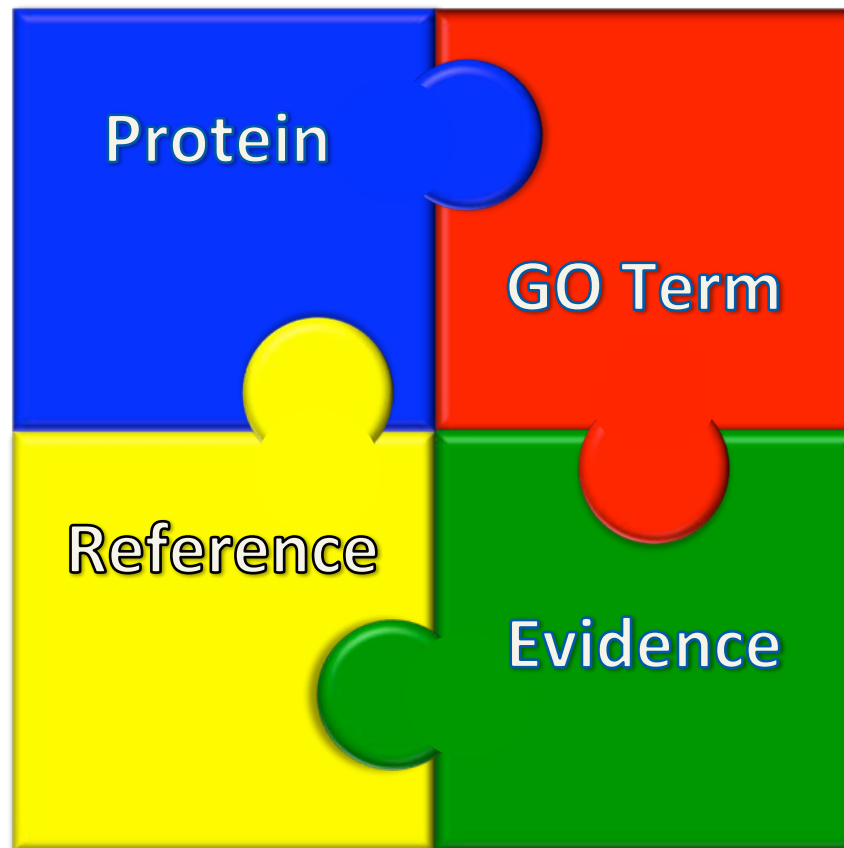


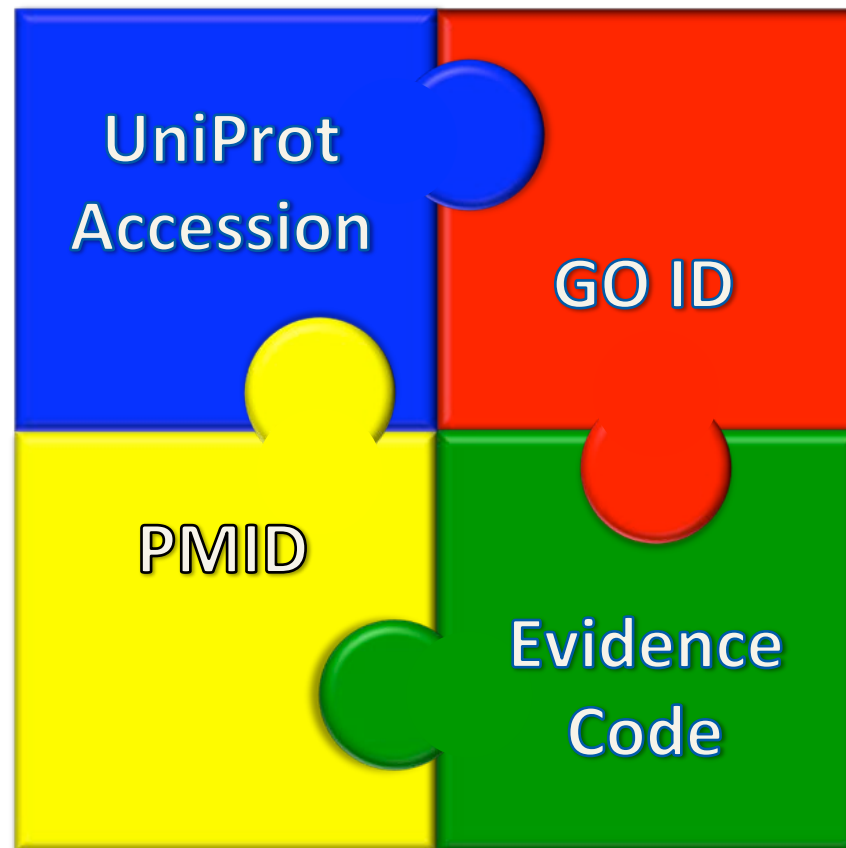
# GO ANNOTATION:

## COMPONENTS OF AN ANNOTATION

# Components of an Annotation: What You Need

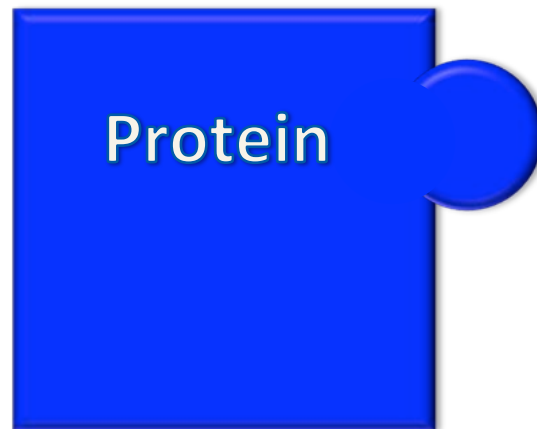


# Components of an Annotation: What You Find

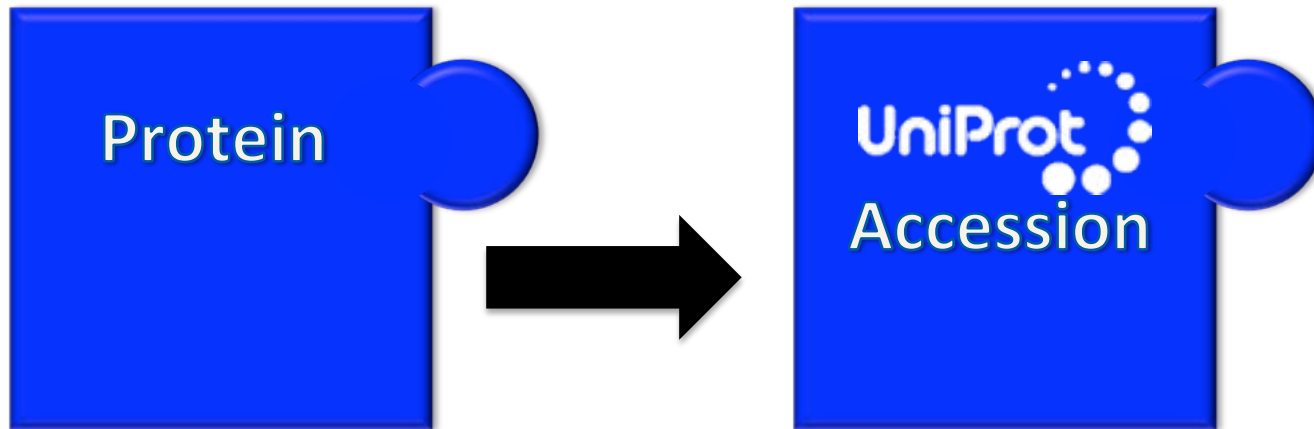


# Components of an Annotation:

Protein (Gene Product)

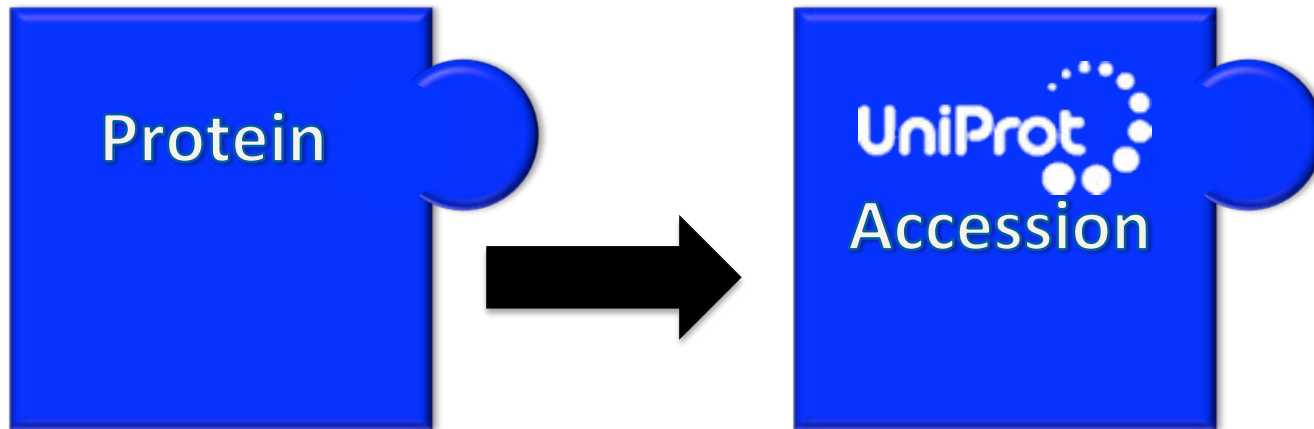


# Components of an Annotation: Protein



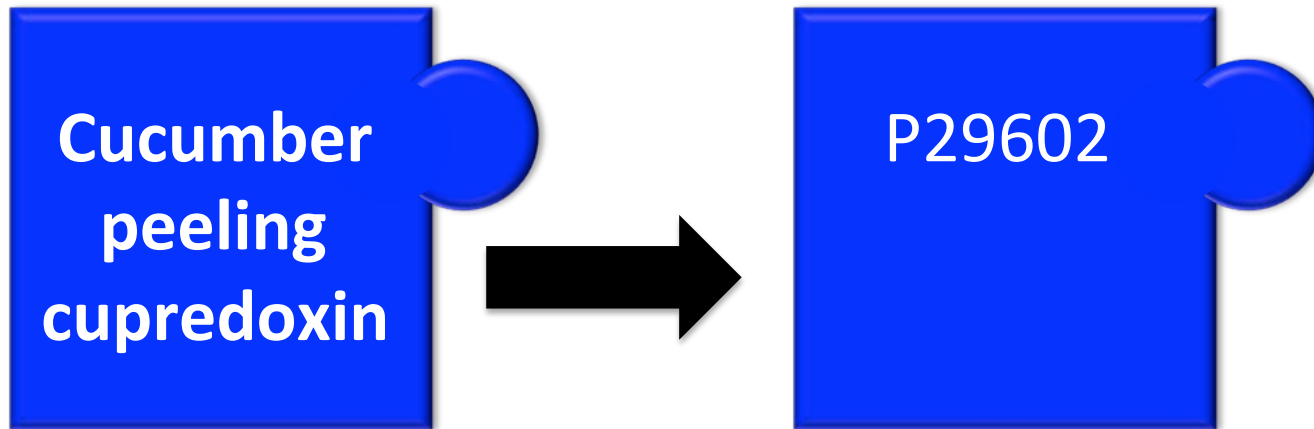
# Components of an Annotation:

## Protein



- Find the UniProt accession using <http://www.uniprot.org/>

# Components of an Annotation: Protein



- Find the UniProt accession using <http://www.uniprot.org/>
- Examples: [C4ZZY9](#), [A0A009DWN1](#)

# Components of an Annotation: Protein Using UniProt

The screenshot shows the UniProt website interface. At the top, there's a search bar with "UniProtKB" entered and a "Search" button. Below the search bar, there are navigation links: "BLAST", "Align", "Retrieve/ID mapping", "Peptide search", "Help", and "Contact". The main heading is "UniProtKB results". On the left, there's a "Filter by" section with two categories: "Reviewed (553,474) Swiss-Prot" and "Unreviewed (73,711,881) TrEMBL". Below this, there's a "Popular organisms" section listing "Human (156,710)" and "Rice (122,826)". The main content area displays a table of search results. The table has columns: "Entry", "Entry name", "Protein names", "Gene names", "Organism", and "Length". There are four rows of results shown.

