Searching and Using MobiDB Resource 6 to Explore Predictions and Annotations for Intrinsically Disordered Proteins

Maria Cristina Aspromonte,[1,4] Federica Quaglia,[1,2,4] Alexander Miguel Monzon,[3] Damiano Clementel,[1] Alessio Del Conte,[1] Damiano Piovesan,[1] and Silvio C. E. Tosatto[1,5]

[1]Department of Biomedical Sciences, University of Padova, Padova, Italy
[2]Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari, Italy
[3]Department of Information Engineering, University of Padova, Padova, Italy
[4]These authors contributed equally to this work
[5]Corresponding author: *silvio.tosatto@unipd.it*

Published in the Bioinformatics section

Intrinsically disordered proteins (IDPs) make up around 30% of eukaryotic proteomes and play a crucial role in cellular processes and in pathological conditions such as neurodegenerative disorders and cancers. However, IDPs exhibit dynamic conformational ensembles and are often involved in the formation of biomolecular condensates. Understanding the function of IDPs is critical to research in many areas of science. MobiDB is a unique resource that serves as a comprehensive knowledgebase of IDPs and intrinsically disordered regions (IDRs), combining disorder annotations from experimental evidence and predictions for a broad range of protein sequences. Over the past decade, MobiDB has evolved with a focus on expanding annotation coverage, standardizing annotation provenance, and enhancing database accessibility. The latest MobiDB, version 6, released in July 2024, includes significant improvements, such as the integration of AlphaFoldDB predictions and a new homology transfer pipeline that has substantially increased the number of entries with high-quality annotations. The user interface has also been updated, highlighting annotation features, clarifying the entry page, and providing an immediate overview of disorder, binding, and disorder functions information in the protein sequence. This protocol guides the user through applications of the MobiDB, including disorder prediction, curated data analysis, and exploration of interaction data. This guide covers how to perform a search in MobiDB annotations using the web interface and the MobiDB REST API for programmatic access. The protocols use a step-by-step walkthrough using the human growth hormone receptor to demonstrate MobiDB's functions for visualization and interpretation of protein disorder data. © 2024 The Author(s). Current Protocols published by Wiley Periodicals LLC.

Basic Protocol 1: Searching MobiDB query formats
Basic Protocol 2: Searching MobiDB selected datasets and selected proteomes
Basic Protocol 3: Performing a search on the Statistics page in MobiDB
Support Protocol: Programmatic access with MobiDB REST API
Basic Protocol 4: Visualizing and interpreting a MobiDB Entry: The GHR use case

Keywords: database • intrinsic disorder predictions • intrinsically disordered proteins • MobiDB

CURRENT PROTOCOLS
A Wiley Brand

Current Protocols e70077, Volume 4
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpz1.70077

How to cite this article:
Aspromonte, M. C., Quaglia, F., Monzon, A. M., Clementel, D., Del Conte, A., Piovesan, D., & Tosatto, S. C. E. (2024). Searching and using mobiDB resource 6 to explore predictions and annotations for intrinsically disordered proteins. *Current Protocols*, *4,* e70077. doi: 10.1002/cpz1.70077

INTRODUCTION

The term "intrinsically disordered proteins" (IDPs) refers to biologically active proteins that lack a stable three-dimensional structure under physiological conditions and do not adopt a fixed 3D conformation. IDPs, and intrinsically disordered regions (IDRs) within proteins generally, are enriched in uncharged and polar amino acids, allowing them to exhibit dynamic behavior (van der Lee et al., 2014). Eukaryotic proteomes contain approximately 30% intrinsically disordered proteins or regions. Many IDPs are involved in protein-protein and protein-DNA/RNA interactions, and they play key roles in biological mechanisms such as liquid-liquid phase separation (Martin et al., 2020) and in various pathological conditions, such as neurodegenerative disorders (Venati & Uversky, 2024) and cancers.

MobiDB is a comprehensive resource for IDPs and IDRs, which has evolved significantly since its development more than a decade ago, providing an integrated database that combines manually curated, experimentally derived, and predicted data on IDPs. MobiDB utilizes disorder predictions propagated through the MobiDB-lite software, which is a distinct component of the MobiDB resource (Necci et al., 2017; Necci, Piovesan, Clementel, et al., 2021). The latest update of MobiDB (version 6, released in July 2024; Piovesan et al., 2024)) represents a substantial enhancement in both its functionality and user interface, further establishing its utility as an essential resource for researchers in the field of IDPs.

A notable enhancement in MobiDB since the version published in 2022 (Piovesan, Del Conte, et al., 2022) is the integration of AlphaFoldDB predictions (Jumper et al., 2021), which have shown high correlation with experimental disorder data and state-of-the-art disorder predictors according to the Critical Assessment of Intrinsic protein Disorder (CAID) assessment (Del Conte et al., 2023), extension of processed data, and propagation of methodological and functional ontologies terms such as the Evidence and Conclusion Ontology (ECO), Gene Ontology (GO) (Gene Ontology Consortium, 2023), and Intrinsically Disordered Proteins Ontology (IDPO). This integration not only enriches the database with accurate predictions but also allows the visualization of positional annotations in protein structures, even in the absence of experimental coordinates from the Protein Data Bank (PDB; Armstrong et al., 2020).

The process of homology transfer in MobiDB involves transferring annotations of particular proteins that have been manually curated to a wider array of proteins based on their homology. As a result, both the specificity and the extent of the annotations in the database have been increased substantially.

The entry page has also undergone substantial improvements. It has been redesigned to provide a more intuitive and informative user experience.

These significant updates to MobiDB motivate the publication of these protocol, which are designed to guide users through the enhanced features for analyzing and characterizing IDPs. This article provides a guide for efficiently accessing and utilizing the vast

data repository within the MobiDB resource. The first two Basic Protocols provide an overview of the MobiDB webpage and illustrate the various search options available on the MobiDB website. In particular, Basic Protocol 1 describes how to perform a search on the MobiDB homepage using different entry formats, whereas Basic Protocol 2 refers to a more specific and advanced search, such as for selected datasets and selected proteomes. Basic Protocol 3 focuses on how to perform searches using the Statistics page of MobiDB. Finally, Basic Protocol 4 describes how to visualize and interpret annotations of a MobiDB entry, providing insights into the structural and functional characteristics of IDPs using the human growth hormone receptor (GHR) protein as a case study. Additionally, a Support Protocol describes the process of accessing MobiDB programmatically through its REST API.

SEARCHING MobiDB QUERY FORMATS

BASIC PROTOCOL 1

The MobiDB homepage offers various search options to find the protein of interest (Fig. 1). Users can enter different types of queries in the "Search box" to retrieve relevant output pages. For proteins, users can search using the UniProt accession, protein name, or gene symbol. For organisms, searches can be performed using the organism name or NCBI taxonomy ID. The following steps guide users through exploring the MobiDB using these different query formats.

Necessary Resources

Hardware

Laptop or desktop computer (best; a smartphone or tablet is an acceptable alternative)
Active and stable internet connection

Software

Internet browser: e.g., Firefox (*http://www.mozilla.org/firefox*), Google Chrome (*http://www.google.com/chrome*), or Safari (*http://www.apple.com/safari*)

Input data

None required

Figure 1 MobiDB homepage and search box.

