Protein structure predictions
Lecture 1-5

Vandana Gupta
Department of Microbiology
Ram Lal Anand College

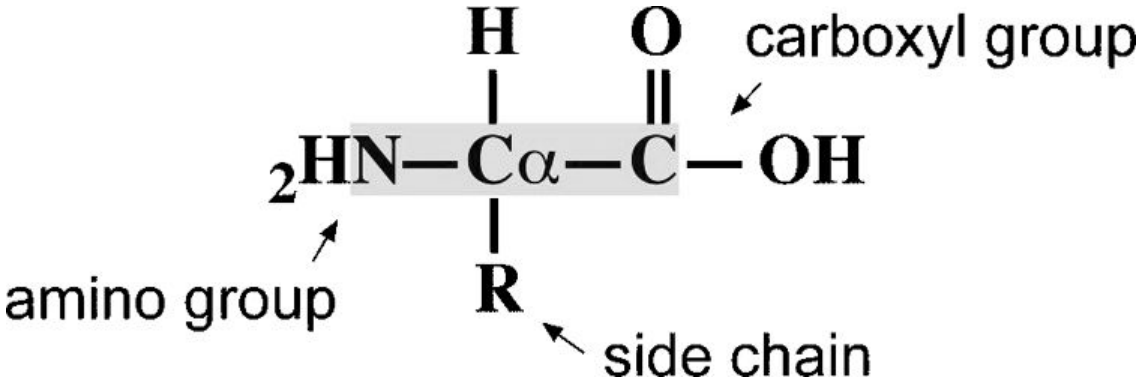
Introduction

Three fundamental properties of proteins in order to play their biological roles:

- Proteins show a variety of structure and function
- All proteins are synthesized at one place, in the ribososme
- Proteins fold spontaneously to active native 3- D states/conformations **(what factors determine the 3d structure of a protein?)**
- though amino acid sequence is specified by nucleotide sequence..... no further role of nucleic acids in folding & 3D structure determination

Structure of proteins

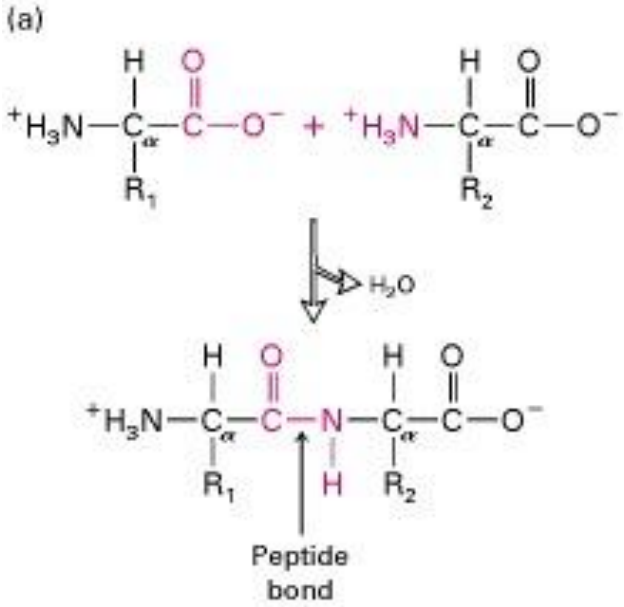
- 20 naturally occurring amino acids
- Differ in R side chains
- Chemical reactivities of the R groups/side chains determine the specific properties of aa



- Amino acids are the monomeric building blocks of proteins.
- The α carbon atom (C_α) of amino acids, which is adjacent to the carboxyl group, is bonded to four different chemical groups: an amino (NH_2) group, a carboxyl (COOH) group, a hydrogen (H) atom, and one variable group, called a *side chain* or *R group*.
- All 20 different amino acids have this same general structure, but their side-chain groups vary in size, shape, charge, hydrophobicity, and reactivity.

bonds

- Nature has evolved a single chemical linkage, the peptide bond, to connect amino acids into a linear, un-branched chain.
- The peptide bond is formed by a condensation reaction between the amino group of one amino acid and the carboxyl group of another.
- **A short chain of amino acids** is **a peptide**; longer peptides are referred to as **polypeptides**.

