Protein Function Classification* Schemes

*: note that “classification” (bio) ≠ “classification” (cs)

specifically: E.C. (Enzyme Classification, Boyer)  
Gene Ontology (GO)  
FunCat (MIPS)

• Accompanying reading:  
Pandey Tech Report, Chp 2 (pp. 6-13)  
(http://www.cs.umn.edu/tech_reports_upload/tr2006/06-028.pdf)

HOW DO WE DEFINE FUNCTION?

• Medical function: e.g. tumor suppressor

• Cellular role / location: pathway  
regulatory network

Molecular function:

- enzymes: catalysis  
substrates, cofactors  
nucleic acids
- regulator proteins  
small molecules, metal ions, electrons
- transport / storage proteins  
proteins  
channels, pores  
(e.g. porin)
- structural proteins  
(e.g. collagen, silk, crystallin)

BINDING AND RELEASENG  
BINDING PARTNER?

GENE ONTOLOGY: Using a common language consistently  
http://www.geneontology.org
Computational methods for function prediction

- accuracy ? (! predictability)
- utility ? (! of interest to user/biologist)

“function” is not as well-defined as “structure”

… in fact, one could argue it isn’t defined at all (and therefore description is as good as it gets)!

The information about the functional attributes of genes and proteins is traditionally hidden in the prose of biological literature (for example, a gene product can be described in terms of its biochemistry, molecular activity, cellular function, and physiological role)

[Human serine protease trypsin: Biochemically it catalyses the hydrolysis of peptide bonds following lysine or arginine residues in peptides, its molecular activity is as a proteolytic enzyme, its cellular function is protein degradation, and its physiological role is to aid digestion.]

Harvesting the full power of automatic information management requires databases to consistently present the information content using standardized vocabularies based on biological concept.

Ref: Rison et. al, Funct integr Genomics, 2000
Protein Function Classification/Categorization Schemes

- **E.C. (Enzyme Classification, Boyer)**
  highly specialized, hierarchical, ~1956

- **FunCat (MIPS)**
  yeast genome “people”, hierarchical, ~1994

- **Gene Ontology (GO)**
  *Drosophila* genome “people”, directed acyclic graph, ~2000

Notes:
- All these efforts are driven by people/communities/sociology
- Devising vocabulary/categories is not providing annotation

**SUBSTRATE - REACTION TYPE**

1. Enzymes given “ase” suffix
2. Substrate first, then reaction type

Example

\[
\begin{align*}
\text{CH}_3\text{CH}_2\text{OH} & \quad \text{CH}_3\text{CHO} \\
\uparrow & \quad \uparrow \\
\text{Substrate is alcohol} & \quad \text{Product is an aldehyde} \\
& \quad \text{Reaction removes hydrogen atoms}
\end{align*}
\]

**Enzyme:**
- **ALCOHOL DEHYDROGENASE**
ENZYME CLASSIFICATION

Systematic Nomenclature

1.0 OXIDOREDUCTASES
2.0 TRANSFERASES
3.0 HYDROLASES
4.0 LYASES
5.0 ISOMERASES
6.0 LIGASES

• Slide by Prof. Edward D Harris, Texas A&M (Lect23 Enz1.ppt)

SYSTEMATIC NOMENCLATURE

\[ \text{CH}_3\text{CH}_2\text{OH} + \text{NAD}^+ \rightleftharpoons \text{CH}_3\text{CHO} + \text{NADH} + \text{H}^+ \]

EC1.1.1.1. To see why, study the Table below.

