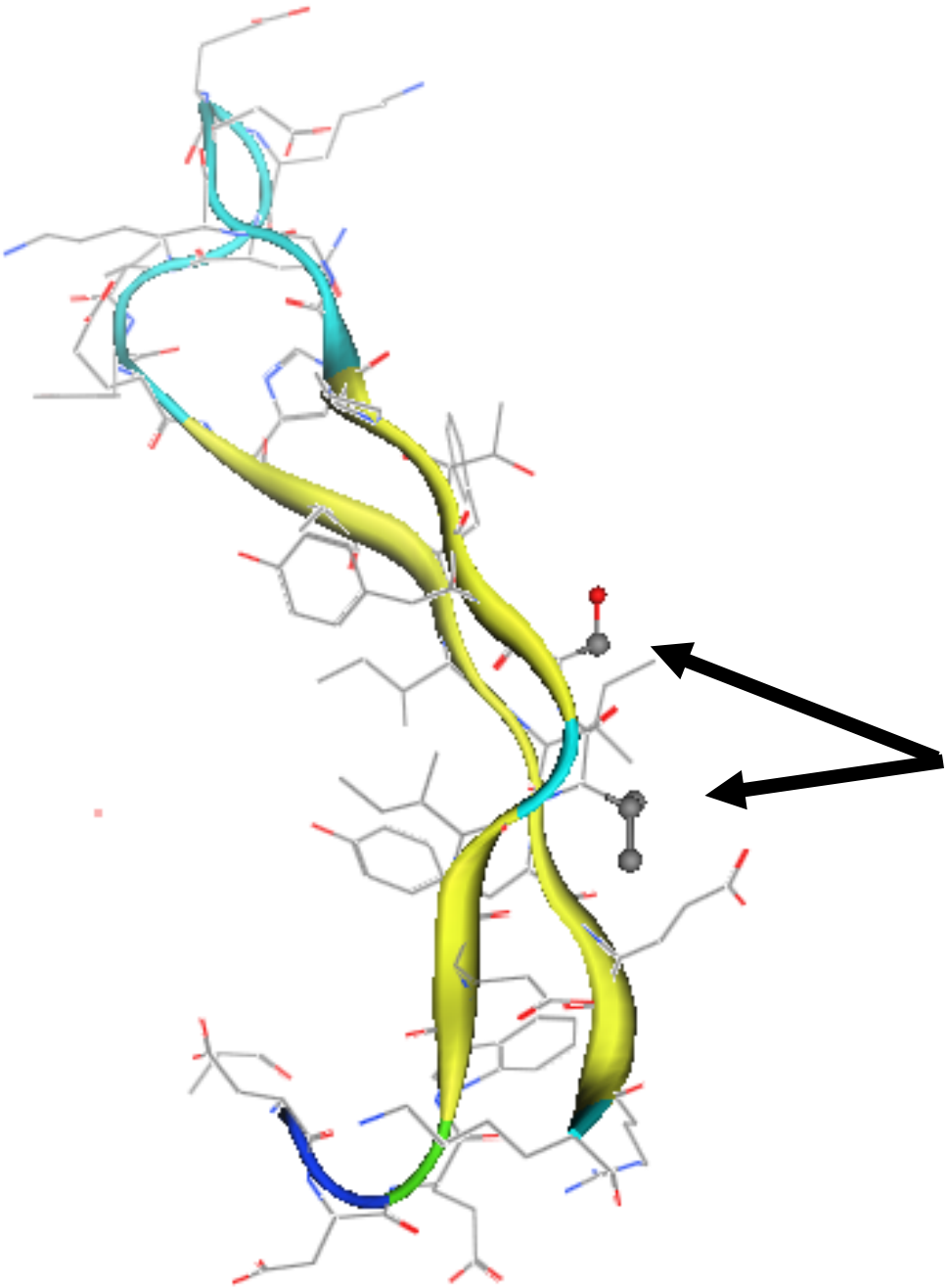
Molecular Modeling 2018
Lecture 9

Automated homology modeling.
Manual re-alignment

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



Beta-bulge :

Two sidechains (usually polar) point to one side of the sheet, instead of just one. This causes a kink in an otherwise continuous pairing.

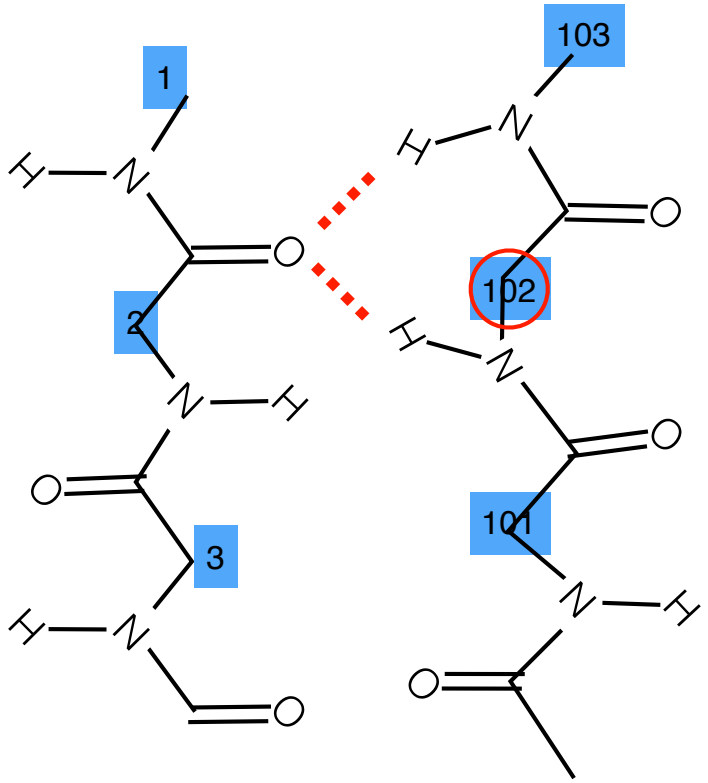
A beta-bulge is almost always a 1-residue insertion, here, usually a polar amino acid, most frequently D.

Occasionally, a many residue insertion occurs here. In that case it is called a "**beta-blowout**."

Hydrogen bonding pattern for beta bulge

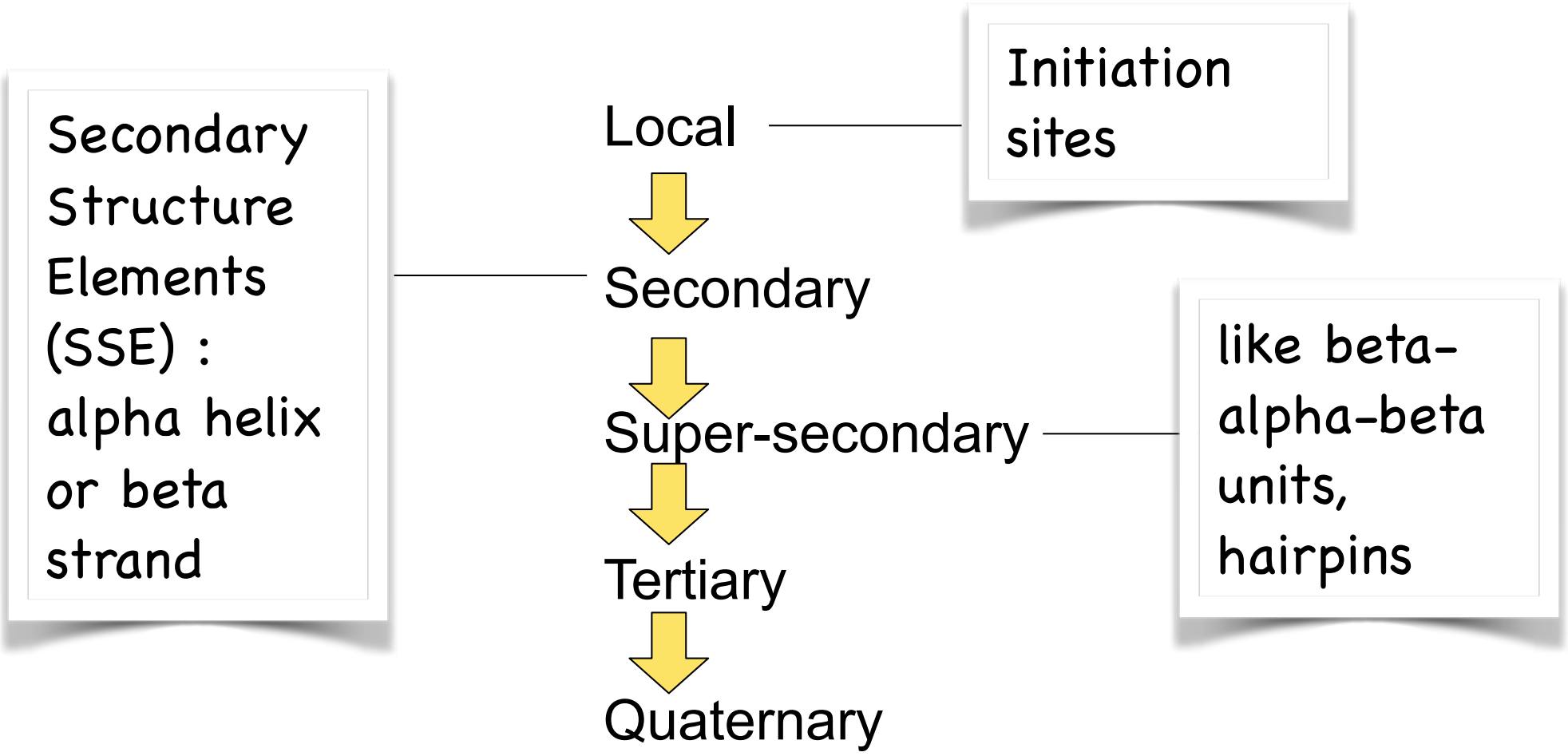
1	0	1
0	1	0

1	0	-1
0	1	0



Bulge residue between two donor in a row.

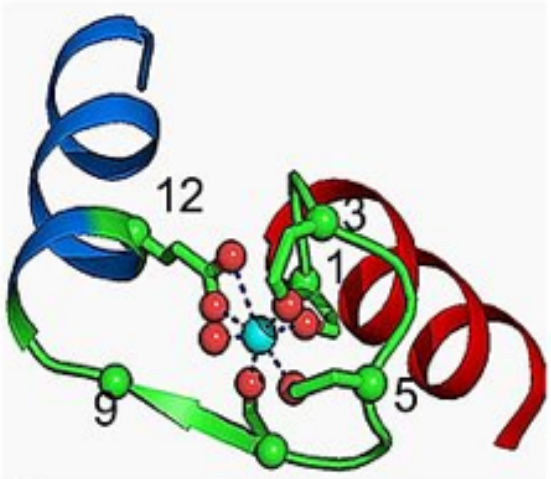
Folding


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

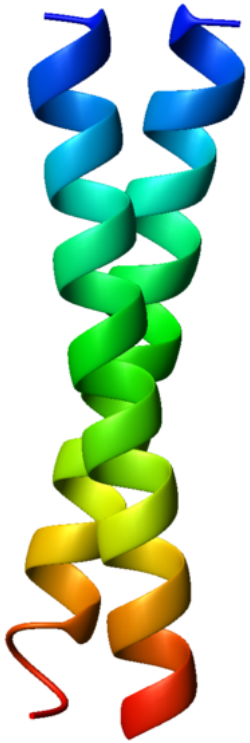
Super-Secondary Structure (SSS)

α

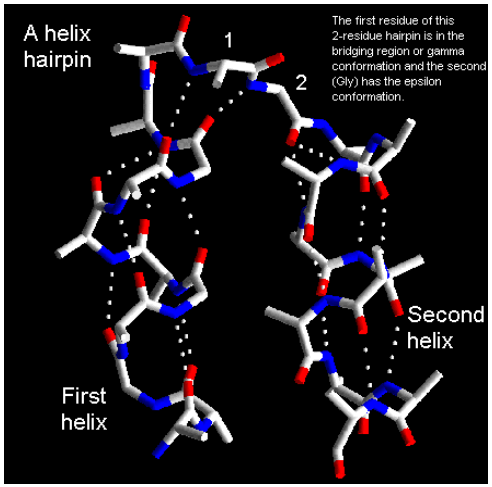
- SSS contains more than one SSE, interacting.
- beta turns and helix caps are usually involved.
- Canonical SSS have names.



