Molecular Modeling 2020
lecture 17 -- Fri Mar 27

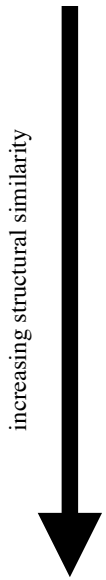
Ramachandran plot

Local structure

Handedness

SCOP -- a hierarchy

- http://scop.berkeley.edu



(1) class

(2) fold

(3) superfamily

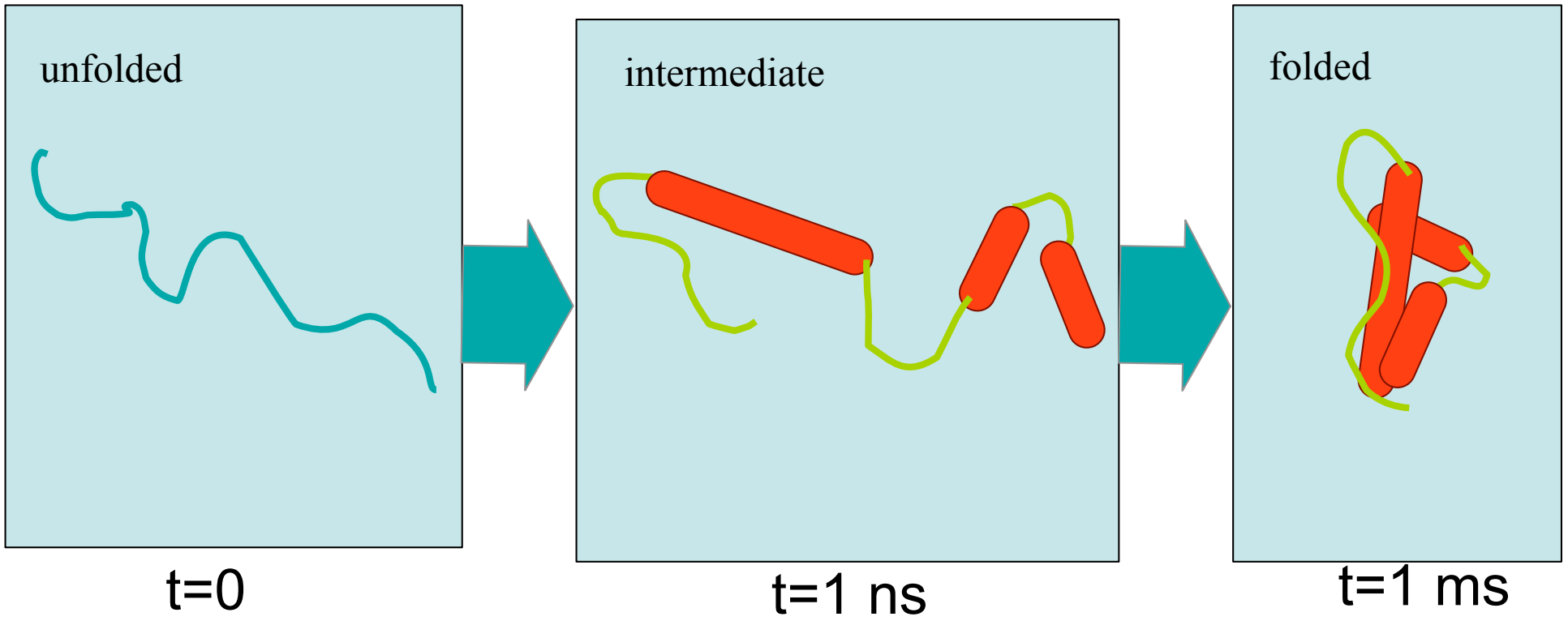
(4) family

(5) protein

(6) species

SCOP's hierarchy is sequence centered

Folding -- another hierarchy?



t = time after leaving the ribosome

More about protein folding in later lectures

Structural classification viewed along the folding hierarchy

Classification

Secondary structure

Local structure

Super-secondary structure

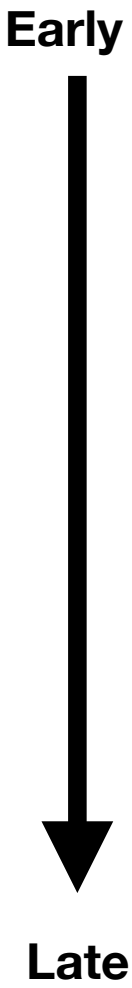
Domains (tertiary structure)

Side chain rotamers

Domain-domain interactions

Quaternary structure

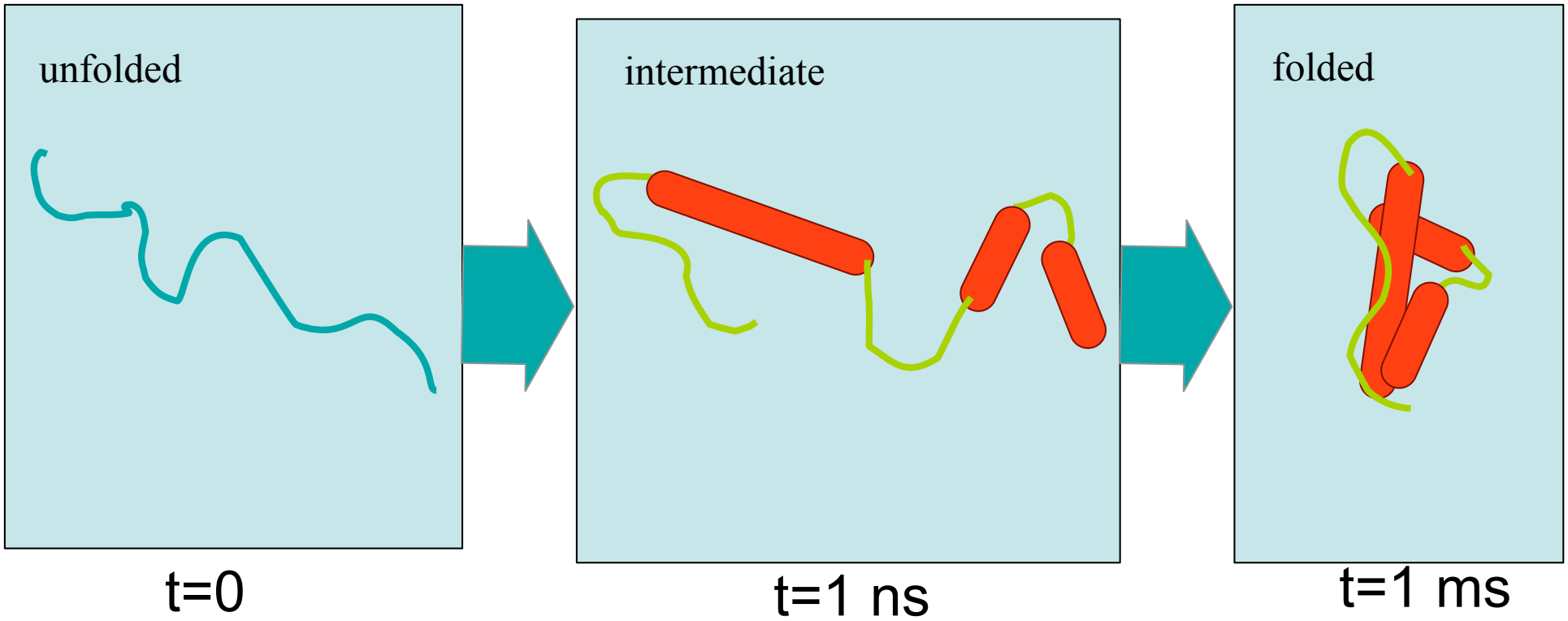
Super-quaternary structure



Why does local sequence predict structure?

Early in the process of folding (nsec timescale)

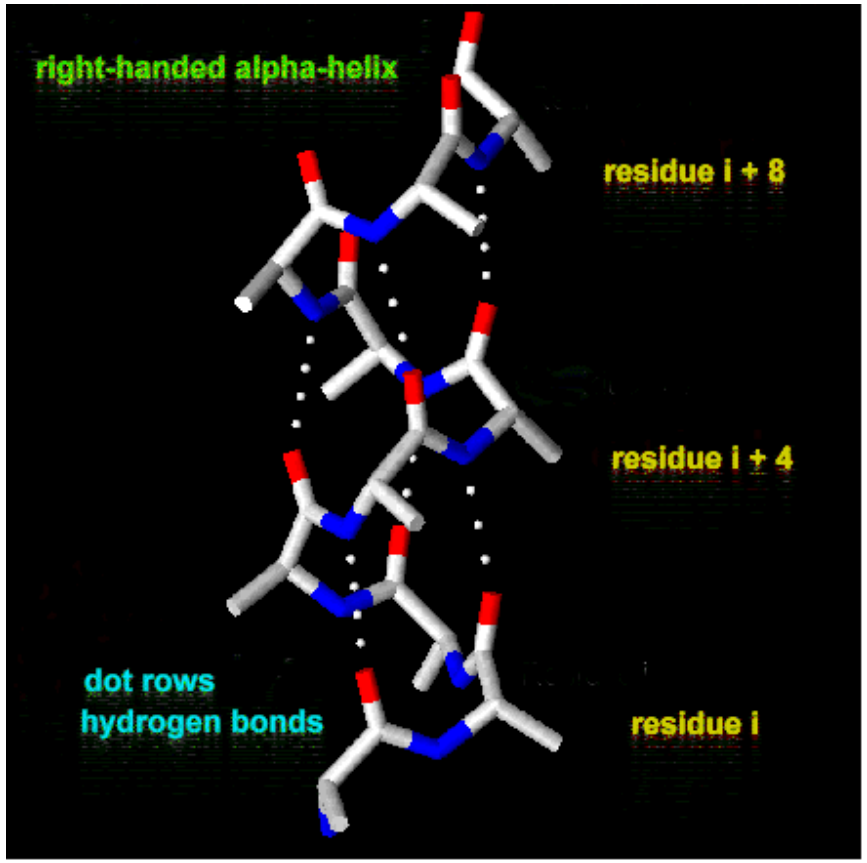
local structures form in the polypeptide chain which guide the formation of tertiary structure.



Alpha helix

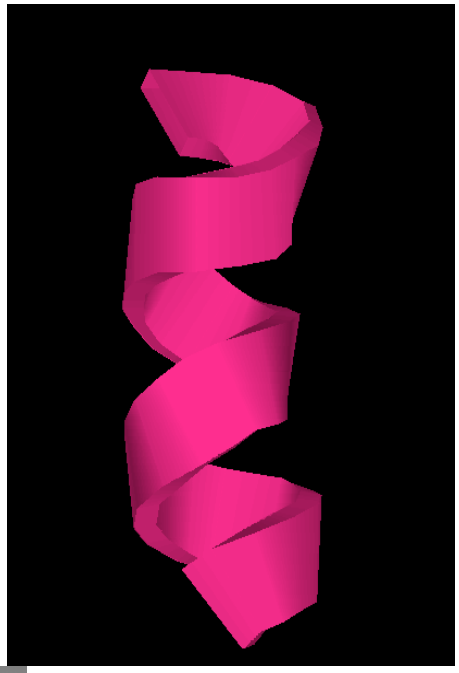
Right-handed helix. H-bond is from the oxygen at *i* to the nitrogen at *i+4*. α-helices have an overall dipole because the H-bonds are all in the same direction. Must be > 3 residues.

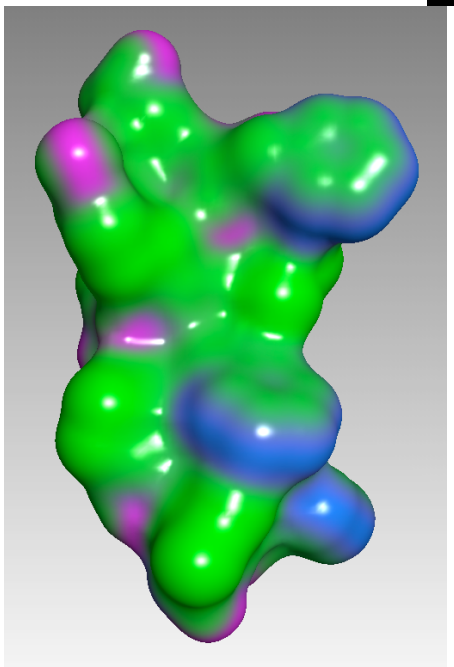
H-bond rule for donor to acceptor (NH->O): i to i+4



dipole

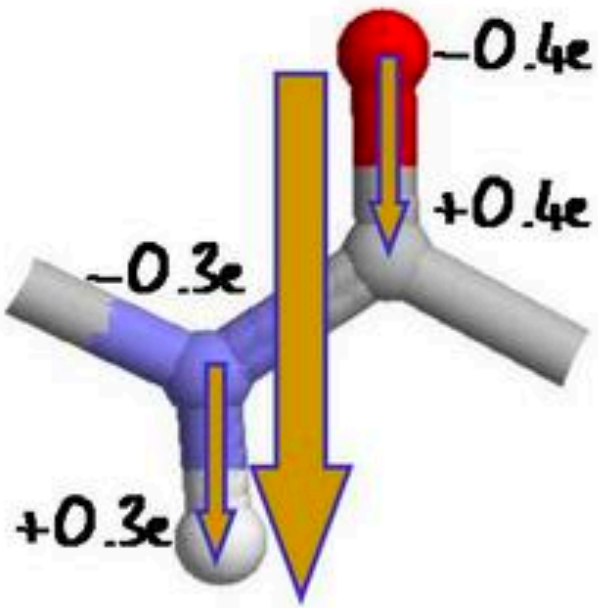
right-handed





Helices do not look “cylindrical”.

ALPHA-HELIX DIPOLE 1



- The peptide group has a strong dipole moment due to partial charges on NH and CO groups.

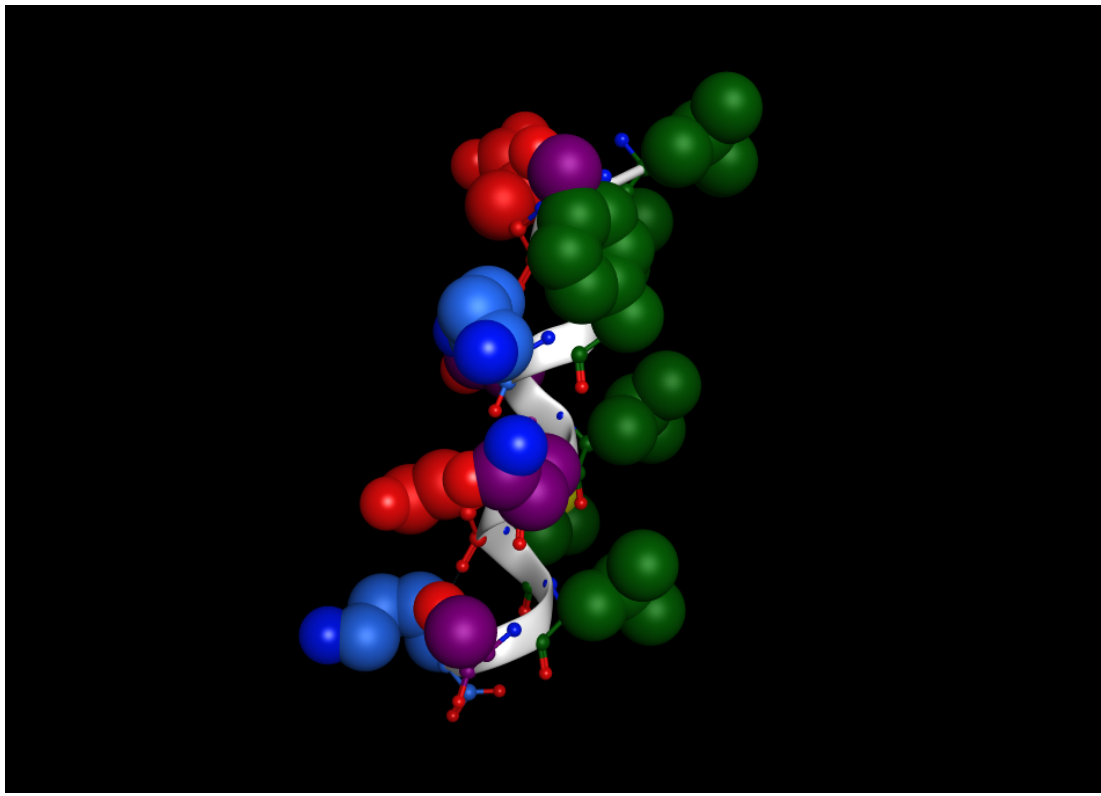
https://web.stanford.edu/class/sbio228/public/lectures/lec2/Lecture2/The_Alpha_Helix/images/Alpha-Helix_Dipole_1.jpg

