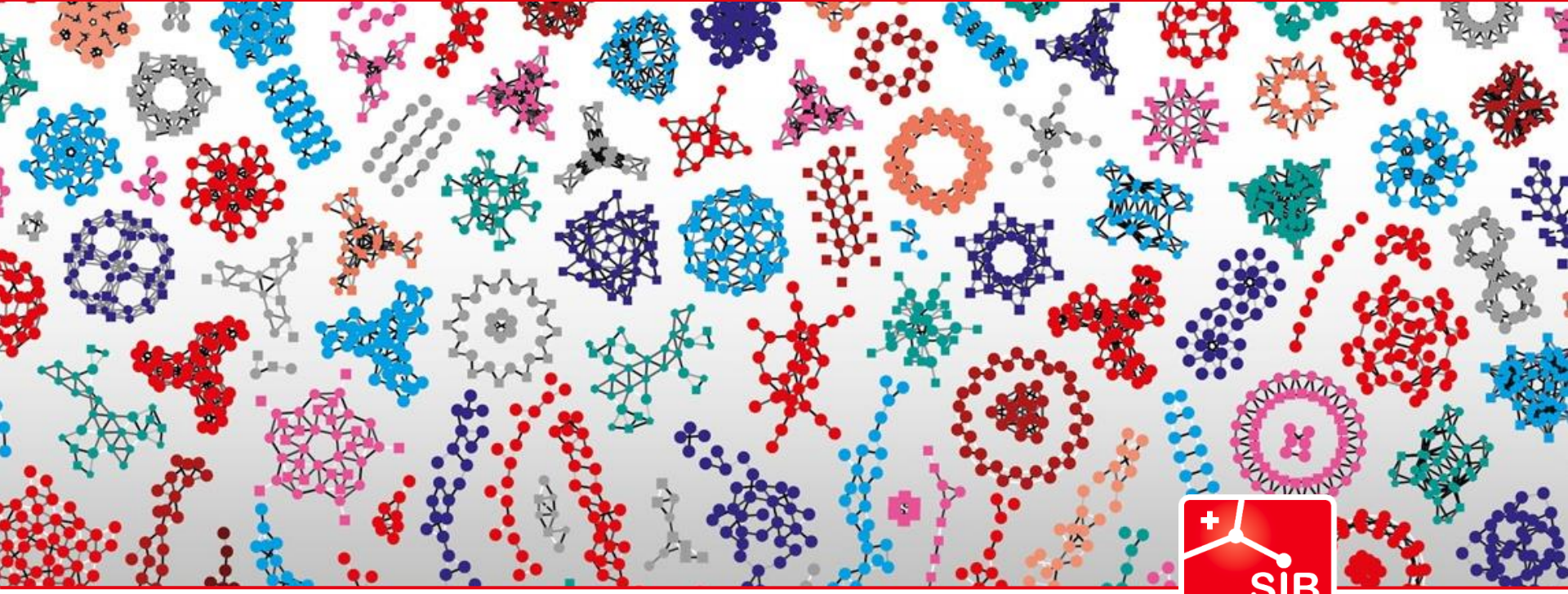
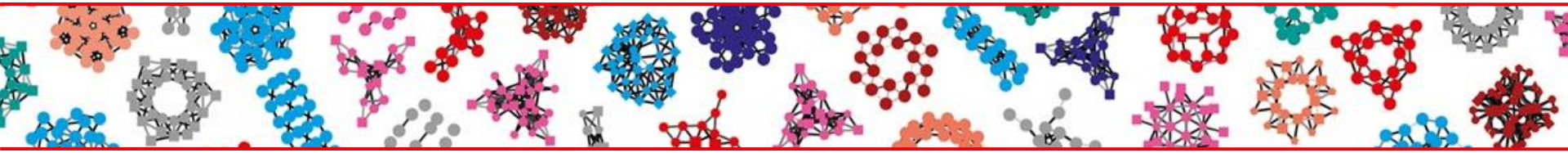


Swiss Institute of Bioinformatics

Automatic annotation in UniProtKB

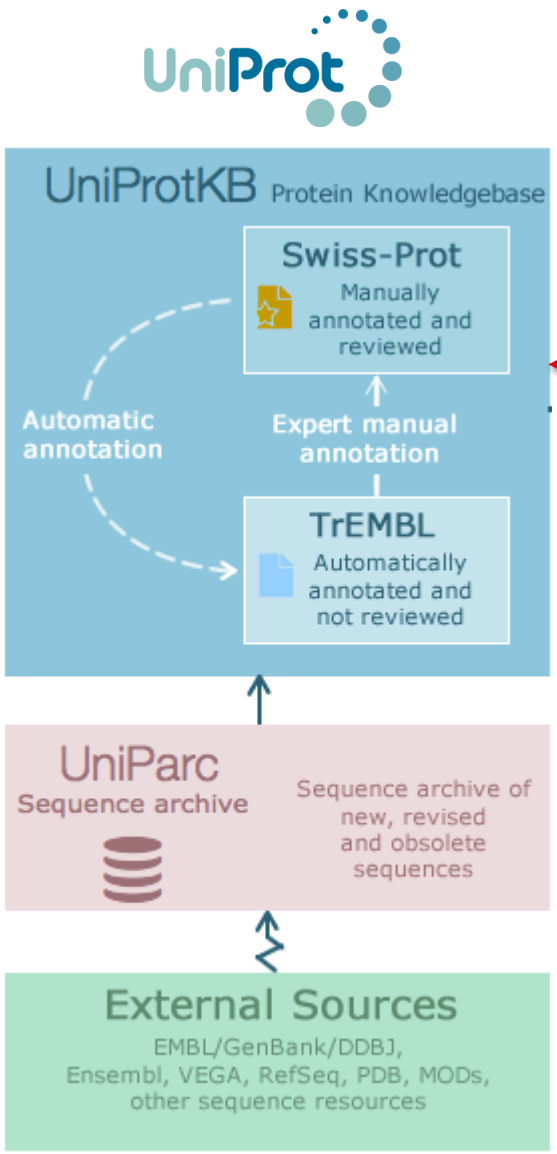
Protein sequence databases and sequence annotation

Streamed from Geneva, 14 October 2022



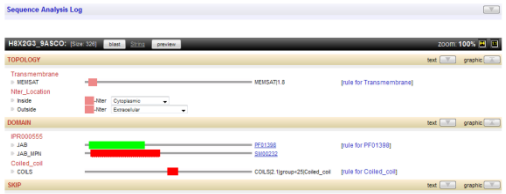
- **Why do we need predictive annotation tools?**
- Protein signatures for homology detection – A short primer
- Annotation rules for functional annotation
- HAMAP and PROSITE - automatic annotation in UniProtKB
- HAMAP and PROSITE - services for external users
- Practical exercises

The UniProt knowledgebase UniProtKB



Literature curation

Curator-supervised sequence analysis



- Motif database scans: Prosite, Pfam, Smart, etc.
- Prediction programs: TMHMM, coils, SignalP, TargetP, Repeat, etc.

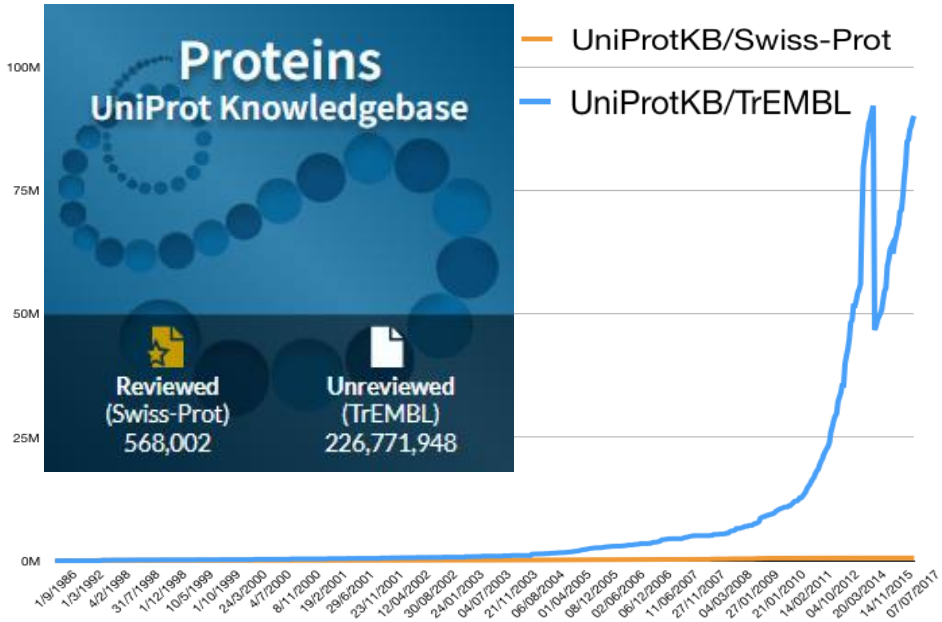
- Protein and gene names
- Function
- Catalytic activity
- Cofactors
- Subcellular location
- Expression
- Protein-protein interactions
- Domains/regions
- Catalytic residues
- PTMs
- Binding sites
- Splice variants

Why do we need predictive annotation tools?

«Continuing advances in next generation sequencing mean that for every experimentally characterized protein, there are now many hundreds of proteins that will never be experimentally characterized in the laboratory !»

http://nar.oxfordjournals.org/content/43/D1/D204

UniProt release 2022_03:



- Currently, reviewed entries make up less than 1% of UniProtKB
- Manual curation is time-intensive and published experimental data focuses on a rather limited number of model organisms
- Problem: The quality of the functional annotation attached to sequences submitted to UniProtKB is very variable (from original submitters).

Annotation propagation in UniProtKB/Swiss-Prot

MT21D_HUMAN

Function

Protein N-lysine methyltransferase that specifically trimethylates 'Lys-315' of VCP/p97; this modification may decrease VCP ATPase activity. 2 Publications

Binding site	43	S-adenosyl-L-methionine (UniProtKB | ChEBI) 1 Publication Manual assertion based on experiment (Inferred from experiment) Ref.9
Binding site	75-77	S-adenosyl-L-methionine (UniProtKB | ChEBI) 1 Publication
Binding site	96	S-adenosyl-L-methionine (UniProtKB | ChEBI) 1 Publication

2 Publications

MT21D_MOUSE

Function

Protein N-lysine methyltransferase that specifically trimethylates 'Lys-315' of VCP/p97; this modification may decrease VCP ATPase activity. By Similarity

Binding site	43	S-adenosyl-L-methionine (UniProtKB | ChEBI) By Similarity Manual assertion inferred from sequence similarity (Inferred from sequence or structural similarity) Q9H867
Binding site	75-77	S-adenosyl-L-methionine (UniProtKB | ChEBI) By Similarity
Binding site	96	S-adenosyl-L-methionine (UniProtKB | ChEBI) By Similarity

By Similarity

Microbial proteome annotation

Statistics

Species	Species code	UniProtKB proteome identifier	Number of entries in UniProtKB		
			All	Swiss-Prot	TrEMBL
Acaryochloris marina (strain MBIC 11017)	ACAM1	UP000000268	8172	322	7850
Accumulibacter phosphatis (strain UW-1)	ACCPU	UP000001619	4438	1	4437
Acetobacter pasteurianus (strain NBRC 3283 / LMG 1513 / CCTM 1153)	ACEP3	UP000000948	2906	4	2902
Acetobacter sp. CAG:267	/	UP000017992	1538	0	1538
Acetobacter sp. CAG:977	/	UP000018259	1775	0	1775
Acetobacterium woodii (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)	ACEWD	UP000007177	3445	4	3441
Escherichia coli (strain K12)	ECOLI	UP000000625	4305	4305	0
Acholeplasma laidlawii (strain PG-8A)	ACHLI	UP000008558	1380	147	1233
Bacillus subtilis (strain 168)	BACSU	UP000001570	4197	4185	12
Acholeplasma sp. CAG:878	/	UP000017930	1145	0	1145
Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv)	MYCTU	UP000001584	3991	2081	1910
Acidaminococcus fermentans (strain ATCC 25085 / DSM 20731 / VR4)	ACIFV	UP000001902	2016	12	2004
Acidaminococcus intestini (strain RyC-MR95)	ACIIR	UP000007093	2386	0	2386
Acidaminococcus sp. CAG:917	/	UP000018048	1276	0	1276
Acidimicrobium ferrooxidans (strain DSM 10331 / JCM 15462 / NBRC 103882 / ICP)	ACIFD	UP000000771	1935	8	1927
Acidiphilium cryptum (strain JF-5)	ACICJ	UP000000245	3521	262	3259
Acidiphilium sp. CAG:727	/	UP000018053	1479	0	1479

Annotation propagation in UniProtKB/Swiss-Prot

MT21D_HUMAN

Function

Protein N-lysine methyltransferase that specifically trimethylates 'Lys-315' of VCP/p97; this modification may decrease VCP ATPase activity. 2 Publications

Binding site	43	S-adenosyl-L-methionine (UniProtKB | ChEBI) 1 Publication Manual assertion based on experiment (Inferred from experiment) Ref.9
Binding site	75-77	S-adenosyl-L-methionine (UniProtKB | ChEBI) 1 Publication
Binding site	96	S-adenosyl-L-methionine (UniProtKB | ChEBI) 1 Publication

2 Publications

MT21D_MOUSE

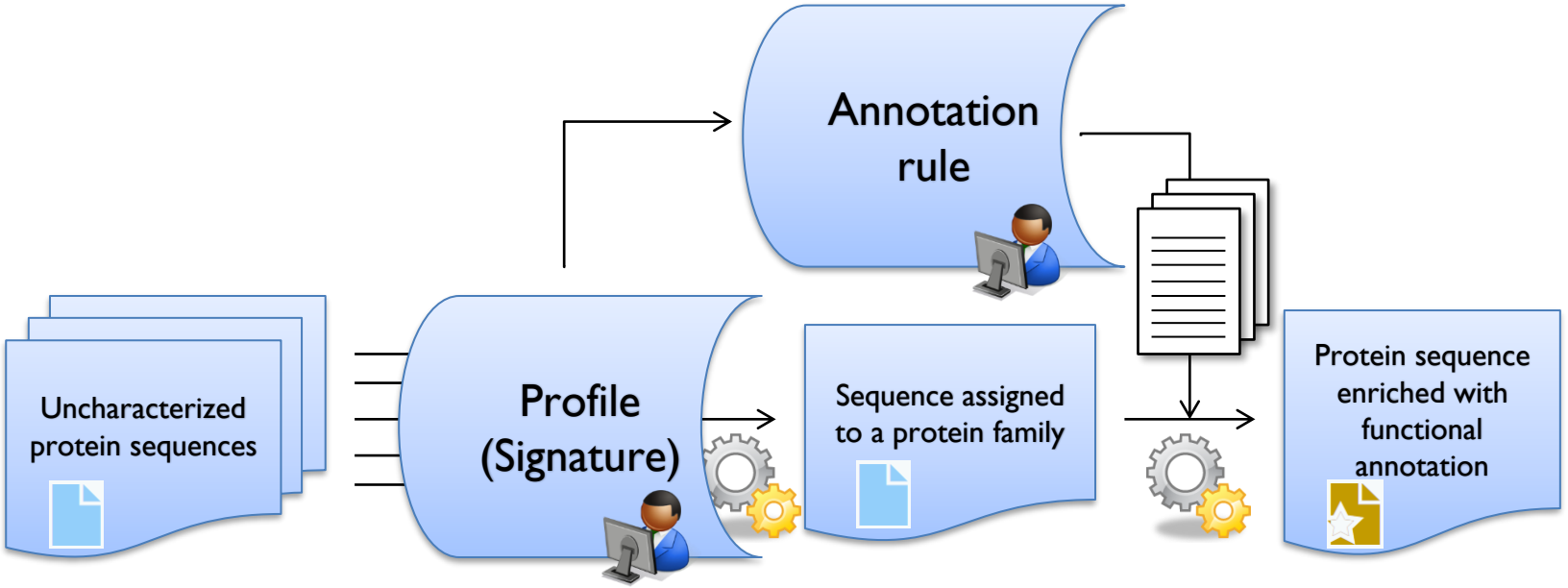
Function

Protein N-lysine methyltransferase that specifically trimethylates 'Lys-315' of VCP/p97; this modification may decrease VCP ATPase activity. By Similarity

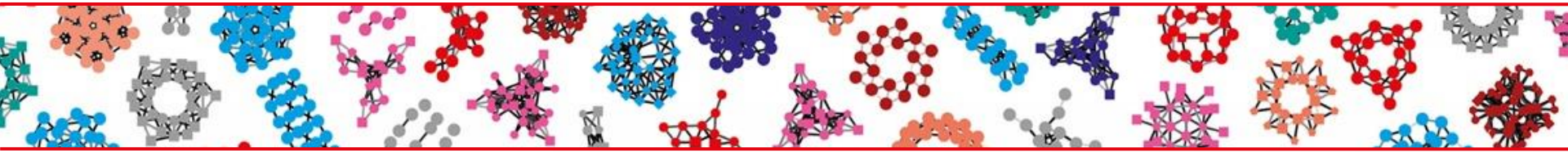
Binding site	43	S-adenosyl-L-methionine (UniProtKB | ChEBI) By Similarity Manual assertion inferred from sequence similarity (Inferred from sequence or structural similarity) Q9H867
Binding site	75-77	S-adenosyl-L-methionine (UniProtKB | ChEBI) By Similarity
Binding site	96	S-adenosyl-L-methionine (UniProtKB | ChEBI) By Similarity

By Similarity

Automatic annotation of protein sequences



- Manually curated **signatures** – a protein matching a specific signature may be annotated according to the contents of the associated rule
- **Manually** created annotation **rules** that specify **annotations** AND the **conditions** under which they may be applied
- **HAMAP** signatures/rules for **full length protein sequences**, **PROSITE** signatures/rules mainly for **domains** and **sites**
- Common format and shared syntax



- Why do we need predictive annotation tools?
- **Protein signatures for homology detection – A short primer**
- Annotation rules for functional annotation
- HAMAP and PROSITE - automatic annotation in UniProtKB
- HAMAP and PROSITE - services for external users
- Practical exercises

Types of sequence similarity

Family

Groups of proteins that are conserved along the whole sequence, sharing a common evolutionary origin, as reflected in their related functions.

Domains

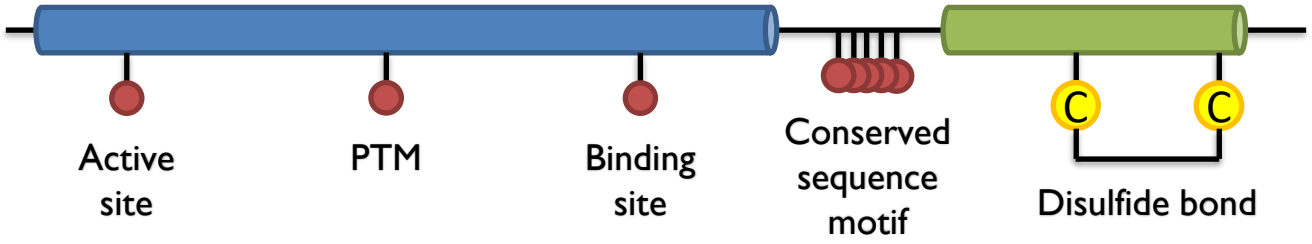
Specific combination of secondary structures organized into a characteristic three dimensional structure or fold, that may exist in a variety of biological contexts.

Repeats

Structural units typically repeated within a protein that assemble into a specific fold. Assemblies of repeats might also be thought of as domains.

Sites & Motifs

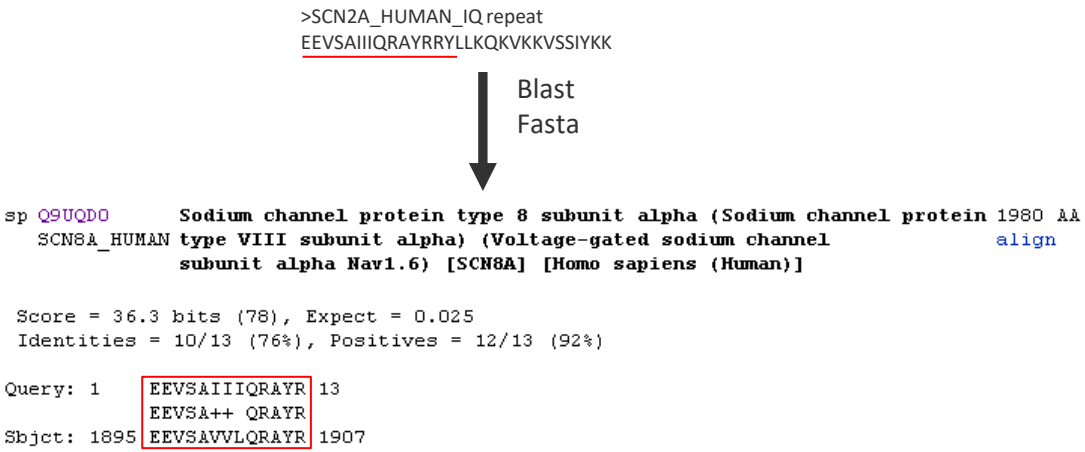
Region of domains containing conserved active-site or binding residues or short conserved regions present outside domains that may adopt folded conformations only in association with their binding ligands



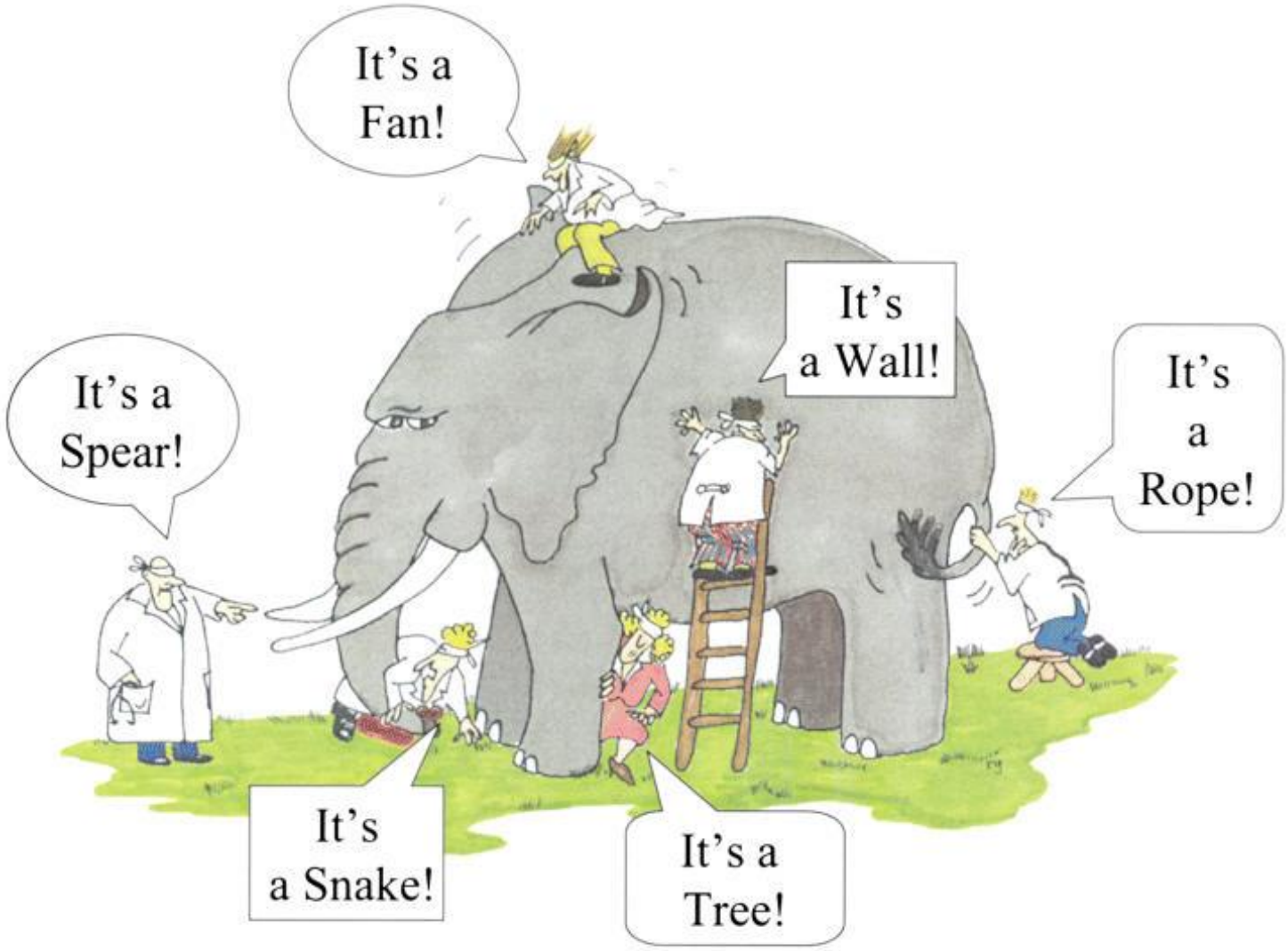
Similarity identification with pairwise alignments

A popular way to identify similarities between proteins is to perform a pairwise alignment (Smith-Watermann, Needlemann-Wunsch, BLAST, …).

Normally, when the identity is higher than 40% this method gives good results.



Only the N-ter of the query sequence matches and with a low score!


It's a Fan!
It's a Spear!
It's a Wall!
It's a Rope!
It's a Snake!
It's a Tree!

Pairwise sequence alignments vs MSAs

Another weakness of the pairwise alignment is that no distinction is made between an amino acid at a crucial position (like an active site) and an amino acid with no critical role.

```
                    *         20          *
SCN2A_HS : EEVSAIIIQRAYRRYLLKQKVKKVSSIYKK : 30
SCN8A_HS : EEVSAVVLQRAYRGHLARRGFICKKTTSNK : 30
           EEVSA666QRAYR  L 4       3   K
```

```
                     *        20         *
SCN2A_HS    : EEVSAIIIQRAYRRYLLKQKVKKVSSIYKK : 30
SCN2A_RN    : EEVSAIVIQRAYRRYLLKQKVKKVSSIYKK : 30
SCN3A_HS    : EEVSAAIIQRNFRCYLLKQRLKNISSNYNK : 30
SCN8A_HS    : EEVSAVVLQRAYRGHLARRGFICKKTTSNK : 30
SCN8A_MM    : EEVSAVVLQRAYRGHLARRGFICRKITSNK : 30
IQGA1_HS_1  : NEGLITRLQARCRGYLVRQEFRSRMNFLKK : 30
IQGA1_HS_2  : QIPAITCIQSQWRGYKQKKAYQDRLAYLRS : 30
IQGA1_HS_3  : HKDEVVKIQSLARMHQARKRYRDRLQYFRD : 30
IQGA1_HS_4  : HINDIIKIQAFIRANKARDDYKTLINAEDP : 30
IQGA1_MM_1  : NEGLITKLQACCRGYLVRQEFRSRMNFLKK : 30
                    6Q   R    4
```

A multiple sequence alignment (MSA) gives a more general view of a conserved region by providing a better picture of the most conserved residues, which are usually essential for the protein function. It can help to identify subfamilies. An MSA contains more information than a pairwise alignment and several tools have been developed to extract this information.

Extracting Information from MSAs

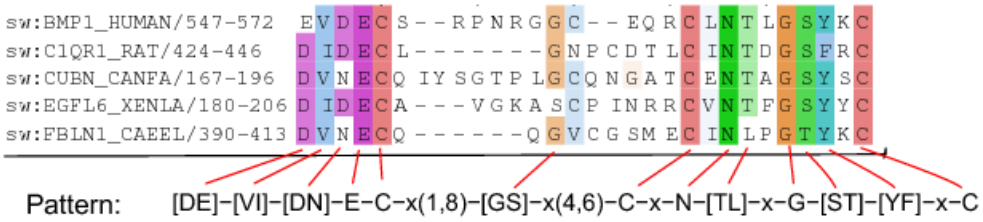
PROSITE patterns use a special syntax to describe the consensus of all the sequences present in the multiple alignment using a single expression.

Used to describe small functional regions:

- Enzyme catalytic sites;
- Prosthetic group attachment sites (heme, PLP, biotin, etc.);
- Amino acids involved in binding a metal ion;
- Cysteines involved in disulfide bonds;
- Regions involved in binding a molecule (ATP, calcium, DNA, etc.) or a protein.

Excellent tool to annotate active sites in combination with profiles (ProRules).

How to build a PROSITE pattern



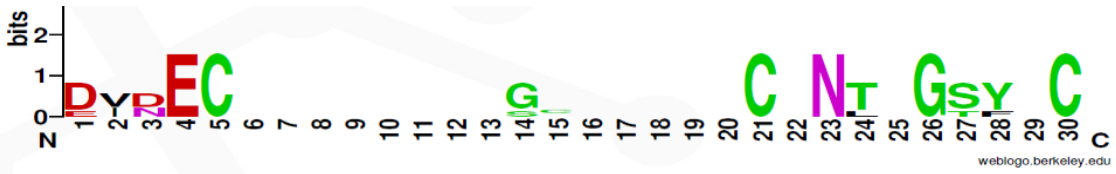
Collect sequences known to contain the signature and produce a multiple sequence alignment of the region of interest.

