Secondary Structure Prediction Software

Figure 11.3 Comparison of secondary structure predictions by various methods. The sequence of flavodoxin, an α/β protein, was used as the query and is shown on the first line of the alignment. For each prediction, H denotes an α helix, E a β strand, T a β turn; all other positions are assumed to be random coil. Correctly assigned residues are shown in inverse type. The methods used are listed along the left side of the alignment and are described in the text. At the bottom of the figure is the secondary structure assignment given in the PDB file for flavodoxin (1OFV, Smith et al., 1983).

Secondary Structure Prediction

- [NN based] PSI-pred, nnPredict (2-layer, feed-forward NN), Pred2ary
- [Consensus Approach] JPRED, SOPMA
- [K-nearest neighbor] NNSSP, PREDATOR
- [HMM] PSA
- ZPRED
- SSP
- PHD (See Sample)

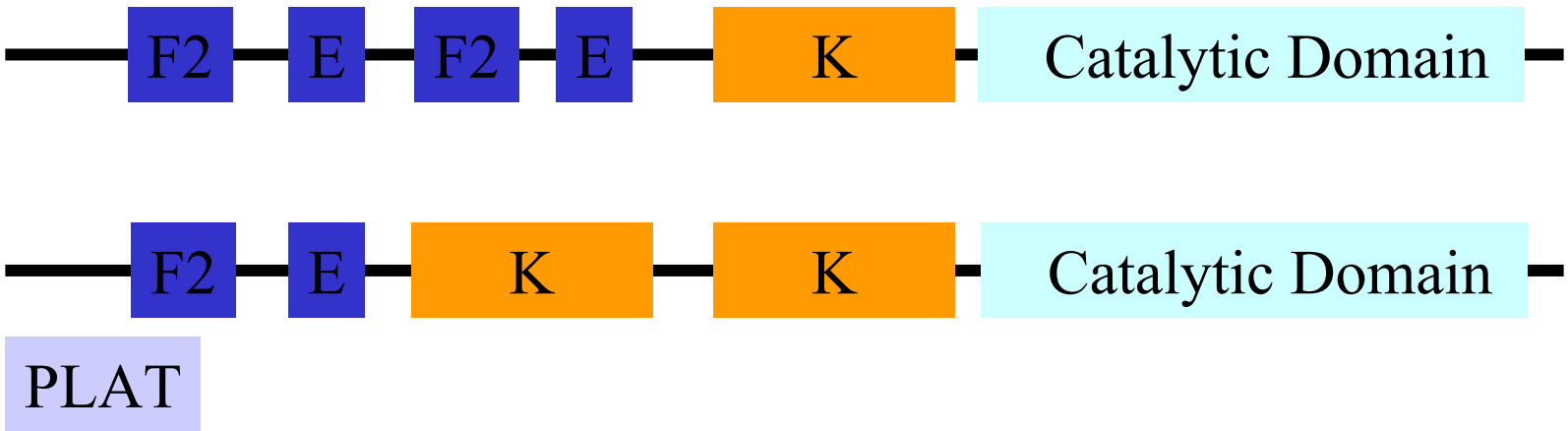
Motif Detection Tools

- PROSITE (Database of protein families & domains)
  - Try PDOC00040. Also Try PS00041
- BLOCKS (multiply aligned ungapped segments for highly conserved regions of proteins; automatically created) Sample Output
- PRINTS Sample Output
- Pfam (Protein families database of alignments & HMMs)
  - Multiple Alignment, domain architectures, species distribution, links: Try
- PIR Sample Protein page
- MoST
- PROBE
- ProDom
- DIP

Modular Nature of Proteins

- Proteins are collections of “modular” domains. For example,

Coagulation Factor XII



Domain Architecture Tools

- CDART
  - Protein AAH24495; Domain Architecture;
  - It's domain relatives;
  - Multiple alignment for 2nd domain
- SMART

Predicting Specialized Structures

- COILS – Predicts coiled coil motifs
- TMPred – predicts transmembrane regions
- SignalP – predicts signal peptides
- SEG – predicts nonglobular regions

Tertiary & Quaternary Protein Structures

- Experimental methods
  - X-ray crystallography [More accurate!]
  - Nuclear Magnetic Resonance Spectroscopy (NMR)
- If protein "unfolded" (denatured) and "released", then it goes back to its native 3-d structure.
- The tertiary structure is a structure of minimum energy.
- Angles ϕ and ψ are constrained.
- Proteins structures often have hydrophobic core.

