Bioinformatics 2 -- lecture 3

Structure classification

TOPS

Domains

3.1 Structure classification

Structure Classification Of Proteins: SCOP

- Contains information about classification of protein structures and within that classification, their sequences.
- Aides in communicating structural information.
- Helps
- http://scop.berkeley.edu

SCOP -- a heirachy

(1) class ← global characteristics (no evolutionary relation)

(2) fold ← Similar “topology” . Distant evolutionary cousins?

(3) superfamily ← Clear structural homology

(4) family ← Clear sequence homology

(5) protein ← functionally identical

(6) species ← unique sequences

- http://scop.berkeley.edu

SCOP -- classes

1. all α (126) ← number of sub-categories (126)
2. all β (81)
3. α/β (87)
4. $\alpha+\beta$ (151)
5. multidomain (21)
6. membrane (21)
7. small (10)
8. coiled coil (4)
9. low-resolution (4)
10. peptides (61)
11. designed proteins (17)

Items 7–11: Not true classes

Item 9: possibly not complete, or erroneous

Proteins of the same class conserve secondary structure content

SCOP -- folds

within α/β 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)

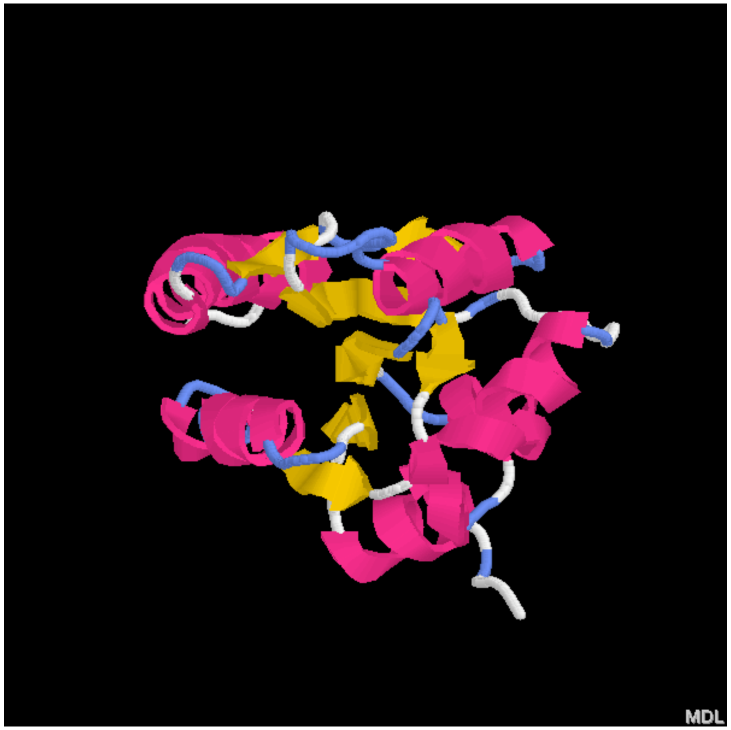
Many folds have historical names. “TIM” barrel was first seen in TIM. These classifications are done *by eye*, by experts.

Proteins of the same Fold conserve topology.

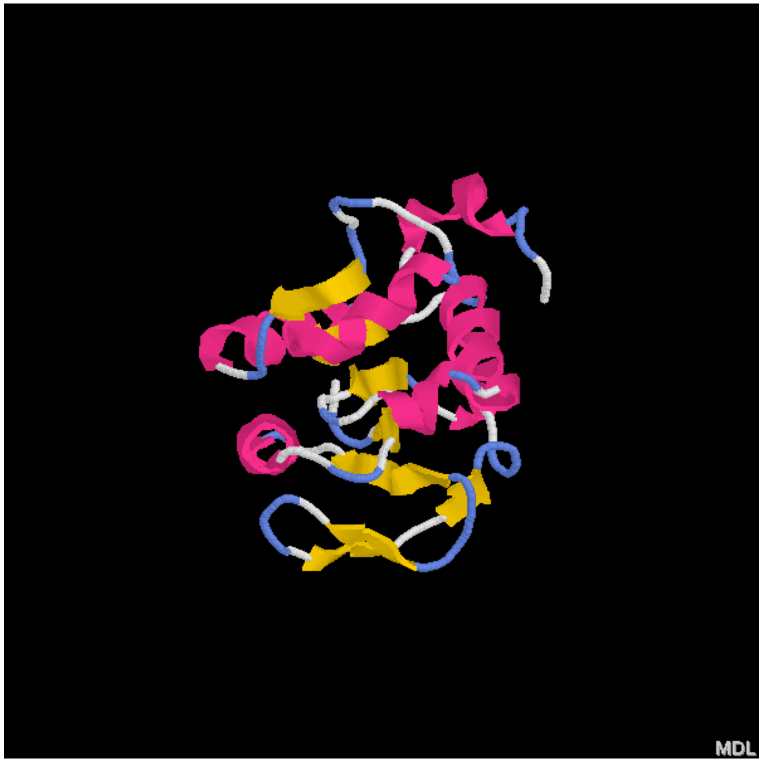
fold-level similarity

common topological features

catalase



flavodoxin



At the fold level, a common core of secondary structure is conserved. Outer secondary structure units may not be conserved.

superfamily

within <u>α/β proteins: flavodoxin-like</u>

Fold Description: 3 layers, α/β/α; parallel beta-sheet of 5 strand, order 21345

Superfamilies:

1. Catalase, C-terminal domain (1)
2. CheY-like (1)
3. Succinyl-CoA synthetase domains (1)
4. Flavoproteins (3)
5. Cobalamin (vitamin B12)-binding domain (1)
6. Ornithine decarboxylase N-terminal "wing" domain (1)
7. Cutinase-like (1)
8. Esterase/acetylhydrolase (2)
9. Formate/glycerate dehydrogenase catalytic domain-like (3)
10. Type II 3-dehydroquinate dehydratase (1)

Note the term: “layers”

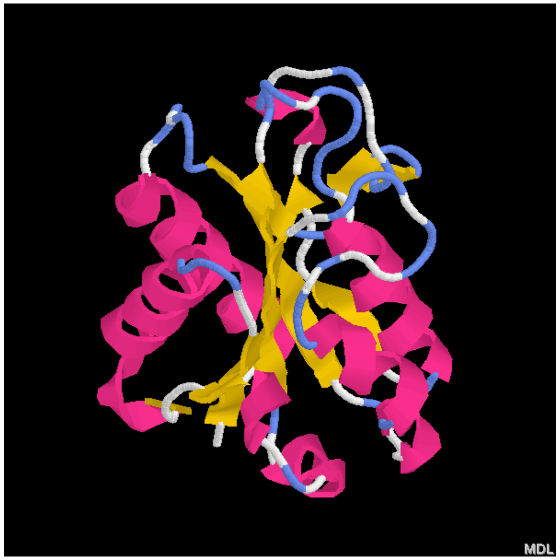
These are not domains. No implication of structural independence.

