

## WEBLEM 5

### Introduction to Domain Databases

**1. Prosite:** PROSITE is a method of determining what is the function of uncharacterized proteins translated from genomic or cDNA sequences. It consists of a database of biologically significant sites and patterns formulated in such a way that with appropriate computational tools it can rapidly and reliably identify which known family of protein (if any) the new sequence belongs to.

The PROSITE database uses two kinds of signatures or descriptors to identify conserved regions, i.e. patterns and generalized profiles, which both have their own strengths 30 and weaknesses defining their area of optimum application.

PROSITE is complemented by ProRule , a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids. The ProRule section of PROSITE is constituted of manually created rules that can automatically generate annotation in the UniProtKB/Swiss-Prot format based on PROSITE motifs.

**2. Pfam:** Pfam is a database of protein domain families. Pfam contains curated multiple sequence alignments for each family, as well as profile hidden Markov models (profile HMMs) for finding these domains in new sequences. Pfam contains functional annotation, literature references and database links for each family. There are two multiple alignments for each Pfam family, the seed alignment that contains a relatively small number of representative members of the family and the full alignment that contains all members in the database that can be detected. All alignments use sequences taken from pfamseq, which is a non-redundant protein set composed of SWISS-PROT and SP-TrEMBL. The profile HMM is built from the seed alignment using the HMMER package, which is then used to search the pfamseq sequence database.

**3. InterPro:** InterPro is an integrative database which was founded 10 years ago when the PROSITE, PRINTS, Pfam and ProDom databases formed a consortium to amalgamate the predictive signatures they individually produced into a single resource. Since then, six other member databases have also joined and their data has been integrated: SMART, TIGRFAMs, PIRSF, SUPERFAMILY PANTHER and Gene3D.

**4. PRINTS:** The last two decades have seen remarkable advances in molecular biology: 20 years ago sequencing a single gene was considered a monumental technical achievement; today, the sequencing of whole genomes has become almost routine. Advances in the fundamental techniques of sequencing, in concert with advances in laboratory automation and robotics, have led to the rapid and unprecedented accumulation of macromolecular sequence data. The challenge resides not just in the management of this huge quantity of information, but also in its analysis. One of the main goals of bioinformatics is to uncover the knowledge implicit within the data.

The decisive step in this knowledge-discovery process is often the identification of the family to which a newly-identified gene belongs; from this devolves a wealth of insights into function. With its links to 3D structure and post-translational modifications, and thus biological function, it is generally thought that the

amino acid sequence, rather than the nucleic acid sequence, is the most appropriate level at which to seek such relationships.

## Reference:

1. <https://prosite.expasy.org/prosuser.html>
2. Bateman, A. (2000). The Pfam Protein Families Database. *Nucleic Acids Research*, 28(1), 263–266. <https://doi.org/10.1093/nar/28.1.263>
3. Hulo, N. (2006a). The PROSITE database. *Nucleic Acids Research*, 34(90001), D227–D230. <https://doi.org/10.1093/nar/gkj063>
4. Hulo, N. (2006b). The PROSITE database. *Nucleic Acids Research*, 34(90001), D227–D230. <https://doi.org/10.1093/nar/gkj063>
5. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., . . . Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research*, 37(Database), D211–D215. <https://doi.org/10.1093/nar/gkn785>

**WEBLEM 5a**  
**Prosite**  
**(URL: <https://prosite.expasy.org>)**

### Aim:

To study Prothrombin domain information using Prosite database.

### Introduction:

PROSITE is a method of determining what is the function of uncharacterized proteins translated from genomic or cDNA sequences. It consists of a database of biologically significant sites and patterns formulated in such a way that with appropriate computational tools it can rapidly and reliably identify which known family of protein (if any) the new sequence belongs to.

The PROSITE database uses two kinds of signatures or descriptors to identify conserved regions, i.e. patterns and generalized profiles, which both have their own strengths 30 and weaknesses defining their area of optimum application.

Cadherins (named for "calcium-dependent adhesion") are a type of cell adhesion molecule(CAM) that are important in the formation of adherens junctions to allow cells to adhere to each other .Cadherins are a class of type-1 transmembrane proteins, and they are dependent on calcium (Ca<sup>2+</sup>) ions to function, hence their name. Cell-cell adhesion is mediated by extracellular cadherin domains, whereas the Intracellular cytoplasmic tail associates with numerous adaptors and signaling proteins, collectively referred as the cadherin adhesome.

### Methodology:

1. Open homepage for Prosite database.
2. Search for query “Prothrombin” using quick scan more of Prosite with help of FASTA sequence.
3. Observe the result for hits by profile and hits by pattern.
4. Interpret the results.

### Observation:

**Fig1. Homepage of Prosite Database**

PROSITE

Home | ScanProsite | ProRule | Documents | Downloads | Links | Funding

**proSite** Database of protein domains, families and functional sites

**News:** SARS-CoV-2 relevant PROSITE motifs

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].

PROSITE is complemented by **ProRule**, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Release 2021\_03 of 02-Jun-2021 contains 1891 documentation entries, 1311 patterns, 1321 profiles and 1333 ProRule.

**Search**

e.g. PDOC00022, PS50089, SH3, zinc finger

**Browse**

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

**Quick Scan mode of ScanProsite**

Quickly find matches of your protein sequences to PROSITE signatures (max 10 sequences). [?] Examples

```
>sp_P55290-CAD13_HUMAN Cadherin 13 OS=Homo sapiens OX=9606
>sp_P55290-CAD13_HUMAN CADHERIN 13 OS=Homo sapiens OX=9606
MOPRTPLVLCVLLSQVLLTSAEDLCTPGFQOKVHINOPAEIFIEDOSILNLTSCKGNDKLRV
NDKLRYEVSSPPFKVNSDGGVALRNTITAVGKTLFVHARTPHAEQDMEILVTVGKD10GS
LQDITFKARTSPVPRKR5IVSP1LIPENORQPFPROVQKVVDSDPRERSKFRLTGKV
DQEPKGIPIR1NTGSVSVTRLTDREVIAYVOLFETTDVNGTKLEGPVPLEVIVDQND
NRP1FREGPYIGHVMEGSPGTGTTVMRKTAFDADPATDNALLRNIRN100TPDKPSPNMFY
IDPEKGD1IVVVSPALLDRETLENPKYELIEAQDMAGLDVGLTGATATIMIDDDKNDKHS
[Scan] [Clear]
 Exclude motifs with a high probability of occurrence from the scan
```

For more scanning options go to [ScanProsite](#)

**Other tools**

- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.

Fig2. Search page for query “Cadherin” using quick scan mode of ScanProsite in Prosite database.

**proSite** ScanProsite Results Viewer

Output format: Graphical view - this view shows ScanProsite results together with ProRule-based predicted intra-domain features [help].

**Hits for all PROSITE (release 2021\_03) motifs on sequence sp-P55290-CAD13\_HUMAN :**

found: 8 hits in 1 sequence

sp-P55290-CAD13\_HUMAN (713 aa)

MQPRTPLVLCVLLSQVLLTSAEDLCTPGFQOKVHINOPAEIFIEDOSILNLTSCKGNDKLRV  
EVSSPYFKVNSDGGVALRNTITAVGKTLFVHARTPHAEQDMEILVTVGKD10GS  
VPRQKRSIVVSP1LIPENORQPFPROVQKVVDSDPRERSKFRLTGKV  
VTRLTDREVIAYVQLFETTDVNGTKLEGPVPLEVIVDQND  
MRNTAFDADPATDNALLRNIRN100TPDKPSPNMFY  
IDPEKGD1IVVVSPALLDRETLENPKYELIEAQDMAGLDVGLTGATATIMIDDDKNDKHS  
RAAHTIINGNPQ5FEIHTNPOTNEGMLSVVKPLDYEISAFHTLLKVENEDPLVPDFSYGPSSSTA  
TVHITLDVNEGPVFYDPMMVTRQEDLSVGSVLLTVNATPDSLQHOTIRYSVYKDPAGWLNP  
INGTVDTTAVLDRRESPFDNSVYTALFLAIDSGNPATGTTLLITLEDVNDNAPF1YPTVAEVCD  
DAKNSLVIILGASDKDLHPNTDPKFEHKQAVPDKVWKISKINNTHALVSLQNLNKANYNLPI  
VTDSGKPPMTNTDLRVQVCSRNKVDCAAGALRFLSPLSVLLSLFSLACL

**Legend:**

- disulfide bridge
- active site
- other 'ranges'
- other sites

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function. For more information about how these graphical representations are constructed, go to <https://prosite.expasy.org/mydomains/>.

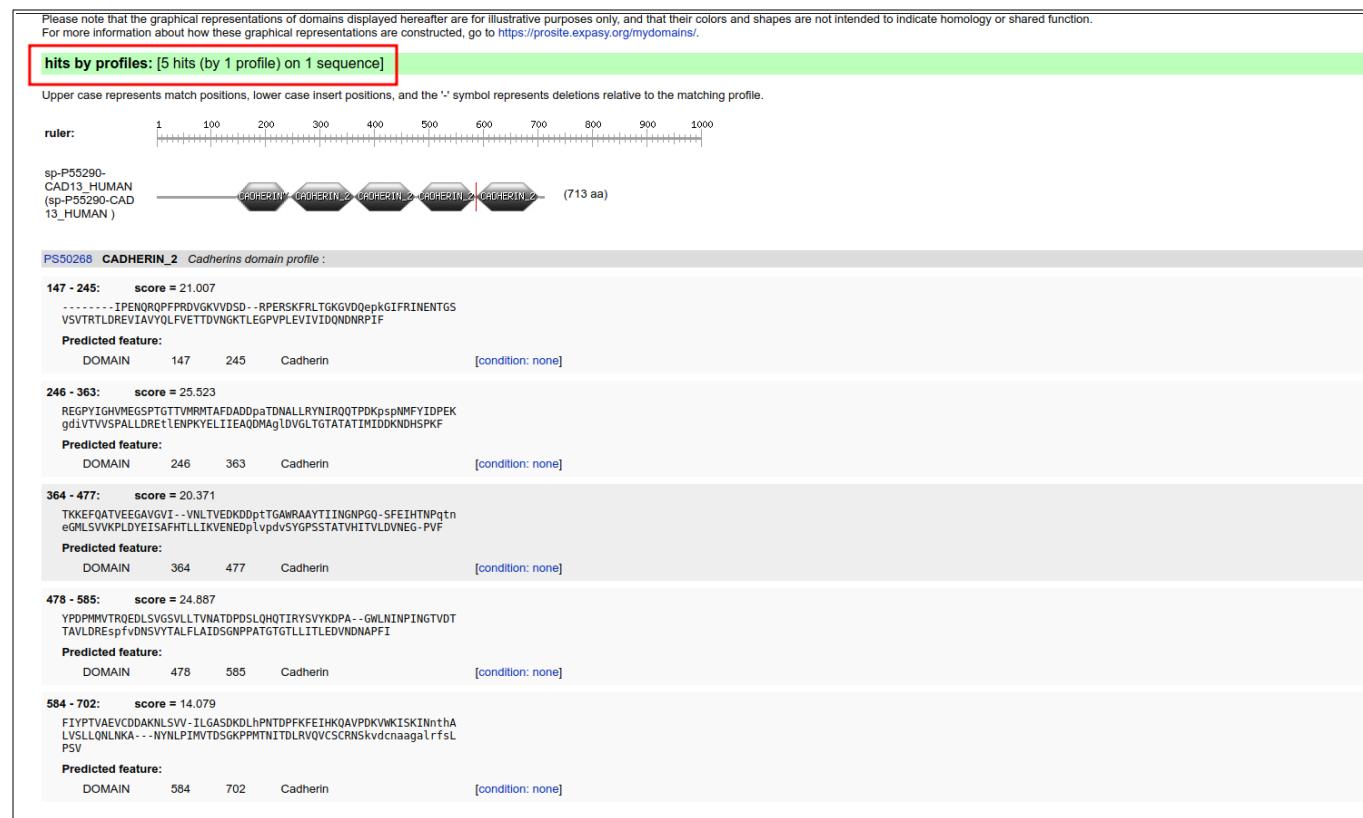
**hits by profiles:** [5 hits (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

