D786–D793 *Nucleic Acids Research, 2007, Vol. 35, Database issue* *Published online 1 December 2006*
doi:10.1093/nar/gkl893

DisProt: the Database of Disordered Proteins

Megan Sickmeier[1], Justin A. Hamilton[1], Tanguy LeGall[1], Vladimir Vacic[2], Marc S. Cortese[1], Agnes Tantos[3], Beata Szabo[3], Peter Tompa[3], Jake Chen[4], Vladimir N. Uversky[1,6], Zoran Obradovic[5] and A. Keith Dunker[1,4,*]

[1]Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46202, USA, [2]Computer Science and Engineering, University of California Riverside, Riverside, CA 92521, USA, [3]Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary, [4]School of Informatics, Indiana University, Indianapolis, IN 46202, USA, [5]Center for Information Science and Technology, Temple University, Philadelphia, PA 19122, USA and [6]Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

Received August 15, 2006; Revised October 5, 2006; Accepted October 10, 2006

ABSTRACT

The Database of Protein Disorder (DisProt) links structure and function information for intrinsically disordered proteins (IDPs). Intrinsically disordered proteins do not form a fixed three-dimensional structure under physiological conditions, either in their entireties or in segments or regions. We define IDP as a protein that contains at least one experimentally determined disordered region. Although lacking fixed structure, IDPs and regions carry out important biological functions, being typically involved in regulation, signaling and control. Such functions can involve high-specificity low-affinity interactions, the multiple binding of one protein to many partners and the multiple binding of many proteins to one partner. These three features are all enabled and enhanced by protein intrinsic disorder. One of the major hindrances in the study of IDPs has been the lack of organized information. DisProt was developed to enable IDP research by collecting and organizing knowledge regarding the experimental characterization and the functional associations of IDPs. In addition to being a unique source of biological information, DisProt opens doors for a plethora of bioinformatics studies. DisProt is openly available at http://www.disprot.org.

INTRODUCTION

The standard sequence-to-structure-to-function paradigm for proteins assumes that each protein first folds into a three-dimensional (3D) structure and that the resulting structure enables function via the lock and key (1) or the induced fit (2) models. Enzymes and their functions, which were traditionally the focus of studies in biochemistry, provided the basis for the lock and key and induced fit models; hence, as expected these models generally and perhaps universally explain enzymatic function. One reflection of the intimate relationship between protein structure and catalytic function is the relatively higher coverage of enzymes in the Protein Data Bank (PDB) as compared with other protein types (3).

Non-catalytic protein functions relating to signaling, regulation and control, such as protein–protein interactions, protein–DNA interactions, protein–RNA interactions, post-translational modifications and linker activities to name a few, are increasingly being studied. Many of these non-catalytic functions have been suggested to depend on, or have been experimentally demonstrated to depend on, proteins that lack fixed 3D structure, with interesting publications on this topic dating up to 70 years ago (4–8).

Functional proteins that lack the relatively fixed structure of enzymes and other globular proteins have been called 'rheomormorphic' (9), 'natively unfolded' (10), 'intrinsically unstructured' (11) and 'natively or intrinsically disordered' (12), among other terms. These proteins or protein regions exist as interconverting, dynamic ensembles of structures instead of folding into a single structure and many of their signaling or regulatory functions depend on their highly flexible nature.

Conformational flexibility facilitates a number of post-translational modifications, such as phosphorylation (13) and ubiquitination (14–16) for example, possibly because similar sequence segments in different proteins can use this flexibility to conform to the active sites of the modifying enzymes.

Many protein-binding interactions important for signaling and regulation involve modular binding domains that often associate with rather short linear motifs (17–21). In many cases, these interactions involve disorder-to-order transitions

*To whom correspondence should be addressed. Tel: +1 317 278 9650; Fax: +1 317 278 9217; Email: kedunker@iupui.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

for at least one of the partners. Such complex formation by coupled folding and binding (22) provides an important mechanism for achieving both high specificity and low affinity (8), which is an ideal combination for signaling and regulation.

Not only can individual disordered proteins and regions bind to multiple partners (4–6,23), but also multiple disordered sequences can each adapt to fit one partner (24). These partnering abilities of disordered proteins suggest their importance and common usage in protein interaction and signaling networks.

Some biological functions involve the flexibility itself, one important example being the ball and chain model for inactivation of voltage-gated ion channels (7). Often the flexibility of disordered regions provides a linker function to enable structured (or unstructured) functional domains to move relative to each other, which can lead to enhanced affinity.