Note how beta sheets are described: number of strands, order (N->C)

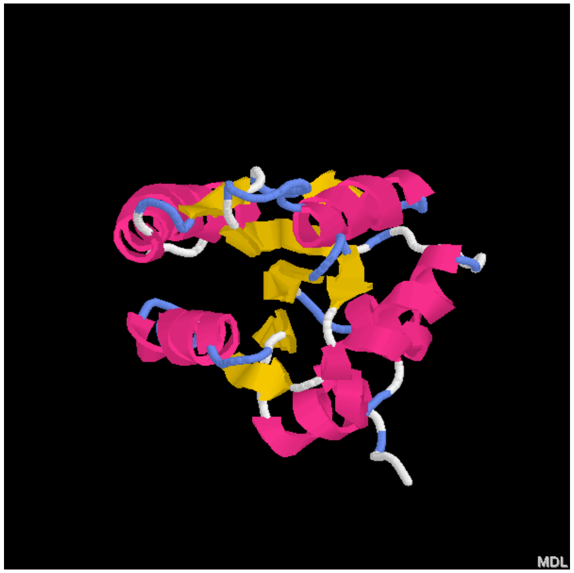
Proteins of the same Superfamily align in space, but may have different function

Superfamily:Flavoproteins

Flavodoxin-related (7)



NADPH-cytochrome p450 reductase, N-terminal domain



Quinone reductase

These molecules do not superimpose well, but side-by-side you can easily see the similar topology. Secondary structures align 1-to-1, mostly.

Family: quinone reductases

binds FAD

Proteins:

quinone
reductase type 2

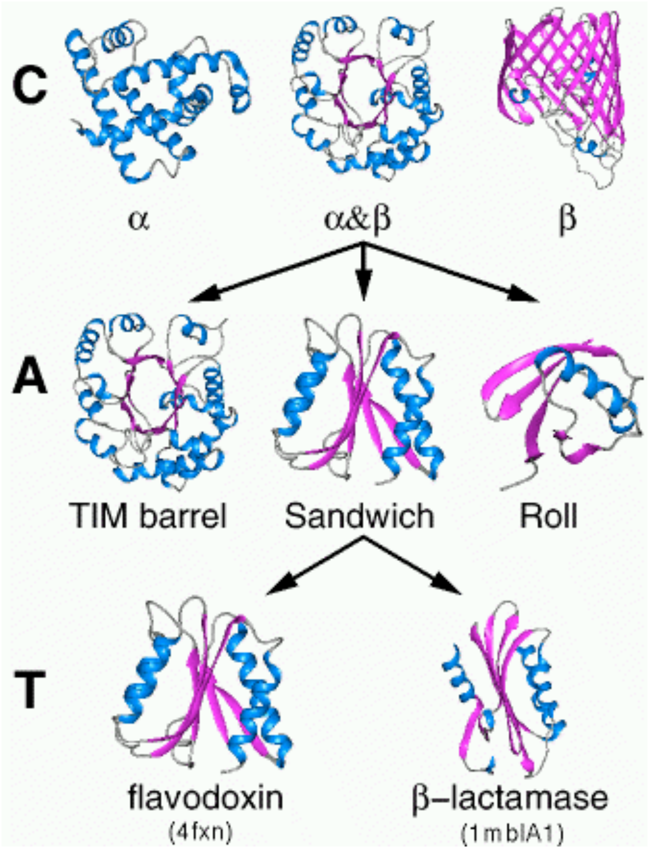
NADPH quinone
reductase

Different members of the same family superimpose well. At this level, a structure may be used as a *molecular replacement model*.

A BLAST sequence search using one member of the family finds all the others.

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

3.2 TOPS

TOPS topology cartoons

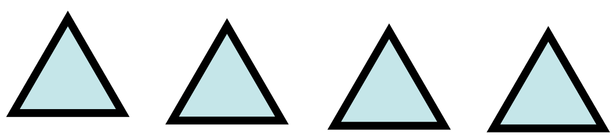
A simple way to draw a protein

beta strand pointing up

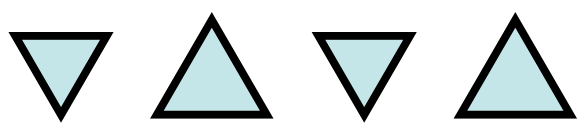
beta strand pointing down

alpha helix

connections



A parallel beta sheet

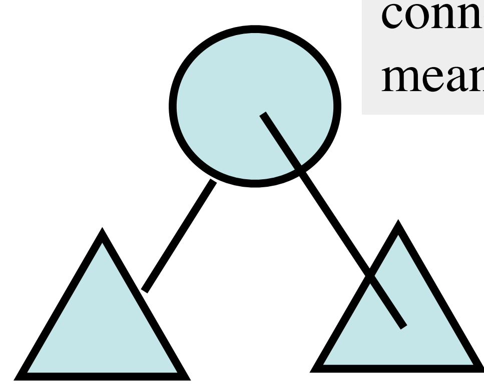


An anti- parallel beta sheet

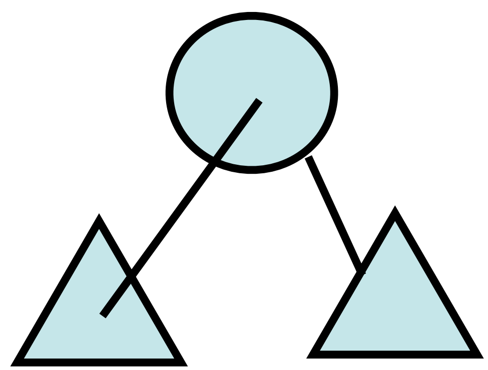
TOPS topology cartoons

connection in middle
means on top.
connection on side
means on bottom.

