Received: 17 June 2022 | Revised: 15 September 2022 | Accepted: 17 September 2022

DOI: 10.1002/pro.4455

FULL-LENGTH PAPER

THE PROTEIN SOCIETY WILEY

Intrinsic protein disorder uncouples affinity from binding specificity

Tamas Lázár[1,2] | **Agnes Tantos**[3] | **Peter Tompa**[1,2,3] | **Eva Schad**[3]

[1]VIB-VUB Center for Structural Biology, Flanders Institute for Biotechnology (VIB), Brussels, Belgium

[2]Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium

[3]Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

Correspondence
Eva Schad, Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary.
Email: schad.eva@ttk.hu

Peter Tompa, Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium.
Email: peter.tompa@vub.be

Funding information
EC H2020-MSCA-RISE Action grant, Grant/Award Number: IDPfun no. 778247; EC H2020-WIDESPREAD-2020-5 Twinning grant, Grant/Award Number: PhasAge no. 952334; Nemzeti Kutatási Fejlesztési és Innovációs Hivatal, Grant/Award Numbers: K124670, K131702; Országos Tudományos Kutatási Alapprogramok, Grant/Award Number: PD-OTKA 108772; Vrije Universiteit Brussel; Strategic Research Program on Microfluidics, Grant/Award Number: SRP51

Review Editor: Nir Ben-Tal

Abstract

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) of proteins often function by molecular recognition, in which they undergo induced folding. Based on prior generalizations, the idea prevails in the IDP field that due to the entropic penalty of induced folding, the major functional advantage associated with this binding mode is "uncoupling" specificity from binding strength. Nevertheless, both weaker binding and high specificity of IDPs/IDRs rest on limited experimental observations, making these assumptions more speculations than evidence-supported facts. The issue is also complicated by the rather vague concept of specificity that lacks an exact measure, such as the K_d for binding strength. We addressed these issues by creating and analyzing a comprehensive dataset of well-characterized ID/globular protein complexes, for which both the atomic structure of the complex and free energy (ΔG, K_d) of interaction is known. Through this analysis, we provide evidence that the affinity distributions of IDP/globular and globular/globular complexes show different trends, whereas specificity does not connote to weaker binding strength of IDPs/IDRs. Furthermore, protein disorder extends the spectrum in the direction of very weak interactions, which may have important regulatory consequences and suggest that, in a biological sense, strict correlation of specificity and binding strength are uncoupled by structural disorder.

KEYWORDS
binding strength, conservation, disordered protein complexes, IDPs, specificity, structural disorder

Abbreviations: ASA, accessible surface area; ΔG, free energies of binding; GO, Gene Ontology; ID, intrinsically disordered; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; IF, interface; K_d, dissociation constant; SLiM, short linear motif; MoRE, molecular recognition elements; PDB, protein data bank; SemSim, semantic similarity.

1 | INTRODUCTION

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions of proteins (IDRs) abound in the proteome of higher organisms and have important functions

mostly related to signaling and regulation.[1–3] IDPs/IDRs lack a stable 3D fold, and they often function by molecular recognition, when they can undergo folding induced upon binding to the partner protein.[4,5] A few of the best-studied examples are binding of eIF4G to eIF4E,[6] p53 to MDM2,[7] E-cadherin to β-catenin,[8,9] p27 to the complex of Cyclin A/Cdk 2,[10] CREB to CBP KIX domain,[11] and inhibitor-2 to PP1.[12,13] Structural analysis of these and other examples has ascertained that the binding mode of IDPs/IDRs differs from that of globular proteins.[14] IDPs do not have flat interfaces, rather, they fit in an extended conformation into the hydrophobic binding cleft of their partners, either via a short recognition motif[15] or via a longer disordered domain.[16] In both binding modes, the interface region is a larger part of the protein, and is better packed, in the case of IDPs.[17,18]

There are many possible functional advantages of binding accompanied by induced folding (disorder-to-order transition), such as a kinetic advantage termed fly casting,[19,20] regulation by posttranslational modification,[21] and structural adaptability to different partners.[22] The advantage most often cited, however, is that induced folding entails a large entropic penalty, which makes the interaction of IDPs/IDRs weaker than that of globular proteins with an interface of similar size, thus "uncoupling" specificity from binding strength.[23–26] This feature might be especially advantageous in regulatory settings, when interactions need to quickly form and break up, which would be difficult with a very strong interaction with low k_{off}.

The key problem with this view is that structural adaptation enabled by the underlying greater degree of plasticity may also be conducive of nonspecific interactions.[25] Actually, intrinsic structural disorder is also often claimed to result in promiscuous binding,[27–29] which may have dire consequences on the cell, such as dosage sensitivity[30] and cancer,[31] but may also confer advantages, such as the evasion of cellular surveillance monitoring protein misfolding[32] or providing raw material for evolutionary innovation.[33–35] Furthermore, although this is a fundamental concept of the IDP field, it relies mostly on preconcepts and misinterpreted data, rather than solid experimental facts. For example, binding of IDPs is not conspicuously weak, K_d values in the μM range are mostly reported for short disordered motifs,[15,21,36] whereas longer, domain-sized disordered binding regions, such as that of E-cadherin, prothymosin-α, p27, and I2, often bind with nM–pM affinities.[8,10,12,37] In addition, considering solely the entropic penalty associated with induced folding is an oversimplification, because IDPs use a higher proportion of hydrophobic residues for binding than globular proteins, which entails a favorable entropic component of water molecules released, and their interfaces are packed better, which suggests an effective enthalpic compensation.[17] Furthermore, their binding-competent conformation may be preferentially sampled in the free state[38] and they often cause less structural change in the partner[18] than globular proteins,[39] which all limit the magnitude of unfavorable entropy associated with structural reorganization upon binding.

This issue is also complicated by the severe conceptual and practical problems with defining and actually measuring specificity. It may be agreed a priori that specificity is the ability of a protein to discriminate between the cognate (native, functional) partner and all other competitors, which must derive not only from the physical–chemical properties of the interacting proteins (thermodynamics and kinetics), but also from biological context not amenable in in vitro experiments, such as posttranslational modifications, cellular expression and localization, the presence of additional partners and last but not least, the (patho) physiological readout of the spectrum of competing interactions.[40] Many of these factors are not mirrored in the binding strength alone, as underscored by a very broad distribution of binding free energies of physiologically relevant (i.e., apparently "specific") protein–protein interactions, spanning from mM to fM affinities.[28,37,39]

To clarify this key concept in the IDP field, a comprehensive collection and analysis of the available information is much overdue, as only the comparison of large datasets can provide enough information on the overall behavior of the different protein classes. To this end, we have collected data on all IDP-partner interactions for which both the structural features and thermodynamic parameters are available. We compared our data to a reference set of folded protein complexes similarly characterized,[39] by exploiting quantitative measures associated with the elusive feature of interaction specificity: (i) evolutionary conservation of the interface, (ii) interface patterning (iPat) specificity, and (iii) functional similarity of interaction partners. By correlating these values, we show that specificity of interactions does not depend on, or correlate with, their binding strength—either for disordered or ordered partners, but disordered proteins extend the spectrum of protein–protein interactions in the direction of weak interactions.

