Review Article

On the specificity of protein–protein interactions in the context of disorder

Kaare Teilum1,2, Johan G. Olsen1,2,3 and Birthe B. Kragelund1,2,3

1The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark; 2Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark; 3REPIN, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

Correspondence: Birthe B. Kragelund (bbk@bio.ku.dk)

With the increased focus on intrinsically disordered proteins (IDPs) and their large interactomes, the question about their specificity — or more so on their multispecificity — arise. Here we recapitulate how specificity and multispecificity are quantified and address through examples if IDPs in this respect differ from globular proteins. The conclusion is that quantitatively, globular proteins and IDPs are similar when it comes to specificity. However, compared with globular proteins, IDPs have larger interactome sizes, a phenomenon that is further enabled by their flexibility, repetitive binding motifs and propensity to adapt to different binding partners. For IDPs, this adaptability, interactome size and a higher degree of multivalency opens for new interaction mechanisms such as facilitated exchange through trimer formation and ultra-sensitivity via threshold effects and ensemble redistribution. IDPs and their interactions, thus, do not compromise the definition of specificity. Instead, it is the sheer size of their interactomes that complicates its calculation. More importantly, it is this size that challenges how we conceptually envision, interpret and speak about their specificity.

Introduction

Molecular communication is at the base of life, and to understand life, we need to understand how information is conveyed by molecules through their interactions. In and between cells, communication operates through a myriad of proteins, which interact with other proteins and molecules, thereby translating information across molecular classes. Because of their key roles in communication, protein interactions have attracted both academic, scientific and industrial interest, developing methods for the detection of complex formation, deriving models to describe the observations, and developing drugs to target interactions connected to pathological states.

Early on, it became clear that a given protein may bind a number of structurally very similar ligands, but prefers one of these ligands by orders of magnitude, measured as substrate turn-over or binding affinity, in work pioneered by Emil Fischer [1]. This property to selectively interact with one ligand was termed specificity (Figure 1A). In the 1950’s, the binding of differently substituted benzoates to rabbit serum revealed a dependency on the chemical nature of the substituted benzoates, which was interpreted as specificity [2]. A similar observation was made for insulin binding to an insulin binding factor in serum where the presence of the many proteins and compounds in serum had no effect on binding of the insulin binding factor [3].

With the development of structural biology, which provided atomic resolution insight into protein–ligand interaction surfaces, the details of molecular communication were brought to light. Interactions at the atomic level could be resolved and their individual contributions to the binding energy determined. In a protein ligand complex, an interaction between e.g. the hydroxyl group of a threonine forming a hydrogen bond to a side chain of an aspartate would typically be termed a specific contact: a contact that would not happen in the absence of chemical complementarity.
Intrinsically disordered proteins (IDP) are proteins without a well-defined globular structure that are rich in polar and charged amino acids, and depleted in hydrophobic ones, leading them to populate ensembles of near-isoenergetic dynamic conformations [4,5]. Around the year 2000, their abundance and importance became acknowledged through bioinformatic scrutiny of the human genome and subsequent experimental confirmation [6–10]. The past 20 years of research have revealed a plethora of properties of IDPs, which are linked to the flexibility and dynamics of the individual structural ensembles. These dynamic properties expand the communication toolbox. Not only are IDPs capable of adapting to binding surfaces through coupled folding and binding [11–15], they can also remain disordered within a complex [16,17], even to a degree where specific chemical contacts are lacking [18]. Similarly, dynamics enables ultra-sensitivity and switch functionality as exemplified by the interaction of the cyclin-dependent kinase inhibitor Sic-1 with Cdc4. This interaction occurs through any one of a number of Sic-1 phosphoryl groups in an apparent non-specific way. However, a cumulative effect from increasing the number phosphorylations provides a threshold for binding [19,20] (Figure 1B).

The same binding region in an IDP may have the capacity to bind several different partners with very similar affinities [21]. This property has suggested that most IDPs are multispecific; a term which has been...
conceptually useful, as it contrasts the concept of being specific, where one ligand is preferred over many others. Nevertheless, the quantitative meaning of multispecificity is not well established. For transcriptional activators and co-activators, a prevailing view has been that these act non-specifically, the argument being that the density of hydrophobic and negatively charged residues combined with the dynamic nature of their complexes would not provide the necessary specificity and hence would not lead to selectivity (Figure 1B). Recent work, however, suggests that specific interactions do exist in these dynamic complexes, and that redistribution of the conformational ensemble of the complex and interactions with residues outside the main interaction surface play key roles [22]. This, and many other studies have shown that the context of which given binding site is found in is emerging as an important dimension in IDP interactions [23–25].

Proteins have more than one binding partner and are thus multispecific (Figure 1C,D) [26]. This property is essential for the function of communication foci in the cytoplasm where hub-proteins interact with a multitude of ligands, in some cases hundreds [27,28]. This raises the question of how specificity can be maintained. In hubs, structural disorder dominates [29–31], where either the hub protein itself is disordered, as exemplified by the protein p53 [21], or the ligands are, as observed for the folded hub proliferating cell nuclear antigen (PCNA) [32,33] or the group of 14-3-3 proteins [34,35]. Thus, structural disorder appears to be a molecular tool for obtaining multispecificity, and understanding how specificity is achieved as well as its quantification remain important.

