

Annotation propagation in UniProtKB

Alan Bridge, Swiss-Prot group

alan.bridge@isb-sib.ch

June 16th 2010



Swiss Institute of
Bioinformatics

Talk outline

- What?

Annotation in UniProtKB, and what do we propagate?

- Why?

Need to propagate annotation in UniProtKB

- How?

Annotation propagation systems in UniProtKB

- Current status and future plans

Talk outline

- **What?**

Annotation in UniProtKB, and what do we propagate?

- **Why?**

Need to propagate annotation in UniProtKB

- **How?**

Annotation propagation systems in UniProtKB

- **Current status and future plans**

What? UniProtKB annotation content

★ Reviewed, UniProtKB/Swiss-Prot **B1ITB3** (IDI_ECOLC)

Last modified April 20, 2010. Version 19.  [History...](#)

Contribute

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

 Clusters with [100%](#), [90%](#), [50%](#) identity |  [Documents](#) (2) |  [Third-party data](#) |  [Customize display](#)

[TEXT](#) [XML](#) [RDF/XML](#) [GFF](#)
[FASTA](#)

Length: 135

[Names and origin](#) · [Protein attributes](#) · [General annotation \(Comments\)](#) · [Ontologies](#) · [Sequence annotation \(Features\)](#) · [Sequences](#) · [References](#)
· [Cross-references](#) · [Entry information](#) · [Relevant documents](#)

Names and origin [Hide](#)

[Hide](#) | [Top](#)

Protein names	<i>Recommended name:</i> Isopentenyl-diphosphate Delta-isomerase Short name=IPP isomerase EC= 5.3.3.2 <i>Alternative name(s):</i> Isopentenyl pyrophosphate isomerase IPP:DMAPP isomerase
Gene names	Name: idi Ordered Locus Names:EcolC_0820
Organism	Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) [Complete proteome] [HAMAP]
Taxonomic identifier	481805 [NCBI]
Taxonomic lineage	Bacteria > Proteobacteria > Gammaproteobacteria > Enterobacteriales > Enterobacteriaceae > Escherichia

What? UniProtKB annotation content

★ Reviewed, UniProtKB/Swiss-Prot **B1ITB3** (IDI_ECOLC)

Last modified April 20, 2010. Version 19.  [History...](#)

Contribute

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

General annotation (Comments) [Hide](#)

[Hide](#) | [Top](#)

Function	Catalyzes the 1,3-allylic rearrangement of the homoallylic substrate isopentenyl (IPP) to its highly electrophilic allylic isomer, dimethylallyl diphosphate (DMAPP) By similarity . HAMAP MF_00202
Catalytic activity	Isopentenyl diphosphate = dimethylallyl diphosphate. HAMAP MF_00202
Cofactor	Binds 1 magnesium ion per subunit. The magnesium ion binds only when substrate is bound By similarity . HAMAP MF_00202 Binds 1 manganese ion per subunit By similarity . HAMAP MF_00202
Pathway	Isoprenoid biosynthesis ; dimethylallyl diphosphate biosynthesis ; dimethylallyl diphosphate from isopentenyl diphosphate: step 1/1 . HAMAP MF_00202
Subunit structure	Homodimer By similarity . HAMAP MF_00202
Subcellular location	Cytoplasm By similarity . HAMAP MF_00202 .
Sequence similarities	Belongs to the IPP isomerase type 1 family . Contains 1 nudix hydrolase domain .

What? UniProtKB annotation content

★ Reviewed, UniProtKB/Swiss-Prot **B1ITB3** (IDI_ECOLC)

Last modified April 20, 2010. Version 19.  [History...](#)

[Contribute](#)

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

Ontologies [Hide](#)

[Hide](#) | [Top](#)

Keywords

Biological process	Isoprene biosynthesis
Cellular component	Cytoplasm
Ligand	Magnesium Manganese Metal-binding
Molecular function	Isomerase
Technical term	Complete proteome

Gene Ontology (GO)

Biological process	isoprenoid biosynthetic process Inferred from electronic annotation. Source: HAMAP
Cellular component	cytoplasm Inferred from electronic annotation. Source: UniProtKB-SubCell
Molecular function	hydrolase activity Inferred from electronic annotation. Source: InterPro isopentenyl-diphosphate delta-isomerase activity Inferred from electronic annotation. Source: HAMAP metal ion binding Inferred from electronic annotation. Source: UniProtKB-KW

What? UniProtKB annotation content

★ Reviewed, UniProtKB/Swiss-Prot **B1ITB3** (IDI_ECOLC)

Last modified April 20, 2010. Version 19.  [History...](#)