It is worth noting that due to the nature of our dataset our analysis is limited to complexes with known 3D structure and our findings cannot be readily extended to fuzzy ID complexes, where a fixed structure is absent even in the complexed state. Our dataset is also devoid of coiled coils, as these complexes often form stable homodimers/homooligomers, and less frequently sample the disordered monomeric state.

2 | RESULTS

2.1 | Binding strength

2.1.1 | Binding strength distributions of complexes

To address the question whether IDPs truly form weaker interactions, we compared complexes of disordered proteins with globular partners (denoted as "disordered" or "ID complexes") (Table S1) and complexes of two globular proteins (denoted as "globular complexes") (Table S2) in terms of their experimentally defined dissociation constants (K_d) (Figure S1), from which we then calculated free energies of binding (ΔG) (Figure 1) (see Section 4 for details). The results revealed that disordered and globular complexes show overlapping but largely different distributions of ΔG. In line with the broadly accepted dogma, the mean ΔG value of ID complexes is significantly lower than that of the globular complexes (7.7 and 10.7 kcal/mol, respectively, $p > 0.0001$, *t*-test, Table S3) and ID complexes indeed present a biased distribution towards weaker interactions. Nevertheless, they are also capable of creating strong complexes in the range of $\Delta G = 3.50$–14.03 kcal/mol ($K_d = 2.7$ mM–52 pM) clearly showing that IDPs do not invariably engage in weak interactions, as they are often generally referred to.[23–26] Figure 2 shows the wide variety of structures and affinities of IDP complexes.

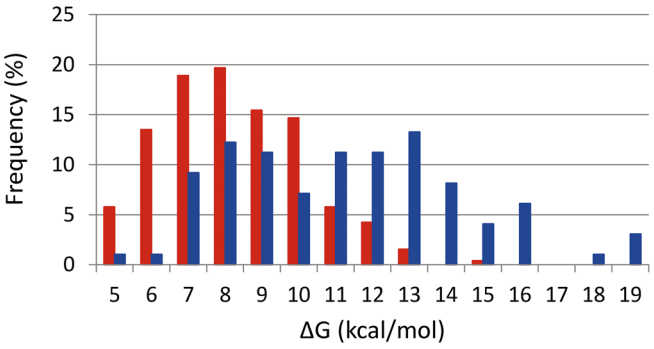

FIGURE 1 Distribution of free energy of binding in ID (red) and globular (blue) complexes. ID, intrinsically disordered

It is also to be noted that globular complexes have a wider range in interaction strength and exceptionally tight binding can mostly be achieved by two globular proteins, while extremely weak interactions almost invariably require a disordered partner.

2.1.2 | Interface size distribution of complexes

In general, interface size correlates with binding strength, but the distribution of total interface size (Figure 3a) does not follow strictly that of the ΔG of the complexes. Apparently, ID complexes have a wider range of interface size distribution than globular ones, that is, they can form complexes with both very small and extraordinarily large interfaces (cf. Figure 2). On the average, ID complexes have smaller interface sizes (which might simply follow from an overrepresentation of complexes established by short motifs), but interestingly, almost all large interfaces are represented in ID complexes (reaching 5000 Å^2, e.g., cadherin - β-catenin, Figure 2a; inhibitor-2 - PP1, Figure 2b; hif1-α - CBP, Figure 2c). There is only one instance where the interface of a globular protein complex exceeds 5000 Å^2 (PDB code: 2oza).

When comparing the disordered and the globular partners in ID complexes, the interface of the IDP is larger than that of its globular partner (*p*-value < 0.0001,

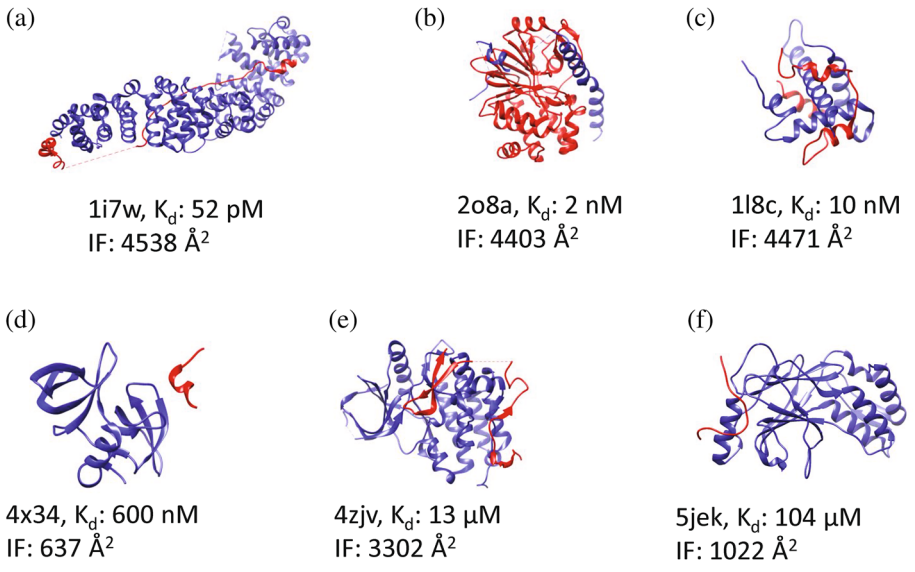

FIGURE 2 Examples of disordered complexes with a broad range of K_d values and interface sizes: (a) cadherin - β-catenin, (b) inhibitor-2 - PP1, (c) hif1-α - CBP, (d) p53 - p53-binding protein 1, (e) ERBB receptor feedback inhibitor 1 - EGFR, and (f) MAVS - IRF3. ID parts are named first. PDB codes and K_d values and interface (IF) sizes are indicated under the corresponding structures. Red: ID chain; blue: globular chain. ID, intrinsically disordered

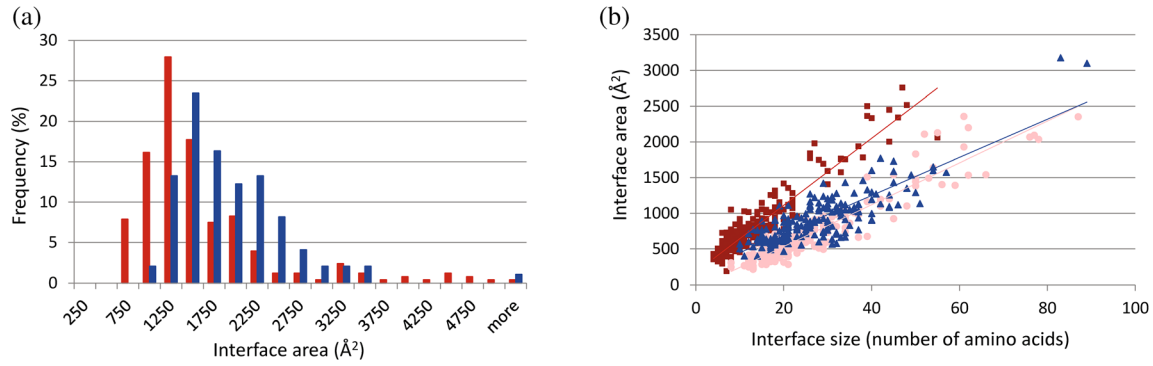

FIGURE 3 Interface sizes of ID and globular complexes. (a) Interface size distribution of disordered (red) and globular (blue) complexes. (b) Relationship between the size and the number of constituent amino acids of half interfaces (ID complex IDP part: red square; ID complex globular part: rose circle; globular complex: blue triangle). ID, intrinsically disordered

t-test) (Figure S2). This type of comparison can only be made for the ID complexes, as in globular complexes the two parts cannot be distinguished. When the interface size is defined as the number of constituent amino acids, we find that IDPs usually need fewer amino acids to create a certain interface area than their globular partners (p-value < 0.0001, *t*-test), or proteins in globular complexes (p-value < 0.0001, *t*-test) (Figure 3b). This observation is related to earlier findings that IDPs use a larger proportion of their amino acids for interaction.[41]