Protein Folding

Unfolded

Rapid (< 1s)

Molten Globule State

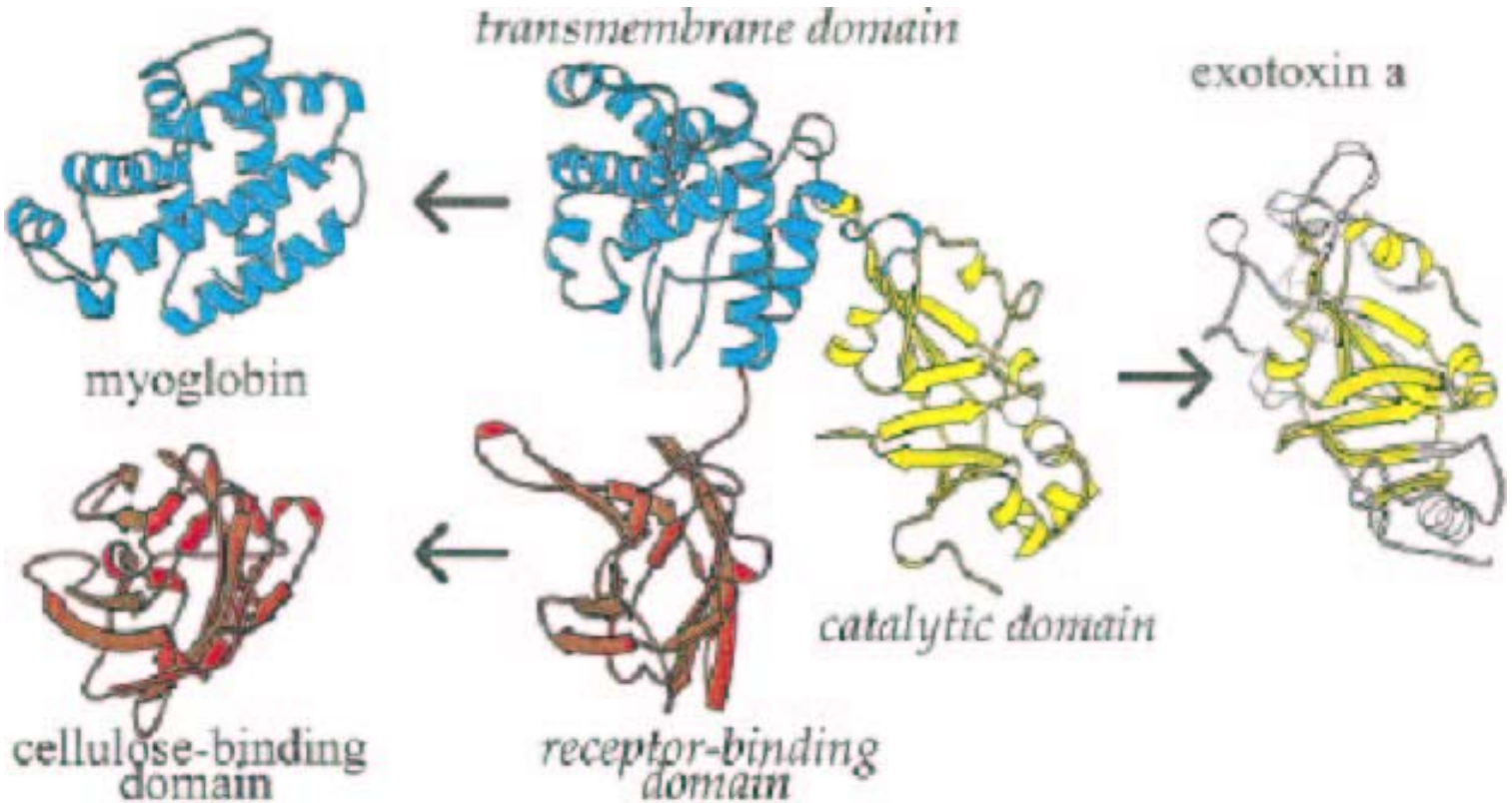
Slow (1 – 1000 s)

