Relationship between interfaces and binding strength of disordered proteins

Project closing report

The aim of the work

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions of proteins (IDRs) often function by molecular recognition, in which they undergo folding induced by the partner. Prevailing idea in the IDP field, that binding of IDPs/IDRs to globular ones "uncouples" specificity from binding strength due to the entropic penalty of induced folding, but the evidence for this idea is rather limited in the literature.

Main purpose of this work was to systematically investigate interactions of intrinsically disordered proteins in particular regarding to binding strength and specificity and test the generalizability of the prevailing idea mentioned above.

Our work aimed to achieve the following goals:
1. Collection of ID complexes with known PDB structure and free energy of interaction
2. Characterization of the collected complexes: binding strength, size of interface/hydrophobic interface, segmentation, type and number of interactions
3. Introduction of two novel measures of "specificity"
    i. evolutionary conservation of the interface
    ii. information content of the interface
4. A new database of disordered protein interactions (in collaborations)

Detailed results

1. Collection of data

Trough literature search we manually collected 75 complexes of intrinsically disordered proteins (IDPs) in which one partner is disordered while the other is usually globular. Disordered nature of the proteins is always proved by experimental analysis, pdb structure and experimentally measured dissociation constant (K_d) are also available for all the collected complexes. Redundancy was filtered out. In cases of similar complexes (with same partners) we choose the one where the construct used for the pdb structure was most similar to the one used for the K_d measurement. The complexes were categorized into three groups: absolutely appropriate complexes (48 complexes), complexes with short participants: less than 15 visible amino acids in the pdb structure (19 complexes) and complexes with proteins having highly different construct for the pdb structure and the K_d measurement (8 complexes). We used a reference database of globular complexes made by the group of A. M. J. J. Bonvin. This collection contains 144 complexes but we only used 99 of them, those containing only two protein components.

2. Physicochemical characterization

The very first steps were to compare the binding strengths, i.e., the experimentally defined dissociation constants of globular and intrinsically disordered complexes and the calculated free energies of binding (dG =-R*T*ln K_d) (Fig. 1.). Disordered and globular complexes have similar dG (K_d) distribution, but IDPs are shifted towards weaker interactions. This shows that in general globular proteins do have stronger interactions (the average of globular dG values is extremely significantly higher than the average dG of IDP complexes), but in some cases IDPs can create strong complexes too.

