

# Secondary Structure Prediction Software

254



**Figure 11.3** Comparison of secondary structure predictions by various methods. The sequence of flavodoxin, an  $\alpha/\beta$  protein, was used as the query and is shown on the first line of the alignment. For each prediction, H denotes an  $\alpha$  helix, E a  $\beta$  strand, T a  $\beta$  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](#))

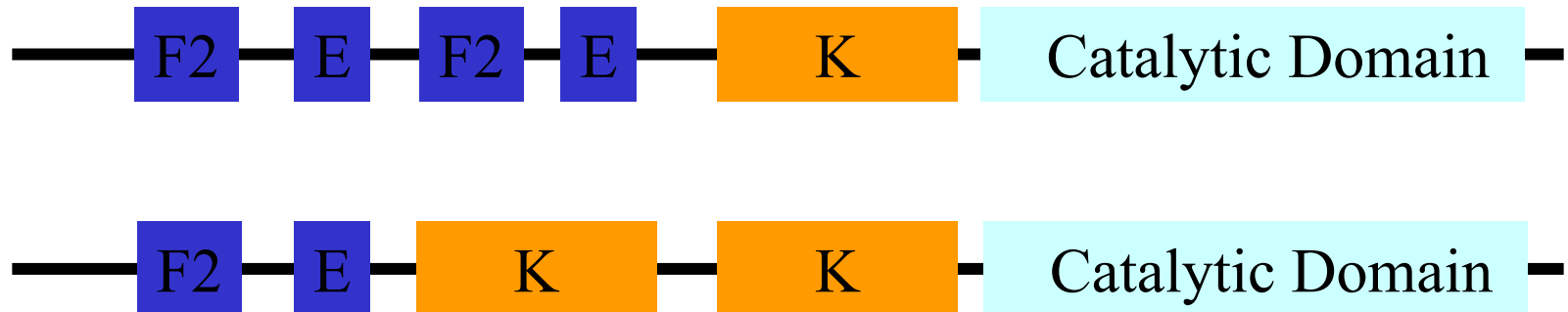
# Motif Detection Tools

- 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



PLAT

# Domain Architecture Tools

- CDART
  - Protein [AAH24495](#); [Domain Architecture](#);
  - It's [domain relatives](#);
  - Multiple [alignment](#) for 2<sup>nd</sup> 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  $\phi$  and  $\psi$  are constrained.
- Proteins structures often have hydrophobic core.

# Protein Folding

Unfolded



Rapid ( $< 1\text{ s}$ )

Molten Globule State



Slow ( $1 - 1000\text{ s}$ )