2.1.3 | Interface size and binding strength

For globular proteins, there is a linear relationship between interface size and binding strength (adding roughly 0.02 kcal/mol/Å^2, based on globular trendline in Figure 4), which suggests that cooperativity between different binding segments within the interface increases the binding energy by a favorable entropic contribution (Figure 4.).

For ID interfaces, increasing interface size adds much less to binding strength, thus there is a crossover between the two types of complexes at about 9 kcal/mol (1500 Å^2). Below this threshold the same binding free energy can be realized by a smaller ID interface than a globular one, but above this value the ID interface must be larger to achieve the same binding strength. It is also clear that ID complexes are unable to form interactions above a certain strength (around 13 kcal/mol).

The same trends are seen when only the hydrophobic part of interface area is taken into account (Figure S3, Table S3).

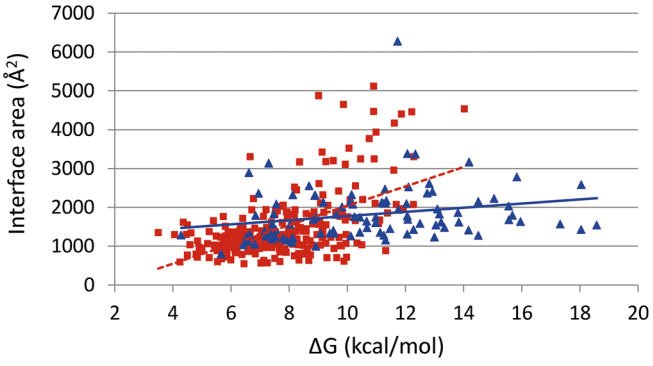

FIGURE 4 Relationship between free energy of binding and the total interface size in disordered (red square) and globular (blue triangle) complexes. Dashed line in case of disordered complexes means that it is not a correct trendline but we could not fit better

2.1.4 | Interaction types in different protein complexes

Given the potential predictive power of hydrophobicity on binding strength and the apparent reliance of IDPs on hydrophobic interactions,[15,38] we analyzed the different types of interactions between the partners in the case of strong (K_d < 1 μM, ΔG > ~8.2 kcal/mol) and weak ($K_d \geq 1$ μM) complexes.

The percental occurrence of the two main types of interactions (hydrophobic interactions and hydrogen bonds) are significantly different as the proportion of hydrophobic interactions is significantly higher in ID, than in globular complexes (p-value = 0.0002, *t*-test). Interestingly, there is no significant difference between strong and weak complexes in either case (Figure 5a).

In both types of complexes, the interface is significantly larger for those binding with strong affinity (for ID strong and weak complexes: 1932 Å^2 vs. 1183 Å^2, respectively; for globular complexes: 1944 Å^2 vs. 1414 Å^2, respectively; p-values < 0.0001, for both cases), but in

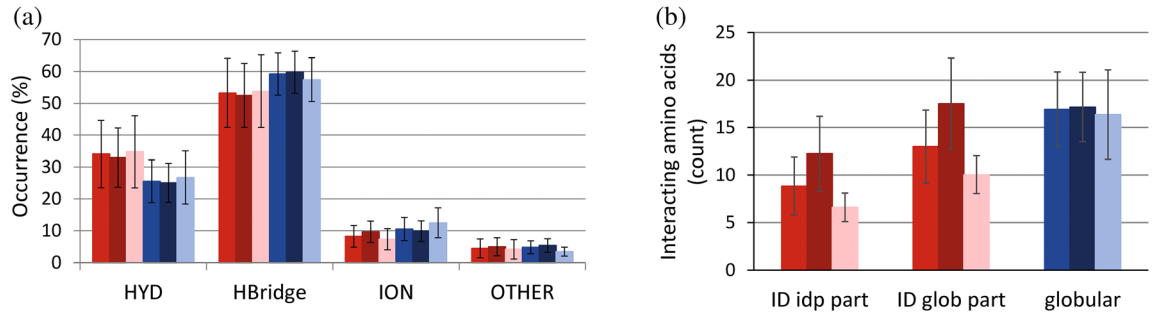

FIGURE 5 Interaction types of ID and globular complexes. (a) Percental distribution of interaction types (HYD: hydrophobic; HBridge: hydrogen bond; ION: ionic; OTHER: disulfide bridges, aromatic-aromatic, aromatic-sulfur, and cation-pi interactions) in strong and weak complexes (red: all ID complexes; dark red: strong; rose: weak ID complexes; blue: all globular complexes; dark blue: strong; light blue: weak globular complexes). (b) Number of interacting amino acids in ID and globular complexes. In the ID complexes, disordered and ordered partners are counted separately: ID idp part and ID glob part (red: all ID complexes; dark red: strong; rose: weak ID complexes; blue: all globular complexes; dark blue: strong; light blue: weak globular complexes). Error bars represent mean ± SD. ID, intrinsically disordered; SD, standard deviation

case of globular proteins the number of interactions remains the same (Figure 5b), meaning that globular proteins achieve stronger binding by increasing only the interface area but not the number of interactions.

As IDPs are typically characterized by high net charges and often participate in ionic interactions[1] it is important to assess the role of electrostatic interactions in IDP binding energetics. To achieve this, we calculated the number of ionic interactions in each ID complex. Similarly to the total number of interactions ($r = 0.57$, Table S3), the number of ionic interactions showed a positive correlation with the binding strength ($r = 0.42$, Table S3, Figure S4A). Nevertheless, this interaction type does not appear to be a determining factor for stronger binding, as the percentage of ionic interactions is not correlated with ΔG ($r = 0.18$, Table S3, Figure S4B).

2.1.5 | Segmentation of interfaces

Segmentation of the binding interfaces is important, as it would be plausible that interfaces assembled from more segments are entropically more unfavorable than interfaces assembled from less segments. Our results, in accordance with previous work[17] show that IDPs have fewer (Figure S5A) and longer (Figure S5B) interacting segments. This underlines the view that IDPs mostly interact with a few (mostly one) short linear motifs (SLiM) or molecular recognition elements (MoRE)[42] as opposed to globular proteins, which form interfaces that are, on average, made up of four to five segments. Nonetheless, these SLiM/MoRE segments of IDPs are usually longer than the binding segments of globular proteins (p values < 0.0001, t-test). More importantly, the correlation between segmentation and binding strength is weak ($r = 0.24$, Table S3): the mean binding free energy of IDPs binding with a single segment is around 7.6 kcal/mol, while of those binding with multiple segments is around 9.5 kcal/mol.

