# Bioinformatics Lunch Seminar (Summer 2014)

UNIVERSITY of WISCONSIN
UWMILWAUKEE
School of Freshwater Sciences

- Every other Friday at 1 pm. 20-30 minutes plus discussion
- Informal, ask questions anytime, start discussions
- Content will be based on feedback
- Targeted at broad audience of various levels of backgrounds and education
- Emphasis on Genomics Center

Contact:
Raymond Hovey
Genomics Center - SFS
rhovey@uwm.edu
414-382-1774
http://www.greatlakesgenomics.uwm.edu/

**Proteins with similar structures often have similar functions.**

**Protein structure is divided into these 4 levels:**

**1. Primary Structure- Polypeptide backbone**

**2. Secondary Structure- Local Hydrogen bonds along the backbone**

**3. Tertiary structure- Long distance bonding involving the AA side chains**

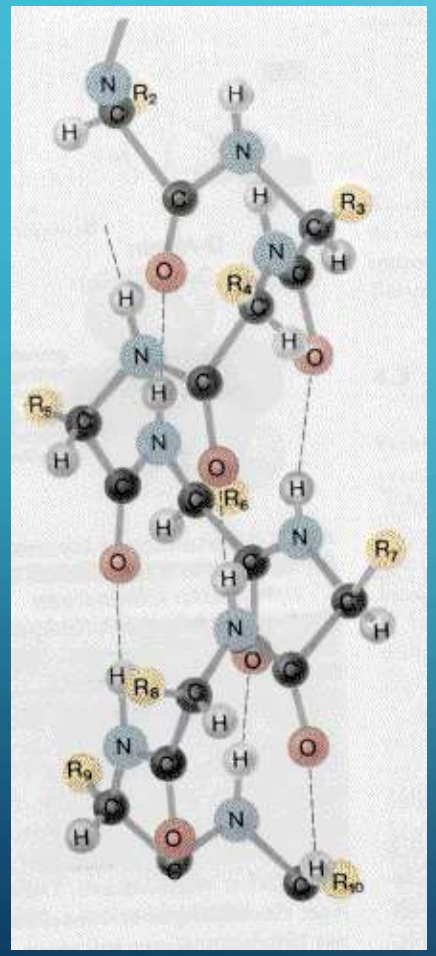
**4. Quaternary structure- Protein-Protein interactions leading to formation of dimers, tetramers, etc.**

**Motifs** and **domains** are combinations of secondary structures. Motifs only consist out of few secondary structures. They may but need not have a function. A domain is more complex. It is usually defined as a modular functional unit folding independently.

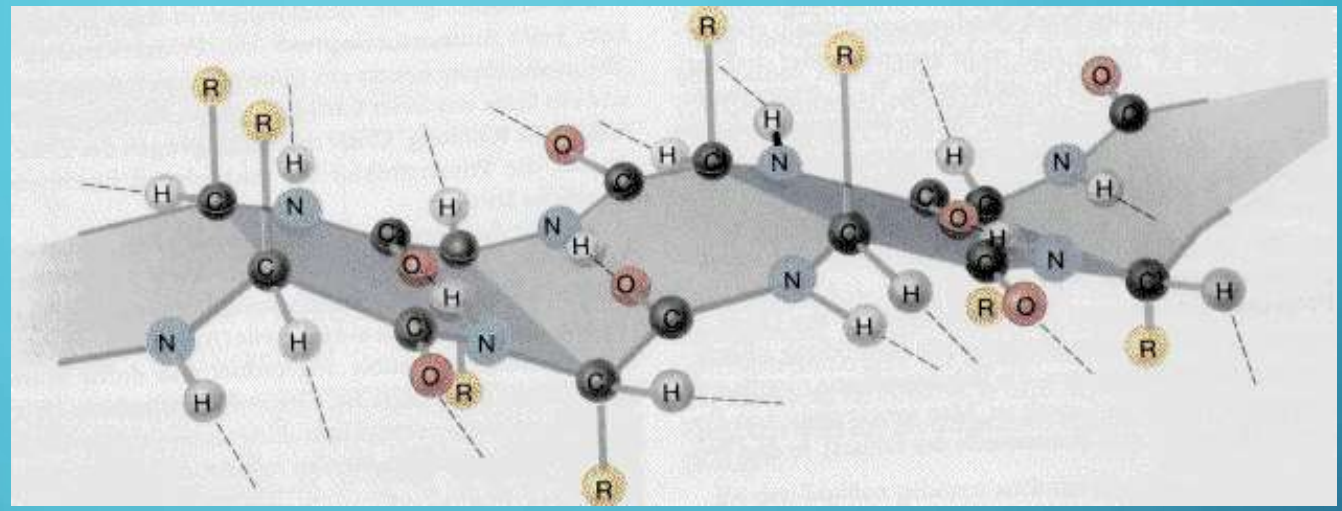
# Chapter 4.2: Three dimensional protein structure

Primary amino acid sequence determines the 3D structure of proteins

Secondary structure elements



α-helix



β-sheet

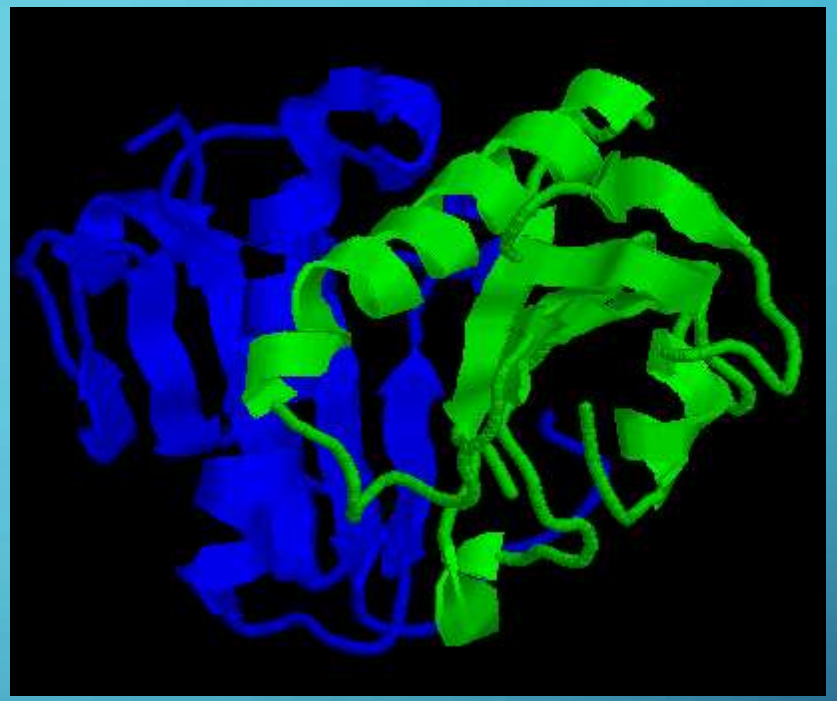
Major structures: H = α-helix
E = β-sheet (extended strands)

Minor structures: G = 3-helix
I = 5-helix
T = H-bonded turn
S = bend

Crambin (1CRN)

```
    TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIIPGATCPGDYAN
     EE SSHHHHHHHHHHHHTT  HHHHHHHHS EE SSS   TTS
```

Tertiary structure
(folding of  helices and β-sheets and loops)



Quaternary structure (two or more subunits)

