



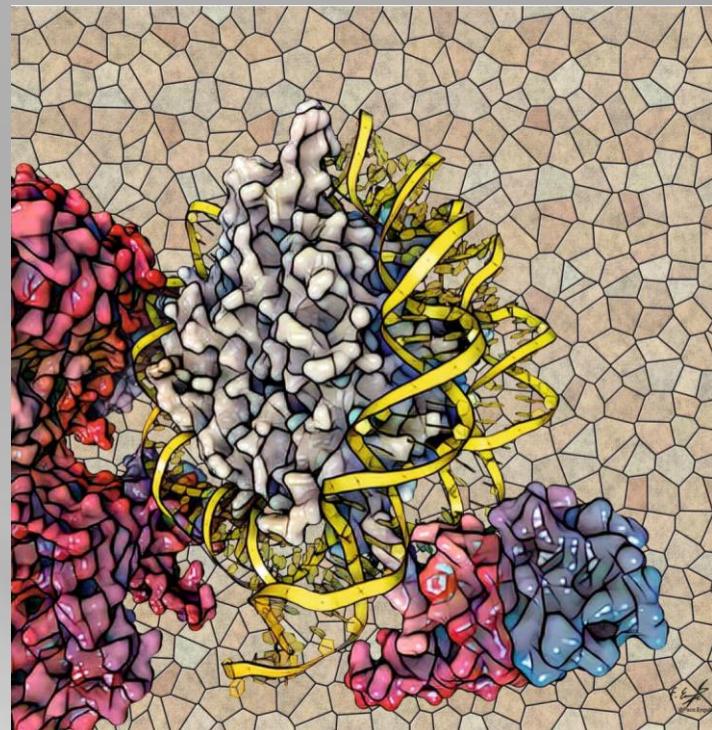
L.B.B

Laboratory of Systems Biology
and Bioinformatics

Introduction to Protein Databases: Part1

Fereshteh Noroozi

Fall 2023



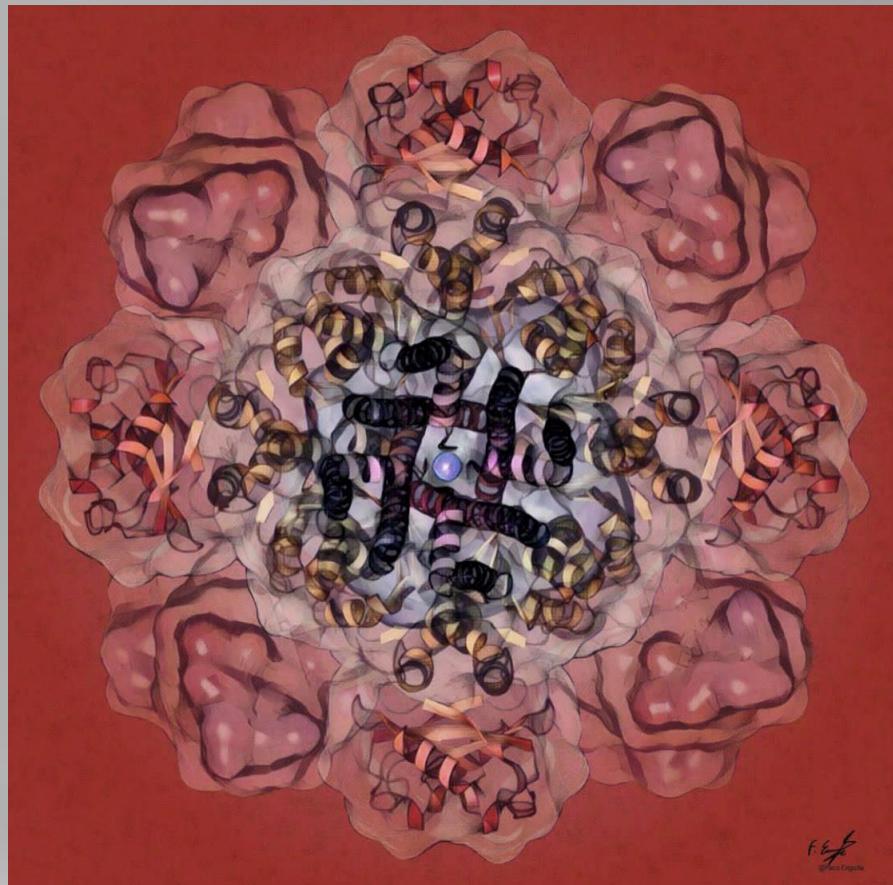


L.B.B

Laboratory of Systems Biology
and Bioinformatics

Importance of protein databases in research

- Large-scale protein information
- Accelerating research
- Functional annotation
- Structure prediction and modeling
- Comparative analysis
- Data-driven discoveries
- Bioinformatics tools and resources



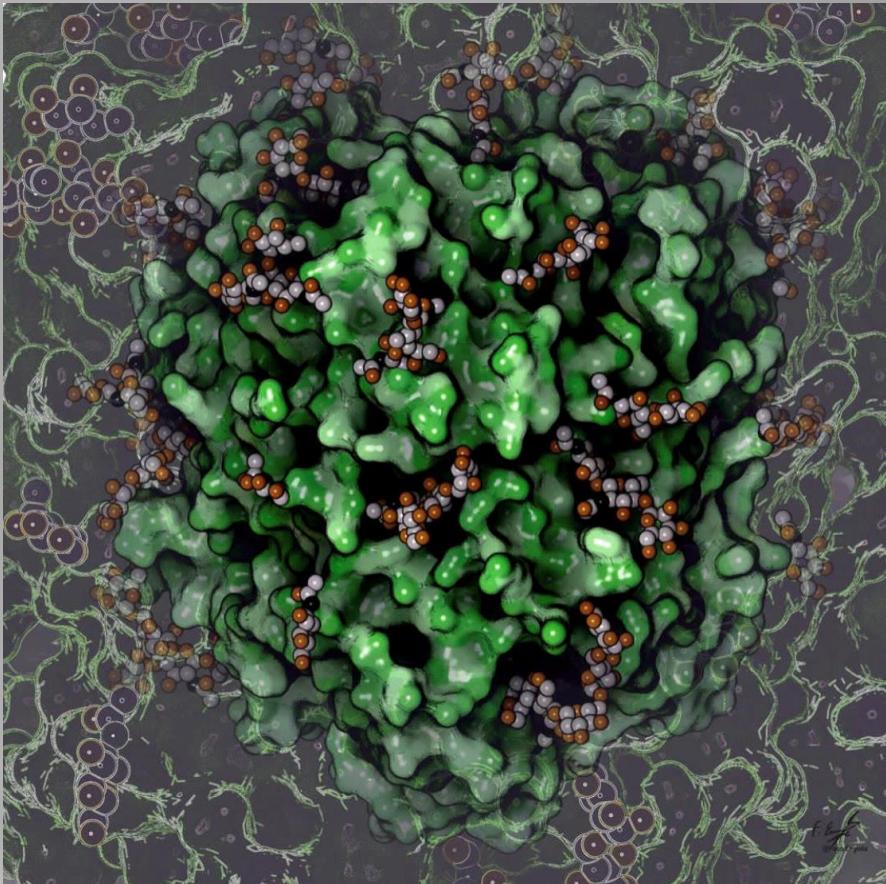


L.B.B

Laboratory of Systems Biology
and Bioinformatics

Key resources

- TmAlphaFold
- BioGRID
- HIPPIE
- dcGO
- MatrisomeDB
- DEPICTER2
- DescribePROT





L.S.B.

Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

TmAlphaFold Transmembrane Protein Structure Database

[Home](#) [Usage](#) [Method](#) [FAQs](#) [Statistics](#) [API](#) [Downloads](#)

TmAlphaFold Transmembrane Protein Structure Database

Developed by Protein Bioinformatics Research Group, RCNS

Search for protein, gene, UniProt accession or organism

[Search](#)

Examples: [tmem](#) [sico3al](#) [P33527](#) [help](#)

[Help](#)

[License](#)

[Feedback](#)

[Disclaimer](#)

TmAlphaFold database provides open access to the membrane orientation of 215844 alpha-helical transmembrane





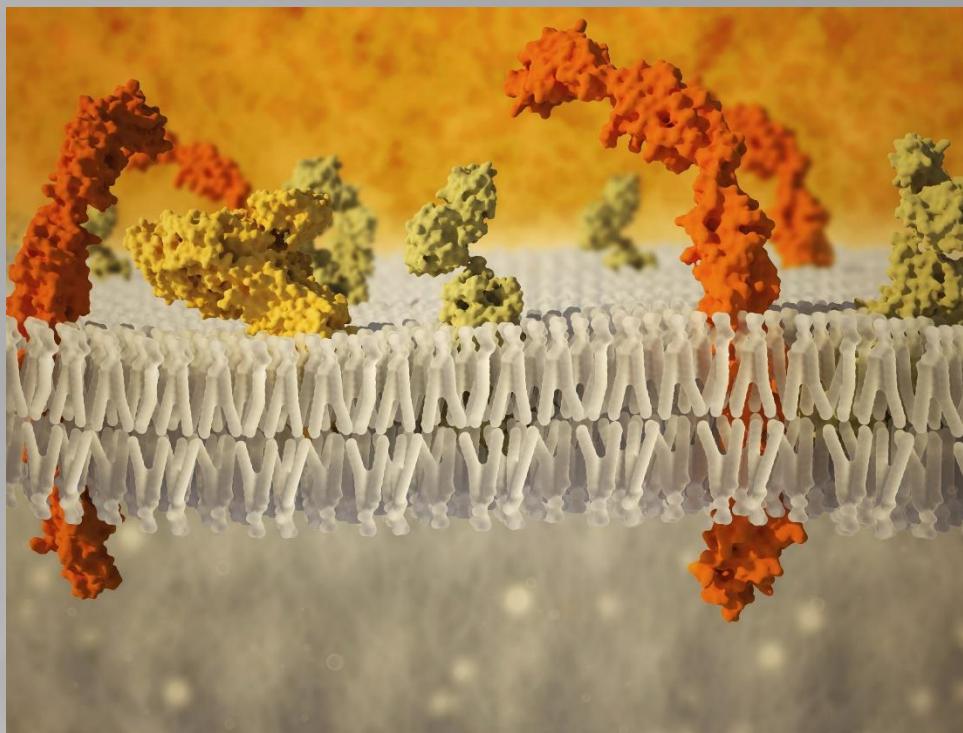
L.B.B

Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

Transmembrane proteins

- Cell membrane
- Molecular transport
- Signal transmission
- Structural integrity
- Lipid bilayer
- Cellular processes





L.B.B

Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

Search modes:

- **Gene name**
- **UniProt Accession**
- **UniProt ID**
- **Species**

Examples: [tmem](#) [slco3a1](#) [P33527](#) [help](#) [Help](#)





Search results:

- Protein name
- Gene name
- Organism
- UniProt AC (UniProt Accession)
- AlphaFold Database hyperlink
- PDB-EBI hyperlink
- Filter options (organism, quality of the structure, topology prediction evidence, number of transmembrane helices)
- Navigation bar (on the left side) for further filtering and navigation

TmAlphaFold

Showing search results for tmem

Showing results 1..20 of 44

Filters:

- Organism
 - Arabidopsis thaliana (1)
 - Caenorhabditis elegans (?)
 - Danio rerio (4)
 - Drosophila melanogaster (1)
 - Plasmodium falciparum (3)
 - Saccharomyces cerevisiae (1)
 - Dictyostelium discoideum (1)
- Evaluation
 - Excellent (10)
 - Good (14)
 - Fair (9)
 - Poor (1)
- # Number of transmembrane segments
 - 1 (14)
 - 2 (3)
 - 3 (5)
 - 4 (9)
 - 5 (2)
 - 6 (2)
 - 7 (1)
- CCTOP Evidence levels
 - SD (0)
 - TOPDOM (1)
 - Exists (2)
 - Prediction (41)

TMEM (Human TransMEMbrane protein) homolog (O01870_CAEEL)

Info: Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Plasmodium falciparum, Saccharomyces cerevisiae, Dictyostelium discoideum
Links: UniProt, UniProt_CAEEL, EMBL-EBI, UniProt_CAEEL

TMEM (Human TransMEMbrane protein) homolog (O76687_CAEEL)

Info: Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Plasmodium falciparum, Saccharomyces cerevisiae, Dictyostelium discoideum
Links: UniProt, UniProt_CAEEL, EMBL-EBI, UniProt_CAEEL

TMEM (Human TransMEMbrane protein) homolog (Q23481_CAEEL)

Info: Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Plasmodium falciparum, Saccharomyces cerevisiae, Dictyostelium discoideum
Links: UniProt, UniProt_CAEEL, UniProt_CAEEL, UniProt_CAEEL

TMEM (Human TransMEMbrane protein) homolog (Q2A695_CAEEL)

Info: Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Plasmodium falciparum, Saccharomyces cerevisiae, Dictyostelium discoideum
Links: UniProt, UniProt_CAEEL, UniProt_CAEEL, UniProt_CAEEL

Tmem115 protein (Q1LVQ9_DANRE)

Info: Danio rerio
Links: UniProt, UniProt_CAEEL, UniProt_CAEEL

TMEM121 domain-containing protein, putative (O95246_PLAFT)

Info: Plasmodium falciparum
Links: UniProt, UniProt_CAEEL, UniProt_CAEEL

Tmem129 protein (B1WSE8_RAT)

Info: Rattus norvegicus
Links: UniProt, UniProt_CAEEL



L.S.B.

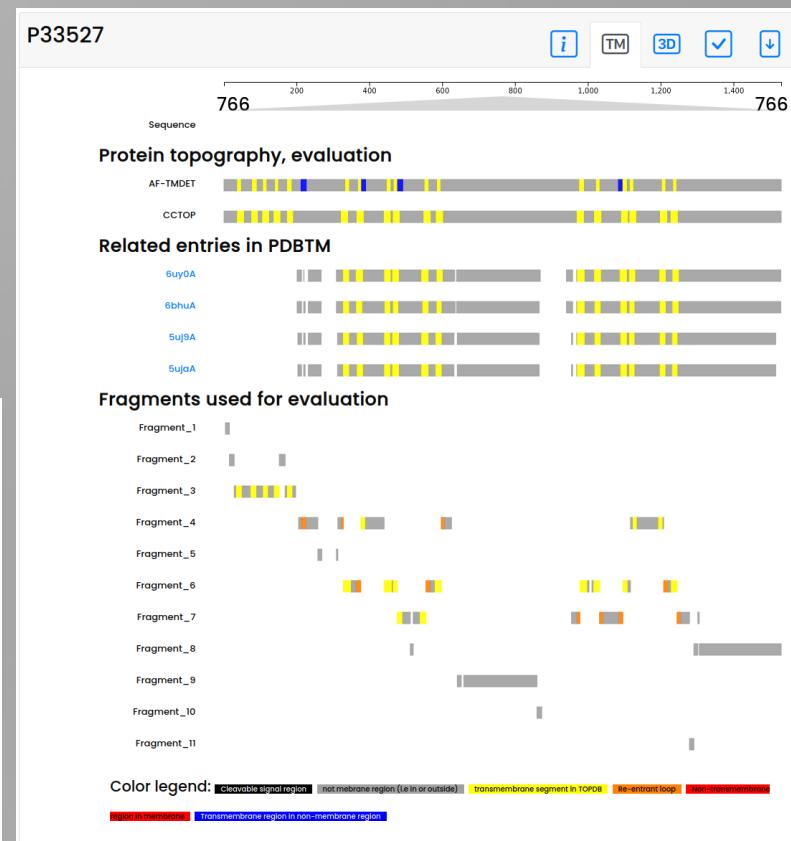
Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

Protein page:

- **Information panel**
- **Topology panel**

P33527	
<input type="button" value="i"/>	<input type="button" value="TM"/>
<input type="button" value="3D"/>	<input checked="" type="checkbox"/>
<input type="button" value=""/>	<input type="button" value=""/>
<input type="button" value=""/>	<input type="button" value=""/>
Gene name(s)	ABCC1
Source	Homo sapiens
Organism	
Subcellular localisation	Cell membrane
UniProt	MRP1_HUMAN go to UniProt
AlphaFold at EBI	go to AlphaFold DB
TMDET result	qValue: 70.7
CCTOP result	Evidence level: Experiment, Reliability: 86.67
Evaluation result	Good





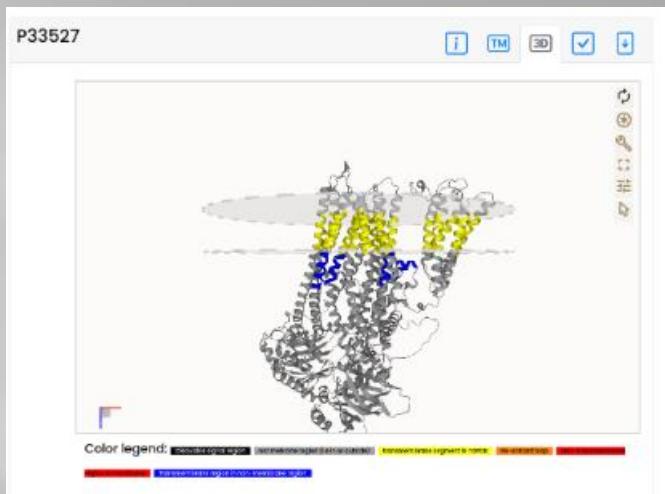
L.B.B

Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

Protein page:

- **3D structure panel**
- **Evaluation panel**



P33527

Test	Passed
Detecting membrane plane ⓘ	✓
Signal ⓘ	✓
Full structure ⓘ	✓
Short helix ⓘ	✓
Masked segment in membrane plane ⓘ	✓
Missing transmembrane part ⓘ	✗
Domain in membrane plane ⓘ	✓
Overpredict cctop ⓘ	✓
Underpredict cctop ⓘ	✓
Membrane plane cctop ⓘ	✓

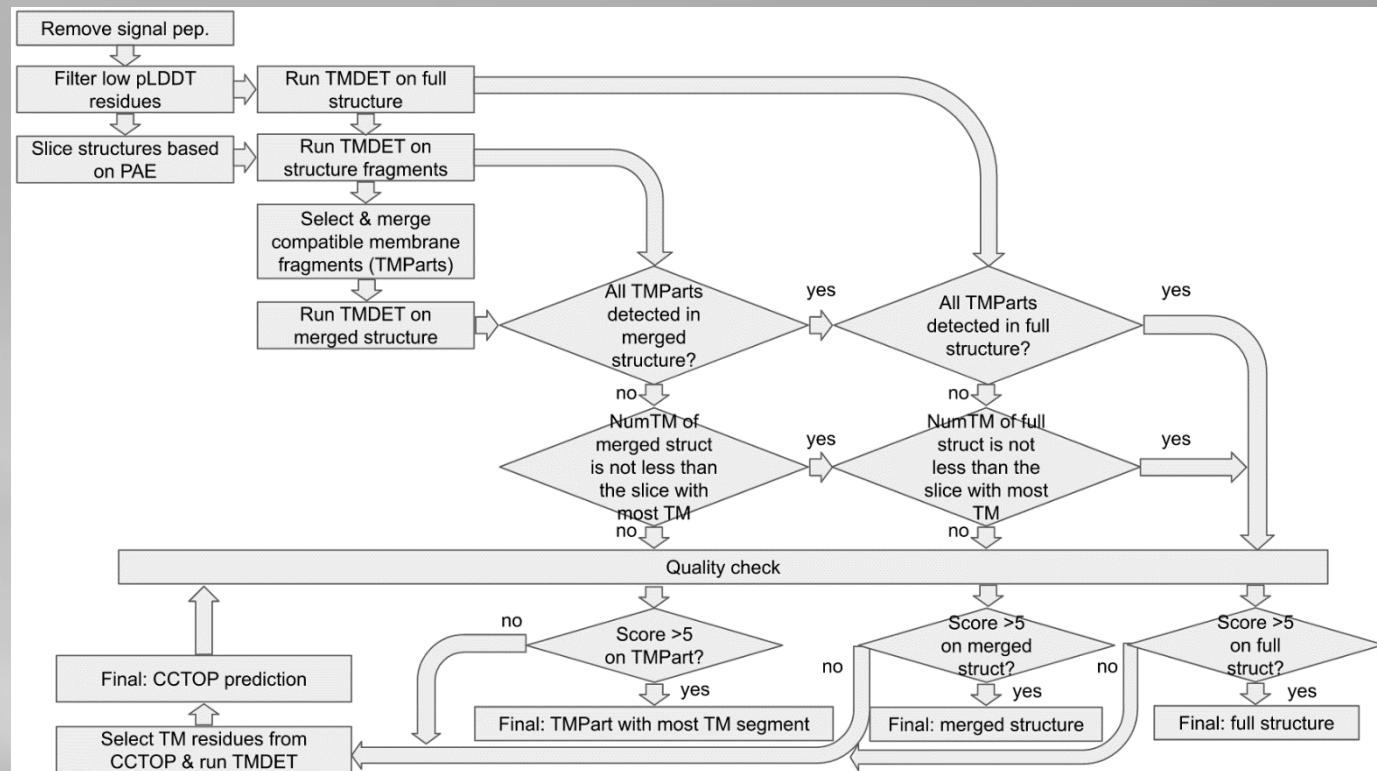


L.B.B

Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

Method





L.B.B.

Laboratory of Systems Biology
and Bioinformatics

BioGRID

BioGRID 4.4

[home](#) [help](#) [wiki](#) [projects](#) [tools](#) [contribute](#) [stats](#) [downloads](#) [partners](#) [about us](#) [Twitter](#)

Welcome to our Database of Protein, Genetic and Chemical Interactions

BioGRID is a biomedical interaction repository with data compiled through comprehensive curation efforts. Our current index is version **4.4.226** and searches **83,191** publications for **2,650,520** protein and genetic interactions, **30,725** chemical interactions and **1,128,339** post translational modifications from major model organism species. All data are freely provided via our search index and available for download in many standardized formats.

[BioGRID Statistics](#) [Latest Downloads](#)

Q Search BioGRID: By Protein/Gene

Search by Protein/Gene Identifiers ...

All Organisms

Submit Identifier Search **Q**

[Advanced Search](#) [Helpful Search Tips](#) [Featured Datasets](#)

BioGRID COVID-19 Coronavirus Curation Project
Search BioGRID for [SARS-CoV-2 Protein Interactions](#) | Download [SARS-CoV-2 and Coronavirus-Related Interactions](#)

Related Resources

BioGRID ORCS - An open repository of CRISPR screens
The BioGRID Open Repository of CRISPR Screens (ORCS) is a publicly accessible [database of CRISPR screens](#) compiled through comprehensive curation of all genome-wide CRISPR screen data reported in the biomedical literature. ORCS is updated on a quarterly basis and is fully searchable by gene/protein, phenotype, cell line, authors, and other attributes. Each screen recorded in ORCS is accompanied by structured metadata annotation that captures salient CRISPR experimental details. All data in ORCS can be downloaded in standard formats.
[Learn more](#)

BioGRID Themed Curation Projects
BioGRID themed curation projects focus on specific biological processes with disease relevance. Core Projects

Partners
 OFFICE OF RESEARCH INFRASTRUCTURE PROGRAMS
 THE HOSPITAL FOR SICK CHILDREN

BioGRID

The Biological General Repository for Interaction Datasets (BioGRID):

Overview:

- An open-access database.
- Houses genetic and protein interactions.
- Curated from primary biomedical literature.
- Encompasses major model organism species and humans.