2.2 | Specificity

As suggested above, specificity is a contextual property of protein–protein interactions that cannot be simply deduced from their binding strength. Here we suggest that it can be approached by indirect features that are each conducive of the possible dominance of the given interaction over those with competing partners.

Therefore, we suggest and evaluate three novel measures of specificity: evolutionary conservation, patterning specificity of the interfaces of complexes, and functional similarity of interaction partners.

2.2.1 | Evolutionary conservation of the interface

Conservation of amino acid sequence is a definite marker for functional importance, as it can be assumed that functionally important (specific) amino acids should have a higher level of conservation.[43–47] Therefore, determining the level of evolutionary conservation of the interface amino acids is a reasonable measure of interaction specificity (see Section 4 for details).

A special aspect of such comparisons is that disordered proteins and regions tend to evolve faster than globular proteins,[48] with the exception of important interaction sites, where sequence conservation is often seen in IDPs too.

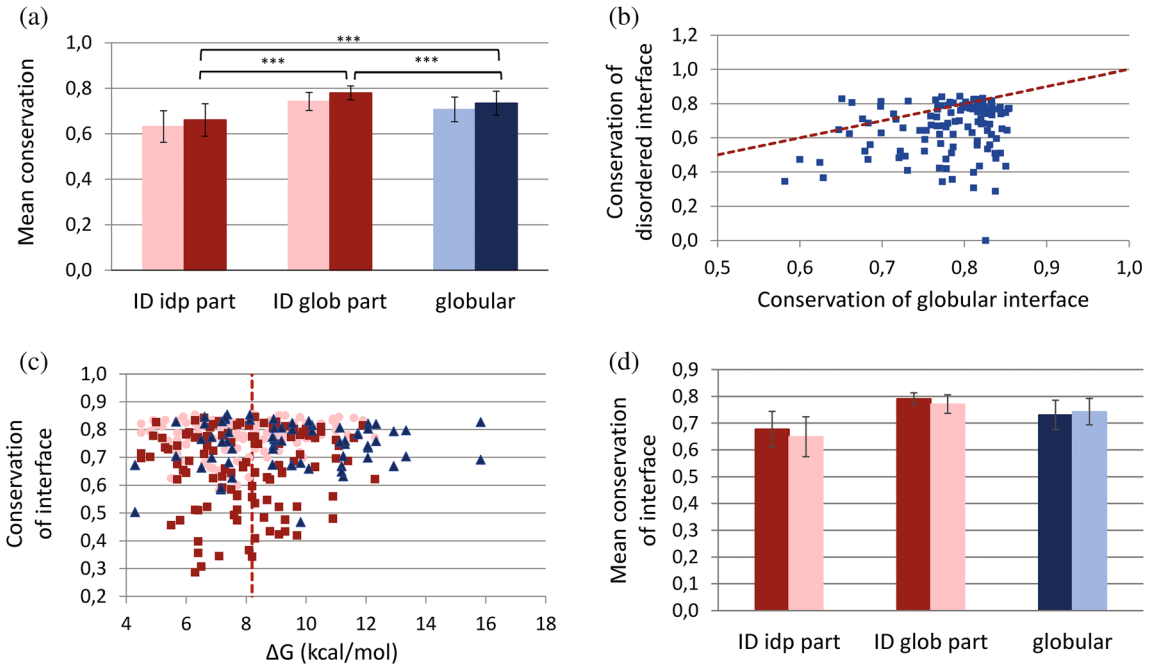

FIGURE 6 Interface conservation. (a) Mean conservation of protein surfaces and interfaces in ID (interface: red; surface: rose) and globular (interface: dark blue; surface: light blue) complexes. Significances are indicated with asterisks above the columns. (b) Relationship between conservation values of interfaces in disordered and globular chains in ID complexes. The diagonal $y = x$ is shown as guiding line in red. (c) Conservation and binding strength. ID complex idp part: red square; ID complex globular part: rose circle; globular complex: blue triangle. Vertical line ($\Delta G = 8.2$ kcal/mol) separates the weak (on the left) and strong (on the right) complexes. (d) Interface conservation of strong and weak ID and globular complexes. In ID complexes, disordered and ordered partners are counted separately: ID idp part and ID glob part (red: strong; rose: weak ID complexes; dark blue: strong; light blue: weak globular complexes.) Error bars represent mean ± SD. ID, intrinsically disordered; SD, standard deviation

Comparison of the interface and surface amino acids of ID and globular complexes (Figure 6a) reveals that the interface amino acids are always more conserved than other surface residues, a clear sign of the importance of the interaction they mediate, which is a correlate of specificity. However, these differences are significant only for the globular partners in ID complexes.

Not surprisingly, interfaces of disordered proteins are less conserved than interfaces of their globular partners in approximately 80% of the cases (Figure 6b). Strikingly, the interfaces of globular proteins that bind an IDP are even more conserved than those of which have globular partner (Figure 6a, p-value = 0.0003, t-test). The type of molecular interaction between the binding partners does not appear to be a major factor in determining conservation, as the number of ionic interactions did not show correlation with conservation (Figure S6, Table S3).

Although stereotypically IDPs tend to form weaker complexes (Figure 1), no correlation can be observed between the extent of conservation and binding strength (Figure 6c, Table S3), meaning that weak-binding complexes can still be conserved and, on the other hand, even a fast-evolving, variable ID chain is capable of forming strong interactions.

This is further supported by the fact that there is no significant difference between the conservation of weak and strong complexes of the same types of protein chains (Figure 6d, Table S3).

2.2.2 | iPat

iPat is a pseudo-potential describing local patterning of the physico-chemical properties of amino acids. Its application for approximating specificity derives from the idea that the more unique the surface amino-acid pattern of an interface is, the more specific the given interaction is[49] (see Section 4 for details).

For iPat, the higher the value, the more interface-like and less noninteracting surface-like the local environment of a residue is.

By comparing the mean iPat scores (<iPat>) for the interfaces and surfaces of both chains in ID complexes and in globular complexes (Figure 7a, Figure S7), we found that interface values are always significantly higher than surface values (p-values < 0.0001, t-test), validating that iPat distinguishes between interfaces and the general surface, that is, that it is a reasonable proxy to specificity. This