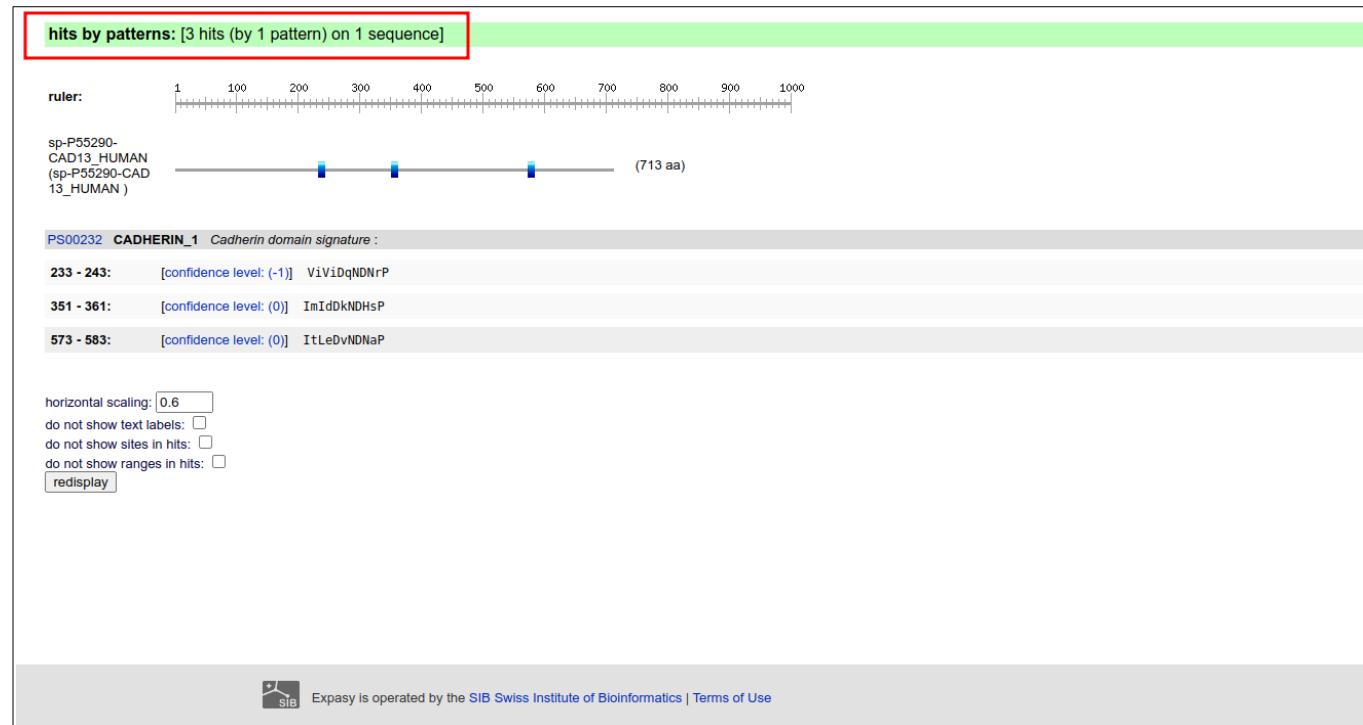
**ruler:**

sp-P55290-CAD13\_HUMAN  
(sp-P55290-CAD13\_HUMAN )

Fig3. Result page for query “Cadherin” in Prosite database.



**Fig4. Result page for query “Cadherin” in Prosite database showing hits by profiles.**



**Fig5. Result page for query “Cadherin” in Prosite showing hits by profiles.**

ExPasy  PROSITE

Home | Contact

 PROSITE documentation PDOC00205 [for PROSITE entry PS50268]

Cadherin domain signature and profile

Description Technical section References Copyright Miscellaneous

**Description**

Cadherins [1,2] are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known, for example:

- Epithelial (E-cadherin) (CDH1).
- Neural (N-cadherin) (CDH2).
- Placental (P-cadherin) (CDH3).
- Retinal (R-cadherin) (CDH4).
- Vascular endothelial (VE-cadherin) (CDH5).
- Kidney (K-cadherin) (CDH6).
- Cadherin-8 (CDH8).
- Cadherin-9 (CDH9).
- Osteoblast (OB-cadherin) (CDH11).
- Brain (BR-cadherin) (CDH12).
- T-cadherin (truncated cadherin) (CDH13).
- Muscle (M-cadherin) (CDH15).
- Kidney (Ksp-cadherin) (CDH16).
- Liver-intestine (LI-cadherin) (CDH17).

Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub-divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. It is suggested that the calcium-binding region of cadherins is located in the extracellular repeats.

Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins:

- Desmoglein 1 (desmosomal glycoprotein I).
- Desmoglein 2.
- Desmoglein 3 (Pemphigus vulgaris antigen).

Other proteins that include cadherin domains are:

## Fig6. Result page for “Cadherin Domain signature” and profile

Description Technical section References Copyright Miscellaneous

**Description**

Cadherins [1,2] are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known, for example:

- Epithelial (E-cadherin) (CDH1).
- Neural (N-cadherin) (CDH2).
- Placental (P-cadherin) (CDH3).
- Retinal (R-cadherin) (CDH4).
- Vascular endothelial (VE-cadherin) (CDH5).
- Kidney (K-cadherin) (CDH6).
- Cadherin-8 (CDH8).
- Cadherin-9 (CDH9).
- Osteoblast (OB-cadherin) (CDH11).
- Brain (BR-cadherin) (CDH12).
- T-cadherin (truncated cadherin) (CDH13).
- Muscle (M-cadherin) (CDH15).
- Kidney (Ksp-cadherin) (CDH16).
- Liver-intestine (LI-cadherin) (CDH17).

Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub-divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. It is suggested that the calcium-binding region of cadherins is located in the extracellular repeats.

Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins:

- Desmoglein 1 (desmosomal glycoprotein I).
- Desmoglein 2.
- Desmoglein 3 (Pemphigus vulgaris antigen).

Other proteins that include cadherin domains are:

- Drosophila fat protein [3], a huge protein of over 5000 amino acids that contains 34 cadherin-like repeats in its extracellular domain.
- Homologs of fat are found in mammals.
- Protocadherins (6 copies).
- Proto-oncogene tyrosine-protein kinase receptor ret (1 copy).

The signature pattern we have developed for the repeated domain is located in it the C-terminal extremity which is its best conserved region. The pattern includes two conserved aspartic acid residues as well as two asparagines; these residues could be implicated in the binding of calcium. We have also developed a profile that spans the complete domain.

Note:

This pattern is found in the first, second, and fourth copies of the repeated domain. In the third copy there is a deletion of one residue after the second conserved Asp.

Last update:

## Fig7. Description page for “Cadherin Domain signature”

Description    Technical section    References    Copyright    Miscellaneous

**Technical section**

PROSITE methods (with tools and information) covered by this documentation:

**CADHERIN\_2, PS50268; Cadherins domain profile (MATRIX)**

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 329
  - detected by PS50268: 329 (true positives)
  - undetected by PS50268: 0 (false negative or 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50268: 2 false positives.
- Domain architecture view of Swiss-Prot proteins matching PS50268

**CADHERIN\_2**

- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50268
- View ligand binding statistics of PS50268
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**CADHERIN\_1, PS00232; Cadherin domain signature (PATTERN)**

- Consensus pattern:  
[LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P  
The 2 D's and the N are involved in calcium binding
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 325
  - detected by PS00232: 310 (true positives)
  - undetected by PS00232: 15 (15 false negatives and 0 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00232: NONE.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00232
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00232
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00232
- View ligand binding statistics of PS00232
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**Fig8. Result page showing Technical section for “Cadherin\_2:PS50268”**

Description    Technical section    References    Copyright    Miscellaneous

**CADHERIN\_2**

- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50268
- View ligand binding statistics of PS50268
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**CADHERIN\_1, PS00232; Cadherin domain signature (PATTERN)**

- Consensus pattern:  
[LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P  
The 2 D's and the N are involved in calcium binding
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 325
  - detected by PS00232: 310 (true positives)
  - undetected by PS00232: 15 (15 false negatives and 0 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00232: NONE.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00232
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00232
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00232
- View ligand binding statistics of PS00232
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**References**

1	Authors	Takeichi M.
	Title	Cadherins: a molecular family important in selective cell-cell adhesion.
	Source	Annu. Rev. Biochem. 59:237-252(1990).
	PubMed ID	<a href="#">2197976</a>
	DOI	<a href="#">10.1146/annurev.bi.59.07190.001321</a>
2	Authors	Takeichi M.
	Source	Trends Genet. 3:213-217(1987).