UniProtKB results

Filter by:

- Reviewed (553,474) Swiss-Prot
- Unreviewed (73,711,881) TrEMBL

Popular organisms:

- Human (156,710)
- Rice (122,826)

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9D2V5	AAR2_MOUSE	Protein AAR2 homolog	Aar2	Mus musculus (Mouse)	384
Q811W1	AARD_MOUSE	Alanine and arginine-rich domain-co...	Aard	Mus musculus (Mouse)	167
P86434	AAS1_HUMAN	Putative uncharacterized protein AD...	ADORA2A-AS1 C22orf45	Homo sapiens (Human)	159
Q5X097	AASD1_RAT	Alanyl-tRNA editing	Aarsd1	Rattus norvegicus	412



# Components of an Annotation:

## Protein

### Using UniProt

The screenshot shows the UniProt website interface. At the top, the browser address bar displays 'www.uniprot.org/uniprot/'. Below the address bar, the UniProt logo is on the left, and a search bar is on the right, highlighted with a red rectangle. The search bar contains the text 'UniProtKB' and has a 'Search' button. Below the search bar, there are links for 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. On the right side of the header, there are links for 'Help' and 'Contact'. Below the header, the text 'UniProtKB results' is displayed. On the left side, there is a 'Filter by' section with two categories: 'Reviewed (553,474) Swiss-Prot' and 'Unreviewed (73,711,881) TrEMBL'. Below this, there is a 'Popular organisms' section with 'Human (156,710)' and 'Rice (122,826)'. The main content area shows a table of search results. The table has columns for 'Entry', 'Entry name', 'Protein names', 'Gene names', 'Organism', and 'Length'. The first four rows of the table are visible, showing results for 'AAR2\_MOUSE', 'AARD\_MOUSE', 'AAS1\_HUMAN', and 'AASD1\_RAT'. The table is paginated, showing '1 to 250 of 74,265,355' results.

www.uniprot.org/uniprot/

UniProt

UniProtKB

Search

BLAST Align Retrieve/ID mapping Peptide search

Help Contact

UniProtKB results

About UniProtKB

Basket

Filter by

Reviewed (553,474) Swiss-Prot

Unreviewed (73,711,881) TrEMBL

Popular organisms

Human (156,710)

Rice (122,826)

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9D2V5	AAR2_MOUSE	Protein AAR2 homolog	Aar2	Mus musculus (Mouse)	384
Q811W1	AARD_MOUSE	Alanine and arginine-rich domain-co...	Aard	Mus musculus (Mouse)	167
P86434	AAS1_HUMAN	Putative uncharacterized protein AD...	ADORA2A-AS1 C22orf45	Homo sapiens (Human)	159
Q5X097	AASD1_RAT	Alanyl-tRNA editing	Aarsd1	Rattus norvegicus	412

1 to 250 of 74,265,355

Show 250

# Components of an Annotation: Protein Using UniProt

UniProtKB - cupredoxin

BLAST Align Retrieve/ID mapping Peptide search

UniProtKB results

Filter by:

- Reviewed (723)
- Unreviewed (127,681)

Popular organisms:

- Human (448)
- Rice (225)
- A. thaliana (204)
- Zebrafish (66)

Entry	Entry name	Protein names	Gene names	Organism	Length
P00451	F8B_HUMAN	Coagulation factor VIII	F8B	Homo sapiens (Human)	2,351
P00N32	CUPR1_ARTEC	Probable GPI-anchored cupredoxin AR...	ARB_05732-1	Arthroderma benhamiae (strain ATCC MYA-4681 / CBS 112375) (Trichophyton mentagrophytes)	186
P04377	AZUP_ALCFA	Pseudoazurin		Alcaligenes faecalis	146
P29602	CPC_CUCSA	Cucumber peeling		Cucumis sativus	137

- Verify the organism (and strain)

# Components of an Annotation: Protein Using UniProt

UniProt

UniProtKB - cupredoxin cucumber

Advanced search Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

About UniProtKB Basket

## UniProtKB results

Filter by

- Reviewed (4) Swiss-Prot
- Unreviewed (157) TrEMBL

Popular organisms

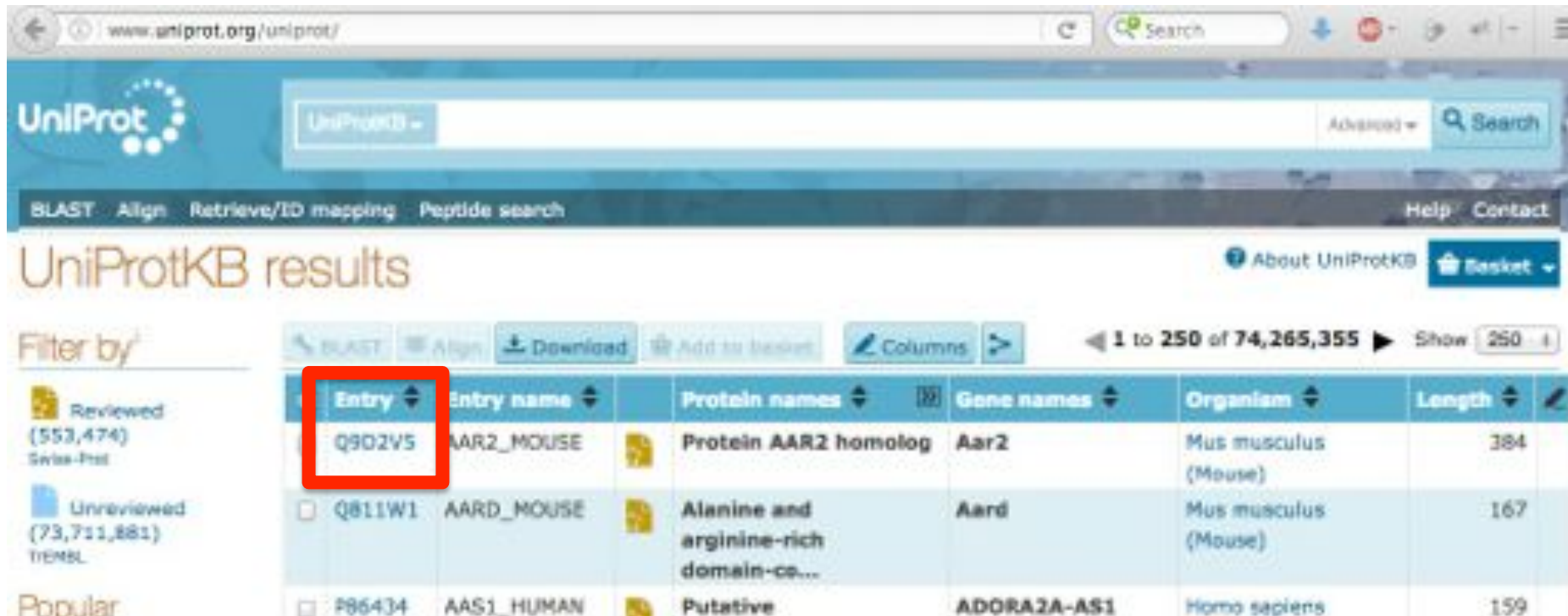
- CUCSA (94)
- GLYSO (1)
- RICCO (6)
- TRFJA (1)
- ANACO (3)

1 to 25 of 161 Show 20

Entry	Entry name	Protein names	Gene names	Organism	Length
P29602	CPC_CUCSA	Cucumber peeling cupredoxin		Cucumis sativus (Cucumber)	137
P14133	ASO_CUCSA	L-ascorbate oxidase		Cucumis sativus (Cucumber)	587
P00303	BABL_CUCSA	Basic blue protein		Cucumis sativus (Cucumber)	96
A0A0A0K7P6	A0A0A0K7P6_CUCSA	Laccase	Csa_6G009490	Cucumis sativus (Cucumber)	555
A0A0A0LF08	A0A0A0LF08_CUCSA	Laccase	Csa_3G732490	Cucumis sativus (Cucumber)	559
A0A0A0L192	A0A0A0L192_CUCSA	Laccase	Csa_4G308480	Cucumis sativus	556

- Verify protein name

# Components of an Annotation: Protein Using UniProt



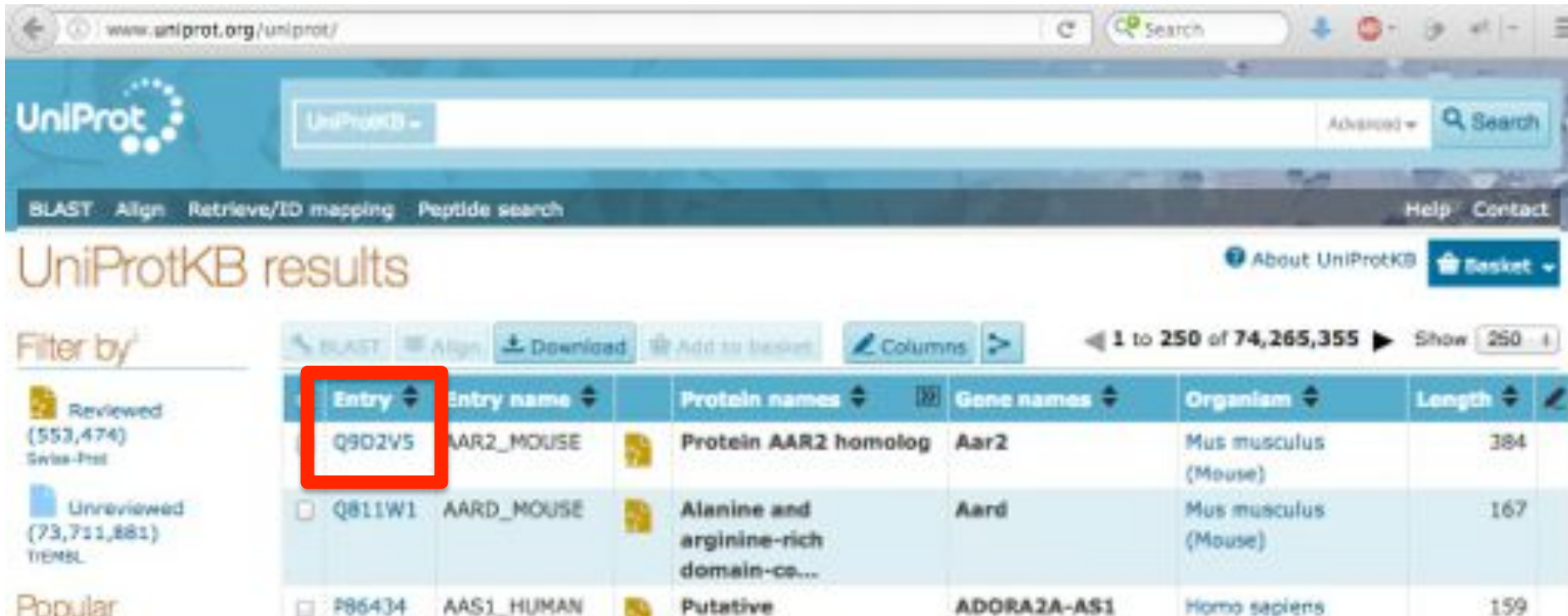
The screenshot shows the UniProt website interface. The browser address bar displays [www.uniprot.org/uniprot/](http://www.uniprot.org/uniprot/). The UniProt logo is on the left, and a search bar with a magnifying glass icon is on the right. Below the search bar, there are links for BLAST, Align, Retrieve/ID mapping, and Peptide search. The main heading is "UniProtKB results". On the left, there are filters for "Reviewed (553,474) Swiss-Prot" and "Unreviewed (73,711,881) TrEMBL". On the right, there are links for "About UniProtKB" and a "Basket" icon. The search results are displayed in a table with columns: Entry, Entry name, Protein names, Gene names, Organism, and Length. The first row is highlighted with a red box, showing the entry Q9D2V5, AAR2\_MOUSE, Protein AAR2 homolog, Aar2, Mus musculus (Mouse), and a length of 384. The second row shows Q811W1, AARD\_MOUSE, Alanine and arginine-rich domain-co..., Aard, Mus musculus (Mouse), and a length of 167. The third row shows P86434, AAS1\_HUMAN, Putative, ADORA2A-AS1, Homo sapiens, and a length of 159.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9D2V5	AAR2_MOUSE	Protein AAR2 homolog	Aar2	Mus musculus (Mouse)	384
Q811W1	AARD_MOUSE	Alanine and arginine-rich domain-co...	Aard	Mus musculus (Mouse)	167
P86434	AAS1_HUMAN	Putative	ADORA2A-AS1	Homo sapiens	159

- Copy the accession...



# Components of an Annotation: Protein Using UniProt



The screenshot shows the UniProt website interface. The top navigation bar includes the UniProt logo, a search bar, and links for BLAST, Align, Retrieve/ID mapping, Peptide search, Help, and Contact. The main heading is "UniProtKB results". Below this, there are filters for "Reviewed (553,474) Swiss-Prot" and "Unreviewed (73,711,881) TrEMBL". A table of search results is displayed, with columns: Entry, Entry name, Protein names, Gene names, Organism, and Length. The first row is highlighted, and the "Entry" column is circled in red.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q9D2V5	AAR2_MOUSE	Protein AAR2 homolog	Aar2	Mus musculus (Mouse)	384
Q811W1	AARD_MOUSE	Alanine and arginine-rich domain-co...	Aard	Mus musculus (Mouse)	167
P86434	AAS1_HUMAN	Putative	ADORA2A-AS1	Homo sapiens	159

- Copy the accession... *or get more information*

# Components of an Annotation:

## Protein

- Protein record shows more info

UniProtKB - P29602 (CPC\_CUCSA)

Display

Entry Publications Feature viewer Feature table

Function Names & Taxonomy Subcell. location Pathol./Biotech. PTM / Processing Expression Interaction Structure Family & Domains Sequence Cross-references

BLAST Align Format Add to basket History Help video

**Protein** | **Cucumber peeling cupredoxin**

**Gene** | N/A

**Organism** | *Cucumis sativus* (Cucumber)

None

**Function**

Sites

Feature key	Position(s)	Description	Actions	Graph
Metal binding <sup>1</sup>	46	Copper		
Metal binding <sup>1</sup>	89	Copper		
Metal binding <sup>1</sup>	94	Copper		
Metal binding <sup>1</sup>	99	Copper		

**GO - Molecular function**<sup>1</sup>

- electron carrier activity ⓘ Source: InterPro
- metal ion binding ⓘ Source: UniProtKB-KW

**GO - Biological process**<sup>1</sup>

- oxidation-reduction process ⓘ Source: UniProtKB-KW

Complete GO annotation...

# Components of an Annotation: Protein

- Protein record shows more info
- Summary of existing GO annotations

UniProtKB - P29602 (CPC\_CUCSA)

Display

Entry Publications Feature viewer Feature table

Function Names & Taxonomy Subcell. location Pathol./Biotech. PTM / Processing Expression Interaction Structure Family & Domains Sequence Cross-references

Protein | **Cucumber peeling cupredoxin**

Gene | N/A

Organism | *Cucumis sativus* (Cucumber)

Status | Reviewed - Annotation score: ●●●○○○ - Experimental evidence

Function<sup>1</sup>

Sites

Feature key	Position(s)	Description	Actions	Graph
Metal binding <sup>2</sup>	46	Copper		
Metal binding <sup>2</sup>	89	Copper		
Metal binding <sup>2</sup>	94	Copper		
Metal binding <sup>2</sup>	99	Copper		

GO - Molecular function<sup>1</sup>

- electron carrier activity ⓘ Source: InterPro
- metal ion binding ⓘ Source: UniProtKB-KW

GO - Biological process<sup>1</sup>

- oxidation-reduction process ⓘ Source: UniProtKB-KW

Complete GO annotation...