The experimental data describing intrinsically disordered proteins (IDPs) are growing rapidly due in part to the increasing interest in signaling, regulation and control. The rapidly increasing number of IDP examples has generated the need for a publicly accessible repository. To facilitate efficient management and annotation of IDP information, the Database of Disordered Proteins (DisProt) was created. As of Release 3.4 (August 15, 2006), DisProt contained 460 IDPs and 1103 disordered regions, encompassing 35 functional categories—all based on published experimental data.

FUNCTIONS DEPENDING ON IDPs

Intrinsically disordered regions and proteins carry out a number of vital functions in the living cell. A new structure–function paradigm that extends the aforementioned sequence-to-structure-to-function model by including disorder as a type of structure provides the basis for describing the functions that depend on disorder (11). The 'Protein Trinity' hypothesis suggests that functional proteins can exist in one of three conformational states: the solid-like ordered state (globular proteins), the liquid-like collapsed disordered state (molten globule) and the gas-like extended disordered state (25). One more extended disordered conformation, the pre-molten globule state, was added to complete the 'The Protein Quartet' model (26). The pre-molten globule contains specific regions that transiently form regular secondary structure while extended disorder lacks a significant amount of such regions and behaves more like a typical random coil. Because the set of natively unfolded proteins probably forms a continuum ranging from little or no transient secondary structure to significant amounts of transient secondary structure but still without the compactness of the molten globule, it is uncertain whether the unfolded realm ought to remain as one category or be partitioned into two separate regions as suggested above. A difficulty with the partitioning is that it is unclear how to carry out this separation in a consistent manner. So given this uncertainty, function is then proposed to arise from any of these three (Trinity Model) or four (Quartet Model) states, or from transitions between them.

Currently, seven IDP-related high-level functional classifications have been proposed and are included in DisProt (Table 1). These are chaperone, entropic chain, metal sponge, modification site, molecular assembly, molecular recognition effectors and molecular recognition scavengers. More specific function annotations, referred to here as functional subclasses, are also provided (Table 1). Only a few of the currently identified 35 functional subclasses attributed to IDPs are included in Table 1. As additional biological processes and functions are continuously identified as being dependent upon protein intrinsic disorder, we anticipate that functional classes and subclasses will be expanded upon over time. Indeed, a recent bioinformatics study suggested that additional functions associated with protein disorder are evident in the literature (see Discussion). Mining this new source of information will add to the IDP functional classifications listed in DisProt.

EXAMPLE DisProt ENTRY

An illustrative example of typical DisProt entry is shown in Figure 1A. DisProt provides users with a number of tools to carry out a variety of biological and computational analyses. Some of these tools are as follows: disordered region sequence download, homologous protein sequence retrieval, functional narratives, graphical ordered and disordered region maps, isoform display, author-verified entries and a comprehensive bibliography for disordered proteins. From every protein entry page (Figure 1), the sequence of the protein and disordered region(s) can be downloaded via convenient links provided at the top of the page. Homologues included in DisProt obtained using the CD-HIT clustering program with a 50% identity threshold, are also accessible via links. A manually annotated description of the protein and its functional role(s) is provided in the functional narrative section. Information on the protein family, cellular localization, or relation to cancer or any other disease is provided in the narrative. When a protein includes both ordered and disordered regions and when the ordered segment(s) are in the PDB, the relevant PDB links are included. The protein and region map provides a visual representation of the location of the ordered and disordered regions in the context of the entire protein. Author-verified protein entries are proteins in which the author(s) of the referenced papers have reviewed and verified all disorder information available for that entry. At the time of this writing, DisProt contains 37 author-verified proteins. Efforts are being continually made to increase the number of author-verified entries.

A searchable bibliography (Figure 1B) that includes all the papers referenced by DisProt, together with papers that have cited several key references on disordered proteins and some other papers found by keyword searches. Although not displayed, abstract text is included in keyword searches in order to increase the usefulness of this collection. In addition, a link to the PubMed abstract is provided for every applicable paper. As of now, the total number of papers in the bibliography is 2289. The number of papers published per year in this field growing rapidly as evident in Figure 2.

