

BIOINFORMATICS

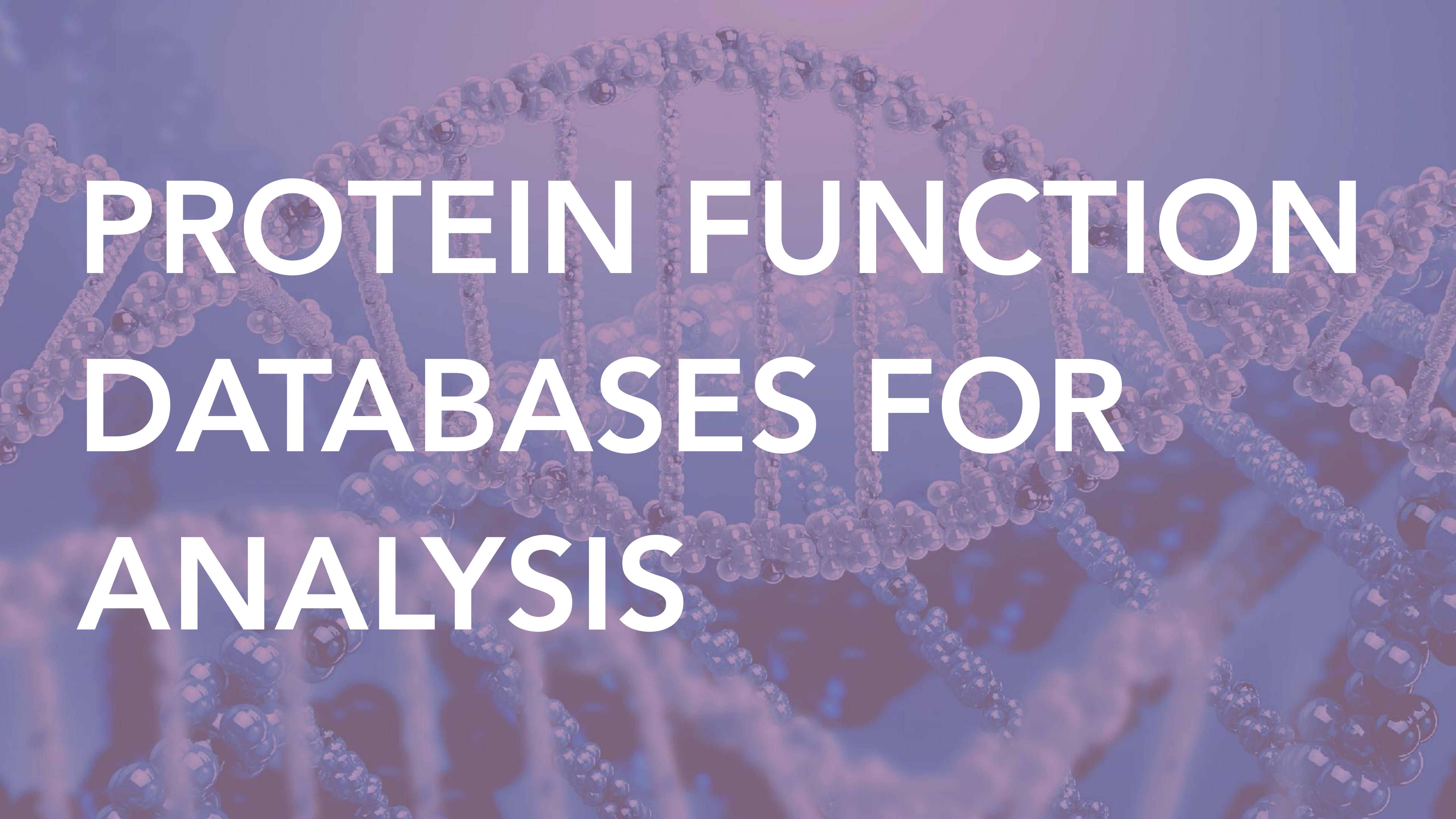
(FOR COMPUTER SCIENTISTS)