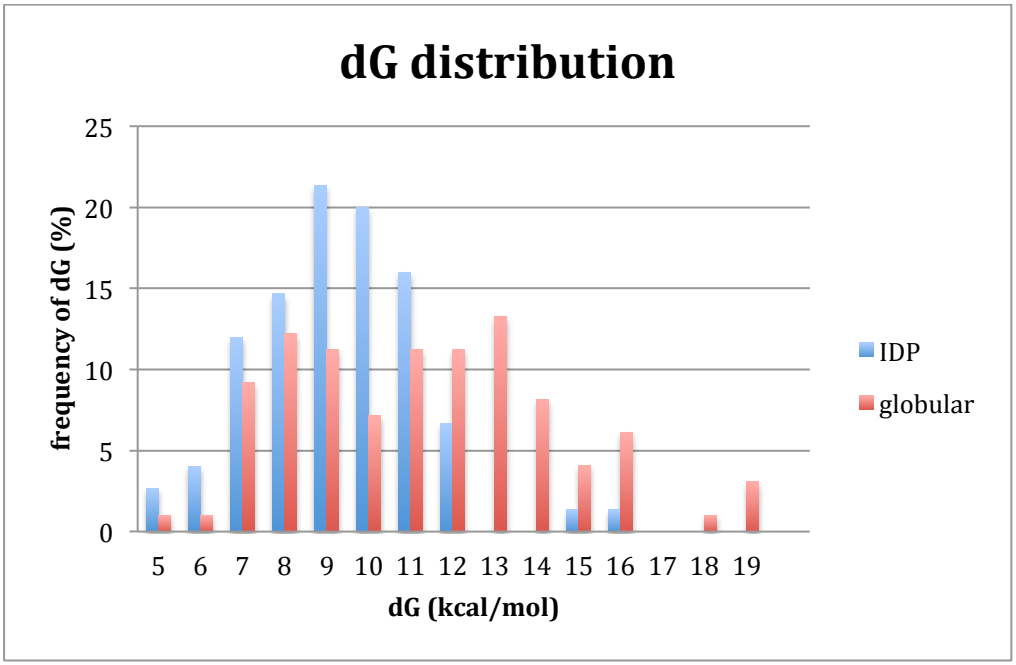

Figure 1. Free energy distribution of disordered and globular complexes

Next step was to define complex interfaces to see whether weaker interactions mean smaller interface sizes. (Interface size was defined as the change in accessible surface area (dASA) of the protein chains during complex formation.) Interface size distribution (Fig. 2.) shows that IDPs have similar but wider interface size range: complexes of disordered proteins have smaller and larger interfaces, too. Interestingly, larger interfaces are represented in almost only in disordered complexes. If we count only the hydrophobic part of interface, the situation is quite similar only IDP interfaces have large hydrophobic contribution above 1500 $Å^2$.