An additional feature of DisProt related to individual protein entries that is not illustrated in Figure 1A is that isoforms included in DisProt (produced primarily through alternative splicing) are annotated as a sub-entry to the original protein. For example, information about isoform 1 of Calcineurin, DP00092, is coded in DP00092_A001. Since alternatively spliced segments of pre-mRNA were recently shown to

Table 1. Examples of IDP-related functional classes and subclasses

Functional class	Functional subclass	Functional subclass description	DisProt ID: region	Name	Structure/function type
Chaperone	Protein–Protein Binding	Binds to the target sequence of protein partner(s)	DP00412: Region 1	10 kDa chaperonin	Function arises via a disorder-to-order transition
Entropic Chain	Flexible Linker/Spacer	Provides separation and permits movement between adjacent domains	DP00231: Region III	POU domain, class 2, transcription factor 1	Function arises from the disordered state
	Entropic Bristle	A disordered region that creates a zone of exclusion by its entropic movement	DP00441: Region I	Glutamic acid-rich protein	Function arises from the disordered state
Metal Sponge	Metal Binding	Store and help remove/neutralize heavy metals for detoxification	DP00205: Region I	Small metal binding protein	Function arises via a disorder-to-order transition
Modification Site	Phosphorylation	Guides the addition of a phosphate to the protein	DP00199: Region I	Beta casein [Precursor]	Function arises from the disordered state
	Acetylation	Guides the addition of an acetyl group during chemical modification of the protein.	DP00501: Region 1	Positive cofactor 4	Function arises from the disordered state
Molecular Assembly	Protein-Protein Binding	Binds to the protein partner(s)	DP00230: Region 1	Growth factor receptor-bound protein 14	Function arises via a disorder-to-order transition
	Protein-DNA Binding	Binds to DNA	DP00423: Region 1	Hho1p	Function arises via a disorder-to-order transition
	Polymerization	Facilitates polymerization	DP00200: Region 1	T-cell surface glycoprotein CD3 zeta chain [Precursor]	Function arises from the disordered state
	Transactivation (Transcriptional Activation)	Mediates transcriptional activation	DP00492: Region 1	Androgen receptor	Function arises via a disorder-to-order transition
Molecular Recognition Effectors	Substrate/Ligand Binding	Binds to substrate(s) and/or ligand(s)	DP00112: Region I	Dessication-related protein	Function arises via a disorder-to-order transition
Molecular Recognition Scavenger	Metal Binding	Binds small molecules (metals)	DP00359: Region I	CP12	Function arises via a disorder-to-order transition

The first column contains one of each of the high-level functional classes currently included in DisProt. Information in the remaining columns includes illustrative details for exemplary annotated proteins from the database with disordered regions that perform the indicated functions. Regions may be annotated with multiple subclasses (this is not reflected in the table).

code for regions of intrinsic disorder much more frequently than they code for regions of 3D structure (27), this feature is likely to become increasingly important over time.

Although the data in DisProt are based on experimental data, predictors of protein disorder have been found to be useful for the analysis of relationships among primary amino acid sequence, structure and function of proteins (28,29). DisProt contains references and URLs for 15 different predictors of intrinsic disorder. This list is updated as new predictors become available.

Through display and tabulation of known disorder information of individual proteins combined with relationships among isoforms and similar proteins, DisProt can supply useful examples for crystallographers and structural biologists as an aid to solving the structures of target proteins that contain unstructured regions.

DETECTION AND CHARACTERIZATION OF IDPs

Disorder can typically be characterized by X-ray crystallography, NMR spectroscopy, CD spectroscopy (both far and near UV) and protease sensitivity in addition to several other less frequently used experimental techniques. A comprehensive list of all detection methods currently used to characterize IDPs and the descriptions of these techniques can be found using convenient links provided in the database. Each protein entry includes the method(s) used for disorder characterization as well as the specific experimental conditions. Clicking on a detection method link brings up a list of all proteins in the database that have been characterized using that particular method.

DATABASE STRUCTURE AND WEB INTERFACE

The DisProt database is implemented as a relational database using PostgreSQL. A simplified representation of DisProt can be found in Figure 3. DisProt is supported by an Apache web server with the web interface implemented using PHP and JavaScript. DisProt is available to the public and can be accessed at http://www.disprot.org/.

