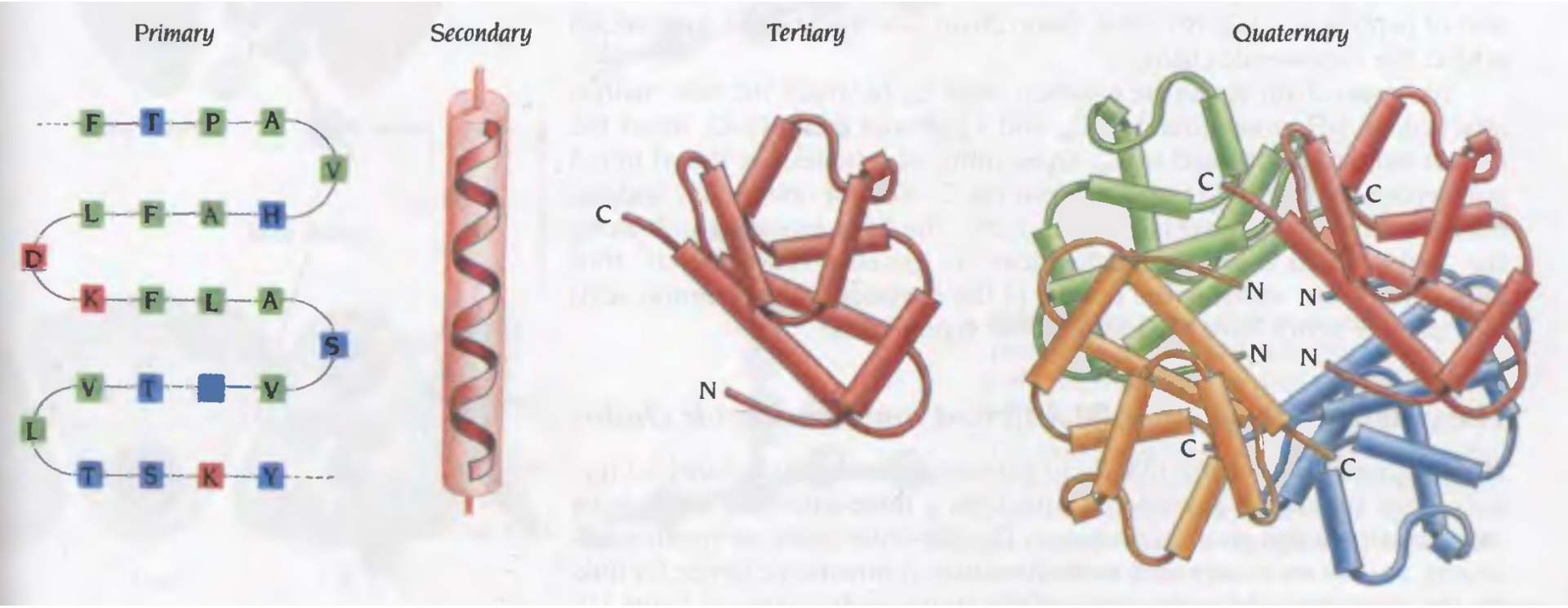
Lecture 06-07

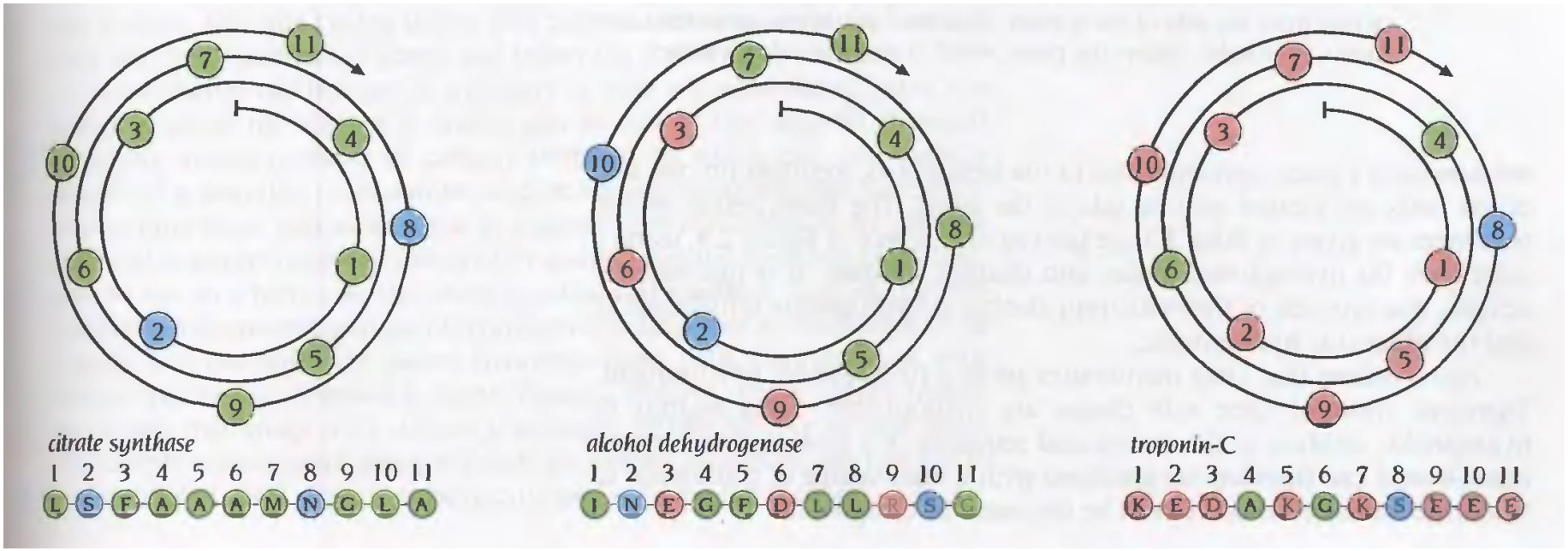
Proteins


Primary
Secondary
Tertiary
Quaternary
F T P A V
L F A H
D
K F L A S
V T V
L
T S K Y
C
N
C
C
N N
N N
C
C

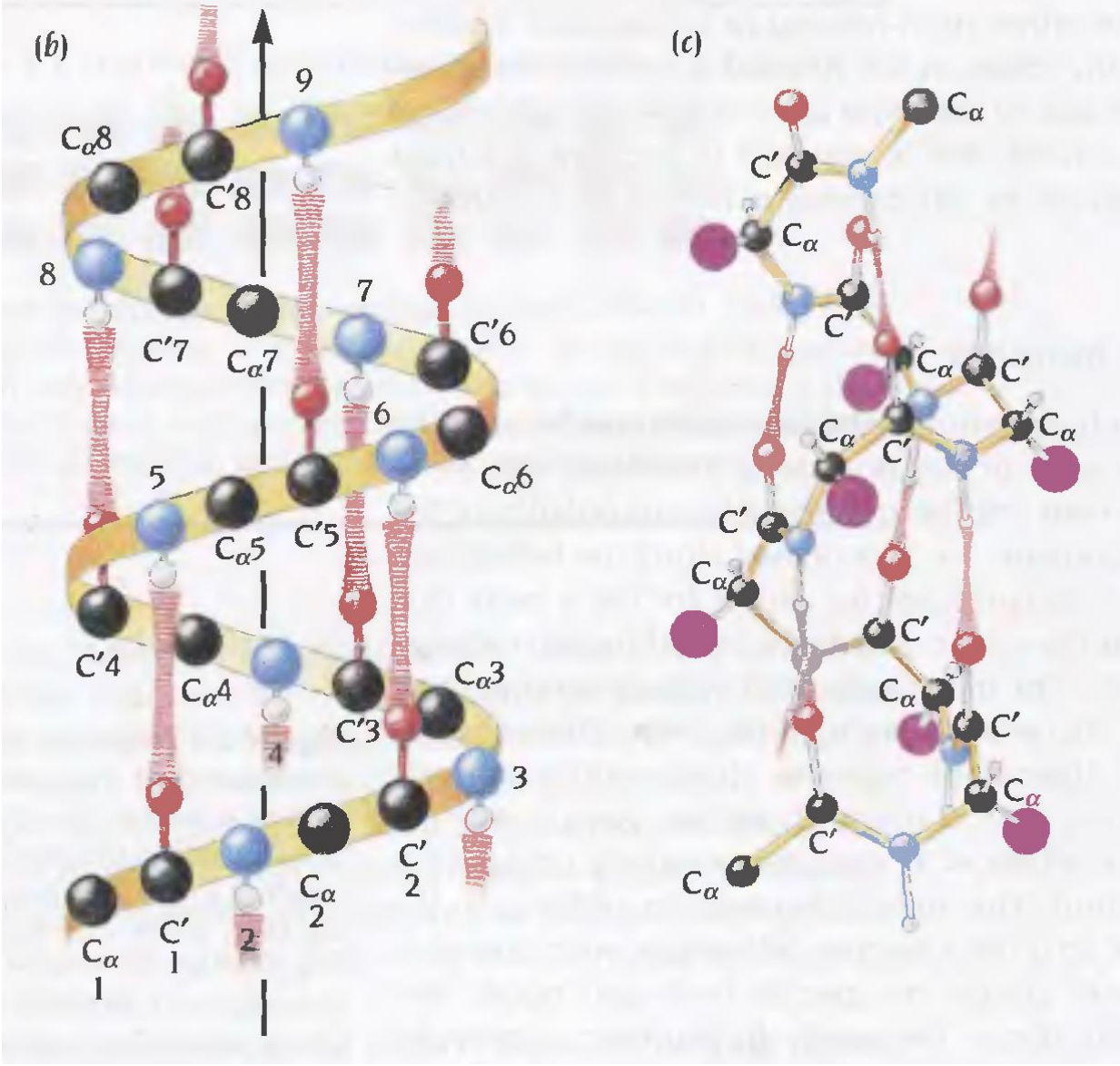
Preference of AA in helix

Table 2.1 Amino acid sequences of three α helices

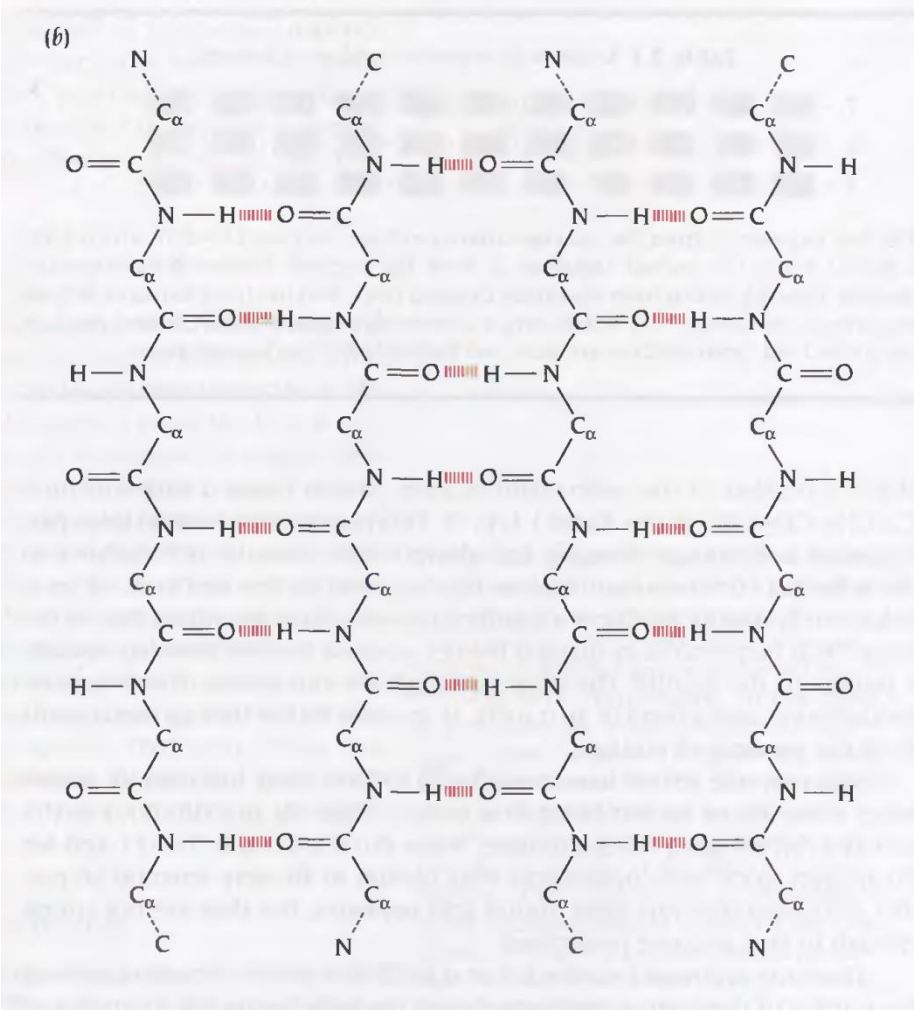
1. - Leu - Ser - Phe - Ala - Ala - Ala - Met - Asn - Gly - Leu - Ala -
2. - Ile - Asn - Glu - Gly - Phe - Asp - Leu - Leu - Arg - Ser - Gly -
3. - Lys - Glu - Asp - Ala - Lys - Gly - Lys - Ser - Glu - Glu - Glu -

