

MobiDB: intrinsically disordered proteins in 2021

Damiano Piovesan¹, Marco Necci¹, Nahuel Escobedo², Alexander Miguel Monzon¹,
Andr as Hatos¹, Ivan Mi eti ¹, Federica Quaglia¹, Lisanna Paladin¹,
Pathmanaban Ramasamy^{3,4,5,6,7}, Zsuzsanna Doszt nyi⁸, Wim F. Vranken^{3,4,5},
Norman E. Davey⁹, Gustavo Parisi², Monika Fuxreiter¹ and Silvio C.E. Tosatto^{1,*}

¹Dept. of Biomedical Sciences, University of Padua, Via Ugo Bassi 58/B, Padua 35121, Italy, ²Dept. of Science and Technology, Universidad Nacional de Quilmes, Buenos Aires, Argentina, ³Interuniversity Institute of Bioinformatics in Brussels, ULB/VUB, Triomflaan, BC building, 6th floor, CP 263, 1050 Brussels, Belgium, ⁴Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium, ⁵Centre for Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium, ⁶VIB-UGent Center for Medical Biotechnology, VIB, Ghent 9000, Belgium, ⁷Department of Biomolecular Medicine, Faculty of Health Sciences and Medicine, Ghent University, Ghent 9000, Belgium, ⁸Dept. of Biochemistry, ELTE E tv s Lor nd University, Budapest, Hungary and ⁹Division of Cancer Biology, The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK

Received September 15, 2020; Revised October 16, 2020; Editorial Decision October 19, 2020; Accepted November 19, 2020

ABSTRACT

The MobiDB database (URL: <https://mobidb.org/>) provides predictions and annotations for intrinsically disordered proteins. Here, we report recent developments implemented in MobiDB version 4, regarding the database format, with novel types of annotations and an improved update process. The new website includes a re-designed user interface, a more effective search engine and advanced API for programmatic access. The new database schema gives more flexibility for the users, as well as simplifying the maintenance and updates. In addition, the new entry page provides more visualisation tools including customizable feature viewer and graphs of the residue contact maps. MobiDB v4 annotates the binding modes of disordered proteins, whether they undergo disorder-to-order transitions or remain disordered in the bound state. In addition, disordered regions undergoing liquid-liquid phase separation or post-translational modifications are defined. The integrated information is presented in a simplified interface, which enables faster searches and allows large customized datasets to be downloaded in TSV, Fasta or JSON formats. An alternative advanced interface allows users to drill deeper into features of interest. A new statistics page provides information at database and proteome levels. The new MobiDB version presents state-of-the-art knowledge on disordered proteins and improves data accessibility for both computational and experimental users.

INTRODUCTION

Intrinsically disordered regions (IDRs) of proteins do not adopt a highly populated structure in isolation, but sample a wide range of conformations. IDR-mediated interactions are implicated in a variety of cellular processes from signal transduction and liquid-liquid phase transition (1–5). IDRs are subjected to extensive pre- and post-translational regulation to modulate protein function in response to cellular stimuli (6–9). Many functions of IDRs, such as entropic springs, flexible linkers or spacers are directly associated with their structural attributes (10,11).

Proteins containing intrinsically disordered regions are present in a considerable fraction of the proteome of eukaryotic organisms (e.g. 43.6% in humans based on MobiDB-lite predictions). In the human proteome, for example, IDRs are highly enriched in interaction interfaces and post-translational modification sites, which serve as regulatory switches of a variety of biochemical pathways (12). Short linear motifs in the IDRs of viruses lead to functional promiscuity and enable compact viral proteomes to extensively rewire their host cells (13). Given the lack of a requirement for a stable globular fold, interfaces in IDRs also appear to have higher evolutionary plasticity allowing them to proliferate in a proteome by ex nihilo evolution. As a result, IDR interfaces are possibly more widespread than ordered interfaces (14). Despite their central roles in key cellular regulatory processes, only a small fraction of disordered regions have been experimentally characterised (15,16). While difficulties in protein expression, purification and structural characterisation hamper experimental characterization, assigning functional modules to dynamic conformational ensembles presents a technical problem for a database.