**Fig9. Result page for “Cadherin” which shows consensus pattern for Cadherin Domain signature**

Description
Technical section
References
Copyright
Miscellaneous

• Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ...
[\[ALL\]](#)

**References**

1	Authors	Takeichi M.	
	Title	Cadherins: a molecular family important in selective cell-cell adhesion.	
	Source	Annu. Rev. Biochem. 59:237-252(1990).	
	PubMed ID	<a href="#">2197976</a>	
2	Authors	Takeichi M.	
	Source	Trends Genet. 3:213-217(1987).	
	3	Authors	Mahoney P.A. Weber U. Onofrechuk P. Biessmann H. Bryant P.J. Goodman C.S.
		Title	The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily.
Source		Cell 67:853-868(1991).	
PubMed ID		<a href="#">1959133</a>	

**Copyright**

PROSITE is copyrighted by the SIB Swiss Institute of Bioinformatics and distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND 4.0) License, see [prosite\\_license.html](#).

**Miscellaneous**

[View entry in original PROSITE document format](#)  
[View entry in raw text format \(no links\)](#)

**Fig10. Result page for “Cadherin” showing references**

Description
Technical section
References
Copyright
Miscellaneous

• Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ...
[\[ALL\]](#)

**References**

1	Authors	Takeichi M.	
	Title	Cadherins: a molecular family important in selective cell-cell adhesion.	
	Source	Annu. Rev. Biochem. 59:237-252(1990).	
	PubMed ID	<a href="#">2197976</a>	
2	Authors	Takeichi M.	
	Source	Trends Genet. 3:213-217(1987).	
	3	Authors	Mahoney P.A. Weber U. Onofrechuk P. Biessmann H. Bryant P.J. Goodman C.S.
		Title	The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily.
Source		Cell 67:853-868(1991).	
PubMed ID		<a href="#">1959133</a>	

**Copyright**

PROSITE is copyrighted by the SIB Swiss Institute of Bioinformatics and distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND 4.0) License, see [prosite\\_license.html](#).

**Miscellaneous**

[View entry in original PROSITE document format](#)  
[View entry in raw text format \(no links\)](#)

**Fig11. Result page for Cadherin domain showing Copyright and miscellaneous**

ExPasy  PROSITE

Home ScanProsite ProRule Documents Downloads Links Funding

**prosite** PROSITE documentation PDOC00205 [for PROSITE entry PS00232]

Cadherin domain signature and profile

Description Technical section References Copyright Miscellaneous

**Description**

Cadherins [1,2] are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known, for example:

- Epithelial (E-cadherin) (CDH1).
- Neural (N-cadherin) (CDH2).
- Placental (P-cadherin) (CDH3).
- Retinal (R-cadherin) (CDH4).
- Vascular endothelial (VE-cadherin) (CDH5).
- Kidney (K-cadherin) (CDH6).
- Cadherin-8 (CDH8).
- Cadherin-9 (CDH9).
- Osteoblast (OB-cadherin) (CDH11).
- Brain (BR-cadherin) (CDH12).
- T-cadherin (truncated cadherin) (CDH13).
- Muscle (M-cadherin) (CDH15).
- Kidney (Ksp-cadherin) (CDH16).
- Liver-intestine (LI-cadherin) (CDH17).

Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub-divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. It is suggested that the calcium-binding region of cadherins is located in the extracellular repeats.

Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins:

- Desmoglein 1 (desmosomal glycoprotein I).
- Desmoglein 2.
- Desmoglein 3 (Pemphigus vulgaris antigen).

Other proteins that include cadherin domains are:

## Fig12.Result page for Cadherin domain signature and profile

Description Technical section References Copyright Miscellaneous

**Description**

Cadherins [1,2] are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known, for example:

- Epithelial (E-cadherin) (CDH1).
- Neural (N-cadherin) (CDH2).
- Placental (P-cadherin) (CDH3).
- Retinal (R-cadherin) (CDH4).
- Vascular endothelial (VE-cadherin) (CDH5).
- Kidney (K-cadherin) (CDH6).
- Cadherin-8 (CDH8).
- Cadherin-9 (CDH9).
- Osteoblast (OB-cadherin) (CDH11).
- Brain (BR-cadherin) (CDH12).
- T-cadherin (truncated cadherin) (CDH13).
- Muscle (M-cadherin) (CDH15).
- Kidney (Ksp-cadherin) (CDH16).
- Liver-intestine (LI-cadherin) (CDH17).

Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub-divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. It is suggested that the calcium-binding region of cadherins is located in the extracellular repeats.

Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins:

- Desmoglein 1 (desmosomal glycoprotein I).
- Desmoglein 2.
- Desmoglein 3 (Pemphigus vulgaris antigen).

Other proteins that include cadherin domains are:

- Drosophila fat protein [3], a huge protein of over 5000 amino acids that contains 34 cadherin-like repeats in its extracellular domain. Homologs of fat are found in mammals.
- Protocadherins (6 copies).
- Proto-oncogene tyrosine-protein kinase receptor ret (1 copy).

The signature pattern we have developed for the repeated domain is located in the C-terminal extremity which is its best conserved region. The pattern includes two conserved aspartic acid residues as well as two asparagines; these residues could be implicated in the binding of calcium. We have also developed a profile that spans the complete domain.

Note:

This pattern is found in the first, second, and fourth copies of the repeated domain. In the third copy there is a deletion of one residue after the second conserved Asp.

Last update:

## Fig13. Result page for Cadherin domain signature and profile showing Description

Description    Technical section    References    Copyright    Miscellaneous

**Technical section**

PROSITE methods (with tools and information) covered by this documentation:

**CADHERIN\_1, PS00232; Cadherin domain signature (PATTERN)**

- Consensus pattern:  
[LV]-x-[LV]-x-D-x-N-D-[NH]-x-P  
The 2 D's and the N are involved in calcium binding
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 325
  - detected by PS00232: 310 (true positives)
  - undetected by PS00232: 15 (15 false negatives and 0 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00232: NONE.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00232
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00232
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00232
- View ligand binding statistics of PS00232
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**CADHERIN\_2, PS50268; Cadherins domain profile (MATRIX)**

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 329
  - detected by PS50268: 329 (true positives)
  - undetected by PS50268: 0 (false negative or 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50268: 2 false positives.
- Domain architecture view of Swiss-Prot proteins matching PS50268
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50268
- View ligand binding statistics of PS50268
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**Fig14. Result page for Cadherin domain signature and profile showing Technical Section and consensus atther**

Description    Technical section    References    Copyright    Miscellaneous

**CADHERIN\_2, PS50268; Cadherins domain profile (MATRIX)**

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 329
  - detected by PS50268: 329 (true positives)
  - undetected by PS50268: 0 (false negative or 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50268: 2 false positives.
- Domain architecture view of Swiss-Prot proteins matching PS50268
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:  
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50268
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50268
- View ligand binding statistics of PS50268
- Matching PDB structures: 1EDH 1FF5 1L3W 1NCG ... [ALL]

**References**

1	Authors	Takeichi M.
	Title	Cadherins: a molecular family important in selective cell-cell adhesion.
	Source	Annu. Rev. Biochem. 59:237-252(1990).
	PubMed ID	<a href="#">2197976</a>
	DOI	<a href="https://doi.org/10.1146/annurev.bi.59.070190.001321">10.1146/annurev.bi.59.070190.001321</a>
2	Authors	Takeichi M.
	Source	Trends Genet. 3:213-217(1987).
	Authors	Mahoney P.A. Weber U. Onofrechuk P. Biessmann H. Bryant P.J. Goodman C.S.
	Title	The fat tumor suppressor gene in <i>Drosophila</i> encodes a novel member of the cadherin gene superfamily.

**Fig15. Result page for Cadherin domain signature showing CADHERIN\_2**

Description
Technical section
References
Copyright
Miscellaneous

- Matching PDB structures: [1E0H](#) [1F55](#) [1L3W](#) [1NCG](#) ... [\[ALL\]](#)

References

1	Authors	Takeichi M.
	Title	Cadherins: a molecular family important in selective cell-cell adhesion.
	Source	Annu. Rev. Biochem. 59:237-252(1990).
	PubMed ID	<a href="#">2197976</a>
	DOI	<a href="#">10.1146/annurev.bi.59.070190.001321</a>
2	Authors	Takeichi M.
	Source	Trends Genet. 3:213-217(1987).
3	Authors	Mahoney P.A. Weber U. Onofrechuk P. Biessmann H. Bryant P.J. Goodman C.S.
	Title	The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily.
	Source	Cell 67:853-868(1991).
	PubMed ID	<a href="#">1959133</a>

Copyright

PROSITE is copyrighted by the SIB Swiss Institute of Bioinformatics and distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND 4.0) License, see [prosite\\_license.html](#).

Miscellaneous

[View entry in original PROSITE document format](#)  
[View entry in raw text format \(no links\)](#)

**Fig16. Result page for Cadherin domain signature showing References**

Description
Technical section
References
Copyright
Miscellaneous

- Matching PDB structures: [1E0H](#) [1F55](#) [1L3W](#) [1NCG](#) ... [\[ALL\]](#)

References

1	Authors	Takeichi M.
	Title	Cadherins: a molecular family important in selective cell-cell adhesion.
	Source	Annu. Rev. Biochem. 59:237-252(1990).
	PubMed ID	<a href="#">2197976</a>
	DOI	<a href="#">10.1146/annurev.bi.59.070190.001321</a>
2	Authors	Takeichi M.
	Source	Trends Genet. 3:213-217(1987).
3	Authors	Mahoney P.A. Weber U. Onofrechuk P. Biessmann H. Bryant P.J. Goodman C.S.
	Title	The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily.
	Source	Cell 67:853-868(1991).
	PubMed ID	<a href="#">1959133</a>

Copyright

PROSITE is copyrighted by the SIB Swiss Institute of Bioinformatics and distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND 4.0) License, see [prosite\\_license.html](#).

Miscellaneous

[View entry in original PROSITE document format](#)  
[View entry in raw text format \(no links\)](#)

**Fig17. Result page for Cadherin Domain signature showing Copyright and miscellaneous**

## **Result:**

The profile and pattern information for CADHERIN\_1 and CADHERIN\_2 was retrieved

## **Conclusion:**

PROSITE database provides resources and tools for easy retrieval of profiles and patterns of various protein domains.