MPCS56420
AUTUMN 2020
SESSION 5



THE UNIVERSITY OF
CHICAGO

PROTEIN FUNCTION DATABASES FOR ANALYSIS



ExPASY

ExPASY

- Expert Protein Analysis System
- Prominent databases:
 - SWISS-PROT
 - PROSITE
 - ENZYME
 - SWISS-MODEL Repository

The screenshot shows the ExPASY website interface. At the top, there's a navigation bar with links for Home, About, and Contact. The main header features the SIB logo and the text "ExPASY Bioinformatics Resource Portal". Below the header is a search bar with a dropdown menu set to "Query all databases" and a search button. To the right of the search bar is a "help" link.

A sidebar on the left contains several red-themed buttons labeled "Visual Guidance", "Categories", "Resources A-Z", and "Links/Documentation". The "Categories" button is currently active, showing a list of life science fields: proteomics, genomics, structural bioinformatics, systems biology, phylogeny/evolution, population genetics, transcriptomics, biophysics, imaging, IT infrastructure, and drug design.

The main content area includes a "Featuring today" section highlighting "MSight" (Mass Spectrometry Imager) with a thumbnail image and a "details" link. Below this is a "How to use this portal?" section with a question mark icon and a list of three items: "Features and updates", "New to ExPASY", and "Experienced ExPASY users: what is different".

On the right side, there are two boxes: "Popular resources" listing UniProtKB, SWISS-MODEL, STRING, and PROSITE, and "Latest News" listing OMA orthology DB redesign, UniProt Knowledgebase release 2014_10, and other news items.

At the bottom of the page, there's a footer with links to "SIB Swiss Institute of Bioinformatics | Disclaimer" and a note about displaying a menu for the current URL.

ExPASY

- SWISS-PROT - Protein knowledgebase
 - An annotated protein sequence database established in 1986 at the Department of Medical Biochemistry of the University of Geneva
 - Now maintained at the Swiss Institute of Bioinformatics (SIB) and the European Bioinformatics Institute (EBI)
 - Minimal level of redundancy
 - High level of integration with other databases (currently cross-referenced with about 45 different databases)

The screenshot shows a web browser window for the ExPASY Bioinformatics Resource Portal. The title bar reads "web.expasy.org". The main content area is titled "Documents" and "UniProtKB/Swiss-Prot". A sub-section header "UniProtKB/Swiss-Prot" is followed by a detailed description of the database: "UniProtKB/Swiss-Prot is the manually annotated and reviewed section of the UniProt Knowledgebase (UniProtKB). It is a high quality annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusions. Since 2002, it is maintained by the UniProt consortium and is accessible via the [UniProt website](#)". Below this are links to "List of UniProtKB/Swiss-Prot (reviewed) entries", "Download - UniProt FTP sites", and "Statistics". A section titled "Additional Information:" lists links to "Why is UniProtKB composed of 2 sections, UniProtKB/Swiss-Prot and UniProtKB/TrEMBL?", "Biocuration in UniProt", "How do we manually annotate a UniProtKB entry?", "UniProt manual annotation program", "UniProt general documentation", and "FAQ". Another section titled "Around UniProtKB/Swiss-Prot" lists links to "NextProt - Exploring the universe of human proteins", "Viralzone - A portal to viral UniProtKB/Swiss-Prot entries", "HAMAP - High-quality Automated and Manual Annotation of Proteins", "SwissVar - A portal to Swiss-Prot diseases and variants", "UniPathway - A metabolic door to UniProtKB/Swiss-Prot", and "Enzyme - A repository of information relative to the nomenclature of enzymes". Further down are links to "Swiss-Shop - A service that allows you to automatically obtain (by email) new UniProtKB/Swiss-Prot entries relevant to your field(s) of interest", "Protein Spotlight - One month, one protein", "Proteomics tools", "Contact the UniProt helpdesk", and the URL "http://www.uniprot.org/contact". At the bottom, there is a footer with links to "SIB Swiss Institute of Bioinformatics | Disclaimer" and "Back to the Top".

