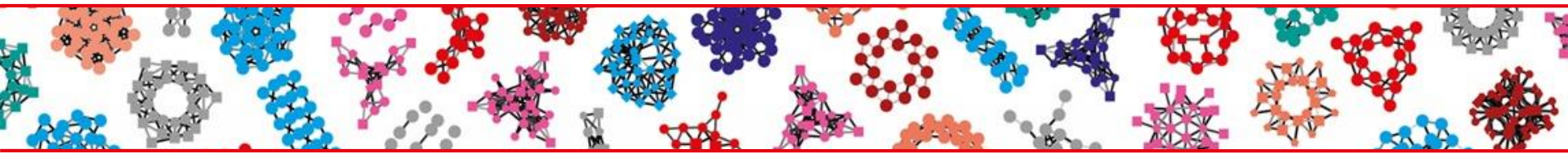


Swiss Institute of
Bioinformatics

PROSITE, HAMAP, InterPro and automated annotation in UniProtKB

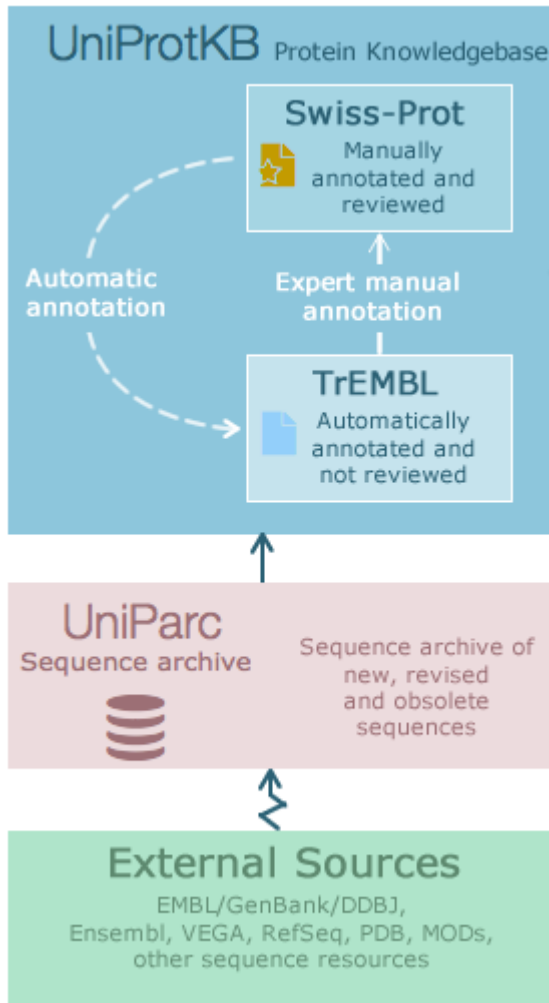
Protein Bioinformatics: Sequence-Structure-Function

Basel 14 November 2018



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

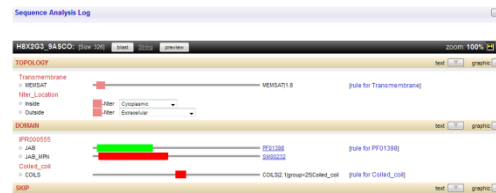
The UniProt knowledgebase UniProtKB



Literature curation



Curator-supervised
sequence analysis



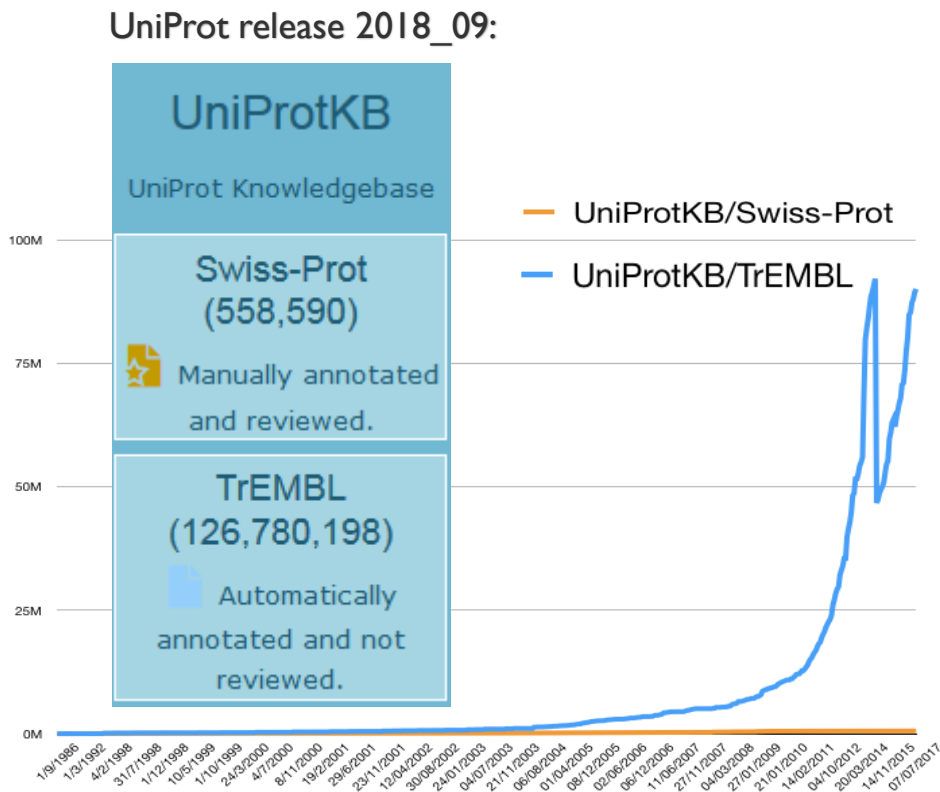
- Motif database scans: Prosite, Pfam, Smart, etc.
- Prediction programs: TMHMM, coils, SignalP, TargetP, Repeat, etc.

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>



- 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

Characterized template entry




```

ADE_YEAST      -----MVSVEFLQELPKCEHLLHLEGTLEPDLLFPLAKRNDIILPE---GFPKSV
ADE_CANGA      -----MVPESFLELPKCEHLLHLEGTLEPDLLFPLAKRNNIQLPD---HFPQTP
ADE_KLULA      --MAKFECTDEVNTNFLTLPKCEHLLHLEGTLEPELLFQLVERNGVQLPG---TFPKTV
ADE_CANAL      --MAGYECSEHMENFLRELKCEHVVHLEGTLEPSLLFKLAKRNNITLPE---TFPKTV
ADE_ASPFU      MC-----QSPLHDFLHGLPKCEHVVHLEGGVTPPELIFQLAKNNIQLPNATHPAYASV
ADE_ASPOR      MC-----KSDLHDFLHGLPKCEHVVHLEGGCLAPDLIFELAKRNNVSLFN---EPAYESI
ADE_EMENI      MCPFNTPYQSQWHAFLHSLPKCEHVVHLEGGLEPPLIFSMARKNNVSLFRSSNPAYTSV
ADE_SCHPO      MS-----NLPITYNFIRKLPKCEHVVHLEGGCLSDLVFRLAKKNGITLPRDAAATTP
ADE           -----ODAAATTP-----PVPYESA
ADE           -----PVPYESA-----SI
ADE           -----SI-----SI
ADE_RHORT      -----MAVDPAFLLHALPKVELLHLHIEGSLPEMVMVALAERNGLRLEP-----SV
ADE_STRCO      -----MKRPYDALMPLPKAELHLHIIEGTLPELAFALAARNGVSLP-----DE
ADE_BURFP      -MTTITVTPTPLAECTALAPKAELHLHIIEGSLPELIFALAERNGVKLAYD-----SI
ADE_BURXL      -MTTITVTSTPLAECTVLAPKAELHLHIIEGSLPELIFALAERNGVKLAYD-----SI
ADE_CUPTR      ----MTIDAALAEQIRRTPKAELHVHIIEGTLPELIFRLAQRNQVALPYP-----SV
ADE_CUPNH      ----MTIDAALAEQIRRTPKAELHVHIIEGTLPELIFRLAQRNQVALPYP-----SV
ADE_CUPPJ      ----MTIDAALADKIRRTPKAELHVHIIEGTLPELIFRLAQRNNVNLPP-----SV
ADE_RALME      ----MTIDAALADKIRRTPKAELHVHIIEGTLPELIFRLAQRNNVKLAYP-----DV
ADE_RALPJ      ----MPISSALAERIATSPKAELHLHIIEGSLPELMFALAERNGVKLPYA-----SV
ADE_RALSO      ----MPISSPALAERIATSPKAELHLHIIEGSLPELMFALAERNGVKLPYA-----SV
ADE_GEOLS      -MNLTNIPRQALPELLCRMPKAELHLHIIEGSLPELIFALAERNRLQLAYP-----TI
ADE_GBOUR      -MNFDCIPREDLHGILCHMPKAELHLHIIEGSLPELIFELATRNNIQLPYP-----TI
    
```

Propagation by sequence similarity

Uncharacterized protein family members in other species



MT2ID_HUMAN

Functionⁱ

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

Sites

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Binding site ⁱ	43 – 43	1	S-adenosyl-L-methionine; via amide nitrogen 1 Publication		
Binding site ⁱ	96 – 96	1	S-adenosyl-L-methionine 1 Publication		
Binding site ⁱ	126 – 126	1	S-adenosyl-L-methionine; via amide nitrogen 1 Publication		

MT2ID_MOUSE

Functionⁱ

Protein-lysine N-methyltransferase that specifically trimethylates 'Lys-315' of VCP/p97; this modification may decrease VCP ATPase activity. [By similarity](#)

Sites

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Binding site ⁱ	43 – 43	1	S-adenosyl-L-methionine; via amide nitrogen By similarity		
Binding site ⁱ	96 – 96	1	S-adenosyl-L-methionine By similarity		
Binding site ⁱ	126 – 126	1	S-adenosyl-L-methionine; via amide nitrogen By similarity		

Microbial proteome annotation

Statistics

Species	Species code	UniProtKB proteome identifier	Number of entries in UniProtKB		
			All	Swiss-Prot	TrEMBL
<i>Acaryochloris marina</i> (strain MBIC 11017)	ACAM1	UP000000268	8172	322	7850
<i>Accumulibacter phosphatis</i> (strain UW-1)	ACCPU	UP000001619	4438	1	4437
<i>Acetobacter pasteurianus</i> (strain NBRC 3283 / LMG 1513 / CCTM 1153)	ACEP3	UP000000948	2906	4	2902
<i>Acetobacter</i> sp. CAG:267	/	UP000017992	1538	0	1538
<i>Acetobacter</i> sp. CAG:977	/	UP000018259	1775	0	1775
<i>Acetobacterium woodii</i> (strain ATCC 29683 / DSM 1030 / JCM 2381 / KCTC 1655)	ACEWD	UP000007177	3445	4	3441

<i>Escherichia coli</i> (strain K12)	ECOLI	UP000000625	4305	4305	0
--------------------------------------	-------	-------------	------	------	---

<i>Acholeplasma laidlawii</i> (strain PG-8A)	ACHLI	UP000008558	1380	147	1233
--	-------	-------------	------	-----	------

<i>Bacillus subtilis</i> (strain 168)	BACSU	UP000001570	4197	4185	12
---------------------------------------	-------	-------------	------	------	----

<i>Acholeplasma</i> sp. CAG:878	/	UP000017930	1145	0	1145
---------------------------------	---	-------------	------	---	------

<i>Mycobacterium tuberculosis</i> (strain ATCC 25618 / H37Rv)	MYCTU	UP000001584	3991	2081	1910
---	-------	-------------	------	------	------

<i>Acidaminococcus fermentans</i> (strain ATCC 25085 / DSM 20731 / VR4)	ACIFV	UP000001902	2016	12	2004
---	-------	-------------	------	----	------

<i>Acidaminococcus intestini</i> (strain RyC-MR95)	ACIIR	UP000007093	2386	0	2386
--	-------	-------------	------	---	------

<i>Acidaminococcus</i> sp. CAG:917	/	UP000018048	1276	0	1276
------------------------------------	---	-------------	------	---	------

<i>Acidimicrobium ferrooxidans</i> (strain DSM 10331 / JCM 15462 / NBRC 103882 / ICP)	ACIFD	UP000000771	1935	8	1927
---	-------	-------------	------	---	------

<i>Acidiphilium cryptum</i> (strain JF-5)	ACICJ	UP000000245	3521	262	3259
---	-------	-------------	------	-----	------

<i>Acidiphilium</i> sp. CAG:727	/	UP000018053	1479	0	1479
---------------------------------	---	-------------	------	---	------

Annotation propagation in UniProtKB/Swiss-Prot

Characterized template entry




```

ADE_YEAST      -----MVSVEFLQELPKCEHLLHLEGTLEPDLLFPLAKRNDIILPE---GFPKSV
ADE_CANGA     -----MVPESFLELPKCEHLLHLEGTLEPDLLFPLAKRNNIQLPD---HFPQTP
ADE_KLULA     --MAKFECTDEVNTNFLTLPKCEHLLHLEGTLEPELLFQLVERNGVQLPG---TFPKTV
ADE_CANAL     --MAGYECSEHMENFLRELKPEHHVHLEGTLEPSLLFKLAKRNNITLPE---TFPKTV
ADE_ASPFU     MC-----QSPLHDFLHGLPKCEHHVHLEGGVTPPELIFQLAKNNIQLPNATHPAYASV
ADE_ASPOR     MC-----KSDLHDFLHGLPKCEHHVHLEGGCLAPDLIFELAKRNNVSLFN---EPAYESI
ADE_EMENI     MCPFNTPYQSQWHAFLHSLPKCEHHVHLEGGLEPPLIFSMARKNNVSLFRSSNPAYTSV
ADE_SCHPO     MS-----NLPYINFRIRLKPCEHHVHLEGGCLSDLVFRLAKKNGITLPRDAAATTP
ADE           -----ODAAATTP-----SV
ADE           -----ODAAATTP-----SV
ADE           -----ODAAATTP-----SI
ADE_RHORT     -----MAVDPAFLHALPKVELLHLHIEGSLEPEMVALAERNGLRLEP-----SV
ADE_STRCO     -----MKRPYDALMPLPKAELHLHIIEGTLPELAFALAARNGVSLP-----DE
ADE_BURFP     -MTTIVTPTPLAECTALAPKAELHLHIIEGSLEPELIFALAERNQVKLAYD-----SI
ADE_BURXL     -MTTIVTSTPLAECTVLAAPKAELHLHIIEGSLEPELIFALAERNQVKLAYD-----SI
ADE_CUPTR     -----MTIDAALAEQIRRTPKAELHVHIIEGTLPELIFRLAQRNQVALPYP-----SV
ADE_CUPNH     -----MTIDAALAEQIRRTPKAELHVHIIEGTLPELIFRLAQRNQVALPYP-----SV
ADE_CUPPJ     -----MTIDAALADKIRRTPKAELHVHIIEGTLPELIFRLAQRNHNVLNLPY-----SV
ADE_RALME     -----MTIDAALADKIRRTPKAELHVHIIEGTLPELIFRLAQRNNVKLAYP-----DV
ADE_RALPJ     -----MPISSALAERIATSPKAELHLHIIEGSLEPELMFALAERNQVKLPYA-----SV
ADE_RALSO     -----MPISSPALAERIATSPKAELHLHIIEGSLEPELMFALAERNQVKLPYA-----SV
ADE_GBOLS     -MNLINIPRQALPELLCRMPKAELHLHIIEGSLEPELIFALAERNRLQLAYP-----TI
ADE_GBOUT     -MNFDCIPREDLHGILCHMPKAELHLHIIEGSLEPELIFELATRNNIQLPYP-----TI
    
```

Propagation by sequence similarity

Uncharacterized protein family members in other species



MT2ID_HUMAN

Functionⁱ

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

Sites

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Binding site ⁱ	43 – 43	1	S-adenosyl-L-methionine; via amide nitrogen 1 Publication		
Binding site ⁱ	96 – 96	1	S-adenosyl-L-methionine 1 Publication		
Binding site ⁱ	126 – 126	1	S-adenosyl-L-methionine; via amide nitrogen 1 Publication		