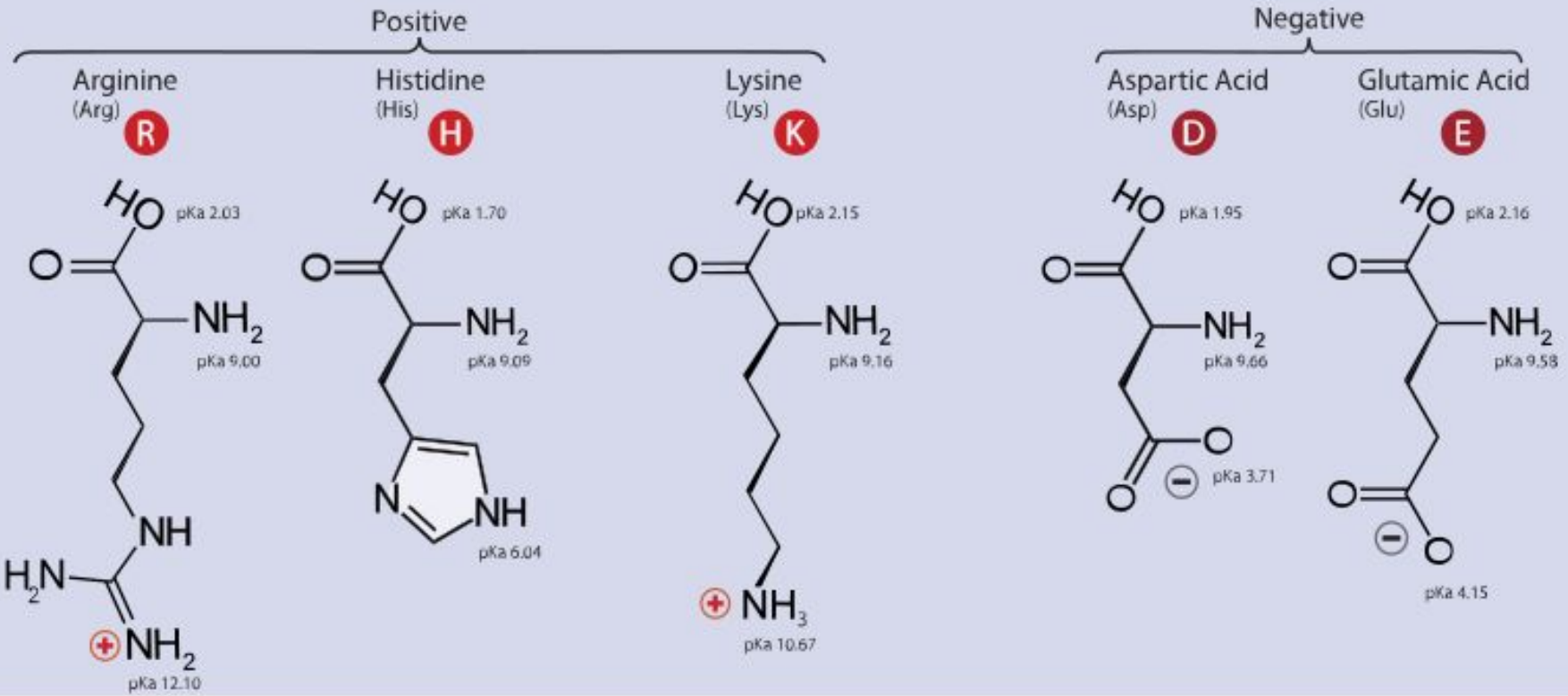


Twenty-One Amino Acids

⊕ Positive ⊖ Nega

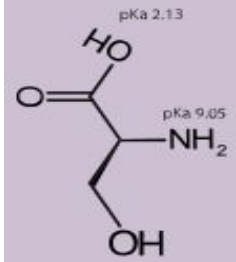
• Side chain charge at physiological pH

A. *Amino Acids with Electrically Charged Side Chains*

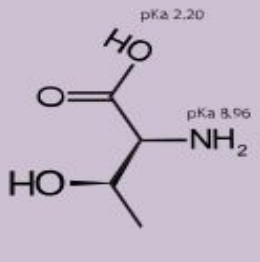


B. Amino Acids with Polar Uncharged Side Chains

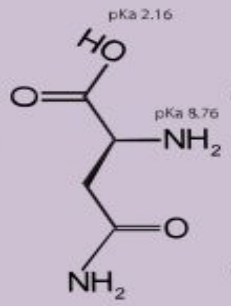
Serine (Ser) S



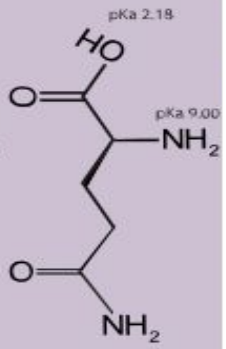
Threonine (Thr) T



Asparagine (Asn) N

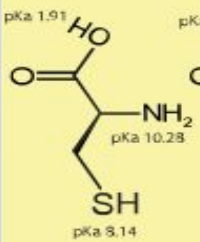


Glutamine (Gln) Q

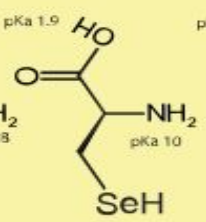


C. Special Cases

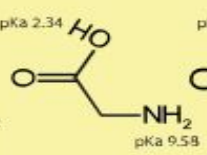
Cysteine (Cys) C



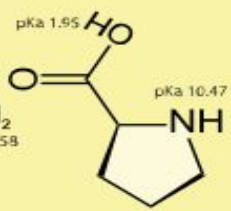
Selenocysteine (Sec) U



Glycine (Gly) G

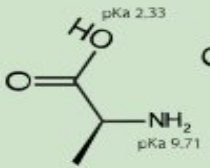


Proline (Pro) P

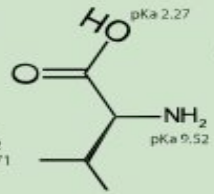


D. Amino Acids with Hydrophobic Side Chain

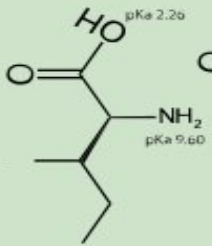
Alanine (Ala) A



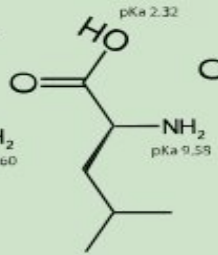
Valine (Val) V



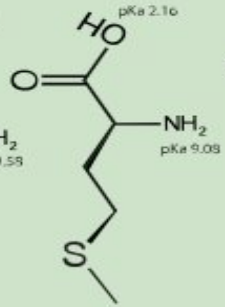
Isoleucine (Ile) I



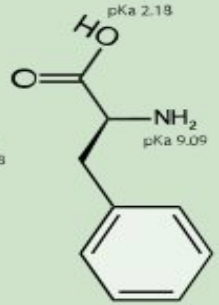
Leucine (Leu) L



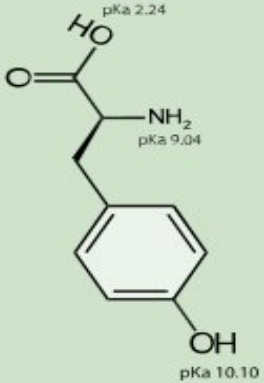
Methionine (Met) M



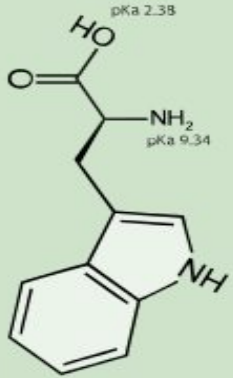
Phenylalanine (Phe) F



Tyrosine (Tyr) Y



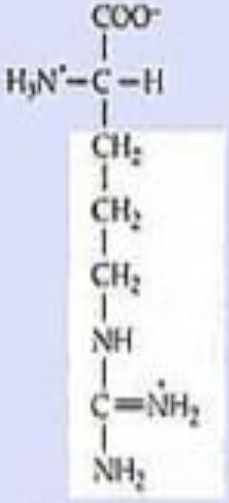
Tryptophan (Trp) W



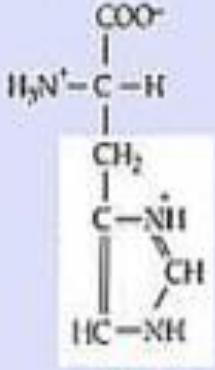
A. *Amino acids with electrically charged side chains*

Positive

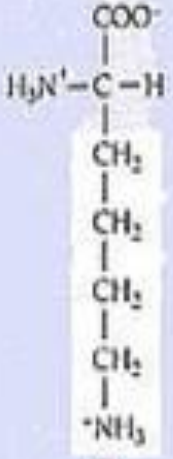
Arginine (Arg)



Histidine (His)

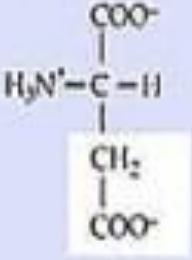


Lysine (Lys)

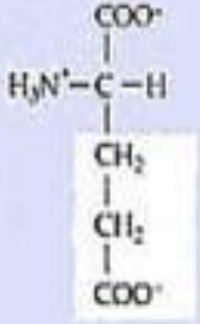


Negative

Aspartic acid (Asp)

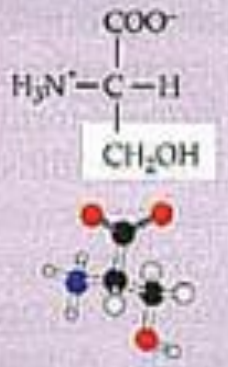


Glutamic acid (Glu)

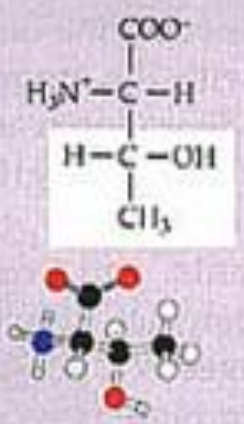


B. *Amino acids with polar but uncharged side chains*

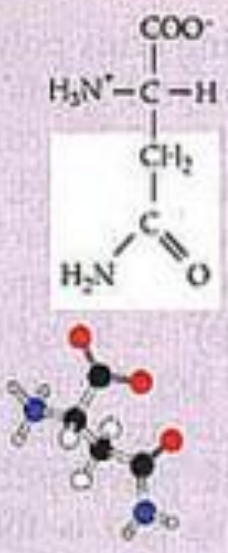
Serine (Ser)



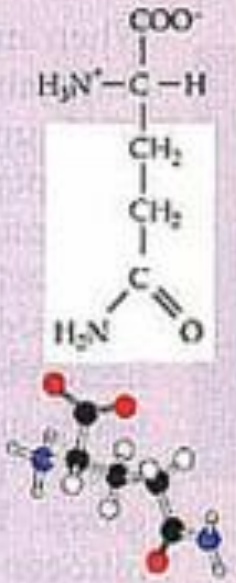
Threonine (Thr)



Asparagine (Asn)

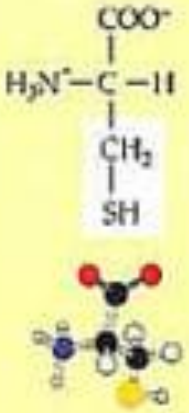


Glutamine (Gln)

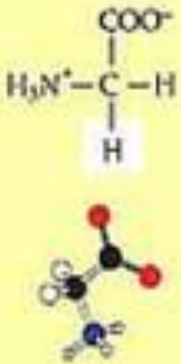


C. *Special cases*

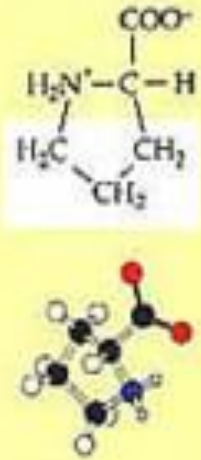
Cysteine (Cys)



