

Databases and ontologies

DIBS: a repository of disordered binding sites mediating interactions with ordered proteins

Eva Schad¹, Erzsébet Fichó¹, Rita Pancsa^{2,1}, István Simon¹,
Zsuzsanna Dosztányi^{3,*} and Bálint Mészáros^{1,3,*}

¹Research Centre for Natural Sciences, Institute of Enzymology, Hungarian Academy of Sciences, Budapest H-1117, Hungary, ²MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK and ³MTA-ELTE Momentum Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, Budapest H-1117, Hungary

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on August 3, 2017; revised on September 29, 2017; editorial decision on October 5, 2017; accepted on October 6, 2017

Abstract

Motivation: Intrinsically Disordered Proteins (IDPs) mediate crucial protein–protein interactions, most notably in signaling and regulation. As their importance is increasingly recognized, the detailed analyses of specific IDP interactions opened up new opportunities for therapeutic targeting. Yet, large scale information about IDP-mediated interactions in structural and functional details are lacking, hindering the understanding of the mechanisms underlying this distinct binding mode.

Results: Here, we present DIBS, the first comprehensive, curated collection of complexes between IDPs and ordered proteins. DIBS not only describes by far the highest number of cases, it also provides the dissociation constants of their interactions, 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.

Availability and implementation: DIBS is freely accessible at <http://dibs.enzim.ttk.mta.hu/>. The DIBS application is hosted by Apache web server and was implemented in PHP. To enrich querying features and to enhance backend performance a MySQL database was also created.

Contact: dosztanyi@caesar.elte.hu or bmeszaros@caesar.elte.hu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Intrinsically Disordered Proteins (IDPs) play crucial roles in biological systems most notably in regulatory and signaling networks (Wright and Dyson, 2015). IDPs do not exhibit a well-defined tertiary structure in their isolated form even *in vivo*; however, in the vast majority of their interactions the interacting disordered segments adopt a stable structure (Sugase *et al.*, 2007). The study of bound IDP structures revealed distinct modes of interaction compared to those previously described for globular protein complexes (Mészáros *et al.*, 2007). The biophysical and structural characterization of even a relatively limited number of known complexes between IDPs and ordered proteins opened the way to the

development of dedicated prediction algorithms (Mészáros *et al.*, 2009; Malhis *et al.*, 2016; Meng *et al.*, 2017). In addition to structural characterization, interactions between an ordered domain and a short, flexible protein region are often described using alternative approaches, such as short linear motifs that define the residues in the flexible partner, essential for mediating the interaction with a specific domain. These consensus motifs were shown to generally reside in disordered regions (Fuxreiter *et al.*, 2007), and while their sequence-based definition fundamentally differs from the structure-centric definition of disordered binding sites, they most often describe the same biological interactions (Mészáros *et al.*, 2012).

While the structural studies of some select cases of IDP interactions even led to successful pharmaceutical targeting (Corbi-Verge and Kim, 2016), systematic analyses focusing on the general description of the underlying structural/functional principles remain scarce. While several disorder-specific databases exist, such as DisProt (Piovesan et al., 2017) or IDEAL (Fukuchi et al., 2012), these typically focus on the identification of disorder at the sequence level in general. Other databases, such as DisBind (Yu et al., 2017) and ELM (Dinkel et al., 2016) focus on the interactions IDPs mediate, however, they also use a sequence-based approach and lack detailed, atomic level structural information. The recently published MFIB database (Fichó et al., 2017) provides this foundation for interactions formed exclusively by IDPs. Furthermore, FuzDB provides examples about cases where IDPs do not fully undergo a disorder-to-order transition upon binding (Miskei et al., 2017). However, interactions between IDPs and ordered partner proteins still lack such an extensive data platform that could bridge the gap between structural details and functional interpretation, lay down the basis of the development of next generation prediction algorithms, or could serve as a starting point in unveiling the link between protein disorder and the emergence of diseases.

2 Construction of DIBS

The main aim of DIBS (Disordered Binding Sites) is to provide an extensive collection of interactions formed by a disordered protein region and one or more ordered protein partners. As the rationale behind DIBS is to enable the structural and functional studies of such complexes, only interactions with determined complex structures available in the PDB were considered. Furthermore, all constituent protein chains of the complex have to have experimental evidence for their disordered or ordered states in their unbound form. The structures contained in DIBS were collected and annotated using sequence-based database mapping between PDB, UniProt, DisProt, IDEAL, ELM and Pfam (Finn et al., 2016), transferring annotations between closely homologous proteins. This annotation procedure was complemented with extensive literature searches (see Supplementary Fig. S1).

A key element of DIBS is the annotation of order and disorder. According to the reliability of the evidence for disorder, annotations are grouped into three categories. Direct proofs of disorder were collected from dedicated databases of disordered proteins such as DisProt or IDEAL, and corresponding proteins are marked as ‘Confirmed’. In addition, many further cases with verified disorder status were found based on literature searches. Apart from direct experimental validation, DIBS also marks proteins as disordered if a close homologue was described to lack intrinsic structure. These cases are marked as ‘Inferred from homology’. In addition, the disordered state could also be inferred for protein regions that bind via a known, short functional motif (either from ELM, UniProt, Pfam or the literature). These entries are labelled as ‘Inferred from motif’ to reflect the less reliable assignment of the disordered status. The novelty of DIBS is apparent from the fact that it shares only a limited overlap with existing disorder databases, ranging between 5% and 48% for ELM, DisProt, IDEAL and DisBind, showing the extent of data originating from the manual processing of the literature.