Sequence pattern for the amphipathic alpha helix

- nppnnpp,
  where n = non-polar, p = polar
- Example:

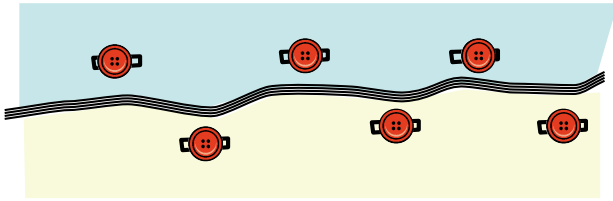
  LSELFKNLQDMLSK

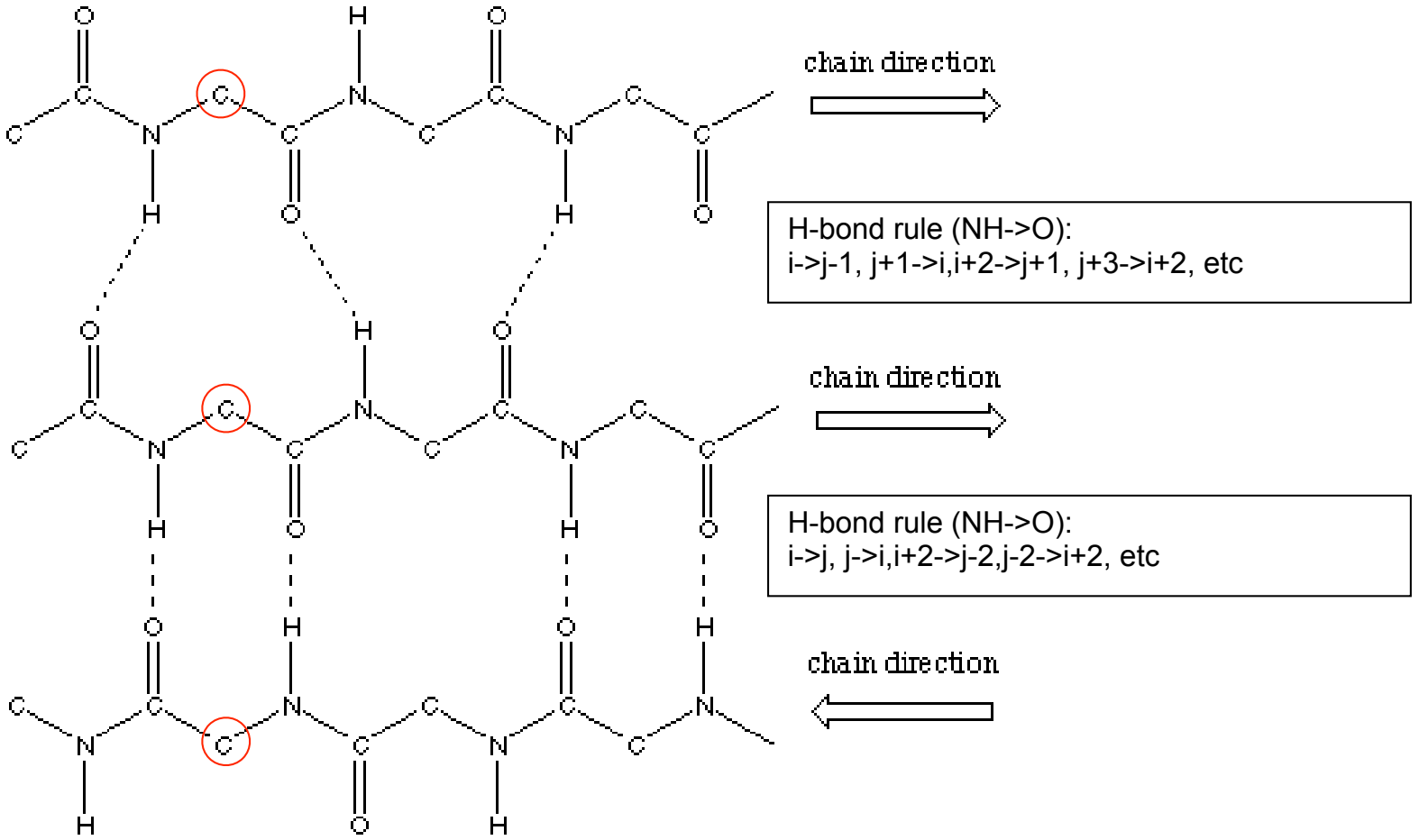
Hydrophobic all on one side



The helix is held together by the hydrophobic effect.
Sticks to other amphipathic helices.

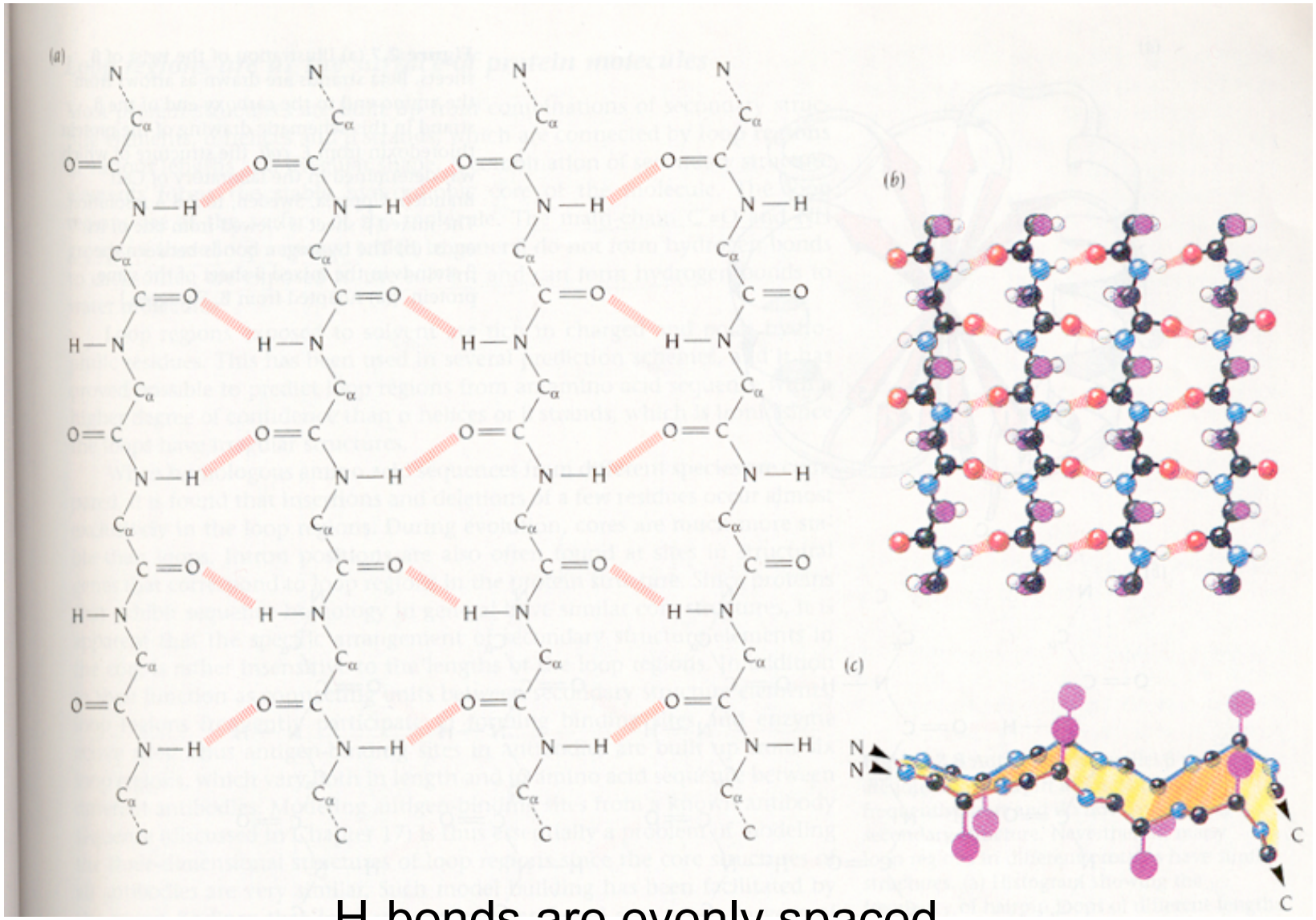
beta sheets





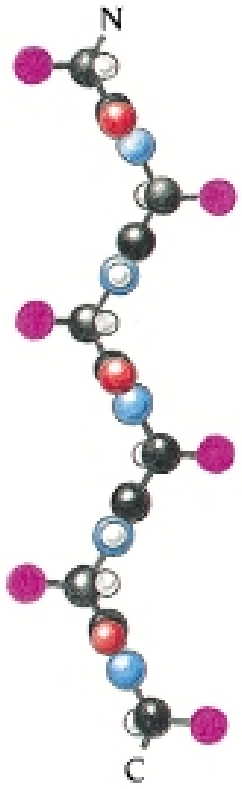
In both parallel and anti-parallel,
sidechains alternate above and below the plane of the sheet.

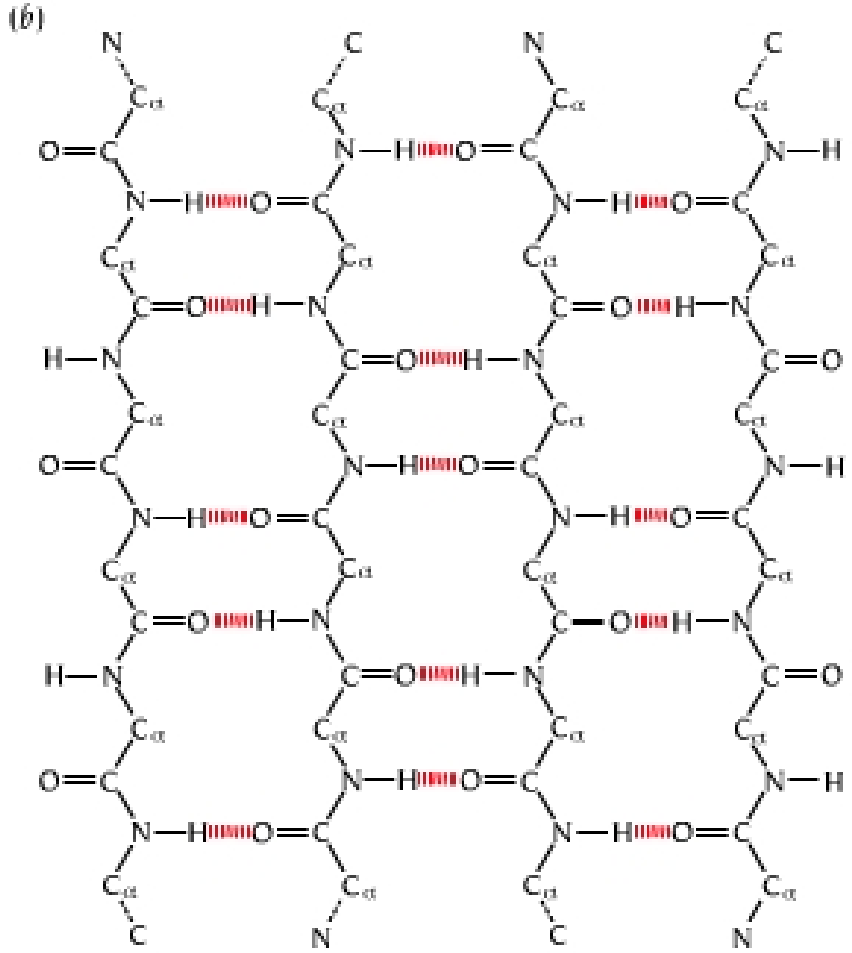
Parallel beta sheet



H-bonds are evenly spaced.
H-bonds are not 90° to the chain.

Anti-parallel beta sheet

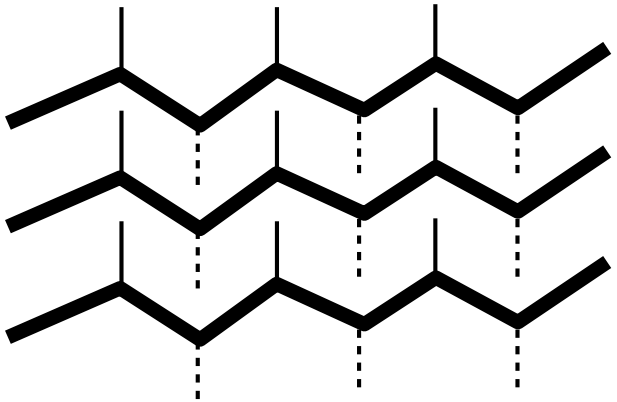




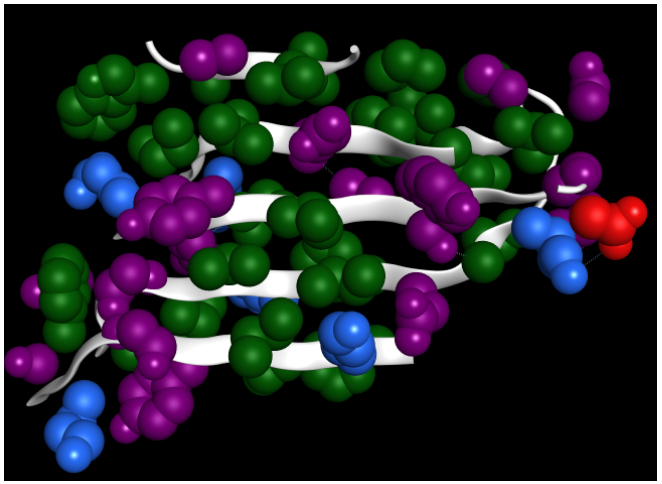
H-bonds are unevenly spaced.
H-bonds are 90° to the chain.

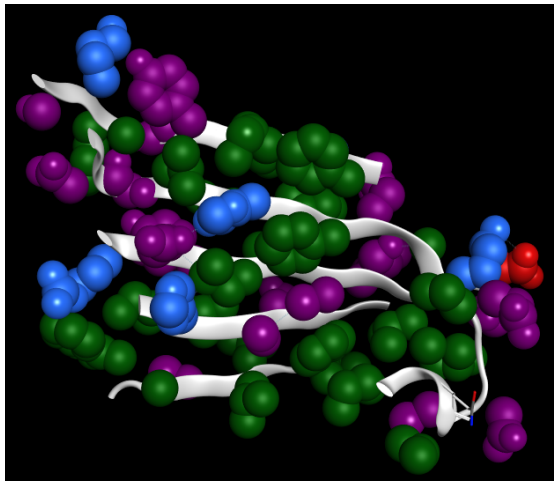
Sequence patterns for beta sheet

- npnp, where n=non-polar, p=polar
- nnnn



Non-polar residues (green, purple) mostly on the face.





Charged residues (blue, red) mostly on the ends.

Secondary structure using matrices

An H-bonding pattern can be expressed using "augmented" matrix notation.

next H-bond donor
next H-bond acceptor

=

multiply by donor	multiply by acceptor	add to donor
multiply by donor	multiply by acceptor	add to acceptor

X

current H-bond donor
current H-bond acceptor

For example, for an alpha helix....

150
146

=

1	0	1
1	0	-3

X

149
145

In a helix, donor NH is always +4 to acceptor O.

