

MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins

Damiano Piovesan^{1,†}, Francesco Tabaro^{1,2,†}, Lisanna Paladin¹, Marco Necci^{1,3,4}, Ivan Mičetić¹, Carlo Camilloni⁵, Norman Davey^{6,7}, Zsuzsanna Dosztányi⁸, Bálint Mészáros^{8,9}, Alexander M. Monzon¹⁰, Gustavo Parisi¹⁰, Eva Schad⁹, Pietro Sormanni¹¹, Peter Tompa^{9,12,13}, Michele Vendruscolo¹¹, Wim F. Vranken^{12,13,14} and Silvio C.E. Tosatto^{1,15,*}

¹Department of Biomedical Sciences, University of Padua, via U. Bassi 58/b, 35131 Padua, Italy, ²Institute of Biosciences and Medical Technology, Arvo Ylpön katu 34, 33520 Tampere, Finland, ³Department of Agricultural Sciences, University of Udine, via Palladio 8, 33100 Udine, Italy, ⁴Fondazione Edmund Mach, Via E. Mach 1, 38010 S. Michele all'Adige, Italy, ⁵Department of Biosciences, University of Milan, 20133 Milano, Italy, ⁶Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Belfield, Dublin 4, Ireland, ⁷UCD School of Medicine & Medical Science, University College Dublin, Belfield, Dublin 4, Ireland, ⁸MTA-ELTE Lendület Bioinformatics Research Group, Department of Biochemistry, Eötvös Loránd University, 1/c Pázmány Péter sétány, H-1117, Budapest, Hungary, ⁹Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, PO Box 7, H-1518 Budapest, Hungary, ¹⁰Structural Bioinformatics Group, Department of Science and Technology, National University of Quilmes, CONICET, Roque Saenz Pena 182, Bernal B1876BXD, Argentina, ¹¹Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK, ¹²Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels 1050, Belgium, ¹³VIB-VUB Center for Structural Biology, Flanders Institute for Biotechnology (VIB), Brussels 1050, Belgium, ¹⁴Interuniversity Institute of Bioinformatics in Brussels, ULB/VUB, 1050 Brussels, Belgium and ¹⁵CNR Institute of Neuroscience, via U. Bassi 58/b, 35131 Padua, Italy

Received September 22, 2017; Revised October 13, 2017; Editorial Decision October 16, 2017; Accepted October 19, 2017

ABSTRACT

The MobiDB (URL: mobidb.bio.unipd.it) database of protein disorder and mobility annotations has been significantly updated and upgraded since its last major renewal in 2014. Several curated datasets for intrinsic disorder and folding upon binding have been integrated from specialized databases. The indirect evidence has also been expanded to better capture information available in the PDB, such as high temperature residues in X-ray structures and overall conformational diversity. Novel nuclear magnetic resonance chemical shift data provides an additional experimental information layer on conformational dynamics. Predictions have been expanded to provide new types of annotation on backbone rigidity, secondary structure preference and disordered binding regions. MobiDB 3.0 contains information for the complete UniProt protein set and synchronization has been improved by covering all UniParc sequences. An advanced search function allows the

creation of a wide array of custom-made datasets for download and further analysis. A large amount of information and cross-links to more specialized databases are intended to make MobiDB the central resource for the scientific community working on protein intrinsic disorder and mobility.

INTRODUCTION

The protein structure-function paradigm is a cornerstone of molecular biology, offering a mechanistic understanding of processes ranging from enzyme catalysis, signal transduction to molecular recognition and allosteric regulation. Underlying this paradigm is the assumption that proteins become functional by assuming a well-defined structure, typically described by the coordinates of all its atoms. A solid foundation of this view is provided by the 130 000 structures of proteins and complexes in the Protein Data Bank, PDB (1). However, it is increasingly recognized that many proteins do not obey this rule. Intrinsically disordered proteins (IDPs) or regions (IDRs) are devoid of order in their native unbound state (2–4). Intrinsic disorder is prevalent

