Sequence- and chemical specificity define the functional landscape of intrinsically disordered regions

Iris Langstein-Skora[1], Andrea Schmid[1], Ryan J. Emenecker[2,3], Mary O.G. Richardson[3,4], Maximilian J. Götz[1,5], Sarah K. Payer[1,6], Philipp Korber[1]*, Alex S. Holehouse[2,3]*

[1] Biomedical Center (BMC), Division of Molecular Biology, Faculty of Medicine, LMU Munich, Martinsried, Germany.

[2] Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA

[3] Center for Science and Engineering Living Systems, Washington University, St. Louis, MO, USA

[4] Present address: Department of Molecular & Cellular Biology, Harvard University, Cambridge, Massachusetts, United States of America

[5] Present address: Gene Center and Department of Biochemistry, LMU Munich, Munich, Germany

[6] Present address: Proteros Biostructures GmbH, Martinsried, Germany

*Corresponding author. Email: pkorber@lmu.de, alex.holehouse@wustl.edu

Abstract:
Intrinsically disordered protein regions (IDRs) pervasively engage in essential molecular functions, yet they are often poorly conserved as assessed by sequence alignment. To understand the seeming paradox of how sequence variability is compatible with function, we examined the functional determinants for a poorly conserved but essential IDR. We show that IDR function depends on two distinct but related properties: sequence- and chemical specificity. While sequence-specificity works via linear binding motifs, chemical-specificity reflects the sequence-encoded chemistry of multivalent interactions through amino acids across an IDR. Unexpectedly, an apparently essential binding motif can be removed if compensatory changes to the sequence chemistry are made, highlighting the orthogonality and interoperability of both properties. Our results provide a general framework to understand the functional constraints on IDR sequence evolution.

One-Sentence Summary: Interactions driven by intrinsically disordered regions can be understood using a two-dimensional landscape that defines binding via motif-dependent and motif-independent contributions.

Main Text:

Intrinsically disordered proteins and protein regions (collectively referred to as IDRs) play important and often essential roles in many biological processes across all three kingdoms of life (*1*). IDRs frequently engage in molecular interactions, and their inherent structural plasticity allows them to bind through a variety of mechanisms (*2*, *3*). As such, understanding how IDR sequences determine the modes of molecular recognition is key to mapping from sequence to function.

Despite their importance for function, IDRs are often poorly conserved as assessed by alignment-based methods (*4–7*). The notable exceptions to this are short linear motifs (SLiMs), conserved stretches of 5 to 15 amino acids that define sequence-specific recognition sites (*8–10*). The modular nature of many SLiMs is demonstrated by their ability to mediate molecular recognition when inserted into otherwise neutral contexts (*10*, *11*). In addition to molecular recognition driven by SLiMs, IDRs that lack SLiMs can engage in distributed multivalent interactions driven by sequence-encoded chemistry (*3*, *12–15*). These interactions can mediate both stoichiometric protein-protein interactions or drive the formation of biomolecular condensates (*16–20*). For these distributed multivalent interactions, the precise sequence order is often unimportant, yet the amino acid composition and patterning are critical (*16*, *17*, *21*). These two modes of interaction are generally considered to drive orthogonal types of molecular recognition (*16*, *22*). With this in mind, a general model for IDR evolution has emerged in which substantial sequence variation can be tolerated assuming SLiMs are conserved or the amino acid composition and patterning (i.e. bulk sequence properties) are maintained (*5*, *23–25*).

Here, using *S. cerevisiae* as a model organism, we sought to push this model to its limits and uncover the determinants of function in an essential IDR. Our results suggest that rather than existing as two distinct modes of interaction, IDR-mediated molecular recognition should be considered on a two-dimensional landscape, whereby SLiMs cooperate with and can even be fully replaced by distributed multivalent interactions. Accordingly, apparently modular sequence motifs may impart function not by acting as *bona fide* SLiMs, but simply by altering the overall sequence chemistry. Our results imply that sequence context and the presence/absence of SLiMs can compensate for and buffer one another, providing a key missing piece in our understanding of the sequence constraints on IDR evolution.

Results

IDRs can evolve more rapidly than folded domains (*4*, *6*). We therefore anticipated that *S. cerevisiae* proteins may possess conserved folded regions alongside less well-conserved IDRs (**Fig. 1A**). To assess this, we performed a systematic analysis of sequence conservation and predicted disorder across the *S. cerevisiae* proteome (see *Methods*) (*26*, *27*). This analysis revealed pervasive disorder and confirmed an expected anti-correlation between conservation and disorder (**Fig. 1B**, **fig. S2-5**). Interestingly, essential proteins in *S. cerevisiae* are on average as disordered as non-essential proteins (**Fig. 1C**) (*28*). As such, "poorly conserved" IDRs are abundant in essential proteins, a result we interpreted to mean that alignment-based sequence conservation need not report on IDR importance. For this reason we hypothesized that poorly conserved IDRs from essential proteins may be functionally conserved nonetheless, e.g., by maintaining amino acid composition. Specifically, we define functional conservation here to mean an IDR from one species can functionally replace the orthologous IDR in another.

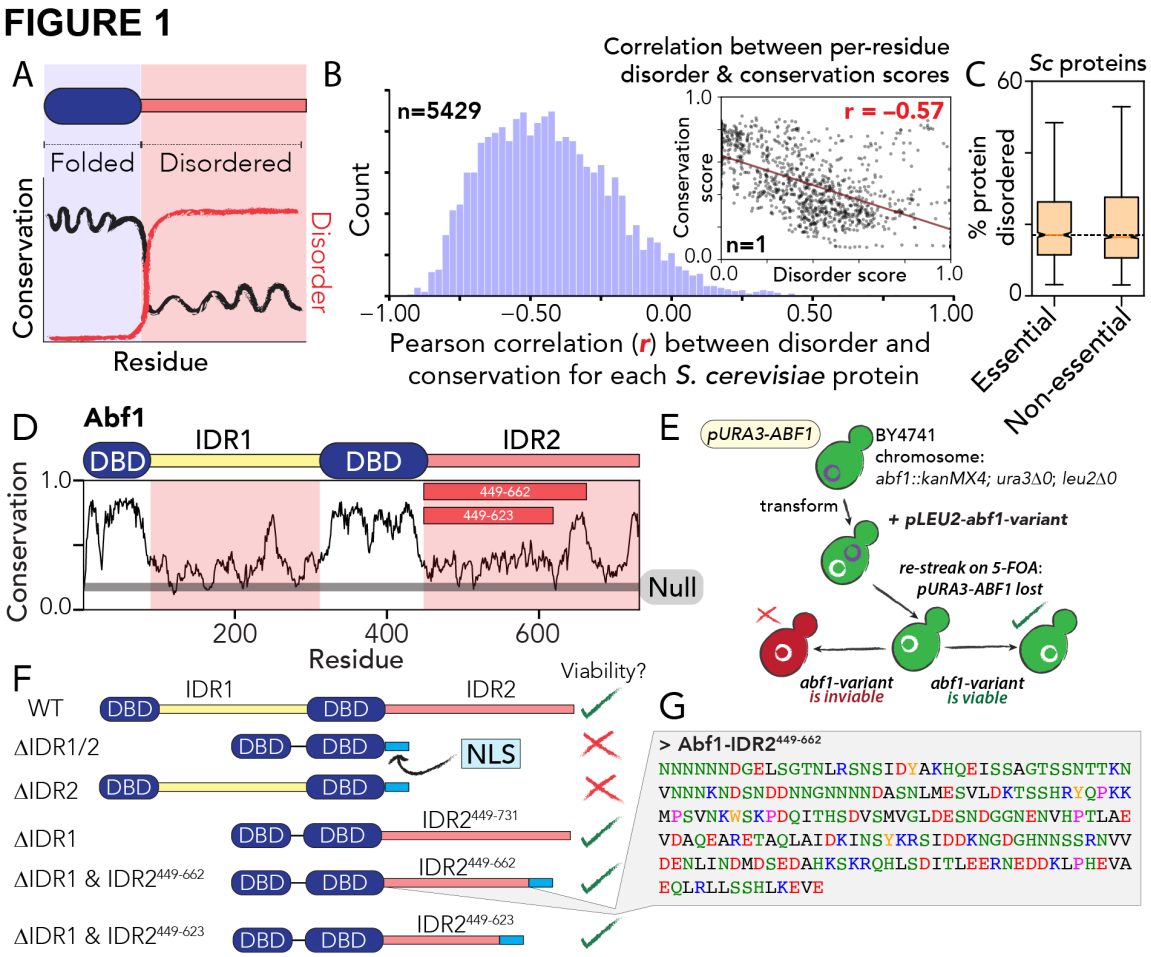

Fig 1: IDRs are poorly conserved as assessed by linear alignment, and Abf1 is an essential protein with a poorly conserved yet essential IDR. (A) Schematic showing conservation and disorder across a hypothetical protein. **(B)** Histogram of per-protein correlations (*r*) between per-residue conservation and disorder scores. Inset shows an example of a single protein with each marker representing one amino acid in this protein. The histogram reports on *r* values for the entire yeast proteome. **(C)** Percentage of the sequence defined as IDRs for essential *vs.* non-essential *S. cerevisiae* proteins. **(D)** Sequence analysis of Abf1 with per-residue conservation shown. The horizontal grey line corresponds to the conservation score expected for a randomly shuffled sequence (Null). **(E)** Schematic of plasmid shuffle assay. **(F)** Domain diagram for Abf1 mutants with their viability shown in the right-hand side column. NLS: SV40 nuclear localization signal **(G)** Abf1-IDR2$^{449\text{-}662}$ amino acid sequence, the focus of this study.

