Many-to-one binding by intrinsically disordered protein regions

Wei-Lun Alterovitz[1*], Eshel Faraggi[1,2,3*], Christopher J. Oldfield[1], Jingwei Meng[1], Bin Xue[1], Fei Huang[1], Pedro Romero[1], Andrzej Kloczkowski[2], Vladimir N. Uversky[1], A. Keith Dunker[1]

[1]*Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 410 W. 10th St, HS5000, Indianapolis, IN46202, USA (kedunker@iupui.edu)*

[2]*Battelle Center for Mathematical Medicine, and the Nationwide Children's Hospital, Department of Pediatrics, The Ohio State University, Columbus, OH43210, USA*

[3]*Research and Information Systems, LLC, 1620 E. 72nd St. Indianapolis, IN46240 USA*

**Contributed equally (weilun.hsu@gmail.com, efaaggi@gmail.com)*

Disordered binding regions (DBRs), which are embedded within intrinsically disordered proteins or regions (IDPs or IDRs), enable IDPs or IDRs to mediate multiple protein-protein interactions. DBR-protein complexes were collected from the Protein Data Bank for which two or more DBRs having different amino acid sequences bind to the same (100% sequence identical) globular protein partner, a type of interaction herein called many-to-one binding. Two distinct binding profiles were identified: independent and overlapping. For the overlapping binding profiles, the distinct DBRs interact by means of almost identical binding sites (herein called "similar"), or the binding sites contain both common and divergent interaction residues (herein called "intersecting"). Further analysis of the sequence and structural differences among these three groups indicate how IDP flexibility allows different segments to adjust to similar, intersecting, and independent binding pockets.

Keywords: Molecular recognition feature; linear motif; disordered binding fragment; binding site.

1. Introduction

High throughput methods have enabled determinations of the protein interactome [1]. While most proteins bind to just one or a very few partners, some proteins, called hubs, bind to many partners. The deletion of the hubs is typically more deleterious than the deletion of non-hubs [2]. Additionally, hubs are inferred to have features that facilitate both their connections to multiple nodes and also their ability to establish new connections [3]. In [4], the following question was raised: what is the special feature that enables protein hubs to bind to multiple partners and to readily bind to new partners? This special feature was proposed to be intrinsically disordered

This work is supported by the grants R01 LM007688-01A1, GM071714-01A2 from the National Institute of Health and EF 0849803 from the National Science Foundation.

protein regions (IDRs) [5,6]. This proposal was based on prior work showing that IDRs enable binding to multiple partners [7,8]. Subsequently, both hub proteins and their binding partners were observed to be enriched in disorder [9–10]. These concepts have been supported by many additional studies [11–17].

Two IDR-based mechanisms have been suggested for enabling hub proteins to interact with multiple partners: (1) one disordered binding region (DBR) associates with many structured partners (one-to-many binding); and (2) many different DBRs associate with one structured partner (many-to-one binding) [5].

The one-to-many and many-to-one mechanisms are particularly well illustrated by two proteins, p53 and 14-3-3, respectively [6]. The p53 molecule uses its IDRs to interact with more than 100 partners, each DBR of p53 has one-to-many interactions with multiple partners by adopting different conformations [6,18,19]. The structured 14-3-3 has many-to-one interactions with more than 200 different IDRs, which bind to a common area on 14-3-3's surface, but with many of these IDRs lacking a common motif. These observations and further experiments showed the importance of IDRs for the binding of these multiple partners to 14-3-3 [20, 21].

Many additional examples of both one-to-many and many-to-one interactions have been assembled through data mining the Protein Data Bank (PDB) [22]; twenty-three of these one-to-many binding complexes were found to show similar results as observed for p53, namely that MoRFs binding to similarly-folded, low sequence identity proteins tended to fold the same way upon binding, that MoRFs binding to differently folded proteins tended to fold differently upon binding, and that post-translational modifications were frequently associated with DBRs found to fold differently upon binding to different proteins [22].

Our focus here is on identifying and characterizing additional many-to-one complexes. In addition to 14-3-3, many other protein-protein interactions are also mediated by the same many-to-one binding mechanism. Well known examples include DBRs that interact with SH3, SH2, PDZ and WW domains [56–58]. However, the true extent and diversity of DBR-mediated interactions is mostly unknown. The main goal here is to enlarge the number of many-to-one examples.

We examined the binding profiles of the many-to-one sets and performed structural analyses on the DBRs. Two general binding profiles were found: (1) independent, where two or more DBRs bind to completely independent sites; and (2) overlapping, where two or more DBRs bind to overlapping sites. For various pairs of DBRs, binding site overlap was observed to range from highly similar to minimally intersecting.