MT2ID_MOUSE

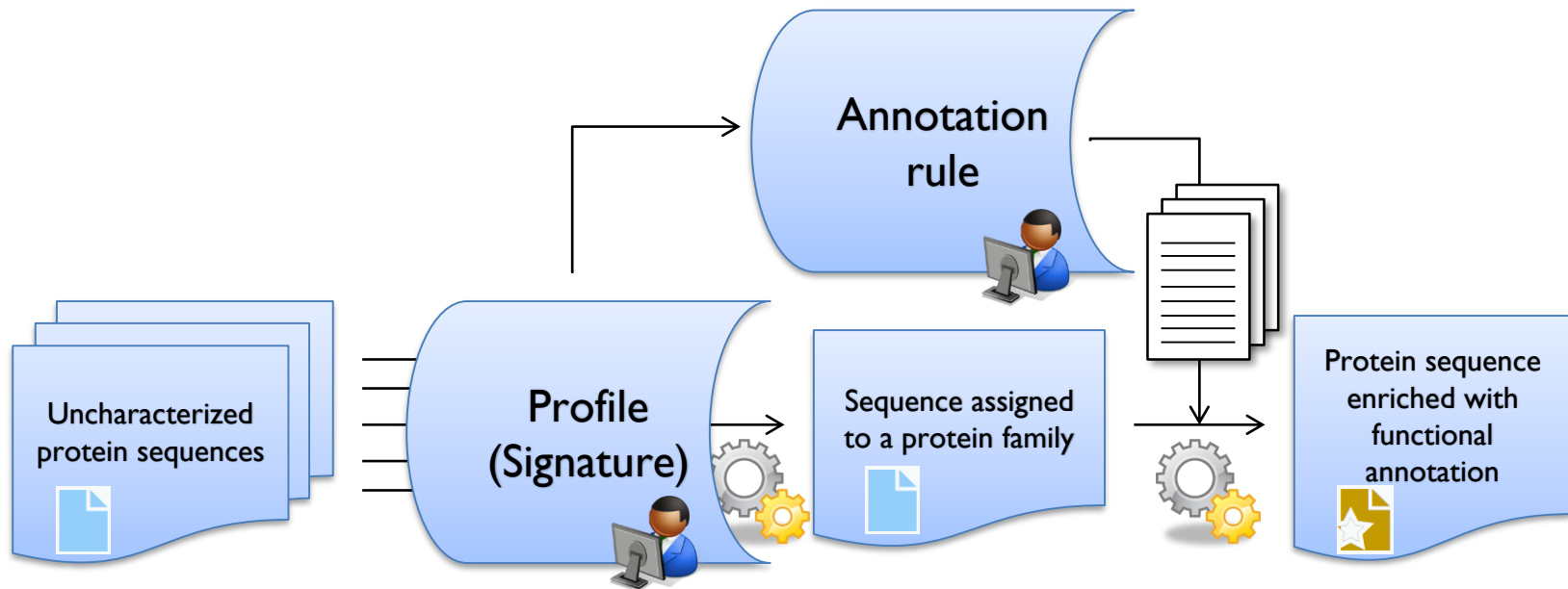
Functionⁱ

Protein-lysine N-methyltransferase that specifically trimethylates 'Lys-315' of VCP/p97; this modification may decrease VCP ATPase activity. [By similarity](#)

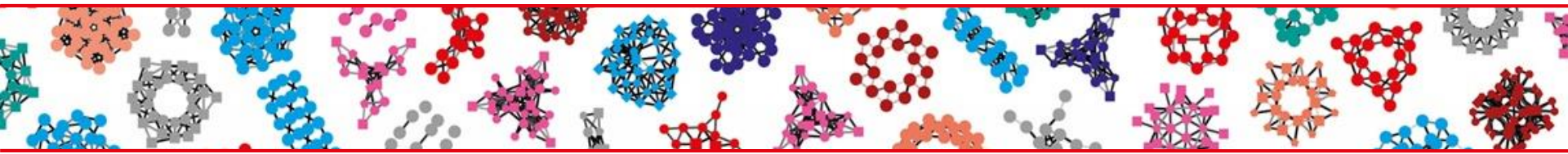
Sites

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Binding site ⁱ	43 – 43	1	S-adenosyl-L-methionine; via amide nitrogen By similarity		
Binding site ⁱ	96 – 96	1	S-adenosyl-L-methionine By similarity		
Binding site ⁱ	126 – 126	1	S-adenosyl-L-methionine; via amide nitrogen By similarity		

Automated 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

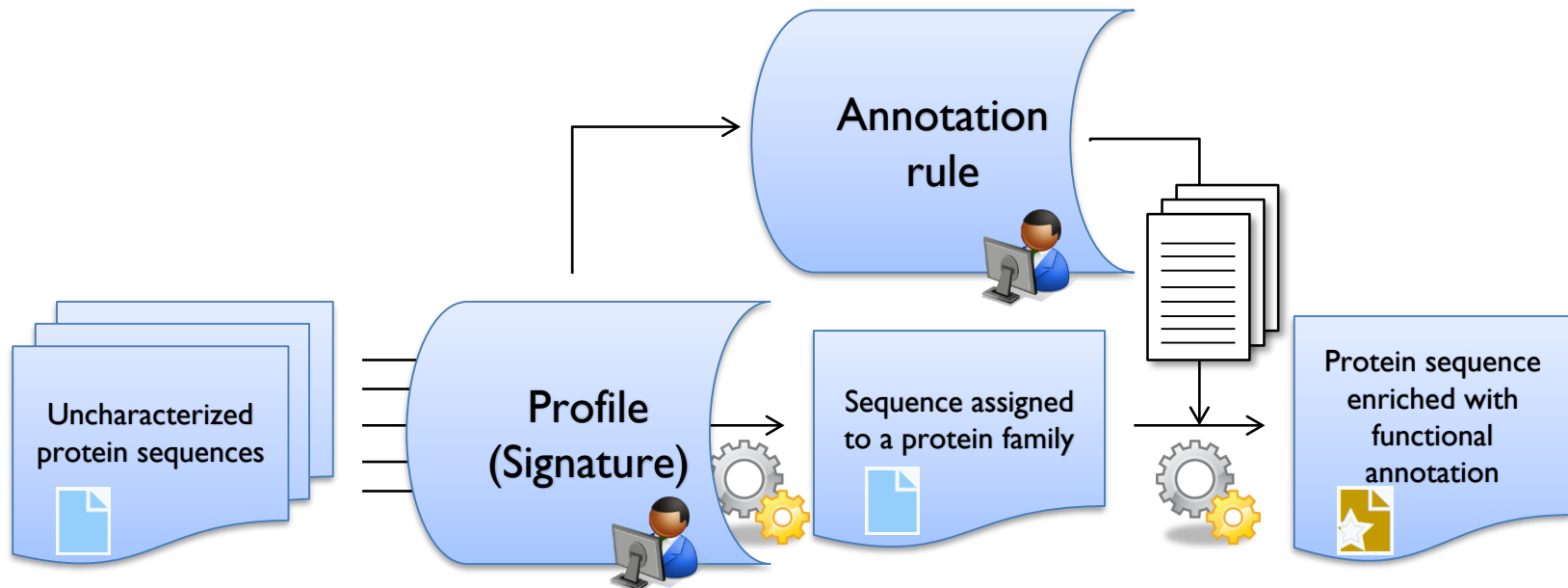


- Why do we need predictive annotation tools?
- **Protein signatures for homology detection – A short primer**
- Annotation rules for functional annotation
- HAMAP and PROSITE - automatic annotation in UniProtKB
- HAMAP and PROSITE - services for external users
- Practical exercises

Patterns and profiles

over to Christian....

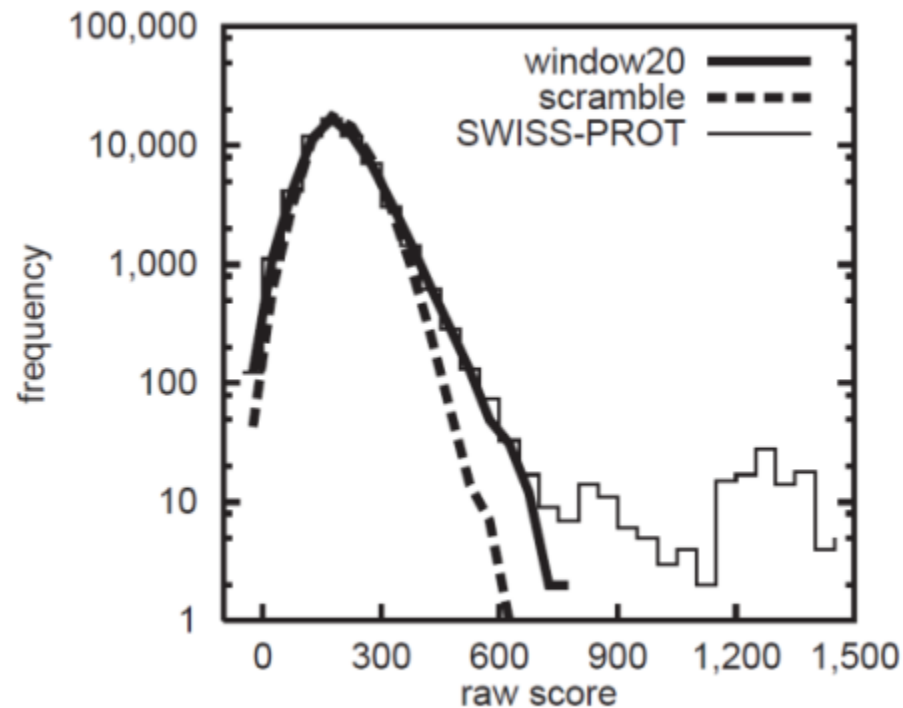
Automatic annotation of protein sequences



- **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**

Profiles: interpretation of the score

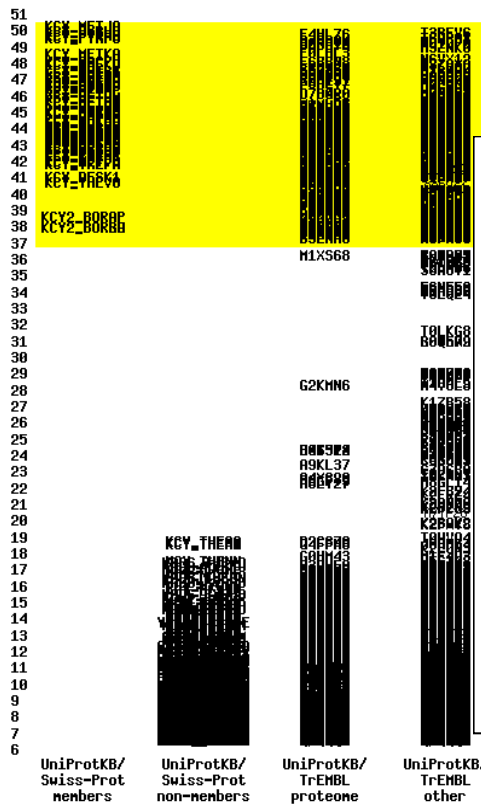
- How do I interpret the score produced by a profile? Which is the lowest score I consider to produce a true match?
- Only biological arguments tell you if a match is true or not.



Validation of a family profile



Score distribution of profile MF_00239 matches in all kingdoms of UniProtKB



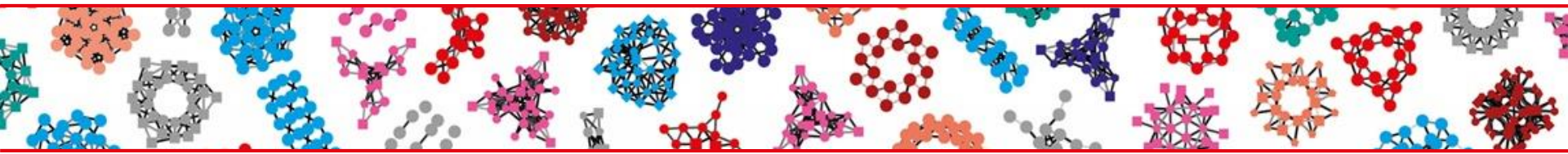
List of UniProtKB/Swiss-Prot true positive matches for the profile MF_00239 (i.e. HAMAP MF_00239 UniProtKB/Swiss-Prot members)

ac	id	kingdom	score	score_diff	description	organism
Q58071	KCY_METIA	Archaea	50.213	+12.972	Cytidylate kinase	Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440)

```

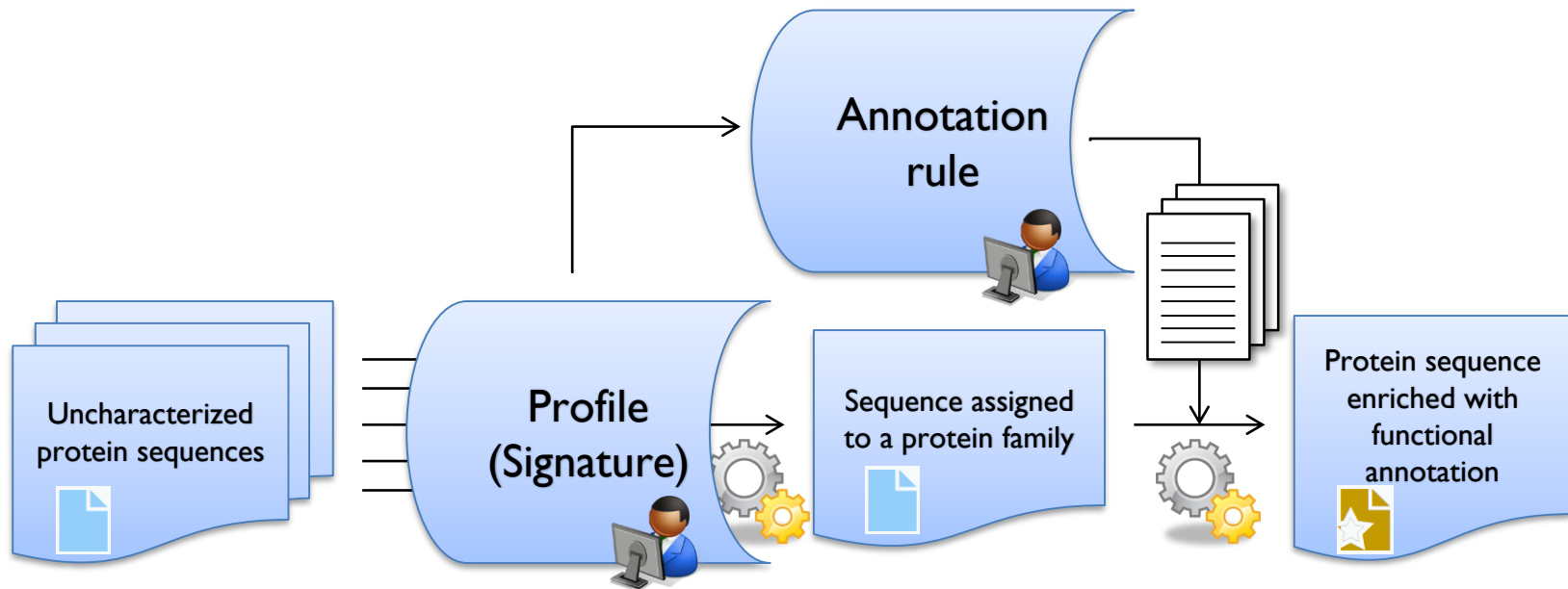
ID Cytidyl_kinase_type2; MATRIX.
AC MF_00239;
DT DEC-2013 (DATA UPDATE).
DE Cytidylate kinase [cmk].
CC /VERSION=4;
MA /GENERAL_SPEC: ALPHABET='ACDEFGHIKLMNPQRSTVWY'; LENGTH=191; LOG_BASE=1.071779; P0=0.9972;
MA P= 7.552363, 1.698108, 5.303439, 6.320015, 4.078187, 6.844419, 2.240667,
MA /DISJOINT: DEFINITION=PROTECT; N1=1; N2=1;
MA /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=17.3045; R2=0.0080365; TEXT='-LogE';
MA /CUT_OFF: LEVEL=1; SCORE=25.4; N_SCORE=37.241; MODE=1; TEXT='!';
MA /CUT_OFF: LEVEL=0; SCORE=25.4; N_SCORE=37.241; MODE=1; TEXT='?';
MA /CUT_OFF: LEVEL=-1; SCORE=-100; N_SCORE=8.5; MODE=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;
    
```

