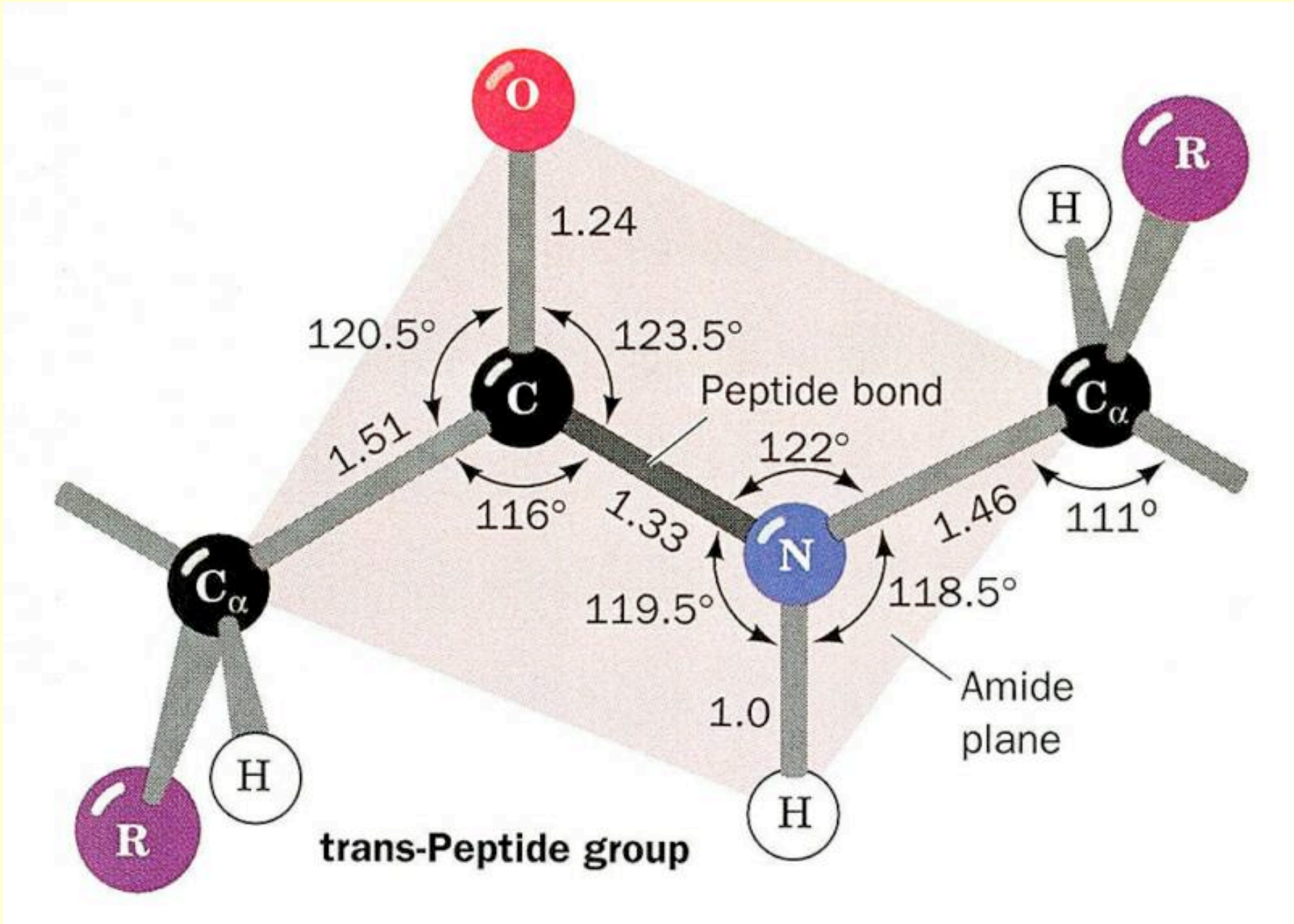
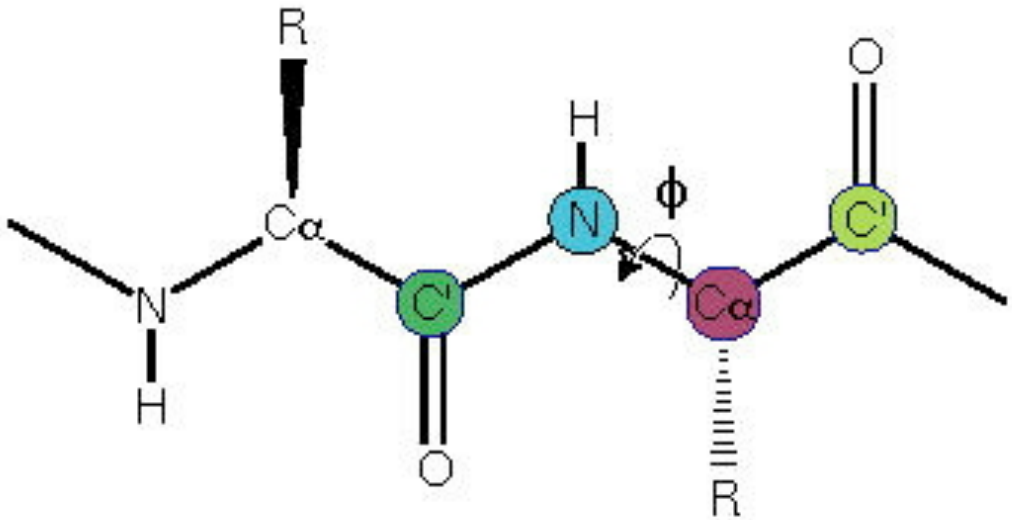
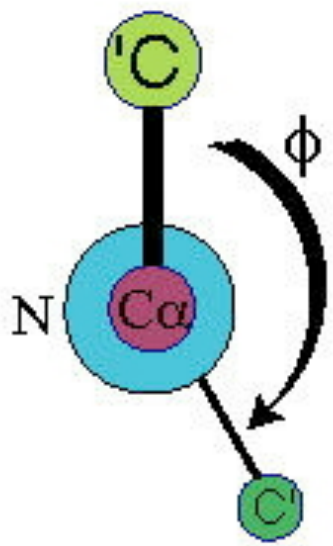
Lecture 5
Secondary Structure Assignment and Prediction

- Peptide bond geometry
- Secondary structure
- Propensities
- Secondary structure assignment
- Secondary structure prediction


O
1.24
120.5°
123.5°
C
Peptide bond
1.51
122°
116°
1.33
1.46
111°
Cα
N
Cα
119.5°
118.5°
1.0
Amide
plane
H
R
H
R
H
trans-Peptide group