## **References:**

1. Maître, J. L., & Heisenberg, C. P. (2013). Three Functions of Cadherins in Cell Adhesion. *Current Biology*, 23(14), R626–R633. <https://doi.org/10.1016/j.cub.2013.06.019>
2. (n.d.). Retrieved from <https://www.uniprot.org/uniprot/P55290.fasta>
3. M., T., P.A., M., U., W., P., O., H., B., P.J., B., & C.S., G. (2004, May 01). Cadherin domain signature and profile. Retrieved from <https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS50268>
4. M., T., P.A., M., U., W., P., O., H., B., P.J., B., & C.S., G. (2004, May 01). Cadherin domain signature and profile. Retrieved from <https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS00232>

## WEBLEM 5b

### Pfam

**(URL: <http://pfam.xfam.org/>)**

#### **Aim:**

To study Cadherin domain information using Pfam database.

#### **Introduction:**

Pfam is a database of protein domain families. Pfam contains curated multiple sequence alignments for each family, as well as profile hidden Markov models (profile HMMs) for finding these domains in new sequences. Pfam contains functional annotation, literature references and database links for each family. There are two multiple alignments for each Pfam family, the seed alignment that contains a relatively small number of representative members of the family and the full alignment that contains all members in the database that can be detected. All alignments use sequences taken from pfamseq, which is a non-redundant protein set composed of SWISS-PROT and SP-TrEMBL. The profile HMM is built from the seed alignment using the HMMER package, which is then used to search the pfamseq sequence database.

Cadherins (named for "calcium-dependent adhesion") are a type of cell adhesion molecule(CAM) that are important in the formation of adherens junctions to allow cells to adhere to each other . Cadherins are a class of type-1 transmembrane proteins, and they are dependent on calcium (Ca<sup>2+</sup>) ions to function, hence their name. Cell-cell adhesion is mediated by extracellular cadherin domains, whereas the Intracellular cytoplasmic tail associates with numerous adaptors and signaling proteins, collectively referred as the cadherin adhesome.

#### **Methodology:**

1. Open homepage for Pfam database. (URL: <http://pfam.xfam.org/>)
2. Search for query “Cadherin” using keyword search.
3. Select a domain and observe the results.
4. Interpret the results.

#### **Observation:**

**Pfam 34.0 (March 2021, 19179 entries)**

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [More...](#)

**QUICK LINKS**

- [SEQUENCE SEARCH](#)
- [VIEW A PFAM ENTRY](#)
- [VIEW A CLAN](#)
- [VIEW A SEQUENCE](#)
- [VIEW A STRUCTURE](#)
- [KEYWORD SEARCH](#)
- [JUMP TO](#)

**YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...**

Analyze your protein sequence for Pfam matches  
View Pfam annotation and alignments  
See groups of related entries  
Look at the domain organisation of a protein sequence  
Find the domains on a PDB structure  
Query Pfam by keywords

Enter any accession or ID  [Go](#) [Example](#)

Or view the [help](#) pages for more information

**Citing Pfam**

If you find Pfam useful, please consider [citing](#) the reference that describes this work:

*Pfam: The protein families database in 2021*; J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladini, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman  
*Nucleic Acids Research* (2020) doi: 10.1093/nar/gkaa913

**Pfam is part of the ELIXIR infrastructure**  
Pfam is an ELIXIR service. [Read more](#)

Comments or questions on the site? Send a mail to [pfam-help@ebi.ac.uk](mailto:pfam-help@ebi.ac.uk).  
European Molecular Biology Laboratory

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#). I agree, dismiss this banner

**Fig1. Homepage of Pfam database**

Pfam 34.0 (March 2021, 19179 entries)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [More...](#)

## QUICK LINKS

SEQUENCE SEARCH

VIEW A PFAM ENTRY

VIEW A CLAN

VIEW A SEQUENCE

VIEW A STRUCTURE

KEYWORD SEARCH

JUMP TO

## QUERY PFAM BY KEYWORD

Search for keywords in text data in the Pfam database.

You can also use the keyword search box at the top of every page.

## Citing Pfam

If you find Pfam useful, please consider [citing](#) the reference that describes this work:

*Pfam: The protein families database in 2021*: J. Misra, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman  
Nucleic Acids Research (2020) doi: 10.1093/nar/gkaa913



This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).  
Waiting for pfam.xfam.org...

**Fig2. Search page for query “Cadherin” Using Keyword search in Pfam database.**

## Keyword search results

We found 363 unique results for your query "Cadherin" in 5 sections of the database. The number of hits in each section is shown below.

Section	Description	Number of hits
Pfam	Text fields for Pfam entries	22
Seq_info	Sequence description and species fields	348
Pdb	HEADER and TITLE records from PDB entries	18
GO	Gene ontology IDs and terms	5
Interpro	InterPro entry abstracts	27

## Matching Pfam families

Your query also appears to match the Pfam entry [Cadherin \(PF00028\)](#). This family is shown in the [highlighted](#) row in the results table below.

This table can be sorted by clicking on the column titles, or restored to the original order [here](#). Only unique Pfam accessions are displayed.

Accession	ID	Description	Pfam	Seq_info	Pdb	GO	Interpro
PF00028	<b>Cadherin</b>	Cadherin domain	✓	✓	✓	✓	✓
PF01049	<b>Cadherin_C</b>	Cadherin cytoplasmic region	✓	✓	✓	✓	✓
PF08266	<b>Cadherin_2</b>	Cadherin-like	✓	✓	✓		✓
PF08258	<b>Cadherin_pro</b>	Cadherin prodomain like	✓	✓	✓		✓
PF17756	<b>RET_CLD1</b>	RET Cadherin like domain 1	✓	✓	✓		✓
PF17812	<b>RET_CLD3</b>	RET Cadherin like domain 3	✓	✓	✓		✓
PF18432	<b>ECD</b>	Extracellular Cadherin domain	✓	✓	✓		✓
PF08374	<b>Protocadherin</b>	Protocadherin	✓	✓			✓
PF12733	<b>Cadherin-like</b>	Cadherin-like beta sandwich domain	✓	✓			✓
PF15351	<b>ICAD</b>	Junctional protein associated with coronary artery disease	✓	✓			✓
PF15974	<b>Cadherin_toll</b>	Cadherin C-terminal cytoplasmic tail, catenin-binding region	✓	✓			✓
PF16492	<b>Cadherin_C_2</b>	Cadherin cytoplasmic C-terminal	✓	✓			✓
PF17803	<b>Cadherin_4</b>	Bacterial cadherin-like domain	✓	✓			✓
PF17813	<b>RET_CLD4</b>	RET Cadherin like domain 4	✓	✓			✓
PF17892	<b>Cadherin_5</b>	Cadherin-like domain	✓	✓			✓
PF16184	<b>Cadherin_3</b>	Cadherin-like	✓	✓			✓
PF12354	<b>Intercardin_N</b>	Bacterial adhesion/invasion protein N terminal	✓		✓		✓
PF06384	<b>ICAT</b>	Beta-catenin-interacting protein ICAT	✓				✓
PF12632	<b>Vezatin</b>	Myosin-binding motif of peroxisomes	✓				✓

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).  
I agree, dismiss this banner

**Fig3. Hit page for query “Cadherin” in Pfam database.**

## Family: Cadherin (PF00028)

Summary  
Domain organisation  
Clan  
Alignments  
HMM logo  
Trees  
Curation & model  
Species  
Structures  
AlphaFold Structures  
trRosetta Structure  
Jump to... ↴  
Enter ID/acc Go

## Summary: Cadherin domain

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

[Wikipedia: Cadherin](#) [Pfam](#) [InterPro](#)

This is the Wikipedia entry entitled "["Cadherin"](#)". [More...](#)

[Cadherin](#) [Edit Wikipedia article](#)

This page is based on a [Wikipedia article](#). The text is available under the [Creative Commons Attribution/Share-Alike License](#).

Pfam is part of the ELIXIR infrastructure  
 Pfam is an ELIXIR service. [Read more](#)  
 Comments or questions on the site? Send a mail to [pfa...@ebi.ac.uk](mailto:pfa...@ebi.ac.uk).  
 European Molecular Biology Laboratory

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).  
 I agree, dismiss this banner

## Fig4. Result page for Cadherin Domain showing summary

## Family: Cadherin (PF00028)

Summary  
Domain organisation  
Clan  
Alignments  
HMM logo  
Trees  
Curation & model  
Species  
Structures  
AlphaFold Structures  
trRosetta Structure  
Jump to... ↴  
Enter ID/acc Go

## Domain organisation

Below is a listing of the unique domain organisations or architectures in which this domain is found. [More...](#)

**There are 3237 sequences with the following architecture: Cadherin\_2, Cadherin x 5, Cadherin\_C\_2**

[OB0500\\_DANRE](#) [Danio rerio (Zebrafish)] Protocadherin 10b (ECO:0000313) Ensembl:ENSDARP00000134223; (1026 residues)

[Show all sequences with this architecture.](#)

**There are 2900 sequences with the following architecture: Cadherin x 5, Cadherin\_C**

[WS2PTD0\\_SHEEP](#) [Ovis aries (Sheep)] Uncharacterized protein (ECO:0000313) Ensembl:ENSOARP00000013711; (794 residues)

[Show all sequences with this architecture.](#)

**There are 2312 sequences with the following architecture: Cadherin**

[WS3P59\\_ANODA](#) [Anopheles darlingi (Mosquito)] Cadherin-domain-containing protein (ECO:0000259) PROSITE:PS50268; (536 residues)

[Show all sequences with this architecture.](#)

**There are 1813 sequences with the following architecture: Cadherin\_2, Cadherin x 5**

[M3YAE8\\_MUSPF](#) [Mustela putorius furo (European domestic ferret)] Protocadherin alpha 3 (ECO:0000313) Ensembl:ENSMUP00000008305; (802 residues)

[Show all sequences with this architecture.](#)

**There are 1668 sequences with the following architecture: Cadherin x 2**

[T1KFE9\\_HELRO](#) [Heloedella robusta (Californian leech)] Uncharacterized protein (ECO:0000313) Ensembl:ESN96612.1, ECO:0000313|EnsemblMetazoa:HelloP150274; (343 residues)

[Show all sequences with this architecture.](#)

**There are 1504 sequences with the following architecture: Cadherin x 3**

[T1PFB6\\_HELRO](#) [Heloedella robusta (Californian leech)] Uncharacterized protein (ECO:0000313) Ensembl:ESO02482.1, ECO:0000313|EnsemblMetazoa:HelloP65492; (557 residues)

[Show all sequences with this architecture.](#)

**There are 1334 sequences with the following architecture: Cadherin x 4**

[A0A645EF664\\_PERFL](#) [Perca fluviatilis (Asian perch)] Uncharacterized protein (ECO:0000313) Ensembl:KAF1373243.1; (496 residues)

[Show all sequences with this architecture.](#)

**There are 1330 sequences with the following architecture: Cadherin x 4, Cadherin\_C**

[ADA0872WUK5\\_SCLFO](#) [Scleropages formosus (Asian bowfin) (Osteoglossum formosum)] Cadherin-10-like (ECO:0000313) Ensembl:KPP67655.1; (Fragment) (670 residues)

[Show all sequences with this architecture.](#)

**There are 1322 sequences with the following architecture: Cadherin\_2, Cadherin x 6, Protocadherin**

[WS2PTD0\\_SHEEP](#) [Ovis aries (Sheep)] Uncharacterized protein (ECO:0000313) Ensembl:ENSOARP00000013711; (794 residues)

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).  
 I agree, dismiss this banner

## Fig5. Result page for Cadherin Domain showing domain organisation

**Summary**

**Domain organisation**

**Clan**

**Alignments**

**HMM logo**

**Trees**

**Curation & model**

**Species**

**Structures**

**AlphaFold Structures**

**trRosetta Structure**

**Jump to...** ↴

Enter ID/acc  Go

**Pfam Clan**

This family is a member of a clan **E-set** (CL0159), which has the following description:

This clan includes a diverse range of domains that have an Ig-like fold and appear to be distantly related to each other. The clan includes: PKD domains, cadherins and several families of bacterial Ig-like domains as well as viral tail fibre proteins. It also includes several fibronectin type III domain-containing families.