Two new measures, one focusing on the DBR and one focusing on the partner, have been developed to provide quantitative estimates of the degree of similarity for two non-identical DBRs binding to the same partner. The first measure is the amount of spatial superposition of the two non-identical DBRs when bound to their common partner, which is expressed as a volume overlap ratio (VOR). The second measure is the degree to which the two non-identical DBRs cover the same surface area upon binding to their common partner, which is expressed as the root-mean-square change in accessible surface area (RMSΔASA) of the partner upon binding by one MoRF as compared to the other. For a group of MoRF pairs, as two non-identical DBRs become more similar in their binding to their common partner, the VOR value tends towards 1 while the RMSΔASA value tends towards 0.

The detailed findings and results regarding the many-to-one binding diversity in protein disorder are described in the following sections, thus leading to a better understanding of MoRF-domain network biology and regulatory mechanisms based on IDRs. The large ranges observed for both the VOR and the RMSΔASA values indicate that non-identical MoRF pairs can bind to common partners using quite dissimilar sets of local interactions. These results have important implications for the characteristics of protein-protein interaction networks and for their evolution from prior networks.

2. Results

2.1. *Many-to-one binding datasets*

The procedure to collect our many-to-one binding dataset involves identifying DBRs bound to globular partners from PDB. The criteria for identifying DBRs are similar to those used in previous work [22,23] and essentially involve searching PDB for short peptides, 5 to 25 residues long, found in association with globular protein partners. These peptides are expected to be mostly DBRs because, with very rare exceptions, such peptides are too short to form independently folded globular domains. Indeed, our previous work revealed that DBR peptides are found for essentially all of the examples identified by this approach [22].

The steps used to create a many-to-one binding dataset, called Dataset_384 are shown in Table 1. The rationale for each step is given in the footnotes to Table 1.

Table 1. Dataset Construction

Dataset_384	DBRs/Part.	Clusters	Average DBRs per cluster
Initial DBR dataset (5-25) [a]	8084		
DBR dataset with biological interaction (>400Å^2) [b]	7064		
Partner dataset with sequence length (>40) [c]	6835		
Partner dataset mapped to UniProt sequence database	4612		
Partner dataset with overlapped region in mapping [d]	4368	514	8.5
Partner dataset without 100% sequence identity in DBRs	2081	384	5.4

[a]DBRs with 5 to 25 residues are the focus of this study

[b]400 Å^2 cutoff was set to filter out the spurious interactions caused by crystal contacts

[c]Binding partners of DBRs are supposed to be globular proteins having more than 40 residues to fold into a certain conformation.

[d]Partners having one or more overlapping residues with each other.

A cluster is defined as one protein with two or more PDB structures showing the given protein in association with two or more DBRs. The DBRs in these structures might or might not have identical sequences. As shown in Table 1, initially 514 clusters were found. When the requirement was added that only clusters with non-identical DBRs are counted, the number of clusters dropped to 384. Finally as shown in Table 1, the 384 clusters contained an overall average of 5.4 DBRs/cluster.

2.2. *Many-to-one binding profiles: independent and overlapping*

Visual inspection of Dataset_384 revealed two main binding profiles that describe many-to-one interactions. MoRF binding sites are located on separated parts of the partner surface (independent) or both common and distinct parts of the partner surface (overlapping). Independent binding sites are entirely separated, whereas the overlapping binding sites can range from very little overlap (intersecting) to almost complete overlap (similar).

Independent binding sites on the same partner often arise by having two or more different DBRs from the same polypeptide chain. Examples of this phenomenon include UPF2 that uses two DBRs to bind to two independent locations on UPF1 [27], Ste5 that uses two DBRs to bind to two independent locations on Fus3 [28], and phosphatase inhibitor 2 that uses two DBRs and one highly extended, mostly helical disordered domain [26] to bind to three different locations on phosphatase 1 [29].

Here the focus is on many-to-one binding for which the two different DBRs exhibit at least one residue of overlap. Two methods were developed for quantifying the extent of interface overlap (see Methods): the VOR and the RMSΔASA. Briefly, the VOR value quantifies the extent to which two DBRs occupy the same space when bound to the same partner, where a value of 1 indicates that the smaller MoRF is completely contained in the larger, and where a value of 0 indicates that the two DBRs are completely separated when bound. The RMSΔASA value quantifies the extent to which the two DBRs bury the same surface area when bound to the same partner, where a value of 0 indicates that the two DBRs bind to exactly the same location on the partner.

For simplicity, we determined the VOR of DBRs pair-by-pair. Neglecting the independent binding pairs, the pairwise VOR distribution follows a normal-like distribution (Figure 1).