Q5IMV8	KCY_NAIPD	Archaea	47.159	+9.918	kinase	2160)
C6A187	KCY_THESM	Archaea	47.159	+9.918	Cytidylate kinase	Thermococcus sibiricus (strain MM 739 / DSM 12597)



- 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

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

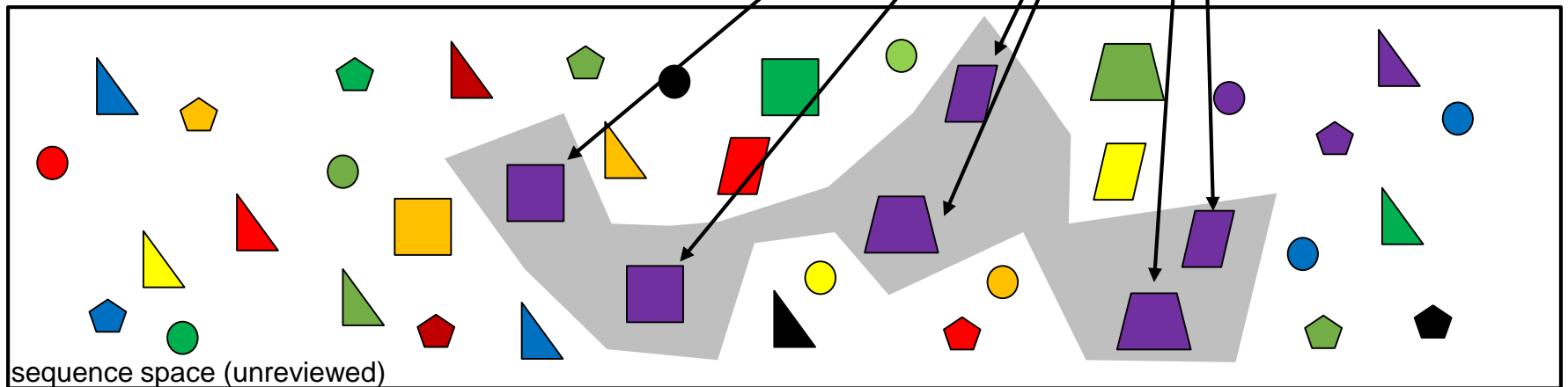
A fictitious rule

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...

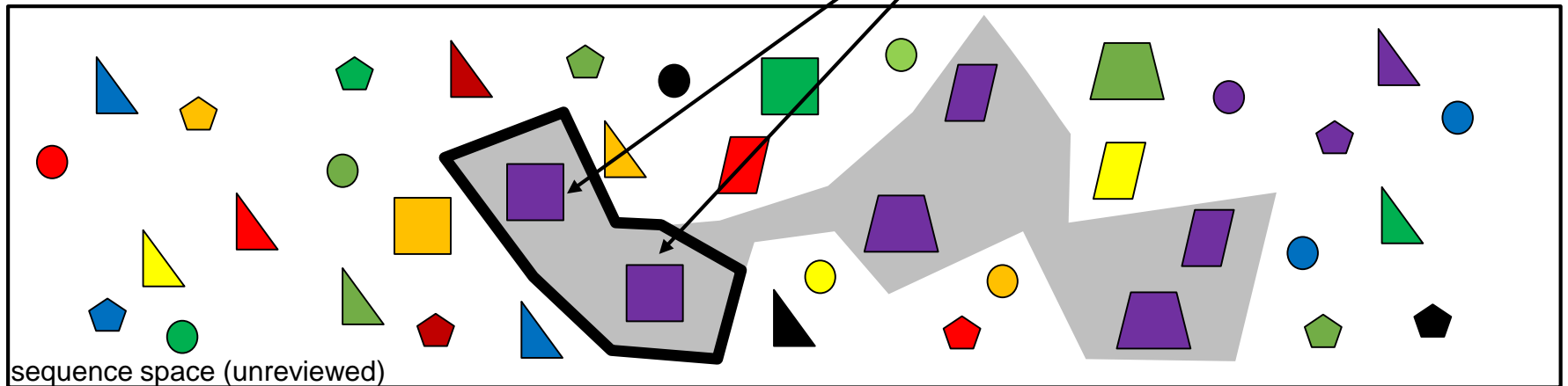
- color: purple
 - has four edges
- and:
- all sides of same length
 - all angles are 90 degrees

... then these annotations are applied

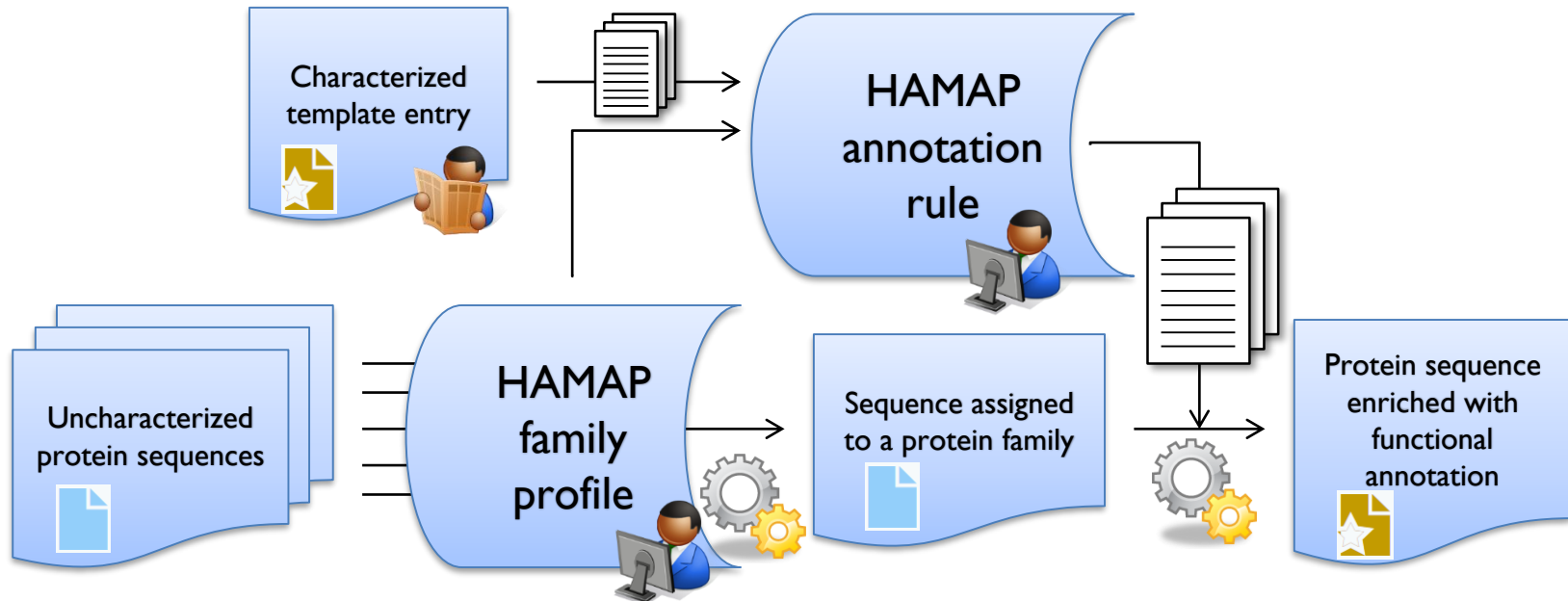
- is a purple quadrilateral

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.

What can be propagated?

Annotation	Propagated?
RecName	Yes
AltName	Yes
Function	Yes
Catalytic activity	Yes
Pathway	Yes
Subunit	Yes
Subcellular location	Yes
Disease	No
Disruption phenotype	No
Polymorphism	No
Alternative products	No

Feature annotation

Annotation	Propagated?
KW	Yes
GO	Yes
Regions of interest	Yes
Active site	Yes
Ligand-binding	Yes
Processing	Yes
PTMs	Yes
Ambiguities	No
Conflicts	No
Natural variants	No
Isoforms	No



Annotation rule MF_00041

[Send feedback](#)

General rule information [?]

Accession	MF_00041
Dates	1-JUN-2001 (Created) 14-NOV-2014 (Last updated, Version)
Name	Cys_tRNA_synth
Scope	Bacteria Archaea
Template	P21888 (SYC_ECOLI)
Triggered by	HAMAP; MF_00041 (Get profile general information and sta)

Annotation applies only to sequences from specified taxa

Annotation applies to sequences matching specified profile

Propagated annotation [?]

Identifier, protein and gene names [?]

Identifier	SYC
Protein name	<p>RecName: Full=Cysteine--tRNA ligase; EC=6.1.1.16;</p> <p>AltName: Full=CysteinyI-tRNA synthetase; Short=CysRS;</p>
Gene name	cysS

A meaningful recommended name which can be safely propagated to orthologs; synonyms



Annotation rule MF_00041

[Send feedback](#)

Propagated annotation [?]

Comments [?]

Catalytic activity	ATP + L-cysteine + tRNA(Cys) = AMP + diphosphate + L-cysteinyl-tRNA(Cys).
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.



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

Keywords [?]

Cytoplasm	<p>Case statements restrict annotation to proteins satisfying these conditions</p>
Aminoacyl-tRNA synthetase	
Protein biosynthesis	
Ligase	
ATP-binding	
Nucleotide-binding	
case <FTGroup:1>	
Metal-binding	
Zinc	
end case	
case <FT:8>	<p>Summary in the form of keywords</p>
Phosphoprotein	
end case	





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 (Q916Y4)

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) ID C4ZN08_THASP Unreviewed; 343 AA.
FT METAL 14 14 Zinc; catalytic. AC C4ZN08;
FT Group: 1; Condition: H
FT METAL 16 16 Zinc; catalytic.
FT Group: 1; Condition: H
FT METAL 194 194 Zinc; catalytic.
FT Group: 1; Condition: H
FT METAL 275 275 Zinc; catalytic.
FT Group: 1; Condition: D
FT METAL 16 16 Zinc; catalytic.
FT METAL 18 18 Zinc; catalytic.
FT METAL 196 196 Zinc; catalytic.
FT METAL 277 277 Zinc; catalytic.
```

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

```
C4ZN08_THASP -----MELEAYVRALPKAEIHHIEGTLPEPMMFALARRNGVALPWA-----SV
ADE_YEAST -----MVSVEFLQELPKCEHHHLEGTLPEDLLFPPLAKRNDIILPE----GFPKSV
ADE_CANGA -----MVPESFLELPLKCEHHHLEGTLPEDLLFPPLAKRNNIQLPD----HFQPT
ADE_KLULA ---MAKFECTDEVTFNLTLPKCEHHHLEGTLPELLFQLVERNGVQLPG----TFPKTV
ADE_CANAL ---MAQYECSEHMENFLRELPKCEHHHLEGTLPESSLFKLAKRNNITLPE----TFPKTV
ADE_ASPFU MC-----QSPLDHDFLHGLPKCEHHHLEGCVTPELIFQLAEKNNIQLPNPATHPAYASV
ADE_ASPOR MC-----KSDLHDFLHGLPKCEHHHLEGCLAPDLIFELAKRNNVSLPN---EPAYESI
ADE_EMENI MCPPNTPYQSQWHAFHLHSLPKCEHHHLEGCLEPPLIFSMARKNNVSLPSPSSNPAYTSV
ADE_SCHPO MS-----NLPIYNFIRKLPKCEHHHLEGCLSPDLVFRLAKKNGITLPS--DDAAYTTP
ADE_GIBZE MC-----KSRVHSFLQALPKVECHHHIEGTLPELLFTLAEKNGIELPN---DPVYESA
ADE_CAUCR -MTDASFAPSASAEFVRGLPKAEIHHHIEGSLEPELMFELAQRNGITLPPA-----SV
ADE_CAUCN -MTDASFAPSASAEFVRGLPKAEIHHHIEGSLEPELMFELAQRNGITLPPA-----SV
ADE_SPHAL -MPDGFASHEERAFAIAGLPKAEIHHHIEGSLEPELLFEFARRNRVAIPFA-----SI
ADE_RHORT -----MAVDPAFLHALPKVEIHHHIEGSLEPEMMVALAERNGLRLPYA-----SV
ADE_STRCO -----MKRPYDALMPLPKAEIHHHIEGTLPELAFALAARNVSLPYA-----DE
ADE_BURPP -MTTTTPTPLAECTALAPKAEIHHHIEGSLEPELIFALAERNVVKLAYD-----SI
ADE_BURXL -MTTTTPTSTPLAECTVLPKAEIHHHIEGSLEPELIFALAERNVVKLAYD-----SI
ADE_CUPTR -----MTIDAALAEQIRRTPKAEIHHHIEGTLPELIFRLAQRNQVALPYP-----SV
ADE_CUPNH -----MTIDAALAEQIRRTPKAEIHHHIEGTLPELIFRLAQRNQVALPYP-----SV
ADE_CUPPJ -----MTIDAALADKIRRTPKAEIHHHIEGTLPELIFRLAQRNHVNLVYP-----SV
ADE_RALME -----MTIDAALADKIRRTPKAEIHHHIEGTLPEPERIFRLAQRNNVVKLAYP-----DV
ADE_RALPJ -----MPISSALAERIATSPKAEIHHHIEGSLEPELMFALAERNVVKLAYP-----SV
ADE_RALSO -----MPISSALAERIATSPKAEIHHHIEGSLEPELMFALAERNVVKLAYP-----SV
ADE_GEOLS -MNLNTNIPRQALPELLCRMPKAEIHHHIEGSLEPELIFALAERNRLQLAYP-----TI
ADE_GEOUR -MNFDCIPREDLHGILCHMPKAEIHHHIEGSLEPELIFELATRNRIQLPYP-----TI
ADE_RHOFD -MTIKPVSQERLPELLRTIPKAEIHHHIEGSLEPELMFALAQRNGVSIYP-----DV
ADE_ZYMMO -----MNLKIFIAALPKAEIHHHIEGSLEPELMFELAKRNVTLFPF-----DV
ADE_ACIAA -----MNQSELIRALPKAEIHHHIEGTFPELMFEIQRNHIDIPYK-----SV
ADE_PSYCK -----MIDLIKRLPKAEIHHHIEGSLEPELMFRLAKKNQIEIPYK-----DI
ADE_VIBPA -----MNAFQQGLPKVEIHHHIEGSLEPELMFKLAKRNGIDIPYS-----SP
ADE_PSEAE -----MYEWLNALPKAEIHHHIEGTLPELLFALAERNRIALPWN-----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 adenosine and AMP deaminases family. Adenine deaminase type 2 subfamily.

Keywords [?]