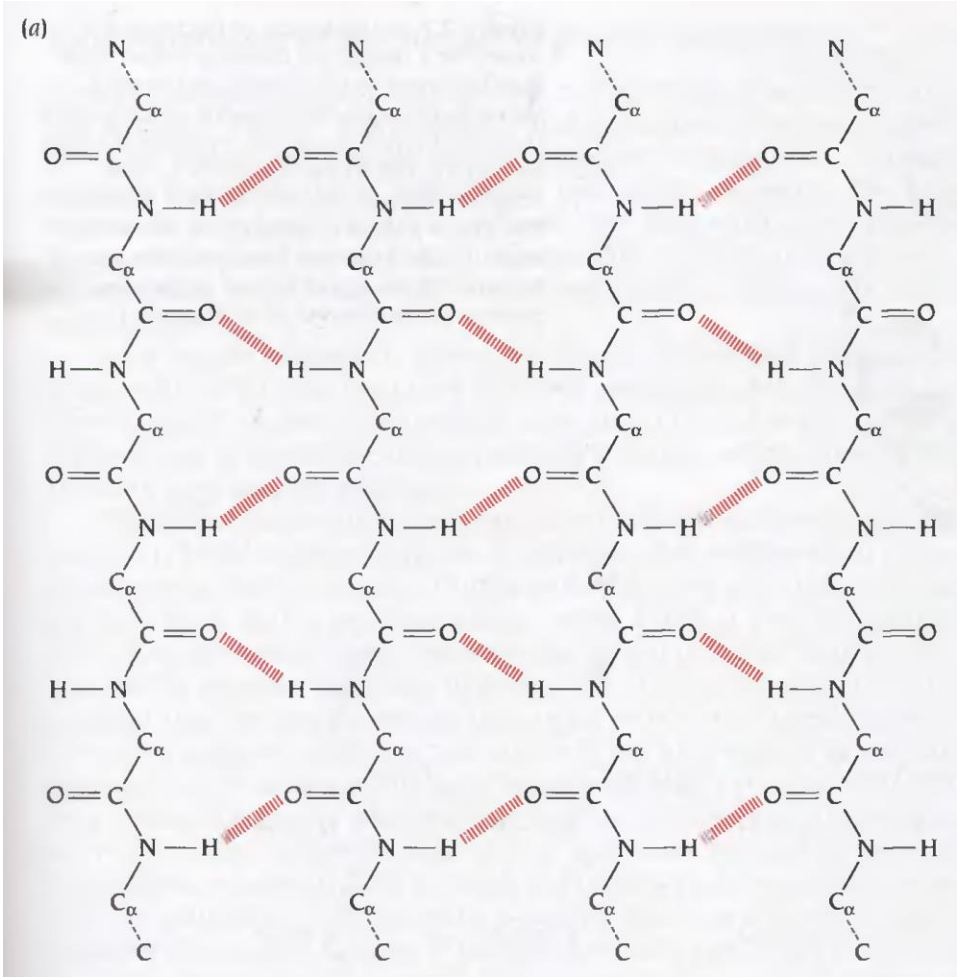


Formation of helix

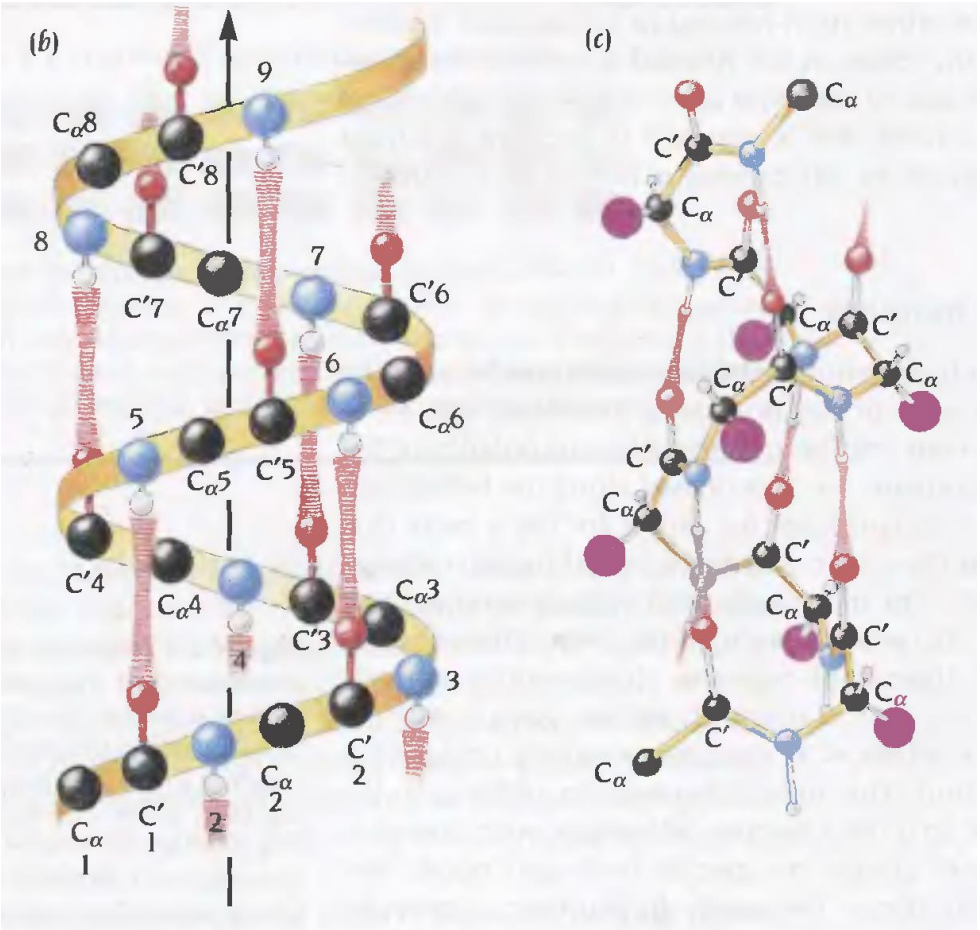


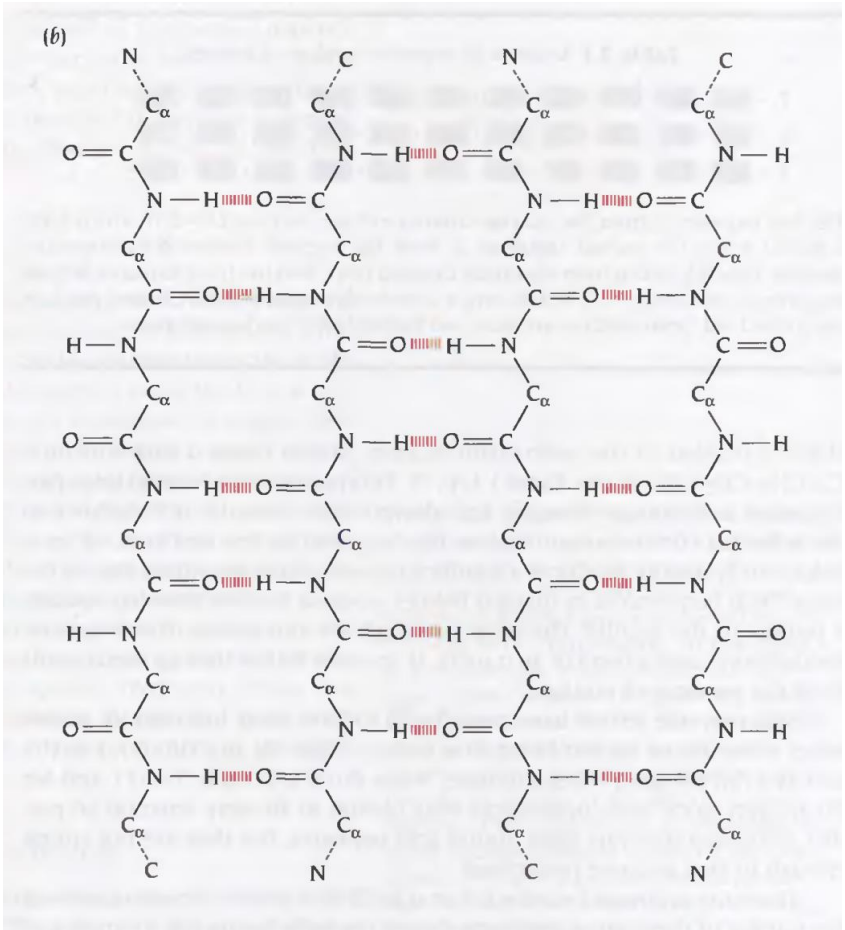
Formation of sheet





Observation based post-processing



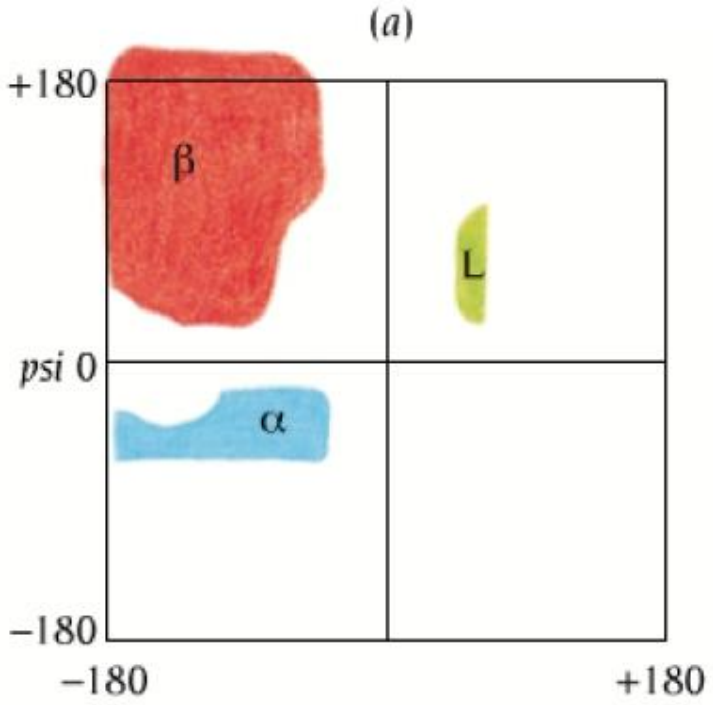


A minimum consecutive amino acids are required for the formation of the helix and sheet.

