo proximity labeling with biotin using TurboID in the filamentous fungus Sordaria macrospora. *Sci. Rep.* **12**, 17727 (2022).
228. Haruki, H., Nishikawa, J. & Laemmli, U. K. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Mol. Cell* **31**, 925–932 (2008).

ACKNOWLEDGEMENTS

First and foremost, I want to thank Prof. Dr. Philipp Korber for the unique opportunity to conduct my PhD in his group. In particular, I wish to thank you for enabling my participation in many workshops and hands-on course. I am sincerely grateful to you for your trust, guidance and unwavering support throughout the years.

I am especially grateful to Prof. Alex S. Holehouse. Thank you for being our ingenious collaboration partner and your continuous enthusiasm on this "journey into the wilderness". Our online meetings via Teams were always a huge inspiration and big boost in motivation.

I wish to thank Prof. Peter Becker, for creating a thriving scientific environment that enables exceptional research.

Moreover, I want to thank Carolin Brieger and the late Edith Müller, whose diligent work lifted a lot of weight off the shoulders of many graduate students including myself.

I would especially like to thank Prof. Dr. Axel Imhof, Dr. Christoph Kurat and Dr. Tamas Schauer, my thesis advisory committee, for their many helpful suggestions and their interest in my PhD project. Dr. Tamas Schauer deserves special thanks for performing the bioinformatic analyses.

Furthermore, I would like to thank Dr. Elizabeth Schröder-Reiter and the IRTG 1064, which made it possible for me to participate in many scientific events and workshops that had a great impact on me.

I would also like to thank the combined Korber and Rupp lab especially Andrea, Elisa, Edith, Janet and Barbara for the familiar atmosphere in the lab. I want to thank Andrea in particular for conducting the ChIP experiments.

I am extremely grateful to Erika, Karolina, Roland and Florian – your encouragement and understanding made a big difference.

I want to express my biggest gratitude to Mama & Papa –you supported me in every way a parent can support their child and that allowed me to be independent and go my own way.

Last but not least I want to thank my husband Tobias: thank you for going this way with me – from learning metabolism pathways with crazy abbreviations on a small balcony in Bochum to unexpectedly shutting off freezer alarms on what were supposed to be relaxing Sunday walks. I could not have done this without you.

ABBREVIATIONS

°C	Degree Celcius
2D	Two-dimensional
3D	Three-dimensional
5-F-dUMP	5-fluorodeoxyuridine monophosphate
5-FOA	5-Fluorootic acid
8-oxo-dGTP	8-Oxo-2'-deoxyguanosine-5'-triphosphate
aa	Amino acid
Abf1	ARS-binding factor 1
ACS	American Chemical Society
ALS	Amyotrophic lateral sclerosis
ARS	Autonomously replicating sequence
bHLH	basic helix-loop-helix
bp	Basepair
Cbf1	Centromere Binding Factor 1
CBP	CREB (Cyclic adenosine monophosphate response element-binding protein) binding protein
Cdc4	Cell division control protein 4
ChIP	Chromatin immunoprecipitation
Cited2	cAMP-responsive element-binding protein (CBP)/p300-interacting transactivators with glutamic acid (E) and aspartic acid (D)-rich tail
CS1/2	C-terminal sequence 1/2
DBD	DNA-binding domain
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dPTP	6H,8H-3,4-Dihydro-pyrimido(4,5-c)(1,2)oxazin-7-one-8-β-D-2'-deoxy-ribofuranoside-5'-triphosphate
dTTP	2′-Desoxthymidine-5′-triphosphate
E. coli	Escherichia coli
ELM	Eukaryotic Linear Motif
EM	Essential motif
FCR	Fraction of charged residues
FUS	Fused in Sarcoma
G418	Kanamycin
Gal4	Galactose metabolism 4
Gcn4	General control nonderepressible4