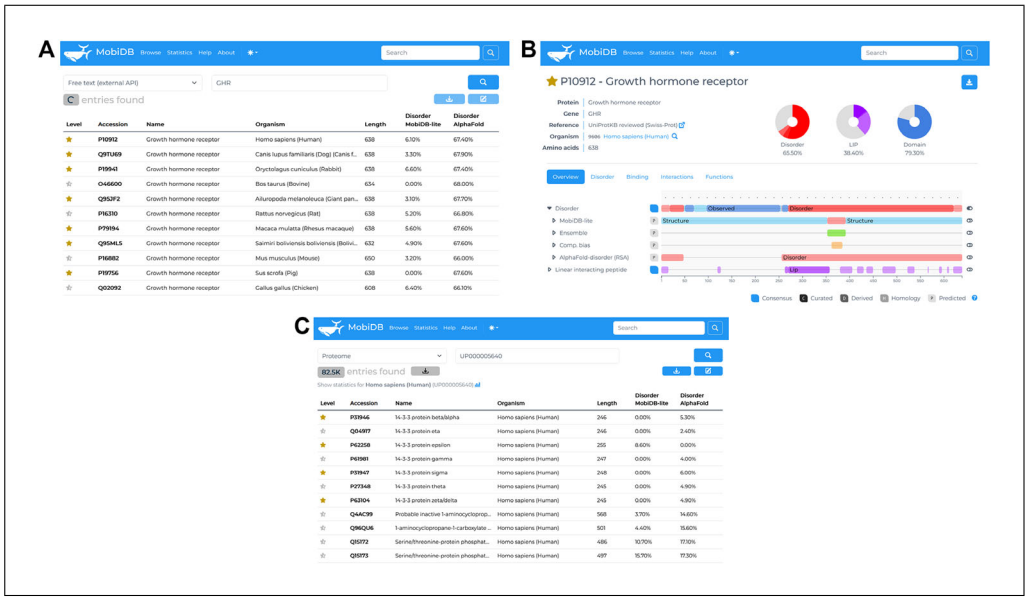

Figure 2 Search results page.

Performing a text search in MobiDB

1. Navigate to the MobiDB homepage at *https://mobidb.org/* using a browser.
2. Enter a search query in the "Search" boxes at the top right or top middle of the MobiDB home page (Fig. 1)—for example, search for "p53" or "P04637," which corresponds to a tumor suppressor protein with experimental evidence of disordered regions, as documented in DisProt for its specific functional relevance. Then click the Search button.
3. The output page can vary depending on the query format used (Fig. 2):
   i. Submitting the gene symbol 'GHR' (Example option) provides the uses with a list of related entries (Fig. 2A).
   ii. Submitting the UniProt accession number 'P10912' (Example option) directs the user to the corresponding MobiDB entry page (Fig. 2B).
   iii. Submitting the organism 'Human proteome' (Example option) provides the user with a list of all entries for that organism (Fig. 2C).
4. Alternatively, it is possible to perform an advanced search by clicking on the "Browse" button located at the top left of the home page (Fig. 1). This redirects users to an advanced search page (Fig. 3), where users can refine their search using specific queries or select predefined terms from a drop-down menu. For example, users might choose "Disorder MobiDB-lite," which defines disorder predictions propagated from MobiDB-lite software.

   Users can refine the search based on the following criteria:

   i. Disorder MobiDB-lite: This defines disorder predictions propagated from MobiDB-lite software. Users can choose from a range spanning 0.0–1.0.
   ii. UniProt IDs (comma separated): This option can be used for multiple UniProt IDs by separating them with a comma (e.g., P40344,Q13188). These UniProt ID examples will retrieve entries corresponding to Autophagy-related protein 3 from *Saccharomyces cerevisiae* (P40344) and to Serine/threonine-protein kinase 3 from *Homo sapiens* (Q13188).
   iii. Level: The levels indicate the reliability of the annotations (Silver and Gold stars). A specific level can be used for filtering the annotations of a selected proteome.
   iv. UniProt Reference Clusters (UniRef): UniRef databases cluster UniProtKB sequences by gathering together proteins based on their sequence similarity. Terms

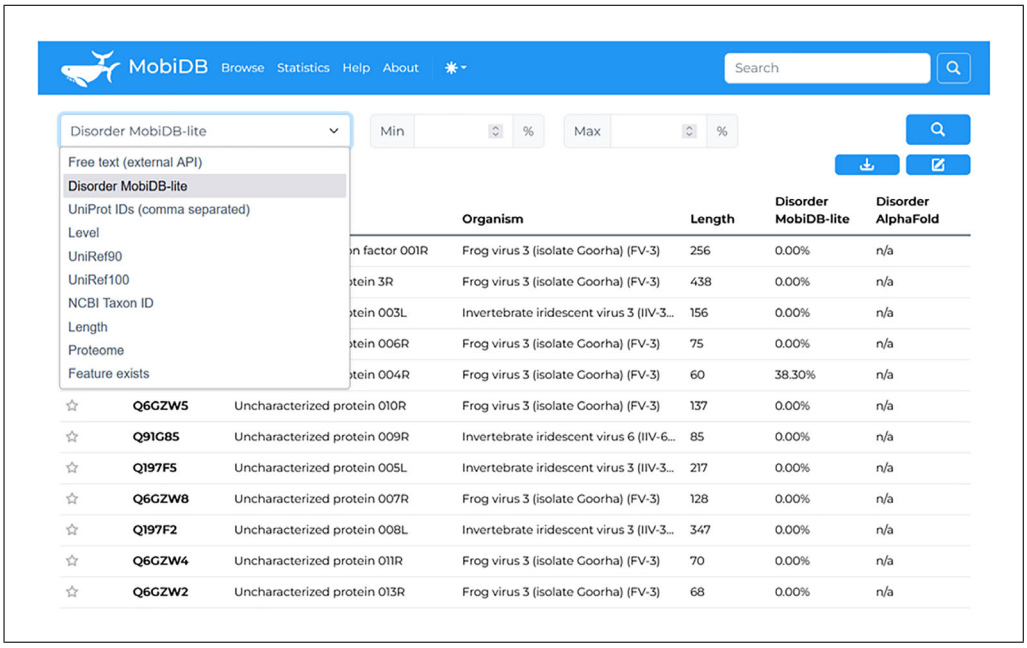

Figure 3 Browse page results.

available are "UniRef90" and "UniRef100" (clustering the sequences at 90% and 100% identity, respectively).

v. NCBI Taxon ID: To search using the NCBI Taxonomy Database identifier: for instance, "9606" to retrieve all *Homo sapiens* MobiDB entries.

vi. Length: To search for a specific protein length by selecting a minimum (min) and maximum (max) length.

vii. Proteome: To search for a specific proteome identifier consisting of "UP" followed by nine digits. For example, looking for "UP000006548" will retrieve all the entries corresponding to the Arabidopsis thaliana proteome.

viii. Feature exists: A set of features that can be associated with a protein.

ix. Free text (external API): Used to search for a term, e.g., VHL; in this case the search will be performed with UniProt APIs, and the UniProt identifiers obtained will be used to retrieve the data in MobiDB.

5. Users can customize the table columns to view more details about entries in the displayed results using the corresponding icon or download the search results using the button at the top-right of the Browse page (Fig. 3). File formats available for download are FASTA, JSON, and TSV.

SEARCHING MobiDB SELECTED DATASETS AND SELECTED PROTEOMES

BASIC PROTOCOL 2

The latest version of MobiDB presents "Selected datasets" and "Selected proteomes" on the MobiDB homepage (Figs. 4 and 5).

The "Selected datasets" are organized and displayed based on different confidence annotation levels (Gold and Silver stars) characterized by the type of "Evidence," which indicates the quality of the annotation (Table 1). Each dataset also includes information on "Features" (such as disorder and other supporting information), consensus sources, the number of proteins associated with the specific dataset, and the percentage of annotated residues (Fig. 4A).

As highlighted in Figure 4A, the two datasets are:

- Gold, representing all manually curated annotations and their homologs (Fig. 4B).
- Silver, representing all derived annotations (PDB) and AlphaFold disorder (Fig. 4C).