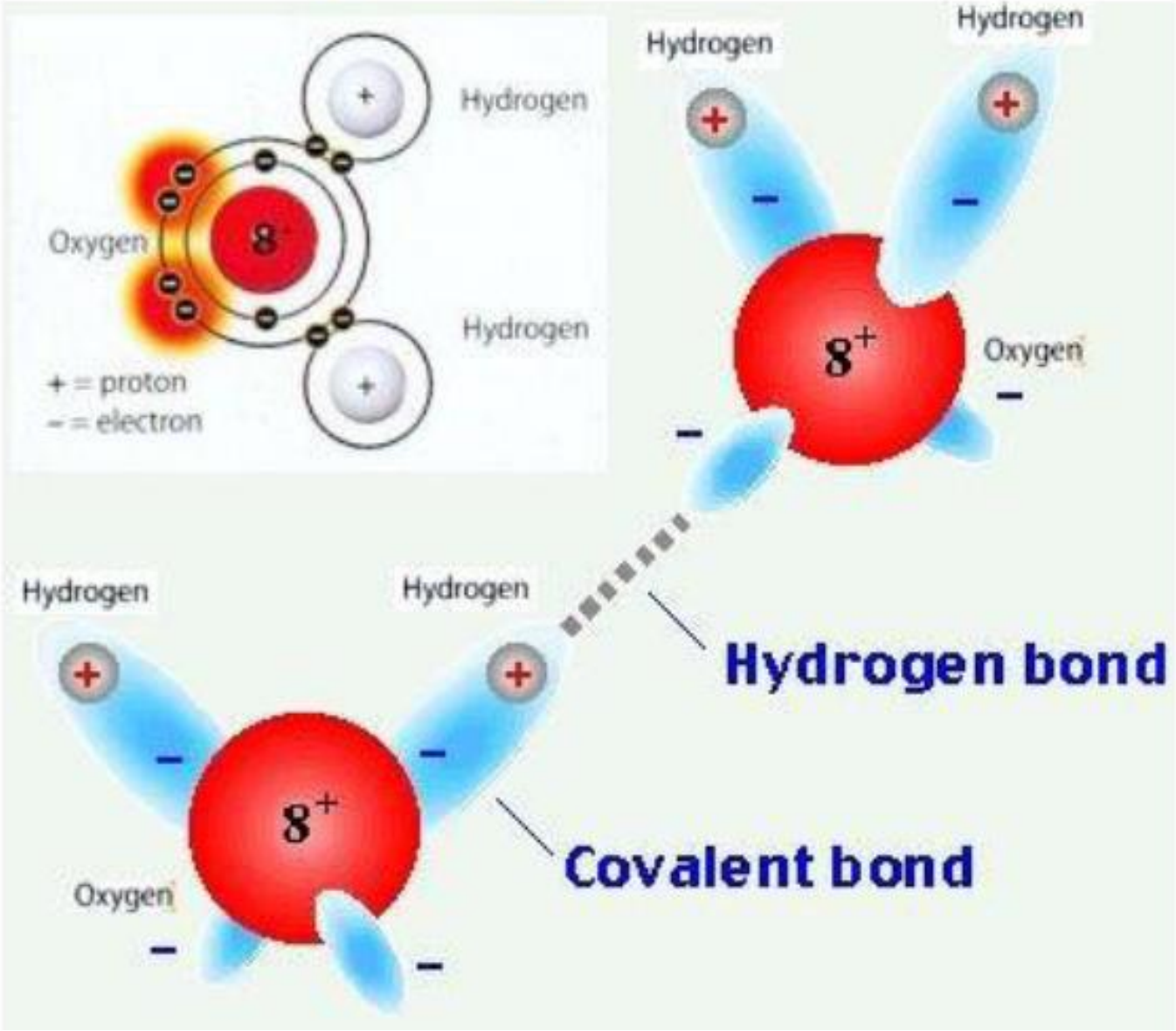
- Improves the accuracy by 3-5%.

What basic structural information is used ?

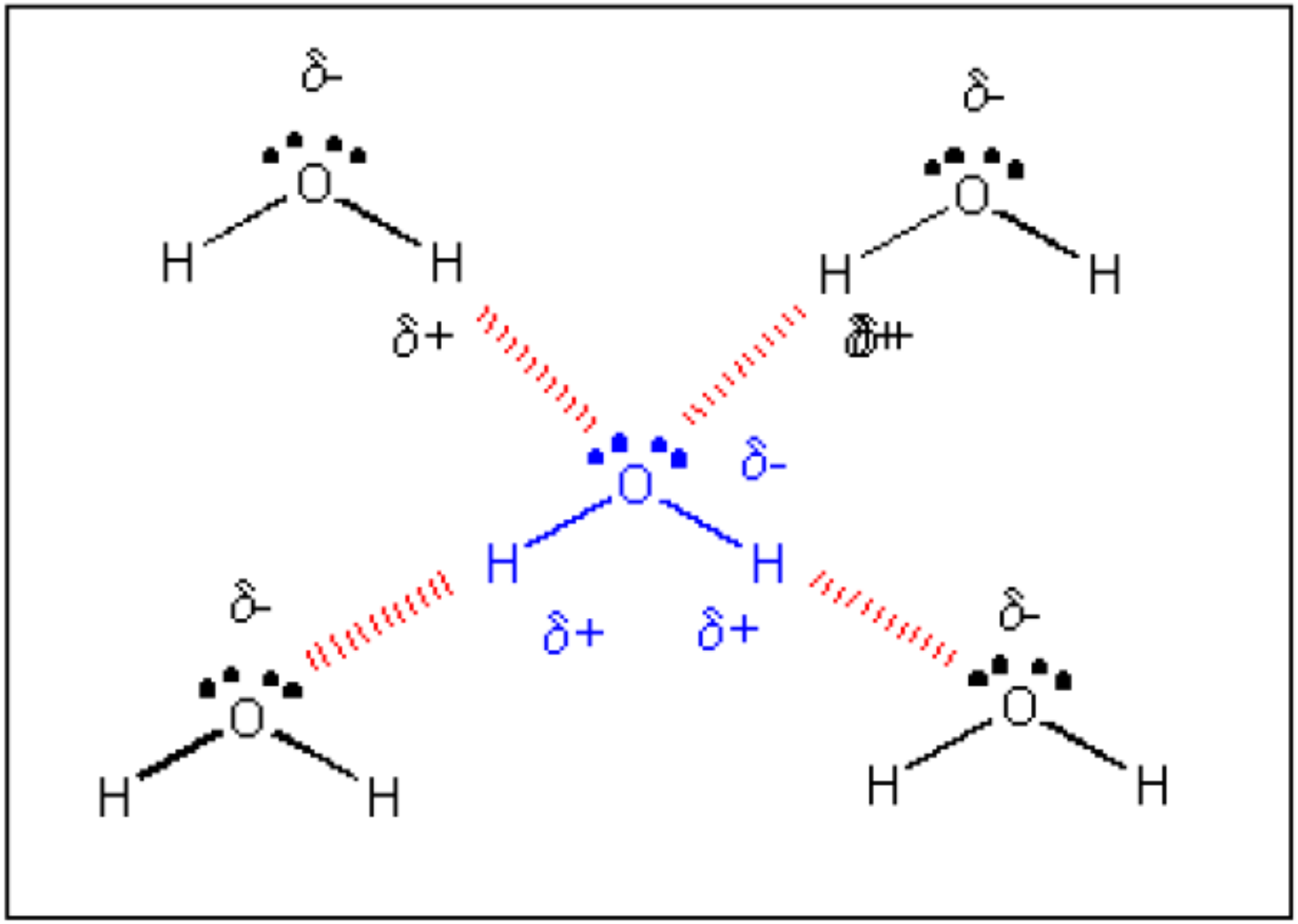
- Hydrogen bond patterns
- Backbone dihedral angles