Contribute

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

Sequence annotation (Features) [Hide](#)

[Hide](#) | [Top](#)

Feature key	Position(s)	Length	Description
-------------	-------------	--------	-------------

Graphical view

Feature identifier










Molecule processing

<input type="checkbox"/>	Chain	1 – 182	182	Isopentenyl-diphosphate Delta-isomerase HAMAP MF_00202		PRO_0000205249
--------------------------	-------	---------	-----	---	---	----------------

Regions

<input type="checkbox"/>	Domain	30 – 164	135	Nudix hydrolase		
--------------------------	--------	----------	-----	-----------------	---	--

Sites

<input type="checkbox"/>	Active site	67	1	HAMAP MF_00202		
<input type="checkbox"/>	Active site	116	1	HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	25	1	Manganese HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	32	1	Manganese HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	67	1	Magnesium; via carbonyl oxygen HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	69	1	Manganese HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	87	1	Magnesium HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	114	1	Manganese HAMAP MF_00202		
<input type="checkbox"/>	Metal binding	116	1	Manganese HAMAP MF_00202		

What? Annotation propagation in UniProtKB

General annotation

Annotation type	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
Biotechnology	No

Feature annotation

Annotation	Propagated?
KW	Yes
GO	Yes
Regions of interest	Yes
Active site	Yes
Ligand-binding	Yes
Processing	Yes
Modified residues	Yes
Ambiguities	No
Conflicts	No
Natural variants	No
Isoforms	No
2D structural elements	No

NB: these lists are not exhaustive

Talk outline

- What?

Annotation in UniProtKB, and what do we propagate?

- **Why?**

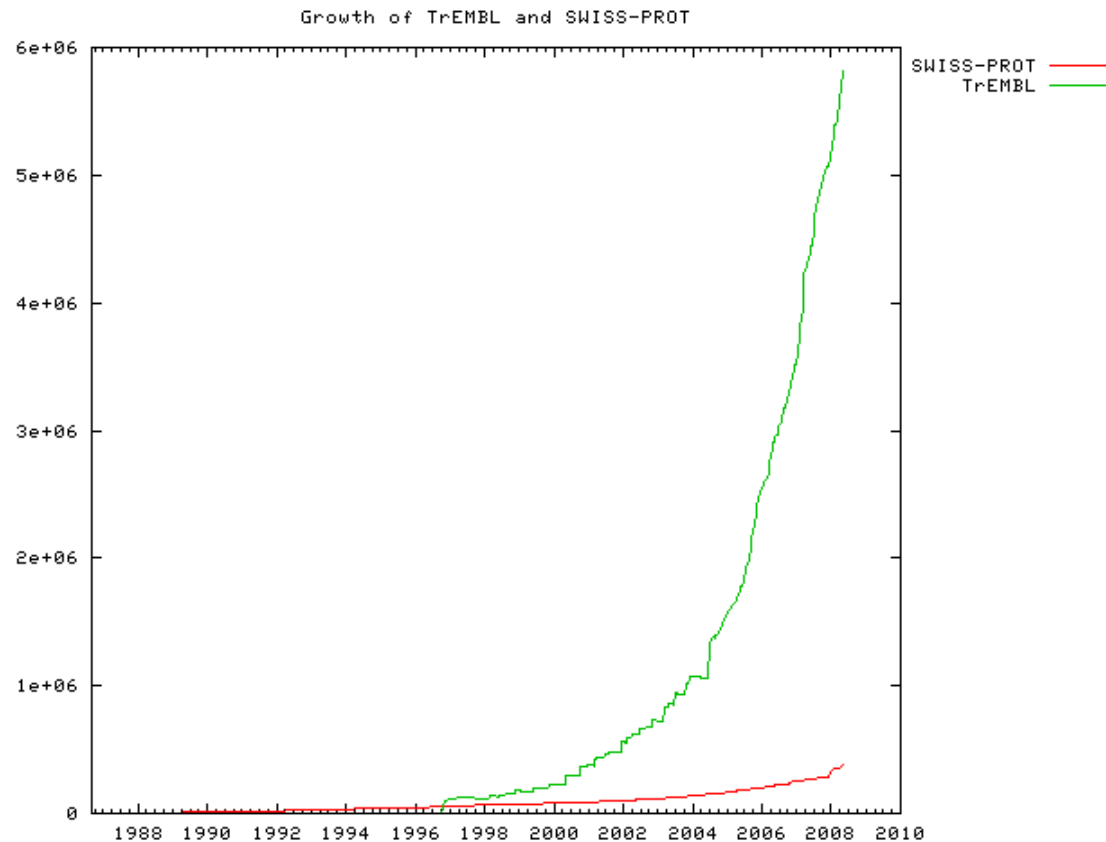
Need to propagate annotation in UniProtKB

- How?