GRF	General regulatory factor
h	Hour
HIF-1	Hypoxia-inducible factor-1
HIF-1	Hypoxia inducing factor 1
HMG	High-Mobility-Group Chromosomal Protein
HML	Homothallism left
HMR	Homothallism right
HMRE	E-region at HMR
IDR	Intrinsically disordered region
ISW1a	ISWI chromatin-remodeling complex ATPase 1a
ISW2	ISWI chromatin-remodeling complex ATPase 2
ISWI	Imitation switch
L	Liter
leu	Leucin
LexA	Locus for X-ray sensitivity A
LLPS	Liquid-liquid phase separation
MAT	Mating type locus
Mcm1	Mini Chromosome Maintenance 1
mg	Milligram
min	Minute
ml	Milliliter
MNase	Micrococcal nuclease
MS	Mass spectrometry
MWCO	Molecular weight cut-off
NCPR	Net charge per residues
NDR	Nucleosome-depleted region
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
ns	Not significant
O/N	Overnight
O1	Ordered region 1
O2	Ordered region 2
OD600	Optical density at 600 nm
ORC	origin recognition complex
Pbs2	Polymyxin B resistance protein 2
PCR	Polymerase chain reaction

PF	Pioneer factor
Pho4	Phosphate system positive regulatory protein 4
PNK	Polynucleotide Kinase
PTM	Post translational modification
PWM	Position weight matrix
qRT-PCR	quantitative Real Time polymerase chain reaction
Rap1	Repressor/activator site-binding protein 1
Reb1	RNA polymerase I enhancer binding protein 1
rpm	Rounds per minute
RSC	Chromatin structure-remodeling complex
RT	Room temperature
S. cerevisiae	Saccharomyces cerevisiae
SARS-Cov-2	Severe acute respiratory syndrome coronavirus type 2
SDS-PAGE	Sodiumdodecyl polyacrylamide gelelectrophoresis
sec	Second
SH2	Src-homology 2
SH3	Src-homology 3
Sho1	High osmolarity signaling protein
Sic1	Cyclin-dependent kinase inhibitor
SLiM	Short linear motif
SPR	Surface plasmon resonance
Sup35	Eukaryotic peptide chain release factor GTP-binding subunit
SV40	Simian virus 40
TAD	Transcription activation domain
Tbf1	TTAGGG repeat-Binding Factor 1
TDP-43	Transactive response DNA binding protein 43 kDa
TF	Transcription factor
ts	Temperature sensitive
U	Units
UASp	Upstream activating sequence phosphate regulated
ura	Uracil
USER	Uracil-Specific Excision Reagent
WT	Wildtype
µg	Microgram
µl	Microliter

APPENDIX

Table S1: Primer used to generate inserts for rational designs.