*To whom correspondence should be addressed. Tel: +39 049 827 6269; Email: silvio.tosatto@unipd.it

Manually curated databases, such as DisProt (17,18) and IDEAL (19), collect annotations from the literature and standardise that information. These databases assign the position of IDRs in the sequence, and aim to assemble disorder-related functional annotations. Other databases focus on functional features: DIBS (15) and MFIB (20) collect regions undergoing folding upon binding, and ELM (16) annotates linear motifs. IDRs forming fuzzy complexes are collected in the FuzDB database (21), which also aims to establish links between structural properties and function. Proteins reported to undergo liquid-liquid phase separation (LLPS) typically contain IDRs, which are available in PhasePro (22), PhaSepDB (23) and LLPSDB (24). Altogether, these specialized datasets exhibit a limited overlap, which presents a bottleneck for comprehensive characterisation of the IDRs contained therein (25).

Large-scale identification of IDRs often depends on primary deposition databases of structural data, like the PDB (26), which can be used to indirectly derive disorder information from missing and mobile residues. Although annotation based on this information is less reliable compared to curated resources, it provides a larger set of examples (27,28). A complementary source of information is provided by sequence based predictors, which typically exploit amino acid biases in regions of proteins to identify IDRs (29). Deriving information from these methods requires a critical assessment of the different techniques, which has been implemented in independent blind tests and in the Critical Assessment of Intrinsic protein Disorder (CAID). These efforts, in particular CAID, enable integrating large-scale predicted information on disordered regions into other core data resources such as InterPro (30), UniProtKB (31) and PDBe (26) via MobiDB-lite (32).

In contrast to structural data, functional annotations related to disordered regions are underrepresented in public databases and ontologies (33). This presents a bottleneck for large-scale functional assignment of disordered regions. Functional predictions are in essence limited to regions that fold upon binding, e.g. ANCHOR (34), DISOPRED3 (35) and MoRFChiBi (66). Recently, prediction of regions that remain disordered upon binding (36) and of regions that exhibit fuzzy binding (37) became available. A new generation of disorder function predictors is also being developed (36) based on the Critical Assessment of Function Annotation challenge (CAFA) (37) and DisProt-based Disorder Ontology (17).

The objective of the new MobiDB (version 4) is to provide stable and sustainable access to data on disordered regions and proteins. As an ELIXIR resource, this is in line with the overall ELIXIR mission and complies with the goals described in the ELIXIR IDP community whitepaper (38). Sustainability has been improved primarily by optimizing the database schema, and by improving the update process and simplifying the data representation formats. The new version incorporates additional curated data from specialized databases and presents novel annotation features including the binding modes derived from processing PDB data as well as an extended list of predictors. It is now possible to download search results and entire genomes with the possibility to specify the format and selected features. A novel statistics page allows the users to compare

annotation coverage of different types of annotations and compare the annotation content at the proteome level for all reference proteomes.

PROGRESS AND NEW FEATURES

Database content

MobiDB data serves both experimental scientists, who are interested in different aspects of disordered protein regions as well as bioinformaticians, who develop softwares for analysis and prediction of disordered regions. In the new version of the database, several new data resources and features were added to increase the coverage and usability of information about protein disorder. Figure 1 gives an overview of the annotations and services provided by MobiDB. All curated disordered regions are mapped across homologs obtained from GeneTree alignments (39) with >80% of sequence similarity and alignment lengths of 10 residues. This expands the curated set of proteins ten fold. Structural and functional properties of disordered regions are based on third party databases and a set of prediction methods.

The full list of software and resources used in MobiDB is provided in Tables 1 and 2. The different types of annotations and predictions are assembled to provide the user a comprehensive view of properties of disordered regions at the residue level. As compared to the previous releases, MobiDB v4 provides annotations on novel functional aspects of disordered proteins, such as binding modes (36), involvement in post-translational modifications and phase separation (3,4,7).

Disorder predictions are provided via the MobiDB-lite software (32) over the entire UniProtKB set of protein sequences. In addition, MobiDB calculates annotations from PDB structures, based on manually curated annotations from specific third party databases and propagating manual curation by homology. MobiDB-lite is also integrated for example into InterProScan (30) which propagates its predictions onto several other EBI resources like UniProtKB, InterPro and PDBe.

