Nucleic Acids Research, 2025, **53**, D487–D494
https://doi.org/10.1093/nar/gkae976
Advance access publication date: 11 November 2024
Database issue

MFIB 2.0: a major update of the database of protein complexes formed by mutual folding of the constituting protein chains

Erzsébet Fichó[1,†], Rita Pancsa[2,†], Csaba Magyar[2], Zsofia E. Kalman[3], Éva Schád[2], Bálint Z. Németh[2], István Simon[2,‡], Laszlo Dobson[2,4,*] and Gábor E. Tusnády[2,4,*]

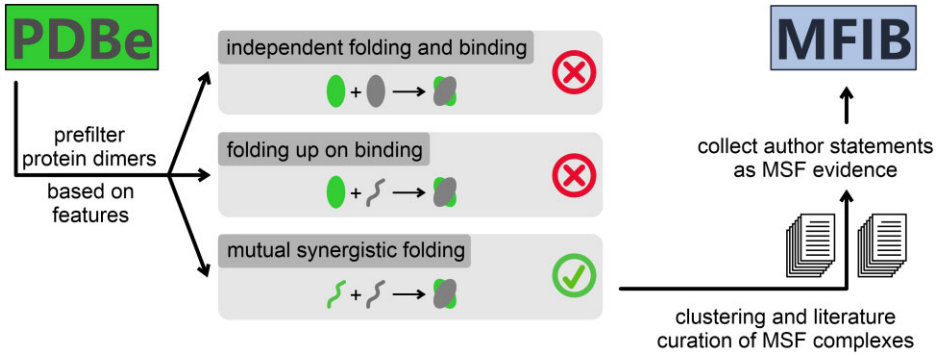
[1]Department of Bioinformatics, Cytocast Hungary Kft, Petőfi Sándor utca 5/A, Budapest 1052, Hungary
[2]Institute of Molecular Life Sciences, Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, Budapest 1117, Hungary
[3]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest 1083, Hungary
[4]Department of Bioinformatics, Semmelweis University, Tűzoltó u. 7, Budapest 1094, Hungary

[*]To whom correspondence should be addressed. Tel: +36 1 3826 727; Email: tusnady.gabor@ttk.mta.hu
Correspondence may also be addressed to Laszlo Dobson. Email: dobson.laszlo@ttk.hu
[†]These authors have contributed equally to the work.
[‡]Deceased.

Abstract

While the majority of proteins with available structures are able to fold independently and mediate interactions only after acquiring their folded state, a subset of the known protein complexes contains protein chains that are intrinsically disordered in isolation. The Mutual Folding Induced by Binding (MFIB) database collects and classifies protein complexes, wherein all constituent protein chains would be unstable/disordered in isolation but fold into a well-defined 3D complex structure upon binding. This phenomenon is often termed as cooperative folding and binding or mutual synergistic folding (MSF). Here we present a major update to the database: we collected and annotated hundreds of new protein complexes fulfilling the criteria of MSF, leading to an almost six-fold increase in the size of the database. Many novel features have also been introduced, such as clustering of the complexes based on structural similarity and domain types, assigning different evidence levels to each entry and adding the evidence coverage label that allowed us to include complexes of multi(sub)domain monomers with partial MSF. The MFIB 2.0 database is available at https://mfib.pbrg.hu.

Graphical abstract



Introduction

Proteins sample a wide spectrum of structural states (1). While folded/globular proteins have a well-defined 3D structure, intrinsically disordered proteins/regions (IDPs/IDRs) lack a stable tertiary structure under physiological conditions, instead they exist as dynamic conformational ensembles (1–3). Many proteins are modular, meaning that they contain both well-folded domains and disordered regions, also, many exploit conformational transitions for their functions, wherein they shift between different structural states in a regulated manner, on some stimulus or partner binding (4). IDPs themselves are also not uniform regarding their structural states; random coil-like extended chains, pre-molten globules and molten globules all classify as IDPs, even though they repre-

Received: August 14, 2024. **Revised:** September 26, 2024. **Editorial Decision:** October 9, 2024. **Accepted:** October 14, 2024

sent markedly different levels of compactness and secondary structure content (5,6).

IDPs fulfill many important functions in a diverse set of biological processes in cells (3,7–9), and most of those are linked to their interaction capacities, i.e. they are interaction specialists (10) and often serve as hubs in interaction networks (11,12) mainly due to their conformational adaptability and ability to interact with multiple partners (13) typically by employing short linear motifs (SLiMs) (14,15). Binding pockets on domain surfaces mediating interactions with disordered regions (e.g. with SLiMs) are attractive targets for therapeutic intervention, both for cancer treatment (16) or for developing antivirals (17).

The initial structural states of the protein chains forming protein complexes can vary widely, and this provides a firm basis for the categorization of the complexes. Considering a binary classification of constituting partners (ordered/disordered), protein complexes may form through autonomous folding and independent binding (AFIB) (when ordered partners interact), coupled folding and binding (disordered partner adapts to ordered partner) (18–20) or mutual synergistic folding (MSF; two disordered partners interact through cooperative folding and binding) (21). Notably, certain IDPs can also engage in so-called fuzzy complexes (22) or liquid–liquid phase separation (23,24) where they do not fold up on binding but remain largely disordered in the complex.

