Molecular Modeling 2018 -- Lecture 8

Local structure

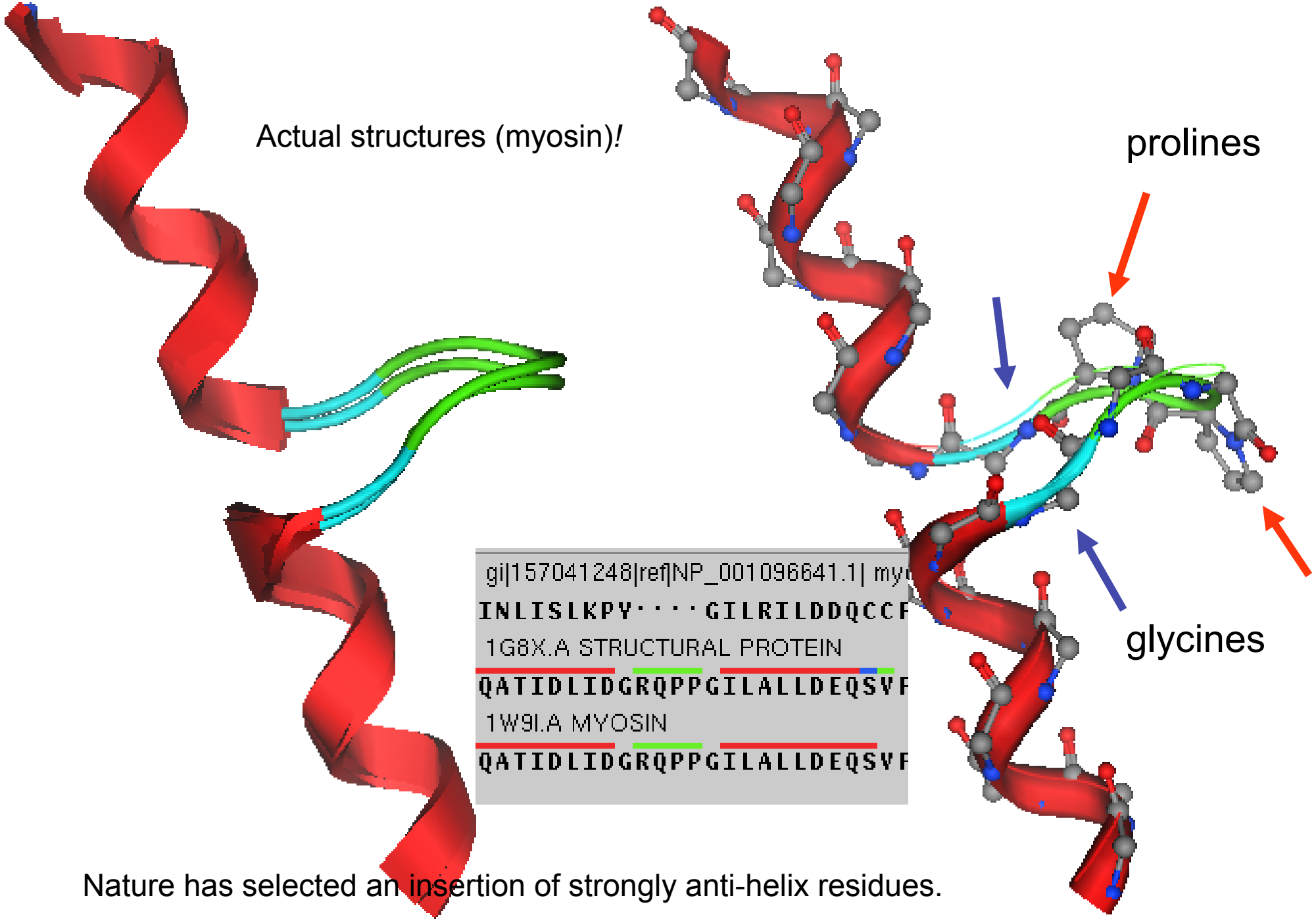
Database search

Multiple alignment

Automated homology modeling

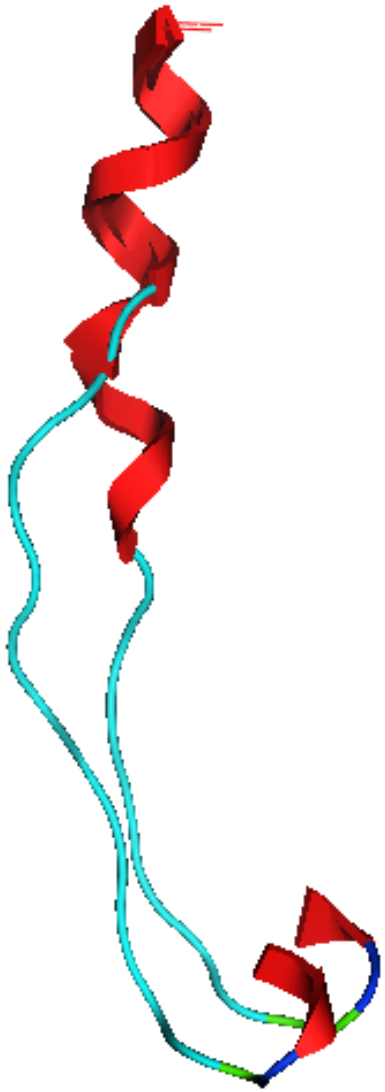
An exception to the *no-insertions-in-helix* rule

Actual structures (myosin)*!*



Nature has selected an insertion of strongly anti-helix residues.

Not an exception to the *no-deletions-in-helix* rule



Deletion in the sequence leads to shorter helix, not a helix with a "*deletion*" in it.

This shows two actual structures. The one with the deletion is more extended, to span the distance.