Feature classification in MobiDB v4

Different levels of reliability and different features are reported as different and independent annotations. Each MobiDB annotation is identified by an 'evidence-feature-source' triplet which uniquely identifies one level of quality, type and source of annotation. Possible values for 'evidence' are: (i) 'curated', the annotation comes from a manually curated database; (ii) 'derived', the annotation is inferred automatically processing primary data, for example PDB structures are processed to detect missing, mobile and lip residues, (iii) 'homology', annotations are obtained aligning curated data and (iv) 'prediction', annotations are automatically extracted from the protein sequence. The 'feature' element of the triplet represents the type (or flavour) of the annotation, for example 'disorder', 'missing_residues', 'binding_mode', 'lip', etc. The 'source' element represents the annotation source which can be the name of a software, a third party database or the type of consensus. Possible values are: 'mobidb-lite', 'disprot', 'merged', 'th_90', etc. The

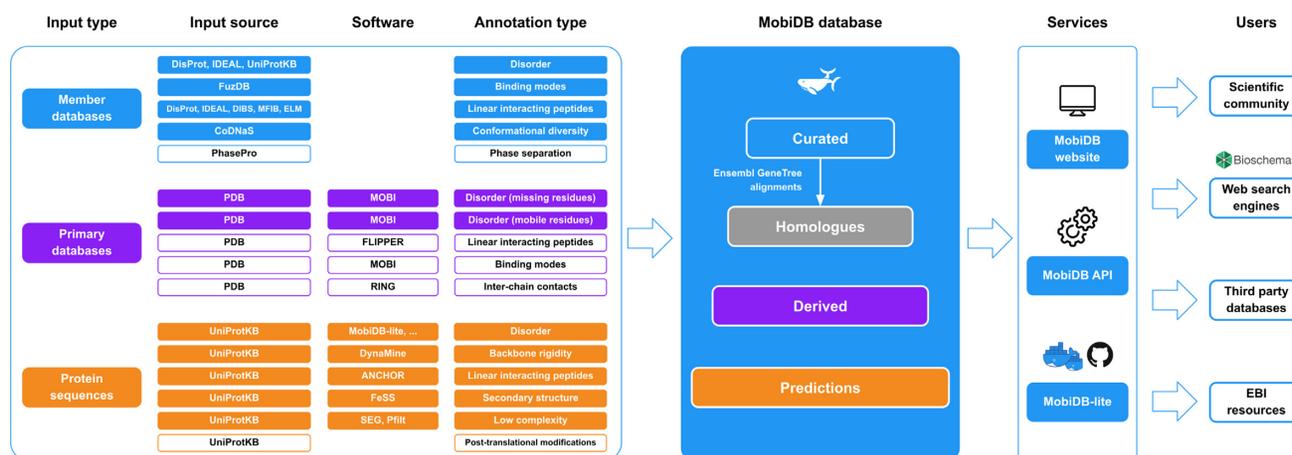


Figure 1. MobiDB data and services. The MobiDB pipeline (left) includes the input source, output features and software used to generate MobiDB data (center). Different background colors indicate different levels of annotation quality. White background in the pipeline (left) indicates novel features, databases and software integrated into MobiDB v4. MobiDB data can be accessed through a website and an API (right). MobiDB predictions can be generated using the MobiDB-lite software which is available both as a Docker container and as a Python package from GitHub. MobiDB web pages are decorated with BioSchemas profiles which allows external search engines to retrieve disorder annotations exploiting structured data.

Table 1. Software programs integrated into MobiDB 4. Pfam and Gene3D (*) annotations are retrieved from the InterPro database, all other software are executed locally in the MobiDB server