We sought to identify a model protein to test our functional conservation hypothesis. Because long IDRs are abundant in the context of chromatin-associated proteins (**fig. S6**) (*29*, *30*) we turned to the class of yeast general regulatory factors (GRFs). GRFs consist of sequence-specific DNA binding domains (DBDs) and long IDRs (**fig. S7**) (*31*), are abundant, and are mostly essential for viability. They are the yeast equivalent of "architectural factors", like mammalian CTCF (*32*, *33*), but with binding sites mostly in promoter regions, and they function as barriers/boundaries/insulators in organizing or regulating chromatin in various contexts (*34–42*). Although often subsumed under the category of "transcription factors" due to their strong effects on transcription, they are distinct from classical transcription factors and mostly do not harbor transactivation domains (*43*). In contrast to transactivation domains, and despite their numerous genetic and physical interactions (*44*, *45*), little is known about their mode of action besides the

general notion that they interact with a range of partners and modulate genomic processes by partitioning chromatin (*34*).

Abf1 is an essential *S. cerevisiae* GRF *(46)* that modulates nucleosome positioning and thereby chromatin accessibility at promoters (*35*, *37*, *39*). Rather than directly driving transcription, it functions as an insulator and mediates unidirectional transcription from potentially bidirectional promoters (*34*, *47*, *48*), as well as participating in roadblock termination of RNA polymerase (*49*, *50*). For our purpose, Abf1 has an ideal prototypical domain architecture, with two highly conserved folded DBDs and two poorly conserved IDRs (IDR1 and IDR2) (**Fig. 1D**). While the DBDs mediate sequence-specific DNA recognition, the IDRs are considered malleable protein-protein interaction modules, based on extensive physical and genetic interaction experiments (*44*, *51–54*).

We first established which Abf1 regions were essential for function. Previous studies on Abf1 focused on the DBD, although truncation experiments identified apparently essential C-terminal sequences (CS1/2) in IDR2 (*55*, *56*). We extended this truncation approach through a classical plasmid shuffling assay (*57*). In this assay, the chromosomal *ABF1* gene was deleted and a wildtype *ABF1* gene provided on a plasmid with *URA3* marker (**Fig. 1E**). Transformation with a plasmid bearing an *abf1* mutant gene and selection on 5-FOA plates for loss of the wildtype *ABF1* gene plasmid (5-FOA is toxic in the presence of *URA3*) assessed if the *abf1* mutant gene supported viability (see *Methods*, **fig. S1** and **Table S1**). As Abf1 is an essential protein, viability in this assay was a clear read out for the essential function of the respective Abf1 mutant protein. Importantly, for the rationally-designed inviable constructs, we performed ChIP assays (see *Methods*, **fig. S8**) to verify that these proteins (i) are expressed, (ii) enter the nucleus, and (iii) bind specifically to Abf1 DNA binding sites. This allowed us to ascribe inviable constructs in terms of a loss of essential IDR2 function, as opposed to spurious mislocalization, impaired DNA binding, or aberrant protein degradation.

Our molecular dissection revealed that IDR2 but not IDR1 is essential (**Fig. 1F**). The maximal C-terminal IDR2 truncation, IDR2$^{449\text{-}662}$ (**Fig. 1G**), that was reported to be viable in the presence of IDR1 (*56*) was also viable in the absence of IDR1 (**Fig. 1F**). We used IDR2$^{449\text{-}662}$ as our "wildtype" IDR reference sequence and simply refer to it as IDR2 for the remainder of the work. IDR2 is poorly conserved across all orthologs (**Fig. 1D**), yet essential for viability, at least in *S. cerevisiae*. The previously identified CS1/2 region that corresponds to a small island of conservation (peak in **Fig. 1D** at residue ~650) was previously shown to be a nuclear localization sequence (NLS) (*54*). We confirmed that this region was not strictly necessary if a heterologous SV40 NLS was included (**Fig. 1F**, bottom). To avoid scoring effects on nuclear localization, we included the SV40 NLS in all *abf1* mutant constructs.

Based on prior work, we anticipated that conservation in IDRs could be considered in terms of compositional and linear sequence conservation (*5*, *16*, *18*, *58*). Using this framework we can identify proteins that are well conserved in terms of linear sequence (and hence composition), by composition alone, or by neither (**Fig. 2A**). A comprehensive analysis of conservation across *S. cerevisiae* IDRs and folded domains confirmed that many IDRs, despite being poorly conserved in terms of linear sequence, are well-conserved in terms of composition (**Fig. 2B, fig. S9-S10, Table S12)**. Importantly, this analysis revealed that IDR2 is more conserved in terms of charged and hydrophobic residues than most IDRs with similar degree of sequence conservation (**Fig. 2B, fig. S9**). In contrast, polar residue composition appeared less well conserved.

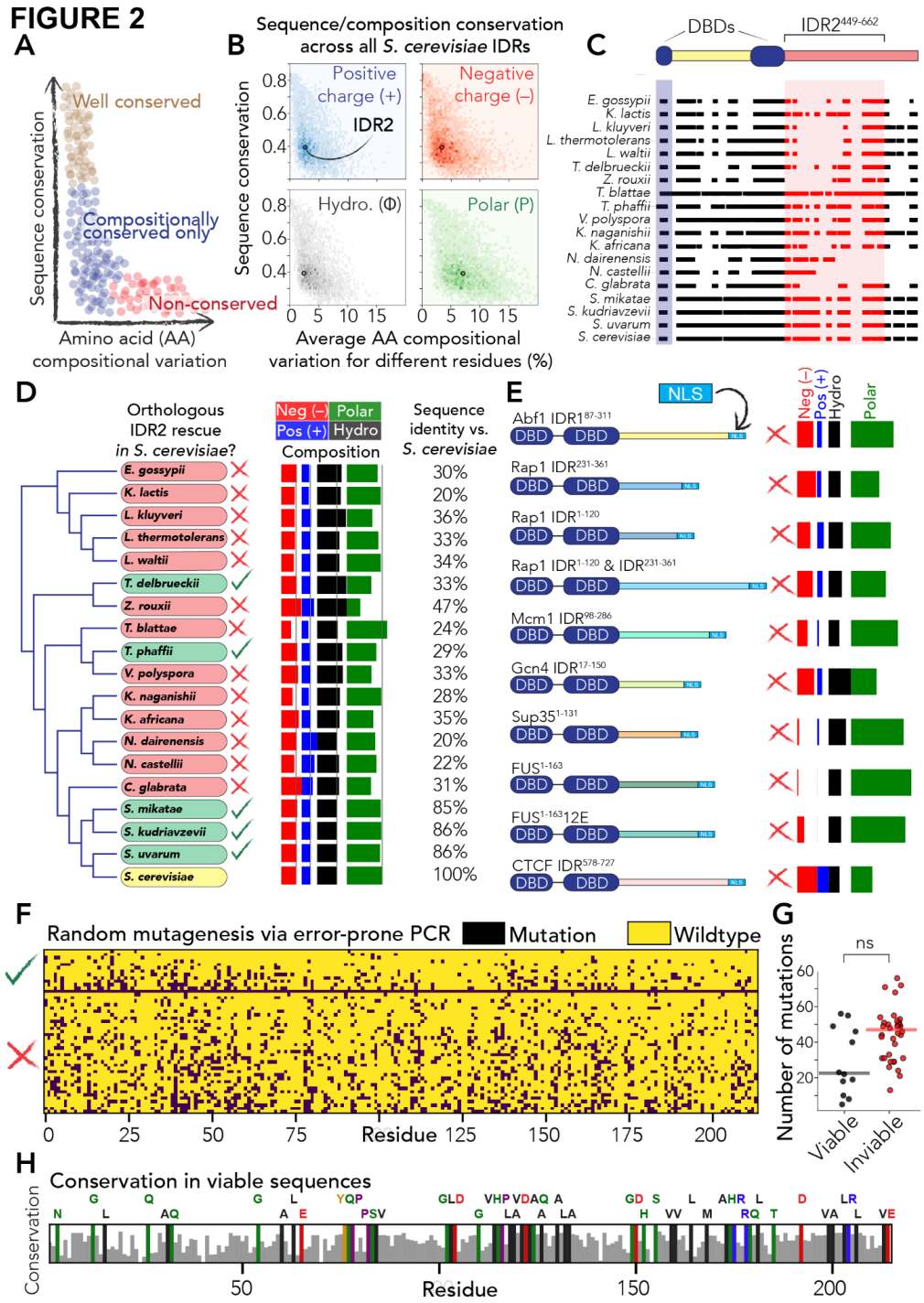

Fig 2: IDR2 shows relatively conserved composition, yet most orthologous sequences cannot rescue function. **(A)** Schematic of compositional and sequence conservation. **(B)** An analysis of all *S. cerevisiae* IDRs reveals Abf1-IDR2 (colored dot) is relatively conserved in terms of charge and hydrophobicity composition, but less so with respect to polar composition. See also **fig. S9**. **(C)** Schematic of orthologous IDR2s shown across full sequence alignment. **(D)** Phylogenetic tree (for whole organism) vs. IDR2 rescue ability, composition (vertical grey lines mark *S. cerevisiae* IDR2 composition), and sequence identity compared to *S. cerevisiae* IDR2. Most IDR2 orthologs do not support viability in *S. cerevisiae*. **(E)** Alternative IDRs tested including their amino acid composition. **(F)** Random mutagenesis of IDR2 with viable(top) and inviable (bottom) mutants. Each row is a separate sequence,

each column a residue along the sequence. **(G)** Comparison of mutational burden in viable vs. inviable sequences (ns, not significant). **(H)** Conservation analysis for viable sequences, with conserved residues (residues that are never mutated in any of the viable sequences) highlighted above.

