



mpMoRFsDB: A database of Molecular Recognition Features (MoRFs) in membrane proteins.

Introduction

Molecular Recognition Features (MoRFs) are short, intrinsically disordered regions in proteins that undergo a disorder-to-order transition upon binding to their partners. MoRFs are implicated in protein-protein interactions, which serve as the initial step in molecular recognition.


The aim of this work was to collect, organize and store all membrane proteins that contain MoRFs. We focused in membrane proteins, as they constitute one third of fully sequenced proteomes and are responsible for a wide variety of cellular functions. Data were initially collected from Protein Data Bank (PDB) and Uniprot and were managed with Perl scripts. MoRFs were classified according to their secondary structure, after interacting with their partners. We identified MoRFs both in transmembrane and peripheral proteins. The position of transmembrane protein MoRFs was determined relative to a protein's topology.

All information was stored in a publicly available mySQL database with a user-friendly web interface. A Jmol applet is integrated for

vizualization of the structures. The utility of the database is the provision of information related to disordered based protein-protein interactions in membrane proteins. Such proteins play key roles in crucial biological functions and ca. 50% of them are putative hubs in protein interaction networks. The database will be updated on a regular basis by an automated procedure.

Home Page

In order to visit mpMoRFsDB, user should enter one of the following addresses: <http://bioinformatics.biol.uoa.gr/mpMoRFsDB/> or <http://bioinformatics.biol.uoa.gr/mpMoRFsDB/index.php>. The page loaded (see below) contains general information about the database and some statistics.



A database of Molecular Recognition Features
in Membrane Proteins

[Home](#) [Search](#) [Browse](#) [Blast Search](#) [Manual](#) [Download](#) [Contact](#)

mpMoRFsDB: A database of molecular recognition features (MoRFs) in membrane proteins.


Molecular Recognition Features (MoRFs) are short, intrinsically-disordered regions in proteins that undergo a disorder-to-order transition upon binding to their partners. MoRFs are implicated in protein-protein interactions, which serve as the initial step in molecular recognition.

The aim of this work was to collect, organize and store all membrane proteins that contain MoRFs. We focused in membrane proteins, as they constitute one third of fully sequenced proteomes and are responsible for a wide variety of cellular functions. Data were initially collected from Protein Data Bank (PDB) and Uniprot and were managed with Perl scripts. MoRFs were classified according to their secondary structure, after interacting with their partners. We identified MoRFs both in transmembrane and peripheral proteins. The position of transmembrane protein MoRFs was determined relative to a protein's topology.

All information was stored in a publicly available MySQL database with a user-friendly web interface. A Jmol applet is integrated for visualization of the structures. The utility of the database is the provision of information related to disordered based protein-protein interactions in membrane proteins. Such proteins play key roles in crucial biological functions and ca. 50% of them are putative hubs in protein interaction networks. The database will be updated on a regular basis by an automated procedure.

Statistics

Database Version:	1.0
Revision Date:	27 May 2013
Total Proteins:	173
Total MoRFs:	244

 University of Athens
Faculty of Biology
Biophysics & Bioinformatics Laboratory

Search data

In order to search database information, user should press the search button. A form appears with multiple options.

Protein Type: Single-Spanning
 Multi-Spanning
 Peripheral
 Any

Protein Contains MoRF Type: Alpha MoRF
 Beta MoRF
 Irregular MoRF
 Complex MoRF

MoRF Length: from to Residues

Protein Name:

Gene Name:

Organism Name:

Uniprot Accession:

PDB ID:

Note: If no filter is used, all entries will be retrieved.

The choices are:

- Search membrane proteins according to type.
 - Single-Spanning (Transmembrane)
 - Multi-Spanning (Transmembrane)
 - Peripheral
 - All the above
- Search proteins containing specific types of MoRFs. MoRFs are separated in four categories, according to their secondary structure upon binding to their partner.
 - Alpha MoRF
 - Beta MoRF
 - Irregular MoRF
 - Complex MoRF
- Search proteins according to MoRF length. The length varies between 10 and 70 residues.
- Search proteins based on protein name.
- Search proteins based on gene name.
- Search proteins based on organism.
- Search proteins based on Uniprot Accession.
- Search proteins based on PDB ID.

The search, based on protein name, gene name and organism does not require specific words. For example if user enters the word “hom”, the result is all proteins containing the word “hom” in the field organism.