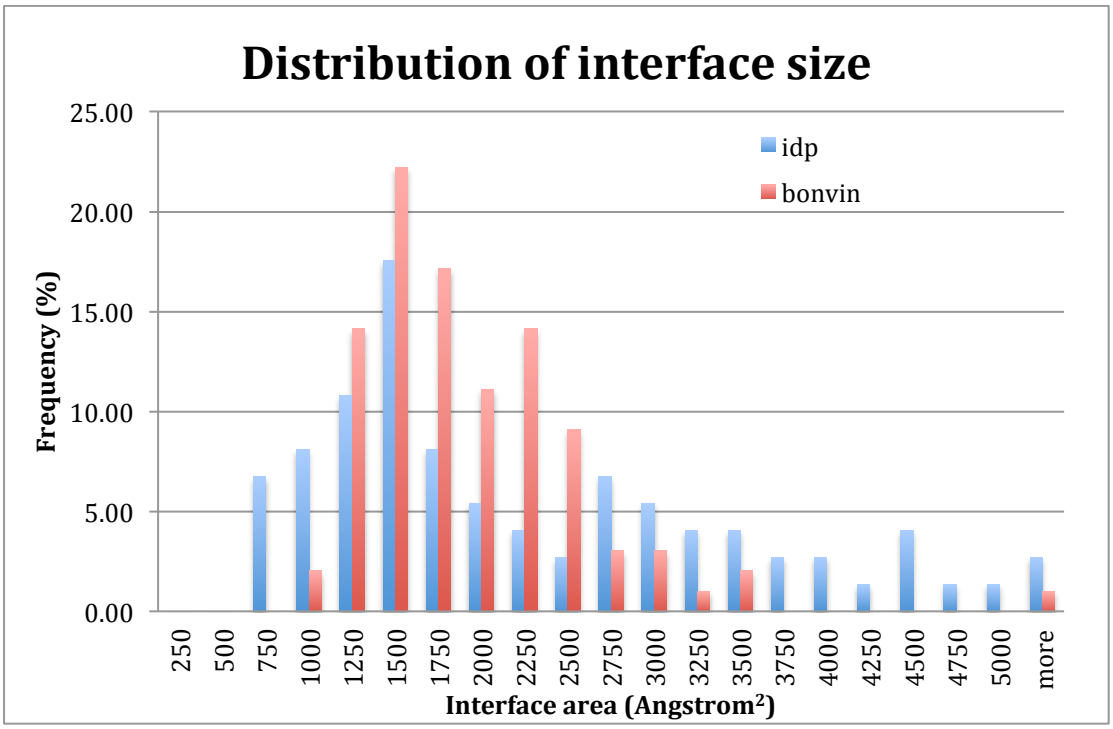

Figure 2. Interface size distribution of disordered and globular complexes

We investigated the connection between the free energy of complex formation and the interface size we wanted to know whether larger interfaces mean stronger interactions (Fig. 3.). For globular proteins, there is a rather linear relationship. For stronger binding, they apply larger and larger interfaces, although the dependence is rather weak. For example, upon going from 6 to 12 to 18 kcal/mol, the interface does not become 3 times bigger (only from 1500 to 2200). This probably means the operation of a "chelate" effect, i.e. that segments within the site cooperate because there is no additional decrease in entropy. For ID interfaces, the situation is very different. First, for free energies above 8 kcal/mol, the ID interface has to be larger than the globular interface. Apparently, unfavorable entropy makes the interaction weaker than expected from the size of the interface. This negative effect is so strong, that to go from 6 to 12 kcal/mol in free energy, the interface has to be 4 times larger (1000 to 4000 A^2). Actually, there seems to be a limit of about 12 kcal/mol, with an ID interface no more free energy can be realized. Further, for the ID interfaces, the linearity of free energy with interface size does not seem to apply, there is an upward curvature and "saturation" effect. An interesting feature of the interfaces is a crossover at about 7 kcal/mole (1500 A^2). Although there are not much data for small(er) globular interfaces, extrapolation suggests that here the same free energy can be realized by a smaller ID interface than globular interfaces. The reason is not apparent, could be that here backbone entropy is not dominant any more (side chain reorganization entropy is the same for both IDPs and globular proteins), and favorable enthalpy coming from better fit (packing density).

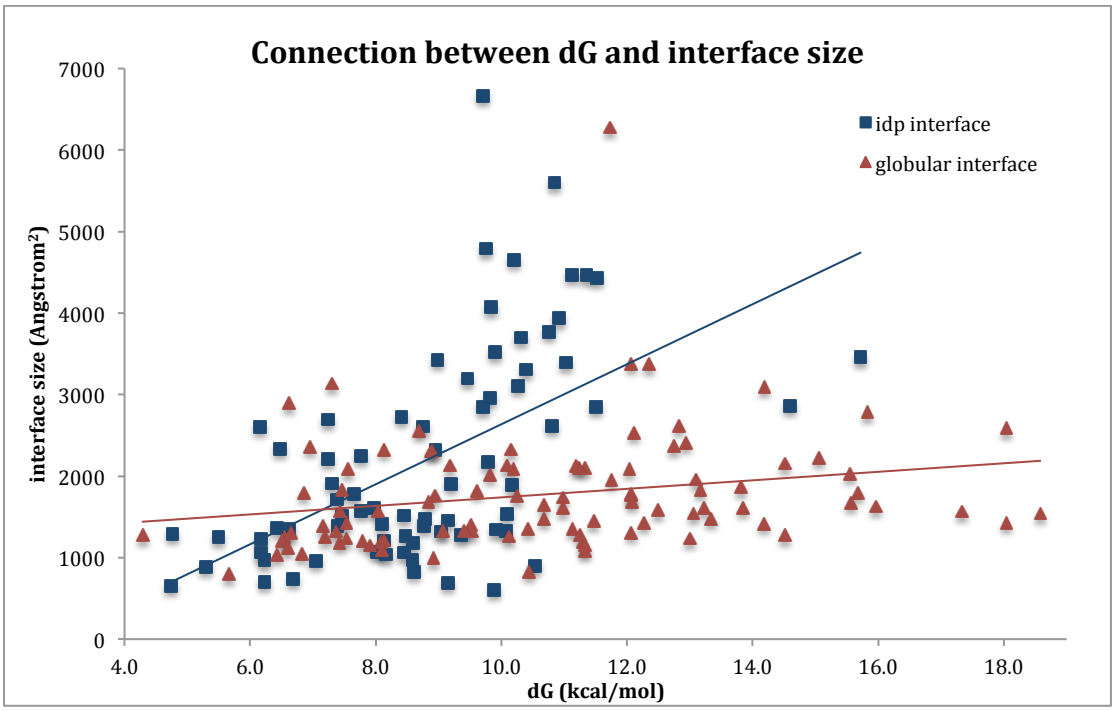

Figure 3. Relationship between free energy of binding and interface size

A conclusion relevant with the issue of specificity is that ID interfaces are significantly weaker than globular interfaces of similar size. The clearest case in the figure is the interface of 2000 A^2, which results in 8 kcal/mole for IDPs but 15 kcal/mole for globular proteins (Kd = 10^{-6} vs. 10^{-9} M). Now the big question is, can we claim they are of similar specificity (or even more specific for IDPs)? I think this figure and the related conclusions are the most important findings in this part of the work.