Annotation propagation systems in UniProtKB

- Current status and future plans

Why? Data increase in UniProtKB



Talk outline

- What?

Annotation in UniProtKB, and what do we propagate?

- Why?

Need to propagate annotation in UniProtKB

- **How?**

Annotation propagation systems in UniProtKB

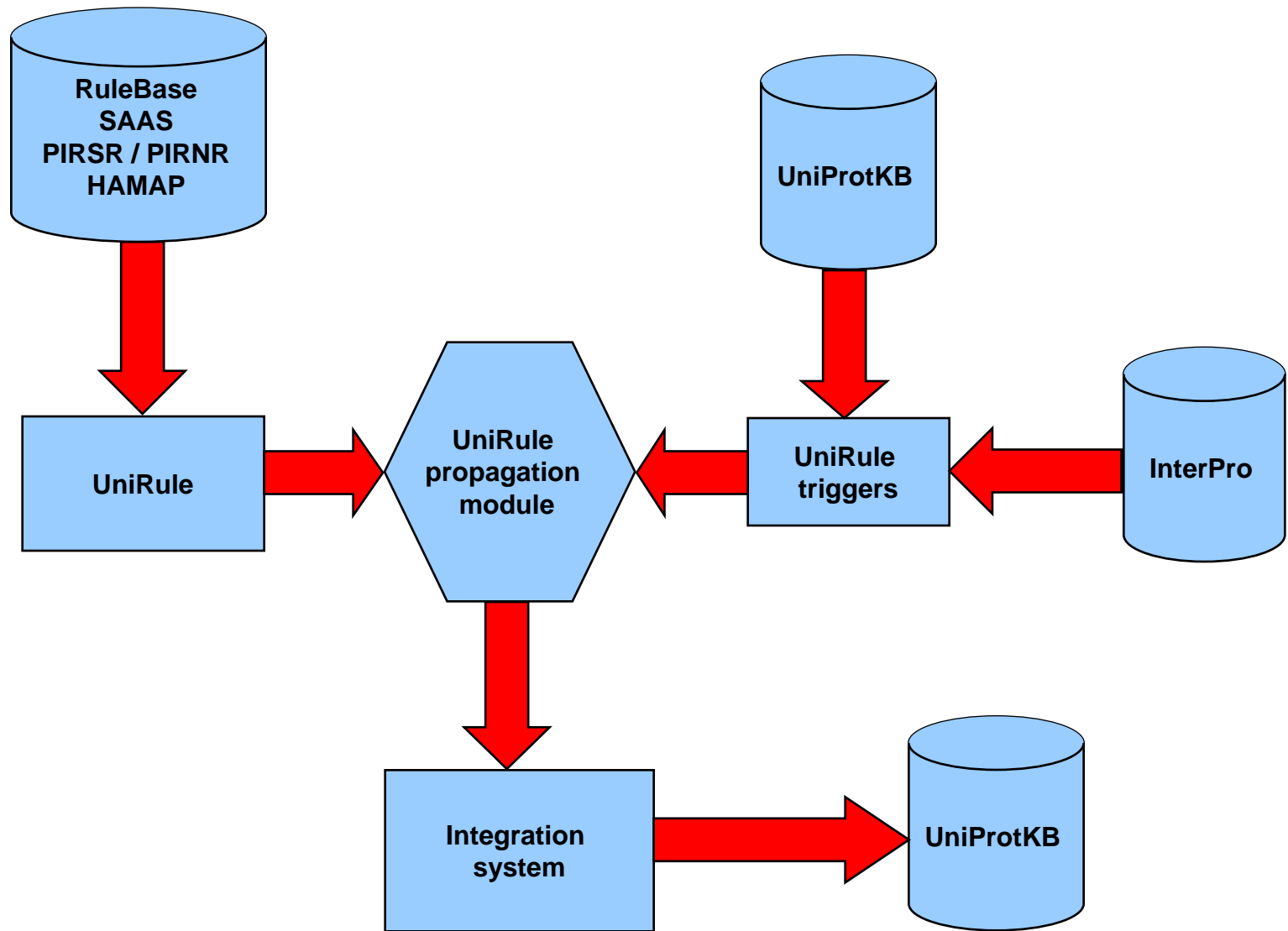
- Current status and future plans

How? UniProtKB annotation propagation

	Rule creation	Trigger	Annotations				
			Protein names	Comments	KW	GO terms	Features
RuleBase	Manual	InterPro	Yes	Yes	Yes	No	No
PIR Name / Site rules	Manual	PIRSF	Yes	Yes	Yes	No	Yes
HAMAP	Manual	HAMAP	Yes	Yes	Yes	Yes	Yes
SAAS	Automatic*	InterPro	No	Yes	Yes	No	No

* C4.5 decision tree algorithm applied each release. Can form starting point for manual rules.