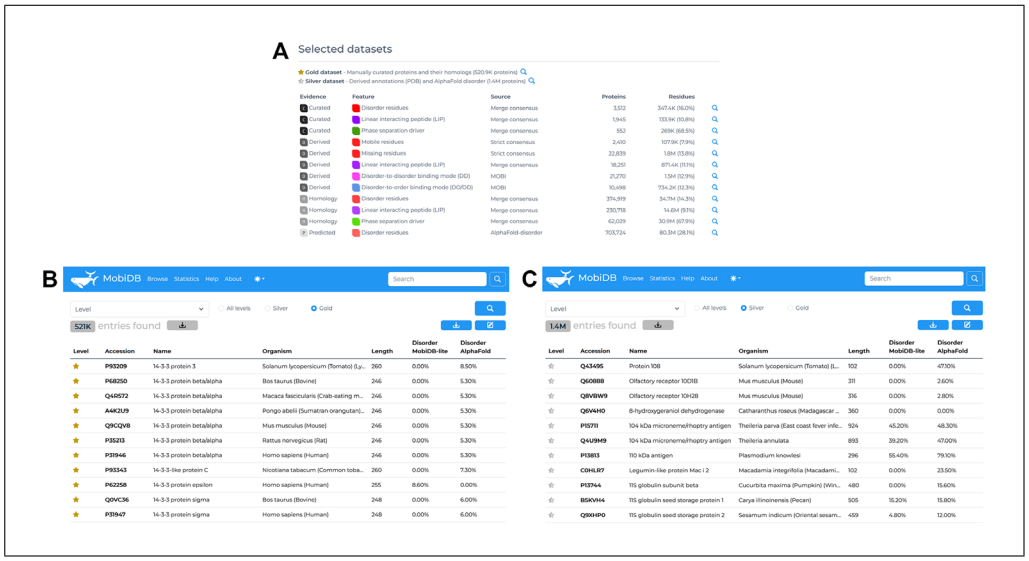

Figure 4 MobiDB homepage—Selected datasets section.

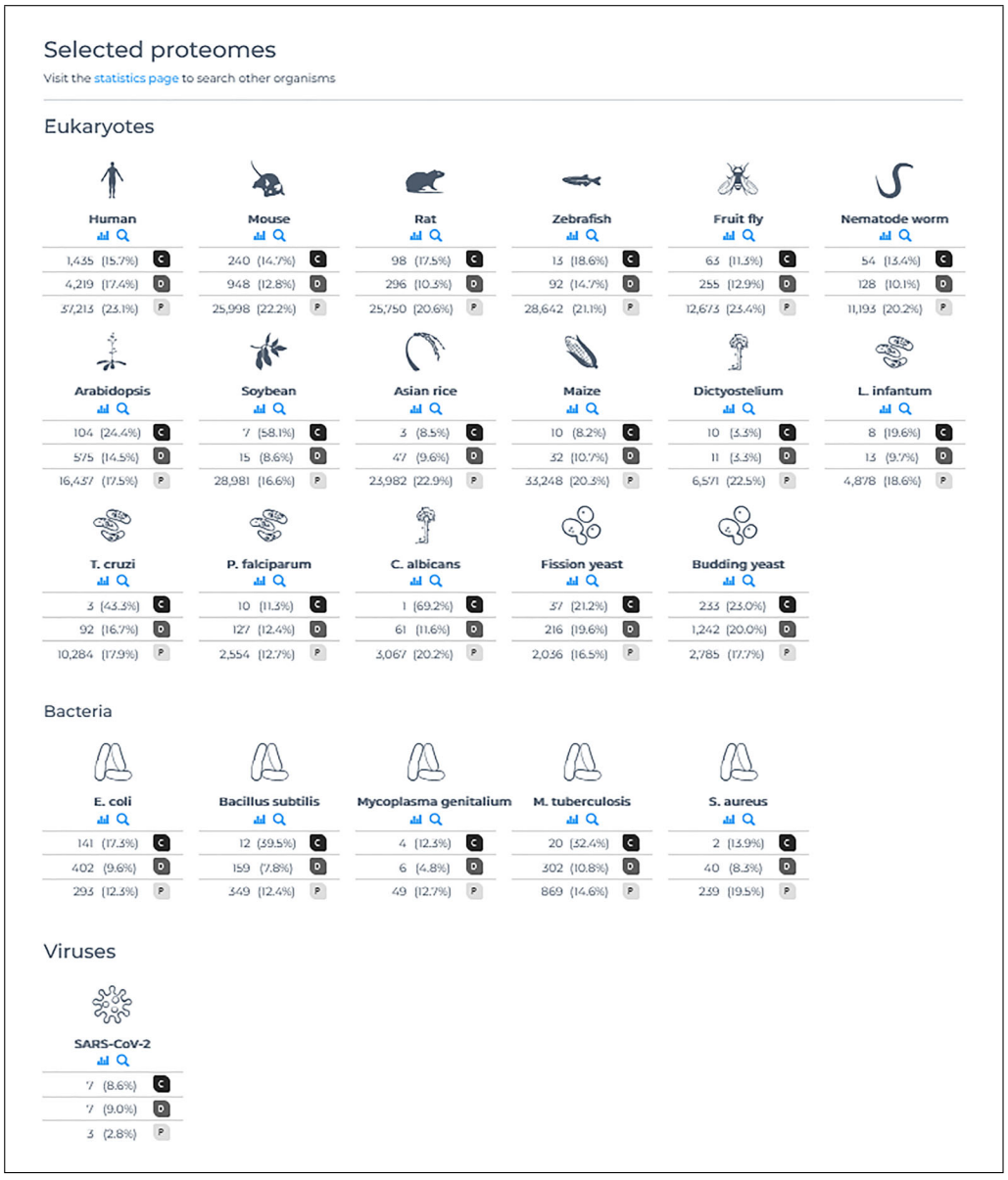

Figure 5 MobiDB homepage—Selected proteomes section.

Table 1 MobiDB Evidence Description

Evidence	Description
Curated (C)	Experimental annotation integrated from manually curated databases: e.g., DisProt and IDEAL for IDPs/IDRs (Aspromonte et al., 2024; Fukuchi et al., 2012) or liquid-liquid phase separation from PhasePro database (Mészáros et al., 2020)
Homology (H)	Annotations propagated by sequence-based homology inference
Predicted (P)	Predictions provided via other methods, sources, or software: e.g., MobiDB-lite software (Necci et al., 2017; Necci, Piovesan, Clementel et al., 2021).
Derived (D)	Experimental annotations derived from primary data: e.g., Missing Residues from PDB structure

The "Selected proteomes" are categorized into the three domains of living organisms: eukaryotes, bacteria, and viruses. For each domain and species (e.g., human), information about protein disorder is visible (Fig. 5).

The following steps describe how to explore and search proteins among the different "Selected datasets" or "Selected proteomes" within MobiDB.

Necessary Resources

Hardware

Laptop or desktop computer (best; a smartphone or tablet is an acceptable alternative)
Active and stable internet connection

Software

Internet browser: e.g., Firefox (*http://www.mozilla.org/firefox*), Google Chrome (*http://www.google.com/chrome*), or Safari (*http://www.apple.com/safari*)

Input data

None required

Performing a search on the MobiDB selected datasets

1a. Visit the MobiDB homepage at *https://mobidb.org/* using a browser.

2a. Scroll down the homepage to view the "Selected datasets" section of MobiDB, and select the dataset of interest using the magnifying glass (Fig. 4B or 4C).

3a. Users can search for protein in MobiDB based on the Curated (C), Derived (4), Homology (H), or Predicted (P) evidence, and search for features of interest.

The color of each icon reflects the Feature, Evidence Source, and Category type: Disorder (red), Linear Interacting Peptide (LIP; violet), Binding mode disorder to disorder (fuchsia), Binding mode disorder to order (light blue) and Phase separation (green). Detailed descriptions of all the types of Category, Feature, Evidence, and Source implemented in MobiDB are provided in the Help section "Controlled vocabulary" (https://mobidb.bio.unipd.it/help#vocabulary).

4a. The user can download or explore the data reported for each dataset:

i. Click on the JSON icon (at right) to download all the proteins included in that dataset in JSON format.
ii. Click on the TSV icon (at right) to download all the proteins included in that dataset in TSV format.
iii. Click on the magnifying glass icon (at right) to access the whole protein list of the dataset one and select the protein of interest.

Performing a search on the MobiDB selected proteomes

1b. Visit the MobiDB homepage at *https://mobidb.org/* using a browser.

2b. Scroll down the homepage to view the "Selected proteomes" section of MobiDB and select the proteome of interest (Fig. 5).

For each selected proteome, the number and the corresponding percentage (%) of proteins annotated as disordered, compared to the total number of proteins in the entire proteome of that species, are displayed, as well as the numbers and percentages specific to the three different types of evidence—Curated, Predicted, and Derived (Fig. 5).

3b. User have two options for exploring the data reported for each proteome (Fig. 5):

i. Click on the Statistic icon to explore the statistical values for all annotations included in that proteome.
ii. Click on the magnifying glass icon to access the complete list of all entries annotated for the selected species and select the one that is of interest.

