Protein Structure Databases and Classification

- SCOP, CATH classification schemes, what they mean.
- Motifs: classic turn types. Extended turn types.
- TOPS: drawing a protein molecule

The SCOP database

- Contains information about classification of protein structures and within that classification, their sequences
- Go to http://scop.berkeley.edu

SCOP classification heirarchy

(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

protein classes

1. all α (126) ← number of sub-categories
2. all β (81)
3. α/β (87)
4. α+β (151)
5. multidomain (21)
6. membrane (21)
7. small (10)
8. coiled coil (4)
9. low-resolution (4) ← possibly not complete, or erroneous
10. peptides (61)
11. designed proteins (17)

class: α/β proteins

Mainly parallel beta sheets (beta-alpha-beta units)

Folds:

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, mostly.

fold: flavodoxin-like

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”

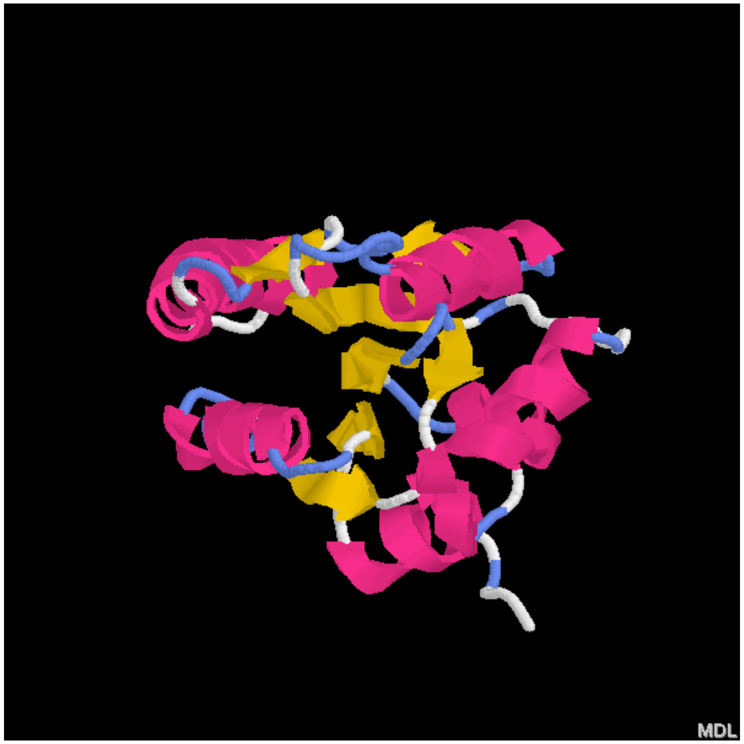
These are not domains. No implication of structural independence.

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

fold-level similarity

common topological features

catalase

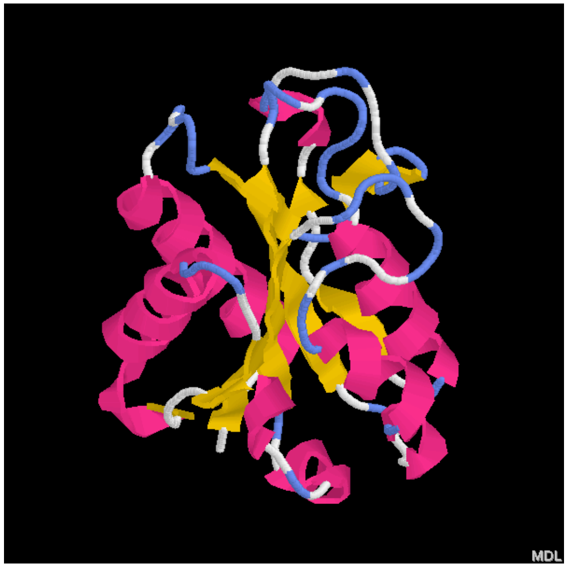


flavodoxin

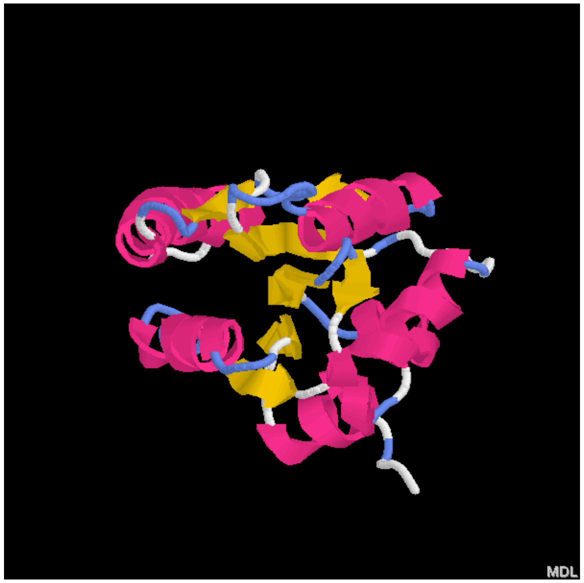
At the fold level, a common core of secondary structure is conserved. Outer secondary structure units may not be conserved.

Superfamily: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. Sec struct's align 1-to-1, mostly.

Family: quinone reductases

binds FAD

Proteins:

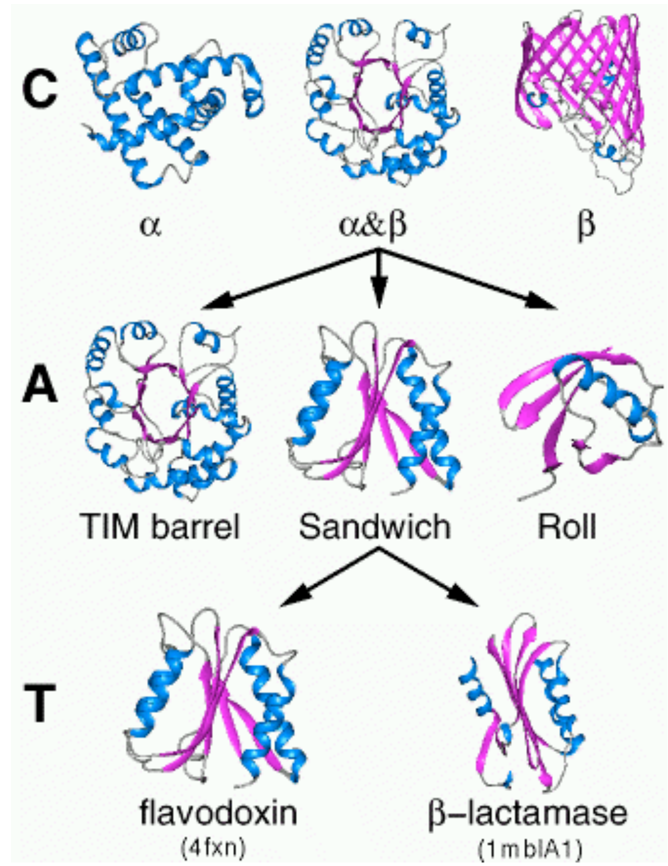
quinone
reductase type 2

NADPH quinone
reductase

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

CATH

- Class
- Architecture
- Topology
- Homology

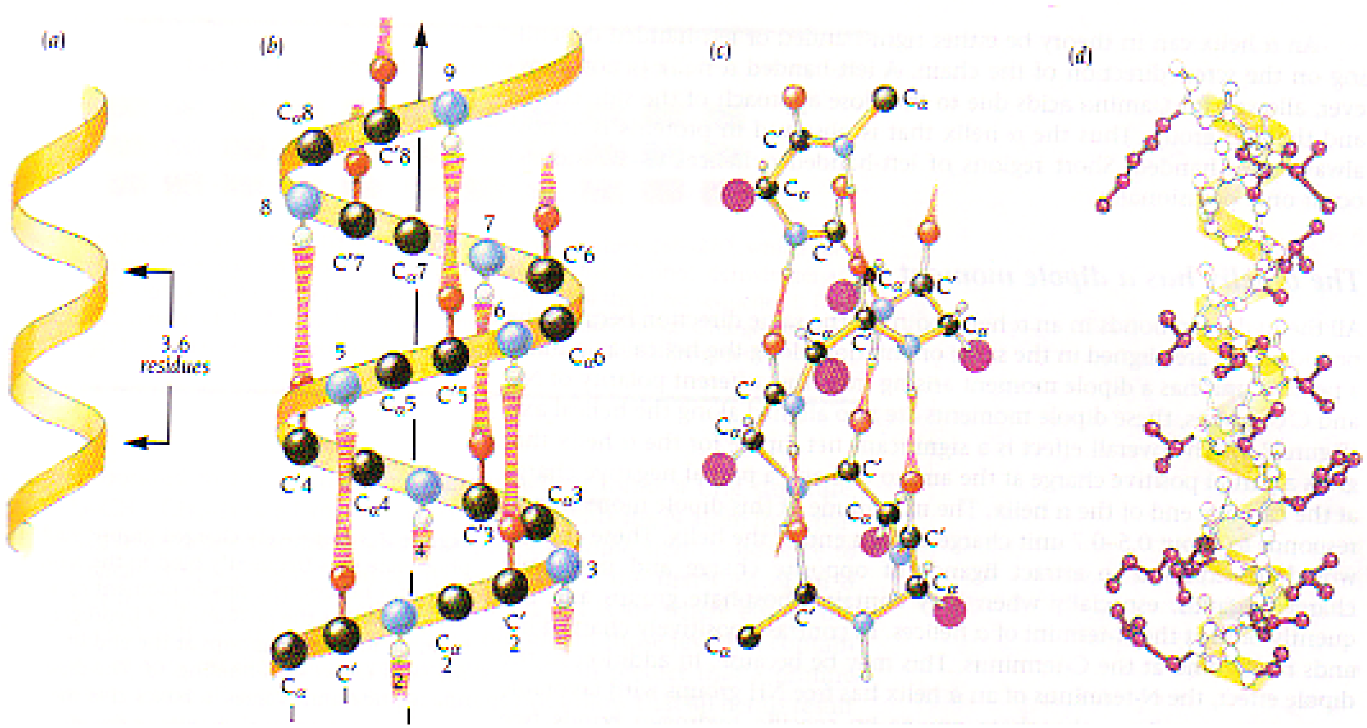


http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html

Structural heirarchy of proteins

- Primary structure
- Secondary structure
- Local structure
- super-secondary structure
- domains, folds
- Global, multi-domain (tertiary structure)
- Quaternary structure

Secondary structure



Alpha helix

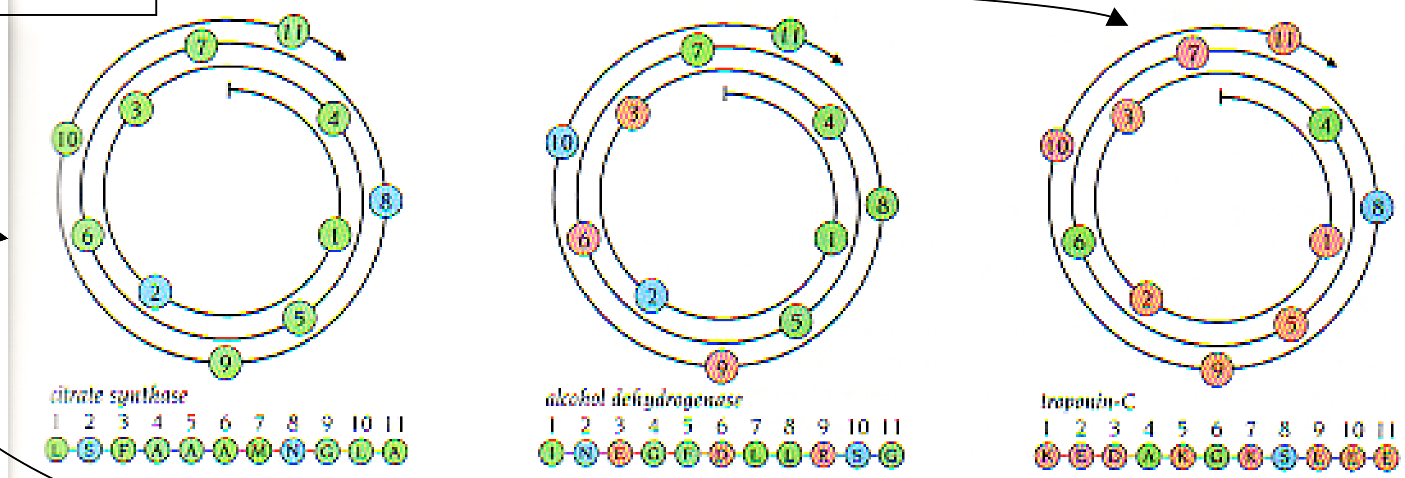
Right-handed
3.6 residues/turn
i->i+4 H-bonds

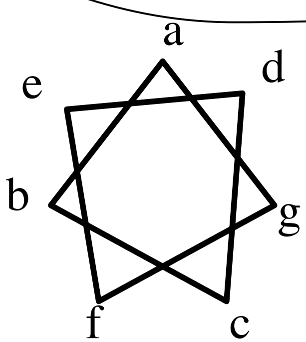
Overall dipole N+->C-

3 types of Alpha helix

Table 2.1 Amino acid sequences of three α helices

1.	non-polar	Leu	Ser	Phe	Ala	Ala	Ala	Met	Asn	Gly	Leu	Ala
2.	amphipathic	Ile	Asn	Glu	Gly	Phe	Asp	Leu	Leu	Arg	Ser	Gly
3.	polar	Lys	Glu	Asp	Ala	Lys	Gly	Lys	Ser	Glu	Glu	Glu



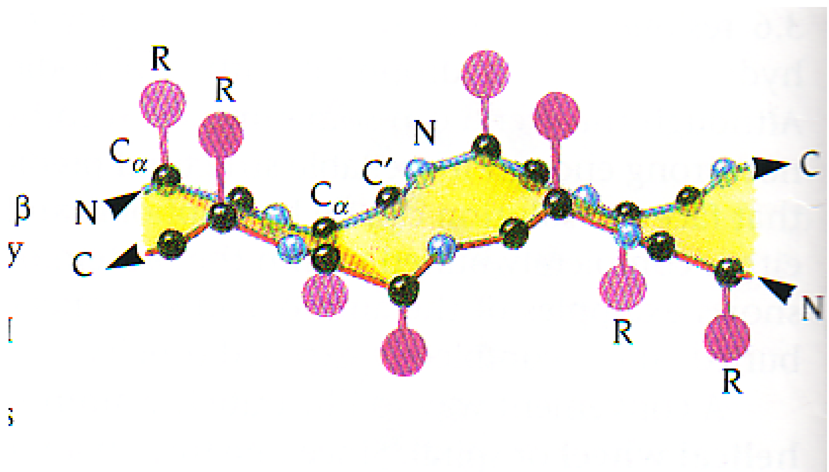


Two ways to display position of sidechain on a helix.

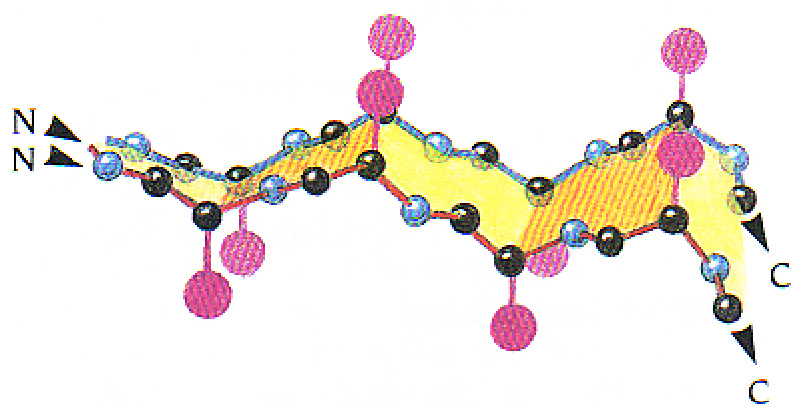
For amphipathic and non-polar, sidechains line up in a favorable way.

beta-strand

Antiparallel:

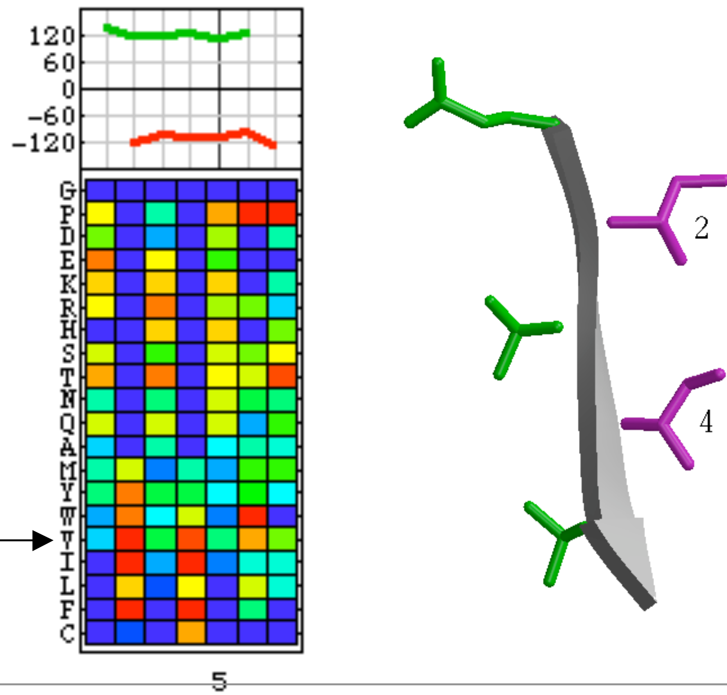
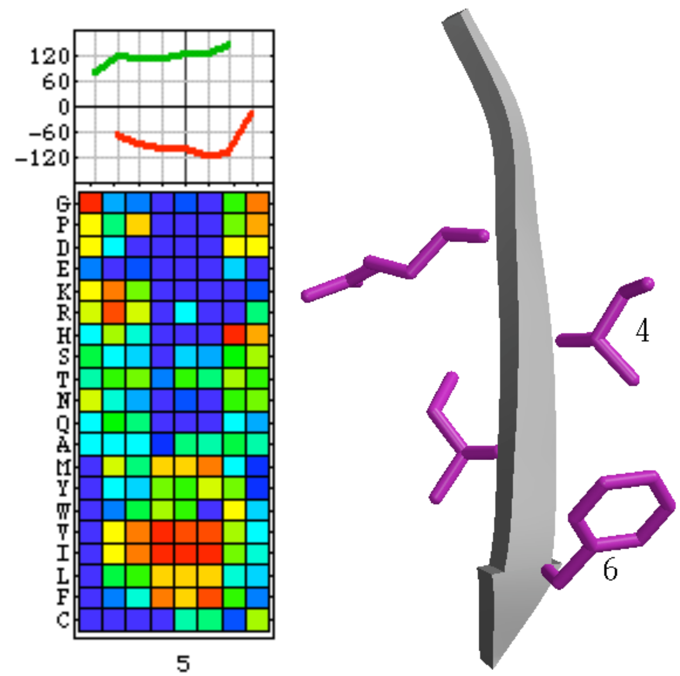


Parallel:

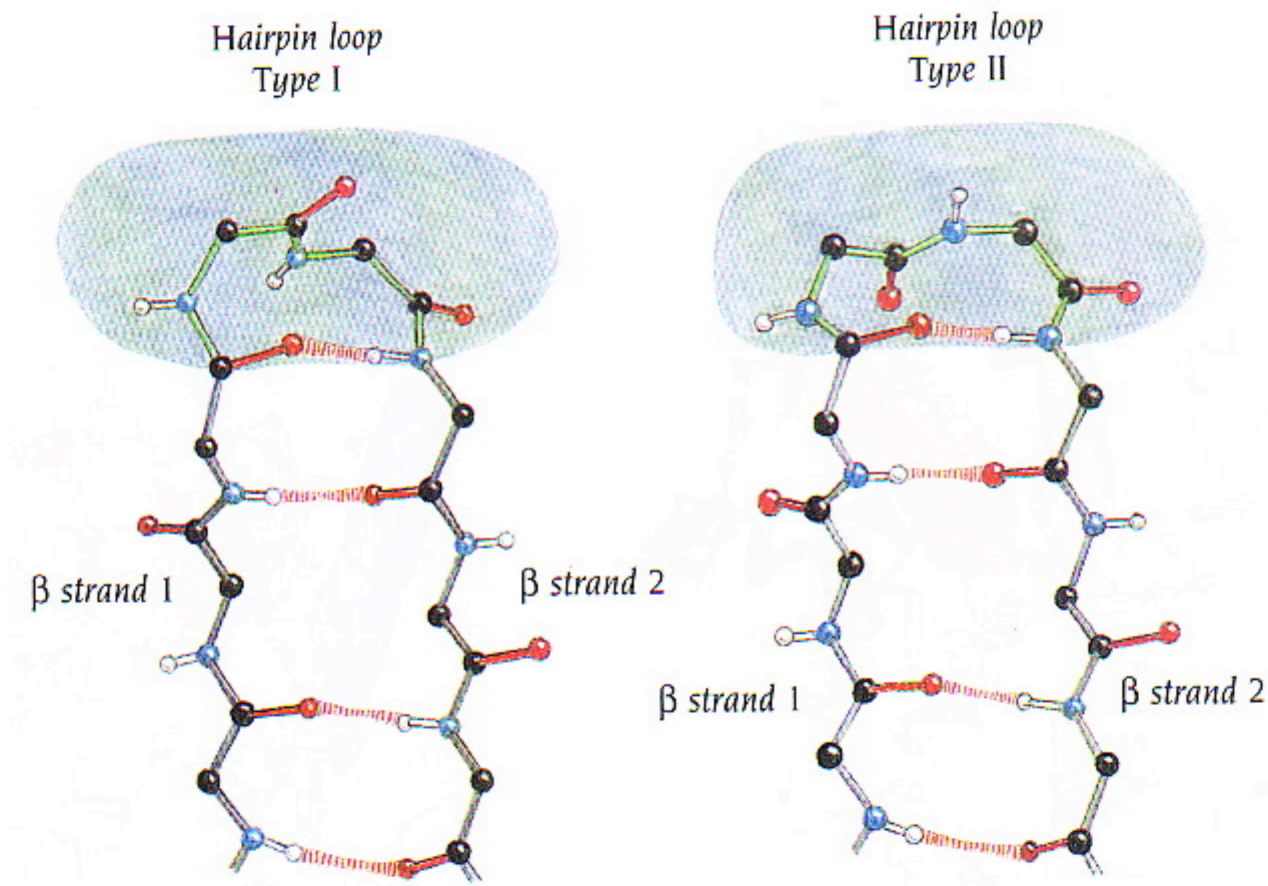


Two types of beta strand

Note preference for beta-branched aa's: I,V,T

Amphipathic	Hydrophobic
	
Found at the edges of a sheet, or when one side of the sheet is exposed to solvent (i.e. 2-layer proteins).	Found in the buried middle strands of sheets in 3-layer proteins.

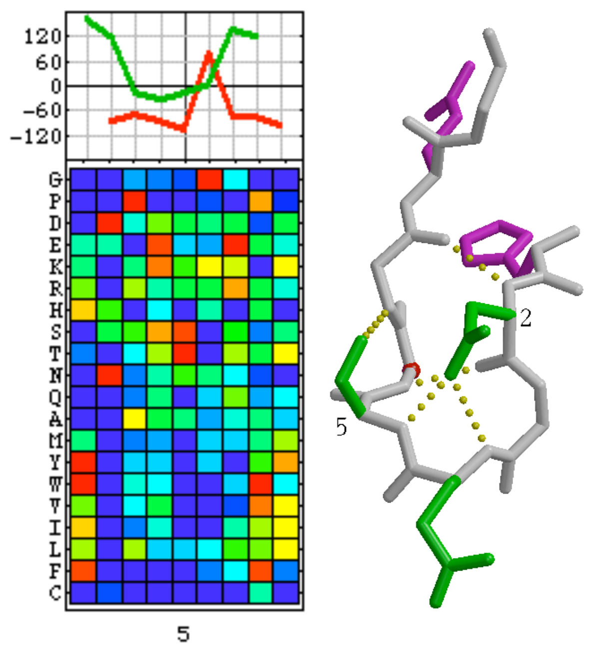
Local structure: beta hairpins



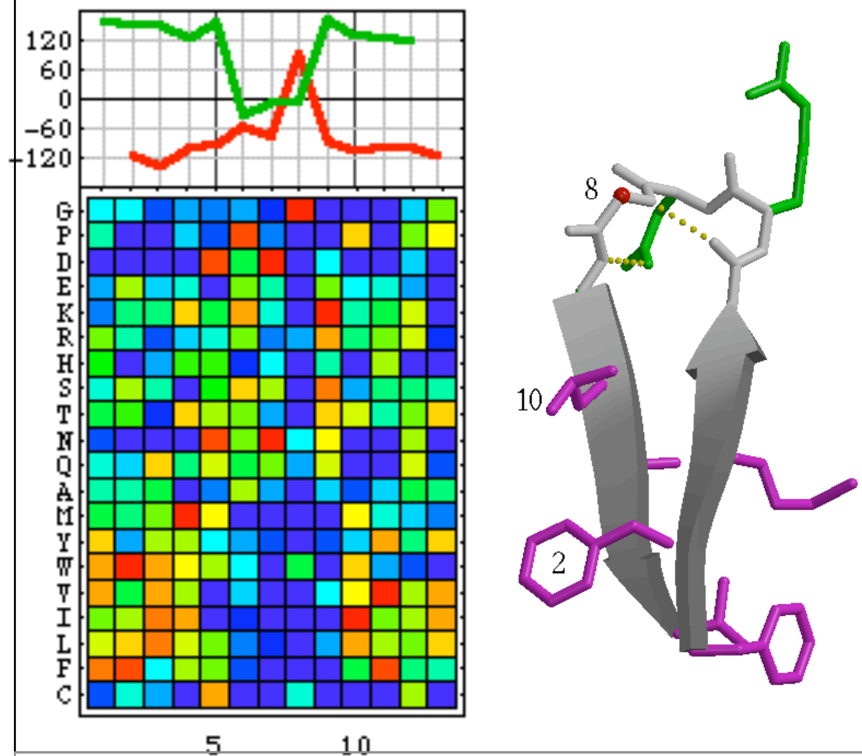
Two adjacent antiparallel beta strands = a beta hairpin