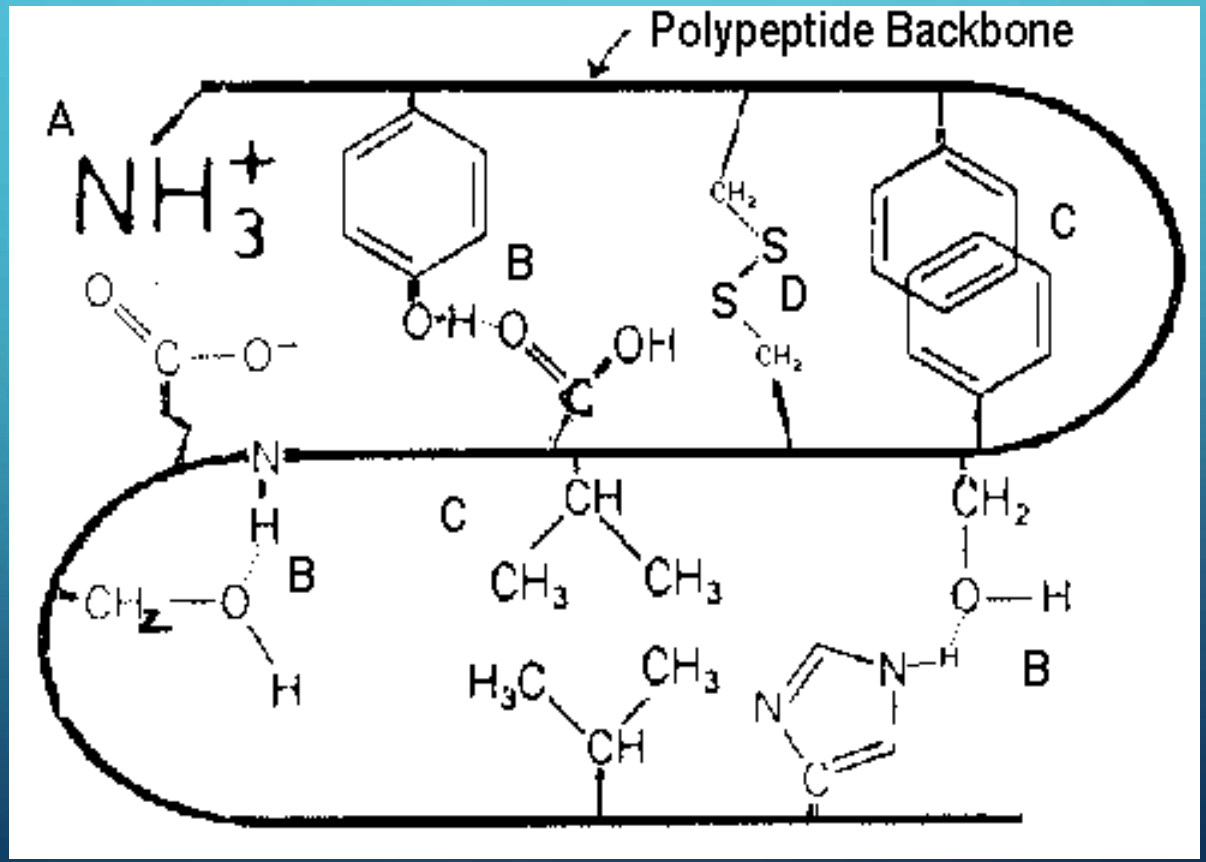
Tertiary structure is formed by long distance interactions of the amino acid side chains.
There are several types of bonds:

A) Ionic bonds

B) Hydrogen bonds

C) Hydrophobic interactions

D) Disulfide bonds -- weak covalent bond between 2 Cysteins (R-Cys-S-S-Cys-R)



**Hydrogen Bonds in Tertiary Structure**

Hydrogen bonds in tertiary structure involve polar non-charged amino acid side chains:

Alcohols -- Ser Thr Tyr

Amides --- Asn Gln

**Ionic bonds in Tertiary Structure**

Ionic bonds formed between polar, charged AA side chains:

Negative -- Glu Asp

Positive -- Lys Arg His

**Hydrophobic Interactions in Tertiary Structure**

Hydrophobic interactions involve the side chains of Non-Polar amino acids:

Hydrocarbon -- Ala Val Leu Ile Pro Met

Aromatics ----- Phe Trp (Tyr)

**Bonds in quaternary structure (between subunits):**

**Commonly found:**

- Ionic Bonds
- Hydrogen Bonds

**Less Commonly found:**

- Hydrophobic interfaces
- Interchain Disulfides

- Protein Structure is mostly stabilized by weak, non-covalent bonds. The energy difference between the denatured and native structure of proteins is small (important—many proteins change conformation during function; but also make them easy to unfold or even denature)

- This is one reason why the prediction of a protein structure is so difficult (no very low energy status)

- There are some structures in the secondary structure that can be predicted with certain probability

-But how the entire structure looks like is unpredictable– possibilities of interaction of amino acids are infinite

- Levinthal's paradox: a 100 aa protein has $3^{200}$ possible backbone configurations - many orders of magnitude beyond the capacity of the fastest computers

- Strategy –> explore the 3D structure of protein classes/ domains by X-ray defraction structure analysis and predict structure of other proteins by modeling

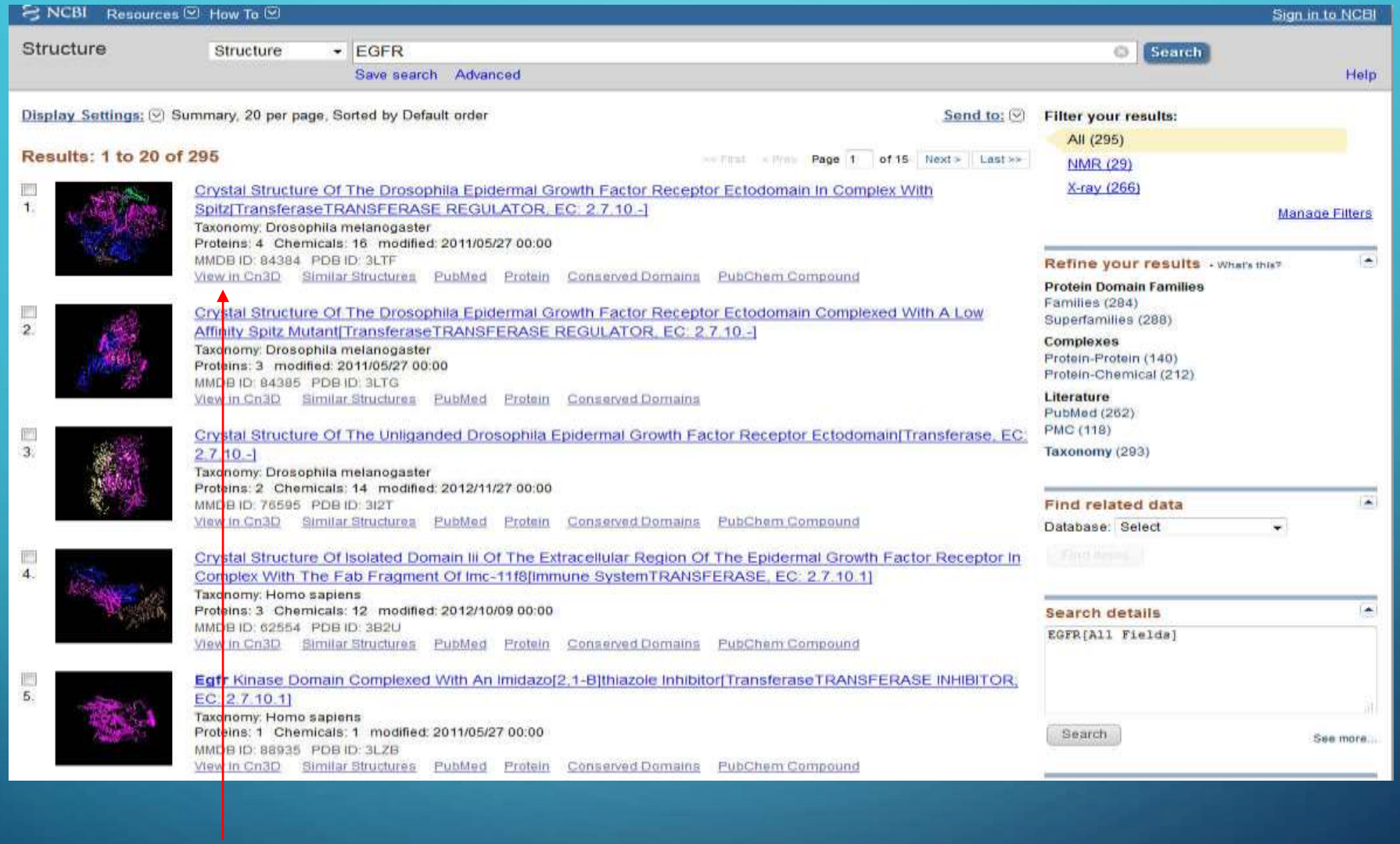
# How to find the 3D structure of a protein of interest