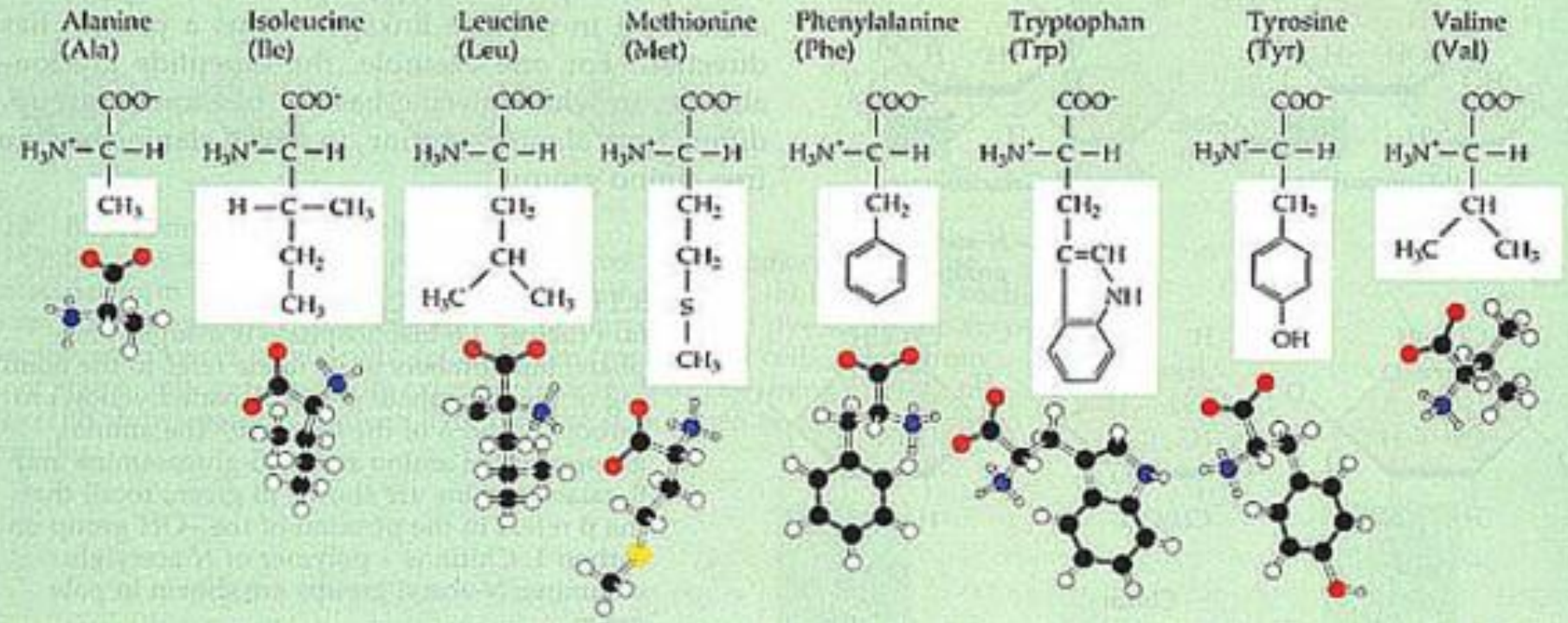
Glycine (Gly)



Proline (Pro)



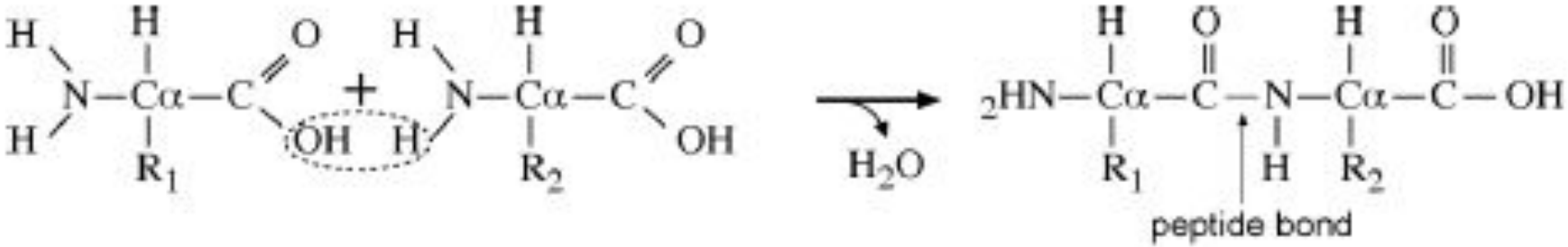
D. *Amino acids with hydrophobic side chains*



20 standard amino acids grouped by their common side –chain features

Amino Acid Group	Amino Acid Name	Three- and One-Letter Code	Main Functional Features
Small and nonpolar	Glycine Alanine Proline	Gly, G Ala, A Pro, P	Nonreactive in chemical reactions; Pro and Gly disrupt regular secondary structures
Small and polar	Cysteine Serine Threonine	Cys, C Ser, S Thr, T	Serving as posttranslational modification sites and participating in active sites of enzymes or binding metal
Large and polar	Glutamine Asparagine	Gln, Q Asn, N	Participating in hydrogen bonding or in enzyme active sites
Large and polar (basic)	Arginine Lysine Histidine	Arg, R Lys, K His, H	Found in the surface of globular proteins providing salt bridges; His participates in enzyme catalysis or metal binding
Large and polar (acidic)	Glutamate Aspartate	Glu, E Asp, D	Found in the surface of globular proteins providing salt bridges
Large and nonpolar (aliphatic)	Isoleucine Leucine Methionine Valine	Ile, I Leu, L Met, M Val, V	Nonreactive in chemical reactions; participating in hydrophobic interactions
Large and nonpolar (aromatic)	Phenylalanine Tyrosine Tryptophan	Phe, F Tyr, Y Trp, W	Providing sites for aromatic packing interactions; Tyr and Trp are weakly polar and can serve as sites for phosphorylation and hydrogen bonding

Dipeptide formation



- The residues in a peptide or polypeptide are numbered beginning with the residue containing the amino group, referred to as the *N-terminus*, and ending with the residue containing the carboxyl group, known as the *C-terminus*.

Polypeptide

- A polymer of more than 50 residues - ***polypeptide***.
- A polypeptide- having a well-defined three-dimensional arrangement-- called a protein
- Polymer with < fifty residues is usually called a peptide
- **The actual sequence of amino acid residues in a polypeptide determines its ultimate structure and function.**
- The atoms involved in forming the peptide bond are referred to as the ***backbone/mainchain atoms.*** *They are the nitrogen of the amino group, the α carbon to which the side chain* is attached and carbon of the carboxyl group

How does the amino acid sequence encode the 3D protein structure??

- Any folding of the mainchain places different residues in contact.
- Interactions of the side-chains and main-chain, with one another and with the solvent, and the restrictions placed on sidechain mobility, determine the relative stabilities of different conformations.

- Governed by 2nd law of thermodynamics - systems at constant temperature & pressure find an equilibrium state that is a compromise between comfort (low enthalpy, H), and freedom (high entropy, S), to give a minimum Gibbs free energy

  G= H-TS, in which T is the absolute temperature.

Unusual amino acids

Of particular interest within the twenty amino acids are glycine and proline.

- Glycine, the smallest amino acid, has a hydrogen atom as the R group. It can therefore adopt more flexible conformations that are not possible for other amino acids.
- Proline is on the other extreme of flexibility. Its side chain forms a bond with its own backbone amino group, causing it to be cyclic and making it very rigid, unable to occupy many of the main chain conformations adopted by other amino acids
- In addition certain amino acids are subjected to modification after a protein is translated in a cell (Post translational modifications)

Side chains (LESK)

Side chains offer the physiochemical versatility required to generate all the different folding patterns. The side chains of the diff aa vary in the following ways:

1. Size: Glycine smallest with H atom, phenylalanine largest with benzene ring
2. Electric charge: charged residues of opposite sign can form attractive pair wise interactions called salt bridges

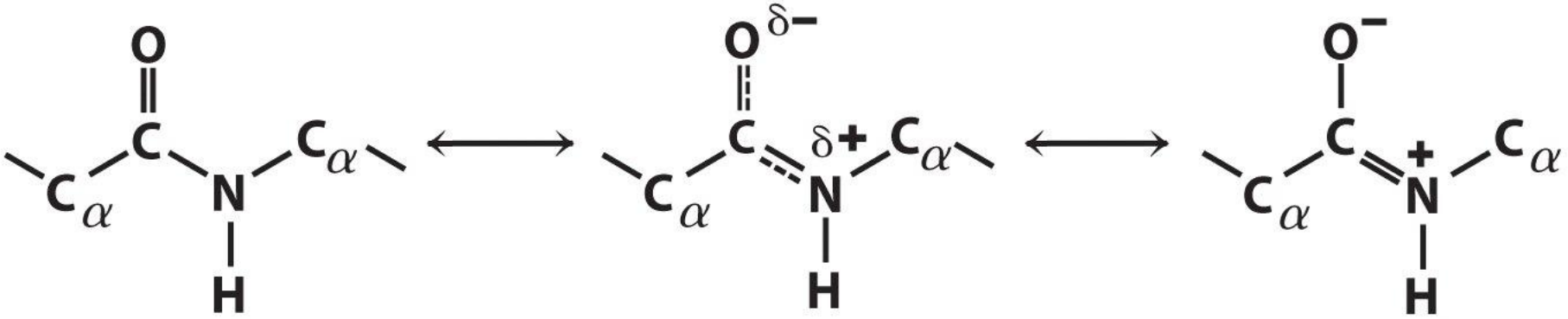
1. Polarity: polar sidechains can form hydrogen bonds to other polar sidechains or to the water.
2. Some side chains have chemical groups related to ordinary hydrocarbons like methane or benzene. They are hydrophobic as they have thermodynamically unfavourable interactions with water. Congregation of hydrophobic residues in protein interiors, was predicted by WJ Kauzmann even before the first protein structure was determined

Glu to Asp or Leu to Ile are conservative changes, may cause minor change in proteins structure and function.

