<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Physical Chemistry, Chemical Physics - PCCP*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Dogan, J., Gianni, S., Jemth, P. (2014)

The binding mechanisms of intrinsically disordered proteins.

Physical Chemistry, Chemical Physics - PCCP, 16(14): 6323-6331

<http://dx.doi.org/10.1039/c3cp54226b>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-224159>

The binding mechanisms of intrinsically disordered proteins

Jakob Dogan^{1*}, Stefano Gianni^{2,3*} and Per Jemth^{1*}

¹Department of Medical Biochemistry and Microbiology, Uppsala University, BMC Box 582, SE-75123 Uppsala, Sweden.

²Dipartimento di Scienze Biochimiche “A. Rossi Fanelli” and Istituto di Biologia e Patologia Molecolari del CNR, Università di Roma “La Sapienza”, P.le A. Moro 5, 00185, Rome, Italy.

³Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Abstract

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) of proteins are very common and instrumental for cellular signaling. Recently, a number of studies have investigated the kinetic binding mechanisms of IDPs and IDRs. These results allow us to draw conclusions about the energy landscape for the coupled binding and folding of disordered proteins. The association rate constants of IDPs cover a wide range (10^5 - 10^9 $\text{M}^{-1}\text{s}^{-1}$) and are largely governed by long-range charge-charge interactions, similarly to interactions between well-folded proteins. Off-rate constants also differ significantly among IDPs (with half-lives up to several minutes) but are usually around 0.1 - 1000 s^{-1} , allowing for rapid dissociation of complexes. Likewise, affinities span from pM to μM suggesting that the low-affinity high-specificity concept for IDPs is not straightforward.

Overall, it appears as binding precedes global folding although secondary structure elements such as helices may form before the protein-protein interaction. Short IDPs bind in apparent two-state reactions whereas larger IDPs often display complex multi-step binding reactions. While the two extreme cases of two-step binding (conformational selection and induced fit) or their combination into a square mechanism is an attractive model in theory, it is too simplistic in practice. Experiment and simulation suggest a more complex energy landscape in which IDPs bind targets through a combination of conformational selection before binding (e.g., secondary structure formation) and induced fit after binding (global folding and formation of short-range intermolecular interactions).

1. Introduction

The field of intrinsically disordered proteins (IDPs) has exploded during the last few years.¹⁻⁴ In fact, the discovery that many proteins, or portions of them, display an extreme conformational freedom came as a big surprise considering the well-established connection between precise structure and function found in for example pre-organized enzyme active sites. Clearly, the very high frequency of disordered stretches in proteins suggests a functional role of such regions. Further, it has also been found that intrinsic disorder is particularly present in proteins implicated in diseases.⁵

Proteins are typically minimally frustrated entities, characterized by a folded and relatively rigid native state at the bottom of a smooth energy landscape. IDPs, on the other hand, sample a multitude of conformations with no single well-defined structure in the unbound state and have a much more weakly funnelled and rugged landscape, which makes it particularly challenging to describe the conformational space sampled by IDPs, and its relation to function. In this respect, IDPs have given a new dimension to the already complex energy landscape of protein-protein interactions. In this perspective article, we will approach this issue by looking at the binding reactions involving purified IDP domains and intrinsically disordered regions (IDRs), to see what they can teach us about the energy landscape of IDP-protein interactions.

What signifies protein-protein interactions involving IDPs? Are there for example mechanistic differences for ordered and disordered proteins, that is, fundamental differences in the energy landscape for binding? While there is still relatively little quantitative data published on disordered proteins we attempt to make some general conclusions using the data at hand and comment on proposed advantages of being disordered.

2. Rate constants of protein-protein interactions

One of the aims of this review is to analyze the reaction mechanisms involving the recognition of IDPs, based on measurements of rate constants for binding. Therefore, we will first recapitulate some of the key equations that are needed to analyze quantitatively the kinetics of binding. In the simplest scheme describing a protein-protein interaction there are only two states, the free state and the bound state (Fig. 1a). The rate constant for association is usually called k_{on} (or sometimes k_1 or k_+) (unit: $\text{M}^{-1}\text{s}^{-1}$) and that for dissociation k_{off} (k_{-1} , k_2 or k_-) (unit: s^{-1}). The equilibrium dissociation constant (unit: M) is the quotient of the two (Eq. 1).

$$K_d = k_{\text{off}}/k_{\text{on}} \quad \text{Eq. 1}$$

If the two proteins are mixed under pseudo-first order conditions, where one protein (P_2) is in large excess over the other, they will bind to each other according to single exponential kinetics (Eq. 2).

$$[P_1P_2] = [P_1]_0 (1 - \exp[-k_{\text{obs}}t])$$

The observed rate constant in Eq. 2, k_{obs} , is the sum of the forward and reverse rate constants at a certain concentration of P_2 (Eq. 3).

$$k_{\text{obs}} = k_{\text{on}} \times [P_2] + k_{\text{off}}.$$

Thus, to obtain the microscopic rate constants k_{on} and k_{off} , k_{obs} is measured at different $[P_2]$ and the data fitted to a straight line; the slope is k_{on} and the intercept is k_{off} . However, k_{off} is often too low (compared to the lowest measured k_{obs}) to be accurately estimated by extrapolation. The k_{off} value can then be determined in a separate displacement experiment, in which a large excess of a competing ligand, with a distinct change in signal upon binding, is mixed with the pre-formed complex. At high concentration of competing ligand, k_{obs} will approach k_{off} . There are several methods available to measure rate constants of which the currently most common for IDPs (and folded proteins) are stopped-flow spectroscopy⁶ and surface plasmon resonance,⁷ but temperature jump,⁸ and fluorescence correlation spectroscopy⁹ have also been used. NMR transverse relaxation dispersion^{10, 11} has also emerged as a powerful method to obtain site resolved information about the thermodynamics, kinetics, and chemical shift differences for proteins undergoing conformational exchange on the μs -ms timescale. This method has been utilized in systems such as the association between the intrinsically disordered phosphorylated kinase-inducible domain (pKID) from the transcription factor CREB and the KIX domain from CREB binding protein.¹² In that study, dispersion curves were measured as a function of the concentration ratio of the two binding components, and rate constants, k_{on} and k_{off} were then extracted from the fitting of the curves. In the present paper we have chosen to discuss data on IDP-protein interactions measured by solution methods (e.g., stopped flow and NMR), since these are our main areas of expertise.