Method	Input	Features	Reference
Mobi	Structure	Binding modes, missing residues, mobile residues, high temperature residues	(43)
FLIPPER	Structure	Linear interacting peptides	Unpublished
RING	Structure	Residue Interaction Network	(40)
MobiDB-lite	Sequence	Disorder, cysteine rich, proline rich, polar, negative polyelectrolyte, positive polyelectrolyte, polyampholyte	(32)
ESpritz-DisProt	Sequence	Disorder	(54)
ESpritz-NMR	Sequence	Disorder	(54)
ESpritz-Xray	Sequence	Disorder	(54)
IUPred-Long	Sequence	Disorder	(55)
IUPred-Sort	Sequence	Disorder	(55)
VSL2b	Sequence	Disorder	(56)
DisEMBL-465	Sequence	Disorder	(57)
DisEMBL-HotLoops	Sequence	Disorder	(57)
GlobPlot	Sequence	Disorder	(58)
JRONN	Sequence	Disorder	(59)
ANCHOR	Sequence	Linear interacting peptides	(42)
FeSS	Sequence	Secondary structure	(60)
DynaMine	Sequence	Backbone rigidity	(61)
Pflit	Sequence	Low complexity	(62)
SEG	Sequence	Low complexity	(63)
Gene3D (*)	Sequence	Conserved domains	(64)
Pfam (*)	Sequence	Conserved domains	(65)

full list and explanation of all available triplets is provided on the about page of the website.

Novel functional annotations in MobiDB v4

The main difference in MobiDB v4 as compared to the previous releases of MobiDB is the addition of novel functional aspects related to disordered regions, which are listed below and with a white background on the first block of Figure 1. These contribute to establishing links between conformational ensembles and their biological roles.

Binding modes

Binding modes of disordered regions refer to the conformational transitions of IDRs upon interacting with spe-

cific partners. IDRs can undergo disorder-to-order transitions and fold upon the template, or remain disordered (disorder-to-disorder transition) in a partner-bound form (36). Some IDRs exhibit both behaviors, and exhibit context-dependent binding with different partners or cellular conditions (37). These binding modes were assigned from collecting all bound-state experimental evidence from PDB, as defined in (36).

MobiDB implementation identifies inter-chain binding regions and the mono-/multimeric state using the RING software (40). RING generates residue level contact-networks by identifying all types of non-covalent interactions at atomic level in a protein structure (PDB). The residue interaction network generated by RING is combined with data on disorder derived from missing

Table 2. Third party databases integrated into MobiDB 4

Database	Features	URL
CoDNAs	Conformational diversity	http://ufq.unq.edu.ar/codnas/
DIBS	Linear interacting peptides	http://dibs.enzim.ttk.mta.hu/
DisProt	Disorder, linear interacting peptides	https://www.disprot.org/
ELM	Linear interacting peptides	http://elm.eu.org/
FuzDB	Binding modes	http://protdyn-database.org/
IDEAL	Disorder, linear interacting peptides	https://www.ideal-db.org/
MFIB	Linear interacting peptides	http://mfib.enzim.ttk.mta.hu/
PDBe	Protein structures	https://www.ebi.ac.uk/pdbe/
PhasePro	Phase separation	https://phasepro.elte.hu/
UniProtKB	Disorder, transmembrane, coiled coil, signal peptide	https://www.uniprot.org/

residues. Comparison of disordered residues in free and bound form of the protein is used to define the different binding modes: disorder-to-order, disorder-to-disorder and context-dependent binding.

Linear interacting peptides (LIPs)

Interactions of disordered regions are realized via short peptide motifs (41). These linear motifs can be collected from the sequence within the Eukaryotic Linear Motif database (16) or predicted using automatic methods such as ANCHOR (42). MobiDB provides a different approach, by exploiting a new software, Fast Linear Interacting Peptides Predictor (FLIPPER), which is able to detect linear interacting peptides (LIPs) from protein structures (PDB). FLIPPER is a random forest classifier which exploits geometrical and physicochemical properties of linear interacting peptides. These properties include linearity, solvent accessibility difference in the bound and unbound state and the ratio of inter- vs. intra-chain contacts. The algorithm has been trained on the same examples used by ANCHOR and validated against DIBS and PixelDB examples. The software is a reimplementation of a published component of the previous MobiDB pipeline (43) and it is freely available from GitHub (<https://github.com/BioComputingUP/FLIPPER>). After processing the entire PDB, FLIPPER recognized about 6800 different proteins with at least one LIP. Approximately 4.5% of the residues in these proteins are contained in LIPs (see MobiDB statistics) which provide a bona fide dataset for the implementation of novel sequence based predictors and statistical/functional analyses.