Statistics as of October 18, 2020:

- Interactions: 1,928 million.
- Publications: 63,083.
- Model Organisms: 71.
- Growth by January 2021:
- Biological Interactions: >2,0 million.
- Chemical-Protein Interactions: 29,023.
- Post-Translational Modifications: 506,485.
- Publications: 75,988.
- Species: >80.



L.B.B

Laboratory of Systems Biology
and Bioinformatics

BioGRID

BioGRID Database Statistics

This page shows the latest snapshot of statistics for the BioGRID database. All previous snapshots are archived below separated by year.

- **Raw Interactions** - Each unique combination of interactors A and B, experimental system and publication is counted as a single interaction. Reciprocal interactions ($A \rightarrow B$ and $B \rightarrow A$) are counted twice.
- **Non-Redundant Interactions** - Each unique combination of interactors A and B are counted as a single interaction, regardless of directionality, experimental system and publication.

Current Build Statistics (4.4.227) - November 2023



L.B.B

Laboratory of Systems Biology
and Bioinformatics

BioGRID

Physical and Genetic Interaction Statistics

Organism	Experiment Type	Raw Interactions	Non-Redundant Interactions	Unique Genes	Unique Publications
<i>Anas platyrhynchos</i>	PHYSICAL	3	3	3	2
	GENETIC	0	0	0	0
	COMBINED	3	3	3	2
<i>Anopheles gambiae (PEST)</i>	PHYSICAL	2	1	2	2
	GENETIC	0	0	0	0
	COMBINED	2	1	2	2
<i>Apis mellifera</i>	PHYSICAL	1	1	2	1
	GENETIC	0	0	0	0
	COMBINED	1	1	2	1
<i>Arabidopsis thaliana (Columbia)</i>	PHYSICAL	82,254	73,962	11,731	2,361
	GENETIC	362	294	322	164
	COMBINED	82,616	74,163	11,778	2,442
<i>Bacillus subtilis (168)</i>	PHYSICAL	18	11	9	4
	GENETIC	0	0	0	0
	COMBINED	18	11	9	4
<i>Bos taurus</i>	PHYSICAL	644	573	593	224
	GENETIC	0	0	0	0
	COMBINED	644	573	593	224
<i>Caenorhabditis elegans</i>	PHYSICAL	40,963	37,301	8,607	1,518
	GENETIC	2,348	2,281	1,140	37
	COMBINED	43,311	39,528	8,809	1,538
<i>Candida albicans (SC5314)</i>	PHYSICAL	1,558	1,327	948	173
	GENETIC	531	454	376	39
	COMBINED	2,089	1,756	1,210	191



L.B.B

Laboratory of Systems Biology
and Bioinformatics

BioGRID

Chemical Interaction Statistics

Organism	Raw Interactions	Non-Redundant Interactions	Unique Genes	Unique Publications	Unique Chemicals
<i>Bacillus subtilis</i> (168)	164	85	49	15	79
<i>Caenorhabditis elegans</i>	8	4	4	2	1
<i>Candida albicans</i> (SC5314)	108	25	7	50	24
<i>Escherichia coli</i> (K12)	5	2	2	3	2
<i>Escherichia coli</i> (K12/MG1655)	1,590	834	287	137	583
<i>Homo sapiens</i>	28,409	12,124	2,268	9,469	5,699
<i>Human Herpesvirus 1</i>	65	24	3	24	21
<i>Human Herpesvirus 4</i>	10	3	3	6	3
<i>Human Immunodeficiency Virus 1</i>	105	53	2	48	53
<i>Leishmania major</i> (Friedlin)	8	4	1	2	4
<i>Middle-East Respiratory Syndrome-related Coronavirus</i>	4	4	3	4	4
<i>Mus musculus</i>	2	1	1	2	1
<i>Mycobacterium tuberculosis</i> (CDC1551)	3	3	3	1	3
<i>Mycobacterium tuberculosis</i> (H37Rv)	1	1	1	1	1
<i>Mycoplasma pneumoniae</i> (M129)	2	1	1	2	1
<i>Plasmodium falciparum</i> (3D7)	4	1	1	4	1
<i>Rattus norvegicus</i>	2	2	1	2	2
<i>Saccharomyces cerevisiae</i> (S288c)	7	7	2	2	7
<i>Severe acute respiratory syndrome coronavirus</i>	8	8	5	7	7
<i>Severe acute respiratory syndrome coronavirus 2</i>	52	50	6	30	47
<i>Streptococcus pneumoniae</i> (ATCCBAA255)	82	72	8	11	29
<i>Streptococcus pneumoniae</i>	81	69	10	10	10



L.B.B

Laboratory of Systems Biology
and Bioinformatics

BioGRID

Post Translational Modification (PTM) Statistics

Organism	PTM	Raw Sites	NR Sites	Un-assigned Sites	Unique Proteins	Unique Genes	Unique Publications
<i>Arabidopsis thaliana</i> (Columbia)	SUMOYLATION	0	0	2	0	2	1
	UBIQUITINATION	0	0	8	0	8	8
	COMBINED	0	0	10	0	10	9
<i>Bos taurus</i>	SUMOYLATION	0	0	2	0	2	2
	UBIQUITINATION	0	0	1	0	1	1
	COMBINED	0	0	3	0	3	3
<i>Caenorhabditis elegans</i>	SUMOYLATION	0	0	1	0	1	1
	COMBINED	0	0	1	0	1	1
<i>Chlorocebus sabaeus</i>	UBIQUITINATION	0	0	1	0	1	1
	COMBINED	0	0	1	0	1	1
<i>Cricetulus griseus</i>	UBIQUITINATION	0	0	1	0	1	1
	COMBINED	0	0	1	0	1	1
<i>Danio rerio</i>	SUMOYLATION	0	0	1	0	1	1
	UBIQUITINATION	0	0	4	0	4	3
	COMBINED	0	0	5	0	5	4
<i>Drosophila melanogaster</i>	NEDDYLATION	0	0	2	0	2	1
	SUMOYLATION	0	0	2	0	2	2
	UBIQUITINATION	3,534	3,534	7	3,343	1,045	5
	COMBINED	3,534	3,534	11	3,343	1,047	7
<i>Gallus gallus</i>	SUMOYLATION	0	0	1	0	1	1
	UBIQUITINATION	0	0	1	0	1	1
	COMBINED	0	0	2	0	1	1
<i>Hepatitis C Virus</i>	UBIQUITINATION	0	0	2	0	1	2
	COMBINED	0	0	2	0	1	2



L.B.B

Laboratory of Systems Biology
and Bioinformatics

BioGRID

Download

Name

[« Back](#)

- [Current-Release](#)
- [Previous-Release](#)
- [Release-Archive](#)
- [Published-Datasets](#)
- [Other-Datasets](#)
- [Latest-Release](#)
- [External-Database-Builds](#)
- [Cytoscape-Plugin](#)



BioGRID

Online Tools and Resources

- **BioGRID ORCS - The BioGRID Open Repository of CRISPR Screens (ORCS)**
- **PhosphoGRID**
- **Yeast Kinome - *Saccharomyces cerevisiae* Kinase and Phosphatase Interactome (KPI) Resource**
- **Pathway Commons**



BioGRID Example

BioGRID 4.4

Result Summary

BioGRID COVID-19 Coronavirus Curation Project
Search BioGRID for SARS-CoV-2 Protein Interactions | Download SARS-CoV-2 and Coronavirus-Related Interactions

NSP5 Severe acute respiratory syndrome coronavirus 2

ORF1ab, ORF1ab-nsp5, SARS-CoV2 nsp5, SARS-CoV-2 nsp5, Mpro, 3CLpro, 3C-like proteinase, R1AB_SARS2, PRO_0000449623, GU280_gp01_nsp5

3C-like proteinase; Non-structural protein 5

COVID-19 Coronavirus Project

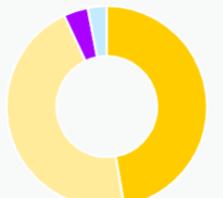
GO Process (0) GO Function (0) GO Component (0)