The web interface allows users to browse through the list of proteins. It is also possible to query the database using amino acid sequence, keywords, protein name, organism name or accession numbers from UNIPROT, SWISSPROT, NCBI and other databases. Query sequences are searched, using RPS-Blast, against a database containing the profiles of all the proteins in the current release. Sequence homologues within DisProt are found using the CD-HIT-2D program

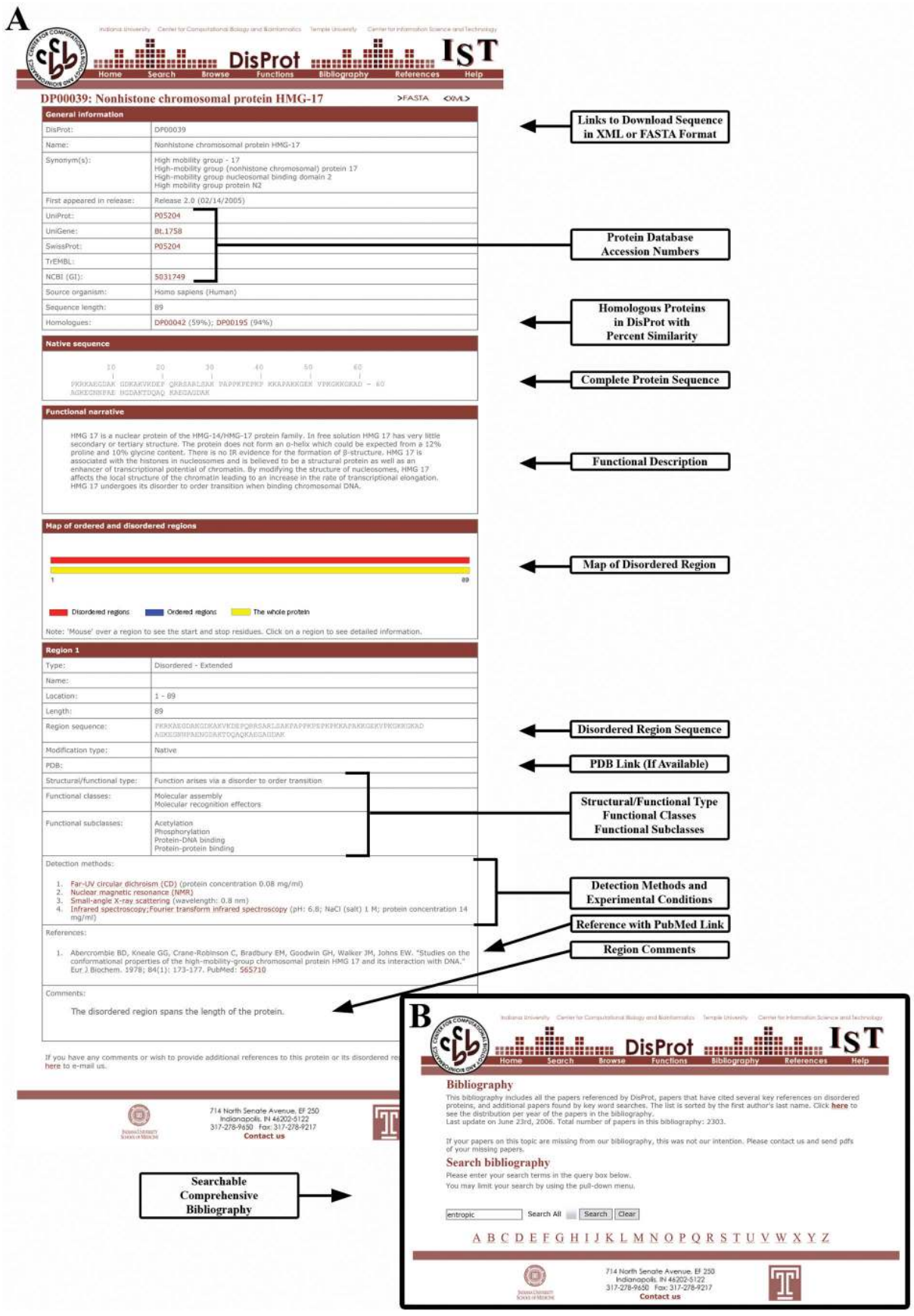

Figure 1. DisProt screen captures. (**A**) protein display for DP000039, Non-histone chromosomal protein HMG-17, and (**B**) bibliography query page.