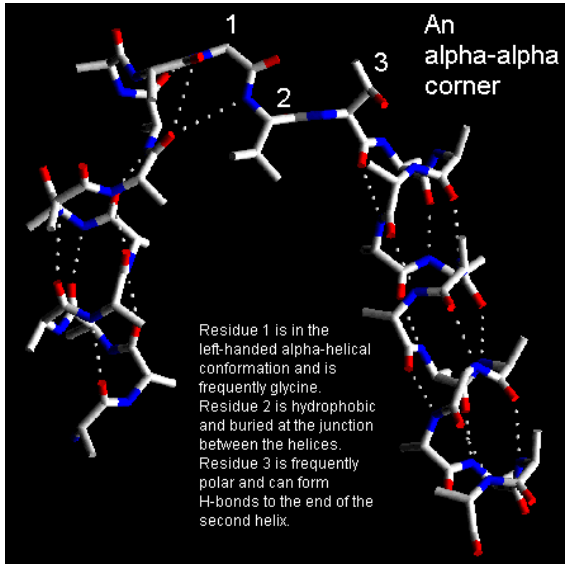
EF hand



Coiled-coil



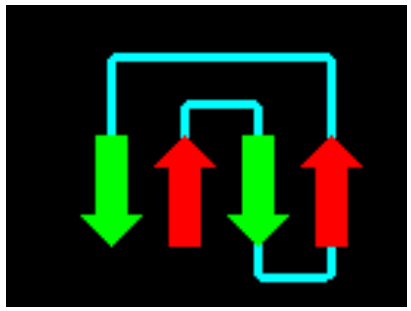
Helix hairpin



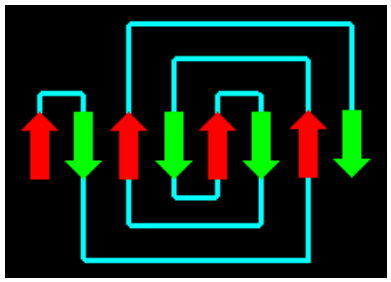
alpha-alpha corner

Super-secondary structure.

β

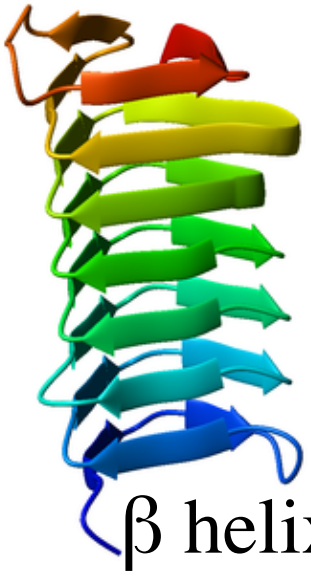


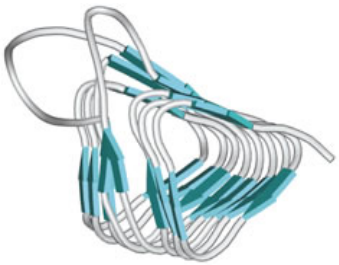
"greek key"



hairpin

meander



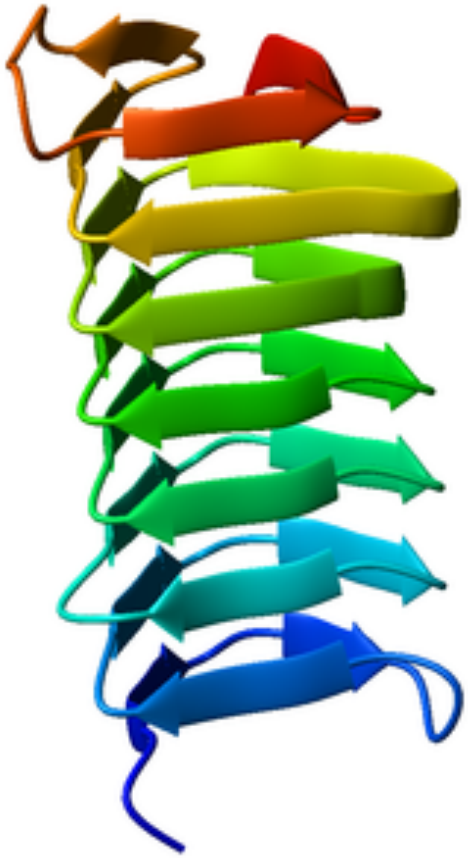


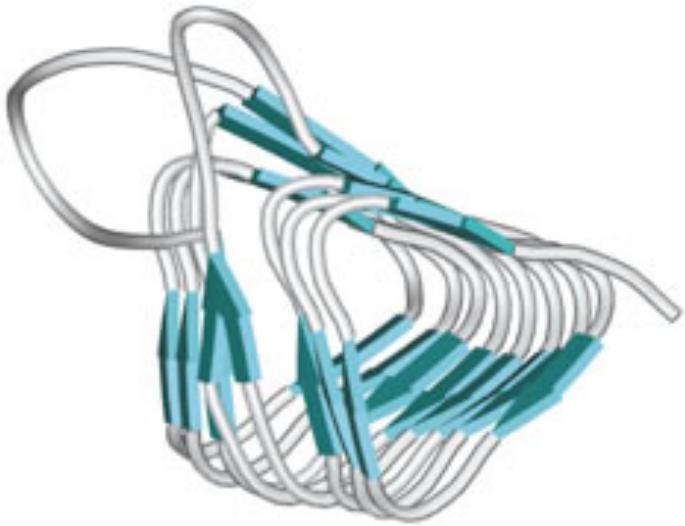
β helix

wikipedia.org/wiki/Beta_sheet

Super-secondary structure.

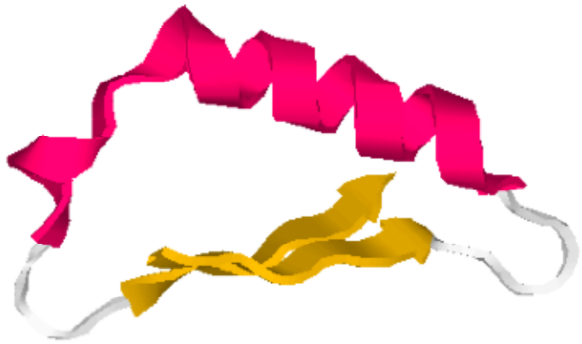
β helix may be right or left handed.

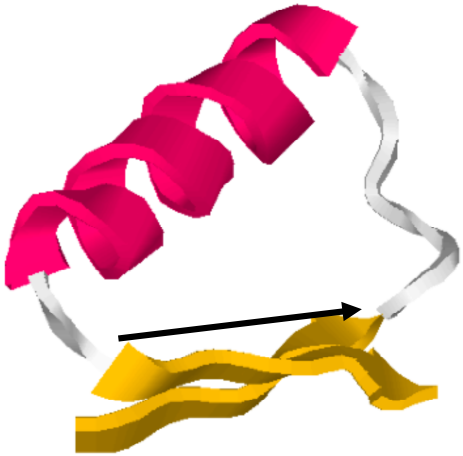




Super-secondary structure. αβ

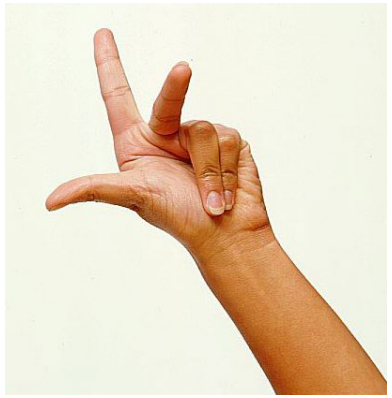
βαβ supersecondary structure units are mostly right-handed





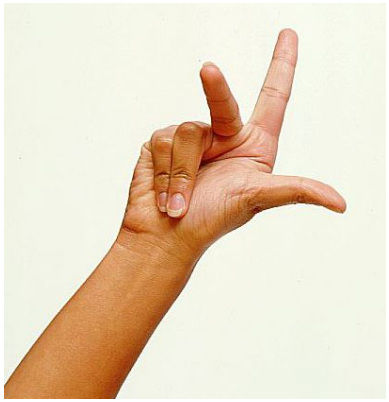
L-handed βαβ

1.5%

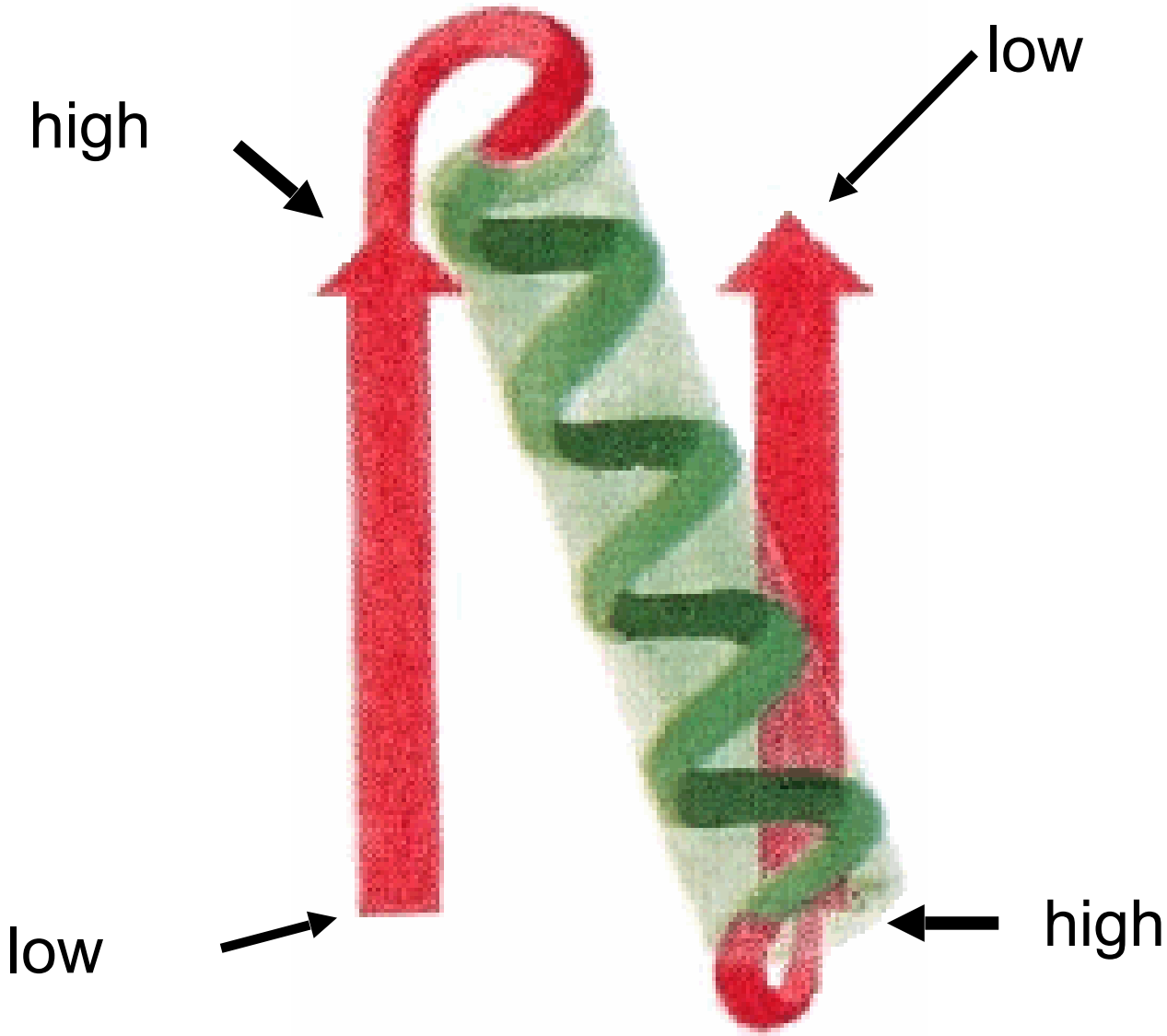


R-handed βαβ

98.5%



Theories for why βαβ units are right-handed.



Sternberg & Thornton: Twist of beta sheet makes right-handed crossover more of a straight line.

Theories for why βαβ units are right-handed.