Construct	Primer insert forward (5'-->3')	Primer insert reverse (5'-->3')
C. glabrata	ACAACCTTCTTCATCTTCCTCTGGTAATAGAC CACCAAGAATCTCTAG	CTTCCTCTTCTTCTTGGGCTCGACATCCTT GAAATGAGAGGATAGC
essential motif only	GGCAGCGGCAGCGGCAGCGAAAACGTTCATCC AACTTTGGCTG	ACTCCCACTCCCACTCCCAGCAGTTTCTCT AGCTTCTTGAGC
CTCF IDR 578-727	ACAACCTTCTTCATCTTCCGCTGGTCCAGATG GTGTTGAAG	CTTCCTCTTCTTCTTGGGCCTATCCATCAT GGACAAGATCATTTCTGG
E. gossypi	ACAACCTTCTTCATCTTCCGATGATGATGTTT CCGTTCCAGAACACG	CTTCCTCTTCTTCTTGGGTTCGACTTCTTT CAAGTGAGAGGACAGC
FUS 1-163 12E	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAGC	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + TDP-43 helix	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + TDP-43 helix & no aro	ACAACCTTCTTCATCTTCCATGGCTTCTAACG ATGGTGAACAAC	CTTCCTCTTCTTCTTGGGAGAATTTTGTTG ATTTTGCTGTCCACCG
FUS 1-163	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATACTCAACAAGC	CTTCCTCTTCTTCTTGGGGGAGTTGTACTG ATTTTGCTGACCATATCC
FUS 1-163 12E + Gal4 G4 motif (Y->L)	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTTGGAACAAC	CTTCCTCTTCTTCTTGGGGGAGTTCAACTG ATTTTGTTGACCTAGACC
FUS 1-163 12E + Gal4 G4 motif distr.	ACAACCTTCTTCATCTTCCATGGCTTCTACCA ACGATTACGATG	CTTCCTCTTCTTCTTGGGGGAGTTGTACTG ATTTTGAAACTGACCG
WT Abf1 IDR2 + EM (Hydro -> polar)	ACAACCTTCTTCATCTTCCAACAACAACAACA ACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTT CAAGTGAGAGGATAGC
FUS 1-163 12E + Gal4 G4 - all acidic residues	ACAACCTTCTTCATCTTCCATGGCCTCTAATT CTTATGGTCAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGTCCGTATC
FUS 1-163 12E + p65	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + Gal4 G4 + aromatic clusters	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTACTACTACTACG	CTTCCTCTTCTTCTTGGGGGAGTTCTGGTT TTGTTGACCG
FUS 1-163 12E + GR	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + GR distr.	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGAGAGTTGTACTG ATTCAACTGTTGTCCG
FUS 1-163 12E + Gal4 M2	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + p65 distr.	ACAACCTTCTTCATCTTCCATGGCCTCTTTGA ACGATTACGAAC	CTTCCTCTTCTTCTTGGGGGAGTTGTACTG ATTCTGGAATTGACC
FUS 1-163 12E + VP16	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + EM	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGTTGGCCGTATCC
FUS 1-163 12E + distributed EM	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTACGAACAACAAAACGC	CTTCCTCTTCTTCTTGGGAGAGTTGTACTG ATTTTGCAATTGACCG
FUS 1-163 12E + Abf1 G4 motif	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + Gal4 G4	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
FUS 1-163 12E + Y/M	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGAGAGTTGTACTG ATTTTGCATTTGTCCGTAGC
FUS 1-163 12E + Gal4 G4 motif (Y->S)	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTCTGAACAAC	CTTCCTCTTCTTCTTGGGGGAATTAGATTG ATTTTGTTGTCCGCC
FUS 1-163 12E + GCN4	ACAACCTTCTTCATCTTCCATGGCCTCTAACG ATTATGAACAACAAG	CTTCCTCTTCTTCTTGGGGGAGTTATACTG ATTTTGCTGACCGTATCC
Gal4 IDR 768-881	ACAACCTTCTTCATCTTCCGCCAATTTTAATC AAAGTGGGAATATTGC	CTTCCTCTTCTTCTTGGGCTCTTTTTTTGG GTTTGGTGGGG
Gcn4 IDR 17-150	TACAACCTTCTTCATCTTCCTCACCATTGGAT GGTTCTAAATCAACC	CTTCCTCTTCTTCTTGGGCAGAGAAACTTC TTCAGTGGATTCAATTGC
Abf1 IDR1 87-311 + Gal4 G4 motif	ACAACCTTCTTCATCTTCCGAAACCTCTTCTC CATCTGCTAACAAC	CTTCCTCTTCTTCTTGGGGGTGTCTCTGAT AGATGATGGAGAATGAG
Abf1 IDR 87-311	ACAACCTTCTTCATCTTCCGAAACCTCATCTC CTTCTGCAAATAATAATACC	CTTCCTCTTCTTCTTGGGGTCTCGTATGGA GGATGGTGAGTGGG
K. africana	ACAACCTTCTTCATCTTCCAACGAACCTACCG ATGTTTTGGAAGAAC	CTTCCTCTTCTTCTTGGGTTCGACATCCTT GAAATGAGAGGACAAC

