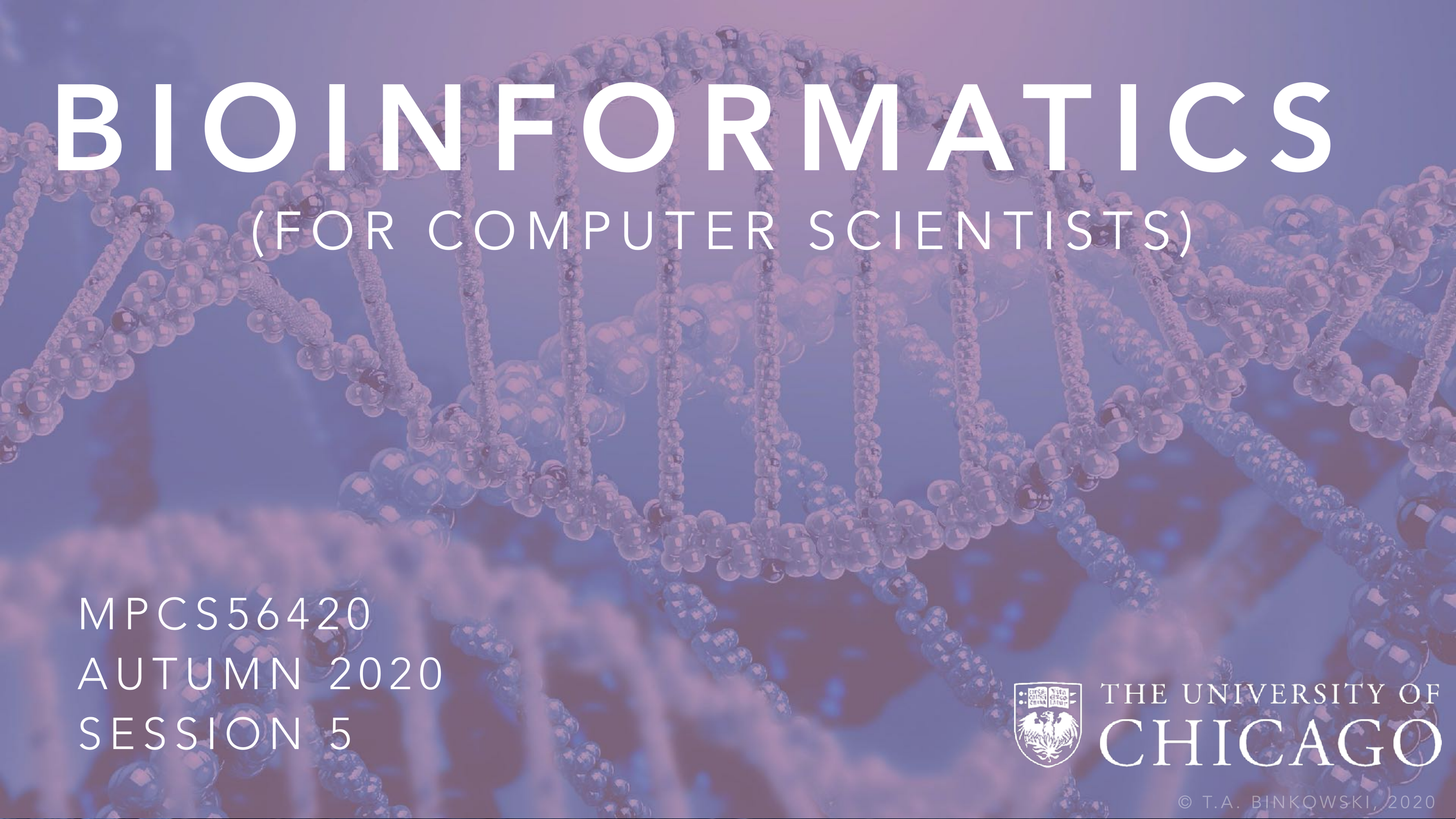
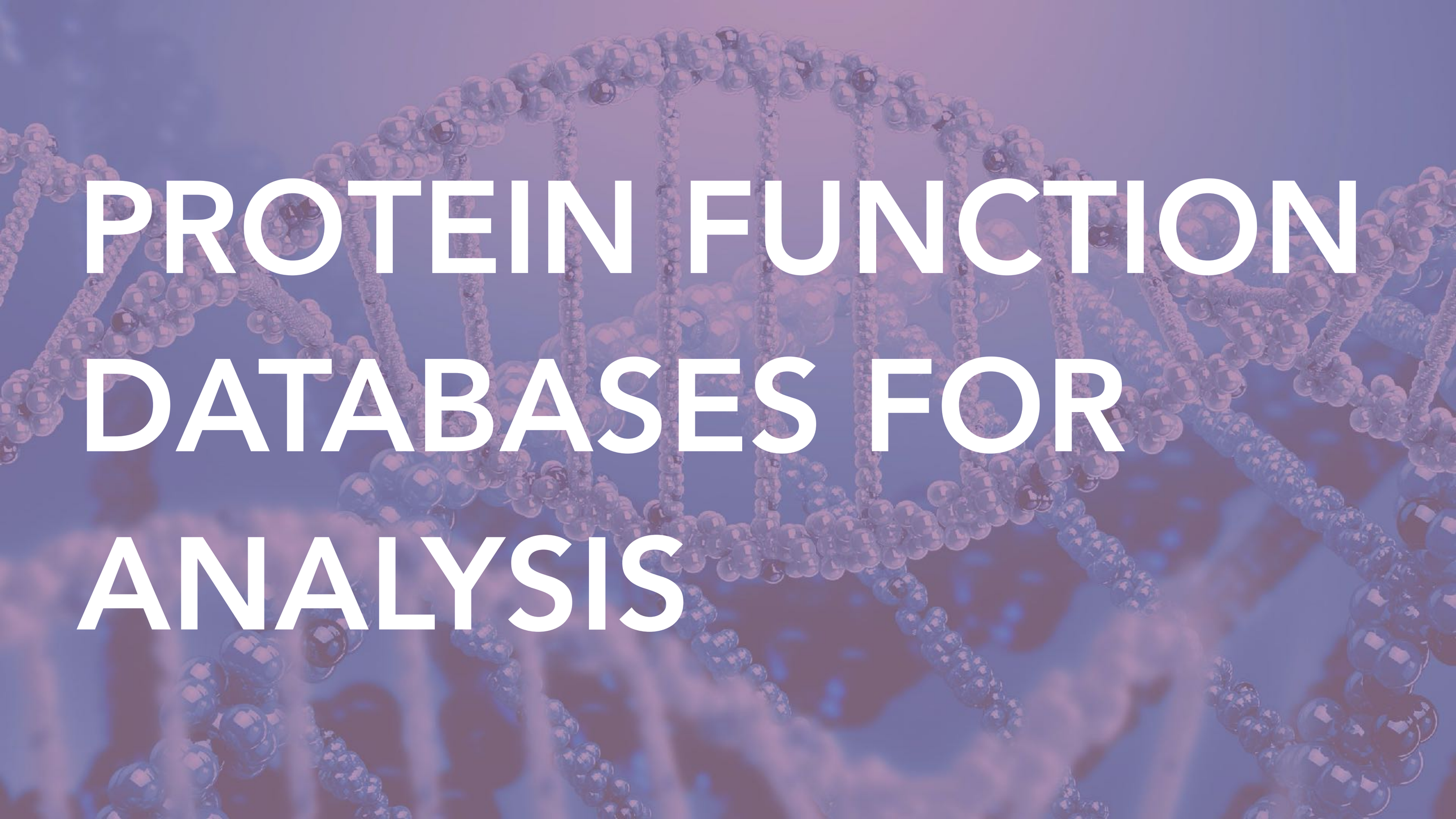