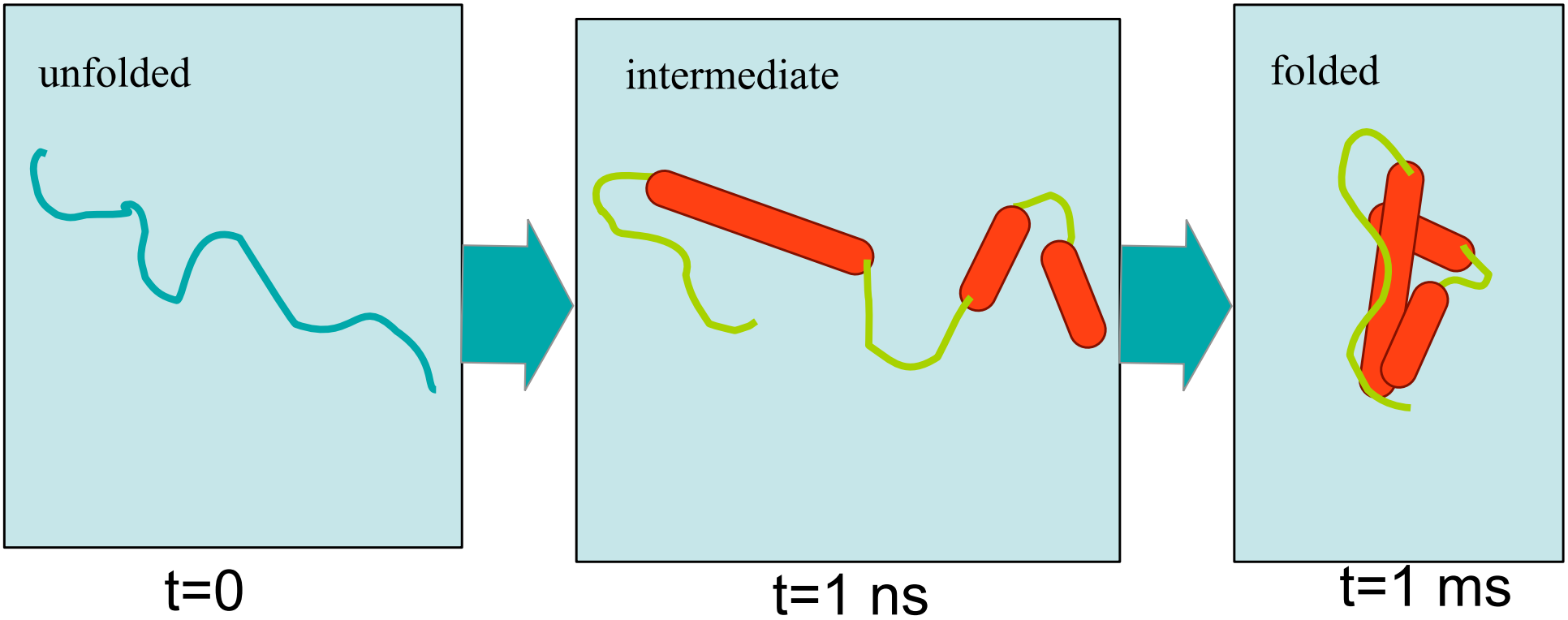
Alignment

```
AFADNQ·PCINLIS
TFIDFGLDSQATID
```

What is local structure?

Early in the process of folding (nsec timescale)

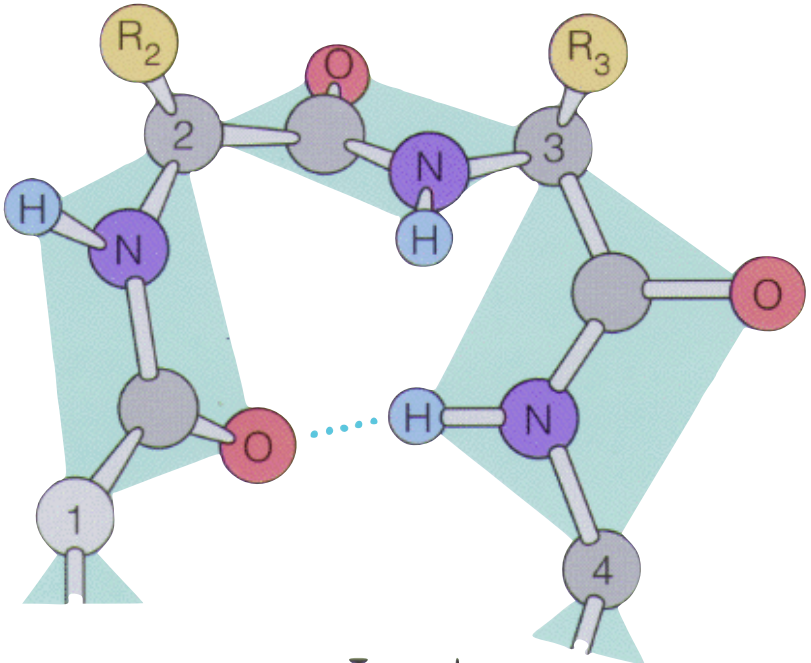
local structures form in the polypeptide chain which guide the formation of tertiary structure.



beta turns

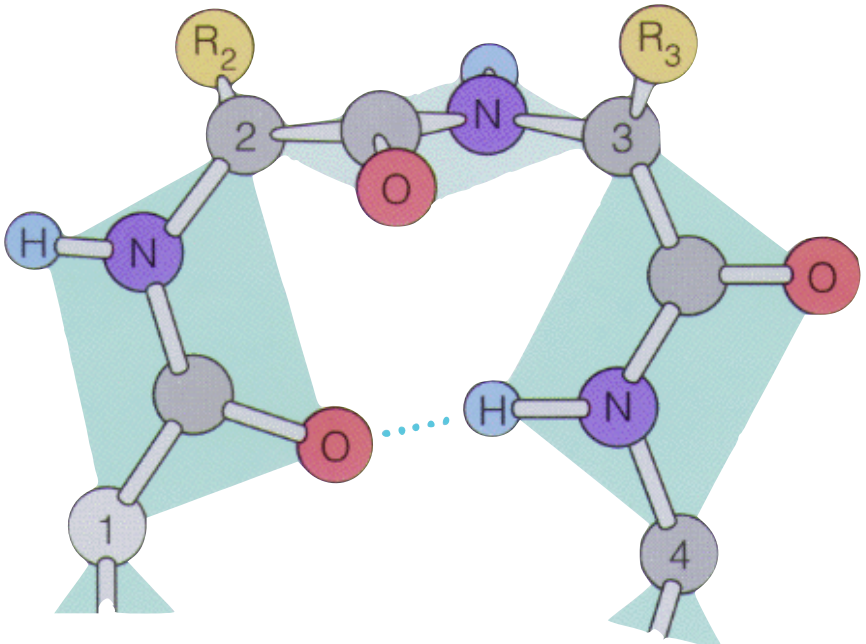
4-residues

Residue 1 hydrogen bonds to residue 4



Type I

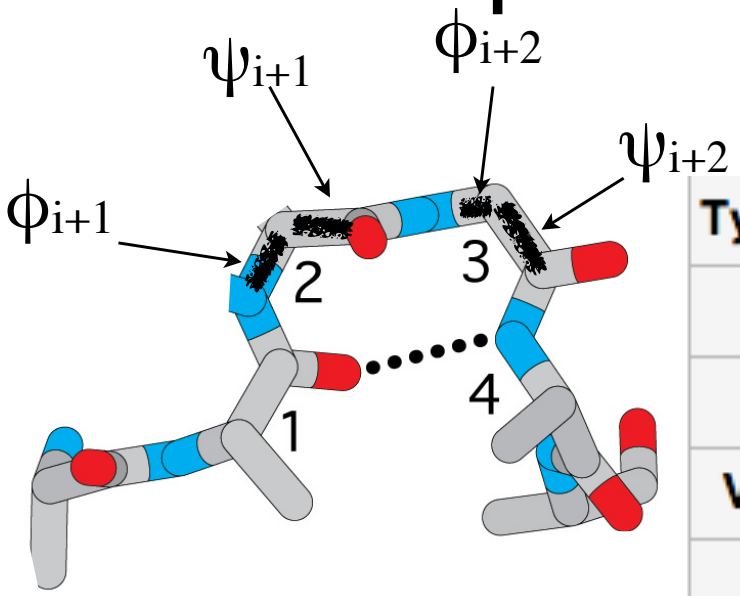
Type I (most common). Oxygen points away, viewed clockwise.