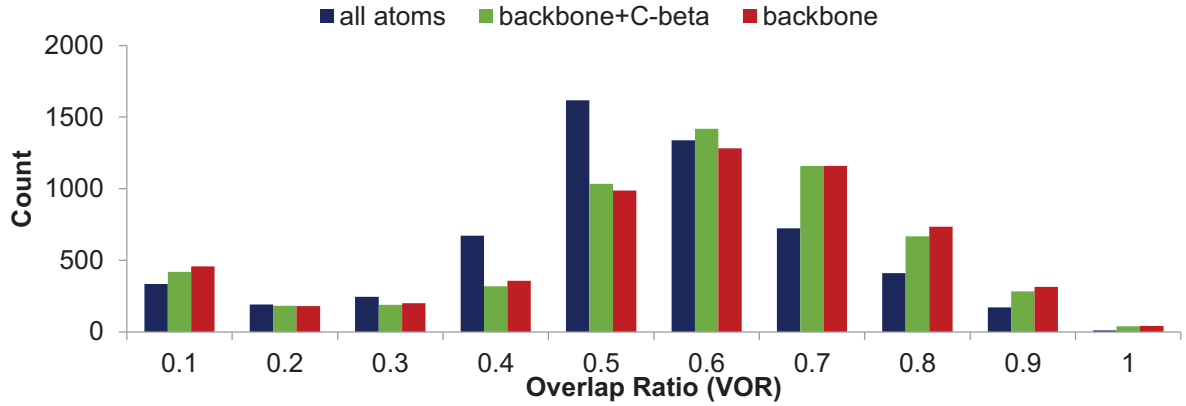


Fig. 1. Histograms of VORs calculated for each pair of DBRs in each many-to-one cluster of Dataset_160. Results from three different VOR calculations are based on different sets of atoms: all atoms (blue), backbone atoms + C_β (green), or backbone atoms only (red). Independent binding pairs are not included here.

The VORs were calculated in three different ways using various set of atoms, including all atoms, backbone + Cβ or only backbone. Two-sample Kolmogorov-Smirnov tests [66] were performed to examine whether the VORs calculated by different approaches have the same

distribution at 5% significance level. The results show that the VOR calculations for which all atoms are included exhibit significant differences from the calculations for which backbone only or backbone + Cβ are included, with p-values of $1.42x10^{-68}$, and $6.23x10^{-72}$, respectively. Similar distributions are found for VOR calculations carried over only backbone atoms, with and without Cβ atoms, with a p-value of 0.070. These results suggest that side chain identity and/or conformational differences have a critical effect on the VOR calculations. Since the volume of the side chains can account for a large portion of the volume of a protein this is to be expected.

2.3 Comparing VOR (with backbone only) and RMSΔASA Values

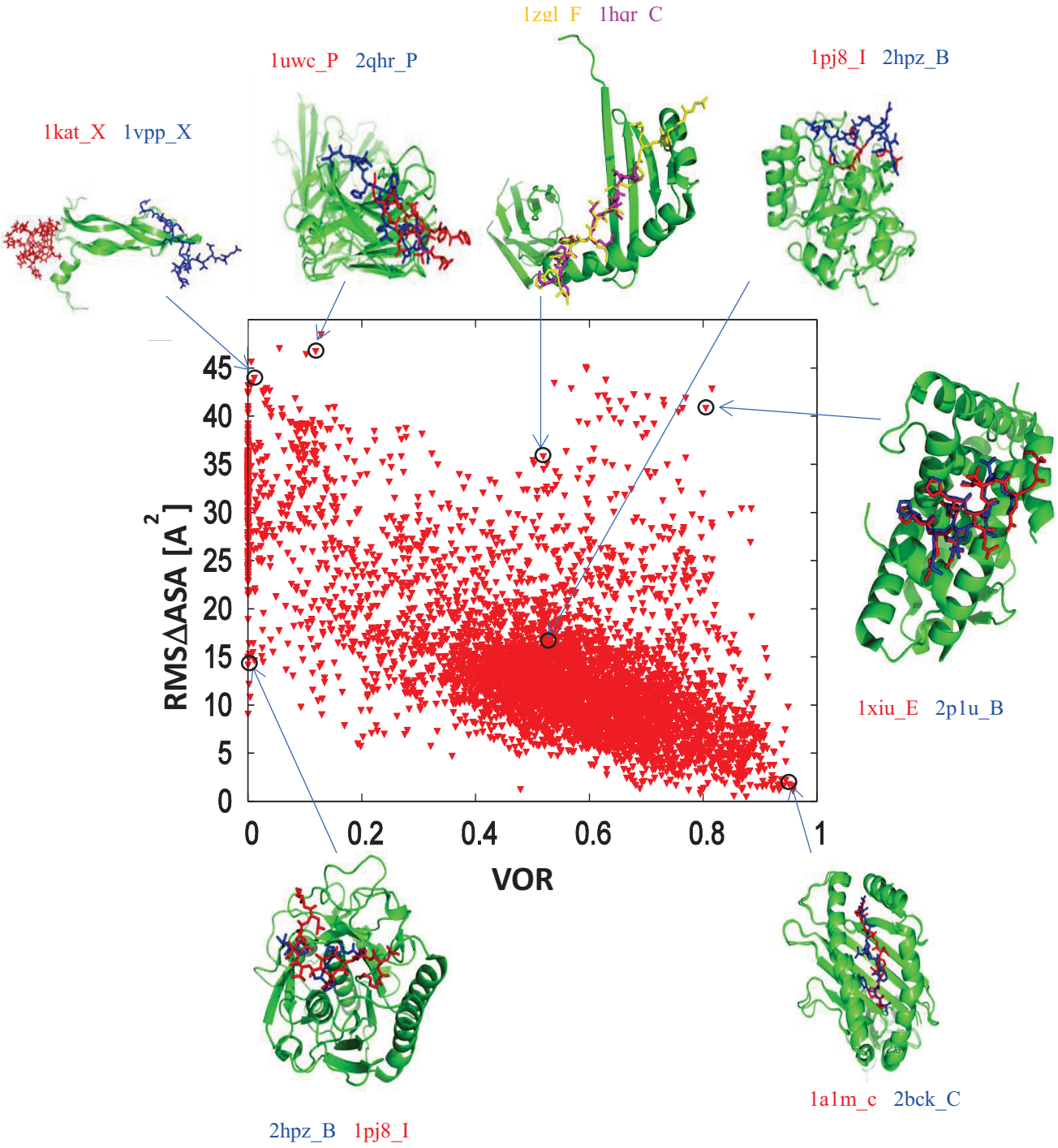
Comparison of the VOR and RMSΔASA with each other (Figure 2A, next page) shows a negative correlation, indicating as expected, highly overlapping DBRs tend to bind to the same residues on the partner, and less overlapping DBRs tend to bind to different residues on the partner. The Pearson correlation coefficient between the RMSΔASA and the VOR is -0.66 indicating a significant trend.