How? Annotation propagation by UniRule



How? UniProtKB annotation propagation

	Rule creation	Trigger	Annotations				
			Protein names	Comments	KW	GO terms	Features
RuleBase	Manual	InterPro	Yes	Yes	Yes	No	No
PIR Name / Site rules	Manual	PIRSF	Yes	Yes	Yes	No	Yes
HAMAP	Manual	HAMAP	Yes	Yes	Yes	Yes	Yes
SAAS	Automatic*	InterPro	No	Yes	Yes	No	No

* C4.5 decision tree algorithm applied each release. Can form starting point for manual rules.

How? HAMAP – a UniRule component for UniProtKB/Swiss-Prot

- High-quality Automated and Manual Annotation of microbial Proteomes
- Description: A semi-automated pipeline system, dedicated to high-throughput, high-quality annotation of proteins from complete microbial proteomes.
- Allows us to annotate automatically, yet with a very high level of quality – includes multiple consistency checks designed to mimic those performed by human curators.
- Scope is currently Bacteria, Archaea, plastids, also being extended to other taxonomic groups.

How? Components of the HAMAP system

1. Two databases
 - a) The proteomes database
 - b) The family database
2. The annotation pipeline

1a) Proteomes database

Principle:

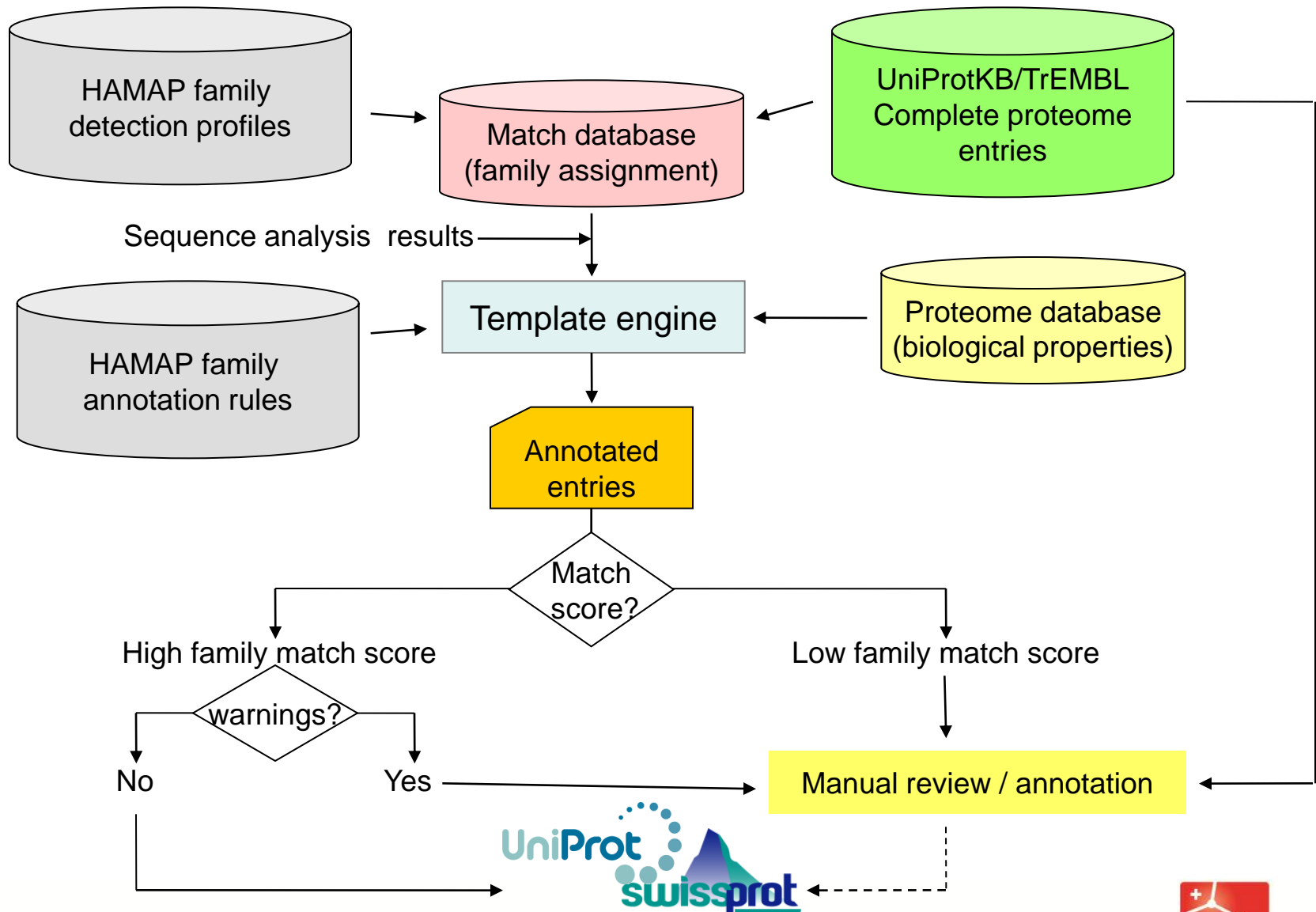
- An “identity card” is created for each completely sequenced genome in UniProtKB/TrEMBL.
- This includes information on the biology, genome and taxonomy of the organism concerned.
- The information contained in this identity card can be used during the process of annotation.
- The proteomes database includes only fully sequenced and assembled genomes submitted to the public databases, i.e. no whole genome shotgun sequences.