We assumed that the relative compositional conservation of IDR2 would provide the molecular basis for our functional conservation hypothesis and anticipated that orthologous IDRs would support viability in *S. cerevisiae*. To test this, we took eighteen Abf1 orthologs, identified their IDRs corresponding to *S. cerevisiae* IDR2$^{449\text{-}662}$ from the full-length proteins (**Fig. 2C**), and replaced the *S. cerevisiae* IDR2$^{449\text{-}662}$ in our test plasmid with each of these orthologous IDR2s. We then tested the resulting chimeric constructs in our plasmid shuffle assay (**Fig. 2D**). Unexpectedly, outside of the *sensu stricto S. cerevisiae* complex (the bottom four species in **Fig. 2D**), only two of the fifteen orthologous IDR2s were viable (**Fig. 2D**), with no obvious relationship between sequence composition, sequence identity, or sequence length and function (**Fig. 2D, fig. S11**). To our surprise, our expectation that compositionally conserved IDRs would be functionally conserved proved incorrect.

We next wondered whether IDRs from proteins with similar functions might confer viability. We tested several candidates with similar IDR amino acid compositions, including IDRs from Abf1 (IDR1), other GRFs (Rap1, Mcm1), a yeast transactivator (Gcn4), and a human insulator (CTCF) (**Fig. 2E, fig. S12**). We also tested unrelated but compositionally similar low-complexity IDRs from the human RNA binding protein FUS and the yeast translation termination factor Sup35 (**Fig. 2E, fig. S12**). All of these IDRs failed to confer viability, suggesting that Abf1's IDR2 provides specific molecular recognition.

To further understand how variations in IDR2 influence function, we generated a library of 48 randomly mutagenized IDR2 variants (**Fig. 2F**, **fig. S13-S14**). Surprisingly, this analysis simultaneously revealed remarkable robustness with respect to some mutations and sensitivity with respect to others. A variant with 56 point mutations was viable and another with 13 was not. Statistically speaking, in the limit of our sample sizes, the mutational burden is not strongly predictive of viability (**Fig. 2G, fig. S15**), and a linear conservation analysis reveals "conserved" residues distributed approximately evenly across the sequence (**Fig. 2H**). Paradoxically, our results thus far imply that IDR2 is (i) simultaneously robust and sensitive to mutations, and (ii) compositionally relatively well-conserved yet cannot be replaced by most orthologs.

Although most orthologous IDR2s could not rescue viability, we unexpectedly identified several IDRs from other *S. cerevisiae* proteins that conferred viability in place of IDR2 (**Fig. 3A**). These included compositionally similar IDRs from the yeast transactivators Gal4 and Pho4, and from the GRF Reb1 (**fig. S16**). Gal4, Pho4, and Reb1 are DNA binding proteins that can trigger chromatin opening *in vivo* (*35–40*, *59–61*). These results illustrate that, perhaps paradoxically, while many orthologs and nearly identical IDRs are inviable (**Fig. 2D, F**), there exist IDRs that differ substantially in length and linear sequence that are viable.

Considering Gal4, Pho4, and Reb1 can mediate chromatin remodeling (one of Abf1's functions), we wondered if a common SLiM for recruiting the requisite machinery may be shared across these IDRs. Given SLiMs depend on their specific linear sequence (*8*), we reasoned that shuffling a SLiM would disrupt its function. As an initial test, we designed three globally shuffled variants with the conserved positions in **Fig. 2H** held fixed (**Fig. 3B**). All global shuffles were inviable, demonstrating unequivocally that IDR2-like composition is insufficient for viability, implicating linear sequence-specific regions that must be essential. In support of this inference, almost all of the composition-dependent sequence features are calculated for the inviable vs. viable sequences in **Fig. 2F** were statistically indistinguishable from one another

(**fig. S13**). Taken together, our evolutionary comparisons and random mutagenesis all paint a picture in which composition may matter, but is not sufficient for function.

These analyses – and especially the global shuffle variants – implied the presence of a non-conserved motif. To identify this putative motif we developed an unbiased approach termed sequential sequence shuffling (**Fig. 3C**). IDR2 was subdivided into non-overlapping 30-residue windows, and the sequence in each window was locally shuffled. This revealed two central windows that did not tolerate shuffling, which we confirmed by shuffling everything except the central 60-residue subsequence (**Fig. 3C**). We then repeated the procedure using 10-residue windows within the 60-residue subsequence. We identified a 20-residue subsequence (the essential motif, EM) that could not be shuffled and was essential for IDR2 function (**Fig. 3C**). Gratifyingly, the essential motif overlaps with the most conserved region in our random mutagenesis (**Fig. 3D**). While this region is unremarkable with respect to other sequence properties and not conserved across orthologs, it is predicted to form a transient helix, a feature frequently associated with IDR-mediated interactions (**fig. S17**) (*22*, *62*).

Given modular SLiMs should confer function when inserted into a neutral context, we tested if the essential motif met this definition. As our neutral context, we selected the phosphomimetic variant of the low-complexity IDR from the human RNA binding protein FUS (FUS12E) (*63*) (**fig. S18**). FUS12E is a compositionally-uniform low-complexity disordered region that lacks secondary structure or known binding motifs (*63*, *64*). However, FUS12E contains uniformly spaced hydrophobic (aromatic) and acidic residues, making it an ideal neutral IDRs with similar sequence properties to IDR2. While FUS12E alone was inviable (**Fig. 2E**), insertion of the essential motif into the FUS12E context "transformed" the FUS12E sequence to make it viable. This result confirmed the validity of the essential motif as a true modular SLiM: it cannot be shuffled and can confer function when placed in another context (**Fig. 3E**).

Might regions homologous to the essential motif exist in the other functional IDRs identified in **Fig. 3A**? A global sequence alignment between IDR2 and $Gal4^{768\text{-}881}$ was relatively poor, with only one sub-region showing alignment (**fig. S19**). Despite its low identity, this alignment revealed two remotely homologous regions, which we named $Abf1^{G4\text{-}like}$ and $Gal4^{G4}$ (**Fig. 3F**).

We wondered if the $Abf1^{G4\text{-}like}$ or $Gal4^{G4}$ subsequence contained *bona fide* modular SLiMs. To test this, as with the essential motif, we inserted $Abf1^{G4\text{-}like}$ or $Gal4^{G4}$ into the FUS12E context where both conferred viability (**Fig. 3G**). To confirm this result we tested the $Gal4^{G4}$ in three more otherwise inviable IDR contexts (**Fig. 3G**). Convincingly, in each case the 17-residue $Gal4^{G4}$ subsequence conferred viability.

How could all three of these quite different subsequences confer viability? If the essential motif includes a hydrophobic-faced transient helix (**fig. S17**), we speculated that an alternative hydrophobic-faced transient helix might also work. Accordingly, we designed an IDR2 alternative by inserting the 24-residue transient helix from the human RNA binding protein TDP-43 (**fig. S20**) into FUS12E (*65*, *66*). This rationally designed chimeric protein was viable, and suggested that we have uncovered at least one determinant of Abf1 IDR2 function, i.e., a binding motif that consists of a transient helix with hydrophobic interface (**Fig. 3H**).

While the essential motif, $Abf1^{G4\text{-}like}$, and $Gal4^{G4}$ appeared to confer viability, what (if any) is the role of the broader sequence context? To answer this question, we designed variants of FUS12E + $Gal4^{G4}$ where the contextual aromatic residues were converted to either serine or leucine (**Fig. 3I**). These aromatic residues are important for FUS-dependent IDR interaction in other systems, where they function via distributed multivalent interactions (*17*, *64*). Despite the presence of

Gal4^{G4}, these variants were inviable (**Fig. 3I**). Furthermore, if Gal4^{G4} was inserted into a glutamine-rich IDR from the yeast transcriptional co-repressor Ssn6, this variant was also inviable (**Fig. 3I**). Together, these results illustrate that the IDR context can also play an essential role in determining function.