Type II

Type II (less common). Oxygen points toward, viewed clockwise.

Backbone angles and preferred sequence of beta turns



Glycine rules turn propensity

position / type \	1	2	3	4
I		P	D/N/S/T	**G**
II	P	P	**G**	
VIII	G/P	P		P
I'		G	**G**	
II'		**G**		

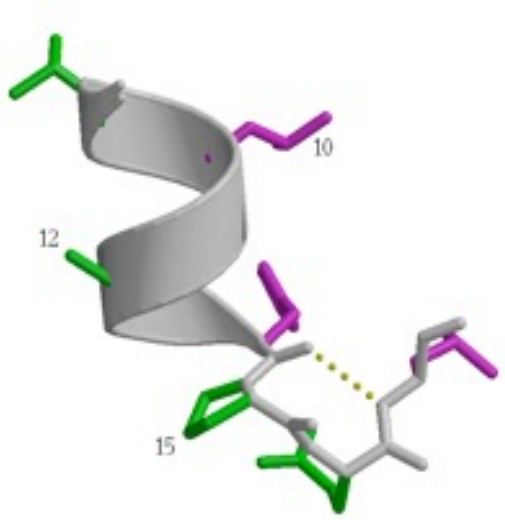
Backbone angles ±30°

Type	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
II	-60	120	80	0
VIII	-60	-30	-120	120
I'	60	30	90	0
II'	60	-120	-80	0
VIa1	-60	120	-90	0*
VIa2	-120	120	-60	0*
VIb	-135	135	-75	160*
IV	turns excluded from all the above categories			