Post-translational modifications

Posttranslational modifications of disordered regions serve as regulatory points of many biochemical pathways (44). Therefore annotating PTM sites is crucial for functional annotation of disordered regions. Data on all types of PTMs are derived from UniProtKB which propagates a small set of well annotated modifications by sequence similarity. Additional phosphorylation sites are provided from the Scop3P database. The Scop3P sites are extracted by large scale re-processing of 36 projects from the PRIDE database (45) with the ionbot approach (<https://ionbot.cloud/>).

Liquid-liquid phase separation

Regions forming dynamic liquid droplets are a new facet of disordered proteins (46). Recent data indicates that weak

interactions involving disordered regions often govern protein phase separation (47). Although the general biological roles of phase separation remain to be elucidated, systematic annotation of regions undergoing LLPS will contribute to elucidating the underlying sequence-codes.

Regions associated with phase separation processes in MobiDB are derived from the PhasePro database (22). This dataset not only assembles proteins, which were reported to phase separate *in vitro* or *in vivo*, but also specifies those regions, which mediate this process.

Data generation pipeline and updates

From the technical point of view, the major change is the new format of the entry document in the database. Multi-class annotations, e.g. secondary structure, have been converted to binary classifications by splitting classes into different fields. The grouping of different features on a common field is deprecated since it encodes an arbitrary interpretation of the type of prediction directly in the entry document. Now all predictions appear at the same level and have the same format. It is now possible to integrate new annotations- without refactoring the entry document. This strategy provides a greater flexibility simplifying updates and database maintenance. This enables us to keep MobiDB annotations up to date with UniProtKB releases, with a minimal delay of a few weeks. For new UniProtKB releases MobiDB generates predictions for new sequences by running MobiDB-lite and other tools, updates protein metadata (UniRef IDs, taxonomy, etc.) and removes obsolete entries.

MobiDB website

The MobiDB website has been completely redesigned in order to improve user experience and satisfy both general use and detailed computational analyses. The graphical user interface retrieves data from a public API which is now well documented and will remain stable for several years.

Entry page

The entry page has been completely renewed as compared to MobiDB v3 and is now divided in two views: simple and advanced. The simple view of the entry page offers an overview of the entry, focusing on a few annotations. Switching to the advanced view, the user accesses the full collection of data associated with an entry, with dedicated

tools for their visualization. In particular, both views revolve around the feature-viewer (48) to organize and dynamically report annotations along the sequence. An interactive network map and a Mol* instance (49) allows an integrated in-depth exploration of residue contacts and structural features.

The feature-viewer combines annotations into a tree of tracks. Parent tracks summarize the information of children tracks by different consensus strategies. Some tracks, e.g. MobiDB-lite tracks are automatically expanded on page load to highlight disordered sub-regions. For each track a squared icon indicates the evidence quality level of the annotation. When a track has children its name gets bold on mouse over and the evidence icon is outlined. On the advanced view, it is possible to customize the feature viewer by hiding specific tracks. By clicking on a region, it is possible to visualize the corresponding annotation in the PDB structure and on the contacts network.

Statistics page

MobiDB v4 has a new statistics page which provides information about all types of annotations for the entire database and for all proteomes separately. The statistics page provides coverage of a given feature both at the database and proteome level. A search bar on the top allows searching for an organism by providing the identifier of the reference proteome or a string which is evaluated against the organism name and kingdom.

Searching and downloading data

The browse page is designed upon the MobiDB server API. MobiDB provides programmatic access to perform a search through a RESTful web service API. A single entry can be retrieved by using UniProtKB identifiers, while database searches can be performed by specifying query fields directly as URL parameters in the HTTP request. Protein metadata (UniProtKB accession, gene, organism, NCBI taxon ID, etc.) can be searched for exact matches, whereas the disorder content and sequence length accept a range of values. In the case of a free text search (free_text parameter) the server uses the EBI Proteins API (50) to retrieve proteins containing the input string in the protein name. In that case the result order is provided by the UniProtKB annotation score (https://www.uniprot.org/help/annotation_score). Both from the browse page and from the API it is possible to download the full search result in different formats, TSV, FASTA and JSON. Entire genomes can be searched and downloaded in minutes.