http://www.ncbi.nlm.nih.gov/Structure/MMDB/docs/mmdb_search.html

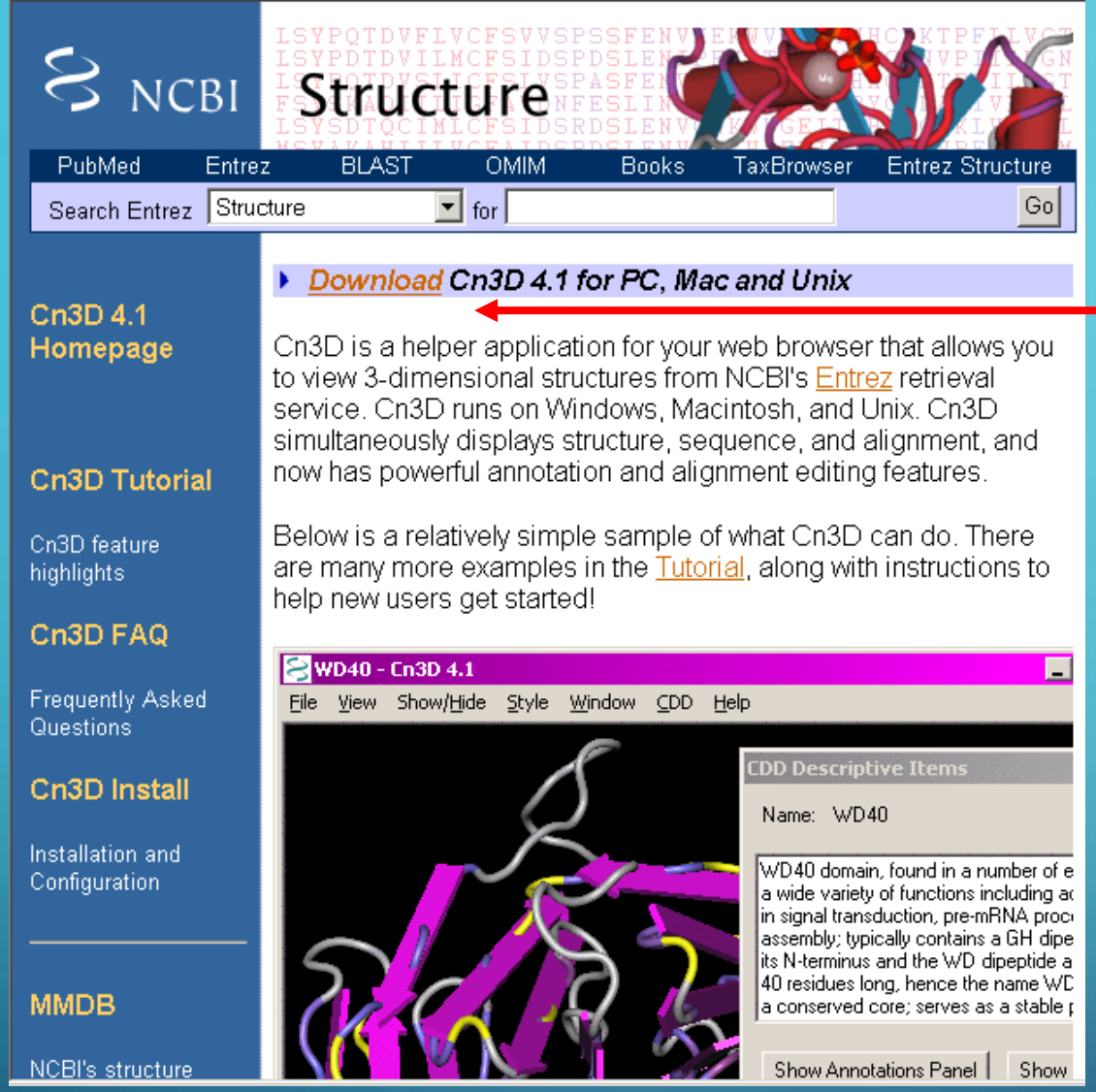
**Because 3D structure is strongly conserved in molecular evolution, one can often make valid predictions concerning structure and function by examining the structures of related proteins.**

**To find a structure for a protein or related proteins, use the following procedure:**

**Find the protein of interest using an accession number or keyword search in Entrez's structure database.**
**Use blastp if the sequence is new and not in Entrez.**