# Components of an Annotation: Protein

- Protein record shows more info
- Summary of existing GO annotations
- Link to QuickGO for complete set of existing annotations

UniProtKB - P29602 (CPC\_CUCSA)

Display

Entry  
Publications  
Feature viewer  
Feature table

None

☒ Function  
☒ Names & Taxonomy  
☐ Subcell. location  
☐ Pathol./Biotech.  
☒ PTM / Processing  
☐ Expression  
☐ Interaction  
☒ Structure  
☒ Family & Domains  
☒ Sequence  
☒ Cross-references

BLAST Align Format Add to basket History Help video

Protein | **Cucumber peeling cupredoxin**  
Gene | N/A  
Organism | *Cucumis sativus* (Cucumber)  
Status | Reviewed - Annotation score: ●●●○○○ - Experimental evidence

### Function<sup>1</sup>

#### Sites

Feature key	Position(s)	Description	Actions	Graph
Metal binding <sup>2</sup>	46	Copper		
Metal binding <sup>2</sup>	89	Copper		
Metal binding <sup>2</sup>	94	Copper		
Metal binding <sup>2</sup>	99	Copper		

#### GO - Molecular function<sup>1</sup>

- electron carrier activity ⓘ Source: InterPro
- metal ion binding ⓘ Source: UniProtKB-KW

#### GO - Biological process<sup>1</sup>

- oxidation-reduction process ⓘ Source: UniProtKB-KW

Complete GO annotation...



# Components of an Annotation: Protein

- Protein record shows more info
- Sequence & length
- Cross References
- Etc.

The screenshot displays a protein annotation interface. On the left, a 'FEATURE VIEWER' sidebar lists various categories with checkboxes: Function, Names & Taxonomy, Subcell. location, Pathol./Biotech., PTM / Processing, Expression, Interaction, Structure, Family & Domains, Sequence, Cross-references, Entry information, and Miscellaneous. The 'Sequence' category is selected. The main content area shows the protein 'PS51485. PHYTOCYANIN. 1 hit, [Graphical view]'. Below this, the 'Sequence' section indicates 'Sequence status: Complete.' and provides a download link for 'P29602-1 [UniParc]' in FASTA format, along with an 'Add to basket' button. A 'Hide' link is also present. The sequence is displayed in a grid format with residue numbers 10, 20, 30, 40, 60, 70, 80, 90, 110, 120, and 130. The sequence data is as follows:

10	20	30	40
QSTVHIWGDN	TGNSVPSSPN	FYSQWAAGKT	FRVGDLSLQFN
60	70	80	90
METKQSPDAC	NFVMSDNQVE	RTSPVIERLD	ELGMHYFVCT
110	120	130	
LSINVVAANA	TVSMPPPSBS	PPSSVMPPPV	MPPPSPS


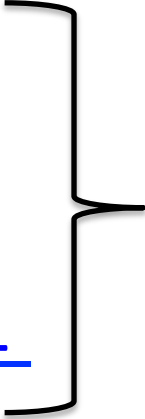
Below the sequence, the 'Sequence databases' section lists 'PIR<sup>1</sup> S27034, SSKV.'. The 'Cross-references' section also lists 'PIR<sup>1</sup> S27034, SSKV.'.

# Components of an Annotation: Protein

- Look in the paper for any accession/ID
  - UniProt accession
    - [E2RF02](#)
  - NCBI Gene ID, RefSeq, etc.
    - [478211](#)
    - [XP\\_853095.1](#)
    - [NP\\_001153791.1](#)



# Components of an Annotation: Protein

- [E2RF02](#) 
  - You MUST use the UniProt accession when adding the annotation to CACAO
- [478211](#)
- [XP\\_853095.1](#)
- [NP\\_001153791.1](#) 
  - Papers may use identifiers other than UniProt accessions
  - Use these to find the UniProt accession

# Components of an Annotation: Protein

## UniProt Warnings

- Sometimes UniProt has multiple entries for the same protein
  - Gold star = SwissProt = reviewed
  - Blank star = TrEMBL = computational entry
- **Ask for help**

UniProtKB consists of two sections:

 **Reviewed (Swiss-Prot) - Manually annotated**  
Records with information extracted from literature and curator-evaluated computational analysis.

 **Unreviewed (TrEMBL) - Computationally analyzed**  
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

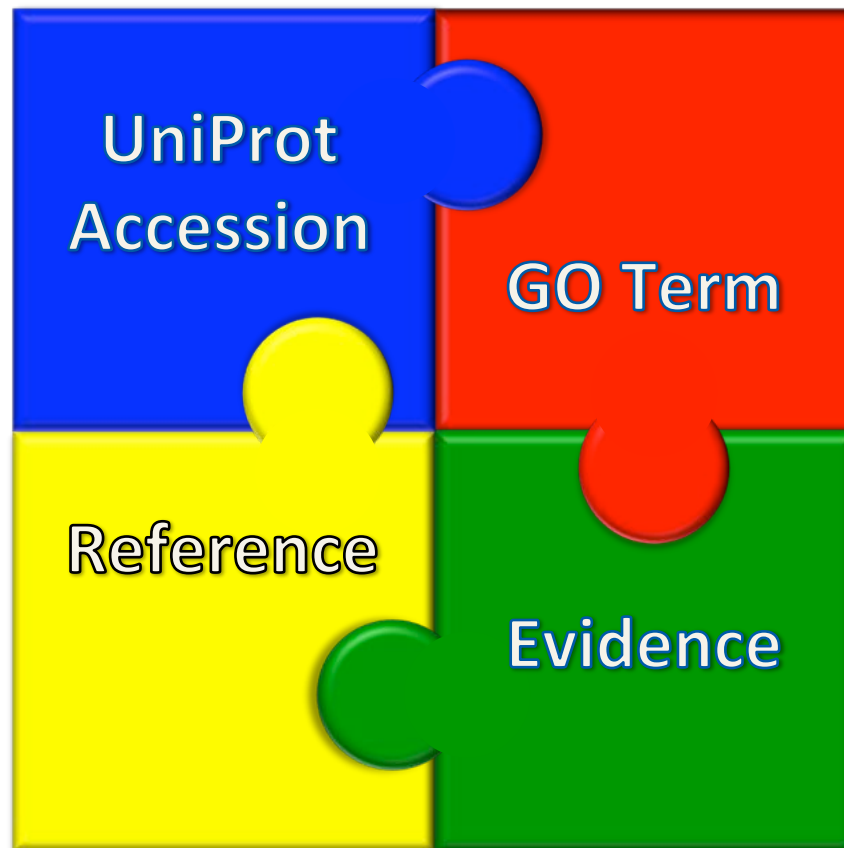
# Components of an Annotation:

## Protein

### UniProt Warnings

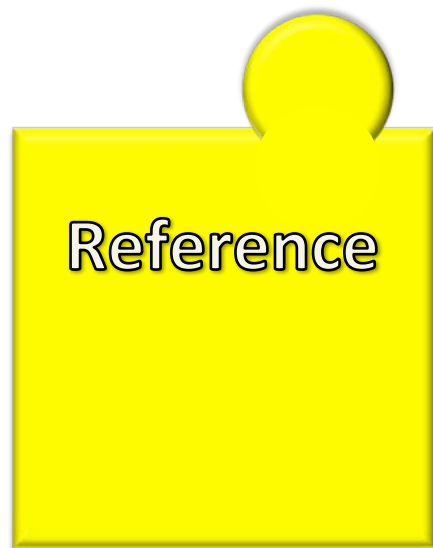
- Right species/strain
- Not a fragment
- Sometimes the protein you want is not in UniProt
  - May want to find another paper/protein
- **Ask for help**
  - OK to email the UniProt help desk
  - Check your reasoning with us!

# Components of an Annotation: What We've Covered



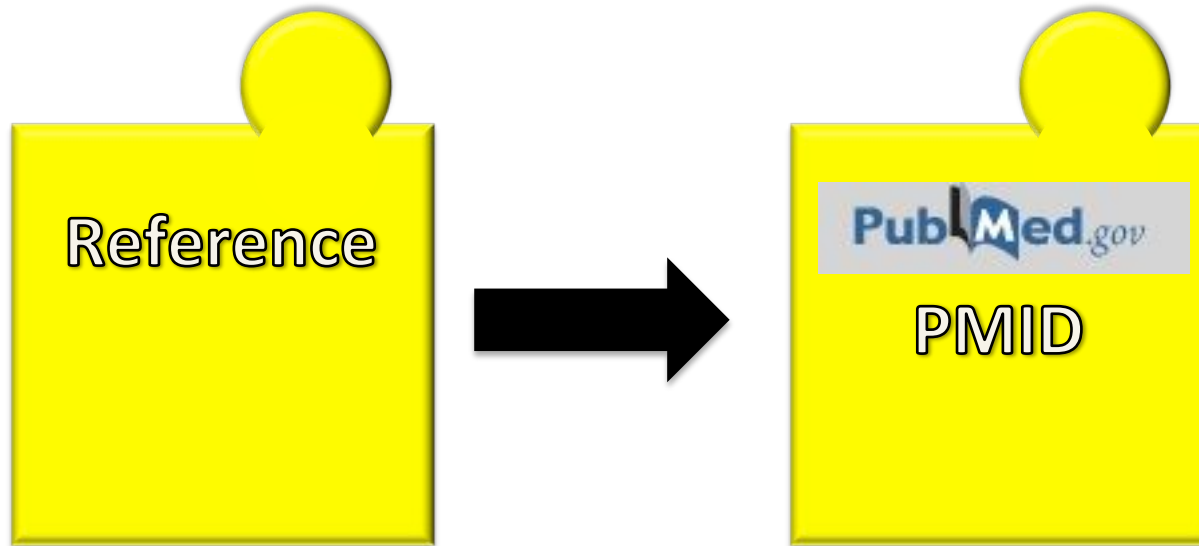
# Components of an Annotation:

## Reference



# Components of an Annotation:

## Reference



- Find the PMID using [pubmed.gov](https://pubmed.ncbi.nlm.nih.gov/)  
(<https://www.ncbi.nlm.nih.gov/pubmed/>)
- Examples: **PMID:22110029**, **PMID:20473289**



# Components of an Annotation:

## Reference

- Need a scientific paper with experimental data
- No review articles, no books, no textbooks, no Wikipedia articles, no class notes...
  - BUT it is GOOD to start with those!
- DON'T start with the first paper you see from a random PubMed search

# Components of an Annotation:

## Reference

- Need a scientific paper with experimental data
  - PubMed review?
  - Refer to the paper through the PMID number
    - *Not* the full citation

NCBI Resources How To

PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed Hu AND McIntosh

RSS Save search Limits Advanced

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to:

Results: 10

☐ [GONUTS: the Gene Ontology Normal Usage Tracking System.](#)

1. Renfro DP, McIntosh BK, Venkatraman A, Siegele DA, Hu JC.  
Nucleic Acids Res. 2012 Jan;40(1):Q1282-9. Epub 2011 Nov 22.  
PMID: 22110029

[Related citations](#)

22110029

# Components of an Annotation:

## Reference

U.S. National Library of Medicine  
National Institutes of Health

Advanced

Help

Display Settings: ☺ Abstract

Send to: ☺

Full text links

Molecular Systems Biology **FREE** **Full** **PM** **PMC** Full text archive

Mol Syst Biol. 2012 May 8;8:561. doi: 10.1038/msb.2012.13.

### Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours.

Yamada T<sup>1</sup>, Wader AS, Raes J, Zeleniak A, Perschil N, Perret A, Selaouat M, Paul KE, Weisenbach J, Bork P.