The clan contains the following 238 members:

A2M	A2M_BRD	A2M_recep	A9	Adeno_GPI19K	AlaCBM31	Ala-aminope_N	Alba_adaptin2	Alba_E2_glycop	Anth_Ig
Arch_flagellin	Arf1B	Arylsulfotran_N	ASF1_hist_chap	ATG19	BACON	BACON_2	BalD	Bal_Ig	Bal_Ig
BIG_10	BIG_11	BIG_12	BIG_13	BALB	BALB	BALB_2	BALB_3	BALB_4	BALB_5
BIG_4	BIG_5	BIG_6	BIG_7	BALB_2	BALB_2	BALB_3	BIPBP_C	BIPBP_C	bMG10
bMG3	bMG5	Bs1A	Bs1P	Bs1P	Bs1P	Bs1P	Bile_Hdr_Trans	Bile_Hdr_Trans	Cadherin_3
Cadherin_5	Cadherin_BRD	CaNk	CaNk_hex_C	Cannabin_A_N	CARD8	CARD8	CARD8_X2	CARD8_X2	CARD_N
Ceropside_alk_C	CERB_HEX_C	Ceropside_A_N	Ceropside_N	Cannabin_A_N	CERD6	CERD6	CD45	CD45	CD45
CshA_repeat	Cyc-mallobact_N	Cytomega_U53	DsbC	CD45R	CD45R	CD45R	CD45R	CD45R	CD45R
DUF3322	DUF3416	DUF3458	DUF3501	DUF11	DUF1410	DUF1425	DUF1929	DUF2271	DUF3244
DUF3458	DUF3509	DUF3530	DUF3583	DUF3823_C	DUF1859	DUF4165	DUF4175	DUF4426	DUF4469
DUF5115	DUF525	DUF5443	DUF6383	DUF6383	DUF6383	DUF6398	DUF6398	DUF6501	DUF6511
Expansin_C	Filamin	FlixG_C	Flavi_glycop_C	DUF916	FBR	FCC	Fcgr2	Fcgr2	Fcgr2
F11C	F11D	F11D_2	F11D_200c	F11D_200c	F11D_200c	F11D_200c	Fcgr3	Fcgr3	Fcgr3
GPI-anchored	Hanta_G1	HCV_N	HCV_N	F11D_200c	F11D_200c	F11D_200c	Fcgr3a	Fcgr3a	Fcgr3a
Ig_mannosidase	IL12p40_C	IL13Ra_Ig	IL13Ra_Ig	F11D_200c	F11D_200c	F11D_200c	Fcgr3b	Fcgr3b	Fcgr3b
Inhibitor_171	Inlk_B	Integrin_alpha2	Integrin_alpha2	F11D_200c	F11D_200c	F11D_200c	Fcgr3c	Fcgr3c	Fcgr3c
LRP_N	LRP_N	LRP_N	LRP_N	F11D_200c	F11D_200c	F11D_200c	Fcgr3d	Fcgr3d	Fcgr3d
MG4	Mo-co_dimer	N_BRCA1_IG	N_K-ATPass	NEAT	NEAT	NEAT	Neuroxophilin	Neuroxophilin	Neuroxophilin
Peptides_C25_C	Phibio_G2_C	PhoD_N	PKD	PKD_2	PKD_3	PKD_3	NPCBM	NPCBM	NPCBM
Phibio_180p	Phibio_G2_C	PhoD_N	PKD	PKD_2	PKD_3	PKD_3	PapD	PapD	PapD
RET_CLD4	RET_CLD4	RGL_Iyse	RHD_dimer	RHD_dimer	Q9_Am_d_4II	Q9_Am_d_4II	PapD	PapD	PapD
Sox2	SorB	SusE	SVA	Rho_GDI	Rib38T	Rib38T	Pax5	Pax5	Pax5
TIG	TIG_2	TIG_melin	TIG_SOH	Tissue_fec	Tobacco_C	Tobacco_C	SCAB3g	SCAB3g	SCAB3g
UL16	Wav6	WIF	WIF	Y	Y	Y	TarS_C1	TarS_C1	TarS_C1

**Pfam is part of the ELIXIR infrastructure**  
ELIXIR is an ELIXIR service. Read more

Comments or questions on the site? Send a mail to [pfam-help@ebi.ac.uk](mailto:pfam-help@ebi.ac.uk).  
European Molecular Biology Laboratory

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).  
I agree, dismiss this banner

Fig6. Result page showing Cadherin Domain showing its clan

**Family: Cadherin (PF00028)**

**Alignments**

We store a range of different sequence alignments for families. As well as the seed alignment from which the family is built, we provide the full alignment, generated by searching the sequence database ([reference proteomes](#)) using the family HMM. We also generate alignments using four representative proteomes (RP) sets and the UniProtB sequence database. [More...](#)

**View options**

We make a range of alignments for each Pfam-A family. You can see a description of each [above](#). You can view these alignments in various ways but please note that some types of alignment are never generated while others may not be available for all families, most commonly because the alignments are too large to handle.

	Seed (55)	Full (212635)	Representative proteomes				UniProt (358710)
			RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (248981)	
Jalview	✓	✓	✓	✓	✓	✓	✓
HTML	✓	—	✗	✗	✗	✗	✗
RP/heatmap	X	1	—	✗	✗	✗	✗

<sup>1</sup>Cannot generate RP/heatmap alignments for seeds; no RP data available

Key: ✓ available, ✗ not generated, — not available.

**Format an alignment**

	Seed (55)	Full (212635)	Representative proteomes				UniProt (358710)
			RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (248981)	
Alignment:	<input checked="" type="radio"/> Selex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Format:	Selex						
Order:	<input checked="" type="radio"/> Tree	<input type="radio"/> Alphabetical					
Sequence:	<input checked="" type="radio"/> Inserts lower case	<input type="radio"/> All upper case					
Gaps:	Gaps as “-” or “*” (mixed) <input checked="" type="checkbox"/>						
Download/view:	<input checked="" type="radio"/> Download	<input type="radio"/> View					
Generate	<input type="button" value="Generate"/>						

**Download options**

We make all of our alignments available in Stockholm format. You can download them here as raw, plain text files or as [gzip](#)-compressed files.

	Seed (55)	Full (212635)	Representative proteomes				UniProt (358710)
			RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (248981)	
Seed	55	212635	RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (248981)	UniProt (358710)

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).  
I agree, dismiss this banner

Fig7. Result page for Cadherin Domain Showing Alignments view options.

**Clan**

**Alignments**

**HMM logo**

**Trees**

**Curation & model**

**Species**

**Structures**

**AlphaFold Structures**

trRosetta Structure

We store a range of different sequence alignments for families. As well as the seed alignment from which the family is built, we provide the full alignment, generated by searching the sequence database ([reference proteomes](#)) using the family HMM. We also generate alignments using four [representative proteomes](#) (RP) sets and the UniProtKB sequence database. [More...](#)

**View options**

We make a range of alignments for each Pfam-A family. You can see a description of each [above](#). You can view these alignments in various ways but please note that some types of alignment are never generated while others may not be available for all families, most commonly because the alignments are too large to handle.

	Seed (55)	Full (212635)	Representative proteomes				UniProt (358710)
			RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (249981)	
Jalview	✓	✓	✓	✓	✓	✓	✓
HTML	—	✗	✗	✗	✗	✗	✗
PP/heatmap	✗	—	✗	✗	✗	✗	✗

Cannot generate PP/Heatmap alignments for seeds; no PP data available

**Key:** ✓ available, ✗ not generated, — not available.

**Jump to...** ↻

[enter ID](#) [Go](#)

**Format an alignment**

	Seed (55)	Full (212635)	Representative proteomes				UniProt (358710)	
			RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (249981)		
Alignment:	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
Format:	<input type="radio"/> Selex	<input type="radio"/> <a href="#">SelEX</a>						
Order:	<input checked="" type="radio"/> Tree	<input type="radio"/> Alphabetical						
Sequence:	<input checked="" type="radio"/> Inserts lower case	<input type="radio"/> All upper case						
Gaps:	<input type="radio"/> Gaps as "-" or "-" (mixed)							
Download/view:	<input checked="" type="radio"/> Download	<input type="radio"/> View						
Generate	<a href="#">Download options</a>							

**Download options**

We make all of our alignments available in Stockholm format. You can download them here as raw, plain text files or as [gzip](#)-compressed files.

	Seed (55)	Full (212635)	Representative proteomes				UniProt (358710)
			RP15 (27455)	RP35 (75038)	RP55 (187536)	RP75 (249981)	
Raw Stockholm	✓	✓	✓	✓	✓	✓	—
Gzipped	✓	✓	✓	✓	✓	✓	—

You can also [download](#) a FASTA format file containing the **full-length sequences** for all sequences in the full alignment.

Fig8. Result page for Cadherin Domain showing download options for alignments.