*To whom correspondence should be addressed. Tel: +39 49 827 6269; Fax: +39 49 827 6363; Email: silvio.tosatto@unipd.it

†These authors contributed equally to the paper as first authors.

in the human proteome (5), appears to play important signaling and regulatory roles (2) and is frequently involved in disease (6). The discovery of intrinsic disorder and its prevalence and functional importance is transforming the field of molecular biology. As intrinsic disorder is emerging as a general phenomenon, databases are collecting and presenting disorder related data in a systematic manner. MobiDB has been a major contributor by providing consensus predictions and functional annotations for all UniProt proteins, driving the field ahead (7,8). The MobiDB upgrade we present in this paper is essential for several reasons.

There is a rapid advance in the functional understanding of intrinsic disorder. The functional classification of IDPs/IDRs is becoming ever more elaborate, with several newly recognized functional mechanisms (9). For example, the central role of intrinsic disorder in the formation of membraneless organelles, such as nucleoli and stress granules, by liquid-liquid phase separation has been characterized recently (10–13). A wide range of experimental observations on the structure-function relationship of IDPs/IDRs is furthering our understanding of disordered states and of the manners in which they function (14–16). These developments have also played a central role in the recent update of the DisProt database (17), the central repository of experimentally characterized IDPs and IDRs. The re-curated version of this database contains experimental observations of disorder for more than 800 protein entries and a renewed functional ontology schema. The experimental evidence on which it rests has also been significantly augmented to include a broad range of biophysical techniques. DisProt is the basis for most developments in disorder predictors (18,19), and its recent update is a major motivation for a new version of MobiDB.

Additional developments in the field make this release timely. A major source of intrinsic disorder is the identification of residues with missing atomic coordinates in the PDB, which can now be augmented by cryo-electron microscopy (cryo-EM) data. This is having a tremendous impact on structural biology (20,21). Structural descriptions of IDPs and IDRs under physiological conditions have also greatly advanced and are starting to appear in dedicated databases such as IDEAL (22), DIBS (23) and MFIB (24). IDPs and IDRs can perform key roles in molecular recognition by folding upon binding of short linear motifs (SLiMs) covered in the ELM database (25). Generally, the full functional characterization of IDPs and IDRs requires the description not just of their free (disordered) states (26,27), but also of their residual dynamics in the bound states (28). Fuzzy (disordered) complexes can be found in FuzDB (29) and structural ensembles describing the free form (30) in the protein ensemble database (PED (31)). Techniques such as in-cell Nuclear magnetic resonance (NMR) spectroscopy (32,33) and single-molecule fluorescence (34) will soon help study these structures in the physiological state. In reflection of all these developments, we are now launching a significantly updated version of our database, MobiDB 3.0. The new version incorporates additional curated data from specialized databases. Novel annotation features include disorder derived from publicly available NMR chemical shift data (35) and an extended list of predictors. Database