Proofs for order were derived from the PDB. The interacting partner of the disordered segment was required to have a determined structure in the monomeric form for at least a close homologue. If the disordered partner interacts with an oligomer, then either all partner chains are required to be ordered in isolation or to form a stable complex without the disordered chain. In the former case, all proteins are marked as ‘Ordered’. In the latter case, chains of the ordered complex were labeled as ‘Ordered component’.

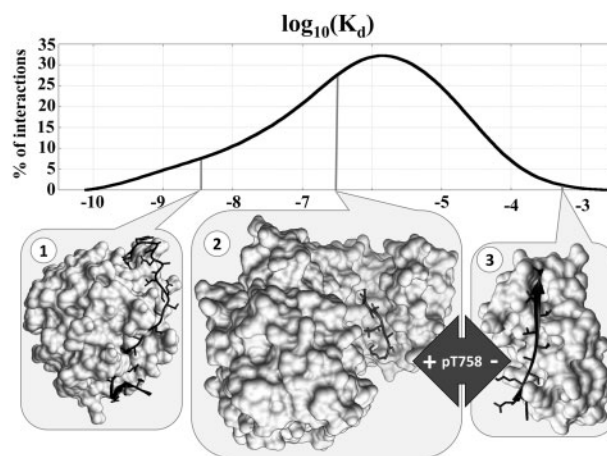


Fig. 1. K_d values in DIBS. Distribution of K_d values of interactions in DIBS (top), together with three selected examples with differing biological functions and, correspondingly, differing K_d s (bottom). Interactions 2 and 3 contain the same disordered protein segment in different post-translational modification states binding to different partners

IDPs cover a wide range of functions and their interactions are optimized in both specificity and binding strength. To indicate the biological functions of interactions in DIBS, annotations from the Gene Ontology are provided (The Gene Ontology Consortium, 2015). To better describe specificity, interactions are grouped according to the domain type of the ordered partner(s) to allow the analysis of various recognition mechanisms of a given protein fold by IDP partners. An extremely valuable information of DIBS is that it also describes the binding strength of the interactions by specifying the corresponding dissociation constants (K_d) where available, gathered from the literature in an exhaustive manual search by database curators. Figure 1 shows the distribution of K_d values of the 488 interactions for which such information is available, covering a wide range between approximately 10^{-3} M and 10^{-11} M. Figure 1 also shows three example interactions with markedly different K_d values. Interaction 1 shows the complex between anophelin—a blood-clotting inhibitor from mosquito—and α thrombin with a K_d of 3.65×10^{-9} M, indicating a remarkably tight, yet reversible interaction. The other two examples both involve integrin $\beta 2$, bound to 14-3-3 ζ (interaction 2) and bound to filamin A (interaction 3). Both interactions are transient, in line with their signaling roles, yet there is still three orders of magnitude difference between the two K_d values (2.61×10^{-7} M vs. 5.25×10^{-4} M). However, there is no direct competition between the two interactions as they are coordinated via a post-translational modification (PTM). Interaction 2 requires a phosphorylation at T758, while interaction 3 requires an unmodified integrin tail. As PTMs often confer specificity and can heavily affect binding strength in general, DIBS also includes PTM annotations for the disordered partners in all included interactions.

DIBS currently contains a total of 1577 structures grouped into 773 entries (merging structures describing essentially the same interactions). The majority of available complexes also feature known K_d values for structures with both direct disorder and motif-only disorder annotations, as shown in Figure 2. Figure 2 also shows the taxonomic distribution of DIBS entries. While interactions mediated by IDPs are prevalent in eukaryotic organisms, the wide coverage of DIBS is apparent from the inclusion of a large number of bacterial and cross-domain interactions, where the interacting protein chains come from organisms of different taxonomic domains.

All entries in DIBS, together with their annotations and related structures are available through a dedicated web-server. Each of the

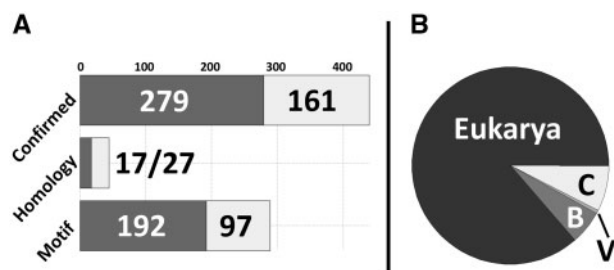


Fig. 2. Distribution of data in DIBS. (A) Distribution of DIBS entries with regards to various types of disorder annotation. Dark grey boxes represent entries for which K_d values are available and light grey boxes mark entries without known K_d values. (B) Taxonomic distribution of the 773 complexes in DIBS (B—Bacteria, V—Viruses, C—Cross-domain interactions). While archaeal proteins form parts of cross-domain interactions, no purely archaeal interactions are currently available

773 entries is assigned a separate page detailing structural and functional annotations, evidence for order and disorder, and a list of highly similar interactions. Entries containing the same interactors but different PTMs are linked to enable the efficient study of molecular switching mechanisms. DIBS also incorporates pages to aid browsing and searching (e.g. see [Supplementary Fig. S2](#)), basic database statistics as well as an extensive help section. DIBS is also available for download in basic text and XML formats together with format guides and corresponding structures.