Shown are “tight turns”, 2 residues in the loop region (shaded). Hairpins can have as many as 20 residues in the loop region.

hairpin sequence motifs

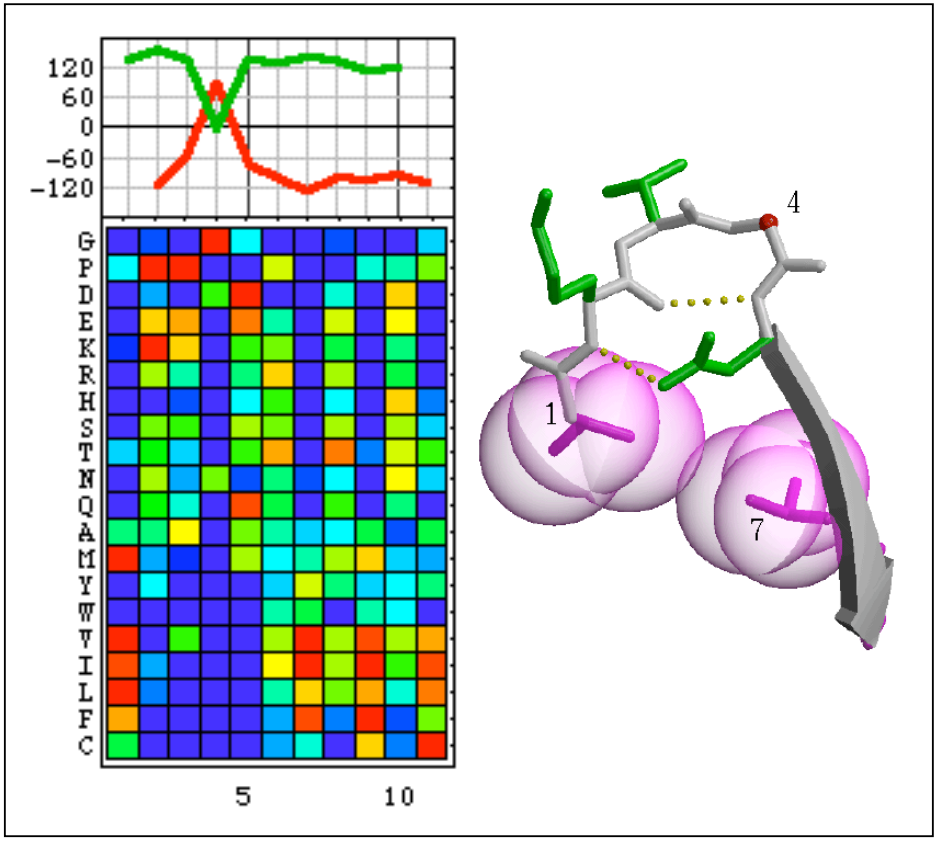


“Serine beta-hairpin” (also called an “alpha turn”). A specific pattern (DPESG) forms an alpha-helical turn 4-residues long.



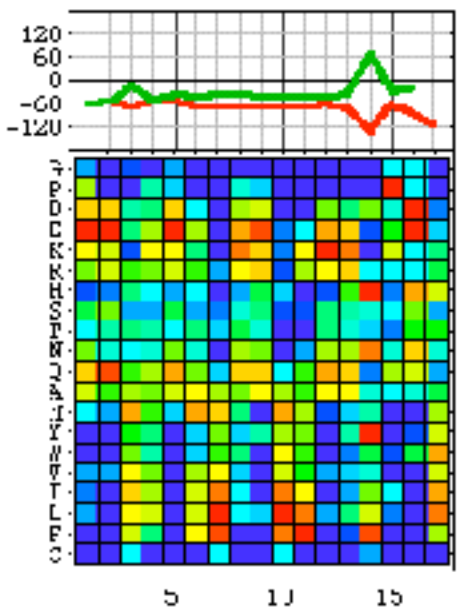
“Extended Type-1 hairpin”. A type-1 “tight turn” has only 2 residues in the turn. This motif, more common than the tight turn, has an additional Pro or polar sidechain. Pattern: PDG.

diverging turn motif



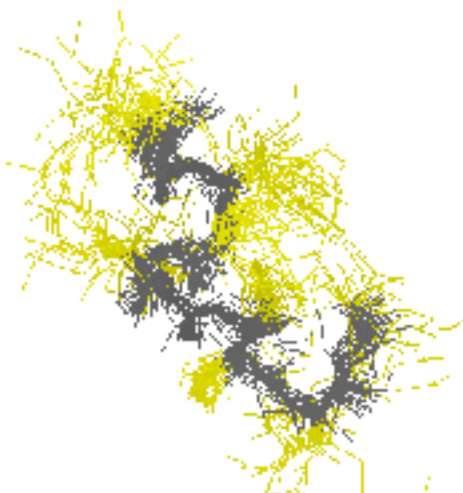
"Diverging turns" have a Type-2 beta turn and two strands that do not pair. The consensus sequence pattern is PDG. The residue before G can be anything polar, but not a D or an N.