In protein-protein interactions, including those of IDPs, there is often a conformational change involved in the binding reaction and we then end up with a three-state system. The conformational change could occur prior to binding (Fig. 1b,

conformational selection) or after binding (Fig. 1c, induced fit). In these cases the binding follow a biphasic behavior (double exponential kinetics), in which one of the phases is hyperbolic and the other roughly linear with increasing protein concentration. However, it is important to remember that often only one of the kinetic phases is experimentally visible, for example when the other kinetic phase occurs on a much faster timescale. The system thus appears as two-state, unless the hyperbolic dependence is clear. The k_{on} and k_{off} values obtained from data are then apparent or composite rate constants. However, they reflect the overall binding reaction and therefore Eq. 1 still holds and K_d can be calculated from the ratio of $k_{\text{off}}^{\text{app}}$ and $k_{\text{on}}^{\text{app}}$.

3. Rate constants of IDPs.

3.1 Magnitude of rate constants

Rate constants for IDP-protein or IDR-protein interactions from the literature (stopped-flow and NMR) display a wide range: k_{on} , 10^5 - $10^9 \text{ M}^{-1}\text{s}^{-1}$ and k_{off} , 0.1 - 1000 s^{-1} (Table 1). We therefore argue here that the rate constants of IDPs appear very similar to those of ordered proteins despite an apparent wish in the field to hold up their fast association and dissociation rate constants. Indeed, small differences (3-fold) in association and dissociation rate constants for IDPs versus ordered proteins may be visualized using literature data.^{13, 14} But, other factors such as ionic strength and pH play a much larger role than disorder for the rate of the protein-protein interaction (see next section). In addition, the use of different probes and how they are attached can yield different rate constants as seen for c-Myb/KIX^{15, 16} and even more dramatically for the S-peptide/S-protein interaction, for which labeling resulted in a charge reversal of the S-peptide and a 40-fold difference in k_{on} .^{17, 18} Finally, the

selection of experimental system by the investigator is likely skewed by factors such as how difficult it is to express and purify the proteins and how well behaved they are in binding experiments (this applies both to IDPs and ordered proteins). Caution must therefore be exercised when comparing limited data sets of rate constants determined under different conditions (temperature, ionic strength and pH) since it could lead to incorrect conclusions. Thus, we conclude that the experimental kinetic data at hand do not provide conclusive evidence that the magnitude of rate constants for IDPs is different from those involving only ordered proteins. To understand the biophysical properties it is usually more instructive to perturb a given system under a defined set of conditions in a controlled way.

3.2 Salt and pH dependence of rate constants

It is clear that ionic strength strongly affects the association rate constants for proteins,¹⁹ including IDPs. Several of the published studies on IDPs have checked the dependence of rate constants on ionic strength. Such experiments reveal the role of charges in the binding, in particular the effect of non specific long-range charge-charge interactions. For example, k_{on} for ACTR and NCBD goes from roughly $30 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$ at low ionic strength ($I=0.02 \text{ M}$, pH 7.4) to $3 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$ at more physiological conditions ($I=0.2 \text{ M}$, pH 7.4).²⁰ A similar pattern is seen for the binding of GB1-PUMA to Mcl-1,²¹ HPV E7/Rb,²² and WASp/Cdc42.²³

Except for S-peptide/S-protein¹⁷ the result of increasing ionic strength for published IDP-protein interactions is invariably a decrease in k_{on} , suggesting that attractive charge-charge interactions is a common strategy for IDPs to modulate their affinity.

Too high binding affinity is not beneficial for IDPs that are involved as hubs in signaling processes. The physiological ionic strength, which is around 0.2 M, reduces k_{on} significantly and prevent IDPs from binding their targets too strongly. The dissociation rate constant is rarely affected by ionic strength unless Hofmeister effects are at play.²⁴ The important role of electrostatics in the association of IDPs have been previously noted and even suggested to promote subsequent folding.²⁵ Furthermore, in the case of NCBD an interesting observation was made using voltage-driven translocation and single-molecule detection of this domain through a nanopore. At around 0.2 M ionic strength, and at pH=8.0, there is a charge reversal in NCBD, which goes from being positive, as calculated from amino acid composition, to negative.²⁰ This phenomenon, which was corroborated by zeta potential measurements, is probably due to the shell of chloride ions surrounding the highly positive NCBD. At higher concentrations of chloride the protein attracts more negative ions than its own charge due to dispersion forces, resulting in a charge reversal. Such effects are not well studied neither for IDPs nor ordered proteins,²⁶ but may have a large impact on k_{on} , especially when the protein has a high theoretical positive charge.

The dependence of chemical reactions on pH can also be very instructive, since pH modulates the charge of individual amino acid residues. For example, the k_{on} and k_{off} values for the reaction between c-Myb and KIX show distinct pH dependencies, demonstrating that Asp/Glu residues are involved in long-range attractive electrostatic interactions ($\text{p}K_{\text{a}} = 4.2$ for pH dependence of k_{on}) and that a residue (probably His) needs to be protonated for maximum affinity ($\text{p}K_{\text{a}} = 7.6$ for pH dependence of k_{off}). The pH dependence of rate constants is thus a powerful tool to assess the role of

charged residues in binding reactions, in particular in combination with structure-based mutagenesis. This approach has been extensively used in enzymology²⁷ but relatively few studies on IDPs have been performed.