ExPASY

- PROSITE - Protein families and domains
 - Database of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family
 - Based on the observation that proteins can be grouped on the basis of similarities in their sequences (signature for a protein family or domain)
 - The protein signatures are provided in PROSITE format
 - This format can also be used to do similarity searching by using PHI-BLAST/NCBI

The screenshot shows a web browser window for the ExPASY Bioinformatics Resource Portal. The title bar reads "web.expasy.org". The main content area is titled "Documents" and features a section for "UniProtKB/Swiss-Prot". It describes UniProtKB/Swiss-Prot as the manually annotated and reviewed section of the UniProt Knowledgebase, maintained by the UniProt consortium since 2002. It includes links for "List of UniProtKB/Swiss-Prot (reviewed) entries", "Download - UniProt FTP sites", and "Statistics". Below this is a "Additional Information:" section with links to "Why is UniProtKB composed of 2 sections, UniProtKB/Swiss-Prot and UniProtKB/TrEMBL?", "Biocuration in UniProt", "How do we manually annotate a UniProtKB entry?", "UniProt manual annotation program", "UniProt general documentation", and "FAQ". Further down are sections for "Around UniProtKB/Swiss-Prot" with links to NextProt, Viralzone, HAMAP, SwissVar, UniPathway, and Enzyme, and "Swiss-Shop" for email alerts. At the bottom, there's a footer with links to "SIB Swiss Institute of Bioinformatics | Disclaimer" and "Back to the Top".

Documents

UniProtKB/Swiss-Prot

UniProtKB/Swiss-Prot is the manually annotated and reviewed section of the UniProt Knowledgebase (UniProtKB). It is a high quality annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusions. Since 2002, it is maintained by the [UniProt consortium](#) and is accessible via the [UniProt website](#).

[List of UniProtKB/Swiss-Prot \(reviewed\) entries.](#)
[Download - UniProt FTP sites.](#)
[Statistics.](#)

Additional Information:

- Why is UniProtKB composed of 2 sections, UniProtKB/Swiss-Prot and UniProtKB/TrEMBL?
- Biocuration in UniProt.
- How do we manually annotate a UniProtKB entry?
- UniProt manual annotation program.
- UniProt general documentation.
- FAQ.

Around UniProtKB/Swiss-Prot

[NextProt](#) - Exploring the universe of human proteins.
[Viralzone](#) - A portal to viral UniProtKB/Swiss-Prot entries.
[HAMAP](#) - High-quality Automated and Manual Annotation of Proteins.
[SwissVar](#) - A portal to Swiss-Prot diseases and variants.
[UniPathway](#) - A metabolic door to UniProtKB/Swiss-Prot.
[Enzyme](#) - A repository of information relative to the nomenclature of enzymes.

[Swiss-Shop](#) - A service that allows you to automatically obtain (by email) new UniProtKB/Swiss-Prot entries relevant to your field(s) of interest.

[Protein Spotlight](#) - One month, one protein.
[Proteomics tools](#)

[Contact the UniProt helpdesk](#)
<http://www.uniprot.org/contact>

SIB Swiss Institute of Bioinformatics | [Disclaimer](#)

[Back to the Top](#)

EXPASY

- ENZYME - Enzyme nomenclature database
 - Repository of information relative to the nomenclature of enzymes
 - Based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB)
 - Describes each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided

The screenshot shows the ENZYME page on the ExPASy Bioinformatics Resource Portal. At the top, there's a logo for SIB (Swiss Institute of Bioinformatics) and ExPASy. The main title is "ENZYME" and the subtitle is "Enzyme nomenclature database". Below this, a large "swissprot" logo is displayed. A text block explains that ENZYME is a repository of enzyme nomenclature based on IUBMB recommendations, mentioning EC numbers and preliminary entries. It also notes the release date (29-Oct-14) and the count of active entries (5543). A search bar and various access options like EC number, enzyme class, and description are shown. Below the search area, sections for "Documents" (user manual, obtaining ENZYME), "Services" (report forms, download via FTP), and "Related tools and databases" (links to BRENDA, IUBMB ExplorEnz, KEGG, MetaCyc, IUBMB Enzyme Nomenclature, and BioCarta) are present.

