



L.B.B

Laboratory of Systems Biology
and Bioinformatics

Introduction to Protein Databases: Part1

Fereshteh Noroozi

Advised by:

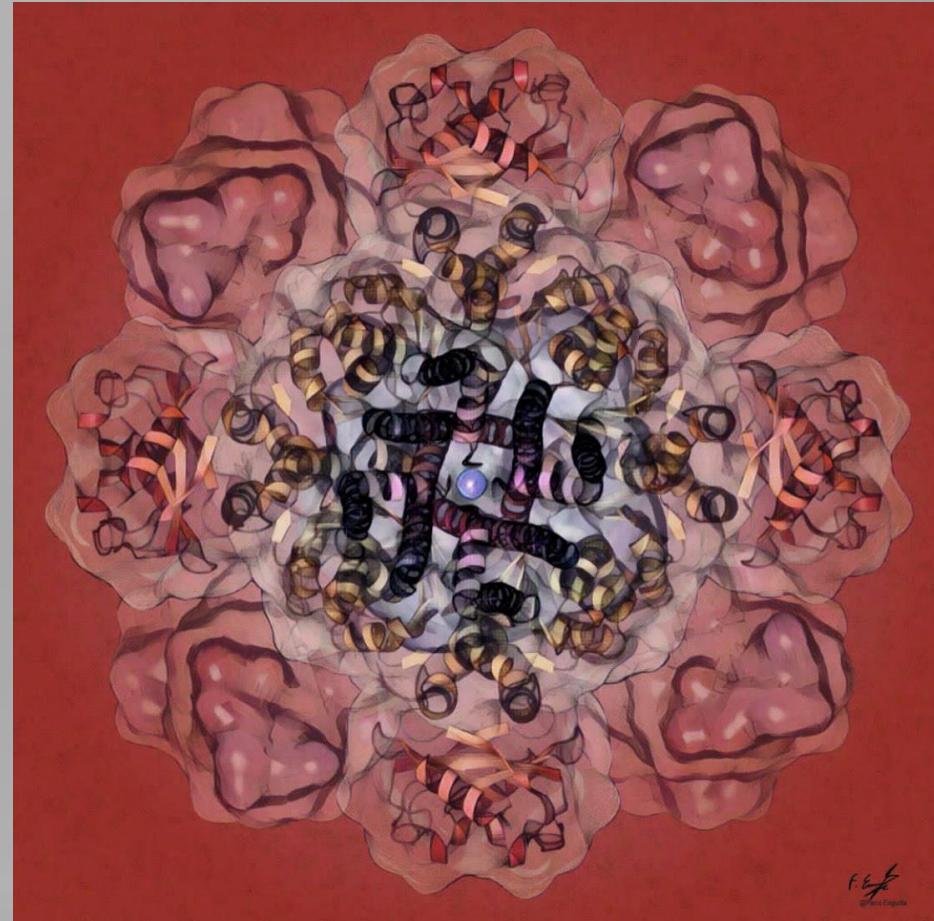
Prof. Ali Masoudi-Nejad

Fall 2023



Importance of protein databases in research

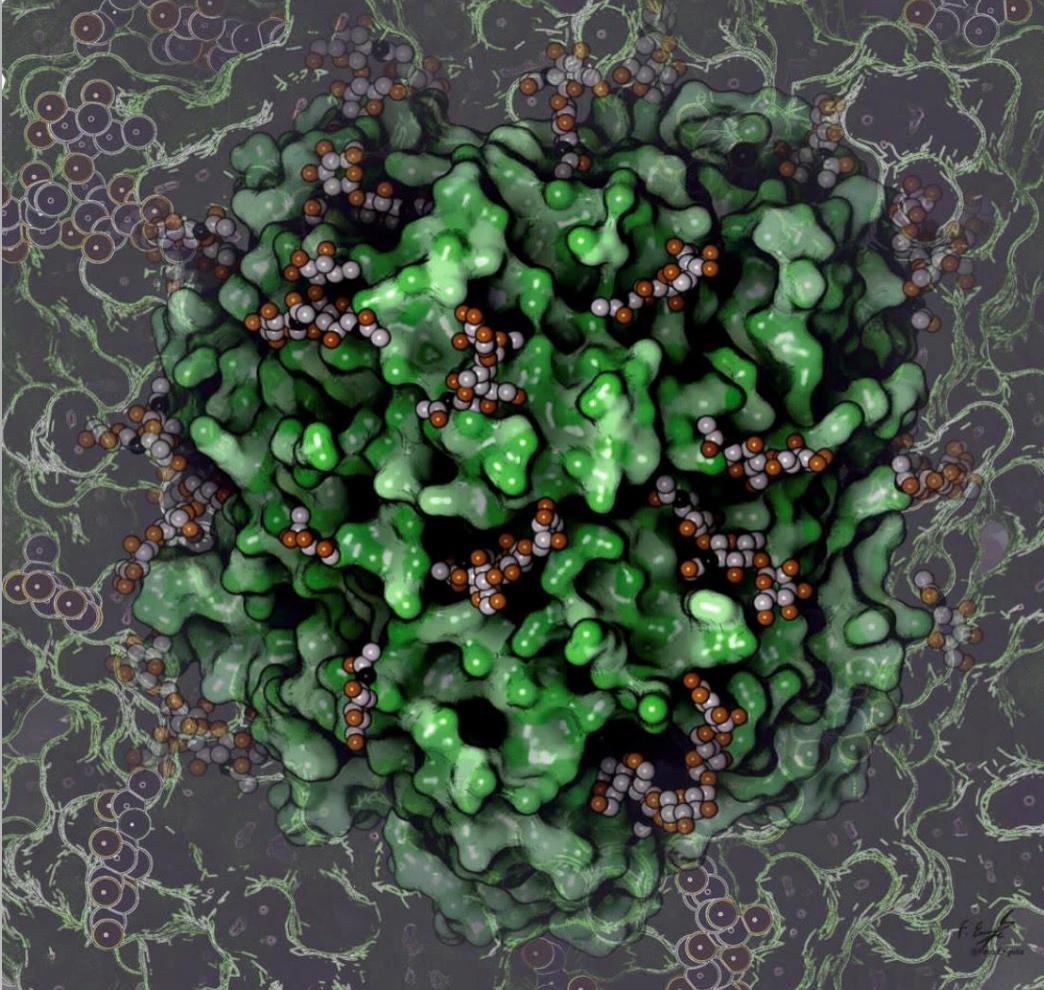
- Large-scale protein information
- Accelerating research
- Functional annotation
- Structure prediction and modeling
- Comparative analysis
- Data-driven discoveries
- Bioinformatics tools and resources





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

Key resources





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

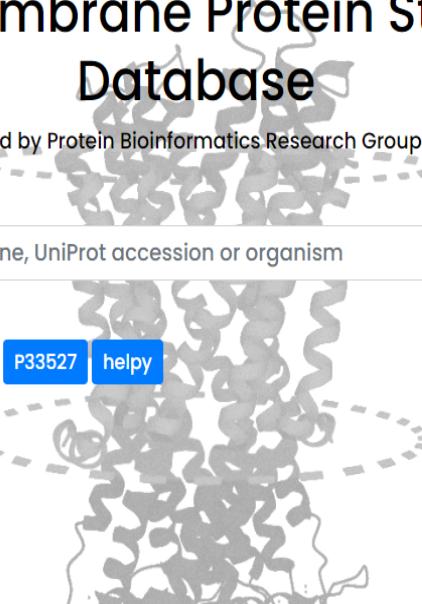
Developed by Protein Bioinformatics Research Group, RCNS

Search for protein, gene, UniProt accession or organism

Search

Examples: [tmem](#) [sico3a1](#) [P33527](#) [helpy](#)

[Help](#)



[License](#)

[Feedback](#)

[Disclaimer](#)