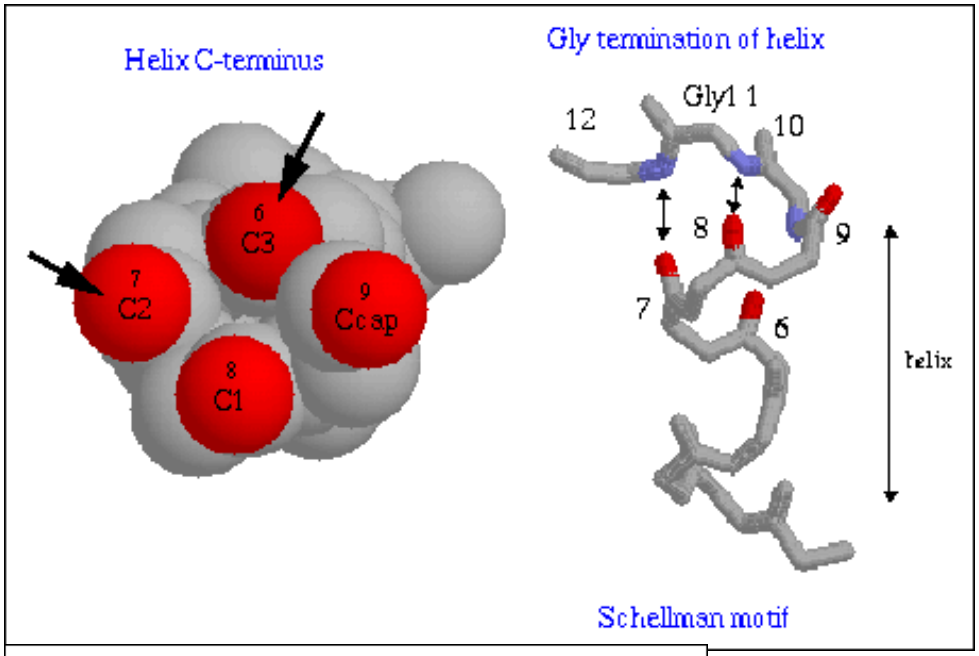
*have cis-peptide bond at $i+2$

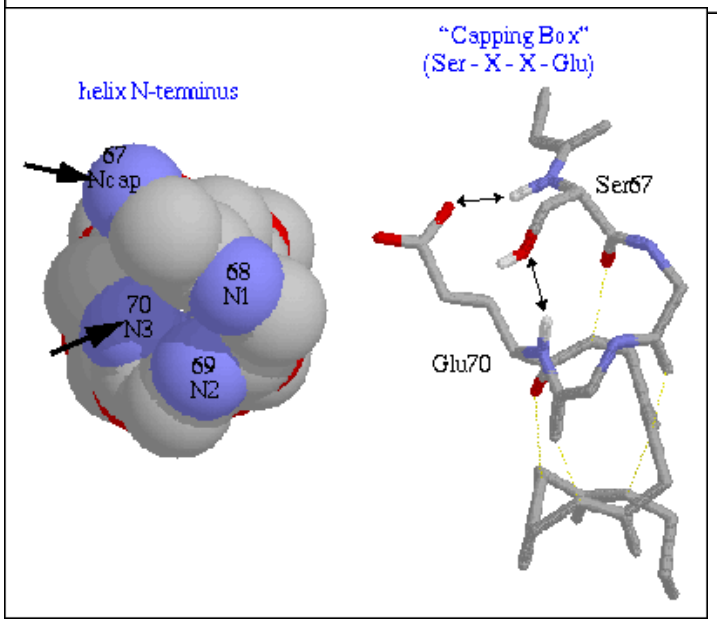
http://www.ebi.ac.uk

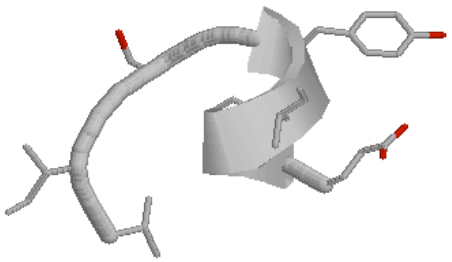
Other local structures: Helix caps



Proline helix C-cap



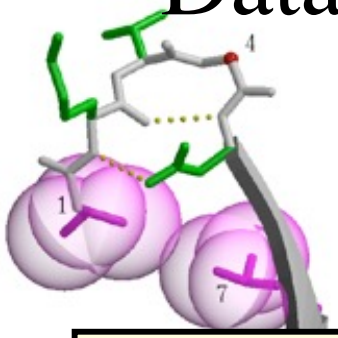




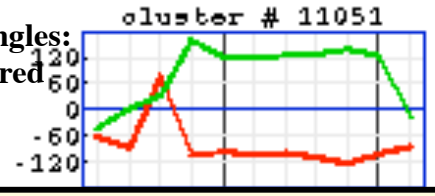
glycine helix N-cap

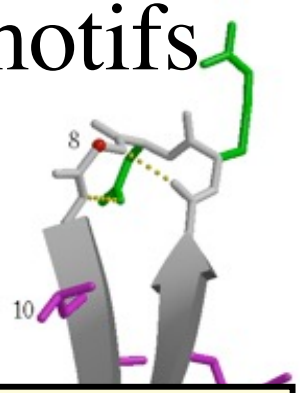
http://www.cryst.bbk.ac.uk

Datamining for local structure motifs



Backbone angles:
ψ=green, ϕ=red





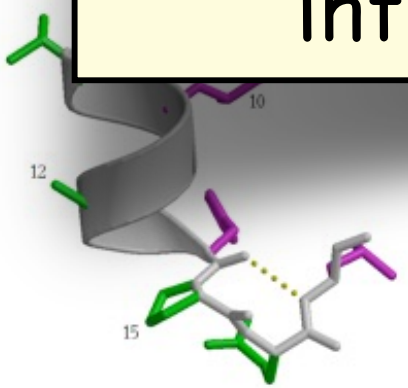
Type-I hairpin

dive
turn

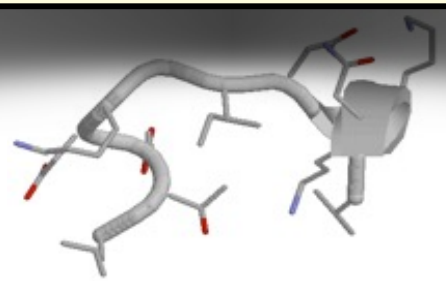
Structures from non-homologous proteins (not same family) were data-mined for correlated sequence/ structure patterns. Strongest correlations were called "folding intiation site" (I-sites) motifs.

Serine hairpi

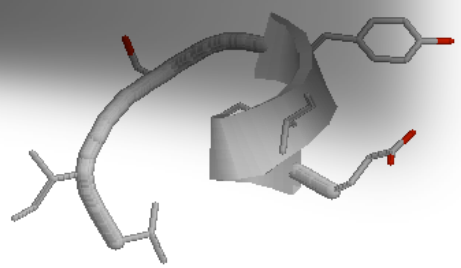
Frayed helix



Proline helix C-cap



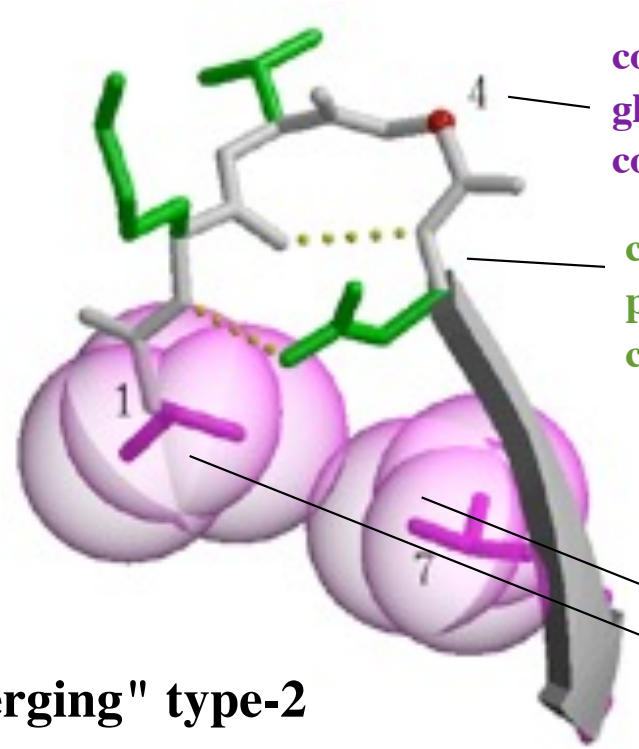
alpha-alpha corner



glycine helix N-cap

http://www.bioinfo.rpi.edu/bystrc/Isites2/

Biophysics of an I-sites motif

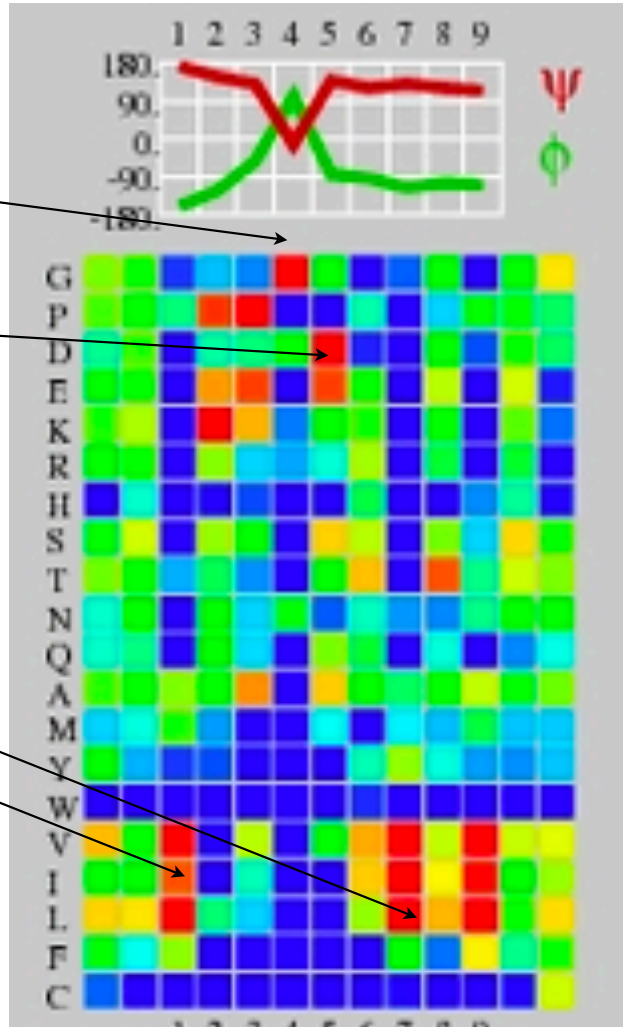


conserved glycine in α_L conformation

conserved polar side chain

conserved non-polar side chains

"Diverging" type-2 turn. A 7-residue peptide forms this structure! [2]



conserved motif backbone angles phi, psi

red= >1 log unit more likely than chance

blue= >1 log unit less likely than chance

[1]Bystroff C & Baker D. (1998). Prediction of local structure in proteins using a library of sequence-structure motifs. *J Mol Biol* 281, 565-77.