The excessive data scatter was explored by examining several representative examples of ribbon structures of pairs of DBRs bound in each case to the same partner (Figure 2A). In general we found two competing effects: the amount of close overlap between the MoRF pairs and the region of the structural partner they cover. Generally speaking if the MoRF pair's VOR is high then they cover similar regions on the structural partner's surface and have small RMSΔASA. However, since the VOR measures the overlap of the smaller MORF within the larger one, in some cases one MoRF has additional residues that interact with a distinct part of the structural partner's surface. This region may cause relatively high RMSΔASA values even for pairs with a high VOR. Similarly, low VOR and relatively low RMSΔASA can be observed for overlapping DBRs with the same binding site but different sequences, which results in non-overlapping backbone volumes.

To further understand the relationship between the VOR and the RMSΔASA we examined the special case of myelin basic protein (MBP) bound to HLA-DR2A (PDB ID 1ZGL) The DBR of MBP is in 93 pairs in a single DBR cluster (cluster ID 57). MBP has three chains with different residues with defined density, which makes each unique in our dataset, in the asymmetric unit of this structure, chains C, F, and I, each with 33, 31, and 29 pairs, respectively. Due to the slightly different sequences and the large number of HLA-DR2A partners, MBP/HLA-DR2A has the most pairs of any protein in our database.

The data pertaining to MBP/HLA-DR2A pairs shows several trends (Figure 2B). The first is that three groups of points appear in the plot, we label these A, B, and C as shown. The three groups contain points from different chains; hence they are not simply due to the three separate chains appearing in our database. Rather, the three clusters represent different scenarios of interaction pair similarity. In general, group C is composed of comparisons between the different MoRF chains of MBP in 1ZGL. The difference between groups A and B however is less trivial.