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





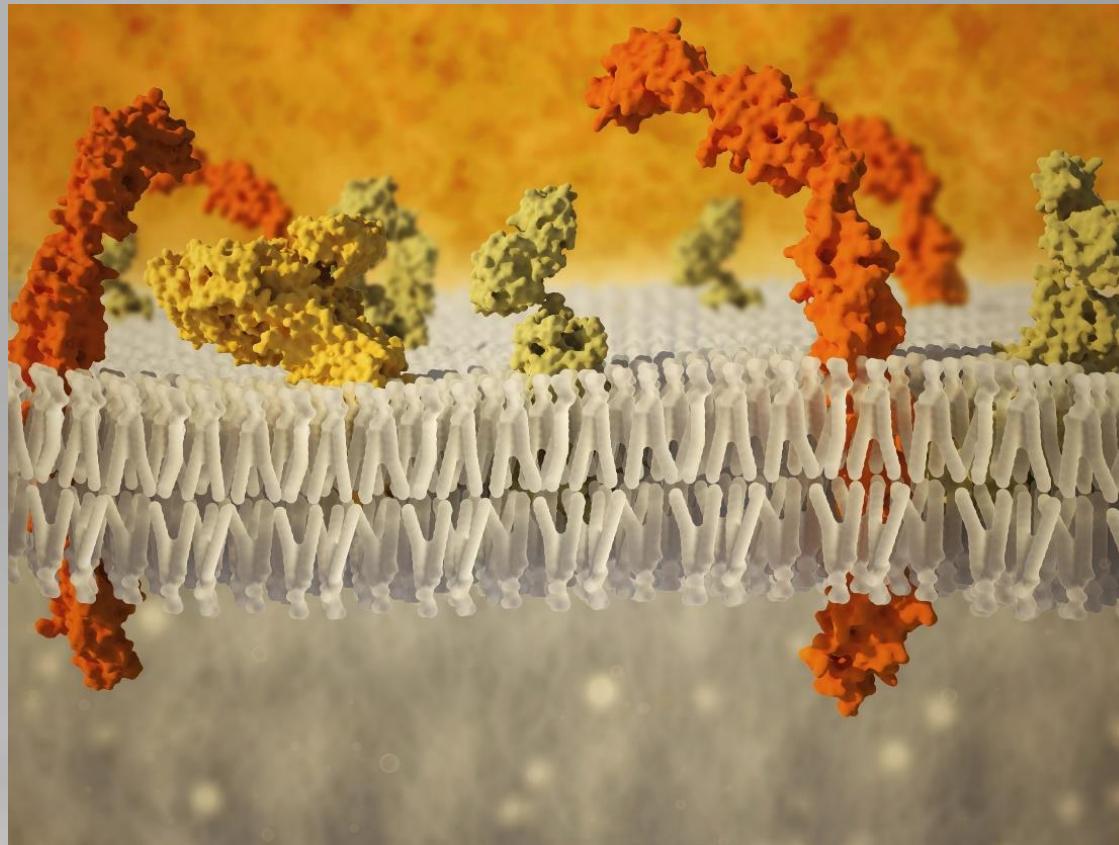
L.B.B

Laboratory of Systems Biology
and Bioinformatics

TmAlphaFold

Transmembrane proteins

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





L.S.B

Laboratory of Systems Biology
and Bioinformatics

Search modes

- Gene name
- UniProt Accession
- UniProt ID
- Species

Search for protein, gene, UniProt accession or organism

Search

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

[Help](#)





L.S.B

Laboratory of Systems Biology
and Bioinformatics

- Protein name
- Gene name
- Organism
- UniProt AC (UniProt Accession)
- AlphaFold Database hyperlink
- PDB-EBI hyperlink
- Filter options
- Navigation bar

Search results

Showing search results for tmem

Showing results 1 .. 20 of 44

Filters:

Organism

<input type="checkbox"/>	Arabidopsis thaliana (1)
<input type="checkbox"/>	Caenorhabditis elegans (7)
<input type="checkbox"/>	Danio rerio (4)
<input type="checkbox"/>	Drosophila melanogaster (1)
<input type="checkbox"/>	Plasmodium falciparum (3)
<input type="checkbox"/>	Saccharomyces cerevisiae (1)
<input type="checkbox"/>	Dictyostelium discoideum (1)

Evaluation

<input type="checkbox"/>	Excellent (10)
<input type="checkbox"/>	Good (14)
<input type="checkbox"/>	Fair (19)
<input type="checkbox"/>	Poor (1)
<input type="checkbox"/>	Failed (0)

Number of transmembrane segments

<input type="checkbox"/>	1 (14)
<input type="checkbox"/>	2 (3)
<input type="checkbox"/>	3 (5)
<input type="checkbox"/>	4 (9)
<input type="checkbox"/>	5 (2)
<input type="checkbox"/>	6 (2)
<input type="checkbox"/>	7 (1)

CCTOP Evidence levels

<input type="checkbox"/>	3D (0)
<input type="checkbox"/>	Expasy (0)
<input type="checkbox"/>	TOPDOM (1)
<input type="checkbox"/>	Exists (2)
<input type="checkbox"/>	Prediction (41)

TMEM (Human TransMEMbrane protein) homolog (001870_CAEEL)

	Caenorhabditis elegans		Prediction		Mood		CSM		TMEM-38
--	------------------------	--	------------	--	------	--	-----	--	---------

TMEM (Human TransMEMbrane protein) homolog (076687_CAEEL)

	Caenorhabditis elegans		Prediction		Mood		CSM		TMEM-120
--	------------------------	--	------------	--	------	--	-----	--	----------

TMEM (Human TransMEMbrane protein) homolog (Q23481_CAEEL)

	Caenorhabditis elegans		Prediction		Mood		CSM		TMEM-17
--	------------------------	--	------------	--	------	--	-----	--	---------

TMEM (Human TransMEMbrane protein) homolog (Q2A955_CAEEL)

	Caenorhabditis elegans		Prediction		Mood		CSM		TMEM-210
--	------------------------	--	------------	--	------	--	-----	--	----------

Tmem115 protein (Q1LVDS_DANRE)

	Danio rerio		Prediction		Mood		CSM		TMEM-115
--	-------------	--	------------	--	------	--	-----	--	----------

TMEM121 domain-containing protein, putative (O96246_PLAF7)

	Plasmodium falciparum		Prediction		Mood		CSM		TMEM-121
--	-----------------------	--	------------	--	------	--	-----	--	----------

Tmem129 protein (BIWBR6_RAT)

	Rattus norvegicus		Prediction		Mood		CSM		TMEM-129
--	-------------------	--	------------	--	------	--	-----	--	----------





L.S.B

Laboratory of Systems Biology
and Bioinformatics

Protein page

Information panel

P33527



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





Protein page

Topology panel

P33527

Sequence 200 400 600 800 1,000 1,200 1,400 766 766

Sequence

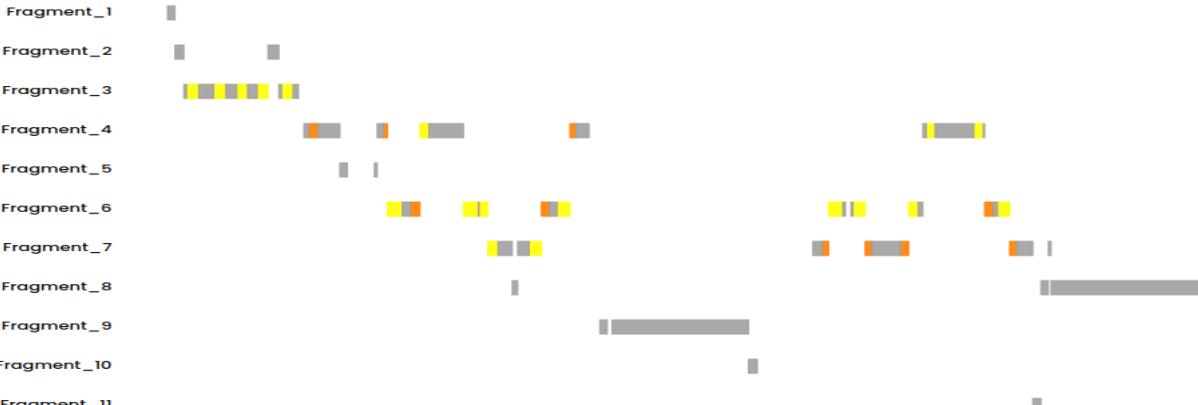
Protein topography, evaluation



Related entries in PDBTM



Fragments used for evaluation



Color legend: cleavable signal region not membrane region (i.e. in or outside) transmembrane segment in TOPDB Re-entrant loop Non-transmembrane