1b) Family database

Principle:

- For each protein family a **profile** and an associated **annotation rule** are created.
 - The **profile** is used to detect putative matches and for the propagation of annotated sequence features.
 - The **annotation rule** defines annotations and specifies the conditions that must be satisfied if they are to be applied.
- Therefore a protein entry must match the profile and satisfy the conditions that are specified in the rule before annotation occurs.
- Note that a protein family is generally composed of a set of presumed orthologs, but may include paralogs (duplications), providing these match the profile and satisfy the conditions specified by the rule.

2) HAMAP annotation pipeline



How? HAMAP annotation rule creation

- All **characterized proteins** that belong to the family are **manually annotated** according to UniProtKB standards.
- **Similarity searches** and **manual selection** of a set of member **sequences** that will be used to build the **seed alignment**.
- **Manual correction of problematic sequences** (erroneous gene model predictions, initiators and frameshifts) - exclude fragments.
- **Profile construction and testing (scan vs UniProtKB)**
 - common function of profile matches
 - clear cut-off between family members and non-members
- Assuming a specific profile can be generated, then extract common annotations and encode them in a rule.

How? HAMAP annotation rule contents

Propagated annotation

Identifier, protein and gene names

Identifier IDI

Protein name RecName: Full=Isopentenyl-diphosphate Delta-isomerase;
Short=IPP isomerase;
EC=5.3.3.2;

Comments

FUNCTION: Catalyzes the 1,3-allylic rearrangement of the homoallylic substrate isopentenyl (IPP) to its highly electrophilic allylic isomer, dimethylallyl diphosphate (DMAPP) (By similarity).

CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl diphosphate.

Keywords

Cytoplasm, Isomerase, Isoprene biosynthesis.

Gene Ontology

GO:0004452; Molecular function: isopentenyl-diphosphate delta-isomerase activity.

Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	

How? HAMAP annotation rule conditions

Sequence alignment showing HAMAP annotation rule conditions. Red arrows point to specific positions (100 and 160) in the alignment.

```

      *      80      *      100      140      *      160      *
IDI_ECOLI : F---NAKGQLIVTRRALSKKAMPGVWITNSVCGHP : 70 : PD--FRYRAT-DPSGIVENEVCPVFAARTTSA--- : 128
IDI_HALMA : F---DENDRILLACRAAMKRLWDTHDGTVAHP : 81 : TDRFEYKRYRYENAGL-EUEVCANLQATLHDT-- : 142
IDI_HALSA : F---DEDGRVLLAQRADRKRLWDTHDGTVAHP : 96 : DR-FEYKRRYLDEGL-EUEVCANLQATLHDT-S- : 157
IDIi_PHOLL : I---NSKNEVYIQRAQSKLLMPGYWNSYCSHP : 68 : K--EYKREKYKDVGY-EHELCHVEVVFDTTP--- : 126
IDI_JANSC : L---RDGDVLLQRRAMCKYHTPGLWANTCCTHP : 65 : RDTVEYRAD-VGGGLIEHEVWDIEVGEPSG--- : 122
IDI_RHOCA : ----TRGNKVLLQQRALSKYHTPGLWANTCCTHP : 66 : MGQLEYRAD-VNNGMIEHEVWEVETAEAEPEGIE- : 125
IDI_ROSDO : L---KGDVLLMQRAMCKYHTPGLWINTCCTHP : 65 : RHHLEYRAD-VGGGLIEHEVWDIEVVAEADETLV- : 124
IDI_SILPO : V---RDMDILLQRRALCKYHTPGLWANTCCTHP : 65 : RHRLEYHAD-VGNMGVENEVWDIEVLAHVRGPLO- : 124
IDI_DINSH : M---RGPETLLQRRALCKYHTPGLWANTCCTHP : 65 : RDRIEYRAD-VGNGLIEHEVWDIEVIAEAPANLK- : 124
IDI_RHOSS : M---AGESVLLQRAAGKYHTPGLWANTCCTHP : 65 : ADQVEYRAD-VGSGLIEHEVWDIEVVAEAPQDLP- : 124
IDI_RHOS4 : M---AGEAVLLQRAAGKYHTPGLWANTCCTHP : 65 : ADRVEYRAD-VGNGLIEHEVWDIEVVAEAPSDLP- : 124
  