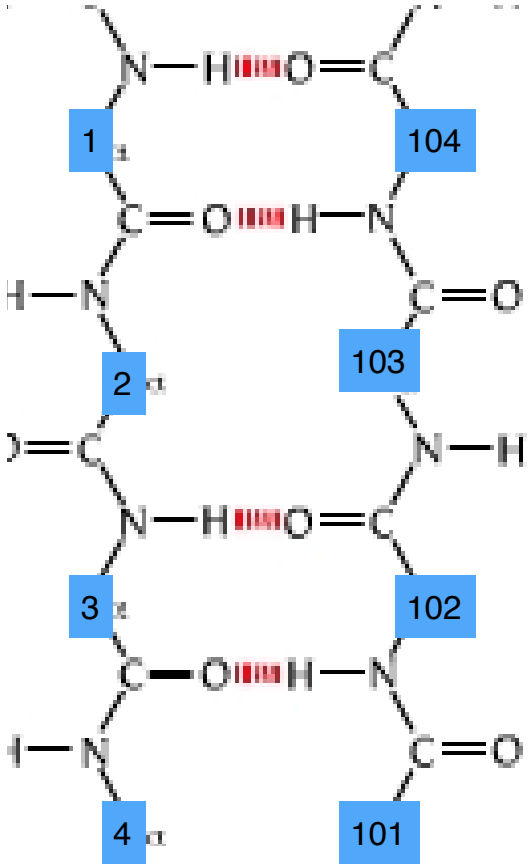
Secondary structure using matrices: antiparallel sheet

handshake

$$\left[\begin{array}{cc|c} 0 & 1 & 0 \\ 1 & 0 & 0 \end{array}\right]$$

skip

$$\left[\begin{array}{cc|c} 0 & 1 & 2 \\ 1 & 0 & -2 \end{array}\right]$$



Use the augmented matrices to find the next H-bond before/after (donor,acceptor)=(102, 3) in a antiparallel sheet

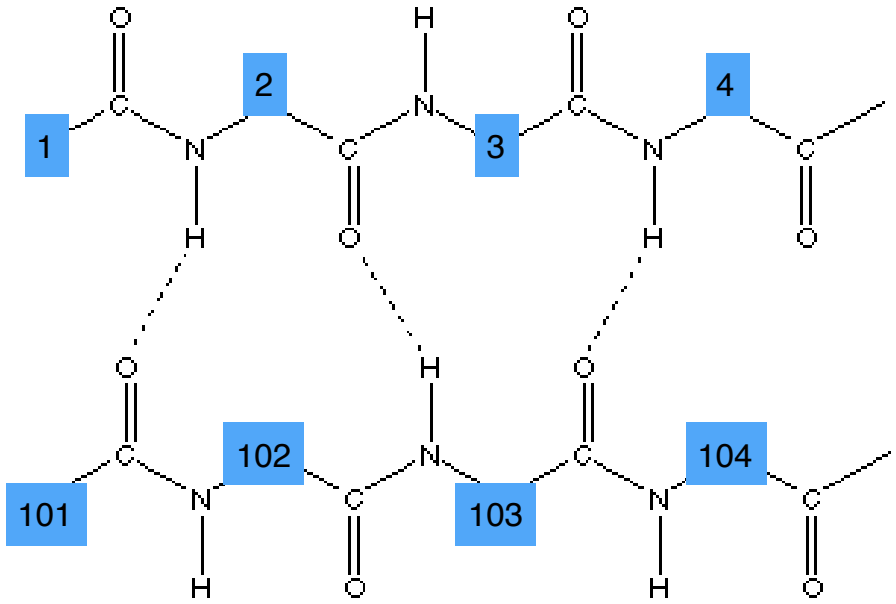
Secondary structure using matrices: parallel sheet

C

0	1	2
1	0	0

N

0	1	0
1	0	-2



Use the augmented matrix to find the next H-bond before/after (donor,acceptor)=(103, 2) in a parallel sheet

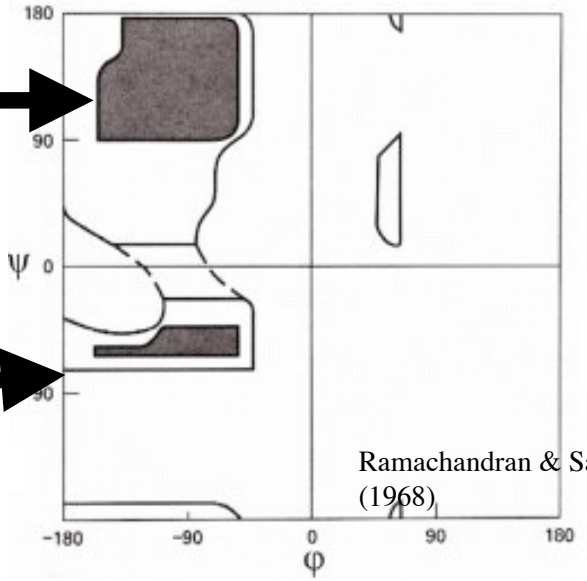
The Ramachandran Plot

beta sheet

-180° < Φ < 0°
90° < Ψ < 180°

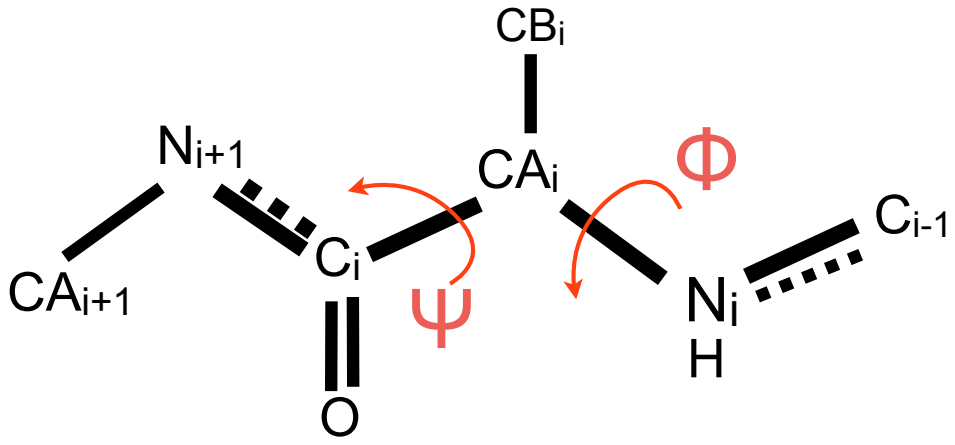
alpha helix

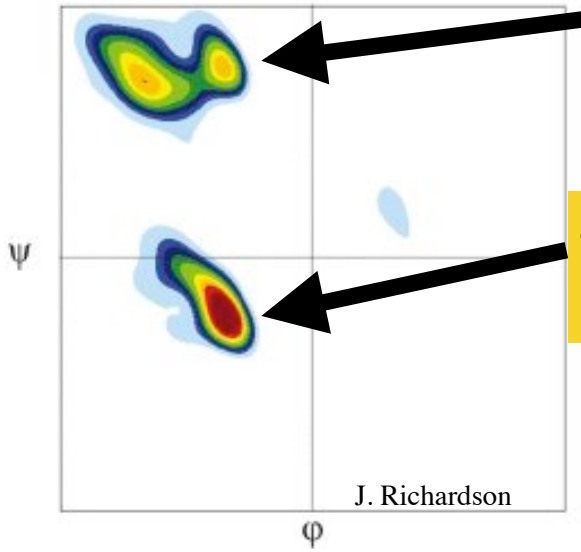
-100° < Φ < -40°
-80° < Ψ < -30°



Ramachandran & Sasisekharan (1968)

(a)





J. Richardson

Φ = -139°
Ψ = 135° β

Φ = -58°
Ψ = -47° α

Predicting secondary structure from primary structure

assumes

1. Secondary structures have sequence patterns
2. Those patterns are conserved across homolog proteins.

predicting burial

- Early methods for predicting the structure of a protein used the chemical characteristics of amino acids -- hydrophobic versus hydrophilic. If a stretch of aminoacids was hydrophobic, it was most often found in the core of the protein, and the opposite was true if a stretch was hydrophilic. Two scales were proposed -- Kyte-Doolittle and Hopp-Woods.

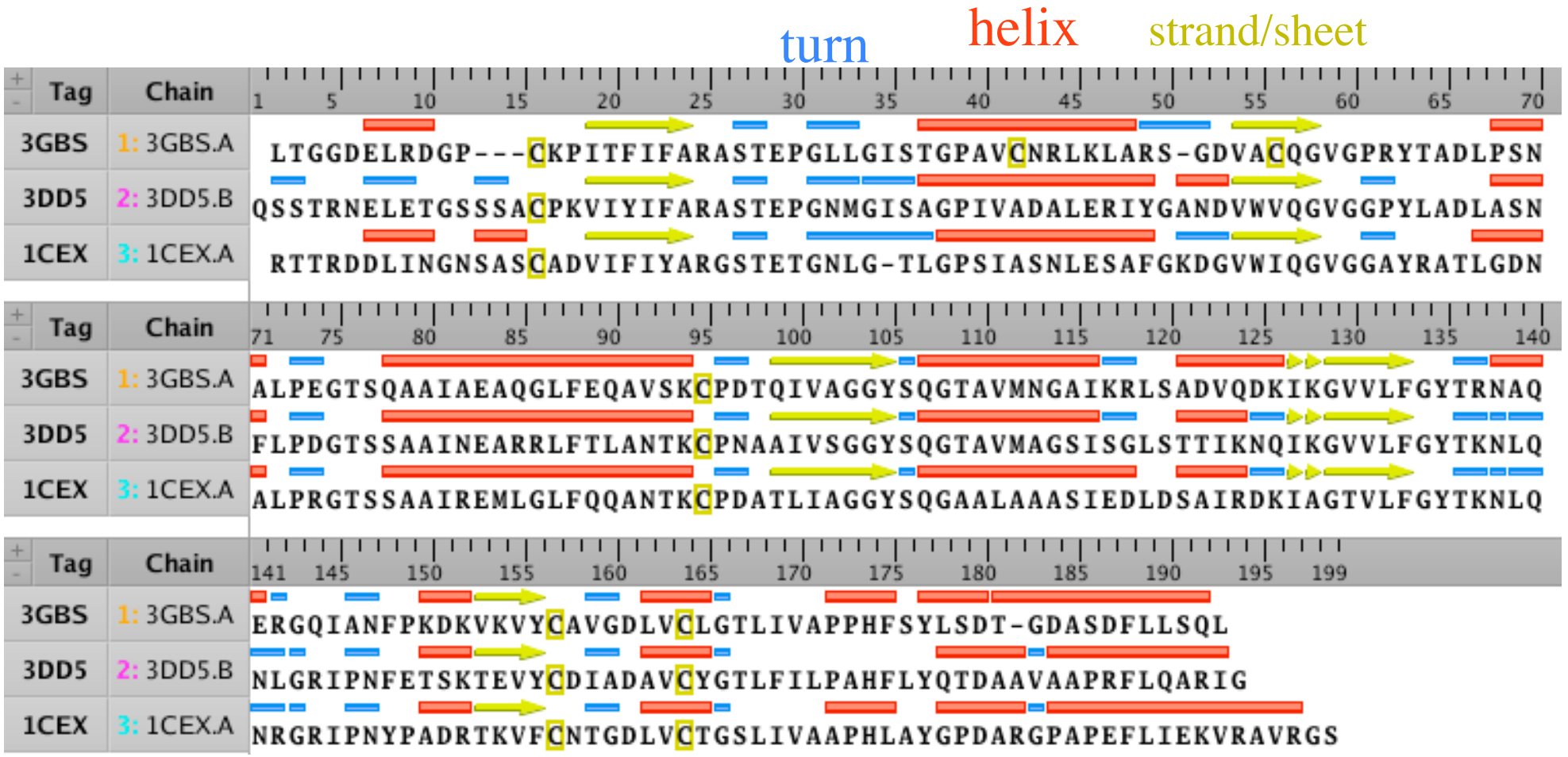
	Kyte-Doolittle	Hopp-Woods
Alanine	1.8	-0.5
Arginine	-4.5	3.0
Asparagine	-3.5	0.2
Aspartic acid	-3.5	3.0
Cysteine	2.5	-1.0
Glutamine	-3.5	0.2
Glutamic acid	-3.5	3.0
Glycine	-0.4	0.0
Histidine	-3.2	-0.5
Isoleucine	4.5	-1.8
Leucine	3.8	-1.8
Lysine	-3.9	3.0
Methionine	1.9	-1.3
Phenylalanine	2.8	-2.5
Proline	-1.6	0.0
Serine	-0.8	0.3
Threonine	-0.7	-0.4
Tryptophan	-0.9	-3.4
Tyrosine	-1.3	-2.3
Valine	4.2	-1.5

Hopp TP and Woods KR: Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78:3824, 1981.

Kyte J and Doolittle RF: A simple method for displaying the hydropathic character of a protien. J Mol Biol 157:105, 1982.

Try it: http://www.vivo.colostate.edu/molkit/hydropathy/

Secondary structure is strongly conserved among even remote homologs.

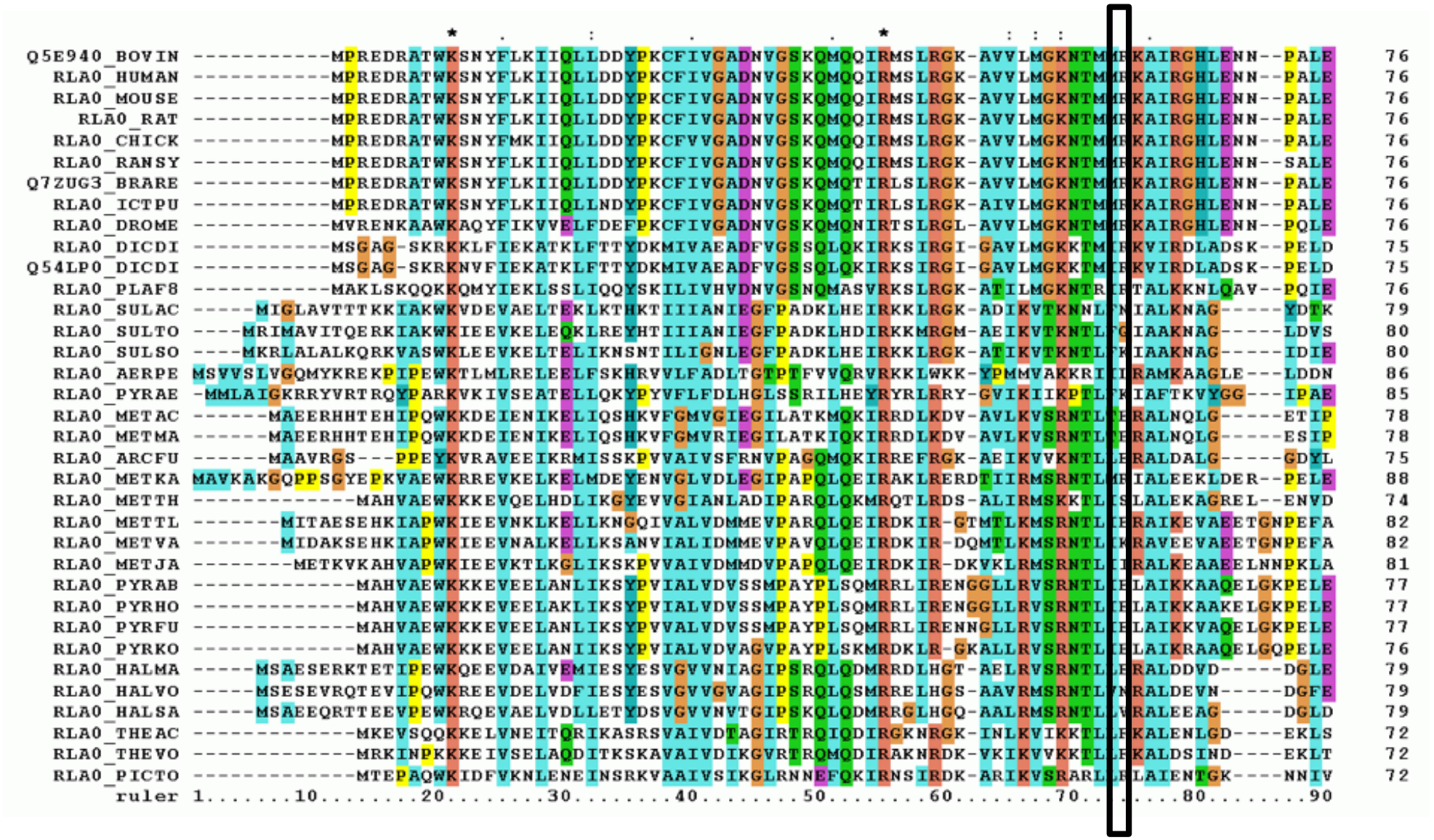


cutinases, 48 - 53% sequence identity.