Proline helix C-cap motif

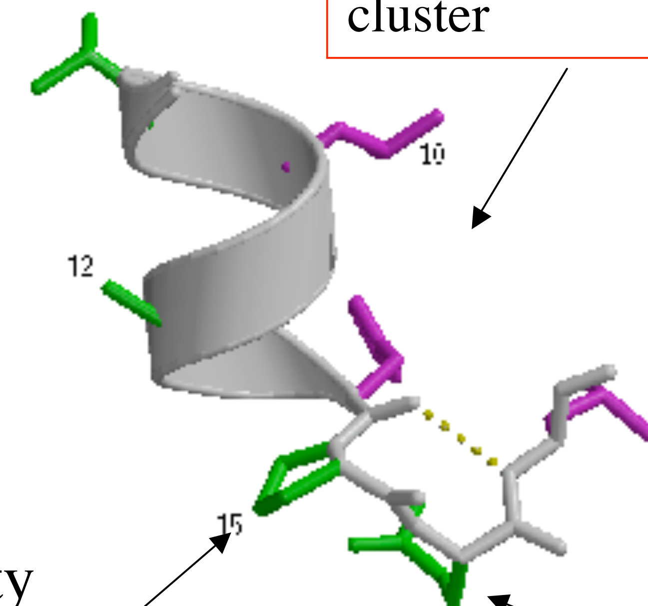


Sequence pattern=
...nppnnpp[HNYF]P[DE]n

“n”=non-polar
“p”=polar
[...]=alternative aa’s



structural variability



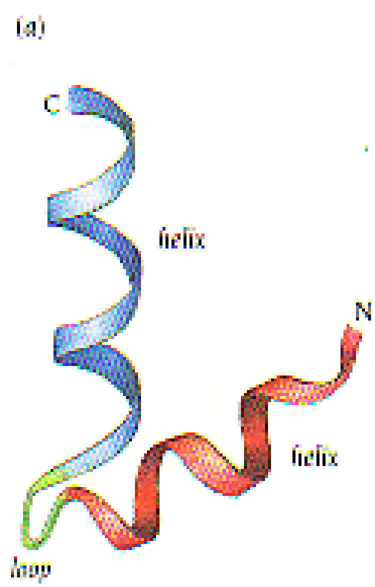
note:
hydrophobic
cluster

Pro blocks helix

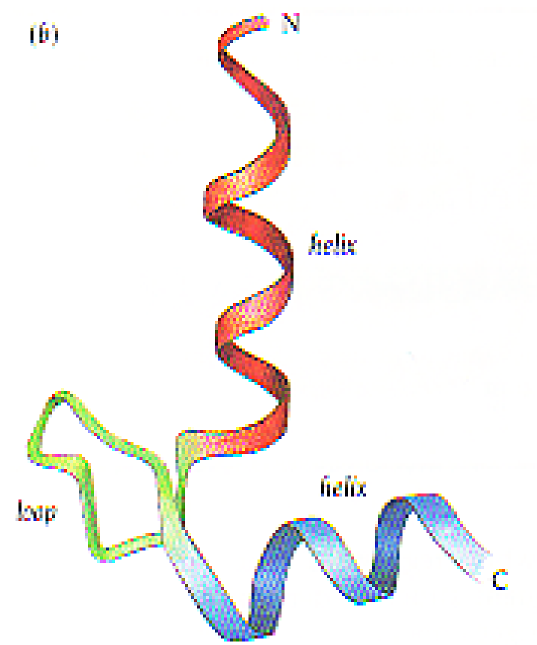
D or E stabilizes
tight turn

Locations of non-polar (magenta) and polar (green) sidechains

Supersecondary: Two Helix motifs



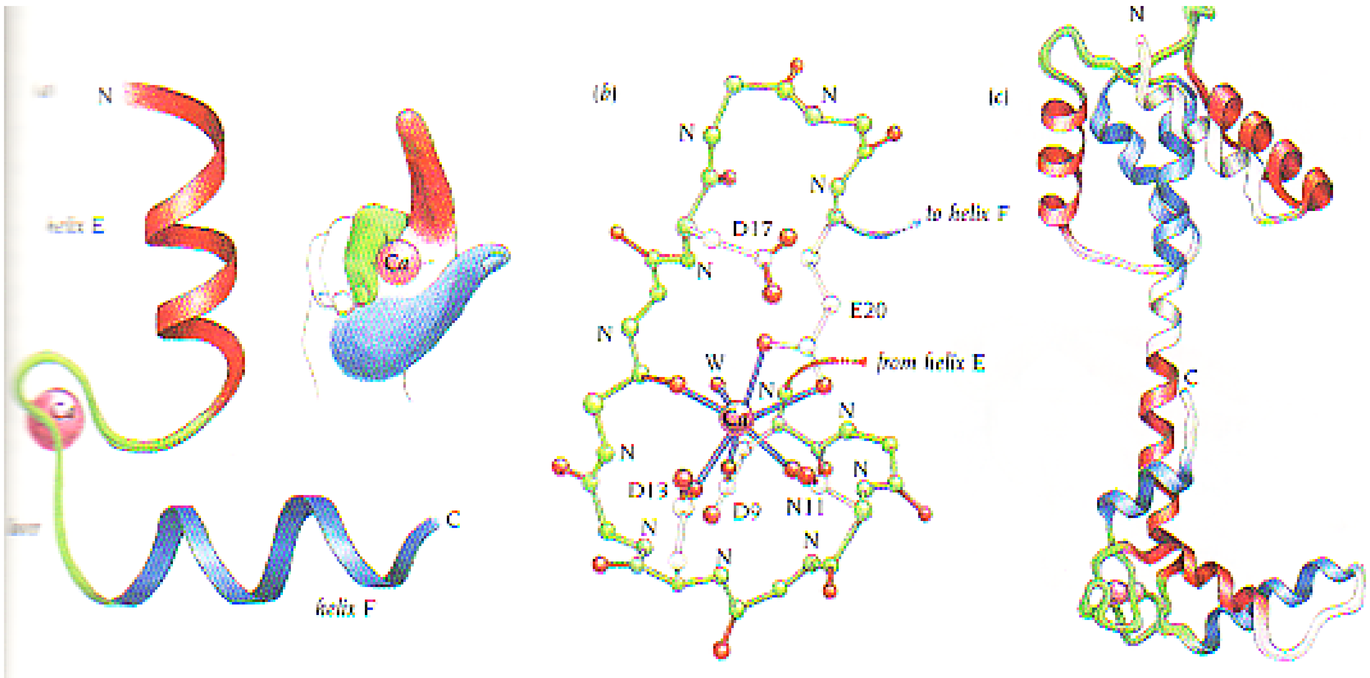
α–α corner
(helix-turn-helix)



EF-hand

(binds Ca^{2+})

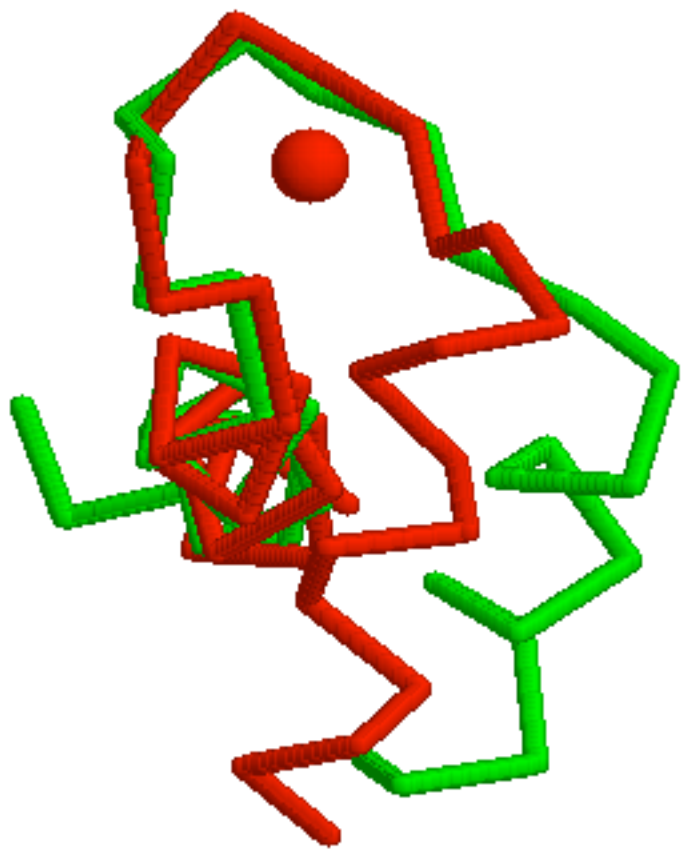
The EF-hand



N
helix E
Ca
C
helix F
(b)
N
to helix F
D17
E20
from helix E
W
Ca
D13
D9
N11
(c)
N
C

Supersecondary motifs are plastic.

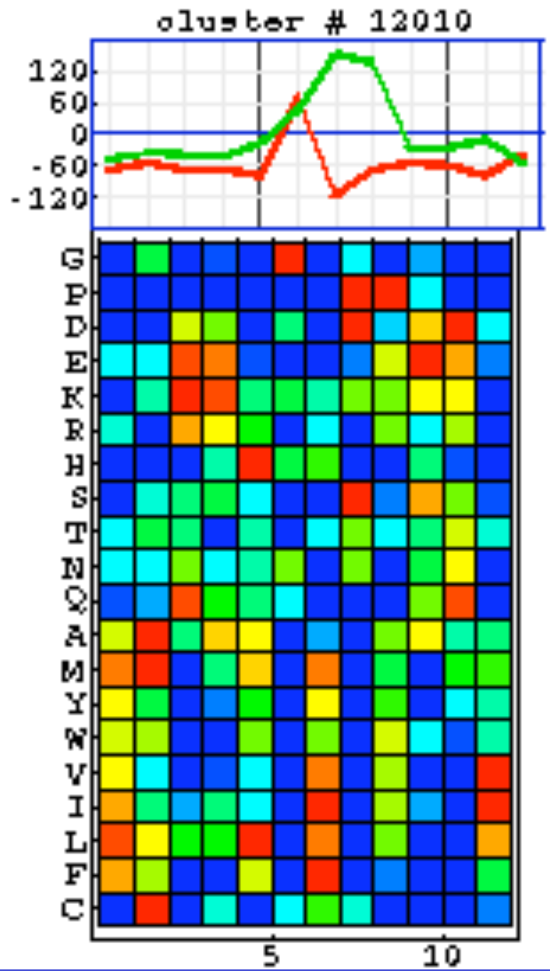
red=1CLL



green=1cfd

Without Ca2+ bound, the helices have more contact. Backbone angles do not change very much when Ca2+ binds.

α–α corner motif



backbone angles:
green=psi
red=phi

red=favorable
blue = unfavorable
green=neutral

"n"=non-polar
"p"=polar
[...]=alternative aa's

motif pattern (summary):

nnpp[nH]Gn[PDS][Px]pn

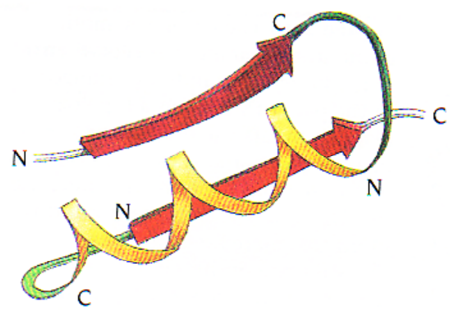
Can you see the α–α motif in the sequence?

```
  1 GLSDGEWQQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDKFKHLK
    ___HHHHHHH HHHHHHHHHT HHHHHHHHHH HHHHHTHHHH HTTTTTTT__

 51 TEAEMKASED LKKHGTVVLT ALGGILKKKG HHEAELKPLA QSHATKHKIP
    SHHHHHTTHH HHHHHHHHHH HHHHHHHTTT __HHHHHHHH HHHHHTS___

101 IKYLEFISDA IIHVLHSKHP GDFGADAQGA MTKALELFRN DIAAKYKELG
    HHHHHHHHHH HHHHHHHHST TSS_HHHHHH HHHHHHHHHH HHHHHHHHHT

151 FQG
```