[Hydrolase](#)

[Nucleotide metabolism](#)

case <FTGroup:1>

[Metal-binding](#)

[Zinc](#)

end case

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

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 rule MF_03125

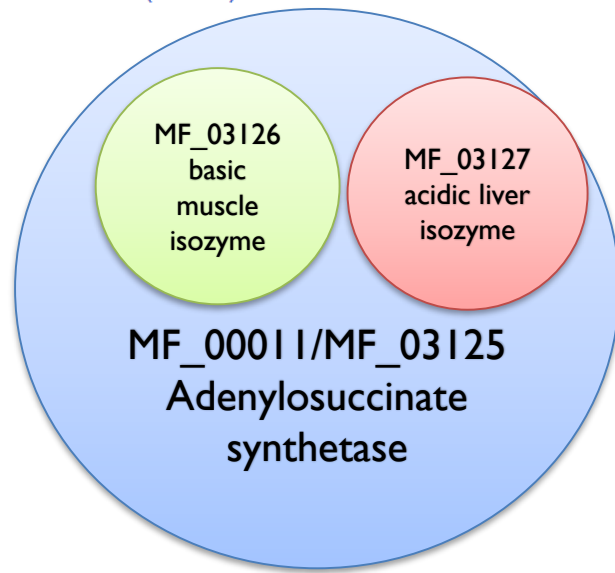
[Send feedback](#)

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





Overview

[Proteins matched](#) (31019)

[Domain architectures](#) (41)

[Pathways & interactions](#)

[Species](#)

[Structures](#)

[Literature](#) (7)

[Cross-references](#) (5)

F Family

Adenylosuccinate synthetase (IPR001114)

Short name: Adenylosuccinate_synthetase

Overlapping homologous superfamilies i

[H P-loop containing nucleoside triphosphate hydrolase](#) (IPR027417)

Family relationships

[F Adenylosuccinate synthetase](#) (IPR001114)

[F Adenylosuccinate synthetase isozyme 1, chordates](#) (IPR027509)

[F 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 γ -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](#) i

[TIGR00184](#) (purA)

[Pfam](#) i

[PF00709](#)

(Adenylsucc_synt)

[CDD](#) i

[cd03108](#) (AdSS)

[PANTHER](#) i

[PTHR11846](#)

(PTHR11846)

[SMART](#) i

[SM00788](#)

(Adenylsucc_synt)

[HAMAP](#) i

[MF_00011](#)

(Adenylosucc_synt)



Overview

[Proteins matched](#) (93)

[Domain architectures](#) (1)

[Pathways & interactions](#)

[Species](#)

[Structures](#)

[Literature](#) (1)

[Cross-references](#) (1)

Family

Adenylosuccinate synthetase isozyme 1, chordates (IPR027509)

Short name: AdSS_1_vert

Overlapping homologous superfamilies ⓘ

ⓘ [P-loop containing nucleoside triphosphate hydrolase \(IPR027417\)](#)

Family relationships

ⓘ [Adenylosuccinate synthetase \(IPR001114\)](#)

ⓘ **[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

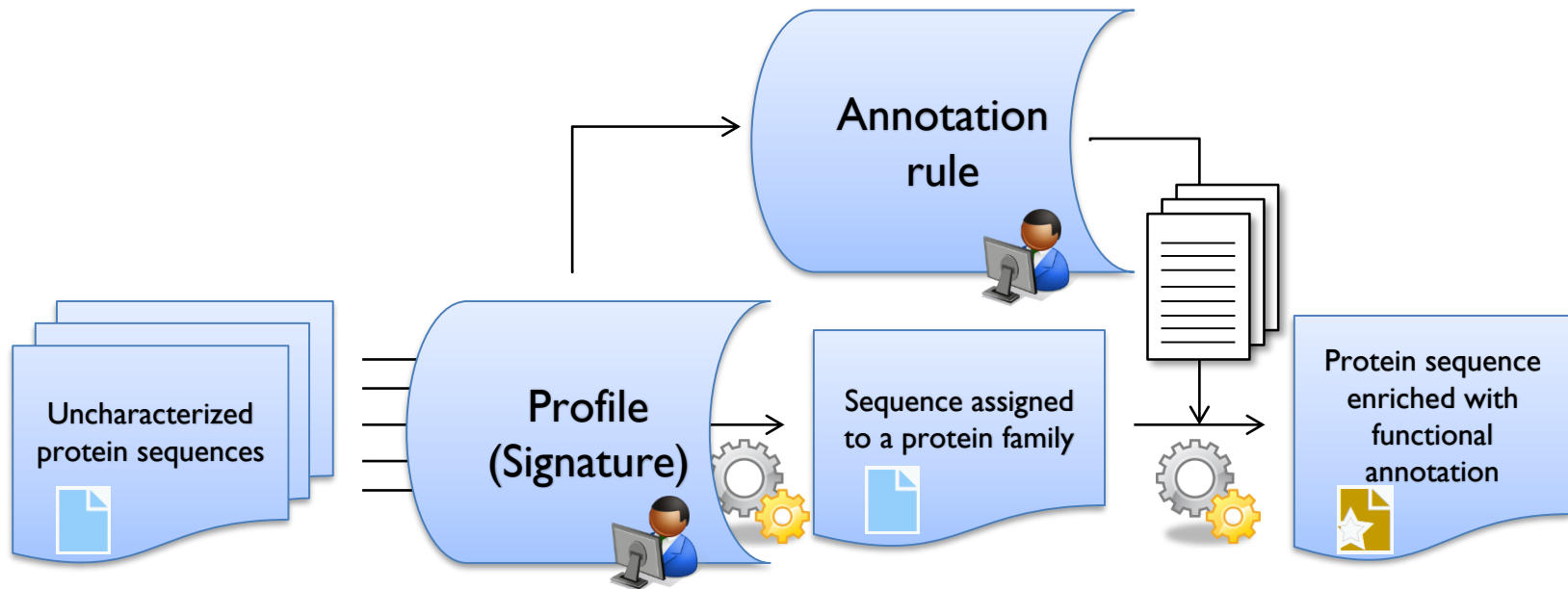
Signatures from InterPro member databases are used to construct an entry.

■ **HAMAP** ⓘ

[MF_03126](#)

(Adenylosucc_synth_vert_basic)

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

Annotation output

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

```
ID  ADE_THASP                Unreviewed;      343 AA.
AC  C4ZN08;
..
DR  HAMAP; MF_01962; Adenine_deaminase; 1.
KW  Hydrolase; Metal-binding; Nucleotide metabolism; Zinc.
FT  ACT_SITE    199    199    Proton donor.
FT  METAL       16     16     Zinc; catalytic.
FT  METAL       18     18     Zinc; catalytic.
FT  METAL       196    196    Zinc; catalytic.
FT  METAL       277    277    Zinc; catalytic.
FT  BINDING     278    278    Substrate.
FT  SITE        220    220    Important for catalytic activity.
**
**  ##### INTERNAL SECTION #####
**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 3.1.3; MF_01962.6; MF_01962; 18-MAR-2016 15:37:54.
SQ  SEQUENCE    343 AA;  37571 MW;  E2DA4A77193E5C4B CRC64;
    MELEAYVRAL PKAELHLHIE GTLEPEMMFA LARRNGVALP WASVEAVRAA YAFTDLQSFL
    DLYYAGAAVL VREQDFFELE FAYFERAHAD GVVHAELFFD PQTHTARGVA LETVLDGLER
    ACVEARARWG IGSRLILCFL RHLSEEEGFA TLQQALPHLS RIDGVGLDSS ERGHPPAKFA
    RLFARCRELG LHVVAHAGEE GPPAYIVDAL DLLKAERIDH GVRCTEDPAL VGRLVREQVP
    LTVCP LSNVK LCVFPDLARH NLGQLFAAGL KVTINSDDPA YFGGYVAKNY VDTARALGLG
    RAELRRIARN SLEASFVSAA ERAPW LARLD 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 ACT_SITE 163 163 Proton donor.
FT BINDING 242 242 Substrate.
FT SITE 184 184 Important for catalytic activity.
```

```
**
** ##### INTERNAL SECTION #####
**HA FAM; Method MF_01962; ADE; Trusted match; 75.102 (+21.3).
**HA FAM; Method MF_00540; ADD; Weak match; 22.498 (-27.3).
**HA SAM; Annotated by HAMAP 3.1.3; MF_01962.6; MF_01962; 18-MAR-2016 15:37:54.
```

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

```
 catalytic|_FT_3,_FTGRP_1,_HAS_PATTERN_COND,H)),PositionalPattern(Positional|METAL_BINDING|{194
//}|{194//}|Zinc;
 catalytic|_FT_4,_FTGRP_1,_HAS_PATTERN_COND,H)),PositionalPattern(Positional|METAL_BINDING|{275
```

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

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

```
RIDHGVRCLL DAALTAARLAH ERIPETVCPV SNEKERVVDQ LADHNLQLL DAGLEVVTVNS
DDPAYFGGYM NDNFIATFEA LPLTRAHARL LARNSFDAAF LDGATRHKYL AEVDAFFDGQ
PA
```

```
//
```

Annotation output validation

CLUSTAL W (1.83) multiple sequence alignment template=ADE_PSEAE

```
ADE_YEAST      -----MVSVEFLQELPKCEHHLHLEGTLEPDLLFPLAKRNDIILPE----GFPKSV
ADE_CANGA      -----MVPESFLLLELPKCEHHLHLEGTLEPDLLFPLAKRNNIQLPD----HFPQTP
ADE_KLULA      --MAKFECTDEVTNFLTLPKCEHHLHLEGTLEPELLFQLVERNGVQLPG----TFPKTV
ADE_CANAL      --MAQYECSEHMENFLRELPKCEHVVHLEGTLEPSLLFKLAKRNNITLPE----TFPKTV
ADE_ASPFU      MC-----QSPLHDFLHGLPKCEHVVHLEGCVTPELIFQLAEKNNIQLPNPATHPAYASV
ADE_ASPOR      MC-----KSDLHDFLHGLPKCEHVVHLEGCCLAPDLIFELAKRNNVSLPN---EPAYESI
ADE_EMENI      MCPPNTPYQSQWHAFLHSLPKCEHVVHLEGCLEPPLIFSMARKNNVSLPSPSSNPAYTSV
ADE_SCHPO      MS-----NLPIYNFIRKLPKCEHVVHLEGCCLSPDLVFRLLAKKNGITLPS--DDAAYTTP
ADE_GIBZE      MC-----KSRVHSFLQALPKVEQHLHIEGTLEPELLFTLAEKNGIELPN---DPVYESA
ADE_CUPTR      -----MTIDAALAEQIRRTPKAELHVVHIEGTLEPELIFRLAQRNQVALPYP-----SV
ADE_CUPNH      -----MTIDAALAEQIRRTPKAELHVVHIEGTLEPELIFRLAQRNQVALPYP-----SV
ADE_CUPPJ      -----MTIDAALADKIRRTPKAELHVVHIEGTLEPELIFRLAQRNHVNL PYP-----SV
ADE_RALME      -----MTIDAALADKIRRTPKAELHVVHIEGTLEPERIFRLAQRNNVKLAYP-----DV
ADE_RALPJ      -----MPISSALAERIATSPKAEHLHIHIEGSLEPELMFALAERNGVKLPYA-----SV
ADE_RALSO      -----MPISPALAERIATSPKAEHLHIHIEGSLEPELMFALAERNGVKLPYA-----SV
F1VZU6_9BURK -----MALAFE-----SV
ADE_BURPP      -MTTTVTPTPLAECTALAPKAEHLHIHIEGSLEPELIFALAERNGVKLAYD-----SI
ADE_BURXL      -MTTTVTSTPLAECTVLAPKAEHLHIHIEGSLEPELIFALAERNGVKLAYD-----SI
ADE_GEOLS      -MNL TNIPRQALPELLCRMPKAEHLHIHIEGSLEPELIFALAERNRLQLAYP-----TI
ADE_GEOUR      -MNFDCIPREDLHGILCHMPKAEHLHIHIEGSLEPELIFELATRNNRIQLPYP-----TI
ADE_RHOFD      -MTIKPVSQERLPELLRTIPKAEHLHIHIEGSLEPELMFALAQRNGVSI PYP-----DV
ADE_ZYMMO      -----MNNLIKFIAALPKAEHLHLHIEGSLEPELMFELAKRNVKTL PFP-----DV
ADE_CAUCR      -MTDASFAPSASAEFVRGLPKAEHLMHIEGSLEPELMFELAQRNGITLPFA-----SV
ADE_CAUCN      -MTDASFAPSASAEFVRGLPKAEHLMHIEGSLEPELMFELAQRNGITLPFA-----SV
```


Annotation output validation

```
FT BINDING 278 278 Substrate.  
FT SITE 222 222 Important for catalytic activity (By  
FT similarity).
```

```
**
```

```
** ##### INTERNAL SECTION #####
```

```
**HA FAM; Method MF_01962; ADE; Trusted match; 61.674 (+7.4).
```

```
**HA FAM; Method MF_01962; ADE; Trusted match; 61.674 (+7.4).
```

```
**HA FAM; Method MF_00540; ADD; Weak match; 23.344 (-26.9).
```

```
**HA SAM; Annotated by praise 1.1.47; MF 01962; MF 01962.4; Mon Apr 08 05:54:51 CEST 2013.
```

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

```
catalytic|BY_SIMILARITY|,H));POSITIONALPATTERN(Positional|METAL_BINDING|{275//trigger}|{275//trigger}|Zinc;catalytic|BY_SIMILARITY|,D)))
```

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

```
SQ SEQUENCE 332 AA; 37491 MW; 701E1C335EA545DB CRC64;  
MAHEALVQAL PKVELHVHIE GTLEPELKFK LAQRNGIKLP FSSEQEVKDS YTFNDLASFL  
DAYYDGMNVL LASEDFYDLA MAYYRKAASQ GLRYAEIFFD PQAHTCRGVS FHTVISGLRR  
AQLEAEQTLG VFSQFIMCFL RDFQPEYAMA TLLES LPYKQ 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      ---LKWIDGIGLDSSEVGHPPVKFREVEAAAA--LGLRRVAHAGEE--GPPDYVWQAIE
ADE_RHOPB      ---KDQIIAIGMGAELGNPPAKFARFFKAARD--RGFRTTVHAGEE--GPAAYVREALE
ADE_STRCO      ---LDRITGVGLDSAIEVGHPPVKFREVEAAAA--LGLRRVAHAGEE--GPPAYVVEALD
G9PG85_9ACTO   ---KQWIVGVGLDSDIETDHEPAKFAAVFARARR--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_AGR5T      ---HPLVTGFNMAGEERMGRVADYARAFDIARD--AGLGLTIHAGEV--CGPESVADALD
ADE_RHIME      ---HPLVTGFNLAGEERMHSVAEFSRAFDIVRD--AGLGLTIHAGEE--SGAFSVRDALD
ADE_RHISN      ---HPLVTGFNLAGEERMHSVAEFARAFDIARD--AGLGLTIHAGEE--SGAFSVRDALD
ADE_SINMW      ---HPLVTGFNLAGEERMHSVAEFSRAFDIVRD--AGLGLTIHAGEE--SGAFSVRDALD
ADE_AGRVS      ---HPLITGFNMAGEERMNRVADYAPAFDIARE--AGYGITIHAGEE--CGAFSVRDALD
ADE_MESSB      ---HPLVTGFGMAGDERAGHPRDFAYAFDIARE--AGLGISIHAGEE--GGAESVEAALD
ADE_RHILO      ---NPLVTGFGVAGDERVGEDEYVRAFEIARE--AGLGITIHAGEE--TGWETVQAALD
```

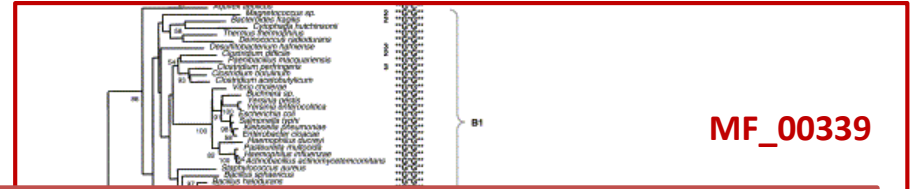
```
. . . . *      :      *..      :      :.
```