ⓘ Author information

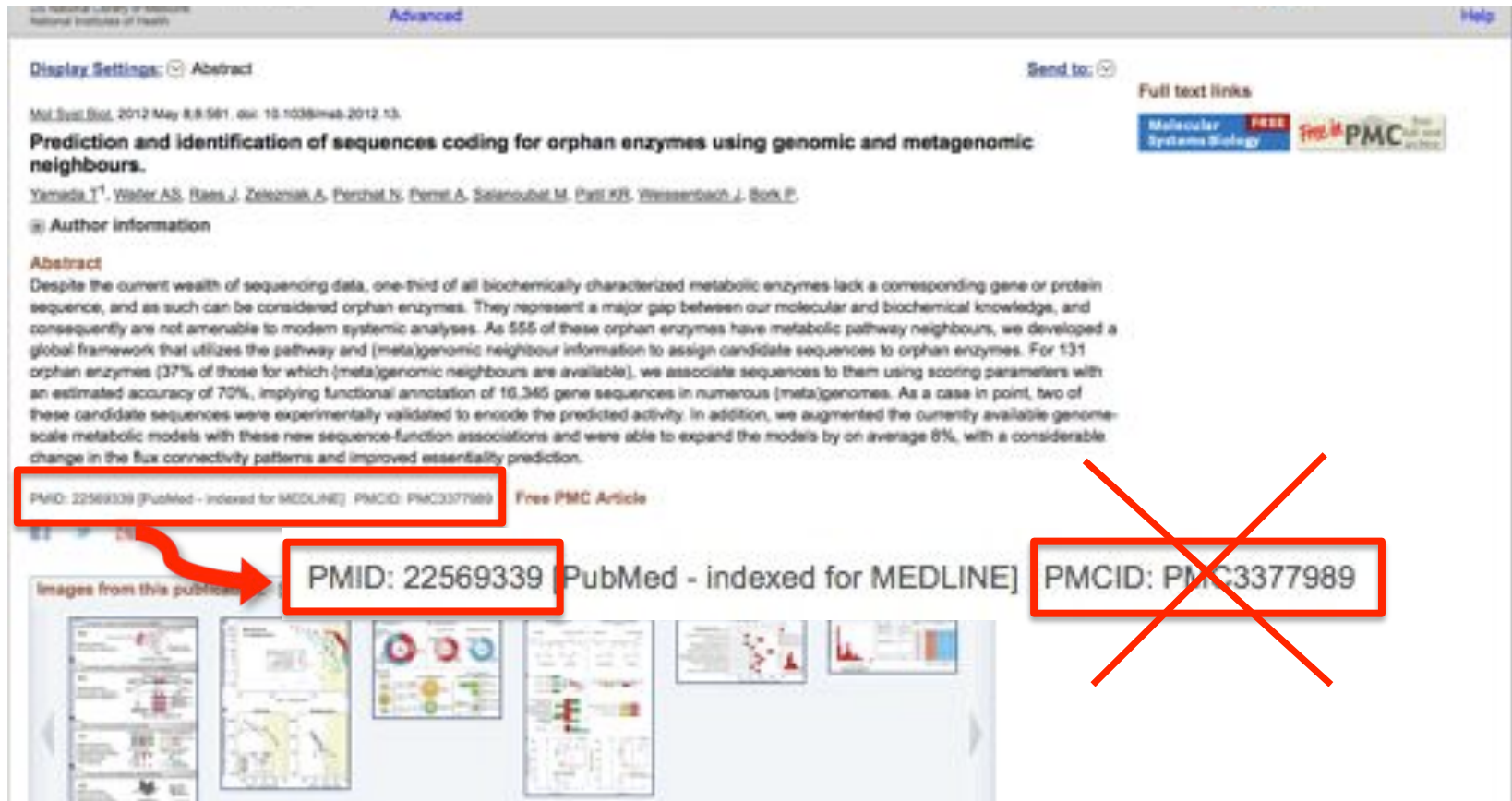
#### Abstract

Despite the current wealth of sequencing data, one-third of all biochemically characterized metabolic enzymes lack a corresponding gene or protein sequence, and as such can be considered orphan enzymes. They represent a major gap between our molecular and biochemical knowledge, and consequently are not amenable to modern systemic analyses. As 555 of these orphan enzymes have metabolic pathway neighbours, we developed a global framework that utilizes the pathway and (meta)genomic neighbour information to assign candidate sequences to orphan enzymes. For 131 orphan enzymes (37% of those for which (meta)genomic neighbours are available), we associate sequences to them using scoring parameters with an estimated accuracy of 70%, implying functional annotation of 16,345 gene sequences in numerous (meta)genomes. As a case in point, two of these candidate sequences were experimentally validated to encode the predicted activity. In addition, we augmented the currently available genome-scale metabolic models with these new sequence-function associations and were able to expand the models by on average 8%, with a considerable change in the flux connectivity patterns and improved essentiality prediction.

PMID: 22569339 [PubMed - indexed for MEDLINE] PMID: PMC3377989 Free PMC Article

Images from this publication

PMID: 22569339 [PubMed - indexed for MEDLINE] PMID: PMC3377989



# Components of an Annotation:

## Reference

U.S. National Library of Medicine  
National Institutes of Health

Advanced

Help

Display Settings: ☺ Abstract

Send to: ☺

Full text links

Molecular Systems Biology **FREE** **PMC** Full text and archive

Met. Syst. Biol., 2012 May 8;8:561. doi: 10.1038/msb.2012.13.

**Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours.**

Yamada T<sup>1</sup>, Walter AS, Bass J, Zeleniak A, Perthal N, Perret A, Salanoubat M, Paul KS, Weissertach J, Bork P.

ⓘ Author information

**Abstract**

Despite the current wealth of sequencing data, one-third of all biochemically characterized metabolic enzymes lack a corresponding gene or protein sequence, and as such can be considered orphan enzymes. They represent a major gap between our molecular and biochemical knowledge, and consequently are not amenable to modern systemic analyses. As 555 of these orphan enzymes have metabolic pathway neighbours, we developed a global framework that utilizes the pathway and (meta)genomic neighbour information to assign candidate sequences to orphan enzymes. For 131 orphan enzymes (37% of those for which (meta)genomic neighbours are available), we associate sequences to them using scoring parameters with an estimated accuracy of 70%, implying functional annotation of 16,345 gene sequences in numerous (meta)genomes. As a case in point, two of these candidate sequences were experimentally validated to encode the predicted activity. In addition, we augmented the currently available genome-scale metabolic models with these new sequence-function associations and were able to expand the models by on average 8%, with a considerable change in the flux connectivity patterns and improved essentiality prediction.

PMID: 22569339 [PubMed - indexed for MEDLINE] **PMCID: PMC3377989** **Free PMC Article**

ⓘ

Images from this publication. See all images (8) **Free text**



- The abstract is not enough for an annotation
  - But, may be enough to reject a paper!!!

# Components of an Annotation: Reference

The screenshot shows a PubMed article page. At the top, it says 'National Library of Medicine National Institutes of Health' and 'Advanced'. The article title is 'Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours.' by Yamada T<sup>1</sup>, Walter AS, Bass J, Zeleniak A, Perhat N, Perret A, Salanoubat M, Paul KS, Weissbach J, Bork P. The abstract text is visible, starting with 'Despite the current wealth of sequencing data, one-third of all biochemically characterized metabolic enzymes lack a corresponding gene or protein sequence...'. Below the abstract, there are links for 'Full text links', 'Molecular Systems Biology', and 'PMC'. At the bottom, there is a section titled 'Images from this publication' showing several thumbnail images of figures and tables.

Display Settings: ☺ Abstract Send to: ☺

Full text links

Molecular Systems Biology **FREE** **PMC** Full text article

Met Syst Biol, 2012 May 8;8:561. doi: 10.1038/msb.2012.13.

**Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours.**

Yamada T<sup>1</sup>, Walter AS, Bass J, Zeleniak A, Perhat N, Perret A, Salanoubat M, Paul KS, Weissbach J, Bork P.

Ⓜ Author information

**Abstract**

Despite the current wealth of sequencing data, one-third of all biochemically characterized metabolic enzymes lack a corresponding gene or protein sequence, and as such can be considered orphan enzymes. They represent a major gap between our molecular and biochemical knowledge, and consequently are not amenable to modern systemic analyses. As 555 of these orphan enzymes have metabolic pathway neighbours, we developed a global framework that utilizes the pathway and (meta)genomic neighbour information to assign candidate sequences to orphan enzymes. For 131 orphan enzymes (37% of those for which (meta)genomic neighbours are available), we associate sequences to them using scoring parameters with an estimated accuracy of 70%, implying functional annotation of 16,345 gene sequences in numerous (meta)genomes. As a case in point, two of these candidate sequences were experimentally validated to encode the predicted activity. In addition, we augmented the currently available genome-scale metabolic models with these new sequence-function associations and were able to expand the models by on average 8%, with a considerable change in the flux connectivity patterns and improved essentiality prediction.

PMID: 22569339 [PubMed - indexed for MEDLINE] PMID: PMC3377989 **Free PMC Article**

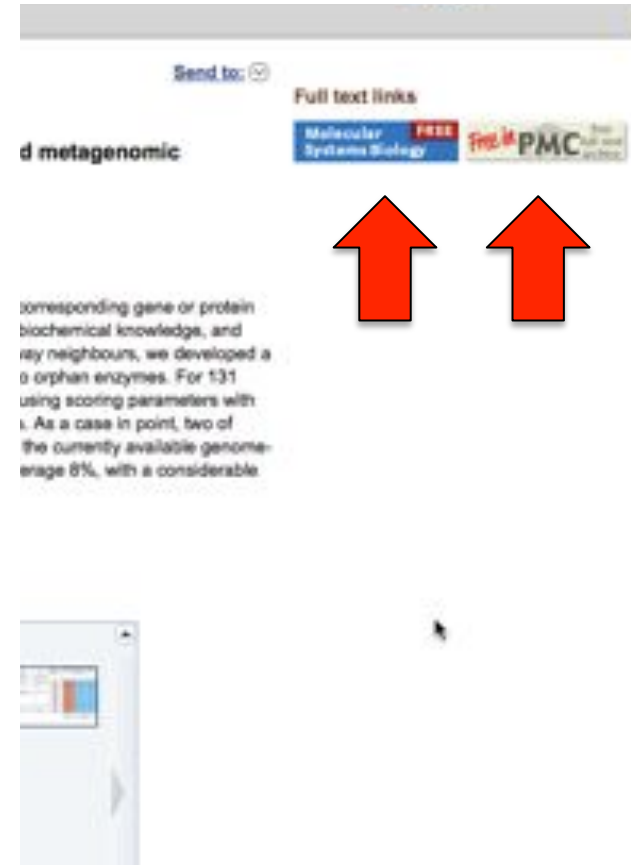
Images from this publication. See all images (8) **Free text**

- PubMed has links to full paper

# Components of an Annotation:

## Reference

- PubMed has links to full papers
  - Links vary between papers
- Some papers are open access
  - Pubmed Central
  - Journal sites
  - TAMU students:  
<http://library.tamu.edu/>
- Others are pay only
  - Don't pay real \$\$!
    - Your library may have subscriptions
    - Pick a different paper
    - Email the author and ask for a pdf
      - Send us a copy





# Components of an Annotation:

## Reference

## Warnings

- Good science  $\neq$  good for annotation

## Second Extracellular Loop of Human Glucagon-like Peptide-1 Receptor (GLP-1R) Differentially Regulates Orthosteric but Not Allosteric Agonist Binding and Function<sup>\*[5]</sup>

Received for publication, September 30, 2011, and in revised form, November 29, 2011. Published, JBC Papers in Press, December 6, 2011, DOI 10.1074/jbc.M111.309369

Cassandra Koole<sup>+</sup>, Denise Wootten<sup>+</sup>, John Simms<sup>+</sup>, Emilia E. Savage<sup>+</sup>, Laurence J. Miller<sup>5</sup>, Arthur Christopoulos<sup>+1</sup>, and Patrick M. Sexton<sup>+2</sup>

From the <sup>+</sup>Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences and Department of Pharmacology, Monash University, Parkville, Victoria 3052, Australia and the <sup>5</sup>Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Scottsdale, Arizona 85259

**Background:** The ECL2 of the GLP-1R is critical for GLP-1 peptide-mediated selective signaling.

**Results:** Mutation of most ECL2 residues to alanine results in changes in binding and/or efficacy of oxyntomodulin and exendin-4 but not allosteric agonists.

**Conclusion:** ECL2 of the GLP-1R has ligand-specific as well as general effects on peptide agonist-mediated receptor activation.

**Significance:** This work provides insight into control of family B GPCR activation transition.

# Components of an Annotation:

## Reference

## Warnings

- Good science  $\neq$  good for annotation

Cell Stem Cell  
**Short Article**

**Cell**  
PRESS

## Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells

Miguel Angel Esteban,<sup>1,6</sup> Tao Wang,<sup>1,6</sup> Baoming Qin,<sup>1,6</sup> Jiayin Yang,<sup>1</sup> Dajiang Qin,<sup>1</sup> Jinglei Cai,<sup>1</sup> Wen Li,<sup>1</sup> Zhihui Weng,<sup>1</sup> Jiekai Chen,<sup>1</sup> Su Ni,<sup>1</sup> Keshi Chen,<sup>1</sup> Yuan Li,<sup>1</sup> Xiaopeng Liu,<sup>1</sup> Jianyong Xu,<sup>1</sup> Shiqiang Zhang,<sup>1</sup> Feng Li,<sup>1</sup> Wenzhi He,<sup>1</sup> Krystyna Labuda,<sup>2</sup> Yancheng Song,<sup>3</sup> Anja Peterbauer,<sup>4</sup> Susanne Wolbank,<sup>2</sup> Heinz Redl,<sup>2</sup> Mei Zhong,<sup>5</sup> Daozhang Cai,<sup>3</sup> Lingwen Zeng,<sup>1</sup> and Duanqing Pei<sup>1,\*</sup>

<sup>1</sup>Stem Cell and Cancer Biology Group, Key Laboratory of Regenerative Biology, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510663, China

<sup>2</sup>Erwin Rottmann Institute for Clinical and Experimental Transplantation, Austrian Cluster for Tissue Regeneration, Vienna 1200, Austria



# Components of an Annotation:

## Reference

## Warnings

- Good science  $\neq$  good for annotation

## Robust design and optimization of retroaldol enzymes

**Eric A. Althoff,<sup>1,2</sup> Ling Wang,<sup>1</sup> Lin Jiang,<sup>1,3</sup> Lars Giger,<sup>4</sup> Jonathan K. Lassila,<sup>5</sup> Zhizhi Wang,<sup>1</sup> Matthew Smith,<sup>1</sup> Sanjay Hari,<sup>1</sup> Peter Kast,<sup>4</sup> Daniel Herschlag,<sup>5</sup> Donald Hilvert,<sup>4</sup> and David Baker<sup>1\*</sup>**

<sup>1</sup>Department of Biochemistry, University of Washington and HHMI, Seattle, Washington 98195

<sup>2</sup>Arzeda Corp., Seattle, Washington 98102

<sup>3</sup>Department of Biological Chemistry, UCLA, Los Angeles, California 90095

<sup>4</sup>Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland

<sup>5</sup>Department of Biochemistry, Stanford University, Stanford, California 94305