Figure 2A



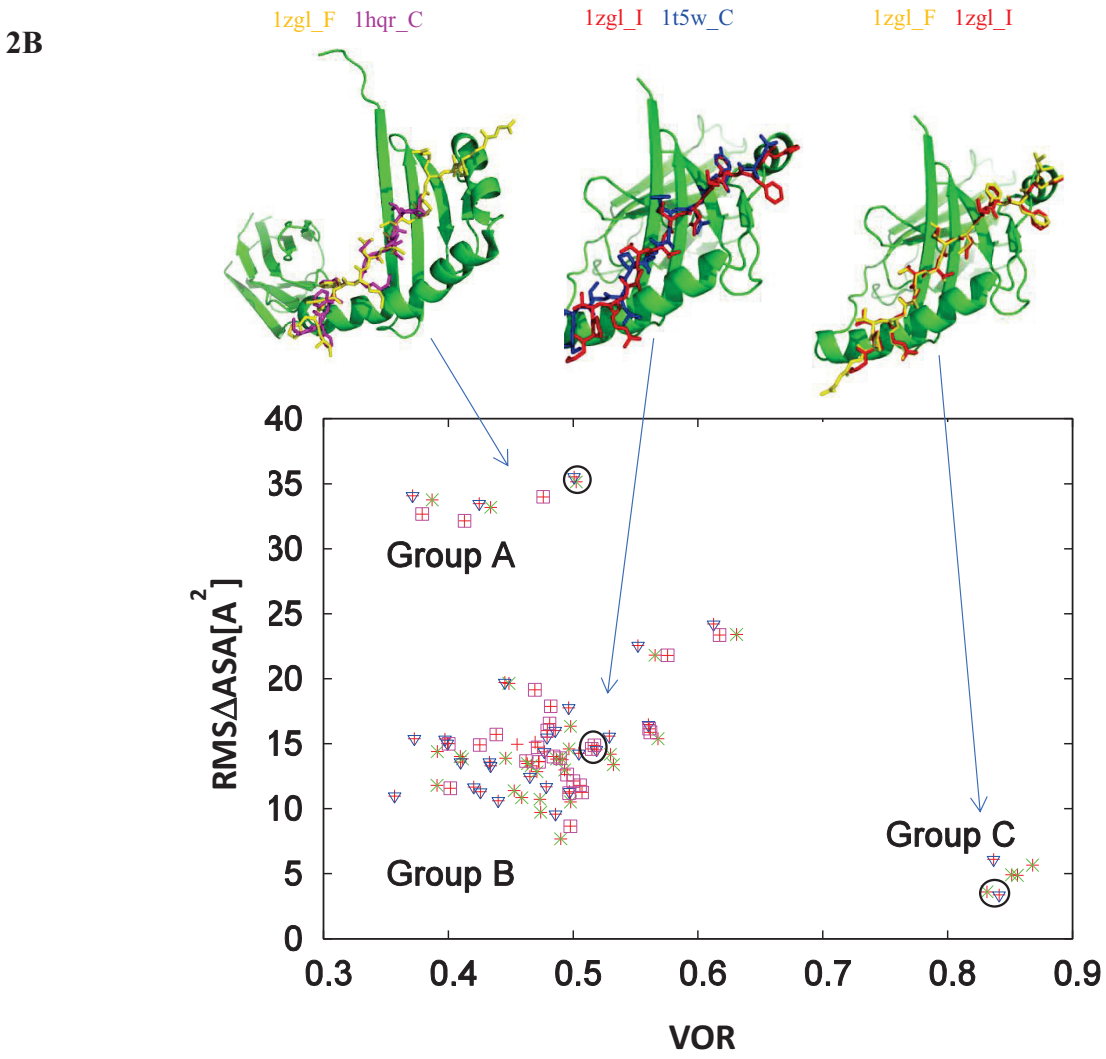


Fig. 2. Comparison of partner interface surface differences with MoRF chain overlap. (A) Plot of RMSΔASA versus backbone VOR for Dataset_160, with representative cartoons showing partners (green ribbons) and MoRF pairs (red and blue sticks) for various points in the plot. Arrows and black circles indicate the locations of these structure pairs in the plot. (B) Plot of RMSΔASA versus backbone VOR for an example cluster, the HLA-DR2A cluster (cluster ID 57). Representative cartoons showing partners (green ribbons) and MoRF pairs (sticks with other colors) for each labeled group of pairs (groups A, B, and C) show the main differences between the three groups. Actual points on plot where representative cartoons were made are marked with a black circle.

Group A is composed of MoRF pairs that differ in both volume overlap and area of interaction. Group B is composed of MoRF pairs whose backbones overlap for only about half their volume, however, they still interact with the same area of their HLA-DR2A. The reason that group A differs in area of interaction is exemplified by an example pair structure (Figure 2B, right most structure). In this structure, one MoRF partner has extra tail residues (Figure 2B, yellow structure).

While this additional tail does not affect the value of the VOR, it does significantly increase the ASA difference, since for one partner in the pair there is no buried surface area in that region at all. In this way, distinctions between the RMSΔASA and the VOR can be used to distinguish

various physical, chemical, and biological features of the DBRs. Finally, note that there is a baseline change in RMSΔASA that results from random variations. An approximation for the baseline value for this particular case can be obtained from cluster C with a value of about 5 A^2.

2.3. *Selected many-to-one case studies*

Several representative pairs of complexes with different binding profiles in peptide-protein interactions and protein-protein interactions are presented here (Figure 3). In addition, an example of reverse sequence binding, termed a retro-DBR is presented.