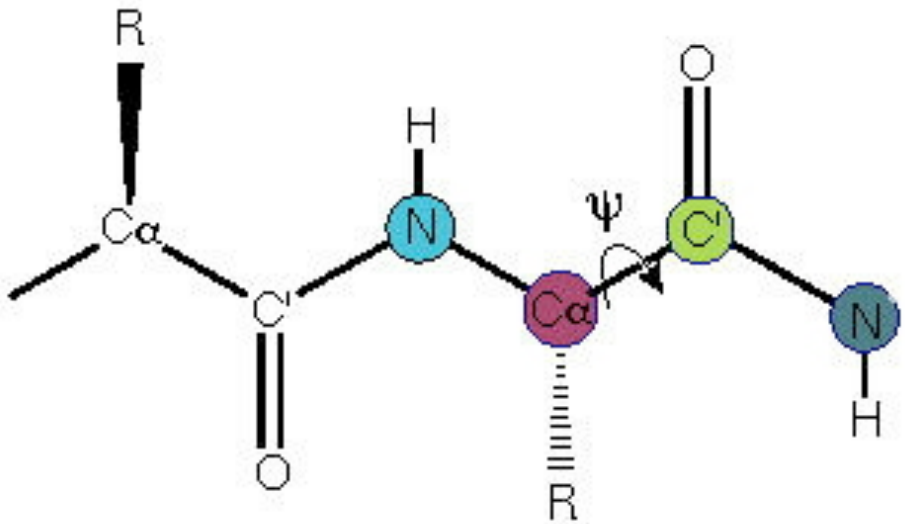
Torsion Angle Phi (ϕ)

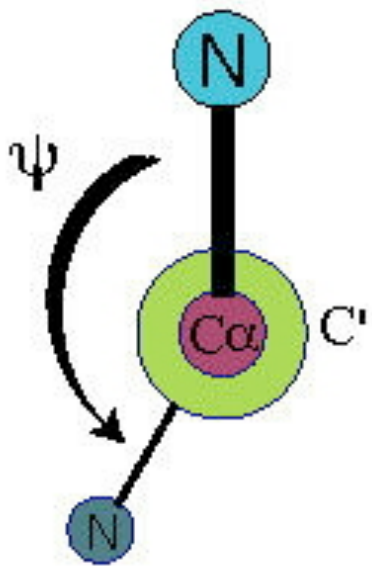




This torsion angle ϕ is right handed and is therefore positive (+130°)

Torsion Angle Psi (ψ)