Build a pattern.

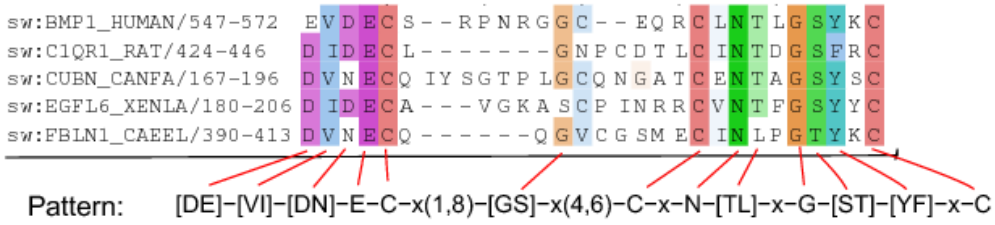
- By hand

- You can use automatic methods (e.g. http://web.expasy.org/pratt/) or a sequence logo to guide you

Example using a sequence logo (http://weblogo.berkeley.edu/logo.cgi):



Tricks to build a PROSITE pattern



- For the construction of the pattern, it is useful to consider residues and regions proved/thought to be important to the biological function of that group of proteins (e.g. enzyme catalytic sites, etc.).
- A first pattern is built from the MSA of the most conserved residues. It is used to scan the database.
- If it picks up too many false positives, it is modified to make it more stringent.
- The difficulty resides in achieving a pattern which does not pick up too many false positives yet does not miss too many sequences (false negatives).
- In some cases this result can not be achieved and an optimal sequence pattern can not be built.

How to estimate the quality of a pattern

- We can not estimate the quality of a match with a pattern: PATTERNS don't produce a score, they match or not!
- But we can estimate the quality of the pattern.
- Two parameters can be computed to estimate the quality of a pattern: precision and recall.

False positives = known false hits.

False negatives = known missed hits.

Precision = true hits/(true hits + false positives).

"how useful the search results are"

Precision = 1 ⇒ no false positive.

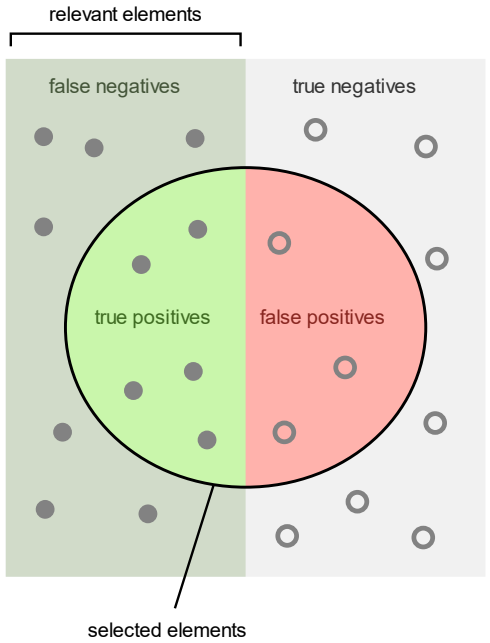
Precision = 0.8 ⇒ 20% false positives.

Recall = true hits/(true hits + false negatives).

"how complete the results are"

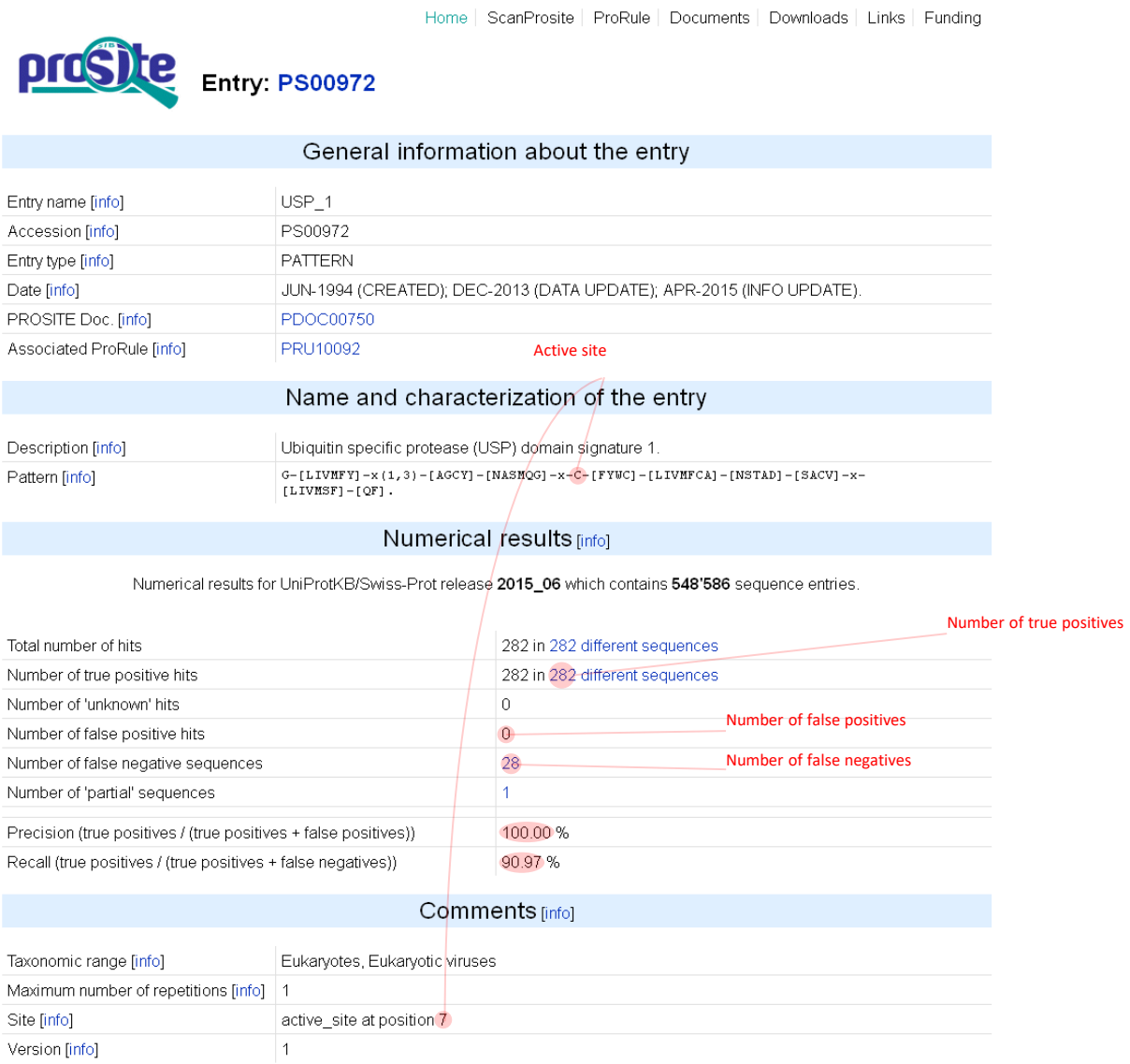
Recall = 1 ⇒ No missed hits.

Recall = 0.8 ⇒ 20% missed hits.



- To obtain these measures we require a well annotated protein databases (PROSITE uses UniProtKB/Swiss-Prot).

PROSITE patterns: example of a pattern entry



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

Entry: PS00972

General information about the entry

Entry name [info]	USP_1
Accession [info]	PS00972
Entry type [info]	PATTERN
Date [info]	JUN-1994 (CREATED); DEC-2013 (DATA UPDATE); APR-2015 (INFO UPDATE).
PROSITE Doc. [info]	PDOC00750
Associated ProRule [info]	PRU10092

Active site

Name and characterization of the entry

Description [info]	Ubiquitin specific protease (USP) domain signature 1.
Pattern [info]	G-[LIVMFY]-x(1,3)-[AGCY]-[NASMQG]-x-C-[FYWC]-[LIVMFCA]-[NSTAD]-[SACV]-x-[LIVMSF]-[QF].

Numerical results [info]

Numerical results for UniProtKB/Swiss-Prot release **2015_06** which contains **548'586** sequence entries.

Number of true positives

Total number of hits	282 in 282 different sequences
Number of true positive hits	282 in 282 different sequences
Number of 'unknown' hits	0
Number of false positive hits	0
Number of false negative sequences	28
Number of 'partial' sequences	1
Precision (true positives / (true positives + false positives))	100.00 %
Recall (true positives / (true positives + false negatives))	90.97 %

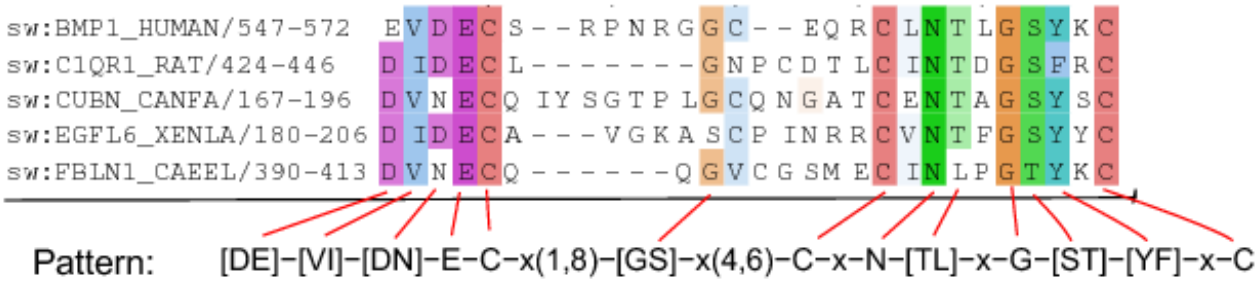
Number of false positives

Number of false negatives

Comments [info]

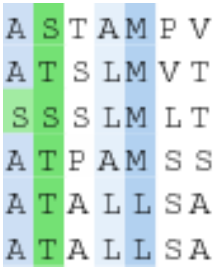
Taxonomic range [info]	Eukaryotes, Eukaryotic viruses
Maximum number of repetitions [info]	1
Site [info]	active_site at position 7
Version [info]	1

Limitations of PROSITE pattern



- OK to detect and annotate very conserved regions, but poor gap models
- residues at one position are considered equivalent in their frequencies
- if a symbol is not present at one position, this will exclude variants that have not yet been observed from being detected
- no score of the match is produced (you match or not)

Profiles



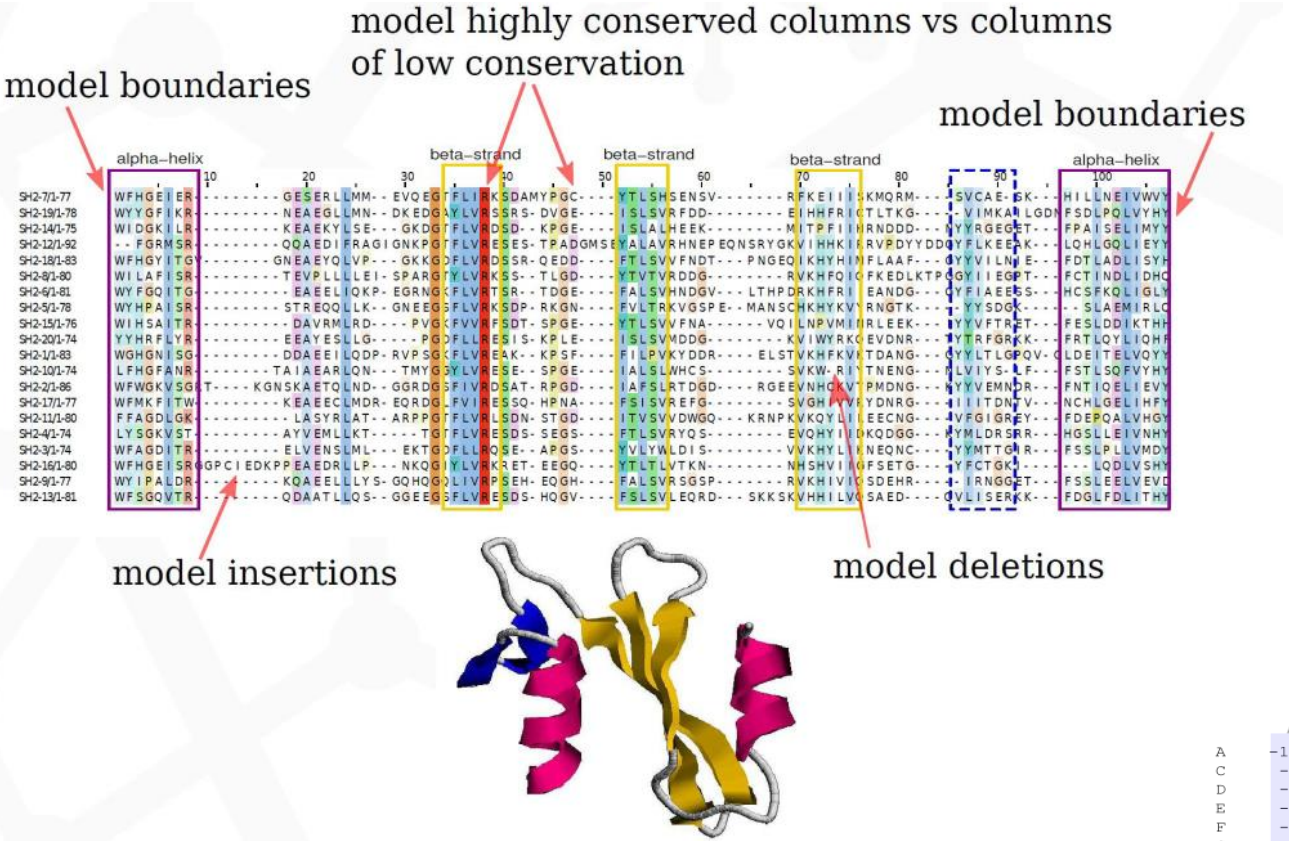
Profiles turn a multiple sequence alignment into a position-specific scoring system suitable for searching databases for remotely homologous sequences

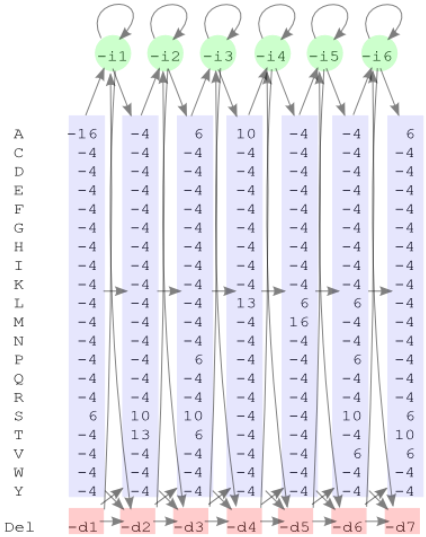
- Sequence weighting: correct sampling bias.
- Residue counts: get the frequency of each residue at each position of the MSA.
- Pseudo-counts: avoid frequencies of 0 ⇒ avoid exclusion of residues.
- Build the final scoring matrix: used to build and score alignments.

	A	S	T	A	M	P	V
w=0.2	A	S	T	A	M	P	V
w=0.2	A	S	S	L	M	V	T
w=0.2	S	S	S	L	M	L	T
w=0.2	A	T	P	A	M	S	S
w=0.1	A	T	A	L	L	S	A
w=0.1	A	T	A	L	L	S	A
	1	2	3	4	5	6	7
A	0.8	0.0	0.2	0.4	0.0	0.0	0.2
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	0.0	0.0	0.0	0.0
F	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0
H	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.0	0.0	0.0
K	0.0	0.0	0.0	0.0	0.0	0.0	0.0
L	0.0	0.0	0.0	0.6	0.2	0.2	0.0
M	0.0	0.0	0.0	0.0	0.8	0.0	0.0
N	0.0	0.0	0.0	0.0	0.0	0.0	0.0
P	0.0	0.0	0.2	0.0	0.0	0.2	0.0
Q	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S	0.2	0.4	0.4	0.0	0.0	0.4	0.2
T	0.0	0.6	0.2	0.0	0.0	0.0	0.4
V	0.0	0.0	0.0	0.0	0.0	0.2	0.2
W	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0

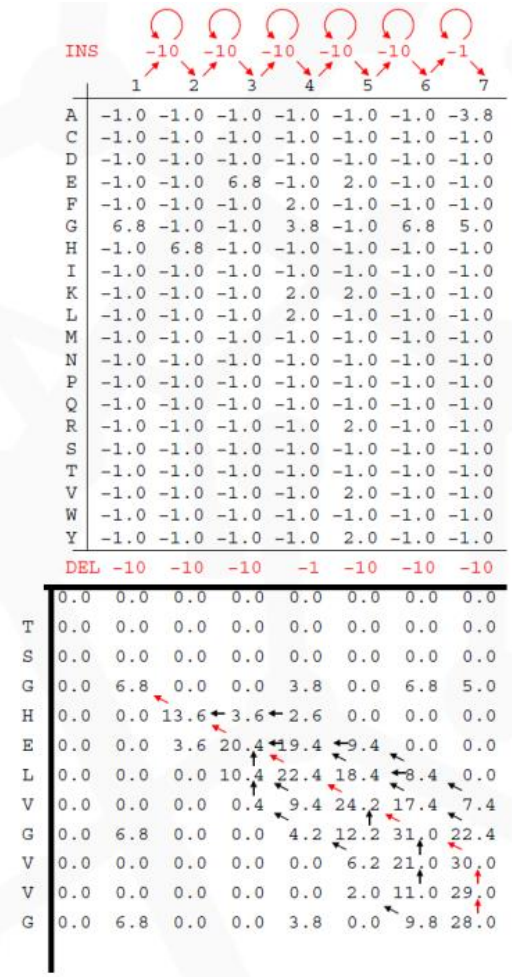
	1	2	3	4	5	6	7
A	16	-4	6	10	-4	-4	6
C	-4	-4	-4	-4	-4	-4	-4
D	-4	-4	-4	-4	-4	-4	-4
E	-4	-4	-4	-4	-4	-4	-4
F	-4	-4	-4	-4	-4	-4	-4
G	-4	-4	-4	-4	-4	-4	-4
H	-4	-4	-4	-4	-4	-4	-4
I	-4	-4	-4	-4	-4	-4	-4
K	-4	-4	-4	-4	-4	-4	-4
L	-4	-4	-4	13	6	6	-4
M	-4	-4	-4	-4	16	-4	-4
N	-4	-4	-4	-4	-4	-4	-4
P	-4	-4	6	-4	-4	6	-4
Q	-4	-4	-4	-4	-4	-4	-4
R	-4	-4	-4	-4	-4	-4	-4
S	6	10	10	-4	-4	10	6
T	-4	13	6	-4	-4	-4	10
V	-4	-4	-4	-4	-4	6	6
W	-4	-4	-4	-4	-4	-4	-4
Y	-4	-4	-4	-4	-4	-4	-4

Modeling of profiles





Scoring: aligning a sequence to a profile



	INS	1	2	3	4	5	6	7
		-10	-10	-10	-10	-10	-1	
A		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-3.8
C		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
D		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
E		-1.0	-1.0	6.8	-1.0	2.0	-1.0	-1.0
F		-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
G		6.8	-1.0	-1.0	3.8	-1.0	6.8	5.0
H		-1.0	6.8	-1.0	-1.0	-1.0	-1.0	-1.0
I		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
K		-1.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0
L		-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
M		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
N		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
P		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Q		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
R		-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
S		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
T		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
V		-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
W		-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Y		-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
	DEL	-10	-10	-10	-1	-10	-10	-10

	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	0.0	6.8	0.0	0.0	3.8	0.0	6.8	5.0
H	0.0	0.0	13.6	3.6	2.6	0.0	0.0	0.0
E	0.0	0.0	3.6	20.4	19.4	9.4	0.0	0.0
L	0.0	0.0	0.0	10.4	22.4	18.4	8.4	0.0
V	0.0	0.0	0.0	0.4	9.4	24.2	17.4	7.4
G	0.0	6.8	0.0	0.0	4.2	12.2	31.0	22.4
V	0.0	0.0	0.0	0.0	0.0	6.2	21.0	30.0
V	0.0	0.0	0.0	0.0	0.0	2.0	11.0	29.0
G	0.0	6.8	0.0	0.0	3.8	0.0	9.8	28.0

The similarity of new sequences to an existing profile can be tested by comparing each new sequence to the profile using a modification of the Smith/Waterman algorithm

The comparison of a sequence symbol to any row of the profile defines a specific value or "profile comparison value."

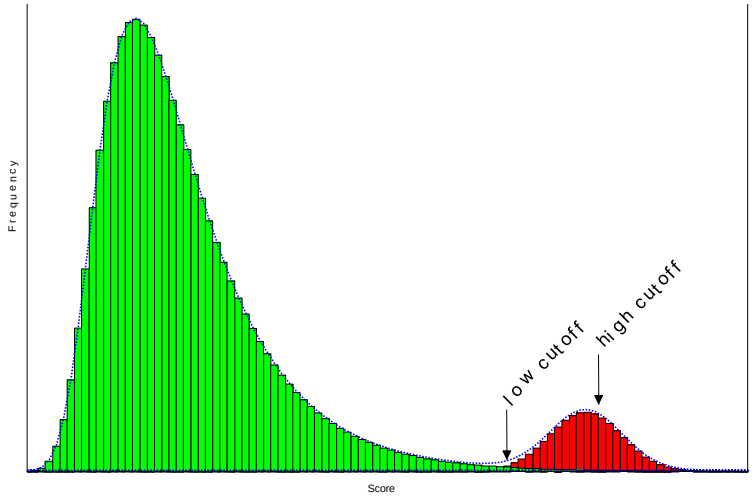
The best alignments of a sequence to a profile are found by aligning the symbols of the sequence to the profile in such a way that the sum of the profile comparison values minus the gap penalties is maximal

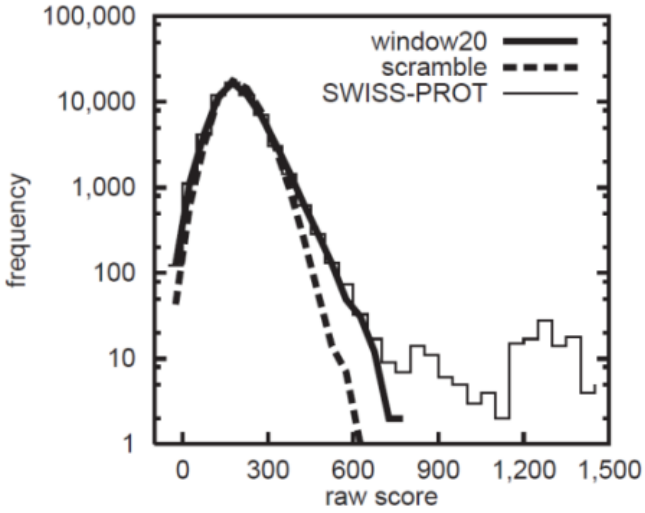
```
ALIGNMENT: score 28

     1 2 3 4 5 6 - - 7
     | | | | | |     |
T S G H E L V G V V G
```

Profiles: interpretation of the score

- How do I interpret the score produced by a profile? Which is the lowest score I consider to produce a true match?
- Only biological arguments tell you if a match is true or not.
- However, a statistical analysis can help us decide if a match is statistically significant (true positive) or not (false positive).





- The score distribution of a profile on unrelated sequences is approximated by an Extreme Value Distribution (EVD) (green bars).
- This property permits to calculate the E-value: the number of matches that we expect to occur by chance with a score ≥ a given cut off.

Summary about patterns and profiles

Patterns

- model a multiple sequence alignment using a compact string
- suited to model short and well conserved motifs
- good to describe functional residues
- easy to build, but not producing a score

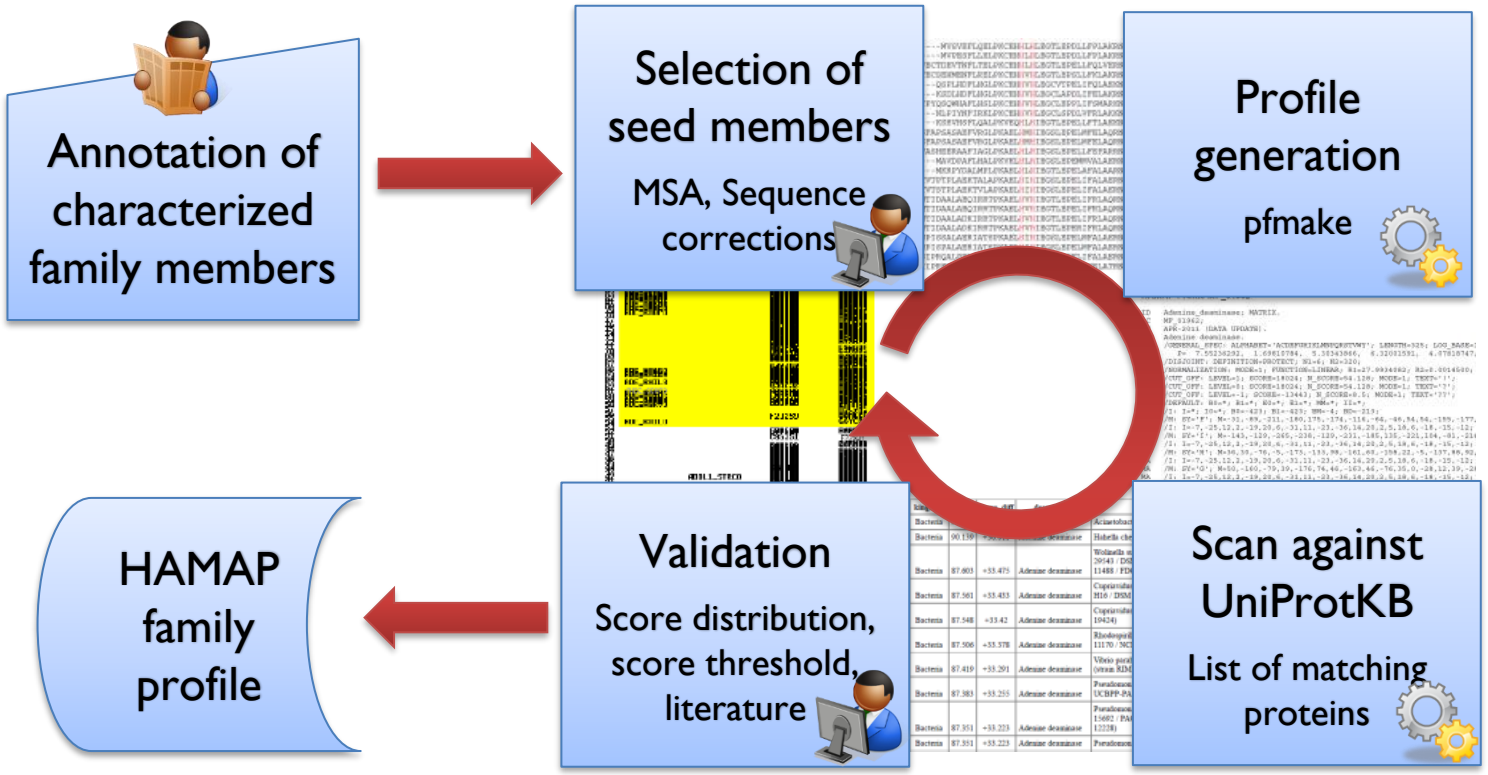
Profiles

- model a multiple sequence alignment using a numerical matrix representing the position-specific distribution of the residues
- suited to model protein domains and gapped motifs
- excellent technology to detect distant homologies
- matches produce a score that can be interpreted using statistical methods

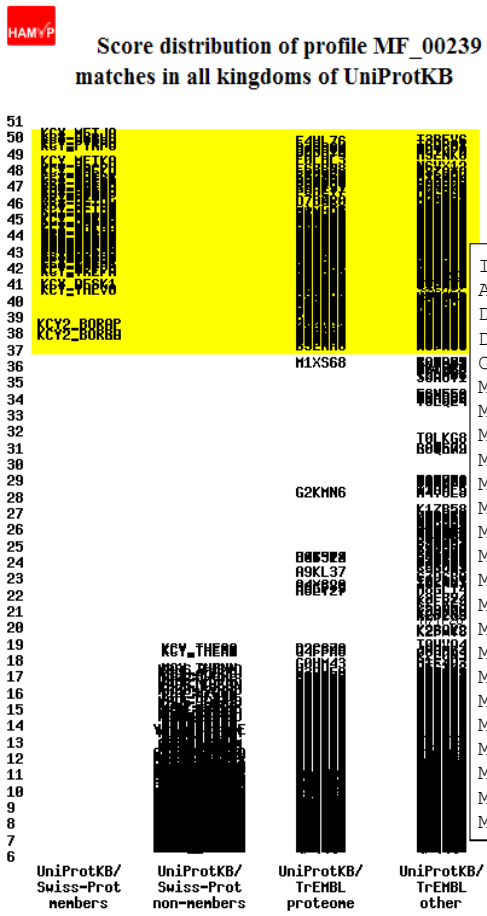
Profiles and pattern can be used together (rules) to produce precise annotation

HAMAP family profiles

- **HAMAP family profiles** are automatically generated from manually curated **seed alignments** of trusted family members and stored in PROSITE format.
- Selection and level of detail for protein family annotation is often based on expert opinion in the scientific literature including phylogenomic considerations.