region in membrane Transmembrane region in non-membrane region

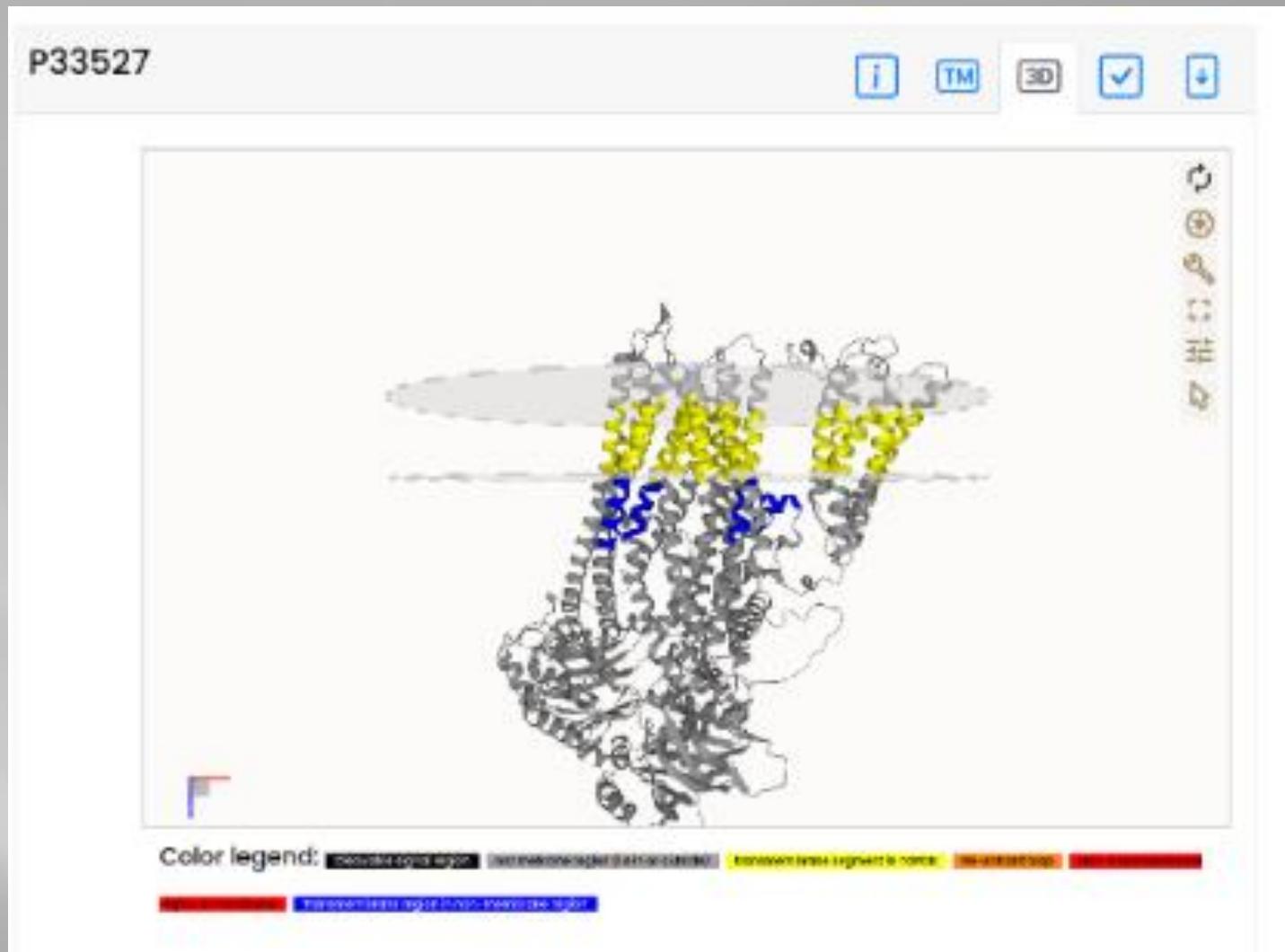


L.B.B

Laboratory of Systems Biology
and Bioinformatics

Protein page

3D structure panel





L.S.B

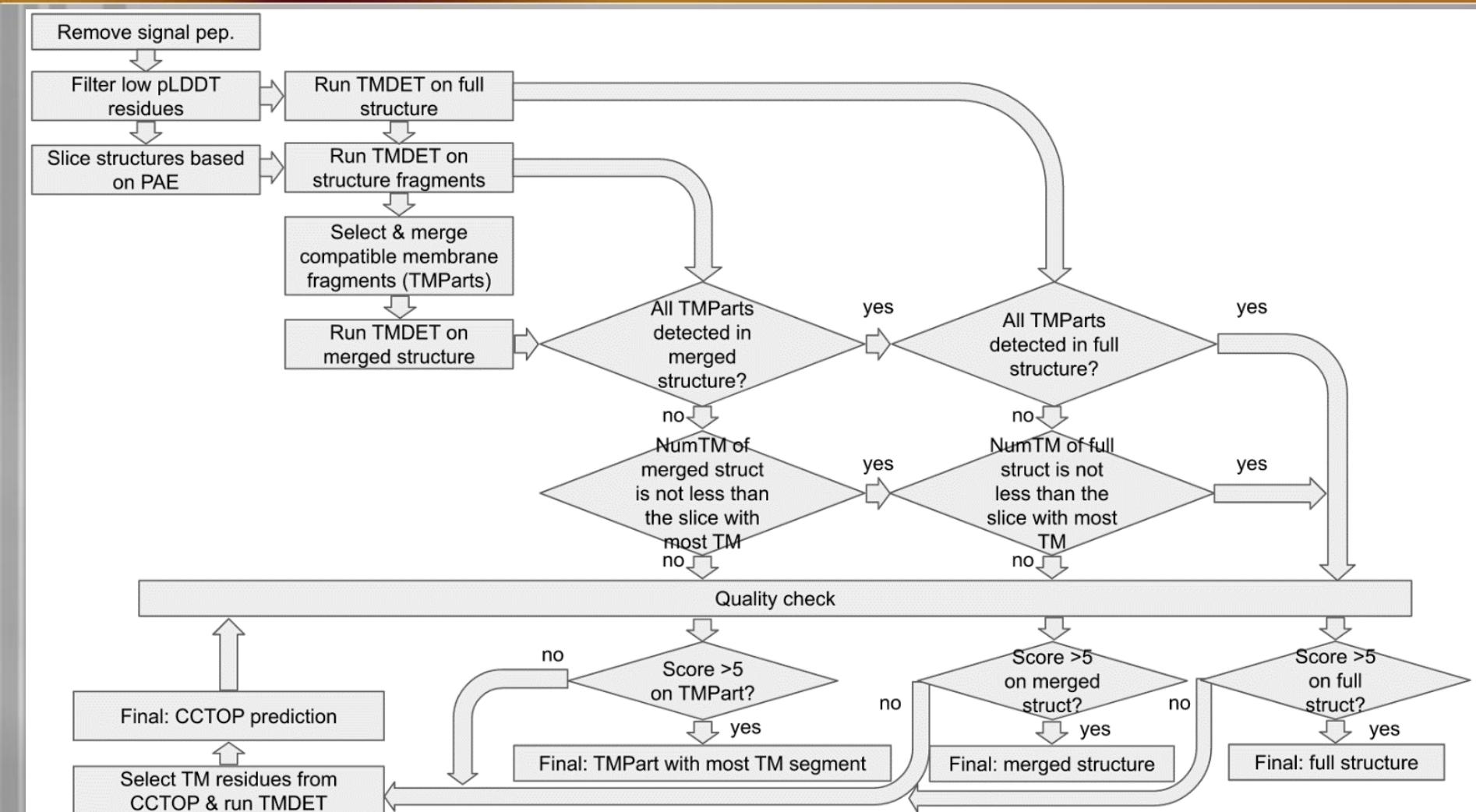
Laboratory of Systems Biology
and Bioinformatics

Protein page

Evaluation panel

P33527	<i>i</i>	TM	3D	<input checked="" type="checkbox"/>	
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				✓	

Method





Overview

<i>Quality Check Criteria</i>	<i>Criteria Met</i>
Membrane Plane Detection	Membrane plane detected and constructed
Signal Peptide Assessment	No signal peptide detected
Full Structure Utilization	All fragments used to successfully define the membrane plane
Transmembrane Helix Length	All transmembrane helices are longer than 10 amino acids
Masked Segment Analysis	No masked regions are embedded in the membrane
Consistency in Structure	Assembly and fragments are consistent, accounting for potential exclusions
Domain and Beta Structure Check	No interference from non-transmembrane pieces or appearance of beta structures
CCTOP Overprediction	No extra transmembrane segments in TmAlphaFold compared to CCTOP prediction



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](#)

Search BioGRID:

By Protein/Gene

Search by Protein/Gene Identifiers ...