Given the importance of context for Gal4^{G4} (**Fig. 3I**), we wondered if context mattered for other motifs. While context shuffling was tolerated in wildtype IDR2 (**Fig. 3C**), a rationally designed IDR2 variant with reduced hydrophobicity outside of the essential motif was inviable, as was a serendipitous mutant generated in our random mutagenesis where the essential motif was unaltered (**Fig. 3J**). Similarly, for our rationally designed FUS12E+TDP-43 variant, the aromatic residues in the context were essential for viability (**Fig. 3J**). Collectively, our results confirm that appropriate context chemistry is essential, revealing a second determinant of Abf1 IDR2 function.

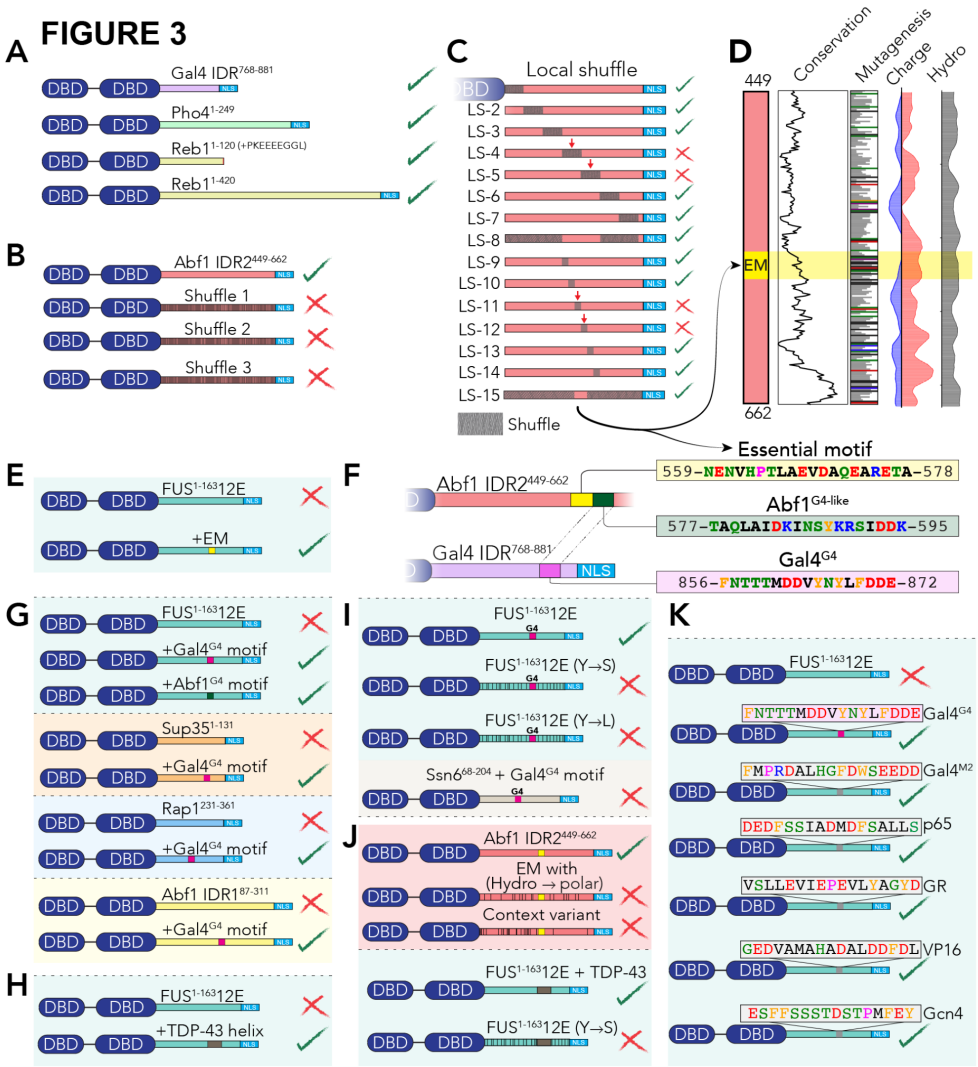

Fig 3: Sequence motifs play a key role in IDR2 function. (A) While most orthologs are inviable, several entirely unrelated IDRs can confer viability. **(B)** Sequence shuffles with 'conserved' positions identified by random mutagenesis (**Fig. 2H**) fixed are inviable, demonstrating that even with the protection of potentially conserved residues, IDR2 composition alone is insufficient for viability. **(C)** Sequential sequence shuffling pin-points an

essential motif (EM) in the center of IDR2. **(D)** The essential motif is not conserved across orthologs, is indistinct in terms of sequence features, yet it emerges as the most conserved subsequence in our random mutagenesis (solid bars in mutagenesis subfigure). **(E)** Insertion of the essential motif into a non-functional IDR (FUS12E$^{1-163}$) transforms that IDR to be functional. **(F)** Three putative motifs are shown in the context of their IDRs. Abf1$^{G4\text{-}like}$ and Gal4^{G4} were identified by sequence alignment of IDR2 and Gal4$^{768-881}$ (**fig. S21**). **(G)** Abf1$^{G4\text{-}like}$ and Gal4^{G4} confer viability introduced into FUS12E. **(H)** The transient, hydrophobic helix from TDP-43 also confers viability inserted into FUS12E. **(I)** The FUS12E IDR context with Gal^{G4} present is a key determinant of function. **(J)** The IDR2 sequence context around the essential motif is also a key determinant of function. **(K)** Compositionally matched subsequences taken from a range of transcription factors also provide viability inserted into FUS12E.

To confirm that Abf1$^{G4\text{-}like}$ and Gal4^{G4} contained *bona fide* SLiMs, we reasoned that an essential control experiment would be to take unrelated but length-matched subsequences with Abf1$^{G4\text{-}like}$/Gal4^{G4} -like amino acid compositions and demonstrate that these were inviable. We identified five subsequences that were compositionally similar to Gal4^{G4} from both yeast and non-yeast transcription factor IDRs (**fig. S21**). If IDR2 function is SLiM-dependent, then these non-alignable subsequences from another species should be inviable, given that - other than composition - they were randomly selected. To our surprise, all six of these 17-residue subsequences were viable in the FUS12E context (**Fig. 3K)**.

This result prompted us to step back and reconsider our data. Conventionally speaking, the ability to insert a short (<20 residue) sequence into a non-functional IDR context and confer function is interpreted as a simple and unambiguous demonstration of a *bona fide* modular motif (*e.g.,* a SLiM). Given sequence-specific motifs are often defined by two characteristics (an inability to tolerate shuffling and autonomous modular activity), the essential motif is clearly a true motif (**Fig. 3C**, **Fig. 3E**). However, the finding that subsequences that were compositionally similar but unrelated in terms of the absolute sequence were also functional implied that we either had a remarkable ability to identify *de novo* motifs, or that something more general was at play.

Our results thus far identified two determinants of function: (i) the presence of a motif, and (ii) the presence of a sequence context that we interpret to mediate distributed multivalent interactions (*16*, *17*, *67*) as hydrophobic residues were important (**Fig. 3I,J**). Generally, these two modes of interaction are discussed separately. Motifs are considered in the context of specific stoichiometric interactions, while distributed multivalent binding is mostly associated with biomolecular condensates (*10*, *16*, *18*, *22*). However, our results prompted us to wonder if these two interaction modes might instead exist on a combined two-dimensional landscape (**Fig. 4A)**.

To guide our intuition, we performed simple coarse-grained simulations to quantify 1:1 binding for an IDR with its partner as a function of motif and context strength (**Fig. 4B, D**, **fig. S22**). The landscape illustrates that a bound state can be achieved via many combinations of motif and context binding strengths. Sequence changes can alter the context (**Fig. 4B,** from 1-to-2), alter the motif (**Fig. 4B,** 1-to-3), or alter both. Accordingly, we sought to test this conceptual framework via further rational sequence design.

Given motifs are - by definition - sequence-specific, i.e., they depend on their contiguous linear amino acid sequence, it should not be possible to distribute a motif across the IDR context and maintain function. In keeping with this, a variant with the essential motif residues distributed across the FUS12E context was inviable (**Fig. 4C**). This variant can be interpreted as simultaneously disrupting the motif but also (modestly) enhancing the context through, for example, the hydrophobic residues that are redistributed (**Fig. 4D**, 1-to-2-to-3).