The Mutual Folding Induced by Binding (MFIB) database, established in 2017 (25), focuses on complexes formed through MSF. Such complexes also show structural and functional diversity. The protein chains forming MSF complexes can be classical IDPs that also exist in cells in a monomeric form, but form transient complexes with the right partner(s) under specific circumstances (e.g. the ACTR–NCBD interaction (26), or the phosphorylation-controlled intrachain interaction within FOXM1 (27)). Such heteromeric transient complexes are mainly involved in signaling. In other cases, the protein chains do not exist and function in a monomeric form because they are absolutely dependent on the formation of permanent complexes (mainly obligate homooligomers (28)). Proteins belonging to the latter group are typically not fully disordered, but molten globule-like nearly ordered structures (29) that would be unstable on their own. They are often aggregation-prone as isolated monomers due to exposing hydrophobic surfaces to the solvent (30,31). Accordingly, such proteins usually form highly stable complexes (often referred to as two-state dimers) and therefore the monomeric forms are not present in cells and are literally impossible to study *in vitro* (30,32,33).

While certain experimental approaches can conclusively prove MSF, e.g. if the complex follows a two-state unfolding process from a folded multimer to unfolded monomers when varying some external factor (most commonly the temperature or the concentration of some denaturing agent) (21,34), such measurements are only available for a limited number of protein complexes. Often, there is little or no information on the structural states of the monomers (especially in the case of the above-mentioned obligate homooligomers), and/or the results of several publications need to be combined to conclude on MSF for a complex. This makes collecting such information difficult and laborious. The original release of the database mainly contained complexes where there is experimental evidence for the disordered state of the monomers, which resulted in a relatively limited set of 205 entries (25).

Here, we present a major update to the MFIB database, where we not only collected new entries along the old terms but also considered a wider scope of evidence supporting MSF and thereby created a more comprehensive data resource. During data collection and manual curation, we considered many hallmarks of complex structures that are frequently discussed in their accompanying publications and can be indicative of MSF. Some of the selected properties were proved to be capable of discriminating MSF complexes from other types of complexes (35), but all of them can be suggestive of MSF along rational arguments. Besides achieving an almost six-fold increase in the size of the database, we also introduced a new framework to classify entries based on the quality of evidence supporting their MSF, implemented a structural similarity- and domain type-based clustering approach, enabled the inclusion of multi(sub)domain entries exhibiting only partial MSF, and revamped old entries to achieve a high level of integration with the novel ones.

Materials and methods

Data processing

First, we performed a pre-filtering approach on the protein complex structures from PDBe (36), where the most preferred assembly composition is dimeric to get a list of MSF candidate complexes that can then be manually curated. To eliminate possible false positives, we excluded all entries where any of the chains have a solved monomeric PDB structure (SIFTS (37) was used to check if multiple PDB chains belong to the same region of a full-length protein sequence from UniProt). We further excluded proteins with (Charged) Single Alpha Helices using the CSAH server (38) (often cross-predicted with coiled-coils (CCs) and disordered regions (39)) and structures containing transmembrane region(s) (40).

Next, we calculated surface and contact properties/features of all dimeric complexes, in a similar way as described by Mészáros et al. (35); however we replaced the Naccess method (41) with Voronota (42) to calculate surfaces and contacts. We compared the 'autonomous folding and independent binding' set ('AFIB set') (35) and the original MFIB database ('MFIB (2017)') and selected features where the average ± the standard deviation of the two datasets do not overlap, resulting in six characteristic properties (see results for the list of these properties and the Supplementary Table S1 for calculated values). Dimeric protein complexes were then filtered using these features and only those complexes were kept as MSF candidates, for which at least one of the calculated features was within the newly defined average ± the standard deviation range.

To achieve clustering of the similar complex structures before manual curation, we performed a structural search on all (including original and new candidate) entries with Foldseek (43), searching all protein chains against all protein chains. In parallel, we also searched for Pfam (44) domains in all constituting chains. We defined protein chain pairs if two chains shared a TM-Score ≥ 0.3 (where false positive ratio is diminishing according to Xu *et al.* (45)) and 60% sequence coverage, or if their unique lists of Pfam domains were identical. A greedy algorithm (starting with protein chains that had the highest number of chains paired based on Foldseek and Pfam) was used to generate protein groups. Next, we extended the clustering to complexes: two protein complexes belong to the

same cluster, if their individual chains belong to the same (single chain) cluster(s).

As a starting point of the manual annotation procedure, the annotators received the clustered candidate protein complexes and their work was further supported by annotations from some external sources (DisProt (46) annotations on disordered regions and information from SOCKET (47) on CCs). During the manual annotation procedure, the annotators were looking for author statements on a predefined set of features indicative of MSF (listed in the Results section describing the manual annotation of complexes) in the descriptive article of the complex structure, and, if available, in additional papers describing the stability of constituting protein chains. All entries included in MFIB 2.0 were checked by at least two annotators to ensure quality. Supplementary Figure S1 shows the pipeline how data were processed, including prefiltering and manual curation.