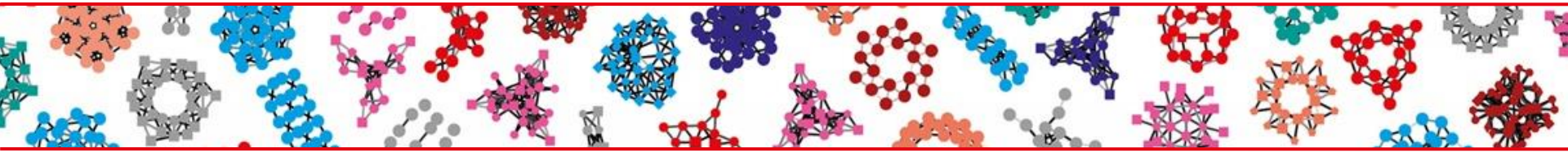
Validation of a family profile



List of UniProtKB/Swiss-Prot true positive matches for the profile MF_00239
(*i.e.* HAMAP MF_00239 UniProtKB/Swiss-Prot members)

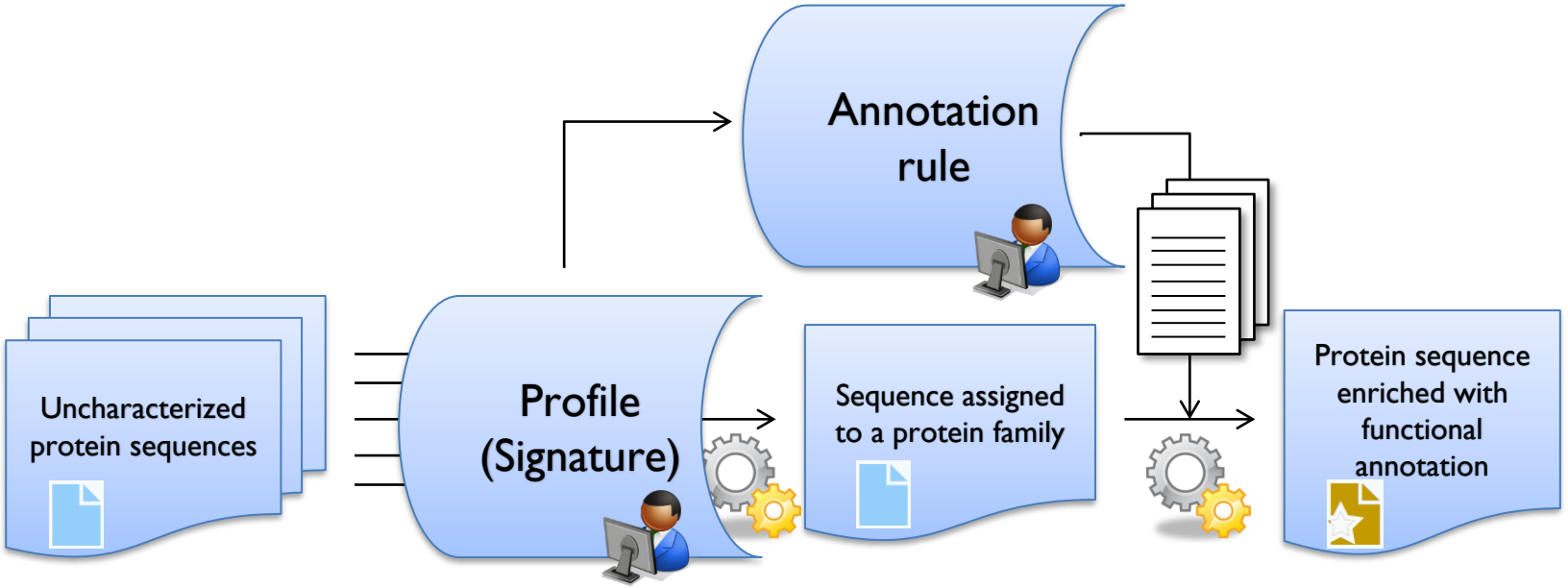
ac	id	kingdom	score	score_diff	description	organism
Q58071	KCY_METJA	Archaea	50.213	+12.972	Cytidylate kinase	Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440)
Q3IMV8	KCY_NATPD	Archaea	47.159	+9.918	Cytidylate kinase	Natronomonas pharaonis (strain ATCC 35678 / DSM 2160)
C6A187	KCY_THESM	Archaea	47.159	+9.918	Cytidylate kinase	Thermococcus sibiricus (strain MM 739 / DSM 12597)

```
ID   Cytidyl_kinase_type2; MATRIX.
AC   MF_00239;
DT   DEC-2013 (DATA UPDATE).
DE   Cytidylate kinase [cmk].
CC   /VERSION=4;
MA   /GENERAL_SPEC: ALPHABET='ACDEFGHIKLMNPQRSTVWY'; LENGTH=191; LOG_BASE=1.071779; P0=0.9972;
MA      P=   7.552363,   1.698108,   5.303439,   6.320015,   4.078187,   6.844419,   2.240667,
MA   /DISJOINT: DEFINITION=PROTECT;
MA   /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=17.[illegible]3045; R2=0.0080365; TEXT='-LogE';
MA   /CUT_OFF: LEVEL=1; SCORE=25[illegible]4; N_SCORE=37.241; MODE=1; TEXT='!';
MA   /CUT_OFF: LEVEL=0; SCORE=25[illegible]4; N_SCORE=37.241; MODE=1; TEXT='?';
MA   /CUT_OFF: LEVEL=-1; SCORE=-10[illegible] N_SCORE=8.5[illegible]ODE=1; TEXT='??';
MA   /DEFAULT: B0=*; B1=*; E0=*; E1=*; MM=*; II=*;
MA   /I: B0=-85; B1=-85; BD=-49;
MA   /M: SY='M'; M=-25,-24,-49,-42,8,-41,-30,-3,-10,-4,46,-37,-41,-35,-8,-32,-6,6,-28,-25; M0=-
MA   /I: MM=0; MI=-100; MD=-111; IM=-9; II=-11; DM=-7; DD=-14;
MA   /M: SY='R'; M=-24,-25,-39,-33,-2,-39,-27,20,0,-1,-17,-3,-39,-4,21,3,-24,11,-29,-25; M0=-12;
MA   /I: MM=0; MI=-100; MD=-111; IM=-9; II=-11; DM=-7; DD=-14;
MA   /M: SY='I'; M=-44,-39,-71,-68,-43,-70,-70,39,-68,-13,-30,-66,-66,-66,-69,-64,-44,4,-62,-57;
MA   /I: MM=0; MI=-100; MD=-111; IM=-9; II=-11; DM=-7; DD=-14;
MA   /M: SY='T'; M=10,11,-59,-53,-34,-51,-43,3,-50,-29,5,-49,-52,-47,-49,-43,31,15,-41,-37; M0=-
MA   /I: MM=0; MI=-101; MD=-111; IM=-9; II=-11; DM=-7; DD=-14;
MA   /M: SY='I'; M=-44,-39,-71,-68,-44,-70,-71,35,-68,-4,-31,-67,-67,-67,-70,-64,-44,20,-63,-58;
M0=-44;
```



- Why do we need predictive annotation tools?
- Protein signatures for homology detection – A short primer
- **Annotation rules for functional annotation**
- HAMAP and PROSITE - automatic annotation in UniProtKB
- HAMAP and PROSITE - services for external users
- Practical exercises (afternoon)

Automatic annotation of protein sequences



- Manually curated **signatures** – a protein matching a specific signature may be annotated according to the contents of the associated rule
- **Manually** created annotation **rules** that specify **annotations** AND the **conditions** under which they may be applied
- **HAMAP** signatures/rules for **full length protein sequences**, **PROSITE** signatures/rules mainly for **domains** and **sites**
- Common format and shared syntax

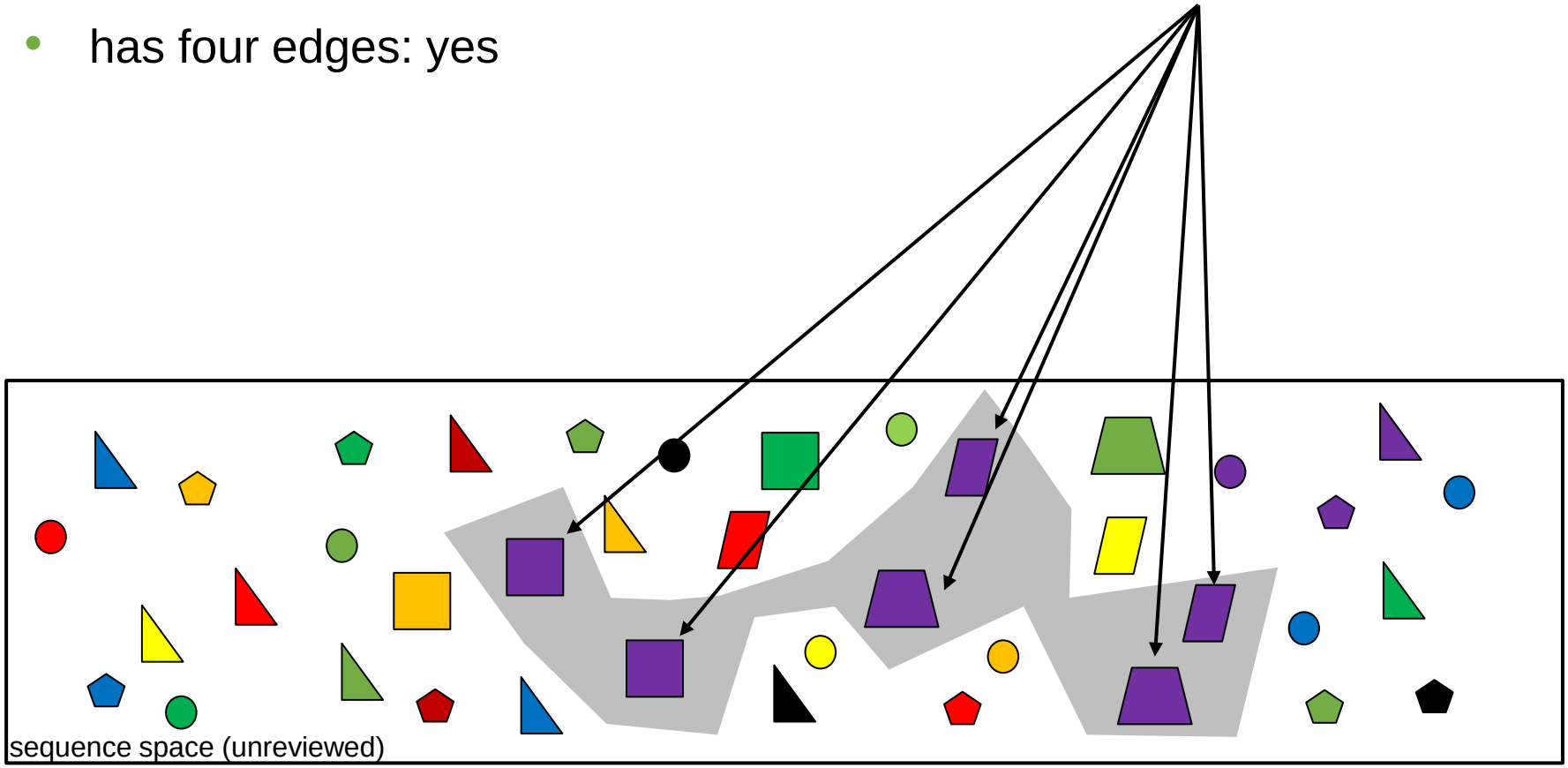
A fictitious rule

If a protein meets these conditions...

- color: purple
- has four edges: yes

... then these annotations are applied

- is a purple quadrilateral



A fictitious rule

If a protein meets these conditions...

- color: purple
- has four ed
- of same length
- a es are 90 degrees

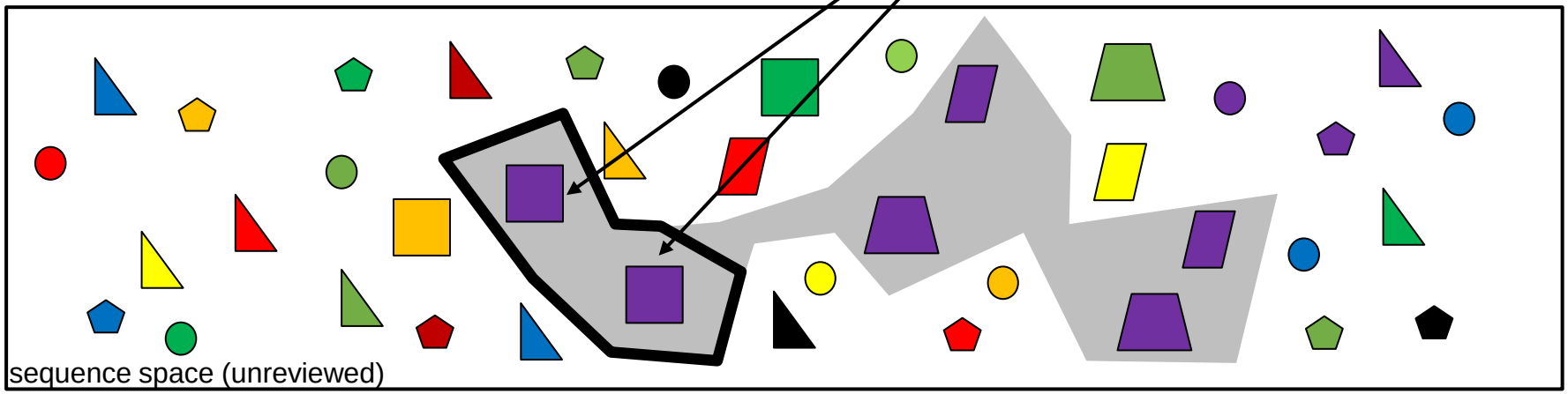
and:

... then these annotations are applied

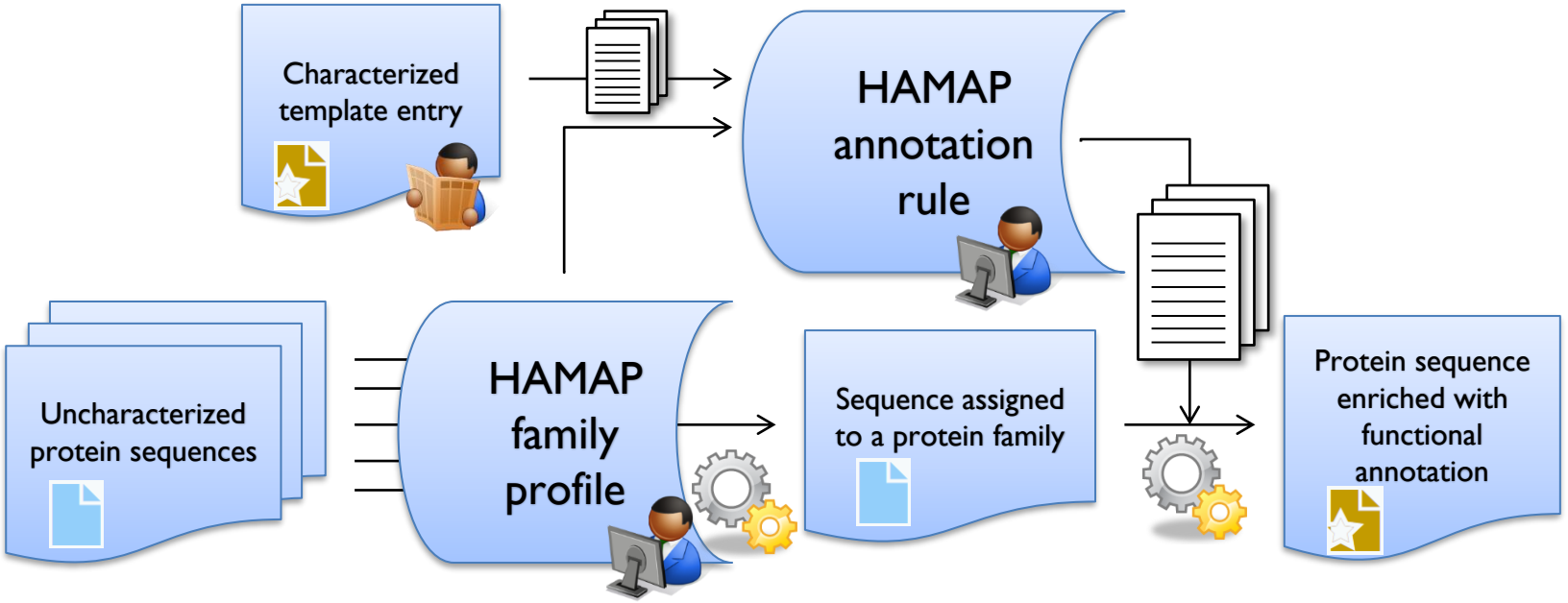
- is a purple quadri
-

CONDITIONS

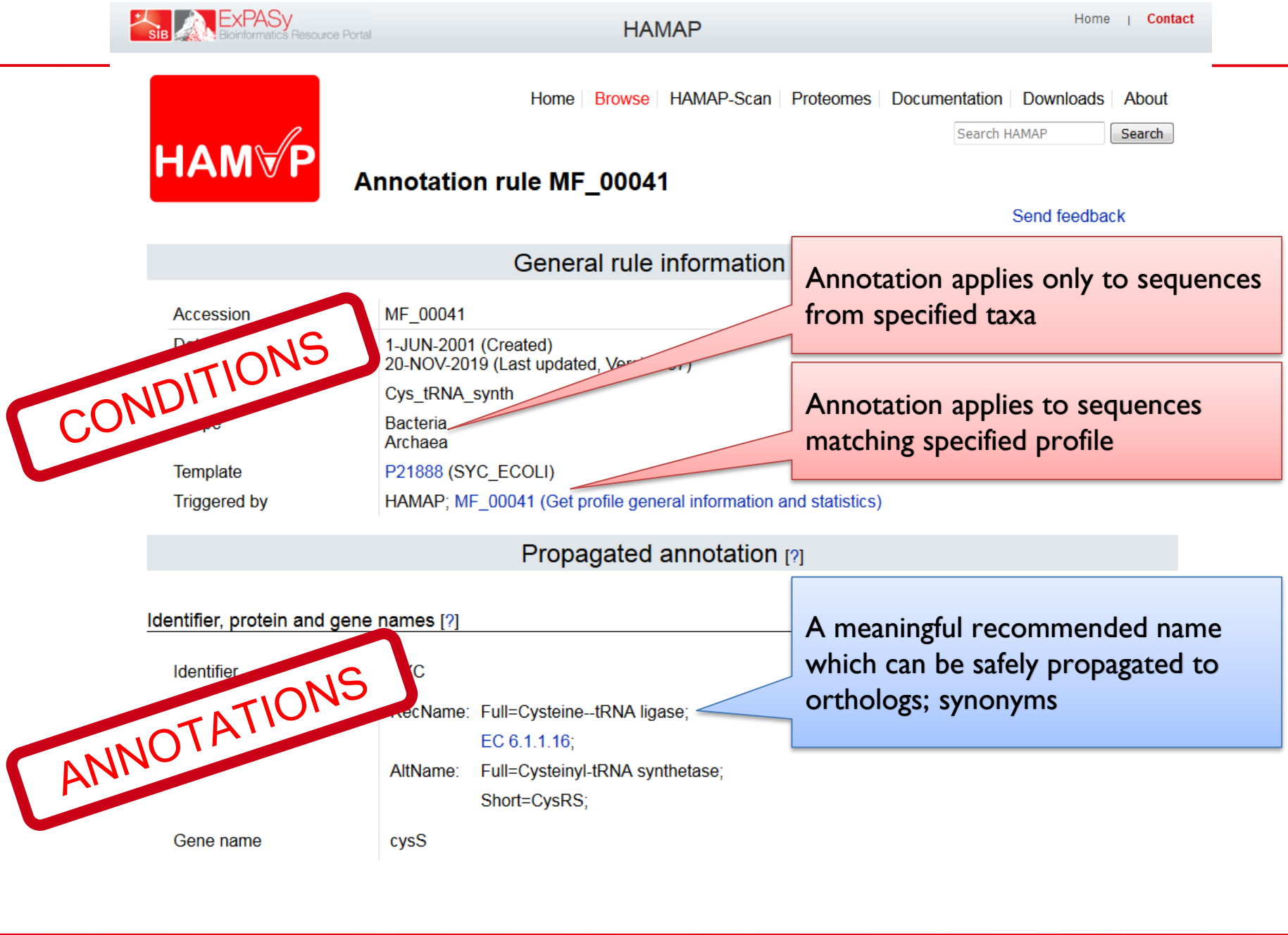
ANNOTATIONS



HAMAP annotation rules



- Rule creation starts with the manual curation of UniProtKB/Swiss-Prot records (=**template entries**).
- Protein names, gene names, functional annotation, GO terms, keywords and sequence features from the templates are selected and combined to build the rule containing the annotation to be propagated.
- Annotations may be subject to control statements that limit their propagation to only those sequences satisfying one or more conditions.


ExPASy Bioinformatics Resource Portal
HAMAP
Home | Contact
HAMAP
Home | Browse | HAMAP-Scan | Proteomes | Documentation | Downloads | About
Search HAMAP
Search
Annotation rule MF_00041
Send feedback
General rule information
Accession
MF_00041
1-JUN-2001 (Created)
20-NOV-2019 (Last updated,
Cys_tRNA_synth
Bacteria
Archaea
Template
P21888 (SYC_ECOLI)
Triggered by
HAMAP; MF_00041 (Get profile general information and statistics)
CONDITIONS
Annotation applies only to sequences from specified taxa
Annotation applies to sequences matching specified profile
Propagated annotation [?]
Identifier, protein and gene names [?]
Identifier
RecName: Full=Cysteine--tRNA ligase;
EC 6.1.1.16;
AltName: Full=Cysteinyl-tRNA synthetase;
Short=CysRS;
Gene name
cysS
ANNOTATIONS
A meaningful recommended name which can be safely propagated to orthologs; synonyms

Annotation rule MF_00041

Send feedback

Propagated annotation [?]

Comments [?]

Catalytic activity	RHEA:17773: ATP + L-cysteine + tRNA(Cys) = AMP + diphosphate + L-cysteinyl-tRNA(Cys) EC 6.1.1.16
case <FTGroup:1>	
Cofactor	Zn(2+) Note: Binds 1 zinc ion per subunit.
end case	
case <OC:Bacteria>	
Subunit	Monomer.
end case	
Subcellular location	Cytoplasm.
Similarity	Belongs to the class-I aminoacyl-tRNA synthetase family.

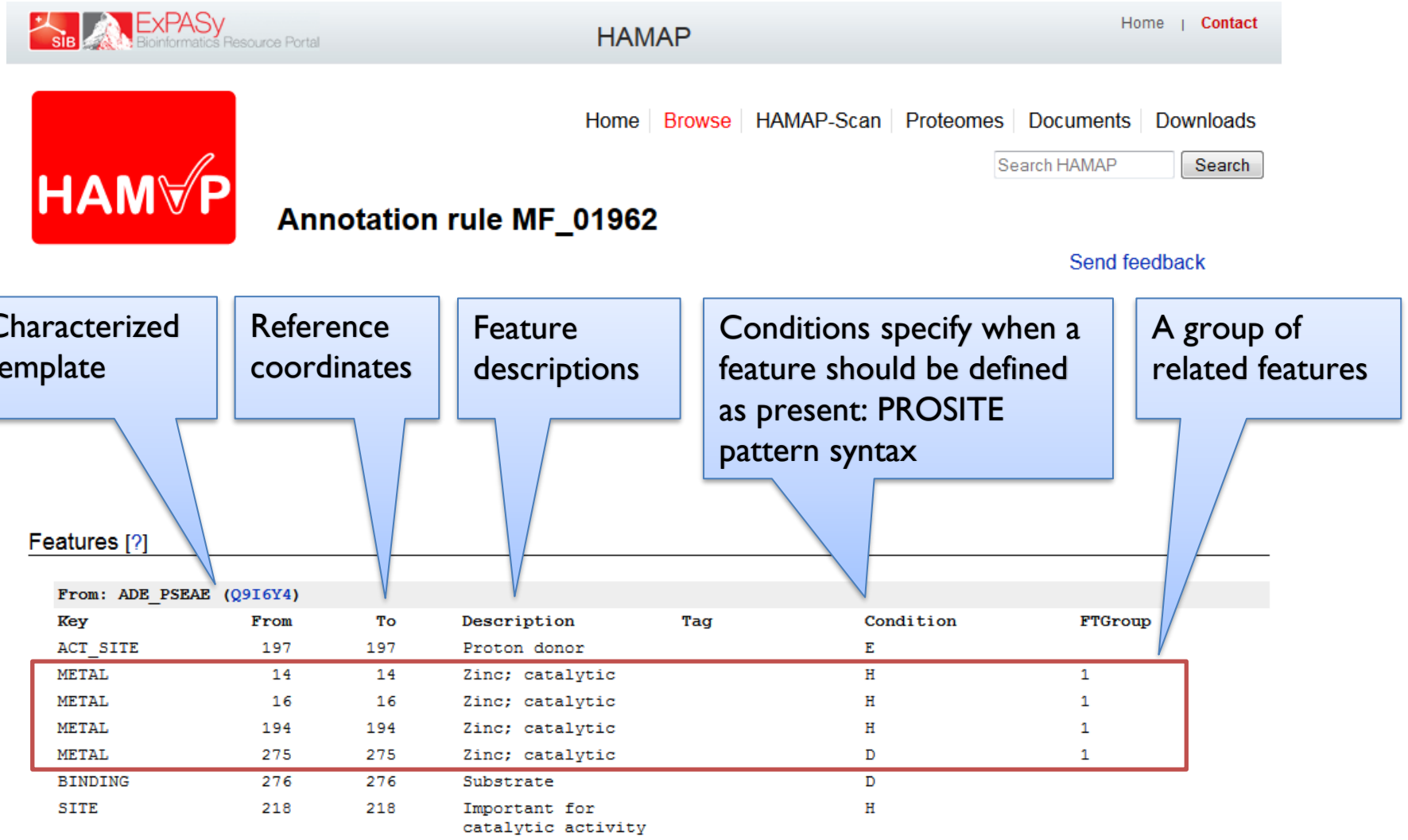
Keywords [?]