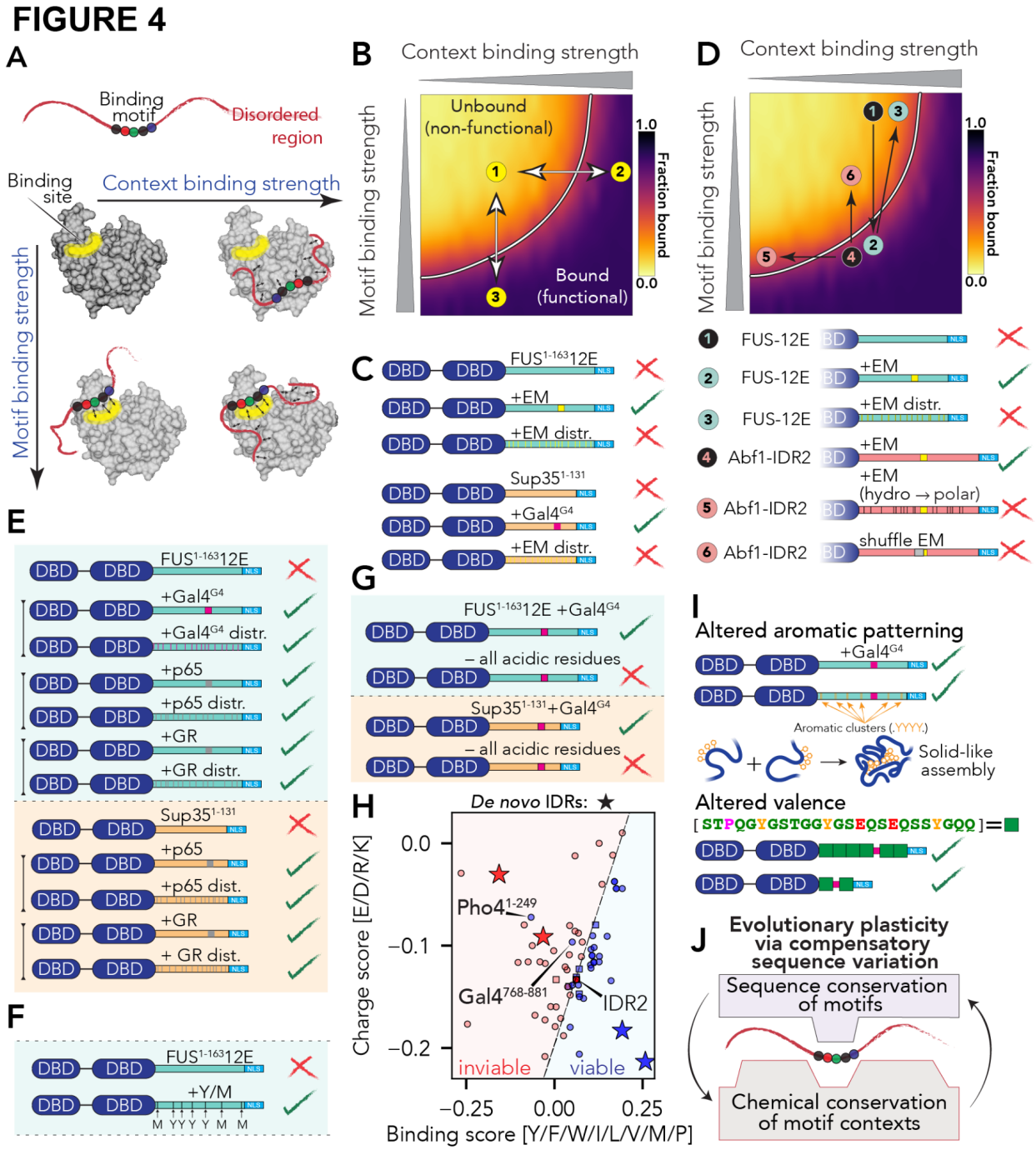

Fig 4: IDR2 mediated interactions can be understood via a two-dimensional binding landscape. **(A)** IDR-mediated interactions can be understood in terms of motif binding and context binding. **(B)** The combination of motif and context binding can be projected onto a simple two-dimensional binding surface. **(C)** The essential motif is a true motif, in that it cannot be distributed across the IDR sequence. **(D)** Previous designs can be interpreted through the two-dimensional binding landscape. **(E)** Variants with distributed motifs identified by composition are functional in both FUS12E and Sup35 backgrounds. **(F)** Rational design of a motif-free FUS12E variant. **(G)** Sufficient acidity in the IDR context is essential for viability. **(H)** Viable and inviable sequences can be classified based on charge and binding scores, parameters based on the weighted sequence composition. Circles are sequences that lack an essential motif (or TDP-43 motif), squares are sequences that have an essential motif (or TDP-43 motif),

and stars are completely synthetic sequences designed to titrate the space (**fig. S25**). Analogous plot for sequences generated by random mutagenesis shown in **fig. S23**. **(I)** Rational designs that should disrupt phase separation are viable, suggesting liquid-liquid phase separation may not play a role in Abf1 function. **(J)** Summary schematic (to be updated).

Our binding landscape model raised an intriguing possibility: What if Gal^{G4} and the other sequences identified in **Fig. 3K** were not *bona fide* motifs, but instead altered the IDR context, albeit very locally, without being a true motif? To test this, we asked if variants where these sequences were distributed were viable. In all cases and over multiple distinct contexts we discovered that these distribution variants were viable, confirming our hypothesis (**Fig. 4E**).

The functional yet motif-free sequences identified in **Fig. 4E** prompted a rational *de novo* design based on basic chemical principles. Given removing hydrophobic residues from contexts abolished viability (**Fig. 3I, J**) and given $Gal4^{G4}$ and the other motifs in **Fig. 4E** must function by modulating the context, we reasoned this likely occurs through an increased number of hydrophobic residues. Accordingly, we asked if a FUS12E variant with additional evenly distributed hydrophobic residues (+4 tyrosine (aromatic), +3 methionine (aliphatic), as in $Gal4^{G4}$) would be viable. Indeed, even though this design was wholly artificial and even though wild-type IDR2 requires a *bona fide* motif, this design was viable (**Fig. 4F**). This result strikingly confirms that SLiM-dependent function and function conferred by sequence chemistry can coexist on the same set of axes.

Finally, we asked if hydrophobicity was the only chemical feature in the IDR context that mattered. Our evolutionary analysis implicated that both hydrophobicity and acidity were similarly conserved (**Fig. 2B**). Accordingly, we tested if acidic residue depletion would compromise viability, which it did (**Fig. 4G**). Based on these principles, we developed a simple compositional-based metric that quantifies IDR sequences in terms of charge and binding score (see *Methods*). These two parameters effectively delineated 88 sequences, where almost the only examples of functional sequences that fall inside the inviable region possess *bona fide* motifs (**Fig. 4H** squares, **fig. S24**). To test the predicted boundaries, we designed a series of completely synthetic IDRs that straddle charge and binding scores (**Fig. 4H** stars**, fig. S25**) and confirmed the predictive power of our compositional-based metric for context viability. In summary, our results support a model in which two orthogonal axes (sequence-specific motifs and distributed multivalent sites) define a two-dimensional landscape in which sequence-to-function mapping in IDR2 can be interpreted.

Recent work has invoked intracellular phase transitions (and specifically liquid-liquid phase separation, LLPS) to explain molecular principles underlying chromatin organization (*68–70*). While FUS12E is highly soluble alone, it can phase-separate with an appropriate partner (*63*, *64*). We therefore wondered if our results could be interpreted though a phase transition model. To address this we followed two orthogonal lines of inquiry. Firstly, clustering of aromatic residues can suppress LLPS and promote aggregation (*16*, *71*, *72*). We re-designed the FUS12E context to cluster aromatic residues while leaving $Gal4^{G4}$ unperturbed (**Fig. 4I, top, fig. S26**). This variant remained viable, providing the first hint that LLPS may not be relevant. Secondly, *bona fide* LLPS of flexible polymers (such as IDRs) is unavoidably dependent on polymer length (*16*, *18*, *73*). As such, a shorter sequence with fewer aromatic residues should suppress assembly through a depletion of valence (*16*, *17*). We designed synthetic, homopolymeric repetitive FUS12E-like IDRs with a single $Gal4^{G4}$. Both 167-residue and 67-residue variants were viable, a pair of IDR lengths that is challenging to reconcile with an LLPS-based mechanism given the

length-dependence of assembly (which predicts a 3-5 order-of-magnitude difference in the molar saturation concentration for these two lengths (**fig. S27**). As such, while we cannot rule out the role of LLPS in this system, we find no strong evidence supporting it.

Discussion:

IDR-mediated interactions have generally been viewed through the lens of either sequence-specific binding motifs (*e.g.,* SLiMs) or distributed multivalent interactions. These interaction modes are determined by sequence-specificity and chemical specificity, respectively, and elegant work from many groups established the functional importance of both (*15*, *19*, *22*, *74*, *75*). Here we uncover the surprising result that – at least in Abf1 – these two modes can be compensatory for one another. A weak context could be compensated by introducing a motif (e.g., **Fig. 3G**) and, more surprisingly, absence of a motif could be compensated by gain of context strength (e.g., **Fig. 4D**, 2-to-3, **Fig. 4E**). Importantly, we show that testing motif shuffling/distribution is essential to identify true motifs/SLiMs.

Our results can be rationally interpreted via a two-dimensional binding landscape (**Fig. 4B**). In this model, the determinants of function reflect how IDR2 engages in intermolecular interactions via some interoperable combination of motif-dependent and context-dependent binding. Based on our molecular understanding, we were able to design *de novo* synthetic IDRs that were functional, although dramatically different and wholly unrelated to the wildtype sequence. These include variants with an established helical binding motif and multiple variants without motifs. The latter multiplicity makes it highly unlikely that we involuntarily generated a new motif each time and is all-the-more surprising as wild-type IDR2 depends on a specific motif. We note that biophysical hints for this model are found in numerous *in vitro* reconstituted systems (*3*, *19*, *20*, *76–78*).