# BIOINFORMATICS

## (FOR COMPUTER SCIENTISTS)

MPCS56420
AUTUMN 2020
SESSION 5

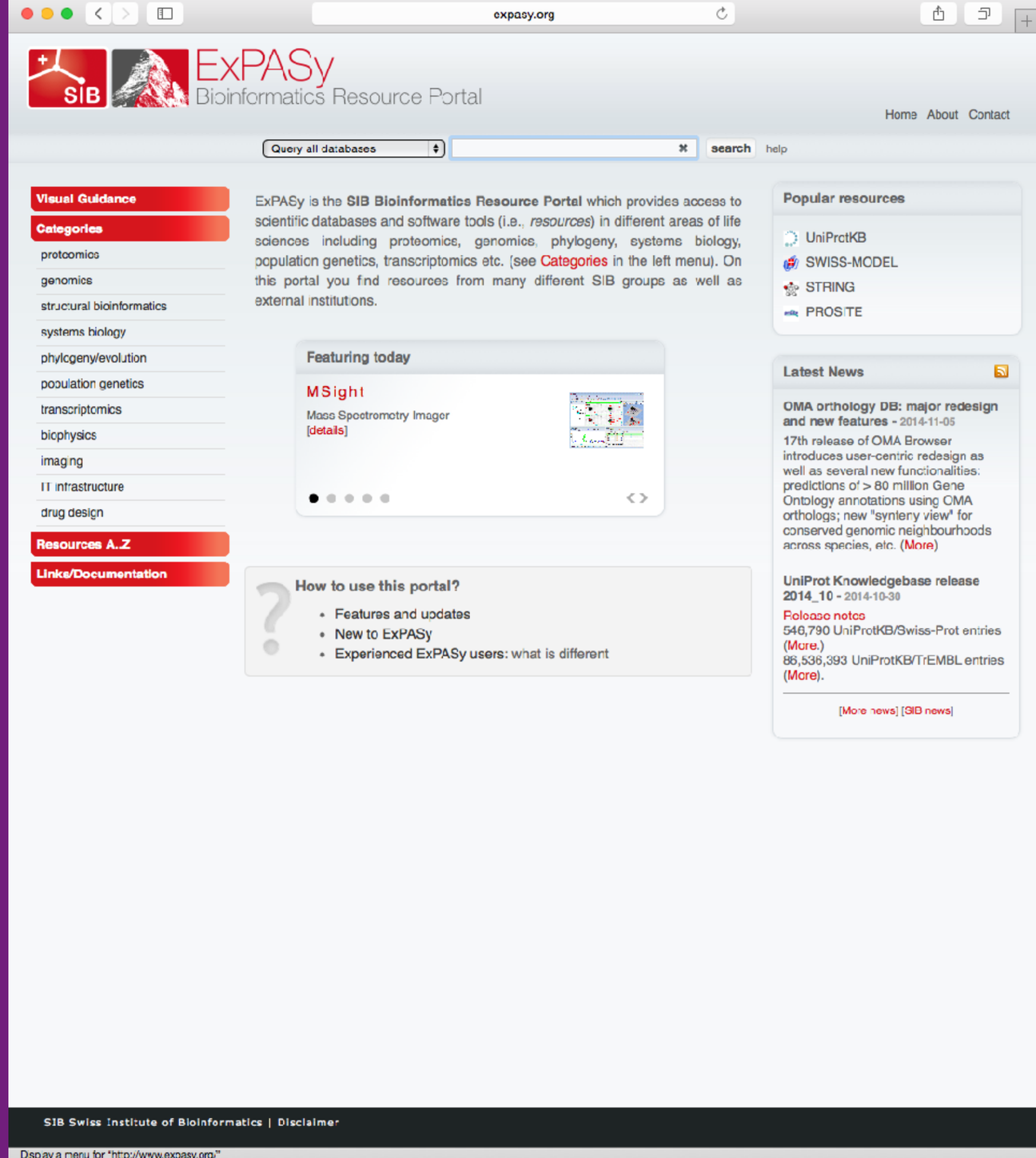
THE UNIVERSITY OF CHICAGO


PROTEIN FUNCTION
DATABASES FOR
ANALYSIS