All Organisms

Submit Identifier Search

[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

Partners



L.B.B

Laboratory of Systems Biology
and Bioinformatics

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.



L.S.B

Laboratory of Systems Biology
and Bioinformatics

BioGRID

- **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.S.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>	21	20	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

Download

Name

[« Back](#)

[Current-Release](#)

[Previous-Release](#)

[Release-Archive](#)

[Published-Datasets](#)

[Other-Datasets](#)

[Latest-Release](#)

[External-Database-Builds](#)

[Cytoscape-Plugin](#)





L.B.B

Laboratory of Systems Biology
and Bioinformatics

Online Tools and Resources

- BioGRID ORCS - The BioGRID Open Repository of CRISPR Screens (ORCS)

BioGRID ORCS 1.1.14.1

Home Browse Help Wiki Tools Statistics Downloads About Us

Welcome to the BioGRID Open Repository of CRISPR Screens (ORCS)

BioGRID ORCS is an open repository of CRISPR screens compiled through comprehensive curation efforts. Our current index is version 1.1.14 and searches 314 publications and 93,130 genes to return 1,772 CRISPR screens from 5 major model organism species, 775 cell lines, and 130 cell types. All screen data are freely provided through our search index and available via download in a wide variety of standardized formats.

Screen Statistics Latest Downloads

Q Search BioGRID ORCS: By Screen

Search by keywords such as autophagy, drug, toxin, virus etc

All Organisms

SEARCH

example searches

Advanced Search Key Search Tips Featured Screens



L.S.B

Laboratory of Systems Biology
and Bioinformatics

Online Tools and Resources

• PhosphoGRID

The screenshot shows the PhosphoGRID website. At the top left is the "phosphoGRID" logo with a blue circle containing a white "P". To its right is a search bar with the placeholder "Search PhosphoGRID..." and a blue "search" button. Above the search bar are four navigation links: "welcome", "contribute", "downloads", and "about us". Below the search bar is a large blue search form with its own "Search PhosphoGRID..." input field and a blue "search" button. To the right of this form is a "NOTICE OF REDIRECTION" box containing text about the website's scope change and redirection to the BioGRID website.

Search PhosphoGRID...

search

welcome contribute downloads about us

NOTICE OF REDIRECTION

PhosphoGRID search results have been permanently re-directed into the BioGRID website as our scope has expanded to cover additional post translational modifications and additional model organisms. These and ALL future efforts of our ongoing curation of post translational modifications are available online now at the [BioGRID Website](#).



L.B.B

Laboratory of Systems Biology
and Bioinformatics

Online Tools and Resources

- Yeast Kinome - *Saccharomyces cerevisiae* Kinase and Phosphatase Interactome (KPI) Resource

YeastKinome.org

Saccharomyces cerevisiae Kinase and Phosphatase Interactome (KPI) Resource

[Home](#) | [Supplementary Materials](#) | [Search Data](#) | [About](#)

Welcome to the Yeast Kinase and Phosphatase Interactome (KPI) Resource

Abstract:

The interactions of protein kinases and phosphatases with their regulatory subunits and substrates underpin cellular regulation. We identified a kinase and phosphatase interaction (KPI) network of 1,044 interactions in budding yeast by mass spectrometric analysis of protein complexes. The KPI network contained many dense local regions of interactions that suggested new functions. Notably, the cell cycle phosphatase Cdc14 associated with multiple kinases that revealed roles for Cdc14 in mitogen-activated protein kinase signaling, the DNA damage response and metabolism, while interactions of the target of rapamycin complex 1 (TORC1) uncovered new effector kinases in nitrogen and carbon metabolism. An extensive backbone of kinase-kinase interactions cross-connects the proteome and may serve to coordinate diverse cellular responses.

NOTE: Additional interactions involving these protein kinases and phosphatases have been continually curated at the BioGRID since initial publication. You can access a full set of the latest interaction data for all yeast kinases and phosphatases directly at the [BioGRID Yeast Kinome Project](#).

[129 Kinases \(click for more details\)](#)



L.S.B

Laboratory of Systems Biology
and Bioinformatics

Online Tools and Resources

- Pathway Commons



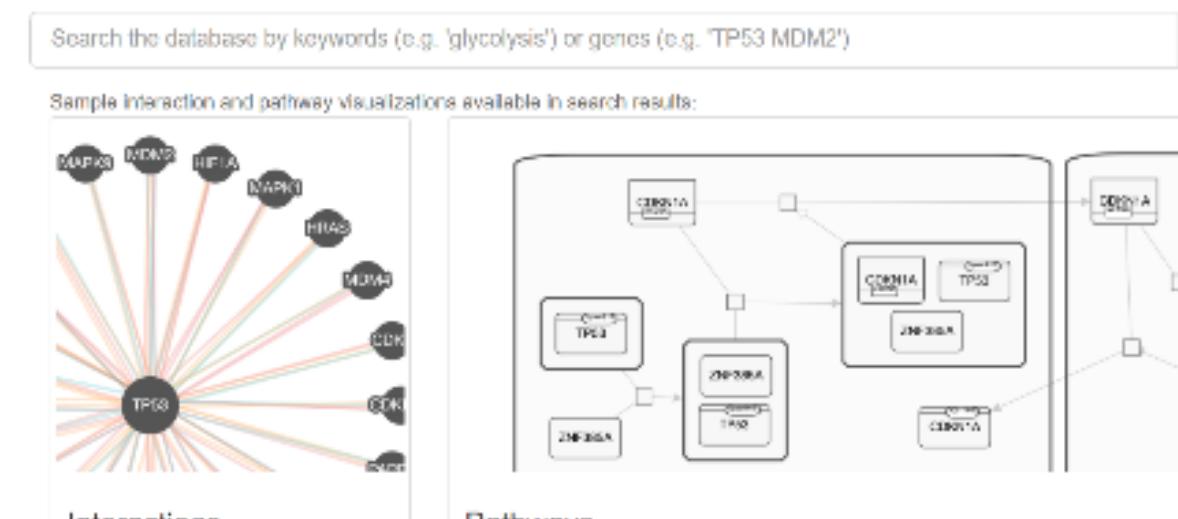
Pathway Commons

Access and discover data integrated from public pathway and interactions databases. 5772 Pathways – 2424055 Interactions – 22 Databases

Pathway Commons 2019 Update. Nucleic Acids Res (2019)
Author-sourced pathway capture using Biofactoid. eLife (2021)

Search the database by keywords (e.g. 'glycolysis') or genes (e.g. 'TP53 MDM2')

Sample interaction and pathway visualizations available in search results:



24



Example

- 3CLpro (protease of SARS-CoV-2)

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 ↗



25



Overview

Aspect	Details
Type of Data	Genetic and protein interactions, chemical-protein interactions, post-translational modifications
Scope	All major model organism species and humans
Data Source	Primary biomedical literature
Distribution	Freely distributed through partner model organism databases and meta-databases; directly downloadable in various formats
Curation System	Interaction Management System (IMS)
Curation Methods	Structured evidence codes, phenotype ontologies, gene annotation; utilization of semi-automated text mining approaches
Enhancements to Architecture	Improved support for various interaction and post-translational modification types, representation of complex multi-gene/protein interactions, accounting for cellular phenotypes through structured ontologies, expedited curation through text mining, enhanced quality control
Species with Virtually Complete Interactions	Budding yeast (<i>Saccharomyces cerevisiae</i>), thale cress (<i>Arabidopsis thaliana</i>), fission yeast (<i>Schizosaccharomyces pombe</i>)



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



Example

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.