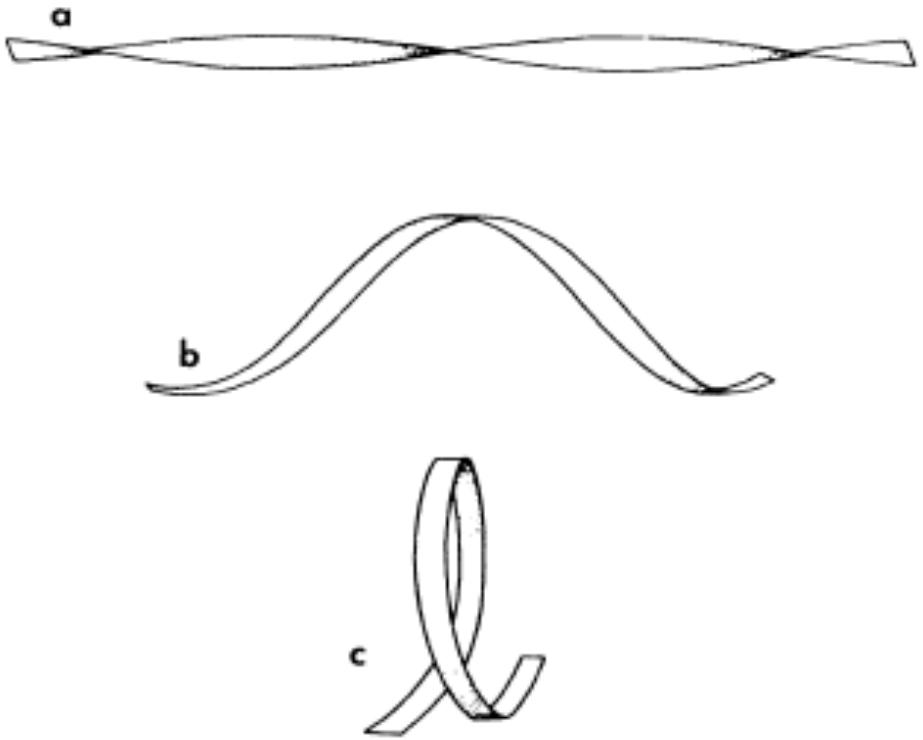


FIG. 4. A possible folding pathway which produces righthanded crossover loops from extended chain. In (a) the section of chain is extended, showing one full turn of the preferred righthanded twist for β strands. In (b) the two ends of this chain segment are moving toward one another, and the ribbon has started to buckle in a right-handed sense constrained by the chain twist. In (c) a complete righthanded loop is formed, with the two ends in position to form parallel β structure.

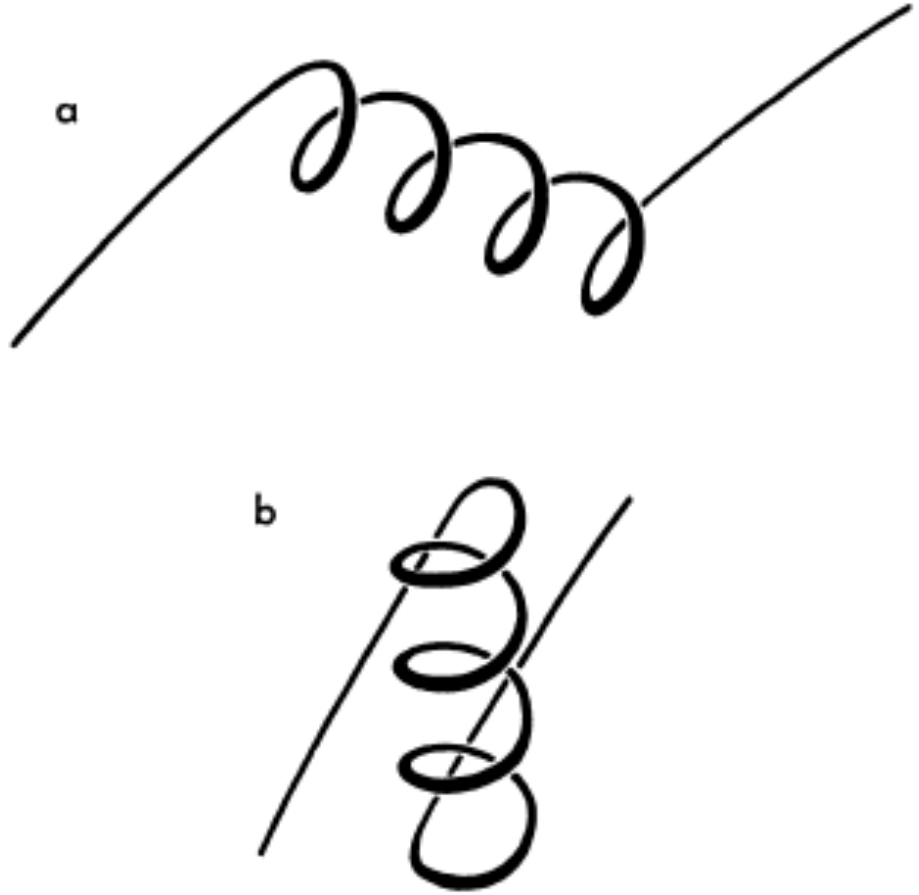
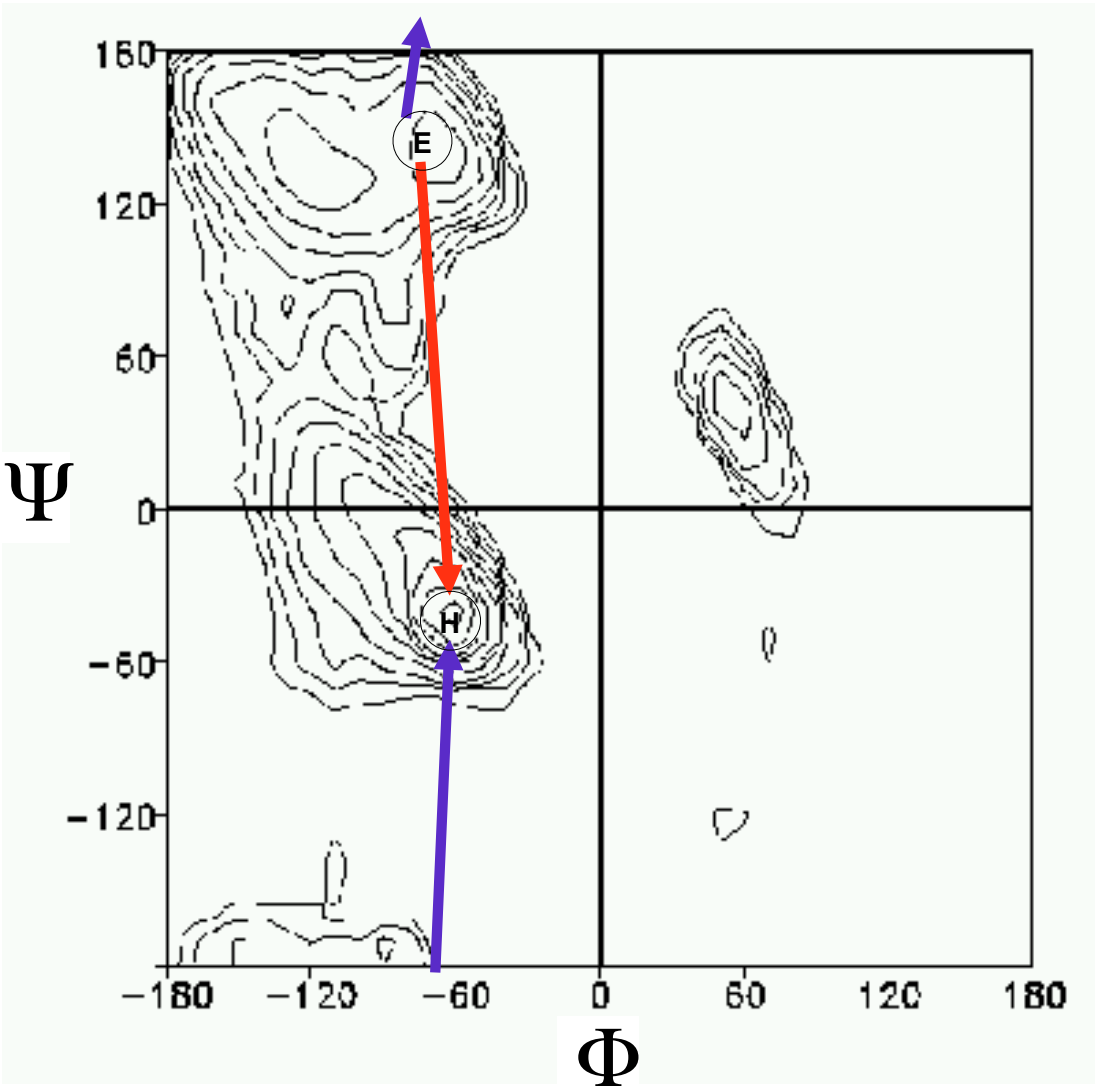


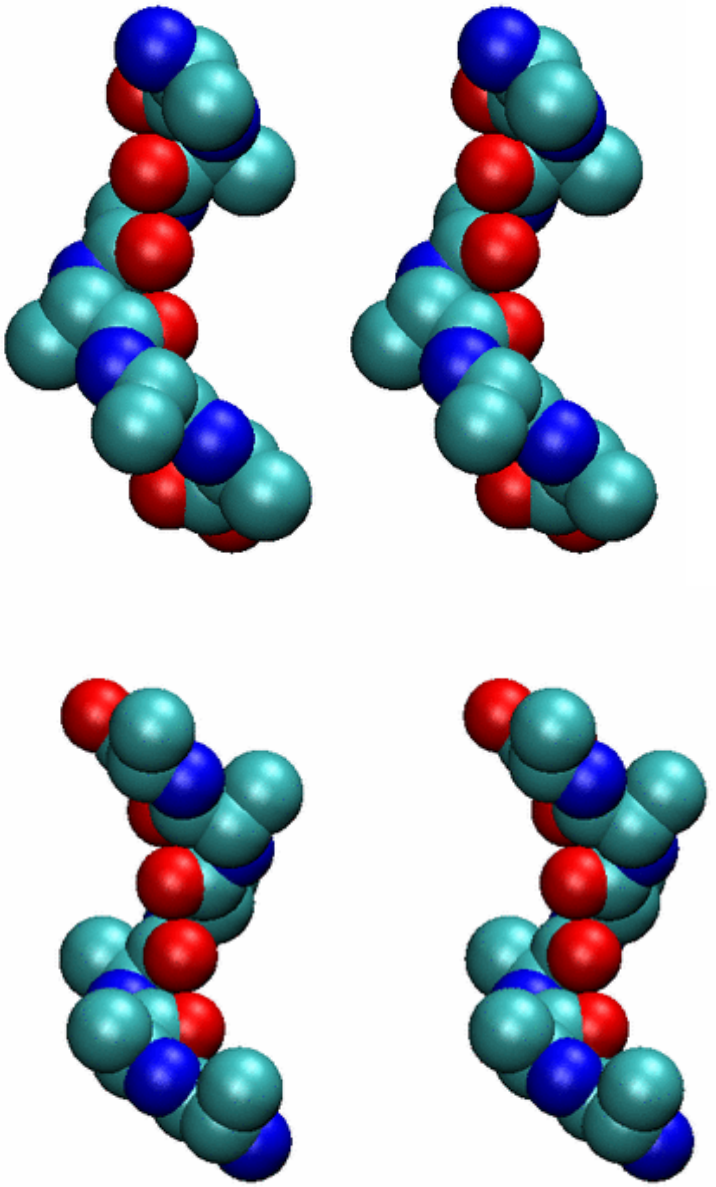
FIG. 5. A possible folding pathway which forms righthanded crossover loops from a righthanded α-helix with a β strand at each end of it.

Richardson, PNAS, 1976: Right-handed crossovers are trapped early in folding

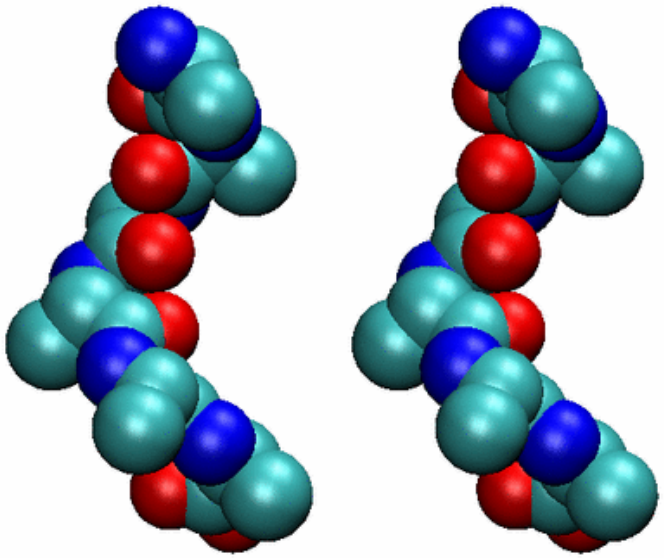
Theories for why βαβ units are right-handed.