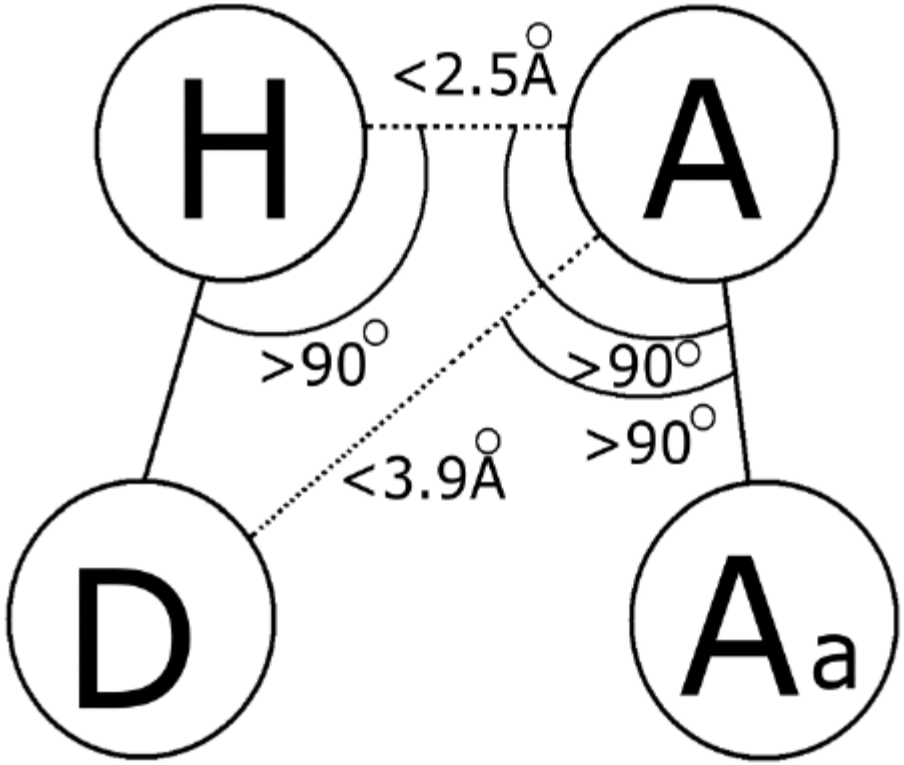
Hydrogen Bond



Hydrogen Bond



Hydrogen Bond



Baker and Hubbard, 1984

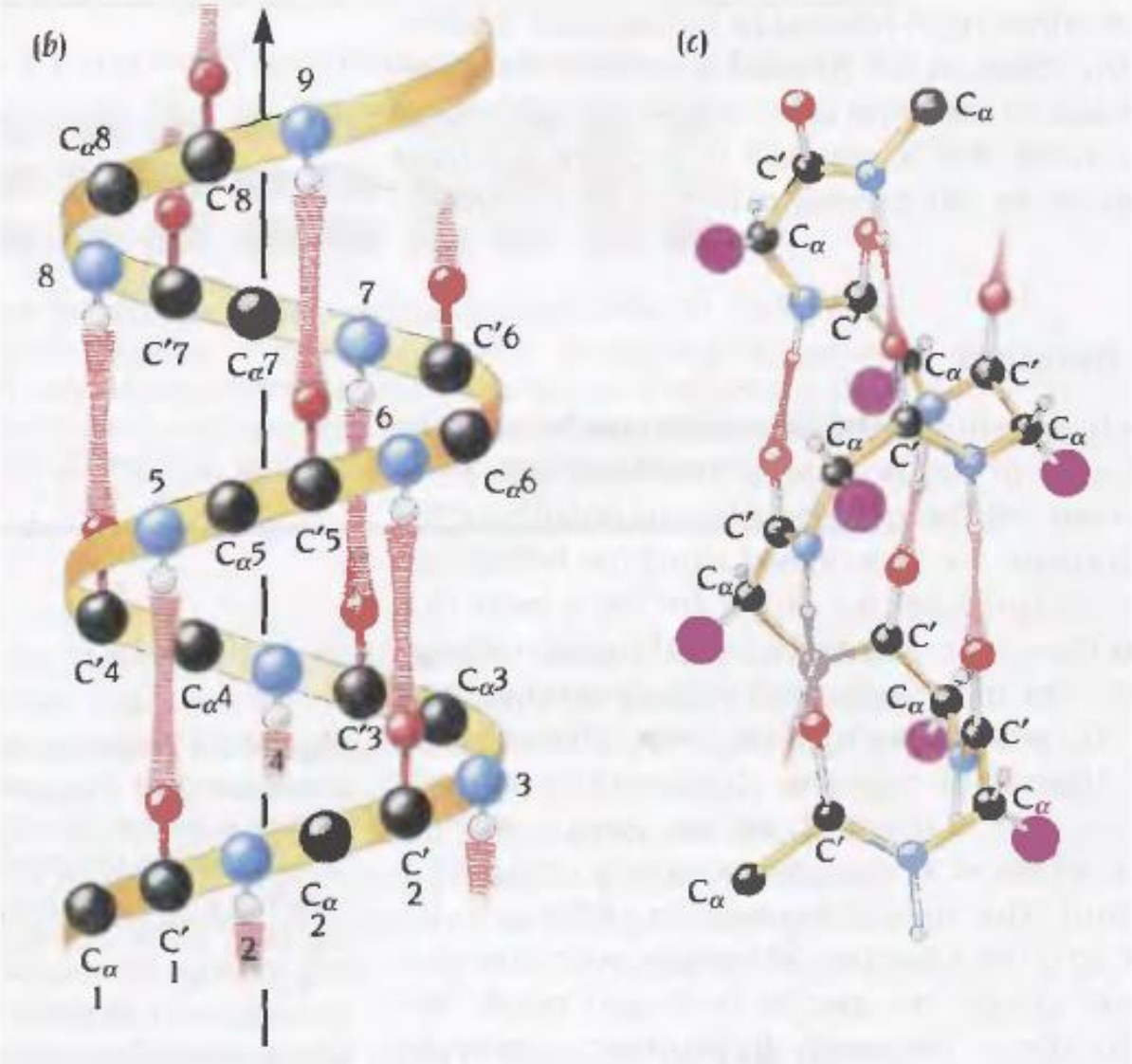
PDB file format

```
ATOM      1  N   ALA A   4      11.751  37.846  29.016  1.00 44.65           N
ATOM      2  CA  ALA A   4      12.501  39.048  28.539  1.00 30.68           C
ATOM      3  C   ALA A   4      13.740  38.628  27.754  1.00 24.74           C
ATOM      4  O   ALA A   4      14.207  37.495  27.890  1.00 25.59           O
ATOM      5  CB  ALA A   4      12.902  39.919  29.730  1.00 16.77           C
ATOM      6  N   TYR A   5      14.235  39.531  26.906  1.00 19.29           N
ATOM      7  CA  TYR A   5      15.552  39.410  26.282  1.00  8.51           C
ATOM      8  C   TYR A   5      16.616  38.913  27.263  1.00  6.11           C
ATOM      9  O   TYR A   5      17.187  37.844  27.068  1.00 17.99           O
ATOM     10  CB  TYR A   5      15.988  40.762  25.702  1.00  2.00           C
ATOM     11  CG  TYR A   5      17.319  40.745  24.982  1.00  2.00           C
ATOM     12  CD1 TYR A   5      17.411  40.331  23.653  1.00 19.29           C
ATOM     13  CD2 TYR A   5      18.476  41.210  25.604  1.00  9.65           C
ATOM     14  CE1 TYR A   5      18.629  40.396  22.953  1.00  2.00           C
ATOM     15  CE2 TYR A   5      19.703  41.271  24.914  1.00  8.78           C
ATOM     16  CZ  TYR A   5      19.763  40.863  23.594  1.00  8.76           C
ATOM     17  OH  TYR A   5      20.971  40.889  22.920  1.00  8.23           O
ATOM     18  N   ILE A   6      16.789  39.630  28.369  1.00 14.56           N
ATOM     19  CA  ILE A   6      17.791  39.281  29.375  1.00 23.27           C
```

Homework

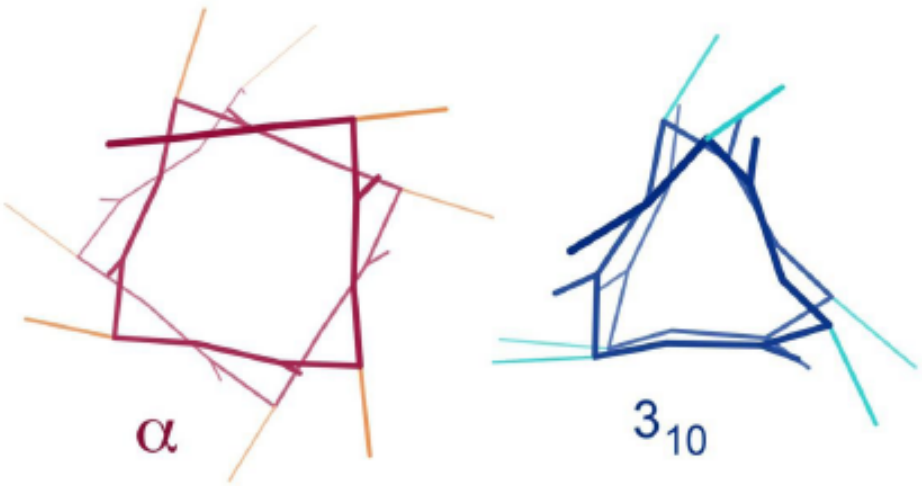
- Write down an algorithm which will check if there is a hydrogen bond between a donor and an acceptor or not.

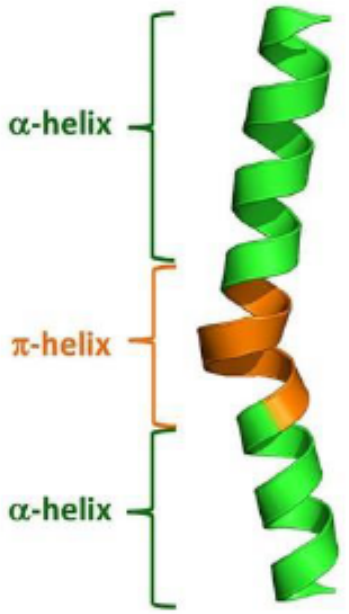
Formation of helix



Book: Branden and Trooze

Helix in protein structure





Facts:

α-helix (3.6_{13}):

Hydrogen bond between i and i+4 residues

100° turn in the helix (3.6 residues per turn)

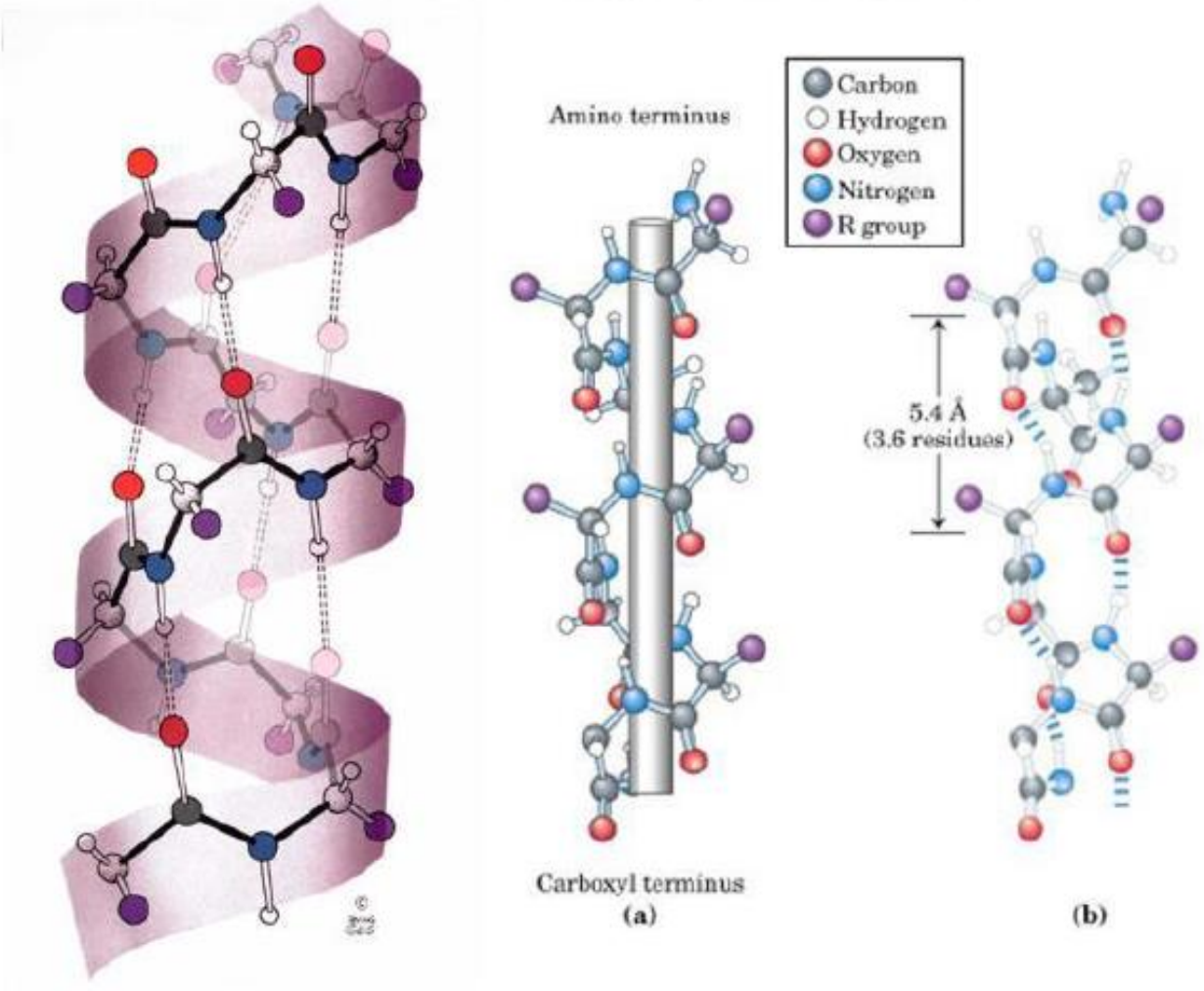
0.15 nm translation along helical axis

3_{10}-helix:

Hydrogen bond between i and i+3 residues

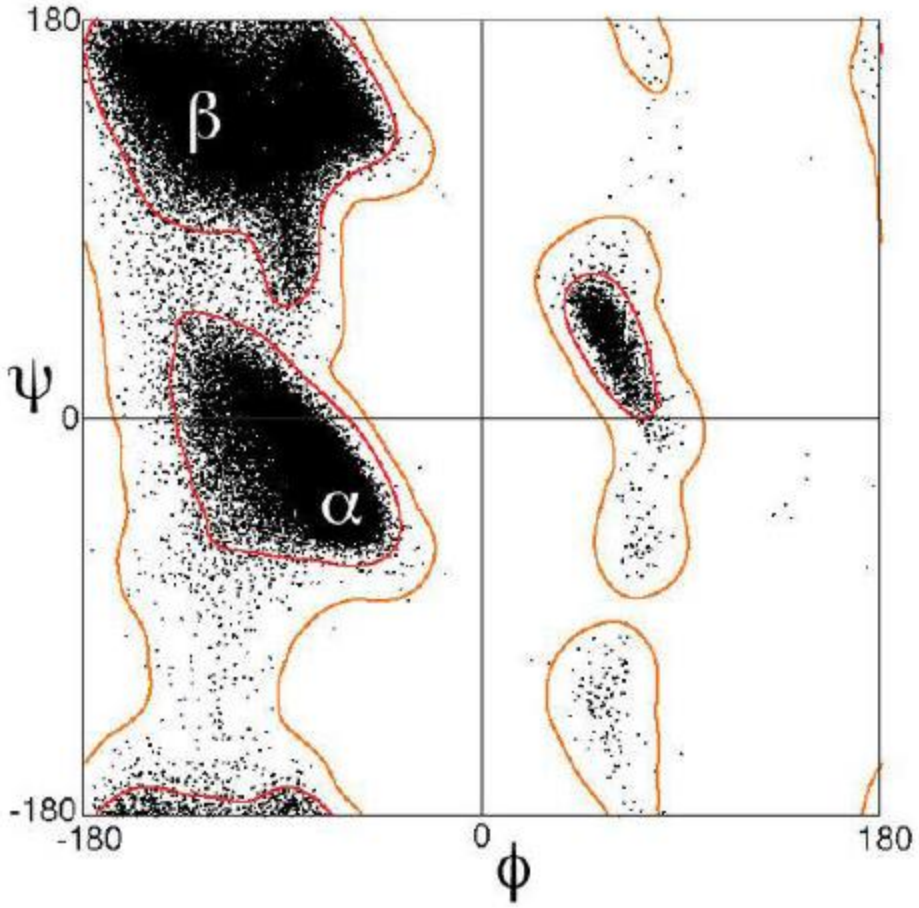
π-helix:

The α-helix

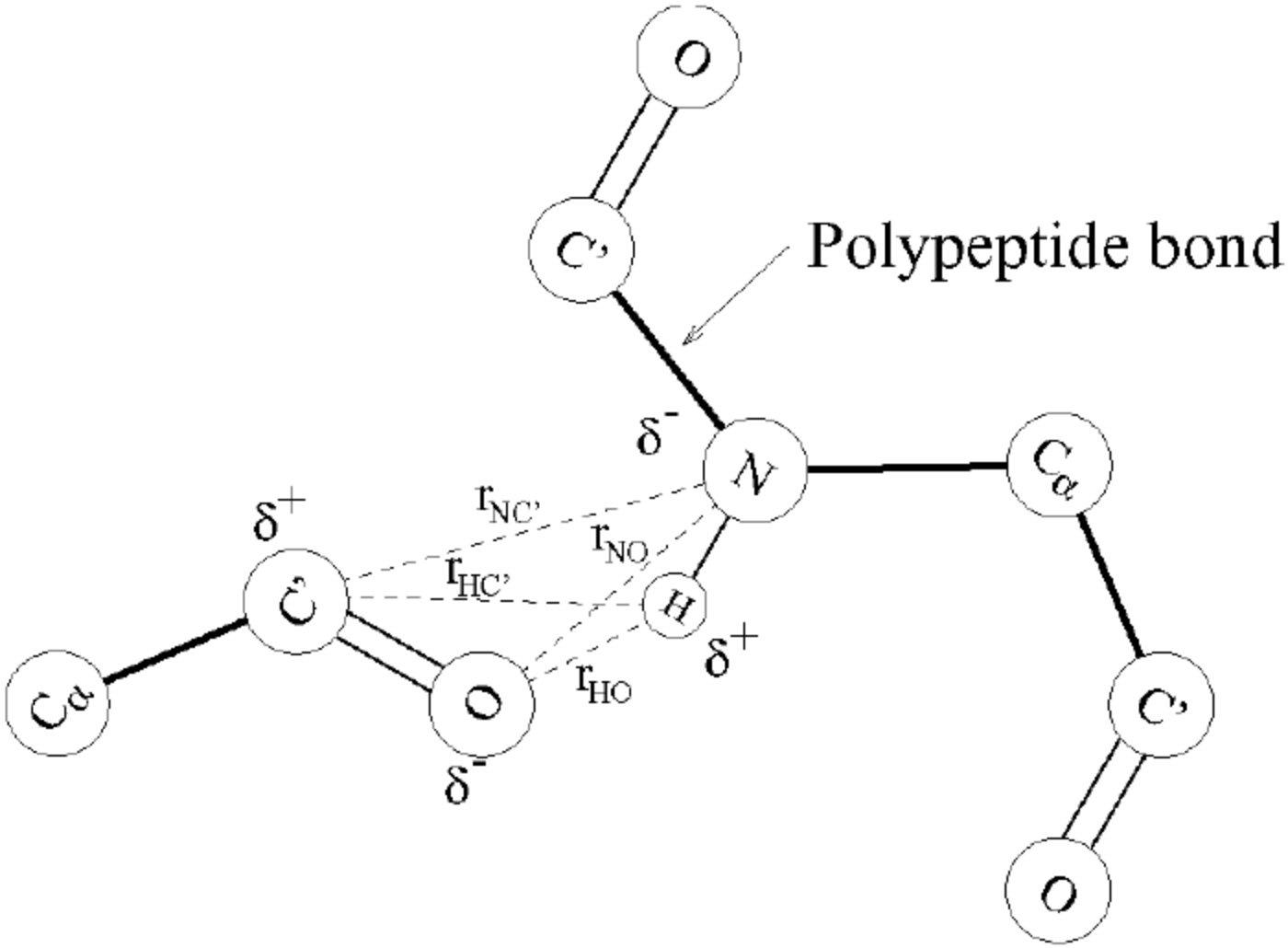


Source: http://xray.bmc.uu.se/Courses/bioinformatik2003/Intro/a-helix.jpg

Ramachandran Plot



$q > 120°$ and $r_{HO} < 2.5$ Å



DSSP Algorithm

- The so-called "Dictionary of Secondary Structure of Proteins" (DSSP) by Kabsch and Sander makes its sheet and helix assignments solely on the basis of backbone-backbone hydrogen bonds.

- The DSSP method defines a hydrogen bond when the bond energy is below -0.5 kcal/mol from a Coulomb approximation of the hydrogen bond energy.

- The structural assignments are defined such that visually appealing and unbroken structures result.

- In case of overlaps, alpha-helix is given first priority.

DSSP Algorithm

- The helix definition does not include the terminal residue having the initial and final hydrogen bonds in the helix.

- A minimal size helix is set to have two consecutive hydrogen bonds in the helix, leaving out single helix hydrogen bonds, which are assigned as turns (state 'T').

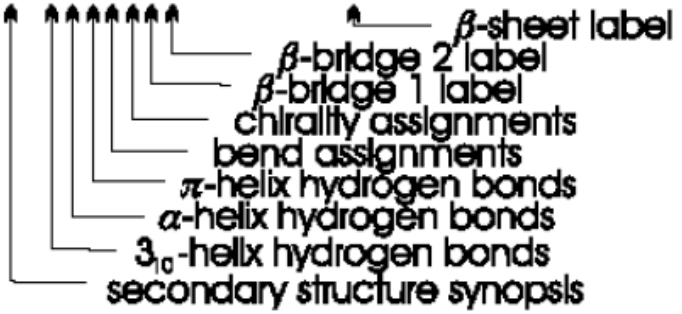
- beta-sheet residues (state 'E') are defined as either having two hydrogen bonds in the sheet, or being surrounded by two hydrogen bonds in the sheet.

- The minimal sheet consists of two residues at each partner segment.

DSSP Algorithm

PDB:1crn

#	RESIDUE	AA	STRUCTURE	BP1	BP2	ACC	N-H-->O	O-->H-N	N-H-->O	O-->H-N
....										
15	15	V	H >< S+	0	0	99	-4,-1.7	3,-1.3	2,-0.2	-2,-0.2
16	16	c	H 3<>S+	0	0	18	-4,-2.5	5,-0.8	1,-0.3	-2,-0.2
17	17	R	H ><5S+	0	0	94	-4,-2.0	3,-1.6	1,-0.2	-1,-0.3
18	18	L	T <<5S+	0	0	144	-3,-1.3	-1,-0.2	-4,-0.6	-2,-0.2
19	19	P	T 3 5S-	0	0	107	0, 0.0	-1,-0.3	0, 0.0	-2,-0.1
20	20	G	T < 5 +	0	0	53	-3,-1.6	-3,-0.2	1,-0.2	-2,-0.1
21	21	T	< -	0	0	37	-5,-0.8	-1,-0.2	1,-0.1	5,-0.1
22	22	P	>> -	0	0	81	0, 0.0	4,-2.2	0, 0.0	3,-0.7
23	23	E	H 3> S+	0	0	70	1,-0.2	4,-2.5	2,-0.2	5,-0.1
24	24	A	H 3> S+	0	0	63	1,-0.2	4,-1.7	2,-0.2	-1,-0.2
25	25	I	H <> S+	0	0	99	-3,-0.7	4,-1.8	2,-0.2	-1,-0.2
26	26	c	H X S+	0	0	0	-4,-2.2	4,-1.9	2,-0.2	6,-0.4
27	27	A	H X S+	0	0	12	-4,-2.5	4,-2.7	-5,-0.2	5,-0.5
28	28	T	H < S+	0	0	120	-4,-1.7	-1,-0.2	1,-0.2	-2,-0.2
29	29	Y	H < S+	0	0	176	-4,-1.8	-1,-0.2	-5,-0.2	-2,-0.2
30	30	T	H < S-	0	0	24	-4,-1.9	-2,-0.2	-3,-0.2	-3,-0.2
31	31	G	S < S+	0	0	35	-4,-2.7	-3,-0.2	1,-0.4	-4,-0.1
32	32	b	-	0	0	5	-5,-0.5	-1,-0.4	-6,-0.4	2,-0.3
33	33	I	E -A	3	0A	51	-30,-2.8	-30,-2.4	-3,-0.1	2,-0.5
34	34	I	E -A	2	0A	78	-2,-0.3	-32,-0.2	-32,-0.2	3, 0.0
....										



STRIDE Algorithm

•The secondary STRuctural IDEntification method by Frishman and Argos uses an empirically derived hydrogen bond energy **and phi-psi torsion angle criteria** to assign secondary structure.

•Torsion angles are given alpha-helix and beta-sheet propensities according to how close they are to their regions in Ramachandran plots.

•The parameters are optimized to mirror visual assignments made by crystallographers for a set of proteins.

•By construction, the STRIDE assignments agreed better with the expert assignments than DSSP, at least for the data set used to optimize the free parameters.

STRIDE Algorithm

•Like **DSSP, STRIDE** assigns the shortest alpha-helix ('H') if it contains at least two consecutive *i* - *i+4* hydrogen bonds.

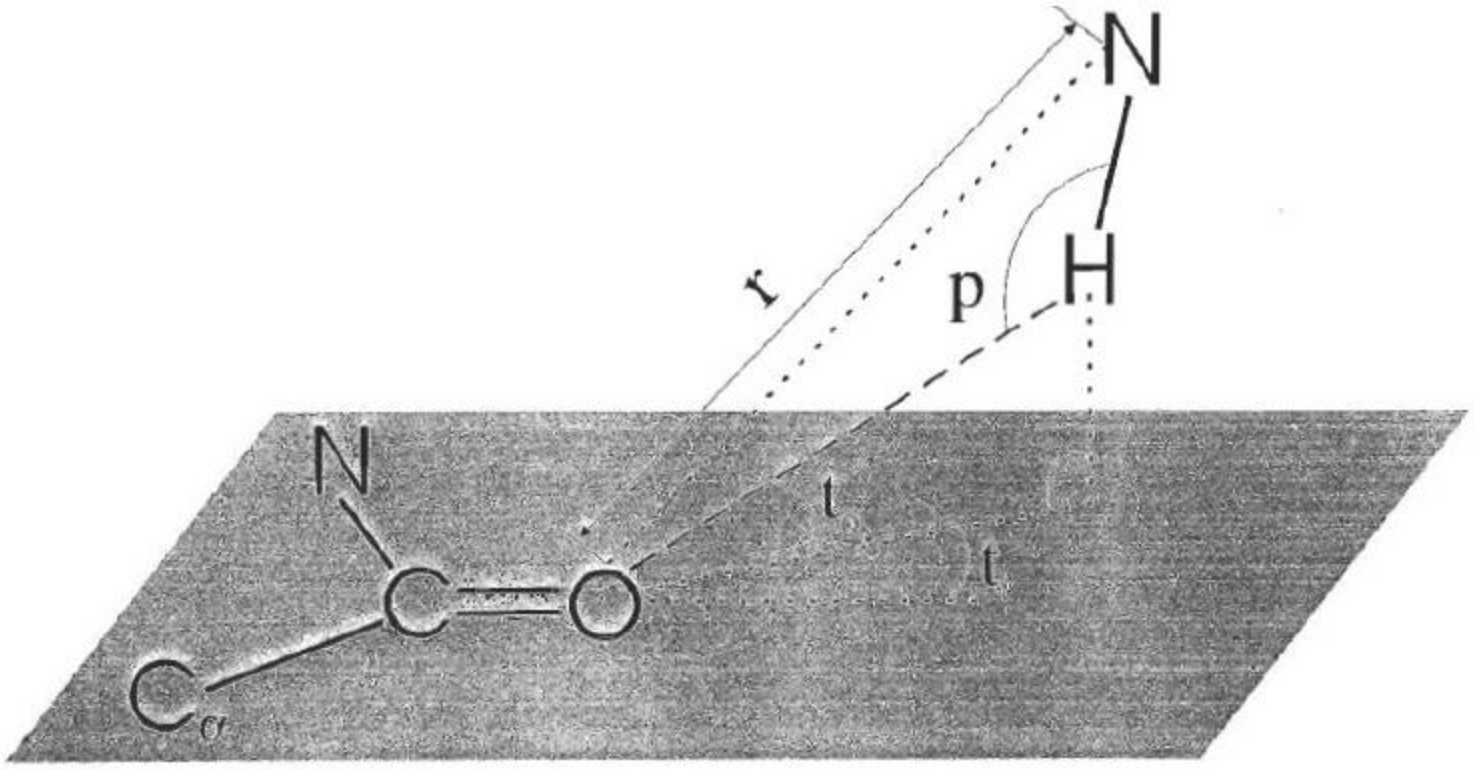
•In contrast to DSSP, helices are elongated to comprise one or both edge residues if they have acceptable phi-psi angles, similarly a short helix can be vetoed.