Updating the web-server backend and frontend

MFIB 2.0 can be accessed through a user-friendly, PHP-based web interface backed by a highly efficient multi-layer SQL database (MariaDB 8.0.32). The current version uses the same front-end style as the original one, using the Bootstrap framework (version 4.6). The page uses the latest versions of the open-source Molstar (version 4.4) and Chart.js (version 4.4.3.) libraries for visualization.

In addition to providing access to the data through the online interface, they can also be downloaded in XML or JSON formats, separately on each entry page, or in a zipped format from the Downloads page, including all entries in plain text format.

The webpage is optimized for desktop view, and it was tested on Windows, Linux and Macintosh operating systems.

Results

Prefiltering the PDB using indicative surface and contact properties

While complexes going through coupled folding and binding can be easily separated from MSF complexes by visual inspection, 'autonomous folding and independent binding' complexes are much more challenging to discriminate. The literature describing protein complexes that could undergo MSF often discusses indicative features that are characteristic of these complexes. Many of these features can be calculated using the PDB structures and show predictive power to identify MSF proteins using the structure alone. As part of our prefiltering approach, we selected the six surface and atomic contact related properties that best discriminated between the different types of complexes (see Methods, Supplementary Table S1), and we calculated these features for dimeric PDB entries. The following fractions were selected as features: buried area/all surface area, interface area/all surface area, interchain atomic contacts/all atomic contacts, interchain hydrophobic-hydrophobic/all atomic contacts, interchain hydrophobic-polar/all atomic contacts and the interchain backbone-backbone/all atomic contacts. Complexes that had at least one of these features in the average ± standard deviation range calculated for the entries in the original MFIB database were considered as MSF candidates and passed on for manual curation after clustering.

Clustering entries based on structure and sequence information

The original MFIB database used UniRef90 for clustering of the complexes and only a single representative structure of each cluster was displayed on the webpage. While collecting the data we realized that the original clustering based on 90% sequence identity was extremely stringent: in many cases, proteins in different clusters exhibited almost identical structures and belonged to the same domain family, furthermore, the descriptive articles of MSF complexes often mentioned close-homologue, structurally very similar complexes that showed much lower sequence identity values. Therefore, based on the assumption that the folding cores and paths are largely conserved between protein chains belonging to the same domain type, we shifted to structure/domain-based clustering and clustered all (old and new candidate) entries using Foldseek and Pfam annotations (see 'Methods' section). The resulting clusters were checked during manual curation and sometimes manually adjusted when a subgroup in the cluster had additional elements or domains compared to other members (e.g. Nudix domain containing proteins were separated into two clusters based on their additional subdomains).

The original MFIB contained 205 separate entries all belonging to different clusters that may still share similarities or belong to the same fold. The 1122 protein complexes of the updated MFIB database are grouped into 395 distinct clusters that all represent different protein domain types (Figure 1/A, left). The most populated cluster is a group of dioxygenases with 64 members, while there are 245 clusters with exactly one element/complex. Since the current release focused on dimeric structures, most of the included clusters are dimeric (almost 95% of all entries), yet higher-order oligomers up to heptamers are also included from the first release.

Manual annotation of complexes and new classifications

All clustered candidate complexes were subjected to a manual annotation phase where the curators of the database made informed decisions (mainly considering evidence in relevant literature) on each complex to be included into the database or rejected, therefore MFIB 2.0 is a fully manually curated database. During this manual annotation phase, diverse lines of evidence were collected supporting MSF of the complexes (at least for one in each cluster). In cases where contradicting evidence was found (e.g. any cluster member exists in a folded monomeric state in solution) or the structures contained any transmembrane regions (for which folding can only be studied and interpreted in the context of lipids), the cluster was rejected. Also, a unique, representative domain was assigned to each cluster that was present in all members. In most cases, the name of the Pfam domain was used as the representative domain, except, when it was lacking and therefore the CATH domain or protein name needed to be used. In some rare cases, distinctive specifications were added to the automatically assigned domain type to distinguish the given cluster from another cluster with a similar domain type. Broader classes and subclasses were also assigned to the complexes as part of the manual curation procedure. The original classes of the MFIB database were mostly kept, and in addition new classes (and subclasses) were also defined, such as BAR domains, Bacterial toxin-antitoxin systems or Cystine-knot cytokines, that each covered several clusters.