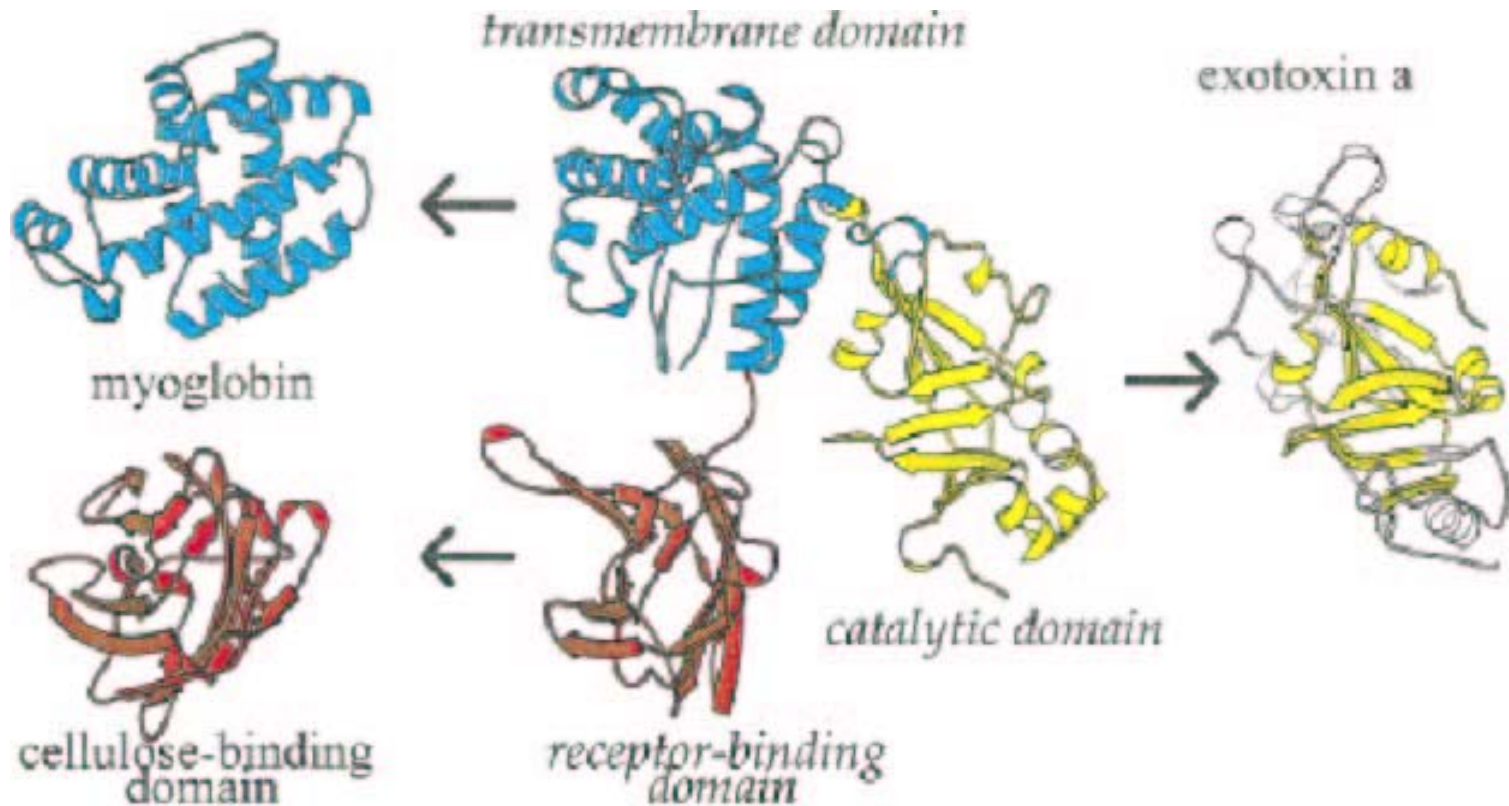
Folded Native State

- 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 structure & 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

**NCSA Mosaic: Document View** **WWW browser (NCSA Mosaic)**

File Options Navigate Annotate help

Document Title: SCOP: Family: Interleukin 8-like

Document URL: <http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.0.004>

Structural Classification of Proteins

scop navigation buttons

**Family: Interleukin 8-like**

Lineage:

1. Root: [scop](#)
2. Class: [Alpha](#)
3. Fold: [Interleu](#)
4. Superfamily: [Interleukin 8-like](#)
5. Family: [Interleukin 8-like](#)

Proteins:

1. Interleukin-8
  1. [108](#)
    1. chain a
    2. chain b
  2. [108](#)
    1. chain a
    2. chain b
2. Platelet factor 4
  1. bovine (*Bos taurus*) (1)
    1. [1nlf](#)
      1. chain a
      2. chain b
      3. chain c
      4. chain d
3. Macrophage inflammatory protein 1beta has different oligomerisation mode
  1. human (*Homo sapiens*) (2)
    1. [1hum](#)
      1. chain a
      2. chain b
    2. [1hum](#)
      1. chain a
      2. chain b

keyword search facility

Enter search key:  Search

Back Forward Home Reload Open... Save As... Clone New Window Close Window

**3-D viewer (RasMol)**

**image viewer (xv)**

Human MIP-1β and Interleukin 8 Dimers

MIP-1β IL-8

**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.
  - Topology (T), topological connections &
  - Homologous Superfamily (H), similar str and functions.