Finally, we note that these effects of charges and ionic strength on kinetics by and large are similar for IDPs and well-structured proteins.¹⁹

4. Mechanism

An experimental picture of the energy landscape of protein-ligand interactions can only be obtained from quantitative measurements of equilibrium and rate constants of the structural states involved. By changing parameters such as ligand concentration, temperature, pH, ionic strength and concentration of osmolytes and denaturants, and in combination with protein engineering, it is then possible to obtain clues about the shape of the landscape.

4.1 Intermediates on the binding pathway

Shorter IDPs/IDRs, which constitute single helices or strands usually display apparent two-state behavior in binding experiments, manifested as a linear dependence between the observed rate constant k_{obs} and protein concentration.^{8, 21, 28, 29} Two-state kinetics mean that the only experimentally visible states are the free unbound species and one bound complex. Note that such two-state behavior does not rule out a multi-step mechanism, but it means that the rate constant(s) of any additional step is larger than the experimentally observed k_{obs} . If a deviation from linearity (hyperbolic) is

observed in a plot of k_{obs} versus concentration of one of the two interacting proteins, then there is at least one more step in the binding reaction. Such three-state (Fig. 1b and c) (or even four-state) mechanism is common among IDPs and in particular some of the longer investigated IDPs, for example, Hirudin/thrombin,³⁰ pKID/KIX,¹² HPV16 E7/Retinoblastoma tumor suppressor²² and ACTR/NCBD.⁶ However, also shorter IDPs like the interactions between S-peptide/S-protein¹⁸ and disordered peptide/PDZ domain³¹ display three-state kinetics at high concentration of the respective peptides. While such hyperbolic kinetic behavior is proof of a conformational change it does not tell whether it takes place before association (conformational selection) or in the complex (induced fit). It is sometimes possible to distinguish these two scenarios by performing experiments in which the concentration of both the interacting molecules is varied.³¹ In the induced fit mechanism the conformational change occurs in the complex and is thus "symmetrical", resulting in a hyperbolic dependence irrespective of which protein is varied. However, if a fast conformational change occurs before binding then k_{obs} will display a hyperbolic behavior only when the species with the conformational change is held constant. Conversely, if the concentration of the species with the conformational change is increased, k_{obs} will increase almost linearly with protein concentration. In practice it may be very hard to conduct this type of decisive experiment, since very high protein concentrations are usually needed and the observed rate constants could require continuous flow, temperature jump or other very rapid kinetic methods. If k_{obs} decreases with increasing protein, it is conclusive evidence that a conformational change precedes binding.²⁷ However, few if any such examples have been reported for IDPs, probably because folding of potential pre-formed elements such as helices is

fast³² compared to binding at μM concentration of protein. (A destabilizing mutation of an ionic interaction in the flexible β -finger of nNOS displays such kinetics.²⁹)

Thus, the order of events is usually deduced or inferred by other methods than from direct kinetic proof. For example, by stabilizing the native state of c-Myb using trifluoroethanol and by observing a clear effect on the dissociation, but not on the association rate constant, Gianni and co-workers demonstrated that c-Myb binds to KIX before folding into a helix.⁸ In fact, most experimental studies on IDPs reach the same conclusion, but on various grounds, namely that binding precedes folding.^{6, 8, 12, 30, 33} For a large and very flexible IDP it is indeed hard to see how an extended and precise structure would be populated in solution. However, smaller secondary structure elements such as alpha helices are likely formed before binding or at least before the rate-limiting step in the binding reaction, and may work as molecular recognition elements.^{34, 35}

4.2 Transition states and folding mechanisms

In order to understand in more detail a binding or folding mechanism it is useful to perturb the experimental system via site-directed mutagenesis in combination with biophysical methods. Such analyses have been applied to enzyme catalysis, protein-ligand interactions and protein folding and together with structural studies they form the basis of our current understanding of structure-function relationships in proteins.²⁷ Data from studies combining site-directed mutagenesis with biophysical experiments serve as benchmarks for computer simulations, which can depict the reaction in

atomistic detail.³⁶ It is therefore obvious to apply the same successful strategy on the binding-induced folding of IDPs.

So far, mutational analysis has been applied to a handful IDPs and only very recently devoted Φ -value analyses^{19, 37} were performed on three protein-protein interactions involving IDPs, namely S-peptide/S-protein,¹⁷ ACTR/NCBD^{38, 39} and c-Myb/KIX.¹⁶ Most interestingly, the results of the three Φ -value analyses underscore that there is no general mechanism for coupled binding and folding of IDPs. While individual Φ -values depict the degree of native bond formation in the transition state for a specific residue it is also useful to combine the data of site-directed mutants into so-called Brønsted (or Leffler) plots as inspired by classical physical-organic chemistry.⁴⁰ Figure 2 shows Brønsted plots for the three IDP systems S-peptide/S-protein, ACTR/NCBD and c-Myb/KIX. This type of Brønsted plot relates the free energy between the free ground states and the rate-limiting barrier for association ($\Delta\Delta G^\ddagger$) with the free energy at equilibrium for the whole binding reaction ($\Delta\Delta G_{\text{Eq}}$). The slope of the plot reflects the overall formation of native contacts in the transition state; if the slope is one it means that all native bonds are fully formed in the transition state, and if the slope is zero there is no native bond formation.