But a mutation in the interior from Leu to Glu would severely damage the structure. It has applications in SNPs by wide scale genetic screening to look for mutants that by changing proteins, cause disease.

- The angle of rotation about the bond is referred to as the *dihedral angle* (*also* called the *tortional angle*).
- Atoms linked to the peptide bond can be moved to a certain extent by the rotation of two bonds flanking the peptide bond.
- This is measured by two dihedral angles. angle along the N–C*α bond, is* phi (*φ*); *and the angle along the Cα–C bond, is* psi (*ψ*). *Various combinations of φ and ψ angles allow the proteins to fold in many* different ways.

The Peptide Bond

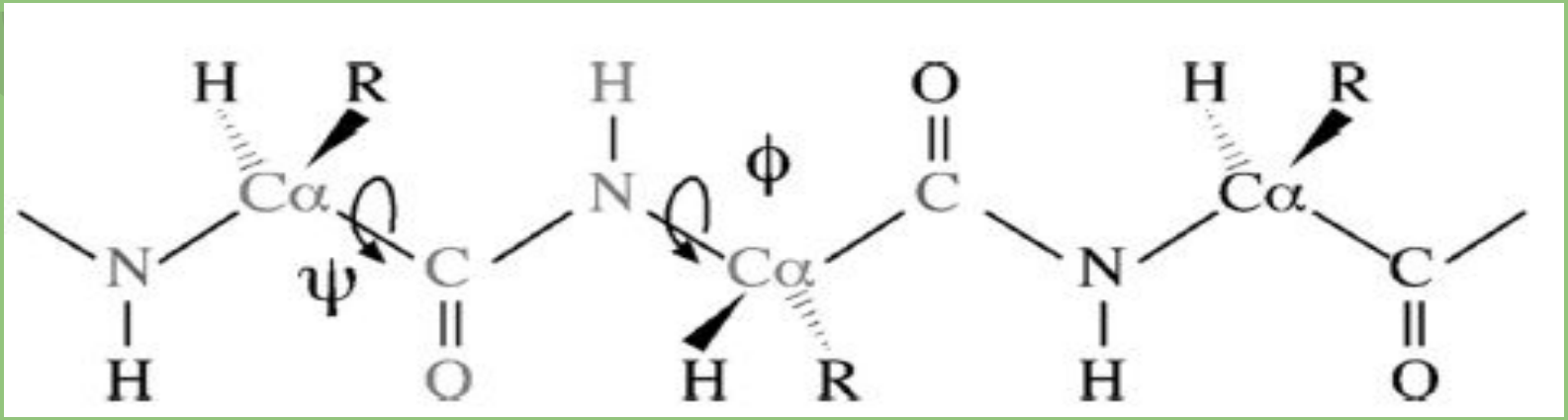


The carbonyl oxygen has a partial negative charge and amide nitrogen a partial positive charge, causing the setting up of a small electric dipole

Dihedral angle

- A **peptide bond** is a **partial double bond :**shared electrons between O=C–N atoms. Rigid double bond structure forces atoms associated with the peptide bond to lie in the same plane, called the ***peptide plane***.
- *Because of the planar* nature of the peptide bond and the size of the R groups, the rotational freedom of the two bonded pairs of atoms around the peptide bond are considerably restricted .
-

Representation of φ and ψ angles



The *φ angle is the rotation about the N–Cα bond, which is measured* by the angle between a virtual plane formed by the C–N–Cα *and the virtual plane by N–Cα–C.*

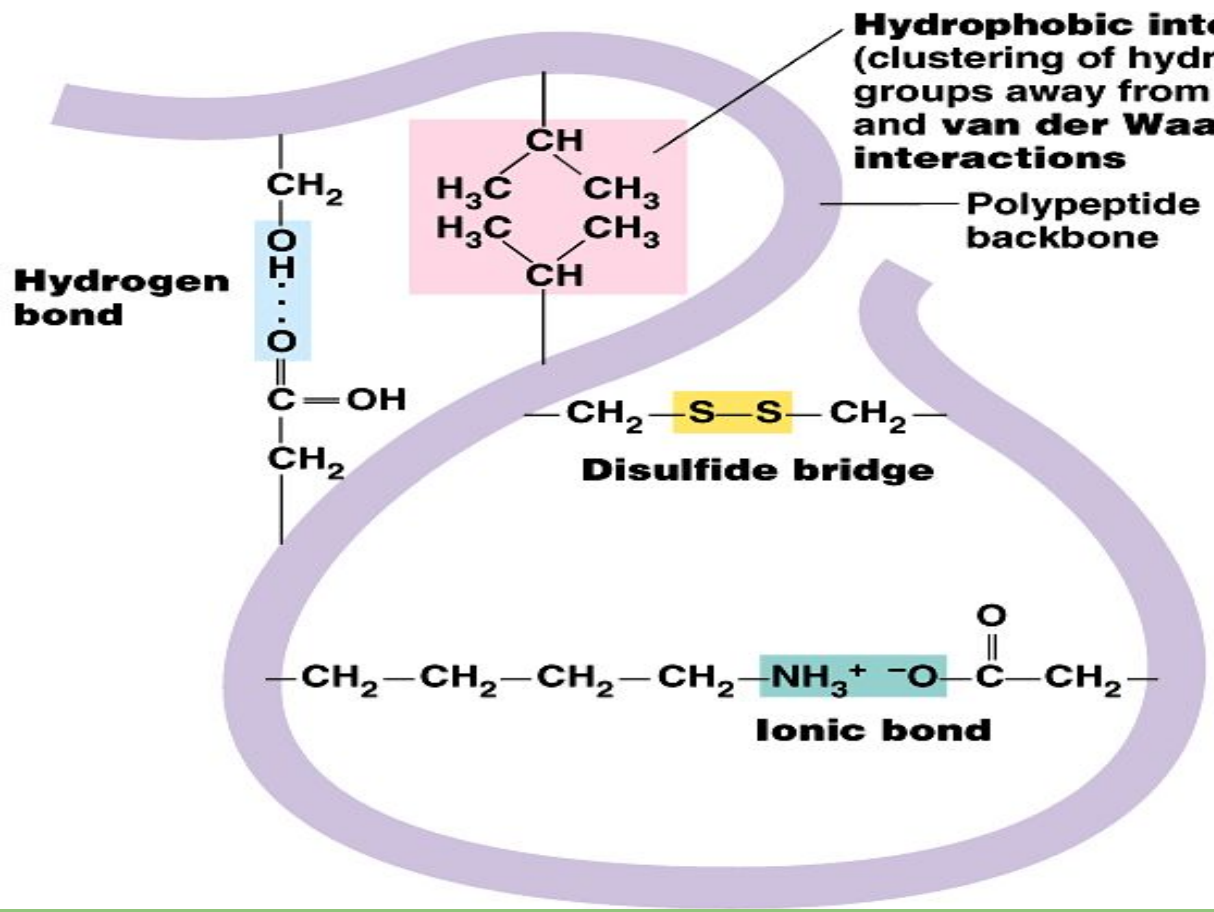
The *ψ angle is the rotation about the Cα–C bond, which is measured by the angle between a* virtual plane formed by the N–Cα*–C and the virtual plane by Cα–C–N*

Different types of residues make different types of interactions, including:

- Hydrogen bonds: The hydrogen bond is an interaction between two polar atoms (oxygen or nitrogen; occasionally sulphur) mediated by a hydrogen atom. Hydrogen bonds between C=O and H–N groups stabilize the structure of proteins.

Types:

- Hydrophobic interactions: Hydrophobic residues have sidechains that are primarily hydrocarbon in nature. They have thermodynamically unfavourable interactions with water. It is energetically favourable to bury hydrophobic side chains in the interior of a protein, where they are not exposed to the solvent
- Van der waal interactions :
- Disulphide bridges: cysteine residues in proteins, with sidechain –CH2SH, can form disulphide bonds: –CH2S– SCH2–. Disulphide bonds contribute to the stability of native states
- Salt bridges:



- Tertiary structure, the next-higher level of structure, refers to the overall conformation of a polypeptide chain, that is, the three-dimensional arrangement of all the amino acids residues

Level of hierarchy in protein structure

- Primary sequence- The amino acid sequence of a polypeptide, listed from N-terminus to C-terminus.
- Secondary structure- (α-helix, β-sheet) Recurring structural feature of proteins stabilized exclusively by hydrogen bonds between peptide bond elements
- Supersecondary structure- Recurring structural feature of proteins composed of two or more secondary structural elements.
- Domain- A segment of protein structure that is autonomously stable.
- Tertiary structure- Three-dimensional structure formed by assembly of secondary structures /A stable, independent protein encoded by a single gene.
- Quaternary structure- A complex structure composed of two or more tertiary structure subunits.