Phone Cord Effect: Northern versus Southern route to helix

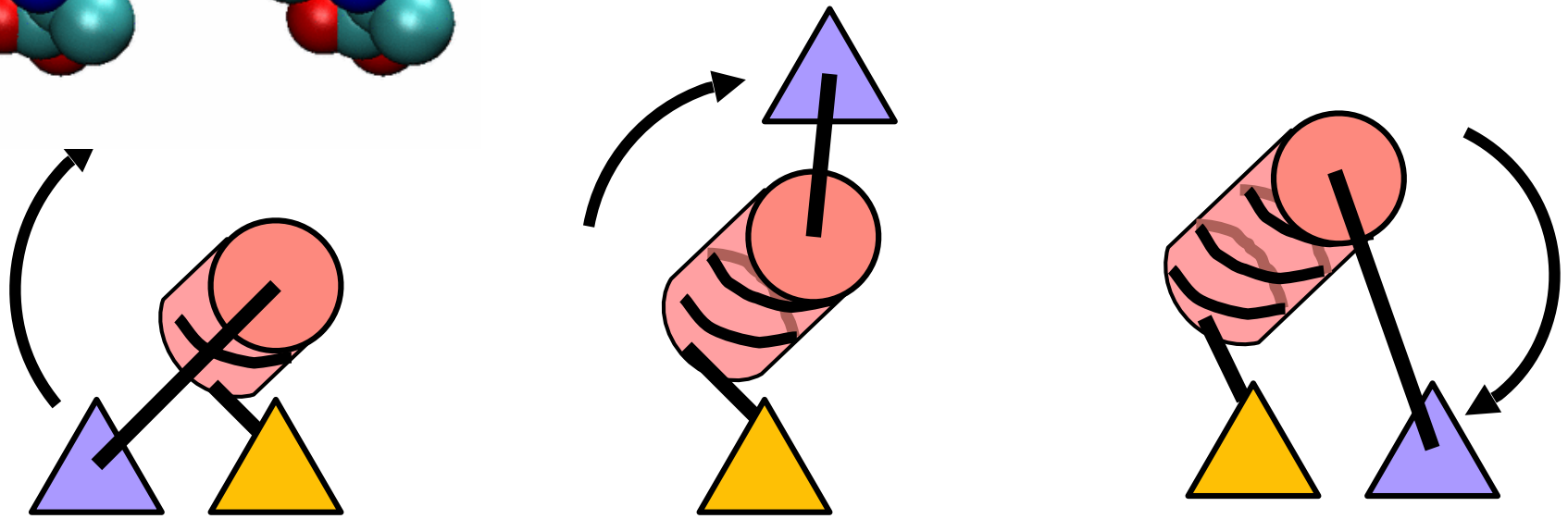




Cole B & Bystroff C. (2009) Alpha helical crossovers favor right-handed supersecondary structures by a kinetic trapping mechanism. The phone cord effect in protein folding. ***Protein Science*** **18(8) 1602 - 1608**



Theories for why βαβ units are right-handed.

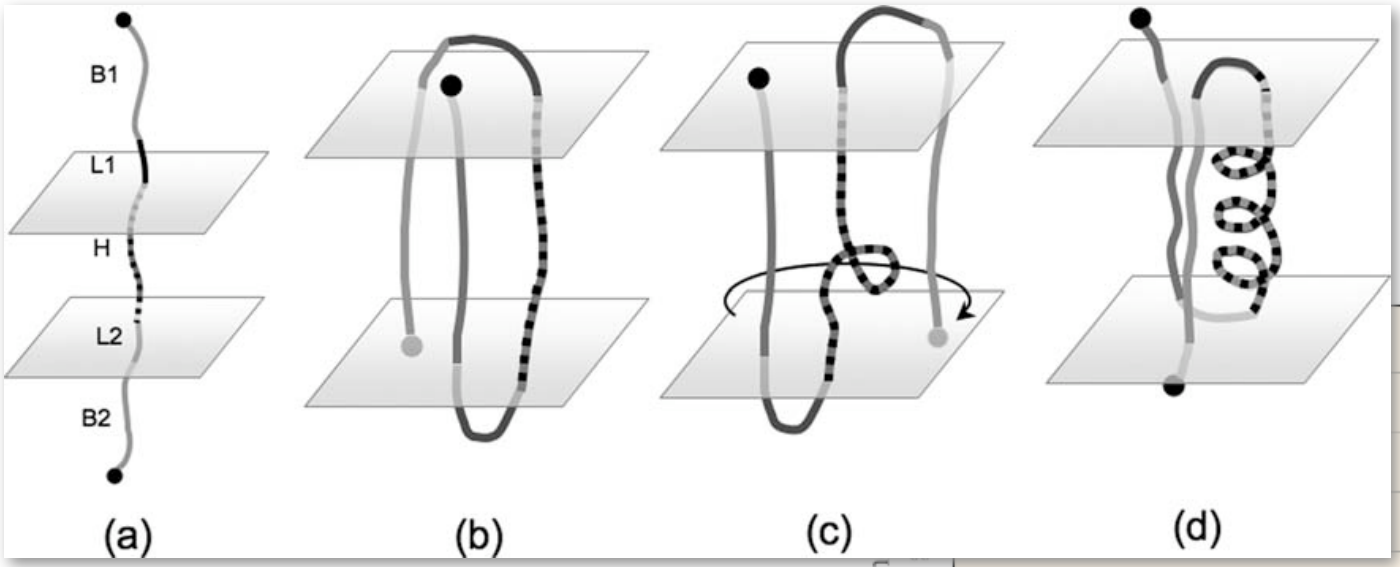


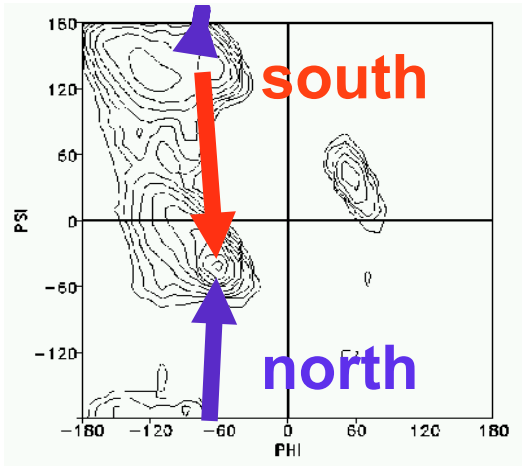
LH ⟶ RH

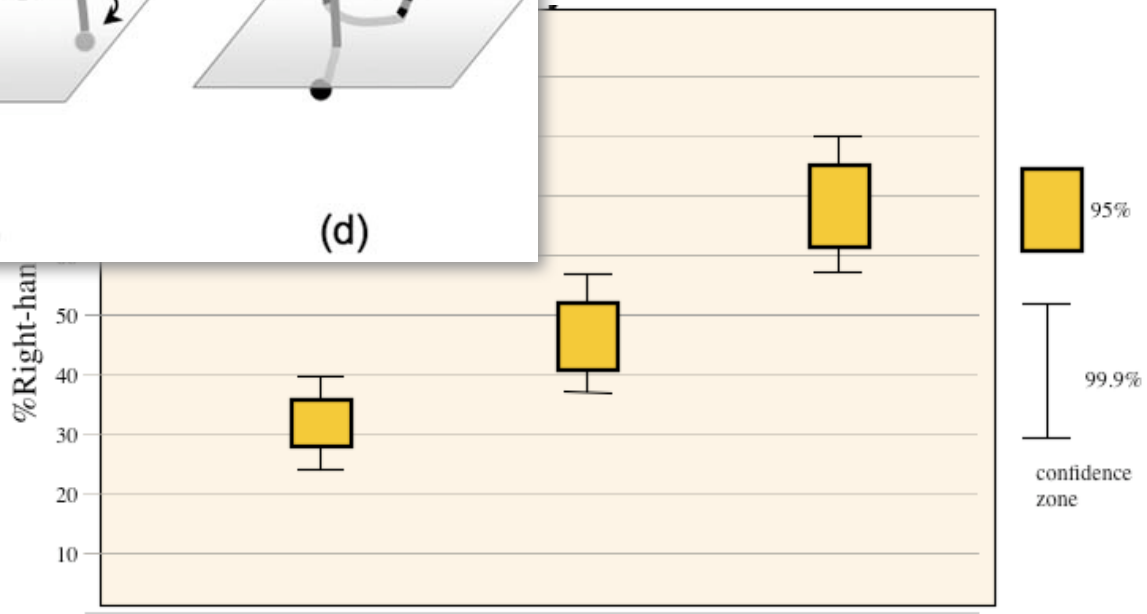
left-handed torque turns left-handed βαβ to right-handed βαβ

Cole B & Bystroff C. (2009) Alpha helical crossovers favor right-handed supersecondary structures by a kinetic trapping mechanism. The phone cord effect in protein folding. ***Protein Science*** **18(8) 1602 - 1608**

Phone cord: Demonstrative Brownian Dynamics Simulations





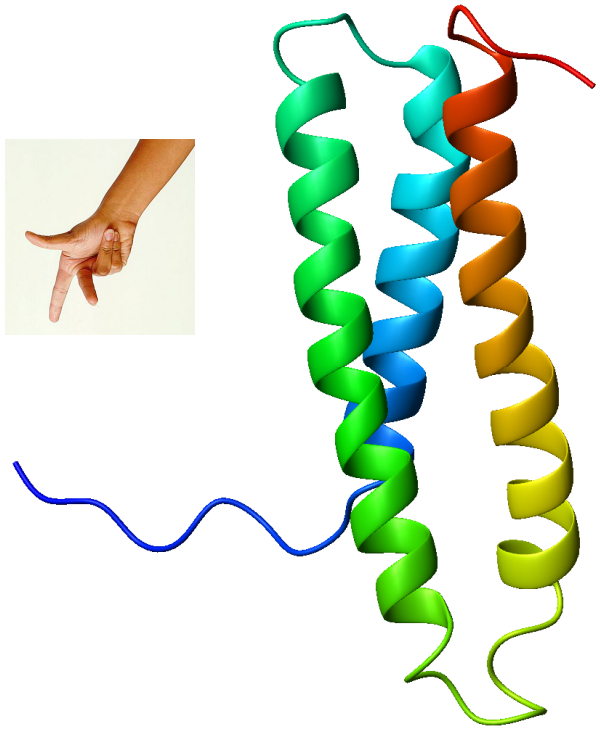


ΔΨ	0% South	50% South	100% South
Trials	2738	1164	501
Collapsed	2540	1066	418
Helical	851	578	286
Ambiguous	456	299	131
Right-handed	124	130	107
Left-handed	271	149	48

http://www.youtube.com/watch?feature=player_embedded&v=6hQYjtmU6E0

3-helix bundles are also right-handed

PDB code/chain	R/L	frac R	Contacts	Helix residue ranges: Helix 1		Helix 2		Helix 3	
1a26A	R	1.00	14	703	721	726	739	755	778
1a9xA	R	1.00	6	420	429	433	445	449	456
1a9xA	R	1.00	7	433	445	449	456	460	479
1a9xA	R	1.00	8	460	479	486	494	499	506
1a9xA	R	1.00	1	486	494	499	506	510	519
1aa7A	R	0.89	28	109	117	121	132	140	157
1aa7A	R	1.00	3	19	33	39	47	54	67
1aa7A	R	1.00	1	39	47	54	67	78	83
1aa7A	R	1.00	18	90	105	109	117	121	132
1abvA	L	0.00	3	23	39	41	47	53	64
1adtA	L	0.00	4	180	194	200	203	212	224
1aepA	L	0.00	9	34	65	69	86	94	121
1aepA	R	1.00	2	69	86	94	121	126	129
1aepA	R	1.00	1	94	121	126	129	131	154
1af7A	R	0.80	5	47	61	66	75	80	88
1agrE	R	1.00	2	53	61	63	68	70	82
1ah7A	L	0.00	3	13	27	34	42	44	54
1ah7A	L	0.00	6	187	190	193	204	206	241
1ah7A	L	0.10	10	106	124	141	151	172	185
1ailA	R	0.86	29	3	24	30	50	54	69
1aorA	L	0.00	1	237	240	243	253	274	280
1aorA	L	0.00	1	274	280	312	317	327	340
1aorA	L	0.00							
1aorA	L	0.00							
1aorA	L	0.14							
1aorA	L	0.20							
1aorA	R	1.00							
1aorA	R	1.00							
1b79A	R	1.00							
1b79A	R	1.00							
1b79A	R	1.00							
1bf5A	R	1.00							
1bf5A	R	1.00							
1bgfA	L	0.00							
1bgfA	R	1.00							
1bmtA	L	0.00							
1bouA	L	0.00							
1bouA	R	1.00							
1bvp1	R	0.90							
1c1kA	L	0.00							
1c1kA	R	1.00							
1c75A	R	1.00							
1cktA	L	0.18							
1crkA	L	0.17							
1cshA	L	0.00							
1cshA	R	1.00							
1cshA	R	1.00							
1cukA	L	0.08							
1d2tA	R	0.96							
1d2tA	R	1.00							
1dbhA	R	1.00							
1dbhA	R	1.00							
1dbhA	R	1.00	2	331	363	367	379	381	392
1dj8A	L	0.14	7	29	39	52	68	74	82
1dj8A	R	1.00	7	18	23	29	39	52	68
1dlwA	R	1.00	1	2	6	9	25	38	52
1dnpA	L	0.00	3	310	318	324	335	341	355
1dnpA	R	0.88	8	204	224	238	243	248	258
1dnpA	R	1.00	2	324	335	341	355	359	369