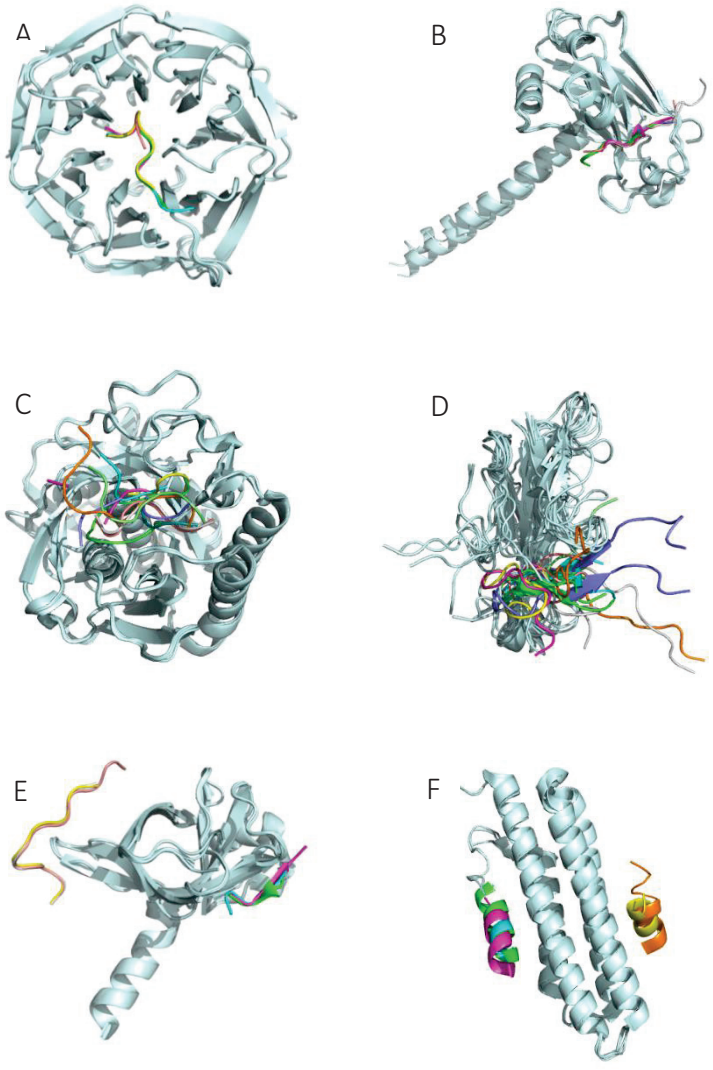


Figure 3. Examples of similar, intersecting, and independent binding sites. Structures are rendered with aligned partners (light blue ribbons) and each MoRF (different color ribbons). Examples of similar binding sites are (A) WD repeat protein 5 (cluster ID 160) and (B) TNF receptor associated factor 3 (cluster ID 136). Examples of intersecting binding sites are (C) proteinase K (cluster ID 52) and (D) α-bungarotoxin (cluster ID 8). Examples of independent binding sites are (E) chymotrypsin (cluster ID 17) and (F) focal adhesion kinase 1 (cluster ID 36). The PDB structures that make up each of these clusters are: WD repeat protein 5 (cluster ID 160) PDB IDs 2H13, 2H6Q, 2H6N, 2H6K, and 2O9K; TNF receptor associated factor 3 (cluster ID 136) PDB IDs 1CZY, 1QSC, 1D0A, 1D00, 1D01, 1D0J, 1CA9, and 1CA9; proteinase K (cluster ID 52) PDB IDs 1BJR, 2HPZ, 2HD4, 2DUJ, 1P7W, 3PRK, 1PFG, 1P7V, 1PJ8, and 2PQ2; α-bungarotoxin (cluster ID 8) PDB IDs 1HAA, 1HC9, 1ABT, 1KL8, 1LJZ, 1HOY, 1RGJ, 1BXP, and 1IDG; chymotrypsin (cluster ID 17) PDB IDs 1GHA, 1CHO, 1AB9, 1YPH, 3GCT, and 1GMD; and focal adhesion kinase 1 (cluster ID 36) PDB IDs 3B71, 3B71, 1OW8, 1OW7, and 1OW6.

For the nonimmune-related subset, example clusters for all three many-to-one binding profiles – similar, intersecting, and independent – are illustrated (Figure 3A/3B, Figure 3C/3D, and Figure

3E/3F, respectively). Examples for similar binding profiles are five DBRs bound to an overlapped sites on WD repeat protein 5 (Figure 3A) and 11 DBRs with 42% sequence identity bound to highly overlapped sites on TNF receptor associated factor 3 (Figure 3B). For intersecting binding profiles, examples are 10 distinct DBRs with 26% sequence identity bound to intersecting sites on proteinase K (Figure 3C) and 10 DBRs with 56% sequence identity bound to intersecting sites on alpha-bungarotoxin (Figure 3D). Finally, for independent binding profiles two different patterns of DBRs bind to two regions of chymotrypsin independently (Figure 3E) and two patterns of DBRs bind to two regions of focal adhesion kinase 1 independently (Figure 3F).