 **HIPPIE** » Human Integrated Protein-Protein Interaction rEference

PROTEIN QUERY NETWORK QUERY BROWSE SCREEN ANNOTATION DOWNLOAD INFORMATION

Construction of a HIPPIE subnetwork from an input query set of proteins or interactions

Input a list of [proteins/interactions](#)

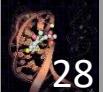
Query set

BRAF
MAP2K1
MAPK3

Example input:
dnmt3a dnmt3b

search (this may take a while)

Alternatively, choose a file to upload
 No file selected.



28



L.B.B

Laboratory of Systems Biology
and Bioinformatics

Example

- Filter (A)
- Filter (B)
- Filter (C)
- 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.

Colon - Sigmoid (selected)

B

Edge directionality (optional)
do not show direction
For shortest path edge direction, please, indicate sources and sinks:
Sources: BRAF
Sinks:

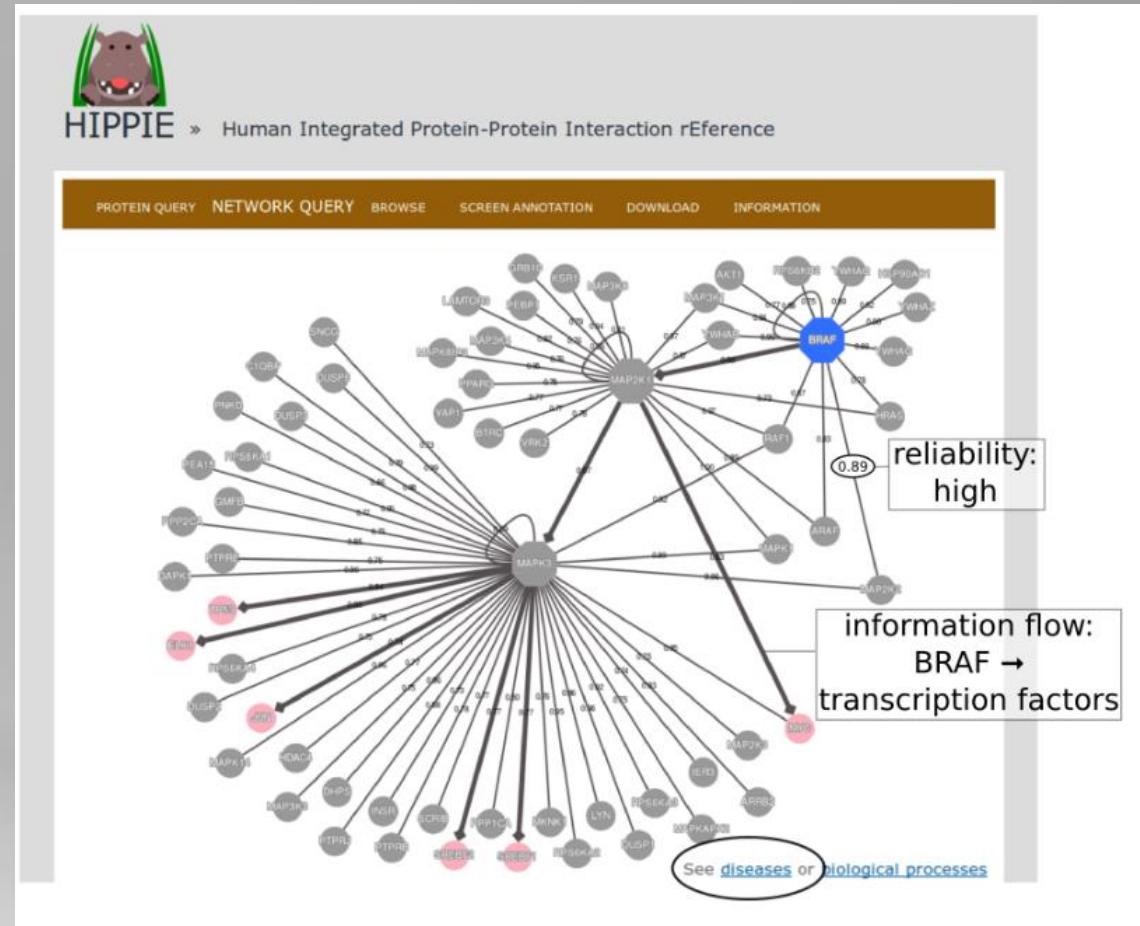
Add receptors to sources Add transcription factors to sinks

C



Example

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





Overview

Feature/Aspect	Details
Network Construction Mode	Users can query multiple proteins or interactions simultaneously in "batch mode" to construct a network. The resulting network can be visualized graphically, listed in a tabular format, or downloaded in various file formats
Input	Query with a list of proteins or interactions. - Max 100 lines allowed. - Pasted or uploaded as a file. - Identifier types: UniProt, gene symbols, or Entrez gene ids
Output Type	Choose: - Show in browser (text/visualization) - HIPPIE tab file format - PSI-MI TAB 2.5 output format
Interaction Layers	Modulate network size: - 0: Only input set interactions. - 1: All interactions with input set members. - Limitations on layers and input set size due to computing time. Script available for local use
Score Filter	Set confidence score threshold: - Custom value or predefined (medium/high). - Applied to HIPPIE interactions, not user-uploaded ones
Curation Methods	Structured evidence codes, phenotype ontologies, gene annotation; utilization of semi-automated text mining approaches
Interaction Type Filter	Distinguish binary and complex interactions using PSI-MI categories from MINT, BioGRID, and IntAct



L.B.B

Laboratory of Systems Biology
and Bioinformatics

dcGO



Faceted Search dcGO

HOME

BACKGROUND

MINING

[Faceted Search](#)

[PSnet](#)

[sTOL](#)

[dcGonet](#)

[dcGO Predictor](#)

[dcGO Enrichment](#)

[dcGO Povo](#)

[dcGOR^{New}](#)

[dcGO](#)

BROWSE

[SCOP Hierarchy](#)

[GO Hierarchy](#)

[BO Hierarchy](#)

[PFAM Hierarchy](#)

ALGORITHM

DOWNLOAD

CITATIONS

LINKOUT

[SUPERFAMILY](#)

<https://en.wikipedia.org/wiki/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.](#)

[How dcGO is built \(Algorithm\)](#)

- 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](#)).



L.B.B

Laboratory of Systems Biology
and Bioinformatics

dcGO

- Modular Design of Proteins
- Contribution of Protein Domains
- Ontological Terms and Protein Domains
- Operational Units and Functional Responsibility
- Associating Terms with Domain Combinations
- Developed Method for Detection

- 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



Example

dcGO christo
on functions, phenotypes, diseases and more

Faceted Search dcGO

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

BACKGROUND

[Faceted Search](#)
[PSnet](#)
[sTOL](#)
[dcGOnet](#)
[dcGO Predictor](#)
[dcGO Enrichment](#)
[dcGO Pevo](#)
[dcGOR^{New}](#)

MINING

[Faceted Search](#)
[PSnet](#)
[sTOL](#)
[dcGOnet](#)
[dcGO Predictor](#)
[dcGO Enrichment](#)
[dcGO Pevo](#)
[dcGOR^{New}](#)

BROWSE

[SCOP Hierarchy](#)
[GO Hierarchy](#)
[BO Hierarchy](#)
[PFAM Hierarchy](#)

ALGORITHM

DOWNLOAD

CITATIONS

LINKOUT

[SUPERFAMILY](#)
[SCOP](#)
[Gene Ontology](#)
[UniProt-GOA](#)
[KeyWords](#)
[PubMed](#)

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](#) [P](#)

3. **BO name:** [HTLV-I Infections](#) (CTD Diseases) [PSnet](#) [sTOL](#) [P](#)

4. **BO name:** [Env-5_glycoprotein, HTLV-I](#) (CTD Chemicals) [PSnet](#) [sTOL](#) [P](#)



Overview

dcGO Database Overview	
Focus	Recognizing the modular nature of proteins, emphasizing the significance of protein domains. Advocating for the association of ontological terms with protein domains for a comprehensive understanding of protein functionality.
Methodology	- Utilizes SCOP classifications at superfamily and family levels. - Relies on structural, sequence, and functional evidence. - Incorporates domain architectures, supra-domains, and co-occurrence analysis.