BioSchemas

To increase MobiDB interoperability and findability, key web pages are decorated with BioSchemas markup (<https://bioschemas.org>) (51) developed within ELIXIR, the European infrastructure for biological data. BioSchemas extends Schema.org (<https://schema.org>) for life sciences making the data contained in the database indexable by search engines and other services such as the Google dataset search tool. MobiDB uses DataCatalog and Dataset profiles in the main page, and DataRecord for the entry pages.

The DataRecord includes Protein and SequenceAnnotation profiles. These profiles were recently implemented in MobiDB which makes the database, together with DisProt, among the first resources exposing sequence region information using BioSchemas.org.

CONCLUSIONS AND FUTURE WORK

The rapidly accumulating experimental data on the structure and function of disordered protein regions highlights the importance of systematic analysis of protein functions related to protein disorder. MobiDBv4 provides a major improvement in this respect as compared to previous releases by adding descriptions of functional aspects of disorder, such detailed information about different binding modes, including both disorder-to-order and fuzzy binding; posttranslational modifications; and regions associated with phase separation processes. Interpretation of residue interaction networks is also facilitated by visualising contact maps in an interactive manner. The database data and programming infrastructure has also been upgraded, facilitating systematic analysis, searches and updates. The new website provides more comprehensive information of each entry in a convenient format for both experimental and computational users.

MobiDB aims to provide gold standard disorder predictions, learning from the results of the Critical Assessment of Intrinsic protein Disorder (CAID, <http://idpcentral.org/caid>) and the Disorder Ontology prediction challenge at the Critical Assessment of protein Function Annotation (CAFA, <https://www.biofunctionprediction.org/cafa>). These initiatives will further contribute to the improvement of the disorder predictions available in MobiDB. MobiDB predictions are also integrated into several EBI resources including InterPro, UniProtKB and PDB. MobiDB is a central resource for the IDP community in ELIXIR.

Thanks to the statistics page, for the first time MobiDB allows to answer questions like for example which is the fraction of disordered residues in a given organism, which is the fraction of residues annotated by the majority of predictors or how many proteins have at least one disordered region.

Further developments of the database aim to improve predictions related to protein disorder. On the one hand, the sequence-based predictions of disordered regions have to be standardised and improved, including the predictions of regions, which can undergo liquid-liquid phase separation. These predictions can be facilitated by evolutionary analysis of disordered regions and vice versa prediction of disordered regions can facilitate evolutionary analysis of proteins with low degrees of sequence similarities (52,53).

The other major line of development is functional predictions for disordered regions, including their roles in interaction networks, conformational switches and organisers of liquid droplets. The novel annotations implemented in MobiDB v4 contribute to these efforts. These functional annotations will also be integrated into CAID and CAFA challenges, and will be related to GO classifications. In the present form, MobiDB provides up-to-date knowledge on disordered proteins and regions, which can be widely used

by both experimental and computational experts as well as by third party services.

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. MobiDB is a service of the ELIXIR IDP community.

FUNDING

European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie [778247]; Italian Ministry of University and Research (MIUR), PRIN [2017483NH8 to S.T.]; P.R. and W.V. acknowledge funding by the Research Foundation Flanders (FWO) [G.0328.16N]; N.D. acknowledges funding by a Cancer Research UK Senior Cancer Research Fellowship [C68484/A28159]; G.P. is a CONICET researcher; Universidad Nacional de Quilmes [PUNQ 1004/11]; ANPCyT [PICT-2014-3430]. Funding for open access charge: IDPfun (Marie Skłodowska-Curie) [778247].

Conflict of interest statement. None declared.