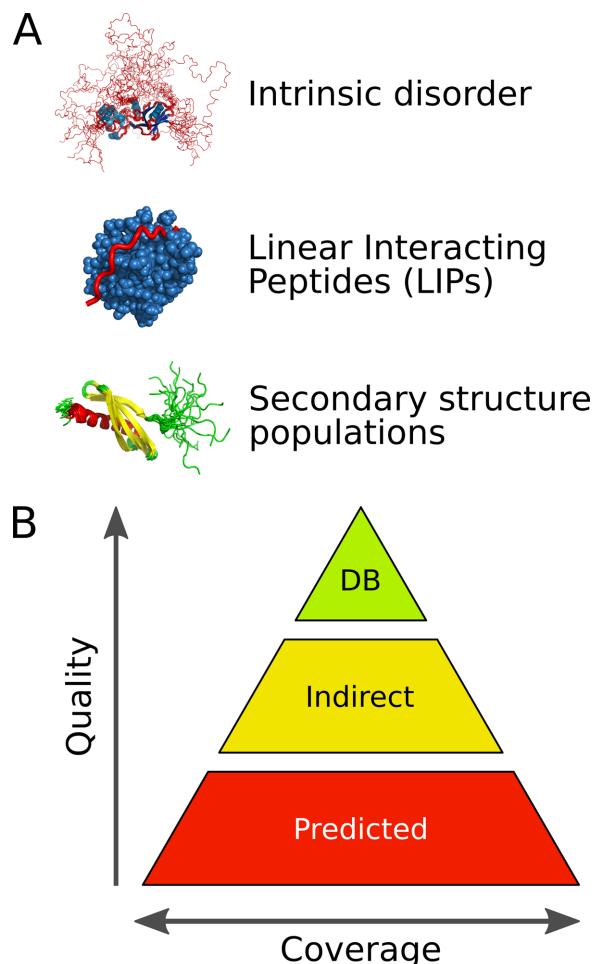


Figure 1. Overview of different annotation data types (A) and levels of accuracy (B) in MobiDB 3.0.

searches are facilitated by an improved search algorithm, pre-calculated data and new sections in the database.

DATABASE DESCRIPTION

MobiDB 3.0 is intended to be a central resource for large-scale intrinsic disorder sequence annotation. This new version is organized by both type of disorder annotation and quality of disorder evidence (Figure 1). Disorder information is grouped in three different sections: disorder, linear interacting peptides (LIPs) and secondary structure populations. The latter represents the conformational heterogeneity of IDPs and IDRs as the ability to populate different secondary structure populations in solution. LIPs are structure fragments that interact with other molecules preserving an elongated structure or folding upon binding. The data in MobiDB is organized hierarchically. The top tier is formed by manually curated data from external databases and represents the highest quality annotations. Annotations derived from experimental data such as X-ray and NMR chemical shifts are indirect but far more abundant. At the bottom, predictions provide disorder annotation at lower confidence than experimental evidence. The main disorder definition in MobiDB is provided by a consensus combin-

Table 1. Overview of databases integrated into MobiDB 3.0

Database	Type	Comment	URL
UniProt	Curated	Disorder	http://www.uniprot.org/
DisProt	Curated	Disorder	http://www.disprot.org/
FuzDB	Curated	Disorder	http://protdyn-database.org/
ELM	Curated	LIPs	http://elm.eu.org/
MFIB	Curated	LIPs	http://mfib.enzim.ttk.mta.hu/
DIBS	Curated	LIPs	http://dibs.enzim.ttk.mta.hu/
IDEAL	Curated	LIPs	http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/
Gene3D	Curated/Prediction	Structure	http://gene3d.biochem.ucl.ac.uk/
Pfam	Curated/Prediction	Domains/Families	http://pfam.xfam.org/
CoDNaS	Indirect	Conformational diversity	http://ufq.unq.edu.ar/codnas/

ing all available sources prioritizing curated and indirect evidences over predictions in analogy to the previous version (8). In the following, we will describe the main recent improvements since the previous release. The database schema, web interface and server have been completely redesigned and the underlying technology updated. The feature viewer showing sequence annotations is now fully dynamic and allows the generation of high quality images for publications with a click. Where available, MobiDB annotation is projected directly onto the structure and shown in a new 3D viewer. The look and feel and organization of the page and loading latency were also improved.