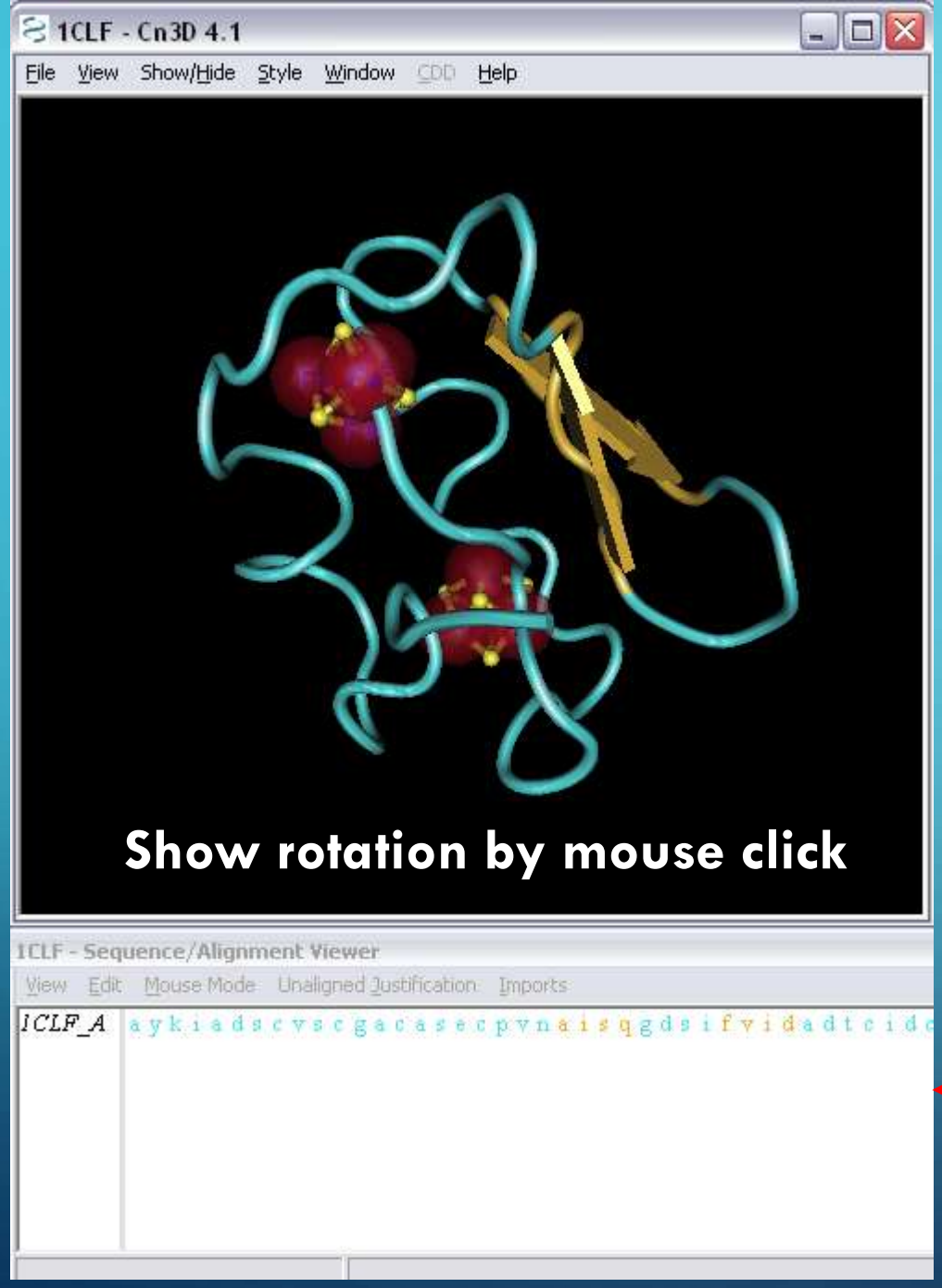

Click to open in Cn3D software




To visualize the 3D structure we need a helper program = Cn3D. Down-load it if it is not installed already

Cn3D is NCBI's 3D structure viewer. It allows you to interactively view 3D structures, sequences, and sequence alignments. Cn3D is available for Windows, MacOS, (and Unix until V4.1). _

http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml



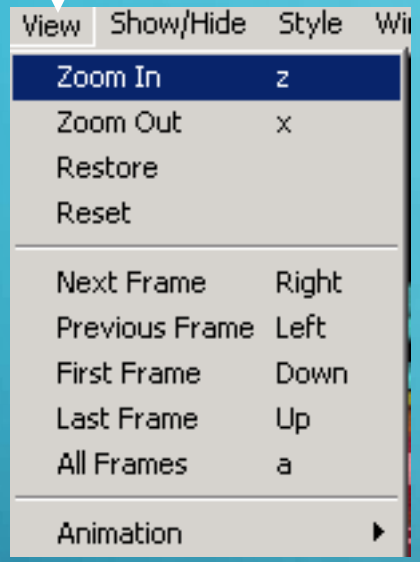

**Example:**
**Ferredoxin of Clostridium Pasteurianum with 2 oxidized [4Fe-4S] (MMDB-ID 55778)**

**Structure viewer**

**Sequence and alignment viewer**

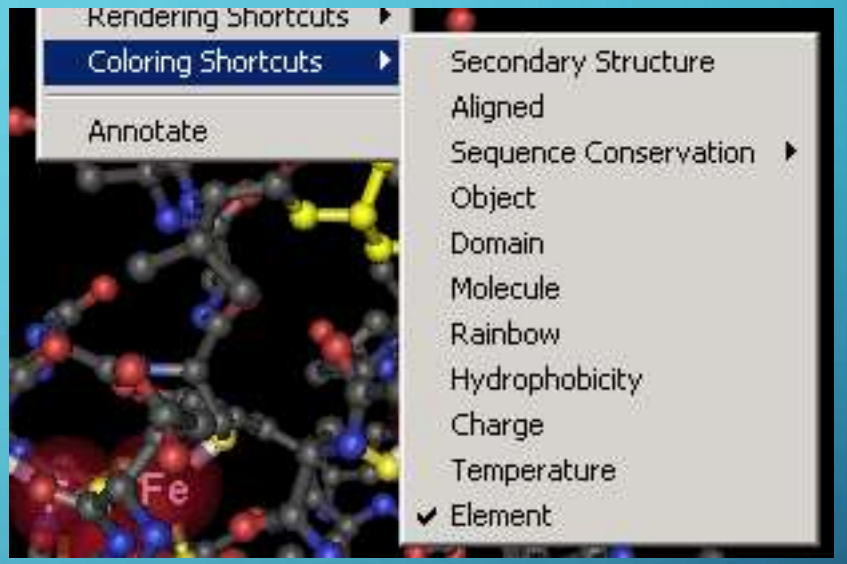
# Many opportunities to change 3D view of the protein

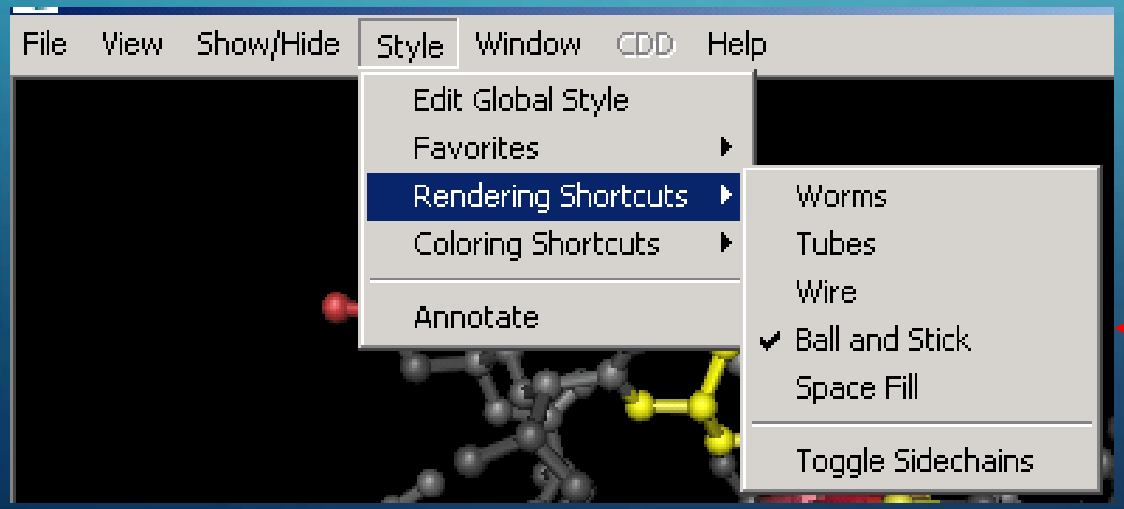
Tutorial at: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dtut.shtml