ENZYME
Enzyme nomenclature database

ENZYME is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) and it describes each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided [More details / References]. ENZYME now includes entries with preliminary EC numbers. Preliminary EC numbers include an 'n' as part of the fourth (serial) digit (e.g. EC 3.5.1.n3).

Release of 29-Oct-14 (5543 active entries)

Access to ENZYME

- by EC number: . . .
- by enzyme class
- by description (official name) or alternative name(s):
- by chemical compound
- by cofactor
- by search in comments lines

Documents

- ENZYME user manual
- How to obtain ENZYME

Services

- Report forms for a new ENZYME entry or for an error/update in an existing entry
- Downloading ENZYME by FTP

Related tools and databases

- BRENDA - Comprehensive Enzyme Information system
- IUBMB ExplorEnz Enzyme database - Explore the IUBMB Enzyme Nomenclature List
- KEGG - Kyoto Encyclopedia of Genes and Genomes
- MetaCyc - Metabolic Encyclopedia of enzymes and metabolic pathways
- IUBMB Enzyme Nomenclature
- BioCarta - Pathways of Life

SIB Swiss Institute of Bioinformatics | Disclaimer
Display a menu ▲ Back to the Top

EXPASY

- ENZYME - Enzyme nomenclature database
 - Format of number
 - Every enzyme code consists of the letters "EC" followed by four numbers separated by periods.
 - Those numbers represent a progressively finer classification of the enzyme
 - e.g. 3.5.1

2. -.-	Transferases.
2. 1. --	Transferring one-carbon groups.
2. 1. 1.-	Methyltransferases.
2. 1. 2.-	Hydroxymethyl-, formyl- and related transferases.
2. 1. 3.-	Carboxyl- and carbamoyltransferases.
2. 1. 4.-	Amidinotransferases.
2. 2. --	Transferring aldehyde or ketone residues.
2. 2. 1.-	Transketolases and transaldolases.
2. 3. --	Acyltransferases.
2. 3. 1.-	Transferring groups other than amino-acyl groups.
2. 3. 2.-	Aminoacyltransferases.
2. 3. 3.-	Acyl groups converted into alkyl on transfer.
2. 4. --	Glycosyltransferases.
2. 4. 1.-	Hexosyltransferases.
2. 4. 2.-	Pentosyltransferases.
2. 4.99.-	Transferring other glycosyl groups.
2. 5. --	Transferring alkyl or aryl groups, other than methyl groups.
2. 5. 1.-	Transferring alkyl or aryl groups, other than methyl groups.
2. 6. --	Transferring nitrogenous groups.
2. 6. 1.-	Transaminases (aminotransferases).
2. 6. 3.-	Oximinotransferases.
2. 6.99.-	Transferring other nitrogenous groups.
2. 7. --	Transferring phosphorous-containing groups.
2. 7. 1.-	Phosphotransferases with an alcohol group as acceptor.
2. 7. 2.-	Phosphotransferases with a carboxyl group as acceptor.
2. 7. 3.-	Phosphotransferases with a nitrogenous group as acceptor.
2. 7. 4.-	Phosphotransferases with a phosphate group as acceptor.
2. 7. 6.-	Diphosphotransferases.
2. 7. 7.-	Nucleotidyltransferases.
2. 7. 8.-	Transferases for other substituted phosphate groups.
2. 7. 9.-	Phosphotransferases with paired acceptors.
2. 7.10.-	Protein-tyrosine kinases.
2. 7.11.-	Protein-serine/threonine kinases.
2. 7.12.-	Dual-specificity kinases (those acting on Ser/Thr and Tyr residues).
2. 7.13.-	Protein-histidine kinases.
2. 7.99.-	Other protein kinases.
2. 8. --	Transferring sulfur-containing groups.
2. 8. 1.-	Sulfurtransferases.
2. 8. 2.-	Sulfotransferases.
2. 8. 3.-	CoA-transferases.
2. 8. 4.-	Transferring alkylthio groups.
2. 9. --	Transferring selenium-containing groups.
2. 9. 1.-	Selenotransferases.
2.10. --	Transferring molybdenum- or tungsten-containing groups.
2.10. 1.-	Molybdenumtransferases or tungstentransferases with sulfide groups.