BASIC PROTOCOL 3

PERFORMING A SEARCH ON THE STATISTICS PAGE IN MobiDB

The Statistics page provides all data about each type of "Evidence," "Feature," and "Source" for the proteins available in MobiDB. Moreover, it allows users to search for statistics of specific organisms and proteomes.

Necessary Resources

Hardware

Laptop or desktop computer (best; a smartphone or tablet is an acceptable alternative)
Active and stable internet connection

Software

Internet browser: e.g., Firefox (*http://www.mozilla.org/firefox*), Google Chrome (*http://www.google.com/chrome*), or Safari (*http://www.apple.com/safari*)

Input data

None required

1. Visit the MobiDB homepage at *https://mobidb.org/* using a browser.
2. Click the Statistics page (*https://mobidb.org/statistics*) (Fig. 1).
3. Users can perform a search by selecting either the "Free Text" or "Proteome" term option and applying filters based on the annotation level (All, Silver, Gold). Figures 6 and 7 show an example of how the list of proteins associated with the selected proteome appears when the Gold filter is applied, enabling users to sort the display according to their preferred order by interacting with the seven columns (Evidence, Feature, Source, Proteins, Fraction proteins, Residues, Fraction residues) and their corresponding sorting arrows.

i. **Free text:** Searching for a specific organism (e.g., Plasmodium), or for a broader set of organisms (e.g., Human virus), will retrieve all the entries corresponding to the organism of interest.
ii. **Proteome:** Searching for a specific proteome identifier (e.g., UP000006548), corresponding to proteome identifiers from UniProt ("UP" followed by nine digits), will retrieve all the entries corresponding to the proteome of interest.

For example, looking for UP000006548, a page about the Arabidopsis thaliana (mouse-ear cress) proteome highlights the "Total Proteins" (39.3K), "Total Residues" (16.6M), "Proteome ID" (UP000006548), and "NCBI taxon ID" (3702) (Fig. 7). Columns display the

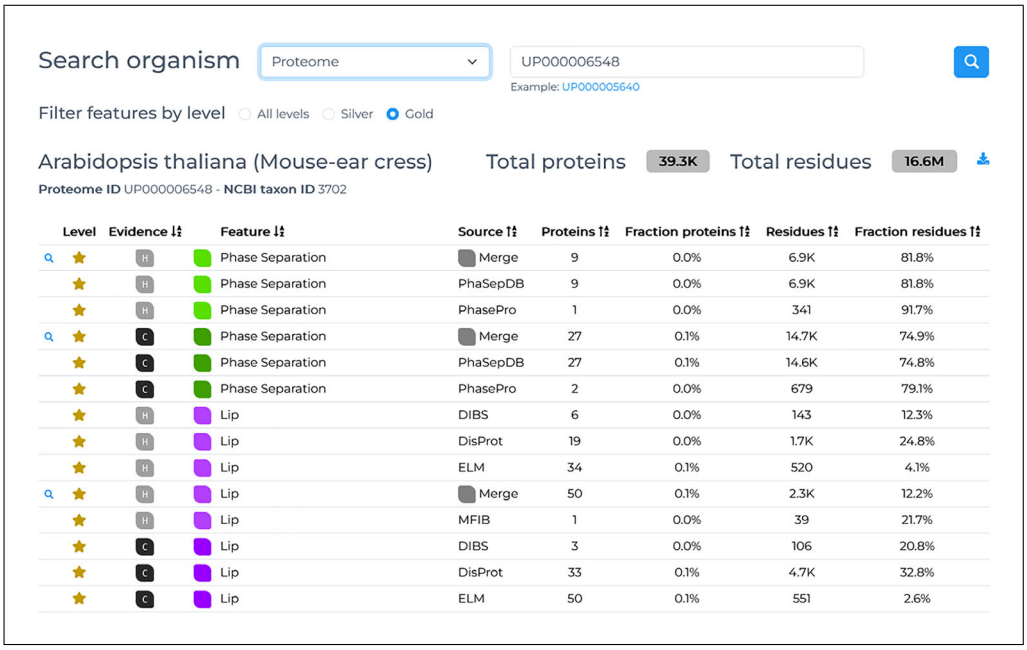

Figure 6 MobiDB Statistics page with an example search using the proteome ID:UP000006548 with the Gold filter applied.

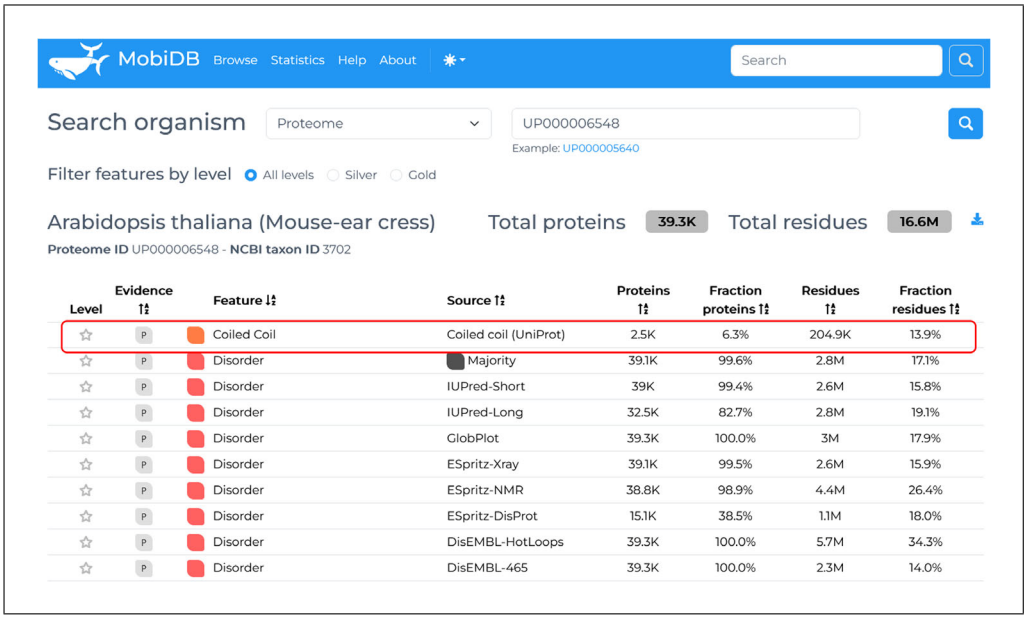

Figure 7 Example from the *Arabidopsis thaliana* proteome (UP000006548) search on the MobiDB Statistics page.

annotation confidence level (Silver or Gold) and the three essential elements characterizing each annotation (Evidence, Feature, and Source). For each level and triplet of elements, the database shows the number of proteins, the fraction of proteins (as a percentage, %), the number of residues, and the fraction of residues (%) (Fig. 7).

4. To allow users to explore the data more easily, the magnifying glass can be used to quickly access the list of proteins annotated as "Merge" (i.e., the consensus obtained when residues are annotated when at least one children is annotated). Specifically, for each "Feature" and "Evidence" type, the consensus and the corresponding annotated protein are provided (Fig. 7). A summary of the different "Features" and their corresponding "Evidence," along with the "Sources" from which the data were obtained, is provided in Table 2.

Table 2 Summary of the Different Features and Their Corresponding Evidence Types and Sources

Evidence	Feature	Source
Curated (C, green)	Binding Mode Context Dependent	AlphaFold, UniProt, DIBS
	Binding Mode Disorder to Disorder	AlphaFold, DisProt, ELM
	Binding Mode Disorder to Order	AlphaFold, DisProt
	Coiled Coil	Gene3D, UniProt
	Conformational Diversity	DisProt, MobiDB-lite
	Cysteine Rich	UniProt, IDEAL
	Disorder	IUPred, DisProt, UniProt
Derived (D, blue)	Domain	Pfam, Gene3D
	Glycine Rich	UniProt, IDEAL
	LIP	FuzDB, PhaSePro
	Low Complexity	SEG, GlobPlot
Homology (H, yellow)	Missing Residues	UniProt, AlphaFold
	Missing Residues Context Dependent	AlphaFold, CoDNaS
Prediction (P, red)	Mobile	MOBI, AlphaFold
	Mobile Context Dependent	AlphaFold, CoDNaS
	Negative Polyelectrolyte	DisProt, UniProt
	Observed	AlphaFold, DIBS
	Phase Separation	PhaSePro, FuzDB
	Plddt	AlphaFold
	Polar	DisProt, UniProt
	Polyampholyte	DisProt, UniProt
	Positive Polyelectrolyte	DisProt, UniProt
	Proline Rich	UniProt, IDEAL
	Signal Peptide	UniProt
	Transmembrane	UniProt, Pfam

Each "Feature" is annotated using multiple integrated resources, including AlphaFold, UniProt, and DisProt, among others, contributing to MobiDB's comprehensive annotation framework.