Folded Native State

- How to find minimum energy configuration?

Modular Nature of Protein Structures

Example: Diphtheria Toxin



Structural Classification of Proteins

- SCOP (Structural Classification of Proteins)
  - Based on structurla & evolutionary relationships.
  - Contains ~ 40,000 domains
  - Classes (groups of folds), Folds (proteins sharing folds), Families (proteins related by function/evolution), Superfamilies (distantly related proteins)

SCOP Family View

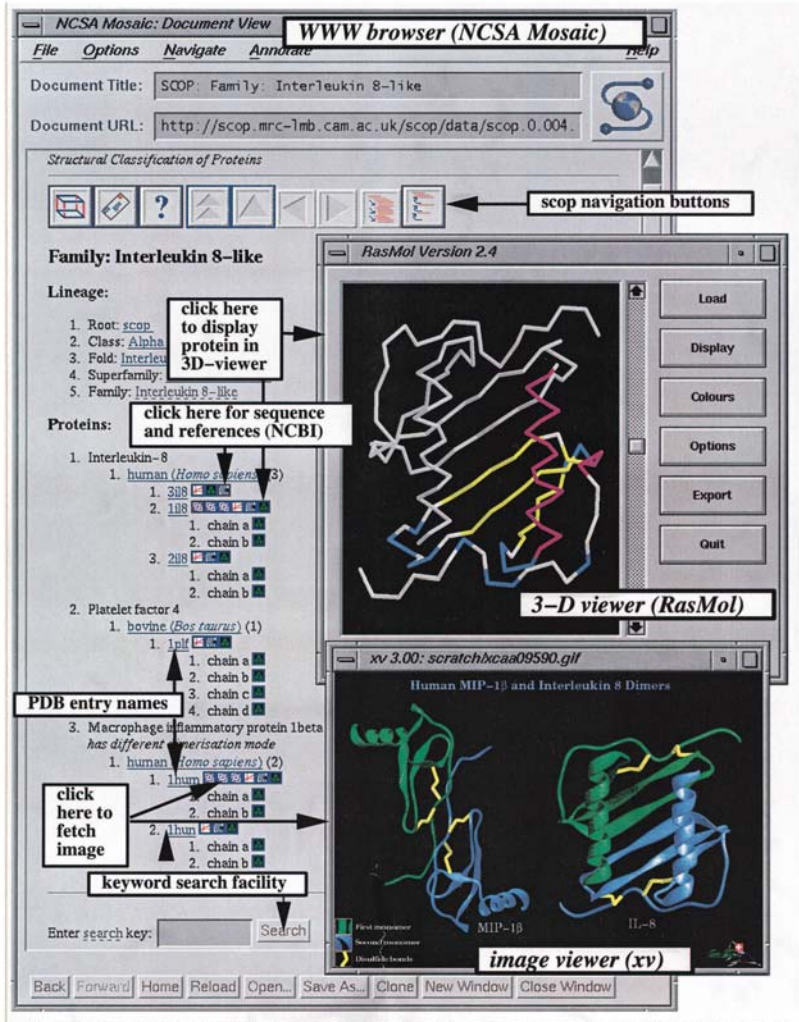


Figure 2. A typical scop session is shown on a unix workstation. A scop page, of the Interleukin 8-like family, is displayed by the *WWW browser program (NCSA Mosaic)* (Schatz & Hardin, 1994). Navigating through the tree structure is accomplished by selecting any underlined entry, by clicking on buttons (at the top of each page) and by keyword searching (at the bottom of each page). The static image comparing two proteins in this family was downloaded by clicking on the icon indicated and is displayed by image-viewer program *xv*. By clicking on one of the green icons, commands were sent to a