The difference in terms of formation of native bonds in the respective rate-limiting transition state for the three IDP systems is striking. All three systems were investigated with regard to secondary structure (alpha helix) and tertiary intermolecular interactions. The S-peptide folds into one alpha helix upon binding to the S-protein. It is clear from the Brønsted plot (and the Φ -values)^{17, 41} that neither the helix nor intermolecular interactions have formed at the top of the rate-limiting

barrier. ACTR forms three helices upon binding to NCBD. While only a few native intermolecular hydrophobic interactions have been formed in the transition state,³⁸ helix one in ACTR displays around 50% native interactions.³⁹ Finally, c-Myb, which folds into an alpha helix upon binding to KIX displays a considerable degree of native bond formation in the rate limiting transition state, both with regard to secondary and tertiary structure.¹⁶ To use mechanistic models developed for protein folding, these three examples suggest that IDPs may use the whole range of strategies for coupled binding and folding, from preformed structure with subsequent docking (framework/diffusion collision)⁴² to a concerted formation of secondary and tertiary structure (nucleation condensation).⁴³

4.3 Single molecule experiments

Kinetic methods may provide detailed information about overall mechanism as described in the previous sections. Further, NMR provides with a wealth of site-resolved information on the structure and dynamics of IDPs, but loss of information takes place due to the ensemble averaging that is inherent in the technique. Furthermore, NMR often requires relatively high protein concentrations due to sensitivity issues, which may result in self-association that could complicate data analysis and interpretation. Single molecule fluorescence resonance energy transfer (smFRET),^{44, 45} pulling experiments using optical tweezers⁴⁶ and nanopore translocation experiments²⁰ are therefore perfect complements to the ensemble techniques. For instance, smFRET experiments are usually performed at as little as subnanomolar protein concentrations and have the last couple of years proven to be a powerful tool to reveal hidden information on systems with multiple coexisting states

such as IDPs. The direct detection and quantification of intermediates without model dependent analysis (which is needed in NMR), was demonstrated for the binding of α -synuclein to SDS molecules.⁴⁴ Importantly, while classical kinetic experiments could sometimes prove the existence of on-pathway intermediates⁴⁷ single molecule experiments have the possibility to directly monitor transitions between distinct intermediates and ground states, in particular in time-resolved experiments.⁴⁸

One potential disadvantage of smFRET is the need to use fluorophores, which may affect the properties of the system (see also section 3.1) and optical tweezers obviously requires physical attachment of the protein to a surface. Therefore, caution should be exercised and if possible controls should be made to make sure that incorporation of labels or tethering do not affect the properties significantly. Nevertheless, in the coming years we expect single molecule experiments to complement the mechanistic studies from "ensemble techniques" and help defining intermediates in IDP-protein interactions and their role on the binding pathway.

5. Discussion

We will here try to summarize a typical IDP-protein interaction. We will also comment on some recurring statements about IDPs, in the light of available kinetic binding data.

5.1. A typical binding reaction for an IDP.

Based on the notions listed in the Results section we propose the following general scenario for coupled binding and folding of IDPs. The initial association is often strongly dependent on ionic strength, demonstrating that specific or non-specific charge-charge interactions govern the rate of association. After association there is a complex in which the IDP adopts one or several non-native conformations. Evidence for this intermediate (or encounter complex) comes from non-linearities in kinetics^{6, 22, 23, 30} as well as NMR experiments on pKID/KIX.¹² We expect the intermediate to be different from the bound state and probably heterogeneous, for the following reasons: (i) kinetic Φ -values for S-peptide and ACTR/NCBD suggest that the rate-limiting transition state (after the intermediate) is largely non-native like; (ii) the NMR experiments on pKID shows that it forms an ensemble of different intermediate structures upon interaction with KIX. After forming the unstable intermediate the complex can either dissociate or re-arrange to find the most stable bound conformation in an induced fit type of mechanism, during which the binding reaction crosses the rate-limiting barrier. The rate of this search depends on the ruggedness or frustration of the energy landscape. Sometimes other intermediates will accumulate and these may be on- or off-pathway.

For IDPs and IDRs, all experimental evidence suggest a late formation of native tertiary interactions along the binding reaction coordinate, while secondary structure may form early. In other words, inter-molecular bonds and conformations that are present in the solved structures of bimolecular complexes are not present in the rate-limiting transition state for the coupled binding and folding reaction. We speculate that this transition state is disordered, but to a lesser extent than the intermediate and the free state. However, c-Myb displays a very ordered transition state,¹⁶ showing that

generalizations with regard to transition state structure at this stage may be premature and that tertiary interactions may form early on the reaction coordinate.

It is also important to remember that for an induced-fit type of mechanism there is one transition state for binding and one for folding, in the simplest case. It is usually not straightforward to deduce which of the two barriers that is rate limiting, particularly in an apparent two-state reaction, but in the cases discussed here, we find it most plausible that the second barrier is the rate limiting one.

5.2 Induced fit, conformational selection and beyond

In most experimental studies, it is thus concluded that an induced fit mechanism is present in IDP-protein interactions. Computational studies also point in the same direction and then the mechanism is often referred to as dock and coalesce (see Zhou et al.⁴⁹ and references therein). However, while it is likely true that induced fit is the prevailing mechanism, this is actually rarely experimentally demonstrated beyond any doubt. Furthermore, we know from experiment that helices form and unfold on a very fast timescale (sub μ s)^{32, 50} and it is plausible that they often bind to the interaction partner in the folded state. For IDPs with more than one helix, such as ACTR, it is possible that one or more of the helices fold before binding³⁹ but that the overall folding and locking-in of all native intermolecular bonds is according to induced fit.⁶ Thus, the multiple possible conformational selections and subsequent binding-induced folding events even for a "simple" IDP system as the two domains ACTR and NCBD may result in a complex reaction scheme as reflected in the binding kinetics.^{6, 38} We think that this is a likely scenario for all IDPs with multiple secondary structure

elements. Therefore, the popular four state mechanism of coupled folding and binding depicted as a square with induced fit in one path and conformational selection in the other one, and variable flux through the respective pathway is probably too simplistic in most cases. A more complex energy landscape with several interconnected routes and concentration-dependent flux⁵¹ through the respective pathway is a likely scenario. A multi-route energy landscape is also suggested by recent computational studies on the nucleoprotein from measles virus.⁵² It gets even more complex with "fuzzy complexes",⁵³ in which there are no clear disorder-to-order transition upon binding. One beautiful example is the interaction between the IDP Sic1, which contains multiple phosphorylation sites, and is a central player in the cell-cycle control, and Cdc4, which only has one binding pocket for a phosphate group.⁵⁴ Sic1 needs to be phosphorylated at several sites in order to bind Cdc4 with sufficient affinity through the avidity effect. Since all phosphate groups are able to bind this site, but only one at a time, the complex is in a multi-pathway dynamic equilibrium. Only those residues that interact experience a transient ordering whereas the rest of Sic1 still is disordered.