For example, in the Arabidopsis thaliana (mouse-ear cress; Fig. 7), the first line accounts for Predictions (P) for the Coiled Coil Feature based on UniProtKB as the source. There are a total of 2484 proteins, constituting 6.3% of fraction proteins, and 204,888 residues, representing 13.9% of fraction residues.

SUPPORT PROTOCOL

PROGRAMMATIC ACCESS WITH MobiDB REST API

MobiDB can be accessed programmatically through its APIs; refer to the API endpoints description page at *https://mobidb.org/help#swagger*. The download_page endpoint makes it possible to perform a search on the whole database according to multiple fields. Among them, a list of UniProt accession numbers can be specified, in order to retrieve specific entries. Similarly, the "count" endpoint allows you to determine the number in the database which matches a specific search query. Finally, "statistics" for the current release of the MobiDB can be retrieved to the endpoint named accordingly. Detailed descriptions of the endpoints, including their parameters, input, and output specifications, are provided below. https://mobidb.org/help#swagger

Necessary Resources

Hardware

Laptop or desktop computer (best; a smartphone or tablet is an acceptable alternative)
Active and stable internet connection

Software

Internet browser: e.g., Firefox (*http://www.mozilla.org/firefox*), Google Chrome (*http://www.google.com/chrome*), or Safari (*http://www.apple.com/safari*)

Input data

None required

1. **Search entries**

Users can retrieve multiple entries by performing a GET request to the */download_page* endpoint with various search parameters. Performing a request without any search parameter will retrieve at most 300,000 entries contained in the whole database. The more search parameters are defined, the stricter the search will be and the fewer entries will be retrieved. The endpoint makes it possible to perform a projection on data to be returned, hence defining the fields retrieved from each entry. This reduces the size of the response. The output format can be specified through the "format" parameter. The default format is JSON, but TSV or FASTA are also available. Specific entries can be retrieved by setting their UniProt accession number in the "acc" field.

The sample code below retrieves entries whose UniProt accession number is either P04637, P10912, or P0DTC9.

```
#!/usr/bin/env python3


import requests
# Define list of accession numbers
acc_list = ['P04637', 'P10912', 'P0DTC9']
# Define request URL
request_url = 'https://mobidb.org/api/download_page'
# Perform GET request, retrieve JSON-formatted response
json_response = requests.get(request_url, params={
'acc': ','.join(acc_list), # Concatenate list of UniProt accession numbers
'format': 'json' # Define output format
})
# Print JSON-formatted response
print(json_response.json()['data'])
```

The following is similar to the example above, except that it searches for a single entry, P04637, by its own UniProt accession number and retrieves it as such:

```
#!/usr/bin/env python3
import requests
# Define single UniProt accession number
uniprot_acc = 'P04637'
# Define request URL
request_url = 'https://mobidb.org/api/download_page'
# Perform GET request, retrieve JSON-formatted response
json_response = requests.get(request_url, params={
```

```
'acc': uniprot_acc, # Define UniProt accession number
'format': 'json' # Define output format
})
# Print JSON-formatted entry
print(json_response.json()["data"].pop())
```

2. **Search results**

   The download_page endpoint makes it possible to specify three different values for the "format" parameter in the GET request, which handles the return format itself. JSON is the default format; in this case, the response will be a list of JSON-formatted objects representing the filtered entries or pieces of them, according to the "projection" parameter. Similarly, if TSV is chosen, a TSV-formatted table will be returned, and if FASTA is chosen, the response will be a FASTA-formatted text for each filtered entry, associated to a UniProt entry through its accession number.

3. **Count**

   The "count" endpoint allows you to search the whole database in the same way the download_page endpoint does. However, this endpoint returns only the number of results matching the query parameters.

   The example code below demonstrates how to perform a text search and return the number of entries which match such a query.

```
#!/usr/bin/env python3
import requests
# Define searched text
search_text = 'Disorder'
# Define request URL
request_url = 'https://mobidb.org/api/count'
# Perform GET request, retrieve response object
json_response = requests.get(request_url, params={
'free_text': search_text # Define text query
})
# Print number of retrieved entries
print('Retrieved %d entries for term "%s"' % (json_response.json()["metadata"].get('count', 0),
search_text))
```

4. **Count results**

   The "count" endpoint returns a JSON-formatted object, containing a single parameter *n*. Its value defines the number of entries matched by the given input query.

5. **Statistics**

   The "statistics" endpoint can be accessed through the GET method and provides information about the data included in the latest MobiDB release. It is possible to retrieve statistics for each database partitions linked to one or more proteomes by specifying their identifiers in the "proteome" parameter of the GET request.

   The example code below demonstrates how to retrieve statistics for the "Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)" and the "Homo Sapiens (Human)" proteome, whose identifiers are UP000464024 and UP000005640, respectively.

```
#!/usr/bin/env python3
import requests
# Define list of proteome identifiers
proteomes_list = ['UP000464024', 'UP000005640']
```

```
# Define request URL
request_url = 'https://mobidb.org/api/statistics'
# Perform GET request, retrieve response object
response_obj = requests.get(request_url, params={
'proteome': ','.join(proteomes_list) # Concatenate list of proteome identifiers
})
# Get JSON-formatted response
json_response = response_obj.json()
# Print number of retrieved entries
print('Retrieved statistics for %d proteomes' % len(json_response))
# Loop through each proteome statistics entry
for i in range(len(json_response)):
# Get identifier of current proteome
print('%d-th proteome is %s' % (i+1, json_response[i]['proteome']))
# Get one of the statistics
print('Proteome %s contains %d proteins, which make up to %d residues' % (json_response[i]
['proteome'],json_response[i]['total_proteins'], json_response[i]['total_residues']))
```

6. **Statistics results**

   The "statistics" endpoint returns a JSON-formatted response consisting of an array of statistics entities for each of the proteomes whose identifier has been set in the GET query parameters. If no proteome was set, statistics are retrieved for all proteomes within the MobiDB. Each statistics entity is an object whose properties define statistics computed on the partition of the MobiDB associated with a proteome. Among those, the "features" property is an array of statistics computed for each feature available in the database.

VISUALIZING AND INTERPRETING A MobiDB ENTRY: THE GHR USE CASE

BASIC PROTOCOL 4

The MobiDB protein entry page, featuring examples such as the *Homo sapiens* growth hormone receptor (GHR), accessible at *https://mobidb.org/P10912*, offers comprehensive information, including the protein description, sequence, and all disorder-related predictions and annotations.

Using GHR as example, we illustrate the organizational structure common to all entries in MobiDB. At the top of the entry page, users can access general protein information with Features Content, Feature viewer, and AlphaFold structure visualization. Two other sections, Subcellular Locations and Cross References, are available at the bottom of the page (Fig. 8).

In the middle of the page, users will find the AlphaFold structure, the corresponding protein sequence, and a visualization of the annotations. The annotations dynamically adjust according to the selected tabs. Users can choose an "Overview" of the annotations or explore more detailed information related to "Disorder," "Binding," "Interactions," and "Functions."