Controlling annotations using cases and conditions

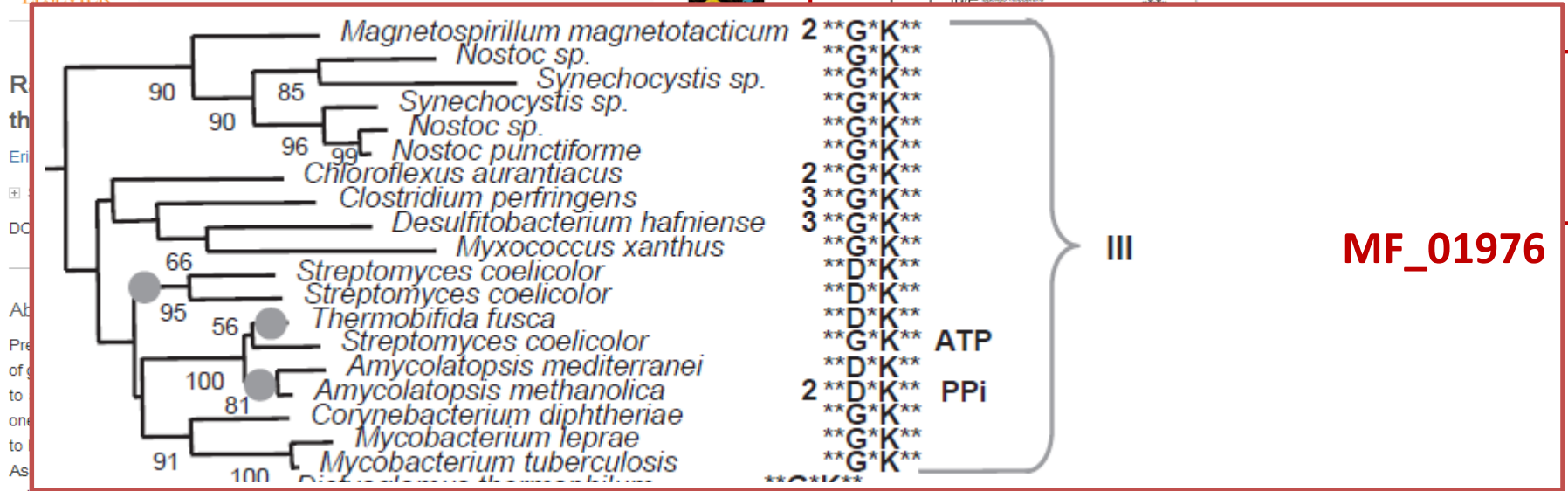


Gene

Volume 318, 30 October 2003, Pages 185–191

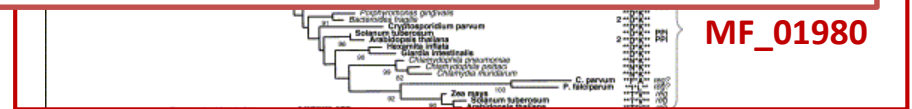


MF_00339



MF_01976

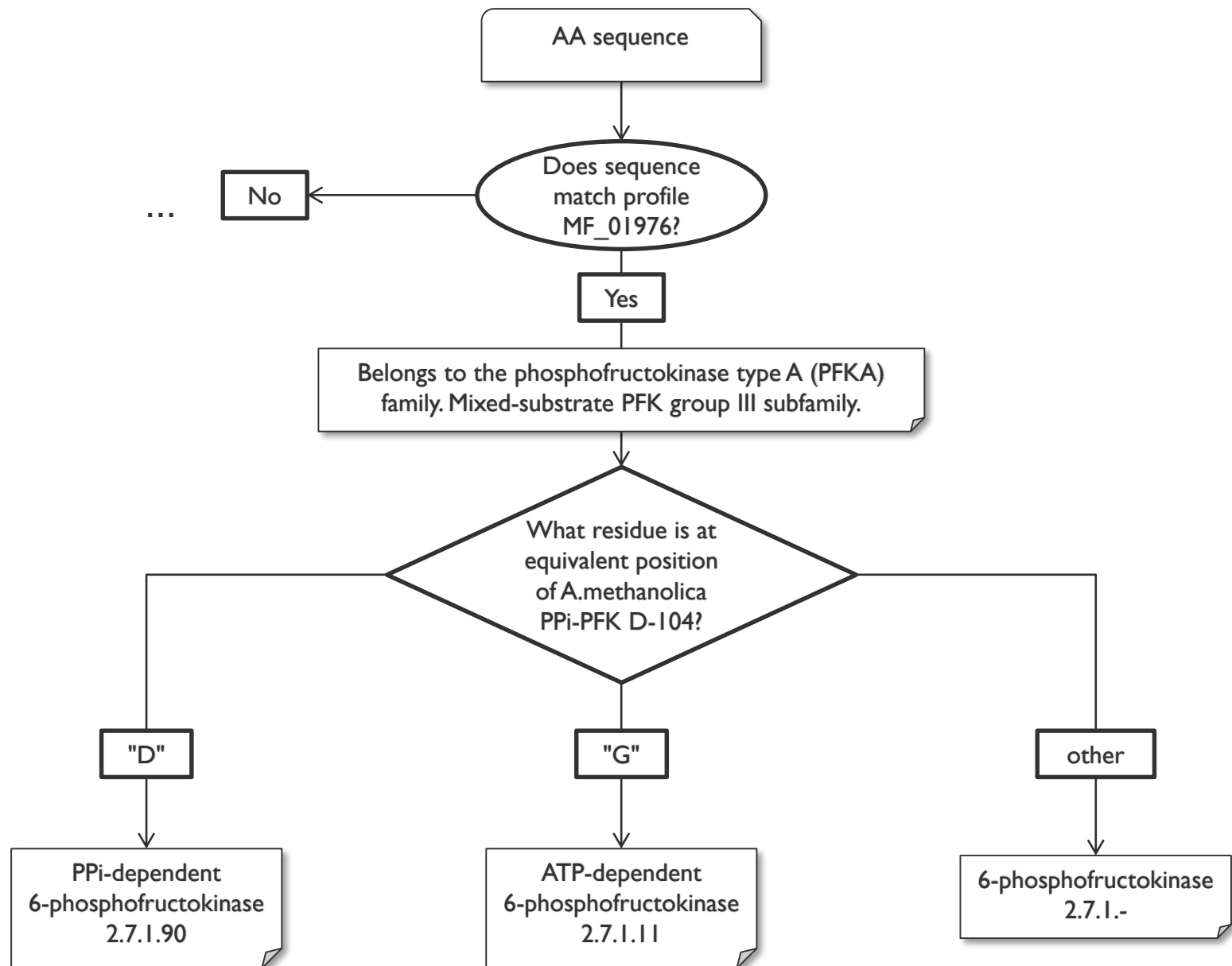
existence of numerous sequences from distantly related species carrying atypical combinations of amino acids. Several adaptive changes of phospho-donors, probably requiring a single mutation at position 104, have likely occurred independently in many lineages. The analysis of this gene suggests the existence of a high rate of both HGT and substitution in its active sites. These rampant HGT events and flexibility in phospho-donor use illustrate the importance of tinkering in molecular evolution.



MF_01980



MF_01981









Annotation rule MF_01976

[Send feedback](#)

Features [\[?\]](#)

From: PFP_AMYPE ([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 P _i ; prevents ATP from binding by mimicking the alpha-phosphate group of ATP	PPI	D 	"D"
other						
 SITE (Optional)	104	104	Important for substrate specificity; cannot use P _i as phosphoryl donor	ATP	G 	"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	

Identifier, protein and gene names [?]

"D"

case <FTTag:PPI>

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

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.



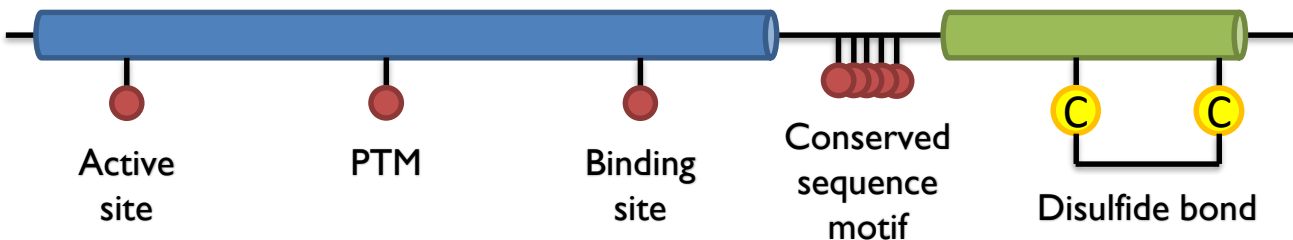
Domains Specific combination of secondary structures organized into a characteristic three dimensional structure or fold, that may exist in a variety of biological contexts.



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



PROSITE

Home | Contact



annotation rule: PRU00297

- Home
- ScanProsite
- ProRule
- Documents
- Downloads
- Links
- Funding

General rule information

Accession	PRU00297
Dates	12-DEC-2003 (Created) 2-DEC-2010 (Last updated, Version 05)
Data class	Domain
Predictors	PROSITE; PS50873; PEROXIDASE_4
Name	Plant heme peroxidase
Function	Removal of H(2)O(2), oxidation of toxic reductants

Different classes of signature - domains, patterns, families

Generic function of a domain

Propagated annotations

Conditional annotations (again)

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.

ProRules

← → ↻ prosite.expasy.org/unirule/PRU00297 ☆



PROSITE

[Home](#) | [Contact](#)

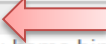


annotation rule: PRU00297

[Home](#) [ScanProsite](#) **[ProRule](#)** [Documents](#) [Downloads](#) [Links](#) [Funding](#)

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

Coordinates refer to positions in profile PS50783 - no template sequence here

case <Feature:PS50873:1=Q> 
Pyrrolidone carboxylic acid
end case

ProRules

← → ↻ prosite.expasy.org/unirule/PRU00297 ☆



PROSITE

Home | [Contact](#)

[Home](#) [ScanProsite](#) [ProRule](#) [Documents](#) [Downloads](#) [Links](#) [Funding](#)



Profile identifier

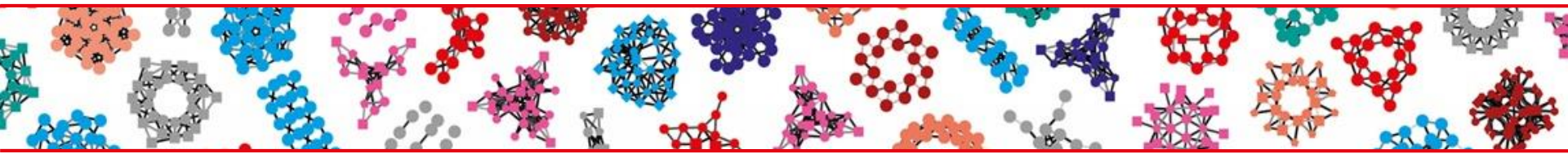
Positions in profile

annotation

Features

From: [PS50873](#)

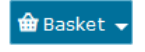
Key	From	To	Description	Tag	Condition	FTGroup
METAL	43	43	Calcium #1 (<i>By similarity</i>)		[DE]	1
METAL	46	46	Calcium #1; via carbonyl oxygen (<i>By similarity</i>)		[VI]	1
METAL	48	48	Calcium #1; via carbonyl oxygen (<i>By similarity</i>)		G	1
METAL	50	50	Calcium #1 (<i>By similarity</i>)		[DE]	1
METAL	52	52	Calcium #1 (<i>By similarity</i>)		[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



Q8N3Z0 - PRS35_HUMAN



Protein | Inactive serine protease 35

Gene | PRSS35

Organism | *Homo sapiens (Human)*

Sequence features | View only features (sites, domains, PTMs ...)

Status | Reviewed - Annotation score: - Experimental evidence at transcript level¹

Display All None [BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#)

[Feedback](#) [Help video](#)

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Publications
- Entry information
- Miscellaneous
- Similar proteins

Sequence¹

Sequence status¹: Complete.

Sequence processing²: The displayed sequence is further processed into a mature form.

Q8N3Z0-1 [UniParc] [FASTA](#) [Add to basket](#)

[< Hide](#)

10	20	30	40	50
MENMLLWLF	FTPGWTLIDG	SEMEWDFMWH	LRKVPRIVSE	RTFHLTSPAF
60	70	80	90	100
EADAKMMVNT	VCGIECQKEL	PTPSLSELED	YLSYETVFEN	GTRTLTRVKV
110	120	130	140	150
QDLVLEPTQN	ITTKGVSVRR	KRQVYGTDSR	FSILDKRELT	NPPFSTAVKL
160	170	180	190	200
STGCSGILIS	PQHVLTAHC	VHDGRDYVKG	SKKLRVGLLK	MRNKSGGKKR
210	220	230	240	250
RGSKRSRREA	SGGDQREGTR	EHLRERAKGG	RRRKSGRGQ	RIAEGRPSFQ
260	270	280	290	300
WTRVKNTHIP	KGWARGGMD	ATLDYDYALL	ELKRAHKKKY	MELGISPTIK
310	320	330	340	350
KMPGGMIHFS	GFDNDRADQL	VYRFCVSDE	SNDLLYQYCD	AESGSTGSGV
360	370	380	390	400
VLDLDDDKK	NWEDKLAH	SGHGFVPLG	YKKNVIAH	LDLKVAGIC

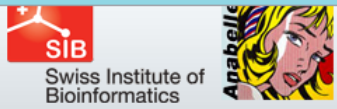
Length: 413

Mass (Da): 47,098

Last modified: September 11, 2007 - v2

Checksum:¹ 818D9C951BD2D6C1

BLAST

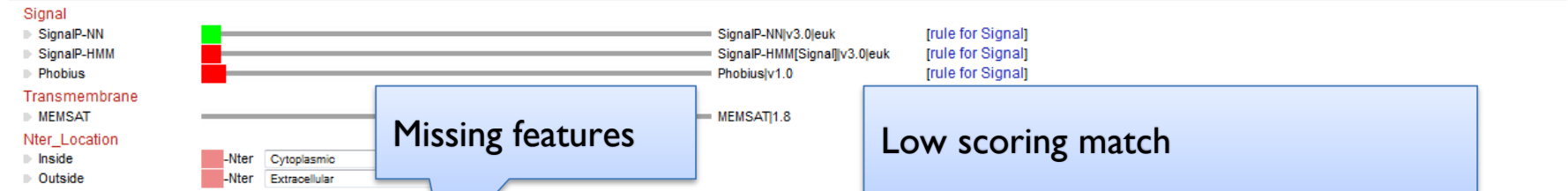


color codes quick help tutorial anabelle full technical doc

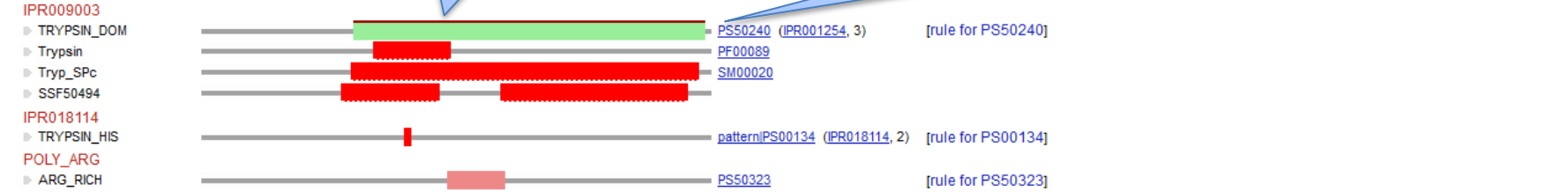
Sequence Analysis Log

PRSS35_HUMAN: [Size: 413] [blast](#) [String](#) [HoverProt links](#) [Family](#) [Alignment](#) [Tree](#) [preview](#) Zoom: 100%

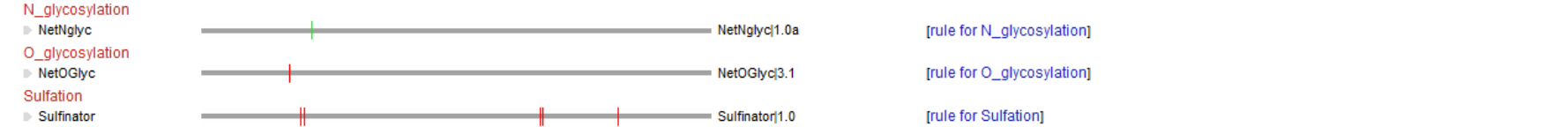
TOPOLOGY



DOMAIN



SITE



SKIP

CURRENT_FT



Display All None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Publications
- Entry information
- Miscellaneous
- Similar proteins

[▲ Top](#)

PTM / Processingⁱ

Molecule processing

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Signal peptide ⁱ	1 - 16	16	Sequence Analysis			Add BLAST
Chain ⁱ	17 - 413	397	Inactive serine protease 35		PRO_0000299358	Add BLAST

Amino acid modifications

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Glycosylation ⁱ	90 - 90	1	N-linked (GlcNAc...) Sequence Analysis			
Disulfide bond ⁱ	154 ↔ 170		By similarity			

Keywords - PTMⁱ

Disulfide bond, Glycoprotein

Proteomic databases

PaxDb ⁱ	Q8N3Z0.
PRIDE ⁱ	Q8N3Z0.

The domain is nevertheless annotated following manual evaluation by the curator...