One of the early characteristics that became associated with IDPs was that they bound with low affinity, but high specificity. However, recent analyses have highlighted that when comparing complexes formed by globular proteins to those formed between an IDP and a globular protein, their affinities covered the same range [36]. It was shown that for complexes where one of the interaction partners was an IDP that folded-upon binding, the binding free energies were moderately compromised, on average by 2.5 kcal mol$^{-1}$, which was suggested to originate from the loss of conformational entropy [36]. Furthermore, tight complexes of nM affinities can form between short linear motifs (SLiMs) in IDPs and folded proteins [37,38], and pM affinities are possible in complexes formed between two IDPs of opposite charge forming highly dynamic complexes [18]. In this light, it appears that IDPs do not adopt a different affinity profile strategy from folded proteins. The question is then how an IDP with a large number of biologically relevant binding partners can be specific in the formal sense of conveying relevant biological information via a relevant binding partner.

Here we address the question if IDPs challenge our concepts of specificity and multispecificity, and if IDPs tend to be more or less specific than their globular counterparts. We describe the concepts of specificity and multispecificity and relate these to how IDPs partake in molecular communication. In the present review, we reason that IDPs are not particularly different from globular proteins in terms of specificity. Rather, it is their larger interactomes [39] that challenge the quantification and perception of their specificity.

**Specificity, affinity and biological relevance**

A protein capable of binding several different ligands is referred to as multispecific or promiscuous. Protein chemists may even talk about ‘very specific’ or ‘highly multispecific’ proteins. While an intuitive understanding of such expressions seems to be shared in the community, a precise definition is lacking. Two extreme degrees of specificity can be envisaged. One extreme would be a protein that binds to one single ligand only, which would make it absolute specific. The other is a protein that binds everything it comes into contact with, which would make it absolute non-specific. Both extremes are likely purely hypothetical, and all proteins exist somewhere along this continuum. An intrinsically disordered hub-protein may for example be able to bind many different ligands with similar affinities. If all those ligands are accessible in equimolar amounts at the same time, then, by definition, the hub protein shows very low specificity. However, the relative concentration levels and accessibility of the ligands are likely tightly adjusted depending on the biological state of the given system. If the hub protein is presented with only one of its ligands in a crowded cellular environment, then this ligand will be the only binding partner, and then, by definition, the hub-protein is highly specific. The affinity of a complex on its own is thus not a measure of specificity although it provides basic knowledge of the energetics in the process. With their high flexibility, IDPs can bind and adapt to a large surface area on a binding partner and in this way make favorable contacts, which result in high affinity [40–43]. The affinity alone, however, does not dictate which one of a series of ligands occupies the binding site of the protein most of the time, and higher affinity does not necessarily result in higher specificity. As we discuss further below, both the concentration of a ligand and the affinity for the ligand relative to the same parameters for other ligands determine the specificity.
Specificity can also be considered in terms of the molecular details of binding, referred to as chemical specificity or complementarity (Figure 1A). The interface between two protein molecules that interact may contain a number of salt bridges, hydrogen bonds, van der Waals’ interactions etc.; each interface requiring from the other a certain distribution of functional groups in three-dimensional space, i.e. the interfaces are complementary. An interface may also, either exclusively or in addition, consist of long-range interactions between complementary charges, the exact distribution of which may not change the energy of binding. For example, certain DNA binding proteins recognize and bind a unique sequence of nucleotides. Such a complementary binding interface is an example of chemical specificity. Assuming a protein that has a single available target sequence on DNA, it can be regarded as highly specific. The chemical specificity is ensured by the sequence preference. However, parts of the protein outside the DNA base-pair binding region interact with the phosphate backbone, contributing energy to the complex. In itself, the phosphate backbone interaction is independent of the base-pair combination, and so it can be argued that this part of the protein does not contribute to the chemical specificity, but because the interaction changes the binding \( K_d \), the chemically unspecific regions may actually increase the specificity. Another typical example is the interaction between a membrane-bound receptor and a soluble ligand, which is specific, and the interaction between the soluble ligand and the membrane itself, which is chemically non-specific (Figure 1A).

Finally, there is the question of biological relevance. Non-specific interactions may very well be biologically relevant. The molecular storm or noise inside a living system must result in molecules continually interacting, and although many of these so-called quinary interactions [44,45] have no immediate biological output, they necessarily must constitute an evolutionary driving mechanism of fundamental importance to biology for developing new communication lines. Thus, even highly non-specific, low affinity and accidental interactions can have biological relevance. This means that the degree of specificity, the magnitude of the \( K_d \)s and biological relevance are not necessarily correlated.