Necessary Resources

Hardware

- Laptop or desktop computer (best; a smartphone or tablet is an acceptable alternative)
- Active and stable internet connection

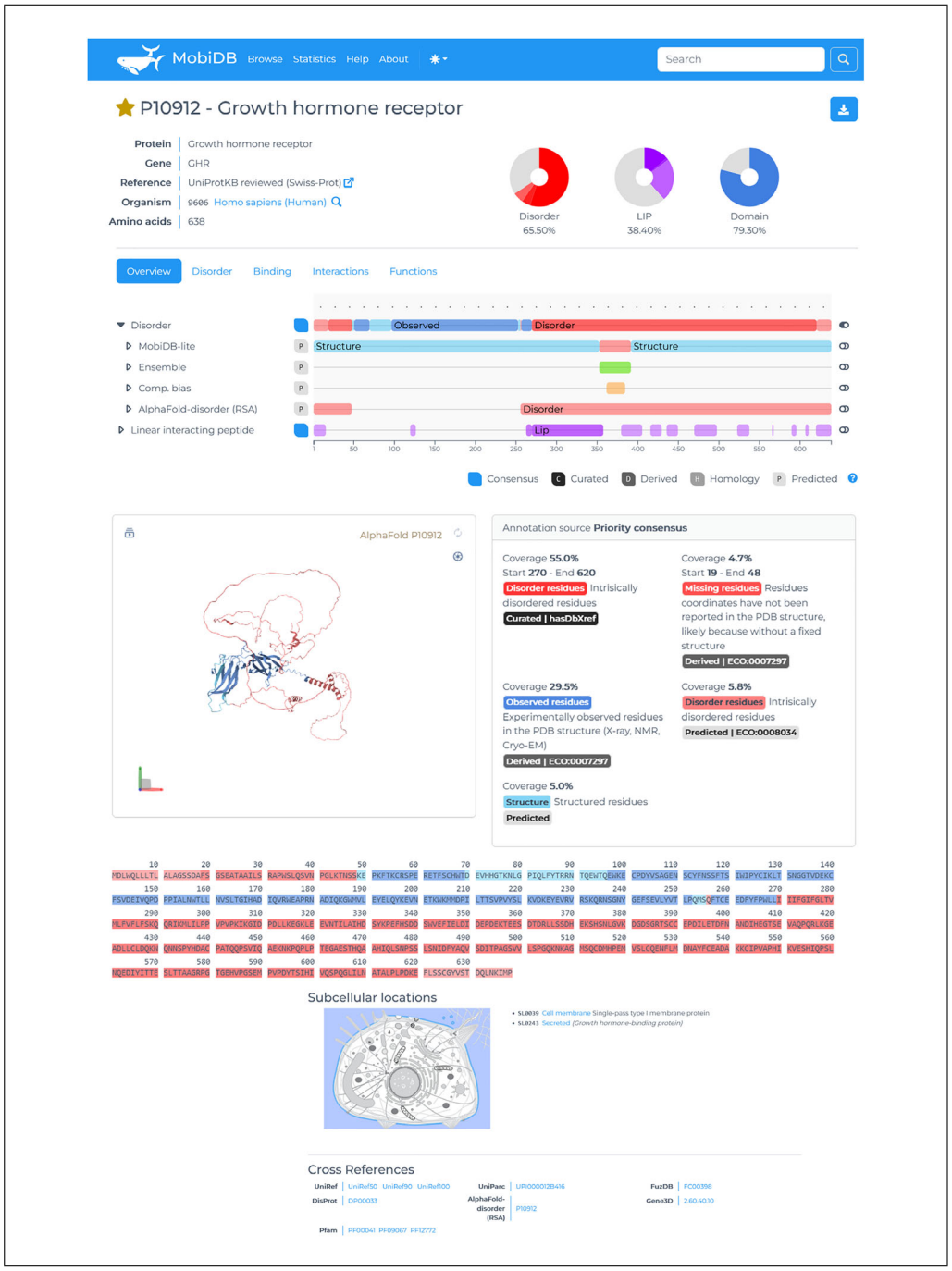

Figure 8 GHR entry page in MobiDB—Overview tab.

Software

Internet browser: e.g., Firefox (*http://www.mozilla.org/firefox*), Google Chrome (*http://www.google.com/chrome*), or Safari (*http://www.apple.com/safari*)

Input data

None required

1. Visit the MobiDB homepage at *https://mobidb.org/* and choose the entry for growth hormone receptor (GHR, P10912). Alternatively, you can find it directly at *https://mobidb.org/P10912*. At the top of the page, you will find a protein summary on the left side, while the right side displays the content of Disordered (red), LIP (violet), and Domain (blue) residues, with circles colored differently based on the type of the

feature shown. Clicking on the differently colored circles (light, dark) allows the user to specifically view the number of Disordered, LIP, and Domain residues related to Curated and Predicted evidence.

2. As you scroll down, you can navigate through the feature viewer with detailed visualization of annotations, the AlphaFold structure, the annotation source, and the sequence. To the right of each row in the feature viewer, it is possible to view the details of each annotation, which appear in the "Annotation Source" box next to the AlphaFold structure (Fig. 8). These elements change based on the tab selected (Overview, Disorder, Binding, Interactions, or Functions).

   i. Clicking on the "Overview" tab will display the AlphaFold structure and sequence, showcasing all the colors of the annotations. The Feature viewer will show disorder consensus, disorder and Ensemble MobiDB-lite predictions, AlphaFold disorder predictions, and the Consensus of LIP.

   ii. Selecting the "Disorder" tab provides a detailed Feature viewer of all disorder annotations in the GHR protein. Annotations can originate from "Curated" (C) databases (e.g., DisProt), "Derived" (D) data annotated as missing residues (e.g.,

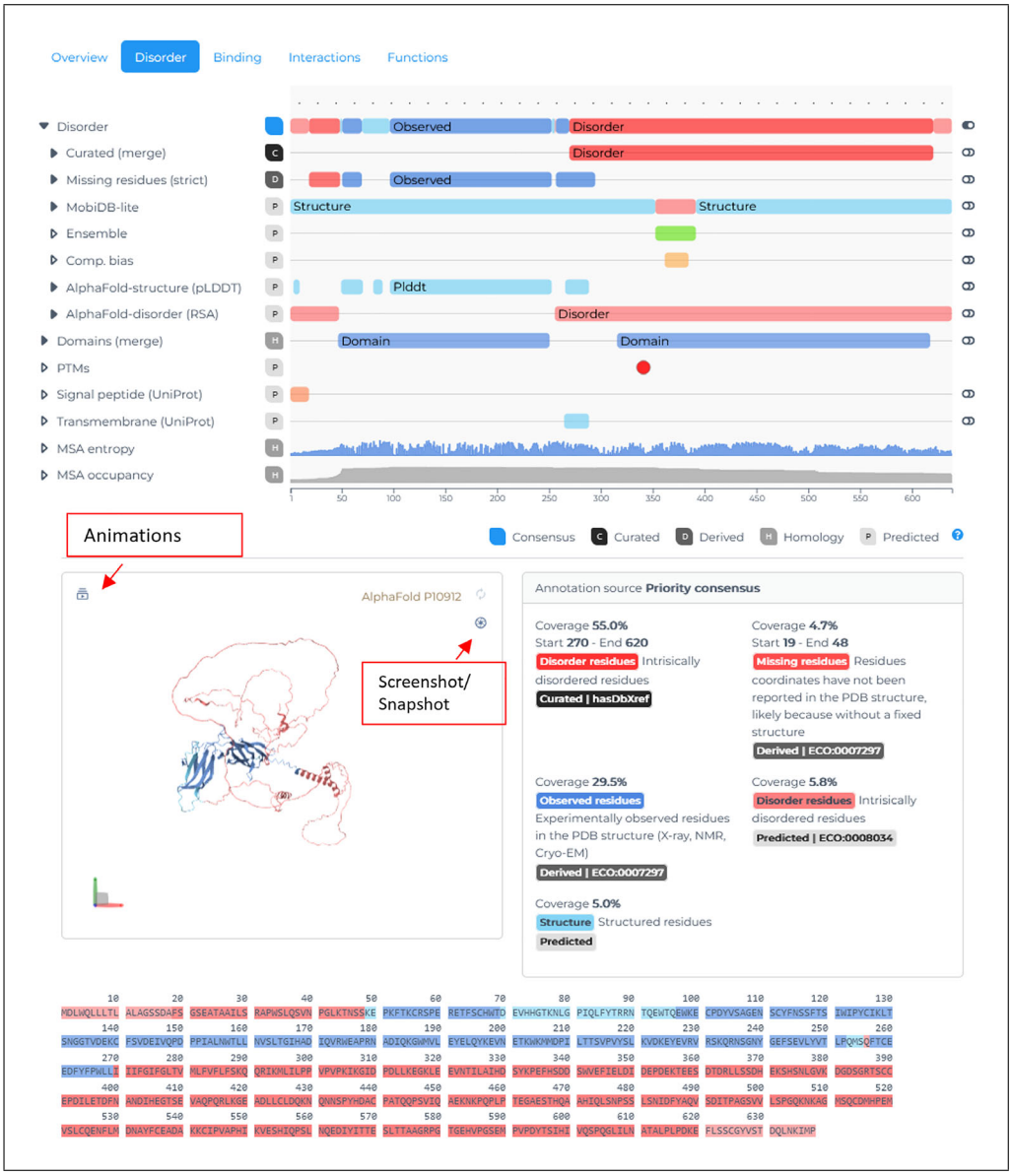

Figure 9 GHR entry page in MobiDB—Disorder tab and Annotation source.