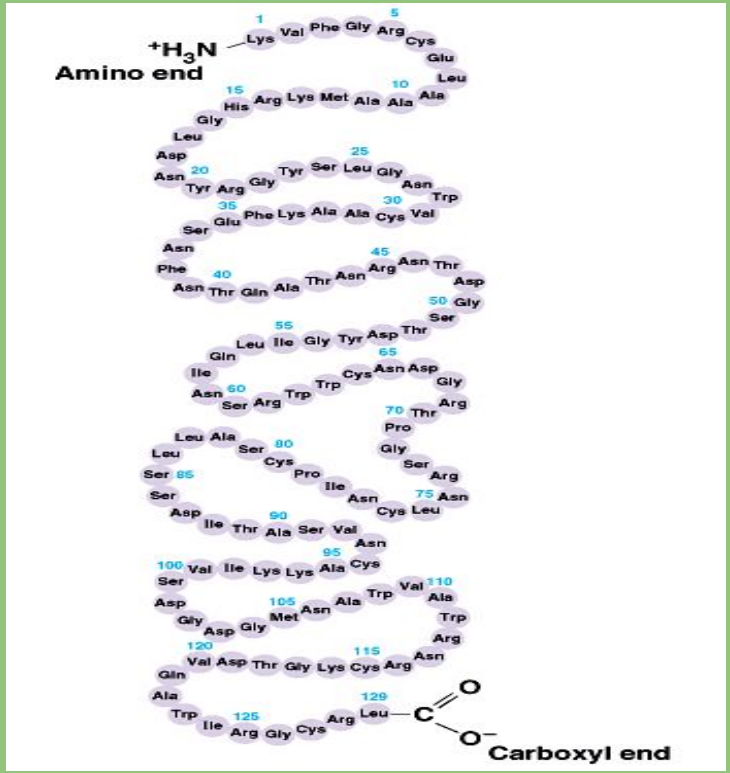
BOX 10.3

Protein structure – basic vocabulary

Polypeptide chain	Linear polymer of amino acids.
Mainchain	Atoms of the repetitive concatenation of peptide groups . . . N–Cα–(C=O)N–Cα–(C=O) . . .
Sidechains	Sets of atoms attached to each Cα of the mainchain. Most sidechains in proteins are chosen from a canonical set of 20.
Primary structure	The chemical bonds linking atoms in the amino acid sequence in a protein.
Hydrogen bond	A weak interaction between two neighbouring polar atoms, mediated by a hydrogen atom.
Secondary structure	Substructures common to many proteins, compatible with mainchain conformations, free of interatomic collisions, and stabilized by hydrogen bonds between mainchain atoms. Secondary structures are compatible with all amino acids, except that a proline necessarily disrupts the hydrogen-bonding pattern.
α-Helix	Type of secondary structure in which the chain winds into a helix, with hydrogen bonds between residues separated by four positions in the sequence.
β-Sheet	Another type of secondary structure, in which sections of mainchain interact by lateral hydrogen bonding.
Folding pattern	Layout of the chain as a curve through space.
Tertiary structure	The spatial assembly of the helices and sheets, and the pattern of interactions between them. (Folding pattern and tertiary structure are nearly synonymous terms.)
Quaternary structure	The assembly of multisubunit proteins from two or more monomers.
Native state	The biologically active form of a protein, which is compact and low energy. Under suitable conditions, proteins form native states spontaneously.
Denaturant	A chemical that tends to disrupt the native state of a protein; for instance, urea.
Denatured state	Non-compact, structurally heterogeneous state formed by proteins under conditions of high temperature, or high concentrations of denaturant.
Post-translational modification	Chemical change in a protein after its creation by the normal protein-synthesizing machinery.
Disulphide bridge	Sulphur–sulphur bond between two cysteine sidechains. A simple example of a post-translational modification.

Primary structure

- The primary structure of a protein is the linear arrangement, or *sequence*, of amino acid residues that constitute the polypeptide chain.



Secondary structure types

- Secondary structure refers to the localized organization of parts of a polypeptide chain, which can assume several different spatial arrangements such as:.
  - Alpha helix
  - Beta strands
  - Loops
  - Reverse turns
  - β-hairpins
  - Coiled coil

Protein stability and conformation

- A conformation that simultaneously solves all the following problems:

1. All residues must have stereo- chemically allowed conformations in both mainchain and sidechains. Stearic collisions would raise the energy of the conformations and make it unstable

2. Buried polar atoms must be H-bonded to other buried polar atoms. If a few H –bonds are missed then protein will prefer denatured state to allow these polar atoms to form H bonds with the solvent

1. Enough hydrophobic surface must be buried, and the interior must be sufficiently densely packed, to provide thermodynamic stability

For most proteins there is a unique solution of these problems called the “native state”

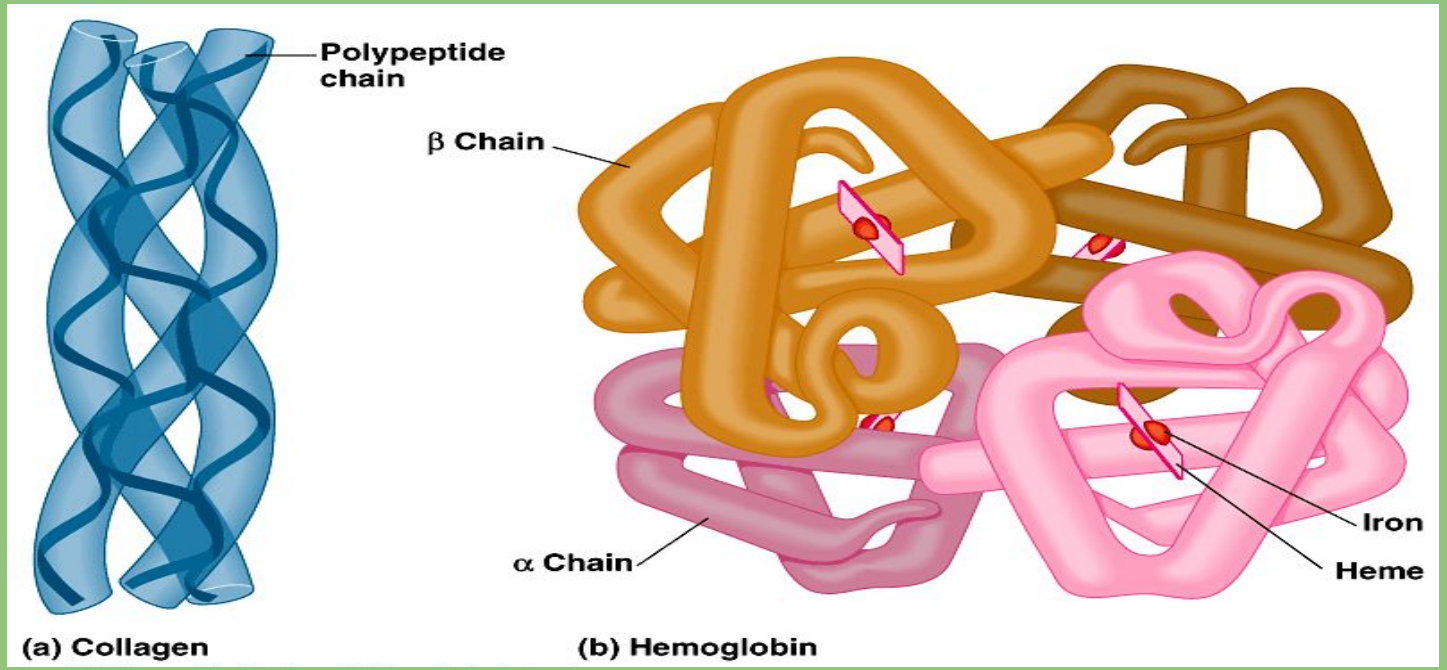
This one conformation of a protein- *the native state*- has substantially greater stability than others.