# ExPASY

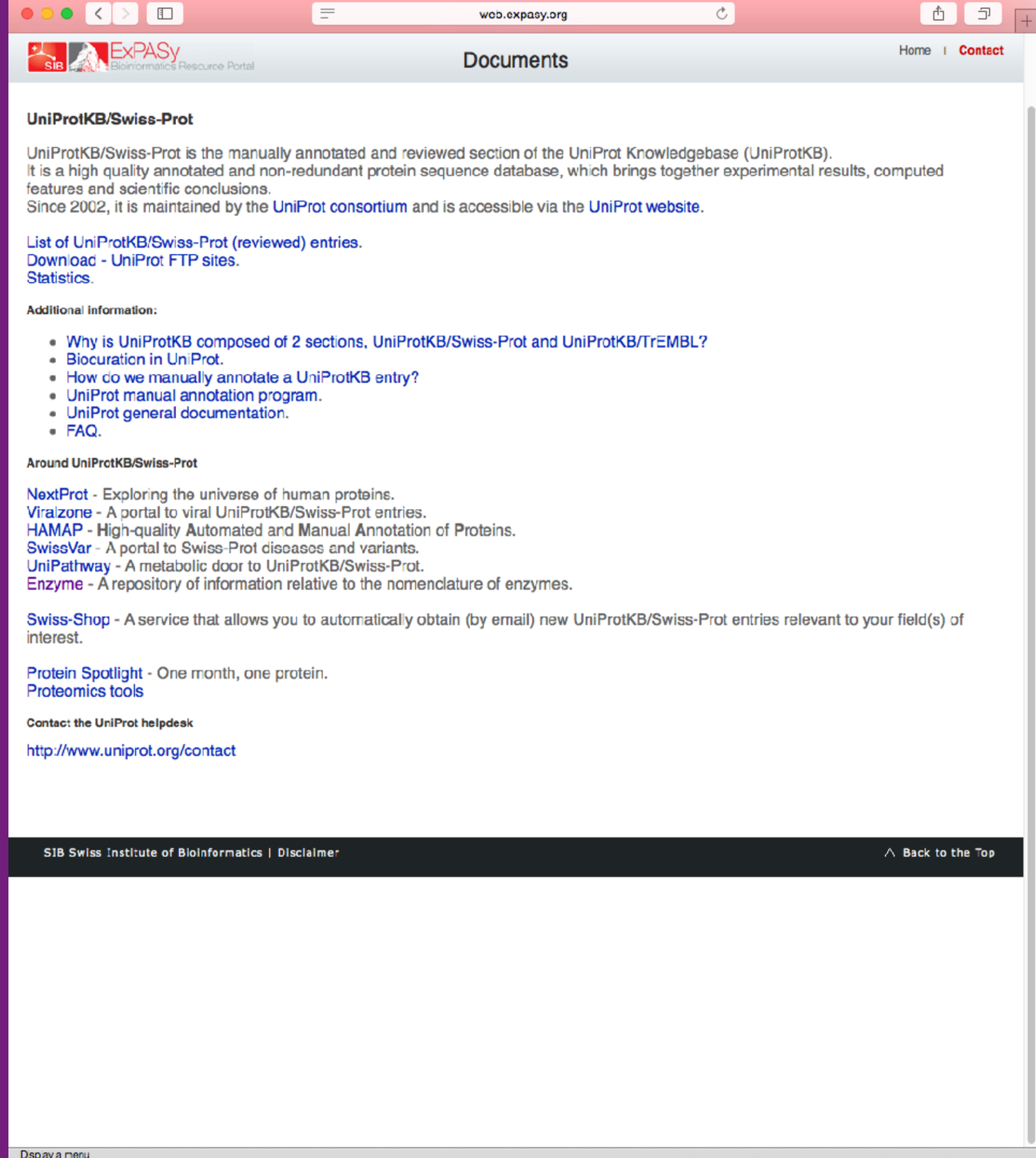
# ExPASY

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



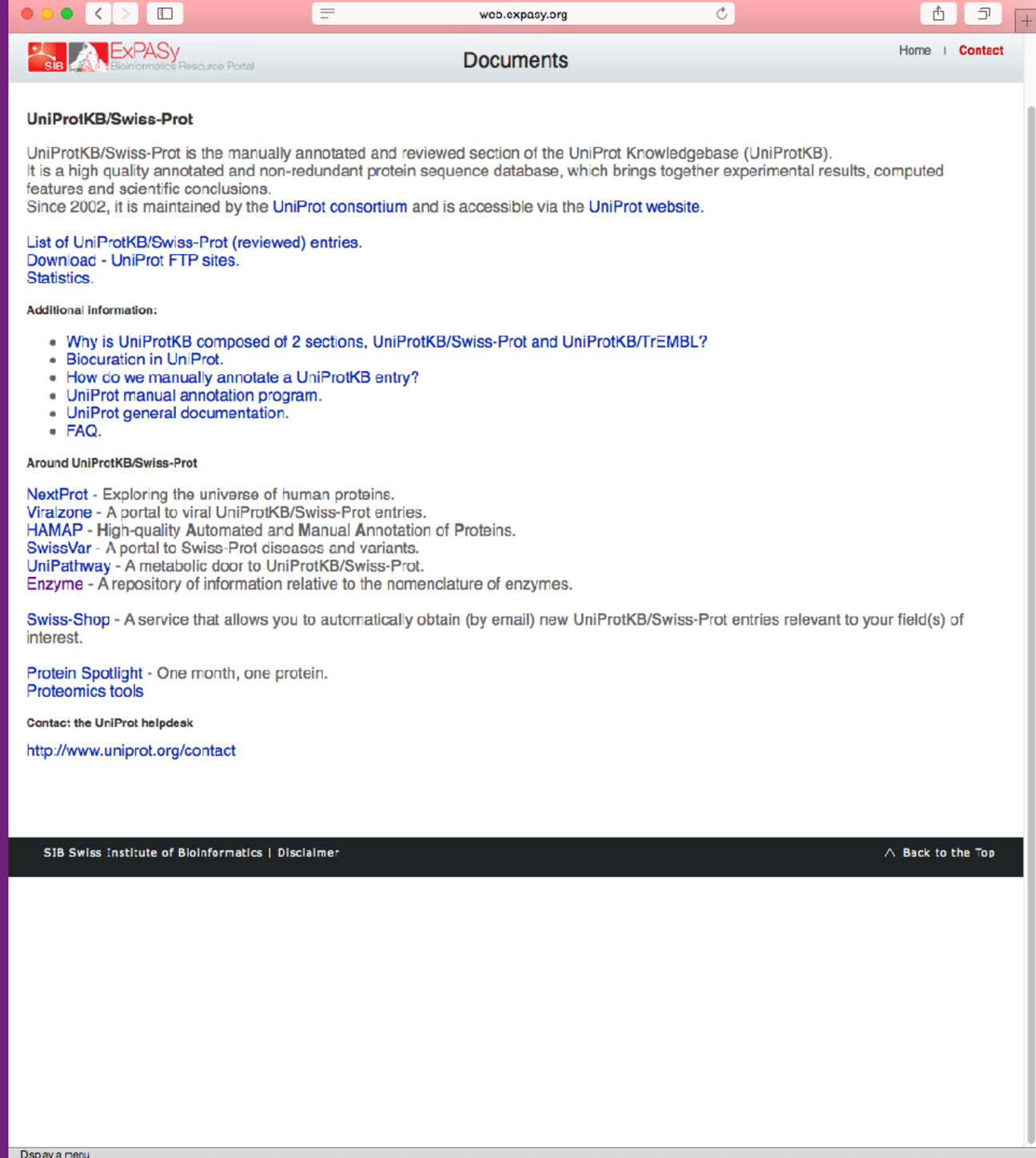
# ExPASY

- 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)