Cytoplasm
Aminoacyl-tRNA synthetase
Protein biosynthesis
Ligase
ATP-binding
Nucleotide-binding
case <FTGroup:1>
Metal-binding
Zinc
end case
case <FT:8>
Phosphoprotein
end case

Protein function, interactions, location, pathways, family membership and others – Swiss-Prot syntax

Summary in the form of keywords

Case statements restrict annotation to proteins satisfying these conditions


ExPASy Bioinformatics Resource Portal
HAMAP
Home | Contact
HAMAP
Home | Browse | HAMAP-Scan | Proteomes | Documents | Downloads
Search HAMAP
Search
Annotation rule MF_01962
Send feedback
Characterized template
Reference coordinates
Feature descriptions
Conditions specify when a feature should be defined as present: PROSITE pattern syntax
A group of related features
Features [?]
From: ADE_PSEAE (Q9I6Y4)
Key From To Description Tag Condition FTGroup
ACT_SITE 197 197 Proton donor E
METAL 14 14 Zinc; catalytic H 1
METAL 16 16 Zinc; catalytic H 1
METAL 194 194 Zinc; catalytic H 1
METAL 275 275 Zinc; catalytic D 1
BINDING 276 276 Substrate D
SITE 218 218 Important for catalytic activity H

Annotation transfer

- Sequence features are propagated by aligning target sequences to the profile and translating feature positions from the template to the target sequence.

```
FT   From: ADE_PSEAE (Q9I6Y4)
FT   METAL           14
FT                   /note="Zinc; catalytic"
FT   Group: 1; Condition: H
FT   METAL           16
FT                   /note="Zinc; catalytic"
FT   Group: 1; Condition: H
```

```
ID C4ZN08_THASP              Unreviewed;       343 AA.
AC   C4ZN08;
```

```
FT   METAL           16
FT                   /note="Zinc; catalytic"
FT                   /evidence="ECO:0000256|HAMAP-Rule:MF_01962"
FT   METAL           18
FT                   /note="Zinc; catalytic"
FT                   /evidence="ECO:0000256|HAMAP-Rule:MF_01962"
```

```
CLUSTAL W (1.83) multiple sequence alignment template=ADE_PSEAE profile_method=hmmbuild

C4ZN08_THASP    ---------MELEAYVRALPKAELHHHIEGTLEPEMMFALARRNGVALPWA-------SV
ADE_YEAST       ---------MVSVEFLQELPKCEHHHHLEGTLEPDLLFPLAKRNDIILPE----GFPKSV
ADE_CANGA       ---------MVPESFLLELPKCEHHHHLEGTLEPDLLFPLAKRNNIQLPD----HFPQTP
ADE_KLULA       --MAKFECTDEVTNFLTELPKCEHHHHLEGTLEPELLFQLVERNGVQLPG----TFPKTV
ADE_CANAL       --MAQYECSEHMENFLRELPKCEHHNHLEGTLEPSLLFKLAKRNNITLPE----TFPKTV
ADE_ASPFU       MC------QSPLHDFLHGLPKCEHHNHLEGCVTPELIFQLAEKNNIQLPNPATHPAYASV
ADE_ASPOR       MC------KSDLHDFLHGLPKCEHHNHLEGCLAPDLIFELAKRNNVSLPN---EPAYESI
ADE_EMENI       MCPPNTPYQSQWHAFLHSLPKCEHHNHLEGCLEPPLIFSMARKNNVSLPSPSSNPAYTSV
ADE_SCHPO       MS------NLPIYNFIRKLPKCEHHNHLEGCLSPDLVFRLAKKNGITLPS--DDAAYTTP
ADE_GIBZE       MC------KSRVHSFLQALPKVEQHHHIEGTLEPELLFTLAEKNGIELPN---DPVYESA
ADE_CAUCR       -MTDASFAPSASAEFVRGLPKAELHNHIEGSLEPELMFELAQRNGITLPFA-------SV
ADE_CAUCN       -MTDASFAPSASAEFVRGLPKAELHNHIEGSLEPELMFELAQRNGITLPFA-------SV
ADE_SPHAL       -MPDGFASHEERAAFIAGLPKAELHHHIEGSLEPELLFEFARRNRVAIPFA-------SI
ADE_RHORT       --------MAVDPAFLHALPKVELHHHIEGSLEPEMMVALAERNGLRLPYA-------SV
ADE_STRCO       --------MKRPYDALMPLPKAELHHHIEGTLEPELAFALAARNGVSLPYA-------DE
ADE_BURPP       -MTTTTVTPTPLAEKTALAPKAELHHHIEGSLEPELIFALAERNGVKLAYD-------SI
ADE_BURXL       -MTTTTVTSTPLAEKTVLAPKAELHHHIEGSLEPELIFALAERNGVKLAYD-------SI
ADE_CUPTR       -----MTIDAALAEQIRRTPKAELHNHIEGTLEPELIFRLAQRNQVALPYP-------SV
ADE_CUPNH       -----MTIDAALAEQIRRTPKAELHNHIEGTLEPELIFRLAQRNQVALPYP-------SV
ADE_CUPPJ       -----MTIDAALADKIRRTPKAELHNHIEGTLEPELIFRLAQRNHVNLPYP-------SV
ADE_RALME       -----MTIDAALADKIRRTPKAELHNHIEGTLEPERIFRLAQRNNVKLAYP-------DV
ADE_RALPJ       -----MPISSALAERIATSPKAELHHHIEGSLEPELMFALAERNGVKLPYA-------SV
ADE_RALSO       -----MPISPALAERIATSPKAELHHHIEGSLEPELMFALAERNGVKLPYA-------SV
ADE_GEOLS       -MNLTNIPRQALPELLCRMPKAELHHHIEGSLEPELIFALAERNRLQLAYP-------TI
ADE_GEOUR       -MNFDCIPREDLHGILCHMPKAELHHHIEGSLEPELIFELATRNRIQLPYP-------TI
ADE_RHOFD       -MTIKPVSQERLPELLRTIPKAELHHHIEGSLEPELMFALAQRNGVSIPYP-------DV
ADE_ZYMMO       --------MNNLIKFIAALPKAELHHHIEGSLEPELMFELAKRNKVTLPFP-------DV
ADE_ACIAD       ---------MNQSELIRALPKAELHNHIEGTFEPELMFEIAQRNHIDIPYK-------SV
ADE_PSYCK       -----------MIDLIKRLPKAELHHHIEGSLEPELMFRLAKKNQIEIPYK-------DI
ADE_VIBPA       -----------MNAFIQGLPKVELHHHIEGSLEPELMFKLAKRNGIDIPYS-------SP
ADE_PSEAE       -----------MYEWLNALPKAELHHHLEGTLEPELLFALAERNRIALPWN-------DV
```

Annotation rule MF_01962

Send feedback

Propagated annotation [?]

case <FTGroup:1>

Cofactor	Zn(2+) Note: Binds 1 zinc ion per subunit.

end case

Similarity	Belongs to the metallo-dependent hydrolases superfamily. Adenosine and AMP deaminases family. Adenine deaminase type 2 subfamily.

Keywords [?]

Hydrolase
Nucleotide metabolism
case <FTGroup:1>
Metal-binding
Zinc
end case

Gene Ontology [?]

GO:0000034; Molecular function: adenine deaminase activity.
GO:0006146; Biological process: adenine catabolic process.
GO:0043103; Biological process: hypoxanthine salvage.
case <FTGroup:1>
GO:0008270; Molecular function: zinc ion binding.
end case

Annotation applied if, and only if, ALL members of the specified group (here, FTGroup1, are present).

Annotation output

- Many checks are performed in order to prevent the propagation of wrong annotation:

```
DR   HAMAP; MF_01962; Adenine_deaminase; 1.
KW   Hydrolase; Metal-binding; Nucleotide metabolism; Zinc.
FT   CHAIN           1..343
FT                   /note="Adenine deaminase"
FT   ACT_SITE        199
FT                   /note="Proton donor"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   METAL           16
FT                   /note="Zinc; catalytic"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   METAL           18
FT                   /note="Zinc; catalytic"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   METAL           196
FT                   /note="Zinc; catalytic"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   METAL           277
FT                   /note="Zinc; catalytic"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   BINDING         278
FT                   /note="Substrate"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   SITE            220
FT                   /note="Important for catalytic activity"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
**
**   ##################    INTERNAL SECTION    ##################
**EV ECO:0000255; HAMAP-Rule:MF_01962; XXX; 02-MAR-2020.
**HA FAM; Method MF_01962; ADE; Trusted match; 85.781 (+32).
**HA FAM; Method MF_00540; ADD; Weak match; 28.113 (-21.7).
**HA SAM; Annotated by HAMAP 3.68.17; MF_01962.9; MF_01962; 02-MAR-2020 12:42:26.
SQ   SEQUENCE   343 AA;  37571 MW;  E2DA4A77193E5C4B CRC64;
     MELEAYVRAL PKAELHLHIE GTLEPEMMFA LARRNGVALP WASVEAVRAA YAFTDLQSFL
     DLYYAGAAVL VREQDFFELA FAYFERAHAD GVVHAELFFD PQTHTARGVA LETVLDGLER
     ACVEARARWG IGSRLILCFL RHLSEEEGFA TLQQALPHLS RIDGVGLDSS ERGHPPAKFA
     RLFARCRELG LHVVAHAGEE GPPAYIVDAL DLLKAERIDH GVRCTEDPAL VGRLVREQVP
     LTVCPLSNVK LCVFPDLARH NLGQLFAAGL KVTINSDDPA YFGGYVAKNY VDTARALGLG
     RAELRRIARN SLEASFVSAA ERAPWLARLD ALGEDCEGEG GAA
//
```

Annotation output validation

```
ID   F1VZU6_9BURK               Unreviewed;       343 AA.
AC   F1VZU6;
..
DR   HAMAP; MF_01962; Adenine_deaminase; 1.
KW   Hydrolase; Nucleotide metabolism.
FT   CHAIN           1..302
FT                   /note="Adenine deaminase"
FT   ACT_SITE        163
FT                   /note="Proton donor"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   BINDING         242
FT                   /note="Substrate"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   SITE            184
FT                   /note="Important for catalytic activity"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
**
**   ##################    INTERNAL SECTION    ##################
**EV ECO:0000255; HAMAP-Rule:MF_01962; XXX; 02-MAR-2020.
**HA SAM; Annotated by HAMAP 3.68.17; MF_01962.9; MF_01962; 02-MAR-2020 11:28:48.
**HW HAMAP: WARNING(0) for [MF_01962]: Conditions for mandatory group [1] not met (in one or more instances of the
catalytic|__FT_5,__FTGRP_1,__HAS_PATTERN_COND,D)))
**HW HAMAP: WARNING(0) for [MF_01962]: Protein size (302) smaller than specified minimum (315)
     MALAFESVDA LRAAYAFTDL QSFLDIYYAG ASVLLTEQDF YDMTRAYLLR AQQDQVRHTE
     PA
//
```

**HW HAMAP: WARNING(0) for [MF_01962]: Conditions for mandatory group [1] not met (in one or more instances of the trigger) (group condition: …

**HW HAMAP: WARNING(0) for [MF_01962]: Protein size (302) smaller than specified minimum (315)

**HW HAMAP: WARNING(0) for [MF_01962]: Trigger [MF_01962] (with MaxNterGaps: 17): N-ter too short by 29 AA

Annotation output validation

```
CLUSTAL W (1.83) multiple sequence alignment template=ADE_PSEAE

ADE_YEAST         ---------MVSVEFLQELPKCEHHLHLEGTLEPDLLFPLAKRNDIILPE----GFPKSV
ADE_CANGA         ---------MVPESFLLELPKCEHHLHLEGTLEPDLLFPLAKRNNIQLPD----HFPQTP
ADE_KLULA         --MAKFECTDEVTNFLTELPKCEHHLHLEGTLEPELLFQLVERNGVQLPG----TFPKTV
ADE_CANAL         --MAQYECSEHMENFLRELPKCEHHVHLEGTLEPSLLFKLAKRNNITLPE----TFPKTV
ADE_ASPFU         MC------QSPLHDFLHGLPKCEHHVHLEGCVTPELIFQLAEKNNIQLPNPATHPAYASV
ADE_ASPOR         MC------KSDLHDFLHGLPKCEHHVHLEGCLAPDLIFELAKRNNVSLPN---EPAYESI
ADE_EMENI         MCPPNTPYQSQWHAFLHSLPKCEHHVHLEGCLEPPLIFSMARKNNVSLPSPSSNPAYTSV
ADE_SCHPO         MS------NLPIYNFIRKLPKCEHHVHLEGCLSPDLVFRLAKKNGITLPS--DDAAYTTP
ADE_GIBZE         MC------KSRVHSFLQALPKVEQHLHIEGTLEPELLFTLAEKNGIELPN---DPVYESA
ADE_CUPTR         -----MTIDAALAEQIRRTPKAELHVHIEGTLEPELIFRLAQRNQVALPYP-------SV
ADE_CUPNH         -----MTIDAALAEQIRRTPKAELHVHIEGTLEPELIFRLAQRNQVALPYP-------SV
ADE_CUPPJ         -----MTIDAALADKIRRTPKAELHVHIEGTLEPELIFRLAQRNHVNLPYP-------SV
ADE_RALME         -----MTIDAALADKIRRTPKAELHVHIEGTLEPERIFRLAQRNNVKLAYP-------DV
ADE_RALPJ         -----MPISSALAERIATSPKAELHIHIEGSLEPELMFALAERNGVKLPYA-------SV
ADE_RALSO         -----MPISPALAERIATSPKAELHIHIEGSLEPELMFALAERNGVKLPYA-------SV
F1VZU6_9BURK      ---------------------------------------------MALAFE-------SV
ADE_BURPP         -MTTTTVTPTPLAEKTALAPKAELHIHIEGSLEPELIFALAERNGVKLAYD-------SI
ADE_BURXL         -MTTTTVTSTPLAEKTVLAPKAELHIHIEGSLEPELIFALAERNGVKLAYD-------SI
ADE_GEOLS         -MNLTNIPRQALPELLCRMPKAELHIHIEGSLEPELIFALAERNRLQLAYP-------TI
ADE_GEOUR         -MNFDCIPREDLHGILCHMPKAELHIHIEGSLEPELIFELATRNRIQLPYP-------TI
ADE_RHOFD         -MTIKPVSQERLPELLRTIPKAELHIHIEGSLEPELMFALAQRNGVSIPYP-------DV
ADE_ZYMMO         --------MNNLIKFIAALPKAELHLHIEGSLEPELMFELAKRNKVTLPFP-------DV
ADE_CAUCR         -MTDASFAPSASAEFVRGLPKAELHMHIEGSLEPELMFELAQRNGITLPFA-------SV
ADE_CAUCN         -MTDASFAPSASAEFVRGLPKAELHMHIEGSLEPELMFELAQRNGITLPFA-------SV
```

Annotation output validation

```
ID   G9PG85_9ACTO            Unreviewed;       332 AA.
AC   G9PG85;
..
DR   HAMAP; MF_01962; Adenine_deaminase; 1.
KW   Hydrolase; Nucleotide metabolism.
FT   CHAIN           1..332
FT                   /note="Adenine deaminase"
FT   BINDING         278
FT                   /note="Substrate"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
FT   SITE            222
FT                   /note="Important for catalytic activity"
FT                   /evidence="ECO:0000255|HAMAP-Rule:MF_01962"
**
**   ##################    INTERNAL SECTION    ##################
**EV ECO:0000255; HAMAP-Rule:MF_01962; XXX; 02-MAR-2020.
**HA SAM; Annotated by HAMAP 3.68.17; MF_01962.9; MF_01962; 02-MAR-2020 11:36:03.
**HW HAMAP: WARNING(0) for [MF_01962]: Conditions for mandatory group [1] not met (in one or more instances of the
```

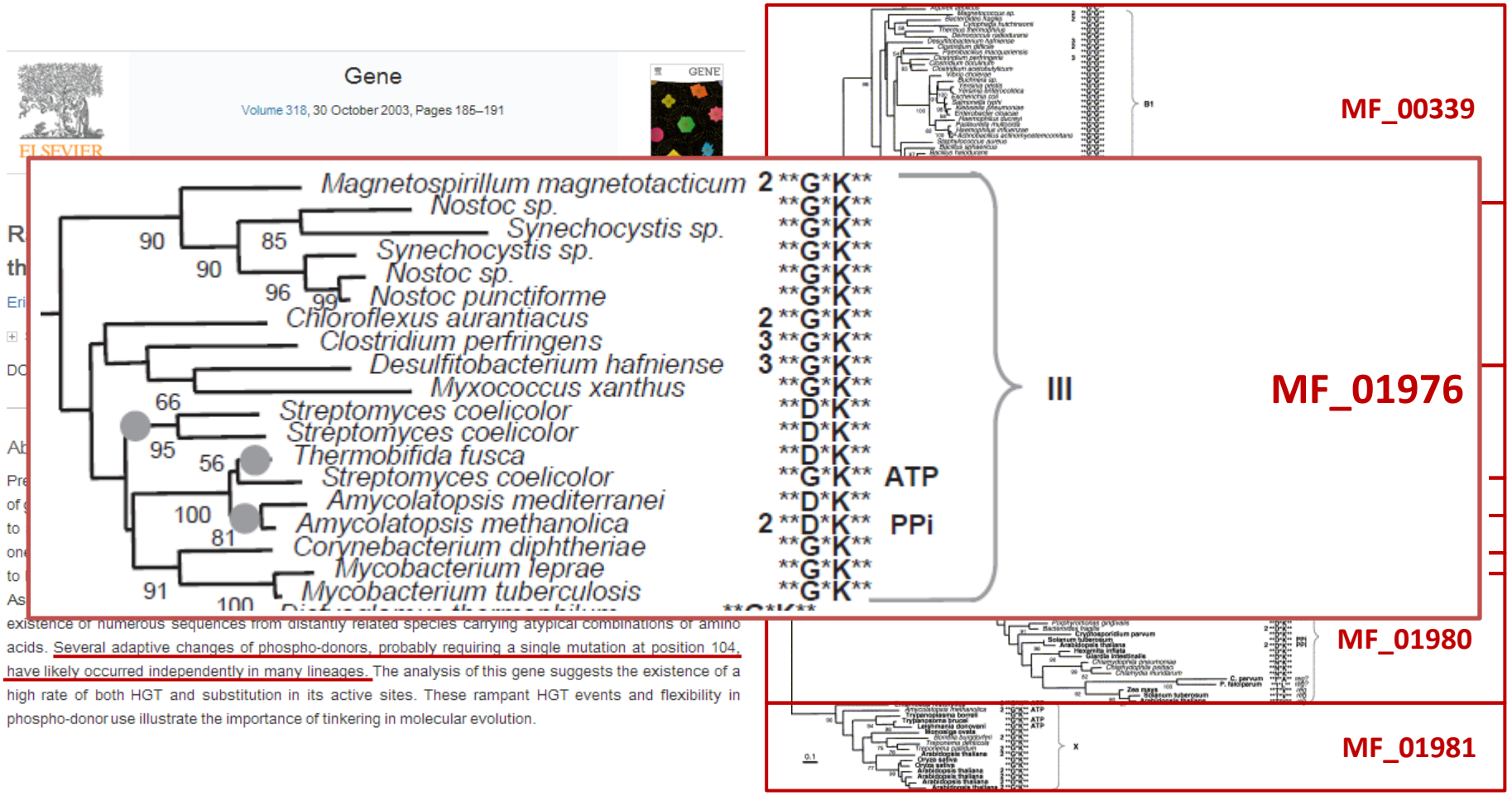
**HW HAMAP: WARNING(0) for [MF_01962]: Mandatory Positional (pattern [E]) not matched (1 or more times for Positional|ACTIVE_SITE|{197//}|{197//}|Proton donor|__FT_1,__HAS_PATTERN_COND)

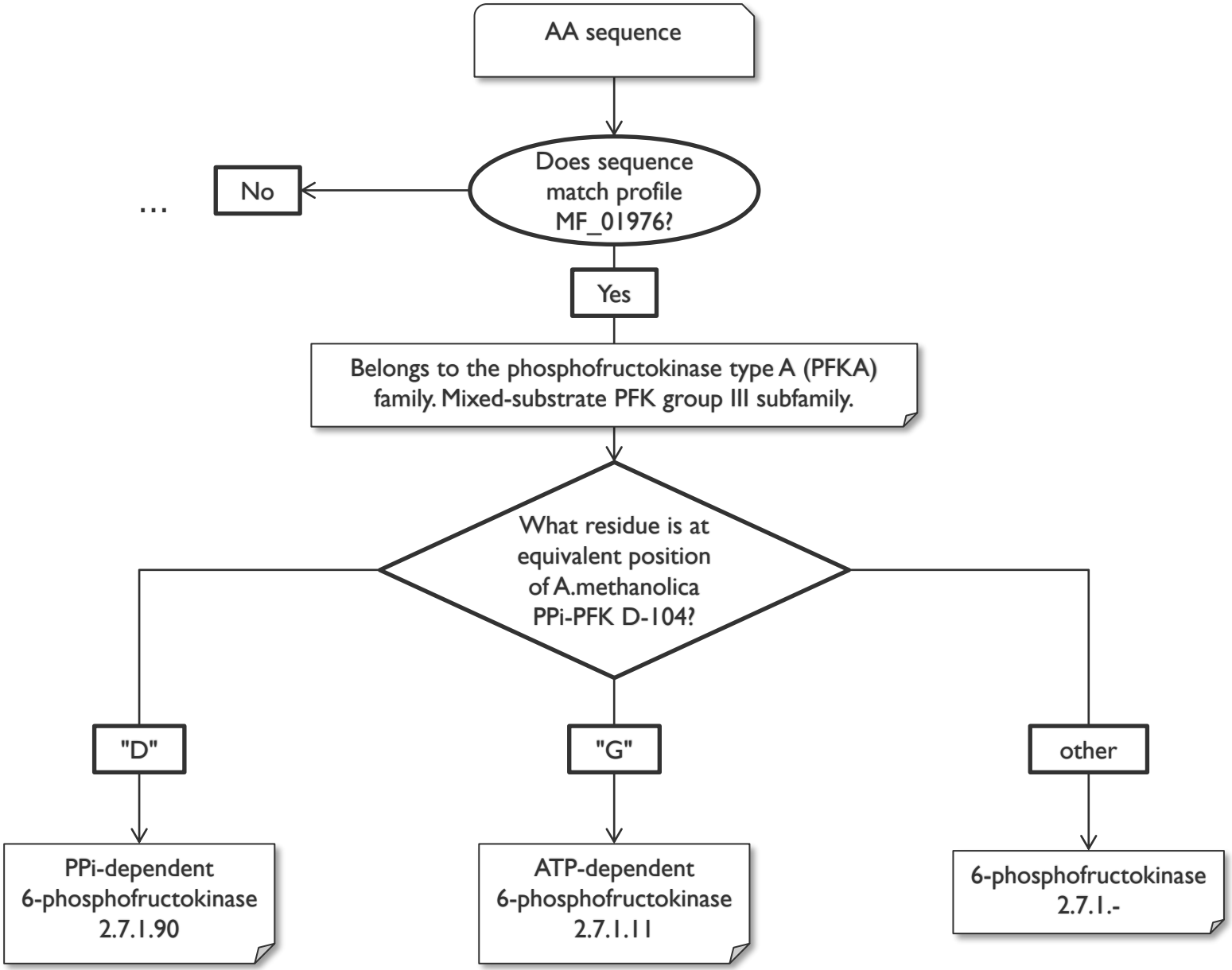
```
SQ   SEQUENCE   332 AA;  37491 MW;  701E1C335EA545DB CRC64;
     MAHEALVQAL PKVELHVHIE GTLEPELKFK LAQRNGIKLP FSSEQEVKDS YTFNDLASFL
     DAYYDGMNVL LASEDFYDLA MAYYRKAASQ GLRYAEIFFD PQAHTCRGVS FHTVISGLRR
     AQLEAEQTLG VFSQFIMCFL RDFQPEYAMA TLLESLPYKQ WIVGVGLDSD ETDHEPAKFA
     AVFARARREG YQLTMHCDVD IKDSIEHIRQ VIEDIQVQRV DHGTNVVENP ALVKYLVEHR
     IGLTSCPISN LWISESDKVE LVKQLVADGV LVTINSDDPA YFGGYIGDNF QRVADHDGVD
     EDFLRALVAN AIEISWAPLA VKQRLREELA AV
//
```

Annotation output validation

```
ADE_RALME          ---RDRFVGVGLDSSERGNPPEKFARVFARAKE--LGLHLVAHAGEE--GPPQYVTDALD
ADE_RALPJ          --PANRIIGVGLDSSERGNPPEKFARVFARCKE--LGLRLVAHAGEE--GPAQYVIDALD
ADE_RALSO          --PANRIIGVGLDSSERGNPPEKFARVFARCKA--LGLRLVAHAGEE--GPAQYVIDALD
ADE_GEOLS          ---RDKFIGVGLDSGEKGNPPEKFSRVFARCRE--LGLRLVAHAGEE--GTAEYIWHALD
ADE_GEOUR          ---RDKFIGVGLDSSERGNPPEKFTRVFARCRE--LGLRLVAHAGEE--GSAEYISHSLD
ADE_RHOFD          ---LDKLVGVGLASSEMGHPPEKFARVFARARE--LGLRLVAHAGEE--GPPAYIWSALD
ADE_ZYMMO          ---LDKIAGVGLDSSEVGNPPSKFRHVFAEARQ--KGLKLVAHAGEE--GDASYIKEALD
ADE_MARMS          ---LKWIDGIGLDSSEVGHPPEKFLRVFEACKN--LGLKVTAHAGEE--GPPDYVWQAIE
ADE_RHOPB          ---KDQIIAIGMGGAELGNPPAKFARFFKAARD--RGFRTTVHAGEE--GPAAYVREALE
ADE_STRCO          ---LDRITGVGLDSAEVGHPPVKFREVYEAAAA--LGLRRVAHAGEE--GPPAYVVEALD
G9PG85_9ACTO       ---KQWIVGVGLDSDETDHEPAKFAAVFARARR--EGYQLTMHCDVDIKDSIEHIRQVIE
ADE_RHIEC          ---NPLITGFNLAGEERMGRVADYIRAFDIARD--AGLGLTIHAGEV--CGAFSVADALD
ADE_RHIE6          ---NPLITGFNLAGEERMGRVADYSRAFDIARD--AGLGLTIHAGEV--CGAFSVADALD
ADE_RHILW          ---NPLITGFNLAGEERMGRVADYARAFDIARD--AGLGLTIHAGEV--CGAFSVADALD
ADE_RHIL3          ---NPLITGFNLAGEERMGRVADYARAFDIARD--AGLGLTIHAGEV--CGAFSVADALD
ADE_AGRRK          ---NPLISGFNMAGEERMGRVADYARAFDIARE--AGLGITIHAGEV--CGAFSVADAVE
ADE_AGRT5          ---HPLVTGFNMAGEERMGRVADYARAFDIARD--AGLGLTIHAGEV--CGPESVADALD
ADE_RHIME          ---HPLVTGFNLAGEERMHSVAEFSRAFDIVRD--AGLGLTIHAGEL--SGAFSVRDALD
ADE_RHISN          ---HPLVTGFNLAGEERMHSVAEFARAFDIVRD--AGLGLTIHAGEL--SGAFSVRDALD
ADE_SINMW          ---HPLVTGFNLAGEERMHSVAEFSRAFDIVRD--AGLGLTIHAGEL--SGAFSVRDALD
ADE_AGRVS          ---HPLITGFNMAGEERMNRVADYAPAFDIARE--AGYGITIHAGEL--CGAFSVRDALD
ADE_MESSB          ---HPLVTGFGMAGDERAGHPRDFAYAFDIARE--AGLGISIHAGEF--GGAESVEAALD
ADE_RHILO          ---NPLVTGFGVAGDERVGEMEDYVRAFEIARE--AGLGITIHAGEL--TGWETVQAALD
                        . ...  . *      :               *..        :   :.
```

Controlling annotations using cases and conditions




AA sequence
Does sequence match profile MF_01976?
No
...
Yes
Belongs to the phosphofructokinase type A (PFKA) family. Mixed-substrate PFK group III subfamily.
What residue is at equivalent position of A.methanolica PPi-PFK D-104?
"D"
"G"
other
PPi-dependent 6-phosphofructokinase 2.7.1.90
ATP-dependent 6-phosphofructokinase 2.7.1.11
6-phosphofructokinase 2.7.1.-

ExPASy Bioinformatics Resource Portal

HAMAP

Home | Contact

Home | Browse | HAMAP-Scan | Proteomes | Documents | Downloads

Search HAMAP | Search

Annotation rule MF_01976

Send feedback

Features [?]

From: PFP_AMYME (Q59126)

Key	From	To	Description	Tag	Condition	FTGroup
SITE (Optional)	104	104	Important for catalytic activity and substrate specificity; stabilizes the transition state when the phosphoryl donor is PPi; prevents ATP from binding by mimicking the alpha-phosphate group of ATP	PPI	D	
SITE (Optional)	104	104	Important for substrate specificity; cannot use PPi as phosphoryl donor	ATP	G	
REGION	125	127	Substrate binding		T-x-D	
REGION	169	171	Substrate binding		M-G-[RH]	
REGION	271	274	Substrate binding		[HY]-x(2)-R	
ACT_SITE	127	127	Proton acceptor		D	

"D"

other

"G"

"D" → case <FTTag:PPI>

Identifier and protein and gene names [?]	
Identifier	PFP
Protein name	RecName: Full=Pyrophosphate--fructose 6-phosphate 1-phosphotransferase; EC=2.7.1.90; AltName: Full=6-phosphofructokinase, pyrophosphate dependent; AltName: Full=PPi-dependent phosphofructokinase; Short=PPi-PFK; AltName: Full=Pyrophosphate-dependent 6-phosphofructose-1-kinase;
Gene name	pfp

"G" → else case <FTTag:ATP>

Identifier	PFKA
Protein name	RecName: Full=ATP-dependent 6-phosphofructokinase; Short=ATP-PFK; Short=Phosphofructokinase; EC=2.7.1.11; AltName: Full=Phosphohexokinase;
Gene name	pfkA

other → else

Identifier	PFP
Protein name	RecName: Full=6-phosphofructokinase; EC=2.7.1.-;

end case

ExPASy Bioinformatics Resource Portal

HAMAP

Home | Contact

HAMAP

Home | Browse | HAMAP-Scan | Proteomes | Documents | Downloads

Search HAMAP | Search

Annotation rule MF_03125

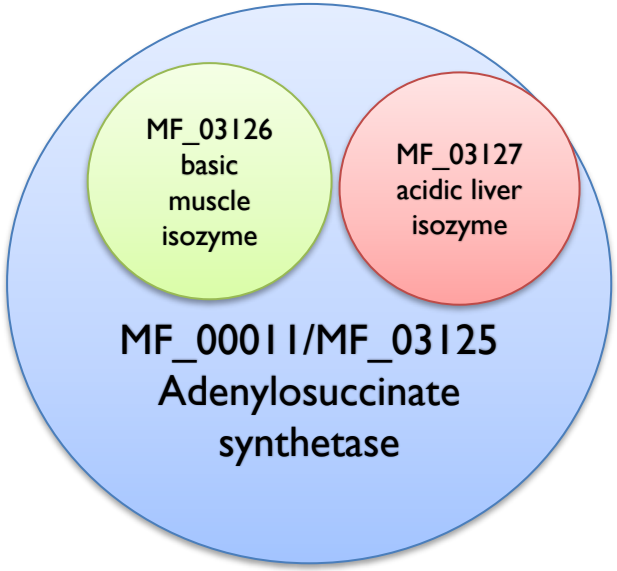
Send feedback

Additional information [?]