New curated data

MobiDB 3.0 includes different sources of manually curated disorder annotations (Table 1). These annotations fall into two categories: disorder and LIPs. LIPs are binding regions presumed or demonstrated to be intrinsically disordered that fold upon binding. These come under different names such as SLiMs or MoREs (molecular recognition elements) in the literature. The IDEAL database calls them ‘protean’ segments (ProS) (22). MobiDB includes both ‘verified’ and ‘possible’ ProS from IDEAL, where verified means disorder has been experimentally observed in the isolated molecule. The Database of Disordered Binding Sites (DIBS, (23)) collects cases where a disordered region folds upon binding with a globular domain and the Mutual Folding Induced by Binding (MFIB, (24)) database includes disordered regions that fold upon binding with another disordered region. ELM (25) provides SLiM annotations involved in binding and post-translational modifications. General disorder annotation, i.e. without any knowledge about transition driven by interactions, is collected from UniProtKB (36), DisProt (17) and FuzDB (29). UniProtKB provides manually curated disorder annotations under the region field in the features section. FuzDB collects cases of fuzzy complexes, where conformational diversity has a functional role in the regulation and formation of protein complexes or higher-order assemblies. DisProt has been recently revamped and MobiDB now propagates DisProt disordered regions by homology transfer. Regions homologous to experimentally characterized IDRs are mapped across homologs obtained from GeneTree alignments (37). Regions with identity and similarity >80% and an alignment of at least 10 residues are retained as homologous IDRs. Gene3D (38) contributes complementary order annotation to the MobiDB consensus calculation, while Pfam (39) is used

to highlight protein domains. Lastly MobiDB also maps CoDNaS information to highlight conformational diversity in globular regions. CoDNaS measures structural differences among conformers of the same protein (40).

New indirect annotations

Previous releases of MobiDB provided indirect annotations from the PDB through missing residues in X-ray structures and mobile regions from NMR ensembles as calculated with the Mobi software (41). In the current release, this annotation has been complemented with additional indirect information from experimental data in the PDB and chemical shifts from the Biological Magnetic Resonance Data Bank (BMRB) (35). The new Mobi 2.0 software (42) is used to extract LIPs and disorder information from PDB files. Disorder is encoded by three different parameters: high-temperature, missing and mobile residues. High-temperature residues are detected from B-factor regions for X-ray and cryo-EM structures using a threshold proportional to the resolution of the structure. Missing residues are available for all experimental types and obtained comparing the experimental sequence (i.e. PDB SEQRES entries) with the observed residues in the structure (i.e. PDB ATOM entries). A mobility estimate is provided for NMR structures by comparing C_{α} displacement and local conformations in different aligned models (41). LIPs are identified by comparing intra- versus inter-chain contacts calculated using RING (43). The closest atoms between two residues are used to establish a contact which is then distinguished by chemical type (e.g. hydrogen bond, salt bridge, π - π stack). LIPs are identified as any region where the number of inter-chain contacts is at least two times the number of intra-chain contacts (42).

MobiDB 3.0 better exploits the power of NMR spectroscopy to probe the structural properties of proteins in solution, as well as their dynamics on a wide range of timescales (44). Chemical shifts quantify structural fluctuations of proteins up to the millisecond timescale and are relatively easy to measure. Using chemical shifts to obtain information about the statistical populations of different structural motifs allows for a more comprehensive structural description of proteins in solution than static structures or binary definitions such as ‘ordered’ and ‘disordered’ (44). MobiDB 3.0 uses chemical shift data from BMRB directly as reported without applying chemical shift re-referencing methods. The software packages δ 2D (45) and Random Coil Index (RCI) (46) are used to calculate

Table 2. Overview of tools used into MobiDB 3.0

Tool	Type	Description
Mobi 2.0	Indirect	Missing, high-temperature and mobile residues from PDB structures
RING 2.0	Indirect	Residue interactions from PDB structures, used to define LIPs
RCI	Indirect	Random coil index from BMRB chemical shifts
δ 2D	Indirect	Secondary structure populations from BMRB chemical shifts
DynaMine	Prediction	Random coil index
FeSS	Prediction	Secondary structure prediction component of FIELDS
MobiDB-lite	Prediction	Long disorder based on consensus
DisEMBL	Prediction	Disorder. Versions: 465, Hot-loops
ESpritz	Prediction	Disorder. Versions: DisProt, NMR, X-ray
IUPred	Prediction	Disorder. Versions: Short, Long
VSL2b	Prediction	Disorder
GlobPlot	Prediction	Globular regions, used as opposite of disorder
SEG	Prediction	Low complexity
Pfilt	Prediction	Low complexity

