Protein classification

SCOP
TOPS
Contact maps
A protein domain...

...is defined as an autonomously-folding substructure of a protein.
...must be > 30 residues, typically < 200
...has a single hydrophobic core
...is usually composed of one chain (occasionally composed of multiple chains)
...is usually composed on one contiguous segment of a chain (occasionally made of discontiguous segments of the same chain)
SCOP -- a hierarchy

http://scop.berkeley.edu

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

Increasing structural similarity:
- similar secondary structure content
- vague structural homology
- Clear structural homology
- Clear sequence homology
- nearly identical sequences
- individual structures
Definition of Family and Superfamily/Fold

A **Family** is the set of homologs we can find by BLAST search.

A **Superfamily** is a set of distant homologs that cannot be easily found by BLAST search.

A **Fold** (as defined by SCOP) is an even more distant homologous relationship.

A **Class** is not a homologous relationship.
Family level similarity

FAD/NAD-linked reductases, N-terminal and central domains [51943]

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.
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 contains a superposable domain of secondary structure elements.
**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.
SCOP -- fold level

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 level 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.

$\beta$ layer

$\alpha$ layer

$\alpha$ layer

2 1 3 4 5
SCOP -- class

1. all $\alpha$ (289)
2. all $\beta$ (178)
3. $\alpha/\beta$ (148)
4. $\alpha+\beta$ (388)
5. multidomain (71)
6. membrane (60)
7. small (98)
8. coiled coil (7)
9. low-resolution (25)
10. peptides (148)
11. designed proteins (44)
12. artifacts (1)

Not true classes of globular protein domains

Proteins of the same class conserve secondary structure content
Exercise 16.1: Superimpose by hand

Do these pairs: 1N9U vs 1N9V  
1WFA_A vs 1WFA_B

File | Open: RCSB PDB: code: xxxx
• Delete parts not needed (e.g. waters). Select Chain label in SEQ. R-mouse/Delete.

Hide | All atoms
Ribbon | Style: oval, Color: chain or terminus
Select | synchronize

In SEQ window:
  Double-click on chain label to select one molecule.

In MOE window:
  Rotate selected: meta-middlemouse-drag.
  Translate selected: shift-meta-middlemouse-drag
  Rotate all: middlemouse-drag
  Translate all: shift-middlemouse-drag

Until molecules are superposed.

Show me your superposition.
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.
• 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
Exercise 16.2: contact map and TOPS cartoon

Open MOE

File | Open: RCSB PDB: codes: 2ptl

Ribbon | Style: oval

Ribbon | Color : structure

Identify SSEs. Draw triangles and circles

Ribbon | Color : terminus

Number and connect SSEs.
• all anti-parallel barrel, closed; \( n=6, S=10 \); greek-key

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.
Exercise 16.3: TOPS cartoon of beta barrel

Open MOE. Open Green Fluorescent Protein

File | Open: RCSB PDB: code: 2b3p

Ribbon | Style: oval

Ribbon | Color : structure

Identify SSEs. Draw triangles and circles

Ribbon | Color : terminus

Number SSEs. Draw connections. Label termini.
Contact maps: proteins in 2D

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

“0” = $D_{ij} > 8\text{Å}$
TOPS and contact maps

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

parallel sheet.

helix contacts.
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
Crude contact map to TOPS diagram

- Alpha-1 (α1)
- Beta-1 (β1)
- Alpha-2 (α2)
- Beta-2 (β2)
- Alpha-3 (α3)
- Beta-3 (β3)
- Alpha-4 (α4)
- Beta-4 (β4)
- Alpha-5 (α5)
- Beta-5 (β5)
Exercise 16.4: TOPS from contact map

Do this on paper.

Draw a TOPS cartoon that has this contact map. SSEs are $\beta\alpha\beta\alpha\beta$. 
Most genes represent multidomain proteins

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

**Most** of all proteins are multidomain. (~60% in uncellular 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.
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.
Exercise 16.5: recognizing domains

1YY3 (easy) ___________________
1IO1 (easy) ___________________
1G71 (hard) ___________________

Retrieve, one at a time, using

File | Open: RCSB: Code:  xxxx

Display ribbon and backbone atoms (for labeling)

Ribbon | Style: oval

If using MOE2013, Show | backbone, so that you can select atoms.
If using MOE2018 you can select atoms on the ribbon.

Select atoms of each domain. Color domains differently.
Review questions

- 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.
- Find 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>