Family & Domainsⁱ

Domains and Repeats

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Domain ⁱ	124 - 408	285	Peptidase S1			Add BLAST

Sequence similaritiesⁱ

Belongs to the [peptidase S1 family](#). [Curated](#)

Contains 1 [peptidase S1 domain](#). [Curated](#)

Keywords - Domainⁱ

Signal

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]	<pre> /GENERAL_SPEC: ALPHABET='ABCDEFGHIJKLMNPQRSTUVWXYZ'; 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; ... </pre> <p>» More</p>

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

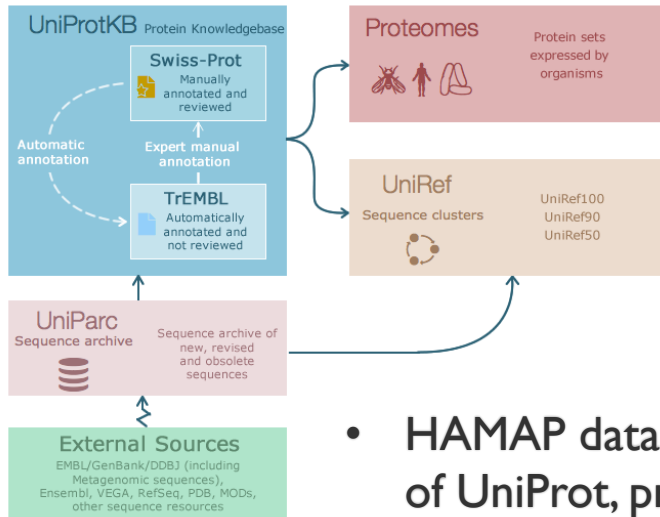
Feature key [info]	DOMAIN
Feature description [info]	Peptidase S1
Version [info]	2

Cross-references [\[info\]](#)

UniProtKB/Swiss-Prot True positive sequences	<p>739 sequences</p> <p>ACH1_LONAC (P23604), ACH2_LONAC (P23605), ACRO_HUMAN (P10323), » More</p> <ul style="list-style-type: none"> Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits: [Clustal format, color, condensed view] [Clustal format, color] [Clustal format, plain text] [Fasta format] Retrieve the sequence logo from the alignment
UniProtKB/Swiss-Prot False negative sequences	<p>17 sequences</p> <p>HTR1B_DANRE (A9JRB3), HTRA1_BOVIN (F1N152), HTRA1_XENLA (A6YFB5), HTRA1_XENTR (A4IHA1), PRS23_BOVIN (Q1LZE9), PRS23_HUMAN (O95084), PRS23_MACMU (Q1WK23), PRS23_MOUSE (Q9BCK6), PRS23_RAT (Q6AY61), PRS35_BOVIN (Q5E9X0), PRS35_HUMAN (Q8N3Z0), PRS35_MACMU (Q1WK24), PRS35_MOUSE (Q8C0F9), PRS35_RAT (Q5E212), Y3671_MYCTO (P9WHR8), Y3671_MYCTU (P9WHR9), YH05_SCHPO (O74325) Less <</p>
UniProtKB/Swiss-Prot 'Partial' sequences	<p>80 sequences</p> <p>ACRO_CAPHI (P10626), CATG_RAT (P17977), CBP_MESMA (P0C8M2), » More</p>
PDB [Detailed view]	<p>1685 PDB</p> <p>1A0H; 1A0J; 1A0L; 1A2C; 1A3B; 1A3E; 1A46; 1A4W; 1A5G; 1A5H; 1A5I; 1A61; » More</p>

[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')

UniProtKB - Q63RY2 (Q63RY2_BURPS)

Display

- [Entry](#)
- [Feature viewer](#)
- [Feature table](#)

All None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Publications
- Entry information
- Miscellaneous
- Similar proteins

[▲ Top](#)

- [BLAST](#)
- [Align](#)
- [Format](#)
- [Add to basket](#)
- [History](#)

- [Feedback](#)
- [Help video](#)
- [Other tutorials and videos](#)

 Protein | **Adenine deaminase**

 Gene | **add**

 Organism | *Burkholderia pseudomallei* (strain K96243)

 Status | [Unreviewed](#) - Annotation score: ●○○○○ - Protein inferred from homologyⁱ

Family & Domainsⁱ

Sequence similaritiesⁱ

 Belongs to the adenosine and AMP deaminases family. Adenine deaminase type 2 subfamily. [UniRule annotation](#)

Phylogenomic databases

eggNOG ⁱ	ENOG410SEKD. Bacteria. COG1816. LUCA.
HOGENOM ⁱ	HOG000218813.
KO ⁱ	K01488.
OMA ⁱ	DHGVRAI.
OrthoDB ⁱ	EOG6BPDDD.

Cross-reference links back to HAMAP profile

Family and domain databases

HAMAP ⁱ	MF_01962. Adenine_deaminase.
InterPro ⁱ	IPR001365. A/AMP_deaminase_dom. IPR028892. ADE. IPR006330. Ado/ade_deaminase. IPR032466. Metal_Hydrolase. [Graphical view]
Pfam ⁱ	PF00962. A_deaminase. 1 hit. [Graphical view]
SUPFAM ⁱ	SSF51556. SSF51556. 1 hit.
TIGRFAMs ⁱ	TIGR01430. aden_deam. 1 hit.



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	

UniProtKB - Q63RY2 (Q63RY2_BURPS)

Display

 Entry

 Feature viewer

 Feature table

None

 Function

 Names & Taxonomy

 Subcellular location

 Pathology & Biotech

 PTM / Processing

 Expression

 Interaction

 Structure

 Family & Domains

 Sequence

 Cross-references

 Publications

 Entry information

 Miscellaneous

 Similar proteins

 Protein | **Adenine deaminase**

 Gene | **add**

 Organism | *Burkholderia pseudomallei* (strain K96243)

 Status | Unreviewed - Annotation score: - Protein inferred from homology¹

Function¹

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

Catalytic activity¹

 Adenine + H₂O = hypoxanthine + NH₃. UniRule annotation

Cofactor¹

 Zn²⁺ UniRule annotation

Note: Binds 1 zinc ion per subunit. UniRule annotation

Sites

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Metal binding ¹	18 - 18		1 Zinc; catalytic UniRule annotation			
Metal binding ¹	20 - 20		1 Zinc; catalytic UniRule annotation			
Metal binding ¹	201 - 201		1 Zinc; catalytic UniRule annotation			
Active site ¹	204 - 204		1 Proton donor UniRule annotation			
Site ¹	225 - 225		1 Important for catalytic activity UniRule annotation			
Metal binding ¹	282 - 282		1 Zinc; catalytic UniRule annotation			
Binding site ¹	283 - 283		1 Substrate UniRule annotation			

Evidence attribution links back to annotation rule

 Automatic assertion according to rules¹

HAMAP-Rule:MF_01962

UniRule: UR000089071

Source Rule HAMAP: MF_01962

[View all proteins annotated by this rule](#) [Remove highlights](#)

 The annotation and conditions in this rule are derived from the following entries: [Q9I6Y4 \(ADE_PSEAE\)](#), [P53909 \(ADE_YEAST\)](#)

If a protein meets these conditions...

Common conditions

If

Matches HAMAP signature [MF_01962](#)
 taxon = [Bacteria](#)
 fragment ≠ the sequence is fragmented

Special conditions

Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#)
 Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#)
 Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#)
 Subsequence at position 16 - 16 aligns to "D" in entry [Q9I6Y4](#)

Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#)
 (applies "Zinc; catalytic") Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#) (applies "Zinc; catalytic") Subsequence at position 194 - 194 aligns to "H" in entry [Q9I6Y4](#) (applies "Zinc; catalytic") Subsequence at position 275 - 275 aligns to "D" in entry [Q9I6Y4](#) (applies "Zinc; catalytic")

Subsequence at position 197 - 197 aligns to "E" in entry [Q9I6Y4](#)
 (individually applies "Proton donor")

CONDITIONS

... then these annotations are applied

Protein name

then

Recommended name:
Adenine deaminase (EC:3.5.4.2)
Short name:
 ADE
Alternative name(s):
 Adenine aminohydrolase
Short name:
 AAH

Sequence similarity

Belongs to the Adenosine deaminases superfamily. Adenosine deaminase type 2 subfamily.

Cofactors

Zn²⁺
Note: Binds 1 zinc ion per subunit.

Catalytic activity

Adenine + H₂O = hypoxanthine + NH₃.

ANNOTATIONS

UniRule: UR000089071

Source Rule HAMAP: MF_01962

[View all proteins annotated by this rule](#)
[Remove highlights](#)

The annotation and conditions in this rule are derived from the following entries: [Q9I6Y4 \(ADE_PSEAE\)](#), [P53909 \(ADE_YEAST\)](#)

If a protein meets these conditions...ⁱ

Common conditions

If

Matches HAMAP signature [MF_01962](#)
 taxon = [Bacteria](#)
 fragment ≠ the sequence is fragmented

Special conditions

and

Subsequence at position 14 - 14 aligns to "H" in entry [Q9I6Y4](#)
 Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#)
 Subsequence at position 194 - 194 aligns to "H" in entry [Q9I6Y4](#)
 Subsequence at position 275 - 275 aligns to "D" in entry [Q9I6Y4](#)

Subsequence at position 14 - 14 aligns to "H" in entry [Q9I6Y4](#) (applies "Zinc; catalytic")
 Subsequence at position 16 - 16 aligns to "H" in entry [Q9I6Y4](#) (applies "Zinc; catalytic")
 Subsequence at position 194 - 194 aligns to "H" in entry [Q9I6Y4](#) (applies "Zinc; catalytic")
 Subsequence at position 275 - 275 aligns to "D" in entry [Q9I6Y4](#) (applies "Zinc; catalytic")

Subsequence at position 197 - 197 aligns to "E" in entry [Q9I6Y4](#) (individually applies "Proton donor")

... then these annotations are appliedⁱProtein nameⁱ

then

Recommended name:
Adenine deaminase (EC:3.5.4.2)
Short name:
 ADE
Alternative name(s):
 Adenine aminohydrolase
Short name:
 AAH

Sequence similaritiesⁱ

Belongs to the [metallo-dependent hydrolases superfamily](#), [Adenosine and AMP deaminases family](#), [Adenine deaminase type 2 subfamily](#).

Cofactorⁱ

Zn²⁺
Note: Binds 1 zinc ion per subunit.

Catalytic activityⁱ

Adenine + H₂O = hypoxanthine + NH₃.

UniProtKB - Q63RY2 (Q63RY2_BURPS)

Display

 Entry

 Feature viewer

 Feature table

None

 Function

 Names & Taxonomy

 Subcellular location

 Pathology & Biotech

 PTM / Processing

 Expression

 Interaction

 Structure

 Family & Domains

 Sequence

 Cross-references

 Publications

 Entry information

 Miscellaneous

 Similar proteins

 Protein | **Adenine deaminase**

 Gene | **add**

 Organism | *Burkholderia pseudomallei* (strain K96243)

 Status | Unreviewed - Annotation score: - Protein inferred from homology¹

Functionⁱ

GO - Molecular functionⁱ

- adenine deaminase activity Source: UniProtKB-UniRule
- zinc ion binding Source: UniProtKB-UniRule

View the complete GO annotation on QuickGO ...

GO - Biological processⁱ

- adenine catabolic process Source: UniProtKB-UniRule
- hypoxanthine salvage Source: UniProtKB-UniRule
- nucleotide metabolic process Source: UniProtKB-KW

View the complete GO annotation on QuickGO ...

Keywordsⁱ

Molecular function	Hydrolase UniRule annotation SAAS annotation Imported
Biological process	Nucleotide metabolism UniRule annotation
Ligand	Metal-binding UniRule annotation SAAS annotation , Zinc UniRule annotation SAAS annotation

GO terms are directly integrated into the UniProt-GOA database and loaded into UniProtKB

GO annotations

[Taxon](#)
[Gene Products](#)
[GO terms](#)
[References](#)
[Aspect](#)
[Evidence](#)
[Extension](#)
[More](#)
[Clear all](#)

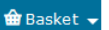
Annotations [Statistics](#)

[Customise](#)
[Export](#)
 11 annotations

Gene Product	Symbol	Qualifier	GO Term	Evidence	Reference	With / From	Taxon	Assigned By	Annotation Extension
UniProtKB:Q63RY2	add	enables	GO:0000034 adenine deaminase activity	ECO:0000256 (IEA)	GO_REF:0000002	InterPro:IPR028892	272560	InterPro	
UniProtKB:Q63RY2	add	enables	GO:0000034 adenine deaminase activity	ECO:0000501 (IEA)	GO_REF:0000003	EC:3.5.4.2	272560	UniProt	
UniProtKB:Q63RY2	add	enables	GO:0000034 adenine deaminase activity	ECO:0000256 (IEA)	GO_REF:0000104	UniRule:UR000089071	272560	UniProt	
UniProtKB:Q63RY2	add	enables	GO:0008270 zinc ion binding	ECO:0000256 (IEA)	GO_REF:0000104	UniRule:UR000089071	272560	UniProt	
UniProtKB:Q63RY2	add	enables	GO:0016787 hydrolase activity	ECO:0000323 (IEA)	GO_REF:0000038	UniProtKB-KW:KW-0378	272560	UniProt	
UniProtKB:Q63RY2	add	enables	GO:0019239 deaminase activity	ECO:0000256 (IEA)	GO_REF:0000002	InterPro:IPR001365 more...	272560	InterPro	
UniProtKB:Q63RY2	add	enables	GO:0046872 metal ion binding	ECO:0000323 (IEA)	GO_REF:0000038	UniProtKB-KW:KW-0479	272560	UniProt	
UniProtKB:Q63RY2	add	involved_in	GO:0006146 adenine catabolic process	ECO:0000256 (IEA)	GO_REF:0000002	InterPro:IPR028892	272560	InterPro	
UniProtKB:Q63RY2	add	involved_in	GO:0006146 adenine catabolic process	ECO:0000256 (IEA)	GO_REF:0000104	UniRule:UR000089071	272560	UniProt	
UniProtKB:Q63RY2	add	involved_in	GO:0009117 nucleotide metabolic process	ECO:0000323 (IEA)	GO_REF:0000038	UniProtKB-KW:KW-0546	272560	UniProt	
UniProtKB:Q63RY2	add	involved_in	GO:0043103 hypoxanthine salvage	ECO:0000256 (IEA)	GO_REF:0000104	UniRule:UR000089071	272560	UniProt	

UniProtKB results

About UniProtKB



Filter byⁱ

Unreviewed (4,099)
TrEMBL

Popular organisms

- ACIHA (2)
- Pseudomonas sp. LAMO17WK12:I10 (1)
- Pseudomonas sp. GR 6-02 (1)
- Pseudomonas sp. IB20 (1)
- Kangiella sp. (1)

Other organisms

View by

Results table

- Taxonomy
- Keywords
- Gene Ontology
- Enzyme class
- Pathway

UniRef

Your results in sequence clusters with identity of: 100%, 90% or 50%

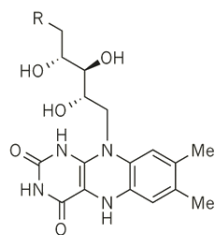
Demo

Help video

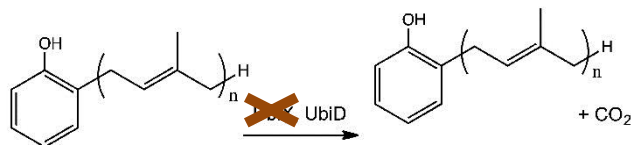
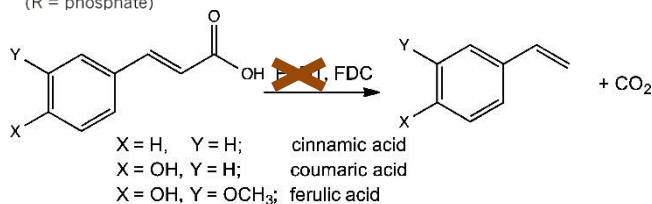
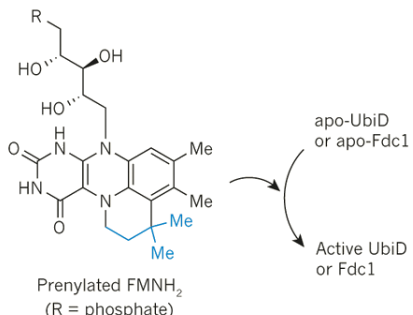
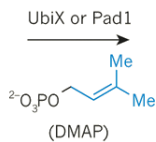
◀ 1 to 25 of 4,099 ▶ Show 25 ▾