Our divergent *de novo* designs demonstrate the depth of our understanding of the functional determinants for Abf1 IDR2 (**Fig. 3, 4**). They also highlight the power of interpreting IDR-mediated interactions via our two-dimensional landscape. We suggest that many IDRs studied previously can be placed somewhere on the landscape, although the boundaries for function will naturally vary in a system-specific way. Finally, to help other groups identify essential subregions, we developed a computational tool for constructing sequential sequence shuffle libraries (https://pipit.readthedocs.io/).

Functions based on molecular interactions generally involve some degree of specificity, both towards a partner of interest as well as against off-target inhibitory interactions. Specificity is typically considered in terms of shape complementarity and chemical compatibility, two features afforded by folded domains and, to a lesser extent, SLiMs (*79*, *80*). Here we find that even for an IDR that depends on a *bona fide* motif, there can exist alternative IDRs where function is upheld by rationally interpretable chemical features in the absence of any motif. We speculate that this chemical specificity offers an evolutionarily plastic mode of molecular recognition (**fig. S28**) (*18*, *19*, *58*).

In light of this evolutionary leeway, the limited functional conservation in IDR2 across orthologs despite the conservation of amino acid composition and sequence features may seem surprising (**Fig. 2, fig. S29**). To verify that full-length Abf1 performs analogous functions in other species, we confirmed that full-length Abf1 from *K. lactis* is viable in *S. cerevisiae* (**fig. S1**) (*81*). With this in mind, we emphasize that our study is focussed on IDR sub-regions orthologous to

IDR2$^{449-662}$, and not orthologous full-length proteins. Therefore, there are several (non-mutually exclusive) explanations for the inviability of orthologs. Firstly, interaction networks co-evolve by coupled evolutionary changes, such that SLiMs in orthologs may be incompatible with partners found in *S. cerevisiae*. An important feature presently absent is the mapping of the interactome for wildtype and Abf1 variants. Identification of the interaction partners is ongoing and will provide insight into the molecular basis for chemical specificity. Secondly, functionally important features can relocate across an IDR-containing protein. In this model, the specific location of a binding motif in the protein may be relatively unimportant, such that motifs can be lost from one region and emerge in another. An intriguing prediction from this model is that we should in fact *expect* motifs to rapidly appear and disappear from a given IDR, a prediction supported and compatible with previous work on *ex nihilo* motif evolution (*82*).

While the essential motif appears poorly conserved, we still expect evolutionarily-conserved SLiMs to be important and ubiquitous across proteomes. In line with this expectation, we identified thousands of short and conserved hydrophobic subsequences within IDRs, with almost twice as many conserved hydrophobic subsequences in essential proteins as non-essential ones (**fig. S30, S31, tables S7-S11**). As a corollary, we wondered if other non-conserved regions of transient structure may be found in the yeast proteome, and upon analysis identified 963 short (<40-residue) subregions in IDRs with predicted transient structure (**Table S14**). Included in this set of *de novo* predictions is the previously identified Pho4 activation domain (Pho4$^{69-94}$) and four separate subregions in the N-terminal IDR of Reb1. While many such regions may be inert, others may offer specific binding interfaces, as is the case in Abf1, or specific helical regions identified in transactivation domains (*83*, *84*). In summary, this analysis offers specific, testable predictions for putative functional modules in IDRs at a proteome-wide scale.

Our rational mutagenesis reveals IDR2 is remarkably sensitive to small perturbations (**Fig. 2G**), even though much larger sequence changes offer alternative and functional variants. We interpret the fact that IDR2 is on the edge of viability as a signature of sensitivity. If IDR2 were deep inside the 'bound' regime, tuning molecular function via PTMs or additional partners may be challenging. Instead, by sitting close to the conceptual midplane on our landscape between bound and unbound, the wildtype sequence is maximally sensitive (**Fig. 4D**).

We focused on function measured by viability in *S. cerevisiae* growing under low-challenge laboratory conditions. This specific growth niche likely does not assess all facets of Abf1 function. The importance of other Abf1 regions and features in alternative growth conditions remains unassessed so far. How might alternative motifs or sequence features contribute to IDR function in Abf1? Ongoing work implies large-scale IDR-dependent remodeling of transcription during glucose starvation, and IDRs have been proposed to function as intrinsic sensors of intracellular state (*85*, *86*). If chemical specificity is defined by IDR chemistry, mechanisms to tune this chemistry (either via post-translational modifications, changes in protonation state, or changes in sidechain solvation properties) offer an attractive route for modulating specificity.

Finally, intracellular phase transitions (and notably LLPS) emerged over the last decade as key principles to explain cellular organization (*87*). While we cannot exclude an LLPS-based Abf1 function, we do not find compelling evidence to support it (**Fig. 4I**). Low-complexity sequences with clustered aromatics have been shown to undergo self-assembly and form irreversible, solid-like condensates *in vitro* and *in vivo* (*16*, *71*, *72*, *88*). Our synthetic IDR2 with clustered aromatic residues is viable, a result that further supports our inference that Abf1 interacts with

other proteins, rather than with itself (**Fig. 4I**). Distributed multivalent interactions were previously interpreted by us and by others as a signature of proteins that were under selection to form biomolecular condensates (*16*, *18*). As an additional (and non-mutually exclusive) explanation, distributed multivalent interactions may offer a convenient platform for protein-mediated interactions, potentiating additional sequence-specific interactions driven by folded domains or SLiMs (*3*, *12*, *89*, *90*). If this is correct, phase separation would be an unavoidable but perhaps relatively benign consequence of these multivalent interactions (*71*, *91*). We emphasize that there are also many examples in which intracellular phase transitions offer compelling functional advantages (*74*, *87*, *92–97*). Delineating functionally important phase transitions from situations in which it is an unavoidable outcome of multivalent proteins is a major and open challenge for the field of cell biology.

Acknowledgments: We thank Rohit V. Pappu for his immediate willingness to collaborate once approached by P.K. and allowing A.S.H. to work independently on this project while completing his postdoctoral work. We thank Rahul Das for his help to jump-start this project and Benoit Kornmann for helping M.O.G.R. with handling his SATAY data at an early stage of the project. We thank Shahar Sukenik, David Moses, and Broder Schmidt for critical comments and feedback on the manuscript. We are grateful to Axel Imhof, Christoph Kurat and Tamas Schauer for their input as thesis advisory committee members for I.L.-S. and especially Axel Imhof for his suggestion to use the ChIP assay as control. We acknowledge funding by the German Research Foundation (grants KO 2945/3-1 and within SFB1064) to P.K., the Longer Life Foundation (an RGA/Washington University Collaboration) to A.S.H., and the MOLSSI via a Seed Fellowship to A.S.H.

Funding:

Longer Life Foundation (ASH)

MOLSSI Seed Fellowship (ASH)

William H. Danforth Foundation Fellowship (RJE)

German Research Foundation (KO 2945/3-1) (PK)

German Research Foundation (within SFB1064) (PK)

Author contributions:

Conceptualization: PK, ASH

Methodology: PK, ILS, ASH, AS, RJE, MOGR

Investigation: PK, ILS, ASH, AS, RJE, MJG, SKP, MOGR

Visualization: ILS, PK, ASH

Funding acquisition: PK, ASH

Project administration: PK, ASH

Supervision: PK, ASH

Writing – original draft: ILS, PK, ASH

Writing – review & editing: ILS, PK, ASH, RJE

Competing interests: ASH is a scientific consultant for Dewpoint Therapeutics. All other authors declare no competing interests.

Data and materials availability: All code and data are available in the main text, supplementary material, or online at https://github.com/holehouse-lab/supportingdata/tree/master/2021/Langstein-Skora_2021. The Python package PIPT (https://pipit.readthedocs.io/) allows for sequence libraries done for sequential sequence shuffling to be generated automatically.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S31

Tables S1 to S14

References (*1–35*)

References

1. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).

2. M. M. Babu, R. W. Kriwacki, R. V. Pappu, Structural biology. Versatility from protein disorder. *Science*. **337**, 1460–1461 (2012).

3. P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).

4. C. J. Brown, A. K. Johnson, G. W. Daughdrill, Comparing models of evolution for ordered and disordered proteins. *Mol. Biol. Evol.* **27**, 609–621 (2010).

5. T. Zarin, B. Strome, G. Peng, I. Pritišanac, J. D. Forman-Kay, A. M. Moses, Identifying molecular features that are associated with biological function of intrinsically disordered protein regions. *Elife*. **10** (2021), doi:10.7554/eLife.60220.

6. C. J. Brown, S. Takayama, A. M. Campen, P. Vise, T. W. Marshall, C. J. Oldfield, C. J. Williams, A. K. Dunker, Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* **55**, 104–110 (2002).

7. A. N. Nguyen Ba, B. J. Yeh, D. van Dyk, A. R. Davidson, B. J. Andrews, E. L. Weiss, A. M. Moses, Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci. Signal.* **5**, rs1 (2012).

8. N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, T. J. Gibson, Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).