PDB code/chain	R/L	frac R	Contacts	Helix residue ranges: Helix 1		Helix 2		Helix 3	
1j0tA	L	0.06	16	16	31	37	40	49	56
1jj2O	L	0.00	9	4	14	28	33	37	45
1jj2O	L	0.00	5	90	111	116	127	134	141
1jr3A	R	0.78	9	278	297	304	308	310	319
1jr3A	R	1.00	11	246	258	261	273	278	297
1jswA	L	0.00	6	47	65	70	83	104	121
1jswA	L	0.00	18	201	226	246	257	275	302
1jswA	L	0.00	6	275	302	331	355	365	388
1jswA	R	1.00	1	147	182	201	226	246	257
1k6kA	R	1.00	2	4	20	27	35	38	46
1k6kA	R	1.00	8	27	35	38	46	51	64
1k8kE	R	1.00	8	63	83	88	100	123	148
1kjsA	R	0.76	38	16	26	34	38	45	62
1kp8A	L	0.00	4	53	59	65	84	89	109
1kp8A	R	0.82	28	10	29	53	59	65	84
1l8wA	L	0.00	11	228	240	255	260	277	289
1lbuA	R	1.00	1	17	25	44	56	67	76
1lkpA	R	1.00	1	56	63	83	112	115	144
1llaA	L	0.00	6	266	282	300	309	317	320
1llaA	R	1.00	5	300	309	317	320	323	332
1llpA	L	0.00	1	166	177	203	209	236	242
		1.00	3	70	73	75	80	87	101
		0.00	9	6	18	24	36	41	54
		0.86							
		1.00							
		1.00							
		0.00							
		0.00							
		0.11							
		1.00							
		1.00							
		1.00							
		0.00							
		0.00							
		1.00							
		1.00							
		1.00							
		1.00							
		1.00							
		1.00							
		0.00							
		1.00							
		1.00							
		0.79							
		1.00							
		0.95							
		1.00							
		0.00							
		1.00	4	66	76	80	83	86	92
		1.00	3	120	129	131	145	151	169
		1.00	4	2	19	22	39	43	63
		0.00	15	33	55	62	67	70	83
1n81A	L	0.00	13	116	142	146	153	156	169
1n81A	R	1.00	3	62	67	70	83	91	108
1n81A	R	1.00	5	146	153	156	169	176	193
1n93X	L	0.00	3	130	149	152	156	159	172
1n93X	L	0.00	2	256	276	278	281	289	293
1n93X	L	0.12	8	71	82	86	93	130	149
1n93X	R	1.00	9	53	59	63	68	71	82

61% R,

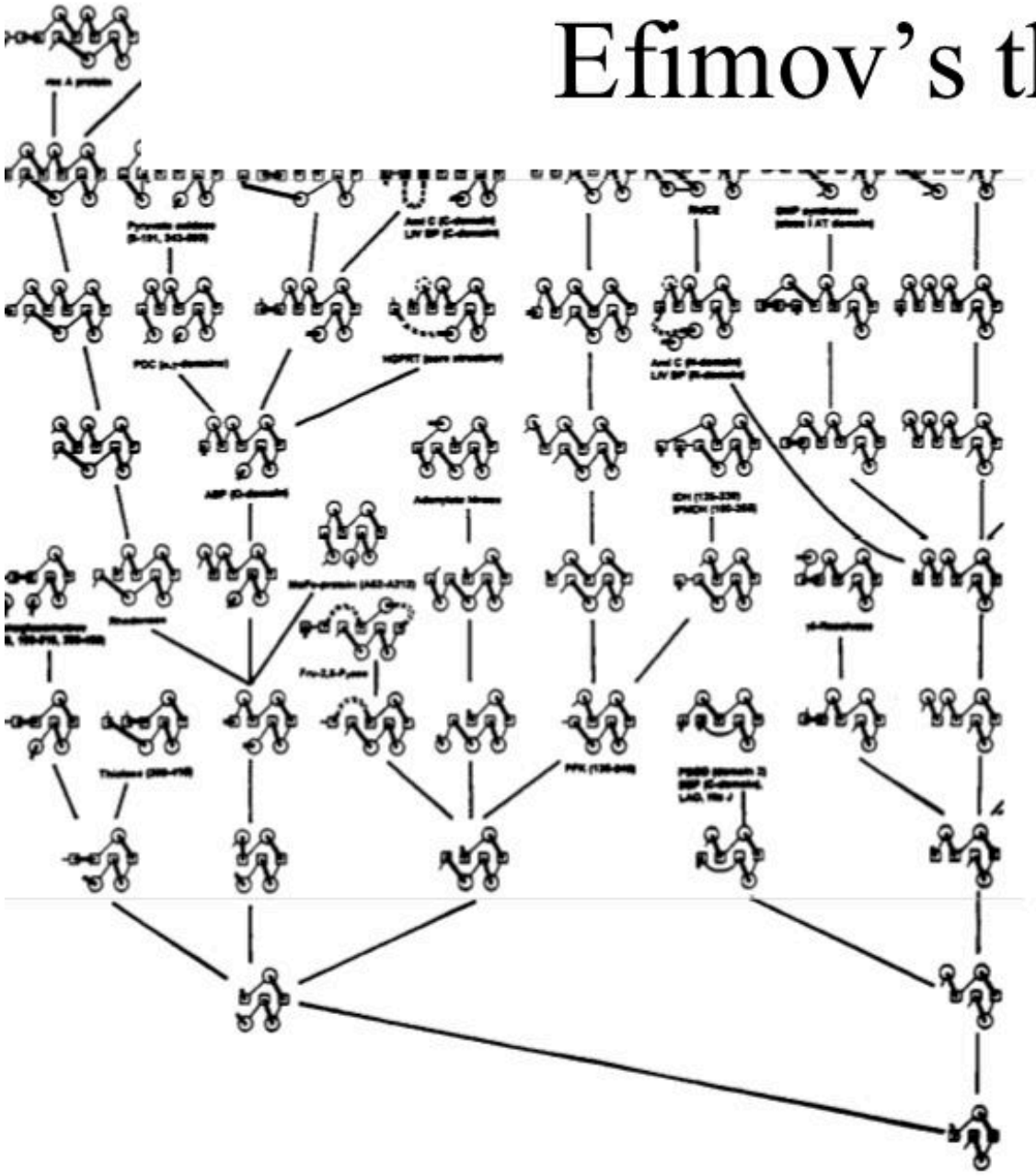
39% L,

n=431,

p(R=0.50)<0.01

PDB code/chain	R/L	frac R	Contacts	Helix residue ranges: Helix 1		Helix 2	
1tm9A	R	0.83	6	57	74	92	106
1tx4A	L	0.10	21	192	203	209	214
1tx4A	R	0.86	14	64	75	90	102
1tx4A	R	1.00	5	165	183	185	188
1tx9A	L	0.00	11	88	97	104	109
1tx9A	R	1.00	2	75	85	88	97
1tzyA	L	0.00	3	17	21	26	37
1u84A	R	1.00	2	28	38	44	58
1ubyA	L	0.00	5	53	67	73	85
1ubyA	L	0.00	13	167	191	204	214
1ubyA	L	0.00	6	204	214	216	231
1ubyA	L	0.00	4	283	291	294	303
1un8A	R	1.00	24	356	371	373	382
1un8A	R	1.00	5	388	404	413	427
1un8A	R	1.00	16	477	490	495	511
1us7B	L	0.00	8	203	226	234	242
1us7B	R	1.00	4	156	164	168	177
1us7B	R	1.00	6	294	300	317	321
1utgA	R	1.00	5	4	14	18	27
1uujA	R	1.00	2	5	21	25	35
1v2zA	R	1.00	7	186	203	211	225
1v2zA	R	1.00	9	211	225	229	246
1v54H	R	1.00	1	26	45	50	63
			9	9	18	22	38
			20	22	38	50	63
			11	68	77	87	93
			8	135	150	160	178
			3	67	78	84	97
			17	43	57	60	72
			10	60	72	77	91
			2	102	116	123	132
			6	17	23	25	45
			14	381	402	412	428
			9	412	428	438	450
			5	438	450	458	475
			2	458	475	477	485
			5	263	297	304	321
			8	364	377	381	390
			2	289	301	305	322
			4	11	44	49	71
			1	92	128	138	161
			5	6	12	16	20
			5	23	41	55	70
			5	16	20	23	41
			1	105	120	132	137
			3	28	36	46	60
			2	242	252	260	281
			14	316	326	332	336
1yfsA	R	0.80	10	260	281	291	309
1yfsA	R	1.00	3	338	370	380	389
1yfsA	R	1.00	14	380	389	394	403
1yfsA	R	1.00	3	394	403	410	423
1ygeA	L	0.00	3	636	640	643	656
1ygeA	L	0.12	8	255	276	286	292
1ygeA	R	1.00	1	410	415	417	422
1ygeA	R	1.00	6	474	516	523	530
1yozA	R	0.77	13	65	81	83	95
1z67A	L	0.00	7	32	41	45	52
1zt2B	L	0.00	2	27	33	35	50

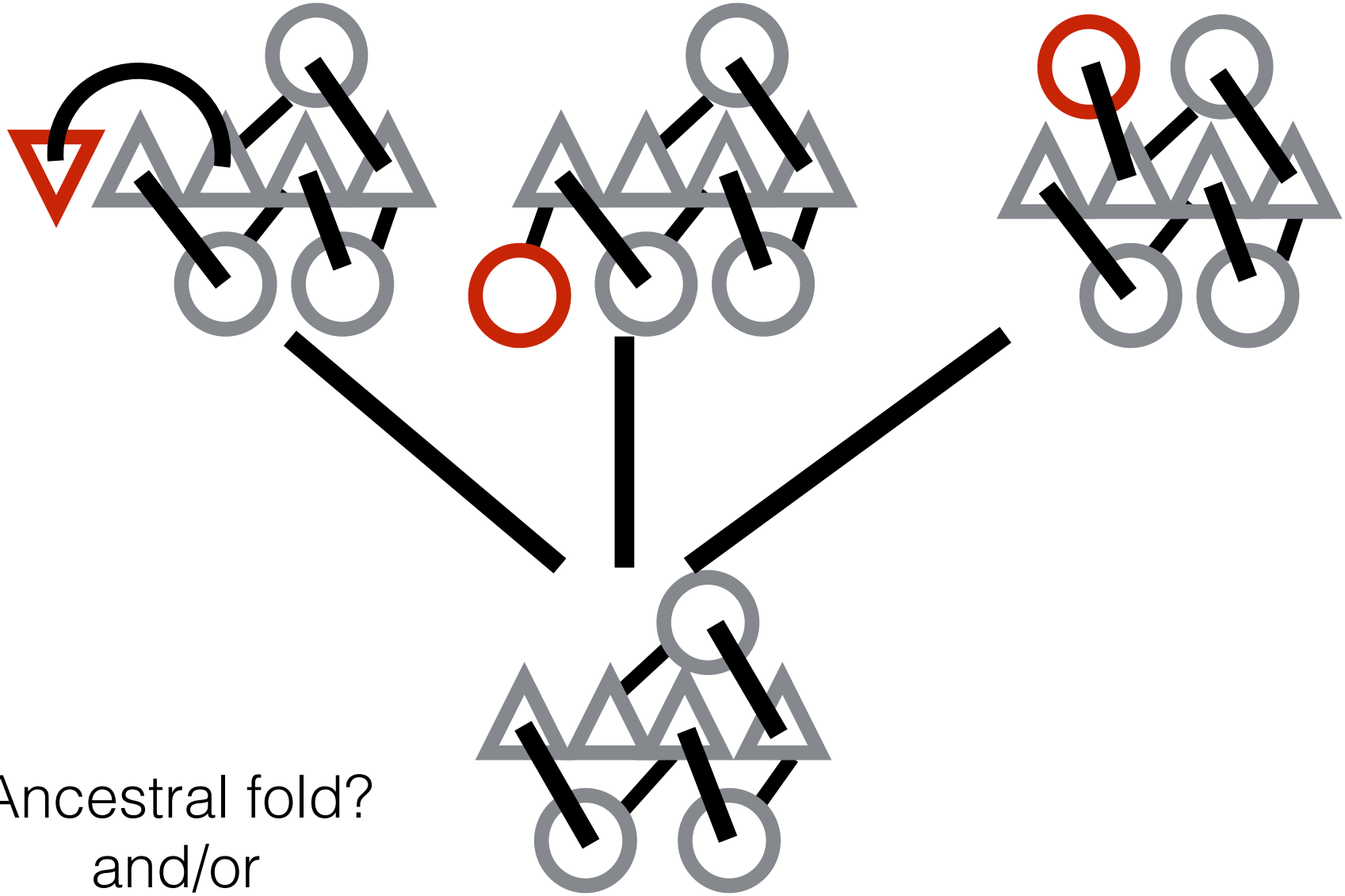
Efimov's theory



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 this recapitulate evolution? the folding pathway?