3 Discussion

Complexes formed between IDPs and ordered proteins present critical elements of protein-protein interaction networks in general, with a particular importance in signaling and regulatory pathways. DIBS presents the first systematic and by far the largest collection of complexes between IDPs and ordered proteins in structural detail supported by high-quality, manually curated annotations. Structural description is connected to binding strength through the incorporation of a large amount of K_d and PTM data, while also providing the biological functions. DIBS also incorporates annotations about functional motifs in disordered partners, further connecting the two complementary models of such interactions. This, on one hand, strengthens the connection between motif occurrence and protein disorder; on the other hand provides a platform for the potential structure-based discovery of novel motifs.

While DIBS serves as a foundation for future analyses, some conclusions regarding the general features of the incorporated interactions are already apparent. One of the routinely quoted hallmark of IDP interactions is that—due to the loss of conformational freedom upon binding—there is heavy entropic penalty acting against their binding, giving rise to transient interactions ([Chu and Wang, 2014](#)). In theory this is undoubtedly true, as evidenced by the heavy involvement of IDPs in regulatory systems. However, as the distribution of K_d values in DIBS shows, K_d values for IDP-ordered protein interactions range from truly transient binding to unexpectedly tight complexes (with K_d s as low as 10^{-11} M), ultimately covering the full spectrum of biologically relevant binding strengths. This shows that binding strength is so heavily dependent on biological function that generic claims have only limited validity.

We believe that DIBS will serve as the basis for a more complete understanding of IDP interactions. DIBS not only integrates data from various databases but also adds novel examples based on extensive manual curation that are currently not recorded in disorder

related databases. This novel database can enhance the development of improved prediction algorithms and aid the future targeting of IDP-mediated interactions for biomedical and therapeutic purposes.

Acknowledgements

The authors would like to thank István Reményi and Gábor E. Tusnády for their help with setting up the DIBS server.

Funding

This work was supported by the PostDoc fellowship of the Hungarian Academy of Sciences for B.M, the European Molecular Biology Organization fellowship (ALTF 702-2015) to R.P., the Hungarian Research and Developments Fund [PD-OTKA 108772 (E.S.), OTKA K115698 (I.S.) and OTKA K108798 (Z.D.)], the “Lendület” grant from the Hungarian Academy of Sciences (LP2014-18) for Z.D., and the Project no. FIEK_16-1-2016-0005 financed under the FIEK_16 funding scheme (National Research, Development and Innovation Fund of Hungary).

Conflict of Interest: none declared.

References

- Chu, X. and Wang, J. (2014) Specificity and affinity quantification of flexible recognition from underlying energy landscape topography. *PLoS Comput. Biol.*, **10**, e1003782.
- Corbi-Verge, C. and Kim, P.M. (2016) Motif mediated protein-protein interactions as drug targets. *Cell Commun. Signal: CCS*, **14**, 8.
- Dinkel, H. et al. (2016) ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.*, **44**, D294–D300.
- Fichô, E. et al. (2017) MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinformatics*. <https://academic.oup.com/bioinformatics/article/4061276/MFIB-a-repository-of-protein-complexes-with-mutual?searchresult=1>
- Finn, R. et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
- Fukuchi, S. et al. (2012) IDEAL: intrinsically disordered proteins with extensive annotations and literature. *Nucleic Acids Res.*, **40**(Database issue), D507–D511.
- Fuxreiter, M. et al. (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23**, 950–956.
- Malhis, N. et al. (2016) MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.*, **44**, W488–W493.
- Meng, F. et al. (2017) Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell Mol. Life Sci.*, **74**, 3069–3090.
- Mészáros, B. et al. (2007) Molecular principles of the interactions of disordered proteins. *J. Mol. Biol.*, **372**, 549–561.
- Mészáros, B. et al. (2009) Prediction of protein binding regions in disordered proteins. *PLoS Computat. Biol.*, **5**, e1000376.
- Mészáros, B. et al. (2012) Disordered binding regions and linear motifs—bridging the gap between two models of molecular recognition. *PLoS One*, **7**, e46829.
- Miskei, M. et al. (2017) FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res.*, **45**, D228–D235.
- Piovesan, D. et al. (2017) DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.*, **45**, D1123–D1124.
- Sugase, K. et al. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, **447**, 1021–1025.
- The Gene Ontology Consortium (2015) Gene ontology consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
- Wright, P.E. and Jane Dyson, H. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
- Yu, J.F. et al. (2017) DisBind: a database of classified functional binding sites in disordered and structured regions of intrinsically disordered proteins. *BMC Bioinform.*, **18**, 206.