Tertiary Structure

- While backbone interactions define most of the secondary structure interactions, it is the side-chains that define the tertiary interactions eg: Myoglobin (Kendrew 1958) and hemoglobin (Perutz 1960) gave us the proven experimental insights into tertiary structure as secondary structures interacting by a variety of mechanisms

Quaternary structure is the overall protein structure resulting from combinations of polypeptide subunits

Quaternary structure describes the number and relative positions of the subunits in a multimeric protein

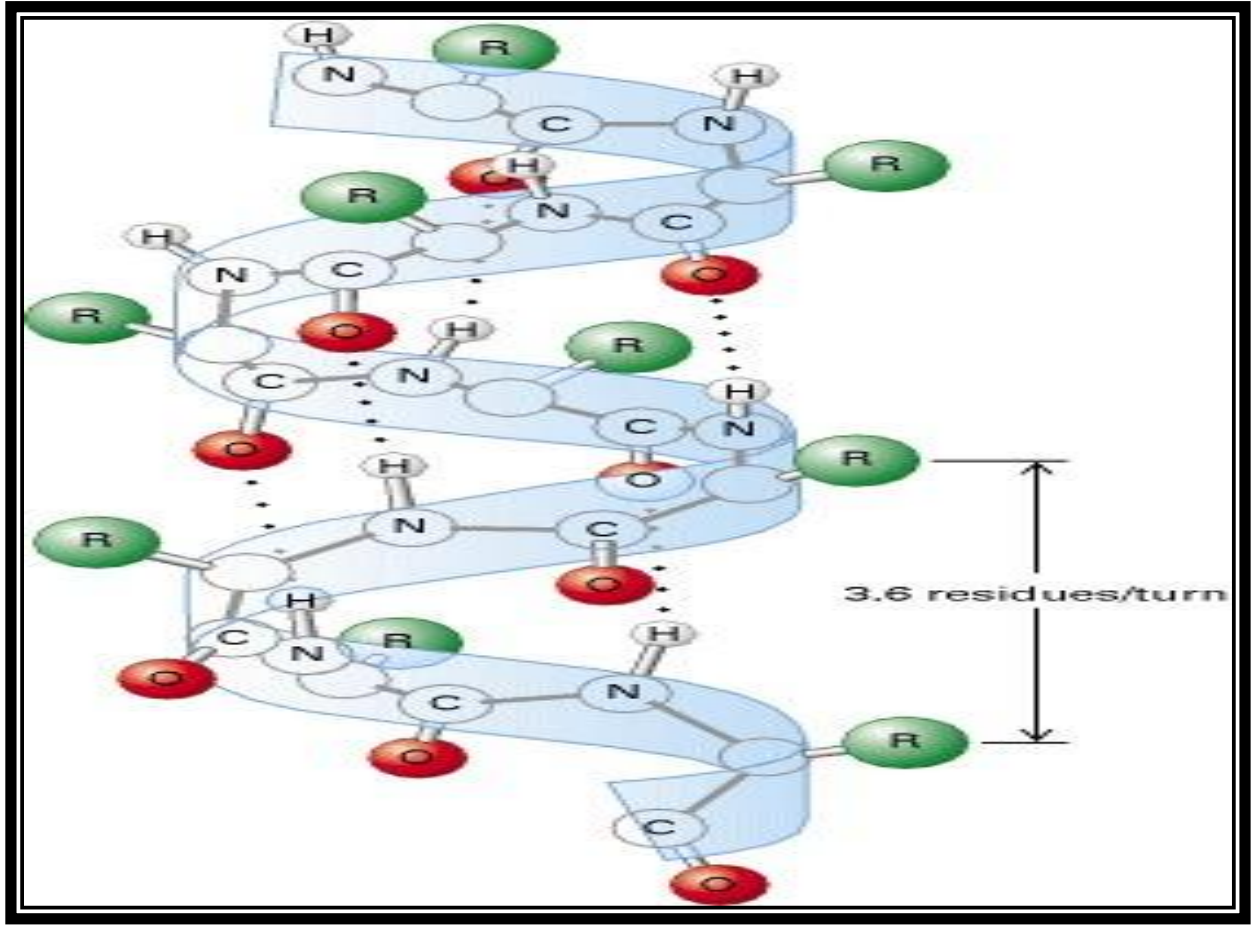


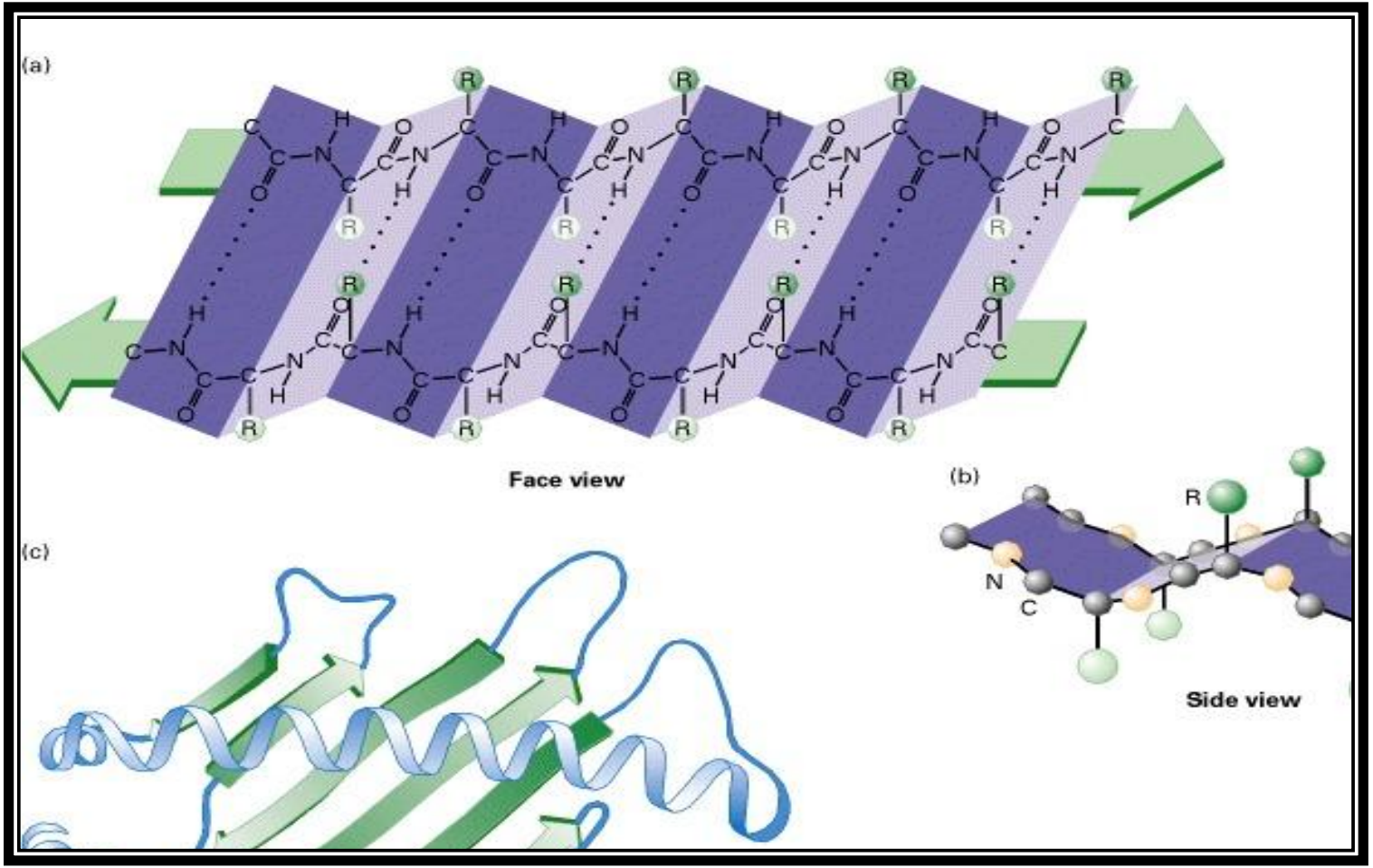
The α Helix and β Sheet

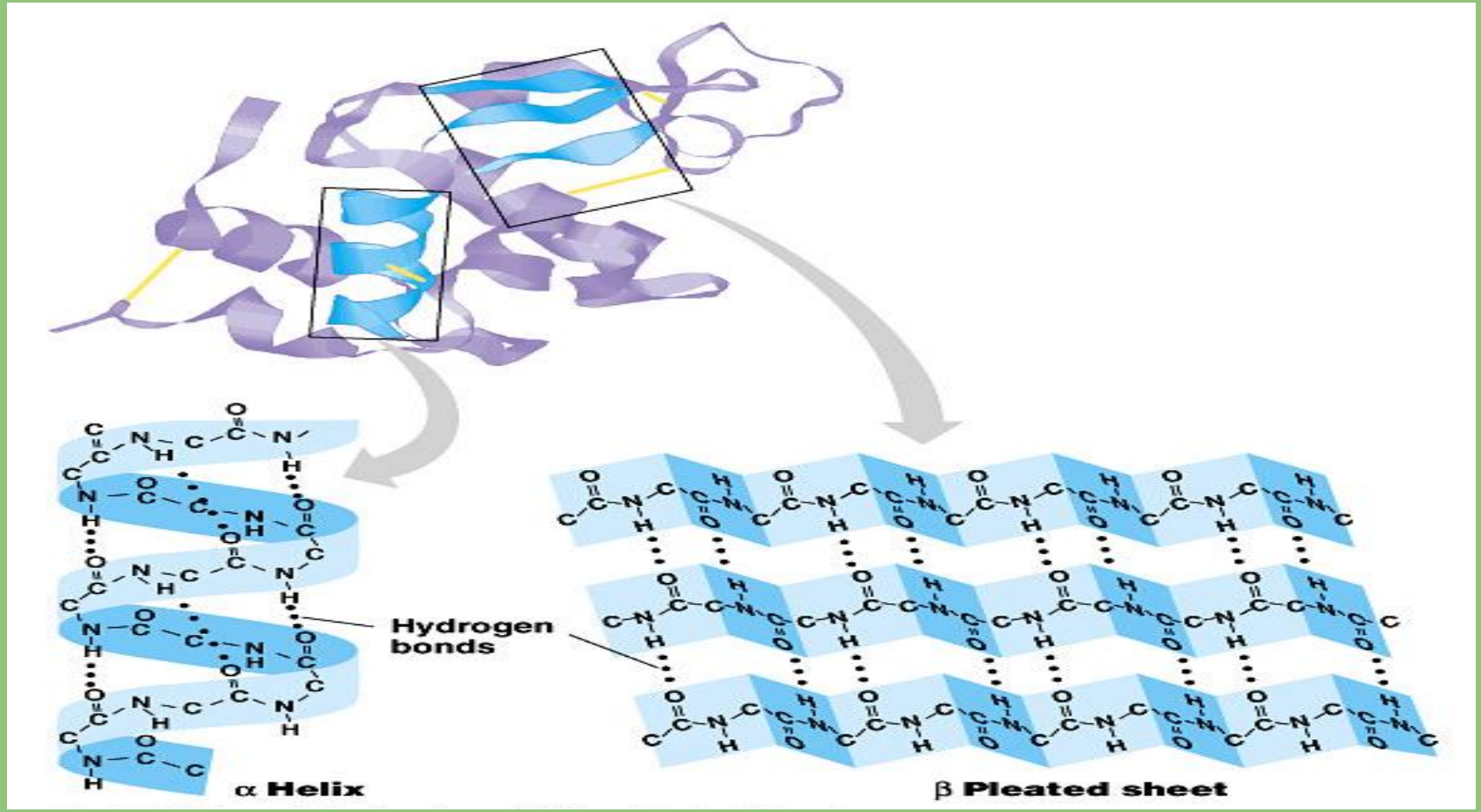
- Polypeptide segments can assume a regular spiral, or helical, conformation, called the α *helix*.

- Another regular secondary structure, the β sheet, consists of laterally packed β strands.

- Each β strand is a short (5–8-residue), nearly fully extended polypeptide chain.


3.6 residues/turn

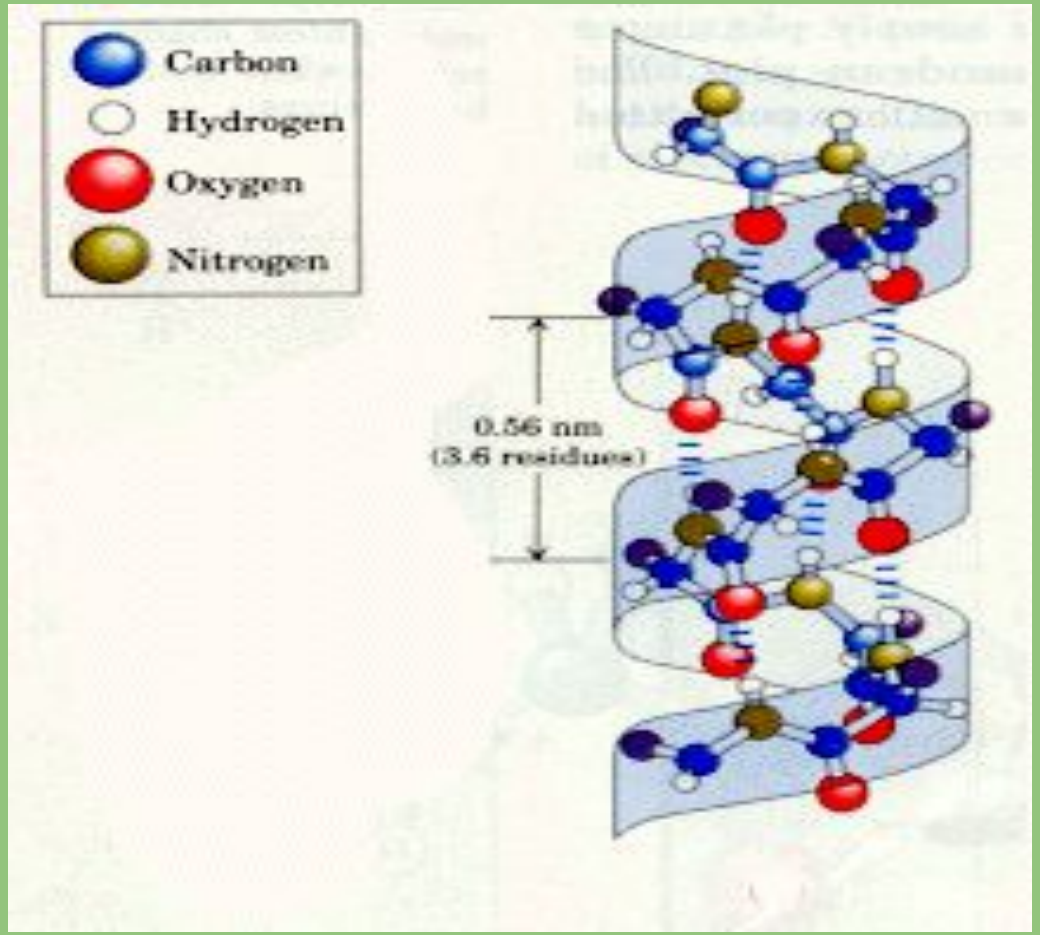

(a)
Face view
(b)
R
N
C
Side view
(c)


Hydrogen bonds
α Helix
β Pleated sheet

Characteristics of alpha helices

- Alpha helices : common structure: characterized by single spiral chain of aa, stabilized by hydrogen bonds
- The structure repeats itself every **5.4 A °** (length of each turn) along the helix axis
- **3.6** aa residues per turn i.e. A 36 aa long helix will have 10 turns with H-bonds between C'=O of residue N and NH of residue n+4.

- hence all NH and C’O groups are joined with hydrogen bonds except the first NH groups & the last C’O groups at the end of the alpha helix. The ends of the helices are polar
- All the aa have -ve ψ and φ angles, typical values between -60 ° and -50 ° respectively
- The peptide planes are roughly parallel with the helix axis i.e. All C=O groups point in the same direction and all N-H groups point in the other way. Side chains point outwards from the helix axis and are generally oriented towards its amino-terminal end.



How Long is an 80 amino acid alpha helix?