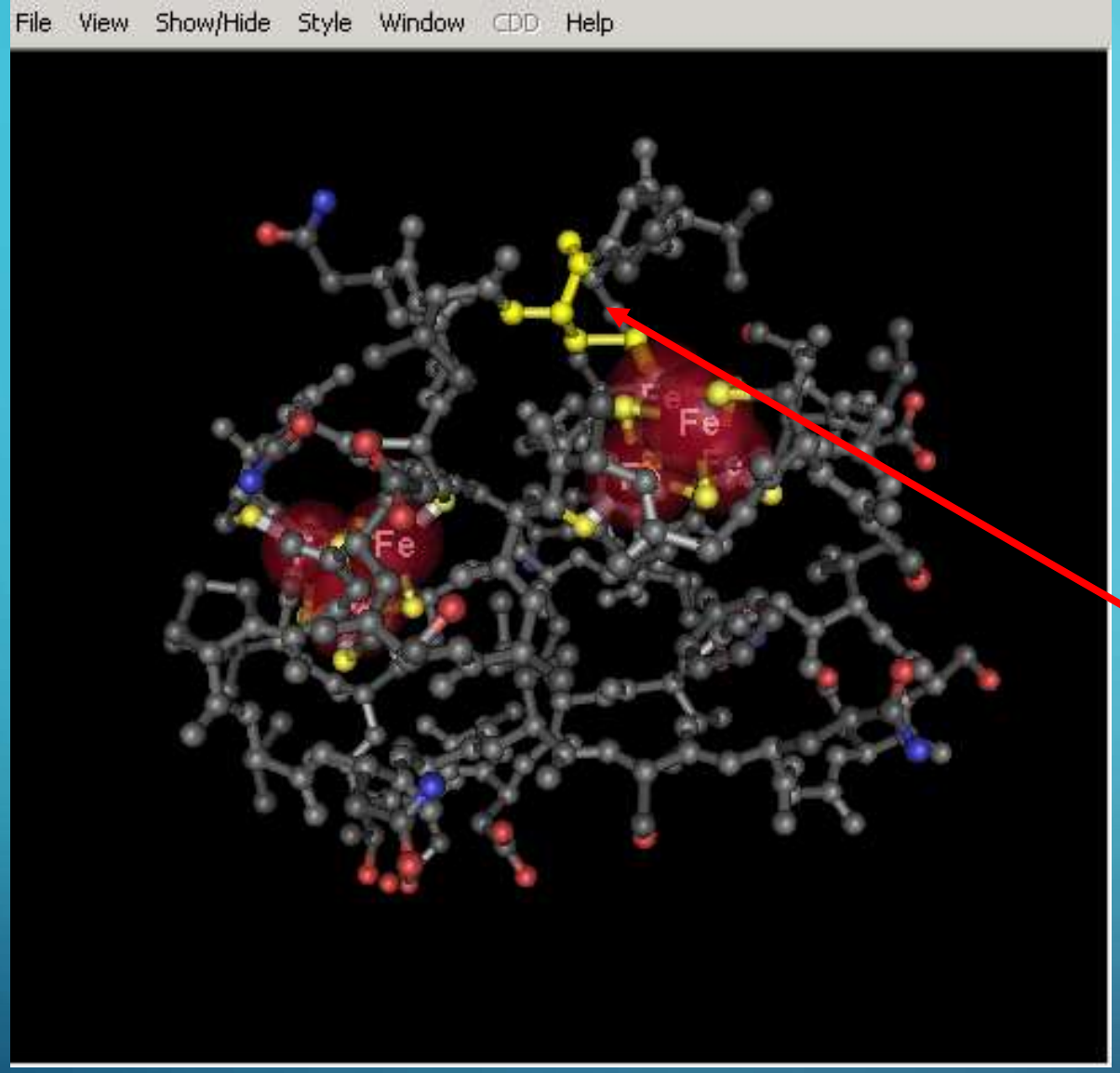

Zoom functions

Animated move of the protein



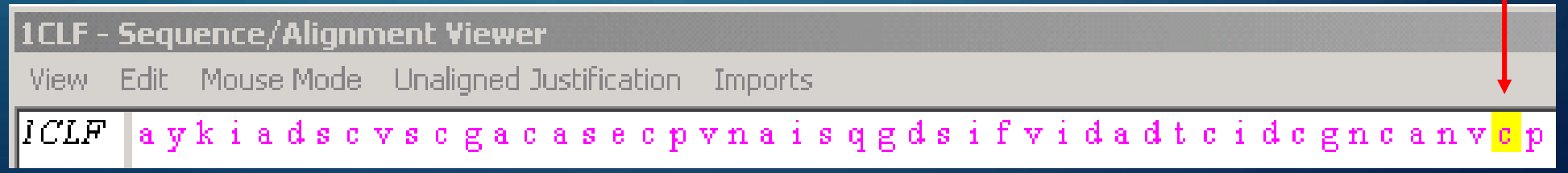

**The style panel contains detailed controls for all drawing styles, colors, and labels**




Enlarge by zoom in (view menu)

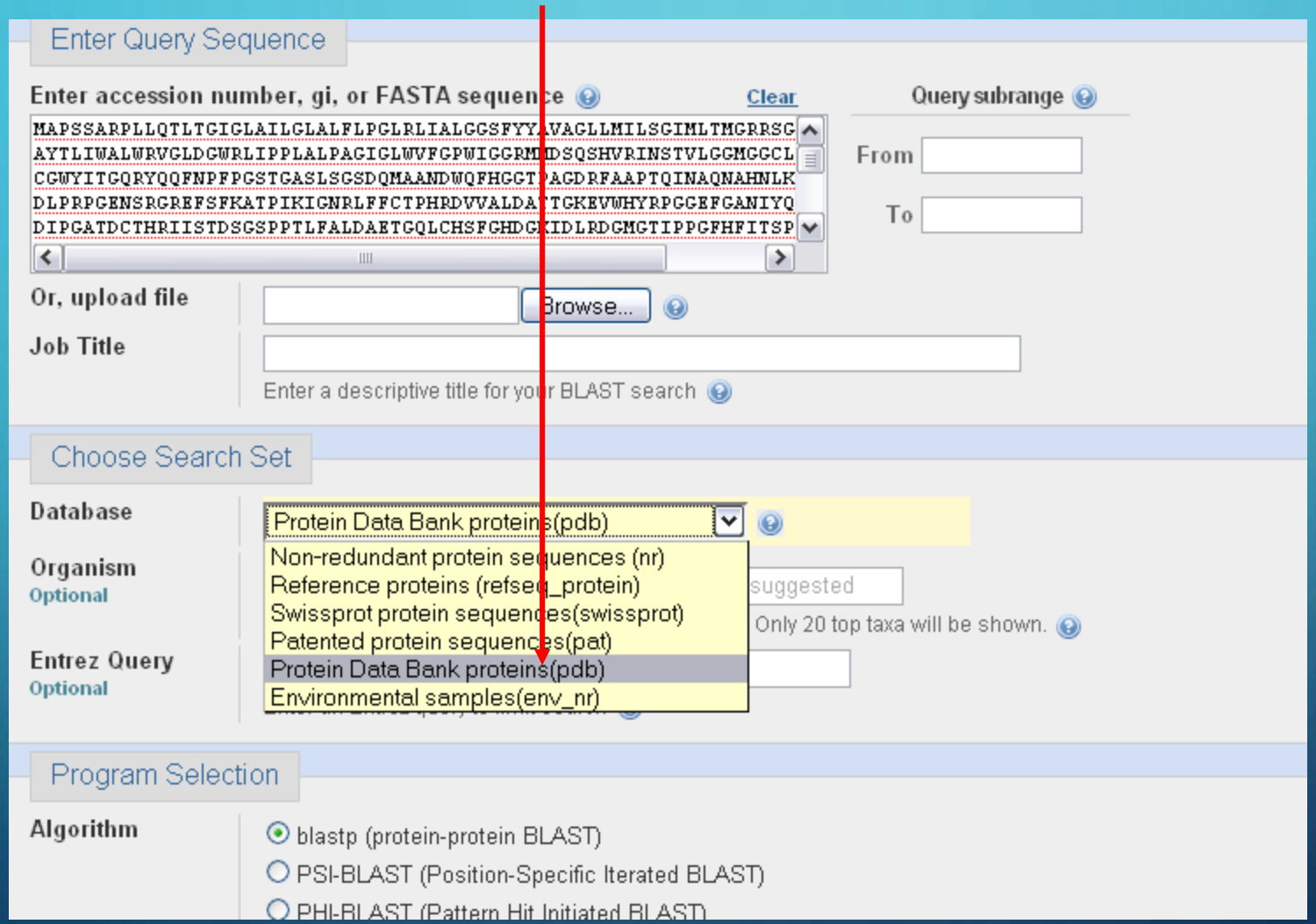
Highlight single amino acids (double click; one of the cysteine residues That coordinate the FeS cluster)