**Specificity can be quantified**

The most basic parameter for describing a binding reaction is the affinity, which is a measure of the binding strength. For a simple one-step reaction involving one protein and one ligand \((P + L \rightleftharpoons PL)\), the strength of the binding is given by the equilibrium constant for the reaction, called the association constant, \(K_a\) or, more frequently, the dissociation constant \(K_d\), which is the equilibrium constant for the reverse reaction (and \(K_a = 1/K_d\)). It has the value of the concentration of free ligand that results in half saturation of the protein.

For a protein binding two different ligands, two binding equilibria exist, each with an associated \(K_d\) value:

\[
P + L_1 \rightleftharpoons PL_1 \quad K_{d1} = [P][L_1]/[PL_1] \\
P + L_2 \rightleftharpoons PL_2 \quad K_{d2} = [P][L_2]/[PL_2]
\]

Here, \(P\) is a protein, and \(L_1\) and \(L_2\) are two different (protein) ligands. The simplest and most straightforward way to compare the binding of the two ligands is by taking the ratio of the two \(K_d\) values. This ratio is called the selectivity or discrimination ratio, \(d = K_{d1}/K_{d2}\) and gives the concentration of ligand \(L_1\) relative to ligand \(L_2\) needed to get the same level of saturation of the protein by the two ligands [46]. The selectivity, however, does not take the actual concentrations of free ligands into account, so the value itself cannot directly be used to assess which of the complexes, \(PL_1\) or \(PL_2\), that will dominate in a given situation and hence the specificity.

In the case where a protein has more than one potential ligand, the quantification of how much one complex is favored over others, the specificity factor, \(\alpha_1\), is defined, in which the free ligand concentrations are included. For the two binding reactions above, we have [47]:

\[
\alpha_1 = [PL_1]/[PL_2] = ([L_1]/K_{d1})/([L_2]/K_{d2})
\]

and in the more general case where several alternative ligands \(L_i\) can bind, this becomes:

\[
\alpha_1 = [PL_1]/\Sigma[PL_{(i>1)}] = ([L_1]/K_{d1})/\Sigma([L_{(i>1)}]/K_{d,(i>1)})
\]

Here, the sum in the denominator is over all ligands except ligand 1. \(\alpha_1\) thus depends both on all the affinities and on the concentrations of all the free ligands. This means that \(\alpha_1\) will change if the free concentration of
any one ligand changes. The specificity factor thus considers the situation where the ligand with the highest affinity may be present at extremely low concentration, and vice versa. This means that in order to evaluate the specificity of a certain protein quantitatively, we need the $K_d$ values and free concentrations of all possible biologically relevant ligands in the cell at any given time. Such information is rarely, if ever, available.

**Specificity in relation to IDPs**

Variations in the free concentrations of a set of ligands capable of binding to the same IDP can affect biological outcome. To illustrate this, we will use the tumor suppressor protein p53 as an example and look at a subset of the many p53 binding partners, specifically the $K_d$ values for p53 binding and their concentrations in the cell. The p53-derived peptides used in the different studies vary in length, but the interaction region is the same when in complex with different folded domains. The purpose here is to illustrate how varying concentrations of the competing ligands change biological outcome.

The p53 tumor suppressor is a transcriptional activator, that contain ordered and disordered discrete domains participating in sequence-specific DNA binding, tetramerization, or transcriptional activation [48] (Figure 2A). One of its domains, the transcriptional activation domain 2 (TAD2) engages in complexes with many differently folded ligands including, but not restricted to, ligands active in transcriptional machinery complexes, chromatin modification and DNA homeostasis (Figure 2B). Their affinities were measured through isothermal titration calorimetry, fluorescence polarization, or nuclear magnetic resonance spectroscopy and range from 7.5 mM to 300 nM [49–54]. Based on the observed affinities alone (Figure 2C), it is hard to imagine how the weakest binding ligand would ever stand a chance. To address the specificity for a selected

![Figure 2. Part of the interactome of p53 and illustration of the role of concentration in specificity.](image-url)