[2] Yi Q, Bystroff C, Rajagopal P, Klevit RE & Baker D. (1998). Prediction and structural characterization of an independently folding substructure in t he src SH3 domain. J Mol Biol283, 293-300.

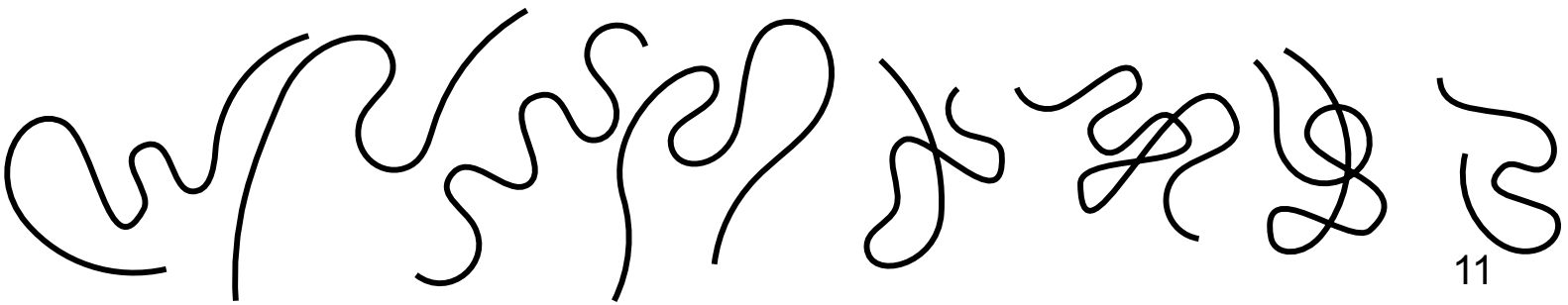
Local structure motifs are marked by **glycines** and **hydrophobic patterns**

	Motif	Average boundaries *mda* (°)	*dme* (Å)	Average *rmsd* (len)	Pattern of conserved non-polar residues
1	Amphipathic α-helix	56	0.71	0.78 (15)	1-4-8, 1-5-8
2	Non-polar α-helix	54	0.58	0.40 (11)	1-4-8, 1-5-8
3	Schellman cap type 1	81	1.01	1.02 (15)	1-6-9-11
4	Schellman cap type 2	76	0.94	0.94 (15)	1-6-8-9
5	Proline α-helix C cap	92	1.07	0.89 (13)	1-2-5-8
6	Frayed α-helix	75	0.96	0.69 (15)	1-5-9-13
7	Helix N capping box	99	0.95	0.65 (15)	1-6-9-13
8	Amphipathic β-strand	89	0.87	0.87 (6)	1-3, 1-3-5
9	Hydrophobic β-strand	101	0.91	0.91 (7)	1-2-3
10	β-Bulge	100	0.97	0.78 (7)	1-4-6
11	Serine β-hairpin	94	0.76	0.81 (9)	1-8
12	Type-I hairpin	80	0.94	1.23 (13)	1-7-8
13	Diverging type-II turn	87	1.04	1.00 (9)	1-7-9

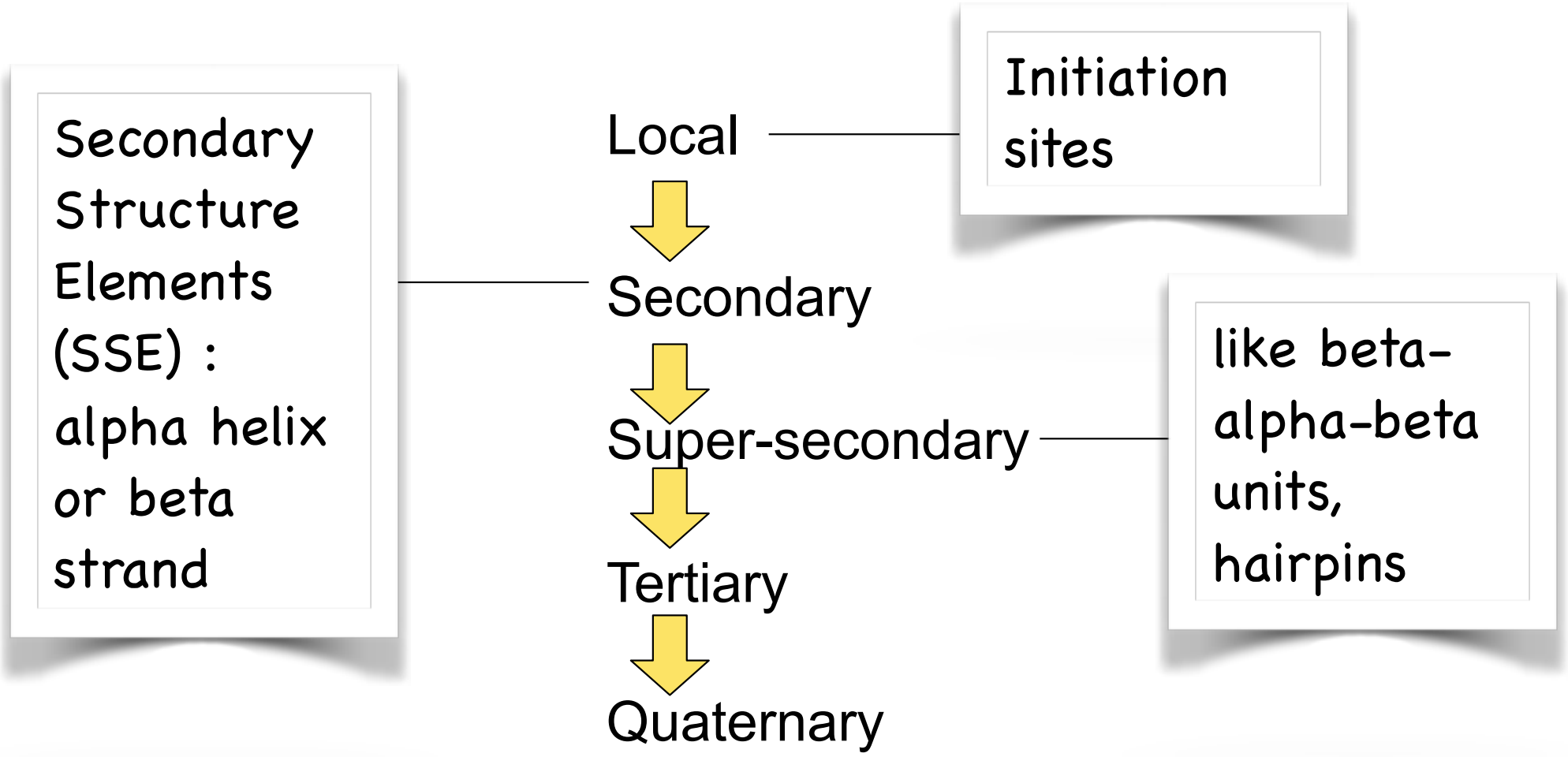
Bystroff C & Baker D. (1998). Prediction of local structure in proteins using a library of sequence-structure motifs. *J Mol Biol* 281, 565-77.