A right-handed βαβ unit

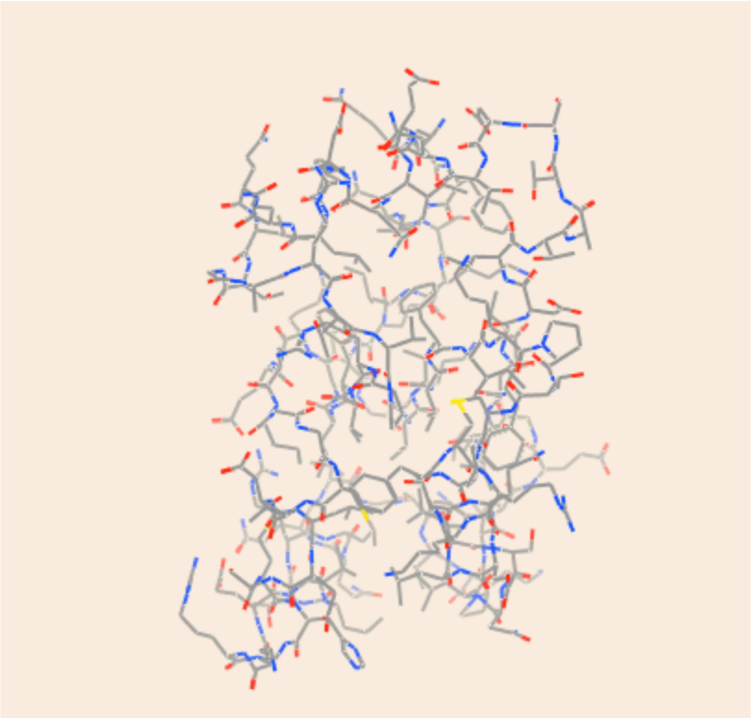


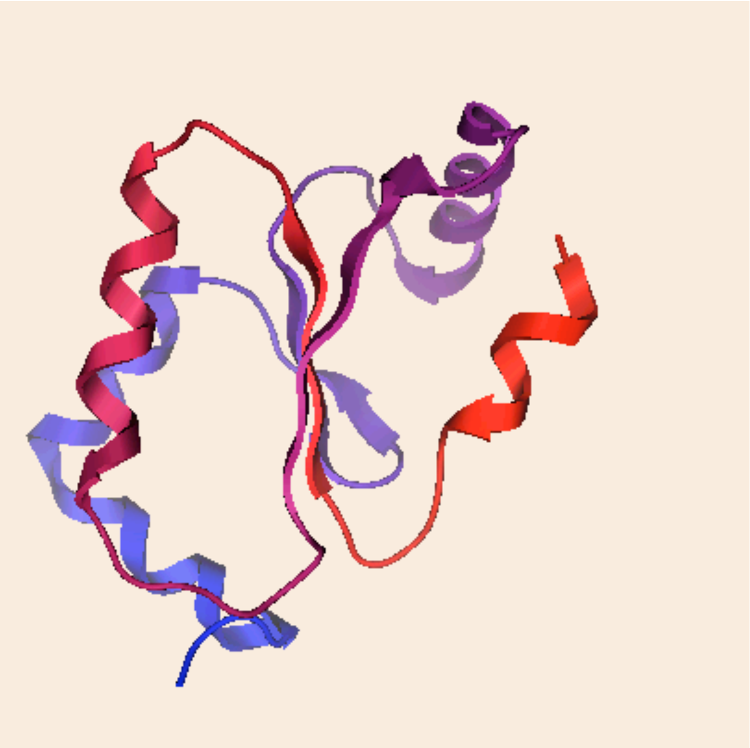
A left-handed βαβ unit
(rarely seen)



How to draw TOPS

Select one molecule and Hide the others.
Render-->Backbone-->Cartoon
Render-->Backbone-->Color-->terminus (to help see the chain direction)

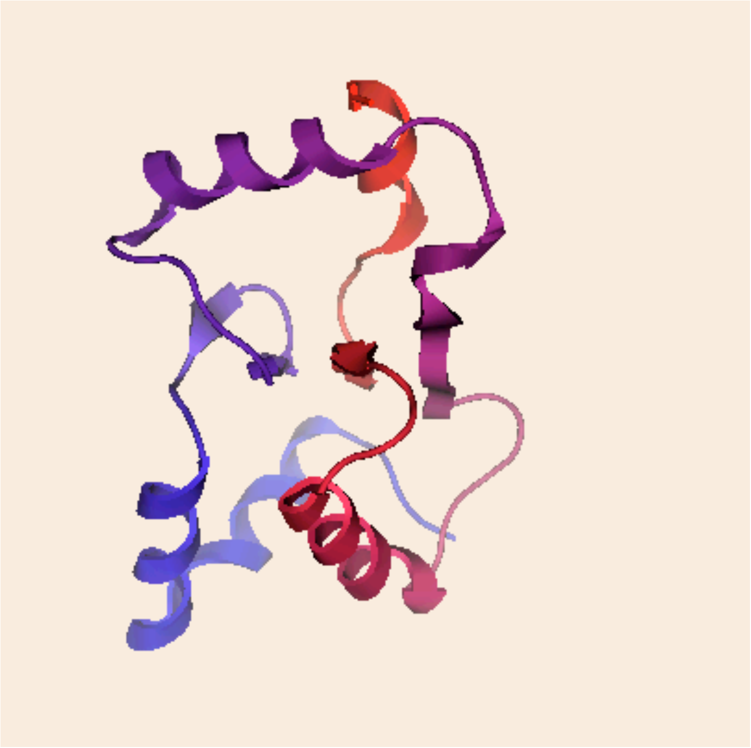




How to draw TOPS

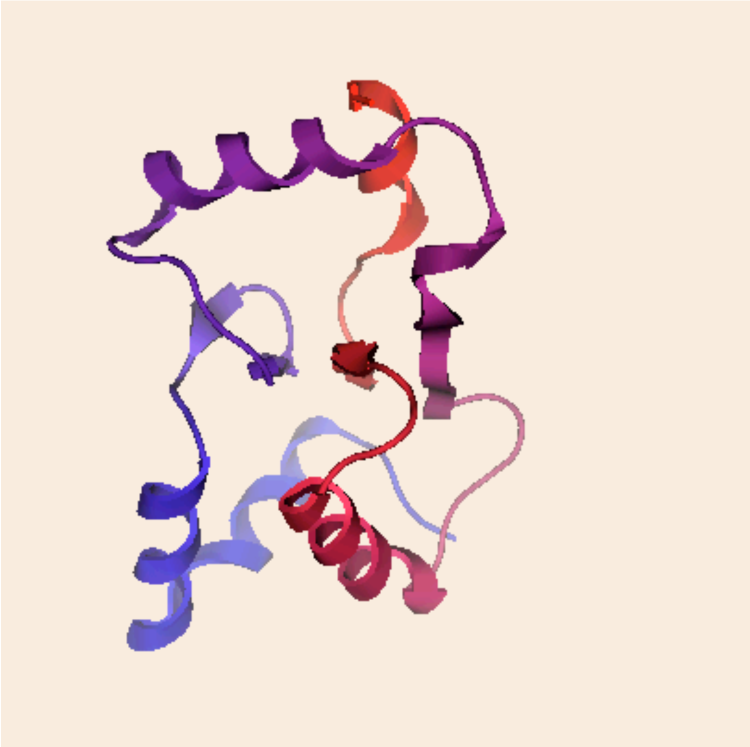
Line up the molecule along the beta sheet, if present.

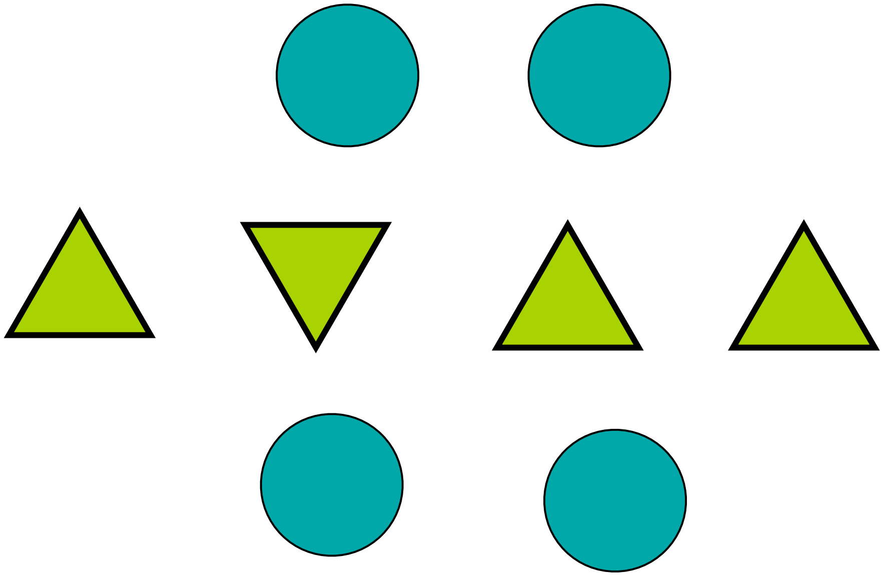
Otherwise choose a direction so that secondary structures are mostly perpendicular to the screen.