(A) domain structure of p53 with its TAD2 domain highlighted in red. (B) Variations in bound-state structures of the p53–TAD2 domain (in red) in complex with different folded domains (in grey) from RPA (the N-terminal DBD domain (DBD-F); PDB 2B3G [49]), TFIH (The TFB1 domain; PDB 2GS0 [50]), CBP (the TAZ2 domain; PDB 2MZD [51]), HMGB1 (the A-box; PDB 2LY4 [52]) and p62 (The PB1 domain; PDB 2RUK [148]). The N-terminus of p53 is to the left in each figure. (C) Association constants $K_a$ for p53 TAD2 from [49–53]. (D) Cellular concentrations taken from PaxDB [55] in two different cell lines HEK293 and HeLa (Geiger) converted from ppm to µM [58]. (E) Specificity of p53 given by the specificity factor $\alpha$ for different ligands of the p53–TAD2 in two different cell lines. HEK293 cells in black and HeLa cells in red.
subset of TAD2-binding ligands (Figure 2A), the cellular concentrations of their free states are required. We simulated these concentrations from the $K_{dS}$ and their total cellular concentrations in two different cell lines (HEK293 and HeLa cells) available in the PaxDB database [55,56], using the software Copasi [57]. The abundance values are reported in ppm (parts-per-million), which were converted to μM as described [58] (Figure 2D). From this, we calculated the specificity factor for each ligand in the presence of the others, disregarding any other contact among the ligands (Figure 2E). Several observations are relevant. First, it is evident that the specificity factor for most ligands is below 1, but dominated by HMGB1, the concentration of which is much higher than any of the other ligands. This could not be concluded from inspection of the affinities alone (Figure 2B). Second, changing the concentrations from those found in HEK cells to those in HeLa cells changed the specificity factor for all ligands, and hence will change the resulting biological outcome. In this particular case, p53 binding to RPA and PC4 is less likely in HeLa cells, which proliferate abnormally fast. Indeed, p53 binding to RPA suppresses homologous recombination [59] and PC4 binding to p53 enhances its tumor suppresser function [60], so different specificity profiles manifest in the phenotype. A more complete model of this system would have to consider all the different ways the available pool of free ligands can be modified. Modification can be achieved via degradation, rate of synthesis, interaction with other molecules or chemical modifications such as phosphorylation and methylation. As an example, acetylation of PC4 enhances, while phosphorylation blocks, its ability to activate p53 DNA binding [60]. In such cases, it becomes increasingly difficult to predict the actual biological output. The example illustrates that especially for IDPs, interactions per se do not compromise the definition of specificity, but the size of their interactomes [39] complicates its quantification.

Multivalent interactions

In the above, we have implicitly assumed that the ligands compete for the exact same binding site. In many cases, however, IDPs have several binding sites next to one another, which complicates the picture somewhat (Figure 1C). The core binding sites often take the form of short linear motifs (SLiMs) that either may be very different in sequence and bind different ligands, or very similar in sequence and bind the same ligand [24,61]. The multiple binding sites or domains make the protein multivalent (Figure 1D) and a quantitative description of the interactions becomes complex [62–65]. Multivalent binding reactions occur stepwise, but not necessary in a predetermined sequence of events. The first binding event may significantly alter the properties of the interacting molecules and perturb the strength of the subsequent binding events, resulting in an overall macroscopic binding equilibrium described by the avidity constant $K_{av}$.

$$K_{av} = \frac{[P_{\text{bound}}]}{([P_{\text{free}}])(L)}$$

If the binding sites in the multivalent system are independent, $K_{av}$ will be the same as $K_d$ for the monomeric interaction as long as $[P]$ refers to the total concentration of binding sites. $K_{av}$ may, however, become much larger than for a monomeric interaction. As much as a $10^{17}$-fold increase has been reported for a fairly simple system of a trivalent peptide binding to vancomycin [66]. The avidity in a system depends on how the individual binding sites are connected. The avidity also depends on how the initial binding and the linker(s) between the binding sites modulate the entropy (dynamics) of the system, and whether the linker becomes structured or not upon binding of a ligand [67]. All these thermodynamic parameters can be rationalized, although they may be hard to quantify experimentally.

Closely connected to the avidity of a multivalent system is the effective concentration, which is a more operational measure of the strengths of binding sites in multivalent interactions. Where the initial binding event is dependent on the concentrations of the interacting molecules, this is not the case for the subsequent intra-complex interactions. The spatial tethering of subsequent binding sites in a multivalent binding reaction infers that the probability of two binding sites forming a constructive interaction is greatly increased. The effect can be quantified by measuring the effective concentration, which is the concentration of the ligand-binding domain in an untethered monomeric form needed to outcompete half of the intra-complex binding of the tethered domains. Methods to experimentally measure the effective concentration have been devised [68,69] and can be used to explain both the thermodynamics and the structural changes in the binding process [70].

A key feature of multivalent binding is that several weak binding sites can — when they are linked together — result in very strong binding and lead to a steep binding transition over a narrow range of ligand...
concentration in what has been named super-selectivity [71]. For multivalent IDPs, extremely strong binding is typically not desirable, as it would imply that the molecules in the complex would practically never dissociate [65]. Multivalent IDPs appear to have evolved to circumvent this disadvantageously strong binding by taking advantage of increased dynamics in the complex. One illustrative example is the behavior of the interaction between the FG-repeat proteins in the nuclear pore transport complex and the molecular cargo that has to be transported through the nuclear pores [72,73]. The FG-repeat proteins are disordered and have multiple (up to 50) similar SLiMs (Phe–Ser–Phe–Gly) connected by disordered linkers. They line the rim of the nuclear pores, where they bind cognate transport factors through multiple interactions with the SLiMs. For one transport factor, NTF2, the interactions with variants harboring a variable number of FSFG SLiMs, was studied. As the number of repeats went up, so did the affinity, but there was no affinity gain by adding more than four repeats [74]. Thermodynamic analysis showed that the favorable enthalpic contribution to the binding continues to increase up to at least 12 repeats. However, this is counteracted by an increasing unfavorable loss of entropy as the flexible linkers joining the repeats become ordered [74]. Based on the thermodynamic parameters, the effective concentrations of the FG-repeats reach more than 32 mM when six repeats are present [74]. It is noteworthy that this is near the actual concentration of FG-repeats estimated, which may be as high as 50 mM in the center of the nuclear pore complexes [73]. The system has evolved to give just the right affinity to keep the transport factors within the nuclear pore transport complex, but in a highly dynamic way that allows the transport factors to slide through the pore by transient interactions with the many FG-repeat proteins [75]. The transport process is, however, not fully understood. It turns out that purified FG-repeat proteins may undergo liquid–liquid phase separation (LLPS) in vitro and form droplets that have similar permeability for proteins known to be transported through the nuclear pore complex [76]. Whether the FG-repeat proteins form the same phase separation in vivo remains to be clarified.