Construct	Primer insert forward (5'-->3')	Primer insert reverse (5'-->3')
K. lactis	ACAACCTTCTTCATCTTCCCATCACCAAGTTCAACATCATCATCACCC	CTTCCTCTTCTTCTTGGGTTCGACTTCTTTCAAGTGAGAGGACAATAATC
K. lactis full length	CATATTTGCAATTTCACAAGGATGTCCCTGTACGAGTACAAACATCC	CTTCCTCTTCTTCTTGGGTTGACCTCTCAATTCTGGCTGAATG
K. naganashii	ACAACCTTCTTCATCTTCCGGTTCTTCTACTGTTACTCCATCAGGTGC	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTGAAGTGAGAGGATAGCAATC
L. kluyveri	ACAACCTTCTTCATCTTCCGCTGCTGCTGCTATTGCTTCTGG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAACAAACGCC
LS-1	ACAACCTTCTTCATCTTCCGAAAACCAGAACTCCAACAAGAACTACG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-2	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTGAG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-3	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTGAG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-4	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTGAG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-5	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTGAG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-6	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTGAG	CTTCCTCTTCTTCTTGGGTTCGACTTCTTTCAAGTGAGAGGACAACAAC
LS-7	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTGAG	CTTCCTCTTCTTCTTGGGTTCGACTTCTTTTTCTGGTTCGATCAAG
L. thermotolerans	ACAACCTTCTTCATCTTCCGCTGCTGCTGCTATTAACGCTAATGG	CTTCCTCTTCTTCTTGGGTTCGACTTCTTTCAAGTGAGAGGACAGC
L. waltii	ACAACCTTCTTCATCTTCCGCTGCTGCTGCTATTAACGCTAATGG	CTTCCTCTTCTTCTTGGGTTCGACTTCTTTCAAGTGAGAGGACAGC
Mcm1 IDR 98-286	ACAACCTTCTTCATCTTCCGATGATGAGGAAGAAGACGAGGAGG	TCCTCTTCTTCTTGGGGTATTGGCCTTGTTGCGGTTC
Altered valcence 2	ACAACCTTCTTCATCTTCCTCTACTCCACAAGGTTATGGTTCTACTGG	CTTCCTCTTCTTCTTGGGCTGTTGACCATAGCTAGATTGTTCACTC
N. castellii	ACAACCTTCTTCATCTTCCTTGAACGTTCCAAAGAAGGTCAGAAGAAAGAGG	CTTCCTCTTCTTCTTGGGTTGGATGTCGTTCAAACCGGAAGTC