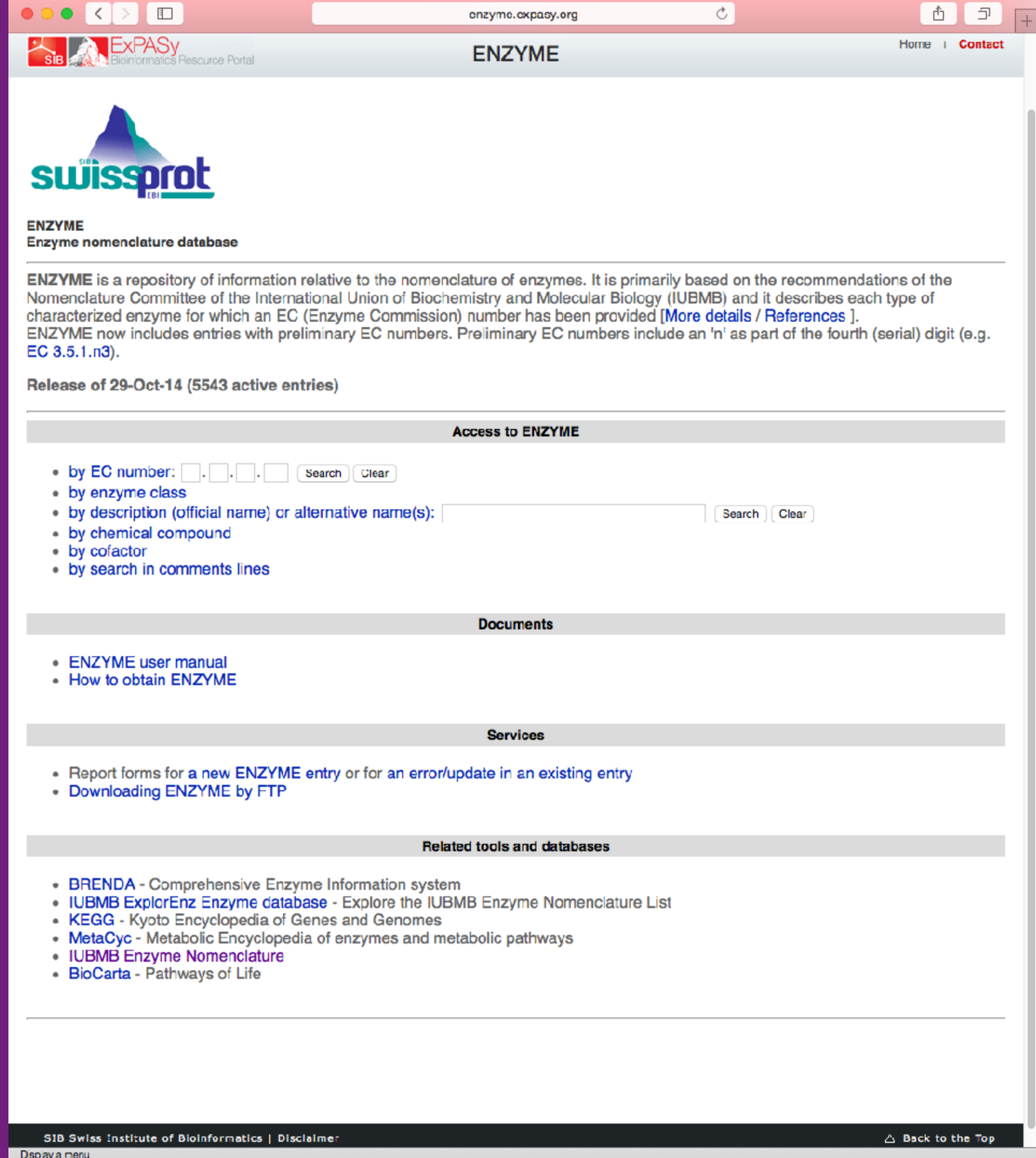
# ExPASY

- 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



# 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



# EXPASY

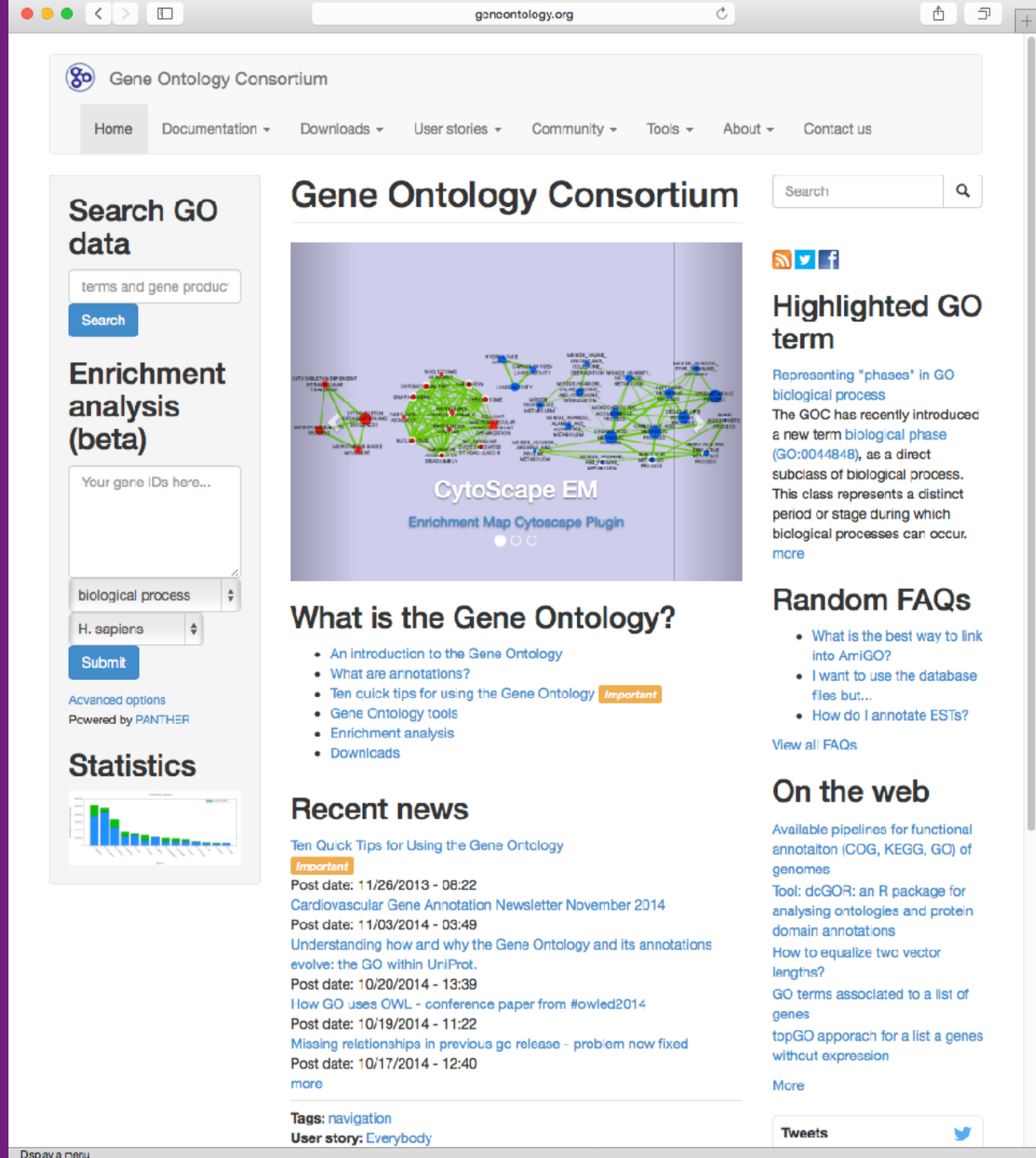
- ENZYME - Enzyme nomenclature database
  - Format of number
  - Every enzyme code consists of the letters "EC" followed by four numbers separated by periods.
  - Those numbers represent a progressively finer classification of the enzyme
    - *e.g. 3.5.1*

```
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 residu
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 grou
```