Amino acid sequence profiles have patterns in them

- Positions in homologs conserve location, side chain conformation, packing environment.
- Evolution has sampled the low energy ways to fill each position.
- Multiple sequence alignments inform us about the nature of the position.
  - buried vs exposed.
  - alpha vs beta vs loop

First make a multiple sequence alignment



Each position in a MSA is a column of AA's representing the evolutionary history of one position.

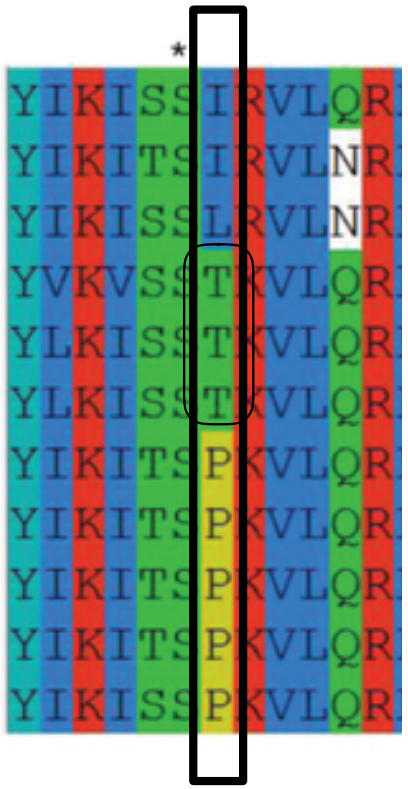
Sequence profiles are calculate from MSAs

$$P(T|7) = \frac{\sum_{i \,\forall\, S_{(7)}=T} w_i}{\sum_{all\ i} w_i}$$

Prob of Thr@ position 7 is the sum of the weights.

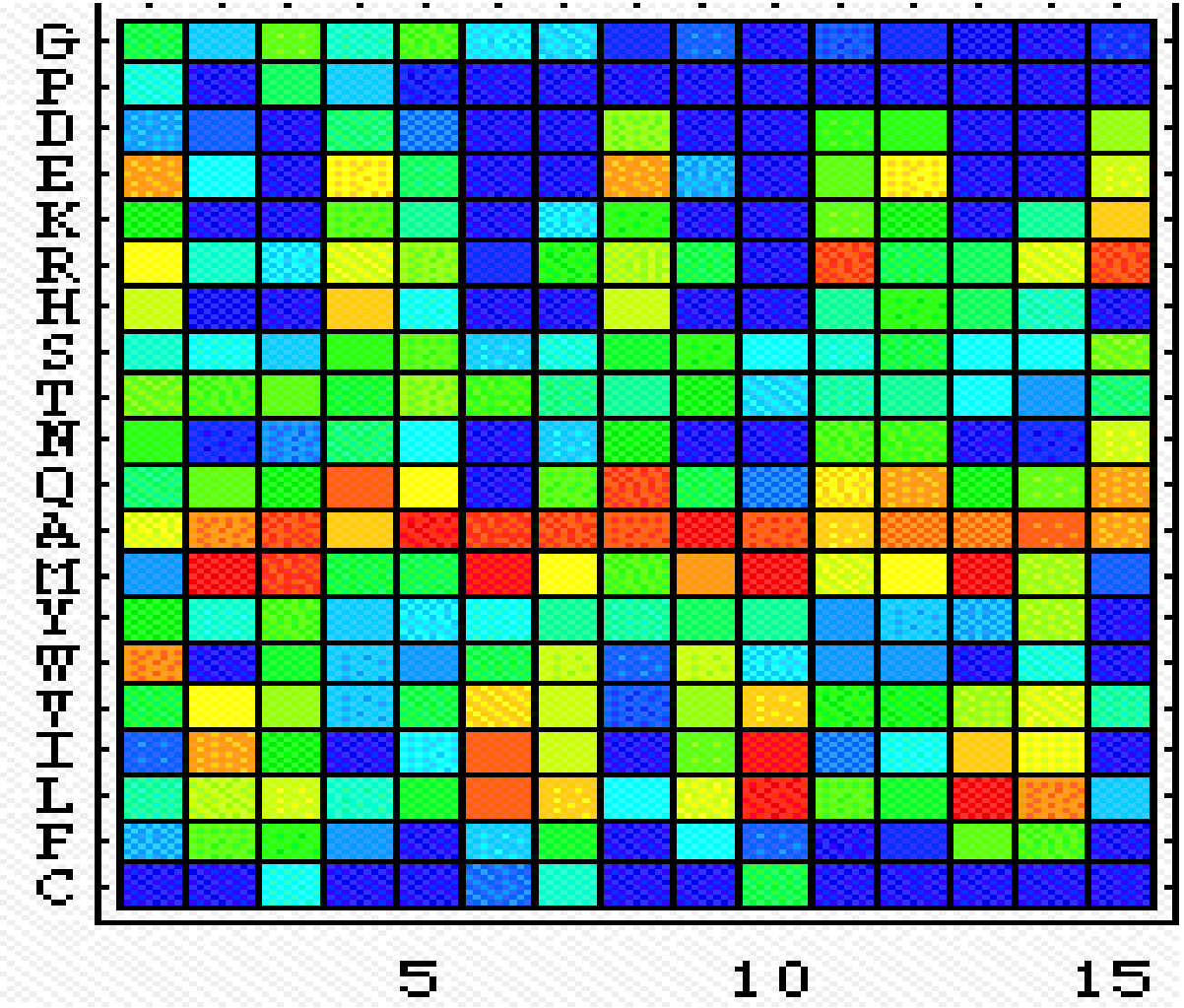
w_1
w_2
w_3
w_4
w_5
w_6
w_6
w_7
w_8
w_9
w_1
w_1

Sequences in the MSA are "weighted".



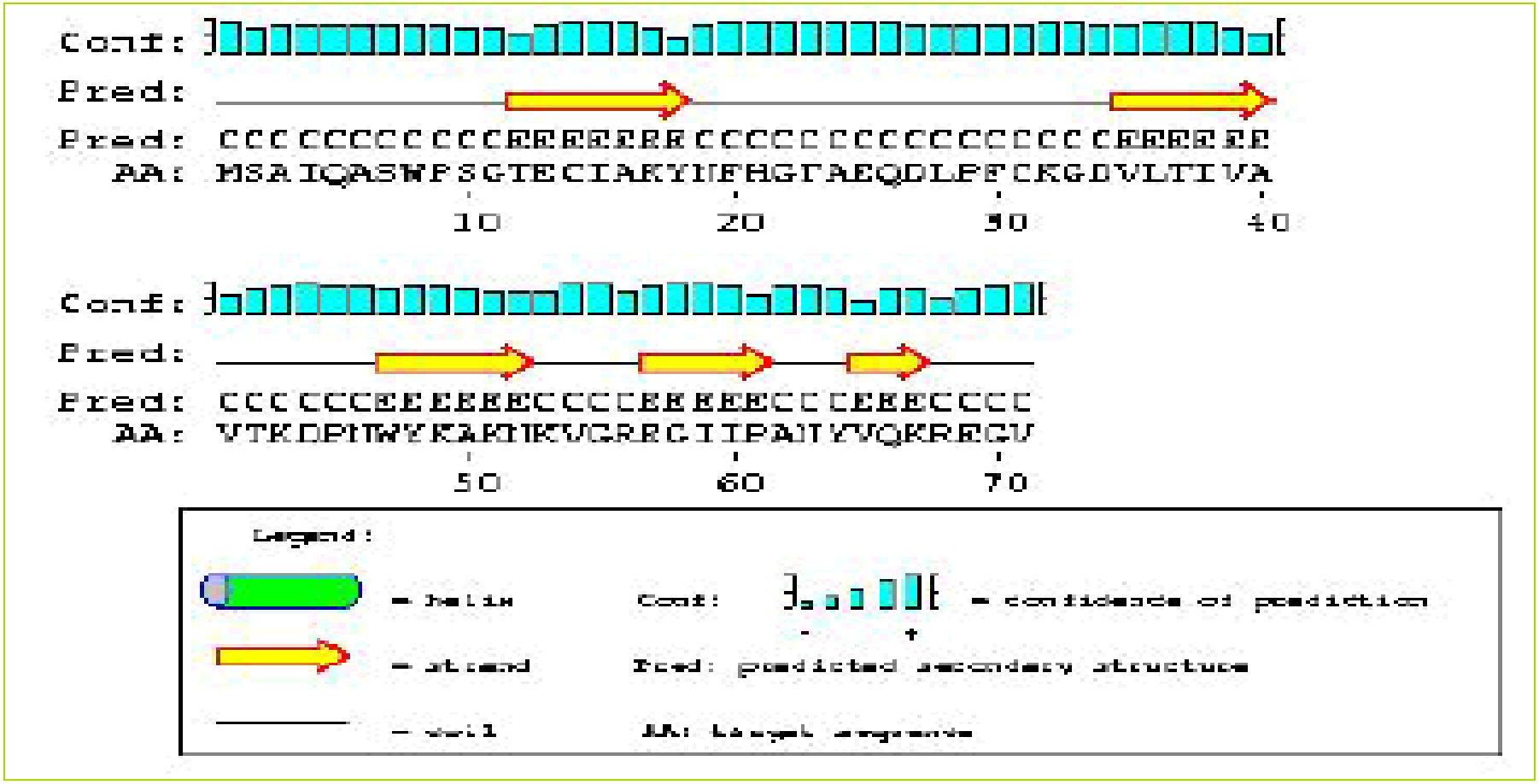
[The probability of amino acid *T* at position *7* is the sum of the sequence weights w_i over all sequences *i* such that the amino acid at position *7* of that sequence is *T*, divided by the sum over the sequence weights w_i.]

A sequence profile is a 20xN matrix of AA probabilities



PSI-pred-- secondary structure from **profiles**

• PSI-PRED (Jones et al.) is currently the best server for secondary structure prediction, based on an artificial neural network that connects a profile (**Psi-Blast** output) with known protein secondary structure Predictions are assigned *confidences*. A window of 15 is used to predict the central residue. Accuracy claimed to be 76-78%.

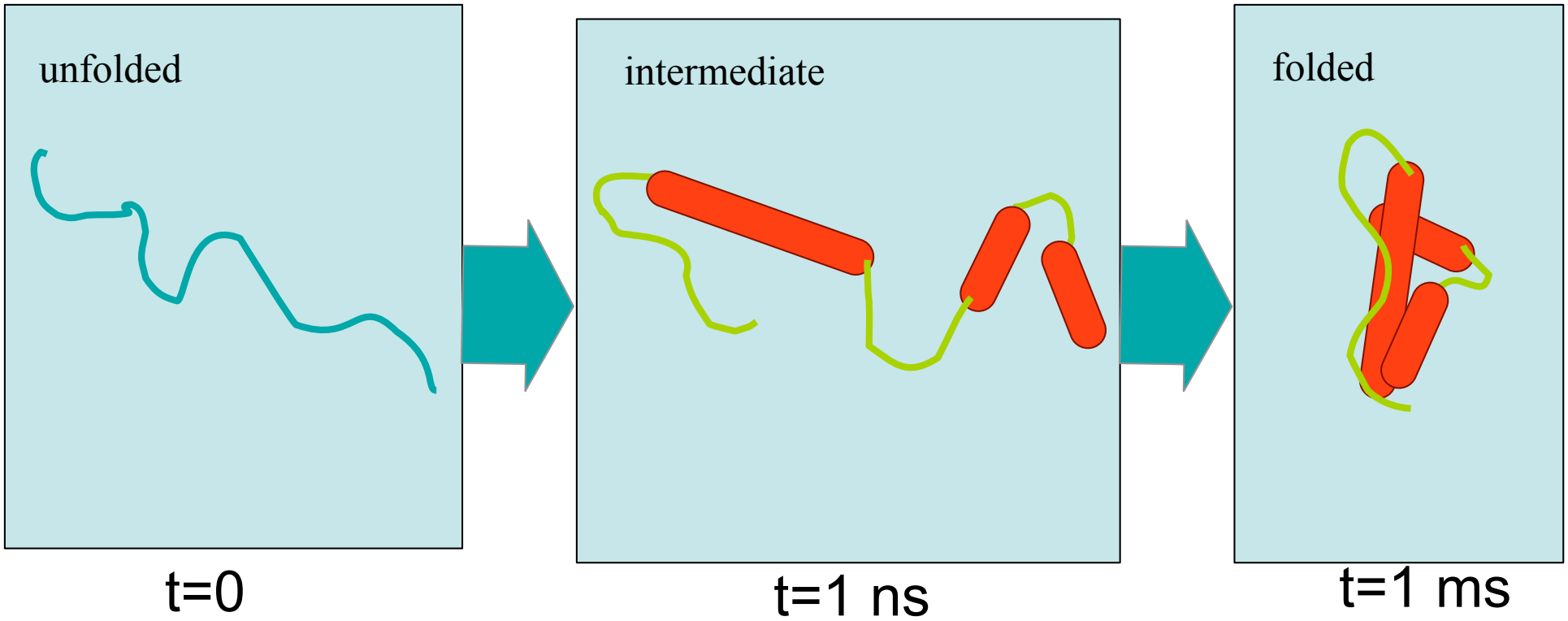


The PSIPRED protein structure prediction server L. J. McGuffin, K. Bryson, D. T. Jones (2000) *Bioinformatics* 16 (4) p. 404-405

Why does local sequence predict structure?

Early in the process of folding (nsec timescale)

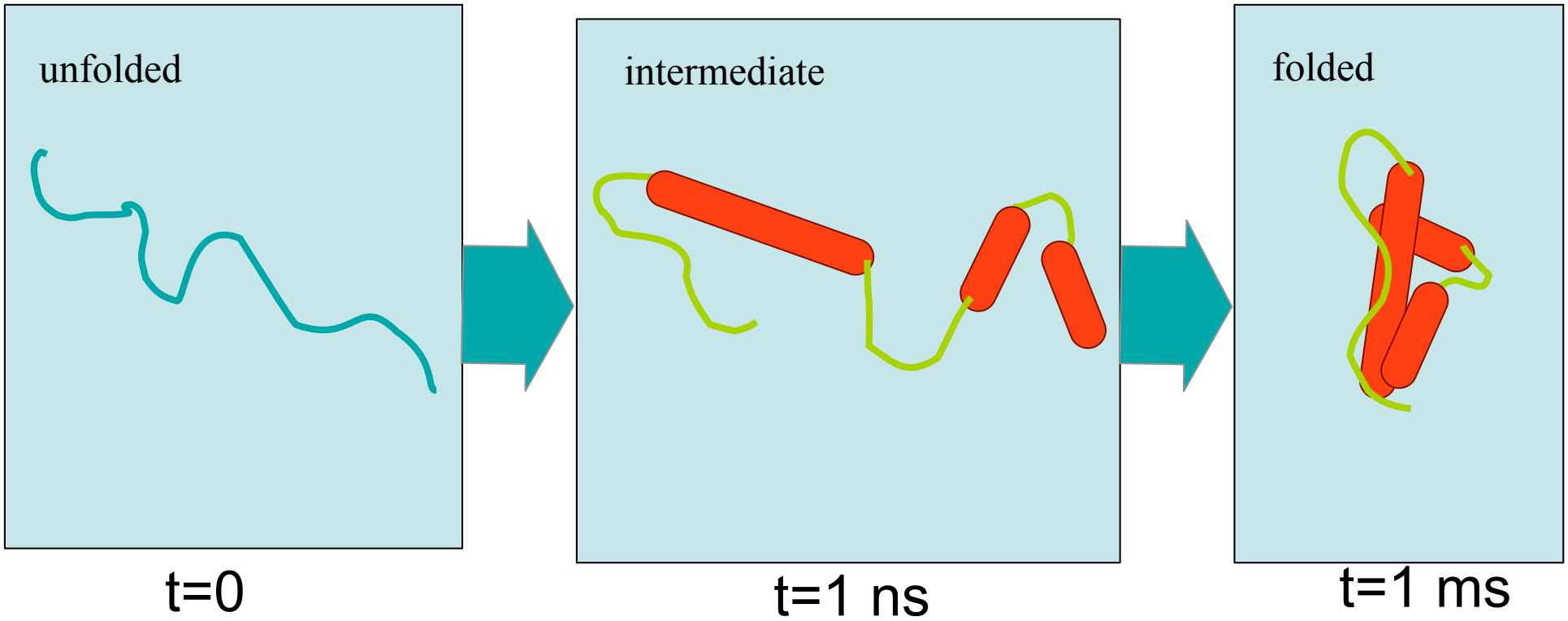
local structures form in the polypeptide chain which guide the formation of tertiary structure.