TOPS diagram

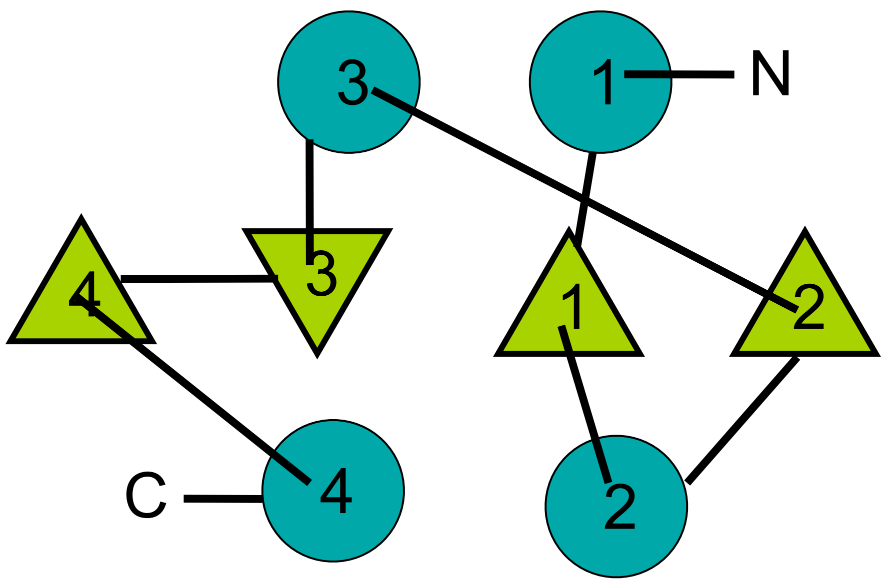
- Draw secondary structures first.