# Gene Ontology (GO) Consortium

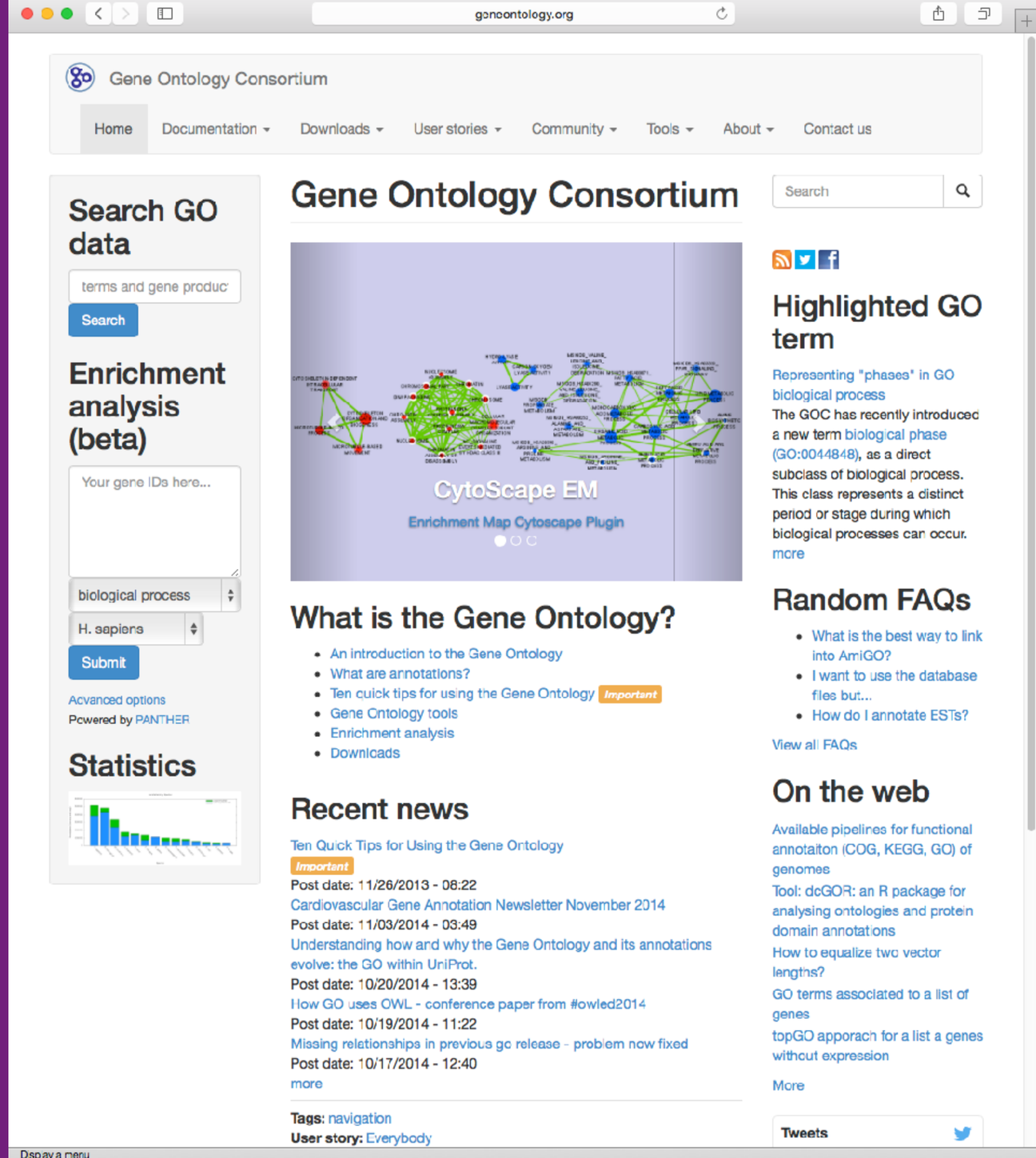
# GENE ONTOLOGY (GO) CONSORTIUM

- 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 proces
  - Cellular compartment



# 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



# GENE ONTOLOGY (GO) CONSORTIUM

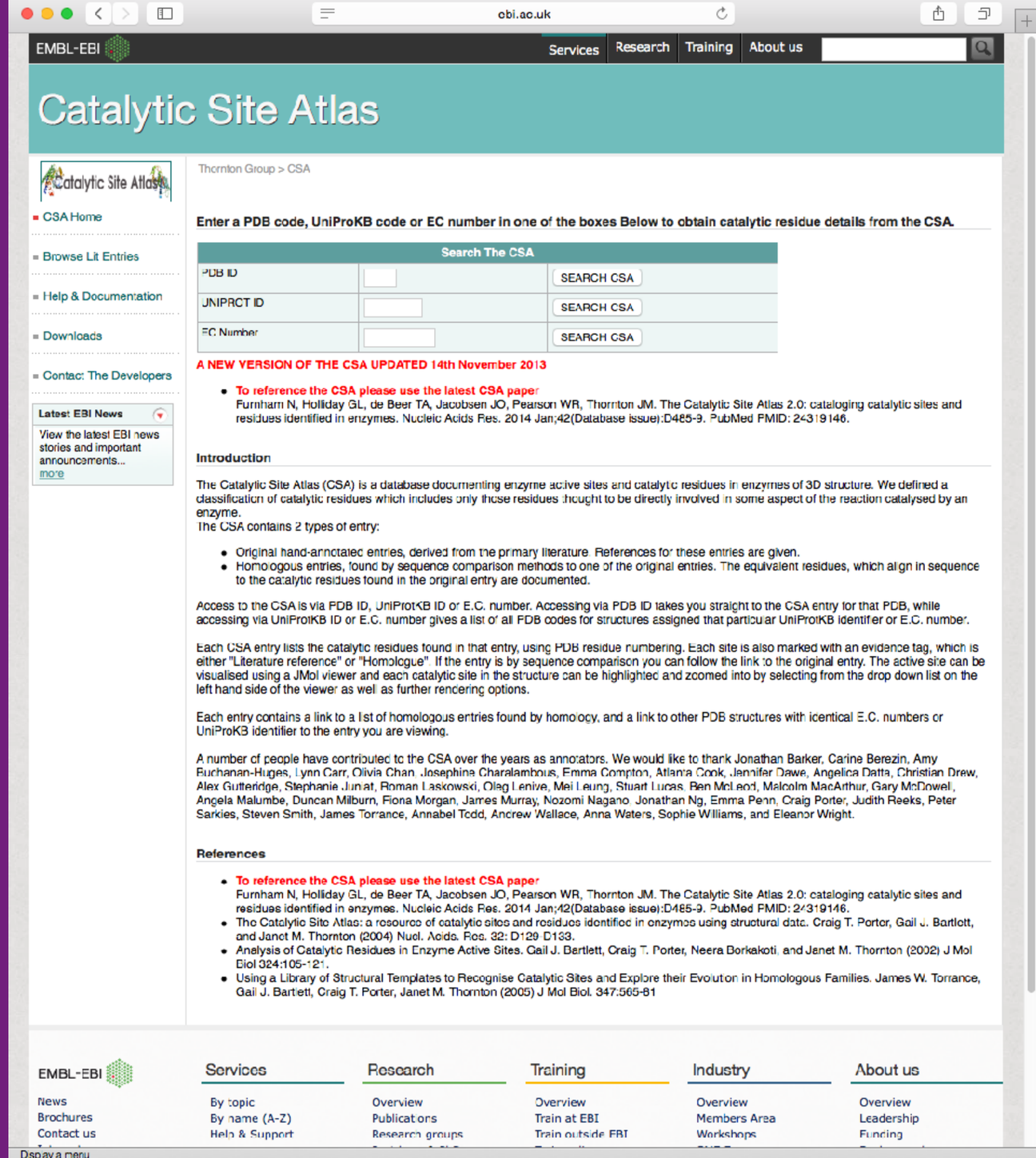
- Evidence codes
  - IC Inferred by curator
  - IDA Inferred from direct assay
  - IEA Inferred from electronic annotation
  - IEP Inferred from expression pattern
  - IGIInferred from genetic interaction
  - IMP Inferred from mutant phenotype
  - IPI Inferred from physical interaction
  - ISS Inferred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