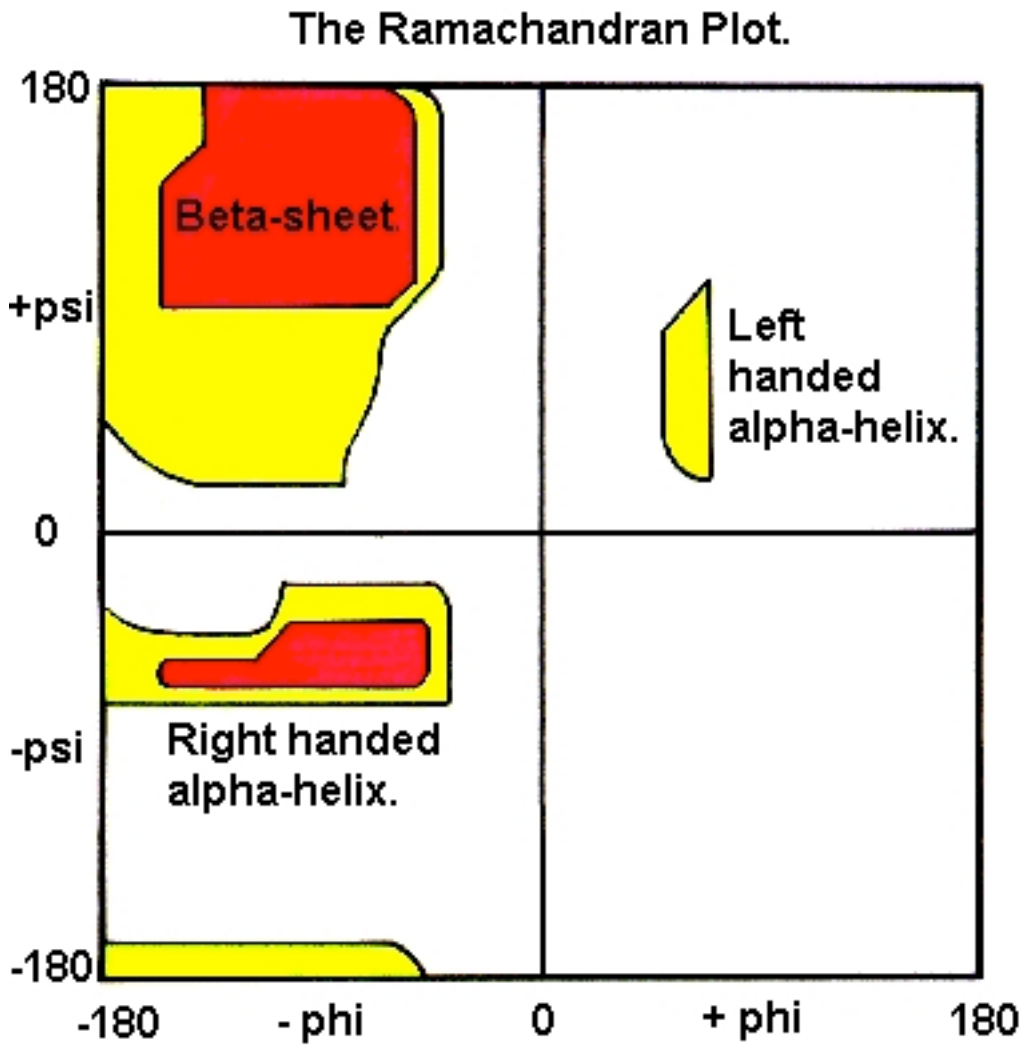




Loren Williams
Georgia Tech

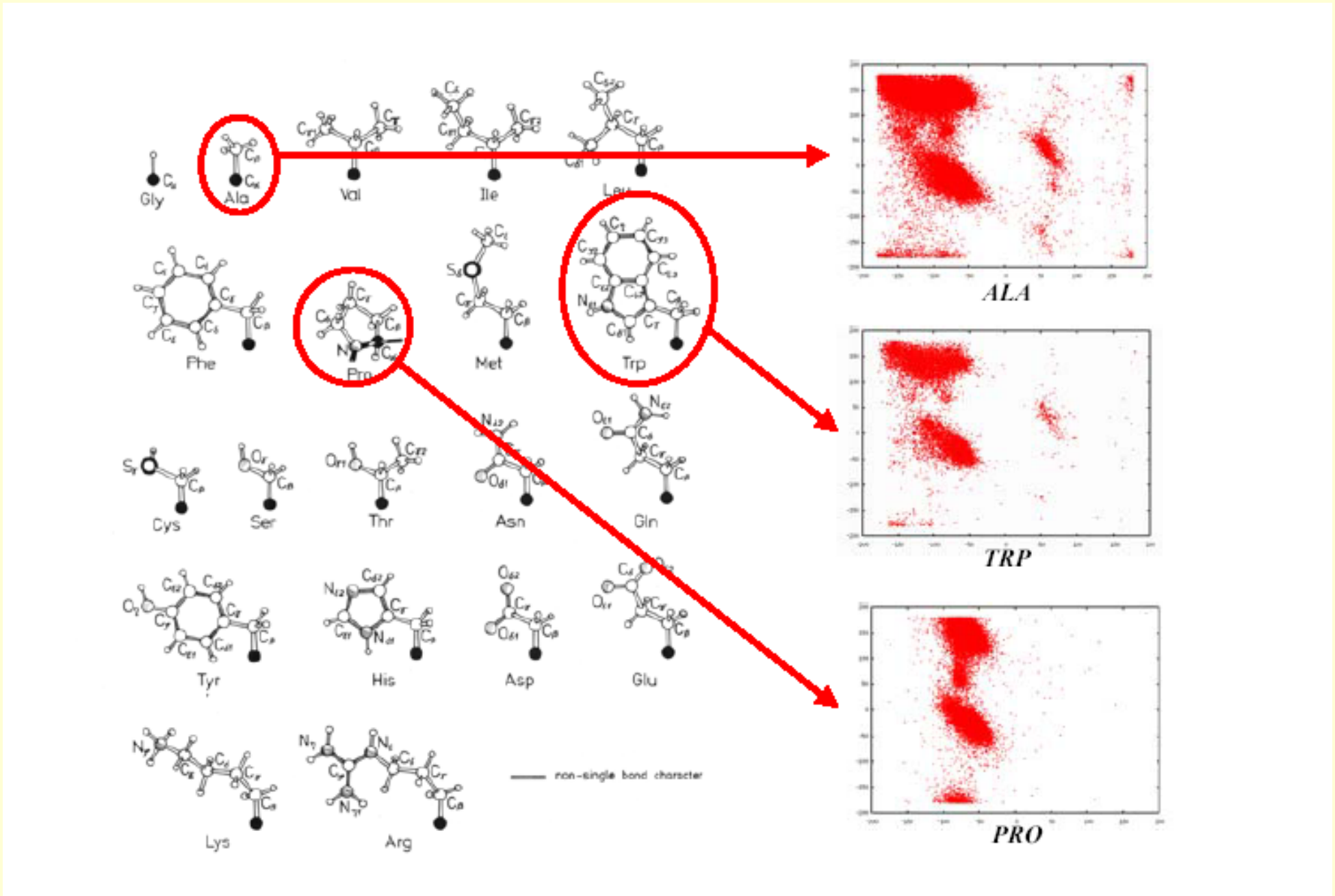
This torsion angle ψ is left-handed and
is therefore negative (-130°)

Ramachandran Plot

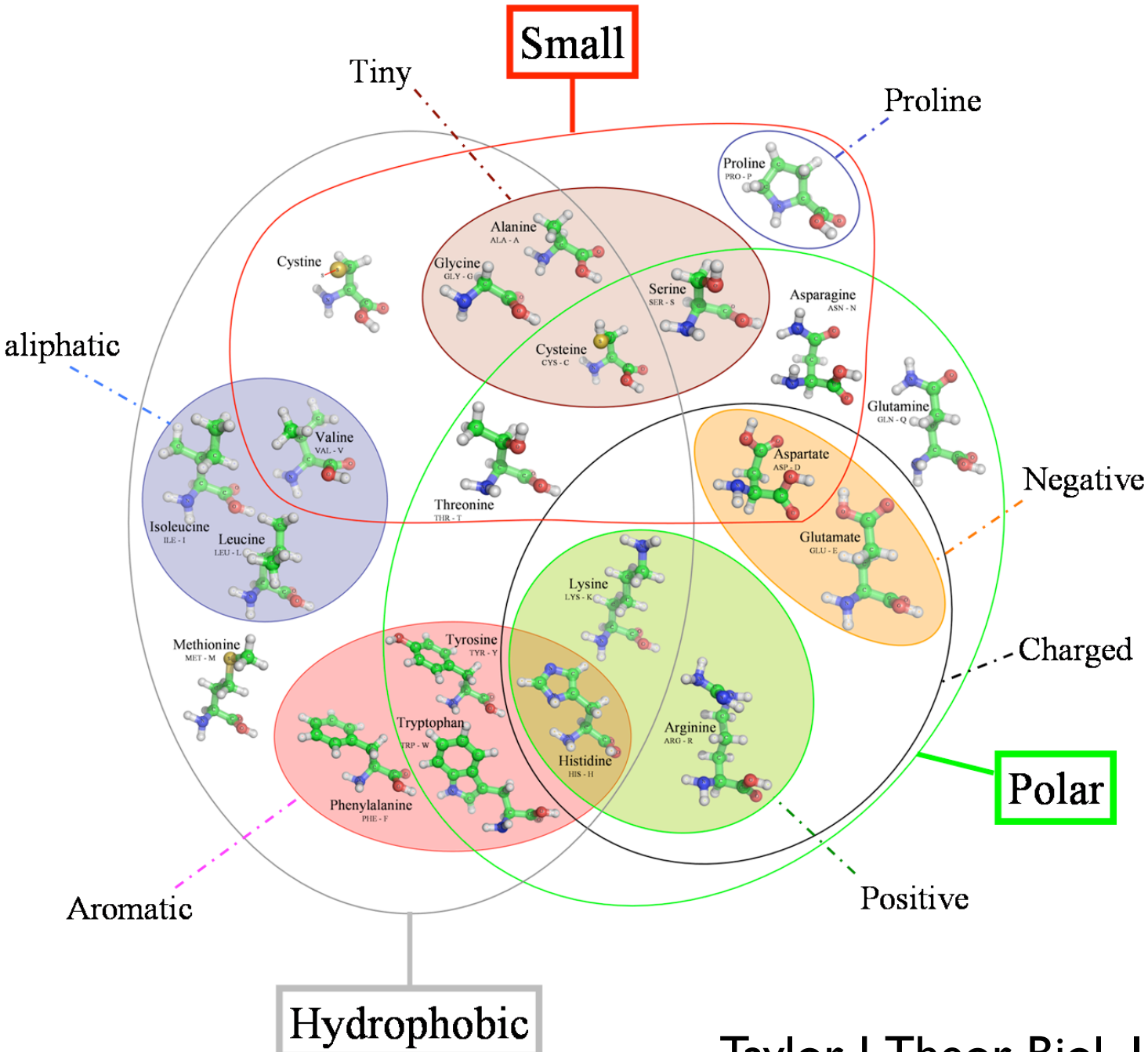


Projection of a torus onto a plane.
Imagine the plot bent such that opposing edges meet.