Size range	411-710 amino acids
Related rules	MF_03126 (PURA1 **supersedes the current rule**); MF_03127 (PURA2 **supersedes the current rule**)
Fusion	None

View rule in raw text format (no links)

We can indicate related HAMAP rules and specify their precedence



Family

Adenylosuccinate synthetase (IPR001114)

Short name: Adenylosuccinate_synthetase

Overlapping homologous superfamilies

P-loop containing nucleoside triphosphate hydrolase (IPR027417)

Family relationships

- **Adenylosuccinate synthetase (IPR001114)**
  - **Adenylosuccinate synthetase isozyme 1, chordates (IPR027509)**
  - **Adenylosuccinate synthetase isozyme 2, chordates (IPR027529)**

Description

Adenylosuccinate synthetase (EC:6.3.4.4) plays an important role in purine biosynthesis, by catalysing the GTP-dependent conversion of IMP and aspartic acid to AMP. IMP and L-aspartate are conjugated in a two-step reaction accompanied by the hydrolysis of GTP to GDP in the presence of Mg2+. In the first step, the r-phosphate group of GTP is transferred to the 6-oxygen atom of IMP. An aspartate then displaces this 6-phosphate group to form the product adenylosuccinate. Adenylosuccinate synthetase has been characterised from

Add your annotation

Contributing signatures

Signatures from InterPro member databases are used to construct an entry.

TIGRFAMs
TIGR00184 (purA)

Pfam
PF00709
(Adenylsucc_synt)

CDD
cd03108 (AdSS)

PANTHER
PTHR11846
(PTHR11846)

SMART
SM00788
(Adenylsucc_synt)

HAMAP
MF_00011
(Adenylosucc_synth)

Family

Adenylosuccinate synthetase isozyme 1, chordates (IPR027509)

Short name: AdSS_1_vert

Overlapping homologous superfamilies

H P-loop containing nucleoside triphosphate hydrolase (IPR027417)

Family relationships

F Adenylosuccinate synthetase (IPR001114)

F **Adenylosuccinate synthetase isozyme 1, chordates (IPR027509)**

Description

This entry represents the adenylosuccinate synthetase isozyme 1 (AdSS1) (EC:6.3.4.4) from vertebrates. It is a component of the purine nucleotide cycle (PNC), which interconverts IMP and AMP to regulate the nucleotide levels in various tissues, and which contributes to glycolysis and ammoniagenesis. It catalyses the first committed step in the

Add your annotation

Contributing signatures

Signatures from InterPro member databases are used to construct an entry.

HAMAP

MF_03126 (Adenylosucc_synth_vert_basic)

Types of sequence similarity

Family

Groups of proteins that are conserved along the whole sequence, sharing a common evolutionary origin, as reflected in their related functions.

Domains

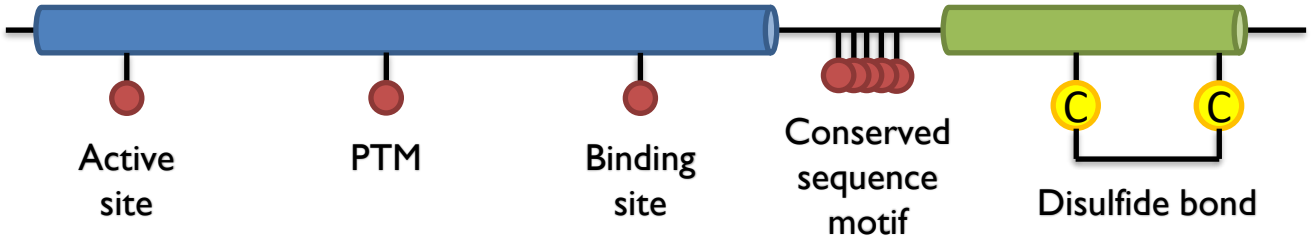
Specific combination of secondary structures organized into a characteristic three dimensional structure or fold, that may exist in a variety of biological contexts.

Repeats

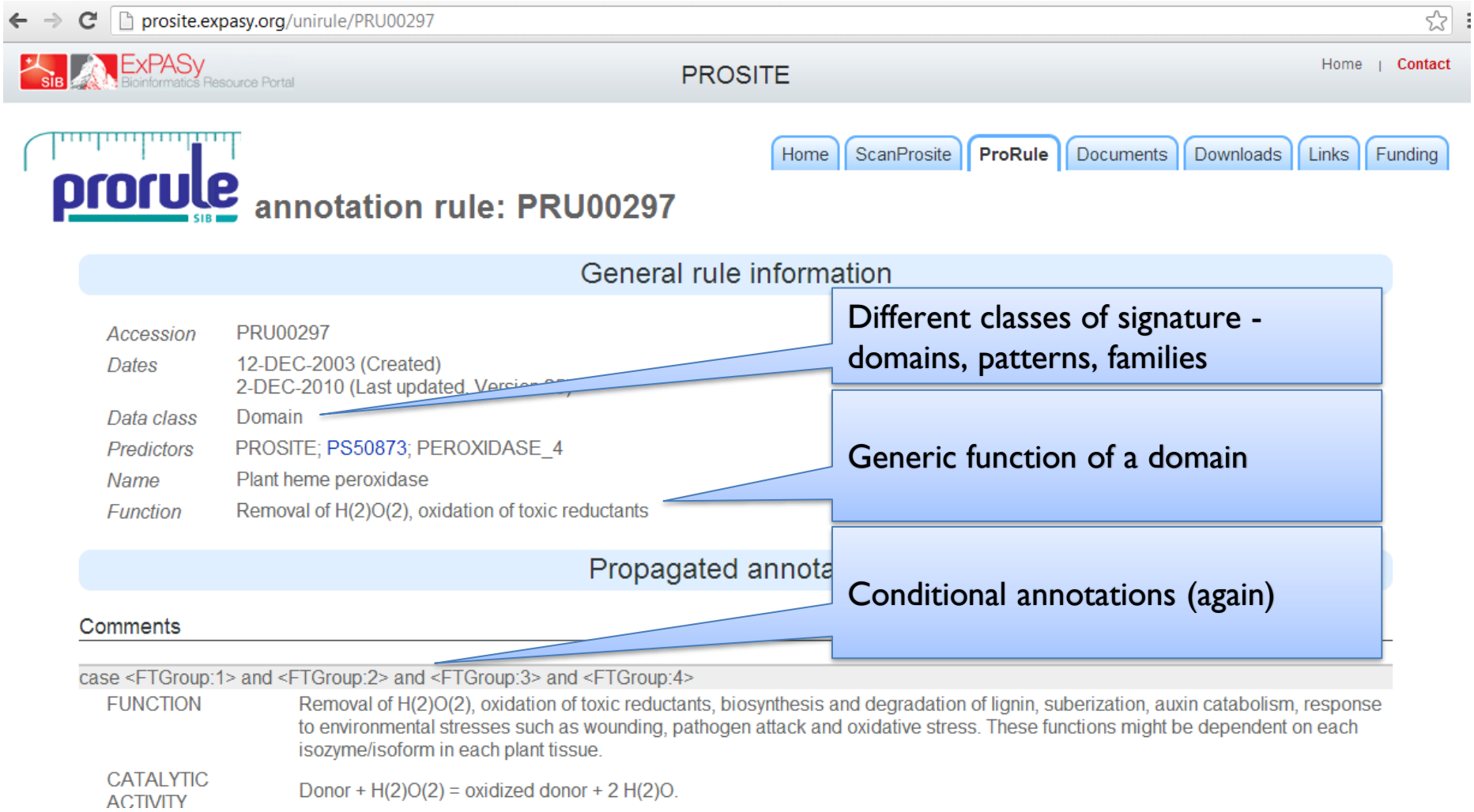
Structural units typically repeated within a protein that assemble into a specific fold. Assemblies of repeats might also be thought of as domains.

Sites & Motifs

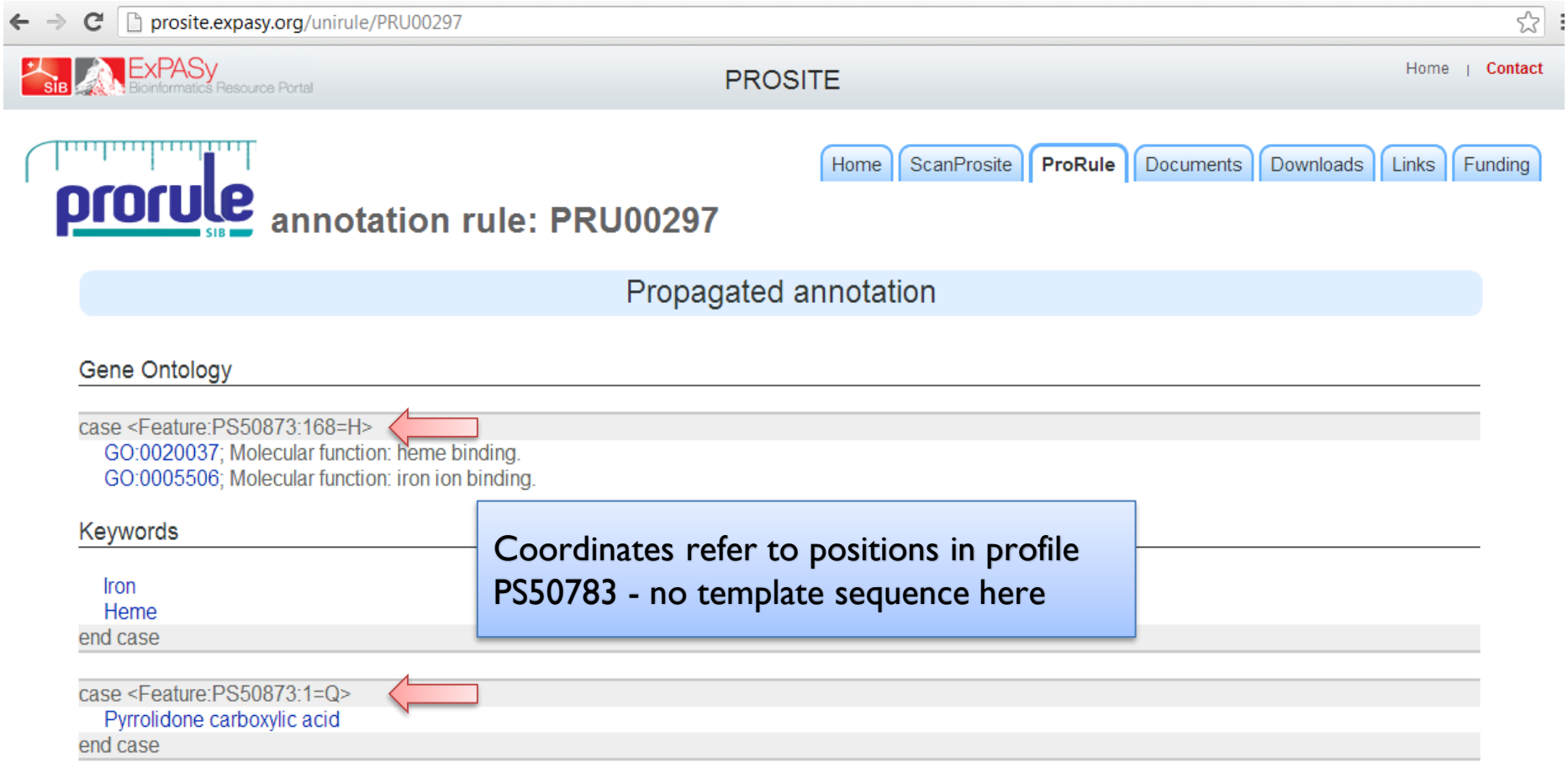
Region of domains containing conserved active-site or binding residues or short conserved regions present outside domains that may adopt folded conformations only in association with their binding ligands



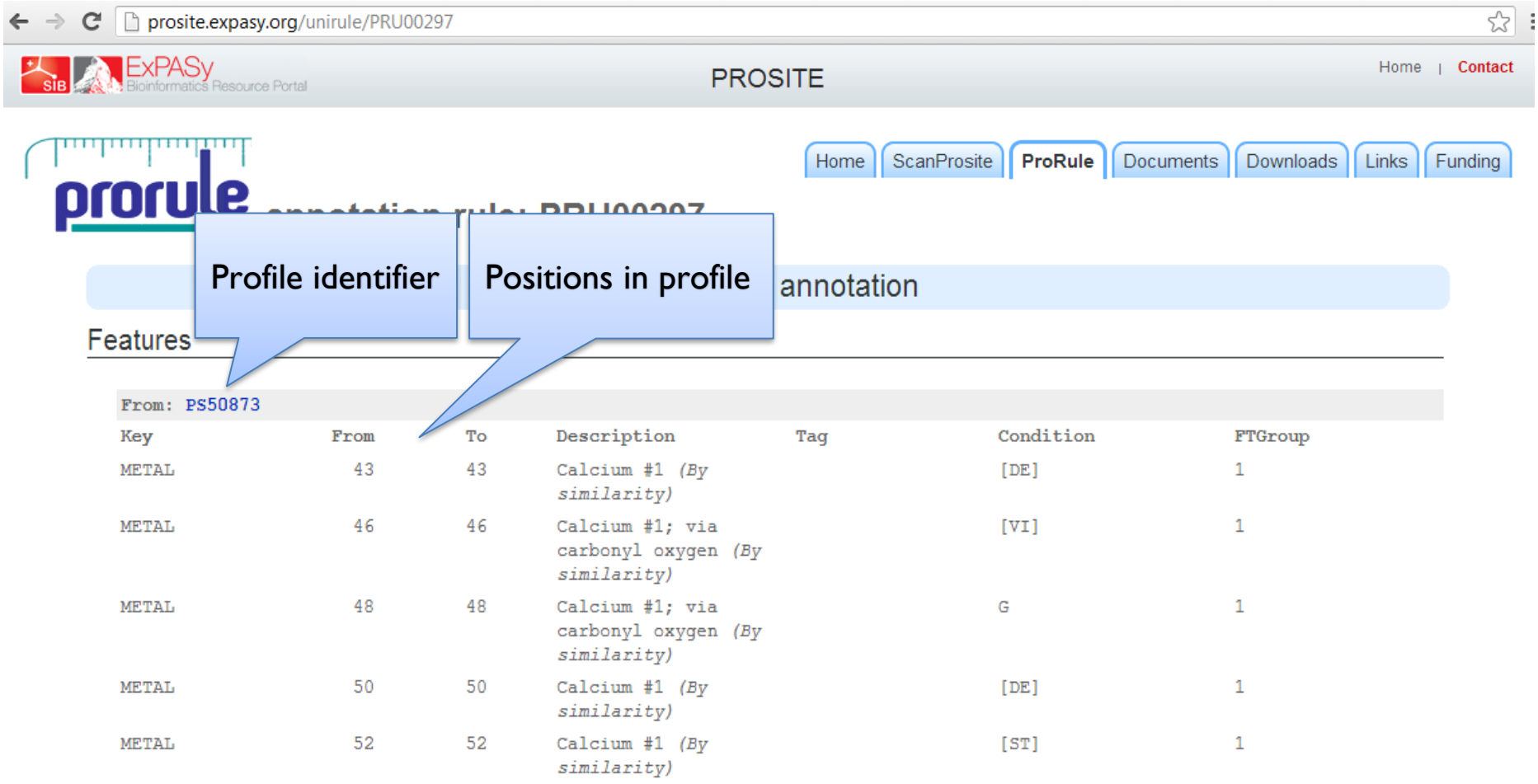
ProRules

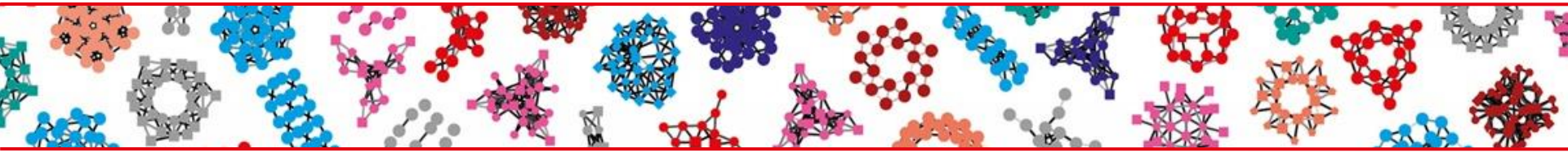

prosite.expasy.org/unirule/PRU00297
ExPASy Bioinformatics Resource Portal
PROSITE
Home | Contact
prorule SIB
Home
ScanProsite
ProRule
Documents
Downloads
Links
Funding
annotation rule: PRU00297
General rule information
Accession PRU00297
Dates 12-DEC-2003 (Created)
2-DEC-2010 (Last updated, Version
Data class Domain
Predictors PROSITE; PS50873; PEROXIDASE_4
Name Plant heme peroxidase
Function Removal of H(2)O(2), oxidation of toxic reductants
Propagated annota
Comments
case <FTGroup:1> and <FTGroup:2> and <FTGroup:3> and <FTGroup:4>
FUNCTION Removal of H(2)O(2), oxidation of toxic reductants, biosynthesis and degradation of lignin, suberization, auxin catabolism, response to environmental stresses such as wounding, pathogen attack and oxidative stress. These functions might be dependent on each isozyme/isoform in each plant tissue.
CATALYTIC ACTIVITY Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
Different classes of signature - domains, patterns, families
Generic function of a domain
Conditional annotations (again)

ProRules


prosite.expasy.org/unirule/PRU00297
ExPASy Bioinformatics Resource Portal
PROSITE
Home | Contact
prorule
Home
ScanProsite
ProRule
Documents
Downloads
Links
Funding
annotation rule: PRU00297
Propagated annotation
Gene Ontology
case <Feature:PS50873:168=H>
GO:0020037; Molecular function: heme binding.
GO:0005506; Molecular function: iron ion binding.
Keywords
Iron
Heme
end case
case <Feature:PS50873:1=Q>
Pyrrolidone carboxylic acid
end case
Coordinates refer to positions in profile PS50783 - no template sequence here

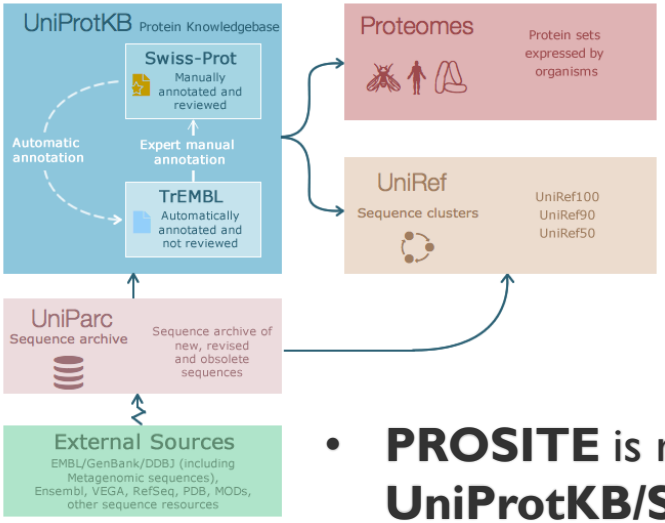
ProRules


prosite.expasy.org/unirule/PRU00297
ExPASy Bioinformatics Resource Portal
PROSITE
Home | Contact
prorule
Home
ScanProsite
ProRule
Documents
Downloads
Links
Funding
Profile identifier
Positions in profile
annotation
Features
From: PS50873
Key From To Description Tag Condition FTGroup
METAL 43 43 Calcium #1 (By similarity) [DE] 1
METAL 46 46 Calcium #1; via carbonyl oxygen (By similarity) [VI] 1
METAL 48 48 Calcium #1; via carbonyl oxygen (By similarity) G 1
METAL 50 50 Calcium #1 (By similarity) [DE] 1
METAL 52 52 Calcium #1 (By similarity) [ST] 1



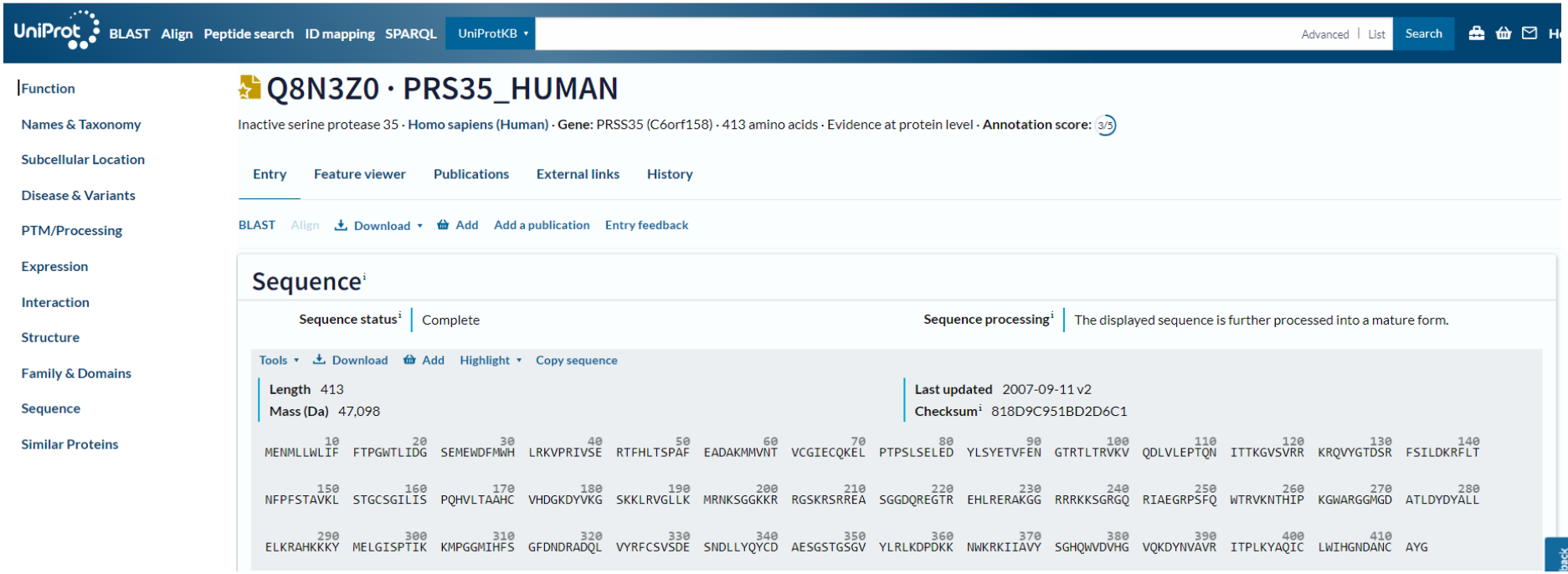
- Why do we need predictive annotation tools?
- Protein signatures for homology detection – A short primer
- Annotation rules for functional annotation
- **HAMAP and PROSITE - automatic annotation in UniProtKB**
- HAMAP and PROSITE - services for external users
- Practical exercises (afternoon)

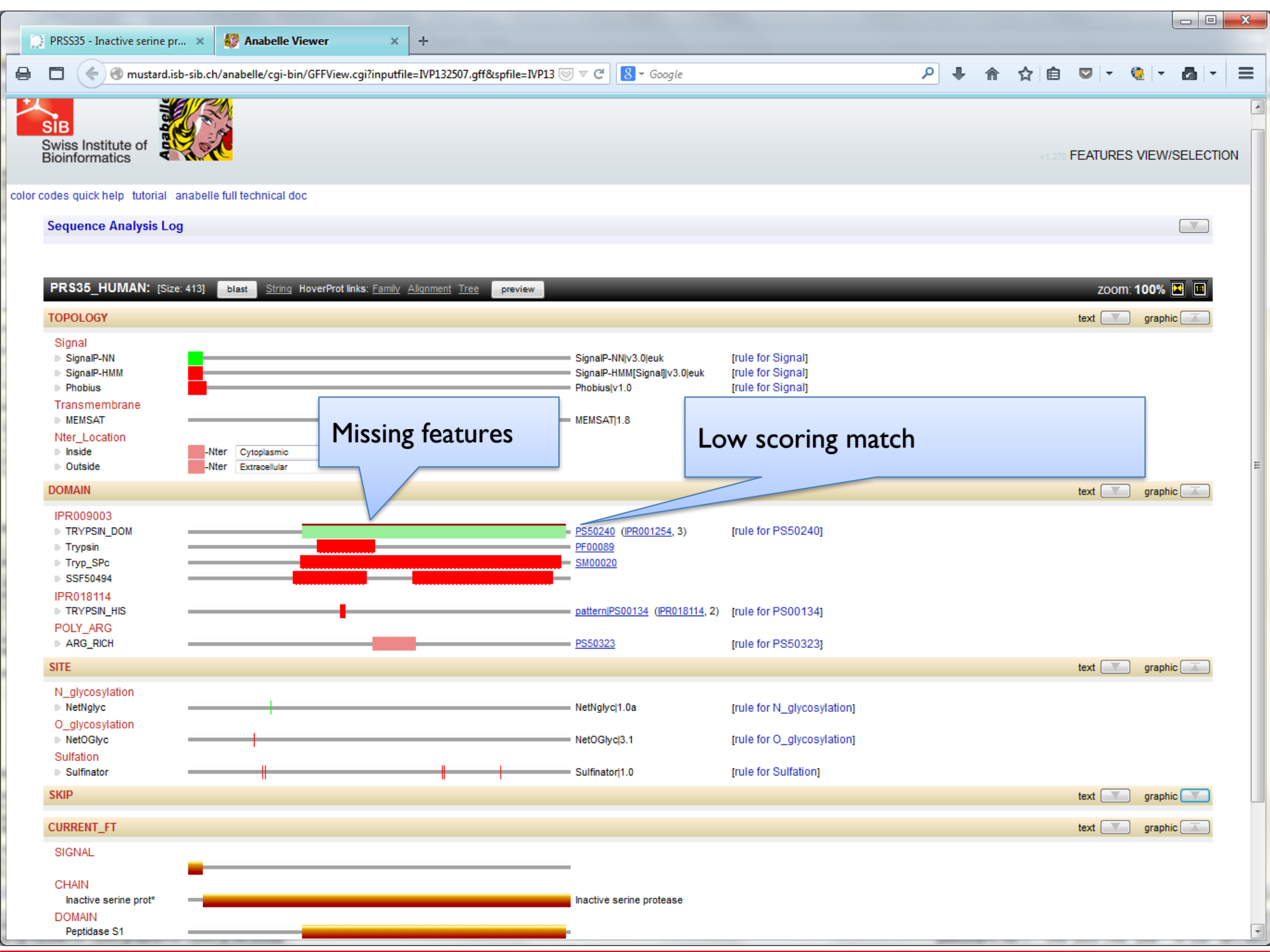
PROSITE annotation in UniProtKB/Swiss-Prot