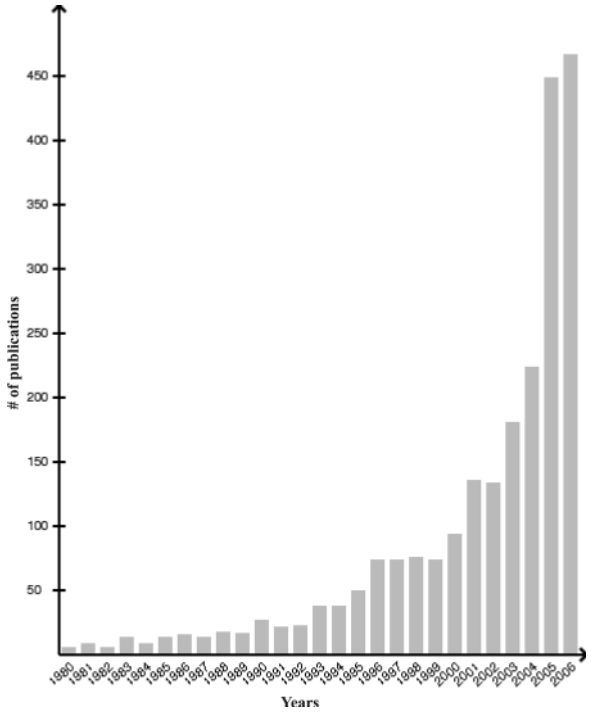

Figure 2. The number of papers tallied by year referencing protein disorder that are included in the searchable bibliography of DisProt.

with a 50% identity threshold. In addition, the complete DisProt database is available for download in the FASTA and XML format (http://www.disprot.org/downloads.php).

DISCUSSION

DisProt is the central repository for structure–function annotations associated with protein intrinsic disorder. The database has been used by researchers from over 35 countries worldwide. These users are involved in many branches of protein science and are from a variety of different organization types, including academic, industry and government.

Interest in IDPs is growing rapidly. The nearly 450 papers published in 2005 is about twice the number published in 2004, and the number of papers in 2006, although the year is not complete as of this writing, shows clear evidence of a continued rapid rise (Figure 2). Because inconsistent nomenclature is used to describe these proteins, the bibliography in DisProt is very useful to researchers in this rapidly growing field. The rapid growth of disorder-related publications argues for the importance of having an organized database such as DisProt. Researchers who study IDPs are encouraged to look over this bibliography and to send us the citations of papers that we have missed.

Our initial attempt to identify disorder-function relationships led to 28 specific functions that we grouped into four classes (12). A different schema was proposed at about the same time that led to additional functions being added (30). The 7 and 35 functional classes and subclasses, respectfully, in the current DisProt came mostly from these prior publications, but with a number of additions that were discovered during the process of annotating proteins.

A major usage of intrinsic disorder is for molecular recognition and binding. A search of the PDB for short segments called molecular recognition features, MoRFs, was carried out. MoRFs undergo disorder-to-order transitions upon binding to their partners, which typically have globular structure. This search yielded 1261 MoRFs that were clustered into 372 sets on the basis of high-sequence identity among members of a given set (31). Many of these MoRFs have experimental data supporting disorder-to-order transitions upon binding. In addition, proteins that undergo disorder-to-order transitions upon complex formation are distinguishable from globular proteins that associate with one another. The disorder-based complexes have larger monomer surface areas and larger interaction surface areas as compared to interacting globular proteins (32). All of the 372 MoRF-partner complexes exhibit these large surface areas for the monomers and interfaces; furthermore, nearly all of the MoRFs have substantial prediction of disorder in their flanking regions as well. Both these observations support the concept that these interactions involve disorder-to-order transitions of the MoRFs (31). Direct experimental evidence in support of this concept has been presented in the case of deacetylation (33) and phosphorylation sites, SH3 interaction motifs (34) and recognition elements of 14-3-3 proteins (24), which all have been found in locally disordered regions of their parent proteins. The possible generality of this mode of protein–protein interactions has also been underlined by predicting the local structural preferences of interaction sites of IDPs (35).

Another approach, based on sequence comparison rather than analysis of structures in PDB, has been used to systematically identify short, linear motifs that bind to protein partners (20,21,36). These sequence-identified segments have been collected and are contained in the Eukaryotic Linear Motif (ELM) server (http://elm.eu.org) (20). Many of the sequence-based ELMs and the structure-based MoRFs are simply different descriptions of the same protein segments, and in these cases the ELMs likely undergo coupled binding and folding upon association with their partners (Fuxreiter, Tompa and Simon, work in progress). The ELMs that have been experimentally verified to be unstructured in the absence of their partners and the ELM-MoRF matches will be added to DisProt with appropriate cross-references to the ELM collection.