Conformational propensities



Venn Diagram of Amino Acids



Taylor, J. Theor. Biol. 119, 205-218 (1986).

Secondary Structure Propensities

(Koehl-Levitt, 1999)

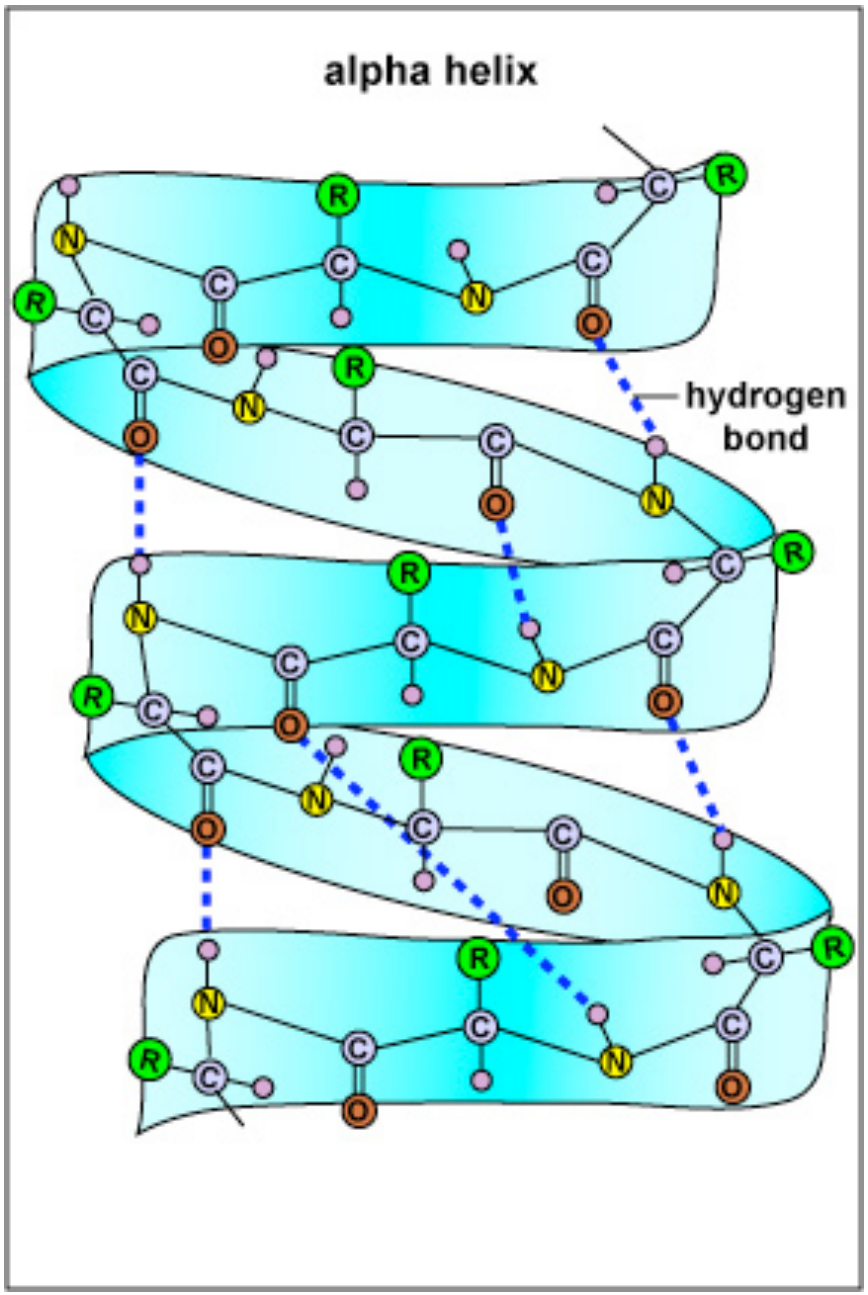
Alpha-helix propensity derived from designed sequences

A/L	R/K	N/M	D/F	C/P	Q/S	E/T	G/W	H/Y	I/V
-0.04	-0.30	0.25	0.27	0.57	-0.02	-0.33	1.24	-0.11	-0.26
-0.38	-0.18	-0.09	-0.01	0.	0.15	0.39	0.21	0.05	-0.06

Beta-sheet propensity derived from designed sequences

A/L	R/K	N/M	D/F	C/P	Q/S	E/T	G/W	H/Y	I/V
-0.12	0.34	1.05	1.12	-0.63	1.67	0.91	0.76	1.34	-0.77
0.15	0.29	-0.71	-0.67	0.	1.45	-0.70	-0.14	-0.49	-0.70

http://www.genome.jp/aaindex/



The amino acids in an α-helix are arranged in a right-handed helical structure, 5.4 Å wide. Each amino acid corresponds to a 100° turn in the helix (i.e., the helix has 3.6 residues per turn). The N-H group of an amino acid forms a hydrogen bond with the C = O group of the amino acid four residues earlier; this repeated hydrogen bonding defines an α-helix.

The α-Helix

Linus Pauling: prediction of α-helix with paper and pencil.