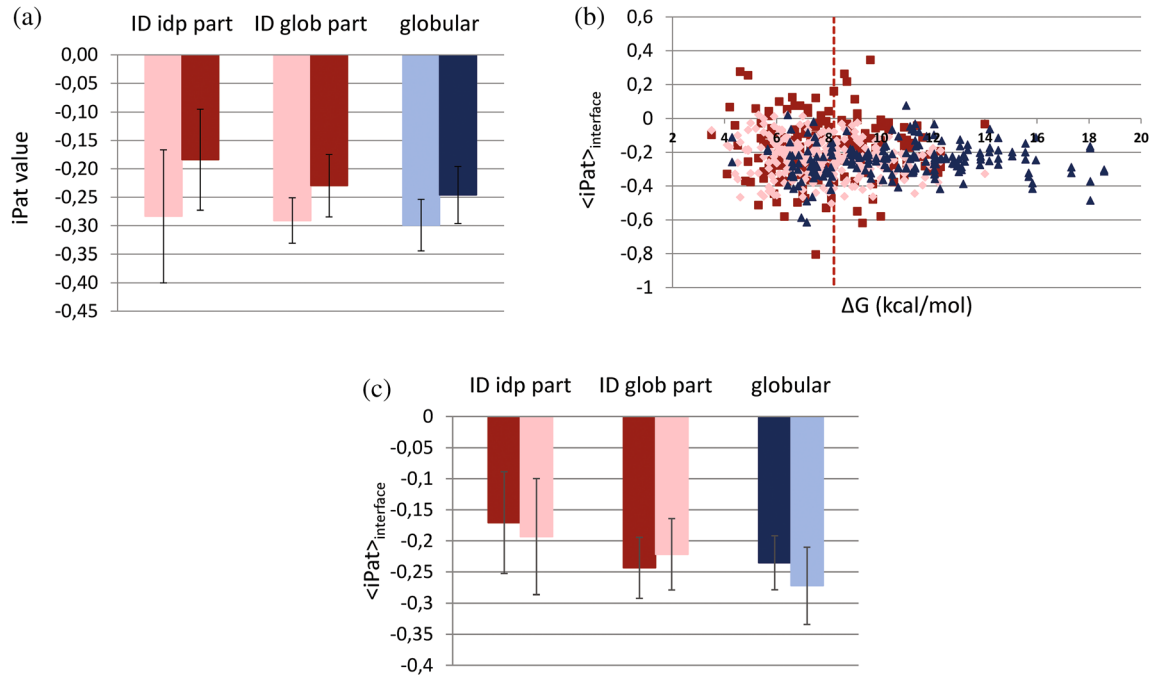

FIGURE 7 iPat scores in ID and globular complexes. (a) Mean iPat scores of protein surfaces and interfaces in disordered (interface: red; surface: rose) and globular (interface: dark blue; surface: light blue) complexes. (b) Correlation between binding strength and interface mean iPat scores (ID complex idp part: red square; ID complex globular part: rose circle; globular complex: blue triangle). The vertical line ($\Delta G = 8.2$ kcal/mol) separates the weak (on the left) and strong (on the right) complexes. (c) Mean interface iPat values of ID and globular complexes. In the ID complexes, disordered and ordered partners are counted separately: ID idp part and ID glob part (red: strong; rose: weak ID complexes; dark blue: strong; light blue: weak globular complexes.) Error bars represent mean ± SD. ID, intrinsically disordered; iPat, interface patterning; SD, standard deviation

is also underscored by the lack of differences between the noninteracting surface values (Figure 7a). <iPat> values for ID interfaces score significantly higher than those of globular proteins, both in ID (p-value = 0.0054, t-test) and globular (p-value = 0.0003, t-test) complexes (Figure 7a), which argues that interfaces of the ID partners carry specific information required for partner recognition. A high number of ionic interactions is accompanied by a lower iPat value (Figure S8, Table S3), which is not surprising, given that the concentration of amino acids with similar physico-chemical properties could reduce the specific patterning of a region.

Mean iPat values of both the interface ($<iPat>_{interface}$) and surface ($<iPat>_{surface}$) show a higher variance in the disordered part of ID complexes than in the globular partner, ranging between [−0.8, 0.4] versus [−0.6, 0.1] (Figure S7), respectively. This suggests that IDPs are capable of forming both more specific and less specific interactions than globular proteins. Furthermore, globular complexes are of lower patterning specificity, signified by the higher number of instances with $<iPat>_{surface}$ greater than $<iPat>_{interface}$ (34.2%) compared to globular chains in ID complexes (29.8%) or the disordered segments (29.6%) (Figure S7).

Similarly to surface area, interface <iPat> values also does not display any correlation with binding strength (Figure 7b, Table S3). When separating the interactions again to "weak" and "strong" ones we could quantitatively show the lack of significant differences (Figure 7c). Interactions of both low- and high-specificity were found among weak-affinity IDPs chains, indicated by interface <iPat>.

We were also unable to detect any remarkable correlations between conservation and specificity of iPat (correlation coefficient $r \sim 0.1$–0.3, Table S3), confirming that conservation and the iPat specificity (measured by iPat) are two orthogonal measures (Figure S9).

2.2.3 | Functional specificity

If pairs of proteins are associated with similar cellular functions, their interaction can be considered to exhibit higher specificity than those involved in different biological pathways. If the latter case is observed, it suggests either a less specific interaction or a case of pathway crosstalk. In practice, functional similarity is commonly defined as semantic similarity[50–53] between sets of Gene

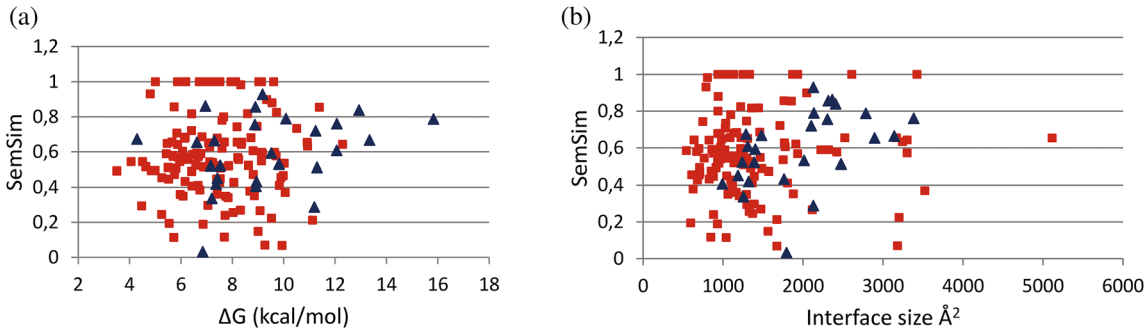

FIGURE 8 Correlations of the binding free energy and interface size with the functional similarity. (a) Correlations between the binding free energy (ΔG) and GOSemSim for ID complexes (red square, $r < 0.01$) and for globular complexes (blue triangle, $r = 0.29$). (b) Correlations between the binding free energy (ΔG) and GOSemSim for ID complexes (red square, $r = 0.03$) and for globular complexes shown (blue triangle, $r = 0.47$). ID, intrinsically disordered; GO, gene ontology; SemSim, semantic similarity

Ontology (GO) terms from the Biological Process (BP) aspect associated with the interacting partners (see Section 4 for details).

Analysis of the functional specificity of the interactions revealed that functional specificity is decoupled from the affinity of ID complexes ($r < 0.01$) and only very weakly correlated ($r = 0.29$, Table S3) in case of protein complexes of globular proteins (Figure 8a). Furthermore, functional specificity of ID complexes is independent of the total and hydrophobic interface area ($r = 0.03$, Table S3), and is only moderately dependent in the case of globular proteins ($r = 0.47$, Table S3), indicating that larger interfaces mediate only marginally more specific interactions (Figure 8b). It is also independent of the ionic interactions, as functional similarity shows no correlation with the number of ionic interactions (Figure S10, Table S3).