molecular viewer program (*RasMol*) written by Roger Sayle (Sayle, 1994), instructing it to automatically display the relevant PDB file and colour the domain in question by secondary structure. Since sending large PDB files over the network can be slow, this feature of scop can be configured to use local copies of PDB files if they are available. Equivalent WWW browsers, image-display programs and molecular viewers are also available free for Windows-PC and Macintosh platforms.

CATH: Protein Structure Classification

- Semi-automatic classification; ~36K domains
- 4 levels of classification:
  - Class (C), depends on sec. Str. Content
  - Architecture (A), orientation of sec. Str.
  - Topolgy (T), topological connections &
  - Homologous Superfamily (H), similar str and functions.

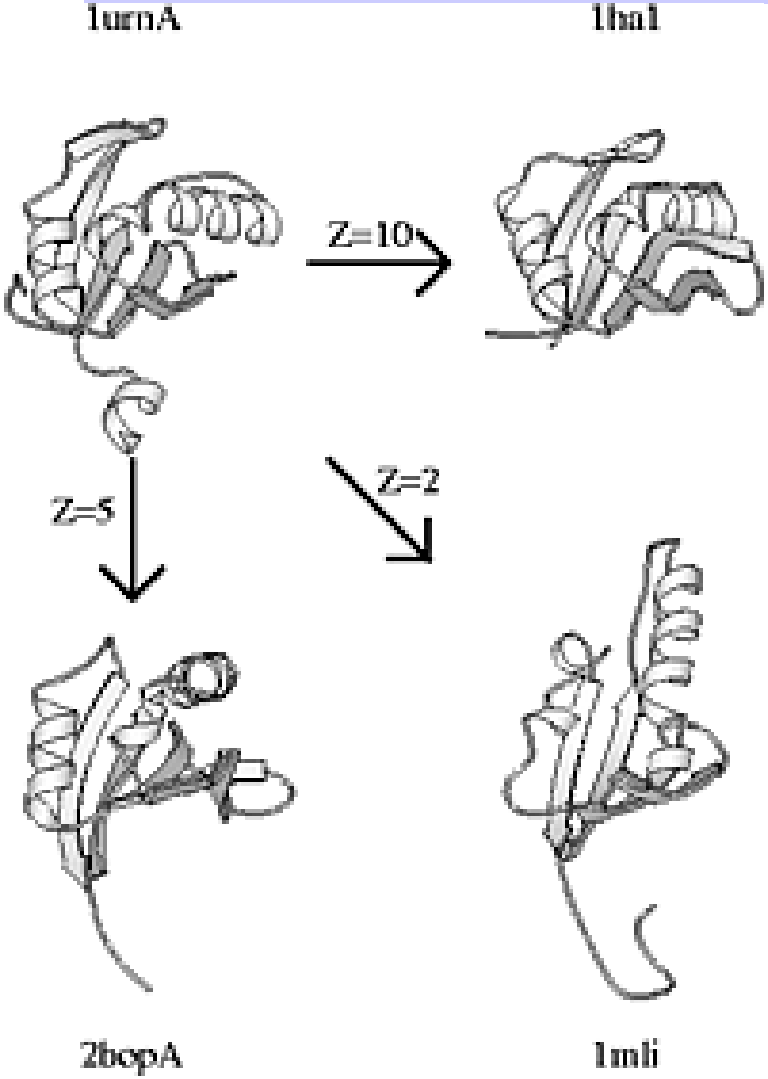
DALI Domain Dictionary

- Completely automated; 3724 domains
- Criteria of compactness & recurrence
- Each domain is assigned a Domain Classification number DC_l_m_n_p representing fold space attractor region (l), globular folding topology (m), functional family (n) and sequence family (p).