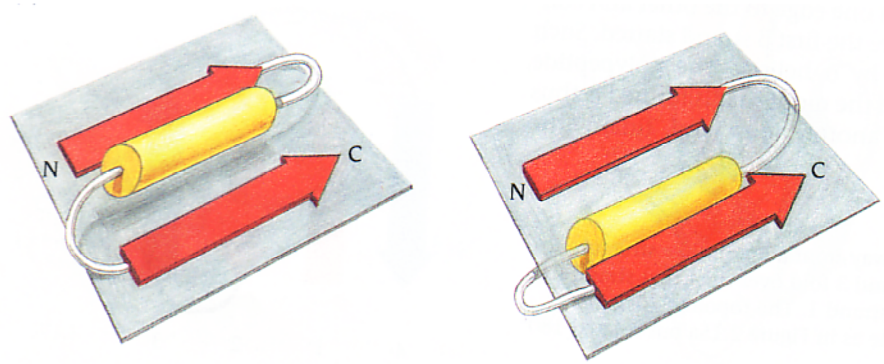
motif pattern (summary):

nnpp[nH]Gn[PDS][Px]pn

βαβ motif

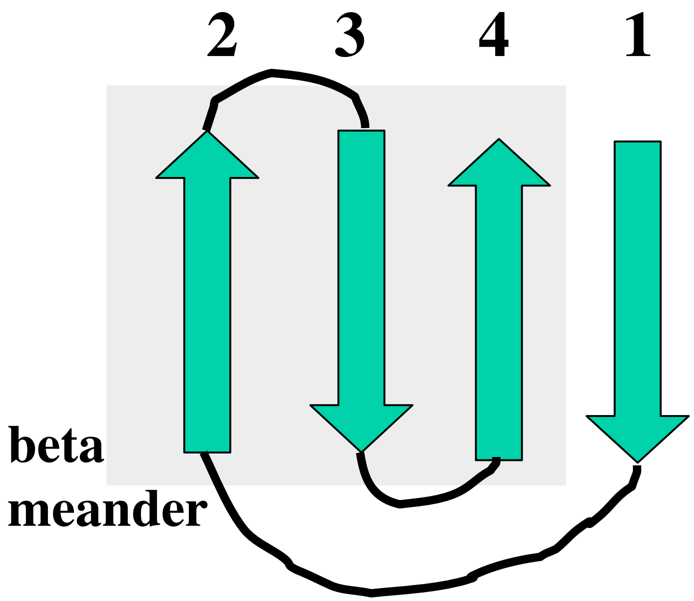


When a helix occurs between two strands, they are often paired in a **parallel** sheet.



The cross-over from one strand to the next is almost always right-handed, possibly for energetic reasons, possibly for kinetic reasons.

Greek key motif



One of the most common arrangements of four strands.

Two permutations. 2341 and 3214

Exercise: Download 2PLT.
Find the Greek key motifs.

TOPS topology cartoons

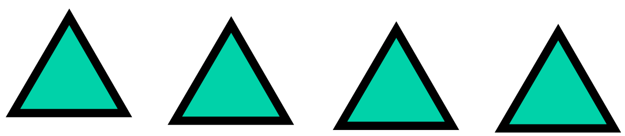
A simple way to draw a protein

beta strand pointing up

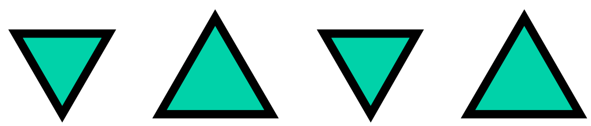
beta strand pointing down

alpha helix

connections



A parallel beta sheet

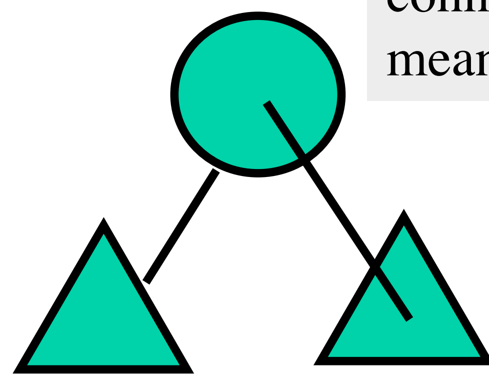


An anti- parallel beta sheet

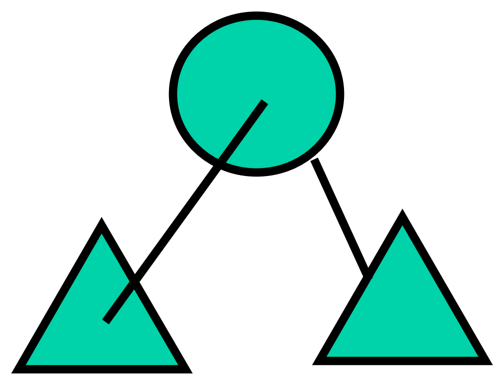
TOPS topology cartoons

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

A right-handed βαβ unit

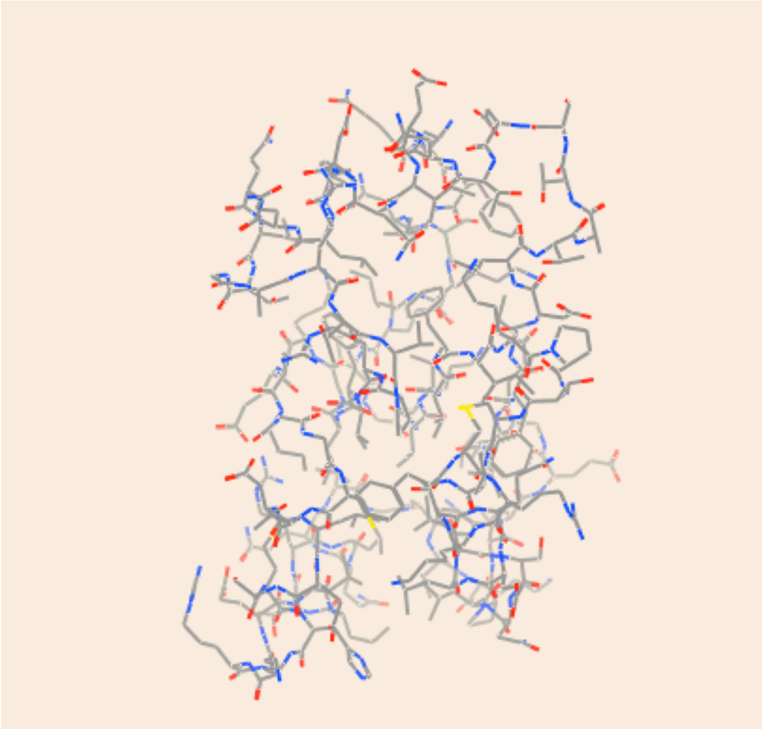


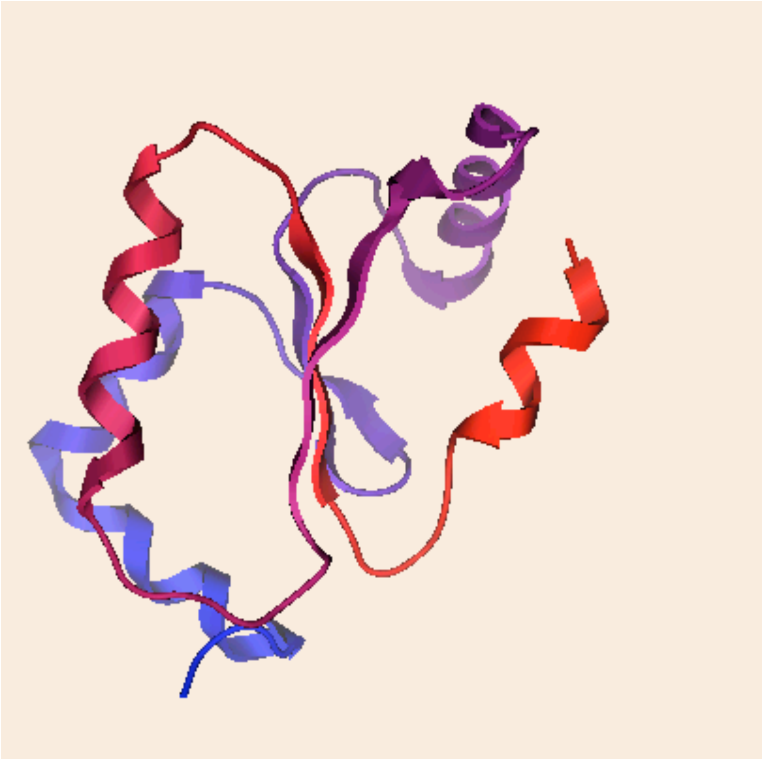
A left-handed βαβ unit
(rarely seen)