We also investigated how segmented the interfaces of IDPs and globular proteins are which means the number of segments consist of an interface. In accordance with previous studies we showed that IDPs have less and longer segments. There is no correlation between segmentation and binding strength.

We analysed the different types of interactions made between the two partners. Among others the most interesting finding was that it seems as if ID complexes couldn't enhance their binding strength beyond a certain limit even by increasing the number of interactions (Fig. 4.).

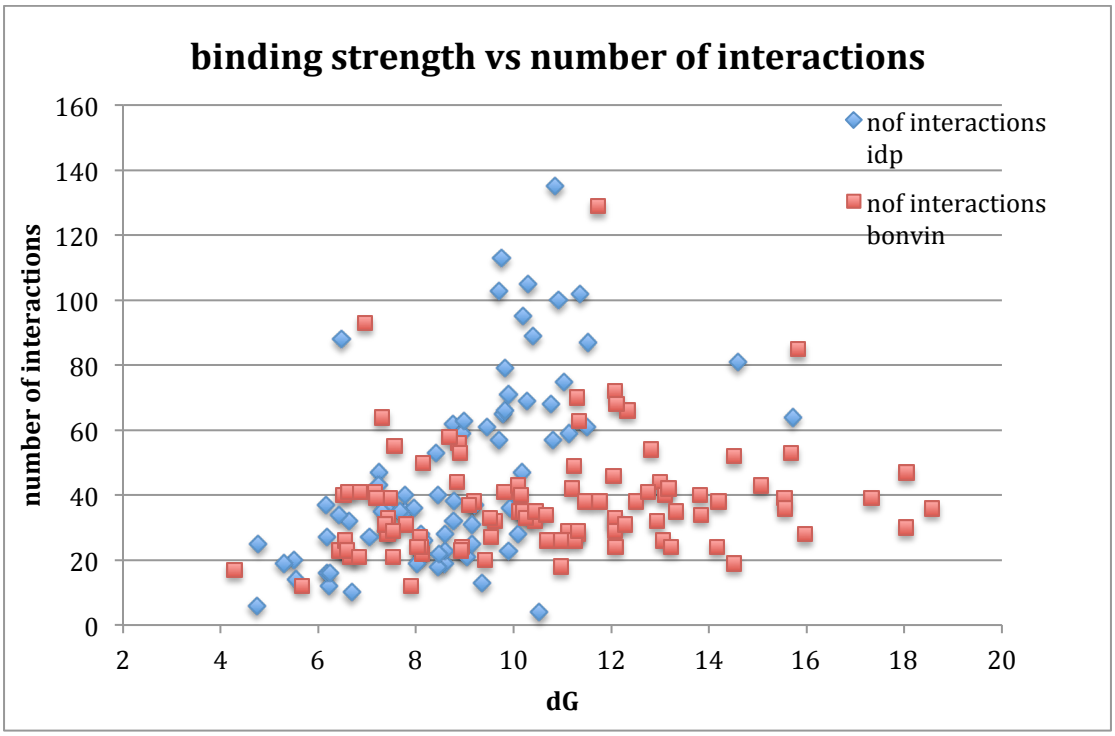

Figure 4. Correlation between binding strength and number of intermolecular interactions within complexes

3. Measure of specificity

It is very complicated to define specificity. Specificity has several contextual elements, which determine the ability to discriminate the cognate partner from all other competing ones. Just looking at thermodynamics of a given interface does not directly tell if a given interaction observed *in vitro* is specific, therefore we decided to invoke two indirect definitions that have contextual flavour. i) The conservation of an interface, which is a direct indication of evolutionary selection pressure, a clear attribute of functionality, i.e. specificity. ii) A new definition, which rests on the assumption that interfaces that are farther away from random, and are less likely to be encountered on other partners are more specific.

i) Conservation

In favour of defining conservation of interface and surface amino acids we first made a BLASTP homology search for each protein followed by multiple alignment using MAFFT (Multiple Alignment using Fast Fourier Transform) program. Conservation scores for each position in the alignment were calculated with the algorithm developed by Capra et al. For ID complexes we considered 25-25 “interface-flanking” amino acids as part of the surface, too, in case they were disordered. For ID complexes disordered and globular partners were distinguished. As expected, interface amino acids are always more conserved than those belong to the surface, but the differences are significant only for disordered part of ID complexes. Not surprisingly the disordered proteins are less conserved than any kind of globular proteins from the same part of the complexes. What is most interesting from these findings is that globular partners in ID complexes are significantly more conserved than proteins of fully

structured complexes. It seems as if for ID complexes the interface of globular partners are responsible for conservation (Fig. 5.).