Protein Type:	<input type="radio"/> Single-Spanning
	<input type="radio"/> Multi-Spanning
	<input type="radio"/> Peripheral
	<input checked="" type="radio"/> Any
Protein Contains MoRF Type:	<input type="checkbox"/> Alpha MoRF
	<input type="checkbox"/> Beta MoRF
	<input type="checkbox"/> Irregular MoRF
	<input type="checkbox"/> Complex MoRF
MoRF Length:	from <input type="text"/> to <input type="text"/> Residues
Protein Name:	<input type="text"/>
Gene Name:	<input type="text"/>
Organism Name:	<input type="text" value="hom"/>
Uniprot Accession:	<input type="text"/>
Pdb ID:	<input type="text"/>
	<input type="button" value="Submit"/> <input type="button" value="Reset"/>

Accession	Protein Name	Organism	Length	Type	
1 000555	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	Homo sapiens	2505	Multi-spanning	View
2 014745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	Homo sapiens	358	Peripheral	View
3 014836	Tumor necrosis factor receptor superfamily member 13B	Homo sapiens	293	Single-spanning	View
4 043521	Bcl-2-like protein 11	Homo sapiens	198	Peripheral	View
5 060331	Phosphatidylinositol 4-phosphate 5-kinase type-1 gamma	Homo sapiens	668	Peripheral	View
6 075154	Rab11 family-interacting protein 3	Homo sapiens	756	Peripheral	View
7 P01130	Low-density lipoprotein receptor	Homo sapiens	860	Single-spanning	View
8 P01133	Pro-epidermal growth factor	Homo sapiens	1207	Single-spanning	View
9 P01135	Protransforming growth factor alpha	Homo sapiens	160	Single-spanning	View
10 P01730	T-cell surface glycoprotein CD4	Homo sapiens	458	Single-spanning	View
11 P01892	HLA class I histocompatibility antigen, A-2 alpha chain	Homo sapiens	365	Single-spanning	View
12 P02686	Myelin basic protein	Homo sapiens	304	Peripheral	View
13 P04233	HLA class II histocompatibility antigen gamma chain	Homo sapiens	296	Single-spanning	View
14 P05067	Amyloid beta A4 protein	Homo sapiens	770	Single-spanning	View
15 P05106	Integrin beta-3	Homo sapiens	788	Single-spanning	View
16 P07204	Thrombomodulin	Homo sapiens	575	Single-spanning	View
17 P08138	Tumor necrosis factor receptor superfamily member 16	Homo sapiens	427	Single-spanning	View
18 P11362	Fibroblast growth factor receptor 1	Homo sapiens	822	Single-spanning	View
19 P11717	Cation-independent mannose-6-phosphate receptor	Homo sapiens	2491	Single-spanning	View
20 P11836	B-lymphocyte antigen CD20	Homo sapiens	297	Multi-spanning	View

All the above search fields can be combined in order to make the search result as specific as possible.

For example if we make the following combined search:

Protein Type:	<input type="radio"/> Single-Spanning <input type="radio"/> Multi-Spanning <input type="radio"/> Peripheral <input checked="" type="radio"/> Any
Protein Contains MoRF Type:	<input type="checkbox"/> Alpha MoRF <input checked="" type="checkbox"/> Beta MoRF <input type="checkbox"/> Irregular MoRF <input type="checkbox"/> Complex MoRF
MoRF Length:	from <input type="text"/> to <input type="text"/> Residues
Protein Name:	<input type="text"/>
Gene Name:	<input type="text"/>
Organism Name:	<input type="text" value="homo"/>
Uniprot Accession:	<input type="text"/>
Pdb ID:	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

We get only one protein with the specific characteristics.

Accession	Protein Name	Organism	Length	Type	
1 Q12879	Glutamate receptor ionotropic, NMDA 2A	Homo sapiens	1464	Multi-spanning	View

Another example is to find all single spanning membrane proteins containing MoRFs from 20 to 30 residues.

Protein Type: Single-Spanning
 Multi-Spanning
 Peripheral
 Any

Protein Contains MoRF Type: Alpha MoRF
 Beta MoRF
 Irregular MoRF
 Complex MoRF

MoRF Length: from to Residues

Protein Name:

Gene Name:

Organism Name:

Uniprot Accession:

Pdb ID:

And the result is

	Accession	Protein Name	Organism	Length	Type	
1	P03437	Hemagglutinin	Influenza A virus (strain A/Aichi/2/1968 H3N2)	566	Single-spanning	View
2	P36941	Tumor necrosis factor receptor superfamily member 3	Homo sapiens	435	Single-spanning	View
3	P11362	Fibroblast growth factor receptor 1	Homo sapiens	822	Single-spanning	View
4	P15702	Leukosialin	Mus musculus	395	Single-spanning	View
5	Q16611	Bcl-2 homologous antagonist/killer	Homo sapiens	211	Single-spanning	View
6	P46097	Synaptotagmin-2	Mus musculus	422	Single-spanning	View
7	Q07817	Bcl-2-like protein 1	Homo sapiens	233	Single-spanning	View
8	P70452	Syntaxin-4	Mus musculus	298	Single-spanning	View
9	O08734	Bcl-2 homologous antagonist/killer	Mus musculus	208	Single-spanning	View
10	Q61391	Nepriylisin	Mus musculus	750	Single-spanning	View
11	P15379	CD44 antigen	Mus musculus	778	Single-spanning	View
12	P01730	T-cell surface glycoprotein CD4	Homo sapiens	458	Single-spanning	View
13	P25445	Tumor necrosis factor receptor superfamily member 6	Homo sapiens	335	Single-spanning	View
14	P18962	Dipeptidyl aminopeptidase B	Saccharomyces cerevisiae (strain ATCC 204508 / S288c)	818	Single-spanning	View
15	Q63155	Netrin receptor DCC	Rattus norvegicus	1445	Single-spanning	View
16	Q01705	Neurogenic locus notch homolog protein 1	Mus musculus	2531	Single-spanning	View
17	Q8NPF91	Nesprin-1	Homo sapiens	8797	Single-spanning	View
18	Q8WXH0	Nesprin-2	Homo sapiens	6885	Single-spanning	View

If we want to select all proteins containing irregular or complex MoRFs, we make the following selection:

Protein Type: Single-Spanning
 Multi-Spanning
 Peripheral
 Any

Protein Contains MoRF Type: Alpha MoRF
 Beta MoRF
 Irregular MoRF
 Complex MoRF

MoRF Length: from to Residues

Protein Name:

Gene Name:

Organism Name:

Uniprot Accession:

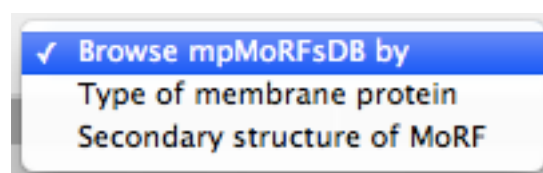
Pdb ID:

And the result is:

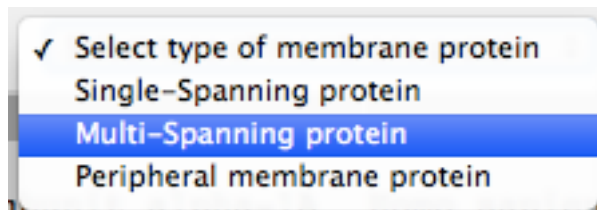
Accession	Protein Name	Organism	Length	Type	
1 P03437	Hemagglutinin	Influenza A virus (strain A/Aichi/2/1968 H3N2)	566	Single-spanning	<input type="button" value="View"/>
2 P26663	Genome polyprotein	Hepatitis C virus genotype 1b (isolate BK)	3010	Single-spanning	<input type="button" value="View"/>
3 P31431	Syndecan-4	Homo sapiens	198	Single-spanning	<input type="button" value="View"/>
4 O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	Homo sapiens	358	Peripheral	<input type="button" value="View"/>
5 P11836	B-lymphocyte antigen CD20	Homo sapiens	297	Multi-spanning	<input type="button" value="View"/>
6 P70452	Syntaxin-4	Mus musculus	298	Single-spanning	<input type="button" value="View"/>
7 Q2YHF0	Genome polyprotein	Dengue virus type 4 (strain Thailand/0348/1991)	3387	Peripheral	<input type="button" value="View"/>
8 O80297	Attachment protein G3P	Enterobacteria phage If1	460	Single-spanning	<input type="button" value="View"/>
9 Q61391	Neprilysin	Mus musculus	750	Single-spanning	<input type="button" value="View"/>
10 P15379	CD44 antigen	Mus musculus	778	Single-spanning	<input type="button" value="View"/>
11 P20963	T-cell surface glycoprotein CD3 zeta chain	Homo sapiens	164	Single-spanning	<input type="button" value="View"/>
12 P18962	Dipeptidyl aminopeptidase B	Saccharomyces cerevisiae (strain ATCC 204508 / S288c)	818	Single-spanning	<input type="button" value="View"/>
13 P03524	Glycoprotein G	Rabies virus (strain ERA)	524	Single-spanning	<input type="button" value="View"/>
14 Q12879	Glutamate receptor ionotropic, NMDA 2A	Homo sapiens	1464	Multi-spanning	<input type="button" value="View"/>
15 P26662	Genome polyprotein	Hepatitis C virus genotype 1b (isolate Japanese)	3010	Single-spanning	<input type="button" value="View"/>
16 Q5UB51	Genome polyprotein	Dengue virus type 3 (strain Singapore/8120/1995)	3390	Peripheral	<input type="button" value="View"/>
17 O88917	Latrophilin-1	Rattus norvegicus	1515	Multi-spanning	<input type="button" value="View"/>

Browse

User can browse data by pressing the browse button. At first all entries appear. User has the ability to apply filters and browse the database by type of membrane protein or by MoRF type.



