Crissan Harris

University of California Santa Cruz

Undergraduate Researcher

Predicting Protein Structure
CASP7

Lead Researcher:

Professor Kevin Karplus

UCSC Bioinformatics Dept.

What is CASP?

- CASP is a community wide experiment where leading researchers in the field of bioinformatics use computational methods to predict the three-dimensional structure of a protein given only the amino-acid sequence.
  - The structures of these target proteins have been solved experimentally but the results are kept unreleased until after CASP is over.
- The main goal of CASP is to obtain an in-depth and objective assessment of the communities current abilities and inabilities in the area of protein structure prediction.

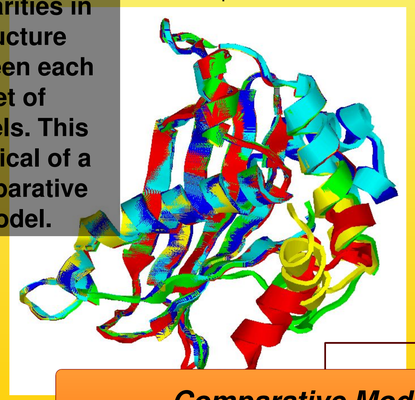
T0379

Each color indicates the location in the chain of top models.

Notice the similarities in structure between each set of models. This is typical of a Comparative Model.

T0364

Each Color indicates a different top model.



Why do we care about predicting protein structures?

- Proteins are the building blocks of living cells.
- Their folded 3-D shape is different for different proteins and is essential to the function of the protein and the cell.
- Knowledge of the structure of proteins is used in:
  - drug design
  - design of synthetic proteins
  - re-engineering of defective proteins
- Because protein structures are expensive to determine experimentally (both in dollars and in time), the availability of a computational method of determination has become a necessity.

Using the sequence of amino acids and a library of proteins with known 3-D structures, know as the "template library", we use our tools to align the parts of the proteins that are similar. This results in one of three cases.

What are the tools used in predictions?

- Undertaker
  - a fragment-packing program
  - prediction of tertiary (3-D) structure using conformation generation and scoring
- SAM
  - the premier suite of hidden Markov Model tools
  - HMM, database query, secondary structure predictions and more
  - remote homology detection and sequence alignments
- DSSP-EHL2
  - uses neural nets to predict local secondary structure (i.e. helix, sheet, etc.) using residue properties that are conserved through evolution
- ProteinShop
  - interactive tool for protein manipulation

Comparative Modeling

~The majority of the Protein can be modeled from existing templates.

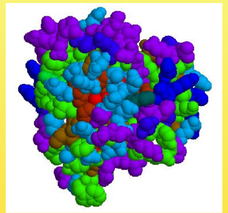
Fold Recognition

~A portion of the Protein can be modeled from existing templates.

Ab-Initio or New Fold

~There is not enough information in the templates to determine the structure.

T0342

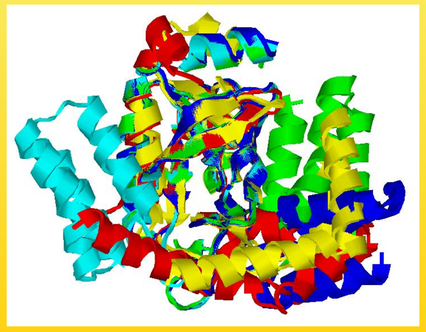


This model is shown in space fill with burial coloring.

The spectrum goes from Purple to Red (exposed to buried).

Notice that the inner barrel core remains the same, but the surrounding helices are different in each model. This is typical of a Fold Recognition model.

All of the top models are shown together here. Each color represents a different model.



One of the models colored by the location in the chain

One of the models shown in DSSP-EHL2 coloring

Pink = Helix

Yellow = Sheet

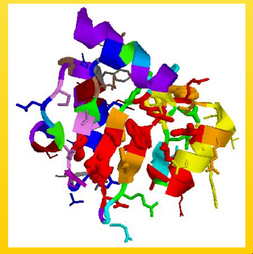
Grey = Coil

T0358

Notice That none of the top models look the same. Each represents a different fold possibility. This is typical of a New Fold

Each color indicates a different top model.

Here, this model is displayed using a residue-residue script which shows amino acids that are predicted to be close to each other (shown by matching colors). The greater the thickness of the residue, the stronger the prediction. This can help in determining a new fold.



Background graphic courtesy of Chromosomal Laboratories Inc.

Summer Undergraduate Research Fellowship – Information Technology

Sponsored by the NSF