Of course, it is important to remember that the only way to interpret kinetic experiments is to apply the simplest scheme that satisfactorily explains data. In practice, to answer a particular question, for example the role of a charge-charge interaction using salt- or pH dependence, it is useful to choose experimental conditions such that the binding reaction is apparent two-state.

5.3 High specificity and low affinity

It is often stated that the binding mechanism of IDPs, coupled folding and binding, gives the possibility of high specificity in combination with low affinity. We argue here that this concept is not straightforward and has to be considered for each particular IDP-protein interaction.

IDPs often bind several targets, they are promiscuous, a phenomenon that is not well understood not only for IDPs but for proteins in general.^{4, 55-58} One potential way for proteins to achieve promiscuity is to interact mainly through hydrophobic interactions, and it has been suggested that this can be accomplished by providing alternative contact points.⁵⁹ Furthermore, it has been shown that the binding interfaces formed by IDPs have a more pronounced hydrophobic character than those formed by ordered proteins.^{25, 60, 61} It is expected that the change in solvent entropy, due to desolvation of the proteins and ligands involved, can be quite large for IDPs in disorder to order transitions, in which a substantial amount of solvent accessible surface upon binding may be buried. In particular, if the buried surface is mostly hydrophobic, then the change in solvent entropy will be very favourable for binding. The resulting increase in solvent entropy will counterbalance the concomitant decrease in conformational entropy. The relative contribution of these two entropic parameters vary from one case to another, and this may give IDPs a greater freedom to tune their binding affinities, up or down.

With this background, we note that disorder can significantly modulate the binding affinity. The conformational entropy change upon complex formation has previously been shown to be an important contributor to the binding free energy,^{62, 63} with magnitudes similar to the change in solvent entropy. If binding results in a significant

loss in conformational entropy, then this will be very unfavourable for binding. However, the crucial point is that this will not necessarily result in a low affinity, since the desolvation entropy may counter balance the loss in conformational entropy.

Moreover, it is important to recognise the enthalpic contribution to affinity. IDPs are able to form large binding interfaces, which could result in many favourable enthalpic interactions resulting in high binding affinities. Thus, even when the overall entropy change is unfavourable the affinity of complexes formed by IDPs could be high. One example is ACTR/NCBD for which the total entropy change (which is dominated by changes in solvent and conformational entropy) at 31 °C is very unfavourable,⁶⁴ meaning that the entropic penalty of becoming more ordered is even larger than the gain in solvent entropy, which is expected to be quite significant considering the hydrophobic character of this large binding interface. The large loss in conformational entropy is also supported by NMR studies, showing that the unbound components, in particular ACTR, are very dynamic, but that the dynamics of the complex is similar to those formed by stable proteins.⁶⁵ However, the enthalpic contribution to the free energy at 31 °C is very favourable, resulting in a nanomolar K_d .

There is of course nothing such as a defined "high affinity" and all functional protein-protein interactions have evolved to fulfil a particular purpose, in competition with other potential interactions in the crowded cell. The energetics of the disorder to order transition upon binding for an IDP is a complex interplay between enthalpy, conformational entropy and desolvation entropy. Available experimental data on IDPs (Table 1) do not support a general view of low affinity-high specificity for IDPs.

5.4 Fly-casting

Fast association kinetics has been frequently cited as one of the potential advantages of disorder, as a larger capture radius of IDPs has been shown on theoretical grounds to increase the rate constant of binding compared to ordered proteins, i.e. the so-called fly-casting hypothesis.⁶⁶ However, the theoretical maximum increase in k_{on} is only a modest 1.6-fold. Also, Huang & Liu suggested that such effects are counterbalanced by the slower translational diffusion of IDPs,¹³ but that an overall increase in association kinetics (again, only a modest 3-fold difference) for IDPs may occur, due to a lower energy barrier in subsequent steps before reaching the native complex. Thus, it is not apparent how the association kinetic properties of IDPs would be much more different from those of ordered proteins. Indeed, experimentally, there is not much support, if any, for the fly-casting hypothesis; disordered proteins adopt a wide range of k_{on} values (Table 1) similarly to rigid proteins. Also, it is difficult to design an experiment that really tests the hypothesis. For example, it could be argued that a destabilizing mutation increases the flexibility of the IDP.²⁹ Such mutations rarely if ever increases k_{on} as would be expected from fly-casting. However, while the mutation destabilizes the complex it does not necessarily destabilize the IDP relative to the transition state (make it more flexible). Destabilization or stabilization of helices by mutation may lead to a more flexible helix in the sense that it populates the unfolded state a larger fraction of the time. But again, destabilizing mutations always decrease k_{on} ^{16, 17} and stabilizing mutations increase k_{on} .³⁹ We thus conclude that kinetic binding experiments in combination with protein engineering performed to date lend no support to the fly-casting hypothesis but neither can they falsify it.