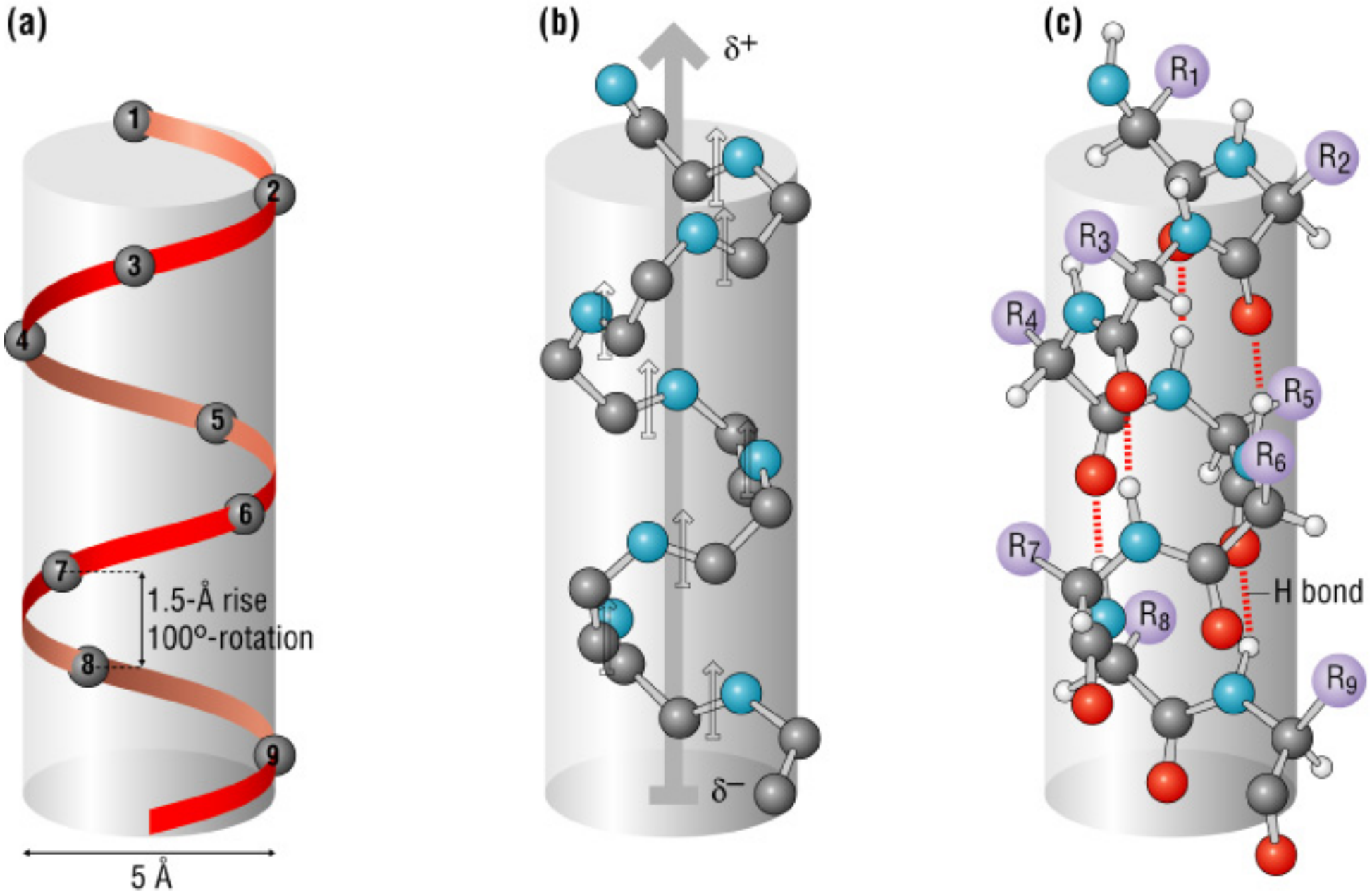
"I was so pleased with the α-helix that I felt sure that it was an acceptable way of folding polypeptide chains and that it would show up in the structure of some proteins when it finally became possible to determine them experimentally."

The α-helix was a breakthrough discovery in the structure of proteins. Pauling had happened upon the structure while lying in bed with a bad cold In 1948, but delayed publishing until 1951 in order to prove to himself and others that the model was correct.

Pauling's spiral, or helical structure, was one of many structures he proposed for polypeptide chains in a flurry of papers in 1951.

http://www.paulingexhibit.org/exhibit/alpha-helix.html

From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe



Helix Types

helix type	Φ / °	Ψ / °	syst. name
alpha (right)	-57	-47	3.6(13)
alpha (left)	57	47	3.6(13)
3.10	-60	-30	3.0(10)
π	-50	-70	4.2(16)
Collagen	-51/-76/-45	153/127/148	$(Gly\text{-}Pro\text{-}Hyp)_x$
Polyproline	-78	149	$(Pro)_x$

Helices

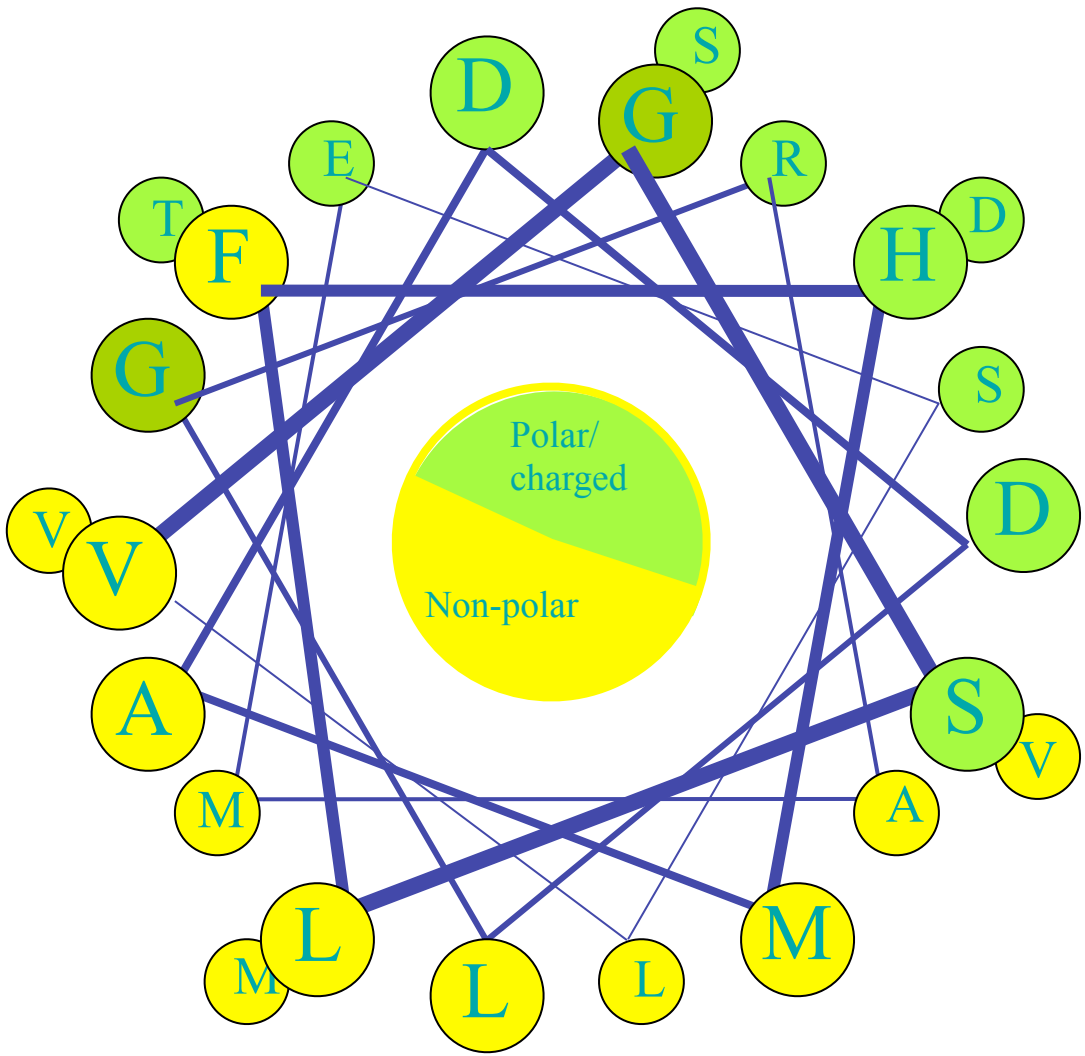
- Backbone wound around (hypothetical) helix axis
- Hydrogen bonds are formed between residues i+3(3.10), i+4 (α) or i+5 (π)
- Hydrogen bonds are parallel to helix axis
- Sidechains are orthogonal to helix axis, pointing outwards
- Helices have a dipole moment due to hydrogen bonds
- Helices can be amphiphatic