REFERENCES

- Ivarsson, Y. and Jemth, P. (2019) Affinity and specificity of motif-based protein-protein interactions. *Curr. Opin. Struct. Biol.*, **54**, 26–33.
- Olsen, J.G., Teilmann, K. and Kragelund, B.B. (2017) Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci. CMLS*, **74**, 3175–3183.
- Mitrea, D.M. and Kriwacki, R.W. (2016) Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal. CCS*, **14**, 1.
- Martin, E.W. and Mittag, T. (2018) Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry*, **57**, 2478–2487.
- Wright, P.E. and Dyson, H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
- Bah, A. and Forman-Kay, J.D. (2016) Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.*, **291**, 6696–6705.
- Csizmok, V. and Forman-Kay, J.D. (2018) Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications. *Curr. Opin. Struct. Biol.*, **48**, 58–67.
- Weatheritt, R.J. and Gibson, T.J. (2012) Linear motifs: lost in (pre)translation. *Trends Biochem. Sci.*, **37**, 333–341.
- Van Roey, K., Gibson, T.J. and Davey, N.E. (2012) Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.*, **22**, 378–385.
- Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.*, **579**, 3346–3354.
- 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.
- Tompa, P., Davey, N.E., Gibson, T.J. and Babu, M.M. (2014) A million peptide motifs for the molecular biologist. *Mol. Cell*, **55**, 161–169.
- Davey, N.E., Travé, G. and Gibson, T.J. (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.*, **36**, 159–169.
- Davey, N.E., Cyert, M.S. and Moses, A.M. (2015) Short linear motifs - ex nihilo evolution of protein regulation. *Cell Commun. Signal. CCS*, **13**, 43.
- Schad, E., Fichó, E., Pancsa, R., Simon, I., Dosztányi, Z. and Mészáros, B. (2018) DIBS: a repository of disordered binding sites mediating interactions with ordered proteins. *Bioinforma. Oxf. Engl.*, **34**, 535–537.
- Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J.A., Bukirova, D., Čalyševa, J. *et al.* (2020) ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.*, **48**, D296–D306.
- Hatos, A., Hajdu-Soltész, B., Monzon, A.M., Palopoli, N., Álvarez, L., Aykac-Fas, B., Bassot, C., Benítez, G.I., Bevilacqua, M., Chasapi, A. *et al.* (2020) DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.*, **48**, D269–D276.
- 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.
- 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.
- Fichó, E., Reményi, I., Simon, I. and Mészáros, B. (2017) MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinforma. Oxf. Engl.*, **33**, 3682–3684.
- 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.
- Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O.P., Kovács, M. *et al.* (2020) PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.*, **48**, D360–D367.
- You, K., Huang, Q., Yu, C., Shen, B., Sevilla, C., Shi, M., Hermjakob, H., Chen, Y. and Li, T. (2020) PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.*, **48**, D354–D359.
- Li, Q., Peng, X., Li, Y., Tang, W., Zhu, J., Huang, J., Qi, Y. and Zhang, Z. (2020) LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro. *Nucleic Acids Res.*, **48**, D320–D327.
- Necci, M., Piovesan, D. and Tosatto, S.C.E. (2018) Where differences resemble: sequence-feature analysis in curated databases of intrinsically disordered proteins. *Database J. Biol. Databases Curation*, **2018**, bay127.
- Mir, S., Alhroub, Y., Anyango, S., Armstrong, D.R., Berrisford, J.M., Clark, A.R., Conroy, M.J., Dana, J.M., Deshpande, M., Gupta, D. *et al.* (2018) PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic Acids Res.*, **46**, D486–D492.
- Monzon, A.M., Necci, M., Quaglia, F., Walsh, I., Zanotti, G., Piovesan, D. and Tosatto, S.C.E. (2020) Experimentally Determined Long Intrinsically Disordered Protein Regions Are Now Abundant in the Protein Data Bank. *Int. J. Mol. Sci.*, **21**, 4496.
- 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.
- 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.
- Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I. *et al.* (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, **47**, D351–D360.
- UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- 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.
- Necci, M., Piovesan, D. and Tosatto, S.C.E. (2016) Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein Sci. Publ. Protein Soc.*, **25**, 2164–2174.
- Mészáros, B., Simon, I. and Dosztányi, Z. (2009) Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.*, **5**, e1000376.
- Jones, D.T. and Cozzetto, D. (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinforma. Oxf. Engl.*, **31**, 857–863.