How to draw TOPS

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

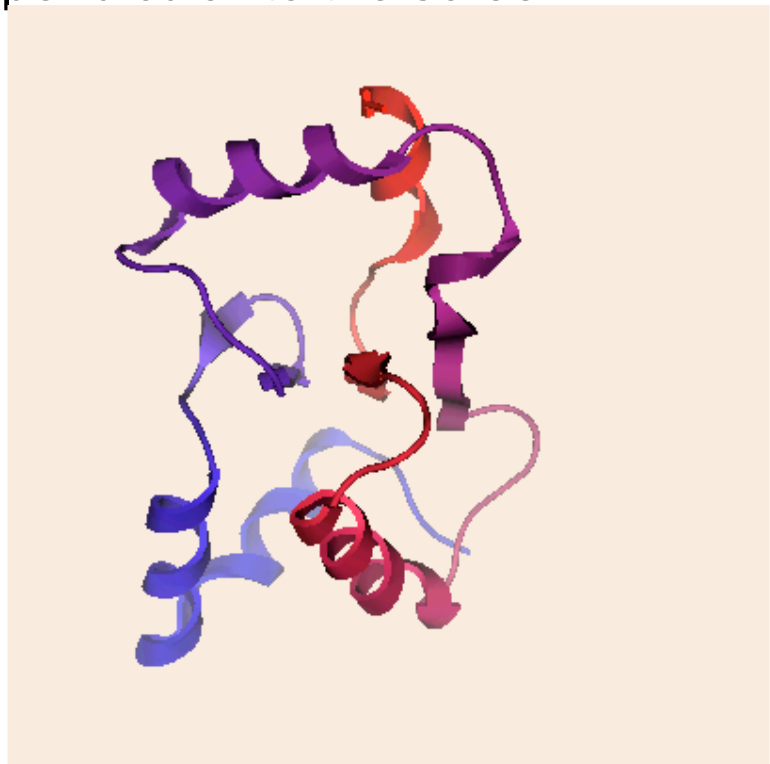




How to draw TOPS

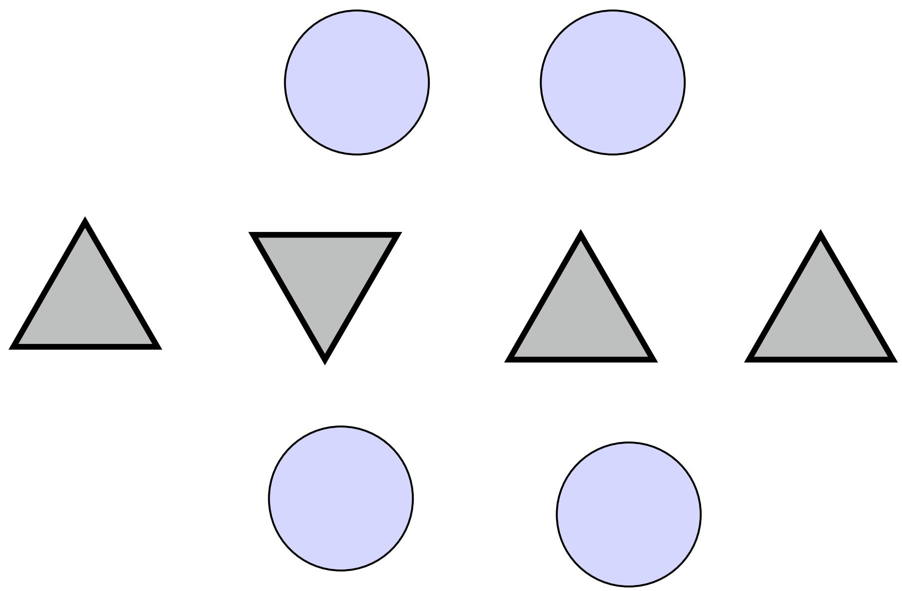
Line up the molecule along the beta sheet, if present.

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



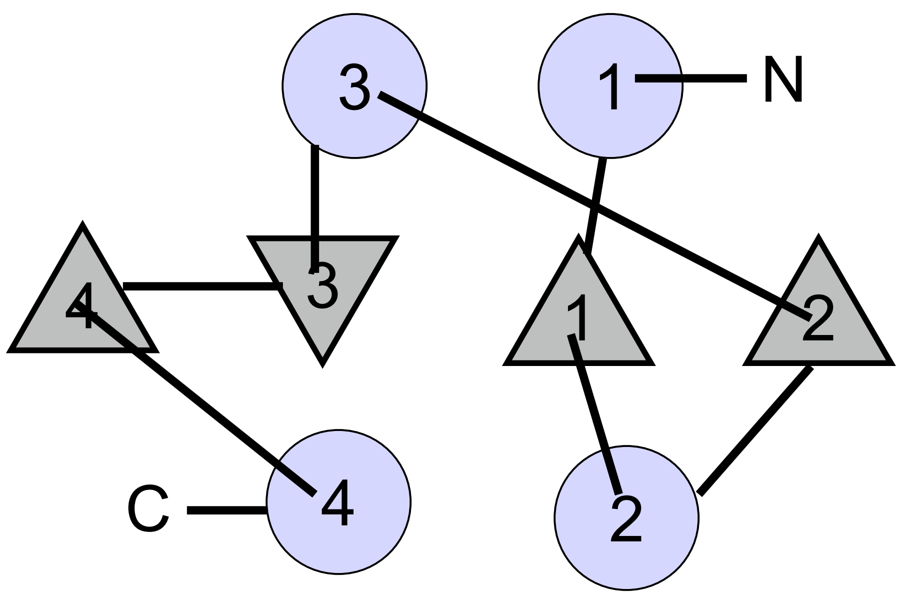
TOPS diagram

- Draw secondary structures first.



TOPS diagram

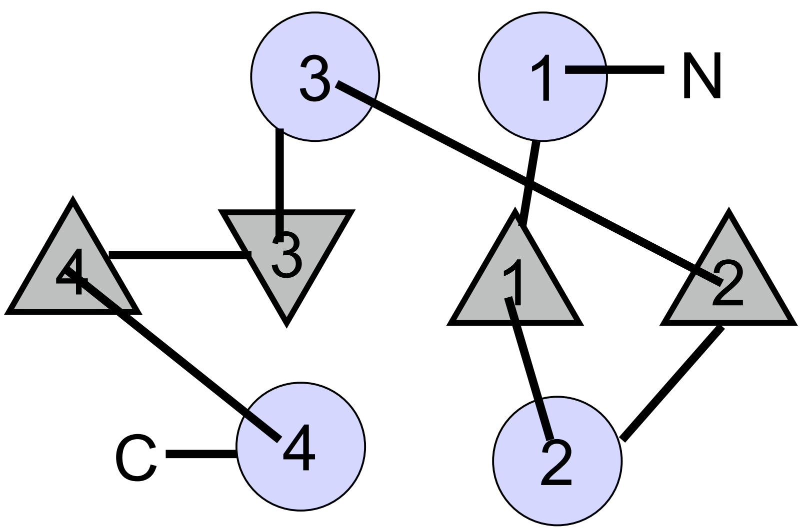
- number them and connect them

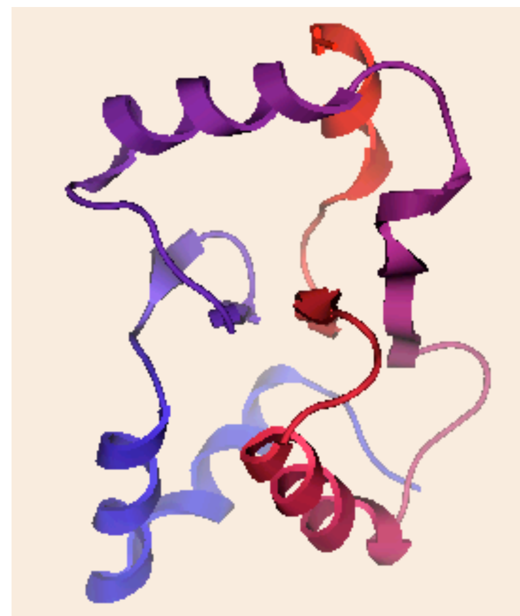


Be careful to draw connections to the center or side, when it is in front or in back, respectively.