5.5 Coupled binding and downhill folding.

Finally, computational studies on IDP-protein interactions,⁵² including that between ACTR and NCBD,⁶⁷ suggest that the folding reaction, following binding, takes place without a significant barrier, so-called downhill folding. Extensive folding studies on the BBL domain have highlighted the extreme difficulty of experimentally proving downhill folding.^{68, 69} Naturally, the same experimental challenge applies for coupled binding and folding of IDPs. A typical way to detect downhill kinetics is to measure the rate constant of reactions at their speed limit.⁷⁰ In fact, in the absence of barriers, folding would be limited by chain diffusion, with observed rate constants typically higher than 10^6 - 10^7 s⁻¹. To the best of our knowledge, there is no experimental evidence for a binding induced folding reaction of an IDP occurring at those time scales. First-order rate constants from hyperbolic binding kinetics for IDPs have been measured for pKID/KIX (2000-3000 s⁻¹),¹² WASp/Cdc42 (ca. 80 s⁻¹),²³ S-peptide/S-protein (ca. 390 s⁻¹),¹⁸ Hirudin/Thrombin (ca. 50 and 300 s⁻¹)³⁰ and ACTR/NCBD (ca. 40 s⁻¹),⁶ suggesting that the downhill hypothesis, while fascinating, does probably not hold true for most IDPs.

6. Concluding remarks

While sequence and charge are defining features of IDPs, their binding reactions appear very similar to those of ordered proteins. We also want to emphasize that the range of published affinities does not suggest that IDP-protein interactions are particularly weak. It may prove more fruitful to appreciate the similarities between disordered and ordered proteins than emphasizing their differences. With respect to the binding reaction, the latter dominate.

Acknowledgements

This work was funded by the Swedish Research Council and the Human Frontiers Science Programme (to P.J.), the Magnus Bergvall Foundation (to J.D.) and from the Italian Ministry of University and Research (PNR-CNR Aging Program 2012– 2014) (to S.G).

References

1. A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner and Z. Obradovic, *J. Mol. Graph. Model.*, 2001, **19**, 26-59.
2. H. J. Dyson, *Q. Rev. Biophys.*, 2011, **44**, 1-52.
3. J. D. Forman-Kay and T. Mittag, *Structure*, 2013, **21**, 1492-1499.
4. V. N. Uversky, *Curr. Pharm. Des.*, 2013, **19**, 4191-4213.
5. V. N. Uversky, C. J. Oldfield and A. K. Dunker, *Annu. Rev. Biophys.*, 2008, **37**, 215-46.
6. J. Dogan, T. Schmidt, X. Mu, Å. Engström and P. Jemth, *J. Biol. Chem.*, 2012, **287**, 34316-34324.
7. E. R. Lacy, I. Filippov, W. S. Lewis, S. Otieno, L. Xiao, S. Weiss, L. Hengst and R. W. Kriwacki, *Nat. Struct. Mol. Biol.*, 2004, **11**, 358-364.
8. S. Gianni, A. Morrone, R. Giri and M. Brunori, *Biochem. Biophys. Res. Commun.*, 2012, **428**, 205-209.
9. J. K. Lum, H. Neuweiler and A. R. Fersht, *J. Am. Chem. Soc.*, 2012, **134**, 1617-1622.
10. J. P. Loria, M. Rance and A. G. Palmer, 3rd, *J. Biomol. NMR*, 1999, **15**, 151-155.
11. M. Tollinger, N. R. Skrynnikov, F. A. Mulder, J. D. Forman-Kay and L. E. Kay, *J. Am. Chem. Soc.*, 2001, **123**, 11341-11352.
12. K. Sugase, H. J. Dyson and P. E. Wright, *Nature*, 2007, **447**, 1021-1026.
13. Y. Huang and Z. Liu, *J. Mol. Biol.*, 2009, **393**, 1143-1159.
14. S. L. Shammash, J. M. Rogers, S. A. Hill and J. Clarke, *Biophys. J.*, 2012, **103**, 2203-2214.
15. S. L. Shammash, A. J. Travis and J. Clarke, *J. Phys. Chem. B*, 2013, **In press**.
16. R. Giri, A. Morrone, A. Toto, M. Brunori and S. Gianni, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 14942-14947.
17. A. Bachmann, D. Wildemann, F. Praetorius, G. Fischer and T. Kiefhaber, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3952-3957.
18. J. M. Goldberg and R. L. Baldwin, *Biochemistry*, 1998, **37**, 2546-2555.
19. G. Schreiber, G. Haran and H. X. Zhou, *Chem. Rev.*, 2009, **109**, 839-860.