What is local structure?

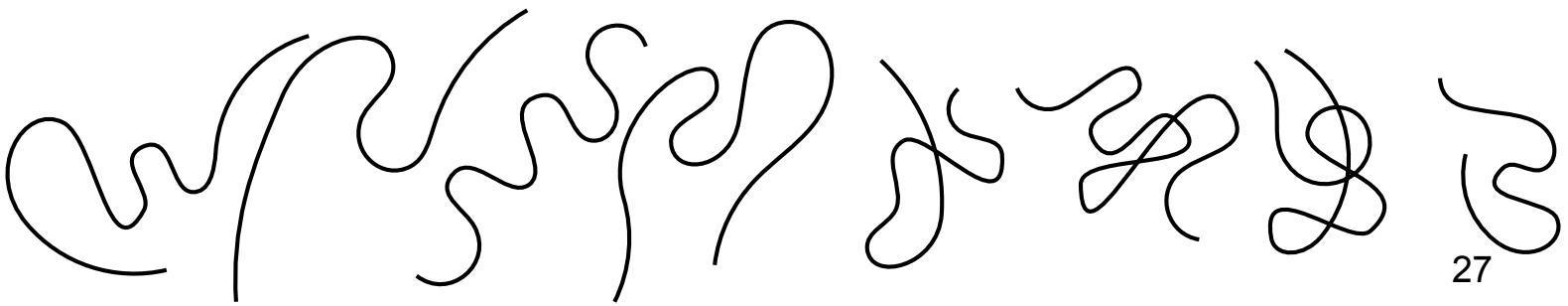
Early in the process of folding (nsec timescale)

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



Local structure formation

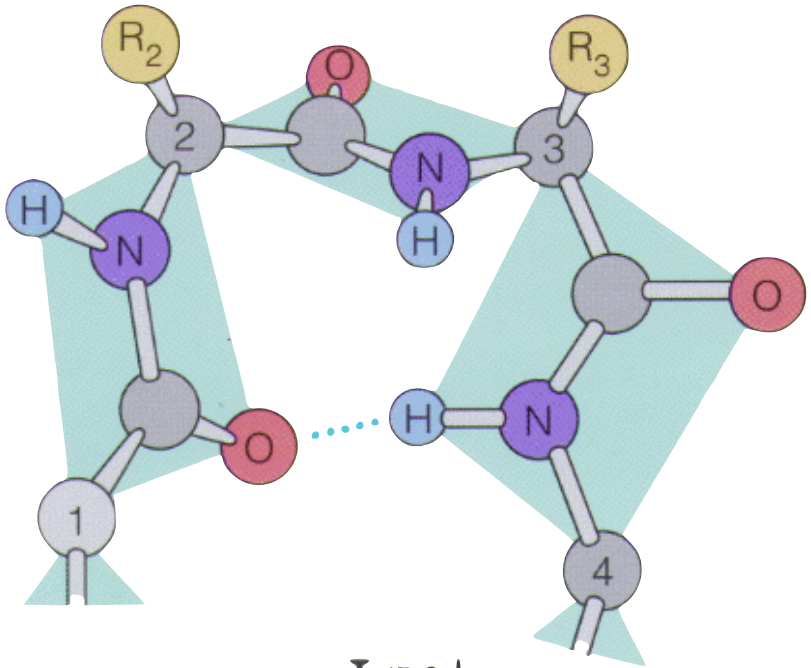
- Short pieces of protein sample conformational space randomly, driven by the hydrophobic effect (mostly).
- Glycines provide points of greater flexibility.



beta turns

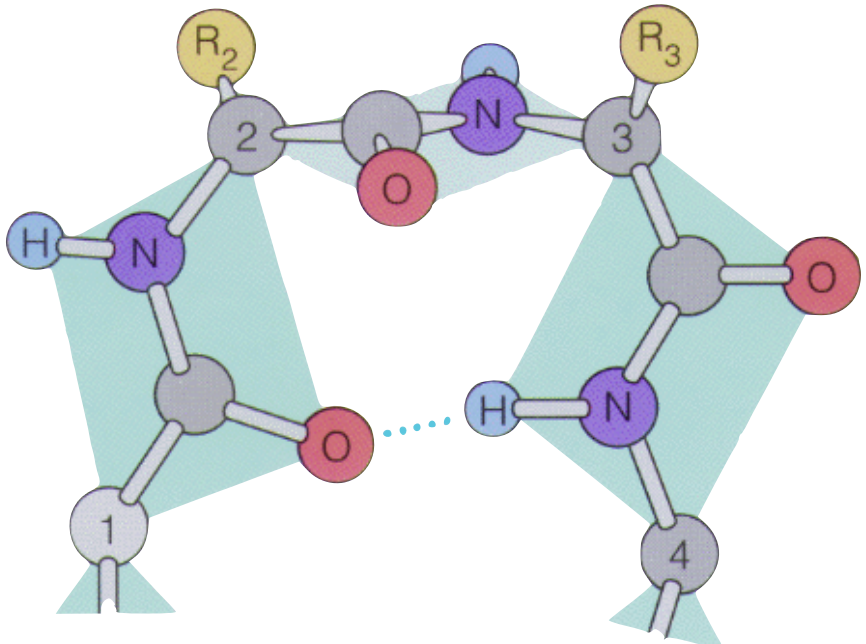
4-residues

Residue 1 hydrogen bonds to residue 4



Type I

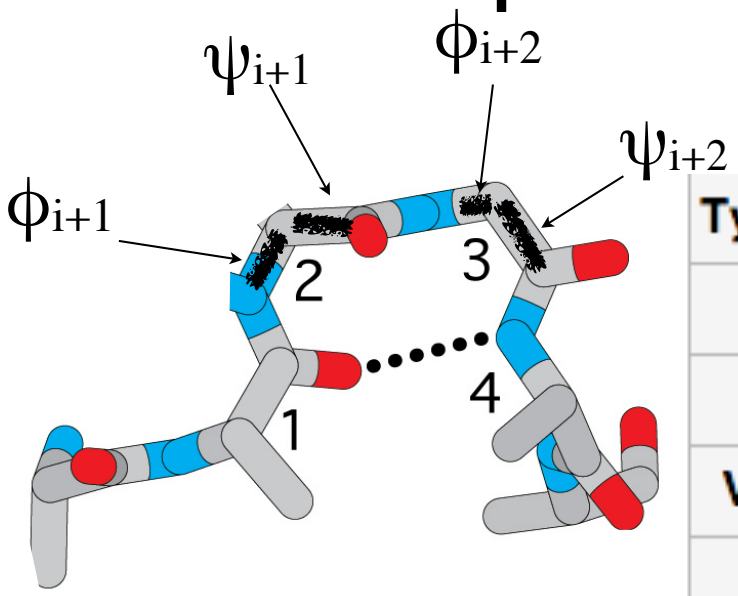
Type I (most common). Oxygen points away, viewed clockwise.



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**		

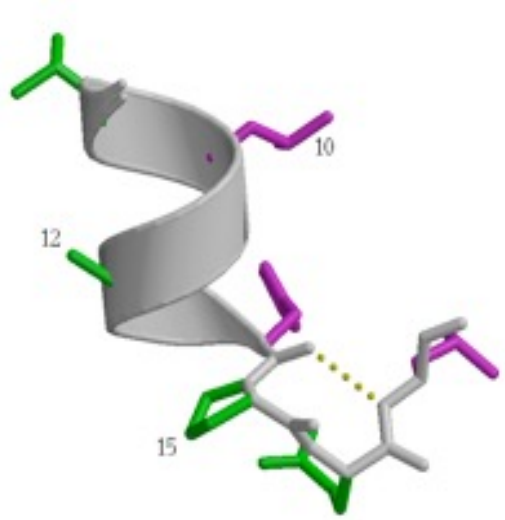
Backbone angles ±30°

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			

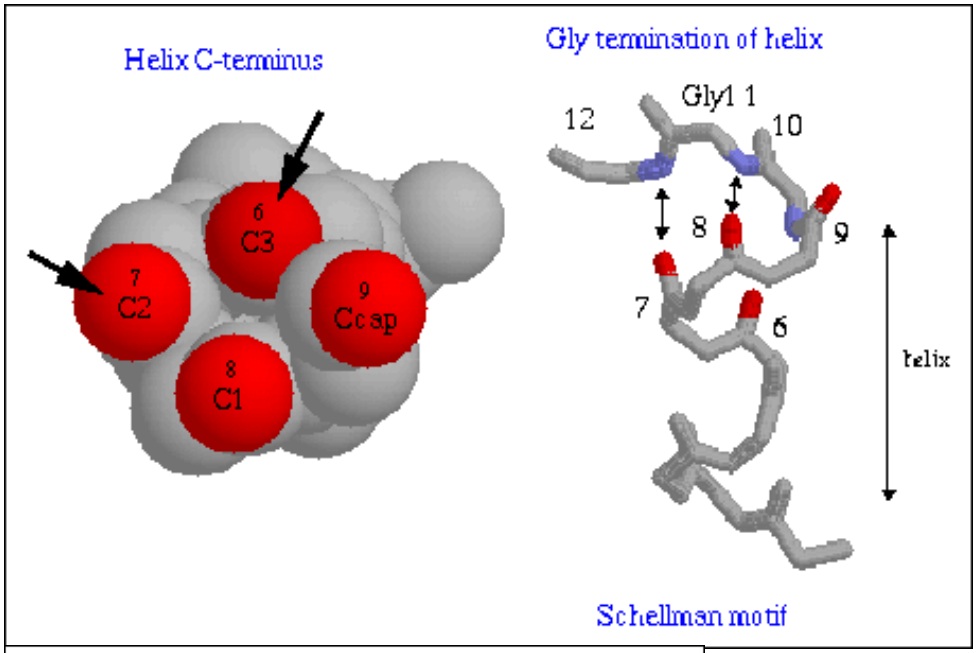
*have cis-peptide bond at *i*+2

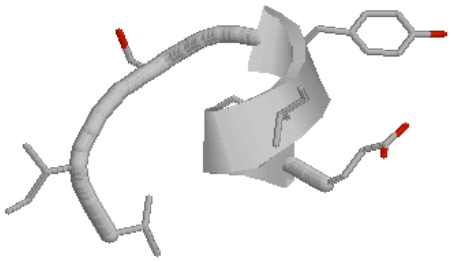
http://www.ebi.ac.uk

Other local structures: Helix caps

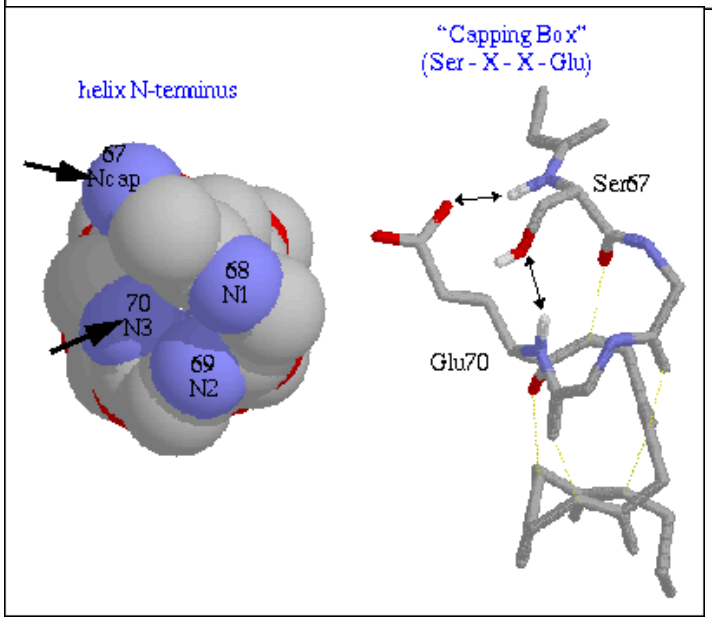


Proline helix C-cap



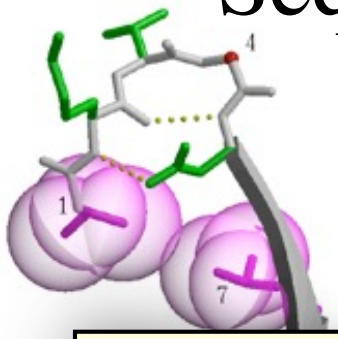


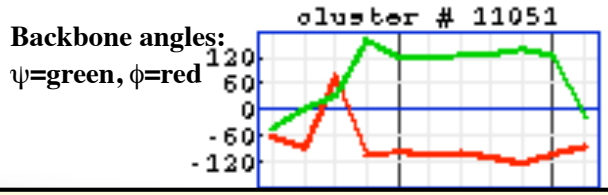
glycine helix N-cap

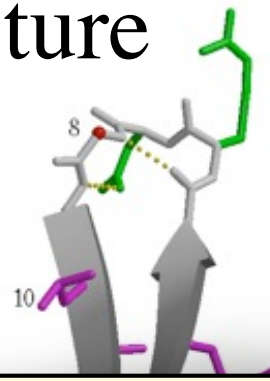


http://www.cryst.bbk.ac.uk

Sequence motifs for local structure







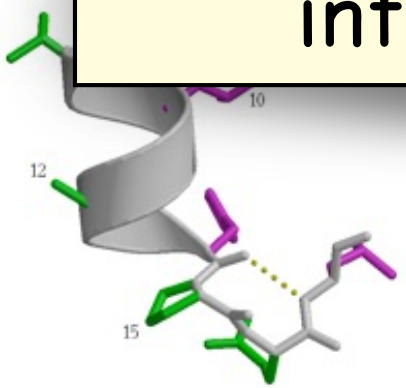
Type-I hairpin

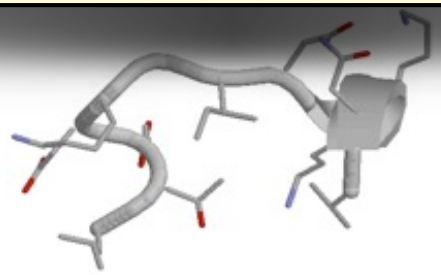
dive
turn

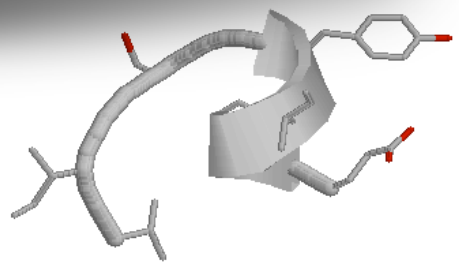
Structures from non-homologous proteins (not same family) were data-mined for correlated sequence/ structure patterns. Strongest correlations were called "folding intiation site" (I-sites) motifs.

Serine
hairpi

Frayed helix





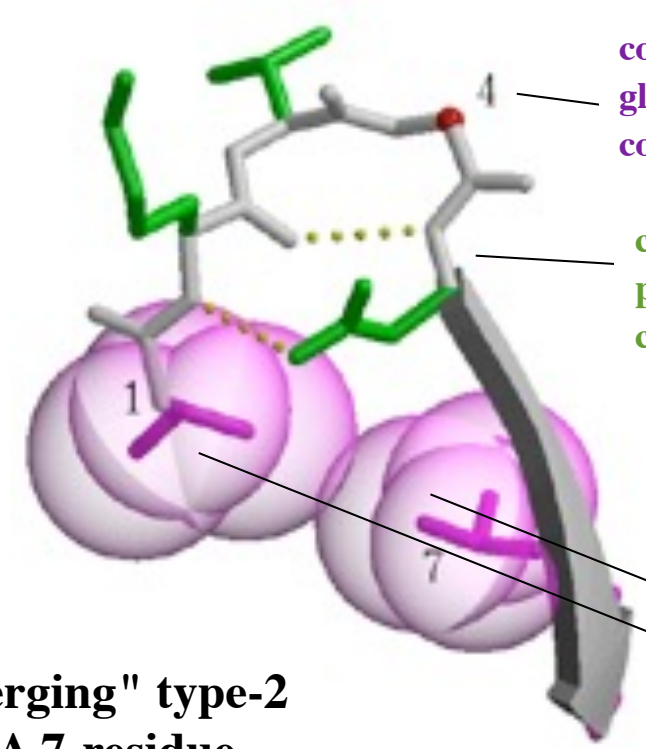


Proline helix C-cap

alpha-alpha corner

glycine helix N-cap

Sequence determines structure!