GENE ONTOLOGY (GO) CONSORTIUM

GENE ONTOLOGY (GO) CONSORTIUM

- The GO Consortium compiles a dynamic, controlled vocabulary of terms related to gene products
 - <http://www.geneontology.org>
- There are three organizing principles:
 - Molecular function
 - Biological process
 - Cellular compartment

geneontology.org

Gene Ontology Consortium

Home Documentation Downloads User stories Community Tools About Contact us

Search

RSS Twitter Facebook

Gene Ontology Consortium

Highlighted GO term

Representing "phases" in GO biological process

The GOC has recently introduced a new term **biological phase** (GO:0044848), as a direct subclass of biological process. This class represents a distinct period or stage during which biological processes can occur.

more

Enrichment analysis (beta)

Your gene IDs here...

biological process H. sapiens Submit Advanced options Powered by PANTHER

Statistics

Ten Quick Tips for Using the Gene Ontology

Important Post date: 11/26/2013 - 08:22

Cardiovascular Gene Annotation Newsletter November 2014

Post date: 11/03/2014 - 03:49

Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt.

Post date: 10/20/2014 - 13:39

How GO uses OWL - conference paper from #owl4ed2014

Post date: 10/19/2014 - 11:22

Missing relationships in previous go release - problem now fixed

Post date: 10/17/2014 - 12:40

more

Tags: navigation User story: Everybody

CytoScape EM Enrichment Map Cytoscape Plugin

Random FAQs

- What is the best way to link into AmiGO?
- I want to use the database files but...
- How do I annotate ESTs?

View all FAQs

Recent news

On the web

Available pipelines for functional annotation (COG, KEGG, GO) of genomes

Tool: dcGOR: an R package for analysing ontologies and protein domain annotations

How to equalize two vector lengths?

GO terms associated to a list of genes

topGO approach for a list of genes without expression

More

Tweets

Display a menu

GENE ONTOLOGY (GO) CONSORTIUM

- There is no centralized GO database
 - Curators of organism-specific databases assign GO terms to gene products for each organism
 - GO terms are assigned to Entrez Gene entries

Display a menu

geneontology.org

Gene Ontology Consortium

Home Documentation Downloads User stories Community Tools About Contact us

Gene Ontology Consortium

Search

RSS Twitter Facebook

Highlighted GO term

Representing "phases" in GO biological process
The GOC has recently introduced a new term **biological phase** (GO:0044848), as a direct subclass of biological process. This class represents a distinct period or stage during which biological processes can occur.

[more](#)

Random FAQs

- What is the best way to link into AmiGO?
- I want to use the database files but...
- How do I annotate ESTs?

[View all FAQs](#)

On the web

Available pipelines for functional annotation (COG, KEGG, GO) of genomes
Tool: dcGOR: an R package for analysing ontologies and protein domain annotations
How to equalize two vector lengths?
GO terms associated to a list of genes
topGO approach for a list of genes without expression
More

Tweets

GENE ONTOLOGY (GO) CONSORTIUM

- Evidence codes
 - IC Inferred by curator
 - IDA Inferred from direct assay
 - IEA Inferred from electronic annotation
 - IEP Inferred from expression pattern
 - IGI Inferred from genetic interaction
 - IMP Inferred from mutant phenotype
 - IPI Inferred from physical interaction
 - ISS Inferred from sequence or structural similarity
 - NAS Non-traceable author statement
 - ND No biological data
 - TAS Traceable author statement

CATALYTIC SITE

ATLAS

CATALYTIC SITE

- The Catalytic Site Atlas (CSA) is a database documenting enzyme active sites and catalytic residues in enzymes of 3D structure
- A classification of catalytic residues which includes only those residues **thought** to be directly involved in some aspect of the reaction catalysed by an enzyme

EMBL-EBI 

Services Research Training About us

Catalytic Site Atlas

Thornton Group > CSA

Enter a PDB code, UniProKB code or EC number in one of the boxes Below to obtain catalytic residue details from the CSA.