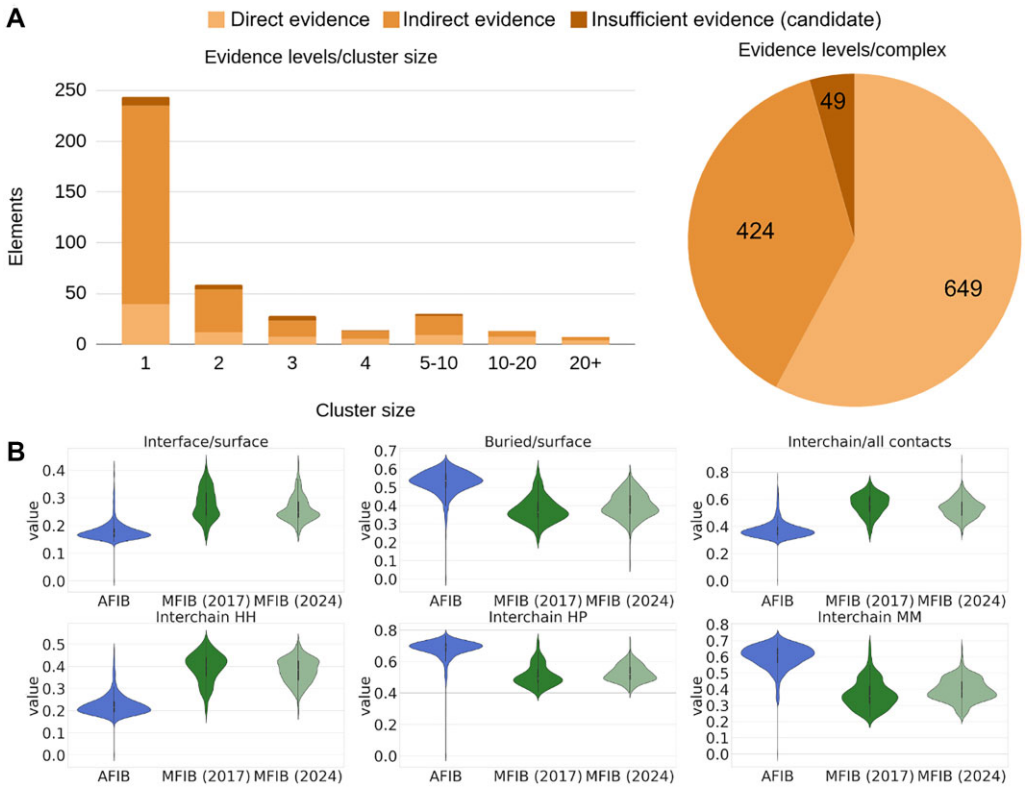


Figure 1. Statistical analysis of various structural features and evidence levels in MFIB 2.0: (**A**) Left: distribution of evidence levels in different cluster sizes; Right: distribution of evidence levels for all MFIB 2.0 complexes. (**B**) Characteristic calculated surface and contact features compared between ordered complexes (AFIB), old (MFIB (2017)) and all MFIB entries (MFIB (2024)) (interface/surface: interface accessible area/surface accessible area, buried/surface: buried area/surface accessible area, interchain/all contacts: interchain atomic contacts/all atomic contacts, interchain HH: interchain hydrophobic-hydrophobic atomic contacts, interchain HP: interchain hydrophobic-polar atomic contacts, interchain MM: main chain-main chain atomic contacts between chains). For each group (ordered complexes: 548 clusters, MFIB (2017): 123 clusters, MFIB (2024): 395 clusters), the distributions of the mean values calculated for each cluster are shown for all the features and used for comparisons.

As a novel feature introduced in MFIB 2.0, the entries are grouped into three different evidence levels (direct evidence, indirect evidence and insufficient evidence (candidate)) based on the types of information supporting their MSF (Figure 1/A, right). 'Direct evidence' label was added to clusters/complexes for which MSF is supported by experimental data proving that all constituent protein chains only adopt a stable structure as a result of complex formation. The following conclusive evidence types were accepted as direct proof of MSF: (1) All protein chains (at least 70% of the interacting regions of the proteins) have been shown to be intrinsically disordered in their monomeric form (based on the DisProt database). (2) The folding/unfolding of the complex was traced while changing some environmental factor (most often the temperature or the concentration of some denaturing agent) and the well-defined tertiary structure of the monomers disappeared exactly when the complex was broken up. Most annotated complexes fulfilled this criterion by following a two-state folding/unfolding behavior from folded oligomers to unfolded monomers, while some had a folding/unfolding path with more than two states, but none of those were a folded monomer. This approach is particularly important for obligate homooligomeric complexes, in which the monomers cannot be studied in isolation. (3) The vast majority of the complex structure was made up by a structural element, such as CC, ribbon-helix-helix or basic helix-loop-helix that has been previously demonstrated to acquire its dimeric folded structure through MSF. In such cases, the complex received 'Direct evidence' label with no further examination required and was classified into the respective dedicated Class/Subclass. (4) The authors claimed that certain features of the complex structure clearly imply that the isolated monomers would not be stable, e.g. because the hydrophobic core of the dimer extends through the monomer–monomer interface.

In many cases, the above listed direct evidence was lacking, but some characteristic features of the complex mentioned in relevant publications implied that it is a case of MSF. Therefore, complexes showing a convincing combination of the following features indicative of MSF (mainly based on author statements but also on observations made by the curators during visual inspection of the complex structures) received the 'Indirect evidence' label: (1) large relative interaction surface, (2) large and hydrophobic buried surfaces, (3) highly intertwined/interdigitated/intimate complex structure and/or extensive domain swapping, (4) beta sheet augmentation and/or helix packing/CC/helix bundle forming interactions take place between segments of the monomers, (5) functional sites (active site, cofactor-binding site, etc.) lie on the monomer–monomer interface, suggesting loss of function on dissociation, (6) only oligomeric states detected in solution by dedicated biophysical methods (such as gel filtration, analytical ultracentrifugation, dynamic light scattering, Sodiumdodecyl sulphate-polyacrylamide gel electrophoresis, etc.) even at low protein concentrations.