```

Propagated annotation

Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	

How? HAMAP annotation rule conditions

```

      *           80           *           100           140           *           160           *
IDI_ECOLI  : F---NAKGQLLVTRRALSKKAMPGVWTTNSVCGHP : 70 : PD-FRYRAT-DPSGIVENEVCPVFAARTTSA--- : 128
IDI_HALMA  : F---DENDRILLACRAAMKRLWDTHMDGTVAASHP : 81 : TDRFEYKRYRYENAGL-EUEVCALQATLHDT-- : 142
IDI_HALSA  : F---DEDGRVLLAQRADRKRLWDTHMDGTVAASHP : 96 : DR-FEYKRRYLDEGL-EUEVCALQATLHDT-S- : 157
IDI1_PHOLL : I---NSKNEVYIQRAQSKLLMPGYWNSYCSHP : 68 : K--FOYREKYKDVGY-EHELCHVEVVFDTTP--- : 126
IDI_JANSC  : L---RDGDVLLQRRAMCKYHTPGLWANTCCTHP : 65 : RDTVEYRAD-VGGGLIEHEVWDIEVGEMPSG--- : 122
IDI_RHOCA  : ----TRGNKVLQQRALSKYHTPGLWANTCCTHP : 66 : MGQLEYRAD-VNNGMIEHEVWEVETAEAEPEGIE- : 125
IDI_ROSDO  : L---KGDVLLQRRAMCKYHTPGLWINTCCTHP : 65 : RHHLEYRAD-VGGGLIEHEVWDIEVVAEADETLV- : 124
IDI_SILPO  : V---RDMDILLQRRALCKYHTPGLWANTCCTHP : 65 : RHRLEYHAD-VGNMGVENEVWDIEVLAHVRGPLO- : 124
IDI_DINSH  : M---RGPETLIQRRALCKYHTPGLWANTCCTHP : 65 : RDRIEYRAD-VGNGLIEHEVWDIEVIAEAPANLK- : 124
IDI_RHOSS  : M---AGESVLIQRAAGKYHTPGLWANTCCTHP : 65 : ADQVEYRAD-VGSGLIEHEVWDIEVVAEAPQDLP- : 124
IDI_RHOS4  : M---AGEAVLIQRAAGKYHTPGLWANTCCTHP : 65 : ADRVEYRAD-VGNGLIEHEVWDIEVVAEAPSDLP- : 124
  
```

Propagated annotation


Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	

How? HAMAP annotation rule conditions

★ Reviewed, UniProtKB/Swiss-Prot **Q5UX45** (IDI_HALMA)

Last modified April 20, 2010. Version 44.  [History...](#)

[Contribute](#)

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

General annotation (Comments) [Hide](#)

[Hide](#) | [Top](#)

Function	Catalyzes the 1,3-allylic rearrangement of the homoallylic substrate isopentenyl (IPP) to its highly electrophilic allylic isomer, dimethylallyl diphosphate (DMAPP) (By similarity) . HAMAP MF_00202
Catalytic activity	Isopentenyl diphosphate = dimethylallyl diphosphate. HAMAP MF_00202
Cofactor	Binds 1 magnesium ion per subunit. The magnesium ion binds only when substrate is bound (By similarity) . HAMAP MF_00202 Binds 1 manganese ion per subunit (By similarity) . HAMAP MF_00202
Pathway	Isoprenoid biosynthesis; dimethylallyl diphosphate biosynthesis; dimethylallyl diphosphate from isopentenyl diphosphate: step 1/1. HAMAP MF_00202
Subcellular location	Cytoplasm (By similarity) HAMAP MF_00202 .
Sequence similarities	Belongs to the IPP isomerase type 1 family. Contains 1 nudix hydrolase domain .
Caution	Could lack activity as the potential active site Cys residue in position 78 is replaced by an Ala.