N. dairenensis	ACAACCTTCTTCATCTTCCTCTGACTTGGAAACTATGAACAACGCCG	CTTCCTCTTCTTCTTGGGTCTGGATCTCTTCAAGTGGTAGGACAAG
NCS21_hydro2polay_2wt	ACAACCTTCTTCATCTTCCAACGACAACAACAACTCTGATGGTG	CTTCCTCTTCTTCTTGGGTTCGACTTCTCTCAAATGTGGAGACAAC
Pho4 70-110	ACAACCTTCTTCATCTTCCGACAGTCTCGCGTTGGACG	CTTCCTCTTCTTCTTGGGCCGCGGCTTGATTGTAGCAG
Pho4 1-249	ACAACCTTCTTCATCTTCCATGGGCCGTACAACTTCTGAGG	CTTCCTCTTCTTCTTGGGATCGTCCACGAGCGCGC
Rap1 IDR 231-361	ACAACCTTCTTCATCTTCCATGTCCTCTCCAGATGATTTTGAAACTGC	CTTCCTCTTCTTCTTGGGTTGTTCTTTCTTTTCGTCAGCTGCAG
Rap IDR 1-120	ACAACCTTCTTCATCTTCCATGTCCTCTCCAGATGATTTTGAAACTGC	CTTCCTCTTCTTCTTGGGAGCCTTGTTATGAGATGGCAAAGC
Rap1 IDR 1-120 & IDR 231-361	ACAACCTTCTTCATCTTCCTTCAGAGAAGTTGTCGACTCCAGATTG	CTTCCTCTTCTTCTTGGGAGCCTTGTTATGAGATGGCAAAGC
Rap1 231-361 + Gal4 G4 motif	ACAACCTTCTTCATCTTCCTTCAGAGAAGTTGTCGACTCCAGATTG	CTTCCTCTTCTTCTTGGGAGCCTTGTTATGAGATGGCAAAGC
Reb1 1-420	ACAACCTTCTTCATCTTCCATGCCTTCAGGTCATAACGATAAAAACGC	CTTCCTCTTCTTCTTGGGATACTCGTTAATAAATCTTTCTAAAGCGGCGTC
Reb1 1-120	ACAACCTTCTTCATCTTCCAACAACAACGACGACTCTAACAGAAACG	CTTCCTCTTCTTCTTGGGCAGCTGATCTCTGTCGATAACATCATTG
Altered valcence 1	ACAACCTTCTTCATCTTCCTCTACTCCACAAGGTTATGGTTCTACTGG	CTTCCTCTTCTTCTTGGGTTGTTGGCCATAACTGCTCTGTTC
LS-9	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-10	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-11	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-12	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-13	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-14	ACAACCTTCTTCATCTTCCAACAACAACAACAACAATGACGGTG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC
LS-8	ACAACCTTCTTCATCTTCCTCCCATATGCATTTGAACGCCTC	CTTCCTCTTCTTCTTGGGAACGTGTTCTGGTCTTTCTTTAGCCTG
LS-15	ACAACCTTCTTCATCTTCCTCCCATATGCATTTGAACGCCTC	CTTCCTCTTCTTCTTGGGAACGTGTTCTGGTCTTTCTTTAGCCTG
S. kudriavzevii	ACAACCTTCTTCATCTTCCAACAATAACAACAACAACAATGACGACGAG	CTTCCTCTTCTTCTTGGGCTCGACTTCTTTCAAGTGAGAGGATAGC