# Catalytic Site Atlas

# CATALYTIC SITE

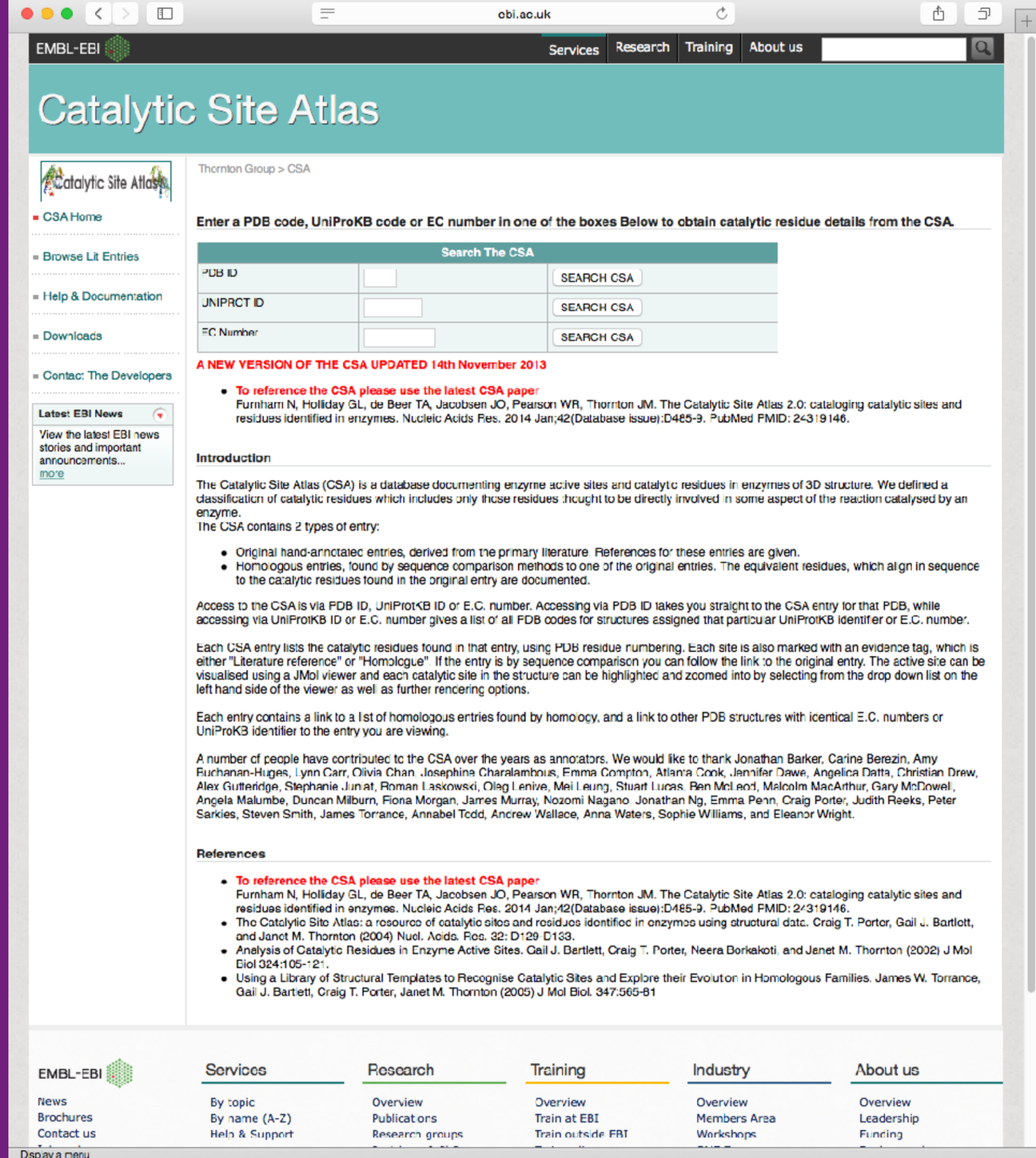
- The Catalytic Site Atlas (CSA) is a database documenting enzyme active sites and catalytic residues in enzymes of 3D structure