Many proteins share common core structures (Efimov cores)



Ancestral fold?
and/or
Folding intermediate?

Methods for Predicting protein structure at various levels.

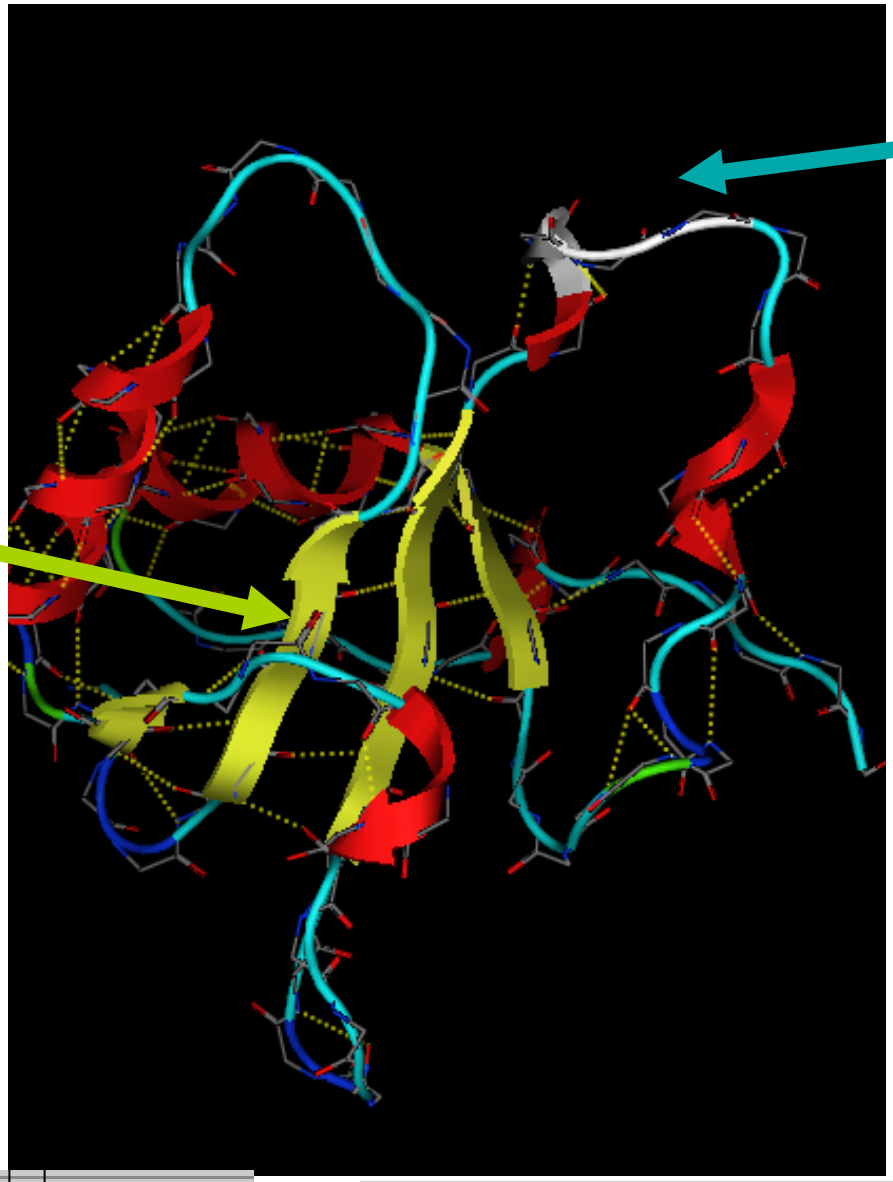
Representation	--	Algorithms used.
Secondary structure--		stats, neural nets, HMMs,
Local structure --		stats, HMMs
Supersecondary structure --		MD, rules, HMMs
Inter-residue contacts --		neural nets, rules, covariance
Tertiary structure--		MD, homology
Sidechain conformation --		MD, homology
Domain-domain interactions --		MD, homology
Quaternary structure --		MD, homology

Homology modeling, continued.

1. Database search
2. Alignment (automatic, manual)
3. Automated homology model script
4. Re-modeling, manual modeling, energy minimization.
5. Sidechain modeling, energy minimization.
6. etc....

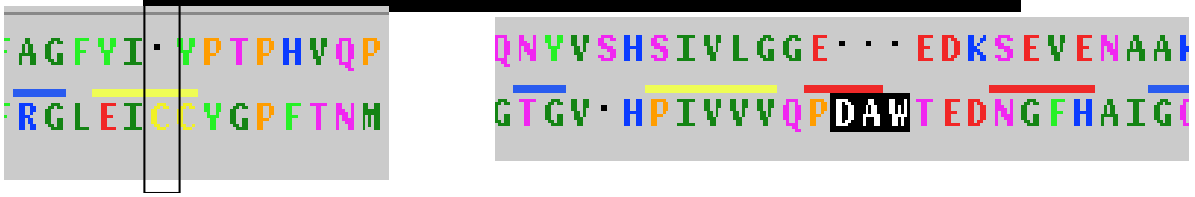
Setup of loop search: **the wrong way**

Deleted cys is here, in the middle of a strand.



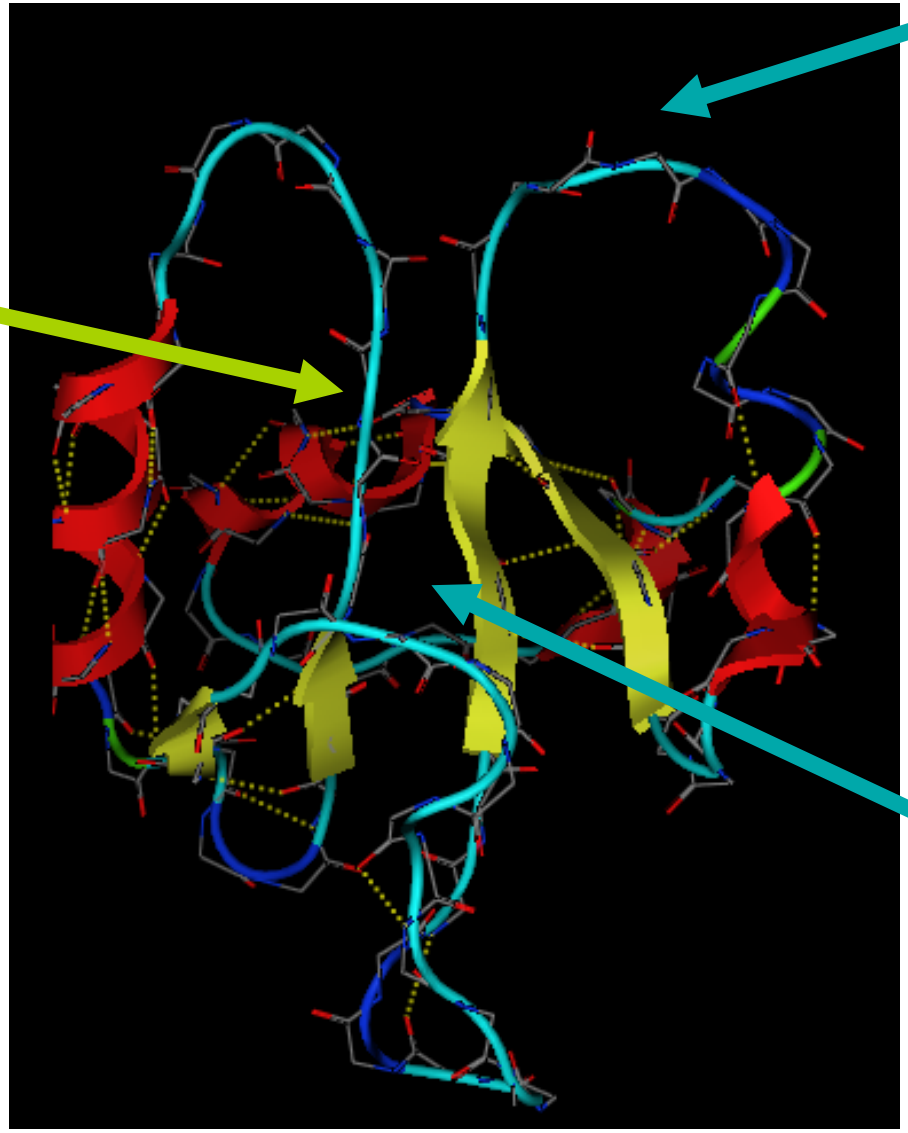
This part of the template will be deleted. A long bond will connect the ends. Energy minimization will cause commotion.

target
template



Results of Loop building: the wrong way

Deletion here results in straightening and shifting of the strand.

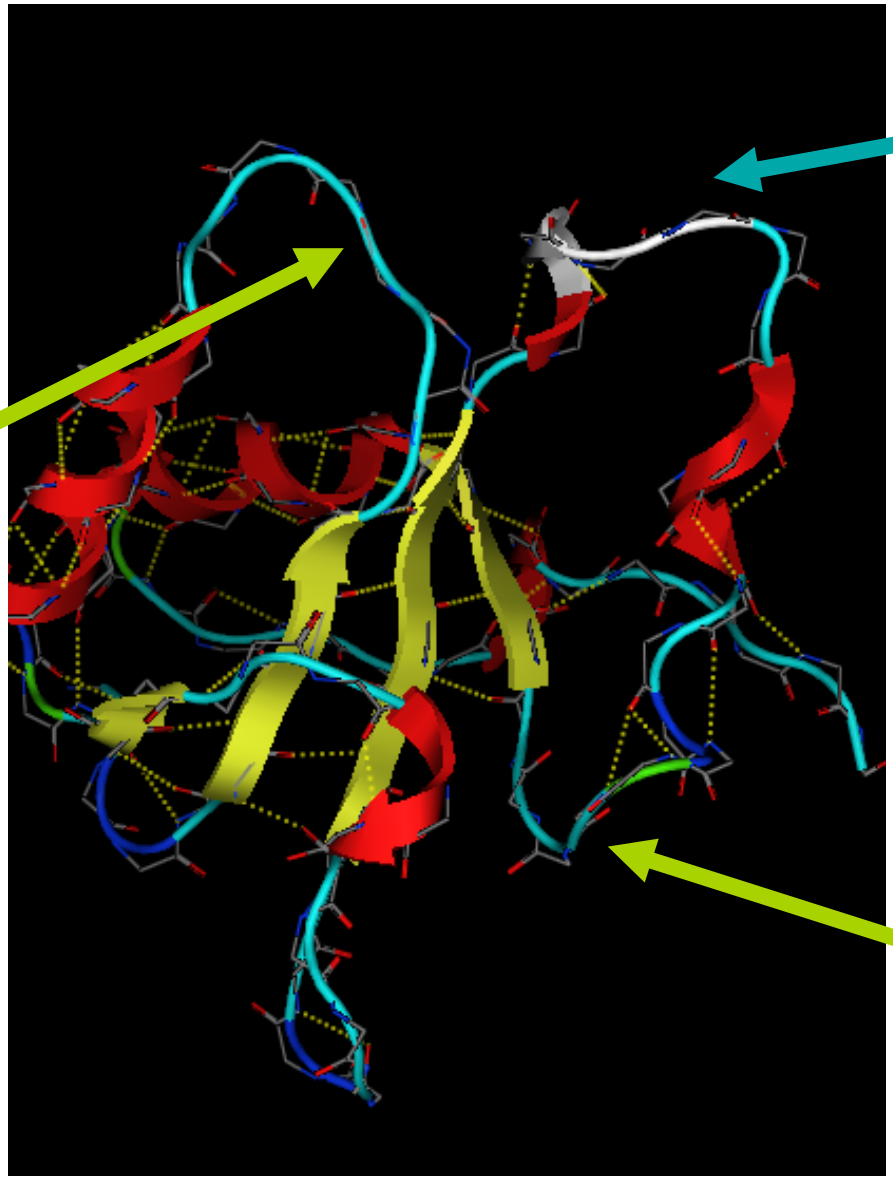


The resulting model has a short loop here.