We recently carried out a bioinformatics study to determine which Swiss Protein keywords were associated with the prediction of long disordered regions and which keywords were associated with the absence of such predictions. Of 711 function-associated keywords for which there were enough protein examples in Swiss Protein to make statistical inferences, 302 keywords were strongly associated with the absence of disorder prediction and 262 were strongly associated with the prediction of disorder. Manual literature searches provided numerous confirmatory examples for which laboratory experiments verified the direct involvement of disordered regions in carrying out the identified functions (H. Xie, S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic and V.N. Uversky, submitted for publication). In the coming year, we will focus our annotation efforts on finding papers that determine whether or not IDPs are directly responsible for carrying out the 262 functions that

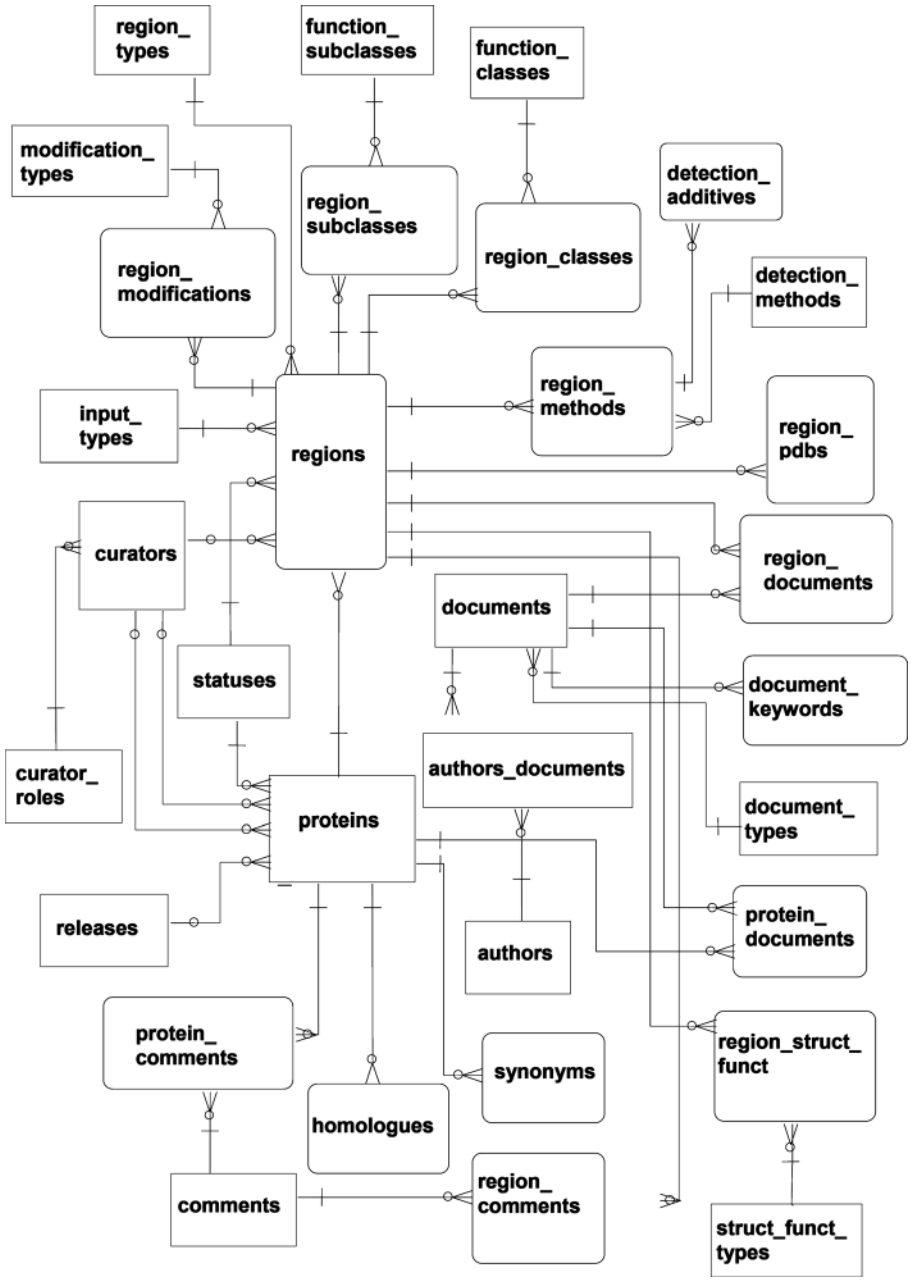

Figure 3. Simplified IDEF1X representation of the DisProt database structure. Boxes with round edges represent tables with at least one foreign key. Dashed lines with an oval at one end and a cross at the other represent mandatory relationships (not null foreign keys), plane lines with an oval at one end and a cross at the other represent identifying relationship (foreign keys, possibly null, for which the field names in the tables are identical), dashed lines with an oval at both ends represent non-identifying optional relationships (foreign keys, possibly null, for which the field names are not identical).

were indicated to be IDP-associated. This bioinformatics-directed DisProt expansion will enable us to rapidly increase the number of experimentally verified disordered protein–function relationships.