36. Miskei, M., Horvath, A., Vendruscolo, M. and Fuxreiter, M. (2020) Sequence-based prediction of fuzzy protein interactions. *J. Mol. Biol.*, **432**, 2289–2303.
37. Horvath, A., Miskei, M., Ambrus, V., Vendruscolo, M. and Fuxreiter, M. (2020) Sequence-based prediction of protein binding mode landscapes. *PLoS Comput. Biol.*, **16**, e1007864.
38. Davey, N.E., Babu, M.M., Blackledge, M., Bridge, A., Capella-Gutierrez, S., Dosztanyi, Z., Drysdale, R., Edwards, R.J., Elofsson, A., Felli, I.C. *et al.* (2019) An intrinsically disordered proteins community for ELIXIR. *FL1000Research*, **8**, ELIXIR-1753.
39. 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.
40. 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**, W367–374.
41. Pancsa, R. and Fuxreiter, M. (2012) Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB Life*, **64**, 513–520.
42. Dosztanyi, Z., Mészáros, B. and Simon, I. (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, **25**, 2745–2746.
43. Piovesan, D. and Tosatto, S.C.E. (2018) Mobi 2.0: an improved method to define intrinsic disorder, mobility and linear binding regions in protein structures. *Bioinforma. Oxf. Engl.*, **34**, 122–123.
44. 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.
45. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M. *et al.* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.*, **47**, D442–D450.
46. Banani, S.F., Lee, H.O., Hyman, A.A. and Rosen, M.K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, **18**, 285–298.
47. Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L. *et al.* (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.*, **28**, D420–D435.
48. Paladin, L., Schaeffer, M., Gaudet, P., Zahn-Zabal, M., Michel, P.-A., Piovesan, D., Tosatto, S.C.E. and Bairoch, A. The feature viewer: a visualization tool for positional annotations on a sequence. *Bioinformatics*, **36**, 3244–3245.
49. Sehnal, D., Rose, A.S., Koča, J., Burley, S.K. and Velankar, S. (2018) Mol*: towards a common library and tools for web molecular graphics. In: *Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data, MolVA '18*. Eurographics Association, Goslar, DEU, pp. 29–33.
50. Nightingale, A., Antunes, R., Alpi, E., Bursteinas, B., Gonzales, L., Liu, W., Luo, J., Qi, G., Turner, E. and Martin, M. (2017) The proteins API: accessing key integrated protein and genome information. *Nucleic Acids Res.*, **45**, W539–W544.
51. Gray, A.J.G., Goble, C. and Jimenez, R.C. (2017) From Potato Salad to Protein Annotation.
52. Zsolyomi, F., Ambrus, V. and Fuxreiter, M. (2020) Patterns of dynamics comprise a conserved evolutionary trait. *J. Mol. Biol.*, **432**, 497–507.
53. Marchetti, J., Monzon, A.M., Tosatto, S.C.E., Parisi, G. and Fornasari, M.S. (2019) Ensembles from ordered and disordered proteins reveal similar structural constraints during evolution. *J. Mol. Biol.*, **431**, 1298–1307.
54. 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.
55. Mészáros, B., Erdős, G. and Dosztanyi, Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
56. Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K. and Obradovic, Z. (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, **7**, 208.
57. 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.
58. Linding, R., Russell, R.B., Neduva, V. and Gibson, T.J. (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.*, **31**, 3701–3708.
59. Yang, Z.R., Thomson, R., McNeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
60. Piovesan, D., Walsh, I., Minervini, G. and Tosatto, S.C.E. (2017) FIELDS: fast estimator of latent local structure. *Bioinformatics*, **33**, 1889–1891.
61. 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.
62. Jones, D.T. and Swindells, M.B. (2002) Getting the most from PSI-BLAST. *Trends Biochem. Sci.*, **27**, 161–164.
63. Wootton, J.C. (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.*, **18**, 269–285.
64. Lewis, T.E., Sillitoe, I., Dawson, N., Lam, S.D., Clarke, T., Lee, D., Orengo, C. and Lees, J. (2018) Gene3D: extensive prediction of globular domains in proteins. *Nucleic Acids Res.*, **46**, D435–D439.
65. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A. *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432.
66. Malhis, N., Wong, E.T.C., Nassar, R. and Gsponer, J. (2015) Computational identification of MoRFs in protein sequences using hierarchical application of Bayes rule. *PLoS One*, **10**, e0141603.