# Components of an Annotation:

## Reference

## Warnings

- Good science  $\neq$  good for annotation

10624 • The Journal of Neuroscience, August 11, 2010 • 30(32):10624–10638

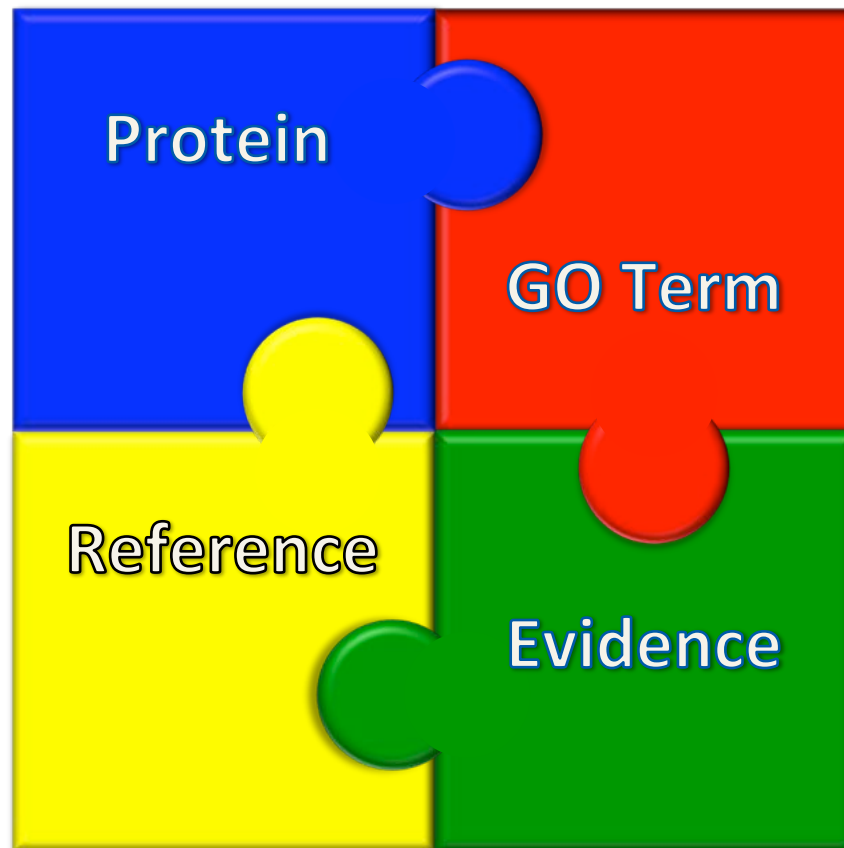
Neurobiology of Disease

## Excess Phosphoinositide 3-Kinase Subunit Synthesis and Activity as a Novel Therapeutic Target in Fragile X Syndrome

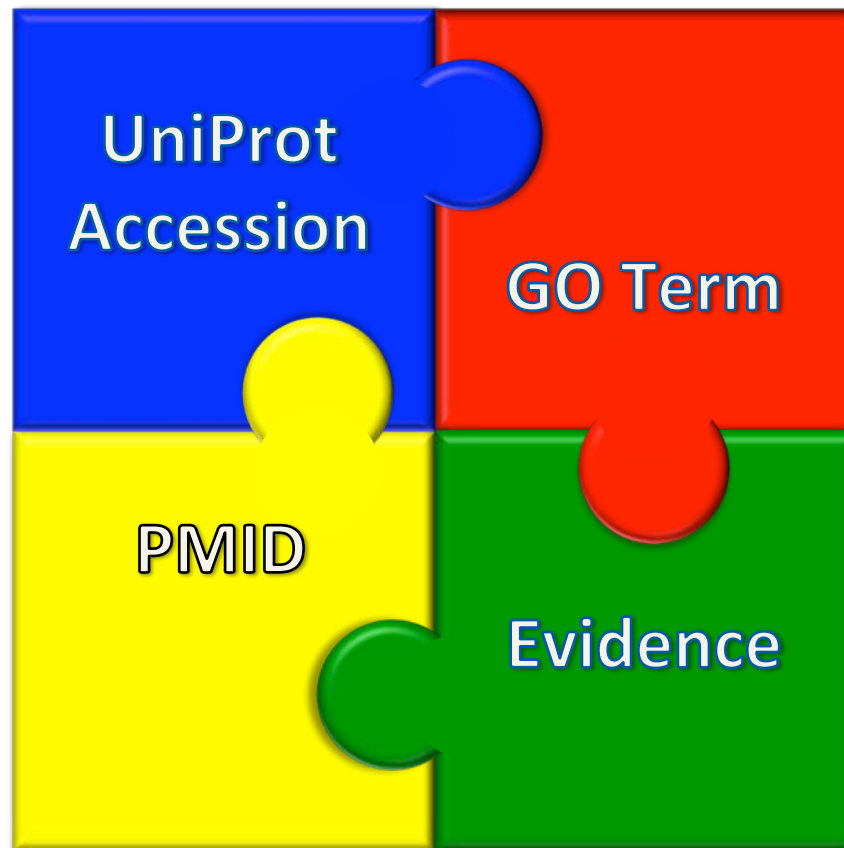
Christina Gross,<sup>1</sup> Mika Nakamoto,<sup>2\*</sup> Xiaodi Yao,<sup>1\*</sup> Chi-Bun Chan,<sup>3</sup> So Y. Yim,<sup>1</sup> Keqiang Ye,<sup>3</sup> Stephen T. Warren,<sup>2,4,5</sup> and Gary J. Bassell<sup>1,6</sup>

Departments of <sup>1</sup>Cell Biology, <sup>2</sup>Human Genetics, <sup>3</sup>Pathology and Laboratory Medicine, <sup>4</sup>Biochemistry, <sup>5</sup>Pediatrics, and <sup>6</sup>Neurology, Emory University School of Medicine, Atlanta, Georgia 30322

# Components of an Annotation: What We've Covered



# Components of an Annotation: What We've Covered



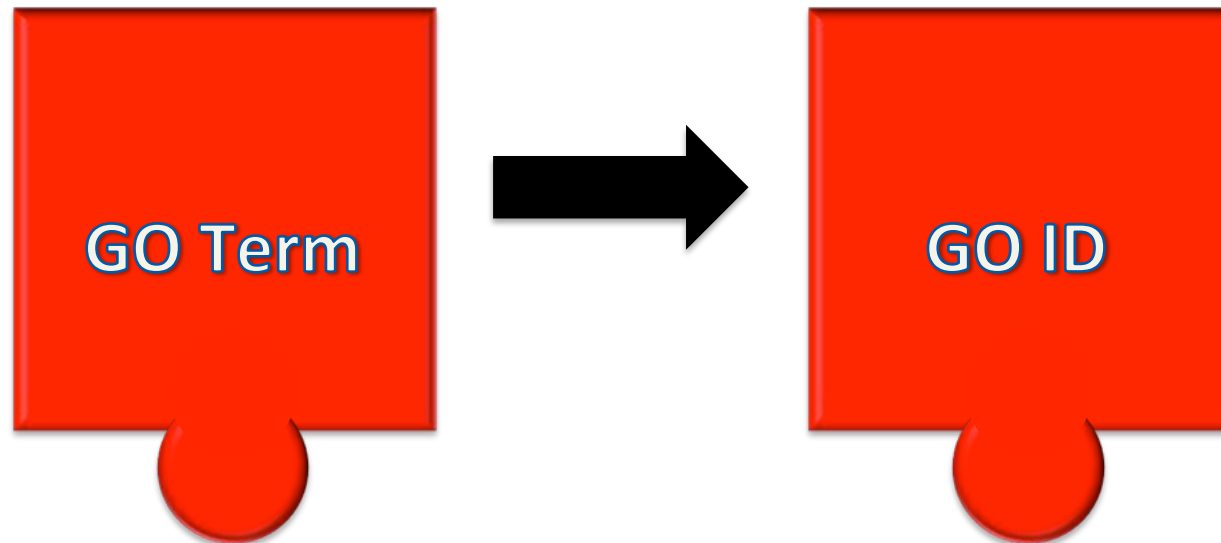
# Components of an Annotation:

## GO Term



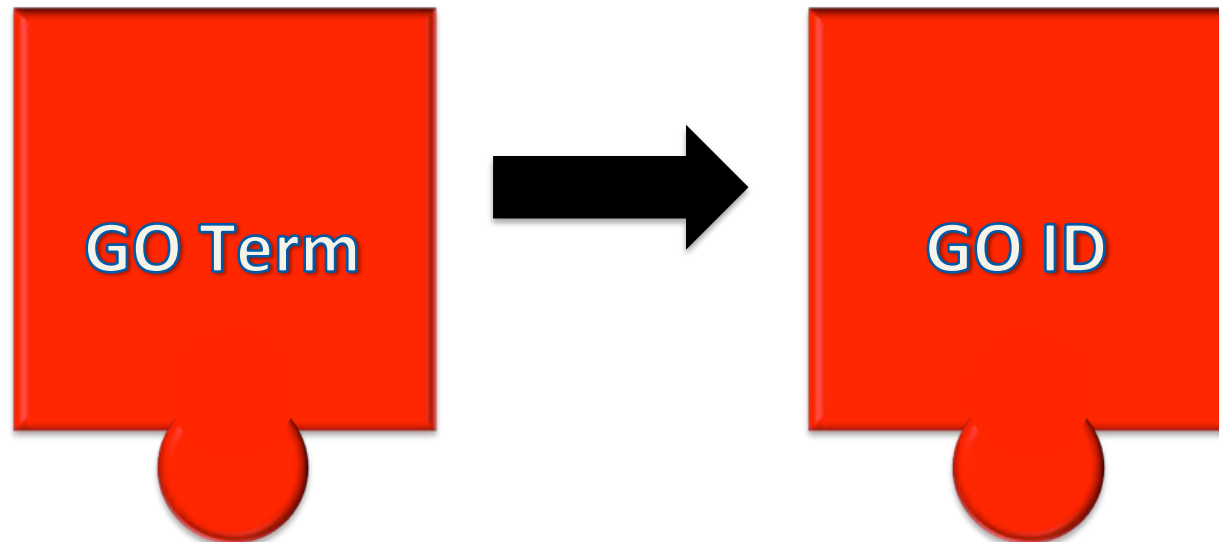
# Components of an Annotation:

## GO Term



# Components of an Annotation:

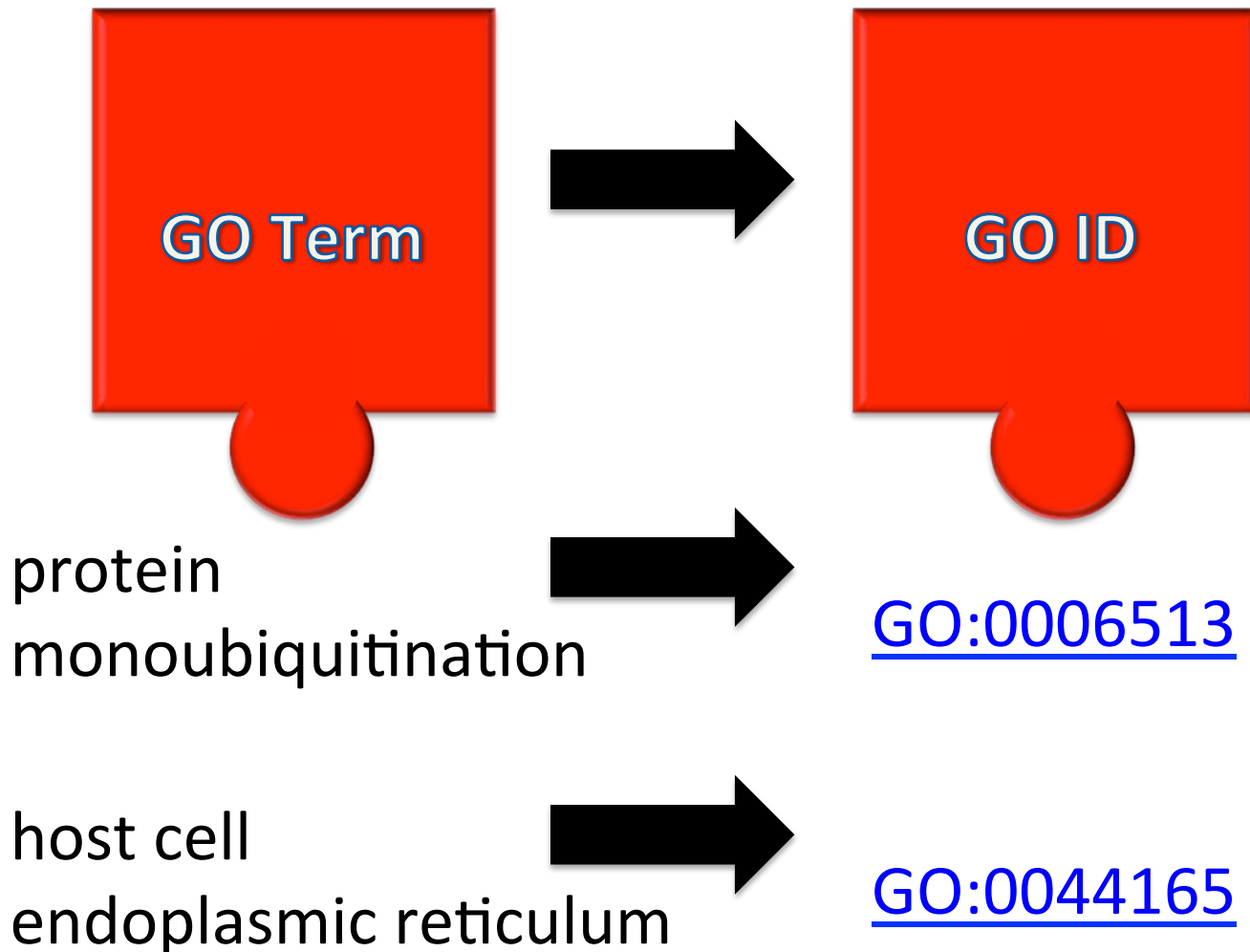
## GO Term



- Find the GO ID using a Gene Ontology browser
  - QuickGO: <http://www.ebi.ac.uk/QuickGO/>
  - AmiGO: <http://amigo.geneontology.org/amigo>
  - GONUTS: <http://gowiki.tamu.edu>

# Components of an Annotation:

## GO Term





# Components of an Annotation: GO Term

go term discussion edit history delete protect switch purge

GONUTS is undergoing some major debugging for Picos. Please expect blank pages and some delays in updating. [Email comments to Daniel.]