3. Discussion

The independent and overlapping binding profiles we observed in the many-to-one dataset provides a novel look at structured binding sites with binding diversity like 14-3-3. This dataset demonstrates that many-to-one domains are common, rather than just curiosities. Furthermore, the increase in identified clusters between dataset_160 and dataset_384 indicate that limited coverage of protein-peptide complexes in PDB is the primary limitation in the identification of many-to-one domains. These observations suggest that many-to-one domains, in combination with one-to-many DBRs [22], are a prevalent mechanism by which IDPs act as and enable PPI hubs.

Our results contribute to a better understanding of the role of DBRs that may serve as protein interaction hubs. Exploring the diverse binding partners of our collected MoRF sets and the corresponding complex conformations provide a Rosetta stone for interpretation of the underlying biological mechanisms and evolutional aptness. The importance and indispensability of hub proteins is apparent as they appear to evolve more slowly and are more likely to be vital for survival. Given their importance, many human disease-associated proteins related to cancer, diabetes, autoimmune disease, neurodegenerative disease and cardiovascular disease are found to have predicted DBRs as we expect. These DBRs associate with other structured partners and may be considered as promising druggable interactions [31]. Since IDPs have high tendency to participate directly in large numbers of pairwise protein-protein interactions, these promiscuous protein interactions are usually toxic when overexpressed; they are dosage sensitive. This is in contrast to a knockout or knockdown model, indicating there is something special about the excess participation specifically in pairwise interactions [32].

In terms of characterizing the nature of many-to-one binding, it is a challenge to quantitatively categorize overlapping binding sites into similar and intersecting subgroups. VOR calculations seems most useful compared to other methods such as RMSD measure of DBRs, sequence similarity of DBRs and structure alignment score of the MoRF-domain complexes. By the VOR measure, these results indicate that most overlapping sites occupy the majority of the same space on their partner, with only a few overlapping sites have very small VOR. These data shows that many-to-one domains with overlapping sites tend to use the same volume at their surface for MoRF binding. Since competitive binding is an important mechanism of cross-talk in cellular signaling, this suggests that high overlap may be required for mutually exclusive MoRF binding.
There are many more ideas to test on the many-to-one binding dataset. For example, the structures of MoRF pairs with different VORs will be examined for the pairwise secondary structure profiles

within different VOR groups (e.g. similar and intersecting). As another example, we have observed post-translation modifications (PTMs) and alternatively spliced events (ASEs) with regard to the DBRs in our one-to-many set and proposed that the three major players (DBRs, PTMs and ASEs) contribute significantly to the highly complex protein interaction networks in eukaryote cells [22]. However, establishing a more systematically computational approach is necessary to validate our hypothesis. Our many-to-one set is also a good resource for tracking possible mutation compensation in between MoRF and its partners which may lead to new concepts with regard to the structural basis of protein interaction networks.

4. Materials and Methods

4.1. *Dataset preparation*

PDB entries released before June 19, 2012 were used to collect many-to-one binding complexes (Dataset_384). Dataset_160 is a subset of Dataset_384 constructed in 2008 with the same searching criteria. The average MoRFs per cluster in Dataset_160 and Dataset_384 are similar (5.7 comparing to 5.4 per cluster). Protein chains with 5 to 25 visible residues in crystallographic electron density maps binding to other globular protein chains with minimum 40 residues are the disorder-order complexes we studied here. A cutoff interface size of 400 $Å^2$ was set to exclude non-biological interactions due to crystal packing[33]. To cluster partner sequences, the Fasta program was used to align the sequence of each partner to the UniProt sequence database to identify the source sequence. Complexes with partners with overlapping mapped source sequences were clustered together. Only identical DBRs (100% sequence identity) within each cluster were removed to keep all possible DBRs and their interaction patterns with the same partner in the data set.

4.2. *MoRF sequence similarity*

Sequence identity calculations in this study are based on the structure alignments. The sequence identities of DBRs within many-to-one clusters were obtained from PRALINE multiple sequence alignment server [34]. The transposed coordinates and multiple structure alignments were generated by MultiProt algorithm [35] using the complex structures including both MoRF and partner.

4.3. *MoRF interface similarity*

Two measures of interface similarity are used to compare the binding of two different DBRs to the same partner: VOR and root-mean-squared buried surface area (RMSΔASA). Both measures quantify the similarity of binding sites for a pair of DBRs, MoRFi and MoRFj, from the same many-to-one binding cluster. The VOR measures the extent to which two DBRs occupy the same space when bound to a common partner, and the RMSΔASA measures the extent to which two DBRs interact with the same residues of a common partner.