# DALI Domain Dictionary

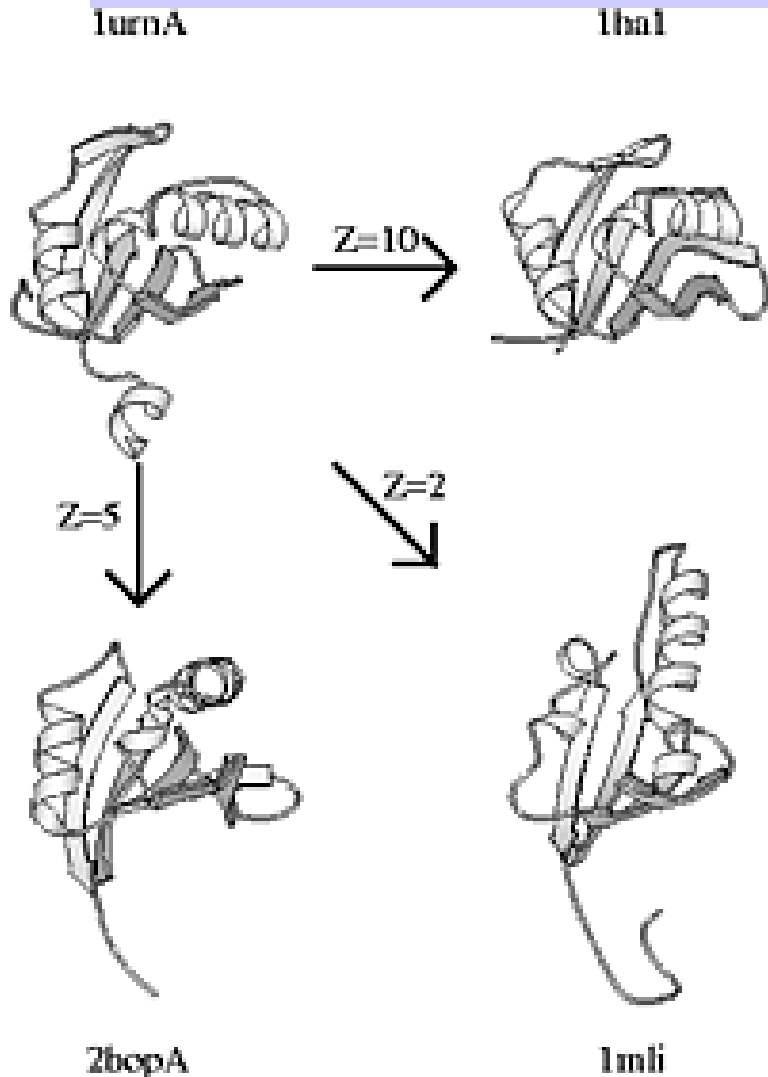
- Completely automated; 3724 domains
- Criteria of compactness & recurrence
- Each domain is assigned a Domain Classification number DC\_l\_m\_n\_p representing fold space attractor region (l), globular folding topology (m), functional family (n) and sequence family (p).

# 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

**B**

```

1urnA  --RPNHTIYINNLNEKI-----KKDELKKSLSLHAIFSRFG---QILDILV-SRS---LKM---
Z=10      *      *      *      *      *      *
1ha1    ahLTVKKIFVGGIKEDT-----EEHHLRDYFEOYG---KIEVIEI-MTDrgsGKK---
Z=5      *
2bopA   ----sCFALIS-GTANQ-----vKCYRFRVKKNHRHR-----YENCTTtWFT---Vadnga
Z=2
1mli    ---mLFHVKMTVKLPvdmdpakatgIkadeKELAQRlgregTWRHLWR-IAG-----

```

```

1urnA   ----RGQAFVIFKEV--SSATNALRSMQGFPPFYDKPMRIQYAKTSDIIAKM-----
Z=10    **  ***  *      *      *
1ha1    ----RGFAFVTFDDH--DSVDKIVIO-kyHTVNGHNCEVRKAL-----
Z=5     *      *      *      *      *      *
2bopA   erggQAQILITFGSP--SORODFLKHVPLPP---GMNISGF-----tASLdf-----
Z=2     *      *      **      *      *
1mli    ----HYANYSVFDVpsvEALHDTLMQLpLFPY---MDIEVD-----gLCRHpssihsddr

```



# Frequent Fold Types



(141) 1hdcA:1  
alpha/beta domain



(85) 1mfA:3  
immunoglobulin fold



(63) 1ceo:2  
TIM barrel



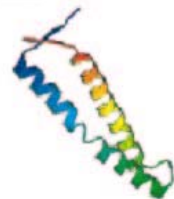
(43) 1befA:1  
helical bundle



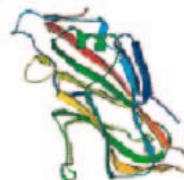
(36) 2pii:2  
alpha/beta-meander



(33) 1vdfA:1  
single helix



(27) 1grj:2  
coiled coil



(25) 1bbt2:1  
beta-meander



(19) 1rro:2  
EF-hand



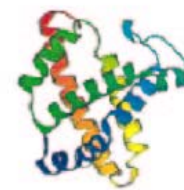
(18) 1oetC:3  
HTH-motif



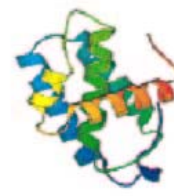
(18) 1ptf:1  
OB-fold



(17) 3grs:2  
FAD/NAD binding domain



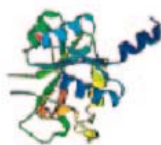
(14) 1mbd:1  
globin fold



(13) 1vin:3  
cyclin fold



(13) 1aozA:15  
blue copper protein



(13) 1lcf:17  
periplasmic binding protein



(12) 1eelA:3



(12) 1epaA:1  
lipocalin fold



(12) 2arcA:4  
beta-roll



(12) 2yhx:3  
actin fold