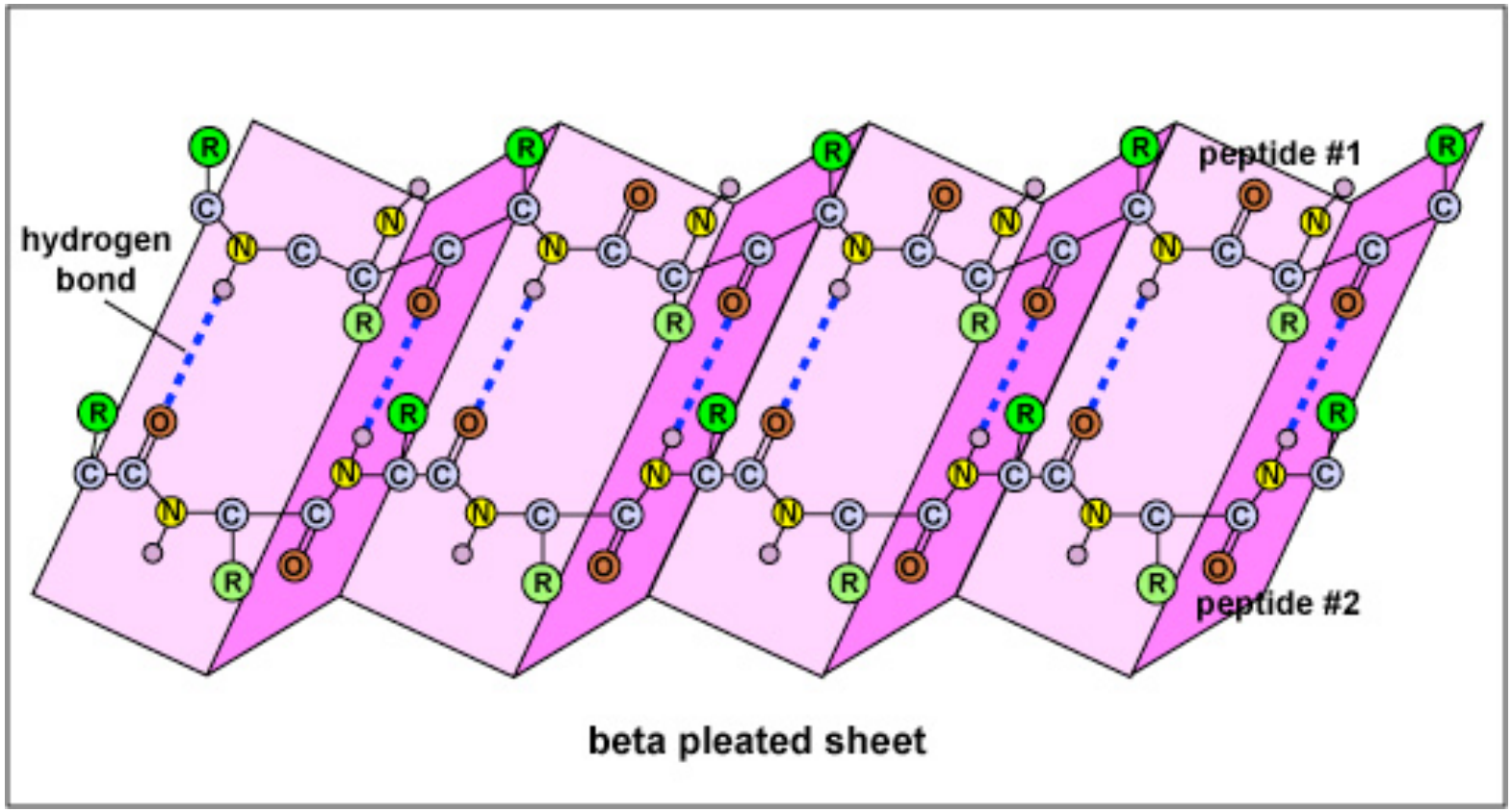
Helical Wheel Projection

HNVGSLFHMADDLGRAMESLVSVMTDEEGAE

Helical Wheel Projection

HNVGSLFHMADDLGRAMESLVSVMTDEEGAE



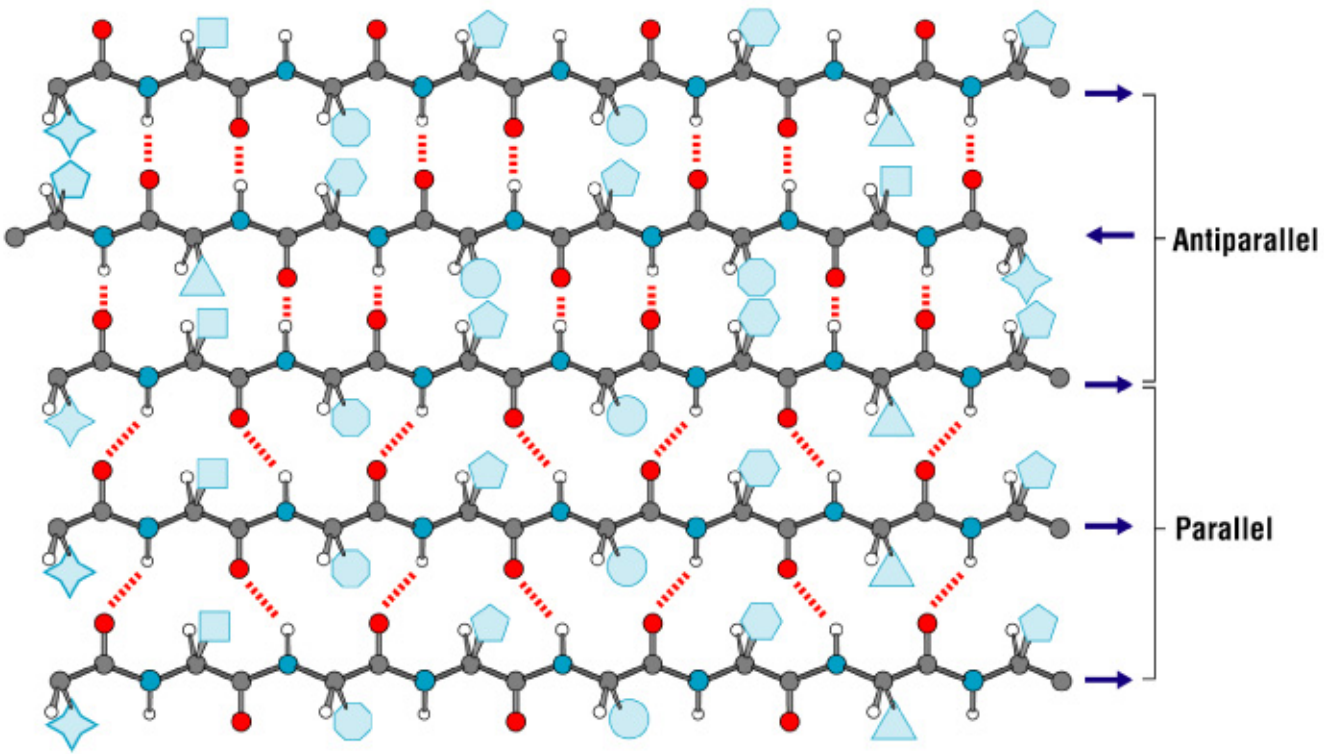


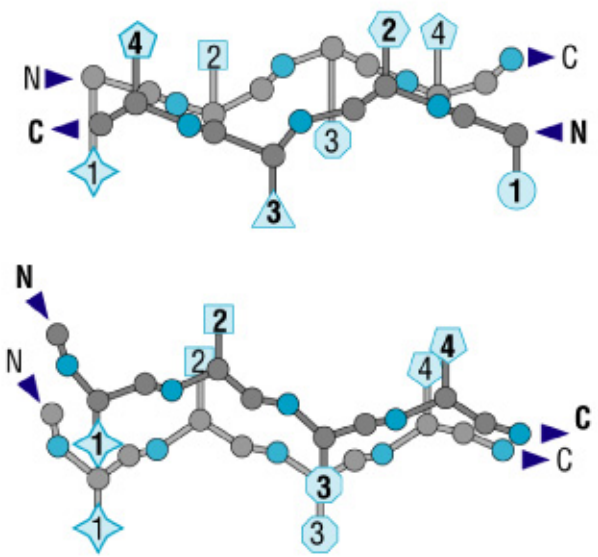
The β sheet (also β-pleated sheet) is consisting of beta strands connected laterally by three or more hydrogen bonds.
A beta strand (also β-strand) is a stretch of amino acids typically 5-10 amino acids long whose peptide backbones are almost fully extended.

Sheet Types

Sheet types	Φ / °	Ψ / °
parallel	-119	113
anti-parallel	-139	135

From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe





Strands and β-Sheets

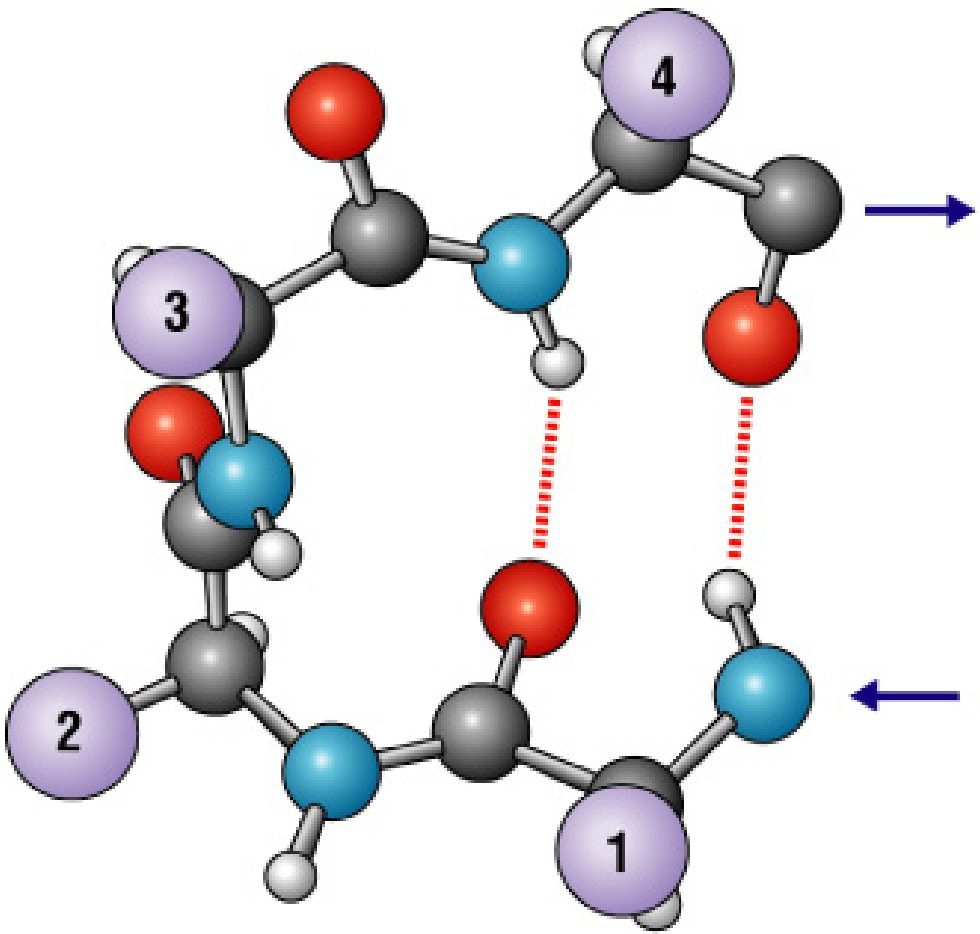
- Extended backbone conformation
- Hydrogen bonds to adjacent strand -> β-sheet
- Hydrogen bonds orthogonal to backbone
- Anti-parallel β-sheet is more stable than parallel
- β-sheets are pleated (not flat) and twisted
- Sidechains form layer on top and bottom

Turn Types

Turn type	i+1 Φ / °	i+1 Ψ / °	i+2 Φ / °	i+2 Ψ / °
I beta	-60	-30	-90	0
I' beta	60	30	90	0
II beta	-60	120	80	0
II' beta	60	-120	-80	0
III beta	-60	-30	-60	-30
III' beta	60	30	60	30
gamma turn	70 ~ 85	-60 ~ -70		
inverse gamma turn	-70 ~ -85	60 ~ 70		

Typical beta-turn

From **Protein Structure and Function**
by Gregory A Petsko and Dagmar Ringe



Secondary Structure Assignment

The DSSP program defines 7 secondary structure states

H : alpha helix
B : residue in isolated beta-bridge
E : extended strand, participates in beta ladder
G : 3-helix (3/10 helix)
I : 5 helix (pi helix)
T : hydrogen bonded turn
S : bend

The secondary structure assignment with DSSP over a database of structures can be used as 'standard of truth' for secondary structure prediction methods.