Search The CSA

PDB ID	<input type="text"/>	<input type="button" value="SEARCH CSA"/>
JNPRCT ID	<input type="text"/>	<input type="button" value="SEARCH CSA"/>
EC Number	<input type="text"/>	<input type="button" value="SEARCH CSA"/>

A NEW VERSION OF THE CSA UPDATED 14th November 2013

- To reference the CSA please use the latest CSA paper
Furnham N, Holliday GL, de Beer TA, Jacobsen JO, Pearson WR, Thornton JM. The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D485-9. PubMed PMID: 24319146.

Introduction

The Catalytic Site Atlas (CSA) is a database documenting enzyme active sites and catalytic residues in enzymes of 3D structure. We defined a classification of catalytic residues which includes only those residues thought to be directly involved in some aspect of the reaction catalysed by an enzyme.

The CSA contains 2 types of entry:

- Original hand-annotated entries, derived from the primary literature. References for these entries are given.
- Homologous entries, found by sequence comparison methods to one of the original entries. The equivalent residues, which align in sequence to the catalytic residues found in the original entry are documented.

Access to the CSA is via PDB ID, UniProtKB ID or E.C. number. Accessing via PDB ID takes you straight to the CSA entry for that PDB, while accessing via UniProtKB ID or E.C. number gives a list of all PDB codes for structures assigned that particular UniProtKB identifier or E.C. number.

Each CSA entry lists the catalytic residues found in that entry, using PDB residue numbering. Each site is also marked with an evidence tag, which is either "Literature reference" or "Homologyle". If the entry is by sequence comparison you can follow the link to the original entry. The active site can be visualised using a JMol viewer and each catalytic site in the structure can be highlighted and zoomed into by selecting from the drop down list on the left hand side of the viewer as well as further rendering options.

Each entry contains a link to a list of homologous entries found by homology, and a link to other PDB structures with identical E.C. numbers or UniProtKB identifier to the entry you are viewing.

A number of people have contributed to the CSA over the years as annotators. We would like to thank Jonathan Barker, Carine Berezin, Amy Buchanan-Huges, Lynn Carr, Olivia Chan, Josephine Charalambous, Emma Compton, Atlanta Cook, Jennifer Dawe, Angelica Datta, Christian Drew, Alex Gutteridge, Stephanie Juniat, Roman Laskowski, Oleg Lenine, Mei Leung, Stuart Lucas, Ben McLeod, Malcolm MacArthur, Gary McDowell, Angela Malumbe, Duncan Milburn, Fiona Morgan, James Murray, Nozomi Nagano, Jonathan Ng, Emma Penn, Craig Porter, Judith Reeks, Peter Sarkies, Steven Smith, James Torrance, Annabel Todd, Andrew Wallace, Anna Waters, Sophie Williams, and Eleanor Wright.

References

- To reference the CSA please use the latest CSA paper
Furnham N, Holliday GL, de Beer TA, Jacobsen JO, Pearson WR, Thornton JM. The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D485-9. PubMed PMID: 24319146.
- The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Craig T. Porter, Gail J. Bartlett, and Janet M. Thornton (2004) *Nucl. Acids. Res.* 32: D129-D133.
- Analysis of Catalytic Residues in Enzyme Active Sites. Gail J. Bartlett, Craig T. Porter, Neera Borkakoti, and Janet M. Thornton (2002) *J Mol Biol* 324:105-121.
- Using a Library of Structural Templates to Recognise Catalytic Sites and Explore their Evolution in Homologous Families. James W. Torrance, Gail J. Bartlett, Craig T. Porter, Janet M. Thornton (2005) *J Mol Biol.* 347:565-81

EMBL-EBI 

Services Research Training Industry About us

News Brochures Contact us By topic By name (A-Z) Help & Support Overview Publications Research groups Overview Train at EBI Train outside EBI Overview Members Area Workshops Overview Leadership Funding