Entrez Gene RefSeq UniprotKB

Download Curated Data for this Protein

Interactor Statistics

Proteins/Genes	Chemicals	Publications
766	24	115

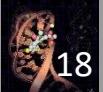


- Interactors w/ Physical (HTP) Evidence (373)
- Interactors w/ Physical (LTP) Evidence (362)
- Interactors w/ More than One Evidence Type (31)
- Chemical Interactors (24)

Switch View: Interactors 790 Interactions 1,003 Chemical Interactions 26 Network

Showing 1 to 300 of 790 unique interactors

Filter Interactions... ADV





L.B.B

Laboratory of Systems Biology
and Bioinformatics

HIPPIE



HIPPIE

» Human Integrated Protein-Protein Interaction rEfERENCE

PROTEIN QUERY NETWORK QUERY BROWSE SCREEN ANNOTATION DOWNLOAD INFORMATION

Welcome to HIPPIE, a web tool to generate reliable and meaningful human protein-protein interaction networks

Please enter a single [UniProt identifier \(accession\)](#), [gene symbol](#) or [Entrez gene id](#)

(e.g. HD_HUMAN, P42858, HTT or 3064)

search

Here, you can query HIPPIE for the interaction partners of a single protein 

Or check out further [query options and examples](#)

NEWS

Apr 29, 2022 We just released HIPPIE v2.3

Feb 14, 2019 A new version of HIPPIE (v2.2) has been released today

Jul 18, 2017 The update to HIPPIE v2.1 contains 52,000 new interactions

Nov 3, 2016 A new [paper](#) is out describing the new functionality and data of HIPPIE v2.0

Jun 24, 2016 HIPPIE v2.0 has been released including new data and analyses options

Sep 01, 2015 We just released a new version of HIPPIE





L.B.B

Laboratory of Systems Biology
and Bioinformatics

HIPPIE

BRAF, MEK1 (MAP2K1) and ERK1 (MAPK3), which are members of the Mitogen-activated protein kinase (MAPK) signaling cascade and activate each other in the stated order.

The screenshot shows the HIPPIE web application. At the top, there is a logo of a cartoon hippo and the text "HIPPIE » Human Integrated Protein-Protein Interaction rEference". Below this is a navigation bar with links: PROTEIN QUERY, NETWORK QUERY, BROWSE, SCREEN ANNOTATION, DOWNLOAD, and INFORMATION. The main content area has a heading "Construction of a HIPPIE subnetwork from an input query set of proteins or interactions". It features a text input field with placeholder text "Insert a list of proteins/interactions" and a list box containing "BRAF", "MAP2K1", and "MAPK3". To the right of this, there is an "Example input:" field with "dnmt3a dnmt3b" and a "Browse..." button with "No file selected.". At the bottom left is a "search" button with the note "(this may take a while)".



L.B.B

Laboratory of Systems Biology
and Bioinformatics

HIPPIE

- Filter (A): High Confidence Interactions
- Filter (B): Colon Tissue Specificity
- Contextual Information
- BRAF Mutation Analysis
- Filter (C): Shortest Path to Transcription Factors from BRAF
- Final Interaction Set

A

Score filter (optional)
Insert a threshold on the HIPPIE confidence score
0.01 [0,1]

Or, choose predefined confidence level
high confidence (0.73)

Tissue filter (optional)
Input of user defined filter set
Alternatively, choose a file to upload
Browse... No file selected.

B

do not show direction
For shortest path edge direction, please, indicate sources and sinks:
Sources: BRAF
Sinks:

C

Edge directionality (optional)
Add receptors to sources Add transcription factors to sinks



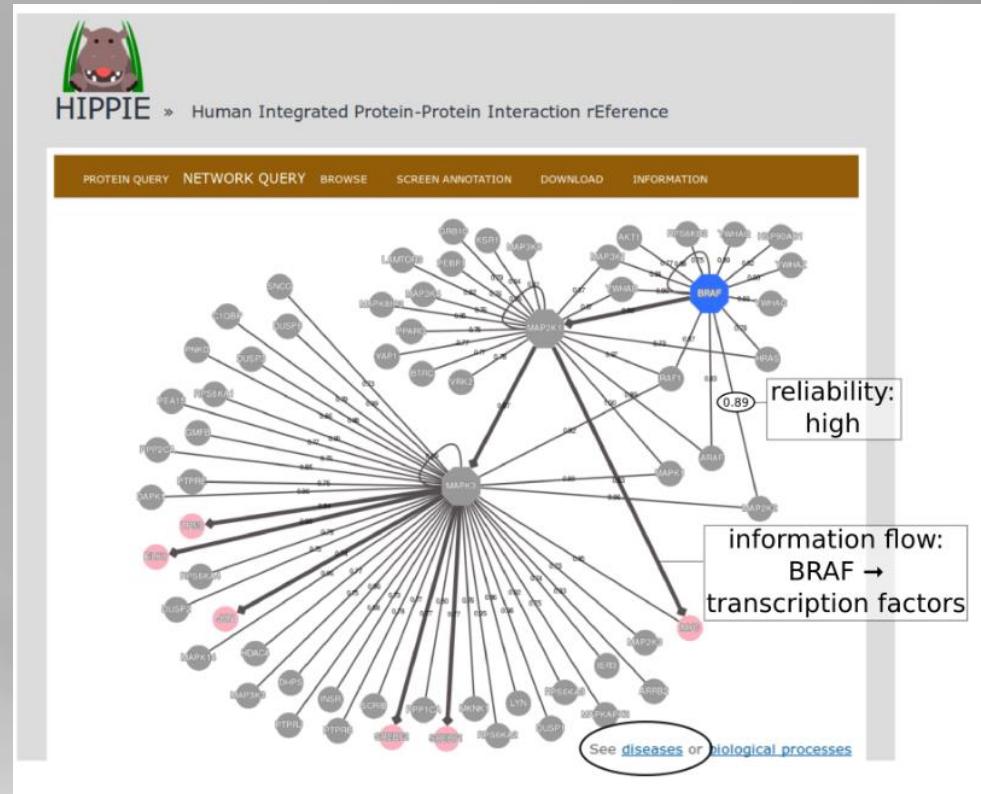


L.B.B

Laboratory of Systems Biology
and Bioinformatics

HIPPIE

- Shortest Paths from BRAF to Transcription Factors
- Terminal Nodes (Transcription Factors)
- Further Analysis Option - Disease Identification
- Tool Utilized
- Interactive Exploration





L.B.B

Laboratory of Systems Biology
and Bioinformatics

dcGO

dcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more

Basic ... Advanced ... Mining hub ...

What is dcGO (Background)

- As a biomedical ontology resource, dcGO integrates knowledge from a variety of contexts, ranging from functional information like Gene Ontology (GO) to others on enzymes and pathways, from phenotype information across major model organisms to information about human diseases and drugs. In dcGO, all Biomedical Ontologies that are not GO are collectively referred to as BO.
- As a protein domain resource, dcGO includes annotations to both the individual domains and supra-domains (i.e., combinations of two or more successive domains). By default, the domain classifications are taken from the Structural Classification Of Proteins (SCOP) at both the superfamily and family levels.
- As a general method, dcGO has an automated procedure for statistically inferring associations between ontological terms and domains or combinations of domains. An automatic pipeline regularly updates dcGO on a fortnightly basis.
- Similar to the concept of GO slim, dcGO has a partition procedure for deriving a reduced, more manageable version of the ontology. In dcGO, each ontology slim contains terms at four levels of increasing granularity (i.e., highly general, general, specific, and highly specific).

How to access dcGO (Download)

- Both flat files and MySQL tables are available for download along with detailed documentation.
- Domain classifications and ontologies are organized in hierarchies, and dcGO includes the facility to browse the hierarchies: SCOP Hierarchy for browsing domains, GO Hierarchy for browsing GO terms, and BO Hierarchy for browsing other terms (mostly phenotypes).
- In addition to SCOP domains, GO annotations to Pfam families are also provided (see PFAM Hierarchy and PFAM2GO Download).

https://en.wikipedia.org/wiki/dcGO



dcGO

- Modular Design of Proteins:

Proteins exhibit a modular design.