Construct	Primer insert forward (5'-->3')	Primer insert reverse (5'-->3')
S. mikatae	ACAACCTTCTTCATCTTCCAATGGTGCCGATG ATGAATTGTCTAACACC	CTTCCTCTTCTTCTTGGGCTCGACTTCTTT CAAGTGAGAGGATAGC
Ssn6 68-204 + Gal4 G4 motif	ACAACCTTCTTCATCTTCCCAACAACAACAGC CAGCTCAAC	CTTCCTCTTCTTCTTGGGTTGTCTTTGTTG CTGCTGTTGTTGC
Sup35 1-131	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATCAAGGTAACAACC	CTTCCTCTTCTTCTTGGGCTGGAAGTCGTT CAAAGACATACCTTG
Sup35 1-131 + GR	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATCAAGGTAACAACC	CTTCCTCTTCTTCTTGGGCTGGAAGTCGTT CAAAGACATACCTTG
Sup35 1-131 + GR dist.	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATTACCAAGGTAACAACC	CTTCCTCTTCTTCTTGGGCTGGAAGTCATT CAAAGACATACCAGC
Sup35 1-131 + Gal4 G4 motif	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATCAAGGTAACAACC	CTTCCTCTTCTTCTTGGGCTGGAAGTCGTT CAAAGACATACCTTG
Sup35 1-131 + Gal4 G4 motif - all acidic residues	ACAACCTTCTTCATCTTCCATGTCTGGTTCCA ATCAAGGTAACAAC	CTTCCTCTTCTTCTTGGGCTGGAAACCGTT CAAAGACATACCTTG
Sup35 1-131 + p65	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATCAAGGTAACAACC	CTTCCTCTTCTTCTTGGGCTGGAAGTCGTT CAAAGACATACCTTG
Sup35 1-131 + p65 dist.	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATCAAGGTAACAACC	CTTCCTCTTCTTCTTGGGTTGGAAGTCGTT CAATGAGGACATACC
Sup35 1-131 + distributed EM	ACAACCTTCTTCATCTTCCATGTCCGACTCTA ATCAGGGTAACAATC	CTTCCTCTTCTTCTTGGGTTGGAAGTCGTT AGCCAAAGAC
S. uvarum	ACAACCTTCTTCATCTTCCGATAACAACGACG ACGAATTGTCCTCTGC	CTTCCTCTTCTTCTTGGGCTCGACTTCTTT CAAGTGAGAGGATAGC
synthetic_1	ACAACCTTCTTCATCTTCCTACCAACAACCAC AGCAAGAATCTTC	CTTCCTCTTCTTCTTGGGTTGAGATTGGCC AGATTGAGCACC
synthetic_2	ACAACCTTCTTCATCTTCCTCTCAGTCATCTA CTTTGGGTAACATGCAAC	CTTCCTCTTCTTCTTGGGTTGTGGTGGAGA TGGTTGAGACATCTGTTGC
T. blattae	ACAACCTTCTTCATCTTCCATCCCAATGGTTA ACTCCTCCGCTAAC	CTTCCTCTTCTTCTTGGGTTCGACATCCTT GTAATGAGAGGACAGC
T. delbruecki	ACAACCTTCTTCATCTTCCCAATCTTACACCA GACATACCAGAAACGC	CTTCCTCTTCTTCTTGGGCTCGACATCCTT GAAATGAGAGGATAGC
T. phaffii	ACAACCTTCTTCATCTTCCGGTAATGCTGTTG GTGTTAACGGTTC	CTTCCTCTTCTTCTTGGGCTCGTCCTTGAA ATGAGAGGATAGCAATC
Shuffle 1	ACAACCTTCTTCATCTTCCACCAACAACATCG AGTCTAAAAAGACCAAC	CTTCCTCTTCTTCTTGGGTTCGACGTAGTT GGACATGGTAGACAATC
Shuffle 2	ACAACCTTCTTCATCTTCCACCAAGAACATCC CACATAGGACTAACATTAAC	CTTCCTCTTCTTCTTGGG CTCGACTCTGACCTTGTTGATAACCAATTC
Shuffle 3	ACAACCTTCTTCATCTTCCCCACACAACTCCG AAAACAACAAAG	CTTCCTCTTCTTCTTGGGTTCGACGTTGTG ATGATCAGATGACAAG
V. polyspora	ACAACCTTCTTCATCTTCCCATCCACAACCTA GATATTCTAGAGCTGCC	CTTCCTCTTCTTCTTGGGCTCGACTTCTTT GAAGTGAGAGGATAGC
Z. rouxii	ACAACCTTCTTCATCTTCCCAACCTAGAGTTA GAAGAGAACCAGAAATTTTGGATGC	CTTCCTCTTCTTCTTGGGCTCGACATCCTT GAAATGAGAGGATAGC