EMBL-EBI 

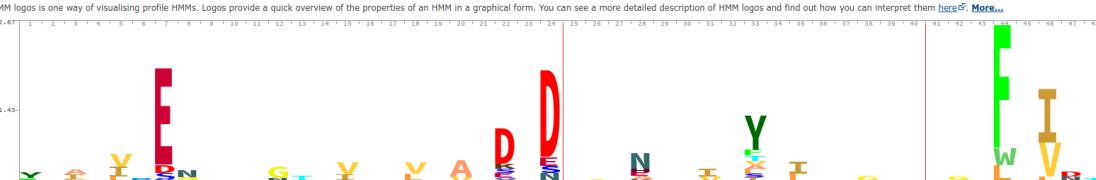
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

**Family: Cadherin (PF00028)**

2182 architectures
212635 sequences
0 interactions
849 species
700 structures

Summary
HMM logo
2.87
1.43
0

HMM logos is one way of visualising profile HMMs. Logos provide a quick overview of the properties of an HMM in a graphical form. You can see a more detailed description of HMM logos and find out how you can interpret them [here](#) & [More](#).



Trees
Curation & model
Species
Structures

AlphaFold Structures
trRosetta Structure

Jump to... 
enter ID  

 **Pfam is part of the ELIXIR infrastructure**

Pfam is an ELIXIR service [Read more](#)

Comments or questions on the site? Send a mail to [pfam-help@ebi.ac.uk](mailto:pfam-help@ebi.ac.uk).

European Molecular Biology Laboratory

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).

[I understand this banner](#)

## Fig9. Result page for Cadherin Domain showing HMM logo

## Family: Cadherin (PF00028)

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation &amp; model

Species

Structures

AlphaFold Structures

trRosetta Structure

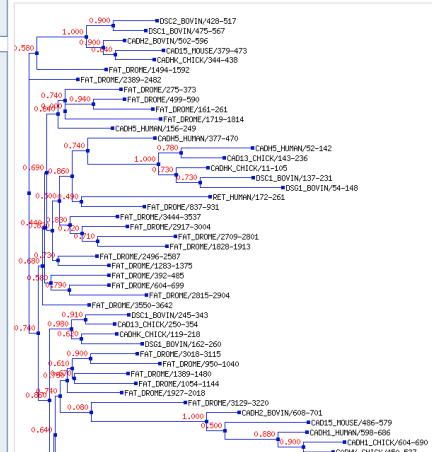
Jump to...

Enter ID/acc

Go

## Trees

This page displays the phylogenetic tree for this family's seed alignment. We use [FastTree](#) to calculate neighbour join trees with a local bootstrap based on 100 resamples (shown next to the tree nodes). FastTree calculates approximately-maximum-likelihood phylogenetic trees from our seed alignment.



This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).

I agree, dismiss this banner

Fig10. Result page for Cadherin Domain showing Phylogenetic tree.

## Family: Cadherin (PF00028)

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation &amp; model

Species

Structures

AlphaFold Structures

trRosetta Structure

Jump to...

Enter ID/acc

Go

## Curation and family details

This section shows the detailed information about the Pfam family. You can see the definitions of many of the terms in this section in the [glossary](#) and a fuller explanation of the scoring system that we use in the [scores](#) section of the help pages.

## Curation

Seed source: Swissprot\_feature\_table  
Previous IDs: cadherin;  
Type: Domain  
Sequence Ontology: SO:0000417

Author: Sonnhammer ELL

Number in seed: 55  
Number in full: 212635  
Average length of the 93.10 aa

Average identity of full 24 % alignments

Average coverage of 44.55 % the sequence by the domain:

## HMM information

HMM build commands: build method: hmmbuild -o /dev/null HMM SEED  
search method: hmmssearch -Z 57096847 -E 1000 --cpu 4 HMM pfamseq

Model details:	Parameter	Sequence Domain
	Gathering cut-off	28.8 28.8
	Trusted cut-off	28.8 28.8
	Noise cut-off	28.7 28.7

Model length: 93

Family (HMM) version: 19

Download: [download](#) the raw HMM for this family

Pfam is part of the ELIXIR infrastructure  
elixir  
Pfam is an Elixir service. [Read more](#)  
Comments or questions on the site? Send a mail to [pfam-help@ebi.ac.uk](mailto:pfam-help@ebi.ac.uk).  
European Molecular Biology Laboratory

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).

I agree, dismiss this banner

Fig11. Result page for Cadherin Domain showing curation and family details.

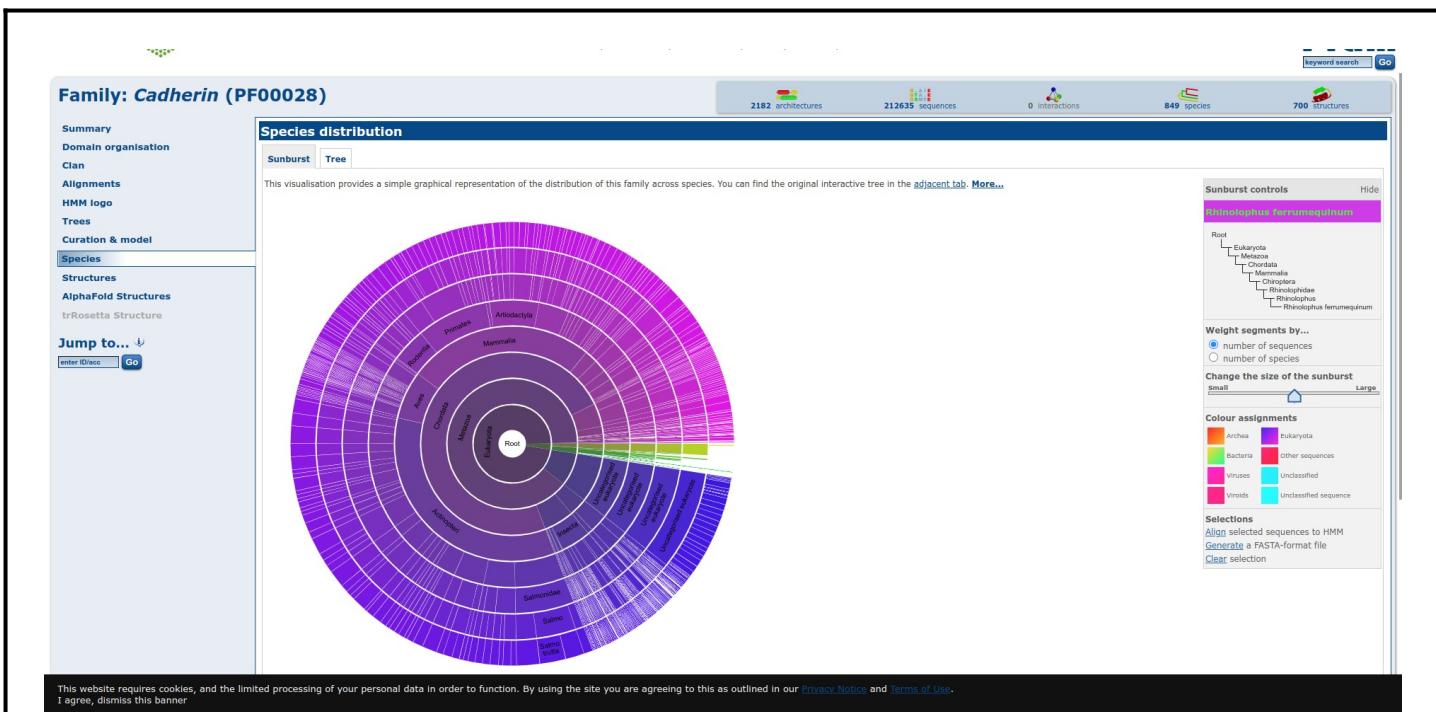


Fig12. Result page for Cadherin Domain showing species distribution in subburst format

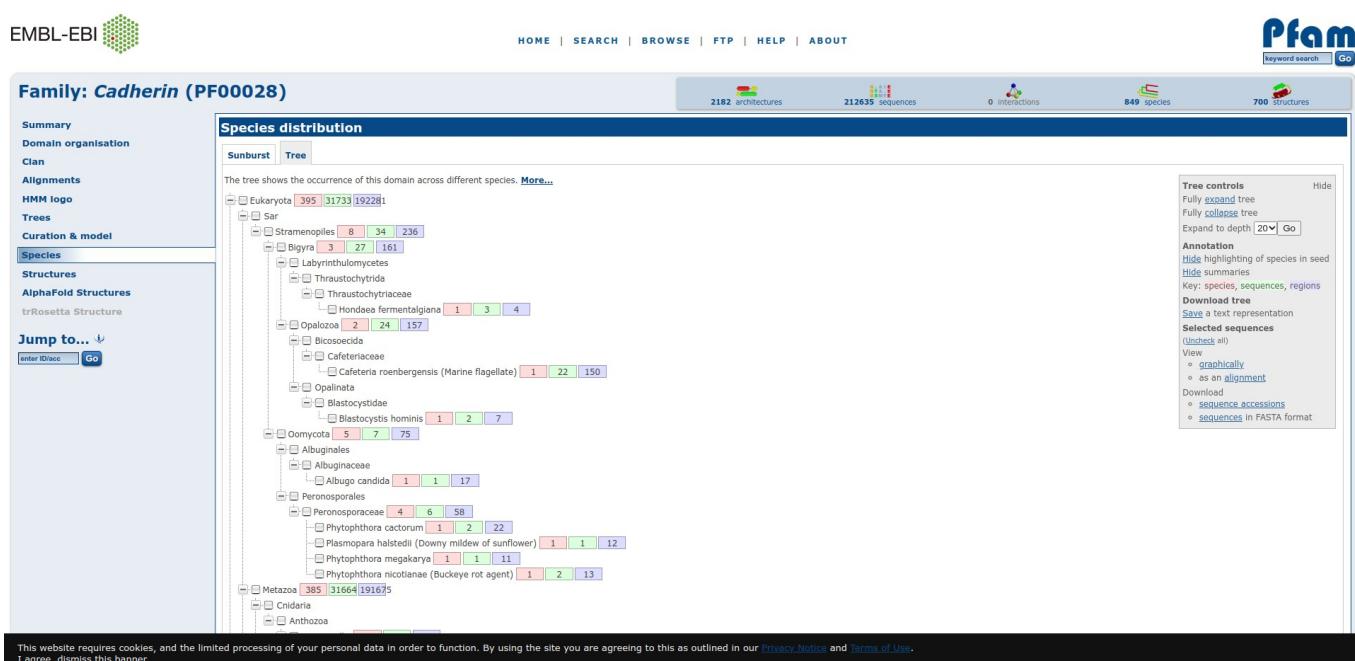


Fig13. Result page for Cadherin Domain showing species distribution in tree format

EMBL-EBI  [Pfam](#) [Keyword search](#) [Go](#)

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

**Family: Cadherin (PF00028)**

 Loading page components (1 remaining)...