Local structure formation

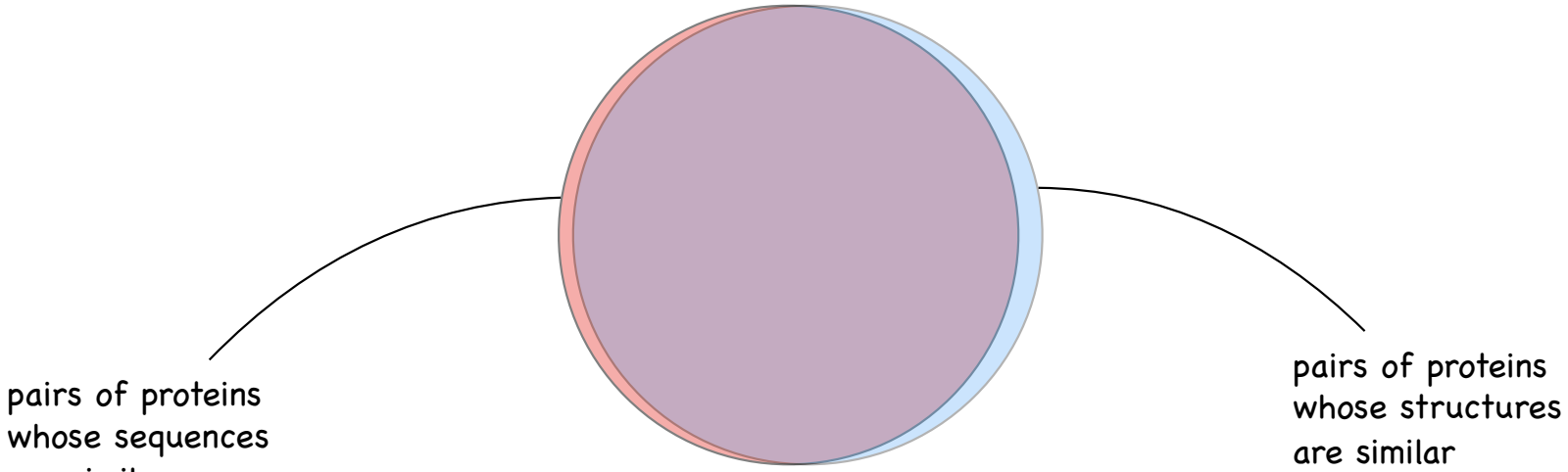
- Short pieces of protein sample conformational space randomly, driven by the hydrophobic effect (mostly).
- Glycines provide points of greater flexibility.



Folding


Secondary Structure Elements (SSE) : alpha helix or beta strand
Local
Initiation sites
Secondary
Super-secondary
like beta-alpha-beta units, hairpins
Tertiary
Quaternary

If the sequence is similar, then the structure is similar.



venn diagram

Searching for a homolog of known structure.

Download "Sequence 1" from **http://www.bioinfo.rpi.edu/bystrc/courses/biol4550/biol4550.html.** Name it **"strepto.fasta"**

Open it in **MOE**

SEQ: Protein > Search > PDB

In the database search window, **Load chain** (select strepto)

Search

Choose the hit with the best e-value

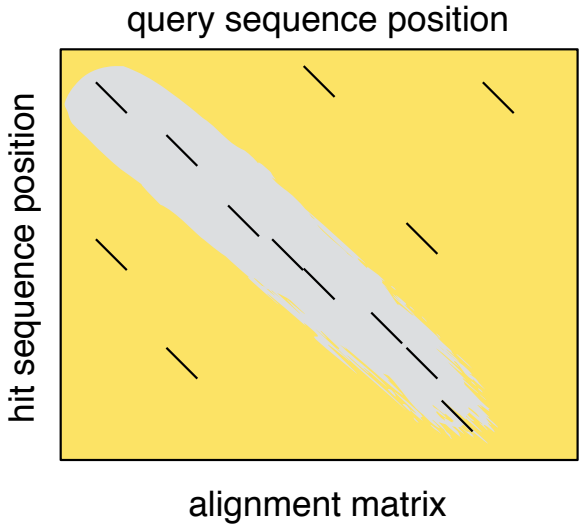
Inspect the alignment

Load the alignment. **Close** the search window,

In **SEQ** window, color residues by similarity (bottom bar **Residues > Similarity**)

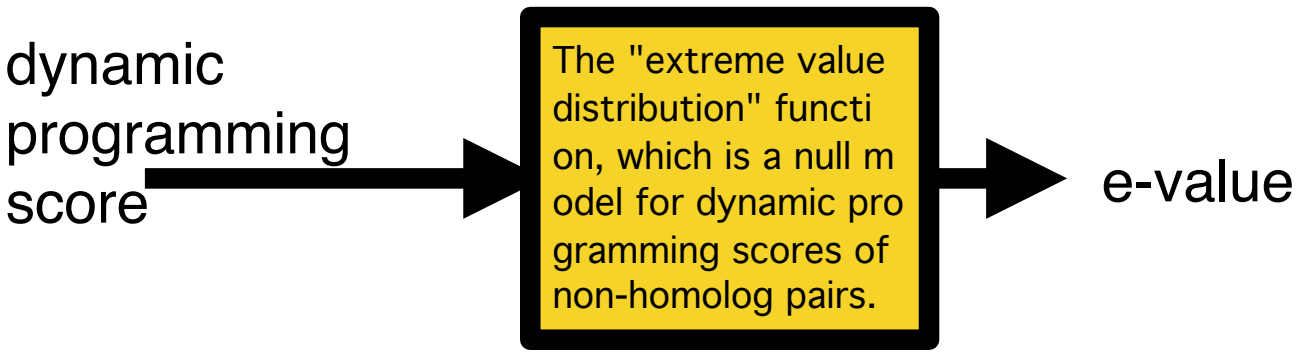
Sequence Database search

- Your sequence (**query**) is chopped up into 3-tuples.
- Every 3-tuple (there are exactly 8000) has its own look-up table, or **index**, of database locations (pdbcode, chain ID, position)
- **Hits** are chains with the most 3-tuples arranged along a diagonal on the **alignment matrix**, query vs hit.
- Hits are aligned to query using the **dynamic programming algorithm** (Smith-Waterman)
- The Dynamic Programming score is converted to a statistic, called the **e-value**.



e-value

- The number of times in a database search that you will get a random, non-homologous hit with the same score or better.