FUTURE DEVELOPMENTS

By expanding the number of different functions associated with experimentally characterized IDPs, DisProt will become increasingly more useful in the field of genome annotation. Since intrinsically disordered regions often show high-sequence variability compared with structured regions in the same proteins, identifying functional homologues by sequence matching will be generally more difficult for IDPs than for structured proteins. We previously showed that disordered regions could be classified based on differences in sequence properties and these disordered regions with differing sequence properties showed differences in function. A major goal in concert with the expansion of DisProt will be to refine the associations between different functions and different sequence properties, thus facilitating the use of DisProt for function annotation. Given the high

frequency of IDPs and the large number of functions carried out by these proteins, future versions of DisProt and the associated tools to be developed will become essential for complete function annotation of proteomes.

Another direction of the DisProt future development will be the elaboration of tools for the IDP sequence analyses. We plan to add features such as BLAST-like analyses and disorder prediction service. The availability of BLAST-like analyses using scoring matrices associated with the IDP-specific sequence features would enable sequence/function relationships amongst IDPs to be more accurately analyzed. Although in its current configuration DisProt contains links to known publicly available disorder predictors, a planned development is a disorder prediction service, in which results are obtained from multiple servers and compiled in one report (i.e. a service resembling the PredictProtein server (37,38). Rather than limiting users to the predictors we have developed, we hope to make an array of predictors available, which can then be used in combination for more detailed analysis as described recently (39). Another useful tool is the recently described disorder score versus sequence complexity plot (40); these plots appear to be extremely useful for comparing IDPs. In these ways, DisProt will evolve into a resource with both information and tools.

