Molecular Modeling 2018
lecture 3

Protein domain classification
TOPS
Contact maps
Domain

- is defined as an autonomously-folding substructure of a protein.
- must be > 30 residues
- has a single core
- is usually composed of one chain (occasionally composed of multiple chains)
- is usually composed on one contiguous segment of the chain (occasionally made of discontiguous segments)
SCOP -- a heirachy

(1) class  
(2) fold  
(3) superfamily  
(4) family  
(5) protein  
(6) species  

global characteristics (no evolutionary relation)

Similar “topology”. Distant evolutionary cousins?

Clear structural homology

Clear sequence homology

functionally identical

unique sequences

http://scop.berkeley.edu
Proteins of the same class conserve secondary structure content
SCOP -- fold

within \( \alpha/\beta \) proteins -- Mainly parallel beta sheets (beta-alpha-beta units)

TIM-barrel (22)
swivelling beta/beta/alpha domain (5)
spoIIaa-like (2)
flavodoxin-like (10)
restriction endonuclease-like (2)
ribokinase-like (2)
chelatase-like (2)

Many folds have historical names. “TIM” barrel was first seen in TIM. These classifications are done by eye, by experts.

Proteins of the same Fold conserve topology.
SCOP fold jargon

example: $\alpha/\beta$ proteins: flavodoxin-like

SCOP Description: 3 layers, $\alpha/\beta/\alpha$; parallel beta-sheet of 5 strand, order 21345

Note the term: “layers”

Rough arrangements of secondary structure elements.

Note the term: “order”

The sequential order of beta strands in a beta sheet.
Fold-level similarity

7-stranded alpha/beta barrel

SSE are in the same order along the chain, and trace roughly the same path through space. Similarity is evident when viewed side-by-side.

But the SSE do not superpose. Some superposition algorithms fail to superpose proteins of the same fold.
Superfamily level similarity

FAD-linked reductases

Members of the same superfamily cannot usually be found in a BLAST search. But can be identified by structural superposition.

Proteins in the same superfamily may look completely different, but upon close inspection they contain a superposable domain of secondary structure elements.
Different members of the same family superimpose well. At this level, a structure may be used as a molecular replacement model for X-ray crystallography.

A BLAST search using one family member finds all other family members.
TOPS topology cartoons

A simple way to draw a protein

beta strand pointing up
beta strand pointing down
alpha helix
connections

A parallel beta sheet
An anti-parallel beta sheet
TOPS topology cartoons

A right-handed $\beta\alpha\beta$ unit

A left-handed $\beta\alpha\beta$ unit (rarely seen)

connection in middle means on top.
connection on side means on bottom.
How to draw TOPS

On course website, find the link "TOPS practice" (tops_practice.moe)

Save it. Open it in moe.
How to draw TOPS

Line up the molecule along the beta sheet, if present. Otherwise choose a direction so that secondary structures are mostly perpendicular to the screen.
TOPS diagram

- Draw secondary structures first.
TOPS diagram

• number them and connect

Be careful to draw connections to the center or side, when it is in front or in back, respectively.
Name it. SCOP-style.

- 3 layers, 2-4-2 $\alpha\beta\alpha$, all parallel, 1234
To draw a barrel, determine strand neighbors, up or down, arrange triangles in a circle. Draw connector lines in front, or in back, of triangles. Ignore extra loops.
Contact maps: proteins in 2D

In a Contact Map: “1” = $D_{ij} < 8 \text{Å}$

“0” = $D_{ij} > 8 \text{Å}$

- Hairpin
- Helix
- Parallel strands
- Anti-parallel strands
TOPS and contact maps

A "contact map" for a $\beta\alpha\beta$ unit.
Contact map for a small protein

A contact map contains enough information to build the 3D structure within ~2Å RMSD.
A crude contact map based on SSEs

(1) Arrange the SSEs along the sequence (a line) in both directions
(2) Draw a line parallel to the diagonal for each helix
(3) For any two SSEs that touch, draw a line parallel to the diagonal if the contacts are parallel, draw a line perpendicular to the diagonal if the contacts are anti-parallel. Draw a dotted line if a helix is involved.
Crude contact map to TOPS diagram

- α1 → α2
- α2 → α3
- α3 → α4
- α4 → β1
- β1 → β2
- β2 → β3
- β3 → β4
- β4 → β5
- β5 → α1

Color codes:
- Blue: alpha
- Purple: beta
- Orange: alpha-beta
- Green: alpha-alpha
Crude contact map to TOPS diagram
Most genes represent multidomain proteins

~40% of known structures (crystal, NMR) are multidomain proteins, but

Most of all proteins are multidomain. (~60% in unicellular organisms, ~90% in eukaryotes).

Domain boundaries can be seen as "weak" connections in the structure.

"Weak" means few contacts and few chain cross-overs.

Domain boundaries can be seen in multiple sequence alignments if the alignments are of whole genes.
C/N-Terminal domain, cut-and-pasted

(research)
Example of two, discontiguous domains seen using a contact map.

Contacts are mostly within domains, not between domains. One domain consists of N and C-terminal parts.
C/N-Terminal domain, cut-and-pasted
Review questions

- Describe a beta turn
- Describe a helix cap
- Describe a beta bulge
- What kind of sequence patterns correlate with local structure?
- Can you draw a greek key made up of beta strands, as arrows?
- Name three types of alpha helical super secondary structure.
- What is a beta-alpha-beta unit?
- Why are baby units right-handed?
- What is a domain?
- What is a “fold” according to SCOP?
- What does “strand order” mean w/respect to SCOP naming?
- What defines a sequence “family”?
- What defines a sequence “superfamily”?
- Draw a beta-alpha-beta unit using TOPS.
- Draw a crude contact maps based on a TOPS diagram.
- How do we see domain boundaries using a contact map?
- How can we infer domain boundaries using a multiple sequence alignment?
Supplementary slides
CATH

• Class
• Architecture
• Topology
• Homology

**Architecture** = conserves arrangement of SSE (secondary structural elements) but not sequential order.

**Topology** = like SCOP Fold.

http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html
**protein structure and representation - a hierarchy or a continuum?**

<table>
<thead>
<tr>
<th>Structure</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary structure</td>
<td>1D, three states</td>
</tr>
<tr>
<td>Local structure</td>
<td>motifs, backbone angles.</td>
</tr>
<tr>
<td>Super-secondary structure</td>
<td>TOPS.</td>
</tr>
<tr>
<td>Inter-residue distances</td>
<td>2D contact maps</td>
</tr>
<tr>
<td>Tertiary structure</td>
<td>3D backbone</td>
</tr>
<tr>
<td>Side chain conformation</td>
<td>rotamers</td>
</tr>
<tr>
<td>Domain-domain interactions</td>
<td>interface maps</td>
</tr>
<tr>
<td>Quaternary structure</td>
<td>poses, interaction maps.</td>
</tr>
</tbody>
</table>