The VOR between DBRi and DBRj was calculated by first aligning the structures of the DBR-partner complexes then applying the formula:

$$\frac{\left(V_i + V_j - V_{ij}\right)}{\min\left(V_i, V_j\right)}$$

where Vi and Vj are the volume of DBRi and DBRj, respectively and Vij is the volume of the union of DBRi and DBRj. Volumes were calculated by setting a 0.1 Angstrom grid in space and counting the number of grid cubes that contain atomic coordinates. The union volume was calculated by counting grid cubes that contain coordinates from both structures. This normalization was selected so that the value of the VOR would be between zero and one. Note that by this choice the VOR measures the amount of overlap the smaller of the two DBRs has with the larger. PERL source and Linux executables used to calculate the VOR are available from Research and Information Systems, LLC at http://www.mamiris.com.

The RMSΔASA was calculated as the RMSD between the ΔASA profiles of the MoRF partner from the two MoRF-partner complexes. The ΔASA profiles were calculated by taking the difference between ASA calculated for the partner without and with the MoRF. ASAs were calculated with DSSP. Given two ΔASA profiles, ΔASAi for the partner of MoRFi and ΔASAj for the partner of MoRFj, then

$$RMS\Delta ASA \;=\; \sqrt{\frac{\sum_k^N \left(\Delta ASA_{i,k} - \Delta ASA_{j,k}\right)^2}{N}}$$

where $\Delta ASA_{x,k}$ is the kth position of profile ΔASAx, N is the number of positions that are non-zero in either the ΔASAi or ΔASAj profiles, and k are the indexes of the N non-zero positions.

References

1. A.-L. Barabási, et al.(2004) Nat Rev Genet. 5: 101–113.
2. H. Jeong, et al. (2004) Lethality and centrality in protein networks, Nature. 411: 41–42.
3. A.-L. Barabási, E. Bonabeau (2003) Sci. Am. 288: 60–69.
4. J. Hasty, J.J. Collins (2001) Nature. 411: 30–31.
5. A.K. Dunker, et al. (2005) FEBS J. 272:5129–5148.
6. C.J. Oldfield, et al. (2008) BMC Genomics. 9 Suppl 1 S1.
7. R.W. Kriwacki, et al. (1996)Proc. Natl. Acad. Sci. U.S.A. 93: 11504–11509.
8. A.K. Dunker, et al., (1998) Pac Symp Biocomput. (1998) 473–484.
9. C. Haynes, et al., (2006) PLoS Comput. Biol. 2: e100.
10. D. Ekmanet et al., (2006) S. Genome Biol. 7 R45.
11. Z. Dosztányi, et al. (2006) J. Proteome Res. 5: 2985–2995.
12. A. Patil, et al. (2006) FEBS Lett. 580: 2041–2045.
13. M. Boxem, et al. (2008) Cell. 134: 534–545.
14. G.P. Singh, et al. (2007) Proteins. 66: 761–765.
15. G.P. Singh, D. Dash, (2007) Proteins. 68: 602–605.

16. P.M. Kim, et al. (2008) Mol. Syst. Biol. 4: 179.
17. A. Mohan, et al. (2006) J. Mol. Biol. 362: 1043–1059.
18. V.N. Uversky, A.K. Dunker (2010) Biochim. Biophys. Acta. 1804: 1231–1264.
19. C.J. Oldfield, A.K. Dunker (2014) Annu. Rev. Biochem. 83: 553–584.
20. D.M. Bustos, A.A. Iglesias. (2006) Proteins. 63: 35–42.
21. D.M. Bustos. (2012) Mol Biosyst. 8: 178–184..
22. W.-L. Hsu, et al. (2013) Protein Sci. 22: 258–273.
23. V. Neduva, R.B. Russell. (2006) Curr. Opin. Biotechnol. 17: 465–471.
24. N. London, et al. (2010) Structure. 18: 188–199.
25. A. Zucconi, et al. (2000) FEBS Lett. 480: 49–54.
26. P. Tompa, et al. (2009) Bioessays. 31: 328–335.
27. M. Clerici, et al. (2009) EMBO J. 28: 2293–2306.
28. R.P. Bhattacharyya, et al. (2006) Science. 311: 822–826.
29. T.D. Hurley, et al. (2007) J. Biol. Chem. 282 (2007) 28874–28883.
30. A. Justel, et al. (1997) Statistics & Probability Letters. 35: 251–259.
31. Y. Cheng, et al. (2006) Trends Biotechnol. 24 (2006) 435–442..
32. E.M. Marcotte, M. Tsechansky. (2009) Cell. 138: 16–18.
33. K. Henrick, J.M. Thornton. (1998) Trends Biochem. Sci. 23: 358–361.
34. V.A. Simossis, J. Heringa, (2005) Nucleic Acids Res. 33 W289-294.
35. M. Shatsky, et al. (2004) Proteins. 56: 143–156.