9. N. E. Davey, D. C. Shields, R. J. Edwards, Masking residues using context-specific evolutionary conservation significantly improves short linear motif discovery. *Bioinformatics*. **25**, 443–450 (2009).

10. C. Benz, M. Ali, I. Krystkowiak, L. Simonetti, A. Sayadi, F. Mihalic, J. Kliche, E. Andersson, P. Jemth, N. E. Davey, Y. Ivarsson, Proteome-scale mapping of binding sites in the unstructured regions of the human proteome. *Mol. Syst. Biol.* **18**, e10584 (2022).

11. C. P. Wigington, J. Roy, N. P. Damle, V. K. Yadav, C. Blikstad, E. Resch, C. J. Wong, D. R. Mackay, J. T. Wang, I. Krystkowiak, D. A. Bradburn, E. Tsekitsidou, S. H. Hong, M. A. Kaderali, S.-L. Xu, T. Stearns, A.-C. Gingras, K. S. Ullman, Y. Ivarsson, N. E. Davey, M. S. Cyert, Systematic Discovery of Short Linear Motifs Decodes Calcineurin Phosphatase Signaling. *Mol. Cell*. **79**, 342–358.e12 (2020).

12. M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, H. S. Chan, Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9650–9655 (2007).

13. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex

Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).

14. A. N. Amin, Y.-H. Lin, S. Das, H. S. Chan, Analytical Theory for Sequence-Specific Binary Fuzzy Complexes of Charged Intrinsically Disordered Proteins. *J. Phys. Chem. B*. **124**, 6709–6720 (2020).

15. K. Bugge, I. Brakti, C. B. Fernandes, J. E. Dreier, J. E. Lundsgaard, J. G. Olsen, K. Skriver, B. B. Kragelund, Interactions by Disorder - A Matter of Context. *Front Mol Biosci*. **7**, 110 (2020).

16. E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, T. Mittag, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science*. **367**, 694–699 (2020).

17. Y. Lin, S. L. Currie, M. K. Rosen, Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* (2017), doi:10.1074/jbc.M117.800466.

18. J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*. **174**, 688–699.e16 (2018).

19. A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best, B. Schuler, Extreme disorder in an ultrahigh-affinity protein complex. *Nature*. **555**, 61–66 (2018).

20. G. Rosenblum, W. Zheng, H. Hofmann, Diffusion of a disordered protein on its folded ligand. *Proceedings of the* (2021) (available at https://www.pnas.org/content/118/37/e2106690118.short).

21. C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, M. K. Rosen, Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell*. **63**, 72–85 (2016).

22. N. E. Davey, The functional importance of structure in unstructured protein regions. *Curr. Opin. Struct. Biol.* **56**, 155–163 (2019).

23. T. Zarin, B. Strome, A. N. Nguyen Ba, S. Alberti, J. D. Forman-Kay, A. M. Moses, Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *Elife*. **8** (2019), doi:10.7554/eLife.46883.

24. M. C. Cohan, M. K. Shinn, J. M. Lalmansingh, R. V. Pappu, Uncovering Non-random Binary Patterns Within Sequences of Intrinsically Disordered Proteins. *J. Mol. Biol.* **434**, 167373 (2022).

25. A. X. Lu, A. X. Lu, I. Pritišanac, T. Zarin, J. D. Forman-Kay, A. M. Moses, Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. *bioRxiv* (2021), p. 2021.07.29.454330.

26. E. Proux-Wéra, D. Armisén, K. P. Byrne, K. H. Wolfe, A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*. **13**, 237 (2012).

27. R. J. Emenecker, D. Griffith, A. S. Holehouse, Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319 (2021).

28. J. M. Cherry, E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E. T. Chan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, B. C. Hitz, K. Karra, C. J. Krieger, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, M. Simison, S. Weng, E. D. Wong, Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–5 (2012).

29. J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, A. K. Dunker, Intrinsic disorder in transcription factors. *Biochemistry*. **45**, 6873–6888 (2006).

30. M. Fuxreiter, P. Tompa, I. Simon, V. N. Uversky, J. C. Hansen, F. J. Asturias, Malleable machines take shape in eukaryotic transcriptional regulation. *Nat. Chem. Biol.* **4**, 728–737 (2008).

31. A. Tsankov, Y. Yanagisawa, N. Rhind, A. Regev, O. J. Rando, Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. *Genome Res.* **21**, 1851–1862 (2011).

32. J.-F. Xiang, V. G. Corces, Regulation of 3D chromatin organization by CTCF. *Curr. Opin. Genet. Dev.* **67**, 33–40 (2021).

33. L. Braccioli, E. de Wit, CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* **63**, 157–165 (2019).

34. G. Fourel, T. Miyake, P.-A. Defossez, R. Li, E. Gilson, General regulatory factors (GRFs) as genome partitioners. *J. Biol. Chem.* **277**, 41736–41743 (2002).

35. P. D. Hartley, H. D. Madhani, Mechanisms that specify promoter nucleosome location and identity. *Cell*. **137**, 445–458 (2009).

36. L. Yu, R. H. Morse, Chromatin opening and transactivator potentiation by RAP1 in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **19**, 5279–5288 (1999).

37. A. Yarragudi, T. Miyake, R. Li, R. H. Morse, Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.* **24**, 9152–9164 (2004).

38. R. M. Raisner, P. D. Hartley, M. D. Meneghini, M. Z. Bao, C. L. Liu, S. L. Schreiber, O. J. Rando, H. D. Madhani, Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*. **123**, 233–248 (2005).

39. G. Badis, E. T. Chan, H. van Bakel, L. Pena-Castillo, D. Tillo, K. Tsui, C. D. Carlson, A. J. Gossett, M. J. Hasinoff, C. L. Warren, M. Gebbia, S. Talukder, A. Yang, S. Mnaimneh, D.

Terterov, D. Coburn, A. Li Yeo, Z. X. Yeo, N. D. Clarke, J. D. Lieb, A. Z. Ansari, C. Nislow, T. R. Hughes, A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. *Mol. Cell*. **32**, 878–887 (2008).

40. M. Ganapathi, M. J. Palumbo, S. A. Ansari, Q. He, K. Tsui, C. Nislow, R. H. Morse, Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. *Nucleic Acids Res.* **39**, 2032–2044 (2011).

41. A. M. Tsankov, D. A. Thompson, A. Socha, A. Regev, O. J. Rando, The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol.* **8**, e1000414 (2010).

42. D. Challal, M. Barucco, S. Kubik, F. Feuerbach, T. Candelli, H. Geoffroy, C. Benaksas, D. Shore, D. Libri, General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. *Mol. Cell*. **72**, 955–969.e7 (2018).

43. A. L. Sanborn, B. T. Yeh, J. T. Feigerle, C. V. Hao, R. J. L. Townshend, E. Lieberman-Aiden, R. O. Dror, R. D. Kornberg, Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *Elife*. **10** (2021), doi:10.7554/eLife.68068.

44. R. Oughtred, C. Stark, B.-J. Breitkreutz, J. Rust, L. Boucher, C. Chang, N. Kolas, L. O'Donnell, G. Leung, R. McAdam, F. Zhang, S. Dolma, A. Willems, J. Coulombe-Huntington, A. Chatr-aryamontri, K. Dolinski, M. Tyers, The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541 (2018).

45. BioGRID entry for Abf1. *Biological General Repository for Interaction Datasets (BioGRID)*, (available at https://thebiogrid.org/34022/summary/saccharomyces-cerevisiae/abf1.html).

46. P. R. Rhode, K. S. Sweder, K. F. Oegema, J. L. Campbell, The gene encoding ARS-binding factor I is essential for the viability of yeast. *Genes Dev.* **3**, 1926–1939 (1989).

47. M. J. Rossi, P. K. Kuntala, W. K. M. Lai, N. Yamada, N. Badjatia, C. Mittal, G. Kuzu, K. Bocklund, N. P. Farrell, T. R. Blanda, J. D. Mairose, A. V. Basting, K. S. Mistretta, D. J. Rocco, E. S. Perkinson, G. D. Kellogg, S. Mahony, B. F. Pugh, A high-resolution protein architecture of the budding yeast genome. *Nature*. **592**, 309–314 (2021).

48. W. J. de Jonge, M. Brok, P. Lijnzaad, P. Kemmeren, F. C. P. Holstege, Genome-wide off-rates reveal how DNA binding dynamics shape transcription factor function. *Mol. Syst. Biol.* **16**, e9885 (2020).

49. K. Roy, J. Gabunilas, A. Gillespie, D. Ngo, G. F. Chanfreau, Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome Res.* **26**, 1363–1375 (2016).

50. T. Candelli, D. Challal, J.-B. Briand, J. Boulay, O. Porrua, J. Colin, D. Libri, High-resolution transcription maps reveal the widespread impact of roadblock termination

in yeast. *EMBO J.* **37**, e97490 (2018).

51. T. Upton, S. Wiltshire, S. Francesconi, S. Eisenberg, ABF1 Ser-720 is a predominant phosphorylation site for casein kinase II of Saccharomyces cerevisiae. *J. Biol. Chem.* **270**, 16153–16159 (1995).