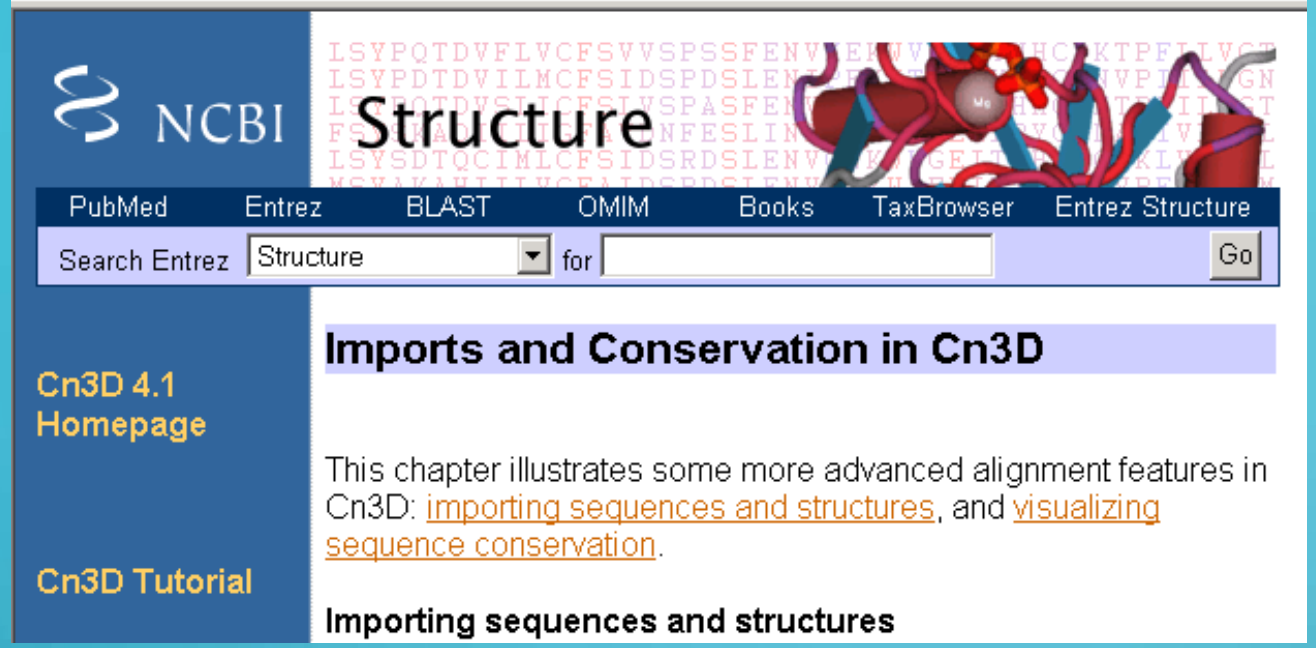
See amino acid in the sequence alignment viewer



# If you are working with an unknown protein do the following:

**Standard protein-protein BLAST [blastp]". Then select "pdb" in the "Database" menu (to search against known structures),**






Live Demonstration:

You have a gene encoding a hypothetical protein that shows homologies to a certain protein family.

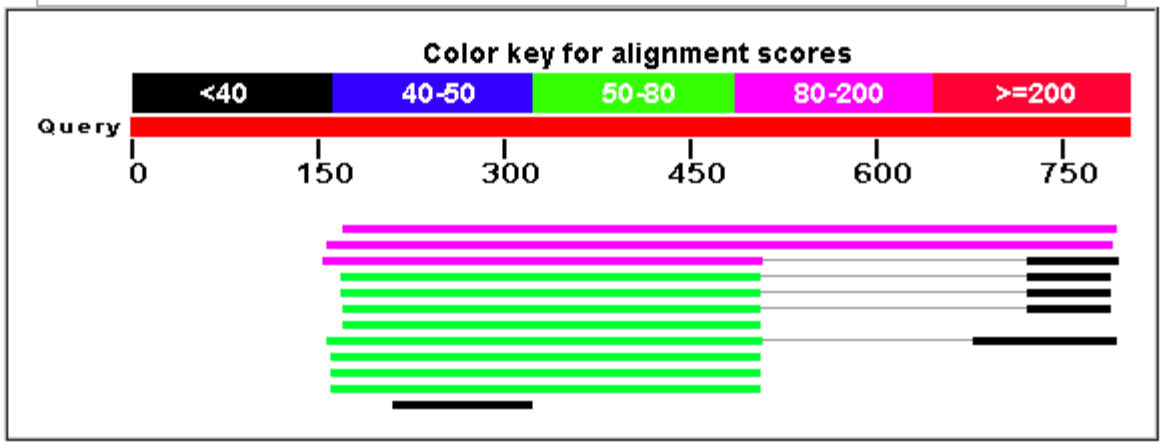
Check this assumption by comparison the predicted 3D-structure

Example is a gene from *Gluconobacter oxydans* that encodes a putative PQQ dependent dehydrogenase (GOX1441)

# Result of BLAST (with pdb database, display may differ depending on your browser)

## Distribution of 17 Blast Hits on the Query Sequence

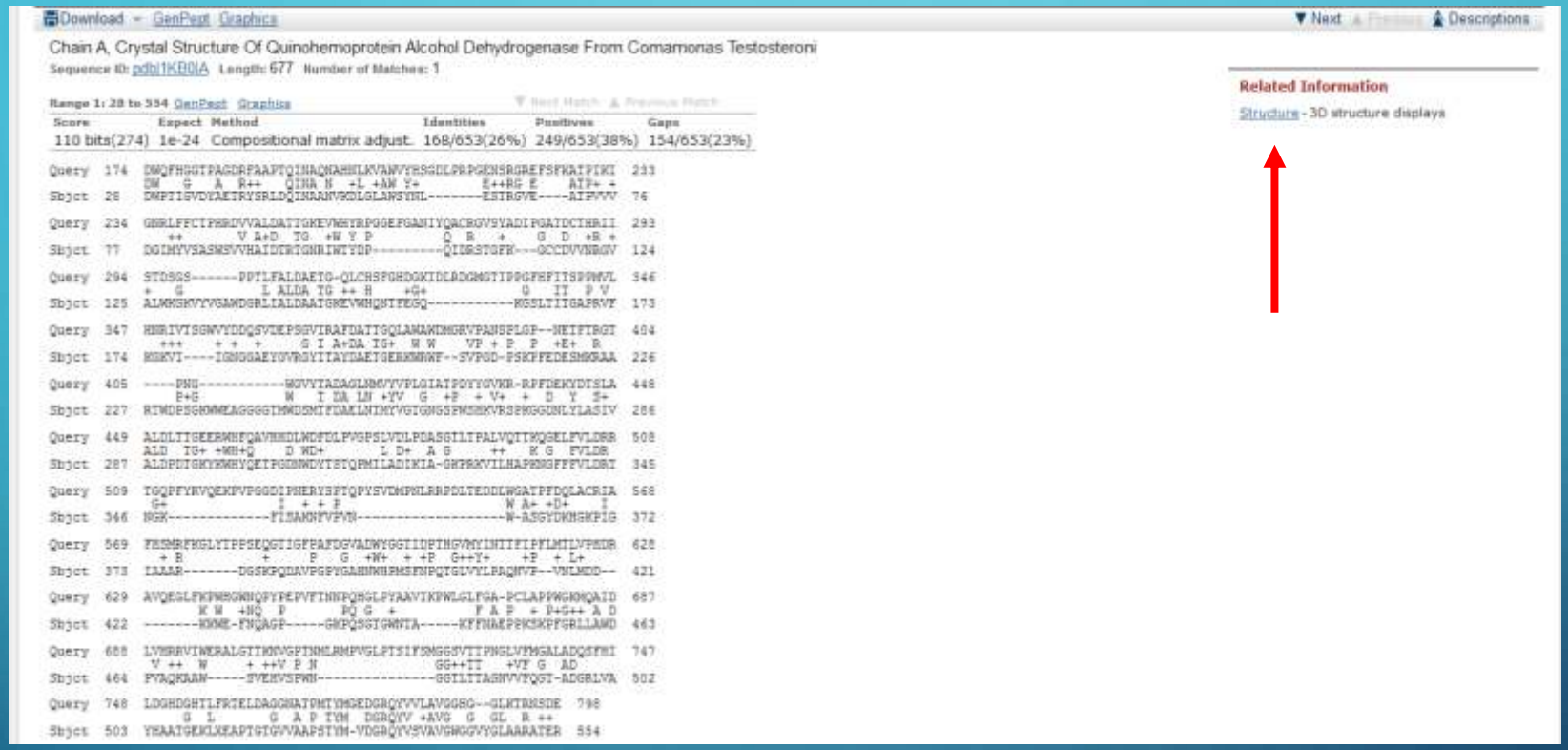
Mouse-over to show defline and scores, click to show alignments