•hydrogen bond patterns may be ignored if the phi-psi angles are unfavorable.

•The sheet category does not distinguish between parallel and anti-parallel sheets. The minimal sheet ('E') is composed of two residues.

•The dihedral angles are incorporated into the final sheet assignment criterion as was done for the alpha-helix.

Hydrogen Bond



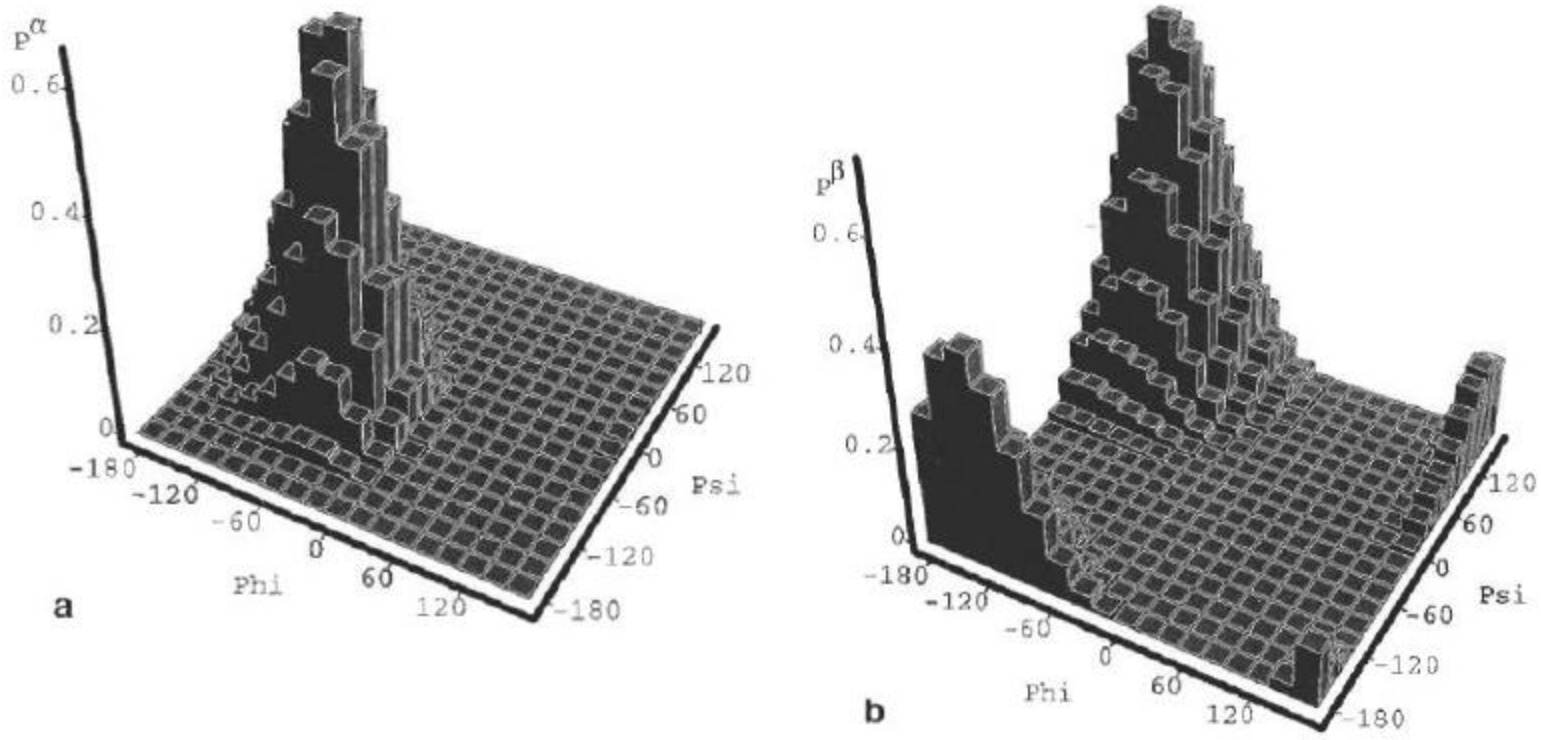
STRIDE

Probabilities for residues in α-helical (a) and β-sheet (b) as calculated in STRIDE

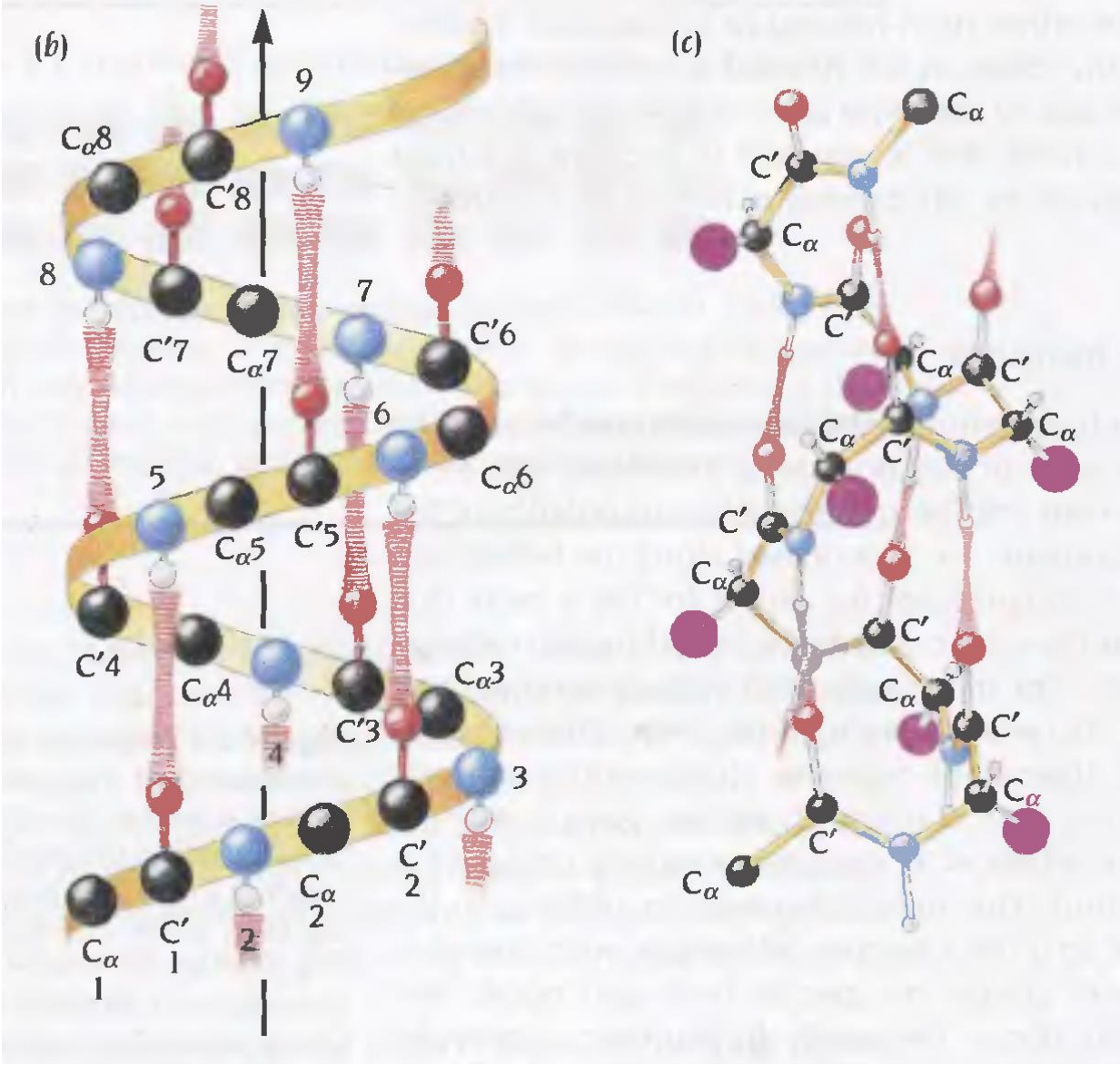
$$P_i^{\alpha} = \begin{cases} \dfrac{N_i^{\alpha}}{N_i^{\text{total}}} & \text{if } -180^\circ < \varphi < 10^\circ \text{ and } -120^\circ < \phi < 45^\circ \\ 0 & \text{otherwise} \end{cases}$$

$$P_i^{\beta} = \begin{cases} \dfrac{N_i^{\beta}}{N_i^{\text{total}}} & \text{if } -180^\circ < \varphi < 0^\circ, -180^\circ < \phi < -120^\circ \text{ or } 45^\circ < \phi < 180^\circ \\ 0 & \text{otherwise} \end{cases}$$

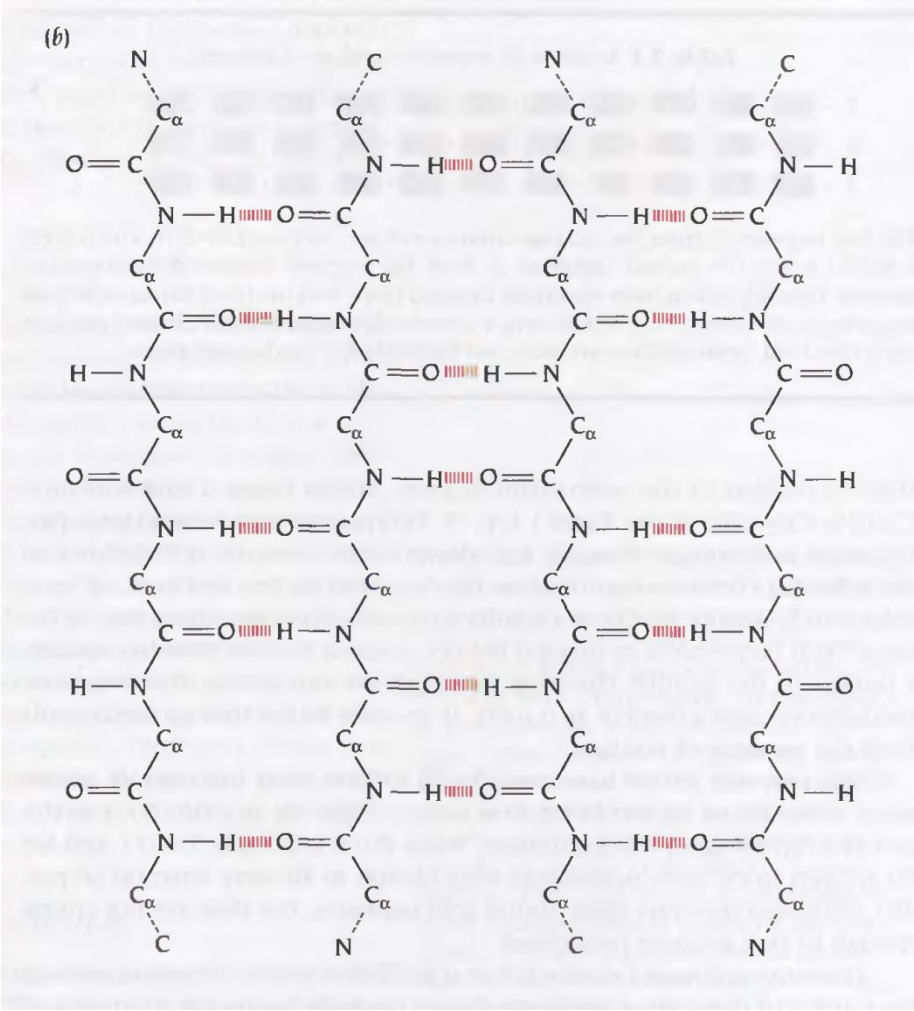
Probabilities for residues in α-helical (a) and β-sheet (b) as calculated in STRIDE