two-dimensional ensembles in terms of secondary structure populations (44) and backbone flexibility. Secondary structure populations are calculated only for residues with at least three atom types with measured chemical shifts, as using fewer chemical shifts results in less accurate mappings of the populations (45). MobiDB 3.0 reports the experimental conditions at which the chemical shifts were measured as the structural properties of some proteins can change drastically between different conditions (e.g. binding partners, lipids, pH) and these can help elucidate protein function (44). When an entry in MobiDB is associated to multiple chemical shifts, an overview of the predominant secondary structure conformation is provided in a consensus track. This can be expanded in the feature viewer to show experimental conditions such as pH, temperature, binding partners, molecular state, sample information and the title of the corresponding BMRB entries.

New predictors

MobiDB 3.0 includes the same set of disorder predictors used in the previous release: ESpritz (47), IUPred (48), DisEMBL (49) and VSL2b (50). Consensus generation is handled by MobiDB-lite (51), which uses a stronger majority threshold and enforces at least 20 consecutive disordered residues to provide highly specific predictions. This is completed by a continuous representation of the fraction of methods predicting disorder for each residue. DynaMine (52), Anchor (53) and FeSS (54) are now also part of the annotation pipeline. DynaMine (52) predicts backbone flexibility where 1.0 means complete order (stable conformation, i.e. rigid) and 0 means fully random bond vector movement (highly dynamic, i.e. flexible). Anchor predicts binding regions located in disordered proteins, providing LIP annotations for all proteins in the database. FeSS is a component of the FIELDS method (54) providing three-state (helix, sheet, coil) secondary structure propensity. FeSS prediction confidence can be interpreted similarly to the dynamic behavior measured by δ 2D in chemical shifts, i.e. a propensity to remain in a given state of secondary structure. The complete list of tools is available in Table 2.

The MobiDB-lite version used in MobiDB 3.0 has been extended to provide a structural characterization of the disorder regions that can help interpret their functional role. It distinguishes different types of disordered regions by mea-

suring the fraction of charged residues and net charge according to a previous classification (55). The different types are: positive polyelectrolytes (D_PPE), negative polyelectrolytes (D_NPE), polyampholites (D_PA) and weak polyampholites (D_WC). A statistical analysis of the different disorder flavours was already performed on the MobiDB 2.0 data (8).

Usage and annotated data

MobiDB now contains all sequences from UniParc, the most comprehensive non-redundant set of protein sequences. Entries are identified also by UniProtKB (36) accession numbers and can be retrieved by organism, taxonomy and other identifiers provided by UniProtKB. Prediction results are combined with indirect disorder evidences derived from PDB data (using Mobi 2) and data extracted from manually curated third party databases. MobiDB annotations are used by DisProt (17) curators to guide the annotation of disorder regions. MobiDB data is made available to the public via a web interface allowing extensive search functionalities and RESTful services for programmatic access. MobiDB 3.0 includes a pre-calculated consensus for all entries allowing real-time statistics and download of entire datasets in different formats directly from the web interface. The new database schema makes it possible to perform complex search queries and to generate custom datasets, for example retrieving all entries with manually curated annotations. The MobiDB update has been automated and is scheduled every three months due to the high computational cost of generating predictions for new sequences.

DISCUSSION

MobiDB 3.0 improves on previous releases by adding descriptions of conformational diversity and disorder-related functions, both in terms of experimental data and predictions. A particular field where it may have a significant impact is the establishment of a long-awaited disorder sequence-function relationship schema. The most reliable proxy to this goal is to assess the function of a protein by homology transfer, i.e. transferring functional annotation based on sequence similarity. Aligning IDR sequences is complicated by their high evolutionary variabil-

ity and often limits evolutionary analysis (56,57). New functional terms introduced in the DisProt update (17), represent non-canonical functions probably only characteristic of IDPs which are not incorporated in functional classification schemes such as GO (58). A large-scale analysis of IDP functional annotations will be necessary to find adequate boundaries for transferring IDP functions by homology. As sufficient data is now available in MobiDB 3.0, we expect a rapid advance in the field of sequence-function correlations of IDPs.