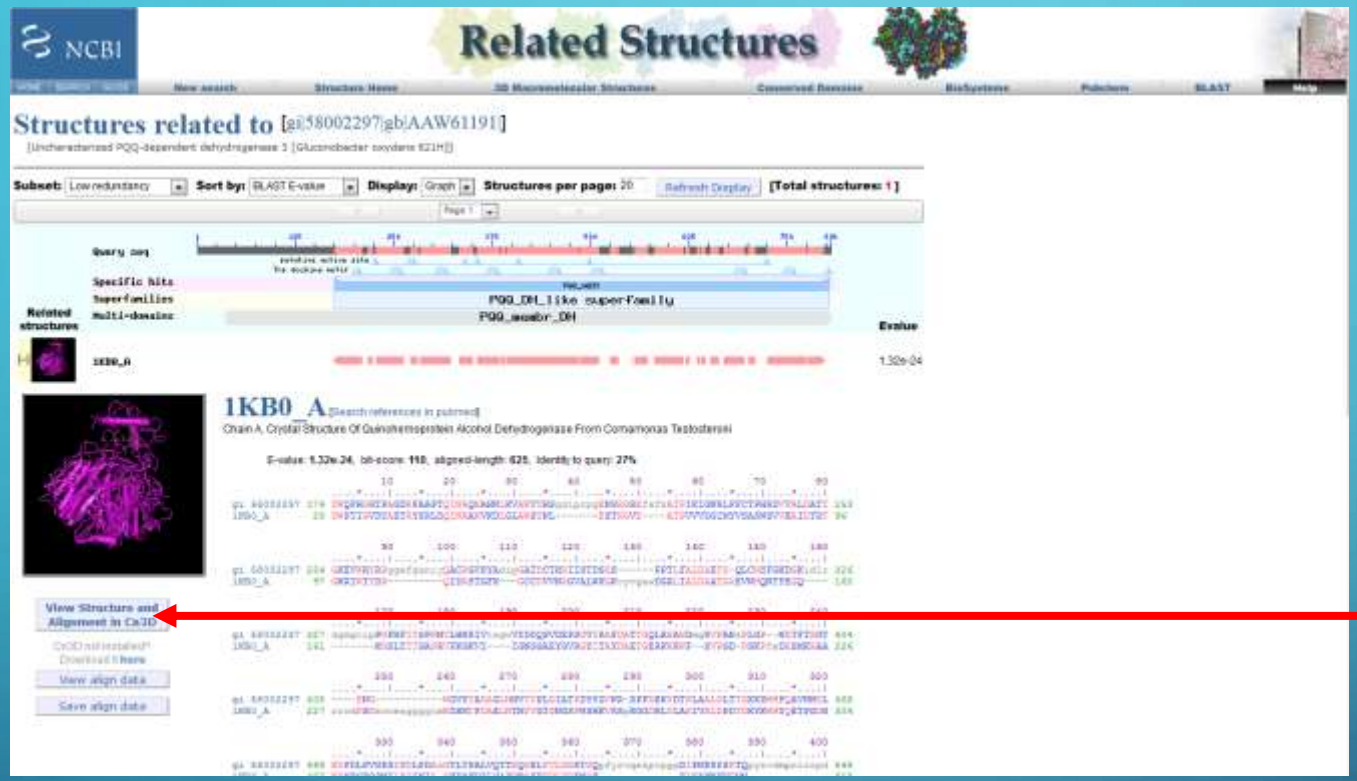

Distance tree of results NEW  Related Structures

| Sequences producing significant alignments: | | Score (Bits) | E Value | |
|---|---|---|---|---|
| pdb\|1KB0\|A | Chain A, Crystal Structure Of Quinohemoprotein Alc... | 102 | 3e-22 | S |
| pdb\|1YIQ\|A | Chain A, Molecular Cloning And Structural Analysis... | 95.9 | 4e-20 | S |
| pdb\|1KV9\|A | Chain A, Structure At 1.9 A Resolution Of A Quinoh... | 88.2 | 9e-18 | S |
| pdb\|4AAH\|A | Chain A, Methanol Dehydrogenase From Methylophilus... | 77.0 | 2e-14 | S |
| pdb\|1G72\|A | Chain A, Catalytic Mechanism Of Quinoprotein Metha... | 77.0 | 2e-14 | S |
| pdb\|2D0V\|A | Chain A, Crystal Structure Of Methanol Dehydrogena... | 74.3 | 1e-13 | S |
| pdb\|1LRW\|A | Chain A, Crystal Structure Of Methanol Dehydrogena... | 69.7 | 3e-12 | S |

**Select the one at the top with the best score and E-value-- leads to the alignment, from which the "Structure" link on the right leads ultimately to the 3d file.**