Condensate formation and LLPS is a direct result of multivalency, either acting through intermolecular interactions between the same protein (homotypic) or between different proteins (heterotypic) [77–79] (Figure 1D). Condensates form in a concentration dependent manner and is driven by low-affinity interactions between several sites, often of low sequence complexity [80–82]. This behavior results in a distribution of affinities in the dense phase, as observed from power-law distributions of dwell-times within the condensates [83]. These membrane-less micrometer-scale compartments have liquid-like properties [84] and are widespread in eukaryotic cells, with the phenomenon being observed and described for different types of proteins and organisms both in vivo and in vitro (for recent reviews, see e.g. [80,82,85–87]). Disordered regions are important co-actors and drivers in these processes, and multivalent IDR-IDR contacts are often seen. However, the process is not restricted to IDPs [88–90], but more to the property of multivalency (Figure 1D), although disordered linkers as well as IDPs play additional roles in this process [91]. RNA and DNA are inherently highly multivalent, and LLPS phenomena involving proteins typically active in RNA and DNA metabolism are currently some of the most thoroughly described. There is, for example, growing evidence for the role of phase separated condensates in transcription [92–95] and in activation domains of transcription factors, such as the estrogen receptor [93] and the glucocorticoid receptor [83]. In all cases, the spatiotemporal confinement and the resulting population increase in the condensates, increase the concentration from typically low μM in the dilute phase, to high μM–mM concentrations in the dense phase [96–100]. Moreover, the condensates are dynamically formed and sensitive to environmental factors such as phosphorylation, pH and temperature, which can lead to their swift dissolution [101–104]. Thus, condensate formation may affect specificity in several ways. One is the compartmentalization, which leads to an effective elimination of potential competing ligands and confinement in a restricted reaction zone. Another is through the large increase in concentration, which greatly increases the specificity factor (as defined above), for ligands available in the condensates. Depending on the mechanism of interaction, the increased concentration also affects the rate of complex formation via an increase in the on-rate for binding, resulting in faster regulation. Thus, condensate formation can be a relevant biological avenue for specificity regulation. Indeed, specificity in condensate formation has been seen [94]. However, how chemical specificity is achieved in condensed phases is still an ongoing question, although patterning, affinity and valency play key roles [105–107].

Multivalency and multispecificity are also properties of globular hub proteins

DNA replication, translesion synthesis, homologous recombination, mismatch repair, chemical modification, transcription and more, are all processes that require the presence of a DNA clamp, a pivotal hub protein.
In eukaryotes, the DNA clamp is the proliferating cell nuclear antigen, PCNA [108]. It is folded into a ring-shaped trimer (Figure 3), which is loaded onto DNA to encircle the double helix [109]. In this position, it can slide along the DNA strands and recruit and position ligands within its three binding sites to enable whatever DNA operating molecular machinery is required, similar to a DNA toolbelt.

A curated list of ∼80 interaction partners was recently collected along with an analysis of PCNA–ligand interactions at the molecular level [32]. Most of these ligands are all either IDPs or intrinsically disordered regions (IDRs) of proteins with disorder in and around the PCNA binding SLiM [32]. The ligands fold into the PCNA interacting protein (PIP) binding pockets, mostly as a single $3_{10}$ helical turn (Figure 3). The PCNA binding SLiM is referred to as the PIP-box [33], although some sequence variants are referred to as a PIP-degron [110,111] or an APIM (AlkB homolog 2 PCNA-interacting motif) [112]. The interactions in the PIP pockets constitute the central binding site. However, despite similar SLiMs, binding affinities for PCNA can vary 3–4 orders of magnitude between ligands. Small peptides comprising only the PIP-box region, have $K_d$ values as high as 3 mM as in the case of an Spd1 derived peptide, and as low as 6 nM for a p21 derived peptide [32].

With >80 ligands harboring more or less the same motif, the specificity for PCNA binding as well as its selectivity may reside elsewhere. Indeed, the binding pockets only constitute a minor part of the large surface of PCNA and recent work has shown that flanking regions of the disordered ligand can modulate affinities by four orders of magnitude [32]. Interactions between the flanking regions of the motif and residues outside the PIP pockets are mainly electrostatic [32]. These interactions are rarely seen in the crystal structures, but they contribute significantly to the binding affinity.