- Contribution of Protein Domains:

Protein domains play a crucial role in understanding proteomic data.

Their contributions can be structural, evolutionary, and functional.

- Ontological Terms and Protein Domains:

Instead of associating ontological terms solely with full-length proteins, it is beneficial to associate terms with protein domains.

- Operational Units and Functional Responsibility:

More than one domain can function as the operational unit responsible for a particular function.

This can occur through collaboration or interaction at the interface between domains.

- Associating Terms with Domain Combinations:

It is useful to associate ontological terms not only with individual domains but also with pairs of domains, triplets, and longer supra-domains.

- Developed Method for Detection:

A general method has been developed to detect functional and phenotypic signals.

This method is based on gene/protein-level annotations and can be explained at the protein domain (and their combinations) levels



L.B.B

Laboratory of Systems Biology
and Bioinformatics

dcGO

- **Source of Domain Classifications**
- **Focus Levels**
- **Criteria for Superfamily Level**
- **Detection and Classification in SUPERFAMILY**
- **Representation of Proteins**
- **Hierarchy within Superfamilies**
- **Co-occurrence and Co-evolution**
- **Supra-domains Definition**
- **Domain Architecture Representation**
- **Exclusion of Gaps in Supra-domains**



L.B.B

Laboratory of Systems Biology
and Bioinformatics

dcGO

Example

dcGO
on functions, phenotypes, diseases and more

Faceted Search dcGO

[HOME](#) > [Faceted search](#) > Faceted Search Results: HTLV-1

BACKGROUND

MINING

Faceted Search
PSnet
sTOL
dcGOnet
dcGO Predictor
dcGO Enrichment
dcGO Pevs
dcGOR New

BROWSE

SCOP Hierarchy
GO Hierarchy
BO Hierarchy
PFAM Hierarchy

ALGORITHM

DOWNLOAD

CITATIONS

LINKOUT

SUPERFAMILY
SCOP
Gene Ontology.
UniProt-GOA
KeyWords
HLB

Keyword Search Results

Results 1-6 of 6 for HTLV-1.

Refine search by:

- SCOP (1)
- Genome names (0)
- Sequence IDs (0)
- GO names (0)
- BO names (5)

1. SCOP classification

Class :	All alpha proteins
Fold :	Retroviral matrix proteins
Superfamily :	Retroviral matrix proteins
Family :	HTLV-II matrix protein

2. BO name: HTLV-II Infections (CTD Diseases) PSnet sTOL PDB

3. BO name: HTLV-I Infections (CTD Diseases) PSnet sTOL PDB

4. BO name: Env-5 glycoprotein, HTLV-I (CTD Chemicals) PSnet sTOL PDB



L.S.B.

Laboratory of Systems Biology
and Bioinformatics

MatrisomeDB

MatrisomeDB

MatrisomeDB is a searchable database that integrates experimental proteomic data on the ECM composition normal and diseased tissues. It also provides live cross-referencing to gene and protein databases for every ECM and ECM-associated genes. If you are interested, please help us to test the new [Matrisome 2.0](#)

Citation: MatrisomeDB: the ECM-protein knowledge database. Shao X, Taha IN, Clauser KR, Gao Y, Naba A. *Nucleic Acids Res.* 2019 Oct 5. pii: gkz849. doi:10.1093/nar/gkz849

Enter gene name (COL), or tissue (Liver) or description word (collagen) to search

Search

Categories

- Core matrisome**
 - Collagens
 - ECM Glycoproteins
 - Proteoglycans
- Matrisome associated**
 - ECM-affiliated Proteins
 - ECM Regulators
 - Secreted Factors

Species

- Human
- Mouse
- Xenograft

Tissues / Tumors

- Human**
 - Blood vessel
 - Breast
 - Breast Cancer
 - Colon
 - Colon Cancer
 - Cranial Bone
 - Eye
 - Kidney





L.B.B

Laboratory of Systems Biology
and Bioinformatics

MatrisomeDB

- Database Overview:

MatrisomeDB is a searchable database integrating experimental proteomic data on extracellular matrix (ECM) composition in normal and diseased tissues.

- Entry Details:

Each entry provides information on the tissues where it was detected.

- Coverage Map

A coverage map illustrates detected peptides for each entry.

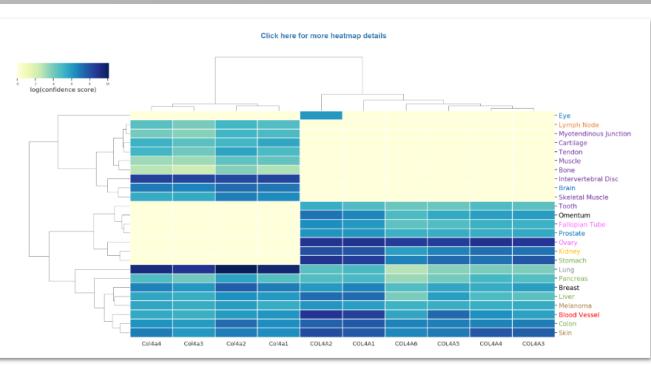
- Post-Translational Modifications:
- Cross-Referencing
- 2019 Release Collaboration
- 2022 Release (MatrisomeDB 2.0) Features



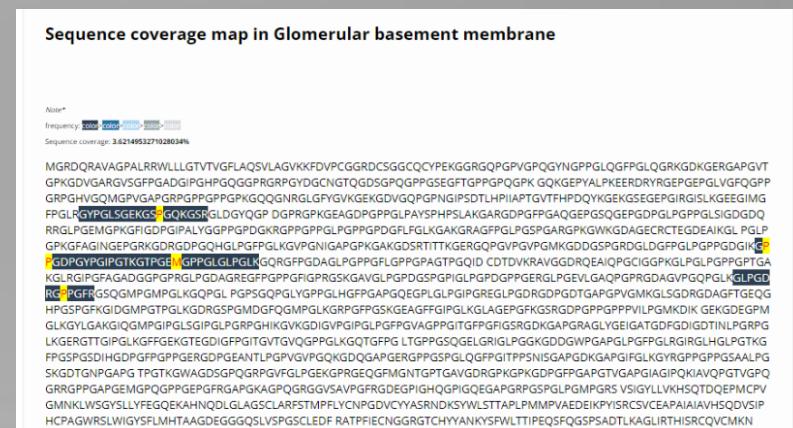
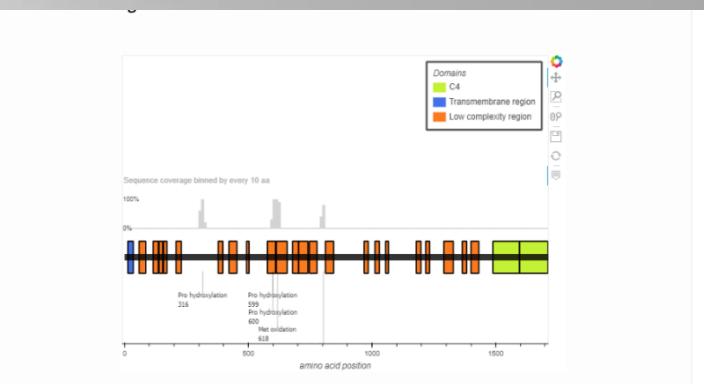
L.B.B
Laboratory of Systems Biology
and Bioinformatics

MatrisomeDB

Result



							Export all results to .tsv file	Export filtered results
Gene	UniProt	Description	Sample type	Tissue	Species	Reference		
COL4A2	P08572	Collagen alpha 2(IV) chain Open in SCV >	Retinal vascular basement membrane	Blood Vessel	Human	Uechi G et al., 2014		
Col4a2	P08122	Collagen alpha 2(IV) chain Open in SCV >	Normal lung (GDSP; ECM)	Lung	Mouse	Schiller HB et al., 2015		
COL4A2	P08572	Collagen alpha 2(IV) chain Open in SCV >	Glomerular basement membrane	Kidney	Human	Lennon R et al., 2014		





L.B.B

Laboratory of Systems Biology
and Bioinformatics

DEPICTER2

DEPICTER2 : DisorderEd Prediction CenTER2

[Tutorials](#) | [Acknowledgments](#) | [Disclaimer](#) | [Biomine](#)

The server is designed to predict intrinsically disordered regions, disordered linkers and disordered binding regions from protein sequences.