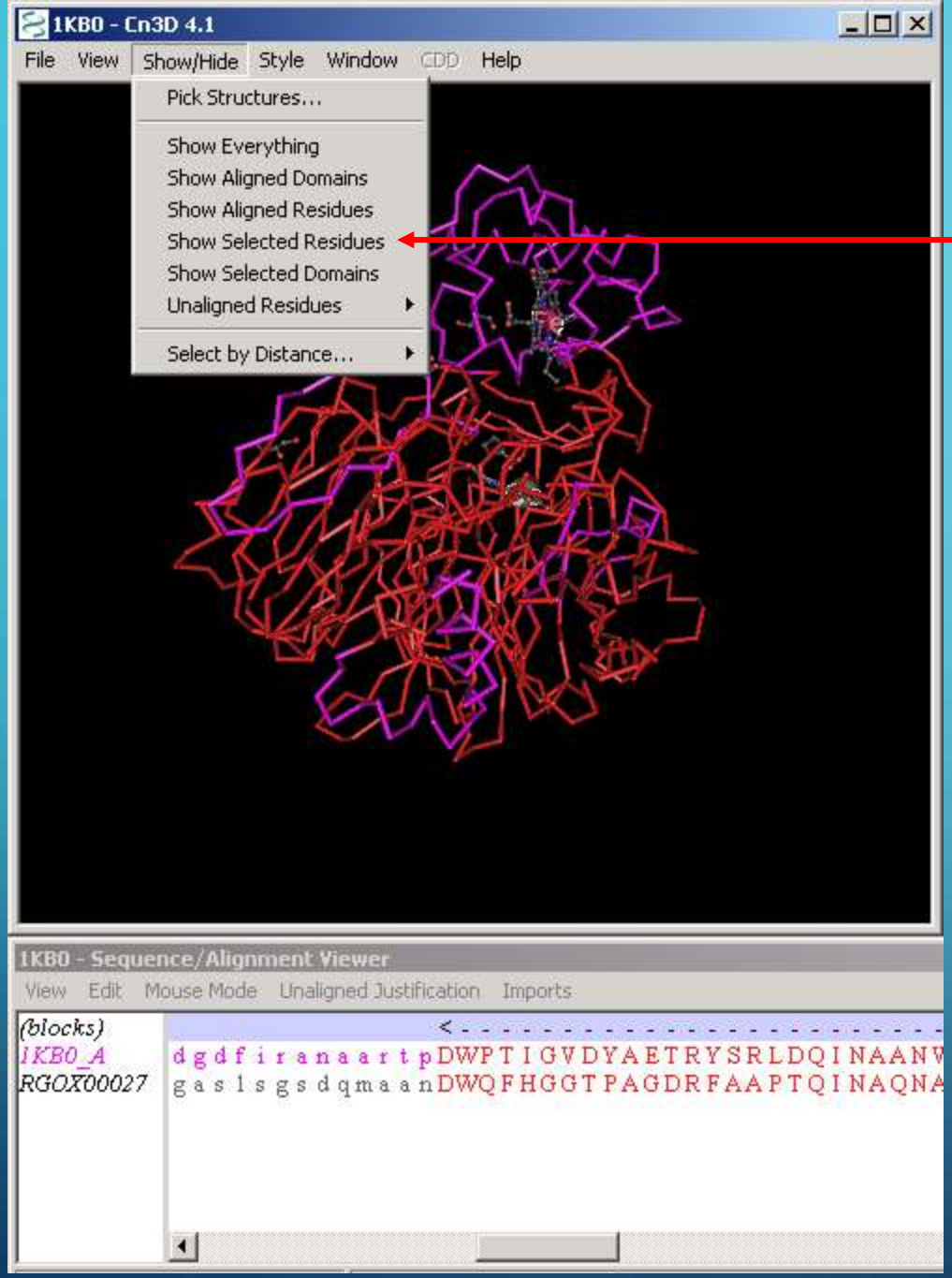


In our case the sequence of the conserved hypothetical protein showed homologies the a PQQ-dependent ethanol dehydrogenase with known 3D structure from Comamonas testosteroni



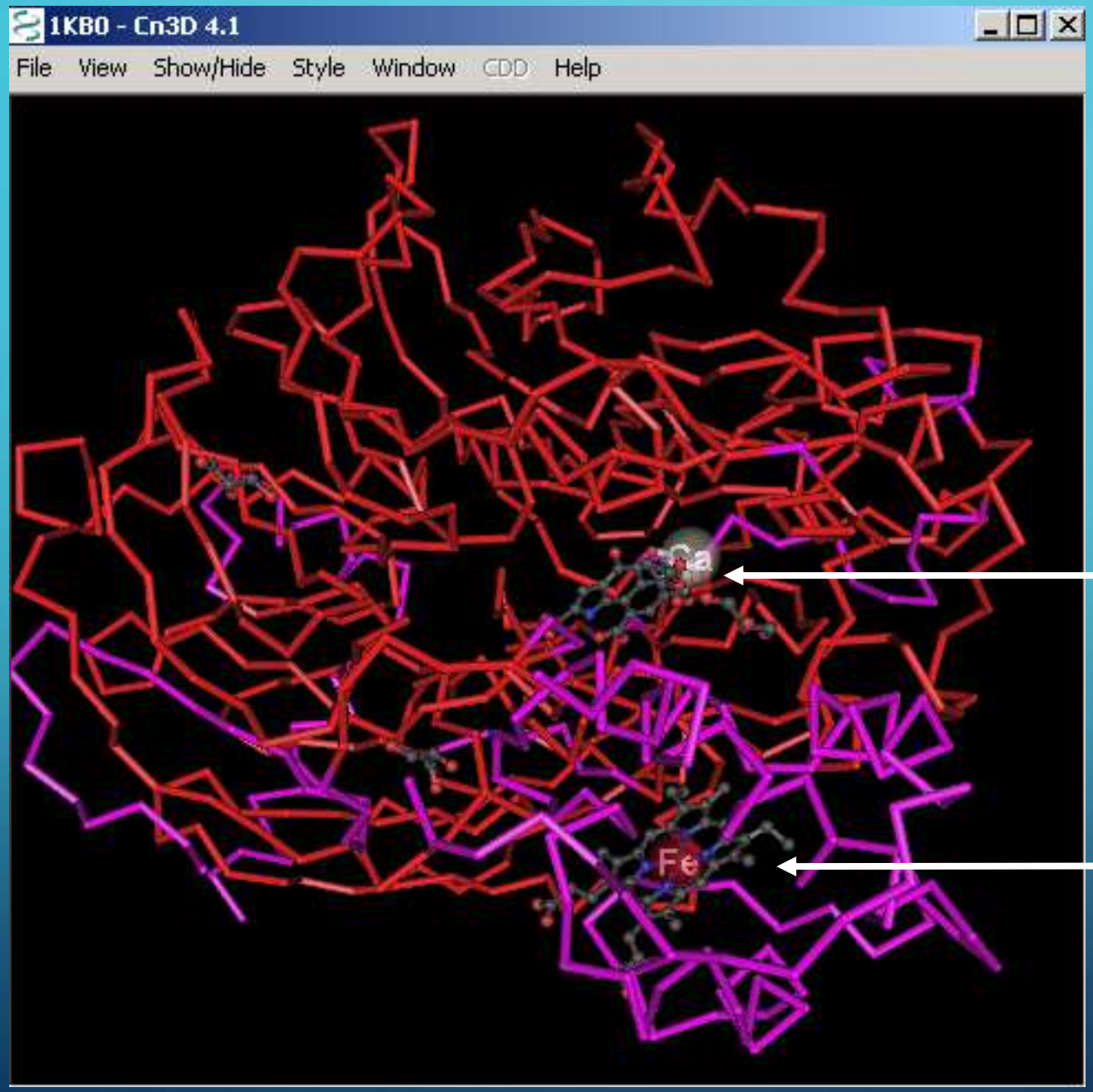

click

**This is a very powerful general method for finding structures whose proteins are related to the query closely enough so that structural properties can be inferred by homology.**




Many opportunities to modify the alignment

**Aligned domains are shown in red and blue, unaligned in gray. Colors are consistent in sequence and structure windows.**




Style-> Coloring Shortcuts -> Aligned

Now aligned is red
and unaligned purple.

PQQ and active center well aligned. So our protein is a PQQ dependent dehydrogenase

It does not contain a heme group and the corresponding heme-binding domain

# Bioinformatics Lunch Seminar (Summer 2014)

UNIVERSITY of WISCONSIN
UWMILWAUKEE
School of Freshwater Sciences

- Every other Friday at 1 pm. 20-30 minutes plus discussion
- Informal, ask questions anytime, start discussions
- Content will be based on feedback
- Targeted at broad audience of various levels of backgrounds and education
- Emphasis on Genomics Center

Contact:
Raymond Hovey
Genomics Center - SFS
rhovey@uwm.edu
414-382-1774
http://www.greatlakesgenomics.uwm.edu/