Display a menu

CATALYTIC SITE

- The CSA contains 2 types of entry:
 - Original hand-annotated entries, derived from the primary literature
 - References for these entries are given.
 - Homologous entries, found by sequence comparison methods to one of the original entries
 - The equivalent residues, which align in sequence to the catalytic residues found in the original entry are documented

The screenshot shows the homepage of the Catalytic Site Atlas (CSA). The header includes the EMBL-EBI logo, a search bar, and links for Services, Research, Training, About us, and cbi.ac.uk. The main title "Catalytic Site Atlas" is displayed prominently. A sidebar on the left provides navigation links for CSA Home, Browse Lit Entries, Help & Documentation, Downloads, and Contact: The Developers. It also features a "Latest EBI News" section with a link to "View the latest EBI news stories and important announcements... more". The main content area displays search fields for PDB ID, UniProt ID, and EC Number, each with a "SEARCH CSA" button. A red banner at the top of the content area states "A NEW VERSION OF THE CSA UPDATED 14th November 2013". Below this, a list of references to the CSA paper is provided, along with the authors' names and publication details. The content area also includes sections for "Introduction", "Access to the CSA", "Each CSA entry lists the catalytic residues found in that entry", "Each entry contains a link to a list of homologous entries", and "A number of people have contributed to the CSA over the years as annotators". At the bottom, there is a "References" section with a list of papers and their authors. The footer contains links for EMBL-EBI services, research, training, industry, and about us, along with a "Display a menu" link.

CATALYTIC SITE

Catalytic Site Atlas



- CSA Home
- Browse Lit Entries
- Help & Documentation
- Downloads
- Contact The Developers

Latest EBI News 
View the latest EBI news stories and important announcements...
[more](#)

Thornton Group CSA > 12as

Search The CSA

PDB ID	<input type="text"/>	<input type="button" value="SEARCH"/>	UNIPROT ID	<input type="text"/>	<input type="button" value="SEARCH"/>	EC Number	<input type="text"/>	<input type="button" value="SEARCH"/>
--------	----------------------	---------------------------------------	------------	----------------------	---------------------------------------	-----------	----------------------	---------------------------------------

CSA LITERATURE entry for 12as

E.C. name	aspartate---ammonia ligase
Species	<i>Escherichia coli</i> (Bacteria)
E.C. Number (IntEnz)	6.3.1.1
CSA Homologues of 12as	11as ,
CSA Entries With UniProtID	P00963
CSA Entries With EC Number	6.3.1.1
PDBe Entry	12as
PDBSum Entry	12as
MACIE Entry	M0075

Show:



Options:

Antialias

You do not have Java applets enabled in your web browser, or you
Check the warning message from your browser and/or e
your web browser preferences, or install the Java Runtime Env

CATALYTIC SITE

Catalytic Sites for 12as

Annotated By Reference To The Literature - Site 1 (Perform Site Search)

Residue	Chain	Number	UniProtKB Number	Functional Part	Function	Target	Description
Asp	A	46	46	macie:sideChain			Deprotonates the ammonia molecule to activate it as a nucleophile, and donates the proton back to AMP. Also stabilises the positive charge on the substrate nitrogen during the transition state.
Gln	A	116	116	macie:sideChain			The NH ₂ group forms part of oxyanion hole to stabilise the negative charge on oxygen during the transition state. The oxygen of Gln116 also stabilises the positive charge on the substrate nitrogen during the transition state.
Arg	A	100	100	macie:sideChain			Forms part of oxyanion hole to stabilise the negative charge on substrate oxygen during the transition state.

Literature References

Notes:

Nakatsu T

Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase.

Nat Struct Biol 1998 5 15-19

PubMed:
[9437423](#)

Format: CODATA

ENTRY
SEQUENCE

P00963

5 10 15 20 25 30

BIOINFORMATICS

(FOR COMPUTER SCIENTISTS)

MPCS56420
AUTUMN 2020
SESSION 5



THE UNIVERSITY OF
CHICAGO