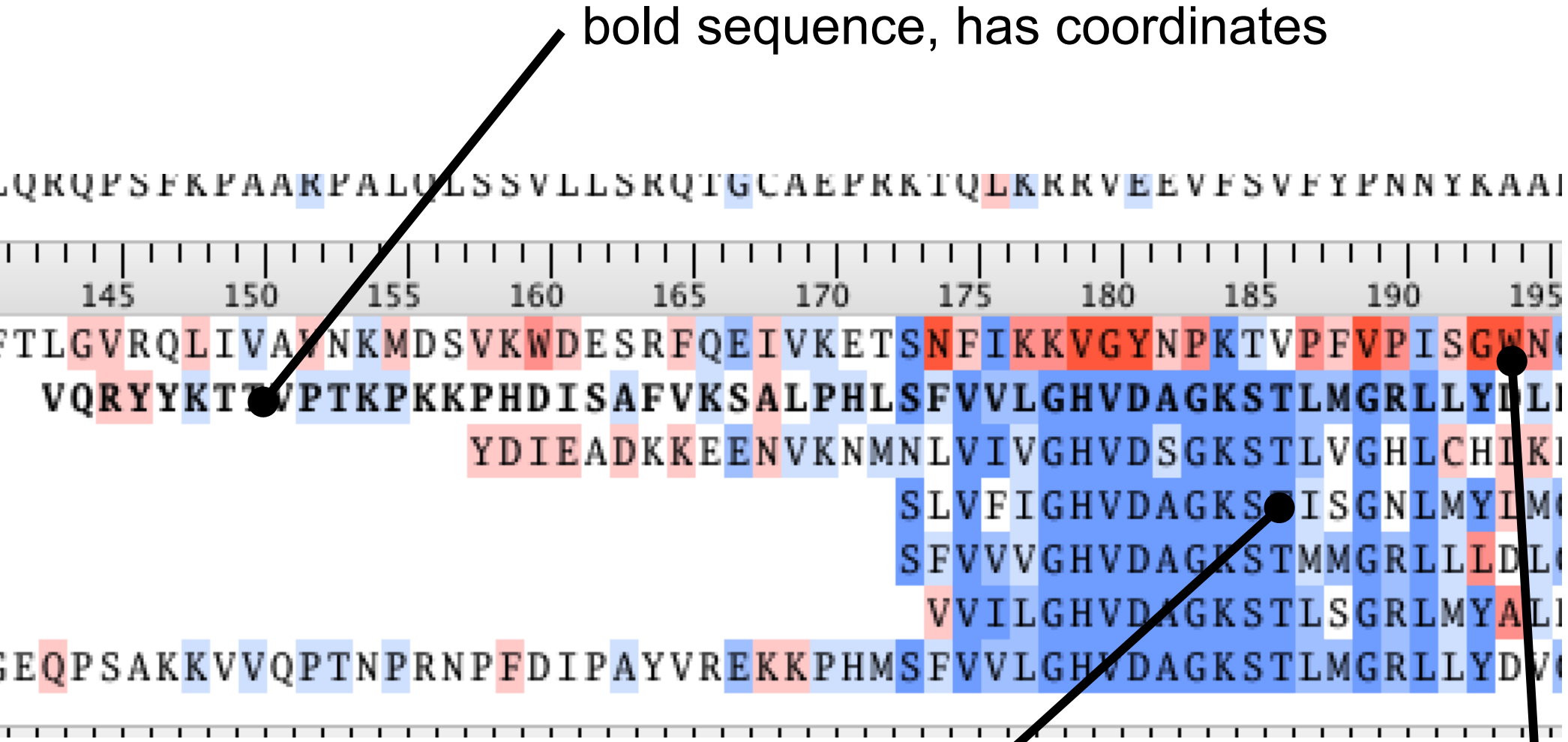
How do I know it's a good alignment?

- In NCBI Blast: Look for a low e-value (<< 1). Lower is better.
- Long strings of contiguous matches is good. Lots of indels, bad.
- Are large portions of the target sequence missing?
- Are the indels "one-sided"? (all deletions in one sequence, all insertions in the other)

How do I know it's a good alignment?

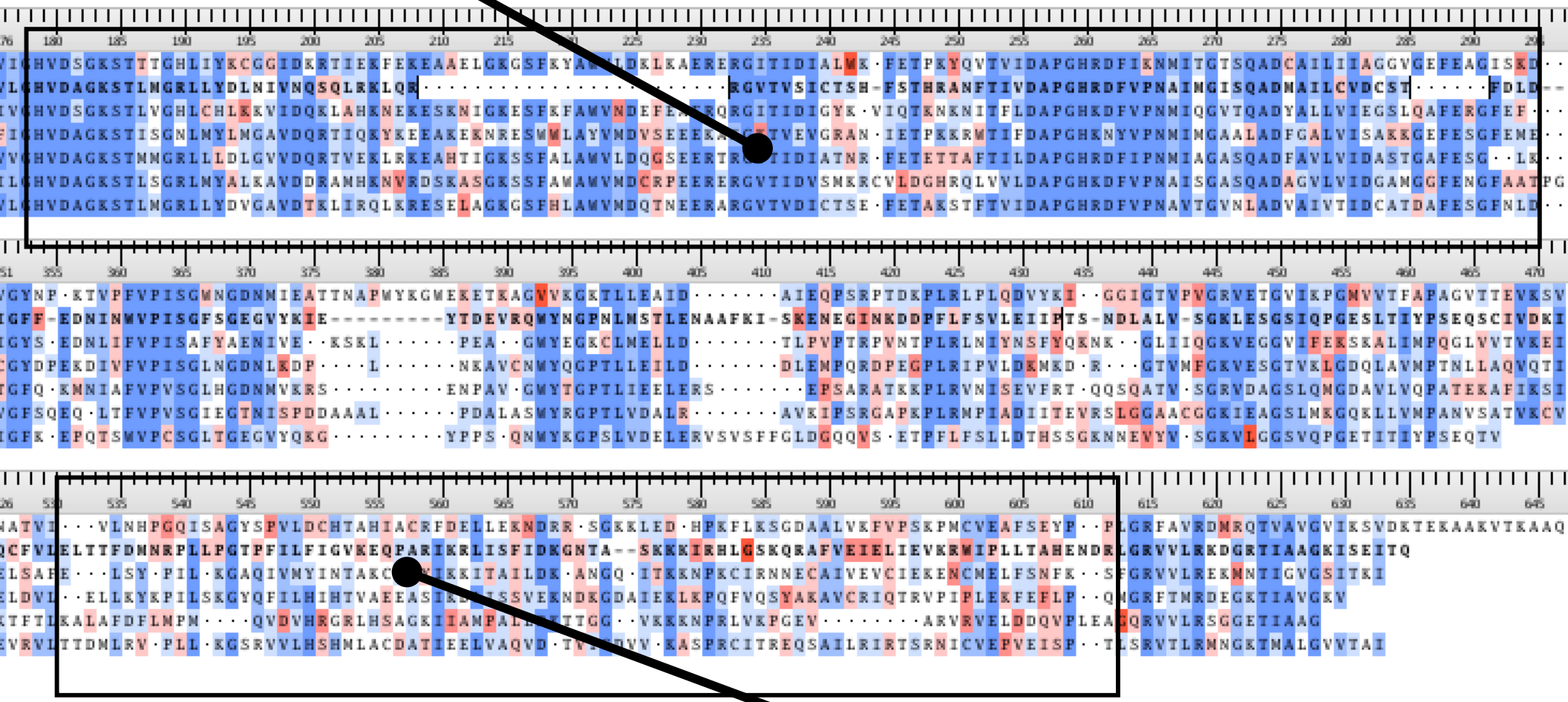
- Look at a multiple sequence alignment. In a good MSA, indel positions tend to be conserved.
- Look at positions around the indel. How conserved are they?
- Check the coverage. Is every part of hte target aligned to a template?