TOPS diagram

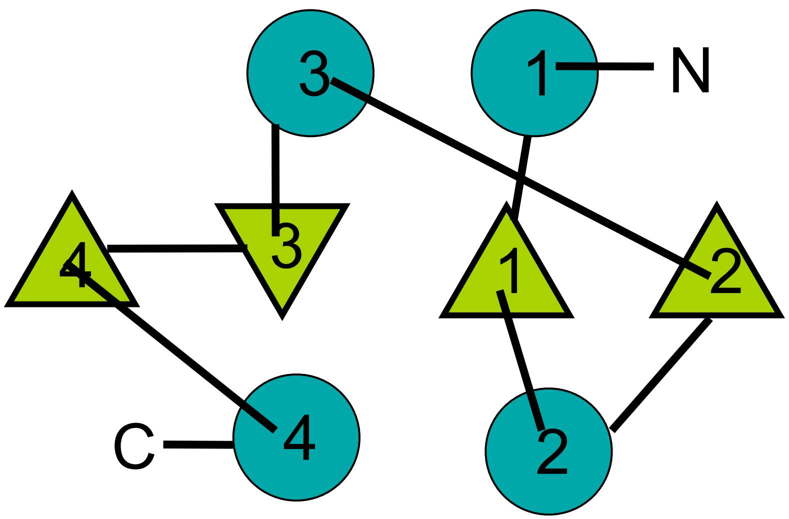
- number them and connect them



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

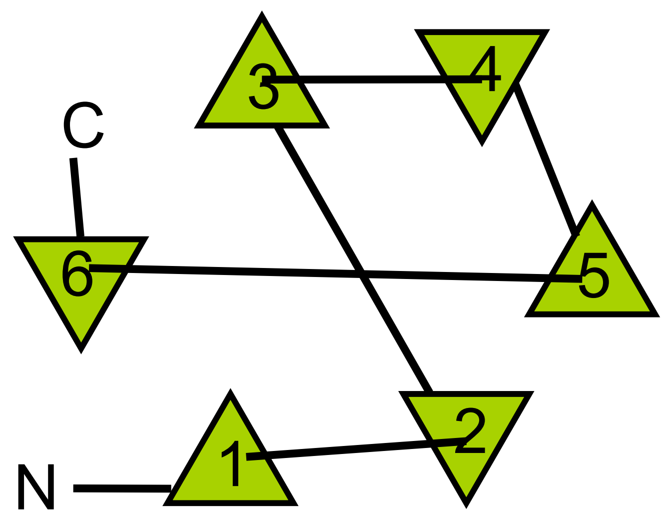
SCOP-style naming

- 3 layers, 2-4-2 αβα, mixed sheet, 4312



SCOP-style naming

- all anti-parallel beta-barrel, closed. n=6

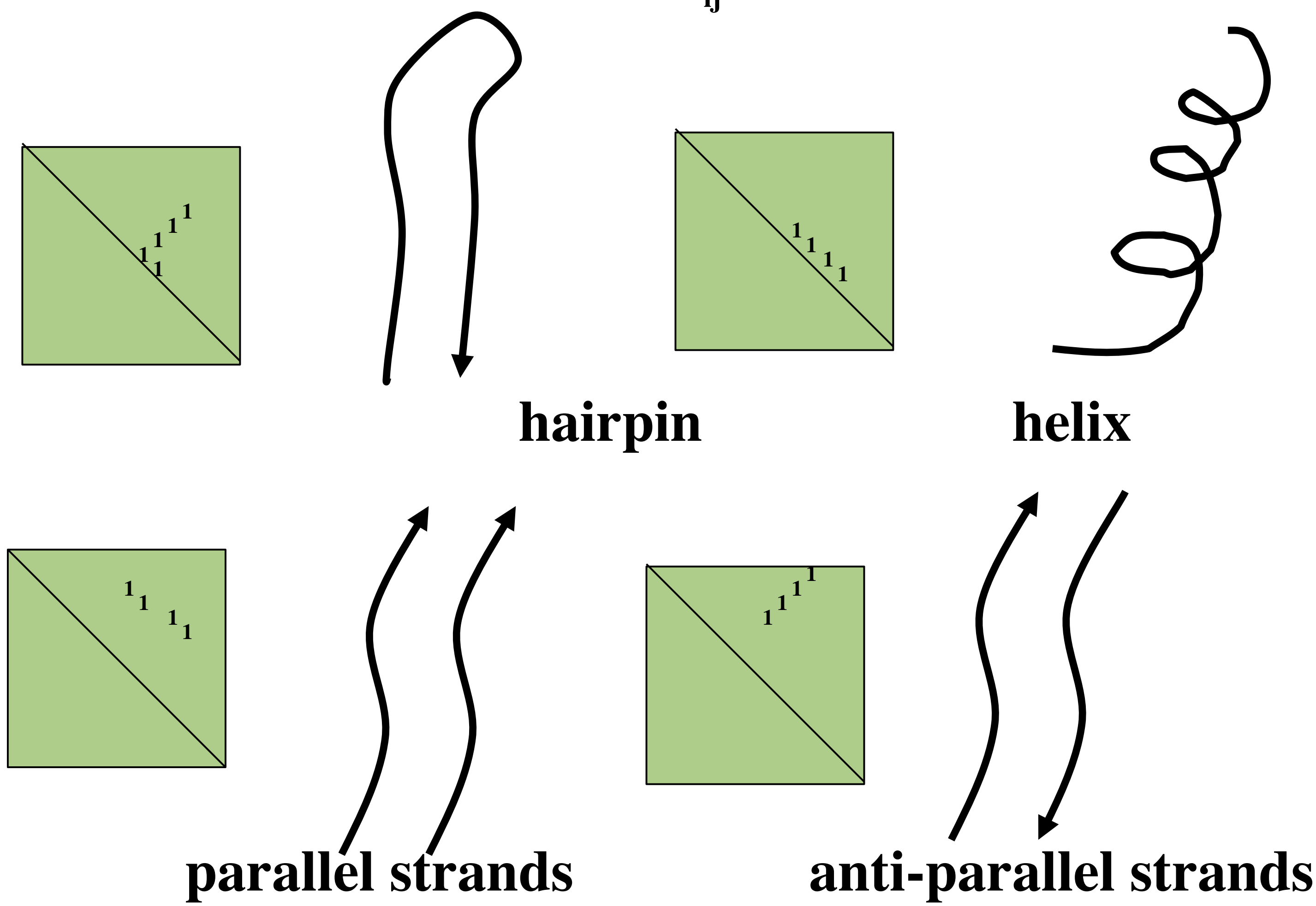


To draw a barrel, put beta strands in a circle. Neighbors on the circle form H-bonds

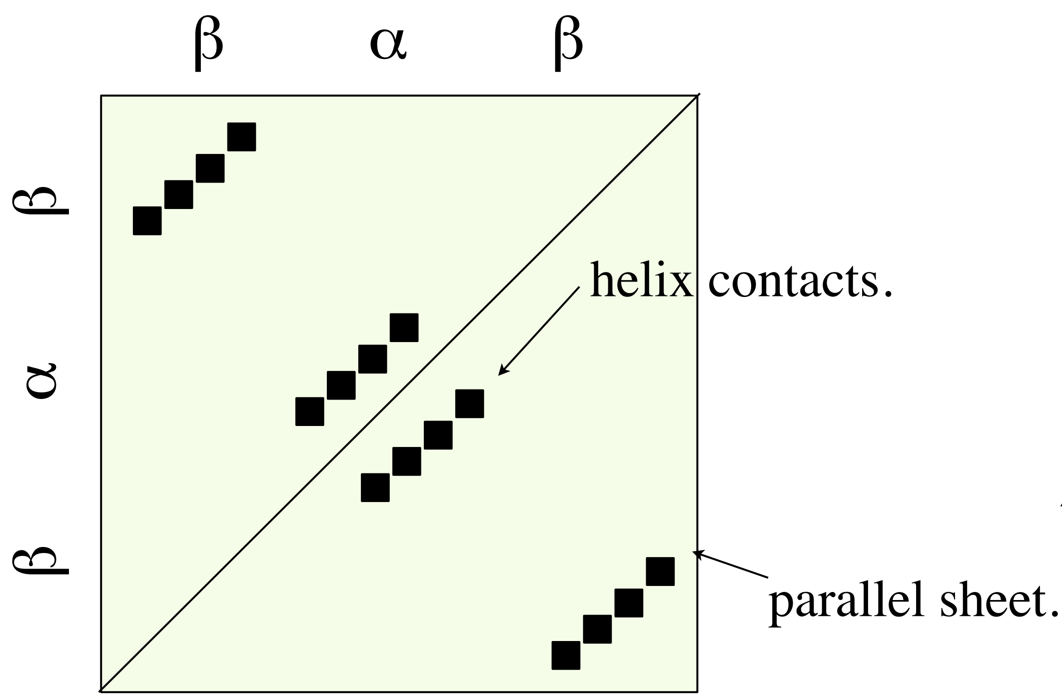
Contact maps: proteins in 2D

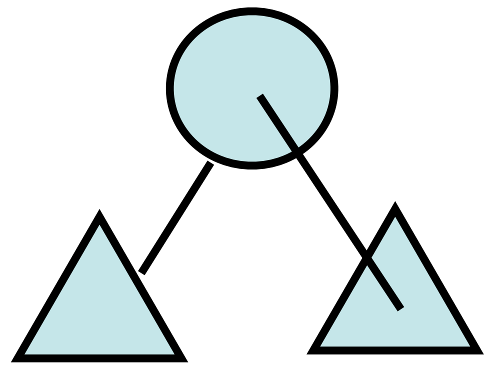
In a Contact Map: “1” = $D_{ij} < 8Å$

“0” = $D_{ij} > 8Å$



TOPS and contact maps

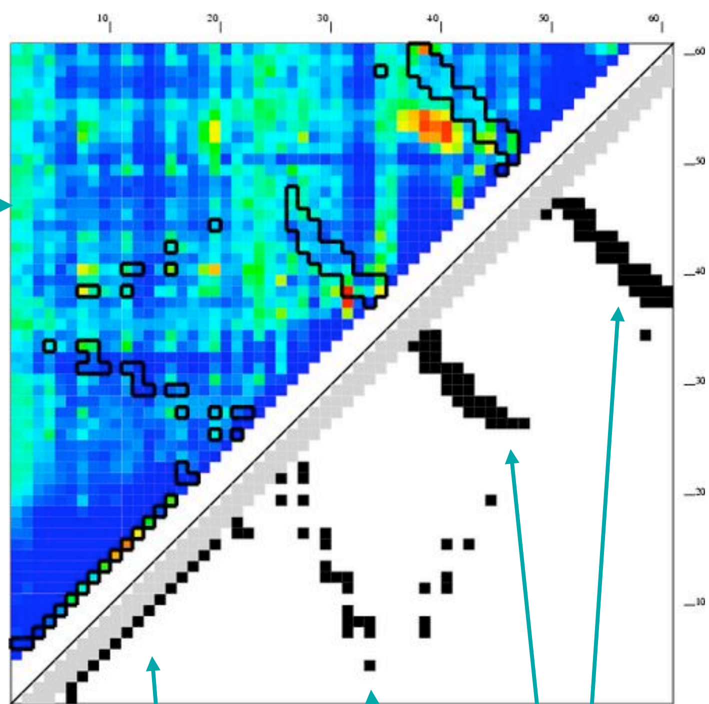




A "contact map" for a βαβ unit.

note-to-self: Update PDB index

Contact map for a small protein

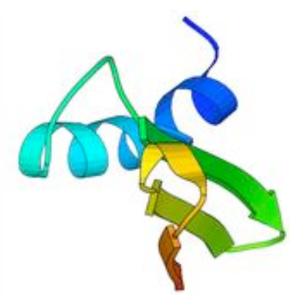


A contact map contains enough information to build the 3D structure within ~2Å RMSD.