Energy minimization has *pulled the sheet apart*.

No H-bonds in middle of sheet!

Setup of Loop building: the right way

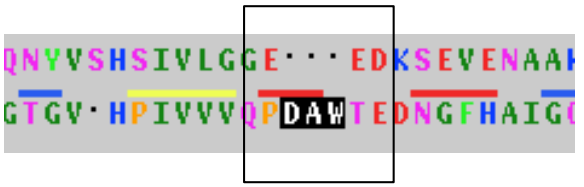


Location of indel moved to here.

1 against 2

1 target residue unaligned, 2 template residues deleted

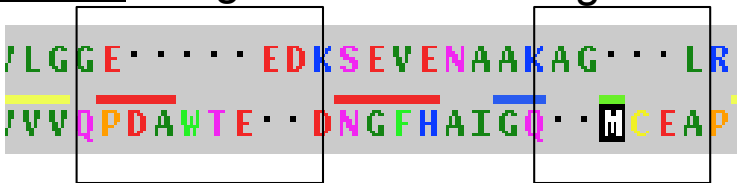
Old deletion
0 against 3



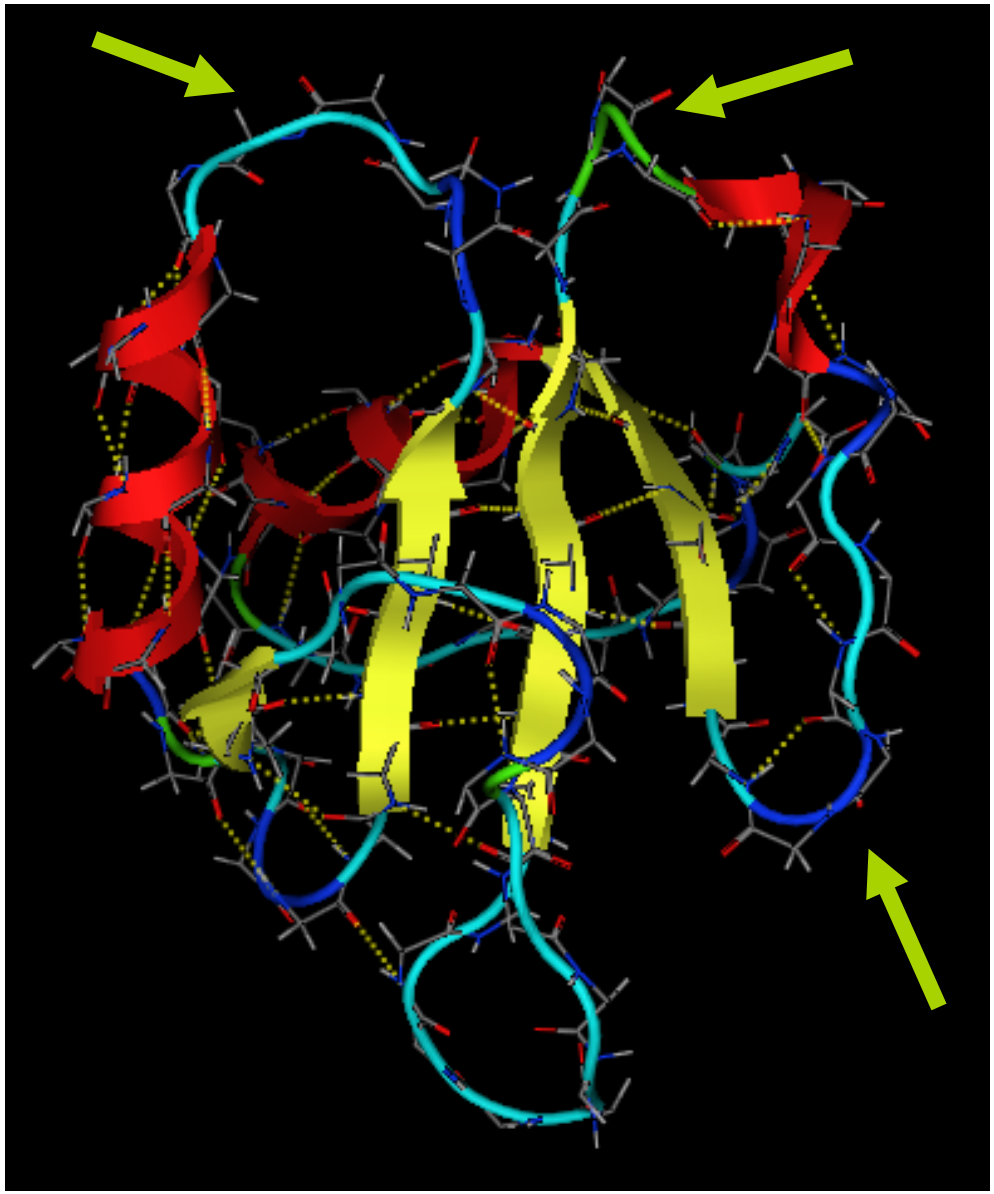
New deletions

2 against 5

2 against 3



Results of Loop building: the right way



New loops are at the ends of SSEs.

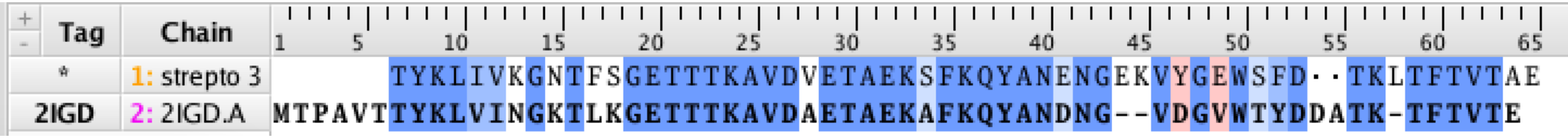
Sheet has retained H-bonds.

Take-home lesson: "Unaligning" makes loop search work better.

Exercise 6

Demonstration of the power of re-alignment

1. Load "**sequence 1**"
2. Run the PDB search if necessary. Load 2IGD.
3. Align (use the defaults)
4. Your alignment should look like this:
5. Save as "before.moe"

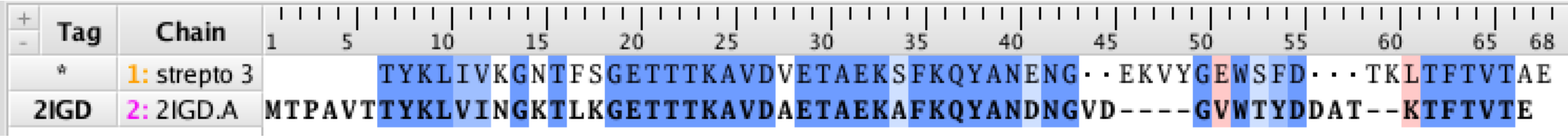


6. **Protein > Homology Model**
   Name **Current System:** automatic.moe
   Name **Output Database:** automatic.mdb
   Sequence: Chain #1
   Template: Chain #2
   Uncheck "disable C-terminal ..."
   Set **Models:** 3
   Forcefield: Amber12
   Leave everything else as is.
   OK

Exercise 6

Demonstration of the power of re-alignment

7. Close
8. Open "before.moe"
9. *Push* the alignment to make it look like this



10. **Protein > Homology Model**
    Name **Current System:** realigned.moe
    Name **Output Database:** realigned.mdb
    Sequence: Chain #1
    Template: Chain #2
    Uncheck "disable C-terminal ..."
    Set **Models:** 3
    Forcefield: Amber12
    Leave everything else as is.
    OK

Exercise 6

Demonstration of the power of re-alignment

11. Close
12. Open "automatic.mdb"
13. *Mouse-over right-mouse > send to MOE* each entry. When it asks whether to clear system, say No.
14. Protein | Geometry | Phi-Psi plot (Ramachandran plot)
15. How many outliers? Where are they?

16. **Look at the results**

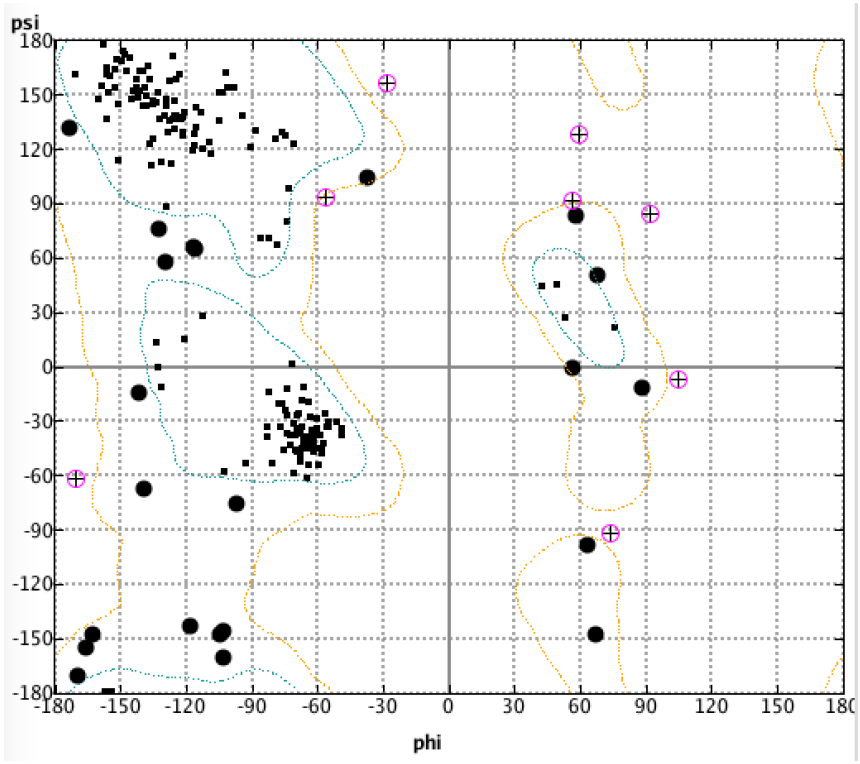
17. Close
18. Open "realigned.mdb"
19. *Mouse-over right-mouse > send to MOE* each entry. When it asks whether to clear system, say No.
20. Protein | Geometry | Phi-Psi plot (Ramachandran plot)
21. How many outliers? Where are they?

22. **Look at the results**

Demonstration of the power of re-alignment

automatic alignment

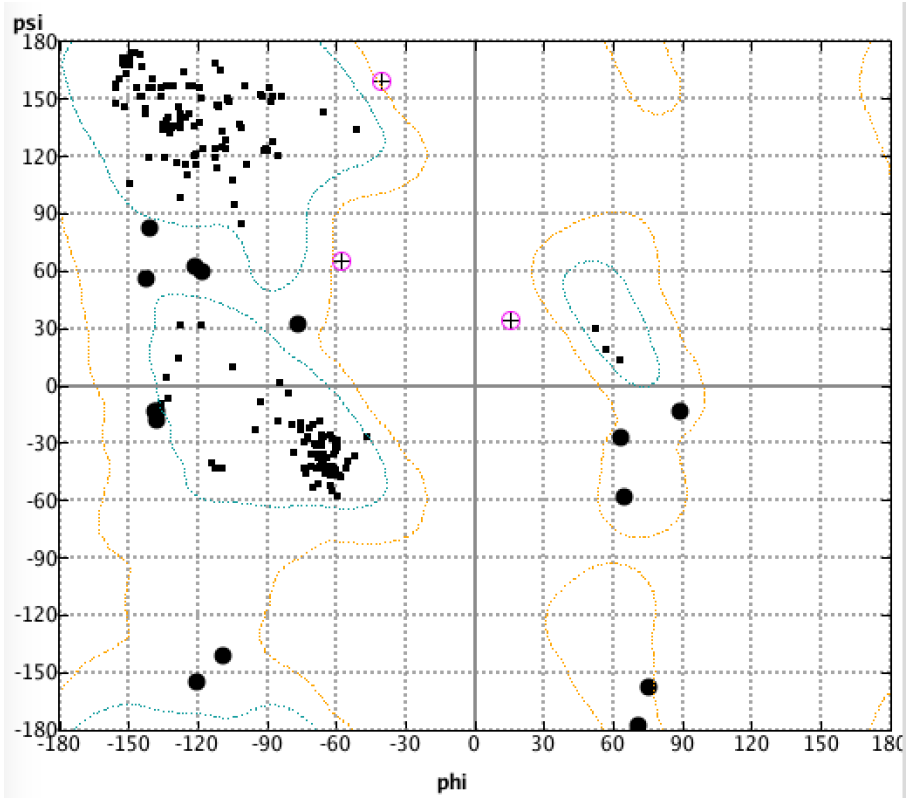
```
ENGEKVYGEWSFD··TKLTFTV
DNG--VDGVWTYDDATK-TFTV
```



Lots of outliers.