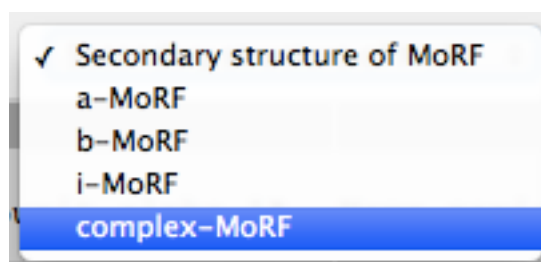
For example if the user selects Multi-Spanning proteins as membrane type.



All Multi-Spanning proteins are retrieved.

All Multi-Spanning proteins						
Accession	Protein Name	Organism	Length	Type		
1 000555	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	Homo sapiens	2505	Multi-spanning		View
2 008675	Proteinase-activated receptor 3	Mus musculus	369	Multi-spanning		View
3 088634	Proteinase-activated receptor 4	Mus musculus	396	Multi-spanning		View
4 088917	Letrophilin-1	Rattus norvegicus	1515	Multi-spanning		View
5 P04480	Colicin-A	Citrobacter freundii	592	Multi-spanning		View
6 P05504	ATP synthase subunit a	Rattus norvegicus	226	Multi-spanning		View
7 P06935	Genome polyprotein	West Nile virus	3430	Multi-spanning		View
8 P07293	Voltage-dependent L-type calcium channel subunit alpha-1S	Oryctolagus cuniculus	1873	Multi-spanning		View
9 P0A921	Phospholipase A1	Escherichia coli (strain K12)	289	Multi-spanning		View
10 P11716	Ryanodine receptor 1	Oryctolagus cuniculus	5037	Multi-spanning		View
11 P11836	B-lymphocyte antigen CD20	Homo sapiens	297	Multi-spanning		View
12 P18011	Invasin IpaB	Shigella flexneri	580	Multi-spanning		View
13 P19634	Sodium/hydrogen exchanger 1	Homo sapiens	815	Multi-spanning		View
14 P20781	Glycine receptor subunit beta	Rattus norvegicus	496	Multi-spanning		View
15 P22002	Voltage-dependent L-type calcium channel subunit alpha-1C	Rattus norvegicus	2169	Multi-spanning		View
16 P25116	Proteinase-activated receptor 1	Homo sapiens	425	Multi-spanning		View
17 P27884	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	Oryctolagus cuniculus	2424	Multi-spanning		View

If the user selects complex-MoRFs.



All proteins containing complex-MoRFs are retrieved.

All membrane proteins containing complex-MoRFs						
Accession	Protein Name	Organism	Length	Type		
1 P03437	Hemagglutinin	Influenza A virus (strain A/Aichi/2/1968 H3N2)	566	Single-spanning		View
2 P31431	Syndecan-4	Homo sapiens	198	Single-spanning		View
3 O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	Homo sapiens	358	Peripheral		View
4 P11836	B-lymphocyte antigen CD20	Homo sapiens	297	Multi-spanning		View
5 P70452	Syntaxin-4	Mus musculus	298	Single-spanning		View
6 Q2YHF0	Genome polyprotein	Dengue virus type 4 (strain Thailand/0348/1991)	3387	Peripheral		View
7 P20963	T-cell surface glycoprotein CD3 zeta chain	Homo sapiens	164	Single-spanning		View

Entries

When user fills the search fields and presses submit a new page appears, with the list of proteins like the image below. It contains the **Uniprot Accession number**, the **protein name**, the **organism**, the **protein length** and the **type** of protein.