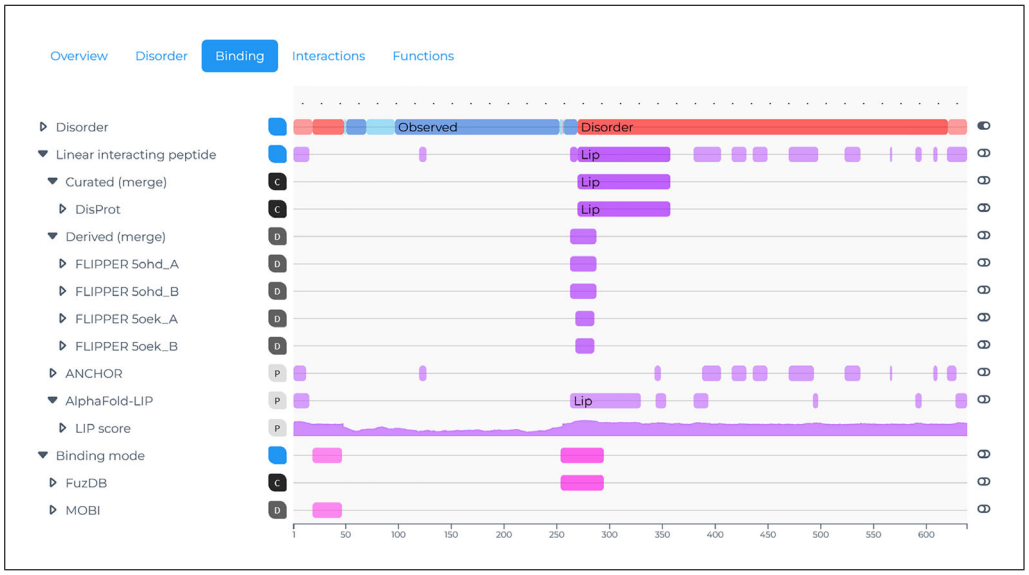

Figure 10 GHR entry page in MobiDB—Interactions tab.

PDB structures), "Predicted" (P) sources (e.g., MobiDB-lite, AlphaFoldDB, etc.), and finally "Homologs" (H). Additional information is provided for Post Translational Modifications (PTMs), Domains (merge), MSA entropy, and MSA occupancy (Fig. 9).

iii. Clicking on the "Binding" tab causes the Feature viewer to provide "Linear interacting peptide" information for the three Evidence types and disordered regions that undergo binding (Fig. 10).

iv. Clicking on the "Interactions" tab provides a detailed Feature viewer that highlights interaction regions and corresponding interactors extracted from PDB complexes using the RING software.

v. Clicking on the "Functions" tab in the Feature viewer displays curated or predicted annotations related to disorder functions. For example, in the case of GHR, MobiDB provides information about lipid binding as disorder function for protein residues 270-357, annotated with GO term, and includes flexible/linker prediction annotated with an IDPO term.

In all Feature viewers, clicking on the arrow next to each feature (e.g., Disorder) will open another Feature viewer offering detailed data.

3. Users can perform various actions on the protein structure, such as structure animation or taking screenshots or state snapshots, as well as modifying settings, including background and lighting (Fig. 9).

4. Priority consensus of the annotation sources is provided for each type of feature (e.g., disorder residues, observed residues, missing residues, etc.), along with the coverage and the type of evidence (Fig. 9).

COMMENTARY

Background Information

MobiDB is a core resource of the IDP Central Consortium and a service of ELIXIR IDP Community (Davey et al., 2019). It was designed to provide structural and functional information about IDPs and IDRs combining annotations from the literature, experimental data, and predictions from multiple resources. MobiDB celebrated ten years of existence as of our previous publication (Piovesan et al., 2022), and for this latest version (v. 6), the updates were designed to enable users to process more data, make the tool faster, and include additional functional features. The recent updates focus on enhancing the database's ability to predict protein ensemble properties, such as conformational diversity and compactness, which are essential

for understanding the structural behavior of disordered regions. Moreover, functional annotations (GO and IDPO) from databases like DisProt are propagated using sequence similarity and protein embeddings, which expand MobiDB's coverage of functional predictions.

In MobiDB, all disorder predictions are generated with MobiDB-lite, a distinct open source component of MobiDB. It was developed in collaboration with the InterPro (Paysan-Lafosse et al., 2023) team and is available at *https://github.com/BioComputingUP/MobiDB-lite*. It integrates most of the other prediction methods and is used as consensus. Integration pipelines in MobiDB also include FLIPPER (Monzon et al., 2021) and the Residue Interaction Network Generator (RING; Del Conte et al., 2024) software for predicting and characterizing LIP in the PDB and noncovalent interactions at the atomic level in protein structures, respectively. The newly designed entry pages and improved visualization tools allow users to explore the disorder annotations more intuitively, making it easier to access specific disorder-related features. To provide a comprehensive description of the MobiDB resource, we created training material in collaboration with other consortia such as ELIXIR and PhaseAge. MobiDB remains a critical resource for researchers investigating the roles of disordered proteins in cellular processes and diseases, contributing to a better understanding of the dynamic nature of IDRs in various biological contexts. The MobiDB homepage allows users to search for proteins using a variety of query formats. Whether the search term is a protein or an organism, users can enter data such as the UniProt accession number, protein name, gene symbol, organism name, or NCBI taxonomy ID into the Search box to generate relevant results. Moreover, the MobiDB home page provides direct access to "Selected datasets" and "Selected proteomes."

Additional information about MobiDB content is available from the Statistics page (*https://mobidb.org/statistics*), including the various types of evidence, features, and sources related to the proteins it catalogs. Additionally, the Help page (*https://mobidb.org/help*) provides a comprehensive guide to MobiDB's key features and functionalities, its structured vocabulary, and how to access its services programmatically. It also includes tutorials on navigating the MobiDB website and utilizing its API for further learning.

Critical Parameters

General considerations

The main factors influencing the effectiveness of the MobiDB resource include the accuracy of computational prediction models, the completeness and reliability of experimental data, and the integration methods used to consolidate diverse data types.

Advanced parameters

MobiDB offers the possibility of customizing search queries and results views based on specific needs or interests, including filtering data by prediction method, disorder type, or specific protein regions. This allows more targeted analyses and can be helpful when performing research that require specific subsets of the database.

Troubleshooting

For a list of potential problems, their causes, and their solutions, please refer to Table 3.