For proteins with sufficient NMR data, MobiDB now features quantitative annotations incorporating structure and equilibrium dynamics in a unified framework. These large-scale quantitative annotations will help understand the biological role of order and disorder, and serve as a basis to construct predictive models. As NMR measurements of proteins in their native complex environments, such as inside living cells, are becoming more common (59), we will be able to address fundamental biological questions with greater physiological relevance (60).

MobiDB is widely used by scientific community and by third party services such as DisProt (17) and ProViz (61). It has recently joined the InterPro consortium to provide disorder annotation alongside protein domains and families (62). MobiDB is becoming a thematic hub for IDPs inside the European sustainable bioinformatics infrastructure (ELIXIR) and we encourage contributions of novel predictors and datasets. Future work will focus on including IDP annotations into core data resources such as UniProt.

ACKNOWLEDGEMENTS

We acknowledge ELIXIR-IIB (elixir-italy.org), the Italian Node of the European ELIXIR infrastructure (elixir-europe.org), for supporting the development and maintenance of MobiDB.

FUNDING

COST Action [BM1405 NGP-net]; FIRC Research Fellowship [16621 to D.P.]; Hungarian Academy of Sciences ‘Lendület’ Grant [LP201418/2016 to Z.D.]; Hungarian Scientific Research Fund [OTKA K 108798 to Z.D.]; Hungarian Academy of Sciences Postdoctoral Fellowship [to B.M.]; Agencia de Ciencia y Tecnología [PICT-2014–3430 to G.P.]; Universidad Nacional de Quilmes [1402/15]; OTKA Grant [PD-OTKA 108772 to E.S.]; Research Foundation Flanders (FWO) Odysseus Grant [G.0029.12 to P.T.]; FWO project [G032816N to W.F.V.]; AIRC IG Grant [17753 to S.T., in part]. Funding for open access charge: COST Action [BM1405].

Conflict of interest statement. None declared.

REFERENCES

- Burley, S.K., Berman, H.M., Kleywegt, G.J., Markley, J.L., Nakamura, H. and Velankar, S. (2017) Protein Data Bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol. Clifton NJ*, **1607**, 627–641.
- Wright, P.E. and Dyson, H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
- Habchi, J., Tompa, P., Longhi, S. and Uversky, V.N. (2014) Introducing protein intrinsic disorder. *Chem. Rev.*, **114**, 6561–6588.