Kabsch & Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 22(12), 2577-2637 (1983). http:/swift.cmbi.ru.nl/gv/dssp/

Secondary Structure Prediction

- Given a sequence, predict the secondary structure for the entire sequence, i.e. each residue.
- Recent methods report a confidence for each residue to be in the predicted (or other) states.
- The number of secondary states is usually between 3 (helix, strand, coil) and 7 (DSSP annotation).
- Recent methods reach an accuracy of 80% on average by using Multiple Sequence Alignments as input.

What we need to do

1) Train a method on a *diverse* set of proteins of known structure

2) Test the method on a test set separate from our training set

3) Assess our results in a useful way against a standard of truth

4) Compare to already existing methods using the same assessment

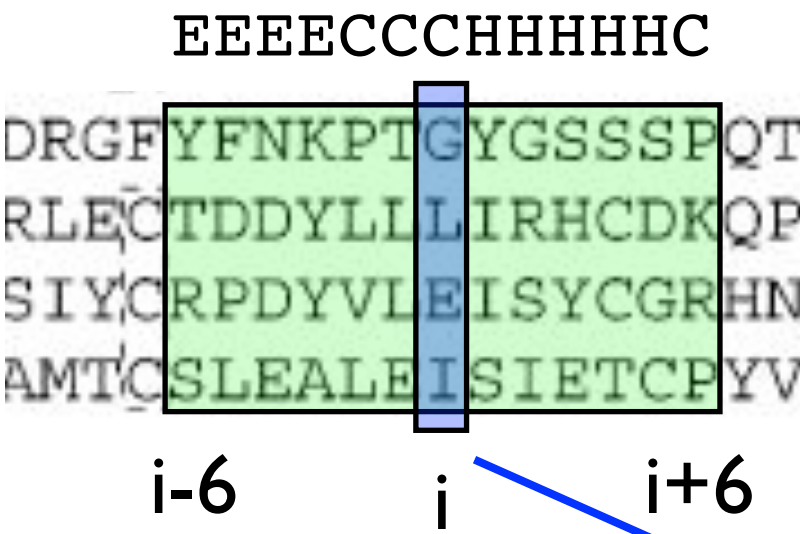
Prediction is Based on Conditional Probability

p(ssH) : probability of being in sec.str. state 'Helix'
p(aaX) : probability of being a.a. residue 'X'

$$p_i(ssH) = \sum_{20} p_i(aaX) * p(ssH \mid aaX)$$

In words: The probability of sequence position 'i' being a helix is equal to the probability of observing amino acid residue 'X' in position 'i' times the recorded (training!) probability of observing a helix if 'X' is present (secondary structure propensity).

Prediction Scheme with Window



sec. str. assignment

multiple alignment

machine learning

Training

$$p_i(ssH) = \sum p_i(aaX) * p(ssH \mid aaX)$$

Prediction

$$p_i(ssH) = \sum p_i(aaX) * p(ssH \mid aaX)$$

Typical Machine Learning Methods

- Neural Network
- Hidden Markov Model
- Support Vector Machine

Prediction Result

```
cd01053 AOX Alternative oxidase (AOX) is a mitochondrial ubiquinol oxidase found in plants and some fungi and
protists. AOX is a member of the ferritin-like diiron-carboxylate superfamily. The plant mitochondrial protein alt
Probab=100.00  E-value=0  Score=520.37  Aligned_cols=163  Identities=44%  Similarity=0.735  Sum_probs=0.0

Q ss_pred              CCCCEEEEHHHHHCCHHHHHHHHHHHHHHCCCCCH-HHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHH
Q ss_conf              96661440132542537999999999985267881-689999999987899999999983788999999999999999
Q Wed_Oct_01_11:     1 NYILRAVFLESVASIPGLVCSNLHHLRCLRRLQPD-SWIKPLVDEAENERMHLLAVRTYTKLTAVQKLFIRITQFSFVTL   79 (203)
Q Consensus          1 ~~~~R~~~LETVA-vPgmv~~~~~HL-sLr~~~~d--wI--ll-EaeNErmHLl-f--l--p-w--R-lv--aQ-vfy--   79 (203)
                     .|+|||||||||||||+||++||+|+|+|++| ||||+||+||||||||||++++||+|++|.++..+|+||||+
T Consensus         4 r~~~R~~~LEtVA-vPgmv~~~~~Hl-slr~~~~d~~~i--ll-EaeNErmHL--f--l--p~~~~r--i~~~q-vfy--   83 (172)
T cd01053           4 RWLRRFIFLETVARVPGMVGGSLLHLYELRGWWRDGGWIKTLLAEAENERMHLLTFEELGGPGLWERRFVAAHQAVYYNA   83 (172)
T ss_pred              CCCCHHEEHHHHHCCHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHHHHHHHHH
T ss_conf              86441010344543626899999999987512184599999998986999999996088599999999999999999


Q ss_pred              HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCHHHCCCC-----CCCCHHHHHHHHHHHHHH
Q ss_conf              99999866899999999889999999999998733555566662208770879-----8774899999987367789
Q Wed_Oct_01_11:    80 FSFLFVFAPRTSHRLVGFLEEHAVDSYTEMIRRIDSNTLENRPATQITKDYWGLP-----EDATLRDALLVIRADEADHR  154 (203)
Q Consensus         80 ~~~~YLisPr-Ahr-vgylEEEAV-TYT--L--id-g~~~~~pAP-iAi-Yw-l-------a-l-Dvi--IRaDEa-Hr  154 (203)
                    |+++||+|||+||||||||||+|||+||+++|+ |+|||+|||+||++|     +++||+|||++||+||++||
T Consensus        84 ~~~~Yl-sPr-Ahr-vgylEeeAV-TYt--l--i--g~~~~~paP-iAi-Yw-l-d--r---a-l-Dvi--IR-DEa-H-  162 (172)
T cd01053          84 FFLLYLISPRMAHRFVGYLEEEAVSTYTHFLKEIEEGLK-DQPAPEIAIDYWRLPDLYRPKIATLRDVIVAIRADEAEHR  162 (172)
T ss_pred              HHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHHHHHHCC-CCCCCHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHH
T ss_conf              99999982007999999999989999999999988600 10-37886789987056544678654499999999986788 99
```

Some Servers

- PSI-pred uses PSI-BLAST profiles
- JPRED Consensus prediction
- PHD home page – all-in-one prediction, includes secondary structure
- nnPredict – uses neural networks
- BMERC PSA Server
- IBIVU YASPIN server
- BMC launcher – choose your prediction program

Learning Outcomes

- Peptide bond geometry
- Ramachandran plot
- Secondary Structure propensities
- Helix, strand, sheet, turn
- Secondary Structure prediction

Suggested readings

1) DSSP: Kabsch W, Sander C.Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 22(12):2577-637.
2) Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol. 1978;47:45-148.
3) Chen H, Gu F, Huang Z. Improved Chou-Fasman method for protein secondary structure prediction. BMC Bioinformatics. 2006 12:7 Suppl 4:S14
4) Rost B, Sander C, Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol. 1993 232(2):584-99.