**Summary**

**Domain organisation**

**Clan**

**Alignments**

**HMM logo**

**Trees**

**Curation & model**

**Species**

**Structures** Structures

For those sequences which have a structure in the [Protein DataBank](#), we use the mapping between [UniProt](#), PDB and Pfam coordinate systems from the [PDB2Pfam](#) group, to allow us to map Pfam domains onto UniProt sequences and three-dimensional protein structures. The table below shows the structures on which the **Cadherin** domain has been found. There are 700 instances of this domain found in the PDB. Note that there may be multiple copies of the domain in a single PDB structure, since many structures contain multiple copies of the same protein sequence.

UniProt entry	UniProt residue	PDB ID	PDB chain ID	PDB residues	View
<a href="#">CAD10_MOUSE</a>	165 - 260	<a href="#">6CG6</a>	A	111 - 206	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	60 - 151	<a href="#">6CG6</a>	A	6 - 97	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	164 - 259	<a href="#">2AEF</a>	A	111 - 206	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
<a href="#">CAD11_MOUSE</a>	244G	A	6 - 97	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>	
	59 - 150	<a href="#">2AEF</a>	A	6 - 97	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	6CG6	A	6 - 97	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>	
<a href="#">CAD13_CHICK</a>	<a href="#">3K55</a>	A	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>	
	143 - 236	<a href="#">3K55</a>	B	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	250 - 354	<a href="#">3K55</a>	A	112 - 216	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
<a href="#">CAD13_HUMAN</a>	143 - 236	<a href="#">2V3Z</a>	A	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	143 - 236	<a href="#">3K5R</a>	A	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	250 - 354	<a href="#">3K5R</a>	B	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
<a href="#">CAD13_MOUSE</a>	143 - 236	<a href="#">2KGF</a>	A	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	250 - 354	<a href="#">3K5R</a>	B	5 - 98	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	112 - 216	<a href="#">3K5R</a>	B	112 - 216	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
<a href="#">CAD20_CHICK</a>	63 - 155	<a href="#">1ZVN</a>	A	5 - 97	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	170 - 265	<a href="#">6CG7</a>	A	111 - 206	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	64 - 156	<a href="#">6CG7</a>	B	111 - 206	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
<a href="#">CAD22_MOUSE</a>	1318 - 1494	<a href="#">5WJ8</a>	A	1295 - 1340	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	1424 - 1518	<a href="#">5WJ8</a>	A	1401 - 1495	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>
	2156 - 2264	<a href="#">5WJ8</a>	A	2156 - 2264	<a href="#">View 3D Structure</a> <a href="#">View in InterPro</a>

Fig14. Result page for Cadherin Domain showing structures of various chains.

Fig15. Result page for Cadherin Domain showing AlphaFold structure predictions

## Result:

The domain organisation, clan, alignment, HMM logo, trees, curation and model, species, structure and AlphaFold structure information related to Cadherin Domain was retrieved using Pfam database.

## **Conclusion:**

Pfam database provides protein domain family information which includes sequence alignments, profile Hidden Markov models which helps understand how much an amino acid is conserved. It also provides functional annotations and literature references. All this information can be easily retrieved from Pdam database by its users.

## **References:**

1. Maître, J. L., & Heisenberg, C. P. (2013). Three Functions of Cadherins in Cell Adhesion. *Current Biology*, 23(14), R626–R633. <https://doi.org/10.1016/j.cub.2013.06.019>
2. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/search/keyword?query=Cadherin>
3. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab1>
4. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab2>
5. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab3>
6. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab4>
7. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab5>
8. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab6>
9. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab7>
10. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab8>
11. Keyword search results. (n.d.). Retrieved from <http://pfam.xfam.org/family/PF00028#tabview=tab9>

## WEBLEM 5c BLOCK (URL: <https://www.ebi.ac.uk/interpro/>)

### Aim:

To study Cadherin domain information using InterPro database.

### Introduction:

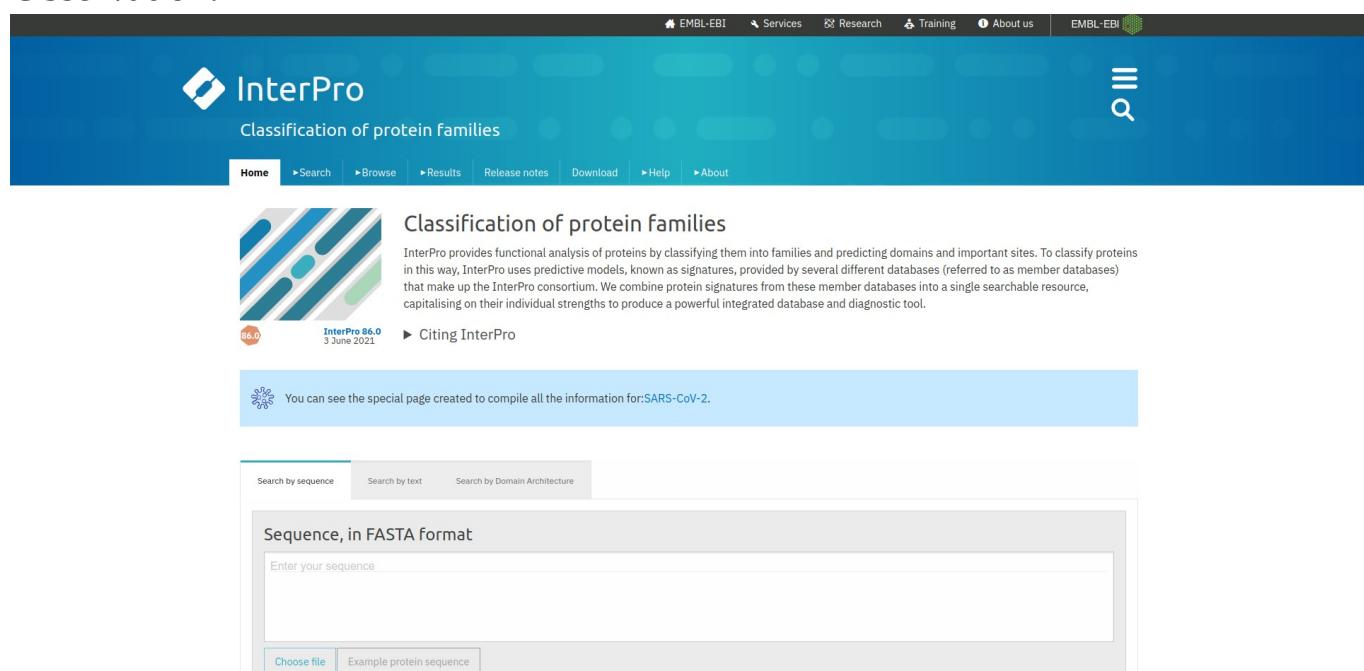
InterPro is an integrative database which was founded 10 years ago when the PROSITE, PRINTS, Pfam and ProDom databases formed a consortium to amalgamate the predictive signatures they individually produced into a single resource. Since then, six other member databases have also joined and their data has been integrated: SMART, TIGRFAMs, PIRSF, SUPERFAMILY PANTHER and Gene3D.

Cadherins (named for "calcium-dependent adhesion") are a type of cell adhesion molecule (CAM) that are important in the formation of adherens junctions to allow cells to adhere to each other. Cadherins are a class of type-1 transmembrane proteins, and they are dependent on calcium ( $\text{Ca}^{2+}$ ) ions to function, hence their name. Cell-cell adhesion is mediated by extracellular cadherin domains, whereas the Intracellular cytoplasmic tail associates with numerous adaptors and signaling proteins, collectively referred as the cadherin adhesome.

### Methodology:

1. Open Homepage for InterPro database. (URL: <https://www.ebi.ac.uk/interpro/>)
2. Search for query "Cadherin" Using search by sequence.
3. Observe the results.
4. Interpret the results.

### Observation:



**Fig1. Homepage for InterPro database.**

You can see the special page created to compile all the information for:SARS-CoV-2.

Search by sequence   Search by text   Search by Domain Architecture

**Sequence, in FASTA format**

```
>sp|P55290|CAD13_HUMAN Cadherin-13 OS=Homo sapiens OX=9606 GN=CDH13 PE=1 SV=1
MQRPTPLVLCVLLSQQVLLLTSAEQLCTPGFOOKVPHINQPAEFLIEDQGSIULNTFSDCKG
NDKLRYEVSSPPYFKNSDGGIVALRNITAVGKTLPHIARTPHAEDEAEILIVIGKGDQGS
LQDIFKPKARTSPVFPVQK851VVSPLIPEMORQGPFPVIGKVVDSDRPERSKPRLGKGV
DQEPKGIGFRINENTGSVYTRPLRBRVIAVYQFLVETTIVVQKTLGEPVPELEVIVIDQND
NRPPIFREGPVIGHVMEGSPITGTVVMTAFADOPATINALLSYNIRQQTOKPSPNMFY
IDPERKGDIVVVSSAIIORLTLENPKYELLEIAQOMAGLQVGLG7GTATATIMIDOKNQHS
PKFTKKEPQATVREGAVGVNLTVTEQDQPTTGAWRAYATIILNGNPQGSFEIHTNPQTN
EGMLSVVKVPDQEYISAFHTLILKVEQDPLVPEVSVYGSSTATTVHTLTVLOVNEGPVVFYPD
PMMVTRQEDLSVSVSLLTVNATOPDLSIOMHOTRISYVYDPAQHINININGTVOTTAVDLO
RESPFVQNSVYTAFLAIDI5GNPPATGTGTLIITLEQDNNAPIYPTVAEVCDQAKNLS
VVI1GAS3KUMLPNTDFKFEIHKKQAVPDKWYKISKINNTHALVSLLCNLHKANYNLPLIM
VTDSSGKPPMTNTIDLRVQWCSCRSNSKVDCNAAGALRFS1LPSVLLLSLFLSACI|
```

Valid Sequence.

Choose file   Example protein sequence

Advanced options

Search   Clear

Powered by InterProScan

**Fig2. Search page for query “Cadherin” using search by sequence**

EMBL-EBI   Services   Research   Training   About us   EMBL-EBI

**Your InterProScan Search Results <sup>1</sup>**

Your InterProScan search results are shown below. Searches may take varying times to complete. You can navigate to other pages and once the search is finished, you will receive a notification. The results will be available for 7 days.