## GO:0004713 | protein tyrosine kinase activity

id: GO:0004713  
name: protein tyrosine kinase activity  
namespace: molecular\_function  
alt\_id: GO:0004718  
def: "Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate." [EC:2.7.10]  
subset: gosubset\_prot  
synonym: "JAK" NARROW []  
synonym: "Janus kinase activity" NARROW []  
synonym: "protein-tyrosine kinase activity" EXACT []  
xref: EC:2.7.10  
xref: MetaCyc:EC-2.7.10  
xref: Rhea:11265 "protein tyrosine kinase activity"  
is\_a: GO:0004672 | protein kinase activity

[AmiGO](#)

Last version checked	Last updated
date: 11-01-2003 12:55	date: 28-10-2003 12:54
saved by: rfoelger	saved by: dph
auto-generated by: OBO-Edt 2.0	auto-generated by: OBO-Edt 2.0

[Gono Ontology Home](#)  
The contents of this box are automatically generated. You can help by adding information to the "Notes" @

### Usage Notes

[edit]

### References


[edit]

See [help:References](#) for how to manage references in GONUTS.

### Child Terms

This term has the following 4 child terms.

- [x] GO:0004714 - transmembrane receptor protein tyrosine kinase activity [13]



# Components of an Annotation: GO Term

## GO:0004713 | protein tyrosine kinase activity

id: GO:0004713

name: protein tyrosine kinase activity

namespace: molecular\_function

alt\_id: GO:0004718

def: "Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate." [EC:2.7.10]

subset: gosubset\_prot

synonym: "JAK" NARROW []

synonym: "Janus kinase activity" NARROW []

synonym: "protein-tyrosine kinase activity" EXACT []

ref: EC:2.7.10

ref: MetaCyc:EC-2.7.10

ref: Rhea:11065 "protein tyrosine kinase activity"

is\_a: GO:0004712 | protein kinase activity

[AmiGO 2](#)

### Last version checked

date: 14.01.2011 17:26

saved-by: rfoelger

auto-generated-by: OBO-Edit 2.0

### Last updated

date: 08.10.2010 13:21

saved-by: dph

auto-generated-by: OBO-Edit 2.0




[Gene Ontology Home](#)

The contents of this box are automatically generated. You can help by adding information to the "Notes" [box](#)

# Components of an Annotation:

## GO Term

**name:** protein tyrosine kinase activity  
**nomenpaces:** molecular\_function  
**id:** GO:0004718  
**def:** "Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate." [EC:2.7.10]  
**subset:** gosubset\_protok  
**synonym:** "JAK NARROW"  
**synonym:** "Janus kinase activity" NARROW  
**synonym:** "protein tyrosine kinase activity" EXACT  
**ref:** EC:2.7.10  
**xref:** MetaCyc:EC:2.7.10  
**xref:** Rfam:PF01265 "protein tyrosine kinase activity"  
**is\_a:** GO:0004672 ! protein kinase activity



**Last version checked:**  
date: 14.01.2011 17:26  
saved by: rlsdger  
auto-generated by: OBO-Edt 2.0

**Last updated:**  
date: 08.10.2010 10:21  
saved by: dph  
auto-generated by: OBO-Edt 2.0

Gene Ontology Home!

The contents of this box are automatically generated. You can help by adding information to the "Notes" at

### Usage Notes

### References

[See also Relationship to other ontology references in OWL file.](#)

### Child Terms

This term has the following 4 child terms.

- [+] GO:0004714 - transmembrane receptor protein tyrosine kinase activity (13)
- [] GO:0004715 - non-membrane spanning protein tyrosine kinase activity
- [+] GO:0004716 - receptor signaling protein tyrosine kinase activity (1)
- [+] GO:0035400 - histone tyrosine kinase activity (1)

Pages in category **GO:0004718 protein tyrosine kinase activity**

The following 200 pages are in this category, out of 732 total.  
Show articles starting with:

(previous 200) (next 200)

<b>C</b>	<b>C cont.</b>	<b>F cont.</b>
+ CHICK.ASM8T8	+ CHICK.G00960	+ FB.Tk4
+ CHICK.ASBVH2	+ CHICK.G00961	+ FB.Tk8
+ CHICK.BTK	+ CHICK.G00962	+ FB.Tkr

# Components of an Annotation:

## GO Term



- Pick the most specific term the paper supports
  - Too vague = wrong
  - Too specific = wrong
- Ask for help. You may need to create a new term
  - “New Term Requests” count as annotation credit for your grade

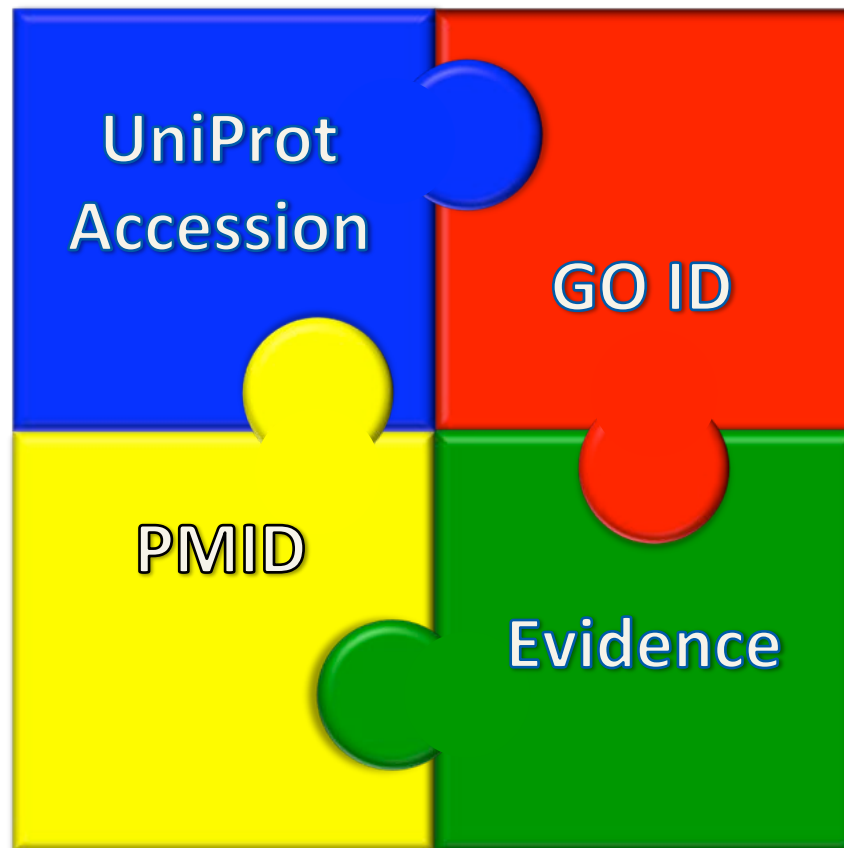
# Components of an Annotation:

## GO Term

### Strategies

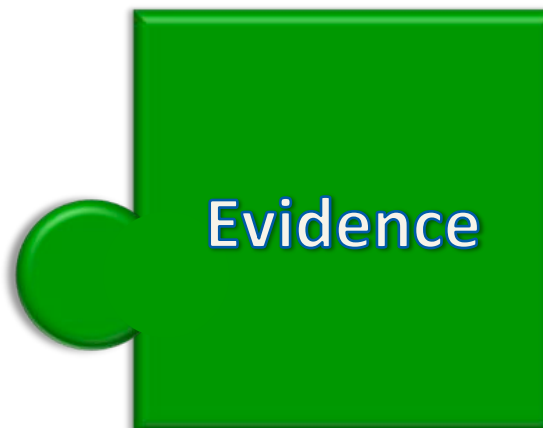
- Search for a keyword and browse the ontology for the right term
  - Look at the parents, children, and relatives
  - Use Google, Wikipedia etc. to find synonyms or alternative search terms
  - In GONUTS only search “Category” namespace if you get too many hits
- Look at terms suggested by others for your protein
  - Computational with the IEA evidence code
  - Curators with TAS or IC
- Look at terms used for homologous proteins in model organisms

# Components of an Annotation: What We've Covered



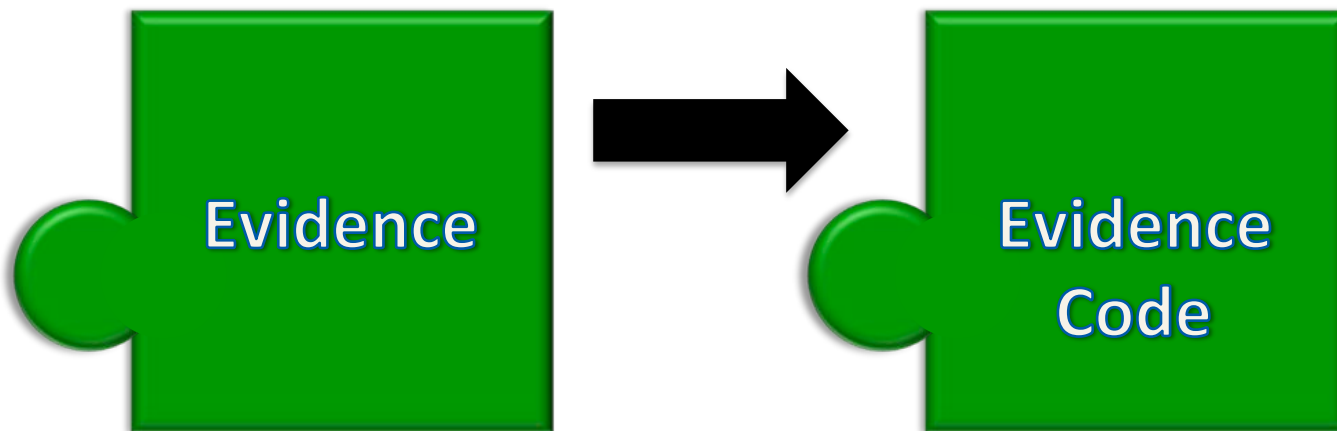
# Components of an Annotation:

## Evidence



# Components of an Annotation:

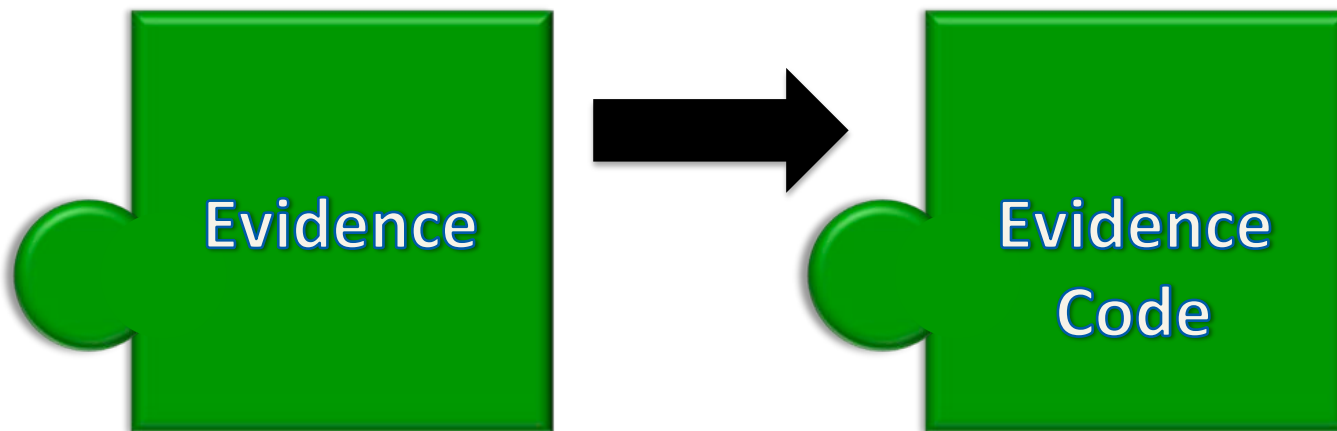
## Evidence





# Components of an Annotation:

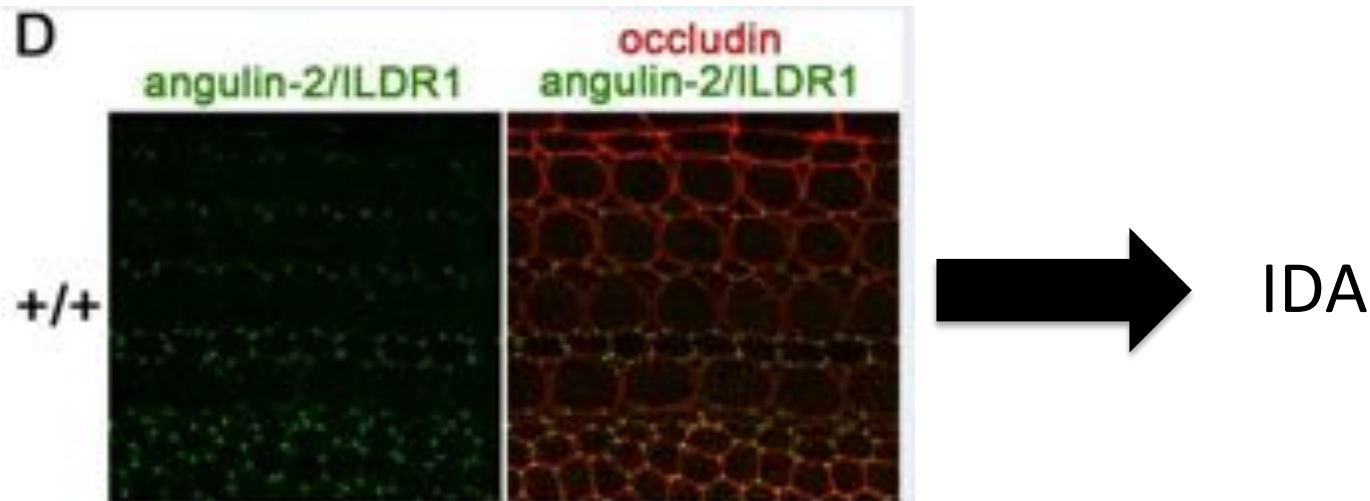
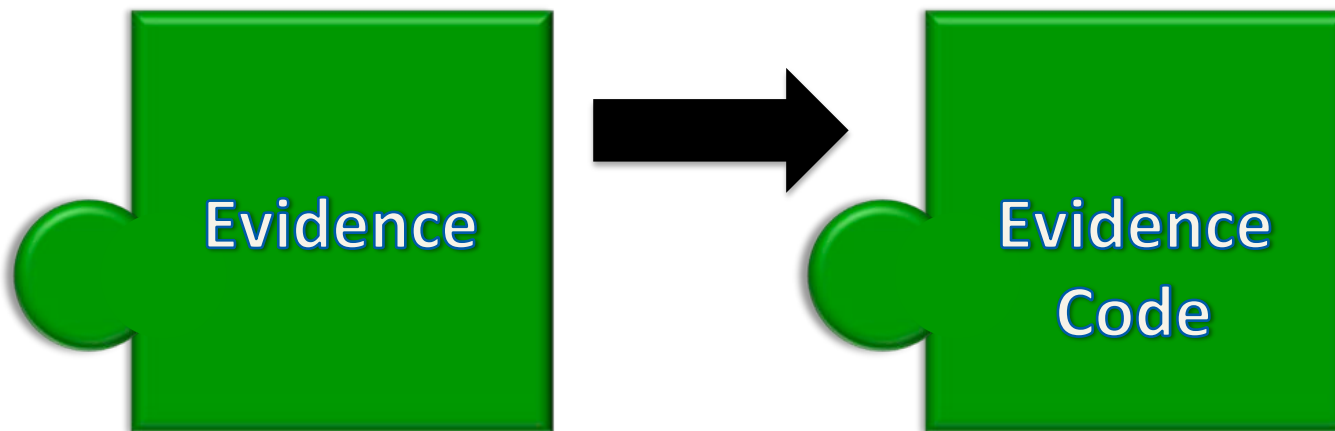
## Evidence



- Evidence codes describe the type of work or analysis done by the authors

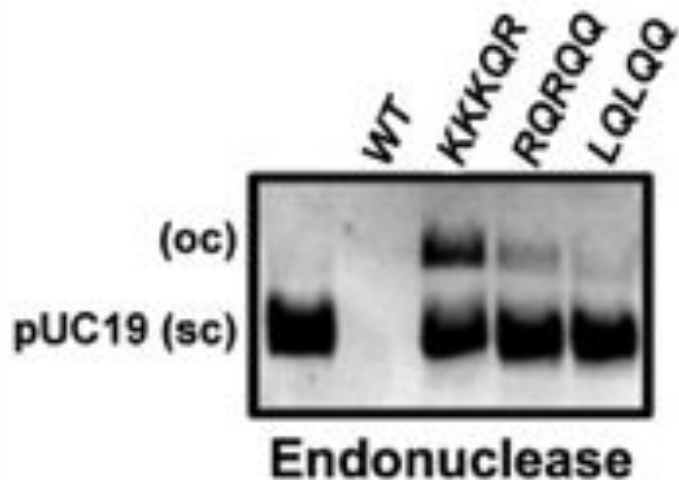
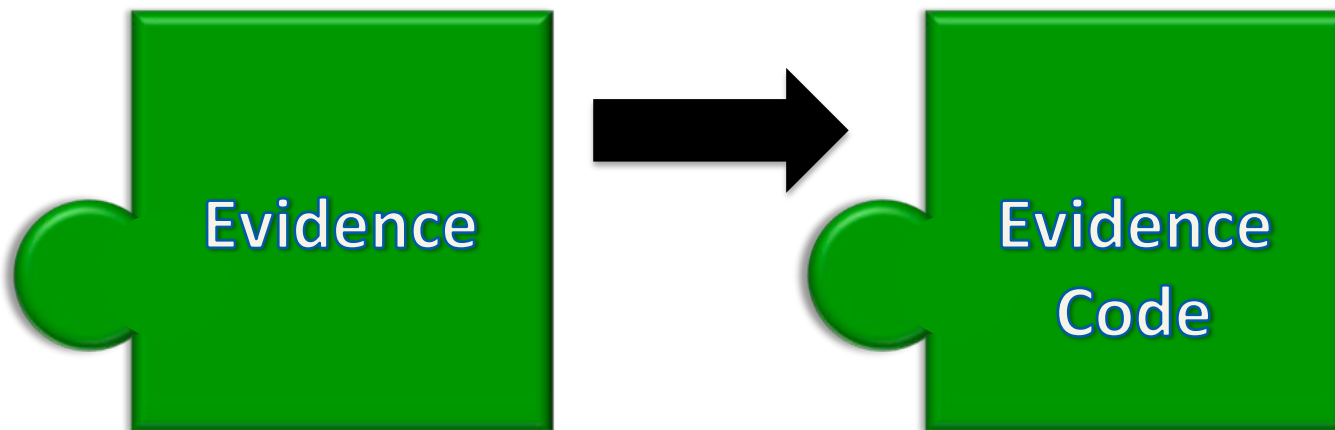
# Components of an Annotation:

## Evidence



# Components of an Annotation:

## Evidence



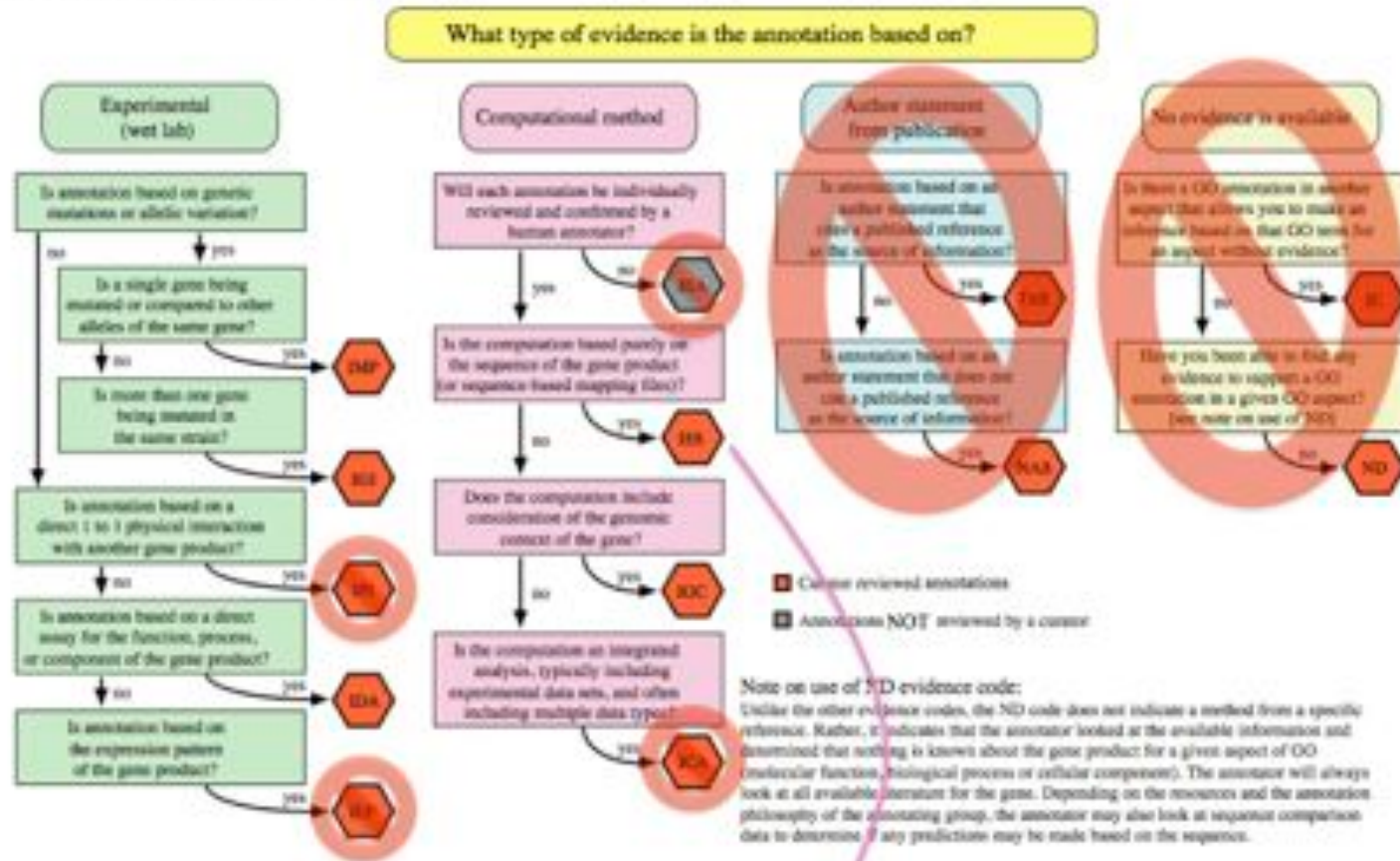
IMP

# Components of an Annotation:

## Evidence

- **IDA: Inferred from Direct Assay**
  - **IMP: Inferred from Mutant Phenotype**
    - NOT just for mutations! Includes inferred from inhibition in vivo by drugs, RNAi, etc.
  - **IGI: Inferred from Genetic Interaction**
  - **ISO: Inferred from Sequence Orthology**
  - **ISA: Inferred from Sequence Alignment**
  - **ISM: Inferred from Sequence Model**
  - **IGC: Inferred from Genomic Context**
- 
- Expert biocurators get to use others, but we restrict them for CACAO.
  - [http://gowiki.tamu.edu/wiki/index.php/evidence\\_codes](http://gowiki.tamu.edu/wiki/index.php/evidence_codes)

# Components of an Annotation: Evidence



## ALLOWED CODES FOR ALL CACAO STUDENTS:

1. **IDA:** Inferred from Direct Assay
2. **IMP:** Inferred from Mutant Phenotype
3. **IGI:** Inferred from Genetic Interaction - requires with/from field to be filled in
4. **ISO:** Inferred from Sequence Orthology - requires with/from field to be filled in
5. **ISA:** Inferred from Sequence Alignment - requires with/from field to be filled in
6. **ISM:** Inferred from Sequence Model - requires with/from field to be filled in
7. **IGC:** Inferred from Genomic Context

Use one of these three codes (ISO, ISA, ISM) if the Decision Tree points you to ISS

# Components of an Annotation:

## Evidence

### Some Evidence Types Require More Information

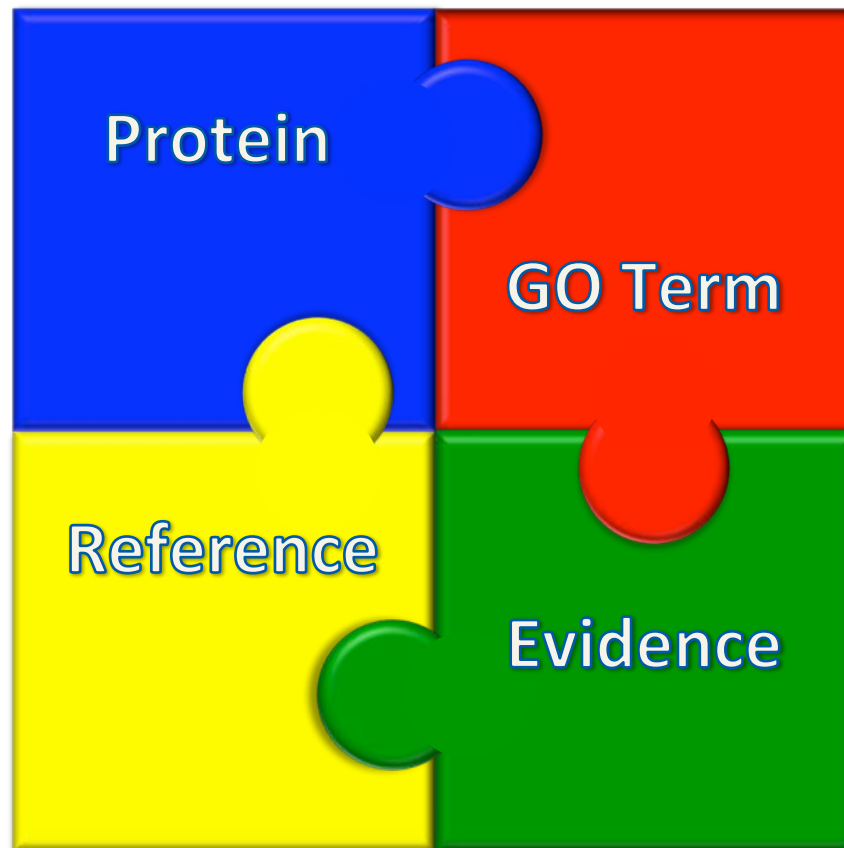
- *With/from*
- Evidence from sequence comparison
  - *With* the protein accession for the protein you are comparing to
    - “Match Protein” = comparison protein
    - “Match Protein” must have experimental annotation to the same GO term
- Evidence from computational analysis
  - *With* the reference for the analysis tool
- Evidence from genetic interaction
  - *With* the other gene(s) your protein is interacting with
- Ask for help if you suspect this applies to your annotation