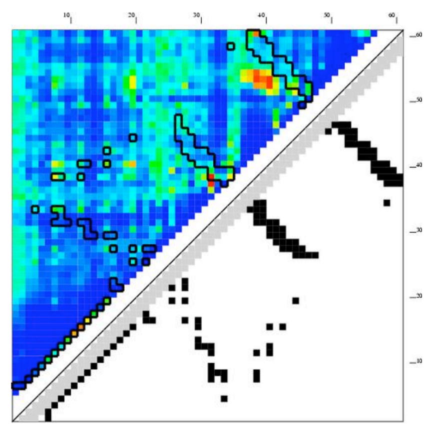
A crude contact map based on SSEs

(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.

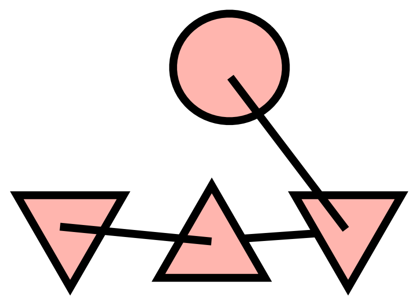
Structure



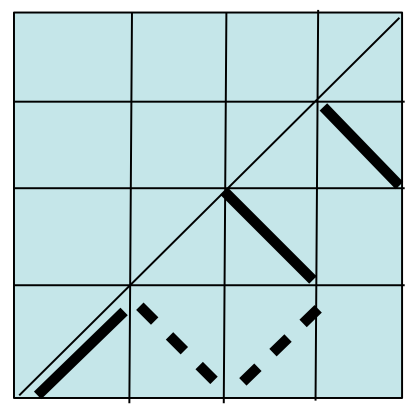
contact map



TOPS



crude contact map



Exercise 3.1: contact map and TOPS cartoon

Open MOE

Search/download **2ptl** from **File/Protein database**

Ribbon/Style:oval

Draw a TOPS diagram.

Make a crude contact map based on secondary structure elements.

3.3 Domains

Most genes represent multidomain proteins

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

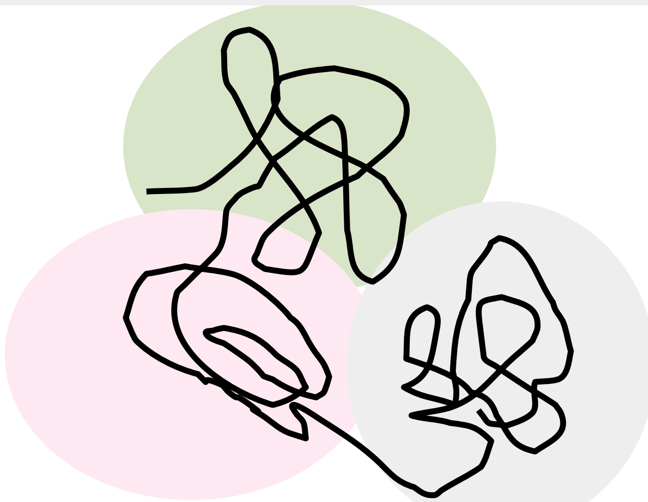
Most of all proteins are multidomain.(~60% in uncellular organisms, ~90% in eukaryotes).

Where do my domains start and end??

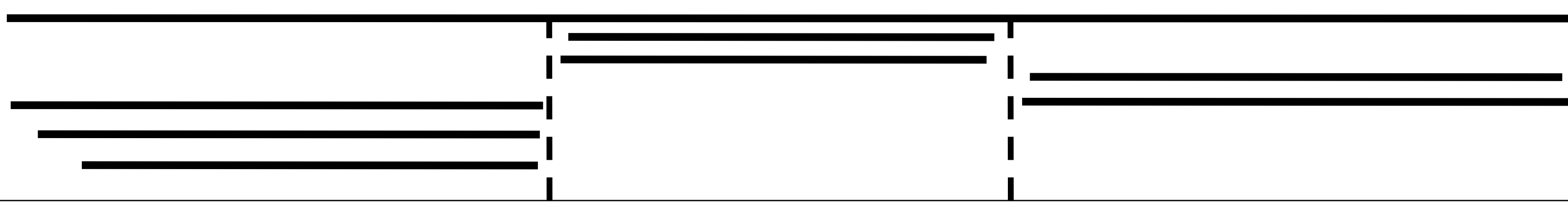
Multidomain proteins

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

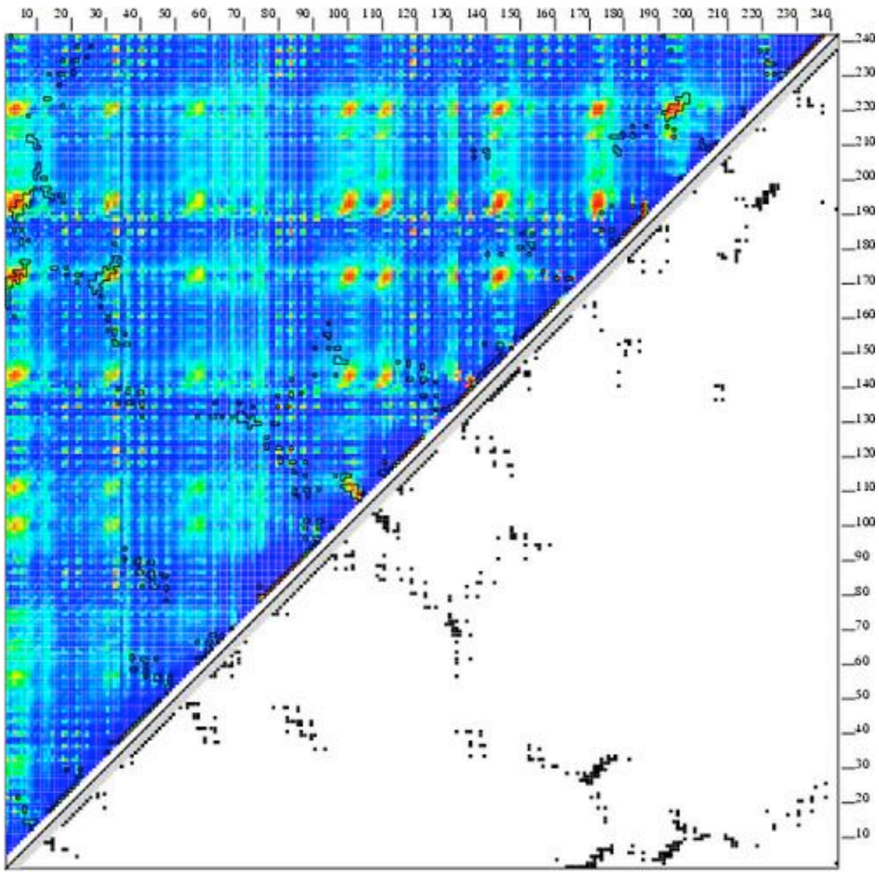
"Weak" means few contacts and few chain cross-overs.