ACKNOWLEDGEMENTS

We would like to thank Predrag Radivojac, Pedro R. Romero, Christopher J. Oldfield, Jie Sun and Joy Nellis for their contributions to the establishment of this database. We would also like to thank all the past and present annotators: William Breidenstein, John Turner, Roger Morse, Elizabeth Patterson, Amy Lewis, Shelly Riggen and Jason P. Baird. This work was supported by NIH Grant RO1 LM007688-01A1 and by the Indiana Genomics Initiative (INGEN), which is funded in part by the Lilly Endowment. P.T. acknowledges the support of the Wellcome Trust International Senior Research Fellowship ISRF 067595. Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Fischer,E. (1894) Einfluss der configuration auf die wirkung derenzyme. *Ber. Dt. Chem. Ges.*, **27**, 2985–2993.
2. Koshland,D.E., Jr, Ray,W.J., Jr and Erwin,M.J. (1958) Protein structure and enzyme action. *Fed. Proc.*, **17**, 1145–1150.
3. Xie,L. and Bourne,P.E. (2005) Functional coverage of the human genome by existing structures, structural genomics targets, and homology models. *PLoS. Comput. Biol.*, **1**, e31.
4. Landsteiner,D.P. (1936) *The Specificity of Serological Reactions.* Dover, New York.
5. Pauling,L. (1940) A theory of the structure and process of formation of antibodies. *J. Am. Chem. Soc.*, **62**, 2643–2657.
6. Karush,F. (1950) Heterogeneity of the binding sites of bovine serum albumin. *J. Am. Chem. Soc.*, **72**, 2705–2713.
7. Armstrong,C.M. and Bezanilla,F. (1972) Inactivation of the sodium channel: II. Gating current experiments. *J. Gen. Physiol.*, **70**, 567–590.
8. Schulz,G.E. (1979) Nucleotide binding proteins. In Balaban,M. (ed.), *Molecular Mechanisms of Biological Recognition*. Elsevier/North Holland Biomedical Press, New York, pp. 79–94.
9. Holt,C. and Sawyer,L. (1993) Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the alphaS1-, beta- and kappa-caseins. *J. Chem. Soc. Faraday Trans.*, **89**, 2683–2692.
10. Weinreb,P.H., Zhen,W., Poon,A.W., Conway,K.A. and Lansbury,P.T., Jr (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*, **35**, 13709–13715.
11. Wright,P.E. and Dyson,H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.*, **293**, 321–331.
12. Dunker,A.K., Brown,C.J., Lawson,J.D., Iakoucheva,L.M. and Obradovic,Z. (2002) Intrinsic disorder and protein function. *Biochemistry*, **41**, 6573–6582.
13. Ruzza,P., Donella-Deana,A., Calderan,A., Filippi,B., Cesaro,L., Pinna,L.A. and Borin,G. (1996) An exploration of the effects of constraints on the phosphorylation of synthetic protein tyrosine kinase peptide substrates. *J. Pept. Sci.*, **2**, 325–338.
14. Dunten,R.L. and Cohen,R.E. (1989) Recognition of modified forms of ribonuclease A by the ubiquitin system. *J. Biol. Chem.*, **264**, 16739–16747.
15. Cox,C.J., Dutta,K., Petri,E.T., Hwang,W.C., Lin,Y., Pascal,S.M. and Basavappa,R. (2002) The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. *FEBS Lett.*, **527**, 303–308.
16. Yoshida,Y., Adachi,E., Fukiya,K., Iwai,K. and Tanaka,K. (2005) Glycoprotein-specific ubiquitin ligases recognize N-glycans in unfolded substrates. *EMBO Rep.*, **6**, 239–244.
17. Pawson,T. and Scott,J.D. (1997) Signaling through scaffold, anchoring, and adaptor proteins. *Science*, **278**, 2075–2080.
18. Kuriyan,J. and Cowburn,D. (1997) Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 259–288.
19. Yaffe,M.B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nature Rev. Mol. Cell Biol.*, **3**, 177–186.
20. Puntervoll,P., Linding,R., Gemund,C., Chabanis-Davidson,S., Mattingsdal,M., Cameron,S., Martin,D.M., Ausiello,G., Brannetti,B., Costantini,A. *et al.* (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.*, **31**, 3625–3630.
21. Neduva,V. and Russell,R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.*, **579**, 3342–3345.
22. Spolar,R.S. and Record,M.T., Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
23. Kriwacki,R.W., Hengst,L., Tennant,L., Reed,S.I. and Wright,P.E. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl Acad. Sci. USA*, **93**, 11504–11509.
24. Bustos,D.M. and Iglesias,A.A. (2006) Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins*, **63**, 35–42.
25. Dunker,A.K. and Obradovic,Z. (2001) The protein trinity—linking function and disorder. *Nat. Biotechnol.*, **19**, 805–806.
26. Uversky,V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.*, **11**, 739–756.
27. Romero,P.R., Zaidi,S., Fang,Y.Y., Uversky,V.N., Radivojac,P., Oldfield,C.J., Cortese,M.S., Sickmeier,M., LeGall,T., Obradovic,Z. *et al.* (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl Acad. Sci. USA*, **103**, 8390–8395.
28. Jin,Y. and Dunbrack,R.L., Jr (2005) Assessment of disorder predictions in CASP6. *Proteins*, **61** (Suppl. 7), 167–175.
29. Wang,G., Jin,Y. and Dunbrack,R.L., Jr (2005) Assessment of fold recognition predictions in CASP6. *Proteins*, **61** (Suppl. 7), 46–66.
30. Tompa,P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.*, **27**, 527–533.
31. Mohan,A., Oldfield,C.J., Radivojac,P., Vacic,V., Cortese,M.S., Dunker,A.K. and Uversky,V.N. (2006) Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.*, **362**, 1043–1059.
32. Gunasekaran,K., Tsai,C.J. and Nussinov,R. (2004) Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J. Mol. Biol.*, **341**, 1327–1341.
33. Khan,A.N. and Lewis,P.N. (2005) Unstructured conformations are a substrate requirement for the Sir2 family of NAD-dependent protein deacetylases. *J. Biol. Chem.*, **280**, 36073–36078.
34. Beltrao,P. and Serrano,L. (2005) Comparative genomics and disorder prediction identify biologically relevant SH3 protein interactions. *PLoS Comput. Biol.*, **1**, e26.

35. Fuxreiter,M., Simon,I., Friedrich,P. and Tompa,P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.*, **338**, 1015–1026.
36. Obenauer,J.C., Cantley,L.C. and Yaffe,M.B. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.
37. Rost,B. and Liu,J. (2003) The PredictProtein server. *Nucleic Acids Res.*, **31**, 3300–3304.
38. Rost,B., Yachdav,G. and Liu,J. (2004) The PredictProtein server. *Nucleic Acids Res.*, **32**, W321–W326.
39. Ferron,F., Longhi,S., Canard,B. and Karlin,D. (2006) A practical overview of protein disorder prediction methods. *Proteins*, **65**, 1–14.
40. Weathers,E.A., Paulaitis,M.E., Woolf,T.B. and Hoh,J.H. (2006) Insights into protein structure and function from disorder-complexity space. *Proteins*, **65**, in press.