SCOP-style naming

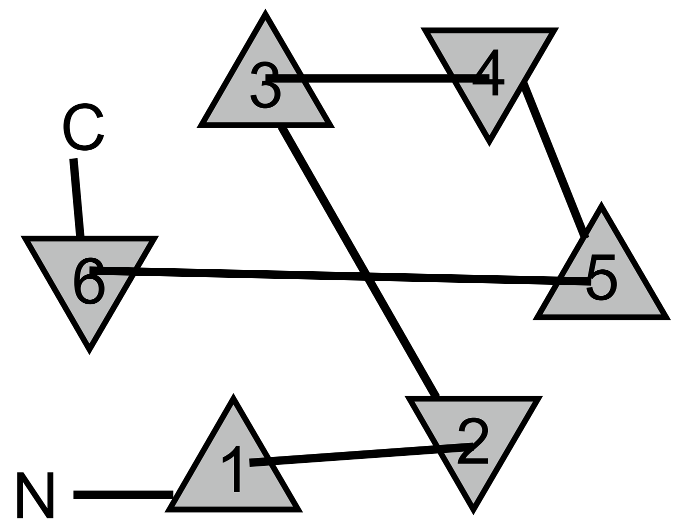
- 3 layers, 2-4-2 αβα, mixed sheet, 4312



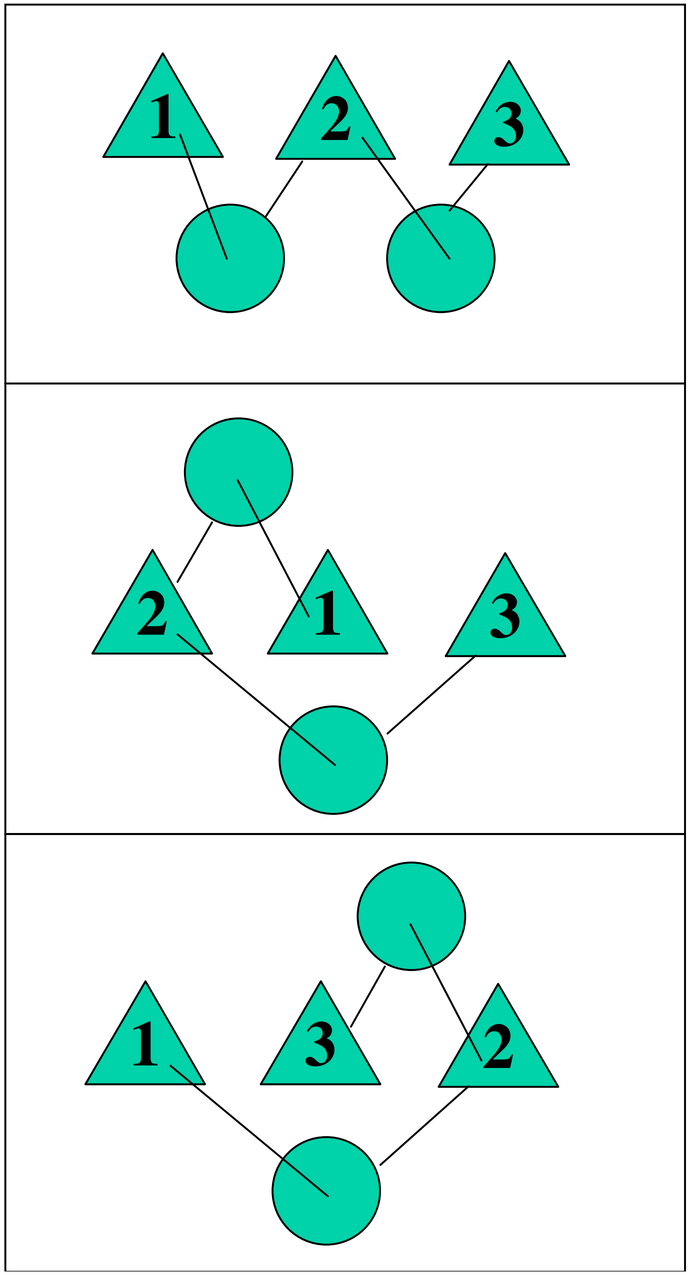


SCOP-style naming

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



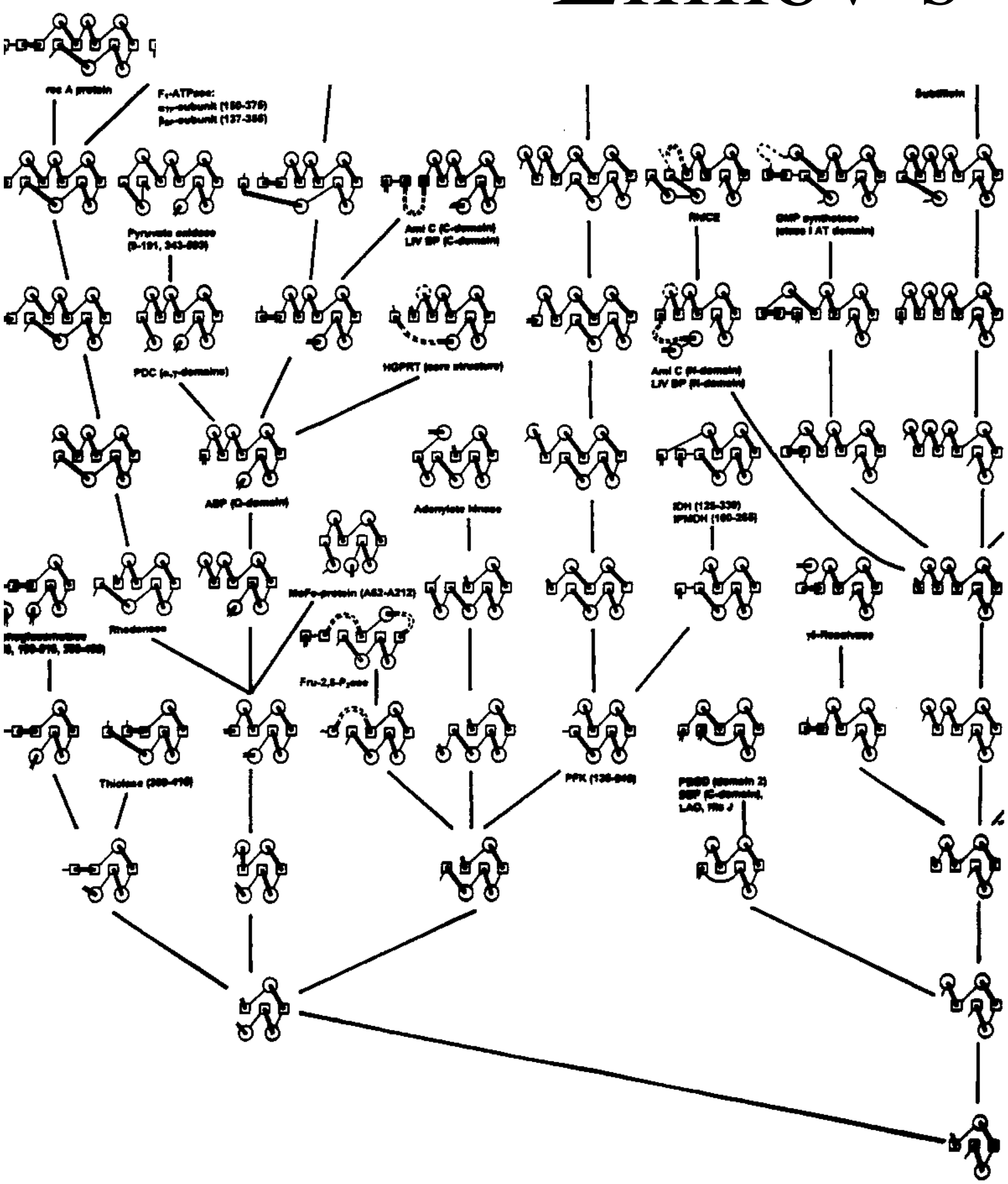
βαβαβ motif



Only 3 are possible.

(with R-handed crossovers)

Efimov’s 7 trees



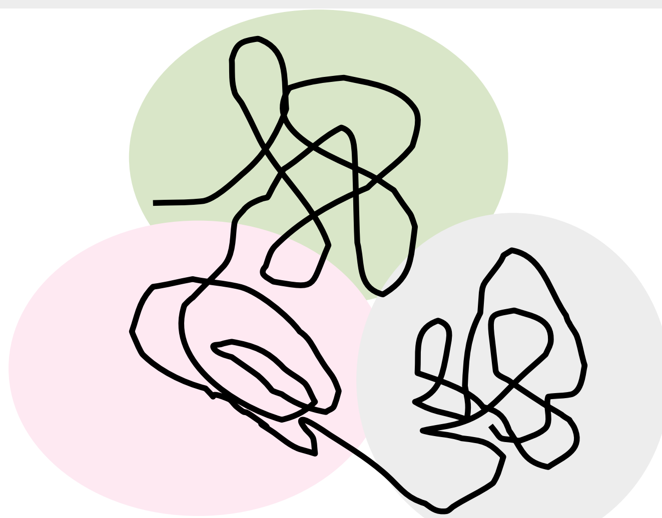
A. Efimov showed that almost all protein structures can be classified as being one of 7 trees, each starting with a motif and “growing” by one secondary structure unit at time.

Does structural phylogeny recapitulate folding?

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.