- 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

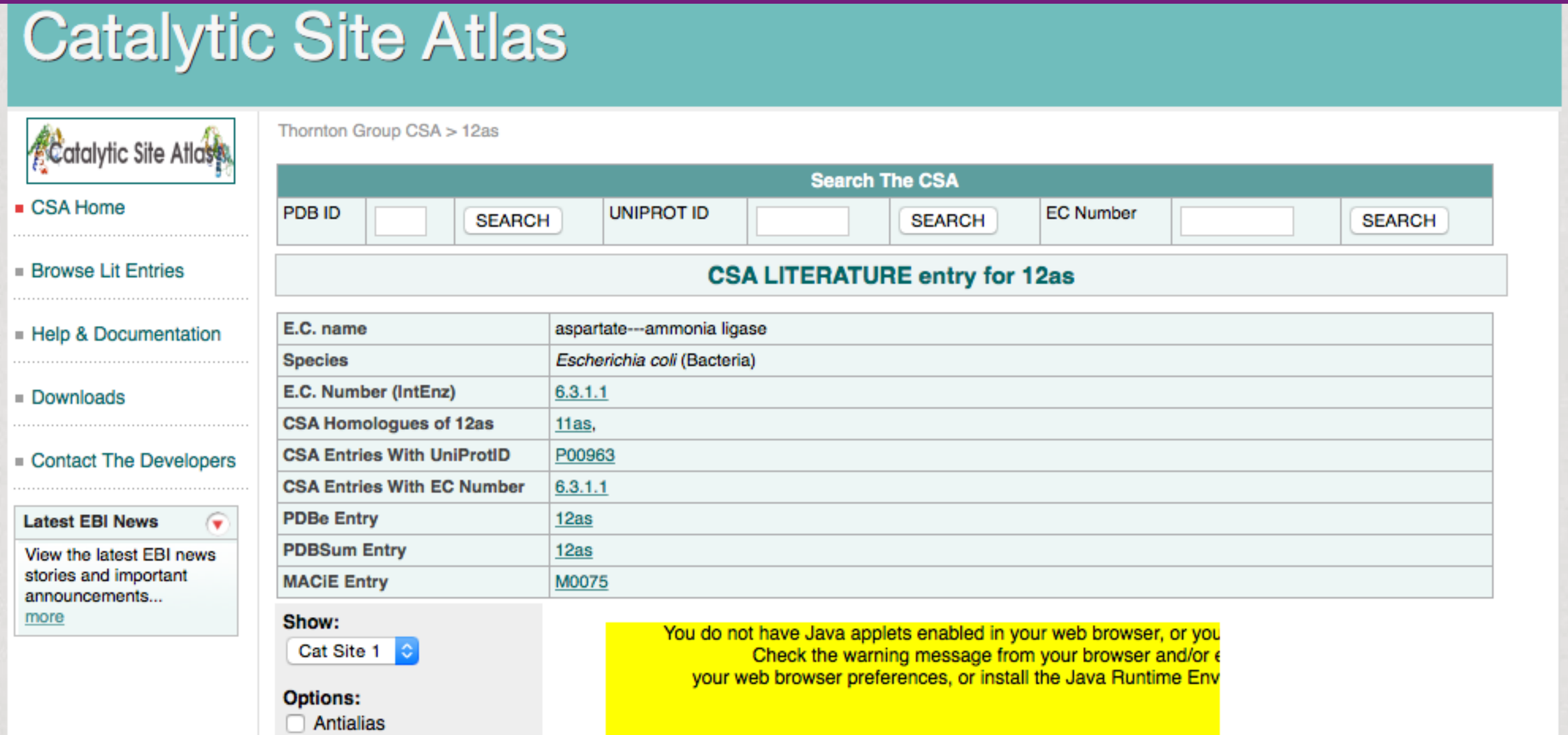


# CATALYTIC SITE

- 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



# CATALYTIC SITE



# 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 NH2 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

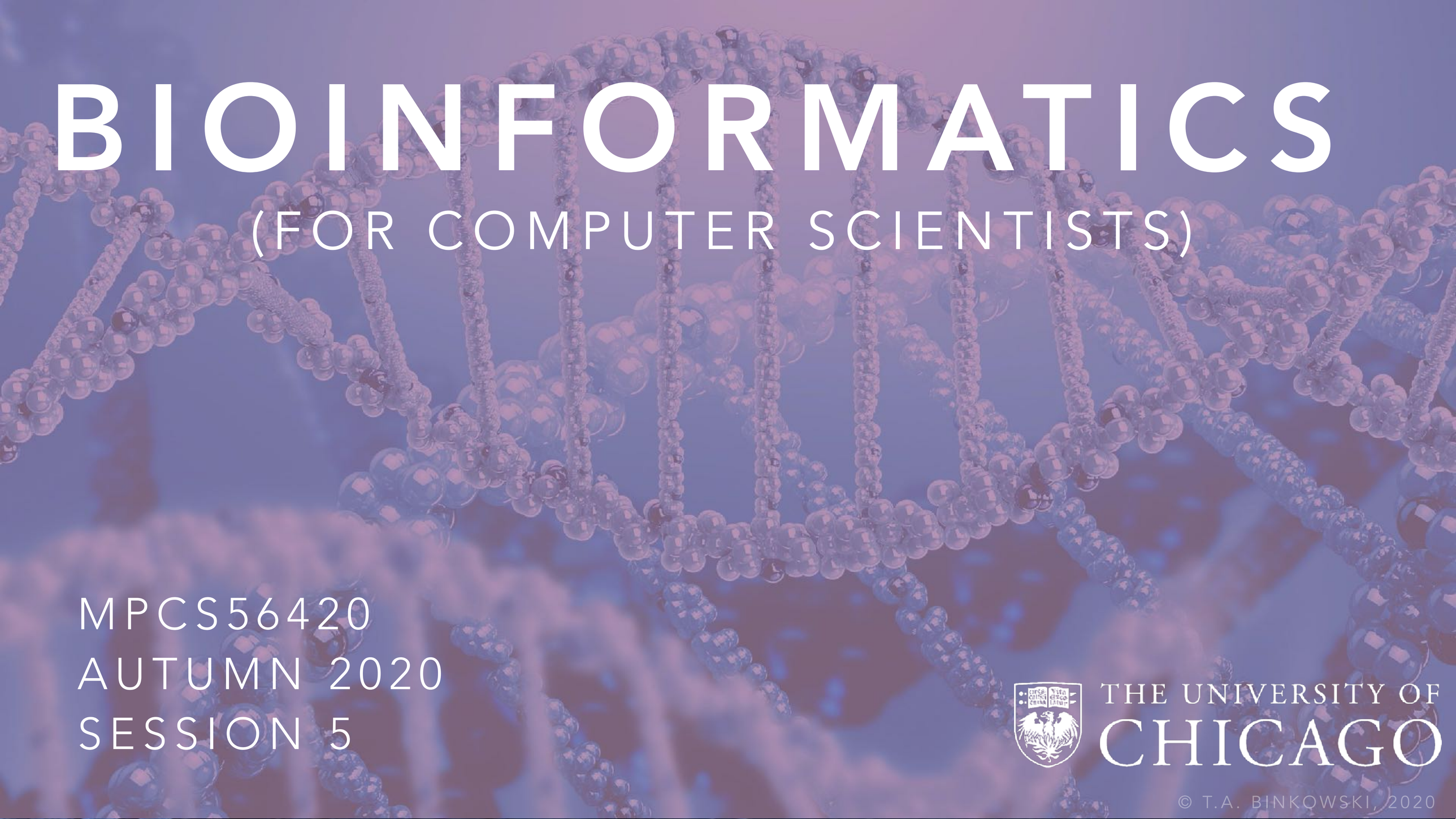
**Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase.**

*Nat Struct Biol 1998 5 15-19*

PubMed: 9437423

Format: CODATA

```
ENTRY           P00963
SEQUENCE
                5         10        15        20        25        30
```



# BIOINFORMATICS

## (FOR COMPUTER SCIENTISTS)

MPCS56420
AUTUMN 2020
SESSION 5

THE UNIVERSITY OF CHICAGO