Accession	Protein Name	Organism	Length	Type	
1 000555	Voltage-dependent P/Q-type calcium channel subunit alpha-1A	Homo sapiens	2505	Multi-spanning	View
2 008675	Proteinase-activated receptor 3	Mus musculus	369	Multi-spanning	View
3 008734	Bcl-2 homologous antagonist/killer	Mus musculus	208	Single-spanning	View
4 014745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	Homo sapiens	358	Peripheral	View
5 014836	Tumor necrosis factor receptor superfamily member 13B	Homo sapiens	293	Single-spanning	View
6 043521	Bcl-2-like protein 11	Homo sapiens	198	Peripheral	View
7 054918	Bcl-2-like protein 11	Mus musculus	196	Peripheral	View
8 060331	Phosphatidylinositol 4-phosphate 5-kinase type-1 gamma	Homo sapiens	668	Peripheral	View
9 070161	Phosphatidylinositol 4-phosphate 5-kinase type-1 gamma	Mus musculus	661	Peripheral	View
10 075154	Rab11 family-interacting protein 3	Homo sapiens	756	Peripheral	View
11 080297	Attachment protein G3P	Enterobacteria phage T4	460	Single-spanning	View
12 088597	Beclin-1	Mus musculus	448	Peripheral	View
13 088634	Proteinase-activated receptor 4	Mus musculus	396	Multi-spanning	View
14 088917	Latrophilin-1	Rattus norvegicus	1515	Multi-spanning	View
15 091734	Genome polyprotein	Echovirus 1 (strain Human/Egypt/Farouk/1951)	2184	Peripheral	View
16 P01130	Low-density lipoprotein receptor	Homo sapiens	860	Single-spanning	View
17 P01133	Pro-epidermal growth factor	Homo sapiens	1207	Single-spanning	View
18 P01135	Protransforming growth factor alpha	Homo sapiens	160	Single-spanning	View
19 P01730	T-cell surface glycoprotein CD4	Homo sapiens	458	Single-spanning	View
20 P01892	HLA class I histocompatibility antigen, A-2 alpha chain	Homo sapiens	365	Single-spanning	View
21 P02686	Myelin basic protein	Homo sapiens	304	Peripheral	View
22 P03300	Genome polyprotein	Poliovirus type 1 (strain Mahoney)	2209	Peripheral	View
23 P03302	Genome polyprotein	Poliovirus type 3 (strains P3/Leon/37 and P3/Leon 12A[1]B)	2206	Peripheral	View
24 P03303	Genome polyprotein	Human rhinovirus 14	2179	Peripheral	View
25 P03313	Genome polyprotein	Coxsackievirus B3 (strain Nancy)	2185	Peripheral	View
26 P03316	Structural polyprotein	Sindbis virus	1245	Single-spanning	View
27 P03437	Hemagglutinin	Influenza A virus (strain A/Aichi/2/1968 H3N2)	566	Single-spanning	View
28 P03524	Glycoprotein G	Rabies virus (strain ERA)	524	Single-spanning	View

Entry

When the user presses the View button he is redirected to the entry page. The entry page contains information about the type and topology of the protein, the total number of MoRFs as well as their position in the protein. Moreover a Jmol Viewer is integrated for visualization of the proteins in 3D.

FASTA TXT XML

Protein Name: Na(+)/H(+) exchange regulatory cofactor NHE-RF1
Gene Name: SLC9A3R1
Organism: Homo sapiens
NCBI Taxonomy: 9606
Sequence: MSADAAGAFLPRLCCLEKPGNGYGFHLHGEKGLGQYIRLVEFGSPAEKAGLLAGDRLVEVNGENVEKETHQQVVSIRAAALNAVRLLV
 VDPEDEQLKRLGVQVRELLRAQEAFGQAEFPAAAEVQAGNENEPREADKSHPEQRRLRPRLCMTMKKGFSGYFNLHSDKSPGGQFIR
 SVDDPFAASGLAQRIVYVNGVCMGKQKQEDVSAIRAGGDETKLLVDRRETFEFKCRVIFPSQHLNGLPLVFFVTNGEIQKENS
 EALAEALLESFRPALVRSASDTSSEELNSQSPFPKQSTAPSSSTSSSDPILDFNLSLMAKERARHKRRSSRAPQMDWSKXNELFSL
Sequence Length: 358 aa
Uniprot Accession: O14745
Type: Peripheral
Total MoRFs: 1



Jmol

PhosphoSite: O14745
Pfam: PF09007 PF00595
PROSITE: P550106
InterPro: IPR015098 IPR017300 IPR001478
SUPFAM: SSF50156
IntAct: O14745
DIP: DIP-29092N
MINT: MINT-4998796
STRING: 9606.ENSP00000262613
OMIM: 604990 612287
DrugBank:

MoRF PDB:Chain	MoRF Start	MoRF End	Type	Partner	
2D10:E	330	357	irregular	2D10:A (P26043)	3D
2D10:F	330	357	complex	2D10:B (P26043)	3D
2D10:G	330	357	complex	2D10:C (P26043)	3D
2D10:H	330	357	complex	2D10:D (P26043)	3D

References

- Hamada K, Shimizu T, Matsui T, Tsukita S, Hakoshima T. Structural basis of themembrane-targeting and unmasking mechanisms of the radixin FERM domain. *EMBO J*.2000 Sep 1;19(17):4449-62. PubMed PMID: 10970839; PubMed Central PMCID:PMC302071.
- Terawaki S, Maesaki R, Okada K, Hakoshima T. Crystallographic characterizationof the radixin FERM domain bound to the C-terminal region of the humanNa⁺/H⁺-exchanger regulatory factor (NHERF). *Acta Crystallogr D Biol Crystallogr*. 2003 Jan;59(Pt 1):177-9. Epub 2002 Dec 19. PubMed PMID: 12499563.
- Hamada K, Shimizu T, Yonemura S, Tsukita S, Tsukita S, Hakoshima T. Structuralbasis of adhesion-molecule recognition by ERM proteins revealed by the crystalstructure of the radixin-ICAM-2 complex. *EMBO J*. 2003 Feb 3;22(3):502-14. PubMed PMID: 12554651; PubMed Central PMCID: PMC140724.
- Terawaki S, Maesaki R, Hakoshima T. Structural basis for NHERF recognition by ERM proteins. *Structure*. 2006 Apr;14(4):777-89. PubMed PMID: 16615918.