How? HAMAP annotation rule case statements

Propagated annotation

Identifier, protein and gene names

Identifier IDI

Protein name RecName: Full=Isopentenyl-diphosphate Delta-isomerase;
Short=IPP isomerase;
EC=5.3.3.2;

Comments

case <FTGroup:1>

COFACTOR: Binds 1 manganese ion per subunit (By similarity).

end case

case <Property:PHOTOSYN>

PATHWAY: Porphyrin biosynthesis; chlorophyll biosynthesis.

end case

case <OC:Enterobacteriaceae>

SUBUNIT: Homodimer (By similarity).

end case

Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	
METAL	25	25	Manganese (By similarity)	H	1
METAL	32	32	Manganese (By similarity)	H	1
METAL	69	69	Manganese (By similarity)	H	1

How? HAMAP annotation rule case statements

Propagated annotation

Identifier, protein and gene names

Identifier IDI

Protein name RecName: Full=Isopentenyl-diphosphate Delta-isomerase;
Short=IPP isomerase;
EC=5.3.3.2;

Comments

case <FTGroup:1>
COFACTOR: Binds 1 manganese ion per subunit (By similarity).
end case

case <Property:PHOTOSYN>
PATHWAY: Porphyrin biosynthesis; chlorophyll biosynthesis.
end case

case <OC:Enterobacteriaceae>
SUBUNIT: Homodimer (By similarity).
end case

Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	
METAL	25	25	Manganese (By similarity)	H	1
METAL	32	32	Manganese (By similarity)	H	1
METAL	69	69	Manganese (By similarity)	H	1

Sequence /
Profile match

How? HAMAP annotation rule case statements

Propagated annotation

Identifier, protein and gene names

Identifier IDI

Protein name RecName: Full=Isopentenyl-diphosphate Delta-isomerase;
Short=IPP isomerase;
EC=5.3.3.2;

Comments

case <FTGroup:1>

COFACTOR: Binds 1 manganese ion per subunit (By similarity).

end case

case <Property:PHOTOSYN>

PATHWAY: Porphyrin biosynthesis; chlorophyll biosynthesis.

end case

case <OC:Enterobacteriaceae>

SUBUNIT: Homodimer (By similarity).

end case

Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	
METAL	25	25	Manganese (By similarity)	H	1
METAL	32	32	Manganese (By similarity)	H	1
METAL	69	69	Manganese (By similarity)	H	1

Proteomes
database

How? HAMAP annotation rule case statements

Propagated annotation

Identifier, protein and gene names

Identifier IDI

Protein name RecName: Full=Isopentenyl-diphosphate Delta-isomerase;
Short=IPP isomerase;
EC=5.3.3.2;

Comments

case <FTGroup:1>

COFACTOR: Binds 1 manganese ion per subunit (By similarity).

end case

case <Property:PHOTOSYN>

PATHWAY: Porphyrin biosynthesis; chlorophyll biosynthesis.

end case

case <OC:Enterobacteriaceae>

SUBUNIT: Homodimer (By similarity).

end case

Entry
taxonomy

Features

From: IDI_ECOLI (Q46822)

Key	From	To	Description	Condition	FTGroup
ACT_SITE	67	67	By similarity	C	
ACT_SITE	116	116	By similarity	E	
METAL	25	25	Manganese (By similarity)	H	1
METAL	32	32	Manganese (By similarity)	H	1
METAL	69	69	Manganese (By similarity)	H	1

Summary

- HAMAP is one component of the UniRule system for UniProtKB
- Based on manually curated profiles and rules and proteome database
- **Designed for high specificity – the primary aim of HAMAP is to preserve accuracy of annotation**
- Conditional, context-dependent annotation with multiple checks and warnings, and manual review
- Propagation not restricted to orthologs but to family members that
 - exceed a defined threshold score
 - satisfy all conditions specified in the rule

Talk outline

- What?

Annotation in UniProtKB

- Why?

Need to propagate annotation in UniProtKB

- How?

Annotation propagation systems in UniProtKB

- **Current status and future plans**

Current status - HAMAP

For the current UniProt release:

- 1614 HAMAP annotation rules
- These annotated 307,162 proteins from 3,296 species in UniProtKB/Swiss-Prot
- HAMAP provided:
 - 651,511 GO annotations for 247,042 UniProtKB/Swiss-Prot entries
 - 942,316 GO annotations to 358,533 UniProtKB/TrEMBL entries
- All GO terms derived from HAMAP rules are assigned the evidence code IEA

Future plans

HAMAP

- Expand to cover other taxa including eukaryotes and viruses – we have over 120 rules affecting over 4500 eukaryotic entries.
- Improvement in the procedures for generation and update of profiles (PROSITE).
- Complete integration of HAMAP into InterPro.