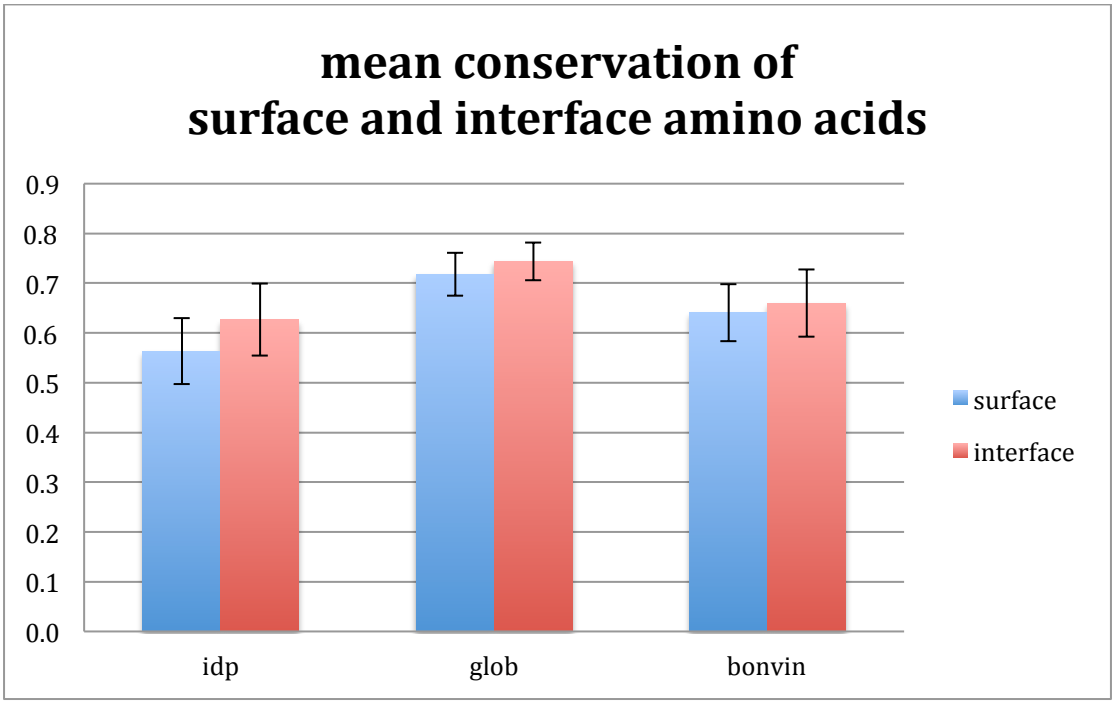

Figure 5. Mean conservation values of surface and interface amino acids

ii) Information content

We developed a method for characterizing interfaces by patterning of physical properties of amino acid pairs in surface patches. We define a statistical potential, termed iPat, to measure the physical patterning of protein surfaces, without taking evolutionary conservation or the interaction partner into consideration. In this regard, iPat provides a measure orthogonal to those used by more traditional algorithms, yet it is effective in scoring protein-protein interactions and discriminating native-like complexes from docking decoys. In addition, iPat is rather insensitive to conformational changes that accompany partner binding, making it the method of choice for characterizing interactions involving intrinsically disordered proteins (IDPs). As iPat can also be combined with other features in docking or interface prediction, it represents a novel tool in analyzing the functional specificity and evolutionary history of protein-protein interactions. (A manuscript describing the method was submitted to PLoS Computational Biology).

Applying this iPat potential to my complex interfaces shows that the interface of IDPs are more interface-like/more specific (has smaller negative value) than that of globular proteins. (Figure 6.)