Figure 3. Multivalency and multispecificity is also a property of globular domains.
The PCNA trimer (in grey) is shown in the central part with three Cdt1 ligands bound (in red). The canonical PIP box motif is shown below. With a few possible exceptions, all ligands have the motif within an intrinsically disordered region and bind to the same site on PCNA forming a $3_{10}$ turn and with varying structures and dynamics in the flanking regions. Here, seven crystal structures of PCNA and small ligand peptides are shown with their PIP box motif indicated below. The bound ligands and the corresponding PDB codes are: Left panel: DNA polymerase iota (2ZVM), RNase H2B (3P87), and p15 (6GWS); central panel: Cdt1 (6QCG); right panel: DNA pol delta (1U76), DNA pol eta (2ZVK), and DNA pol kappa (2ZVL).
PCNA is a fine example of a multivalent folded protein where multispecificity is possible when the ligands are disordered. In this case, the energy of binding to the central complementary PIP-binding pocket is supplemented by contributions from dynamic charge-charge interactions in the flanking regions of the disordered ligands [32]. Thus, different post-translational modifications, primarily phosphorylation, both of PCNA and its ligands, regulate binding [113,114]. So even though p21 is by far the strongest PCNA-binding ligand, obviously DNA replication and all other PCNA dependent processes also happen. The presence of disorder in the ligands will allow for more rapid exchange in the binding pockets (see below) and provide access to modifying enzymes, including those leading to ubiquitination and proteasomal degradation. In this way, the α-factors for the ligands can be modulated both directly and indirectly.

**Specificity of highly charged IDPs**

Although original work suggested IDPs to be more charged and polar than globular proteins [115–117], it is becoming evident that the sequence properties of IDPs are more diverse [118], where some have minimal charges, as seen in the low complexity regions of FUS and hnRNPA1 [119], and others have additional enrichment in proline [120] or glutamine [121,122]. A particular group of IDPs is highly charged with a large surplus of either positive or negative charges, resulting in a large net charge per residue. These IDPs still form many different types of complexes including those with DNA [16,123]. One example is the linker histone H1 and its chaperone prothymosin α that form a high-affinity complex in which both proteins remain as dynamic and disordered as in their free state [18,124]. At low concentration, the complex forms a 1:1 complex of nM affinity. At higher concentration, the dynamics of the complex makes it possible for a second ligand (of either partner) to enter the complex, forming ternary (or higher order) complexes of lower affinity (Figure 4) [125].

The trimers are only transient, and one of the ligands sharing the interaction will remain and reestablish a new nM affinity 1:1 complex. This can happen much faster than one would expect from a ‘normal’ nM affinity interaction with long-lived established contacts at the interface, and with a concentration dependent dissociation rate constant. As a result, the exchange rate in the complex changes more than 45-fold compared with a two-state reaction [125], which has implications for affinity measurement at higher protein concentrations [124]. This mechanism has been referred to as facilitated ligand exchange via competitive substitution [125,126]. The highly charged complexes remain dynamic with apparent non-specific contacts in a mean-field type interaction [18].

A similar concentration dependent behavior was observed for the binding between CuII ions and Aβ, which in its monomeric soluble state is a disordered polypeptide. At certain conditions, Aβ forms fibrils that are the main constituent of plaques in the brains of patients with Alzheimer’s disease [127]. Aβ binds CuII with nM affinity [128]. At low concentrations of Aβ, the CuII-ions dissociate slowly from the complex. However, as the Aβ concentration increases, CuII starts to dissociate faster in a process that depends on the concentration of Aβ through the formation of a ternary Aβ-CuII-Aβ complex [129]. This demonstrates how the components of apparently very stable multimeric species still may exchange in and out of the complex in a highly dynamic way.

Ligand exchange via competitive substitution where transient trimers (or higher order oligomers) are involved in swapping binding partners has been observed in other types of complexes, some also involving folded domains. Examples include facilitated exchange of DNA bound proteins, transcription factors and transcriptional coactivators [130–133], where complex affinity switching has been suggested to be facilitated by dynamics in the folded partner [133,134], and of highly charged antigenic peptides from MHC class II in antibody binding sites, in which case the exchange reaction was termed push-off [135]. In all cases, dynamics at the interface underlies the possibility for enhanced exchange allowing access to interactions for incoming ligands (Figure 4). Thus, the mechanism of facilitated exchange through competitive substitution is not limited to IDPs but requires that the binding site is dynamic.