Overview

Hidden Markov Models (HMMs) in SCOP and SUPERFAMILY

Overview of HMMs

1. Representation of Protein Sequences: Proteins represented as sequences of amino acids. **2. Domain Detection:** HMMs trained to recognize characteristic patterns associated with SCOP domains.

3. Training the HMM: Trained on known protein sequences classified in SCOP. Learns statistical properties of sequences associated with each SCOP domain.

4. Hidden States: Certain states represent the underlying structure or classification of protein sequences. **5. Emission Probabilities:** Each hidden state emits observable symbols (amino acids) with specific probabilities.

6. Viterbi Algorithm: Used to find the most likely sequence of hidden states. **7. Scanning Protein Sequences:** Scans unknown protein sequences, assigns probabilities to hidden states.

8. Classification: Classifies protein sequence into different SCOP domains based on probabilities. **9. Domain Architecture:** Contributes to the definition of domain architectures, representing the sequential order of SCOP domains. **10. Supra-domains:** Identifies combinations of successive domains occurring in more than one distinct architecture.



Overview

SCOP Classifications in dcGO	
Source of Domain Classifications:	dcGO utilizes classifications from the Structural Classification Of Proteins (SCOP).
Focus Levels:	SCOP classifications used at both superfamily and family levels.
Criteria for Superfamily Level:	Based on structure, sequence, and function evidence of a common evolutionary ancestor.
Supra-domains and Co-occurrence	
Co-occurrence:	Certain protein domains tend to co-occur within the same multidomain protein.
Functional Synergy:	Co-occurring domains work synergistically to carry out specific biological functions.
Structural and Functional Coupling:	Co-occurrence related to structural and functional coupling of domains.
Co-evolution:	Evolution of one domain influenced by the evolution of another.



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, Clouser 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
 - Liver
 - Neck
 - Ovary
 - Pancreas
 - Prostate
 - Spleen
 - Stomach
 - Testis
 - Uterus



MatrisomeDB

- **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**

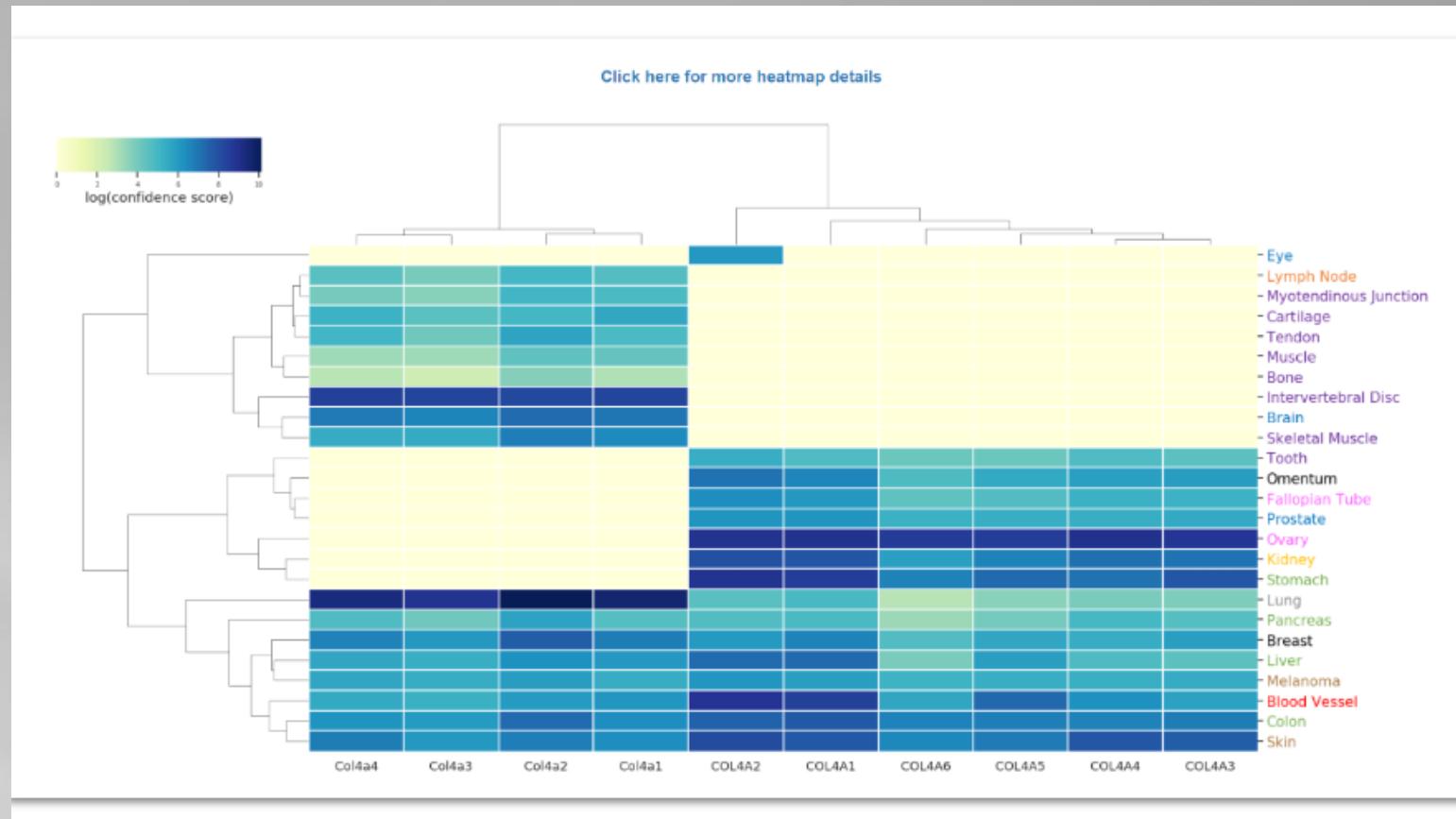
- **2022 Release (MatrisomeDB 2.0) Features**



L.B.B

Laboratory of Systems Biology
and Bioinformatics

Result





L.B.B

Laboratory of Systems Biology
and Bioinformatics

Result

Export all results to .tsv file

Export filtered results

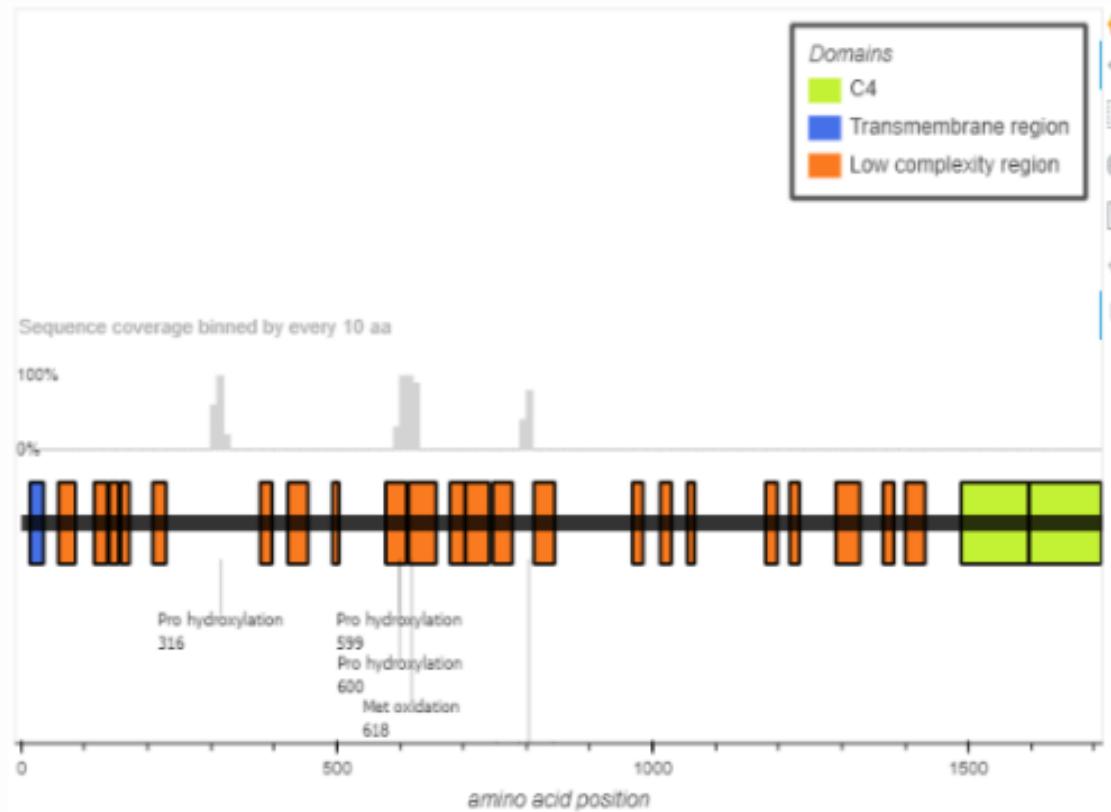
Gene	UniProt	Description	Sample type	Tissue	Species	Reference
COL4A2	P08572	Collagen alpha-2(IV) chain Open in SCV >	Retinal vascular basement membrane	Blood Vessel	Human	Uechi G et al., 2014
Col4a2	P08122	Collagen alpha-2(IV) chain Open in SCV >	Normal lung (QDSP, ECM)	Lung	Mouse	Schiller HB et al., 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