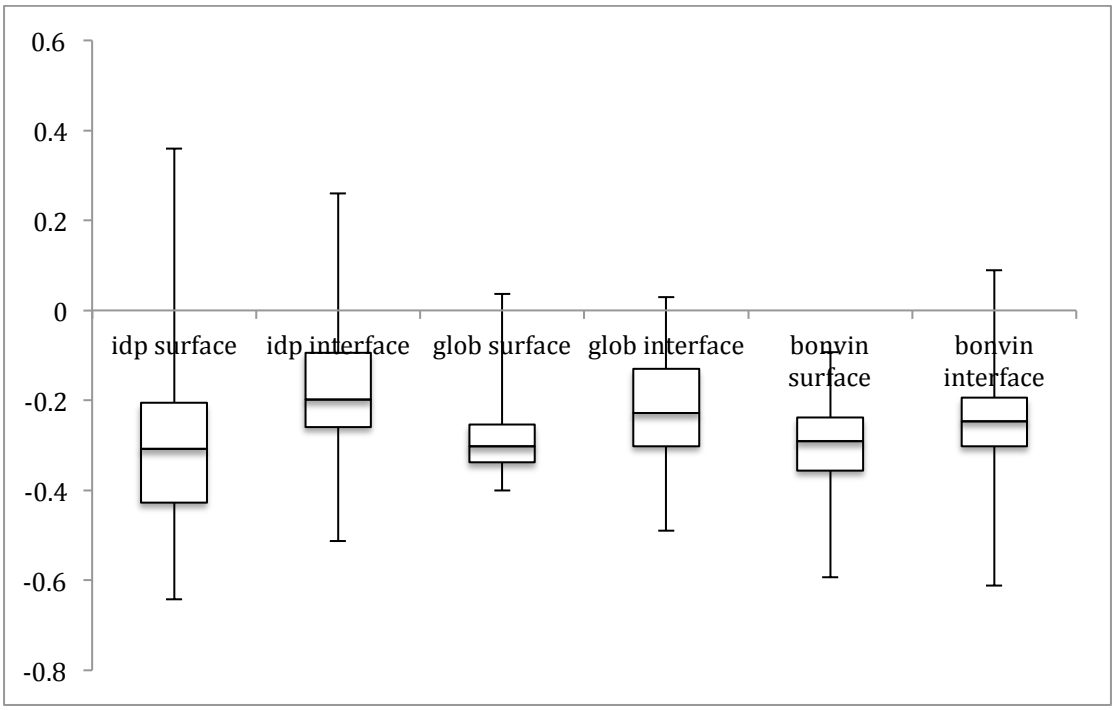

Figure 6. Box plot of iPat values of surface and interface amino acids

Analyzing conservation and information content (iPat) data as well as writing the manuscript are still in progress. Planned title:
Structural disorder uncouples specificity from binding strength in molecular recognition
Eva Schad, Tamas Lázár, Mainak Guharoy, Shoshana Wodak, Joel Janin, Lajos Kalmar and Peter Tompa

4. Database of Disordered Binding Sites

As a further step of the project we created (in collaboration) a new, more extensive database called DIBS (Disordered Binding Sites), which is the first curated dataset that systematically collects interactions formed between IDPs and ordered proteins with their structural complexes. DIBS not only describes by far the highest number of cases, it also provides the dissociation constants of their interactions (773 entries, 488 with Kd values), as well as the description of potential post-translational modifications modulating the binding strength and linear motifs involved in the binding. Together with the wide range of structural and functional annotations, DIBS will provide the cornerstone for structural and functional studies of IDP complexes. The publication about DIBS is almost accepted in Bioinformatics, the manuscript is under second revision. After publication DIBS will be freely accessible at http://dibs.enzim.ttk.mta.hu.
I would like to maintain this collaboration in the future and analyse further this huge amount of data from many other aspects.
Our dataset was also integrated into the new version of MobiDB (manuscript has been submitted to NAR database issue), which is also a sign for its importance.

Summary

Summarizing our findings we can see that by correlating our data and a reference set of folded complexes, we provide evidence that i) binding of an IDP/IDR is significantly weaker than globular protein of similar interface, ii) the specificity of the interactions is commensurable for the two types of interactions, therefore iii) structural disorder uncouples specificity from binding strength.