After loading alignment.



Residues>Similarity colors well-aligned regions blue, poorly aligned red.

Well aligned region. Bluish in *Similarity* coloring. Not so many indels.



Poorly aligned region. Reddish in *Similarity* coloring. Gaps all over the place.

Manual modeling

identity == no changes. do nothing.
similarity == mutate sidechain using **Edit/Mutate**
deletions == remove residues, make new peptide bond using **Builder**, energy minimize.
insertions == make a peptide using **Edit/build/protein**, position the loop. Make two new peptide bonds, energy minimize.

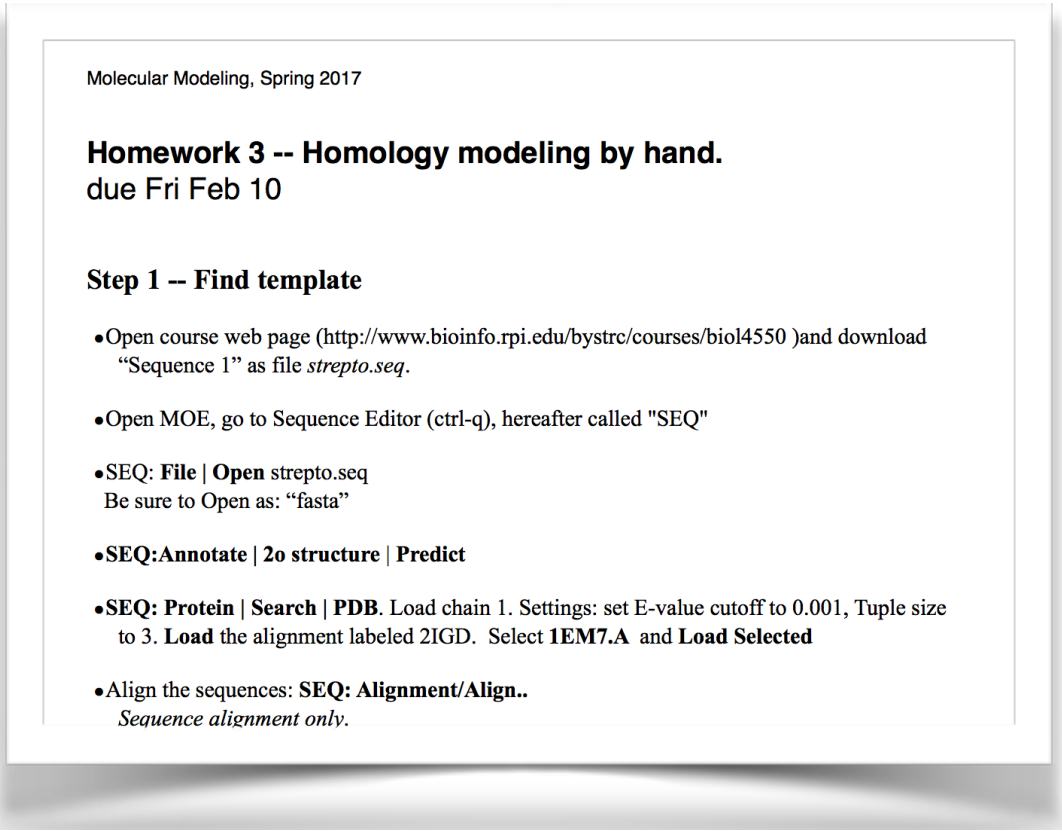
... in that order, because we always model high confidence first.

Manual homology modeling

- When you are finished with Homework 3 you should have mastered...
- Edit/Mutate
- Compute / Prepare / Structure preparation
- Edit / Potential / Fix
- Selection / Invert
- Edit / Potential / Unfix
- Window / Atom Manager (set hybridization)
- SVL: run ‘gizmin.svl'
- Edit / Build / Protein
- SEQ:Edit / Move Chains
- SEQ:Edit / Split chains
- SEQ: Edit / Join chains
- (shift/alt)-middle-mouse drag on atom selection.

Manual homology modeling

- Work on Homework 3
- http://www.bioinfo.rpi.edu/bystrc/courses/biol4550/HW3.pdf



Molecular Modeling, Spring 2017

Homework 3 -- Homology modeling by hand.
due Fri Feb 10

Step 1 -- Find template

- Open course web page (http://www.bioinfo.rpi.edu/bystrc/courses/biol4550)and download "Sequence 1" as file *strepto.seq*.
- Open MOE, go to Sequence Editor (ctrl-q), hereafter called "SEQ"
- SEQ: **File | Open** strepto.seq
  Be sure to Open as: "fasta"
- **SEQ:Annotate | 2o structure | Predict**
- **SEQ: Protein | Search | PDB**. Load chain 1. Settings: set E-value cutoff to 0.001, Tuple size to 3. **Load** the alignment labeled 2IGD. Select **1EM7.A** and **Load Selected**
- Align the sequences: **SEQ: Alignment/Align..**
  Sequence alignment only.