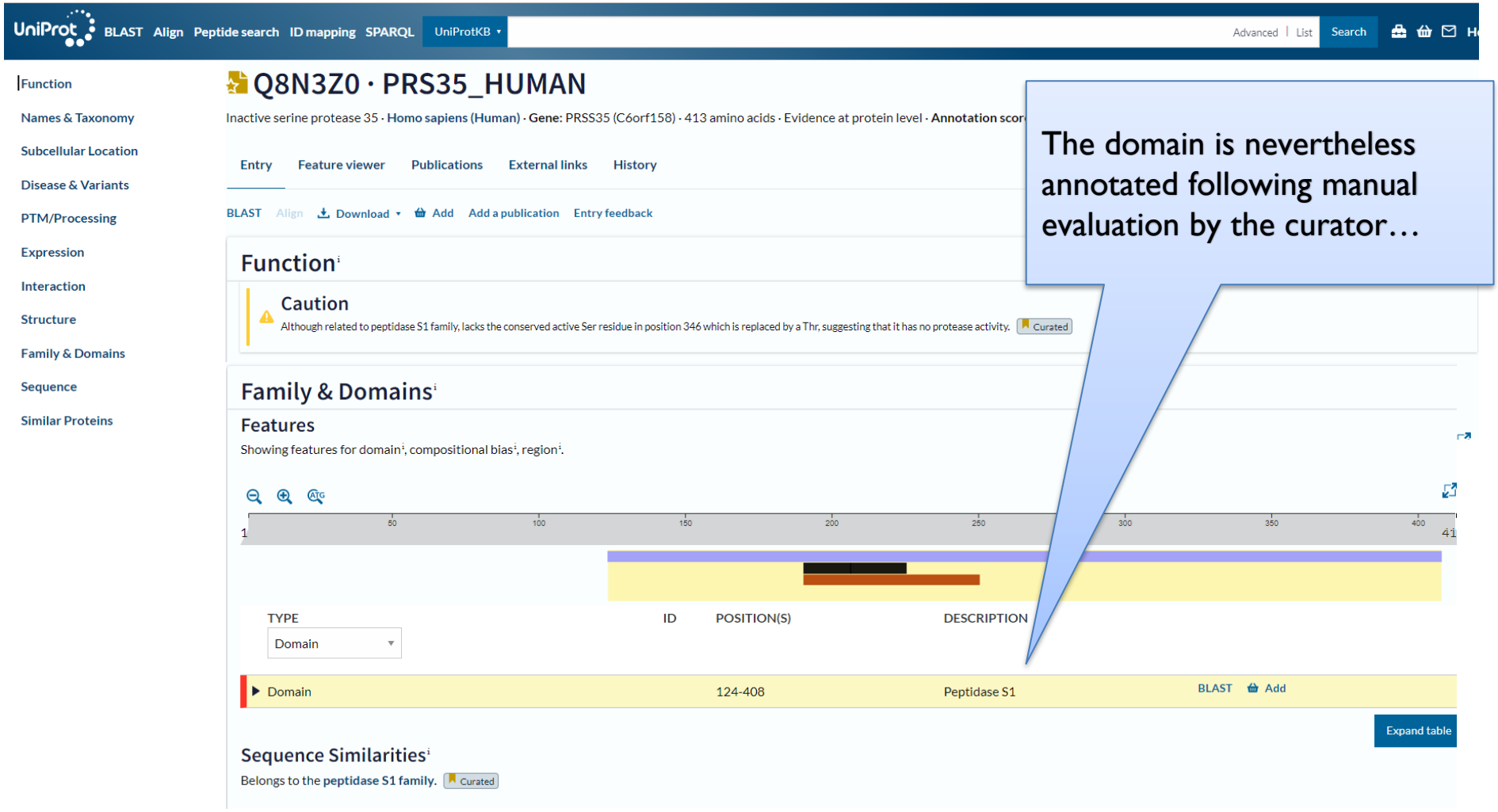


- **PROSITE** is routinely used in the manual annotation in **UniProtKB/Swiss-Prot**
- It forms part of an integrated annotation tool '**Anabelle**' which presents PROSITE results in the context of other predictors for evaluation by Swiss-Prot curators
- PROSITE matches are then **evaluated** during curation and subsequently integrated into UniProtKB/Swiss-Prot records
- This curated information is used to calculate the reliability of patterns and profiles – indicated on PROSITE pages

UniProtKB/Swiss-Prot annotation system




PRSS35 - Inactive serine pr...
Anabelle Viewer
mustard.isb-sib.ch/anabelle/cgi-bin/GFFView.cgi?inputfile=IVP132507.gff&spfile=IVP13
Google
Swiss Institute of Bioinformatics
FEATURES VIEW/SELECTION
color codes quick help tutorial anabelle full technical doc
Sequence Analysis Log
PRS35_HUMAN: [Size: 413] blast String HoverProt links: Family Alignment Tree preview
zoom: 100%
TOPOLOGY
text graphic
Signal
SignalP-NN
SignalP-HMM
Phobius
SignalP-NN|v3.0|euk [rule for Signal]
SignalP-HMM[Signal]|v3.0|euk [rule for Signal]
Phobius|v1.0 [rule for Signal]
Transmembrane
MEMSAT
MEMSAT|1.8
Nter_Location
Inside
Outside
-Nter Cytoplasmic
-Nter Extracellular
Missing features
Low scoring match
DOMAIN
text graphic
IPR009003
TRYPSIN_DOM
Trypsin
Tryp_SPc
SSF50494
PS50240 (IPR001254, 3) [rule for PS50240]
PF00089
SM00020
IPR018114
TRYPSIN_HIS
pattern|PS00134 (IPR018114, 2) [rule for PS00134]
POLY_ARG
ARG_RICH
PS50323 [rule for PS50323]
SITE
text graphic
N_glycosylation
NetNglyc
NetNglyc|1.0a [rule for N_glycosylation]
O_glycosylation
NetOGlyc
NetOGlyc|3.1 [rule for O_glycosylation]
Sulfation
Sulfinator
Sulfinator|1.0 [rule for Sulfation]
SKIP
text graphic
CURRENT_FT
text graphic
SIGNAL
CHAIN
Inactive serine prot*
Inactive serine protease
DOMAIN
Peptidase S1


UniProt
BLAST
Align
Peptide search
ID mapping
SPARQL
UniProtKB
Advanced
List
Search
Function
Names & Taxonomy
Subcellular Location
Disease & Variants
PTM/Processing
Expression
Interaction
Structure
Family & Domains
Sequence
Similar Proteins
Q8N3Z0 · PRS35_HUMAN
Inactive serine protease 35 · Homo sapiens (Human) · Gene: PRSS35 (C6orf158) · 413 amino acids · Evidence at protein level · Annotation scor
Entry
Feature viewer
Publications
External links
History
BLAST
Align
Download
Add
Add a publication
Entry feedback
Function
Caution
Although related to peptidase S1 family, lacks the conserved active Ser residue in position 346 which is replaced by a Thr, suggesting that it has no protease activity.
Curated
Family & Domains
Features
Showing features for domain, compositional bias, region.
1
50
100
150
200
250
300
350
400
TYPE
ID
POSITION(S)
DESCRIPTION
Domain
Domain
124-408
Peptidase S1
BLAST
Add
Expand table
Sequence Similarities
Belongs to the peptidase S1 family.
Curated
The domain is nevertheless annotated following manual evaluation by the curator…

General information about the entry

Entry name [info]	TRYPSIN_DOM
Accession [info]	PS50240
Entry type [info]	MATRIX
Date [info]	DEC-2001 (CREATED); OCT-2013 (DATA UPDATE); APR-2015 (INFO UPDATE).
PROSITE Doc. [info]	PDOC00124
Associated ProRule [info]	PRU00274

Name and characterization of the entry

Description [info]	Serine proteases, trypsin domain profile.
Matrix / Profile [info]	/GENERAL_SPEC: ALPHABET='ABCDEFGHIKLMNPQRSTVWYZ'; LENGTH=234; /DISJOINT: DEFINITION=PROTECT; N1=6; N2=229; /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=0.0169; R2=0.00836256; TEXT='NScore'; /NORMALIZATION: MODE=-1; FUNCTION=LINEAR; R1=6057.56402; R2=28.29509; TEXT='HScore'; /CUT_OFF: LEVEL=0; SCORE=1134; N_SCORE=9.5; H_SCORE=34749; MODE=1; TEXT='!'; /CUT_OFF: LEVEL=-1; SCORE=775; N_SCORE=6.5; H_SCORE=27986; MODE=1; TEXT='?'; /DEFAULT: M0=-9; D=-20; I=-20; B1=-60; E1=-60; MI=-105; MD=-105; IM=-105; DM=-105; ... » More

Numerical results [info]

Numerical results for UniProtKB/Swiss-Prot release **2015_06** which contains **548'586** sequence entries.

Total number of hits	759 in 739 different sequences
Number of true positive hits	759 in 739 different sequences
Number of 'unknown' hits	
Number of false positive hits	
Number of false negative sequences	17
Number of 'partial' sequences	80
Precision (true positives / (true positives + false positives))	100.00 %
Recall (true positives / (true positives + false negatives))	97.81 %

…and tagged as false negative

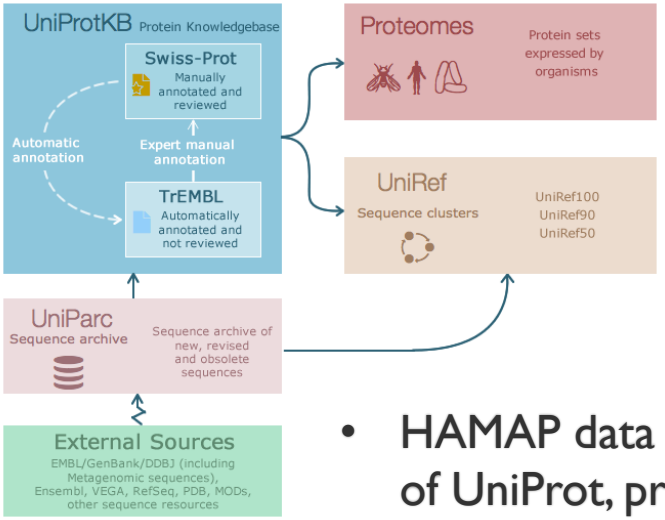
Feature key [info]	DOMAIN
Feature description [info]	Peptidase S1
Version [info]	2

Cross-references [info]

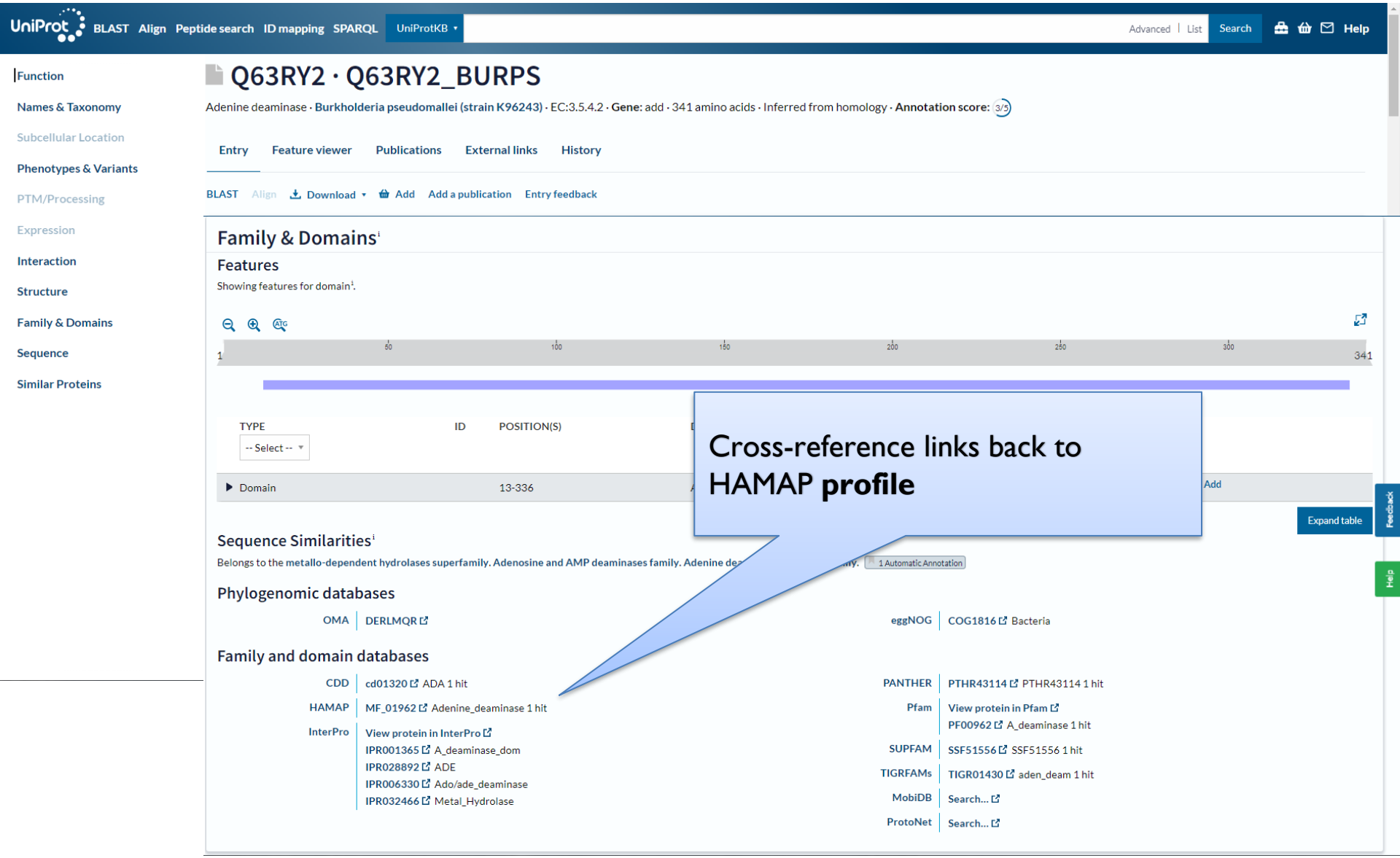
UniProtKB/Swiss-Prot True positive sequences	739 sequences ACH1_LONAC (P23604), ACH2_LONAC (P23605), ACRO_HUMAN (P10323), » More • Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits: [Clustal format, color, condensed view] [Clustal format, color] [Clustal format, plain text] [Fasta format] • Retrieve the sequence logo from the alignment
UniProtKB/Swiss-Prot False negative sequences	17 sequences HTR1B_DANRE (A9JRB3), HTRA1_BOVIN (F1N152), HTRA1_XENLA (A6YFB5), HTRA1_XENTR (A4IHA1), PRS23_BOVIN (Q1LZE9), PRS23_HUMAN (O95084), PRS23_MACMU (Q1WK23), PRS23_MOUSE (Q9D6M6), PRS23_RAT (Q6AY61), PRS35_BOVIN (Q5E9X[illegible]), PRS35_HUMAN (Q8N3Z0), PRS35_MACMU (Q1WK24), PRS35_MOUSE (Q8C0F9), PRS35_RAT (Q5FBZ2), Y3671_MYCTO (P9WHR8), Y3671_MYCTU (P9WHR9), YH05_SCHPO (O74325) Less «
UniProtKB/Swiss-Prot 'Partial' sequences	80 sequences ACRO_CAPHI (P10626), CATG_RAT (P17977), CBP_MESMA (P0C8M2), » More
PDB [Detailed view]	1685 PDB 1A0H; 1A0J; 1A0L; 1A2C; 1A3B; 1A3E; 1A46; 1A4W; 1A5G; 1A5H; 1A5I; 1A61; » More

View entry in original PROSITE format
View entry in raw text format (no links)
Direct ScanProsite submission

HAMAP annotation in UniProtKB



- HAMAP data are incorporated by InterPro and the UniRule pipeline of UniProt, providing annotation of UniProtKB/Swiss-Prot quality for millions of unreviewed protein sequences in UniProtKB/TrEMBL
- Each month, new sequences entering UniProtKB/TrEMBL are scanned against the collection of HAMAP profiles
- Matching sequences are evaluated against corresponding rule conditions and annotated accordingly
- Results are made available in the corresponding UniProtKB/TrEMBL records (search for 'source:HAMAP')


UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search Help
Function
Names & Taxonomy
Subcellular Location
Phenotypes & Variants
PTM/Processing
Expression
Interaction
Structure
Family & Domains
Sequence
Similar Proteins
Q63RY2 · Q63RY2_BURPS
Adenine deaminase · Burkholderia pseudomallei (strain K96243) · EC:3.5.4.2 · Gene: add · 341 amino acids · Inferred from homology · Annotation score: 3/5
Entry Feature viewer Publications External links History
BLAST Align Download Add Add a publication Entry feedback
Family & Domains
Features
Showing features for domain.
1 50 100 150 200 250 300 341
TYPE ID POSITION(S)
-- Select --
Domain 13-336 Add
Expand table
Cross-reference links back to HAMAP profile
Sequence Similarities
Belongs to the metallo-dependent hydrolases superfamily. Adenosine and AMP deaminases family. Adenine de
1 Automatic Annotation
Phylogenomic databases
OMA DERLMQR
eggNOG COG1816 Bacteria
Family and domain databases
CDD cd01320 ADA 1 hit
HAMAP MF_01962 Adenine_deaminase 1 hit
InterPro View protein in InterPro
IPR001365 A_deaminase_dom
IPR028892 ADE
IPR006330 Ado/ade_deaminase
IPR032466 Metal_Hydrolase
PANTHER PTHR43114 PTHR43114 1 hit
Pfam View protein in Pfam
PF00962 A_deaminase 1 hit
SUPFAM SSF51556 SSF51556 1 hit
TIGRFAMs TIGR01430 aden_deam 1 hit
MobiDB Search...
ProtoNet Search...
Feedback
Help

Family profile MF_01962

General profile information [?]

Accession	MF_01962	[View profile]
Entry name	Adenine_deaminase	[View seed alignment]
Entry type	MATRIX	
Date	DEC-2013 (DATA UPDATE).	
Version	4	
Description	Adenine deaminase.	
Taxonomic range	Archaea, Bacteria, Eukaryota	
InterPro	IPR028892 ADE	
Associated rules	MF_01962 name: Adenine_deaminase scope: Bacteria MF_03145 name: Adenine_deaminase_euk scope: Eukaryota	

Statistics [?]

Number of hits in UniProtKB	1,867	[Graphical view of score distribution]
• Number of hits in UniProtKB/Swiss-Prot	70	
• Number of hits in UniProtKB/TrEMBL	1,797	
Taxonomic distribution of hits in UniProtKB		[View taxonomic distribution of UniProtKB matches]
• Archaea	4	[Taxonomic distribution in UniProtKB complete proteomes]
• Bacteria	1,515	
• Eukaryota	347	
• unclassified sequences	1	

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search Help

- Function
- Names & Taxonomy
- Subcellular Location
- Phenotypes & Variants
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Similar Proteins

Q63RY2 · Q63RY2_BURPS

Adenine deaminase · **Burkholderia pseudomallei (strain K96243)** · EC:3.5.4.2 · **Gene:** add · 341 amino acids · Inferred from homology · **Annotation score:** 3/5

Entry | Feature viewer | Publications | External links | History

BLAST | Align | Download | Add | Add a publication | Entry feedback

Function

Catalyzes the hydrolytic deamination of adenine to hypoxanthine. Plays an important role in the purine salvage pathway and in nitrogen catabolism. 1 Automatic Annotation

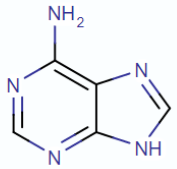
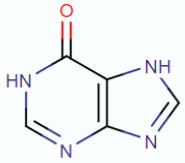
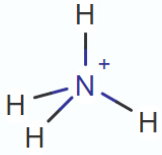
Automatic assertion according to rules (Automatically inferred from sequence model)
UniRule HAMAP-Rule: MF_01962

Evidence attribution links back to **annotation rule**

Catalytic Activity

adenine + H^+ + H_2O = hypoxanthine + NH_4^+ 1 Automatic Annotation
EC:3.5.4.2 (UniProtKB | ENZYME | Rhea)
Source: Rhea 23688

ea reaction

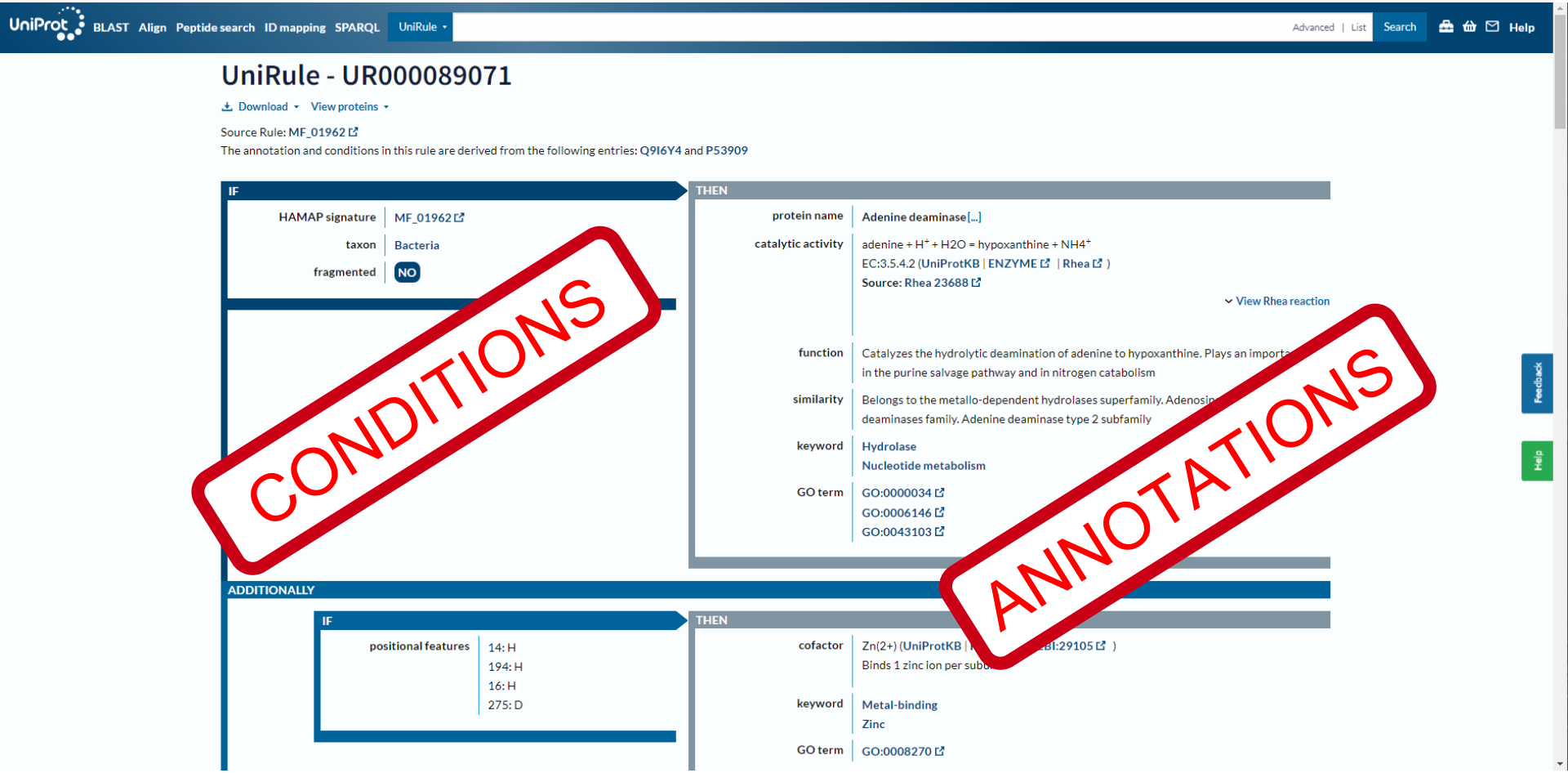
adenine	H^+	H_2O	hypoxanthine	NH_4^+
CHEBI:16708	CHEBI:15378	CHEBI:15377	CHEBI:17368	CHEBI:28938
				
zoom	zoom	zoom	zoom	zoom

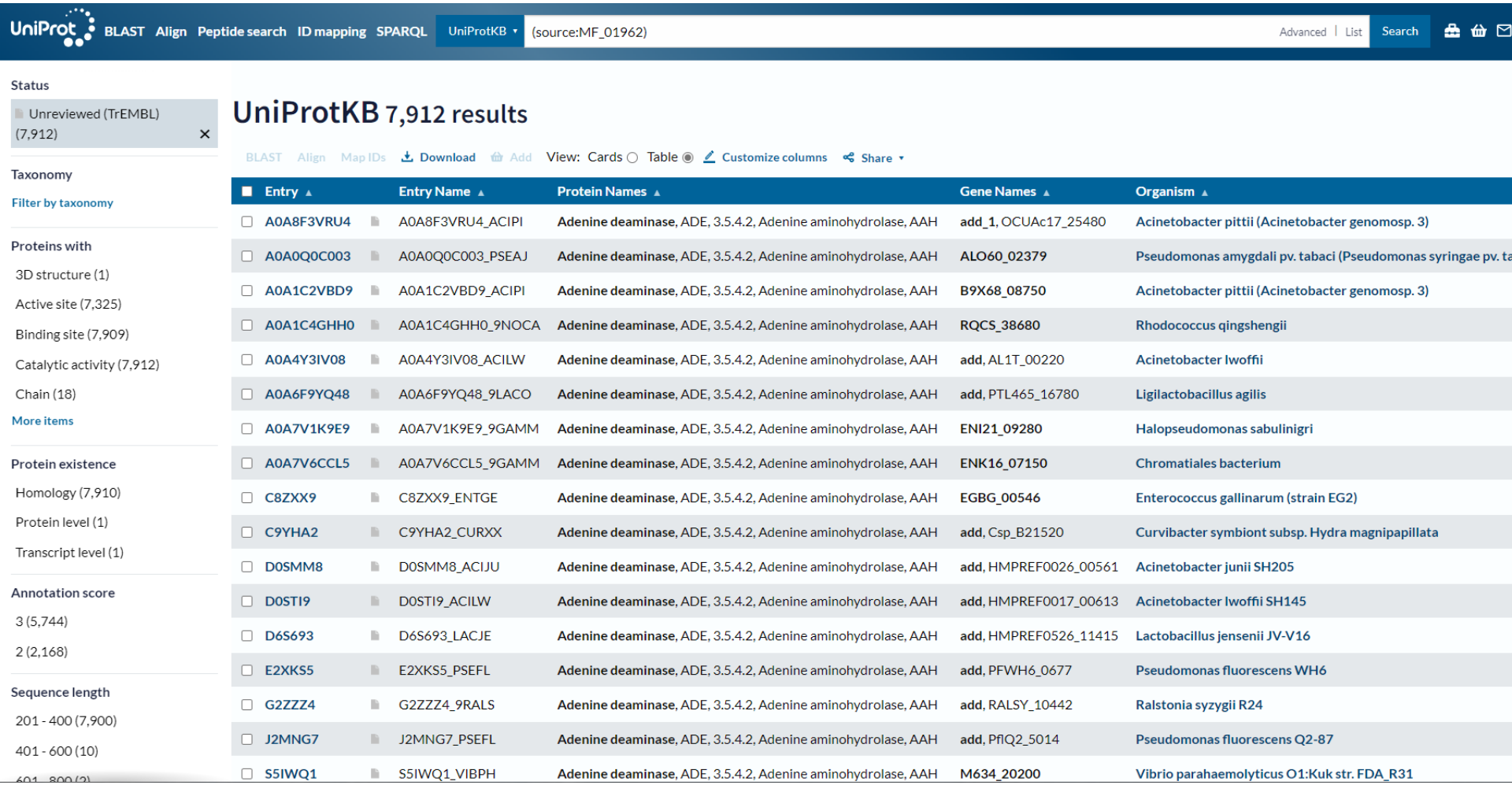
Cofactor

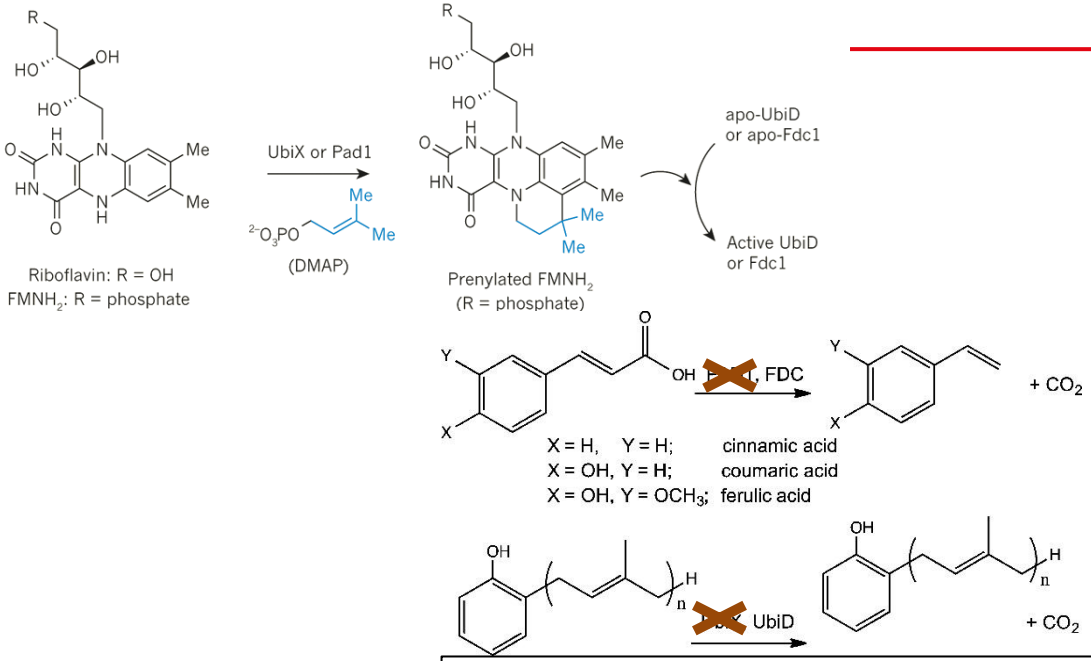
Zn(2+) (UniProtKB | Rhea | CHEBI:29105) 1 Automatic Annotation