20. D. Japrun, J. Dogan, K. J. Freedman, A. Nadzeyka, S. Bauerdick, T. Albrecht, M. J. Kim, P. Jemth and J. B. Edel, *Anal. Chem.*, 2013, **85**, 2449-2456.
21. J. M. Rogers, A. Steward and J. Clarke, *J. Am. Chem. Soc.*, 2013, **135**, 1415-1422.
22. L. B. Chemes, I. E. Sanchez and G. de Prat-Gay, *J. Mol. Biol.*, 2011, **412**, 267-284.
23. L. Hemsath, R. Dvorsky, D. Fiegen, M. F. Carlier and M. R. Ahmadian, *Mol. Cell*, 2005, **20**, 313-324.
24. R. L. Baldwin, *Biophys. J.*, 1996, **71**, 2056-2063.
25. D. Ganguly, S. Otieno, B. Waddell, L. Iconaru, R. W. Kriwacki and J. Chen, *J. Mol. Biol.*, 2012, **422**, 674-684.
26. A. Salis, M. Boström, L. Medda, F. Cugia, B. Barse, D. F. Parsons, B. W. Ninham and M. Monduzzi, *Langmuir*, 2011, **27**, 11597-11604.
27. A. Fersht. (1999). *Structure and Mechanism in Protein Science*, W.H. Freeman and Co., New York.
28. S. R. Haq, C. N. Chi, A. Bach, J. Dogan, Å. Engström, G. Hultqvist, O. A. Karlsson, P. Lundstrom, L. C. Montemiglio, K. Strømgaard, S. Gianni and P. Jemth, *J. Am. Chem. Soc.*, 2012, **134**, 599-605.
29. O. A. Karlsson, C. N. Chi, Å. Engström and P. Jemth, *J. Mol. Biol.*, 2012, **417**, 253-261.
30. M. P. Jackman, M. A. Parry, J. Hofsteenge and S. R. Stone, *J. Biol. Chem.*, 1992, **267**, 15375-15383.
31. S. Gianni, T. Walma, A. Arcovito, N. Calosci, A. Bellelli, A. Engstrom, C. Travaglini-Allocatelli, M. Brunori, P. Jemth and G. W. Vuister, *Structure*, 2006, **14**, 1801-1809.
32. S. Neumaier, A. Reiner, M. Buttner, B. Fierz and T. Kiefhaber, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 12905-12910.
33. D. Ganguly, W. Zhang and J. Chen, *Mol. Biosyst.*, 2012, **8**, 198-209.
34. M. Fuxreiter, I. Simon, P. Friedrich and P. Tompa, *J. Mol. Biol.*, 2004, **338**, 1015-1026.
35. C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2005, **44**, 12454-12470.
36. M. Vendruscolo, E. Paci, C. M. Dobson and M. Karplus, *Nature*, 2001, **409**, 641-645.
37. A. R. Fersht and S. Sato, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 7976-81.
38. J. Dogan, X. Mu, Å. Engström and P. Jemth, *Sci. Rep.*, 2013, **3**, 2076.
39. V. Iešmantavičius, J. Dogan, P. Jemth, K. Teilum and M. Kjaergaard, *Ang. Chem. Int. Ed.*, 2013, **Under revision**.
40. J. E. Leffler, *Science*, 1953, **117**, 340-341.
41. J. M. Goldberg and R. L. Baldwin, *Biochemistry*, 1998, **37**, 2556-2563.
42. M. Karplus and D. L. Weaver, *Protein Sci.*, 1994, **3**, 650-668.
43. L. S. Itzhaki, D. E. Otzen and A. R. Fersht, *J. Mol. Biol.*, 1995, **254**, 260-288.
44. Y. Gambin, V. VanDelinder, A. C. Ferreón, E. A. Lemke, A. Groisman and A. A. Deniz, *Nat. Methods*, 2011, **8**, 239-241.
45. H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 16155-16160.
46. C. Cecconi, E. A. Shank, S. Marqusee and C. Bustamante, *Methods Mol. Biol.*, 2011, **749**, 255-271.

47. P. Jemth, S. Gianni, R. Day, B. Li, C. M. Johnson, V. Daggett and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6450-6455.
48. B. Wunderlich, D. Nettels, S. Benke, J. Clark, S. Weidner, H. Hofmann, S. H. Pfeil and B. Schuler, *Nat. Protoc.*, 2013, **8**, 1459-1474.
49. H. X. Zhou, X. Pang and C. Lu, *Phys. Chem. Chem. Phys.*, 2012, **14**, 10466-10476.
50. B. Fierz, A. Reiner and T. Kiefhaber, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 1057-1062.
51. G. G. Hammes, Y. C. Chang and T. G. Oas, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13737-13741.
52. Y. Wang, X. Chu, S. Longhi, P. Roche, W. Han, E. Wang and J. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **In press**.
53. M. Fuxreiter and P. Tompa, *Adv. Exp. Med. Biol.*, 2012, **725**, 1-14.
54. T. Mittag, S. Orlicky, W. Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers and J. D. Forman-Kay, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17772-17777.
55. A. Cumberworth, G. Lamour, M. M. Babu and J. Gsponer, *Biochem. J.*, 2013, **454**, 361-369.
56. C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky and A. K. Dunker, *BMC genomics*, 2008, **9 Suppl 1**, S1.
57. G. Schreiber and A. E. Keating, *Curr. Opin. Struct. Biol.*, 2011, **21**, 50-61.
58. V. N. Uversky, *Biochim. Biophys. Acta*, 2013, **1834**, 932-951.
59. C. E. Chang, W. A. McLaughlin, R. Baron, W. Wang and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7456-7461.
60. B. Meszaros, P. Tompa, I. Simon and Z. Dosztanyi, *J. Mol. Biol.*, 2007, **372**, 549-561.
61. K. Gunasekaran, C. J. Tsai and R. Nussinov, *J. Mol. Biol.*, 2004, **341**, 1327-1341.
62. M. S. Marlow, J. Dogan, K. K. Frederick, K. G. Valentine and A. J. Wand, *Nat. Chem. Biol.*, 2010, **6**, 352-358.
63. A. J. Wand, *Curr. Opin. Struct. Biol.*, 2013, **23**, 75-81.
64. S. J. Demarest, M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H. J. Dyson, R. M. Evans and P. E. Wright, *Nature*, 2002, **415**, 549-553.
65. M. O. Ebert, S. H. Bae, H. J. Dyson and P. E. Wright, *Biochemistry*, 2008, **47**, 1299-1308.
66. B. A. Shoemaker, J. J. Portman and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 8868-8873.
67. A. N. Naganathan and M. Orozco, *J. Am. Chem. Soc.*, 2011, **133**, 12154-12161.
68. N. Ferguson, T. D. Sharpe, C. M. Johnson, P. J. Schartau and A. R. Fersht, *Nature*, 2007, **445**, E14-15.
69. M. Sadqi, D. Fushman and V. Munoz, *Nature*, 2006, **442**, 317-321.
70. W. Y. Yang and M. Gruebele, *Nature*, 2003, **423**, 193-197.
71. J. Thomson, G. S. Ratnaparkhi, R. Varadarajan, J. M. Sturtevant and F. M. Richards, *Biochemistry*, 1994, **33**, 8587-8593.
72. T. Zor, R. N. De Guzman, H. J. Dyson and P. E. Wright, *J. Mol. Biol.*, 2004, **337**, 521-534.
73. M. Arai, J. C. Ferreón and P. E. Wright, *J. Am. Chem. Soc.*, 2012, **134**, 3792-3803.