Finally, there are some cases (∼5% of the database), where the properties of the structure (interface size, types of atomic contacts between monomers, etc.) indicate that the complex is likely formed through MSF, but there isn't enough direct or indirect evidence to conclude on MSF. At the same time, the curators couldn't identify any evidence against MSF. These complexes received the 'insufficient evidence (candidate)' evidence level label and can be considered as promising candidates for MSF.

In MFIB 2.0, we established the possibility to include structures with multi(sub)domain monomers, where only one of the (sub)domains undergo MSF (see the 'evidence coverage' on the entry pages of MFIB 2.0). For partial MSF cases, the 'complex evidence' text describes which (sub)domain of the complex undergoes MSF. Partial MSF was annotated for 20% of the entries, most often to dimers of winged helix transcription factors that contain independently folded DNA-binding domains and a helical dimerization domain that is typically formed through MSF involving CC formation and/or extensive helix packing interactions.

A comparison of the structural properties of complexes formed exclusively by ordered or disordered protein chains

After finishing the manual annotation round and arriving at the final MFIB (2024) dataset, the same set of 6 features that have been used for prefiltering of the data were calculated on and used to compare the MFIB (2024), MFIB (2017) and ordered complexes (AFIB set) datasets to see how much they differ (Figure 1/B). The distributions of the mean values calculated for each cluster are shown for all the features and used for comparisons (see Supplementary Table S2 for calculated values for structures, Supplementary Table S3 for clusters assignments and Supplementary Table S4 for mean calculated values for clusters). We found that both the original and the new release of MFIB have significantly different distributions compared to the set of ordered complexes taken from Mészáros et al. (35) (Supplementary Table S5). Notably, in some cases, a significant difference (with higher P-value) is also present between the original and the new release of MFIB. The latter difference likely originates from the inclusion of multi(sub)domain complexes where only a part of the complex adheres to MSF, and the inclusion of protein complexes based on a wider range of (direct or indirect) evidence than previously. However, since all entries were manually checked, MSF is certainly well supported by several lines of evidence for all the complexes included in MFIB 2.0.

Webpage of the MFIB 2.0 database

There are several ways to navigate on the webpage. The Home page describes the purpose of the database and provides crosslinks to related resources. In the browser, all entries are visible in a compact and sortable table, with the most relevant information listed: PDB id, the oligomerization state of the assembly, source organism, evidence level and structural (sub)classes. The Protein Map enables users to search for entries in a specific structural group. In the search menu, the free text box can be used to look for full or partial entry names, MFIB accessions, assemblies, organisms, experimental methods, classes and subclasses, PDB IDs or UniProt AC-s.

Using these menus users can navigate to the entry page of the structure of their interest. The entry page consists of five panels. In 'General information', PDB accession and basic structural information is shown, as well as the primary publication describing the complex. On the right side, the PDB structure can be inspected in the molecule viewer. Information related to the particular entry (structure, XML and JSON formatted files) can be downloaded from here. The 'Function and Biology' panel holds GeneOntology (GO) annotations of all three biological aspects. Only those GO terms are listed here that were assigned for each constituting chain of the MSF complex. The 'Structure Summary' panel helps to decide whether all chains of the PDB structure participate in MSF (important if the complex is a trimer or higher oligomer). All MSF chains are listed below: the PDB structure is mapped to their respective UniProt sequence in the sequence viewer and secondary structure, Pfam domains are also displayed aligned to the sequence. During the construction of MFIB 2.0, we changed how evidence was collected, and therefore, the 'Evidence' panel has been extended with new elements. Annotations in MFIB now belong to a representative domain defining the cluster (except for clusters with one entry). All entries are now classified into different 'Evidence levels' (direct evidence, indirect evidence or insufficient evidence (candidate)). While the original MFIB database included manually modified PDB structures to have a better integration, we now always use the most probable PDBe assembly – however when only some parts of the protein participates in the interaction, we highlight this information in the 'evidence coverage' section. Last, but not least, as a novel feature, surface and contact features of the protein complex are also displayed together with the distribution of that feature for the whole database (Figure 2). At the bottom 'Similar structures' shows entries that were clustered together with the inspected entry.

Using the 'Downloads' menu, users can download the whole database in several formats (JSON, XML, plain txt). PDB structures can also be downloaded here. In the statistics menu we highlight some of the interesting features of the MFIB 2.0 database, including the distribution of MFIB data between different oligomeric states, the most represented species or the distribution of entries between the different evidence levels. When showing the statistics on oligomeric states, homo- and hetero-oligomers are also distinguished, highlighting that the vast majority of the database is constituted by homooligomers (mostly homodimers) that can be mostly regarded as permanent/obligatory complexes, while only <10% of the database is made up of heterooligomers (mostly heterodimers) that are usually transient assemblies of proteins. A detailed 'Help' menu includes additional information about MSF proteins and how information was collected and prepared for the resource.