Result





Result

Sequence coverage map in Glomerular basement membrane

Note*

frequency: color > color > color > color > color

Sequence coverage: 3.6214953271028034%

MGRDQRAVAGPALRRWLLLGVTVGFLAQSVLAGVKKFDVPCGGRDCSGGCQCYPEKGGRGQPGPVGPQGYNGPPGLQGFPGLQGRKDGERGAPGVT
GPKGDVGARGVSGFFGADGIPGHPGQGGPRGRPGYDGCNGTQGDSPGQGPGSEGFTGPPGPQGPK GQKGEPYALPKEERDRYRGEPEGLVGFQGPP
GRPGHVGQMGPVGAPGRPGPPGPKGQQQNRLGFYGVKEKGVDVGQPGPNGIPSRTLHPIIAPTGVTFHPDQYKGEKGSEGEPEGIRGISLKGEEGIMG
FPGLRGYPGLSGEKGSPGQKGSRGLDGYQGP DGPRGPKGEAGDPGPPGLPAYSPHPSLAKGARGDPGFGPAQGEPGSQGEPEGDPGLPGLPGLSIGDDQ
RRGLPGEKGFIGDPGIPALYGGPPGPDKRGPPGPPGLPGLPGLFGLKGAKGRAGFPGLPGSPGARGPKGWKGDAECCRTEGDEAIKGL PGLP
GPKGFAGINGEPRKGDRGDPQHGFLPGFPGLKGVPGNIGAPGPKGAKGDSRTITKGERGQPGVPGVMKGGDGSPGRDGLDGFPLPGPPGDGIKCP
PDPGYPGPAGTKGTPGE MGPPGLGLPGLKGQRGFPGDAGLPGPPGFLGPPGPAGTPGQID CTDVKAvggdrqeaipgcicggpkglpglpgppgptga
KGLRGIPGFAGADGGPGPGRGLPGLDAGREGFPGPAGPGRGSKGVAVGLPGPDGSPGPGLPGLPGDPPGERGLPGEVLGAQPGPGRDAGVPGQPLKG
RCPPGFRGSQGMPGMPGLKGQPL PGPSGQPGLYGPPGLHGFPGAPGQEGPLGLPGLPREGGLPDRGDPGDTGAPGPVGMKGLSGDRGDAGFTGEQG
HPGSPGFKGIDGMPGTPGLKGDRGSPGMDGFQGMPGLKGRPGFPGSKGEAGFFGIPGLKGLAGEPGFKGSRGDPGPPGPPPVLPGMKDIK GEKGDEGPM
GLKGYLGAQGIQGMPGIPGLSGIPGLPGRPGHIKGVKGDIGVPGIPGLPGFPGVAGPPGIGFPFIFGSRGDKGAPGRAGLYGEIGATGDFGDIGDTINLPG
LKGERGTTGIPGLKFFFGEKGTEGDIFFGPGITGVGVQGPPGLKGQTGFPG LTGPPGSQGELGRIGLPGKGDDGWPAGPLPGFPGLRGIRGLHGLPGTK
FPGSPGSDIHGDPGPGPPGERGDPGEANTLPGPVGVPGQKGDQGAPGERGPPGSPGLQGFPGITPPSNISGAPGDKGAPGIFGLKGYRGPPGPPGSAALPG
SKGDTGNPGAPG TP GTKGWAGDSPGQGRPGVFLPGEKGPRGEQGFMNTGPTGAVGDRGPKGPKGDGPFGPAGPTVGA PGIAQPKIAVQPGTVGPQ
GRRGPPGAPGEMGPQGPPGEPGFRGAPGKAGPQGRGGVS AVPGFRGDEGPIGHQGPIGQEGAPGRPGSPGLPGMPGRS VSIGYLLVKHSQTDQEPMCPV
GMNKLWSGYSLYFEGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVYYASRNDKSYWLSTTAPLPMMPVAEDEIKPYISRCSVCEAPAIIAVHSQDV
HCPAGWRSLWIGYSFLMHTAAGDEGGGQLVSPGSCLEDF RATPFIECNGGRTCHYYANKYSFWLTIPQSFGSPSADTLKAGLIRTHISRCQVCMIKN



Overview

MatrisomeDB	
Overview	
Database Content	A searchable database integrating experimental proteomic data on the 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	The database includes a list of post-translational modifications identified for each entry.
Cross-Referencing	Live cross-referencing to gene and protein databases is available for ECM and ECM-associated genes.



Overview

Matrisome Project Details	
Project Origin	MatrisomeDB is part of the larger Matrisome Project initiated by Dr. Naba during her postdoctoral training at MIT.
Collaborators	Collaboration involves the Koch Institute for Integrative Cancer Research and the Proteomics Platform of the Broad Institute.
2019 Release Collaboration	The 2019 release resulted from collaboration between the Naba lab at the University of Illinois at Chicago (UIC) and the Gao lab in the UIC College of Pharmacy.
2022 Release Features (MatrisomeDB 2.0)	- Expanded content and enhanced data visualization features. - Peptide and post-translational modification mapping onto domain-based representation. - Inclusion of 3D structures of proteins.
Abstract-Mining Tool	MatrisomeDB 2.0 introduces an abstract-mining tool generating word clouds from study abstracts.
Community Involvement	Researchers are encouraged to contribute to MatrisomeDB by submitting recent ECM proteomics study results.
Contact Information	Researchers can contact matrisomeproject@gmail.com for data integration or contribute datasets and metadata via the provided link.
Collaborative Thriving	MatrisomeDB emphasizes collaborative efforts for the betterment of research in the field, promoting community growth.