Feedback

Help


UniProt
BLAST
Align
Peptide search
ID mapping
SPARQL
UniRule
Advanced | List
Search
Help
UniRule - UR000089071
Download
View proteins
Source Rule: MF_01962
The annotation and conditions in this rule are derived from the following entries: Q9I6Y4 and P53909
IF
HAMAP signature
MF_01962
taxon
Bacteria
fragmented
NO
CONDITIONS
THEN
protein name
Adenine deaminase[...]
catalytic activity
adenine + H+ + H2O = hypoxanthine + NH4+
EC:3.5.4.2 (UniProtKB | ENZYME | Rhea)
Source: Rhea 23688
View Rhea reaction
function
Catalyzes the hydrolytic deamination of adenine to hypoxanthine. Plays an importa
in the purine salvage pathway and in nitrogen catabolism
similarity
Belongs to the metallo-dependent hydrolases superfamily. Adenosin
deaminases family. Adenine deaminase type 2 subfamily
keyword
Hydrolase
Nucleotide metabolism
GO term
GO:0000034
GO:0006146
GO:0043103
ANNOTATIONS
ADDITIONALLY
IF
positional features
14: H
194: H
16: H
275: D
THEN
cofactor
Zn(2+) (UniProtKB |
BI:29105)
Binds 1 zinc ion per subu
keyword
Metal-binding
Zinc
GO term
GO:0008270
Feedback
Help


UniProt
BLAST
Align
Peptide search
ID mapping
SPARQL
UniProtKB
(source:MF_01962)
Advanced | List
Search
Status
Unreviewed (TrEMBL) (7,912)
Taxonomy
Filter by taxonomy
Proteins with
3D structure (1)
Active site (7,325)
Binding site (7,909)
Catalytic activity (7,912)
Chain (18)
More items
Protein existence
Homology (7,910)
Protein level (1)
Transcript level (1)
Annotation score
3 (5,744)
2 (2,168)
Sequence length
201 - 400 (7,900)
401 - 600 (10)
UniProtKB 7,912 results
BLAST Align Map IDs Download Add View: Cards Table Customize columns Share
Entry Entry Name Protein Names Gene Names Organism
A0A8F3VRU4 A0A8F3VRU4_ACIPI Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add_1, OCUAc17_25480 Acinetobacter pittii (Acinetobacter genomosp. 3)
A0A0Q0C003 A0A0Q0C003_PSEAJ Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH ALO60_02379 Pseudomonas amygdali pv. tabaci (Pseudomonas syringae pv. ta
A0A1C2VBD9 A0A1C2VBD9_ACIPI Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH B9X68_08750 Acinetobacter pittii (Acinetobacter genomosp. 3)
A0A1C4GHH0 A0A1C4GHH0_9NOCA Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH RQCS_38680 Rhodococcus qingshengii
A0A4Y3IV08 A0A4Y3IV08_ACILW Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, AL1T_00220 Acinetobacter lwoffii
A0A6F9YQ48 A0A6F9YQ48_9LACO Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, PTL465_16780 Ligilactobacillus agilis
A0A7V1K9E9 A0A7V1K9E9_9GAMM Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH ENI21_09280 Halopseudomonas sabulinigri
A0A7V6CCL5 A0A7V6CCL5_9GAMM Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH ENK16_07150 Chromatiales bacterium
C8ZXX9 C8ZXX9_ENTGE Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH EGBG_00546 Enterococcus gallinarum (strain EG2)
C9YHA2 C9YHA2_CURXX Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, Csp_B21520 Curvibacter symbiont subsp. Hydra magnipapillata
D0SMM8 D0SMM8_ACIJU Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, HMPREF0026_00561 Acinetobacter junii SH205
D0STI9 D0STI9_ACILW Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, HMPREF0017_00613 Acinetobacter lwoffii SH145
D6S693 D6S693_LACJE Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, HMPREF0526_11415 Lactobacillus jensenii JV-V16
E2XKS5 E2XKS5_PSEFL Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, PFWH6_0677 Pseudomonas fluorescens WH6
G2ZZZ4 G2ZZZ4_9RALS Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, RALSY_10442 Ralstonia syzygii R24
J2MNG7 J2MNG7_PSEFL Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH add, PflQ2_5014 Pseudomonas fluorescens Q2-87
S5IWQ1 S5IWQ1_VIBPH Adenine deaminase, ADE, 3.5.4.2, Adenine aminohydrolase, AAH M634_20200 Vibrio parahaemolyticus O1:Kuk str. FDA_R31




```
DT   05-NOV-2015 (Rel. 126, Created)
DT   24-MAR-2016 (Rel. 128, Last updated, Version 4)
XX
DE   Klebsiella pneumoniae 3-octaprenyl-4-hydroxybenzoate carboxy-lyase
XX
.
FT   CDS             CP012753.1:5179010..5179594
FT                   /codon_start=1
FT                   /transl_table=11
FT                   /locus_tag="AN966_25510"
FT                   /product="3-octaprenyl-4-hydroxybenzoate carboxy-lyase"
FT                   /EC_number="4.1.1.-"
FT                   /note="catalyzes the formation of 2-octaprenylphenol from
FT                   3-octaprenyl-4-hydroxybenzoate; Derived by automated
FT                   computational analysis using gene prediction method:
FT                   Protein Homology."
FT                   /inference="EXISTENCE:similar to AA
FT                   sequence:RefSeq:WP_017899515.1"
FT                   /protein_id="ALH88064.1"
FT                   /translation="MKRLIIGISGASGAIYGVRMLQVLRDVPDIETHLILSQAARQTLA
FT                   METDFTVREVQALADVVHDARDIAASISSGSFKTAGMVILPCSMKTLSGIVHSYTDGLL
FT                   TRAADVVLKERRPLVLCVRETPFHLGHLRLLVQAAELGAVIMPPVPAFYHRPQSLDDVI
FT                   NQTVNRVLDQFDISLEQDLFTRWQGSQDCTK"
XX
```

nature International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issu
Archive | Volume 522 | Issue 7557 | Letters | Article
NATURE | LETTER
日本語要約

UbiX is a flavin prenyltransferase re… bacterial ubiquinone biosynthesis

Mark D. White, Karl A. P. Payne, Karl Fisher, Stephen A. Mars…
J. W. Rattray, Drupad K. Trivedi, Royston Goodacre, Stephen…
Scrutton, Sam Hay & David Leys
Affiliations | Contributions | Corresponding author

Nature 522, 502–506 (25 June 2015) | doi:10.1038/nature14559

…UbiX Catalyze Formation of a Novel …d Decarboxylase and 4-Hydroxy-…oxylase

…er†, Xiaoxia Nina Lin*‡, and E. Neil G. Marsh*†§
…ngineering, and §Department of Biological Chemistry, University of

BLAST | Align | Format | Add to basket | History

Feedback | Help video | Other tutorials and videos

Protein **Flavin prenyltransferase UbiX**

Gene **pad1_1**

Organism *Klebsiella pneumoniae*

Status Unreviewed - Annotation score: ●●○○○ - Protein inferred from homology[i]

A0A0C2MA92_KLEPN

Function[i]

Flavin prenyltransferase that catalyzes the synthesis of the prenylated FMN cofactor (prenyl-FMN) for 4-hydroxy-3-polyprenylbenzoic acid decarboxylase UbiD. The prenyltransferase is metal-independent and links a dimethylallyl moiety from dimethylallyl monophosphate (DMAP) to the flavin N5 and C6 atoms of FMN. UniRule annotation

Catalytic activity[i]

Dimethylallyl phosphate + $FMNH_2$ = prenylated $FMNH_2$ + phosphate. UniRule annotation

Automatic assertion according to rules[i]

HAMAP-Rule:MF_01984

Sites

Feature key	Position(s)	Length		Graphical view	Feature identifier	Actions
Binding site[i]	37 – 37	1				
Binding site[i]	123 – 123	1	FMN UniRule annotation			
Binding site[i]	153 – 153	1	DMAP UniRule annotation			

Proteins with
- 3D structure (8)
- Active site (1)
- Beta strand (5)
- Binding site (18,436)
- Biotechnological use (1)

More items

Protein existence
- Homology (18,426)
- Protein level (9)
- Transcript level (1)

Annotation score
- 5 (1)
- 4 (1)
- 3 (32)
- 2 (18,402)

Sequence length
- 1 - 200 (10,266)
- 201 - 400 (8,148)
- 401 - 600 (11)
- 601 - 800 (7)
- >= 801 (4)

UniProtKB 18,436 results

BLAST Align Map IDs Download Add View: Cards ○ Table ◉ Customize columns Share

Entry	Entry Name	Protein Names	Gene Names	Organism
P33751	PAD1_YEAST	**Flavin prenyltransferase PAD1, mitochondrial**, 2.5.1.129, Phenylacrylic acid decarboxylase 1, PAD	**PAD1**, POF1, YDR538W	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)
P0AG03	UBIX_ECOLI	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, dedF, b2311, JW2308	Escherichia coli (strain K12)
Q9HX08	UBIX_PSEAE	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, PA4019	Pseudomonas aeruginosa (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1)
P0AG04	UBIX_ECO57	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, Z3573, ECs3195	Escherichia coli O157:H7
Q9V030	UBIX_PYRAB	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, PYRAB09680, PAB0651	Pyrococcus abyssi (strain GE5 / Orsay)
Q9ZD09	UBIX_RICPR	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, RP541	Rickettsia prowazekii (strain Madrid E)
Q9ZJE3	UBIX_HELPJ	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, jhp_1369	Helicobacter pylori (strain J99 / ATCC 700824) (Campylobacter pylori J99)
P57767	UBIX_THAAR	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**	Thauera aromatica
Q9JXP4	UBIX_NEIMB	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, NMB1945	Neisseria meningitidis serogroup B (strain MC58)
Q9PKH2	UBIX_CHLMU	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, TC_0493	Chlamydia muridarum (strain MoPn / Nigg)
Q9Y8K8	UBIX_SACS2	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, SSO0437, C41_048	Saccharolobus solfataricus (strain ATCC 35092 / DSM 1617 / JCM 11322 / P2) (Sulfolobus solfataricus)
O26011	UBIX_HELPY	**Flavin prenyltransferase UbiX**, 2.5.1.129	**ubiX**, HP_1476	Helicobacter pylori (strain ATCC 700392 / 26695) (Campylobacter pylori)

Feedback

Help

Function
- Names & Taxonomy
- Subcellular Location
- Phenotypes & Variants
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence

A0A0C2MA92 · A0A0C2MA92_KLEPN

Flavin prenyltransferase UbiX · **Klebsiella pneumoniae** · EC:2.5.1.129 · **Gene:** pad1_2 (pad1_1, ubiX) · 194 amino acids · Inferred from homology · **Annotation score:** 2/5

Entry | Feature viewer | Publications | External links | History

BLAST Align Download Add Add a publication Entry feedback

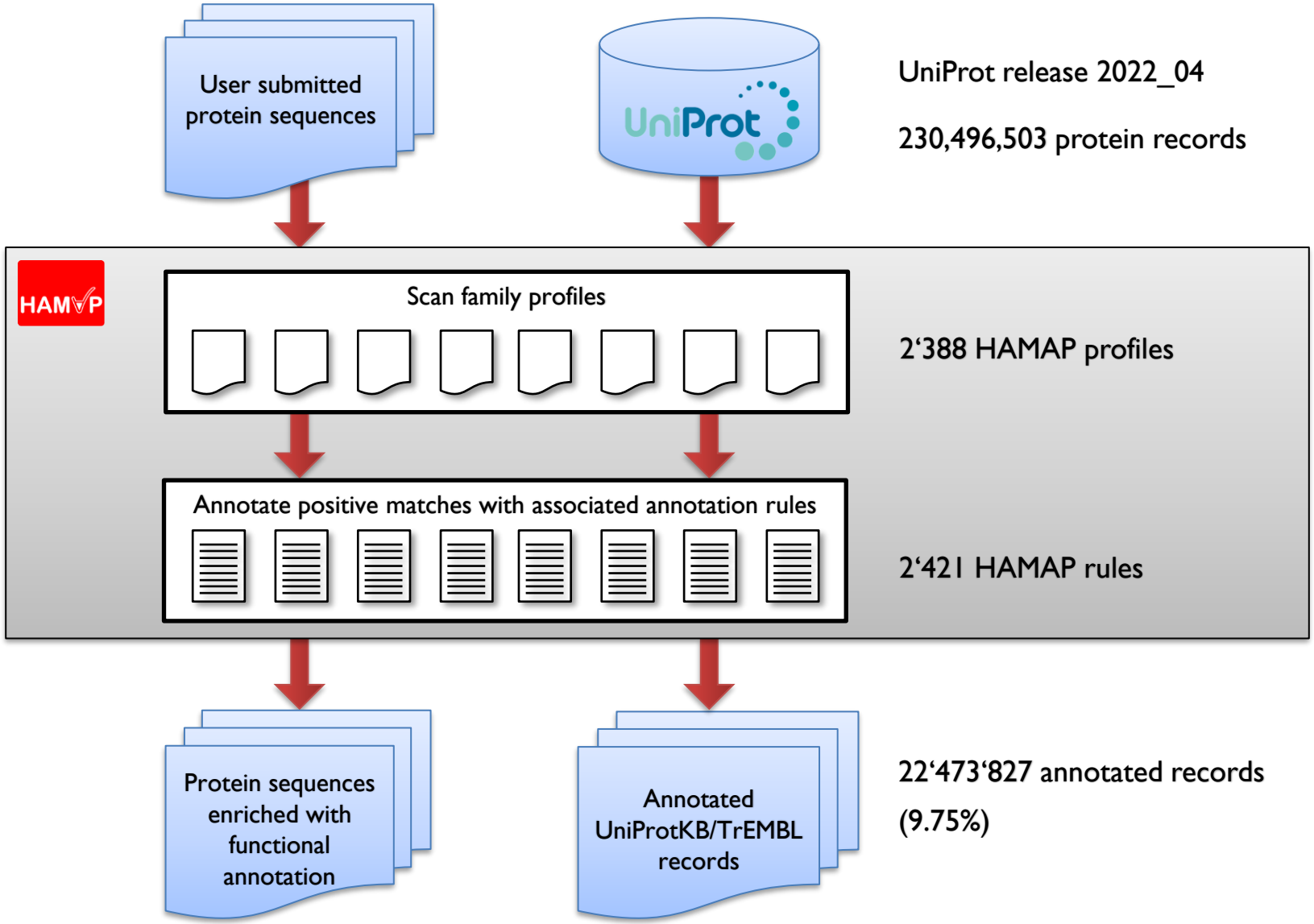
Function

Flavin prenyltransferase that catalyzes the synthesis of the prenylated FMN cofactor (prenyl-FMN) for 4-hydroxy-3-polyprenylbenzoic acid decarboxylase UbiD. The prenyltransferase is metal-independent and links a dimethylallyl moiety from dimethylallyl monophosphate (DMAP) to the flavin N5 and C6 atoms of FMN. 1 Automatic Annotation

Automatic assertion according to rules (Automatically inferred from sequence model)
UniRule HAMAP-Rule: MF_01984

Feedback

Current status



ExPASy Bioinformatics Resource Portal

HAMAP

Home | Contact

Home | Browse | HAMAP-Scan | Proteomes | Documents | Downloads

Search HAMAP Search

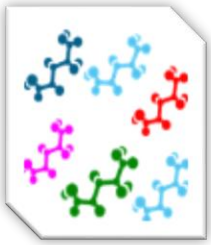
Reference proteomes: archaea

To retrieve information from a particular organism (description, genome reference paper(s), taxonomy or UniProtKB entries), click on the UniProtKB proteome identifier.

Statistics

Species	Species code	UniProtKB proteome identifier	Number of entries in UniProtKB: All	Swiss-Prot	TrEMBL	HAMAP coverage in UniProtKB
Acidianus hospitalis (strain W1)	ACIHW	UP000008458	2329	0	2329	14%
Acidilobus saccharovorans (strain DSM 16705 / JCM 18335 / VKM B-2471 / 345-15)	ACIS3	UP000000346	1499	2	1497	16%
Escherichia coli (strain K12)	ECOLI	UP000000625	4305	4305	0	26%
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Buchnera aphidicola subsp. Acyrthosiphon pisum (strain APS) (Acyrthosiphon pisum symbiotic bacte ...	BUCAI	UP000001806	572	572	0	69%
Archaeoglobus sulfaticallidus PM70-1	/	UP000013307	2213	0	2213	18%
Archaeoglobus veneficus (strain DSM 11195 / SNP6)	ARCVS	UP000008136	2065	0	2065	20%
archaeon GW2011_AR10	/	UP000031777	1339	0	1339	14%
archaeon GW2011_AR15	/	UP000031776	1308	0	1308	13%
archaeon GW2011_AR20	/	UP000031765	1010	0	1010	11%
Caldisphaera lagunensis (strain DSM 15908 / JCM 11604 / IC-154)	CALLD	UP000010469	1477	0	1477	17%

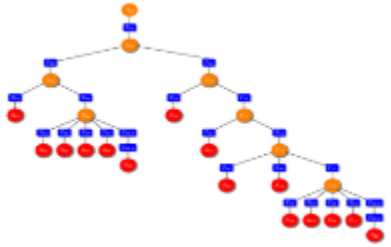
Automatic annotation in UniProtKB



UniRule
(Human informed)

Rules are created by curators.
(HAMAP, PIR name and site rules, and RuleBase rules)

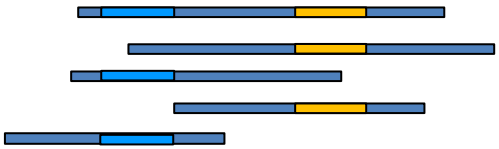
ARBA
(Machine learning)



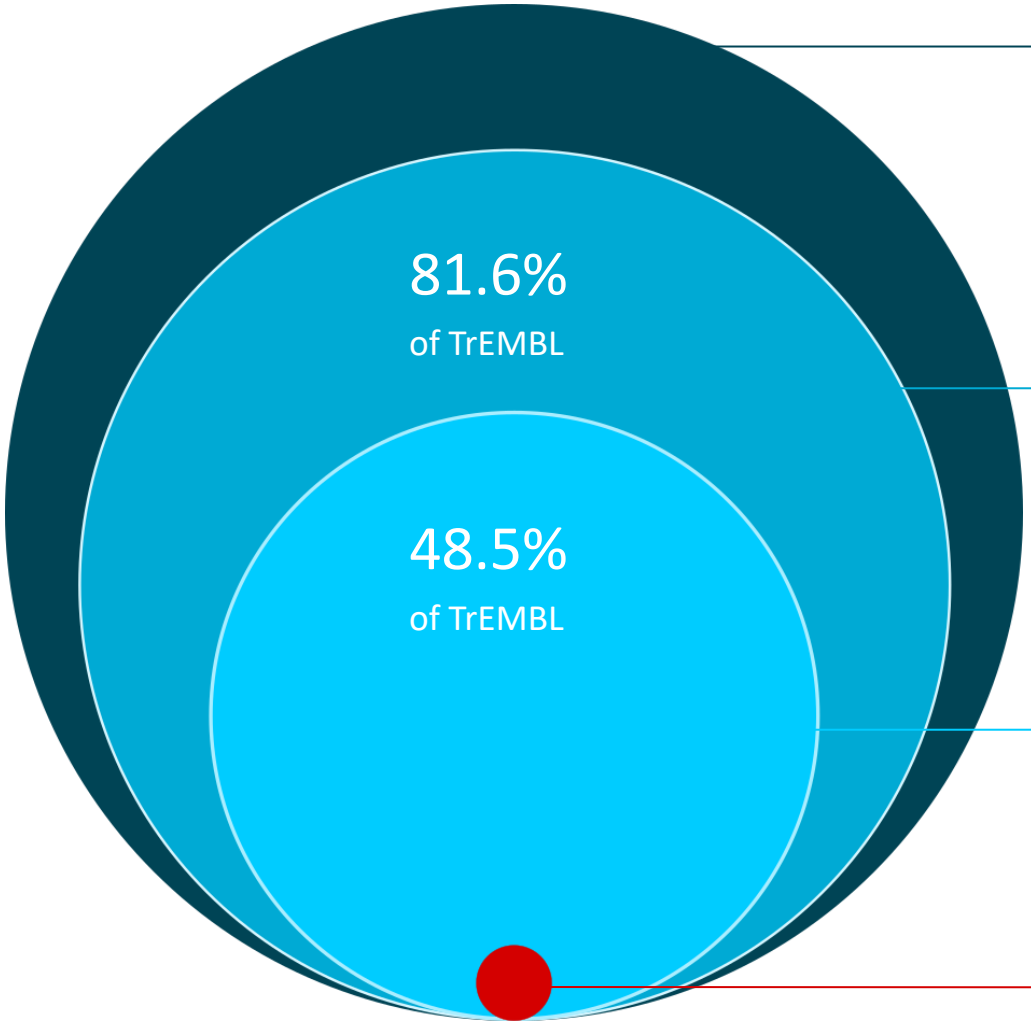
Rules are created by a modified Decision Tree algorithm.

Additional information

Predictor suites
(External software)



Sequence-specific information from external providers such as InterPro or via SAM.



214,406,868 TrEMBL

Non-validated names from submitter; no further annotations

174,854,319

Sequence feature annotation

signal peptides, transmembrane peptides, domains

104,086,334 (UniRule + ARBA)

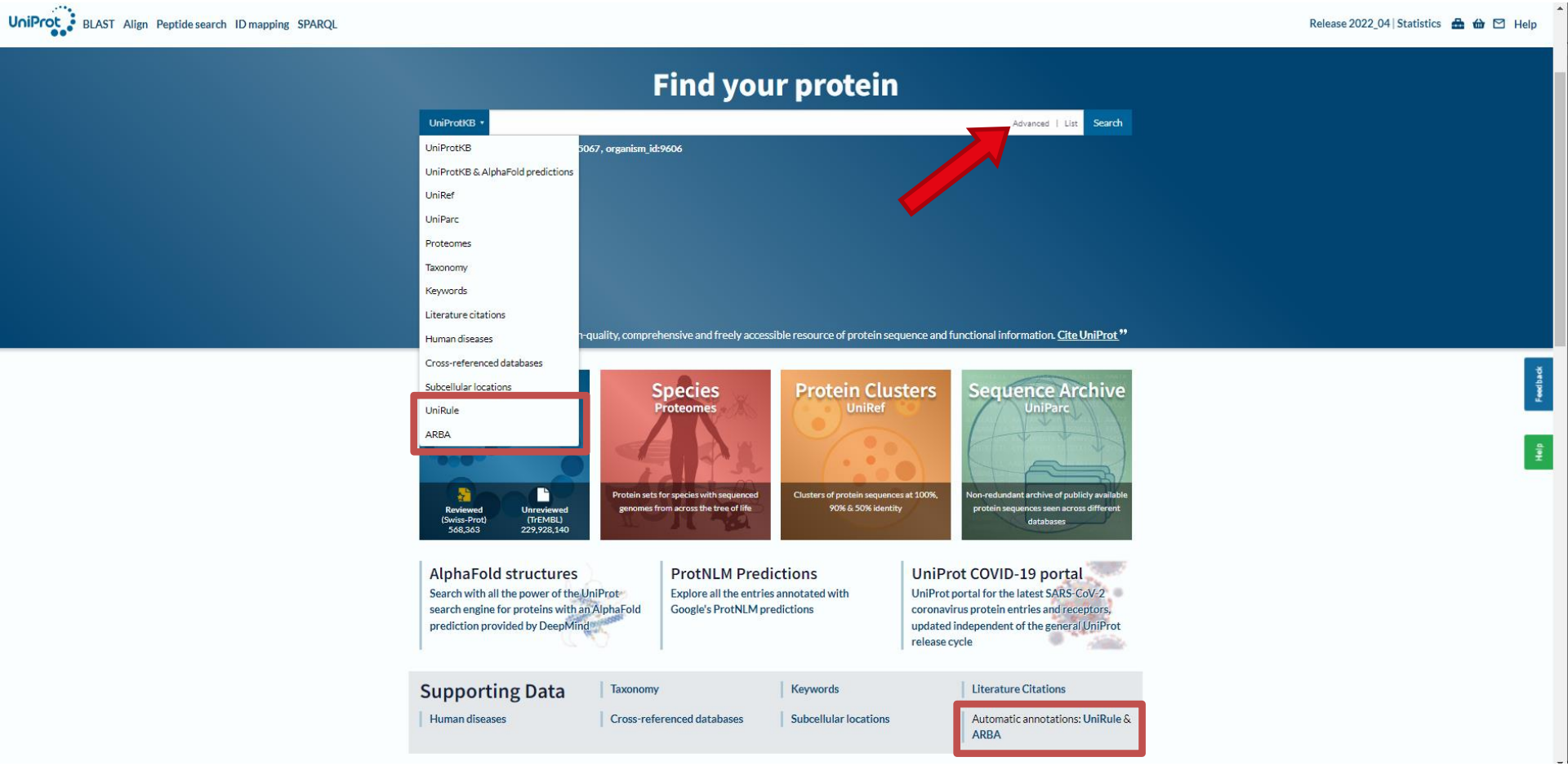
Functional annotation

protein names, enzymatic activity, interactors, subcellular location, GO terms etc.

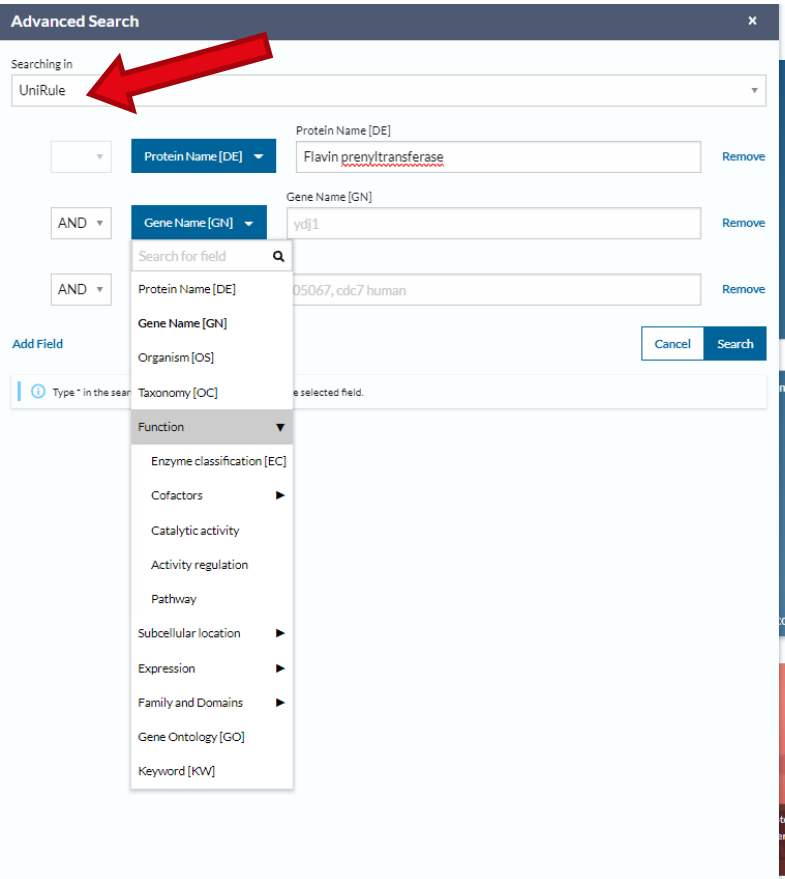
564,638 Swiss-Prot

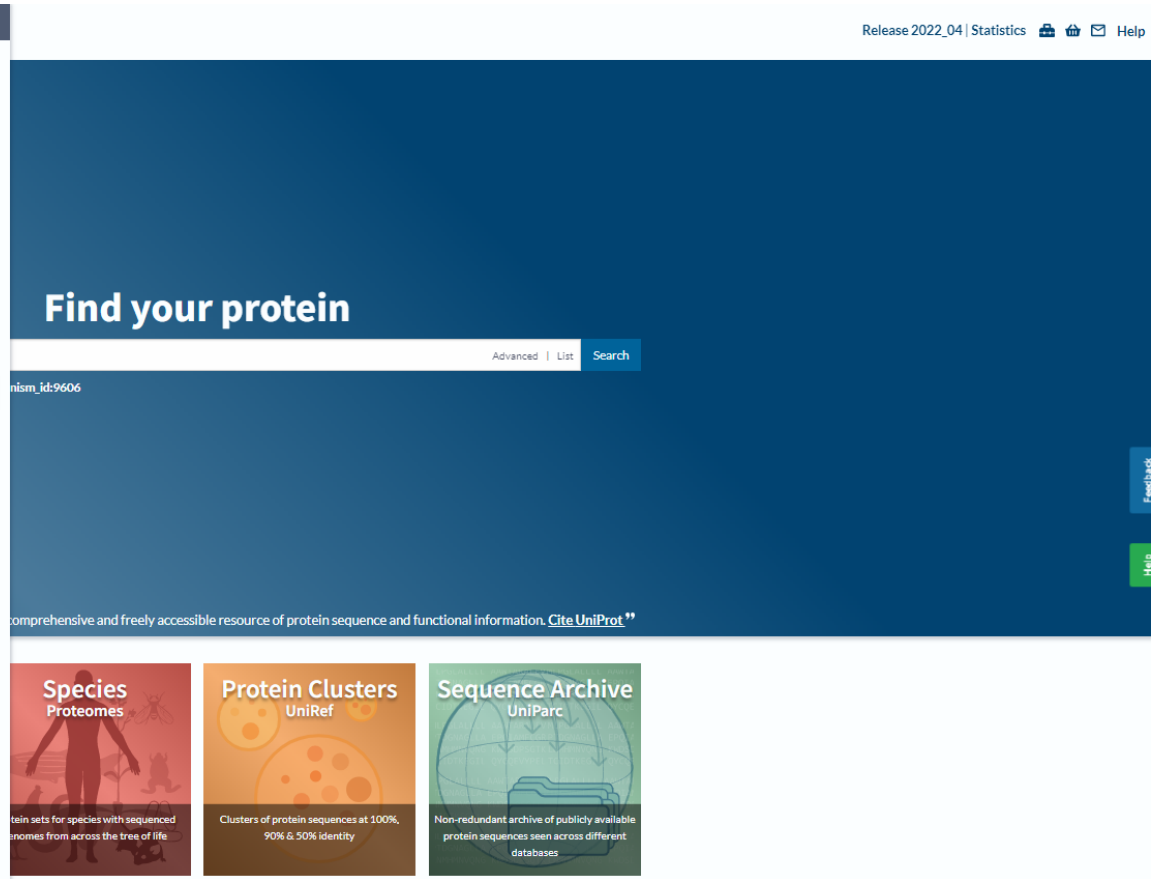
Release 2021_02

How to access automatic annotation data?