The ratio of hydrophobic area of globular protein complexes was found to be weakly anticorrelated ($r = -0.32$, Table S3) with functional similarity, that is, interfaces dominated by hydrophobic patches are associated with lower specificity (Figure S11) — a correlation commonly referenced in the literature.[49]

3 | DISCUSSION

The molecular function of IDPs/IDRs usually results from a combination of disorder (entropic chains) and folding induced upon binding to partner molecule(s).[1,2] The large decrease in backbone entropy that accompanies induced folding gave rise to the prevailing but controversial concept of the IDP field that structural disorder uncouples specificity from binding strength, meaning that IDPs bind weak (or, rather, weaker than their globular counterparts), yet their binding is specific. An often cited but somewhat misinterpreted work that fueled this concept is a classic study on the thermodynamics and specificity of protein–DNA recognition,[54] in which it was found that the unfavorable (positive) entropy change upon folding coupled to binding is essential to specific DNA recognition. It is mostly missed, however, that the paper claims that specificity actually results from an increased complementarity and enthalpic compensation at the protein–DNA interface, that is, from a stronger binding enabled by flexibility. Instead, it persists in the IDP literature as evidence that "structural disorder uncouples specificity from binding strength". Neither element of IDP binding supporting this concept (weak binding and specificity) relies on solid foundation, though. Apparently, IDPs do not, in general, bind their partners weakly, one can find very tight, nM–pM, binders among them.[8,10,12] Even less solid is the claim on specificity of their binding. Specificity can have several distinct conceptual and practical approximations, and it can be argued which one to take into consideration.

To take a closer look and attempt to clarify this issue, we collected data on IDP/IDR interactions for which both the high-resolution structure and thermodynamics of interaction have been characterized. We also invented three measures of specificity (information, conservation, HTS confidence), which can be studied at the level of the isolated protein, but have a strong contextual component.

As a main finding, we present unequivocal proof that IDPs do not invariably engage in weak interactions, as their complexes show a distribution of binding strengths similar to those of globular complexes. Nevertheless, the extremely weak complexes almost always rely on an IDP, while exceptionally tight binding appears to be only achievable by globular proteins. We also found that binding strength and interface size are not strictly related, as large interfaces can also belong to relatively weak complexes.

An interesting observation is that in the low binding strength regime, folded and disordered proteins cross

over, that is, the same binding free energy can be achieved by a smaller IDP interface. It may have several reasons, such as a limited induced folding of pre-formed short binding motifs,[38] better packing of interfaces of IDPs[17] that may contain more backbone H-bonds[18] and also that the entropic penalty is reduced by IDPs inducing less folding of their partner.[18,39] In contrast, for complexes with free energies above 8 kcal/mol, the ID interface must be larger than the interface of a globular complex with similar affinity. Apparently, unfavorable entropy that incurs folding upon binding makes the interaction weaker than expected from the size of the interface. This negative effect is so strong that to go from 4 to 12 kcal/mol in free energy, the interface in an IDP complex must realize a five-fold increase, and within the range 10–12 kcal/mol, ID interfaces are two times larger than globular proteins of similar interaction strength.

Probably as a result of the increase of this entropic penalty, the linearity of free energy with interface size above 10 kcal/mol and 2000 Å^2 does not apply for ID complexes; there is an upward curvature that may indicate a saturation effect at around 13 kcal/mol.

While IDPs and globular proteins apparently rely on similar types of molecular interactions to form complexes, the number of contacts show different behavior between the two groups. Strong ID complexes have significantly more interresidue interactions (and also more interacting amino acids in both partners) than weak ID complexes (p-value < 0.0001, t-test), but interestingly there is no such difference in the globular ones, they seem to have other strategies, such as interface complementarity, to increase their binding strength. We also identified a few ID complexes with very high relative number of intermolecular interactions, many of which exhibited more than five interresidue interactions per interface amino acid. This implies very tight shape complementarity due to the uniquely adaptable nature of IDPs. One of these extremes is the cysteine-rich PDZ-binding protein, that interacts with Disks large homolog 4 with around 8 kcal/mol of free energy[55] using five interface residues that coordinate a total of 51 non-covalent bonds across the interface (Figure S12; PDB code: 5heb).

As the notion that IDPs always form weak interactions cannot be considered generally true, we can consider the same to be applicable for the specificity of interactions. Even though specificity in itself is difficult to define with a simple measure, all three of our specificity-related features show that specificity of weak and strong intercations does not differ even for the complexes of globular proteins. By any of the three measures applied for describing specificity, we found that ID complexes are not consistently weaker, they tend to combine the same specificity with the same binding strength, which probably relies on a different balance of entropic and enthalpic components.

An important observation regarding the conservation of the interface residues is that although surface residues outside the interface are generally considered "noninteracting," their high conservation (0.7–0.75) might indicate their importance in defining specificity and increasing binding strength. If we consider the conservation of globular interactors, the interfaces of globular proteins binding IDPs are significantly more conserved than those of structured proteins (Figure 6a, p-value = 0.0003, t-test). This observation suggests that, in general, ID complexes are at least as specific as globular complexes, and it is the globular partner that is mostly responsible for conservation, while the ID partner is more variable. This can be explained with the adaptability of IDPs, which allows them to recognize a given partner even if their own amino acid sequence has changed over time.

An important limitation of this study is that our conclusions cannot be extended to complexes with extreme fuziness, as these are not present in the structural databases. Improvements in the field of accurately determining structural ensembles will probably make such analyses possible in the future.

In conclusion, instead of confirming the general perception in the field that structural disorder uncouples binding strength from specificity, we shift the focus to the ability of disordered proteins to engage in a wide variety of interactions, including weak but specific ones. This nature of IDPs extends the spectrum of protein–protein interactions towards weak relations probably uniquely amenable for regulation. In fact, we can say that the existence of weak interactions is due to disordered proteins, which could be one of the major functional relevances of this protein class.

4 | METHODS

4.1 | Construction of the datasets

We obtained complexes of IDPs from Disordered Binding Site (DIBS) Database (http://dibs.enzim.ttk.mta.hu)[56] which contains protein complexes formed between disordered and globular proteins. We used only complexes where the disordered nature was directly proven by experimental evidence for the protein (classified as "Confirmed") or a very close homolog (classified as "Inferred from homology"). We selected complexes with PDB structures for which there was information on binding strength (experimentally defined dissociation constant (K_d)). Only complexes with one ordered and one

disordered chain were taken into consideration. Our selection process resulted in 259 complexes (Table S1) out of the total 773 available in the DIBS database. Heterodimeric coiled coils were not included the dataset for IDP complexes.

As a reference database, we used a collection of 144 globular complexes with known K_d values[39] and selected the ones that contain exactly two protein components, which resulted in a dataset of 98 instances (Table S2).

4.2 | Calculation of ΔG

ΔG (change in Gibbs free energy) was calculated from the experimentally measured dissociation constant: $\Delta G = -R \times T \times \ln K_d$, where R is the gas constant = 1.987 cal/(mol K), whereas temperature (T) was arbitrarily set at 25°C (298.15). For easier comparison, ΔG was defined in kcal/mol because of the reference database used in this unit.

4.3 | Calculation of surface and interface size

Interface size was defined as the change in accessible surface area (DASA) of the protein chains upon complex formation. We used the StrucTools web server (link: http://helixweb.nih.gov/structbio/basic.html, now it is only available within the NIH network: https://hpc.nih.gov/helixweb.html). This server uses Gerstein's accessible surface calculator[57] for calculating the contact areas based on the selected chains of the PDB structures of the complexes.