5 Fold Space classes

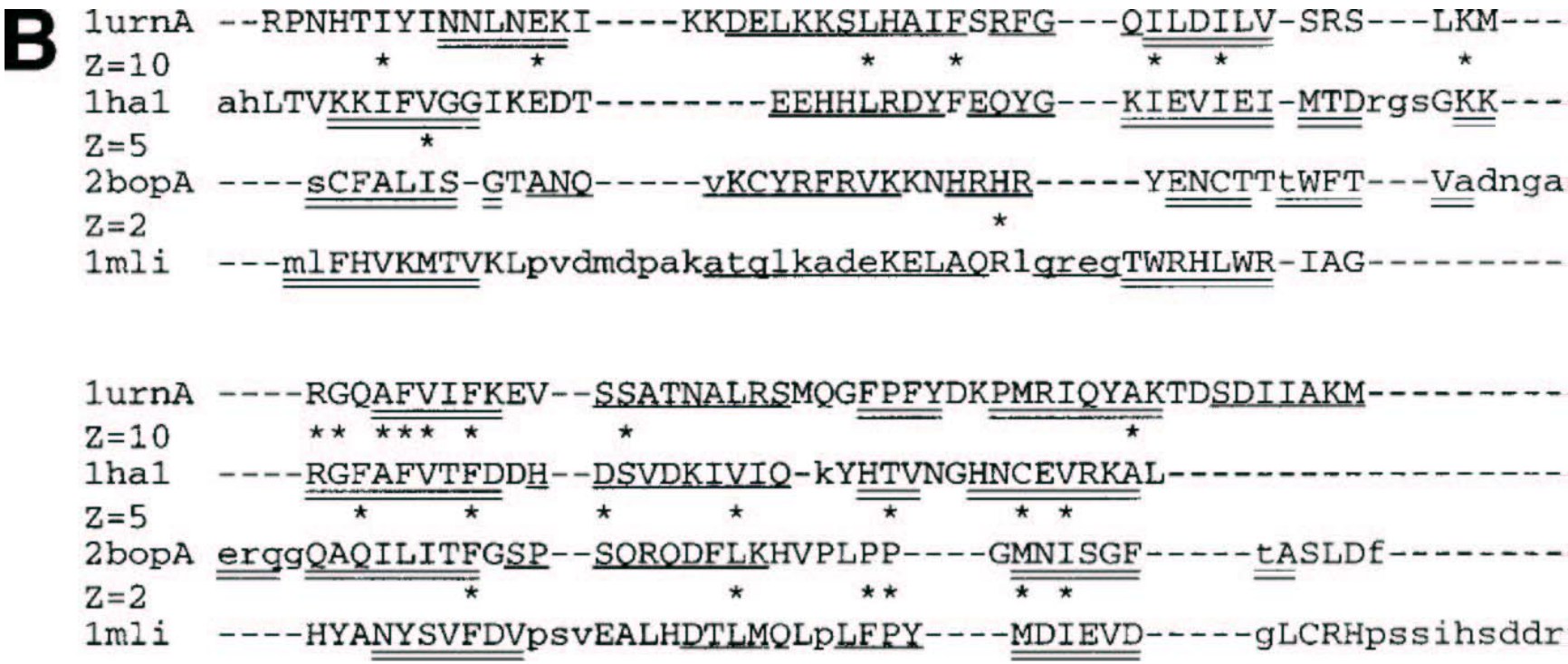
Attractor 1 can be characterized as alpha/beta,
attractor 2 as all-beta, attractor 3 as all-alpha,
attractor 5 as alpha-beta meander (1mli), and
attractor 4 contains antiparallel beta-barrels e.g. OB-fold (1prtF).

Fold Types & Neighbors



Structural neighbours of 1urnA (top left). 1mli (bottom right) has the same topology even though there are shifts in the relative orientation of secondary structure elements.

Sequence Alignment of Fold Neighbors



Frequent Fold Types