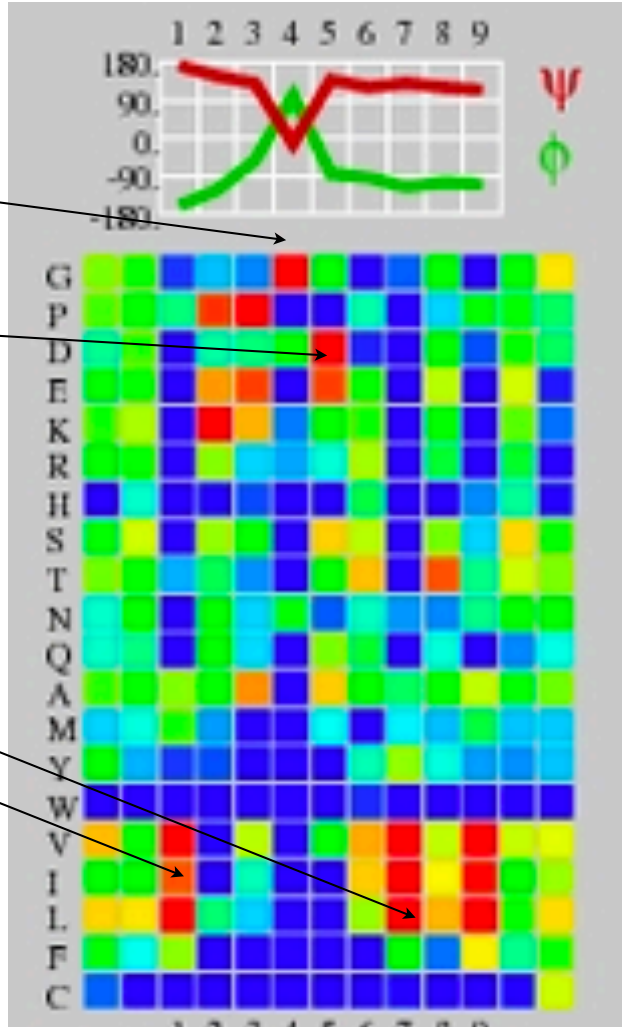


"Diverging" type-2 turn. A 7-residue peptide forms this structure! [2]

conserved glycine in α_L conformation

conserved polar side chain

conserved non-polar side chains



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.

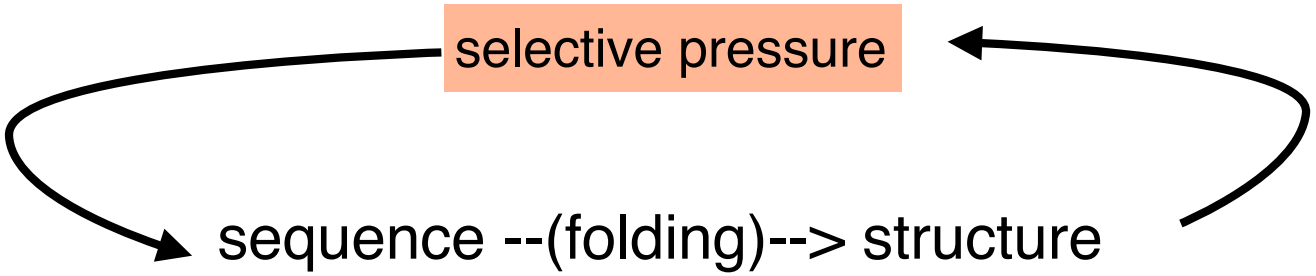
[2] Yi Q, Bystroff C, Rajagopal P, Klevit RE & Baker D. (1998). Prediction and structural characterization of an independently folding substructure in t he src SH3 domain. J Mol Biol283, 293-300.

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.

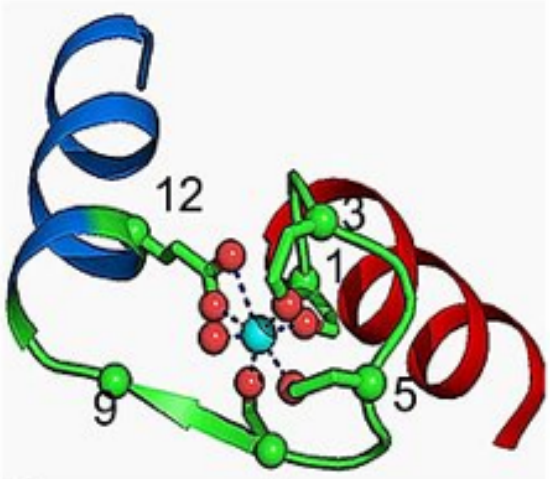
Conserved sequence patterns inform us of the structure.



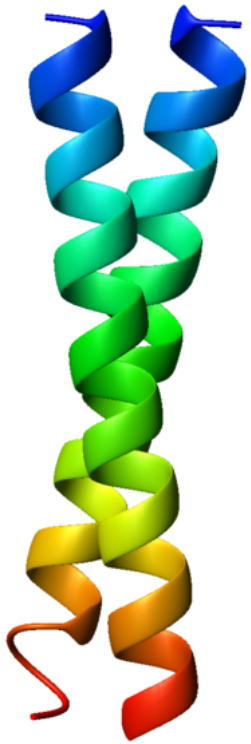
Super-Secondary Structure (SSS)

α

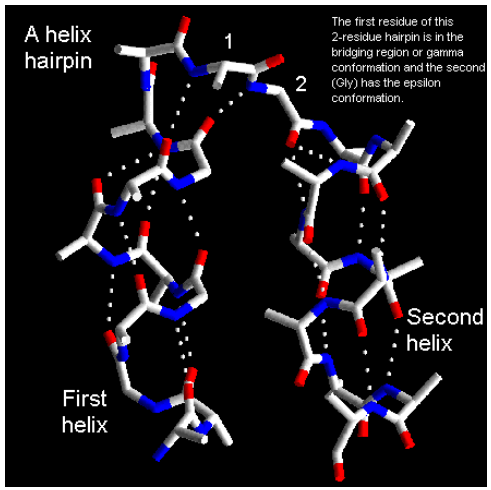
- SSS contains more than one SSE, interacting.
- beta turns and helix caps are usually involved.
- Canonical SSS have names.



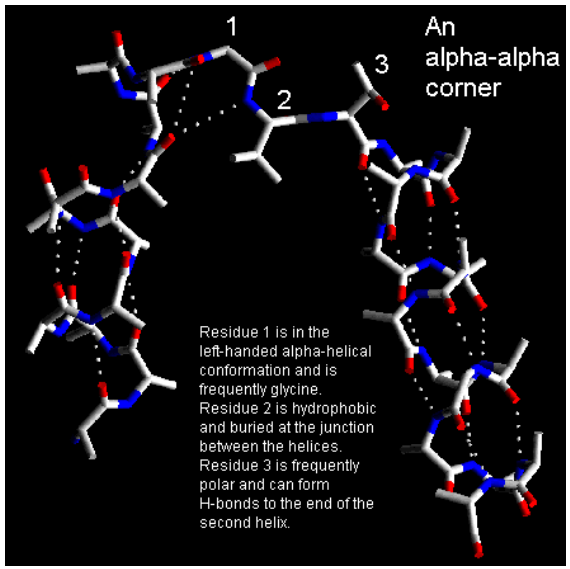
EF hand



Coiled-coil

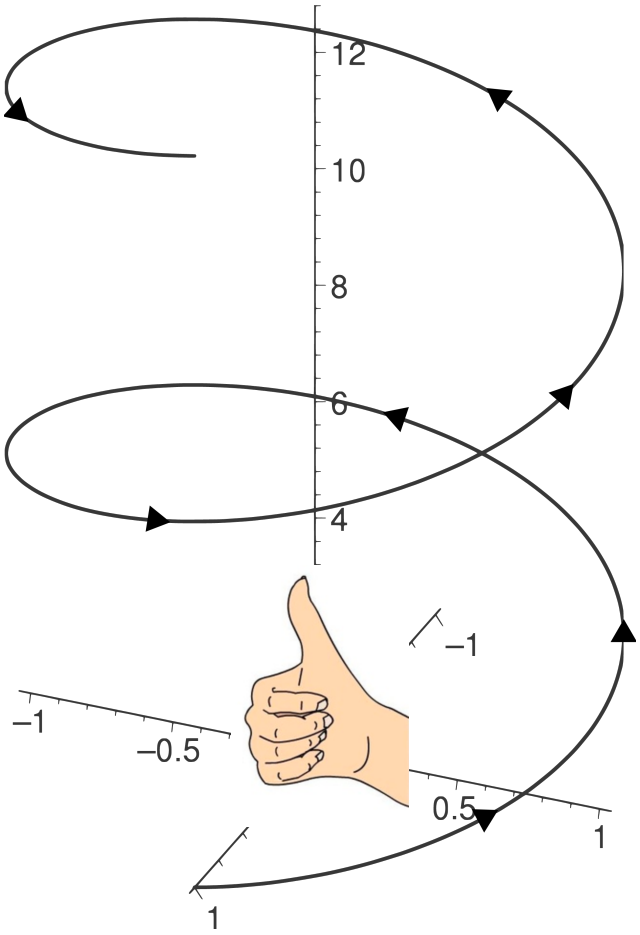


Helix hairpin



alpha-alpha corner

Handedness



<u>Right-handed helix.</u>
Put the thumb of the right hand along the axis of rotation.

As you travel up the helix (going in the direction of your right thumb) the line curve in the direction of your fingers.

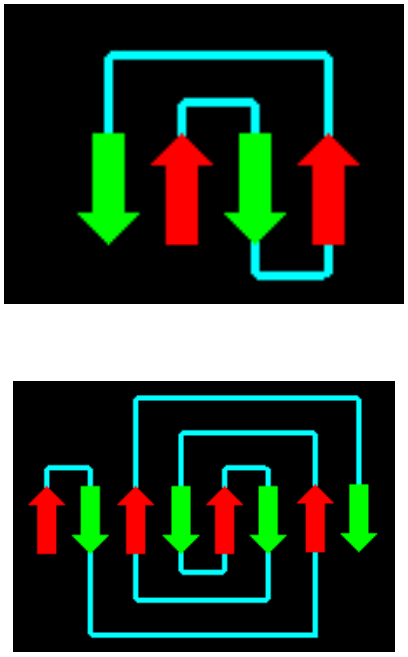
Yes, that means you are turning left when you walk up a right-handed spiral staircase, and right when you are walking up a left-handed spiral staircase.

Super-secondary structure.

β

hairpin

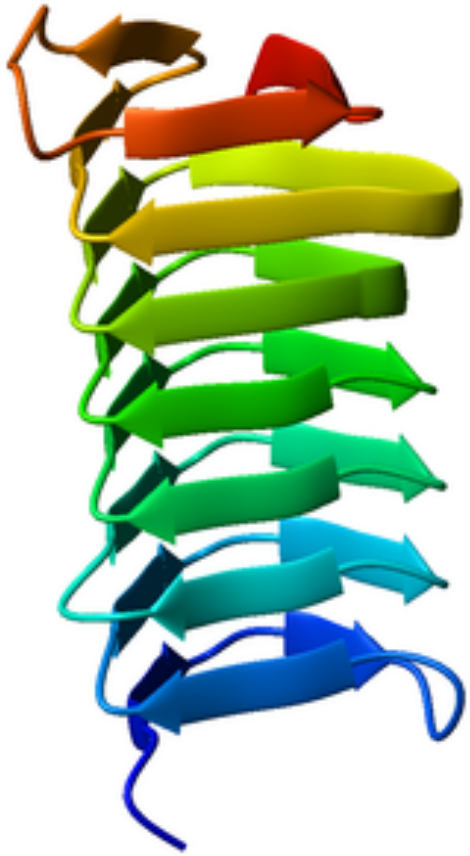
meander

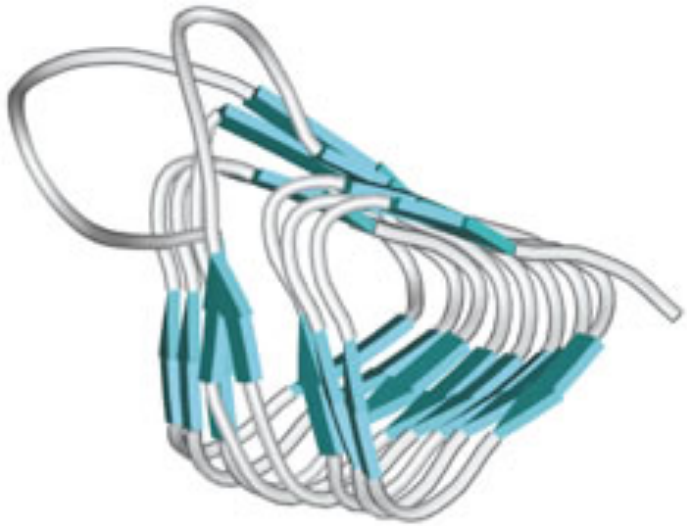


"greek key"

wikipedia.org/wiki/Beta_sheet

Super-secondary structure.

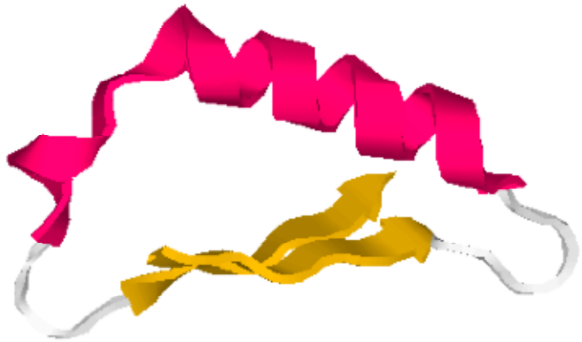


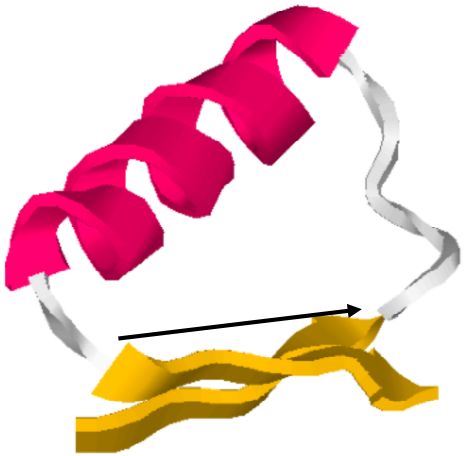


β helix

Super-secondary structure. αβ

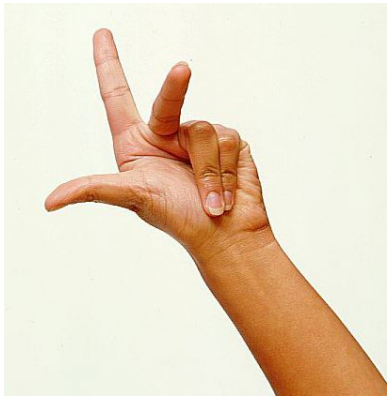
βαβ supersecondary structure units are mostly right-handed