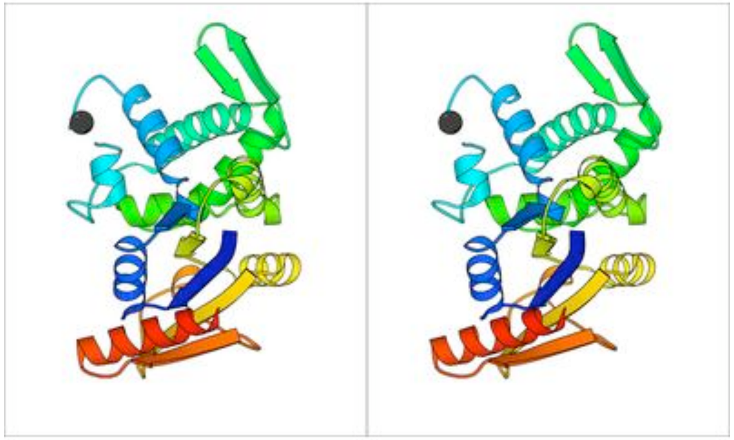


Domain boundaries can be seen in multiple sequence alignments if the alignments are of whole genes.



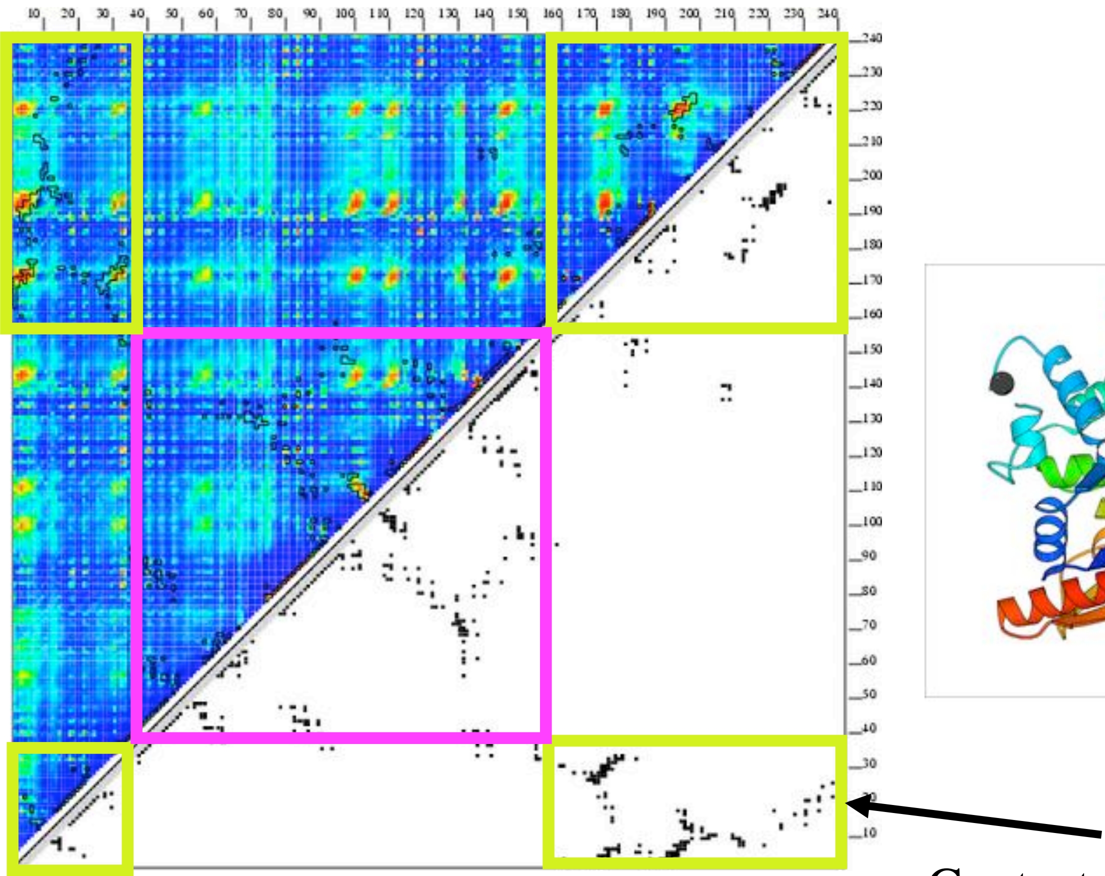
Domains can be seen in contact maps

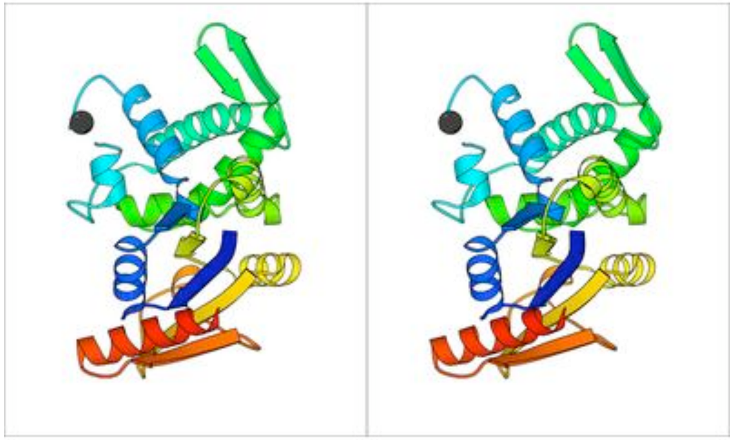




stereo image
(cross your eyes)

Domains can be seen in contact maps





Contacts are mostly within domains, not between domains. One domain consists of N and C-terminal parts

Exercise 3.2: domain boundary

1YY3 (easy)