Show only entries with structural information (1)

<input type="checkbox"/>	Entry	Entry name	Protein names	Gene names	Organism	Length	3D
<input type="checkbox"/>	A0A2G2KTC3	A0A2G2KTC3_9GAMM	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	COA86_16640	Kangiella sp.	334	
<input type="checkbox"/>	A0A173JCL7	A0A173JCL7_9PSED	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	PGR6_06570	Pseudomonas sp. GR 6-02	317	
<input type="checkbox"/>	A0A1L6KNL1	A0A1L6KNL1_ACIHA	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	AHTJS_09940	Acinetobacter haemolyticus	332	
<input type="checkbox"/>	A0A285IN04	A0A285IN04_9PSED	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	SAMN05660489_05132	Pseudomonas sp. LAMO17WK12:I10	317	
<input type="checkbox"/>	A0A263NQV7	A0A263NQV7_9PSED	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	B7453_18085	Pseudomonas sp. IB20	317	
<input type="checkbox"/>	L1P0K8	L1P0K8_9NEIS	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	HMPREF9120_00524	Neisseria sp. oral taxon 020 str. F0370	337	
<input type="checkbox"/>	A0A222FA62	A0A222FA62_9RHOB	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	CHH27_24150	Labrenzia sp. VG12	331	
<input type="checkbox"/>	A0A1G1BH04	A0A1G1BH04_9BACT	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	A2X65_00065	Melioribacter sp. GWF2_38_21	339	
<input type="checkbox"/>	A0A095AU18	A0A095AU18_9SPHN	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	FG91_03173	Sphingopyxis sp. LC81	341	
<input type="checkbox"/>	A0A0W1SH47	A0A0W1SH47_9GAMM	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	AUR68_01945	Idiomarina sp. H105	330	
<input type="checkbox"/>	D6B069	D6B069_9ACTN	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	SSHG_01542	Streptomyces albus J1074	330	
<input type="checkbox"/>	N9Q6E6	N9Q6E6_9GAMM	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	F892_01235	Acinetobacter sp. NIPH 2168	332	
<input type="checkbox"/>	N9MXH8	N9MXH8_9GAMM	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	F897_02596	Acinetobacter sp. NIPH 2171	333	
<input type="checkbox"/>	A0A270B5L3	A0A270B5L3_9SPHN	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	CD928_18145	Sphingopyxis sp. GW247-27LB	341	
<input type="checkbox"/>	I2DKV0	I2DKV0_9BURK	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	MYA_0909	Burkholderia sp. KJ006	341	
<input type="checkbox"/>	A0A2A9LU26	A0A2A9LU26_9BURK	Adenine deaminase (ADE) (EC 3.5.4.2) (Adenine aminohydrolase) (AAH)	BX604_0491	Burkholderia sp. JKS000303	341	



Riboflavin: R = OH
 FMNH₂: R = phosphate



nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issues

Archive | Volume 522 | Issue 7557 | Letters | Article

NATURE | LETTER

日本語要約

UbiX is a flavin prenyltransferase related to 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

```

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="MKRLIIGISGASGAIYGVRLQVLRDVPDIETHLILSQARQTLA
FT METDFTVREVQALADVVDHARDIAASISSGSFKTAGMVILPCSMKTLGSIHVSYTDGLL
FT TRAADVVLKERRPLVLCVRETPFHLGHLRLLVQAAELGAVIMPPVPAFYHRPQSLDDVI
FT NQTVNRVLDQFDISLEQDLFTRWQGSQDCTK"
XX
  
```

Search Cite
 Enter search text / DOI
 ACS Chem. Biol.

Issue Multimedia Submission & Review Open Access

Previous Article

UbiX Catalyze Formation of a Novel 3-Octaprenyl-4-Hydroxybenzoate Decarboxylase and 4-Hydroxybenzoate

ert, Xiaoxia Nina Lin[†], and E. Neil G. Marsh^{†§}
 engineering, and [§]Department of Biological Chemistry, University of

Entry	Entry name	Protein names	Gene names	Organism	Length
P33751	PAD1_YEAST	Flavin prenyltransferase PAD1, mito...	PAD1 POF1,YDR538W	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	242
P0AG03	UBIX_ECOLI	Flavin prenyltransferase UbiX	ubiX dedF,b2311,JW2308	Escherichia coli (strain K12)	189
Q9HX08	UBIX_PSEAE	Flavin prenyltransferase UbiX	ubiX PA4019	Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228)	209
K0NLG5	K0NLG5_DESTT	Flavin prenyltransferase UbiX	ubiX2 ubiX,ubiX,ubiX3,ubiX4,ubiX5	Desulfobacula toluolica (strain DSM 7467 / Tol2)	206
P57767	UBIX_THAAR	Flavin prenyltransferase UbiX	ubiX	Thaueria aromatica	194
A0A0M2WNJ7	A0A0M2WNJ7_9BURK	Flavin prenyltransferase UbiX	bsdB ubiX,VM94_01772	Janthinobacterium sp. KBS0711	204
Q13YE3	Q13YE3_BURXL	Flavin prenyltransferase UbiX	ubiX Bxe_A2072	Burkholderia xenovorans (strain LB400)	218
D8ISN1	D8ISN1_HERSS	Flavin prenyltransferase UbiX	ubiX Hsero_3976	Herbaspirillum seropedicae (strain SmR1)	200
E8RDD9	E8RDD9_DESPD	Flavin prenyltransferase UbiX	ubiX Despr_1795	Desulfobulbus propionicus (strain ATCC 33891 / DSM 2032 / 1pr3)	190
A1KW21	A1KW21_NEIMF	Flavin prenyltransferase UbiX	ubiX NMC1916	Neisseria meningitidis serogroup C / serotype 2a (strain ATCC 700532 / DSM 15464 / FAM18)	223

Protein | Flavin prenyltransferase UbiX
Gene | pad1_1
Organism | *Klebsiella pneumoniae*
Status | Unreviewed - Annotation score: ●○○○ - Protein inferred from homologyⁱ

A0A0C2MA92_KLEPN

Functionⁱ

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 activityⁱ

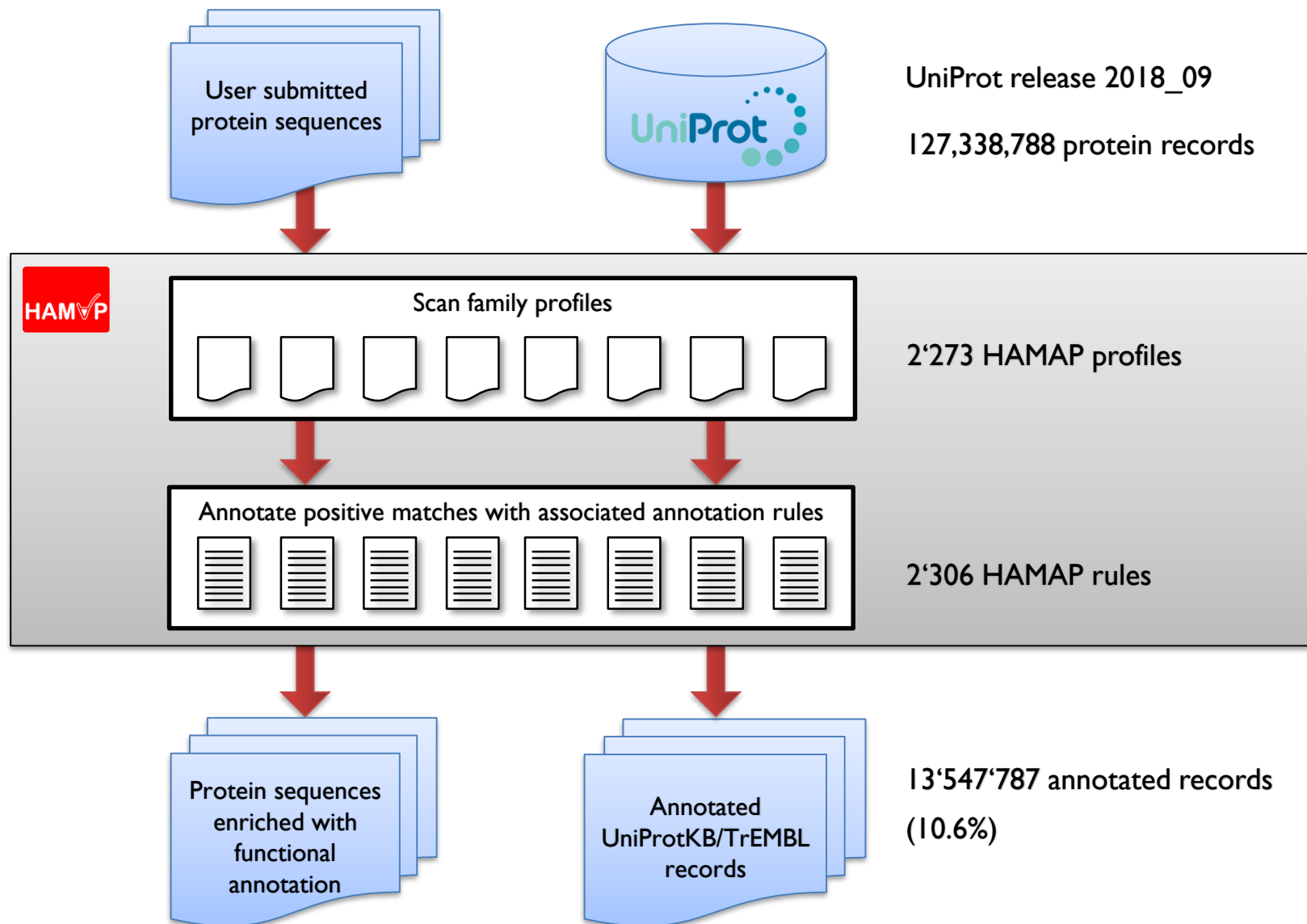
Dimethylallyl phosphate + FMNH₂ = prenylated FMNH₂ + phosphate. UniRule annotation

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	

Automatic assertion according to rulesⁱ
 HAMAP-Rule:MF_01984

Current status





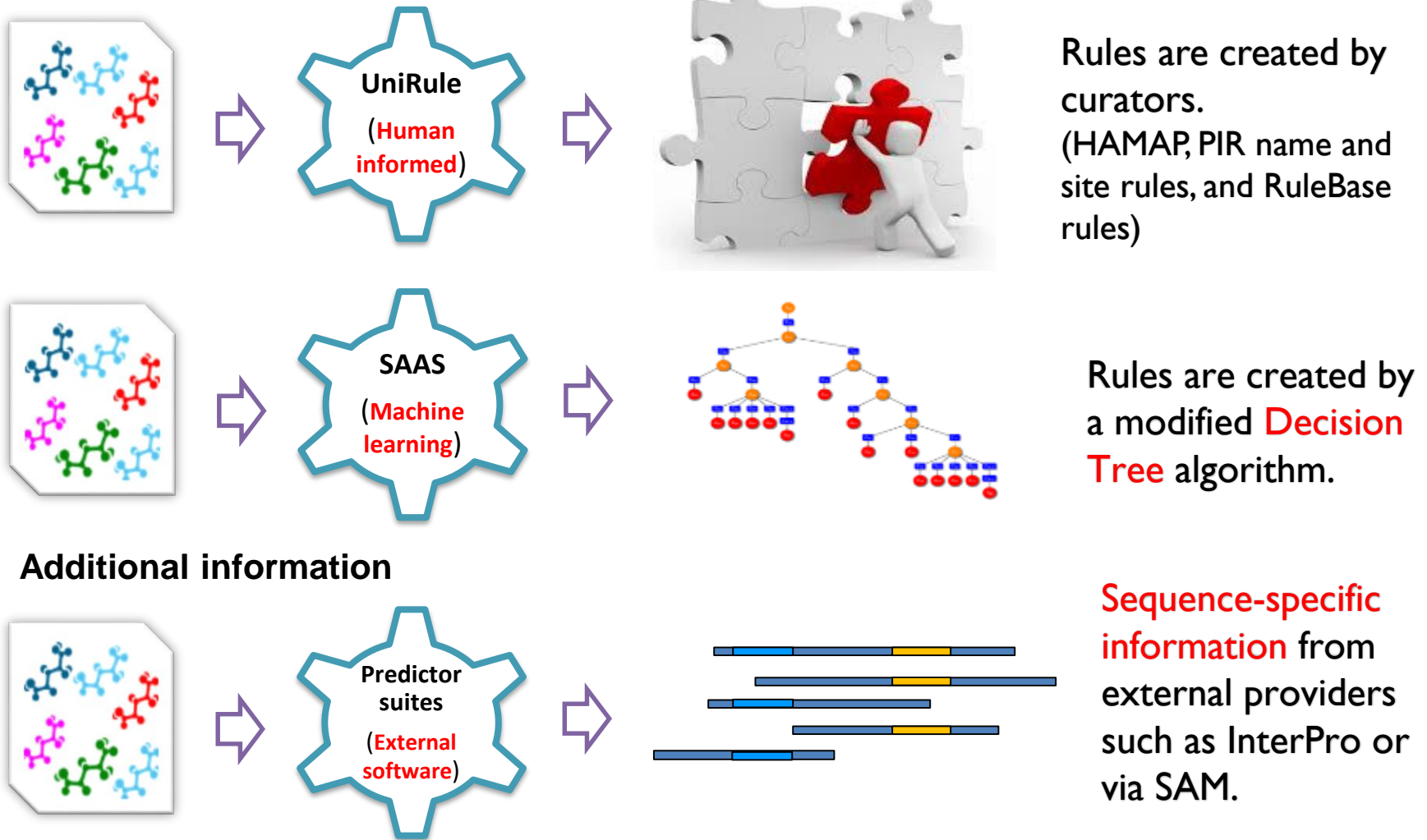
Reference proteomes: archaea

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			HAMAP coverage in UniProtKB
			All	Swiss-Prot	TrEMBL	
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%
Halopyrum pomix (strain ATCC 700557 / DSM 11031 / JCM 3026 / NBRC 100130 / K1)	HALPE	UP000002916	1766	507	1259	20%
Buchnera aphidicola subsp. Acyrthosiphon pisum (strain APS) (Acyrthosiphon pisum symbiotic bacterium)	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

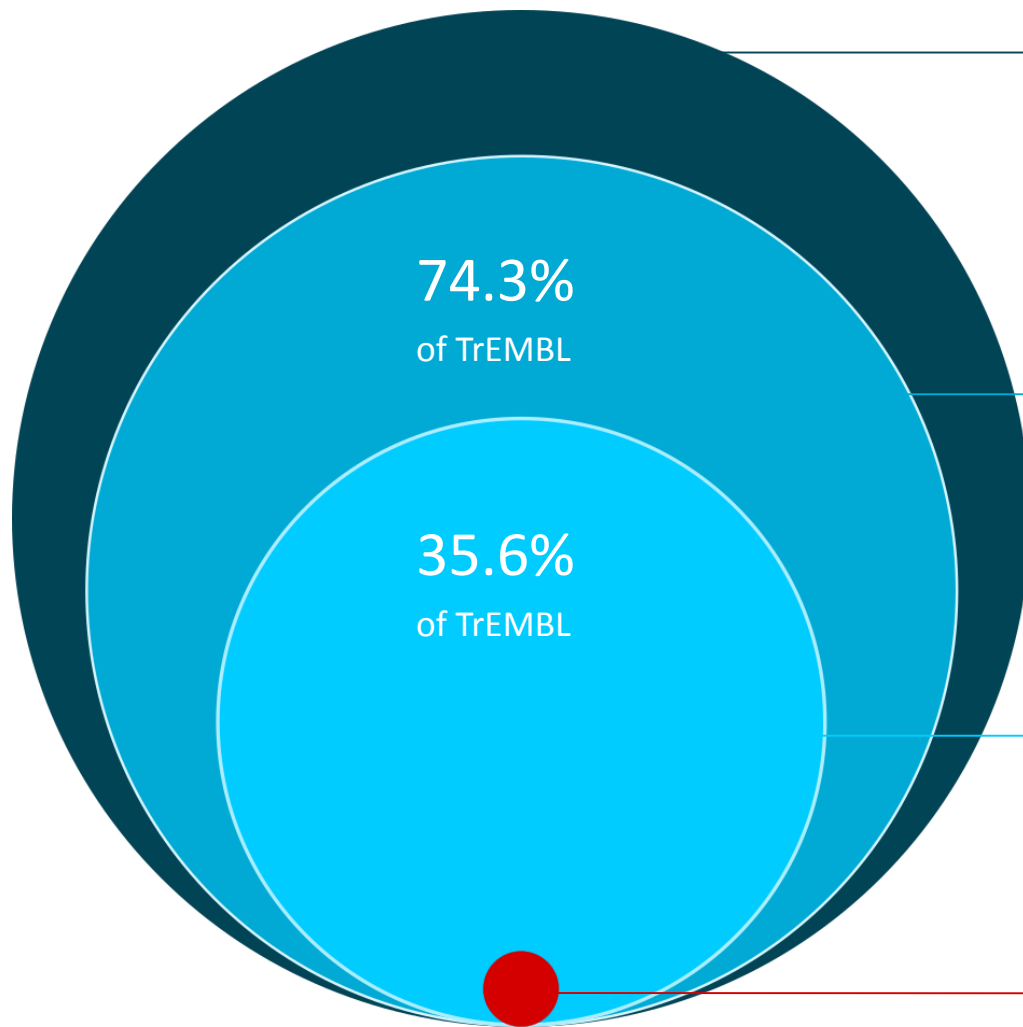


Rules are created by curators.
(HAMAP, PIR name and site rules, and RuleBase rules)

Rules are created by a modified **Decision Tree** algorithm.