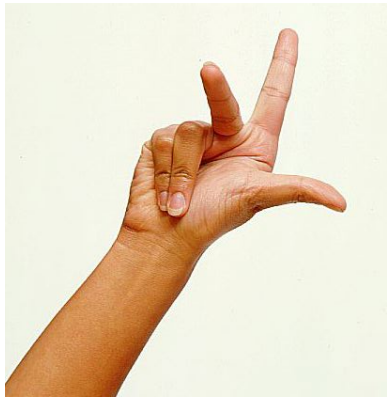
L-handed βαβ

1.5%

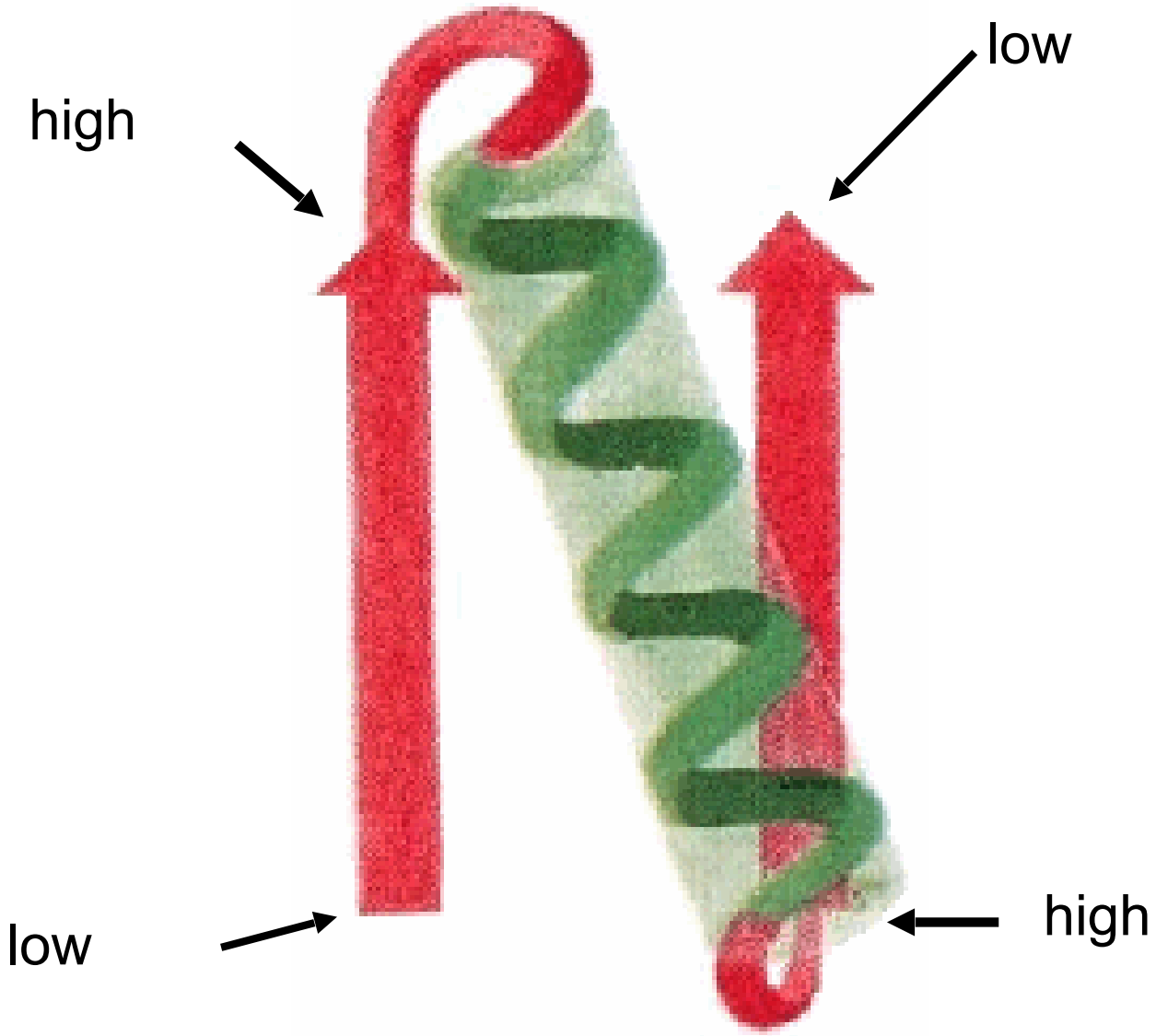


R-handed βαβ

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.

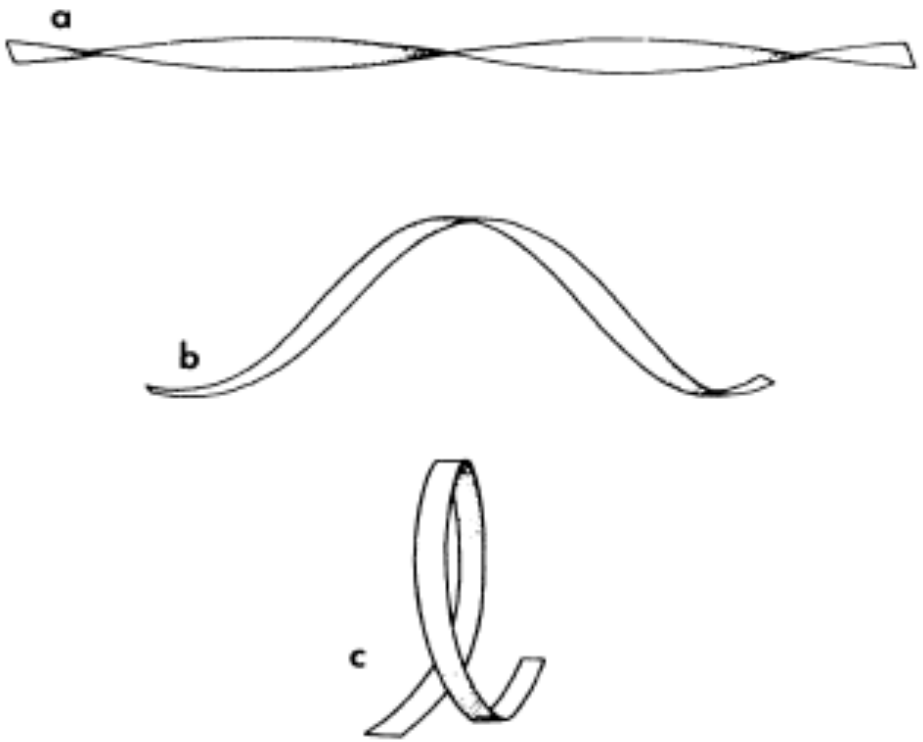


FIG. 4. A possible folding pathway which produces righthanded crossover loops from extended chain. In (a) the section of chain is extended, showing one full turn of the preferred righthanded twist for β strands. In (b) the two ends of this chain segment are moving toward one another, and the ribbon has 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.

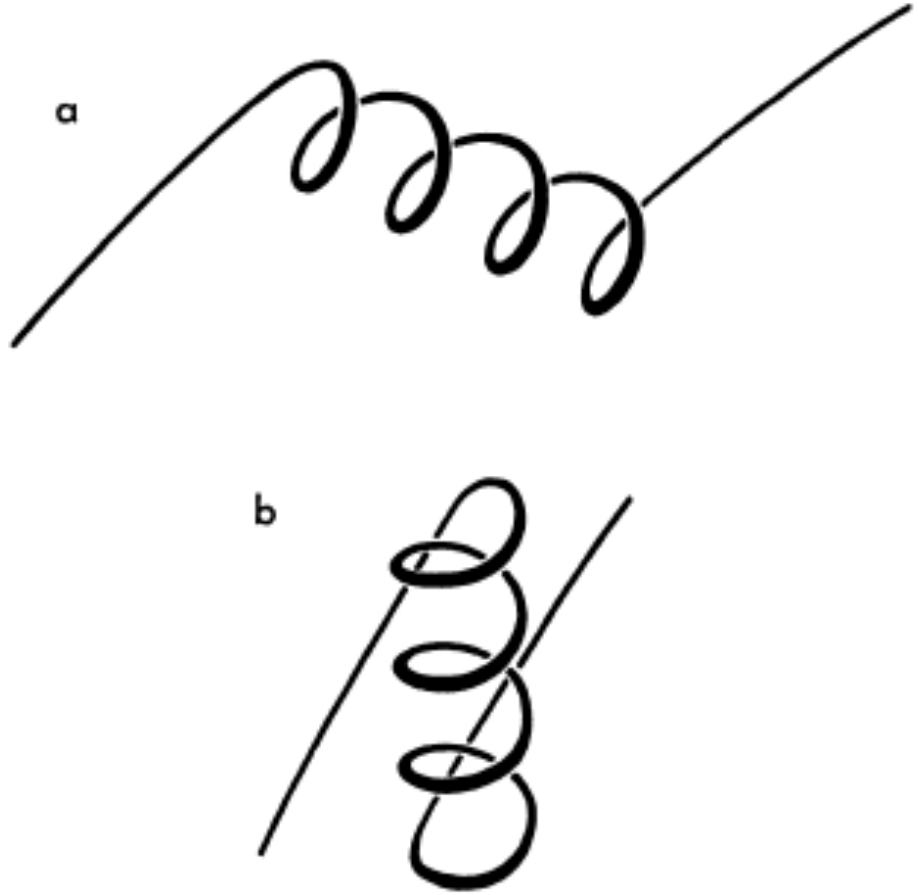
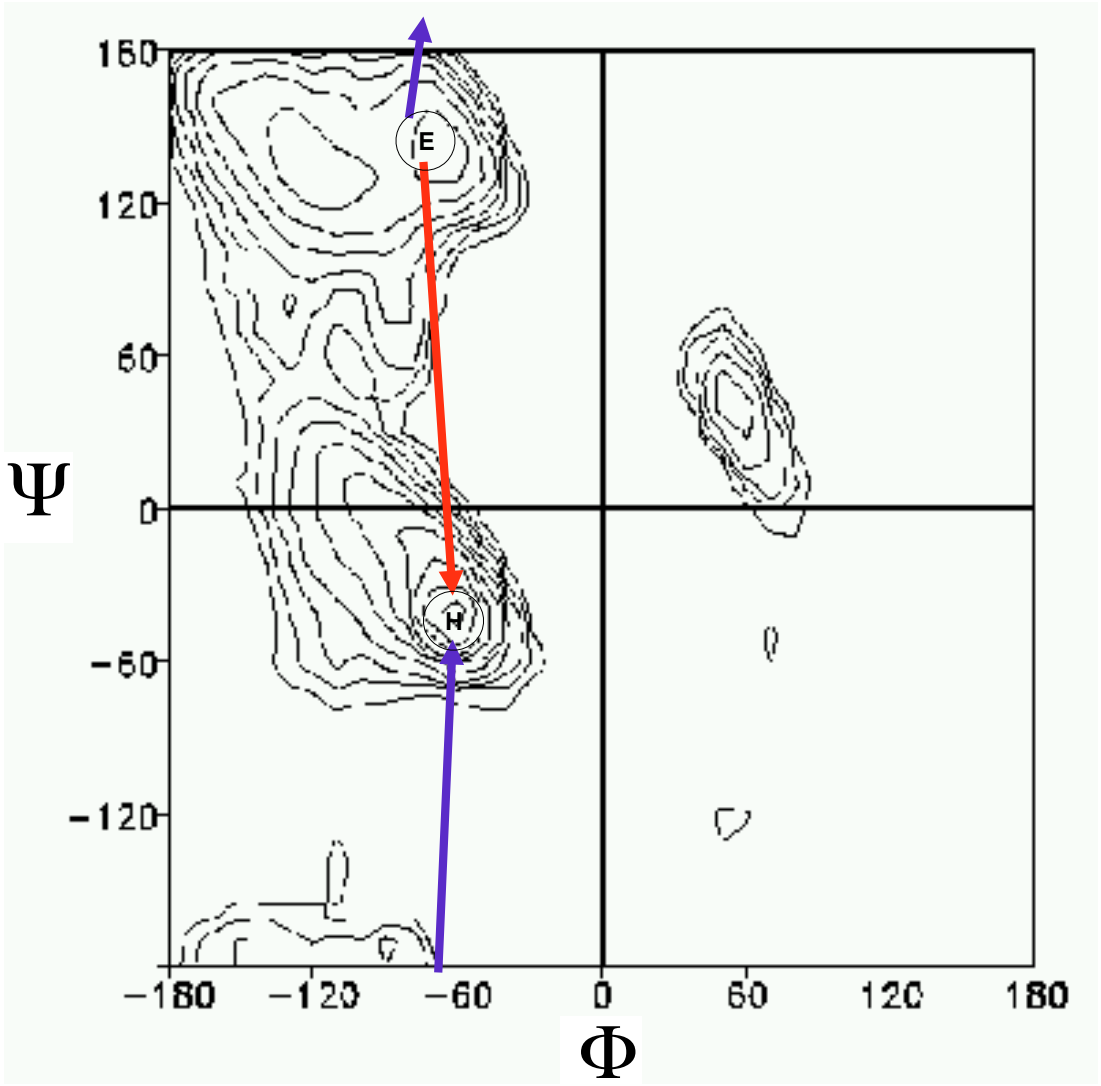


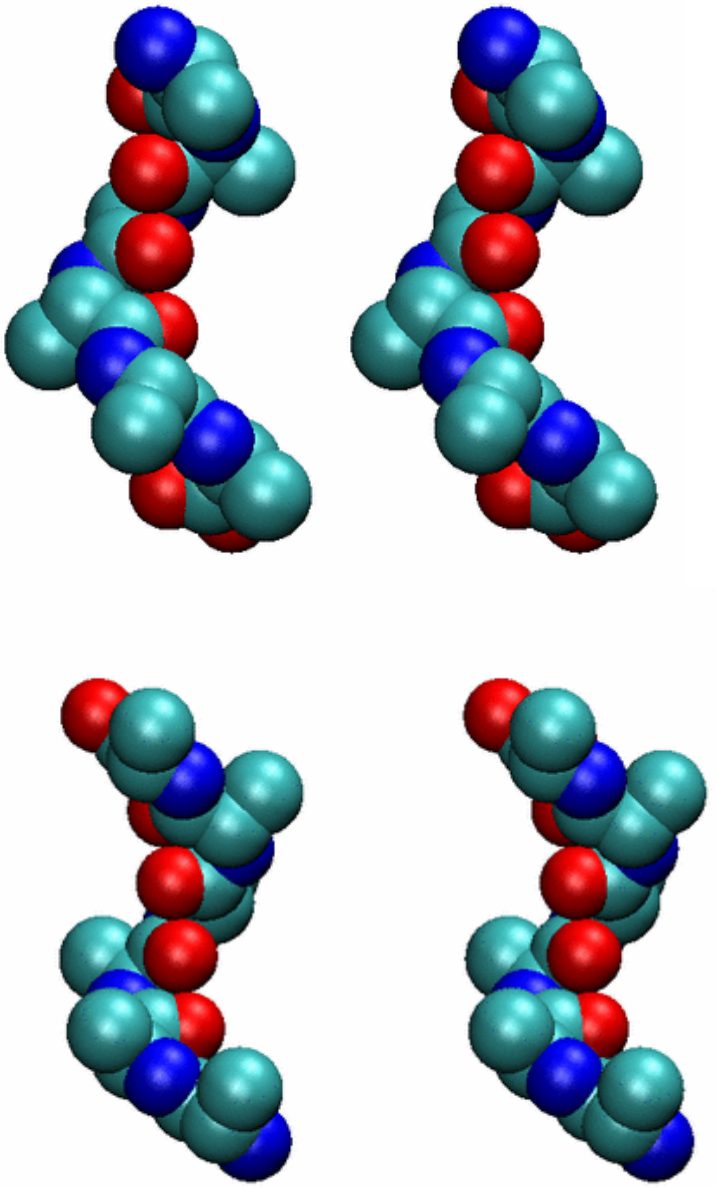
FIG. 5. A possible folding pathway which forms righthanded crossover loops from a righthanded α-helix with a β strand at each end of it.

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

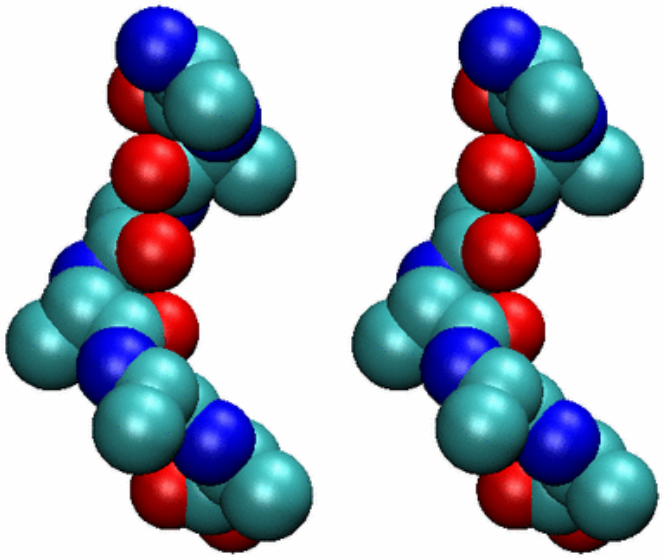
Theories for why βαβ units are right-handed.