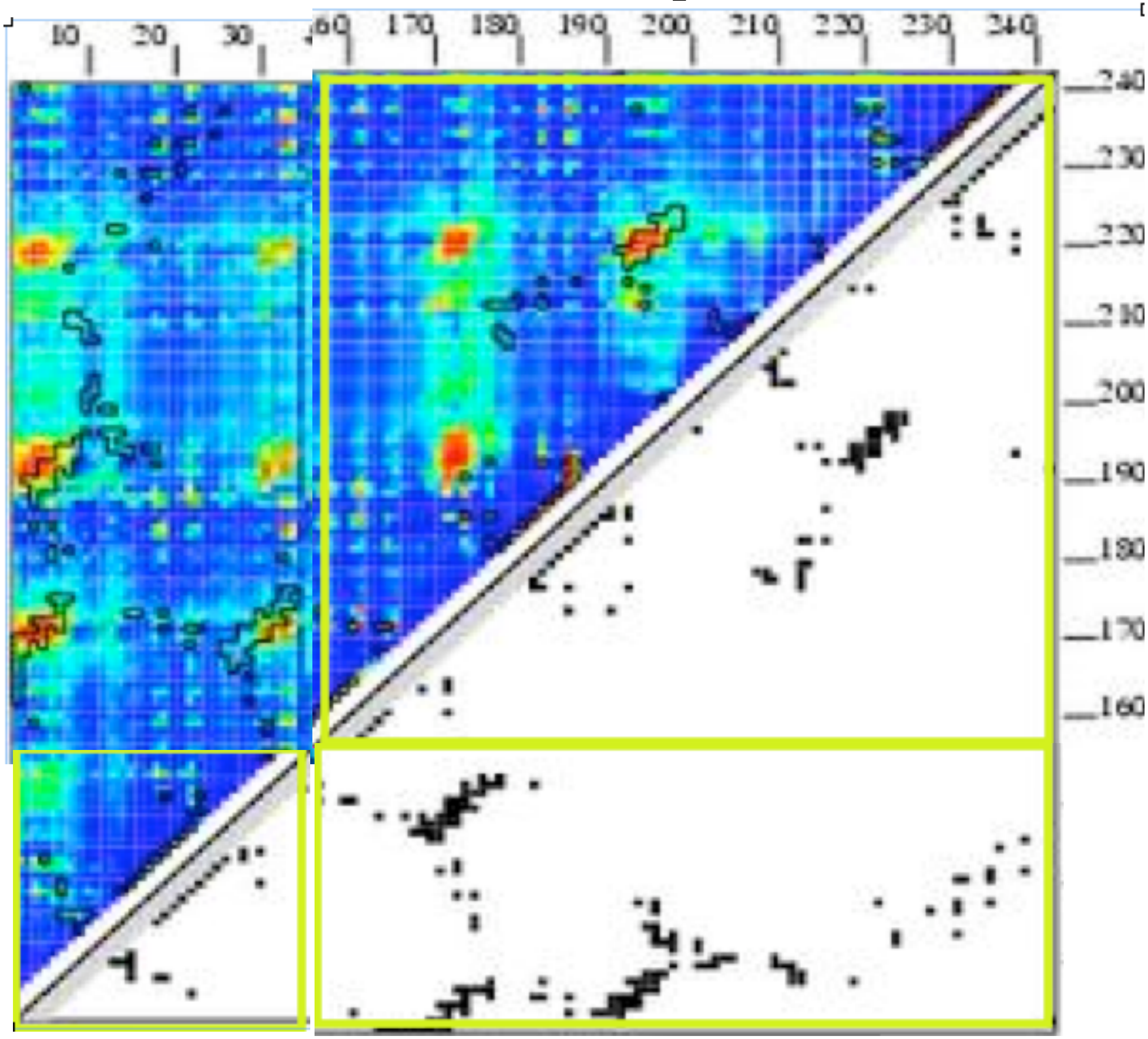
1IO1 (easy)

1G71 (hard)

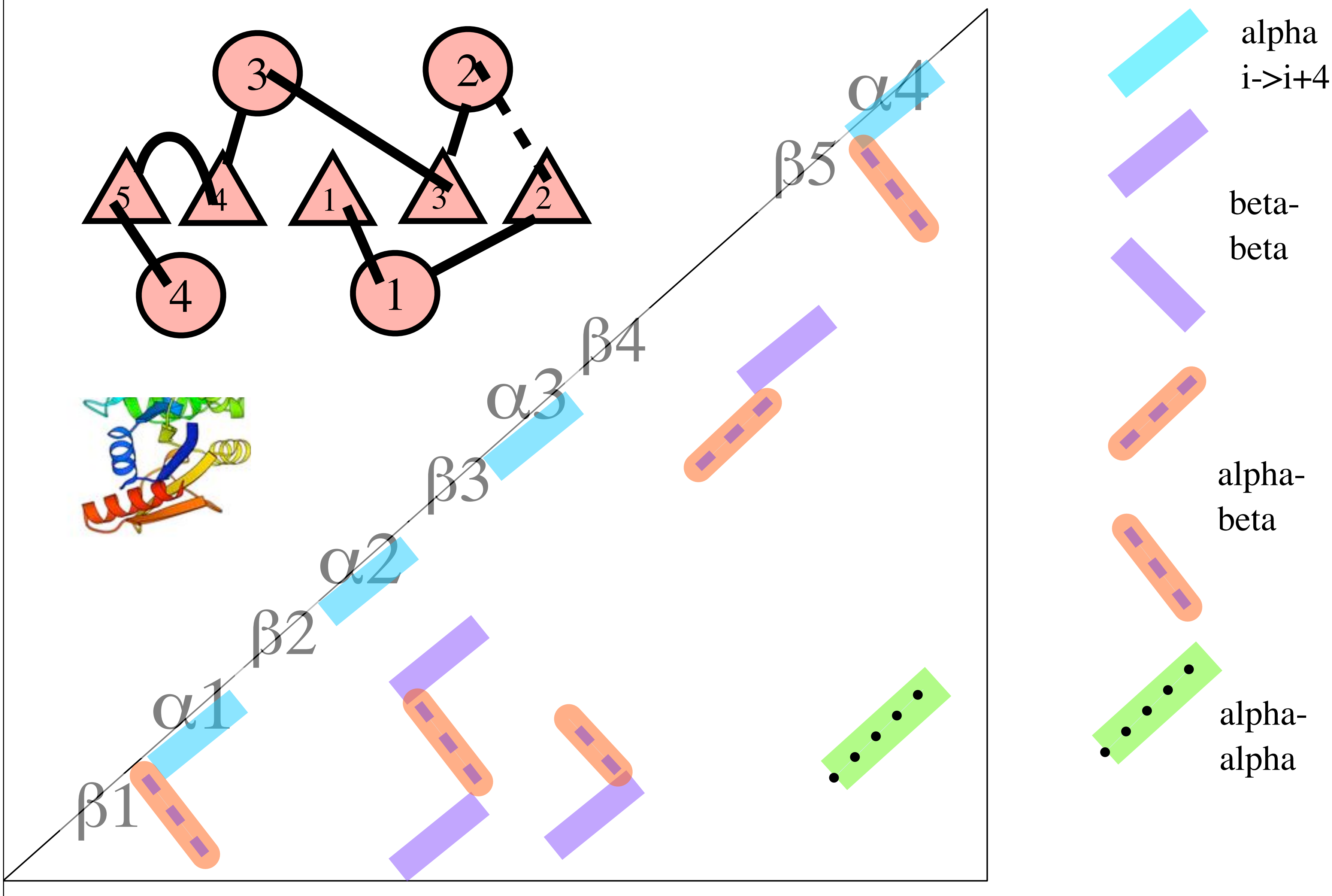
Retrieve using **File/Protein database**.

Use various renderings in MOE to find the points of division within the structure.

Terminal domain, cut-and-paste



Crude contact map to TOPS



Exercise 3.3: TOPS from contact map