Re-alignment

```
ENG··EKVYGEWSFD···TKLTFTV
DNGVD----GVWTYDDAT--KTFTV
```



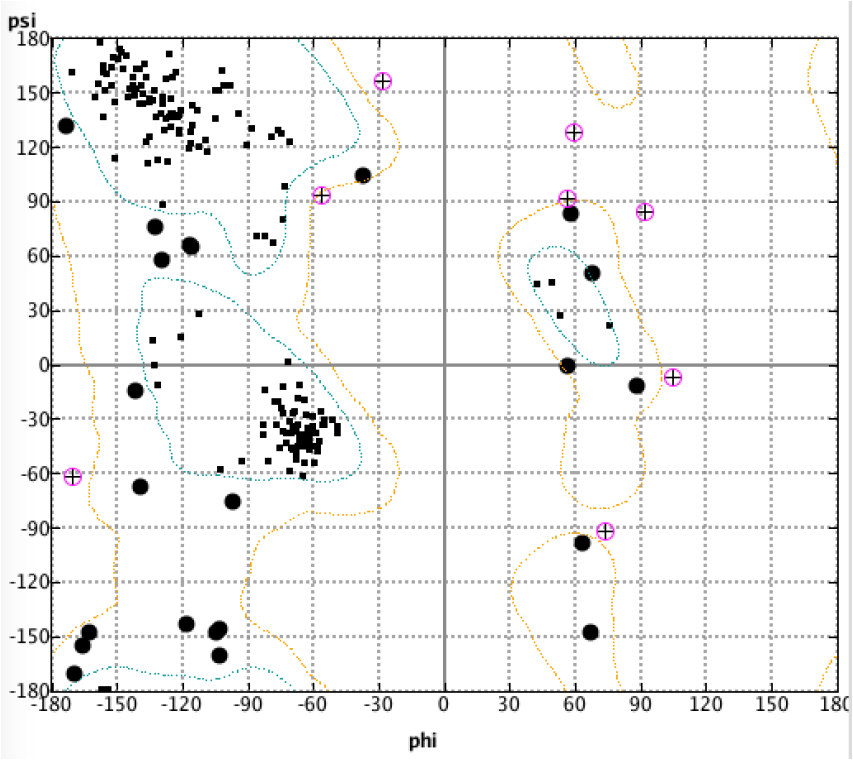
a few outliers.

Exercise 6

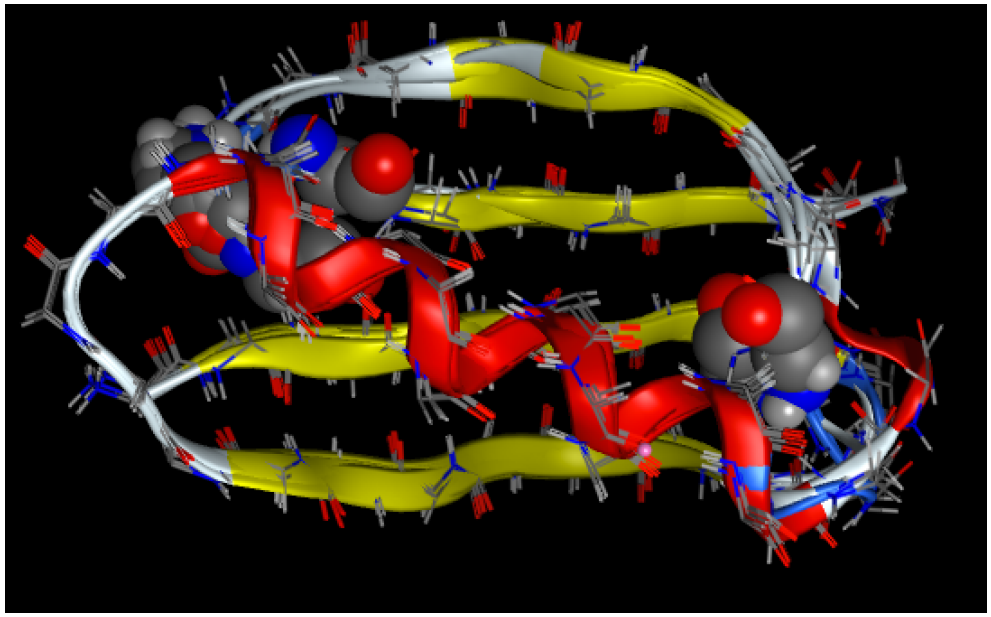
Demonstration of the power of re-alignment

16.**Look at the results: automatic alignment**

Tag	Chain	1 5 10 15 20 25 30 35 40 45 50 55 60 65
*	1: strepto 3	TYKLIVKGNTFSGETTTKAVDVETAEKSFKQYANENGEKVYGEWSFD··TKLTFTVTAE
2IGD	2: 2IGD.A	MTPAVTTYKLVINGKTLKGETTTKAVDAETAEKAFKQYANDNG--VDGVWTYDDATK-TFTVTE



Lots of outliers.

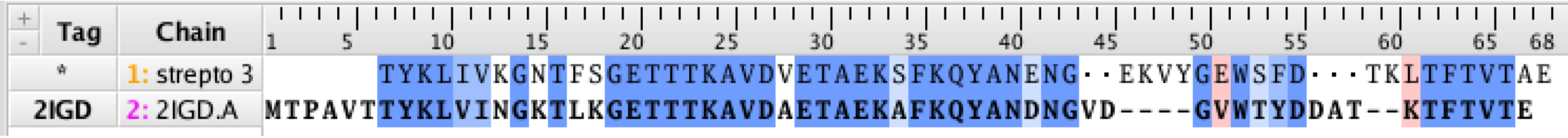


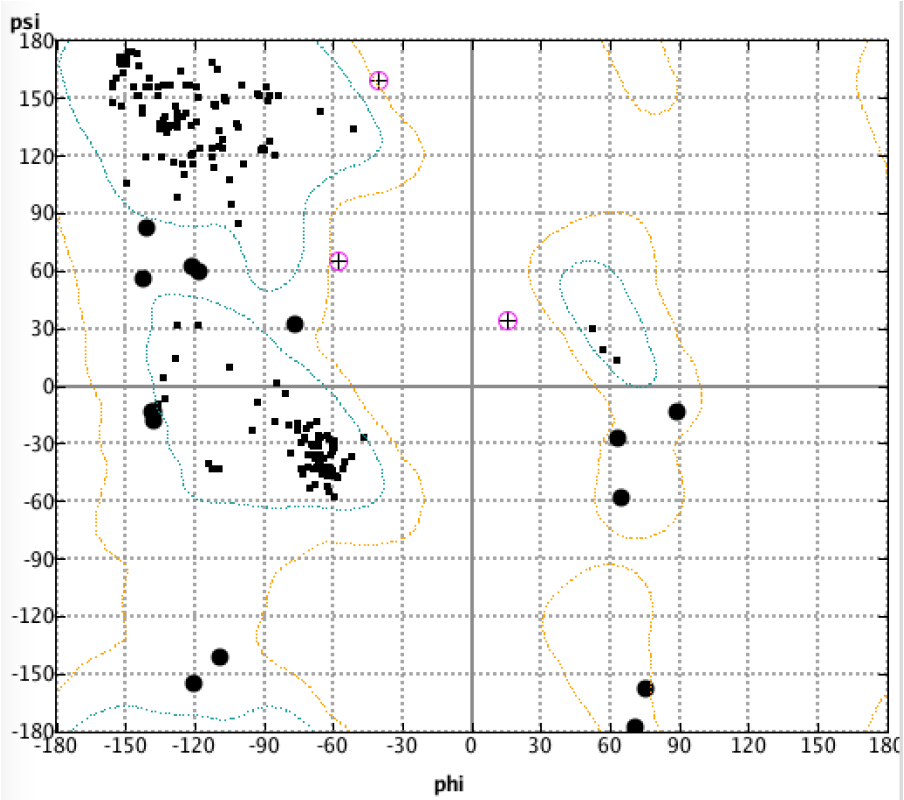
Not coincidentally, in the indels

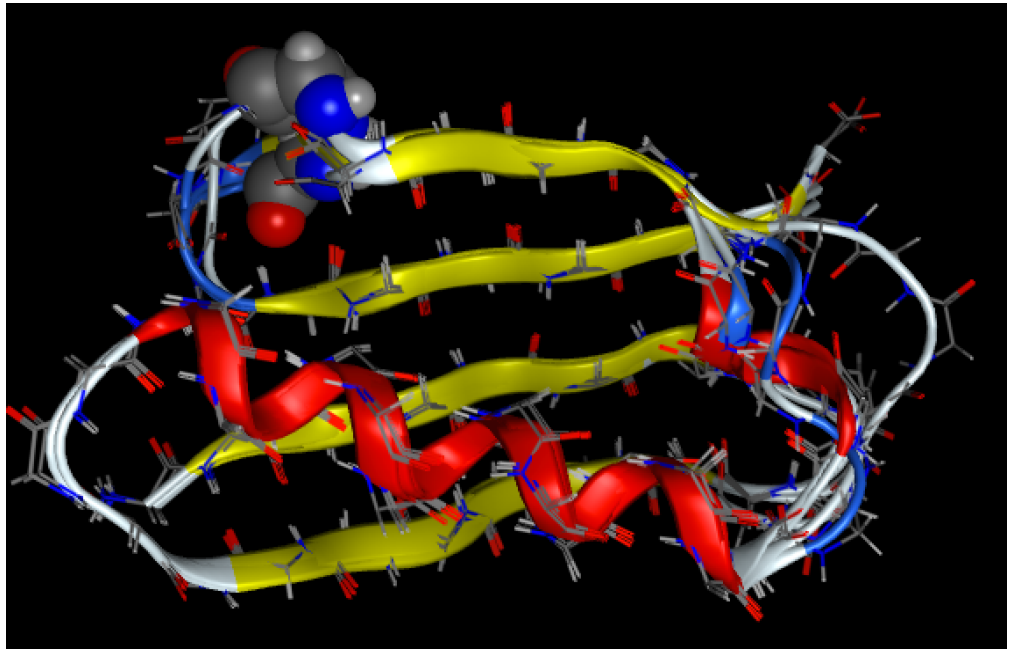
Exercise 6

Demonstration of the power of re-alignment

22. **Look at the results**







Few outliers.

Still in the indels

Exercise 6

Continue working with your model

23. Find outliers in the **phi-psi plot.**
24. Select the sequence around one of the outliers. **Unfix** it. **Invert** selection. **Fix.**
25. **run** '**gizcolorf.svl**'
    Blue means little force. Red means high force, on an atom.
26. **run** '**gizmin.svl**'
27. Pull and push to make regular H-bonds, common secondary and local structures. Make sure it doesn't get too red!
28. Keep **Protein | Geometry | Phi-Psi** plot *on*. You are done when all of the outliers have moved to the allowed regions.
29. Unfix additional residues if necessary.
30. Save and upload the MOE file as "**Exercise 6**". Due Feb 17.

Review questions

- After Homology Model, are we done making a homology model?
- If you have a glycine and it is strictly conserved, where might it lie in the phi-psi (Ramachandran) plot?
- How do you as a protein modeler feel about a buried, unsatisfied, backbone hydrogen bond donor or acceptor?
- How can we guess which donor H-bonds to which acceptor?
- Are conserved sequence regions with no gaps always SCRs?
- Are conserved sequence regions with no gaps usually SCRs?
- Are side chain rotamers conserved in SCRs? (usually, never, always)