Phone Cord Effect: Northern versus Southern route to helix

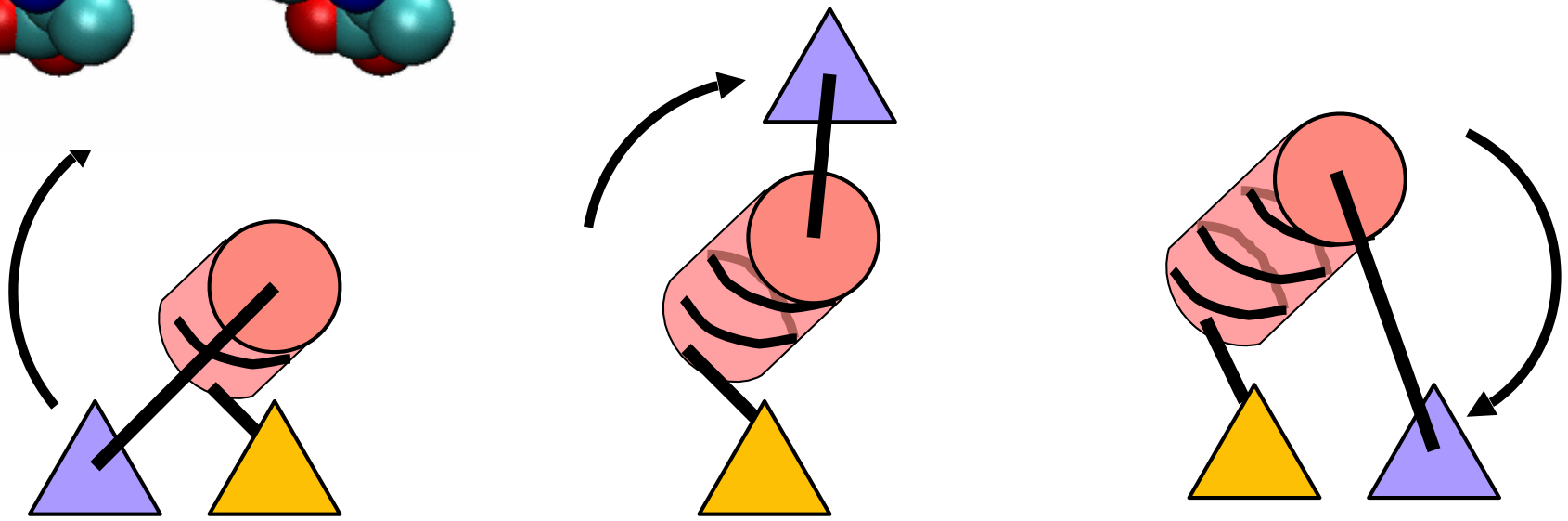




Cole B & Bystroff C. (2009) Alpha helical crossovers favor right-handed supersecondary structures by a kinetic trapping mechanism. The phone cord effect in protein folding. ***Protein Science*** **18(8) 1602 - 1608**



Theories for why βαβ units are right-handed.

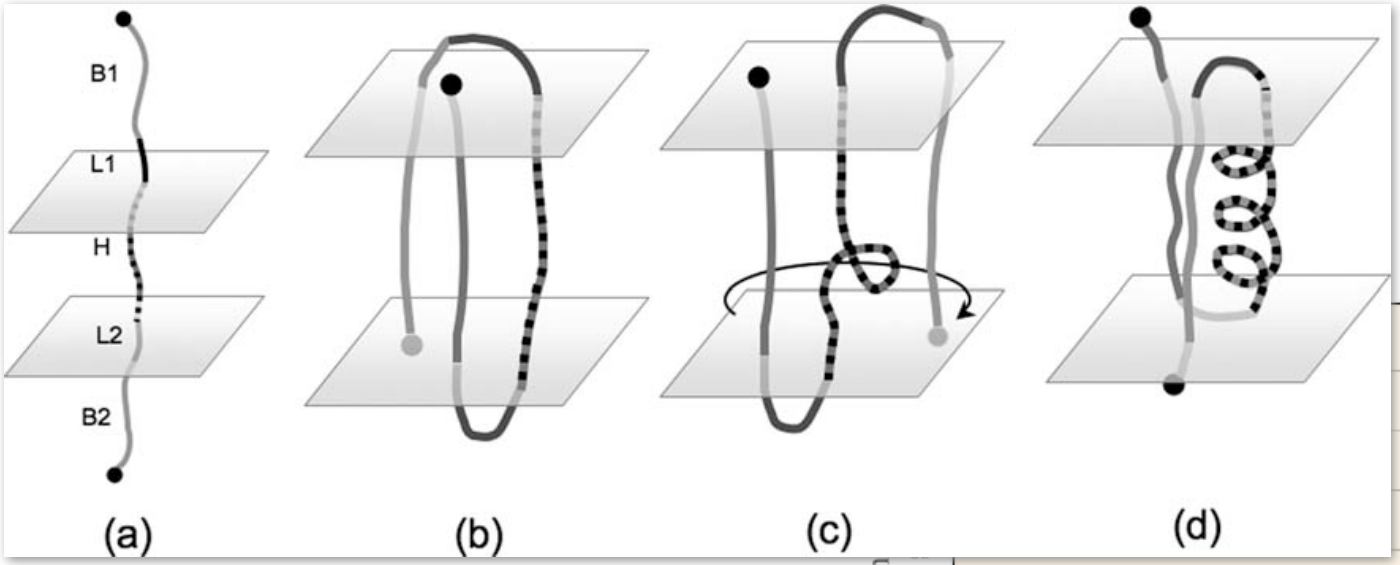


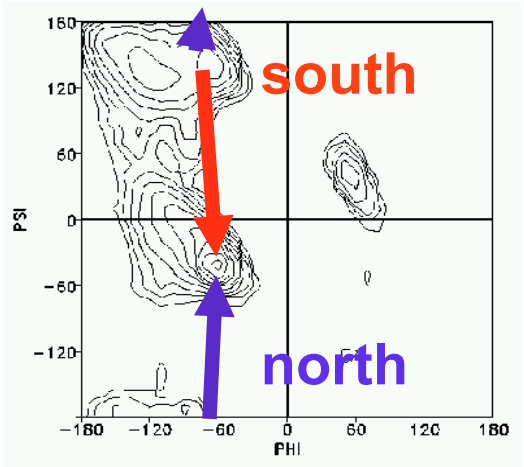
LH ——————————————————— RH

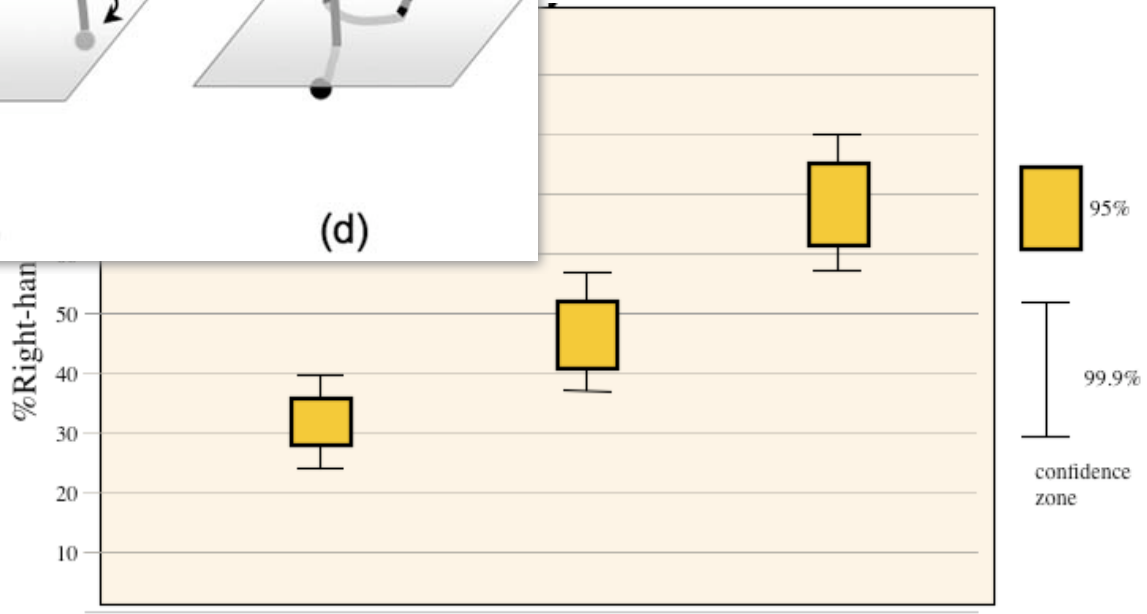
left-handed torque turns left-handed βαβ to right-handed βαβ

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

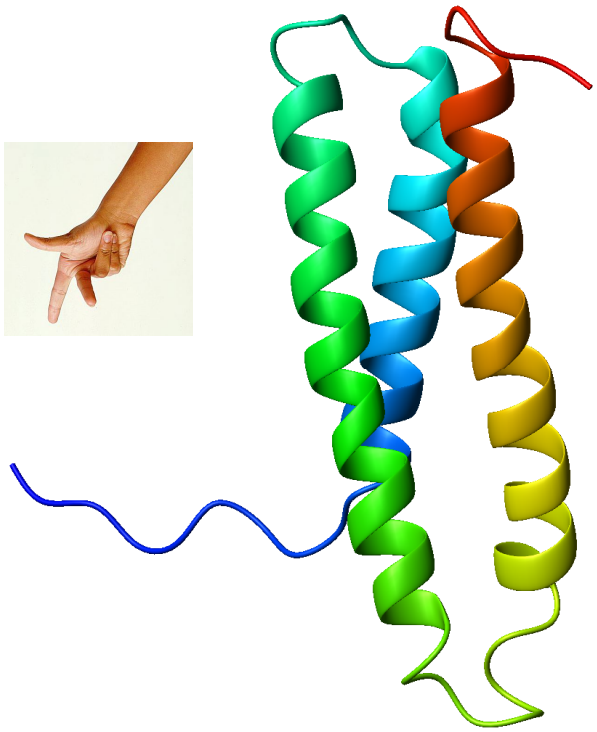
http://www.youtube.com/watch?feature=player_embedded&v=6hQYjtmU6E0

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							
		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

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
		0.00					
		0.86	9	9	18	22	38
		1.00	20	22	38	50	63
		1.00	11	68	77	87	93
		0.00	8	135	150	160	178
		0.00	3	67	78	84	97
		0.11	17	43	57	60	72
		1.00	10	60	72	77	91
		1.00	2	102	116	123	132
		1.00	6	17	23	25	45
		0.00	14	381	402	412	428
		0.00	9	412	428	438	450
		1.00	5	438	450	458	475
		1.00	2	458	475	477	485
		1.00	5	263	297	304	321
		1.00	8	364	377	381	390
		1.00	2	289	301	305	322
		1.00	4	11	44	49	71
		0.00	1	92	128	138	161
		1.00	5	6	12	16	20
		1.00	5	23	41	55	70
		0.79	5	16	20	23	41
		1.00	1	105	120	132	137
		0.95	3	28	36	46	60
		1.00	2	242	252	260	281
		0.00	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



61% R,

39% L,

n=431,

p(R=0.50)<0.01

How to force hydrogen bonds using restraints

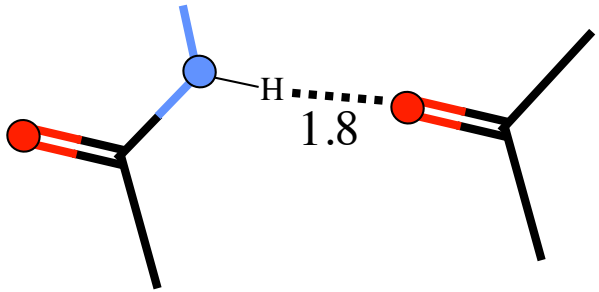
- To add a restraint
  Edit | Potential | Restrain, distance,
  Target 1.8, 1.8, Weight 50
  Pick amide H and carbonyl O.
  Click **Create**.
  Cancel | Restrain (or esc) when done
- Energy minimize

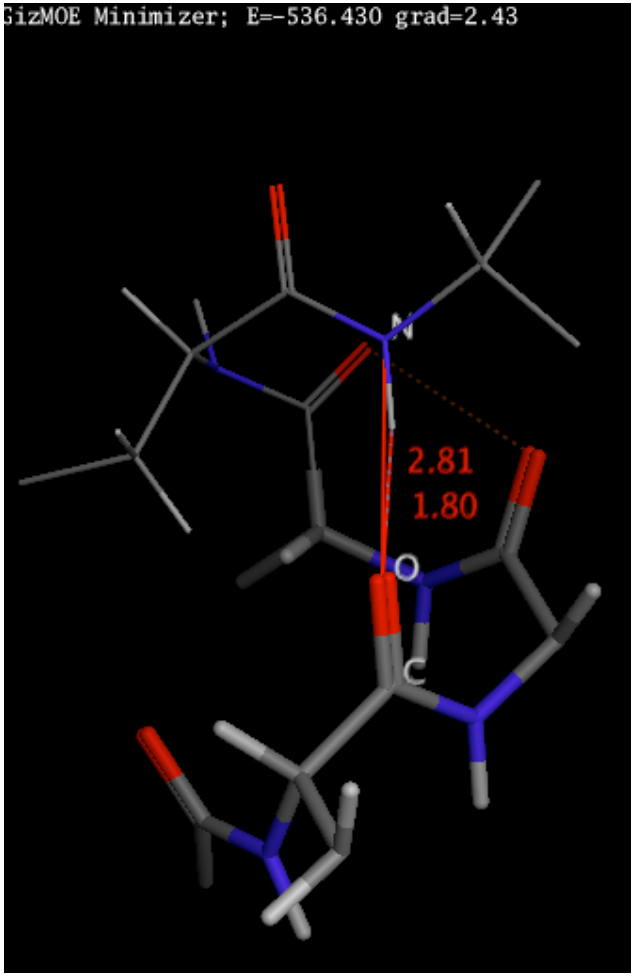
  Compute | prepare | Structure preparation

  Checks for missing atoms, assigns energies.

  SVL: run ‘gizmin.svl’

  When finished, be sure to **Cancel | GizMOE_Minimizer**
- To remove or modify restraints
  Potential setup (button at far lower left)
  Restraints tab





Exercise 18.2

Make a beta hairpin

anti-parallel sheet with valine side chains all on the same side of the sheet.

Edit | Build | Protein, Geometry: **anti-strand**. Residue: **ADVDVKVSPNGVEVKVRA**

Zoom out.

Select the second half of the chain starting with NG.

Rotate and translate it (**shift-alt-middlemouse**) so that the first three valines (3,5,7) are lined up with other there valines (12,14,16), and the valine backbone H-bonding groups (NH and CO) are close to the H-bonding distance (1.8Å from H to O)

Hide side chains to help see the backbone atoms better.

Edit | Potential | Restrain.
Set Target 1.8, 1.8, Weight 50. Select H and O atoms. **Create**.
When done you have 2 restraints for each of the three paired valines for a total of 6 restraints**.**

Compute | Prepare | Structure preparation. Hit **Correct** if necessary. **Protonate3D**.

SVL: run 'gizmin.svl'.

If there are errors in the restraints, **Cancel/GizMOE,** open **Potential Setup** (extreme lower left of the MOE window). **Restraints**. Click on restraints to delete or modify them.
Restart **SVL: run 'gizmin.svl'.**

Look at out the structure.
It should have beta pleating when viewed from the edge of the sheet. Sidechains should alternate up and down in that view.
Residues SPNG form a beta-turn.

Cancel/Gizmin . Remove the restraints. Restart **SVL: run 'gizmin.svl'**.

Does the structure hold together or fall apart?