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

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
MESIFHEKQEGSLCAQHCLNLLQGEYFSPVELSSIAHQLDDEERMRRMAEGGVTSEDYRTFLQQPSGNMDSGFFSIQVISNALK
VWGLEYLILNSPEYQRLRPNIFCNYKEHWFTVRKLGKQWFNLNSLLTGPELISDTYLALFLAQLQQEGYSIFVVKGDLPDCE
ADQLLQMIRVQQMHRPKLIGEALQKERRQRVHKTDLERVLEANDGSGMLDEDEDLQRALALSRSQEIMDMEEEADLRRAIQLS
MQGSSRNISQDMQTSGTNLTSEELRKRRAYFEKQQQQQQQQQQDLSGQSSHPCERTSSGALGSDLGDAMSEE
DMLQAAVTMSLETVRNDLKTGKK
```

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 (fIDPnn)
- B. Disordered linker prediction (DFLpred)
- C. MoRF prediction (MoRFchibiLight)
- 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)

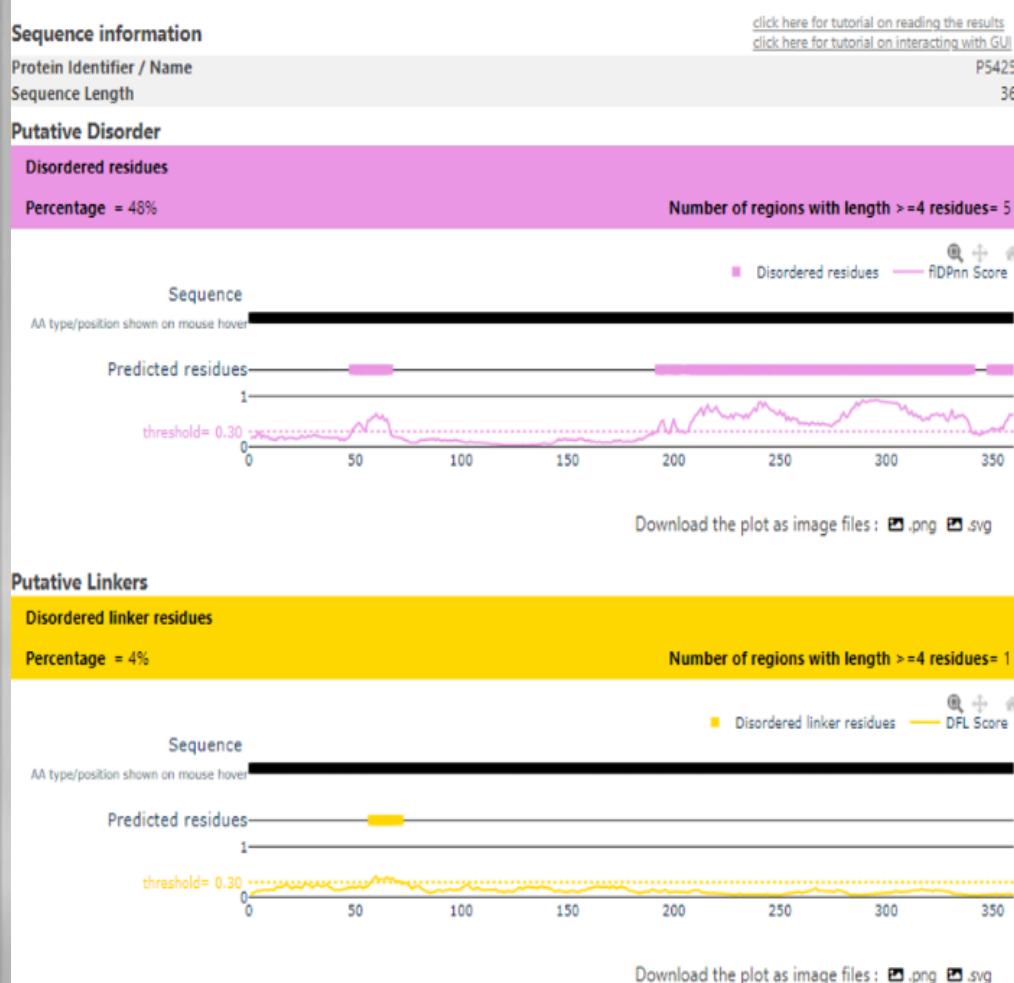




L.B.B

Laboratory of Systems Biology
and Bioinformatics

DEPICTER2



Protein level prediction result

Residue level prediction result



L.B.B

Laboratory of Systems Biology
and Bioinformatics

DEPICTER2

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





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 formated sequence for a query protein.

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



L.S.B.

Laboratory of Systems Biology
and Bioinformatics

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**

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 <input type="checkbox"/>
Putative disordered residues	Content = 47%	click here for details <input type="checkbox"/>
Native secondary structure	Content = H:16% E:24% C:60% (Coverage = 79%)	click here for details <input type="checkbox"/>
Putative secondary structure	Content = H:22% E:20% C:58%	click here for details <input type="checkbox"/>
Native buried residues	Content = 35% (Coverage = 71%)	click here for details <input type="checkbox"/>
Putative buried residues	Content = 24%	click here for details <input type="checkbox"/>

Putative Interactions with Ligands

Percentage of Protein-binding residues	20%	click here for details <input type="checkbox"/>
Percentage of DNA-binding residues	4%	click here for details <input type="checkbox"/>
Percentage of RNA-binding residues	0%	click here for details <input type="checkbox"/>

Other Properties

Signal peptide present	None	click here for details <input type="checkbox"/>
Percentage of highly conserved residues	10%	click here for details <input type="checkbox"/>
Percentage of disordered linker residues	3%	click here for details <input type="checkbox"/>
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 <input type="checkbox"/>



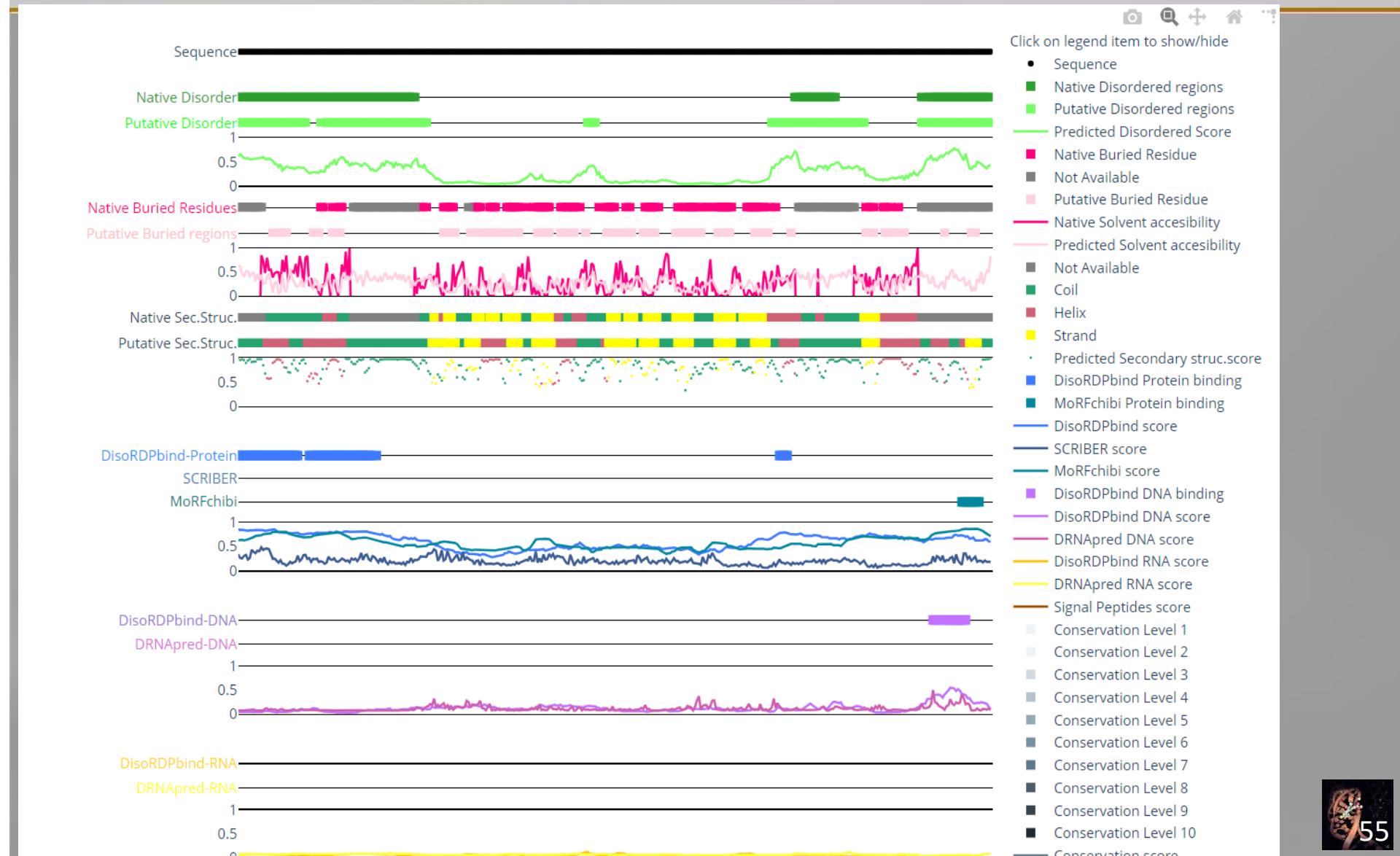


DescribePROT





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.