More specifically:

- In the top right corner user can find three buttons. FASTA, TXT and XML. By pressing these buttons user can download the sequence in FASTA format, all page information in text format or all information in XML format respectively.
- The protein information available is:
 - Protein Name
 - Gene Name
 - Organism
 - NCBI taxonomy
 - Sequence
 - Sequence length
 - Uniprot Accession
 - Protein Type
 - Total number of MoRFs

- The next box contains references to other databases. The databases are:
 - [Phosphosite](#)
 - [Pfam](#)
 - [Prosite](#)
 - [InterPro](#)
 - [SupFam](#)
 - [Intact](#)
 - [DIP](#)
 - [MINT](#)
 - [STRING](#)
 - [OMIM](#)
 - [Drugbank](#)
- The next box contains MoRF related information. It contains the MoRF PDB ID and chain, where the MoRF starts and ends in relation to protein's sequence and the partner of the MoRF.
- The next box contains references related to the PDB IDs appearing in the previous box.
- Last but not least a Jmol 3D viewer is integrated in the page for vizualization of the structures. MoRFs are colored red. User can chose complexes by pressing the "3D" button/s.

The topology of Single Spanning and Multi Spanning proteins was determined as well as the position of MoRFs. In the screenshot below user can view an example of topology (Uniprot Accession: P01730).

```

MNRGVFRHLLLVQLALLPAATQGKRVVLGKGGDTVELTCTASQKKSIFHWKNSNQIKILGNQGSFLTKGPKSLNDRADSRSLWDQGNFPLIINKLKI ESDTYICEVEDQKKEVQLLVFGLTANS DTHLLQGQSLTTLTLESPFGSS
.....
PSVQCRSPRGKNIQGGKTLVSQLELQDSCTWTCTVLQNKQKVEFKIDIVVLAFAQASSIVYKKEGEQVFSFPLAFTVEKLTGSGELAWQAERASSKSWITFDLNKKEVSVKRVTDPKLQMGKLP LHLTLQALPQYAGSGLT LA
.....
LEAKTGKLBQEVNLLVMRATQLQKNTCEVWGPTSPKMLSLKLENKEAKVSKREKAVVWLNPEAGMWQCILLSDSGVLLSNIKVLPTWSTFVQPMALIVLGGVAGLLFI GLGIFFCVRCRHRRRQAERMSQIKRLLSEKKTQCQPHR
.....#####
FQKTCSPF
iiiiiii
-----

```

- s: Signal peptide
- o: Extracellular
- i: Cytoplasmic
- M: Transmembrane
- #: Position of MoRF

BLAST search

With the BLAST search tool, the user may submit a sequence and search the database for finding homologues. The input for the BLAST application is the sequence in standard FASTA format and the user has the ability to specify an e-value cutoff level to use in the query:

BLAST SEARCH

Paste your sequence in **FASTA** format in the field provided

```
>sp|O00555|CAC1A_HUMAN Voltage-dependent P/Q-type calcium channel subunit alpha-1A
OS=Homo sapiens GN=CACNA1A PE=1 SV=2
MARFGDEMPARYGGGGSGAAAGVVVGGGGGRGAGGSRQGGQPGAQRMYKQSMARARTMA
LYNPVVRQNCCLTVNRSLFLFSEDNVVRKYAKKITFEWPPFEYMLATIIANCIVLALAEQH
LPDDDDKTPMSERLDDTEPYFIGIFCFEAGIKIALGFHFKGSYLRNGWNVMDFFVVLTG
ILATVGTFFDLRTLRAVRVLRPLKLVSGIPSLQVVLKSIMKAMIPLLQIGLLLFFAILIF
AIIGLEFYMGKPFHTTCFEEGTDDIQGESAPCGTEEPARTCPNGTKCQPYWEGPNNGITQ
FDNILFAVLTVFQCITMEGWTDLLYNSNDASGNTWNWLYFIPLIHGSFFMLNLVLGVLS
GEFAKERERVENRRAFKLRRQQQIERELNGYMEWISKAEEVILAEDETGEQRHPPFDGA
LRRTTIKKSKTDLNPEEAEDQLADIASVGSPPFARASIKSAKLENSTFFHKKERRMRFYI
RRMVKTQAFYWTVLSLVALNNTLCVAIVHYNQPEWLSDFLYAEFIFLGLFMSEMFIKMYG
LGTRPYFHSSFNCFDCGVHGSIFEVIWAVIKPGTSFGISVLRALRLLRIFKVTKYWASL
RNLVVSLLNSMKSIISLLFLLFLFIVVFALLGMQLFGGQFNFDGTPPTNFDTFPAAIMT
VFQILTGEDWNEVMYDGIKSQGGVQGMVFSYFIVLTLFGNYTLLNVFLAIAVDNLANA
QELTKDEQEEEEANQKLALQKAKEVAEVSPLSAANMSIAVKEQQKNQKPAKSVWEQRTS
EMRKQNLASREALYNEMDPDERWKAAYTRHLRPMKTHLDRPLVVDPPQENRNNNTNKS
AAEPTVDQRLGQQRADFRLKQARYHDRARDPSGSAGLDARRPWAGSQEAELSREGPYGR
ESDHHAREGSLEQPGFWEGEAERKAGDPHRRHVHRQGGRESRSRSGSPRTGADGEHRRHR
AHRRPGEEGPEDKAERRARHREGSRPARGGEGEGEGPDGGERRRRHRHGAPATYEGDARR
EDKERRHRRRKENQGGVPPVSGPNLSTTRPIQQDLGRQDPPLAEDIDNMKNNKLATAESA
APHGSLGHAGLPOSPAKMGNSTDPGPMLAIPAMATNPONAASRRTPNPNPGNPSNPGPPKT
```

Specify an e-value cutoff level to use in your query:

The result page of the BLAST search shows a list of the Blast hits with significant alignment on the query sequence the user has submitted. The list is in a table format including the **mpMoRFsDB_ID** of the target protein, the **Length** of the target sequence and the Query and Target **align range**. The BLAST results can be compared through the **Score** and **E-value** and the **Identities** and **Positives**.

The result page of the above BLAST search is:

Blast search results										INPUT FILE	OUTPUT FILE
◆ splO00555ICAC1A_HUMAN Voltage-dependent P/Q-type calcium channel subunit alpha-1A											
Align with	Length	Hsp hit	Score	E-value	Query align range	Subject align range	Align length	Identities	Positives	Gaps	Show/hide alignment
ID: O00555	2505	1	13549	0	[26-2530]	[1-2505]	2505	2505/2505 (100%)	2505/2505 (100%)	0/2505 (0%)	Show/Hide
ID: P27884	2424	1	10291	0	[26-2232]	[1-2215]	2220	2083/2220 (94%)	2129/2220 (96%)	18/2220 (1%)	Show/Hide
ID: Q05152	2339	1	6970	0	[72-2338]	[44-2221]	2321	1478/2321 (64%)	1680/2321 (72%)	197/2321 (8%)	Show/Hide
ID: Q15878	2313	1	3340	0	[1176-2338]	[1054-2185]	1187	732/1187 (62%)	855/1187 (72%)	79/1187 (7%)	Show/Hide
--	--	2	2891	0	[26-825]	[1-803]	825	585/825 (71%)	664/825 (80%)	47/825 (6%)	Show/Hide
--	--	3	470	6.417e-48	[1248-1845]	[68-715]	687	168/687 (24%)	308/687 (45%)	128/687 (19%)	Show/Hide
--	--	4	302	1.77086e-28	[462-751]	[1440-1732]	309	78/309 (25%)	146/309 (47%)	35/309 (11%)	Show/Hide
--	--	5	281	4.54892e-26	[102-356]	[1132-1385]	265	81/265 (31%)	132/265 (50%)	21/265 (8%)	Show/Hide
--	--	6	208	1.51608e-17	[123-388]	[1478-1726]	294	67/294 (23%)	119/294 (40%)	73/294 (25%)	Show/Hide
--	--	7	181	1.77807e-14	[544-703]	[1186-1382]	199	53/199 (27%)	93/199 (47%)	41/199 (21%)	Show/Hide
--	--	8	145	2.72431e-10	[1258-1586]	[465-760]	339	81/339 (24%)	157/339 (46%)	53/339 (16%)	Show/Hide
ID: Q07652	2222	1	3242	0	[1176-2501]	[1006-2192]	1347	747/1347 (55%)	883/1347 (66%)	181/1347 (13%)	Show/Hide
--	--	2	2811	0	[84-825]	[1-754]	764	559/764 (73%)	637/764 (83%)	32/764 (4%)	Show/Hide
--	--	3	463	3.5201e-47	[1248-1845]	[19-666]	687	169/687 (25%)	310/687 (45%)	128/687 (19%)	Show/Hide
--	--	4	300	3.02064e-28	[462-751]	[1392-1684]	309	78/309 (25%)	146/309 (47%)	35/309 (11%)	Show/Hide
--	--	5	282	3.25809e-26	[102-356]	[1084-1337]	265	82/265 (31%)	132/265 (50%)	21/265 (8%)	Show/Hide
--	--	6	200	1.12304e-16	[123-388]	[1430-1678]	294	66/294 (22%)	119/294 (40%)	73/294 (25%)	Show/Hide
--	--	7	179	2.81351e-14	[544-703]	[1138-1334]	199	53/199 (27%)	93/199 (47%)	41/199 (21%)	Show/Hide
--	--	8	146	2.3444e-10	[1258-1586]	[416-711]	339	82/339 (24%)	157/339 (46%)	53/339 (16%)	Show/Hide
ID: P22002	2169	1	1931	0	[1217-2002]	[887-1668]	809	390/809 (48%)	535/809 (66%)	50/809 (6%)	Show/Hide