Conclusion/discussion

The original release of the MFIB database (25) contained 205 protein complexes proved to undergo MSF. MFIB has already been used by several other resources in its original form despite having a limited number of entries, and to our knowledge, it has remained the sole resource focusing on such data ever since its original publication. Here, a major update to the MFIB database is presented, in which we not only collected and annotated new cases of MSF, achieving a more than 5-fold increase in the size of the database, but also introduced several novel features regard-

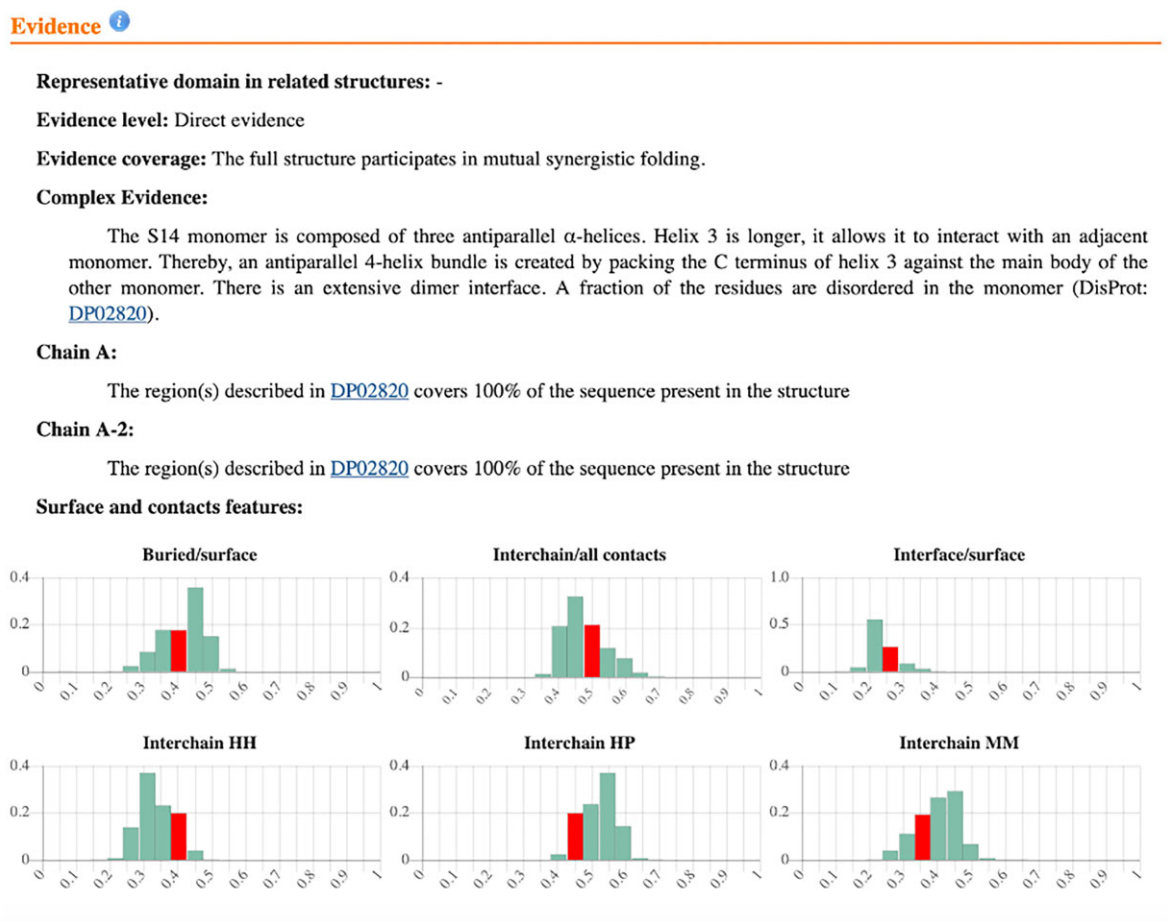


Figure 2. Layout of the evidence panel of MFIB 2.0 entry pages. The 'Evidence' section lists the following information for each entry: representative domain in related structures (if applicable), evidence level, evidence coverage, a free text description of the evidence supporting MSF for the complex, evidence for each chain (if applicable) and six histograms, visualizing six different surface and contacts features (histograms show the distributions of the calculated features for the whole MFIB dataset with the data bin where the given entry belongs being highlighted). The example shows the evidence panel of the MF7000293 entry.

ing the classification of the presented complexes (both old and new). Besides the direct, experimental evidence supporting the disorder/instability of the monomeric forms of the interaction partners forming complexes by MSF that has been accepted in the original release of the database, a well-defined set of indirect evidence has now been considered that were based on author and curator statements on the architecture, interface and detected solution oligomeric state of the complexes.