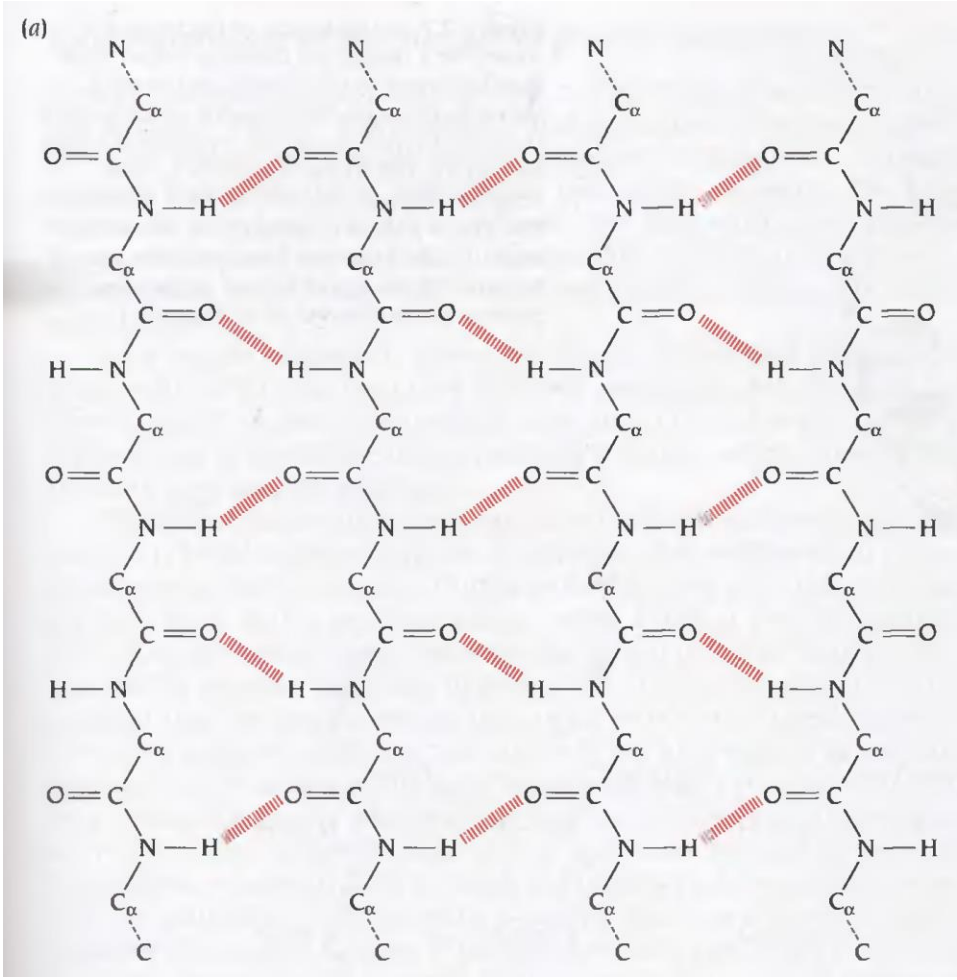


Formation of helix



Formation of sheet





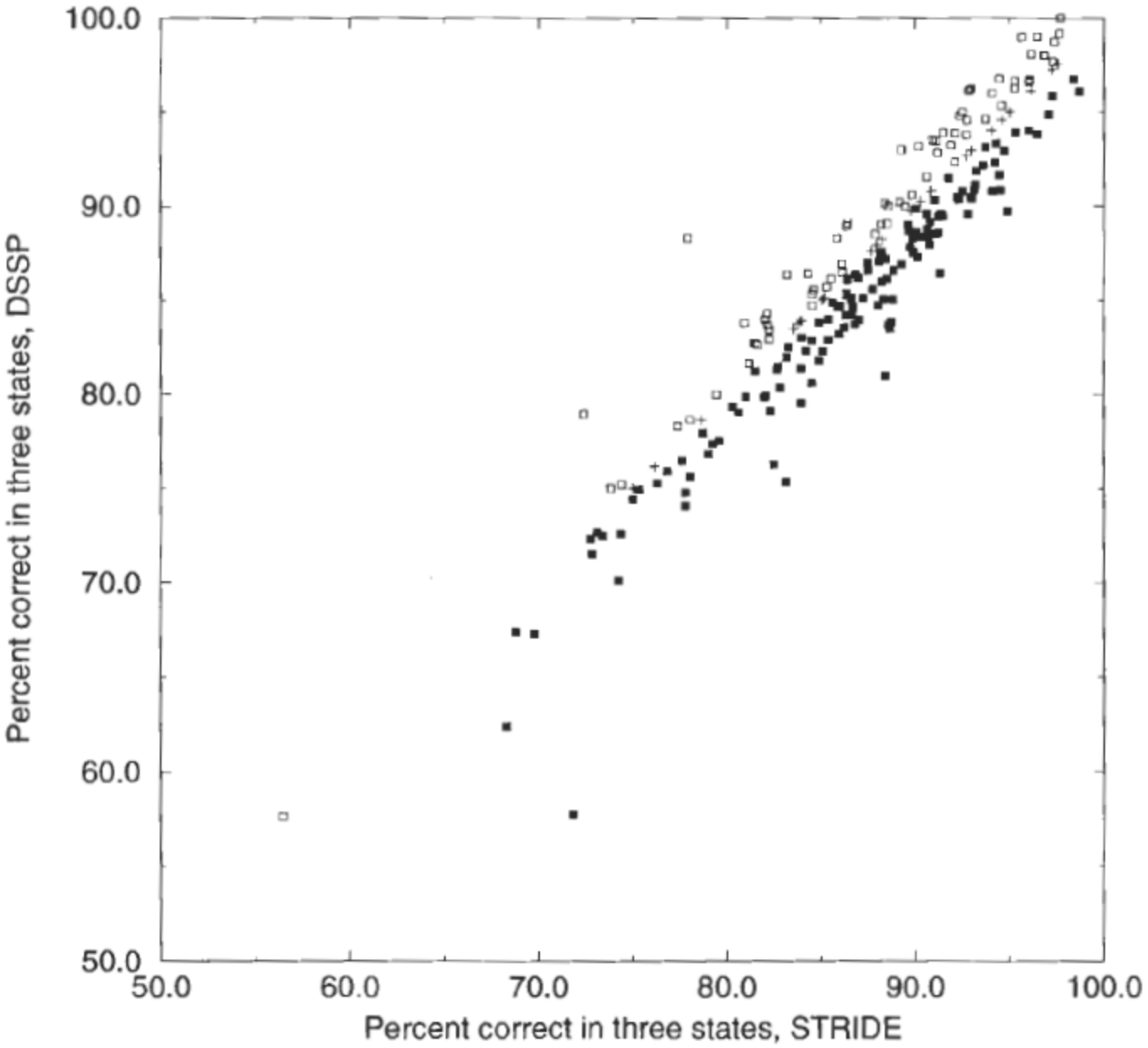
STRIDE Algorithm

PDB:1crn

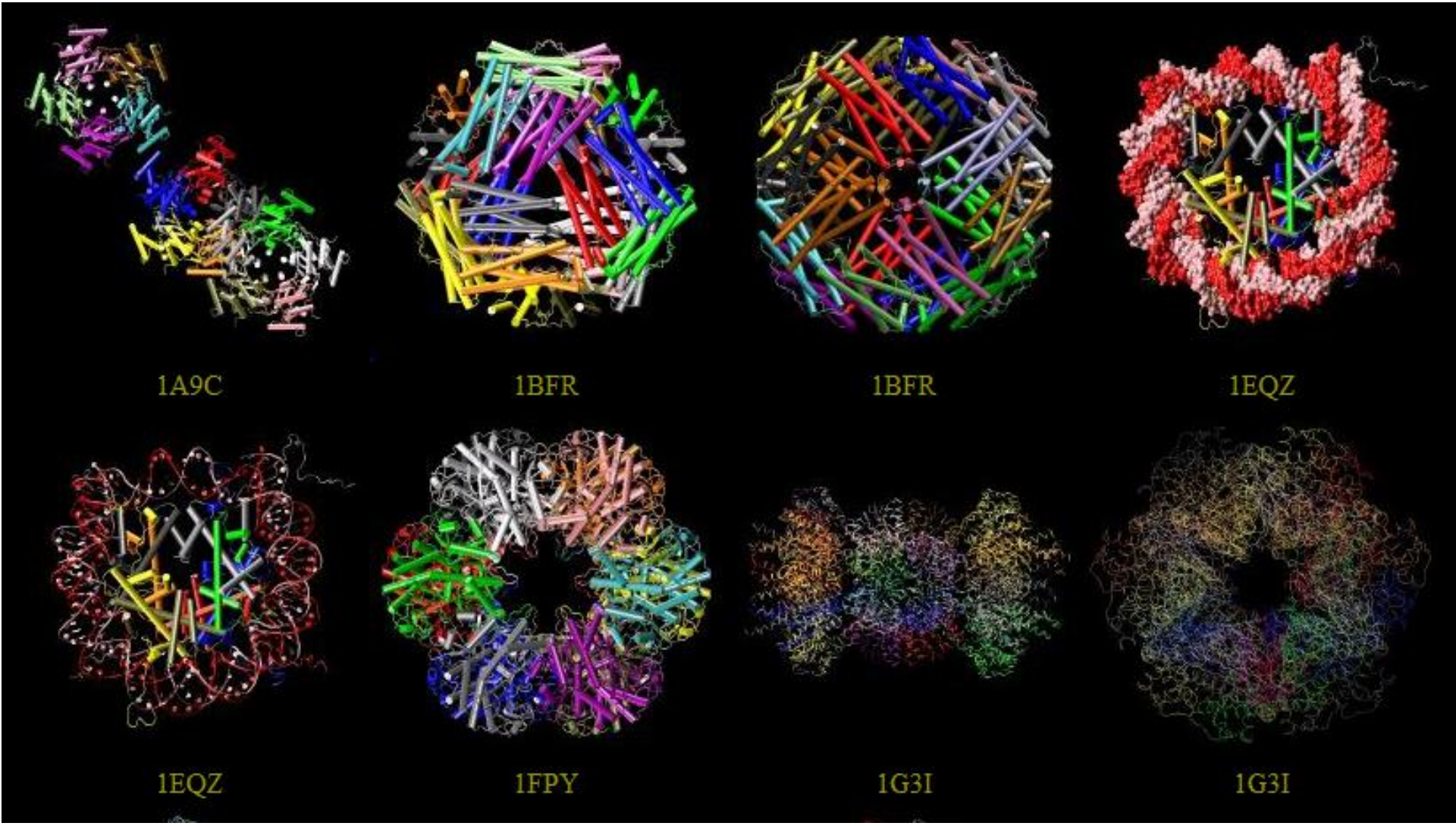
REM	|---Residue---|			|--Structure--|	|-Phi-|	|-Psi-|	|-Area-|	1CRN
....									
ASG	VAL -	15	15	H	AlphaHelix	-69.24	-41.22	93.8	1CRN
ASG	CYS -	16	16	H	AlphaHelix	-56.67	-36.00	18.4	1CRN
ASG	ARG -	17	17	H	AlphaHelix	-77.07	-16.13	94.1	1CRN
ASG	LEU -	18	18	H	AlphaHelix	-53.21	-46.17	143.0	1CRN
ASG	PRO -	19	19	C	Coil	-77.19	-7.60	108.9	1CRN
ASG	GLY -	20	20	C	Coil	106.26	7.31	52.1	1CRN
ASG	THR -	21	21	C	Coil	-52.67	136.34	38.4	1CRN
ASG	PRO -	22	22	C	Coil	-56.98	146.62	81.9	1CRN
ASG	GLU -	23	23	H	AlphaHelix	-56.41	-36.19	68.9	1CRN
ASG	ALA -	24	24	H	AlphaHelix	-63.43	-34.86	61.3	1CRN
ASG	ILE -	25	25	H	AlphaHelix	-74.77	-37.89	98.2	1CRN
ASG	CYS -	26	26	H	AlphaHelix	-64.95	-31.69	0.0	1CRN
ASG	ALA -	27	27	H	AlphaHelix	-62.04	-54.03	11.6	1CRN
ASG	THR -	28	28	H	AlphaHelix	-68.78	-25.49	121.1	1CRN
ASG	TYR -	29	29	H	AlphaHelix	-67.59	-36.30	174.0	1CRN
ASG	THR -	30	30	H	AlphaHelix	-108.96	-18.47	23.4	1CRN
ASG	GLY -	31	31	C	Coil	91.82	-3.07	36.1	1CRN
ASG	CYS -	32	32	C	Coil	-69.52	164.38	4.6	1CRN
ASG	ILE -	33	33	E	Strand	-129.76	157.03	51.0	1CRN
ASG	ILE -	34	34	E	Strand	-111.56	129.59	78.0	1CRN
....									