Table S2: Primer used to generate the plasmid backbone for rational designs.

Construct	Primer backbone forward (5'-->3')	Primer backbone reverse (5'-->3')
General	GGAAGATGAAGAAGGTTGTAAATC	CCCAAGAAGAAGAGGAAGGTG
essential motif only	CTGCCGCTGCCGGAAGATGAAGAAGGTTGTAAATCATTATACTCGC	GAGTGGGAGTCCCAAGAAGAAGAGGAAGGTGGAC
FUS 1-163 12E	CATAATCGTTAGAGGCCATGGAAGATGAAGAAGGTTGTAAATCATTATAC	AGCAAAATCAGTATAACTCCCCCAAGAAGAAGAGGAAGGTG
FUS 1-163	GTATAATCGTTAGAGGCCATGGAAGATGAAGAAGGTTGTAAATCATTATAC	GCAAAATCAGTACAACTCCCCCAAGAAGAAGAGGAAGGTG
Gal4 IDR 768-881	CCACTTTGATTAAAATTGGCGGAAGATGAAGAAGGTTGTAAATCATTATAC	CACCAAACCCAAAAAAAGAGCCCAAGAAGAAGAGGAAGGTG
Gcn4 IDR 17-150	TAGAACCATCCAATGGTGAGGAAGATGAAGAAGGTTGTAAATCATTATACTCGC	CACTGAAGAAGTTTCTCTGCCCAAGAAGAAGAGGAAGGTGGACTATAAGG
K. lactis full length	CCTTGTGAAATTGCAAATATGGTG	CCCAAGAAGAAGAGGAAGGTG
LS-2	TGTTGGAGTTCTGGTTTTCGGAAGATGAAGAAGGTTGTAAATCATTATAC	TCACTTGAAAGAAGTCGAGCCCAAGAAGAAGAGGAAGGTG
LS-3	TGTTGGAGTTCTGGTTTTCGGAAGATGAAGAAGGTTGTAAATCATTATAC	TCACTTGAAAGAAGTCGAGCCCAAGAAGAAGAGGAAGGTG
LS-4	TGTTGGAGTTCTGGTTTTCGGAAGATGAAGAAGGTTGTAAATCATTATAC	TCACTTGAAAGAAGTCGAGCCCAAGAAGAAGAGGAAGGTG
LS-5	TGTTGGAGTTCTGGTTTTCGGAAGATGAAGAAGGTTGTAAATCATTATAC	TCACTTGAAAGAAGTCGAGCCCAAGAAGAAGAGGAAGGTG
LS-6	GTCATTGTTGTTGTTGTTGTTGGAAGATGAAGAAGGTTGTAAATCATTATAC	CTCACTTGAAAGAAGTCGAACCCAAGAAGAAGAGGAAGGTG
LS-7	GTCATTGTTGTTGTTGTTGTTGGAAGATGAAGAAGGTTGTAAATCATTATAC	AACCAGAAAAAGAAGTCGAACCCAAGAAGAAGAGGAAGGTG
NCS21_hydro2polar_2wt	TCAGAGTTGTTGTTGTCGTTGGAAGATGAAGAAGGTTGTAAATCATTATAC	CACATTTGAGAGAAGTCGAACCCAAGAAGAAGAGGAAGGTG
Pho4 70-110	TCCAACGCGAGACTGTCGGAAGATGAAGAAGGTTGTAAATCATTATAC	CTACAATCAAGCCGCGGCCCAAGAAGAAGAGGAAGGTG
Pho4 1-249	AGAAGTTGTACGGCCCATGGAAGATGAAGAAGGTTGTAAATCATTATAC	CGCGCTCGTGGACGATCCCAAGAAGAAGAGGAAGGTG
Shuffle 1	TTAGACTCGATGTTGTTGGTGGAAGATGAAGAAGGTTGTAAATCATTATAC	CATGTCCAACTACGTCGAACCCAAGAAGAAGAGGAAGGTG
Shuffle 2	TATGTGGGATGTTCTTGGTGGAAGATGAAGAAGGTTGTAAATCATTATAC	AACAAGGTCAGAGTCGAGCCCAAGAAGAAGAGGAAGGTG
Shuffle 3	GTTTTCGGAGTTGTGTGGGGAAGATGAAGAAGGTTGTAAATCATTATAC	TGATCATCACAACGTCGAACCCAAGAAGAAGAGGAAGGTG

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Dekanat Medizinische Fakultät
Promotionsbüro

Eidesstattliche Versicherung

Langstein-Skora, Iris

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

Functional determinants of the intrinsically disordered region 2 in the general regulatory factor Abf1 from Saccharomyces cerevisiae

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 27.07.2024
Ort, Datum

Iris Langstein-Skora
Unterschrift Doktorandin/Doktorand