Guidelines for Understanding Results

Over its 10-year history, MobiDB (*https://mobidb.org/*) has evolved into a comprehensive knowledgebase offering a rich repository of aggregated disorder annotations derived from the literature and experimental evidence, alongside predictions for all available protein sequences. Users of MobiDB can access a wide assortment of selected datasets and selected proteomes directly from the home page of the resource. Selected datasets comprise four types of evidence—curated (C), derived (D), homology (D), and predicted (P)—and a variety of features, including disorder, mobile and missing residues, linear interacting peptides (LIP), and disorder-to-disorder (DD) and disorder-to-order (DO) binding modes, from different sources. The selected proteome set provides curated (C), derived (D), and predicted (P) annotations for the most prominent organisms among the eukaryotes, bacteria, and viruses. Additional organisms, as well as their annotations, can be found by searching the Statistics page of MobiDB (*https://mobidb.bio.unipd.it/statistics*). MobiDB now integrates AlphaFoldDB predictions, filling disorder-annotation gaps for those sequences for which experimental coordinates are lacking, and transfers manually curated annotations through its newly implemented homology transfer pipeline. An updated entry page design provides an enhanced overview of the structural disorder annotations and functions associated with its

Table 3 Sources of and Solutions to Potential Errors

Problem	Possible cause	Solution
Inconsistent annotations	Variation in prediction algorithms	Compare annotations from different sources
Missing data	Limited experimental evidence	Look for recent publications or alternative databases
Discrepancies in data	Updates not yet reflected	Check for the latest version of the database
Difficulty interpreting output data	Lack of familiarity with the data format	Refer to the MobiDB tutorials and documentation

binding modes. Users interested in a in-depth exploration of MobiDB and its annotations can find all the essential information about programmatic access, controlled vocabulary, and annotation identifiers, along with user tutorials, on the dedicated Help page (*https://mobidb.bio.unipd.it/help*). Finally, details on citation, governance, funding, licensing, and policies are available on the About page of MobiDB (*https://mobidb.bio.unipd.it/about*).

Acknowledgments

The MobiDB is part of the ELIXIR infrastructure (*https://elixir-europe.org/*) and a service of the IDP Community (*https://elixir-europe.org/communities/intrinsically-disordered-proteins*). This publication is part of a project that has received funding from and funded by the European Union. Horizon Europe MSCA Staff Exchange project IDPfun2, grant agreement no. 101182949 (IDPfun2). Disclaimer: Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. This publication is partially based upon work from COST Action ML4NGP (CA21160), supported by COST (European Cooperation in Science and Technology). Moreover, it received funding from the European Union through the NextGenerationEU PNRR project "ELIXIRxNextGenIT" (IR0000010), PNRR-PRIN 2022 project "PLANS" (2022W93FTW) and the PNRR project CN-G.T.RNA SP. 7 project–National Center for Gene Therapy and Drugs based on RNA Technology (CN3)–CN00000041.

Open access publishing facilitated by Universita degli Studi di Padova, as part of the Wiley - CRUI-CARE agreement.

Author Contributions

Maria Cristina Aspromonte: Conceptualization; formal analysis; investigation; supervision; writing—original draft. **Federica Quaglia:** Conceptualization; formal analysis; writing—original draft. **Alexander Monzon:** Software; conceptualization; formal analysis; **Damiano Clementel:** Software; writing—original draft. **Alessio Del Conte:** Software; writing—original draft. **Damiano Piovesan:** Conceptualization; funding acquisition; project administration; writing—original draft. **Silvio Tosatto:** Conceptualization; funding acquisition; investigation; project administration; writing—original draft.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the protocol are openly available in MobiDB at *https://mobidb.org/*.

Literature Cited

Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Gutmanas, A., Anyango, S., Choudhary, P., Clark, A. R., Dana, J. M., Deshpande, M., Dunlop, R., Gane, P., Gáborová, R., Gupta, D., Haslam, P., Koča, J., Mak, L., Mir, S., Mukhopadhyay, A., Nadzirin, N., … Velankar, S. (2020). PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Research*, *48*(D1), D335–D343. https://doi.org/10.1093/nar/gkz990

Aspromonte, M. C., Nugnes, M. V., Quaglia, F., Bouharoua, A., DisProt Consortium, Tosatto, S. C. E., & Piovesan, D. (2024). DisProt in 2024: Improving function annotation of intrinsically disordered proteins. *Nucleic Acids Research*, *52*(D1), D434–D441. https://doi.org/10.1093/nar/gkad928

Davey, N. E., Babu, M., Blackledge, M., Bridge, A., Capella-Gutierrez, S., Dosztanyi, Z., Drysdale, R., Edwards, R. J., Elofsson, A., Felli, I. C., Gibson, T. J., Gutmanas, A., Hancock, J. M., Harrow, J., Higgins, D., Jeffries, C. M., le Mercier, P., Mészáros, B., Necci, M., … Tosatto, S. C. E. (2019). An intrinsically disordered proteins community for ELIXIR (no. 8:1753). *F1000Research*, *8*, ELIXIR–1753. https://f1000research.com/articles/8-1753/v1

Del Conte, A., Camagni, G. F., Clementel, D., Minervini, G., Monzon, A. M., Ferrari, C.,

Piovesan, D., & Tosatto, S. C. E. (2024). RING 4.0: Faster residue interaction networks with novel interaction types across over 35,000 different chemical structures. *Nucleic Acids Research*, *52*(W1), W306–W312. https://doi.org/10.1093/nar/gkae337

Del Conte, A., Mehdiabadi, M., Bouhraoua, A., Mozon, A. M., Tosatto, S. C. E., & Piovesan, D. (2023). Critical assessment of protein intrinsic disorder prediction (CAID)—results of round 2. *Proteins: Structure, Function, and Bioinformatics*, *91*, 1925–1934. https://doi.org/10.1002/prot.26582

Fukuchi, S., Sakamoto, S., Nobe, Y., Murakami, S. D., Amemiya, T., Hosoda, K., Koike, R., Hiroaki, H., & Ota, M. (2012). IDEAL: Intrinsically disordered proteins with extensive annotations and literature. *Nucleic Acids Research*, *40*(D1), D507–D511. https://doi.org/10.1093/nar/gkr884

Gene Ontology Consortium, Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., … Westerfield, M. (2023). The gene ontology knowledgebase in 2023. *Genetics*, *224*, iyad031. https://doi.org/10.1093/genetics/iyad031

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 7873. https://doi.org/10.1038/s41586-021-03819-2

Martin, E. W., Holehouse, A. S., Peran, I., Farag, M., Incicco, J. J., Bremer, A., Grace, C. R., Soranno, A., Pappu, R. V., & Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science*, *367*(6478), 6478. https://doi.org/10.1126/science.aaw8653

Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O. P., Kovács, M., Tosatto, S. C. E., Tompa, P., Dosztányi, Z., & Pancsa, R. (2020). PhaSePro: The database of proteins driving liquid-liquid phase separation. *Nucleic Acids Research*, *48*(D1), D1. https://doi.org/10.1093/nar/gkz848

Monzon, A. M., Bonato, P., Necci, M., Tosatto, S. C. E., & Piovesan, D. (2021). FLIPPER: Predicting and characterizing linear interacting peptides in the Protein Data Bank. *Journal of Molecular Biology*, *433*(9), 166900. https://doi.org/10.1016/j.jmb.2021.166900

Necci, M., Piovesan, D., Clementel, D., Dosztányi, Z., & Tosatto, S. C. E. (2021). MobiDB-lite 3.0: Fast consensus annotation of intrinsic disorder flavors in proteins. *Bioinformatics*, *36*(22–23), 5533–5534. https://doi.org/10.1093/bioinformatics/btaa1045

Necci, M., Piovesan, D., Dosztányi, Z., & Tosatto, S. C. E. (2017). MobiDB-lite: Fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*, *33*(9), 1402–1404. https://doi.org/10.1093/bioinformatics/btx015

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

Piovesan, D., Del Conte, A., Clementel, D., Monzon, A. M., Bevilacqua, M., Aspromonte, M. C., Iserte, J. A., Orti, F. E., Marino-Buslje, C., & Tosatto, S. C. E. (2022). MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Research*, *51*(D1), D438–D444. https://doi.org/10.1093/nar/gkac1065

Piovesan, D., Del Conte, A., Mehdiabadi, M., Aspromonte, M. C., Blum, M., Tesei, G., von Bülow, S., Lindorff-Larsen, K., & Tosatto, S. C. E. (2024). MobiDB in 2025: Integrating ensemble properties and function annotations for intrinsically disordered proteins. *Nucleic Acid Research*, *gkae969*. https://doi.org/10.1093/nar/gkae969

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., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., & Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews*, *114*(13), 6589–6631. https://doi.org/10.1021/cr400525m

Venati, S. R., & Uversky, V. N. (2024). Exploring intrinsic disorder in human synucleins and associated proteins. *International Journal of Molecular Sciences*, *25*(15), 15. https://doi.org/10.3390/ijms25158399