# Components of an Annotation:

## Evidence

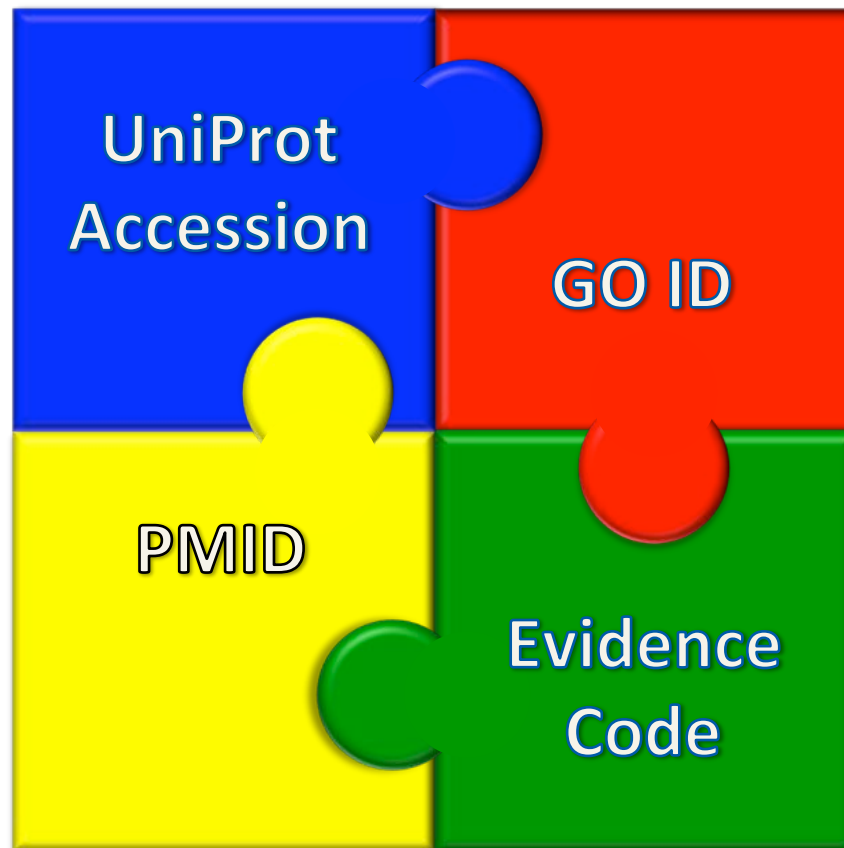
- Picking the right evidence code is important
- Print & Use the evidence code decision tree
  - [http://gowiki.tamu.edu/wiki/images/3/32/CACAO\\_decisiontree.pdf](http://gowiki.tamu.edu/wiki/images/3/32/CACAO_decisiontree.pdf)
- Use the evidence code guidelines at the GO consortium website:
  - <http://www.geneontology.org/GO.evidence.shtml>
- Discuss!

# Components of an Annotation: What You Need





# Components of an Annotation: What You Find



GO ANNOTATION ON GONUTS

# Create/Find a page for your paper

Special page

Search

Interested in participating in CACAO this fall? Sign up now!!!  
Instructors, want your students to compete this session?  
Students, have any questions? Please email us at [ecoliwiki@gmail.com](mailto:ecoliwiki@gmail.com) or  
TAMU students, come see us in Bio/Bio Room 443!

## Create a New Literature Page from PMID

PMID:

Enter a valid PMID ID to create a new literature page in the wiki, containing information fetched from PubMed

Privacy policy About GONUTS Disclaimers

Powered by MediaWiki

Main Page  
Enter GO at the Top  
Help  
Report Bug  
Annotation Jamborees  
Recent Changes  
Login / Create Account  
Cacao  
About CACAO  
Create New Gene Page  
Create New Literature Page  
Fall 2015  
Links  
GO Website  
GO Database Tutorial

# PMID Pages

Main Page  
Enter GO at the Top  
Help  
Report Bug  
Annotation Ambiguities  
Recent Changes  
Login / Create Account

◀ CACAO  
About CACAO  
Create New Gene Page  
Feb 2014  
◀ Links  
GO Website  
GO Ontology Tracker

◀ page contributors  
Jenny  
◀ Tools  
What links here  
Related changes  
Upload file  
Special pages  
Printable version  
Permanent link  
Page information

PMID:24955762

Citation	Korczynska, M, Xiang, DF, Zhang, Z, Xu, C, Narindoshvili, T, Kamat, SS, Williams, HJ, Chang, SS, Kolb, P, Hillerich, B, Sauder, JM, Burley, SK, Almo, SC, Swaminathan, S, Sholchet, BK and Reuschel, FM (2014) Functional annotation and structural characterization of a novel lactonase hydrolyzing D-xylo-1,4-lactone-5-phosphate and L-arabino-1,4-lactone-5-phosphate. <i>Biochemistry</i> <b>53</b> :4727-38	<a href="#">Contents [hide]</a> <a href="#">1 Significance</a> <a href="#">2 Annotations</a> <a href="#">3 Notes</a> <a href="#">4 See also</a> <a href="#">5 References</a>
Abstract	A novel lactonase from <i>Mycoplasma synoviae</i> 53 (MS53_0025) and <i>Mycoplasma agalactiae</i> PG2 (MAG_6390) was characterized by protein structure determination, molecular docking, gene context analysis, and library screening. The crystal structure of MS53_0025 was determined to a resolution of 2.06 Å. This protein adopts a typical amidohydrolase (ββα)-fold and contains a binuclear zinc center located at the C-terminal end of the β-barrel. A phosphate molecule was bound in the active site and hydrogen bonds to Lys217, Lys244, Tyr245, Arg275, and Tyr278. Both docking and gene context analysis were used to narrow the theoretical substrate profile of the enzyme, thus directing empirical screening to identify that MS53_0025 and MAG_6390 catalyze the hydrolysis of d-xylo-1,4-lactone-5-phosphate (2) with <i>K<sub>cat</sub>/K<sub>m</sub></i> values of $4.7 \times 10(4)$ and $5.7 \times 10(4)$ M <sup>-1</sup> s <sup>-1</sup> ) and l-arabino-1,4-lactone-5-phosphate (7) with <i>K<sub>cat</sub>/K<sub>m</sub></i> values of $1.3 \times 10(4)$ and $2.2 \times 10(4)$ M <sup>-1</sup> s <sup>-1</sup> ), respectively. The identification of the substrate profile of these two phospho-furanose lactonases emerged only when all methods were integrated and therefore provides a blueprint for future substrate identification of highly related amidohydrolase superfamily members.	
Links	<a href="#">PubMed</a> <a href="#">PMO4108184</a> <a href="#">Online version:10.1021/bi500595c</a>	
Keywords		
<a href="#">edit table</a>		

Abstract and full text links are automatically created

Significance [\[edit\]](#)

Annotations [\[edit\]](#)

Showing 1 to 2 of 2 entries


Filter Rows:

Evidence: [Any/All](#) [S](#)

Gene product	Qualifier	GO ID	GO term name	Evidence Code	with/from	Aspect	Notes	Status
MYCAP_A51280		<a href="#">GO:0050490</a>	1,4-lactonase activity	IDA: Inferred from Direct Assay		P	Table 3 shows kinetic parameters for Mag6390 with various substrates	complete

Anywhere the reference is used in an annotation is shown

# Create/Find a page for your protein



Special page

**Interested in participating in CACAO this fall? Sign up now!!!**  
**Instructors, want your students to compete this session?**  
**Students, have any questions? Please email us at [ecoliwiki@gmail.com](mailto:ecoliwiki@gmail.com)**  
**TAMU students, come see us in Bio/Bio Room 443!**

## Create New Gene Page

To create a new gene page, please select a database and enter a unique identifier such as an ID or an accession number.  
Please be patient, creating a page may take up to 30 seconds.

[ [edit](#) ]

Main Page  
Enter GO at the Top  
Help  
Report Bug  
Annotation Jamborees  
Recent Changes  
Login / Create Account

▼ Cacao  
About CACAO  
Create New Gene Page  
Create New Literature Page  
Fall 2015


▼ Links  
[r37r Wikidata](#)

# Protein Pages in GONUTS

The screenshot shows the UniProt entry for LAMBD:VLYS. The top section contains basic information: Species (Enterobacteria phage lambda (Bacteriophage lambda) [1]), Gene Name(s) (l), Protein Name(s) (gpG protein Lysis protein S Lysis Inhibitor), and External Links. The bottom section is a table of annotations. A red box labeled "Annotations" points to the table header. Another red box labeled "Annotations" points to the table content. A third red box labeled "Notes" points to the "Notes" column header.

Qualifier	GO ID	GO term name	Reference	Evidence Code	with/from	Aspect	Notes	Status
	GO:0018008	membrane	GO_REF:0000054	IEA: Inferred from Electronic Annotation	SP_KW:KW-0472	C	Seeded From UniProt	
	GO:0005644	host cell membrane	GO_REF:0000054	IEA: Inferred from Electronic Annotation	SP_KW:KW-1043	C	Seeded From UniProt	

# Entering/Editing Annotations



special page

The Spring 2012 s

## TableEdit

### LAMBD:VLYS

Qualifier	<input type="text"/>
GO ID	<input type="text"/>
GO term name	<input type="text"/>
Reference	<input type="text"/>
Evidence Code	<input type="text"/>
with/from	<input type="text"/>
Aspect	<input type="text"/>
Notes	<input type="text"/>
Status	Missing: GO ID, evidence, reference
<input type="button" value="Public"/> <input type="button" value="Refresh"/> <input type="button" value="Save Row"/> <input type="button" value="Cancel"/>	

Public rows can be edited or deleted by any user who can edit  
Private rows can be edited or deleted by their creator, or by admins

navigation

- Main Page
- Enter GO at the top
- Help
- Report Bug
- Update log
- Annotation Jamboree
- Recent changes
- Create New Gene Page
- Login / Create Account

social

- Links about GADAO
- Fall 2011

journal clubs

- Journal Clubs
- Create new literature page

search



# Evidence Pull-Down Menu



[Main Page](#)  
[Enter GO at the Top](#)  
[Help](#)  
[Report Bug](#)  
[Annotation Jamboree](#)  
[Recent Changes](#)  
[Login / Create Account](#)

▼ [Cases](#)  
[About CACAO](#)  
[Create New Case Page](#)  
[Create New Literature Page](#)  
[Full 2010](#)

▼ [Links](#)  
[GO Website](#)  
[GO Ontology Tracker](#)  
[Map GO terms](#)

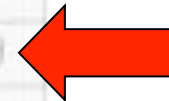
▼ [Tools](#)  
[Upload file](#)  
[Special pages](#)  
[Printable version](#)

Special page

## TableEdit

LAMBDA:VLYS


Qualifier	<input type="text" value="1"/>
GO ID	<input type="text"/>
GO term name	
Reference	<input type="text" value="1"/> <input type="text"/>
Evidence Code	<input type="text" value="1"/>
withfrom	
Aspect	
Notes	<div></div>
Status	Missing: GO ID, evidence, reference
<input type="button" value="Public"/> <input type="button" value="Refresh"/> <input type="button" value="Save Row"/> <input type="button" value="Cancel"/>	



Public rows can be edited or deleted by any user who can edit.  
Private rows can be edited or deleted by their creator, or by admins.



# Note Required for CACAO



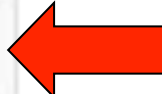
Special page

## TableEdit

### LAMBD:VLYS

Qualifier	<input type="text"/>
GO ID	<input type="text"/>
GO term name	<input type="text"/>
Reference	<input type="text"/>
Evidence Code	<input type="text"/>
withfrom	<input type="text"/>
Aspect	<input type="text"/>
Notes	<div><div></div></div>
Status	Missing: GO ID, evidence, reference

Public rows can be edited or deleted by any user who can edit



# Note Required for CACAO

- Notes **must** include:
  - Figure(s) and/or table(s) you used to choose the Evidence Code
    - You can narrow this down to panels if needed
  - Organism (strain if applicable)
  - Protein name/Gene name as paper refers to it
  - Protein name/Gene name as UniProt refers to it (if different)

# Note Required for CACAO

- Notes **must** include:
  - Figure(s) and/or table(s) you used to choose the Evidence Code
    - You can narrow this down to panels if needed
  - Organism (strain if applicable)
  - Protein name/Gene name as paper refers to it
  - Protein name/Gene name as UniProt refers to it (if different)
- **Annotation will be marked “Unacceptable” if the above are not present**

# Note Required for CACAO

- Notes **must** include:
  - Figure(s) and/or table(s) you used to choose the Evidence Code
    - You can narrow this down to panels if needed
  - Organism (strain if applicable)
  - Protein name/Gene name as paper refers to it
  - Protein name/Gene name as UniProt refers to it (if different)
- Notes should include
  - Explanation of methods if not obvious
- Notes might include
  - Anything else you need to tell us to support your annotation

## Misc. CACAO advice

- Getting help is not cheating!
  - Talk to your teammates
  - Ask us questions
  - Talk to other professors
  - Email authors of papers
- Please complain if things don't work
  - Much of the custom software for CACAO was written by molecular biologists, not software engineers!
  - [ecoliwiki@gmail.com](mailto:ecoliwiki@gmail.com)

# Example Papers

- Powerpoints and pdfs for different examples can be found at:  
[http://gowiki.tamu.edu/wiki/index.php/Category:CACAO\\_training](http://gowiki.tamu.edu/wiki/index.php/Category:CACAO_training)
  - Component annotation based on immunofluorescence in *B. subtilis*
  - Function annotation based on a study of *E. coli* Topoisomerase IV
  - More coming...

# Possible homework

- Find annotations made by professional biocurators
  - Read the paper
  - Discuss what you would put in the note
- Examine past CACAO annotations
  - Discuss whether the annotations make sense
- Pick some papers that you might annotate
  - In class discussion of how you made the choice and why you think it works for GO annotation

