Databases and Ontologies

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

Eva Schad¹, Erzsébet Fichó¹, Rita Pancsa², István Simon¹, Zsuzsanna Dosztányi³,*, Bálint Mészáros¹,³,*

¹Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, 2 Magyar Tudósok krt, Budapest H-1117, Hungary, ²MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, United Kingdom, ³MTA-ELTE Momentum Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, 1/c Pázmány Péter stny, Budapest H-1117, Hungary

*To whom correspondence should be addressed.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

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 general processes underlying this distinct binding mode.

Results: Here we present DIBS, 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, 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: 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: bmeszaros@caesar.elte.hu, dosztanyi@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. 2015) 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 a wealth of 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 unraveling 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 were 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 figure 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 closely homologous protein 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 assignation 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 several proteins, then either all partner chains are required to be ordered in isolation, or they should form a stable complex without the disordered chain. In the former case, all proteins are marked as ‘Ordered’. In the latter case constituent chains of the ordered complex were labelled as ‘Ordered component’.

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 2014). 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 $10^{-8}$M and $10^{-15}$M. Figure 1 also shows three example interactions with markedly different $K_d$ values. Interaction 1 shows the complex between anaphelin – a blood-clotting inhibitor from mosquito – and α-thrombin with a $K_d$ of $3.65 \times 10^{-10}$M, indicating a remarkably tight, yet reversible interaction. The other two examples both involve integrin β2, bound to 14-3-3ζ (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 ($2.61 \times 10^{-8}$M vs. $5.25 \times 10^{-10}$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 1,577 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 constituent chains of the complex originate from organisms of more than one taxonomic domain.

All entries in DIBS, together with their annotations and related structures are available through a dedicated web-server at http://dibs.enzim.ttk.mta.hu/. Each of the 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 con-
DIBS: Disordered Binding Sites

taining 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 (see Supplementary figure S2 for examples), basic database statistics as well as an extensive help section describing the annotation procedure, among others. DIBS is also available for download in basic text and XML formats together with format guides and corresponding structures.

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.

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 annotations about the corresponding 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 it 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 features 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 extremely tight complexes (with $K_d$ as low as $10^{-10}$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 Development 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