1. Oxidoreductases (oxidation-reduction reactions of all types)
   1.1 acting on CH-OH group of substrates
   1.1.1 requires NAD⁺ or NADP⁺ as hydrogen acceptor
   1.1.1 specific substrate is ethyl alcohol

1.0 Oxidoreductases: (Add or remove electrons)

• Slide by Prof. Edward D Harris, Texas A&M (Lect23 Enz1.ppt)
2.0 Transferases

(Transfer group to substrate)

\[
\text{CH}_2\text{OH} \quad \text{HO-C-H} \quad \text{CH}_2\text{OH} + \text{ATP} \rightarrow \text{CH}_2\text{OH} \quad \text{HO-C-H} \quad \text{CH}_2\text{OPO}_3^- + \text{ADP}
\]

Glycerol

Glycerol phosphate

3.0 Hydrolases

(Cleave bonds with H}_2\text{O}

\[
\text{CH}_2\text{OH} \quad \text{HO-C-H} \quad \text{CH}_2\text{OPO}_3^- + \text{H}_2\text{O} \rightarrow \text{CH}_2\text{OH} \quad \text{HO-C-H} \quad \text{CH}_2\text{OH} + \text{HPO}_4^{2-}
\]

Glycerol phosphate

Glycerol

* Slide by Prof. Edward D Harris, Texas A&M (Lect23 Enz1.ppt)

4.0 LYASE (split C-X without water; reverse forms bond without a need for energy)

\[
\text{CH}_2\text{OPO}_3 \quad \text{HO-C-H} \quad \text{H-C-OH} \quad \text{H-C-OH} \quad \text{CH}_2\text{OPO}_3 \\
\text{split} \quad \text{unite} \\
\text{Dihydroxyacetone phosphate} \quad \text{Glyceraldehyde-3-phosphate}
\]

Fructose 1,6 bisphosphate

Split product of the forward or one substrate of the reverse reaction must have a double bond

* Slide by Prof. Edward D Harris, Texas A&M (Lect23 Enz1.ppt)
5.0 ISOMERASES (change groups around)

\[
\begin{align*}
\text{COOH} & \quad \leftrightarrow \\
H-\text{C}-\text{OH} & \quad \leftrightarrow \\
\text{CH}_2\text{PO}_4^- & \quad \leftrightarrow \\
\text{COOH} & \quad \leftrightarrow \\
H-\text{C}-\text{PO}_4^- & \quad \leftrightarrow \\
\text{CH}_2\text{OH} & \\
3\text{-phosphoglycerate} & \quad \leftrightarrow \\
2\text{-phosphoglycerate} & \\
\text{glucose 6-phosphate} & \quad \leftrightarrow \\
\text{fructose 6-phosphate} & \\
\end{align*}
\]

6.0 LIGASES (tie together... need energy)

\[
\begin{align*}
\text{CH}_3-\text{S-CoA} + \text{HCO}_3^- + \text{ATP} & \quad \rightarrow \\
& \quad \text{HOOC-CH}_2-\text{S-CoA} + \text{ADP} + \text{HPO}_4^- \\
& \text{acetyl-CoA} \quad \rightarrow \quad \text{malonyl-CoA} \\
\end{align*}
\]

- Synthetases: Use ATP
- Synthases: No ATP

**Caution:** Synthases could be Lyases
MIPS
(Munich Information Center for Protein Sequences)

- Maintains automatically generated and manually annotated genome specific databases
- Provides tools for comprehensive analysis of protein sequences and systematic schemes for the functional annotation of protein sequences
- MIPS Yeast data categorizes Budding Yeast proteins using 17 (+1) different functional categories at the top level
- FunCat has 26 (+2) categories in total (FunCat 2.1, 2007) http://mips.gsf.de/proj/functcatDB/
- Comprehensive Yeast Genome Database (CYGD) – http://mips.gsf.de/genre/proj/yeast/
  FTP Site: ftp://ftpmips.gsf.de/yeast/
For a protein of your choice, write a 2-3page paper in which you focus on its “function”.

• pick a protein (also specify from which organism)

• what functional classification/annotations can you find for it?

• besides the classification/annotation resources, what biologically relevant aspects can you extract from the literature

• find one thing that is non-standard/unusual/unexpected (to you) with respect to the protein’s function