Although highly charged IDPs may be expected to bind to almost any oppositely charged ligand, they may still be specific. This may be compared with the nuclear transport proteins that travel through the pore down the concentration gradient through many low-affinity encounters with several different proteins [74,75]. However, for the highly charged proteins, specificity in terms of specific contacts may not exist, but may instead depend on other yet to be defined properties such as the number of charges, their distributions along the chains or on surfaces, or on concentration and localization in vivo.
Conclusions

In this paper we have asked if IDPs are different from globular proteins in terms of specificity. The answer is that they are not. Quantitatively, specificity for globular proteins and IDPs is similar. What differs is the size of their interactomes [39] and their adaptability and dynamics. As a result, the values of the specificity factors for protein–ligand systems involving IDPs are much lower for the individual complexes. This is a direct consequence of their large interactomes. Simply by being capable of adapting to more partners, even using the same binding site through composite SLiMs, IDPs increase the number of partners and hence parameters in the system, and this complicates calculation of their specificity factors. Conceptually, however, it is the sheer sizes of their interactome rather than their dynamic nature, that challenge how we envision, interpret and speak about specificity.

In light of these considerations, we can now return to the early statement of low affinity and high specificity as a generic property for IDPs. There may in fact be some truth to it. Although we have seen many IDPs binding with high affinity, even in the low or sub nM range [32,37,38], the higher degree of multispecificity and potentially lower specificity factors, combined with high adaptability, may indeed require low affinity for
generating high specificity. In this case, to increase the value of the specificity factor α for a particular ligand, the affinity for all competing ligands must be even lower. Whereas this will be the case for some IDPs, another suggestion has been that multiple SLiMs present in an IDP will increase specificity. From the above considerations, such SLiM trains will indeed increase the local concentrations of a particular binding site and through this increase the avidity, and even give rise to super-selectivity. However, it does not a priori increase specificity. The protein may still have many partners where similar exploitation of avidity happens and hence in this way, the protein remains multispecific.

Multispecificity appears to be an evolutionary advantage and the functions of many different proteins including enzymes, antibodies, the ubiquitin system, G-coupled protein receptors and more, rely on it [26,136–138]. One of the advantages may be to increase network resilience, as higher specificity in terms of smaller interactomes have been suggested to result in decreased network pliability and malleability [139]. For IDPs, multispecificity is coupled to their flexibility and adaptability, and as important drug targets with key roles in disease [140,141], this cocktail constitutes a challenge for their drugability. Proof-of-concept in targeting of IDPs by small molecules has, however, been made [142–144], and new mechanisms for drug targeting have emerged, such as targeting a folded binding partner or the IDP-complex [145], or via ensemble expansion and entropy optimization [146,147]. Thus, it is not their multispecificity that obstruct targeting of IDPs, but more so their highly dynamic ensembles.

In conclusion, the main difference between IDPs and globular proteins roots back to the property of flexibility and adaption and to the propensity to remain dynamic in a complex. These properties enable a higher degree of multivalency and larger interactomes, and open for new interaction mechanisms through dynamics such as facilitated ligand exchange through competitive substitution, ultra-sensitivity and ensemble redistribution. More importantly, these properties challenge how we envision their specificity and speak about it, and how we unravel their biology and target them in disease.

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

Funding
This is a contribution from REPIN funded by the Novo Nordisk Challenge program (Grant #NNF18OC0033926 to BBK). K.T.E. was funded by Independent Research Fund Denmark (Grant #8020-00099B).

Acknowledgements
We thank all members of REPIN for their continued contributions to discussions and Karen Skrøver for thoughtful feed-back. Asta B. Andersen is thanked for graphics support.

Abbreviations
IDP, intrinsically disordered proteins; LLPS, liquid–liquid phase separation; PCNA, proliferating cell nuclear antigen; PIP, PCNA interacting protein; SLiMs, short linear motifs; TAD2, transcriptional activation domain 2.

References
1 Fischer, E. (1894) Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der Dtsch. Chem. Gesellschaft 27, 2985–2993 https://doi.org/10.1002/cber.18940270364
2 Pressman, D. and Siegel, M. (1953) Specificities of the binding to rabbit serum. Arch. Biochem. Biophys. 45, 41–54 https://doi.org/10.1016/0003-9861(53)90403-2
3 Grodsky, G.M.; Peng, C.T. and Forsham, P.H. (1959) Effect of modification of insulin on specific binding in insulin-resistant sera. Arch. Biochem. Biophys. 81, 264–272 https://doi.org/10.1016/0003-9861(59)90195-X
4 Wright, P.E. and Dyson, H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18–29 https://doi.org/10.1038/nrm3920
5 Bahu, M.M., van der Lee, R., de Groot, N.S. and Gsponer, J. (2011) Intrinsically disordered proteins: regulation and disease. Curr. Opin. Struct. Biol. 21, 432–440 https://doi.org/10.1016/j.sbi.2011.03.011
6 Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., et al. (2001) Intrinsically disordered protein. J. Mol. Graph. Model. 19, 26–69 https://doi.org/10.1016/S1093-3263(00)00138-8
7 Uversky, V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 11, 739–756 https://doi.org/10.1110/ps.4210102
8 Dyson, H.J. and Wright, P.E. (2002) Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. 12, 54–60 https://doi.org/10.1016/S0959-440X(02)00289-0