Please follow the four steps below to make predictions: ([click on this link for tutorials](#))

1. Enter input into the text area

Users can provide either only comma-separated Uniprot accession number(s) or only ([FASTA formatted](#)) protein sequence(s) as input. Each input protein sequence should have minimum length of 26 residues and maximum length of 5000 residues. The server accepts up to 25 sequences when the default/fast methods are selected. It accepts up to two sequences if DisoLipPred is selected. See [Help](#) for details concerning format.

[Example1](#) [Example2](#) [Clear input](#)

2. Provide your e-mail address (optional)

Please enter your email address in the following text area. A link to predictions will be sent to this email address once they are ready. The results will be also available in the browser window.





L.B.B

Laboratory of Systems Biology
and Bioinformatics

DEPICTER2

- **Server Type:** Computational platform or online tool.
- **Prediction Target 1:** Intrinsically disordered regions in protein sequences.
- **Prediction Target 2:** Disordered linkers connecting structured domains.
- **Prediction Target 3:** Disordered binding regions involved in interactions.
- **Purpose:** Analyzing protein sequences for intrinsic disorder-related features.



L.B.B

Laboratory of Systems Biology
and Bioinformatics

DEPICTER2

1. Submission of query using server

1. Enter input into the text area

Users can provide either only comma-separated Uniprot accession number(s) or only ([FASTA formatted](#)) protein sequence(s) as input. Each input protein sequence should have minimum length of 26 residues and maximum length of 5000 residues. The server accepts up to 25 sequences when the default/fast methods are selected. It accepts up to two sequences if DisoLipPred is selected. See [Help](#) for details concerning format.

```
>P54252
MESIFHEKQEGSLCAQHCLNNLLQGEYFSPVLESSIAHQLDDEERMRAEQQVTSYDRTFLQQPSGNMDDSGFFSIQVISNALK
VWGLELILFNSPEYQLRIPINERSFCINYKEHWFTVRKLGKQWFNLNSLLTGPRLISDTVLALFLAQLQQEGYSIVFVKGDLPDCE
ADQLLQMIRVQQMHRPKLIGEELAQLKEQRVHKTDLERVLEANDGSGMLDEDEEDLQLRALALSRQEIDMEDEEADLRRAIQLS
MQGSSRNISQDMTQTSGTNLTEELRKREAYFEKQQQQQQQQQQGDLSGQSSHPERPATSSGALGSQDLGDAMSEE
DMLQAAVTMSLETVRNDLKTEGKK
```

2. Provide your e-mail address (optional)

Please enter your email address in the following text area. A link to predictions will be sent to this email address once they are ready. The results will be also available in the browser window.

3. Select methods that will be included in the prediction

Fast Methods (default)

- A. Disorder prediction (fIDPnm)
- B. Disordered linker prediction (DFLpred)
- C. MoRF prediction (MoRFchibitLight)
- D. Disordered protein-binding prediction (ANCHOR2)
- E. Disordered DNA-binding prediction (DisoRDPbind)
- F. Disordered RNA-binding prediction (DisoRDPbind)

Slow Method

- G. Disordered lipid-binding prediction (DisoLipPred)

4. Predict

Enter protein accession numbers or sequences

'Clear input' will clear the text area

Example1: provides example of comma-separated accession numbers

Example2: provides example of protein sequence in FASTA format

All 'Fast Methods' are selected by default. User can uncheck any box and make selection as per own choice. (Max. sequences allowed = 25)

(Max. sequences allowed = 2)

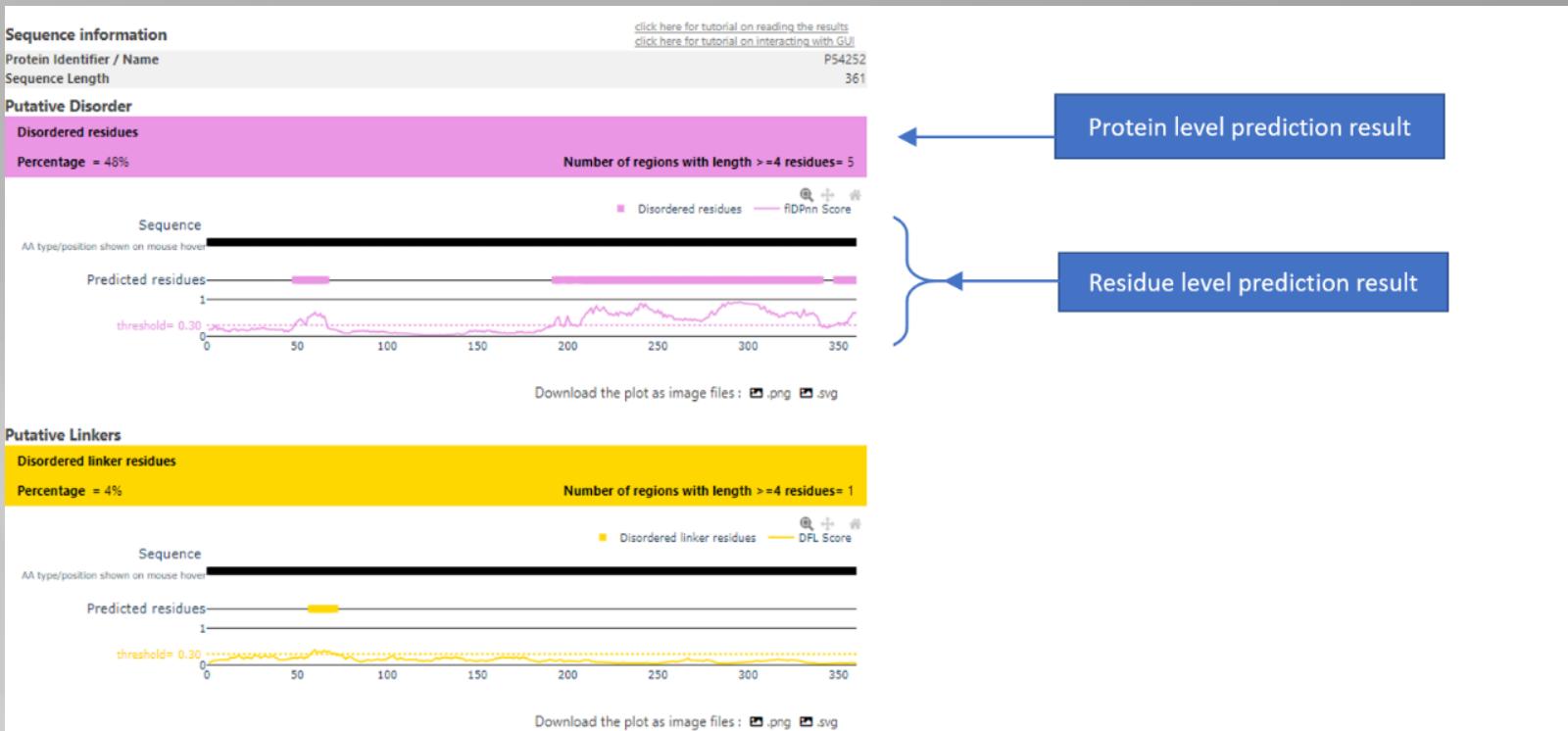




L.B.B

Laboratory of Systems Biology
and Bioinformatics

DEPICTER2





L.S.B.