Additional information

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



126,780,198 TrEMBL

Non-validated names from submitter;
no further annotations

95,576,000

Sequence feature annotation
signal peptides, transmembrane
peptides, domains

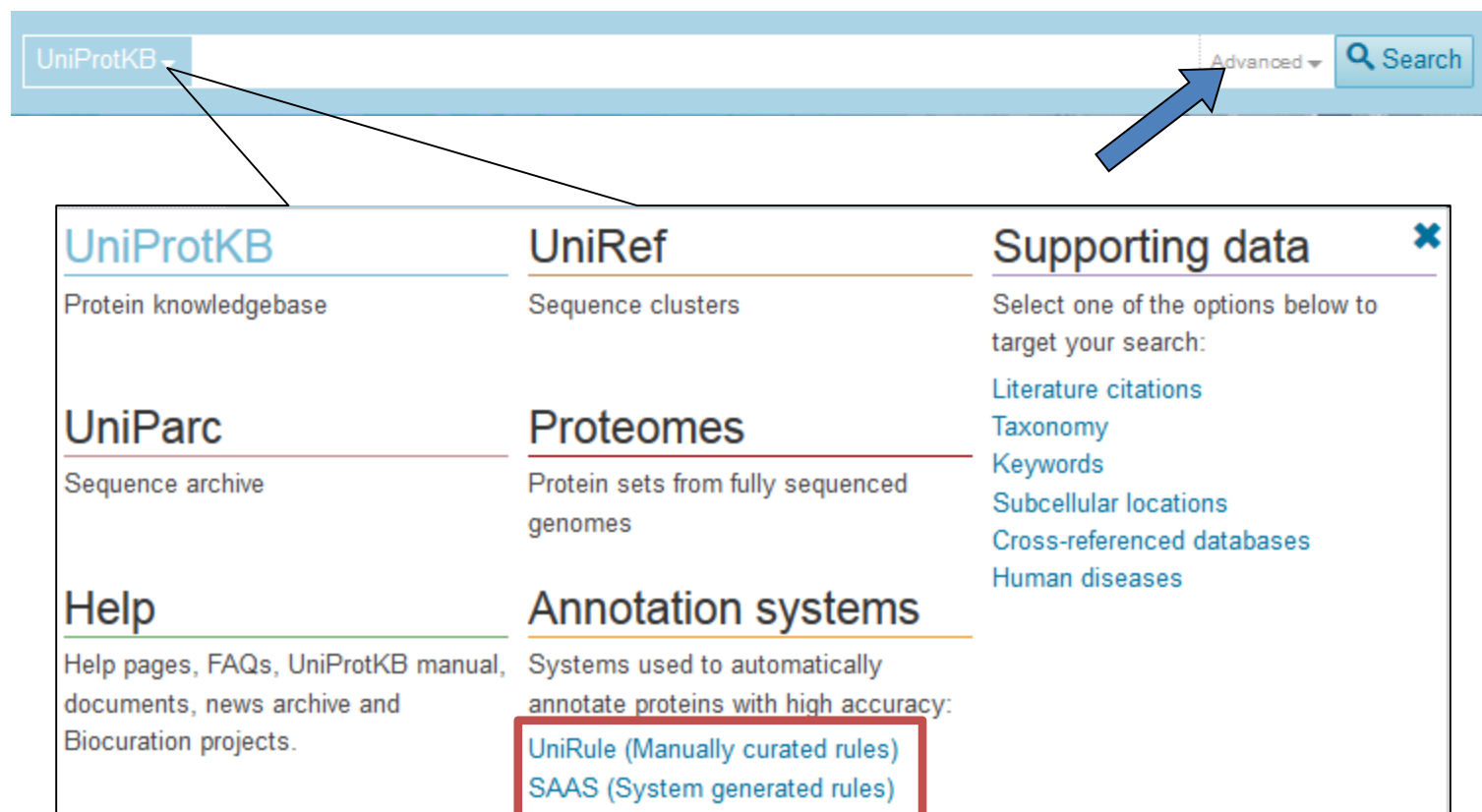
45,756,989

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

558,590 Swiss-Prot

Release 2018_09

How to access automatic annotation data?



The image shows a screenshot of the UniProtKB search interface. At the top, there is a search bar with the text "UniProtKB" on the left, a dropdown menu labeled "Advanced" with a downward arrow, and a "Search" button with a magnifying glass icon. A blue arrow points from the search bar area down to a detailed view of the navigation options. This view is organized into a grid of six panels:

- UniProtKB**: Protein knowledgebase
- UniRef**: Sequence clusters
- Supporting data**: Select one of the options below to target your search:
 - Literature citations
 - Taxonomy
 - Keywords
 - Subcellular locations
 - Cross-referenced databases
 - Human diseases
- UniParc**: Sequence archive
- Proteomes**: Protein sets from fully sequenced genomes
- Annotation systems**: Systems used to automatically annotate proteins with high accuracy:
 - UniRule (Manually curated rules)
 - SAAS (System generated rules)
- Help**: Help pages, FAQs, UniProtKB manual, documents, news archive and Biocuration projects.

Using advanced search in UniRule

Searching in **UniRule (Expertly curated rules)** ✕

Term

Protein name [DE] ▼ D-aminoacyl-tRNA deacylas 🗑️

AND ▼

Term

All ▼

- All
- Protein name [DE]
- Gene name [GN]
- Organism [OS]
- Taxonomy [OC]
- Function**
- Subcellular location
- PTM/Processing
- Expression
- Interaction
- Family and Domains
- Cross-references
- Gene Ontology [GO]
- Keyword [KW]

🗑️ + 🔍



Evaluating information in UniProtKB entries using its source

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

- manual assertions:

 1 Publication ▼  By similarity ▼  UniRule annotation ▼  Imported ▼

- automatic assertions:

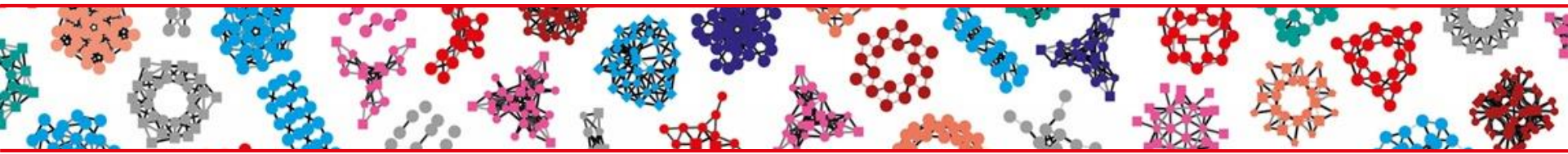
 UniRule annotation ▼  Imported ▼

Filtering UniProtKB results for assertion method

The screenshot shows the UniProtKB search interface. At the top, there is a search bar labeled "Searching in UniProtKB" and a "Help" button. Below the search bar, there are two columns of filters: "Function" and "Function [CC]". The "Function" filter is currently set to "Any". The "Evidence" filter is currently set to "Any assertion method". A dropdown menu is open for the "Evidence" filter, showing the following options:

- Any assertion method
- Any**
- Any assertion method
- Any manual assertion** (highlighted)
- Any automatic assertion
- Any experimental assertion
- Manual Assertions**
- Experimental
- Non-traceable author statement
- Curator inference
- Sequence similarity
- Sequence model
- Combinatorial
- Imported information
- Automatic Assertions**
- Sequence model
- Combinatorial

The interface also includes a search button (magnifying glass icon) and several trash icons for deleting filters. The "AND" operator is selected for the search criteria.



- 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



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

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

[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.





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.

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 [\[help\]](#)

- 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:

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

Receive your results by email



ScanProsite Results Viewer

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

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

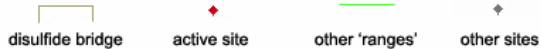
found: 13 hits in 1 sequence

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-catalytic light chain; Full=Enteropeptidase catalytic light chain; Flags: Precursor; . *Homo sapiens (Human)*

```
MGSKRGISSRHHSLSSYEIMFAALFAILVVLCAGLIAVSVCLTIKESQRGAALGQSSHEARATFKITS  
GVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQFENGSIIVVFDLFFAQWVS  
DENVKEELIQGLEANKSSQLVTFHIDLNSVDILDKLTITSHLATPGNVSEICLPGSSPCTDALTCI  
KADLFCDGEVNCPDGSDENKMCATVCDGRFLITGSSGSFQATHYKPKSETSVVCQWIIRVNOGLS  
IKLSFDDFNYYTDILDIEYEGVSSKILRASIWETNPGTIRIFSNQVTATFLIESDESDYVGFNAT  
YTAFNSSSELNNYEKINCNEFDGFCFWVQDLNDDNEWERIQGSTIFSPFTGPNFDHTFGNASGFYIST  
PTGPGGRQERVGLLSPLDPTLEPACLSFWYHMYGENVHKLSINISNDQNMKIVFQKEGNYGDNW  
NYGQVTLNETVFKVAFNAFKNKILSDIALDDISLTYGICNGSLYPEPTLVPTPPPELPTDCGGPF  
ELWEPNITFSSINFPNSYPNLAFVCVWILNAQKGNIQLHFQFDLENINDVVEIRDGEADSLLLA  
VYTGPGFVKDVFSTINRMTVLLITNDVLRGGFKANFTTGYHLGIPPECKADHFQCKNGECVPLVN  
LCDGHLHCEDGSDAECVRFNGTINNNGLVRFRIQSIWHTACAENWTQISNDVQCQLLGLGSGNS  
SKPIFPDGGPFVKLNTAPDGLIILTPSQCLQDSLRLQCNHKSCKGLLAAQDITPKIVGGSNAK  
EGAWFVWVGLYYGGRLLCGASLVSSDWLVSAAHCVYGRNLEPSKWTAILGLHMKSNLTSFQIVPRL  
IDEIVINPHYNRRRKNNDIAMMHLEFKVNYTDYIQPICLPEENQVFPGRNCSIAGWGTVVYQGIT  
ANILQEAADVPLLSNERCQQMPEYNI TENMICAGYEEGGIDSCQDGGSGPLMCQENNRWFLAGVTS  
FGYKCALPNRPGVYARVRSFTEWISFLH
```

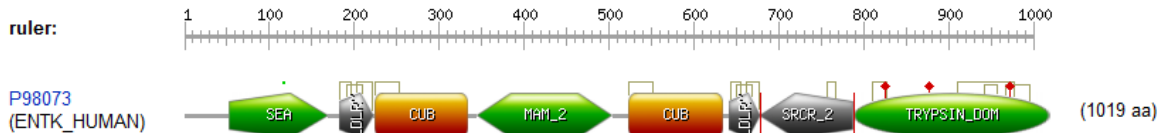
Legend:



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]

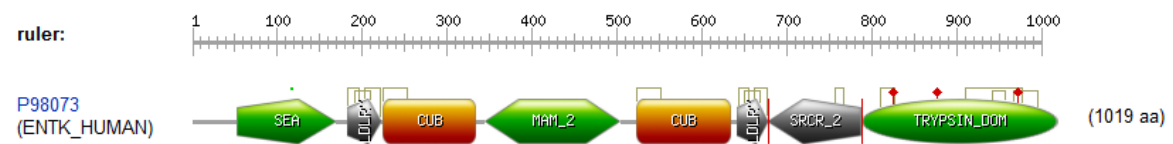
Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



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-catalytic light chain; 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-catalytic light chain; Full=Enteropeptidase catalytic light chain; Flags: Precursor; . *Homo sapiens (Human)*

PS50024 SEA SEA domain profile :

54 - 169: score = 32.979

QSHEARATFKITaGVTYNPNLQDKLSVDFKVLAFDLQQMIDEIFLSSNLKNEYKNSRVLQ
 FEN--GSIIIVVFDLFFAQWVSD--ENVKEELIQGLEANKKsQLVTFHIDLNSVDILDKLT

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

E L P G S S P C T D a L T I K A D L F C D G E V N C P D G S D e d n R M C A

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

P C K A D H F Q C K N G E C V P L V N L C D G H L H C E D G S D E A D C V

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

C D G R F L L I g s S G S F Q A T H Y P K --- P S E t a V V C Q W I I R V N Q G L -- S I K L S F D D - F N T Y ---
 -- Y T D I L D I Y E G v g s S K I L R A S I W E T N P G T I R I F S N Q V T A T F L I E S D E S D y V G F N A T Y T A
 F

Predicted features:

DISULFID	225	253		[condition: C-x*-C]
----------	-----	-----	--	---------------------



Search HAMAP Search

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	Term	Last updated	
All	<input type="text"/>	29-Apr-2015 for UniProt release 2015_05	
<input type="button" value="Search"/> <input type="button" value="Reset"/>		Number of family profiles	2'015
e.g. recA , MF_00191 , Iron		Number of annotation rules	2'035
The wildcard * is supported.			

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

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](#).



Browse annotation rules

Home | [Browse](#) | HAMAP-Scan | Proteomes | Documents | Downloads

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	ksmA / 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

Search HAMAP

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.

*Required field

HAMAP-Scan: 'Scan' output



HAMAP-Scan results

[Home](#) | [Search](#) | [HAMAP-Scan](#) | [Proteomes](#) | [Documents](#) | [Downloads](#)

Showing results for 10 sequences

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

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'



Request code: LJJ

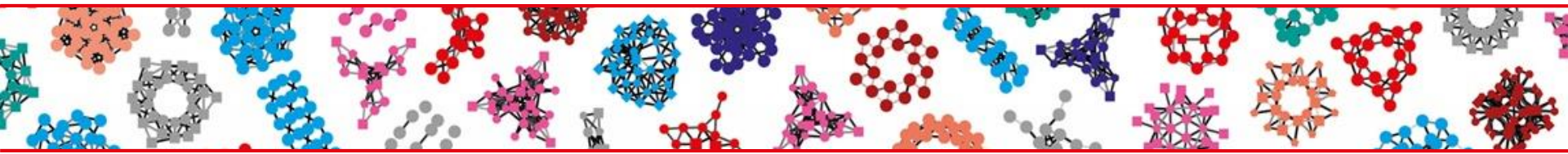
Status: Completed

Input file: 4305 entries

Result files: Annotated matches
"Not annotated entries which app

HA

```
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
LLDGDNVVRHG 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**