UniRule

- Complete integration of all systems into one unified system – this will allow the application of all HAMAP rules to the whole of UniProtKB (fully automatic).
- Improved presentation and access – including provision of integrated web-pages and prediction server.

Acknowledgements

HAMAP

- ✓ **Group Leaders:** Ioannis Xenarios and Lydie Bougueleret
- ✓ **Curators:** Andrea Auchincloss, Elisabeth Coudert, Guillaume Keller, Catherine Rivoire, Ivo Pedruzzi, Emmanuel Boutet, Sylvain Poux
- ✓ **Developers:** especially: Edouard de Castro, Delphine Baratin, Nicole Redaschi, Thomas Kappler
- ✓ **Former contributors:** Corinne Lachaize, Tania Lima, Karine Michoud, Isabelle Phan

PROSITE






- ✓ Nicolas Hulo, Christian Sigrist, Lorenzo Cerutti, Beatrice Cucho

UniRule

- ✓ **Curators and developers at EBI:** Ricardo Antunes, Mark Bingley, Kati Laiho, Sam Patient, Klemens Pichler, Diego Poggioli, Mindi Sehra, Tony Wardell, Maria Jesus Martin, Claire O'Donovan
- ✓ **Curators and developers at PIR:** Chuming Chen, Hongzhan Huang, Peter McGarvey, Natalia Petrova, Cecilia Arighi, Winona C. Barker, Raja Mazumder, Darren Natale, C.R. Vinayaka, Lai-Su L. Yeh, Qinghua Wang, Uzoamaka Ugochukwu
- ✓ **The rest of the UniProt production team**
- ✓ **InterPro**

Questions welcome

A case where no HAMAP family was created

Accession	Entry name	Protein names	Status	Local alignment	Length	Identity (%)	Score	E-value
P53555	BLOK_BACSU	Lysine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.n2) (7- β -diamino-pelargonic acid aminotransferase) (DAPA aminotransferase)	★		448	100	904	0.0
O66557	BIOA_AQUAE	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA aminotransferase) (DAPA AT) (Diaminopelargonic acid synthase)	★		453	51	452	1 e-126
O58696	BIOA_METJA	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA aminotransferase) (DAPA AT) (Diaminopelargonic acid synthase)	★		461	49	446	1 e-124
P22805	BIOA_BACSH	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA aminotransferase) (DAPA AT) (Diaminopelargonic acid synthase)	★		455	43	407	1 e-112
O9ZKM5	BIOA_HELPJ	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA aminotransferase) (DAPA AT) (Diaminopelargonic acid synthase)	★		439	35	266	2 e-70

- **8-amino-7-oxononanoate aminotransferase** catalyzes the formation of the intermediate 7,8-diaminopelargonic acid (DAPA) from 7-keto-8-aminopelargonic acid (KAPA) with lysine or S-adenosylmethionine as the amino donor.
- *B.subtilis* does not utilize S-adenosylmethionine (SAM) as amino donor, but instead utilizes L-lysine. EC number varies according to donor.
- One family/profile, two types of annotation, no unambiguous way to classify the sequences.

Automatic annotation stats from TrEMBL 2010_07

TrEMBL release 2010_07 stats for all automatic annotation systems

11,109,684 TrEMBL entries

- RuleBase – 2,150,953 entries by 1028 rules
- SAAS – 1,606,237 entries
- HAMAP – 220,771 entries by 578 rules
- PIRSR – 10,735 entries by 45 rules
- PIRNR – 4,244 entries by 47 rules

Total coverage around 30% for all systems combined

TrEMBL annotation requires InterPro triggers – e.g. for HAMAP this stands at 621/1633 families.

Corrections and clarifications..

User update requests on HAMAP entries – last 2 years

- 2 families split into new subfamilies with distinct profiles and annotation
 - MF_01849 - Ribosomal RNA large subunit methyltransferase N
MF_01873 - Ribosomal RNA large subunit methyltransferase cfr
(specifically methylates different positions in adenine 2503 in 23S rRNA)
 - MF_00464 - S-adenosylmethionine decarboxylase
MF_01298 - Pyruvoyl-dependent arginine decarboxylase
- All other requests were for simple addition of new characterization information or minor corrections
- Based on comparisons with HOGENOM and other resources we have identified a list of around 100 HAMAP families that could conceivably be split – this is based on comparisons of profile matches – review of functional annotation of matches is proceeding