

# Protein Function Prediction

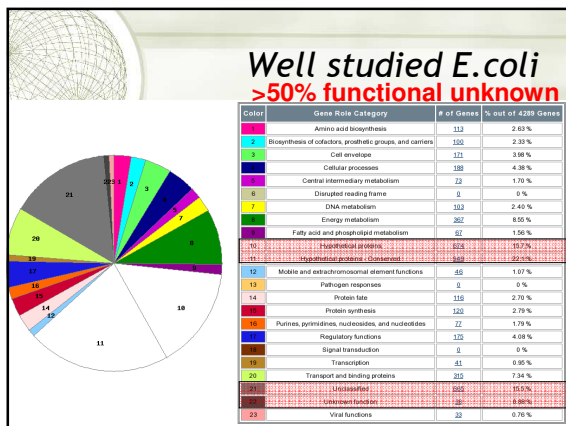
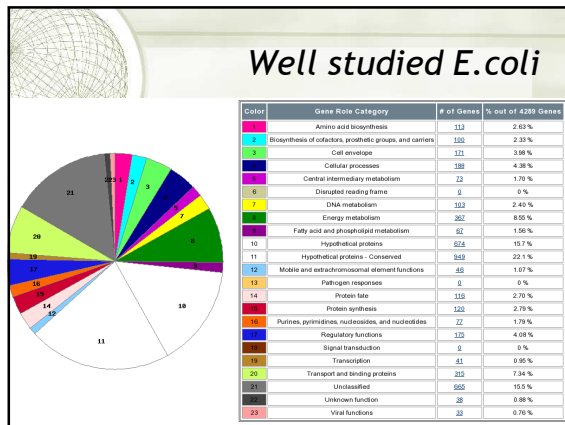
BIOL3004 electives

## What is function?

- “ Molecular function?
- “ Biochemical function?
- “ Cellular function?
- “ phenotypical function?
- “ all of it?

## Relevance of function prediction

“ In a post-genomic, post-transcriptomic, post-proteomic and post-structural-genomic era do we not know all function??



## How to reveal a protein's function?

- “ from sequence
  - “ homology to proteins with known function
  - “ from structure
    - “ similar structures ⇔ similar function?
- “ from genomic context (c.f. operons)
- “ from cellular context (cellular and sub-cellular)
  - “ localisation limits possible function
- “ from evolutionary context

## Function by homology

- “ strategy: Blast, copy and paste
- “ add “-like protein” if you feel like
- “ Problems
  - “ annotation errors in databases
  - “ inheritance of errors
  - “ “chinese whisper”
  - “ a single mutation may make a protein non-functional

## Function by homology

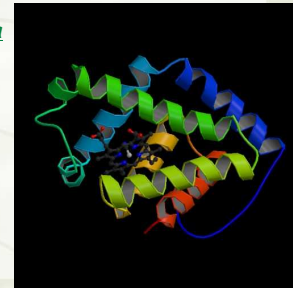
- “ strategy: motif search (e.g. Pfam)
- “ much better than Blast
- “ still relies on detectable sequence similarity
- “ look out for significance of the match!

## Function from structure

- “ function **is** determined by structure
- “ BUT structure **does not** determine function
  - “ paralogs
    - “ function may have changed after gene duplication
  - “ analogs
    - “ Some folds are promiscuous and hold many different functions
- “ Structure **and** sequence determines function!

## hemoglobin

- “ *Vitreoscilla stercoraria* (bacteria) versus *Petromyzon marinus* (eukaryote)
- “ same fold
- “ very similar structure
- “ 8% sequence ID
- “ heme group and HIS residues involved in binding are conserved

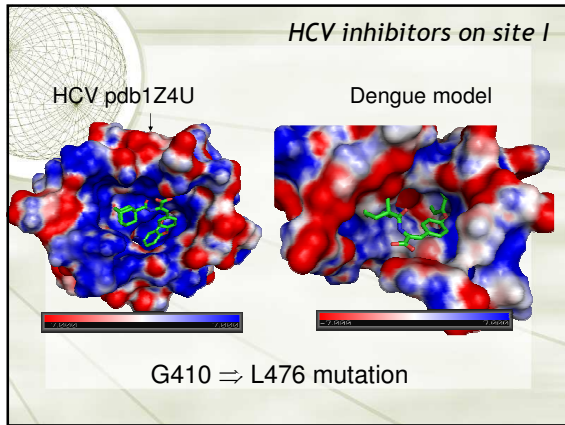


## Combining sequence and structure

- “ compare structures
  - “ how functional promiscuous is the structure?
- “ analyse sequence similarity of related structures to your query sequence
  - “ are functional important residues from proteins with known function conserved in your protein?
- “ extend the sequence analysis to complete family
  - “ are putative functional residues also conserved evolutionary?

## Another look at structure

- “ Biochemical function requires certain physical molecular properties. E.g.
  - “ pockets (increased surface) for binding
  - “ hydrophobic interactions
    - “ non-specific
  - “ charge interactions
    - “ specific
    - “ e.g. positive surface charge of DNA/RNA binding proteins



- ### Protein surfaces
- To highlight surface features
    - high quality visualisation for nice figures in your paper
  - You can calculate them within PyMOL
    - different surface properties (e.g. electrostatic surface)
      - both PyMOL and APBS is on the DVD

- ### Other data supporting function
- genomic context
    - bacterial protein
      - functional units (operons) are conserved
      - analyse functional commonalities of co-locating genes
    - eukaryotic proteins
      - functionally related proteins get often physically joint during evolution
      - look for fusion proteins of your target with other proteins

- ### Other data supporting function
- Protein-protein interactions
    - physical interaction suggest functional interaction
    - interaction networks of proteins (interactomes) are available for several model organisms
    - Data quality varies significantly
      - yeast two hybrid
      - bait tag purification
      - Interaction reports from literature

### Other data supporting function

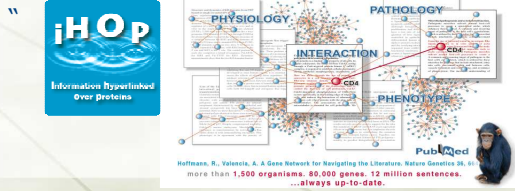
- sub-cellular context
  - Sub-cellular location of proteins can either be predicted or experimentally determined
  - both are available for mouse proteins through the LOCATE database

### cellular context

- cellular function (and to some extent molecular function) are tissue specific
  - for the mouse ortholog of your target there are tissue-specific transcriptional regulation data available through BioInfoWeb
  - microarray data is intrinsically noisy
    - potentially compare regulation data of other genes known to be involved in the putative function

## Literature context

- “ Chances are high that someone has worked on your target
- “ but publication may be hard to find because another name was used



## Summary

- “ Function prediction most accurate when evidence is cumulated
- “ Use holistic, hypothesis-driven approach and try to support (disprove) putative function (alternative functions)