We believe that this major update of the database will (1) remarkably increase its usability through increasing the size of the dataset and enabling users to select subsets of the data based on novel classifications, e.g. evidence levels, (2) increase our understanding of the distinctive characteristics of co-dependent protein chains and cooperative folding, (3) enable the development of machine learning-based prediction methods for the identification of such protein complexes (to narrow down the curators' work to the most promising candidate complexes during annual database updates and to allow future expansion of the database with predicted entries labeled by a dedicated, novel evidence level), and (4) aid the development/improvement of general protein complex structure prediction methods, such as AlphaFold (AF) (48,49). While AF seems to be an excellent predictor for protein disorder (50) and conditional folding (51) and even capable to predict SLiM mediated interactions to some extent (52), disordered protein chains undergoing MSF and therefore appearing in the PDB in a stable form as part of a complex tend to fool all disorder prediction methods trained on the PDB, including those that use AF to assess flexibility. Therefore, there is certainly room for improvement in the prediction of the structure of protein complexes even after the publication of AF 3, and our large, manually curated dataset could be key for the development of methods that can accurately handle MSF cases despite the two-faced behavior of the constituting protein chains.

Data availability

All data can be freely downloaded from the public webpage of the MFIB 2.0 database, which is accessible at https://mfib.pbrg.hu.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

Many thanks to the late Prof. István Simon, without whom this project would never have been possible. We hope we made him proud.

Funding

NKFIH [FK-142285, PD-146564 and K-146314 to R.P, L.D. and G.E.T., respectively]; Hungarian Academy of Sciences [BO/00174/22 to R.P.].

Conflict of interest statement

None declared.

References

1. Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
2. Wright,P.E. and Dyson,H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, **293**, 321–331.
3. Tompa,P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.*, **27**, 527–533.
4. Mendoza-Espinosa,P., García-González,V., Moreno,A., Castillo,R. and Mas-Oliva,J. (2009) Disorder-to-order conformational transitions in protein structure and its relationship to disease. *Mol. Cell. Biochem.*, **330**, 105–120.
5. Uversky,V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.*, **11**, 739–756.
6. Uversky,V.N. (2002) What does it mean to be natively unfolded?*Eur. J. Biochem.*, **269**, 2–12.
7. Csizmok,V., Follis,A.V., Kriwacki,R.W. and Forman-Kay,J.D. (2016) Dynamic protein interaction networks and new structural paradigms in signaling. *Chem. Rev.*, **116**, 6424–6462.
8. Habchi,J., Tompa,P., Longhi,S. and Uversky,V.N. (2014) Introducing protein intrinsic disorder. *Chem. Rev.*, **114**, 6561–6588.
9. 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.
10. Tompa,P., Schad,E., Tantos,A. and Kalmar,L. (2015) Intrinsically disordered proteins: emerging interaction specialists. *Curr. Opin. Struct. Biol.*, **35**, 49–59.
11. Dosztányi,Z., Chen,J., Dunker,A.K., Simon,I. and Tompa,P. (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.*, **5**, 2985–2995.
12. Dunker,A.K., Cortese,M.S., Romero,P., Iakoucheva,L.M. and Uversky,V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.*, **272**, 5129–5148.
13. Tompa,P., Szász,C. and Buday,L. (2005) Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.*, **30**, 484–489.
14. Kumar,M., Michael,S., Alvarado-Valverde,J., Zeke,A., Lazar,T., Glavina,J., Nagy-Kanta,E., Donagh,J.M., Kalman,Z.E., Pascarelli,S., *et al.* (2024) ELM-the eukaryotic Linear Motif resource-2024 update. *Nucleic Acids Res.*, **52**, D442–D455.
15. Van Roey,K., Uyar,B., Weatheritt,R.J., Dinkel,H., Seiler,M., Budd,A., Gibson,T.J. and Davey,N.E. (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.*, **114**, 6733–6778.
16. Corbi-Verge,C. and Kim,P.M. (2016) Motif mediated protein-protein interactions as drug targets. *Cell Commun. Signal.*, **14**, 8.
17. Simonetti,L., Nilsson,J., McInerney,G., Ivarsson,Y. and Davey,N.E. (2023) SLiM-binding pockets: an attractive target for broad-spectrum antivirals. *Trends Biochem. Sci.*, **48**, 420–427.
18. 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. *Bioinformatics*, **34**, 535–537.
19. Mészáros,B., Tompa,P., Simon,I. and Dosztányi,Z. (2007) Molecular principles of the interactions of disordered proteins. *J. Mol. Biol.*, **372**, 549–561.
20. Dyson,H.J. and Wright,P.E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, **12**, 54–60.
21. Rumfeldt,J.A.O., Galvagnion,C., Vassall,K.A. and Meiering,E.M. (2008) Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.*, **98**, 61–84.
22. Fuxreiter,M. and Tompa,P. (2012) Fuzzy complexes: a more stochastic view of protein function. *Adv. Exp. Med. Biol.*, **725**, 1–14.
23. Shin,Y. and Brangwynne,C.P. (2017) Liquid phase condensation in cell physiology and disease. *Science*, **357**, eaaf4382.
24. 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.
25. Fichó,E., Reményi,I., Simon,I. and Mészáros,B. (2017) MFIB: a repository of protein complexes with mutual folding induced by binding. *Bioinformatics*, **33**, 3682–3684.
26. Ganguly,D., Zhang,W. and Chen,J. (2012) Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. *Mol. Biosyst.*, **8**, 198–209.
27. Marceau,A.H., Brison,C.M., Nerli,S., Arsenault,H.E., McShan,A.C., Chen,E., Lee,H.-W., Benanti,J.A., Sgourakis,N.G. and Rubin,S.M. (2019) An order-to-disorder structural switch activates the FoxM1 transcription factor. *eLife*, **8**, e46131.
28. 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.
29. Mészáros,B., Dobson,L., Fichó,E. and Simon,I. (2019) Sequence and structure properties uncover the natural classification of protein complexes formed by intrinsically disordered proteins via mutual synergistic folding. *Int. J. Mol. Sci.*, **20**, 5461.
30. Nishi,H., Hashimoto,K., Madej,T. and Panchenko,A.R. (2013) Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization. *Prog. Mol. Biol. Transl. Sci.*, **117**, 3–24.
31. Ofran,Y. and Rost,B. (2003) Analysing six types of protein-protein interfaces. *J. Mol. Biol.*, **325**, 377–387.
32. Cai,M., Huang,Y., Shen,Y., Li,M., Mizuuchi,M., Ghirlando,R., Mizuuchi,K. and Clore,G.M. (2019) Probing transient excited states of the bacterial cell division regulator MinE by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, **116**, 25446–25455.
33. Ma,L., Jones,C.T., Groesch,T.D., Kuhn,R.J. and Post,C.B. (2004) Solution structure of dengue virus capsid protein reveals another fold. *Proc. Natl. Acad. Sci. USA*, **101**, 3414–3419.
34. Chen,G., Wang,C., Fuqua,C., Zhang,L.-H. and Chen,L. (2006) Crystal structure and mechanism of TraM2, a second quorum-sensing antiactivator of agrobacterium tumefaciens strain A6. *J. Bacteriol.*, **188**, 8244–8251.
35. Mészáros,B., Dobson,L., Fichó,E., Tusnády,G.E., Dosztányi,Z. and Simon,I. (2019) Sequential, structural and functional properties of protein complexes are defined by how folding and binding intertwine. *J. Mol. Biol.*, **431**, 4408–4428.
36. PDBe-KB consortium (2020) PDBe-KB: a community-driven resource for structural and functional annotations. *Nucleic Acids Res.*, **48**, D344–D353.
37. Dana,J.M., Gutmanas,A., Tyagi,N., Qi,G., O'Donovan,C., Martin,M. and Velankar,S. (2019) SIFTS: updated Structure