- Tompa, P. (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.*, **37**, 509–516.
- Panca, R. and Tompa, P. (2012) Structural disorder in eukaryotes. *PLoS One*, **7**, e34687.
- Uversky, V.N., Oldfield, C.J. and Dunker, A.K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.*, **37**, 215–246.
- Di Domenico, T., Walsh, I., Martin, A.J.M. and Tosatto, S.C.E. (2012) MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics*, **28**, 2080–2081.
- Potenza, E., Di Domenico, T., Walsh, I. and Tosatto, S.C.E. (2015) MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.*, **43**, D315–D320.
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T. *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, **114**, 6589–6631.
- Hyman, A.A., Weber, C.A. and Jülicher, F. (2014) Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.*, **30**, 39–58.
- Shorter, J. (2016) Membraneless organelles: phasing in and out. *Nat. Chem.*, **8**, 528–530.
- Toretsky, J.A. and Wright, P.E. (2014) Assemblages: functional units formed by cellular phase separation. *J. Cell Biol.*, **206**, 579–588.
- Brangwynne, C.P., Tompa, P. and Pappu, R.V. (2015) Polymer physics of intracellular phase transitions. *Nat. Phys.*, **11**, 899–904.
- Berlow, R.B., Dyson, H.J. and Wright, P.E. (2017) Hypersensitive termination of the hypoxic response by a disordered protein switch. *Nature*, **543**, 447–451.
- Mylona, A., Theillet, F.-X., Foster, C., Cheng, T.M., Miralles, F., Bates, P.A., Selenko, P. and Treisman, R. (2016) Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation. *Science*, **354**, 233–237.
- Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E. and Forman-Kay, J.D. (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*, **519**, 106–109.
- Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C.J., Aspromonte, M.C., Davey, N.E., Davidović, R., Dosztányi, Z. *et al.* (2017) DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.*, **45**, D1123–D1124.
- Walsh, I., Giollo, M., Di Domenico, T., Ferrari, C., Zimmermann, O. and Tosatto, S.C.E. (2015) Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics*, **31**, 201–208.
- Necci, M., Piovesan, D., Dosztányi, Z., Tompa, P. and Tosatto, S.C.E. (2017) A comprehensive assessment of long intrinsic protein disorder from the DisProt database. *Bioinformatics*, doi:10.1093/bioinformatics/btx590.
- Callaway, E. (2015) The revolution will not be crystallized: a new method sweeps through structural biology. *Nature*, **525**, 172–174.
- Cheng, Y. (2015) Single-Particle Cryo-EM at Crystallographic Resolution. *Cell*, **161**, 450–457.
- Fukuchi, S., Amemiya, T., Sakamoto, S., Nobe, Y., Hosoda, K., Kado, Y., Murakami, S.D., Koike, R., Hiroaki, H. and Ota, M. (2014) IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. *Nucleic Acids Res.*, **42**, D320–D325.
- Schad, E., Fichó, E., Panca, R., Simon, I., Dosztányi, Z. and Mészáros, B. (2017) DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics*, doi:10.1093/bioinformatics/btx640.
- Fichó, E., Reményi, I., Simon, I. and Mészáros, B. (2017) MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinformatics*, doi:10.1093/bioinformatics/btx486.
- Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A. *et al.* (2016) ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.*, **44**, D294–D300.
- Arai, M., Sugase, K., Dyson, H.J. and Wright, P.E. (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 9614–9619.
- Borchers, W., Theillet, F.-X., Katzer, A., Finzel, A., Mishall, K.M., Powell, A.T., Wu, H., Manieri, W., Dieterich, C., Selenko, P. *et al.* (2014)

- Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.*, **10**, 1000–1002.
28. Tompa,P. and Fuxreiter,M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.*, **33**, 2–8.
 29. Miskei,M., Antal,C. and Fuxreiter,M. (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.
 30. Tompa,P. and Varadi,M. (2014) Predicting the predictive power of IDP ensembles. *Structure*, **22**, 177–178.
 31. Varadi,M., Kosol,S., Lebrun,P., Valentini,E., Blackledge,M., Dunker,A.K., Felli,I.C., Forman-Kay,J.D., Kriwacki,R.W., Pierattelli,R. *et al.* (2014) pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.*, **42**, D326–D335.
 32. Theillet,F.-X., Binolfi,A., Bekei,B., Martorana,A., Rose,H.M., Stuver,M., Verzini,S., Lorenz,D., van Rossum,M., Goldfarb,D. *et al.* (2016) Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature*, **530**, 45–50.
 33. Felli,I.C., Gonnelli,L. and Pierattelli,R. (2014) In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins. *Nat. Protoc.*, **9**, 2005–2016.
 34. Sakon,J.J. and Wenginger,K.R. (2010) Detecting the conformation of individual proteins in live cells. *Nat. Methods*, **7**, 203–205.
 35. Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
 36. Consortium,The UniProt (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
 37. Vilella,A.J., Severin,J., Ureta-Vidal,A., Heng,L., Durbin,R. and Birney,E. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.*, **19**, 327–335.
 38. Lees,J., Yeats,C., Perkins,J., Sillitoe,I., Rentzsch,R., Dessailly,B.H. and Orengo,C. (2012) Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Res.*, **40**, D465–D471.
 39. Finn,R.D., Coghill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
 40. Monzon,A.M., Rohr,C.O., Fornasari,M.S. and Parisi,G. (2016) CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. *Database (Oxford)*, **2016**, baw038.
 41. Martin,A.J.M., Walsh,I. and Tosatto,S.C.E. (2010) MOBI: a web server to define and visualize structural mobility in NMR protein ensembles. *Bioinformatics*, **26**, 2916–2917.
 42. Piovesan,D. and Tosatto,S.C.E. (2017) Mobi 2.0: an improved method to define intrinsic disorder, mobility and linear binding regions in protein structures. *Bioinformatics*, doi:10.1093/bioinformatics/btx592.
 43. Piovesan,D., Minervini,G. and Tosatto,S.C.E. (2016) The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.*, **44**, W367–W374.
 44. Sormanni,P., Piovesan,D., Heller,G.T., Bonomi,M., Kucic,P., Camilloni,C., Fuxreiter,M., Dosztányi,Z., Pappu,R.V., Babu,M.M. *et al.* (2017) Simultaneous quantification of protein order and disorder. *Nat. Chem. Biol.*, **13**, 339–342.
 45. Camilloni,C., De Simone,A., Vranken,W.F. and Vendruscolo,M. (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, **51**, 2224–2231.
 46. Berjanskii,M.V. and Wishart,D.S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**, 14970–14971.
 47. Walsh,I., Martin,A.J.M., Di Domenico,T. and Tosatto,S.C.E. (2012) ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics*, **28**, 503–509.
 48. Dosztányi,Z., Csizmok,V., Tompa,P. and Simon,I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
 49. Linding,R., Jensen,L.J., Diella,F., Bork,P., Gibson,T.J. and Russell,R.B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure*, **11**, 1453–1459.
 50. Peng,K., Radivojac,P., Vucetic,S., Dunker,A.K. and Obradovic,Z. (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, **7**, 208.
 51. Necci,M., Piovesan,D., Dosztányi,Z. and Tosatto,S.C.E. (2017) MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*, **33**, 1402–1404.
 52. Cilia,E., Pancsa,R., Tompa,P., Lenaerts,T. and Vranken,W.F. (2013) From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.*, **4**, 2741.
 53. Mészáros,B., Simon,I. and Dosztányi,Z. (2009) Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.*, **5**, e1000376.
 54. Piovesan,D., Walsh,I., Minervini,G. and Tosatto,S.C.E. (2017) FIELDS: fast estimator of latent local structure. *Bioinformatics*, **33**, 1889–1891.
 55. Das,R.K. and Pappu,R.V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 13392–13397.
 56. Brown,C.J., Johnson,A.K., Dunker,A.K. and Daughdrill,G.W. (2011) Evolution and disorder. *Curr. Opin. Struct. Biol.*, **21**, 441–446.
 57. Csizmok,V., Felli,I.C., Tompa,P., Banci,L. and Bertini,I. (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. *J. Am. Chem. Soc.*, **130**, 16873–16879.
 58. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
 59. Plitzko,J.M., Schuler,B. and Selenko,P. (2017) Structural Biology outside the box-inside the cell. *Curr. Opin. Struct. Biol.*, **46**, 110–121.
 60. Gierasch,L.M. and Gershenson,A. (2009) Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nat. Chem. Biol.*, **5**, 774–777.
 61. Jehl,P., Manguy,J., Shields,D.C., Higgins,D.G. and Davey,N.E. (2016) ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucleic Acids Res.*, **44**, W11–W15.
 62. Finn,R.D., Attwood,T.K., Babbitt,P.C., Bateman,A., Bork,P., Bridge,A.J., Chang,H.-Y., Dosztányi,Z., El-Gebali,S., Fraser,M. *et al.* (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.*, **45**, D190–D199.