52. S. H. Reed, M. Akiyama, B. Stillman, E. C. Friedberg, Yeast autonomously replicating sequence binding factor is involved in nucleotide excision repair. *Genes Dev.* **13**, 3052–3058 (1999).

53. L. Feng, B. Wang, A. Jong, Saccharomyces cerevisiae Cdc6 stimulates Abf1 DNA binding activity. *J. Biol. Chem.* **273**, 1298–1302 (1998).

54. C. M. Loch, N. Mosammaparast, T. Miyake, L. F. Pemberton, R. Li, Functional and physical interactions between autonomously replicating sequence-binding factor 1 and the nuclear transport machinery. *Traffic*. **5**, 925–935 (2004).

55. G. Cho, J. Kim, H. M. Rho, G. Jung, Structure-function analysis of the DNA binding domain of Saccharomyces cerevisiae ABF1. *Nucleic Acids Res.* **23**, 2980–2987 (1995).

56. T. Miyake, C. M. Loch, R. Li, 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**, 505–516 (2002).

57. J. D. Boeke, J. Trueheart, G. Natsoulis, G. R. Fink, 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* **154**, 164–175 (1987).

58. T. Zarin, C. N. Tsai, A. N. Nguyen Ba, A. M. Moses, Selection maintains signaling function of a highly diverged intrinsically disordered region. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1450–E1459 (2017).

59. K. D. Fascher, J. Schmitz, W. Hörz, Role of trans-activating proteins in the generation of active chromatin at the PHO5 promoter in S. cerevisiae. *EMBO J.* **9**, 2523–2528 (1990).

60. R. H. Morse, Nucleosome disruption by transcription factor binding in yeast. *Science*. **262**, 1563–1566 (1993).

61. P. C. McAndrew, J. Svaren, S. R. Martin, W. Hörz, C. R. Goding, Requirements for chromatin modulation and transcription activation by the Pho4 acidic activation domain. *Mol. Cell. Biol.* **18**, 5818–5827 (1998).

62. K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, Highly accurate protein structure prediction for the human proteome. *Nature*. **596**, 590–596 (2021).

63. Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R.

O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker, N. L. Fawzi, Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* **36**, 2951–2967 (2017).

64. K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, 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).

65. A. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* (2016), doi:10.1016/j.str.2016.07.007.

66. A. E. Conicella, G. L. Dignon, G. H. Zerze, H. B. Schmidt, A. M. D'Ordine, Y. C. Kim, R. Rohatgi, Y. M. Ayala, J. Mittal, N. L. Fawzi, TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5883–5894 (2020).

67. K. M. Ruff, Y. H. Choi, D. Cox, A. R. Ormsby, Y. Myung, D. B. Ascher, S. E. Radford, R. V. Pappu, D. M. Hatters, Sequence grammar underlying unfolding and phase separation of globular proteins. *bioRxiv* (2021), p. 2021.08.20.457073.

68. B. A. Gibson, L. K. Doolittle, M. W. G. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, M. K. Rosen, Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*. **179**, 470–484.e21 (2019).

69. A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, G. J. Narlikar, Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*. **547**, 236–240 (2017).

70. A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, G. H. Karpen, Phase separation drives heterochromatin domain formation. *Nature*. **547**, 241–245 (2017).

71. H. B. Schmidt, A. Barreau, R. Rohatgi, Phase separation-deficient TDP43 remains functional in splicing. *Nat. Commun.* **10**, 4890 (2019).

72. A. S. Holehouse, G. M. Ginell, D. Griffith, E. Böke, Clustering of Aromatic Residues in Prion-like Domains Can Tune the Formation, State, and Organization of Biomolecular Condensates. *Biochemistry*. **60**, 3566–3581 (2021).

73. M. Rubinstein, R. H. Colby, *Polymer Physics* (Oxford University Press, New York, 2003).

74. J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, D. A. Drummond, Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell*. **168**, 1028–1040.e19 (2017).

75. K. Van Roey, T. J. Gibson, N. E. Davey, Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.* **22**, 378–385 (2012).

76. M. J. Henley, B. M. Linhares, B. S. Morgan, T. Cierpicki, C. A. Fierke, A. K. Mapp, Unexpected specificity within dynamic transcriptional protein-protein complexes. *Proc.*

Natl. Acad. Sci. U. S. A. **117**, 27346–27353 (2020).

77. L. M. Tuttle, D. Pacheco, L. Warfield, J. Luo, J. Ranish, S. Hahn, R. E. Klevit, Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. *Cell Rep.* **22**, 3251–3264 (2018).

78. M. Watson, T. B. Almeida, A. Ray, C. Hanack, R. Elston, J. Btesh, P. A. McNaughton, K. Stott, Anchored and labile pools of Calcineurin are enabled by a disordered AKAP scaffold. *bioRxiv* (2021), p. 2021.05.23.445342.

79. B. E. Eaton, L. Gold, D. A. Zichi, Let's get specific: the relationship between specificity and affinity. *Chem. Biol.* **2**, 633–638 (1995).

80. K. Teilum, J. G. Olsen, B. B. Kragelund, On the specificity of protein-protein interactions in the context of disorder. *Biochem. J*. **478**, 2035–2050 (2021).

81. P. M. Gonçalves, K. Maurer, W. H. Mager, R. J. Planta, Kluyveromyces contains a functional ABF1-homologue. *Nucleic Acids Res.* **20**, 2211–2215 (1992).

82. N. E. Davey, M. S. Cyert, A. M. Moses, Short linear motifs--ex nihilo evolution of protein regulation. *Cell Commun. Signal.* **13**, 43 (2015).

83. M. V. Staller, A. S. Holehouse, D. Swain-Lenz, R. K. Das, R. V. Pappu, B. A. Cohen, A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst*. **6**, 444–455.e6 (2018).

84. M. V. Staller, E. Ramirez, S. R. Kotha, A. S. Holehouse, R. V. Pappu, B. A. Cohen, Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Syst* (2022), doi:10.1016/j.cels.2022.01.002.

85. J. I. Gutierrez, G. P. Brittingham, Y. B. Karadeniz, K. D. Tran, A. Dutta, A. S. Holehouse, C. L. Peterson, L. J. Holt, SWI/SNF senses carbon starvation with a pH-sensitive low complexity sequence. *Elife*. **11** (2022), doi:10.7554/eLife.70344.

86. D. Moses, F. Yu, G. M. Ginell, N. M. Shamoon, P. S. Koenig, A. S. Holehouse, S. Sukenik, Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. *J. Phys. Chem. Lett.* **11**, 10131–10136 (2020).

87. C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Juelicher, A. A. Hyman, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*. **324**, 1729–1732 (2009).

88. E. Boke, M. Ruer, M. Wühr, M. Coughlin, R. Lemaitre, S. P. Gygi, S. Alberti, D. Drechsel, A. A. Hyman, T. J. Mitchison, Amyloid-like Self-Assembly of a Cellular Compartment. *Cell*. **166**, 637–650 (2016).

89. P. Klein, T. Pawson, M. Tyers, Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single receptor site. *Curr. Biol.* **13**, 1669–1678 (2003).

90. J. G. Olsen, K. Teilum, B. B. Kragelund, Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* **74**, 3175–3183 (2017).

91. J. Cubuk, J. J. Alston, J. J. Incicco, S. Singh, M. D. Stuchell-Brereton, M. D. Ward, M. I. Zimmerman, N. Vithani, D. Griffith, J. A. Wagoner, G. R. Bowman, K. B. Hall, A. Soranno, A. S. Holehouse, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **12**, 1936 (2021).

92. J. A. Riback, L. Zhu, M. C. Ferrolino, M. Tolbert, D. M. Mitrea, D. W. Sanders, M.-T. Wei, R. W. Kriwacki, C. P. Brangwynne, Composition-dependent thermodynamics of intracellular phase separation. *Nature*. **581**, 209–214 (2020).

93. D. W. Sanders, N. Kedersha, D. S. W. Lee, A. R. Strom, V. Drake, J. A. Riback, D. Bracha, J. M. Eeftens, A. Iwanicki, A. Wang, M.-T. Wei, G. Whitney, S. M. Lyons, P. Anderson, W. M. Jacobs, P. Ivanov, C. P. Brangwynne, Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell*. **181**, 306–324.e28 (2020).

94. X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, R. D. Vale, Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*. **352**, 595–599 (2016).

95. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, A Phase Separation Model for Transcriptional Control. *Cell*. **169**, 13–23 (2017).

96. T. M. Franzmann, M. Jahnel, A. Pozniakovsky, J. Mahamid, A. S. Holehouse, E. Nüske, D. Richter, W. Baumeister, S. W. Grill, R. V. Pappu, A. A. Hyman, S. Alberti, Phase separation of a yeast prion protein promotes cellular fitness. *Science*. **359** (2018), doi:10.1126/science.aao5654.

97. A. W. Fritsch, A. F. Diaz-Delgadillo, O. Adame-Arana, C. Hoege, M. Mittasch, M. Kreysing, M. Leaver, A. A. Hyman, F. Jülicher, C. A. Weber, Local thermodynamics govern formation and dissolution of Caenorhabditis elegans P granule condensates. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2102772118.