© 2021 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND). 2045
Potoyan, D.A., Zheng, W., Komives, E.A. and Wolynes, P.G. (2016) Molecular stripping in the NF-κB/IκB/DNA genetic regulatory network. Proc. Natl Acad. Sci. U.S.A. 113, 110–115 https://doi.org/10.1073/pnas.1520483112

Berlow, R.B., Dyson, H.J. and Wright, P.E. (2017) Hypersensitive termination of the hypoxic response by a disordered protein switch. Nature 543, 447–451 https://doi.org/10.1038/nature21705

Berlow, R.B., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2019) Role of backbone dynamics in modulating the interactions of disordered ligands with the TAZ1 domain of the CREB-binding protein. Biochemistry 58, 1354–1362 https://doi.org/10.1021/acs.biochem.8b01290

Schmitt, L., Kratz, J.R., Davis, M.M. and McConnell, H.M. (1999) Catalysis of peptide dissociation from class II MHC-peptide complexes. Proc. Natl Acad. Sci. U.S.A. 96, 6581–6586 https://doi.org/10.1073/pnas.96.12.6581

Dasgupta, B., Nakamura, H. and Kinjo, A.R. (2013) Counterbalance of ligand- and self-coupled motions characterizes multispecifity of ubiquitin. Protein Sci. 22, 168–178 https://doi.org/10.1002/pro.2195

Schmitt, L., Kratz, J.R., Davis, M.M. and McConnell, H.M. (1999) Catalysis of peptide dissociation from class II MHC-peptide complexes. Proc. Natl Acad. Sci. U.S.A. 96, 6581–6586 https://doi.org/10.1073/pnas.96.12.6581

Berlow, R.B., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2019) Role of backbone dynamics in modulating the interactions of disordered ligands with the TAZ1 domain of the CREB-binding protein. Biochemistry 58, 1354–1362 https://doi.org/10.1021/acs.biochem.8b01290

Fonin, A.V., Darling, A.L., Kuznetsova, I.M., Turoverov, K.K. and Uversky, V.N. (2019) Multi-functionality of proteins involved in GPCR and G protein signaling: making sense of structure-function continuum with intrinsic disorder-based proteoforms. Cell. Mol. Life Sci. 76, 4461–4492 https://doi.org/10.1007/s00018-019-03276-1

Biou, V. and Cherfils, J. (2004) Structural principles for the multispecificity of small GTP-binding proteins. Biochemistry 43, 6833–6840 https://doi.org/10.1021/bi049630u

Alhindi, T., Zhang, Z., Rueleff, P., Coenen, H., Degroote, H., Iraci, N. et al. (2017) Protein interaction evolution from promiscuity to specificity with reduced flexibility in an increasingly complex network. Sci. Rep. 7, 44948 https://doi.org/10.1038/srep44948

Kulkarni, P. and Uversky, V. (2019) Intrinsically disordered proteins in chronic diseases. Biomolecules 9, 147 https://doi.org/10.3390/biom9044017

Vacic, V. and Iakoucheva, L.M. (2012) Disease mutations in disordered regions—exception to the rule? Mol. Biosyst. 8, 27–32 https://doi.org/10.1039/C1MB05251A

Santofimia-Castaño, P., Xia, Y., Peng, L., Velázquez-Campoy, A., Abián, O., Lan, W. et al. (2019) Targeting the stress-induced protein NUPR1 to treat pancreatic adenocarcinoma. Cells 8, 1453 https://doi.org/10.3390/cells8111453

Neira, J.L., Bintz, J., Arruebo, M., Rizzuti, B., Bonacci, T., Vega, S. et al. (2017) Identification of a drug targeting an intrinsically disordered protein involved in pancreatic adenocarcinoma. Sci. Rep. 7, 39732 https://doi.org/10.1038/srep39732

Estébanez-Perpiñá, E., Arnold, L.A., Arnold, A.A., Nguyen, P., Rodrigues, E.D., Mar, E. et al. (2007) A surface on the androgen receptor that allosterically regulates coactivator binding. Proc. Natl Acad. Sci. U.S.A. 104, 16074–16079 https://doi.org/10.1073/pnas.0708036104

Wolter, M., Valenti, D., Cossar, P.J., Levy, L.M., Hristeva, S., Genski, T. et al. (2020) Fragment-based stabilizers of protein-protein interactions through imine-based tethering. Angew. Chem. Int. Ed. Engl. 59, 21520–21524 https://doi.org/10.1002/anie.202008585

Heller, G.T., Sormanni, P. and Vendruscolo, M. (2015) Targeting disordered proteins with small molecules using entropy. Trends Biochem. Sci. 40, 491–496 https://doi.org/10.1016/j.tibs.2015.07.004

Heller, G.T., Bonomi, M. and Vendruscolo, M. (2018) Structural ensemble modulation upon small-molecule binding to disordered proteins. J. Mol. Biol. 430, 2288–2292 https://doi.org/10.1016/j.jmb.2018.03.015

Okuda, M. and Nishimura, Y. (2014) Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. J. Am. Chem. Soc. 136, 14143–14152 https://doi.org/10.1021/ja506351f