An amino acid was considered as part of the interface if change in its accessible surface area (difference between sum of the accessible surface area of amino acids in the complex and that of the separate chains) was more than 1 Å^2. We calculated the interface areas belonging to each chain in the complex separately, but in this article we usually talk about total interface defined as the sum of the two binding surfaces of the chains. Half-interface means the interface area belongs to single chains in the complexes.

Hydrophobic interface refers to the interface area of the hydrophobic amino acids.

4.4 | Classification of amino acids

We classified the amino acids in four groups based on their physico-chemical properties:

- Hydrophobic: ALA, LEU, MET, PHE, PRO, TYR, TRP, ILE, VAL.
- Polar: ASN, CYS, GLN, SER, THR, GLY.
- Positive: ARG, LYS, HIS.
- Negative: ASP, GLU.

4.5 | Interaction types

The number of different interaction types between the two partners in the complexes was calculated from the relevant chains of the PDB structures using Protein Interaction Calculator (PIC) web server: http://pic.mbu.iisc.ernet.in/job.html[58] for this purpose. For NMR structures only the first model was considered. This method calculates the following intermolecular interaction types: hydrophobic interactions, hydrogen bonds, ionic, and others (disulfide bridges, aromatic-aromatic, aromatic-sulfur, and cation-pi interactions) based on the proper distance of the appropriate amino acids.

4.6 | Weak and strong complexes

We classified a complex as strong if its dissociation constant was less than 1 μM, and weak if K_d was greater (or equal) than 1 μM (separated by $\Delta G = 8.2$ kcal/mol).

4.7 | Segmentation

Part of an interface belongs to the same segment if the distance between the consisting amino acids in the protein chain is no more than five amino acids.

4.8 | Conservation

For measuring evolutionary conservation, the first step was to find orthologs for the proteins of interest (for both globular and disordered proteins) from the OMA ortholog database (https://omabrowser.org/oma/home/).[59] In order to study the conservation of the proteins with the collected orthologs being comparable, the analysis was restricted to proteins from animal species. Only 1:1 orthologs (it means that the protein has only one ortholog in the other species and the orthologous entries have only one ortholog in this species) from the animal kingdom were taken into consideration (histone proteins were eliminated at this point, they had many:1 orthologs, resulting in a less redundant dataset). At least five orthologs were required for each protein originating from the same species for the two partners of a complex. After

these selection criteria, we ended up with 115 disordered and 34 globular complexes, on average with 30 orthologous protein pairs, with a minimum of seven and eight orthologs, respectively. The second step was to make multiple sequence alignments using MAFFT (Multiple Alignment using Fast Fourier Transform)[60] and calculate conservation scores for each position in the alignment with the algorithm developed by Capra and Singh[61] using their webserver (http://compbio.cs.princeton.edu/conservation/score.html). By mapping the conservation scores to the chains of protein structures, average conservation was calculated for both interfaces and surfaces of the protein partners.

In case of disordered complexes, average interface and surface conservation was obtained for 115 complexes (for the IDP chains there are conservation values only for 89 surfaces as for 26 cases there were no non-interface surface amino acids in the PDB structures) and 115 ordered chains. In case of globular complexes, results were obtained for 68 protein chains (34 complexes).

4.9 | iPat specificity

Surface accessible amino acid residues were scored for both chains of protein complexes based on the physico-chemical properties of surrounding residues using the iPat measure.[49] The iPat score measures the average propensity of residue types on the protein surface, grouped into hydrophobic, polar, negatively, and positively charged categories, to be located within a 12 Å radial distance around a central residue with a given property. This score represents the bias towards surface patch compositions typical to recognition sites of globular protein interfaces.

4.10 | Functional specificity characterized by semantic similarity

Semantic similarity was calculated on the "Biological process" GO annotations of both gene products participating in the formation of the protein complex using the GOSemSim v2.18.0[53,62] R package (R-project.org). As a similarity metric, the graph-based Wang method[63] was applied to compare pairs of GO terms. The semantic similarity scores for pairs of GO terms available for the partners were combined by using the average of maximum similarity on each row or column ("rcmax") of the similarity matrix. Only those protein complexes were included in this analysis that are formed between two human proteins, as this organism was by far the most abundant in both datasets. For this reason, we only used GO annotations of human genes loaded into R using the "org.Hs.eg.db" package version 3.14.[64] Only high-quality GO annotations were considered, and annotations having the following evidence codes were not taken into account: "IEA," "IC," "ND," "NAS," "IGC," "ISM," "ISA," "ISO," "ISS," "IRD," "IKR," "IBD," "IBA," "HEP," "IEP," "HGI," "IGI," "IMP," "HMP," "EXP," "HTP." A detailed guide to these evidence codes can be found at http://geneontology.org/docs/guide-go-evidence-codes.

AUTHOR CONTRIBUTIONS

Tamas Lazar: Conceptualization (equal); formal analysis (equal); methodology (equal); software (equal); writing – original draft (equal). **Agnes Tantos:** Supervision (supporting); writing – original draft (supporting). **Peter Tompa:** Conceptualization (equal); funding acquisition (equal); methodology (equal); project administration (equal); supervision (lead); writing – original draft (equal); writing – review and editing (equal). **Eva Schad**: Conceptualization (equal); data curation; visualization; formal analysis (equal); methodology (equal) software (equal); funding acquisition (equal); project administration (equal); writing – original draft (equal).

ACKNOWLEDGMENTS

This work was supported by grants K124670 and K131702 from the National Research, Development and Innovation Office (NKFIH, Hungary, to PT), Strategic Research Program on Microfluidics (SRP51) (VUB, Brussels, Belgium, to PT), an EC H2020-WIDESPREAD-2020-5 Twinning grant (PhasAge, no. 952334, to PT) and EC H2020-MSCA-RISE Action grant (IDPfun, no. 778247, to PT), and by the OTKA grant PD-OTKA 108772 to ES.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Peter Tompa https://orcid.org/0000-0001-8042-9939
Eva Schad https://orcid.org/0000-0002-3006-2910

REFERENCES

1. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci. 2002;27(10):527–533.
2. Tompa P. Unstructural biology coming of age. Curr Opin Struct Biol. 2011;21(3):419–425.
3. van der Lee R, Buljan M, Lang B, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014; 114(13):6589–6631.

4. Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol. 2009;19(1):31–38.
5. Chakrabarti P, Chakravarty D. Intrinsically disordered proteins/regions and insight into their biomolecular interactions. Biophys Chem. 2022;283:106769.
6. von der Haar T, Oku Y, Ptushkina M, et al. Folding transitions during assembly of the eukaryotic mRNA cap-binding complex. J Mol Biol. 2006;356(4):982–992.
7. Kussie PH, Gorina S, Marechal V, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science. 1996;274(5289):948–953.
8. Choi HJ, Huber AH, Weis WI. Thermodynamics of beta-catenin-ligand interactions: the roles of the N- and C-terminal tails in modulating binding affinity. J Biol Chem. 2006;281(2):1027–1038.
9. Gooding JM, Yap KL, Ikura M. The cadherin-catenin complex as a focal point of cell adhesion and signalling: new insights from three-dimensional structures. Bioessays. 2004;26(5):497–511.
10. Lacy ER, Filippov I, Lewis WS, et al. p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. Nat Struct Mol Biol. 2004;11(4):358–364.
11. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell. 1997;91(6):741–752.
12. Hurley TD, Yang J, Zhang L, et al. Structural basis for regulation of protein phosphatase 1 by inhibitor-2. J Biol Chem. 2007;282(39):28874–28883.
13. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W. Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure. 2010;18(9):1094–1103.
14. Morris OM, Torpey JH, Isaacson RL. Intrinsically disordered proteins: modes of binding with emphasis on disordered domains. Open Biol. 2021;11(10):210222.
15. Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. Bioinformatics. 2007;23(8):950–956.
16. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. Bioessays. 2009;31(3):328–335.
17. Meszaros B, Tompa P, Simon I, Dosztányi Z. Molecular principles of the interactions of disordered proteins. J Mol Biol. 2007;372(2):549–561.
18. London N, Movshovitz-Attias D, Schueler-Furman O. The structural basis of peptide-protein binding strategies. Structure. 2010;18(2):188–199.
19. Huang Y, Liu Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J Mol Biol. 2009;393(5):1143–1159.
20. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A. 2000;97(16):8868–8873.
21. Tompa P, Davey NE, Gibson TJ, Babu MM. A million peptide motifs for the molecular biologist. Mol Cell. 2014;55(2):161–169.
22. Tompa P, Szasz C, Buday L. Structural disorder throws new light on moonlighting. Trends Biochem Sci. 2005;30(9):484–489.
23. Dunker AK, Garner E, Guilliot S, et al. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput. 1998;473–484.
24. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol. 1999;293(2):321–331.
25. Huang Y, Liu Z. Do intrinsically disordered proteins possess high specificity in protein-protein interactions? Chemistry. 2013;19(14):4462–4467.
26. Shammas SL, Rogers JM, Hill SA, Clarke J. Slow, reversible, coupled folding and binding of the spectrin tetramerization domain. Biophys J. 2012;103(10):2203–2214.
27. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science. 2008;322(5906):1365–1368.
28. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. Curr Opin Struct Biol. 2011;21(1):50–61.
29. Kragelj J, Orand T, Delaforge E, et al. Enthalpy-entropy compensation in the promiscuous interaction of an intrinsically disordered protein with homologous protein partners. Biomolecules. 2021;11(8):1204.
30. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell. 2009;138(1):198–208.
31. Mahmoudabadi G, Rajagopalan K, Getzenberg RH, Hannenhalli S, Rangarajan G, Kulkarni P. Intrinsically disordered proteins and conformational noise: implications in cancer. Cell Cycle. 2013;12(1):26–31.
32. Kulkarni P, Rajagopalan K, Yeater D, Getzenberg RH. Protein folding and the order/disorder paradox. J Cell Biochem. 2011;112(7):1949–1952.
33. Beltrao P, Serrano L. Specificity and evolvability in eukaryotic protein interaction networks. PLoS Comput Biol. 2007;3(2):e25.
34. Tawfik DS. Messy biology and the origins of evolutionary innovations. Nat Chem Biol. 2010;6(10):692–696.
35. Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B. Protein disorder—a breakthrough invention of evolution? Curr Opin Struct Biol. 2011;21(3):412–418.
36. Neduva V, Russell RB. Linear motifs: evolutionary interaction switches. FEBS Lett. 2005;579(15):3342–3345.
37. Borgia A, Borgia MB, Bugge K, et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature. 2018;555(7694):61–66.
38. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol. 2004;338(5):1015–1026.
39. Kastritis PL, Moal IH, Hwang H, et al. A structure-based benchmark for protein-protein binding affinity. Protein Sci. 2011;20(3):482–491.
40. Landry CR, Levy ED, Abd Rabbo D, Tarassov K, Michnick SW. Extracting insight from noisy cellular networks. Cell. 2013;155(5):983–989.

41. Gunasekaran K, Tsai CJ, Nussinov R. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. J Mol Biol. 2004;341(5):1327–1341.
42. Ivarsson Y, Jemth P. Affinity and specificity of motif-based protein-protein interactions. Curr Opin Struct Biol. 2019;54:26–33.
43. Armon A, Graur D, Ben-Tal N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J Mol Biol. 2001;307(1):447–463.
44. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998;280(1):1–9.
45. Kisters-Woike B, Vangierdegom C, Muller-Hill B. On the conservation of protein sequences in evolution. Trends Biochem Sci. 2000;25(9):419–421.
46. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol. 1996;257(2):342–358.
47. Valdar WS, Thornton JM. Conservation helps to identify biologically relevant crystal contacts. J Mol Biol. 2001;313(2):399–416.
48. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. Curr Opin Struct Biol. 2011;21(3):441–446.
49. Lazar T, Guharoy M, Schad E, Tompa P. Unique physicochemical patterns of residues in protein-protein interfaces. J Chem Inf Model. 2018;58(10):2164–2173.
50. Lord PW, Stevens RD, Brass A, Goble CA. Semantic similarity measures as tools for exploring the gene ontology. Pac Symp Biocomput. 2003;601–612.
51. Wang JZ, du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. Bioinformatics. 2007;23(10):1274–1281.
52. Wu X, Zhu L, Guo J, Zhang DY, Lin K. Prediction of yeast protein-protein interaction network: insights from the gene ontology and annotations. Nucleic Acids Res. 2006;34(7):2137–2150.
53. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010;26(7):976–978.
54. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. Science. 1994;263(5148):777–784.
55. Raman AS, White KI, Ranganathan R. Origins of allostery and evolvability in proteins: a case study. Cell. 2016;166(2):468–480.
56. Schad E, Fichó E, Pancsa R, Simon I, Dosztányi Z, Mészáros B. DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. Bioinformatics. 2018;34(3):535–537.
57. Gerstein M, Richards FM. Protein geometry: volumes, areas and distances. In: Arnold E, Himmel DM, Rossmann MG, editors. International tables for crystallography, Vol. F. Chester, UK: International Union of Crystallography, 2012; p. 703–712.
58. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. Nucleic Acids Res. 2007;35(Web Server issue):W473–W476.
59. Altenhoff AM, Glover NM, Train CM, et al. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. Nucleic Acids Res. 2018;46(D1):D477–D485.
60. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–3066.
61. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics. 2007;23(15):1875–1882.
62. Yu G. Gene ontology semantic similarity analysis using GOSemSim. Methods Mol Biol. 2020;2117:207–215.
63. Pollack CV Jr, Sanders DY, Severance HW Jr. Emergency department analgesia without narcotics for adults with acute sickle cell pain crisis: case reports and review of crisis management. J Emerg Med. 1991;9(6):445–452.
64. Carlson, M., R package 2019, https://doi.org/10.18129/B9.bioc.org.Hs.eg.db.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lázár T, Tantos A, Tompa P, Schad E. Intrinsic protein disorder uncouples affinity from binding specificity. Protein Science. 2022;31(11):e4455. https://doi.org/10.1002/pro.4455