74. P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch and L. E. Kay, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18473-18477.
75. G. Papadakos, N. G. Housden, K. J. Lilly, R. Kaminska and C. Kleanthous, *J. Mol. Biol.*, 2012, **418**, 269-280.
76. K. Sugase, J. C. Lansing, H. J. Dyson and P. E. Wright, *J. Am. Chem. Soc.*, 2007, **129**, 13406-13407.
77. M. Fassolari, L. B. Chemes, M. Gallo, C. Smal, I. E. Sanchez and G. de Prat-Gay, *J. Biol. Chem.*, 2013, **288**, 13110-13123.
78. G. Hultqvist, S. R. Haq, A. S. Punekar, C. N. Chi, Å. Engström, A. Bach, K. Strømgaard, M. Selmer, S. Gianni and P. Jemth, *Structure*, 2013, **21**, 1193-1202.
79. O. Schon, A. Friedler, M. Bycroft, S. M. Freund and A. R. Fersht, *J. Mol. Biol.*, 2002, **323**, 491-501.

Figure legends

Fig. 1. Reaction schemes.

a) One-step (two-state) binding reaction, which gives rise to single exponential kinetics. b) Two-step (three-state) binding with an initial conformational change, which gives rise to double exponential kinetics. c) Two-step (three-state) binding with a conformational change after binding (induced fit), which gives rise to double exponential kinetics. In practice, the schemes in b and c often results in single exponential kinetics because one kinetic phase is too fast. In such cases the observed phase may display a hyperbolic dependence on ligand concentration, which is a tell-tale sign of a three-state mechanism.

Fig. 2. Brønsted plots of the binding reaction between IDPs and target proteins.

a) S-peptide (yellow)/S-protein (grey);⁷¹ b) ACTR (helix 1 in purple, the rest in green)/NCBD (light blue);⁶⁴ c) c-Myb (red)/KIX (blue).⁷² The distinct Brønsted plots in panels a-c likely reflect the large variation in binding mechanism among IDPs.

Table 1. Compilation of binding rate constants for IDPs measured by in solution methods

Disordered protein (residue numbers)	Ligand/buffer	k_{on} ($\mu\text{M}^{-1}\text{s}^{-1}$)	k_{off} (s^{-1})
ACTR ⁶	NCBD 20 mM sodium phosphate, pH 7.4, 150 mM NaCl, 10°C	59	4.5
AD2 1 ⁷³	TAZ2 20 mM Tris/HCl, pH 6.8, 50 mM NaCl, 1 mM DTT, 10% D ₂ O, 35°C	17000	540
AD2 2 ⁷³	TAZ2 20 mM Tris/HCl, pH 6.8, 50 mM NaCl, 1 mM DTT, 10% D ₂ O, 35°C	710	7200
Ark1p peptide, 17 residues ⁷⁴	SH3 50 mM sodium phosphate, pH=7.0, 100 mM NaCl, 1 mM EDTA, 1 mM NaN ₃ , 25°C	630	350
c-Myb ¹⁶	KIX 50 mM sodium phosphate, pH 7.2, 150 mM KCl, 1 mM DTT, 10°C	3.1	24
c-Myb ¹⁵	KIX 20 mM MOPS, pH 7.4, 100 mM NaCl, 0.05% Tween-20, 10°C	11	55
ColE9 TBE ⁷⁵	TolB 50 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM CaCl ₂ , 37°C	0.22	0.2
HIF ⁷⁶	TAZ1 20 mM MES, pH 6.1, 2 mM DTT, 2 mM NaN ₃ , 10% D ₂ O, 25°C	1290	185
Hirudin (52-65) ³⁰	Thrombin 50 mM Tris-HCl, pH 7.8, NaCl to obtain I=0.125M, 0.1% PEG, 37°C	9.3	< 1
HPV16 E7 (36-48) ⁷⁷	M1Fab 10 mM tris-HCl, pH 7.4, 5 mM NaCl, 20°C	64	7.5
HPV16 E7 (16-31) ²²	Retinoblastoma AB domain 20 mM sodium phosphate, pH 7.0, 200 mM NaCl, 2 mM DTT, 0.1% Tween, 20°C	34	0.072
HPV18 E6 (143- 151) ⁷⁸	SAP97 PDZ2 50 mM potassium phosphate, pH 7.5, 10°C	8.7	2.1
nNOS β -finger ²⁹	PSD-95 PDZ2 50 mM potassium phosphate, pH 7.5,	0.40	1.6

	25°C		
p53 (17-26) ⁷⁹	MDM2 10 mM sodium phosphate, pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 15°C	22	1.4
pKID ¹²	KIX 20 mM Tris-acetate, pH 7.0, 50 mM NaCl, 2 mM NaN ₃ , 10% D ₂ O, 30°C	6.3	6
PUMA BH3 ²¹	Mcl-1 10 mM MOPS, pH 7.0, 100 mM NaCl	16	0.0016
S-peptide ¹⁸	S-protein 10 mM MOPS, pH 6.7, 10°C	17	1×10 ⁻⁴
S-peptide ¹⁷	S-protein 50 mM NaOAc, pH 6.0, 25°C	0.44	1.6
WASp ²³	Cdc42 30 mM tris, pH 7.5, 50 mM NaCl, 5 mM MgCl ₂ , 3 mM DTT, 25°C.	22	1.7

a

b

c

a

b

c

It seems that IDPs behave much like any globular protein in many respects with regard to binding!

The distinct Brønsted plots for these 3 cases likely reflect the large variation in binding/folding mechanism - a range of mechanisms may be employed for IDPs

Zero slope for both 2° and 3° structure
→ very little structure in TS