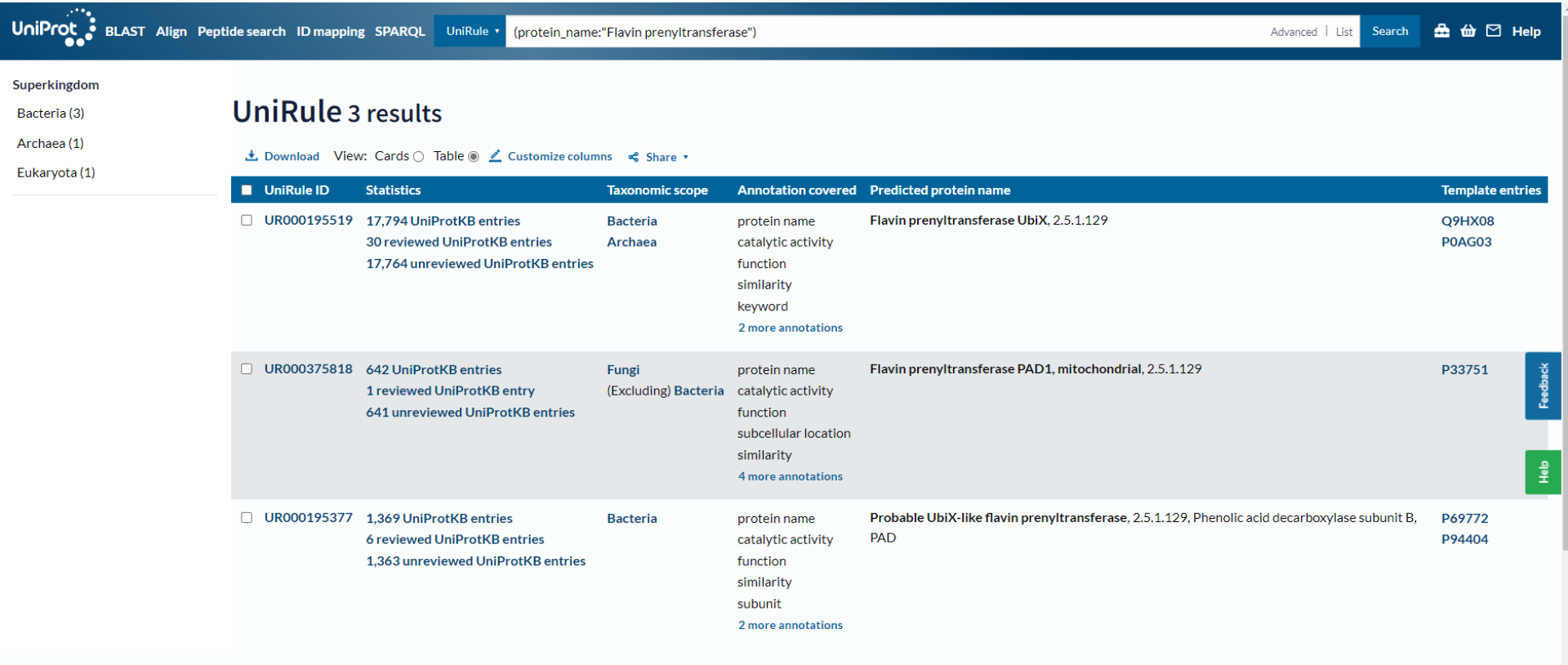


Using advanced search in UniRule





Using advanced search in UniRule

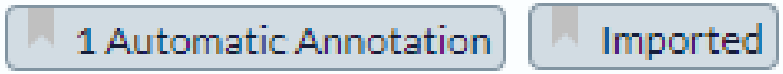


Evaluating information in UniProtKB entries using its source

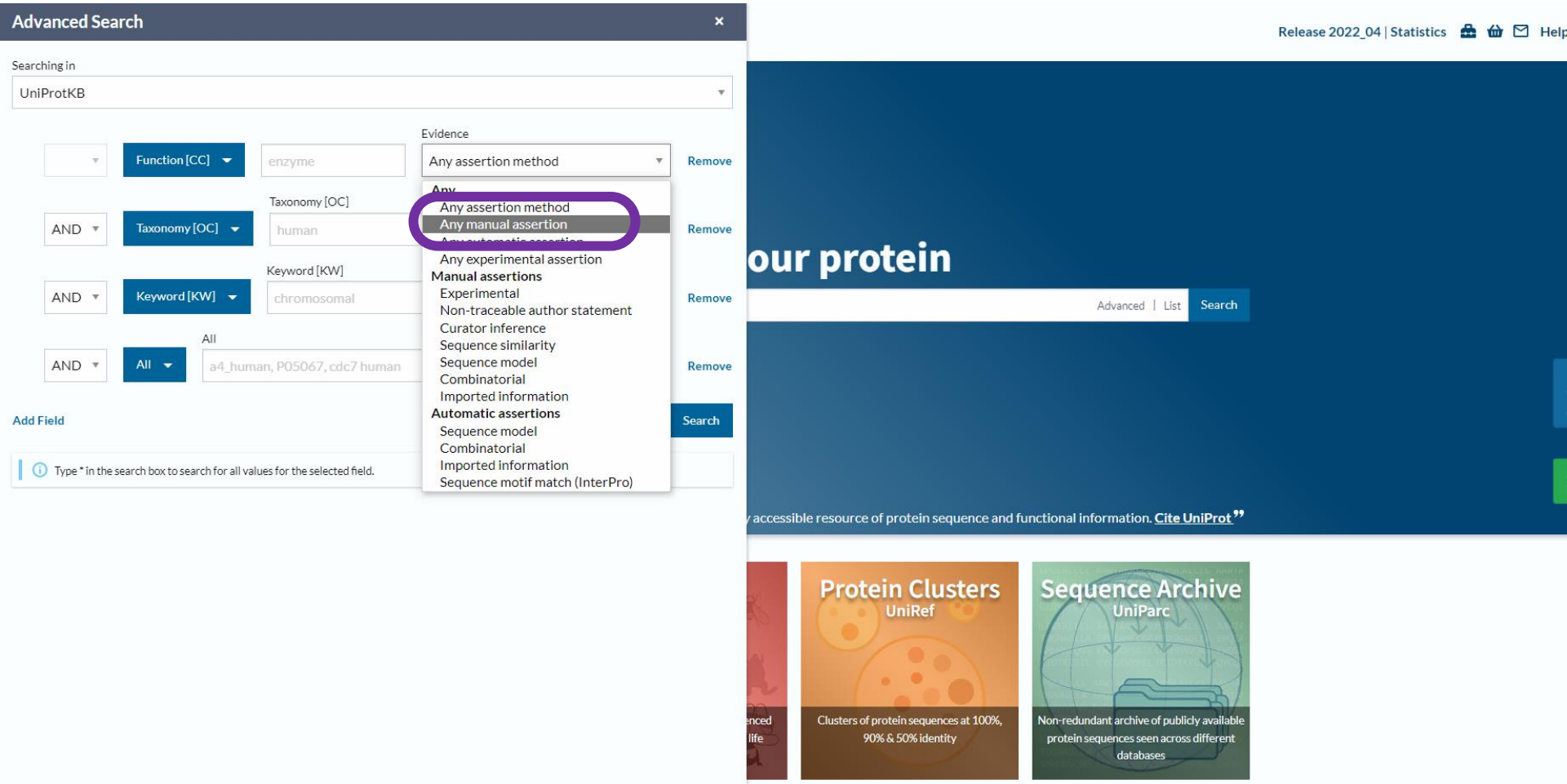
All information in an entry is attributed to its original source.

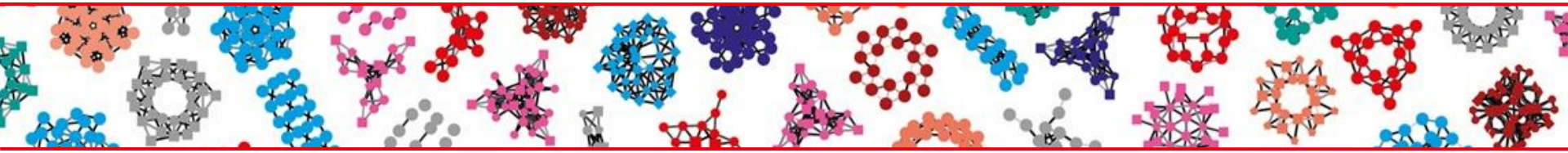
- manual assertions:

- automatic assertions:



Filtering UniProtKB results for assertion method





- Why do we need predictive annotation tools?
- Protein signatures for homology detection – A short primer
- Annotation rules for functional annotation
- HAMAP and PROSITE - automatic annotation in UniProtKB
- **HAMAP and PROSITE - services for external users**
- Practical exercises (afternoon)

Database of protein domains, families and functional sites

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 20.113 of 26-Mar-2015 contains 1718 documentation entries, 1308 patterns, 1112 profiles and 1112 ProRule.

Search

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

Search

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

Enter UniProtKB accessions or identifiers or PDB identifiers or sequences in FASTA format

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.

ScanProsite tool

This form allows you to scan proteins for matches against the PROSITE collection of motifs as well as against your own patterns.

- (•) **Option 1 - Submit PROTEIN sequences to scan them against the PROSITE collection of motifs.**
- ( ) Option 2 - Submit MOTIFS to scan them against a PROTEIN sequence database.
- ( ) Option 3 - Submit PROTEIN sequences and MOTIFS to scan them against each other.

Reset

STEP 1 - Submit PROTEIN sequences [help]

- (•) Submit PROTEIN sequences (max. 10) Examples
- ( ) Submit a PROTEIN database (max. 16MB) for repeated scans (The data will be stored on our server for 1 month).

Supported input:

- UniProtKB accessions e.g. P98073 or identifiers e.g. ENTK_HUMAN
- PDB identifiers e.g. 4DGJ
- Sequences in FASTA format

STEP 2 - Select options [help]

- [x] Exclude motifs with a high probability of occurrence from the scan
- [ ] Exclude profiles from the scan
- [ ] Run the scan at high sensitivity (show weak matches for profiles)

STEP 3 - Select output options and submit your job

Output format: Graphical view

Retrieve complete sequences: [ ] If you choose this option, not all output formats are available.

[ ] Receive your results by email

START THE SCAN Reset

ScanProsite Results Viewer

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

Hits for all PROSITE (release 20.121) motifs on sequence P98073 [UniProtKB/Swiss-Prot (release 2015_12 of 09-Dec-15: 550116 entries)]:

found: 13 hits in 1 sequence

P98073 **ENTK_HUMAN** (1019 aa)
RecName: Full=Enteropeptidase; EC=3.4.21.9; AltName: Full=Enterokinase; AltName: Full=Serine protease 7; AltName: Full=Transmembrane protease serine 15; Contains: RecName: Full=Enteropeptidase non-ca
Full=Enteropeptidase catalytic light chain; Flags: Precursor;. *Homo sapiens (Human)*

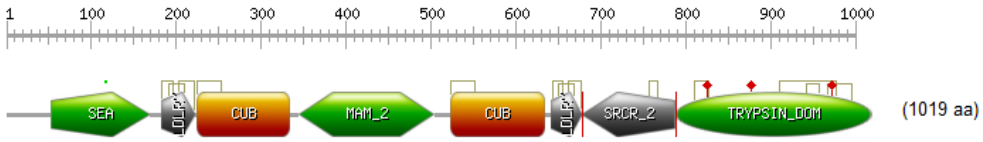
```
MGSKRGISSRHHSLSSYEIMFAALFAILVVLCAGLIAVSCLTIKESQRGAALGQSHEARATFKITS
GVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQFENGSIIVVFDLFFAQWVS
DENVKEELIQGLEANKSSQLVTFHIDLNSVDILDKLTTTSHLATPGNVSIECLPGSSPCTDALTCI
KADLFCDGEVNCPDGSDEDNKMCATVCDGRFLLTGSSGSFQATHYPKPSETSVVCQWIIRVNQGLS
IKLSFDDFNTYYTDILDIYEGVGSSKILRASIWETNPGTIRIFSNQVTATFLIESDESDYVGFNAT
YTAFNSSELNNYEKINCNFEDGFCFWVQDLNDDNEWERIQGSTFSPFTGPNFDHTFGNASGFYIST
PTGPGGRQERVGLLSLPLDPTLEPACLSFWYHMYGENVHKLSINISNDQNMEKTVFQKEGNYGDNW
NYGQVTLNETVKFKVAFNAFKNKILSDIALDDISLTYGICNGSLYPEPTLVPTPPPELPTDCGGPF
ELWEPNTTFSSTNFPNSYPNLAFCVWILNAQKGKNIQLHFQEFDLENINDVVEIRDGEEADSLLLA
VYTGPGPVKDVFSTTNRMTVLLITNDVLARGGFKANFTTGYHLGIPEPCKADHFQCKNGECVPLVN
LCDGHLHCEDGSDEADCVRFFNGTTNNNGLVRFRIQSIWHTACAENWTTQISNDVCQLLGLGSGNS
SKPIFPTDGGPFVKLNTAPDGHLILTPSQQCLQDSLIRLQCNHKSCGKKLAAQDITPKIVGGSNAK
EGAWPWVVGLYYGGRLLCGASLVSSDWLVSAAHCVYGRNLEPSKWTAILGLHMKSNLTSPQTVPRL
IDEIVINPHYNRRRKDNDIAMMHLEFKVNYTDYIQPICLPEENQVFPPGRNCSIAGWGTVVYQGTT
ANILQEADVPLLSNERCQQQMPEYNITENMICAGYEEGGIDSCQGDSGGPLMCQENNRWFLAGVTS
FGYKCALPNRPGVYARVSRFTEWIQSFLH
```

Legend:

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 http://prosite.expasy.org/mydomains/.

hits by profiles: [8 hits (by 6 distinct profiles) on 1 sequence]

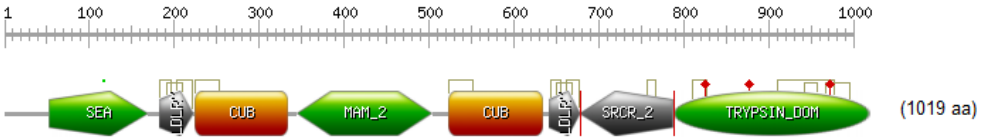
Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



RecName: Full=Enteropeptidase; EC=3.4.21.9; AltName: Full=Enterokinase; AltName: Full=Serine protease 7; AltName: Full=Transmembrane protease serine 15; Contains: RecName: Full=Enteropeptidase non-ca
Full=Enteropeptidase catalytic light chain; Flags: Precursor;. *Homo sapiens (Human)*

hits by profiles: [8 hits (by 6 distinct profiles) on 1 sequence]

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



RecName: Full=Enteropeptidase; EC=3.4.21.9; AltName: Full=Enterokinase; AltName: Full=Serine protease 7; AltName: Full=Transmembrane protease serine 15; Contains: RecName: Full=Enteropeptidase non-ca Full=Enteropeptidase catalytic light chain; Flags: Precursor;. *Homo sapiens (Human)*

PS50024 **SEA** *SEA domain profile* :

54 - 169: **score** = 32.979

QSHEARATFKITsGVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQ
FEN--GSIIVVFDLFFAQWVSD--ENVKEELIQGLEANKssQLVTFHIDLNSVDILDKLT

Predicted features:

DOMAIN	54	169	SEA	[condition: none]
SITE	117	118	Cleavage; by autolysis	[condition: none]

PS50068 **LDLRA_2** *LDL-receptor class A (LDLRA) domain profile* :

183 - 222: **score** = 10.75

ECLPGSSPCTDaLTCIKADLFCDGEVNCPDGSDEdnKMCA

Predicted features:

DOMAIN	183	222	LDL-receptor class A	[condition: none]
DISULFID	184	197		[condition: C-x*-C]
DISULFID	191	210		[condition: C-x*-C]
DISULFID	204	221		[condition: C-x*-C]

642 - 678: **score** = 13.3

PCKADHFQCKNGECVPLVNLCDGHLHCEDGSDEADCV

Predicted features:

DOMAIN	642	678	LDL-receptor class A	[condition: none]
DISULFID	643	655		[condition: C-x*-C]
DISULFID	650	668		[condition: C-x*-C]
DISULFID	662	677		[condition: C-x*-C]

PS01180 **CUB** *CUB domain profile* :

225 - 334: **score** = 13.293

CDGRFLLTgsSGSFQATHYPK---PSEtsVVCQWIIRVNQGL--SIKLSFDD-FNTY---
--YTDILDIYEGvgsSKILRASIWETNPGTIRIFSNQVTATFLIESDESDyVGFNATYTA
F

Predicted features:

				[condition: C-x*-C]

High-quality Automated and Manual Annotation of Proteins

HAMAP is a system for the classification and annotation of protein sequences. It consists of a collection of manually curated family profiles for protein classification, and associated, manually created annotation rules that specify annotations that apply to family members. HAMAP is used to annotate protein records in UniProtKB via UniProt's automatic annotation pipeline. We also provide an interface to scan user sequences against HAMAP family profiles [More...].

Search

Field: All

Term: Search Reset

e.g. recA, MF_00191, Iron
The wildcard * is supported.

Last updated
29-Apr-2015 for UniProt release 2015_05

Number of family profiles	2'015
Number of annotation rules	2'035

HAMAP-Scan

Quickly find matches of your protein sequences to HAMAP family profiles (max. 1'000 sequences). Examples

Enter UniProtKB accessions or identifiers or sequences in FASTA format

Scan Reset

More options

If you want to submit more than 1'000 sequences, or if you want to have your sequences annotated in UniProtKB format by using HAMAP annotation rules, use the

Advanced HAMAP-Scan submission form

Questions? Comments? Please send us your feedback.

"HAMAP in 2015: updates to the protein family classification and annotation system." Nucleic Acids Res. 43:D1064-D1070(2015).

Browse annotation rules

The HAMAP annotation rules are created by expert curators. They are written in the UniRule format and contain data which is used to annotate bacterial, archaeal and eukaryotic proteins in the UniProtKB format [More...].

List all rules | *or* **Browse by:** enzyme class | feature key | GO term | keyword | pathway | taxonomic scope

Name	Gene name	Rule	Profile	Matches
16SrRNA_methyltr_A	ksgA / rsmA	MF_00607	MF_00607	8336
16SrRNA_methyltr_B	rsmB / sun	MF_01856	MF_01856	822
16SrRNA_methyltr_C	rsmC	MF_01862	MF_01862	1756
16SrRNA_methyltr_F	rsmF	MF_01579	MF_01579	962
16SrRNA_methyltr_G	rsmG	MF_00074	MF_00074	7913
16SrRNA_methyltr_H	rsmH	MF_01007	MF_01007	8454
16SrRNA_methyltr_I	rsmI	MF_01877	MF_01877	7638
16SrRNA_methyltr_J	rsmJ	MF_01523	MF_01523	2102
23SrRNA_methyltr_B	rlmB	MF_01887	MF_01887	2319
23SrRNA_methyltr_Cfr	cfr	MF_01873	MF_01873	78
23SrRNA_methyltr_F	rlmF	MF_01848	MF_01848	1479
23SrRNA_methyltr_G	rlmG	MF_01859	MF_01859	1342
23SrRNA_methyltr_H	rlmH	MF_00658	MF_00658	6693
23SrRNA_methyltr_I	rlmI	MF_01857	MF_01857	862
23SrRNA_methyltr_J	rlmJ	MF_00934	MF_00934	2872

HAMAP-Scan

Classify (and annotate) your own protein sequences by using the collection of HAMAP family profiles and annotation rules [help].

=> Retrieve results of a previous scan.

STEP 1 - Enter PROTEIN sequences [?]

(•) Enter PROTEIN sequences (max. 1'000) Examples
or
( ) Upload a file (up to 100'000 sequences)

```
Enter UniProtKB accessions or identifiers or sequences in
FASTA format
```

Supported input:

- UniProtKB accessions e.g. Q8ZHG0 or identifiers e.g. AGUA_YERPE
- Amino acid sequences in FASTA format

STEP 2 - Choose 'Scan' or 'Scan & Annotate' [?]

( ) 'Scan' - Find matches for your sequence(s) to HAMAP family profiles. You will get a list of all (trusted and weak) matches of your sequences along with their match score.
or
(•) 'Scan & Annotate' - Have trusted matches of your sequence(s) to HAMAP family profiles annotated in UniProtKB format by the associated HAMAP annotation rules. To get the best annotation, all your sequences must originate from the same organism and you must indicate the corresponding taxonomic identifier (TaxID).

TaxID: * [ ] Enter the TaxID that represents the organism from which the sequences you're submitting originate from.

STEP 3 - Submit your job [?]

Email: * [ ] You will be notified by email once the job you submitted has completed (even if no match has been found). Your results will be available on our server for 1 month.

Job title: [ ] If available, the job title will be included in the subject of the email.

Password: [ ] If you enter a password, then the same password will be requested before you can download your results.

SUBMIT Reset

*Required field

HAMAP-Scan: ‘Scan’ output

HAMAP-Scan results

Showing results for 10 sequences

Download: [Tab-delimited] [Excel] | **Filter:** ☑ Trusted match ☑ Weak match (below cutoff) ☑ No match Reload

Your sequence [?]	Profile AC [?]	Profile name [?]	Trusted cutoff [?]	Match score [?]	Match region [?]	Match quality [?]
sp|B1LQ15|SYA_ECOSM (876 aa)	MF_00036_B	Ala_tRNA_synth_B	78.0	210.208	1-875	**Trusted**
sp|B1LQ15|SYA_ECOSM (876 aa)	MF_03134	Ala_tRNA_synth_plantC	29.724	20.217	2-876	Weak
sp|B1LQ15|SYA_ECOSM (876 aa)	MF_00036_A	Ala_tRNA_synth_A	18.492	10.689	1-868	Weak
sp|B1LI51|ACP_ECOSM (78 aa)	MF_01217	Acyl_carrier	13.362	22.327	3-77	**Trusted**
sp|B1LGN8|AROE_ECOSM (272 aa)	MF_00222	Shikimate_DH_AroE	26.377	33.106	1-269	**Trusted**
sp|B1LGN8|AROE_ECOSM (272 aa)	MF_01578	Shikimate_DH_YdiB	67.475	14.249	1-270	Weak
sp|B1LI10|MNMA_ECOSM (368 aa)	MF_00144	tRNA_thiouridyl_MnmA	31.753	40.978	6-361	**Trusted**
sp|B1LI10|MNMA_ECOSM (368 aa)	MF_01633	QueC	38.469	8.948	6-196	Weak
sp|B1LF99|LSRB_ECOSM (340 aa)	-	-	-	-	-	No match
tr|B1LFT9|B1LFT9_ECOSM (338 aa)	MF_00492	Transaldolase_1	45.973	52.598	23-338	**Trusted**
tr|B1LFT9|B1LFT9_ECOSM (338 aa)	MF_00496	F6P_aldolase	48.944	10.611	33-299	Weak
tr|B1LFT9|B1LFT9_ECOSM (338 aa)	MF_00494	Transaldolase_3b	42.738	15.62	33-295	Weak
tr|B1LDG3|B1LDG3_ECOSM (197 aa)	MF_01405	Non_canon_purine_NTPase	36.059	51.874	2-196	**Trusted**
tr|B1LDG3|B1LDG3_ECOSM (197 aa)	MF_03148	HAM1_NTPase	49.594	21.502	2-196	Weak
tr|B1LD69|B1LD69_ECOSM (375 aa)	MF_01899	RNase_D	35.982	46.667	5-371	**Trusted**
tr|B1LI94|B1LI94_ECOSM (367 aa)	MF_02121	ASADH	31.464	34.304	1-367	**Trusted**
tr|B1LKL8|B1LKL8_ECOSM (67 aa)	MF_00903	TatE	24.392	27.09	1-67	**Trusted**
tr|B1LKL8|B1LKL8_ECOSM (67 aa)	MF_00236	TatA_E	14.87	21.147	1-62	**Trusted**

HAMAP-Scan: ‘Scan & Annotate’

HA

Request code: LJY

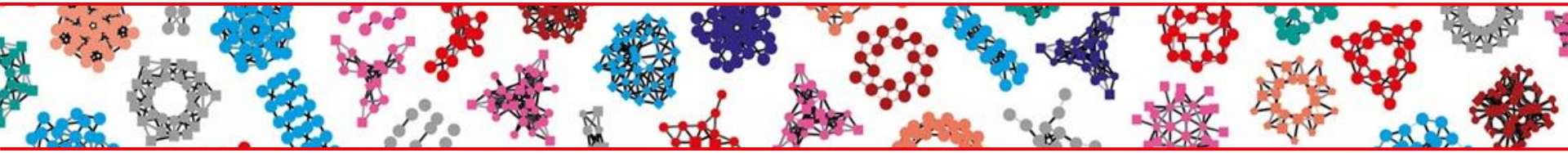
Status: Completed

Input file: 4305 entries

Result files: Annotated matches

"Not annotated entries which app

```
ID   CYSC_ECOLX              Unreviewed;       201 AA.
AC   3LJY468;
DE   RecName: Full=Adenylyl-sulfate kinase;
DE            EC=2.7.1.25;
DE   AltName: Full=APS kinase;
DE   AltName: Full=ATP adenosine-5'-phosphosulfate 3'-phosphotransferase;
DE   AltName: Full=Adenosine-5'-phosphosulfate kinase;
GN   Name=cysC;
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
CC   -!- FUNCTION: Catalyzes the synthesis of activated sulfate.
CC   -!- CATALYTIC ACTIVITY: ATP + adenylyl sulfate = ADP + 3'-
CC       phosphoadenylyl sulfate.
CC   -!- PATHWAY: Sulfur metabolism; hydrogen sulfide biosynthesis; sulfite
CC       from sulfate: step 2/3.
CC   -!- SIMILARITY: Belongs to the APS kinase family.
DR   UniPathway; UPA00140; UER00205.
DR   GO; GO:0004020; F:adenylylsulfate kinase activity; IEA:HAMAP.
DR   GO; GO:0005524; F:ATP binding; IEA:HAMAP.
DR   GO; GO:0000103; P:sulfate assimilation; IEA:HAMAP.
DR   HAMAP; MF_00065; Adenylyl_sulf_kinase; 1.
KW   ATP-binding; Kinase; Nucleotide-binding; Phosphoprotein; Transferase.
FT   NP_BIND      35     42       ATP.
FT   ACT_SITE    109    109       Phosphoserine intermediate.
**
**   ##################    INTERNAL SECTION    ##################
**HA FAM; Method MF_00065; CYSC; Trusted match; 38.735 (+7.3).
**HA SAM; Annotated by praise 1.9.4; MF_00065.30; MF_00065; 06-JAN-2014 11:07:03.
SQ   SEQUENCE   201 AA;  22321 MW;  11E15BB8F9D2FD4B CRC64;
     MALHDENVVW HSHPVTVQQR ELHHGHRGVV LWFTGLSGSG KSTVAGALEE ALHKLGVSTY
     LLDGDNVRHG LCSDLGFSDA DRKENIRRVG EVANLMVEAG LVVLTAFISP HRAERQMVRE
     RVGEGRFIEV FVDTPLAICE ARDPKGLYKK ARAGELRNFT GIDSVYEAPE SAEIHLNGEQ
     LVTNLVQQLL DLLRQNDIIR S
//
```



- Why do we need predictive annotation tools?
- Protein signatures for homology detection – A short primer
- Annotation rules for functional annotation
- HAMAP and PROSITE - automatic annotation in UniProtKB
- HAMAP and PROSITE - services for external users
- **Practical exercises**