Furthermore, the user can have a more detailed view of each alignment through the **Show/Hide** button at the end of each line:

ID: Q07652	2222	1	3242	0	[1176-2501]	[1006-2192]	1347	747/1347 (55%)	883/1347 (66%)	181/1347 (13%)	Show/Hide
Query	1176SGTQNSAKTARKPDHTTVDIPACPPPLNHTVVQVKNANPDPLP-----KKEEEKK1235										
Midline	S TN K + TV IP P ++ TVV ++ + + P ++EEE +										
Target	1006SCMTTNDKATTESVTVVAIPDV-DPLVDSTVVNISNKTGGEASPLKEAETKEEEEEVE1065										
Query	1236EEEEDRGEDGPKMPFPYSSMFILSTTNPLRRLCHYILNLRYPFEMCILMVIAMSSIALAA1295										
Midline	++++ + K M P+SSMFI STTNP+R+ CHYI+NLRYPFEMCIL+VIA SSIALAA										
Target	1066KKKQKKEKRETKAMVPHSSMFPSTTNPIRKACHYIVNLRYPFEMCILLVIAASSIALAA1125										
Query	1296EDPVQPNAPRNNVLRFDYVFTGVFTFEMVIKIDLGLVHQGAYFRDLWNILDFIVVSG1355										
Midline	EDPV N+ RN VLRFDYVFTGVFTFEMVIKID GL+L G+YFRDLWNILDF+VV G										
Target	1126EDPVLTNSEKRVLRFDYVFTGVFTFEMVIKIDQGLILQDGSYFRDLWNILDFVVG1185										
Query	1356ALVAFAFT---GNSKGRDINTIKSLRVLRLRPLKTIKRLPKLKAFCVNVSLKNVFN1415										
Midline	ALVAFA G +KG+DI TIKSLRVLRLRPLKTIKRLPKLKAFCVNV SLKNVFN										
Target	1186ALVAFALANALGTNKRDIKTIKSLRVLRLRPLKTIKRLPKLKAFCVNVSLKNVFN1245										
Query	1416LIVYMLFMPFIFAVVAVQLFKGKFFHCTDESKEPEKDCRGKYLLEYKNEVKARDREWKYE1475										
Midline	LIVY LFMFIFAV+AVQLFKGKFF+CTD SK+ EK+C G Y+ +EKN+++ + REWK++E										
Target	1246LIVYKLFMPFIFAVVAVQLFKGKFFYCTDSSKDETEKICIGNYVDHEKNKMEVKGREWKRHE1305										
Query	1476PHYDNLWALLTLFTVSTGEGWPQVLKHSVDATPENQGPSGYRMEMSIFYVVYFVVFPP1535										
Midline	PHYDN++WALLTLFTVSTGEGWPQVL+HSVD T E++GPS RMEMSIFYVVYFVVFPP										
Target	1306PHYDNIWALLTLFTVSTGEGWPQVLKHSVDVTEEDRGPSSRNRMEMSIFYVVYFVVFPP1365										
Query	1536FFVNIFFVALIIITFQEQGDKMMEESLEKNERACIDFAISAKPLTRHMPQNKQSFQYRMW1595										
Midline	FFVNIFFVALIIITFQEQGDKMMEESLEKNERACIDFAISAKPLTR+MPQN+ +FQYR+W										
Target	1366FFVNIFFVALIIITFQEQGDKMMEESLEKNERACIDFAISAKPLTRYMPQNRHTFQYRVW1425										

Manual

User can view and download the current manual.

Download

User can download all database files in Text or XML format. The option will be activated upon publication.

Contact

Users can contact us for more information at the emails specified at the contact page. Related publications to the current work are also presented.

Technologies

mpMoRFsDB is based on modern technologies. User should have Javascript enabled on the web browser. In order to use the Jmol Viewer user should also have Java installed on his computer.



University of Athens
Faculty of Biology
Biophysics and Bioinformatics Laboratory