Comparison between DSSP and STRIDE

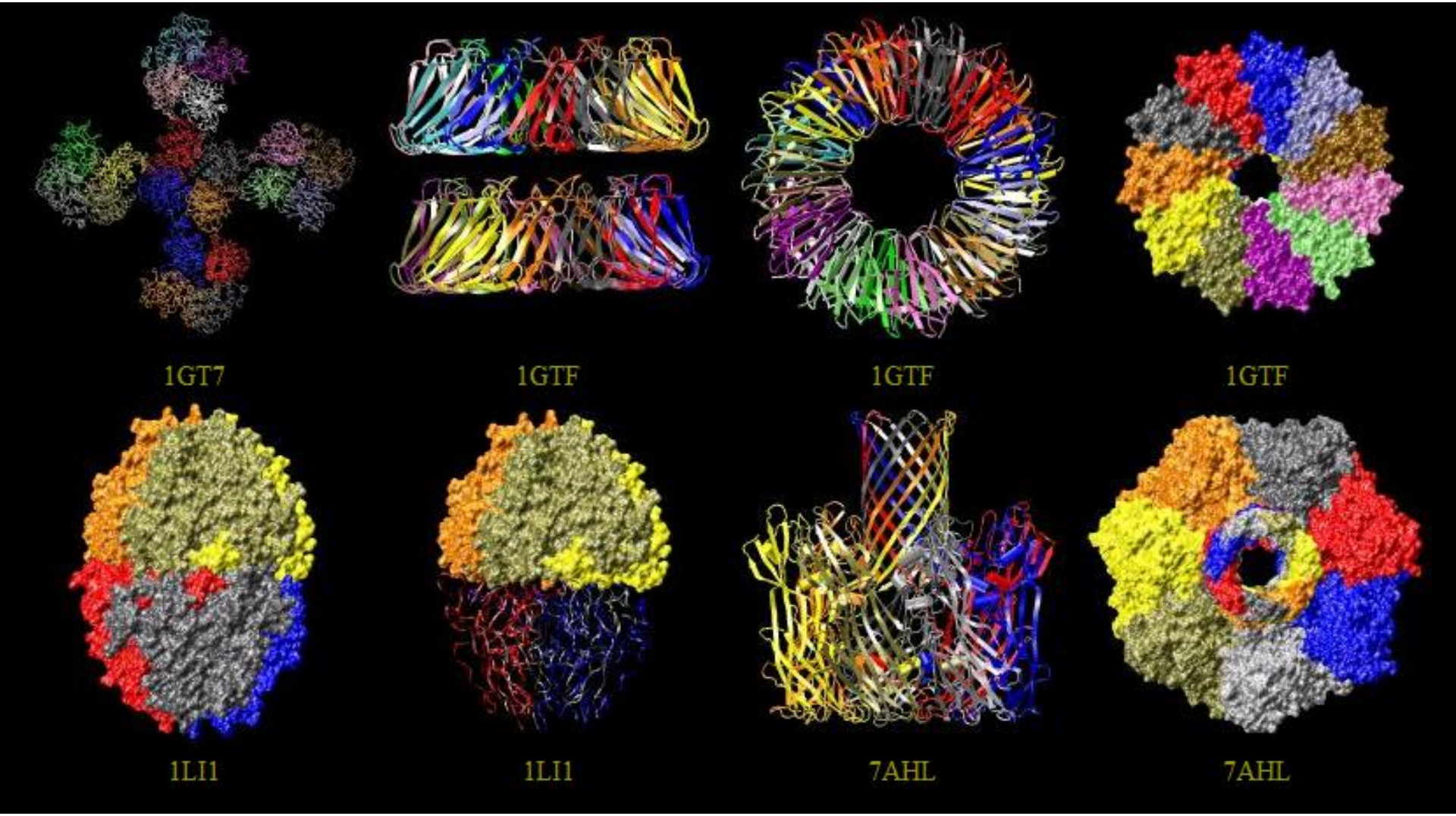
- Size of dataset: 226 protein chains



Visualization of proteins



Visualization of proteins



Protein Data Bank
(http://www.rcsb.org)

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB PROTEIN DATA BANK

An Information Portal to 105499 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 wwPDB EMDataBank NDB Nucleic Acid Database StructuralBiology Knowledgebase

Welcome
Deposit
Search
Visualize
Analyze
Download

A Structural View of Biology

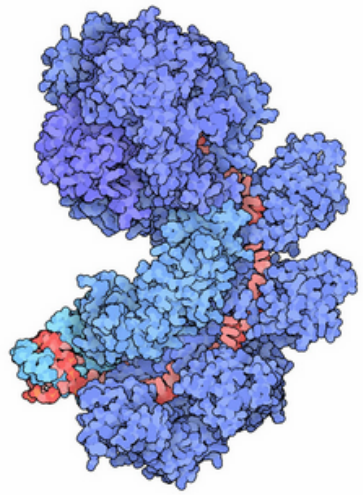
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Structure and Health Focus: Ebola Virus Proteins

January Molecule of the Month

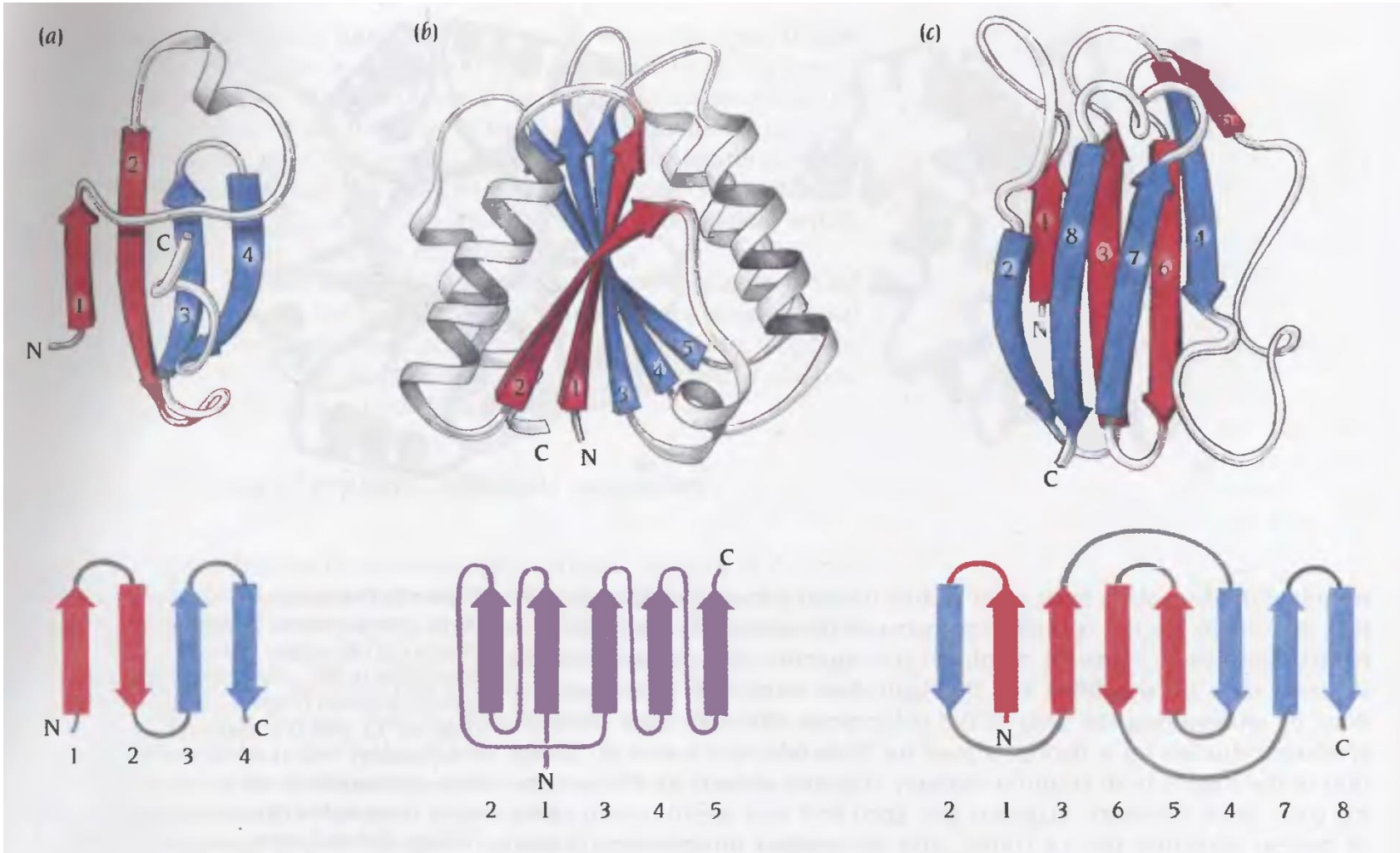


FASTA file format

```
>12AS:A|PDBID|CHAIN|SEQUENCE
MKTAYIAKQRQISFVKSHFSRQLEERLGLIEVQAPILSRVGDGTQDNLSGAEKAVQVKVKALPDAQFEVVHSLA
KWKRQTLGQHDFSAGEGLYTHMKALRPDEDRLSPLHSVYVDQWDWERVMGDGERQFSTLKSTVEAIWAGIKATE
AAVSEEFGLAPFLPDQIHFVHSQELLSRYPDLDAKGRERAIAKDLGAVFLVGIGGKLSDGHRHDVRAPDYDDWS
TPSELGHAGLNGDILVWNPVLEDAFELSSMGIRVDADTLKHQLALTGDEDRLELEWHQALLRGEMPQTIGGGIG
QSRLTMLLLQLPHIGQVQAGVWPAAVRESVPSLL

>12AS:B|PDBID|CHAIN|SEQUENCE
MKTAYIAKQRQISFVKSHFSRQLEERLGLIEVQAPILSRVGDGTQDNLSGAEKAVQVKVKALPDAQFEVVHSLA
KWKRQTLGQHDFSAGEGLYTHMKALRPDEDRLSPLHSVYVDQWDWERVMGDGERQFSTLKSTVEAIWAGIKATE
AAVSEEFGLAPFLPDQIHFVHSQELLSRYPDLDAKGRERAIAKDLGAVFLVGIGGKLSDGHRHDVRAPDYDDWS
TPSELGHAGLNGDILVWNPVLEDAFELSSMGIRVDADTLKHQLALTGDEDRLELEWHQALLRGEMPQTIGGGIG
QSRLTMLLLQLPHIGQVQAGVWPAAVRESVPSLL
```

Topology in Structure



References

- http://kinemage.biochem.duke.edu/teaching/anatax/

- **Introduction to Protein Structure**

  Authors: Carl Branden, John Tooze