Integration with Function, Taxonomy and sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res.*, **47**, D482–D489.
38. Dudola,D., Tóth,G., Nyitray,L. and Gáspári,Z. (2017) Consensus prediction of charged single alpha-helices with CSAHserver. *Methods Mol. Biol.*, **1484**, 25–34.
39. Süveges,D., Gáspári,Z., Tóth,G. and Nyitray,L. (2009) Charged single alpha-helix: a versatile protein structural motif. *Proteins*, **74**, 905–916.
40. Dobson,L., Gerdán,C., Tusnády,S., Szekeres,L., Kuffa,K., Langó,T., Zeke,A. and Tusnády,G.E. (2024) UniTmp: unified resources for transmembrane proteins. *Nucleic Acids Res.*, **52**, D572–D578.
41. Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
42. Olechnovič,K. and Venclovas,C. (2014) Voronota: a fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *J. Comput. Chem.*, **35**, 672–681.
43. van Kempen,M., Kim,S.S., Tumescheit,C., Mirdita,M., Lee,J., Gilchrist,C.L.M., Söding,J. and Steinegger,M. (2024) Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.*, **42**, 243–246.
44. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J., *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
45. Xu,J. and Zhang,Y. (2010) How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, **26**, 889–895.
46. Quaglia,F., Mészáros,B., Salladini,E., Hatos,A., Pancsa,R., Chemes,L.B., Pajkos,M., Lazar,T., Peña-Díaz,S., Santos,J., *et al.* (2022) DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.*, **50**, D480–D487.
47. Kumar,P. and Woolfson,D.N. (2021) Socket2: a program for locating, visualizing and analyzing coiled-coil interfaces in protein structures. *Bioinformatics*, **37**, 4575–4577.
48. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
49. Abramson,J., Adler,J., Dunger,J., Evans,R., Green,T., Pritzel,A., Ronneberger,O., Willmore,L., Ballard,A.J., Bambrick,J., *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, **630**, 493–500.
50. Akdel,M., Pires,D.E.V., Pardo,E.P., Jänes,J., Zalevsky,A.O., Mészáros,B., Bryant,P., Good,L.L., Laskowski,R.A., Pozzati,G., *et al.* (2022) A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.*, **29**, 1056–1067.
51. Alderson,T.R., Pritišanac,I., Kolarić,Đ., Moses,A.M. and Forman-Kay,J.D. (2023) Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proc. Natl. Acad. Sci. USA*, **120**, e2304302120.
52. Lee,C.Y., Hubrich,D., Varga,J.K., Schäfer,C., Welzel,M., Schumbera,E., Djokic,M., Strom,J.M., Schönfeld,J., Geist,J.L., *et al.* (2024) Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. *Mol. Syst. Biol.*, **20**, 75–97.

Received: August 14, 2024. **Revised:** September 26, 2024. **Editorial Decision:** October 9, 2024. **Accepted:** October 14, 2024