Laboratory of Systems Biology
and Bioinformatics

DescribePROT

DescribePROT-DatabasE of StruCtuRe and function residue-Based prEdictions of PROTeins

[Help and Tutorial](#) | [Release Notes](#) | [Statistics](#) | [Download](#) | [References](#) | [Methods](#) | [Acknowledgments](#) | [Biomine](#)

DescribePROT webserver

This server provides 3 experimentally validated structural properties and 19 putative structural and functional properties at the amino acid level for 2,276,602 proteins from 273 complete proteomes of popular/model organisms. Help and Tutorial that explain how to use DescribePROT are available [HERE](#).

Statistics

Number of proteins	2,276,602
Number of amino acids	973,123,229
Number of predictions	21,101,037,225
Number of predicted properties	19
Number of predictors	11
Number of experimentally validated annotations	22,446,340
Number of experimentally validated properties	3
Number of proteomes	273
Number of eukaryotic proteomes <input type="checkbox"/>	92
Number of bacterial proteomes <input type="checkbox"/>	103
Number of viral proteomes <input type="checkbox"/>	61
Number of archaeal proteomes <input type="checkbox"/>	17



L.B.B

Laboratory of Systems Biology
and Bioinformatics

DescribePROT

- **Database Focus:** Predicted amino acid-level descriptors of structure and function of proteins
- **Sequence Conservation**
- **Position-Specific Scoring Matrix (PSSM)**
- **Secondary Structure:** 84.2%
- **Solvent Accessibility:** Pearson's Correlation Coefficient (PCC)
- **Intrinsic Disorder:** AUC = 0.81
- **Disordered Linkers:** Characteristic Curve (AUC), securing AUC = 0.72
- **Signal Peptides:** Matthew's Correlation Coefficient (MCC), showing high predictive quality (MCC ranges between 0.890 and 0.981)
- **MoRFs (Molecular Recognition Features):** AUC of 0.87
- **Interactions with Proteins:** AUCs between 0.64 and 0.72
- **Interactions with DNA:** AUC = 0.78
- **Interactions with RNAs:** AUC = 0.68



L.B.B

Laboratory of Systems Biology
and Bioinformatics

DescribePROT

Enter an [UniProt accession number](#) or [UniProt entry name](#) for a query protein.

[Example1](#)[Example2](#)[Clear input](#)[Search](#)

Enter a [FASTA formatted](#) sequence for a query protein.

[Example](#)[Clear input](#)[Search](#)



DescribePROT

Protein Level Characterization

DescribePROT graphical view is best-viewed in the Chrome, Safari (macOS) and Microsoft Edge (Version 85.0.564.51 or higher) web browsers.

P04637 **Cellular_tumor_antigen_p53** .json .csv

Protein Name

[Cellular tumor antigen p53](#)

UniProt ID

[P04637\(P53_HUMAN\)](#)

Taxonomy ID

9606

Sequence Length

393

Native and Putative Structure

Native disordered residues Content = 38% (Coverage = 100%)

[click here for details](#)

Putative disordered residues Content = 47%

Native secondary structure Content = H:16% E:24% C:60% (Coverage = 79%)

[click here for details](#)

Putative secondary structure Content = H:22% E:20% C:58%

Native buried residues Content = 35% (Coverage = 71%)

Putative buried residues Content = 24%

[click here for details](#)

Putative Interactions with Ligands

Percentage of Protein-binding residues 20%

[click here for details](#)

Percentage of DNA-binding residues 4%

[click here for details](#)

Percentage of RNA-binding residues 0%

[click here for details](#)

Other Properties

Signal peptide present None

[click here for details](#)

Percentage of highly conserved residues 10%

[click here for details](#)

Percentage of disordered linker residues 3%

[click here for details](#)

PTM sites Total no. of PTM residues = 42

Percentage of PTM sites = PH:5% GL:2% UB: 1% SM: 0% AC: 2% ME: 2% PY: 0% PA: 0% HY: 1%

[click here for details](#)

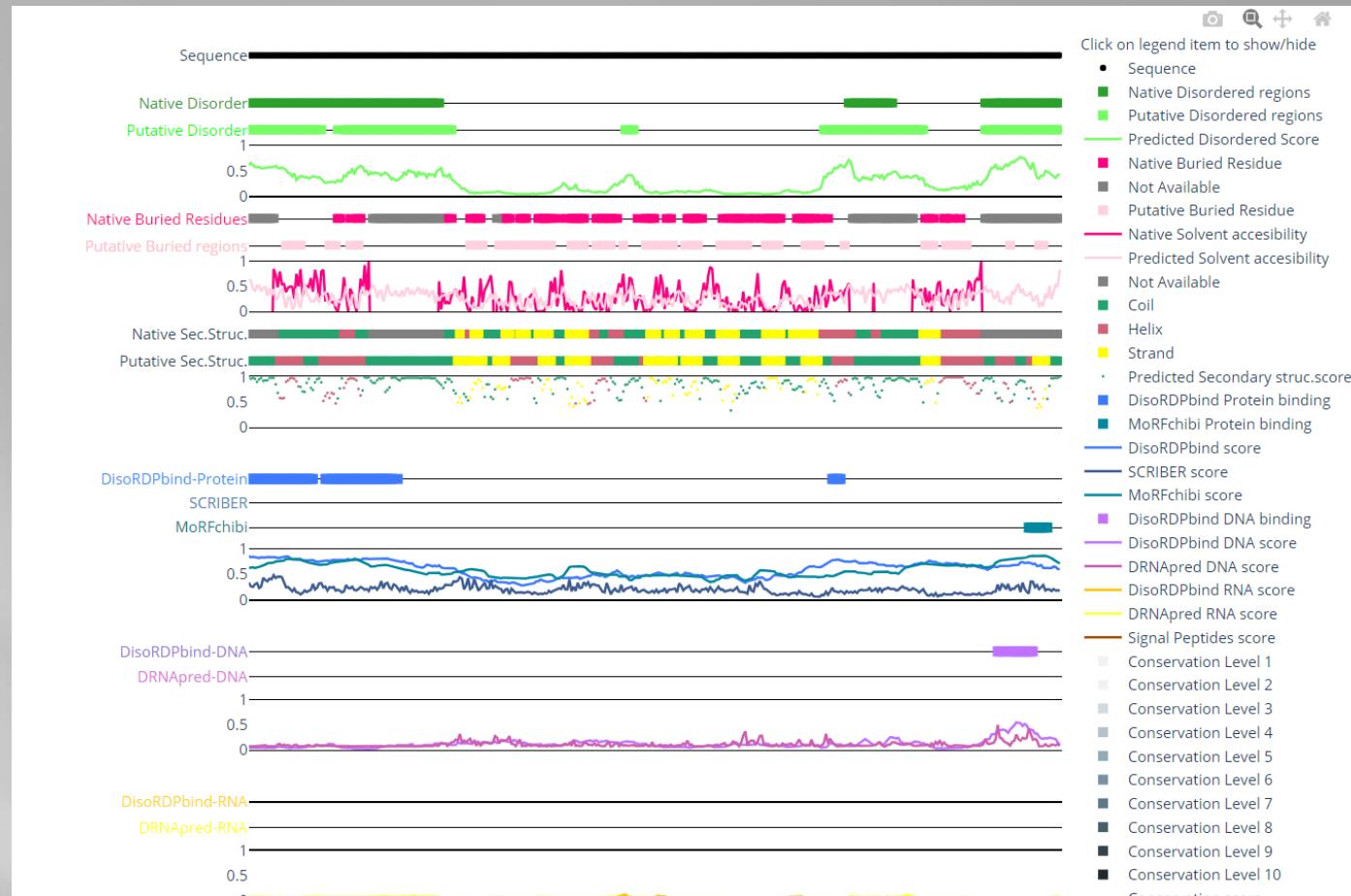




L.B.B

Laboratory of Systems Biology
and Bioinformatics

DescribePROT





L.B.B

Laboratory of Systems Biology
and Bioinformatics

Gratitude to Prof. Francisco J. Enguita at Lisbon Medical School, University of Lisbon, Portugal, for generously granting permission to use his protein images.