Alternatively, you can import the results of an InterProScan run (in JSON format) into this page in order to view your search results interactively.

Submit a new search   Import:

1 - 1 of 1 result

RESULTS	NAME	CREATED	STATUS	ACTION
internal-1634056641591-80	sp P55290 CAD13_HUMAN Cadherin-13 OS=Homo sapiens OX=9606 GN=CDH13 PE=1 SV=1	just now	Searching	<input type="button" value="Delete"/>

Previous   **1**   Next

ELIXIR This service is part of the ELIXIR infrastructure  
InterPro is an ELIXIR Core Data Resource Learn more

**Fig3. Hit page for query “Cadherin” in InterPro database.**

EMBL-EBI Services Research Training About us EMBL-EBI

**InterPro**  
Classification of protein families

Home Search Browse Results Release notes Download Help About

Result / InterProScan / Iprscan5-R20211012-173752-0624-16335657-P1m / Overview

**Overview** Sequence

**InterProScan Search Result**

Title: sp|P55290|CADH13\_HUMAN Cadherin-13 OS=Homo sapiens OX=9606 GN=CDH13 PE=1 SV=1

Job ID: iprscan5-R20211012-173752-0624-16335657-p1m

Length: 713 amino acids

Action:

Status: Searching

**Protein family membership**

None predicted

This service is part of the ELIXIR infrastructure  
InterPro is an ELIXIR Core Data Resource [Learn more](#)

**Fig4. Result page for query “Cadherin”**

InterPro Classification of protein families

Home Search Browse Results Release notes Download Help About

Expires: 10/01/2021

**Protein family membership**

▼ Cadherin (IPR039808)  
   Cadherin-13 (IPR033216)

**Entry matches to this protein**

1 200 400 600 700 713

**Family**

IPR039808  
PTHR24027  
IPR033216  
PTHR24027:SF80

**Domain**

IPR002126  
P550268  
P550269  
PF00028  
SM00112  
IPR014868  
SM01055  
PF08758

**Homologous Superfamily**

IPR015919  
SF49313

**Conserved Site**

IPR020894  
P500232

**Unintegrated**

cd11304  
Ca2+ binding site  
G3DSA:2.60.40.60

**Predictions**

SIGNAL\_PEPIDE\_N\_REGION (1)  
SIGNAL\_PEPIDE\_H\_REGION (2)  
SIGNAL\_PEPIDE\_C\_REGION (3)

**Fig5. Result page for query “Cadherin” showing its family information**

InterPro

Classification of protein families

Home ▶ Search ▶ Browse ▶ Results ▶ Release notes ▶ Download ▶ Help ▶ About

By Entry / InterPro / IPR039808 / Overview

**Cadherin** IPR039808

InterPro entry

Overview Proteins 23k Taxonomy 2k Proteomes 545 Structures 90 AlphaFold 95 Pathways 50

Short name: Cadherin

Family relationships

▼ Cadherin (IPR039808)  
  F Cadherin-1 (IPR030049)  
  ... F N-cadherin (IPR030051)

Description

Cadherins are a group of transmembrane proteins that serve as the major adhesion molecules located within adherens junctions. They can regulate cell-cell adhesion through their extracellular domain and their cytosolic domains connect to the actin cytoskeleton by binding to catenins [1]. These proteins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. They may play an important role in the sorting of different cell types during morphogenesis, histogenesis and regeneration. They may also be involved in the regulation of tight and gap junctions, and in the control of intercellular spacing. Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins.

Structurally, cadherins comprise a number of domains: classically, these include a signal sequence; a propeptide of around 130 residues; a single transmembrane domain and five tandemly repeated extracellular cadherin domains. 4

Fig5.1. Result page for Cadherin family.

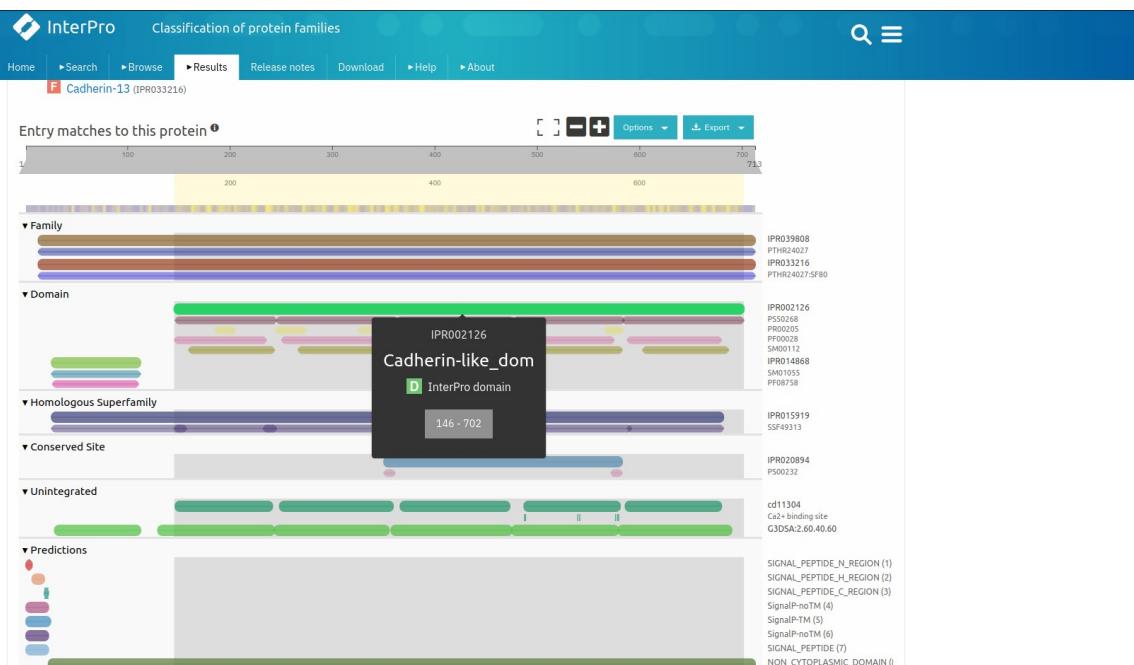


Fig6. Result page for query “Cadherin” showing its domain information

Home / Browse / By Entry / InterPro / IPR002126 / Overview

**D** Cadherin-like IPR002126  
InterPro entry

Overview Proteins 95k Domain Architectures 4k Taxonomy 7k Proteomes 2k Structures 200 AlphaFold 446 Interactions 8 Pathways 108

Short name: Cadherin-like\_dom

Overlapping homologous superfamilies 1

**H** Cadherin-like superfamily (IPR015919)

#### Description

Cadherins are a group of transmembrane proteins that serve as the major adhesion molecules located within adherens junctions. They can regulate cell-cell adhesion through their extracellular domain and their cytosolic domains connect to the actin cytoskeleton by binding to catenins [3]. These proteins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. They may play an important role in the sorting of different cell types during morphogenesis, histogenesis and regeneration. They may also be involved in the regulation of tight and gap junctions, and in the control of intercellular spacing. Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins.

Structurally, cadherins comprise a number of domains: classically, these include a signal sequence; a propeptide of around 130 residues; a single transmembrane domain and five tandemly repeated extracellular cadherin domains, 4 of which are cadherin repeats, and the fifth contains 4 conserved cysteines and a C-terminal cytoplasmic domain [2]. However, proteins are designated as members of the broadly defined cadherin family if they have one or more

### Fig6.1 Result page for Cadherin-like domain

Fig7. Result page for query “Cadherin” Showing its homologous superfamily information

[Home](#) / [Browse](#) / [Entry](#) / [InterPro](#) / [IPR015919](#) / [Overview](#)

**Cadherin-like superfamily** [IPR015919](#)  
InterPro entry

[Overview](#) [Proteins](#) 116k [Taxonomy](#) 12k [Proteomes](#) 4k [Structures](#) 208 [AlphaFold](#) 462 [Pathways](#) 121

Short name: *Cadherin-like\_sf*

#### Overlapping entries

- D Cadherin-like (IPR002126)
- D Dystroglycan-type cadherin-like (IPR006644)
- D Cadherin prodomain (IPR014868)
- F Desmosomal cadherin (IPR009122)

#### Description

This entry represents domains with an immunoglobulin-like beta-sandwich fold, consisting of 7 strands in two sheets with a Greek key topology. Such domains are found in cadherin, as well as at the N-terminal of dystroglycan. Dystroglycan is a cell surface receptor consisting of two subunits: alpha-dystroglycan, extracellular and highly glycosylated, and beta-dystroglycan, spanning the cell membrane. It is a pivotal member of the dystrophin-glycoprotein complex and is involved in a wide variety of important cellular processes such as the stabilisation of the muscle fibre sarcolemma or the clustering of acetylcholine receptors [1, 2, 3].

Cadherins are a family of adhesion molecules that mediate Ca<sup>2+</sup>-dependent cell-cell adhesion in all solid tissues of the organism which modulate a wide variety of processes including cell polarisation and migration [4, 5]. Cadherin-mediated cell-cell junctions are formed as a result of interaction between extracellular domains of identical

### Fig7.1 Result page for Cadherin-Like superfamily

### Fig8. Result page for query “Cadherin” Showing conserved site information

**Fig8.1 Result page for Cadherin conserved site**

**Fig9. Result page for query “Cadherin” showing Unintegrated information**

**Fig9.1 Result page for Cadherin tandem repeat domain Unintegrated information**

**Fig10. Result page for query “Cadherin” Showing predictions**

## Result:

The family, domains, conserved sites, unintegrated information and predictions for Cadherin protein was retrieved from InterPro database

## **Conclusion:**

InterPro database provides its users functional analysis of protein which are classified using predictive models into families and predictive domains. It provides information of protein family, domains, conserved and active sites. This database is useful for cross referencing protein domain information with other databases.

## **References:**

1. Maître, J. L., & Heisenberg, C. P. (2013). Three Functions of Cadherins in Cell Adhesion. *Current Biology*, 23(14), R626–R633. <https://doi.org/10.1016/j.cub.2013.06.019>
2. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/>
3. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/result/InterProScan/#table>
4. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/result/InterProScan/iprscan5-R20211012-173752-0624-16335657-p1m/>
5. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR039808/>
6. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR002126/>
7. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR015919/>
8. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR020894/>
9. (n.d.). Retrieved from <https://www.ebi.ac.uk/interpro/entry/cdd/CD11304/>