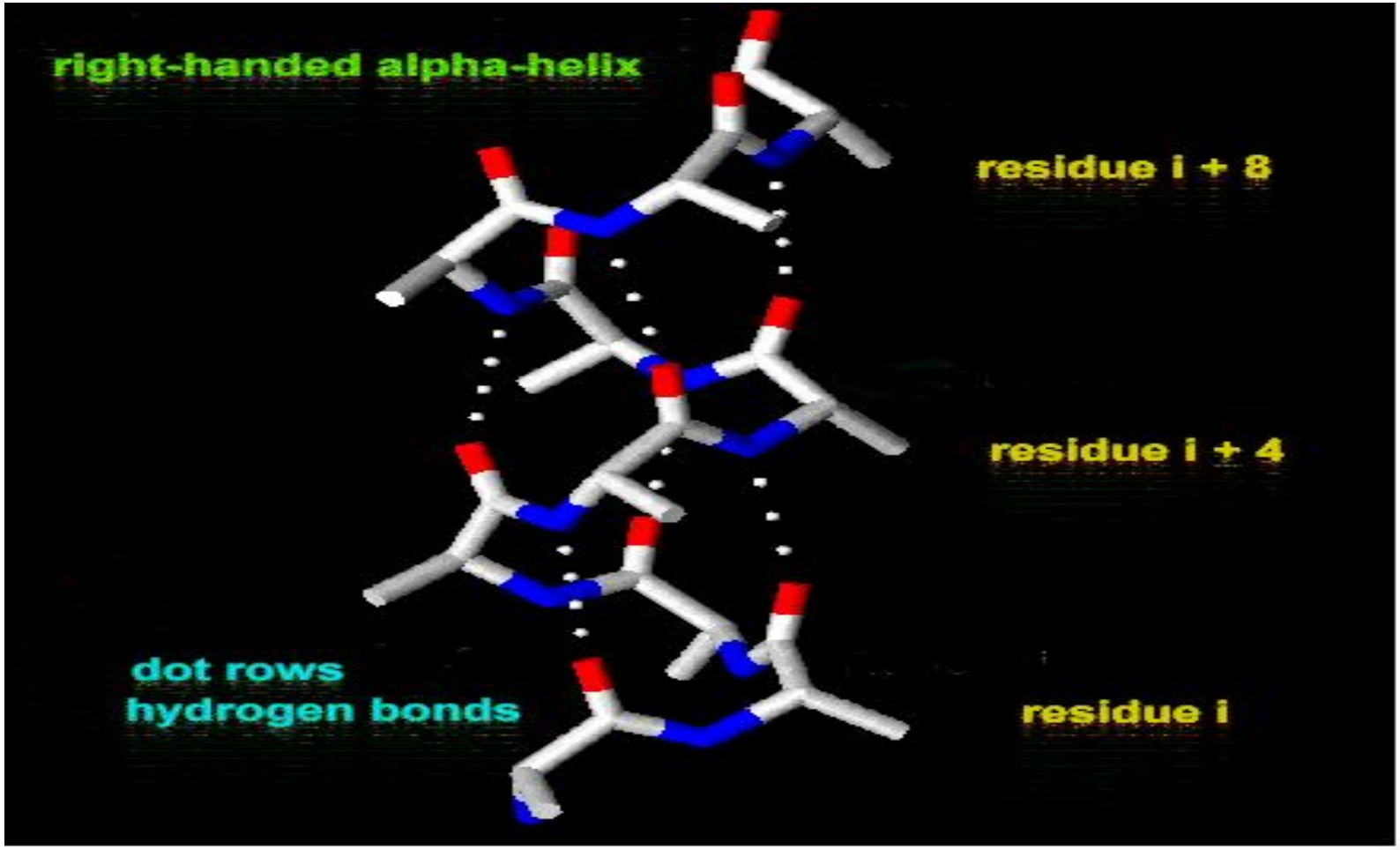
Answer: One turn: 3.6 aa, 5.4 Å long

80 aa / (3.6 aa/turn) = 22.2 turns

22.2 turns x 5.4 Å/turn = 120 Å long

Alpha- helix ...more facts

- Alanine, glutamine, leucine & methionine are good α- helix formers
- While proline, glycine, tyrosine & serine are poor helix formers
- Most common location of a α- helix in a protein structure is outside protein, with one side of the helix facing the solvent and the other side facing the hydrophobic interior of the protein

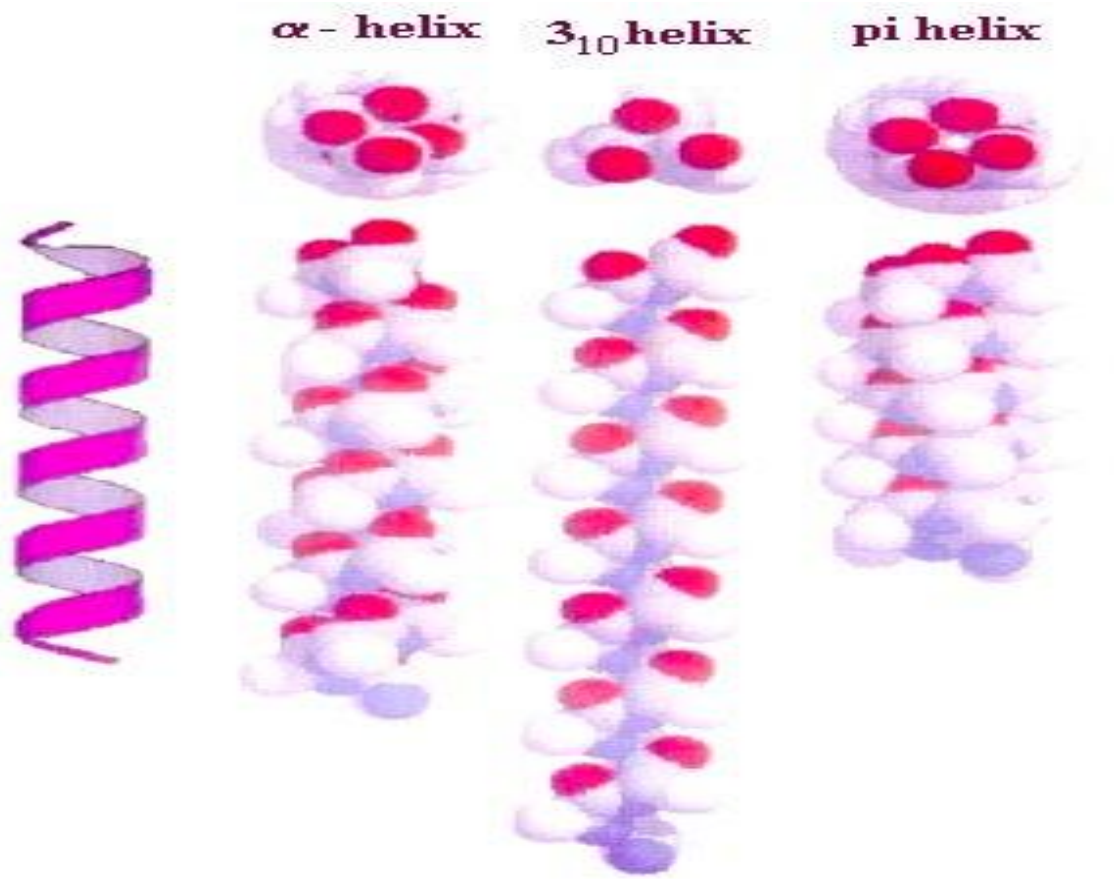

right-handed alpha-helix
residue i + 8
residue i + 4
dot rows
hydrogen bonds
residue i

Types of alpha helices

- Can be left handed or right handed helix based on the coiling of helix in 2 directions.
- **Right handed helix are more commonly observed in experimentally determined structures as they produce less clashes**
- Apart from left handed and right handed helix, several other types of alpha helices are found in a protein like :**3_{10} helices, Pi helix (π)**

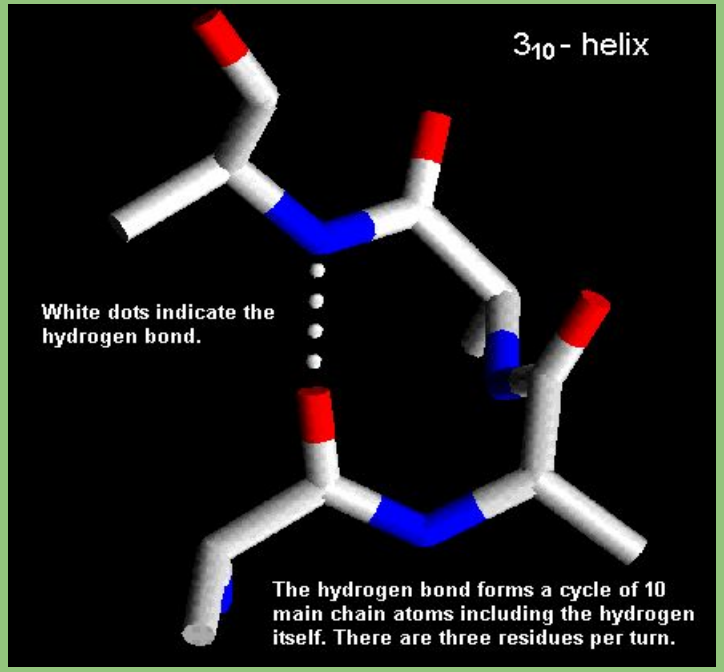
Ψ, φ and ω values for different types of helices

Patterns	Phi (degree)	Psi (degree)	H-bond
Right handed Alpha helix	-57.8	-47.0	i+4
Pi Helix	-57.1	-69.7	i+5
3_{10} helix	-74.0	-4.0	i+3
ω is 180° in all cases			


α - helix
3_{10} helix
pi helix

3_{10} –helix (rare helix)

- 3_{10}- helices form a distinct class of helix but they are always short and frequently occur at the termini of regular α- helices.
- 3 residues per turn and 10 atoms enclosed in a ring formed by each hydrogen bond



- There are main chain hydrogen bonds between residues separated by three residues along the chain (i.e. Hydrogen bonds between residue i and i +3)
- Hence, elongated and more narrower
- The dipoles of 3_{10}- helix are not well aligned as those of the α-helix i.e. **It is a less stable structure** and side chain packing is less favourable.

Pi –helix (rare helix)

- Less favorable geometry
- Pi- helix is a very rare type of helix
- Stabilized by hydrogen bonds of the i residue with the i+5 residue
- Squat and constrained

Sheets and helices

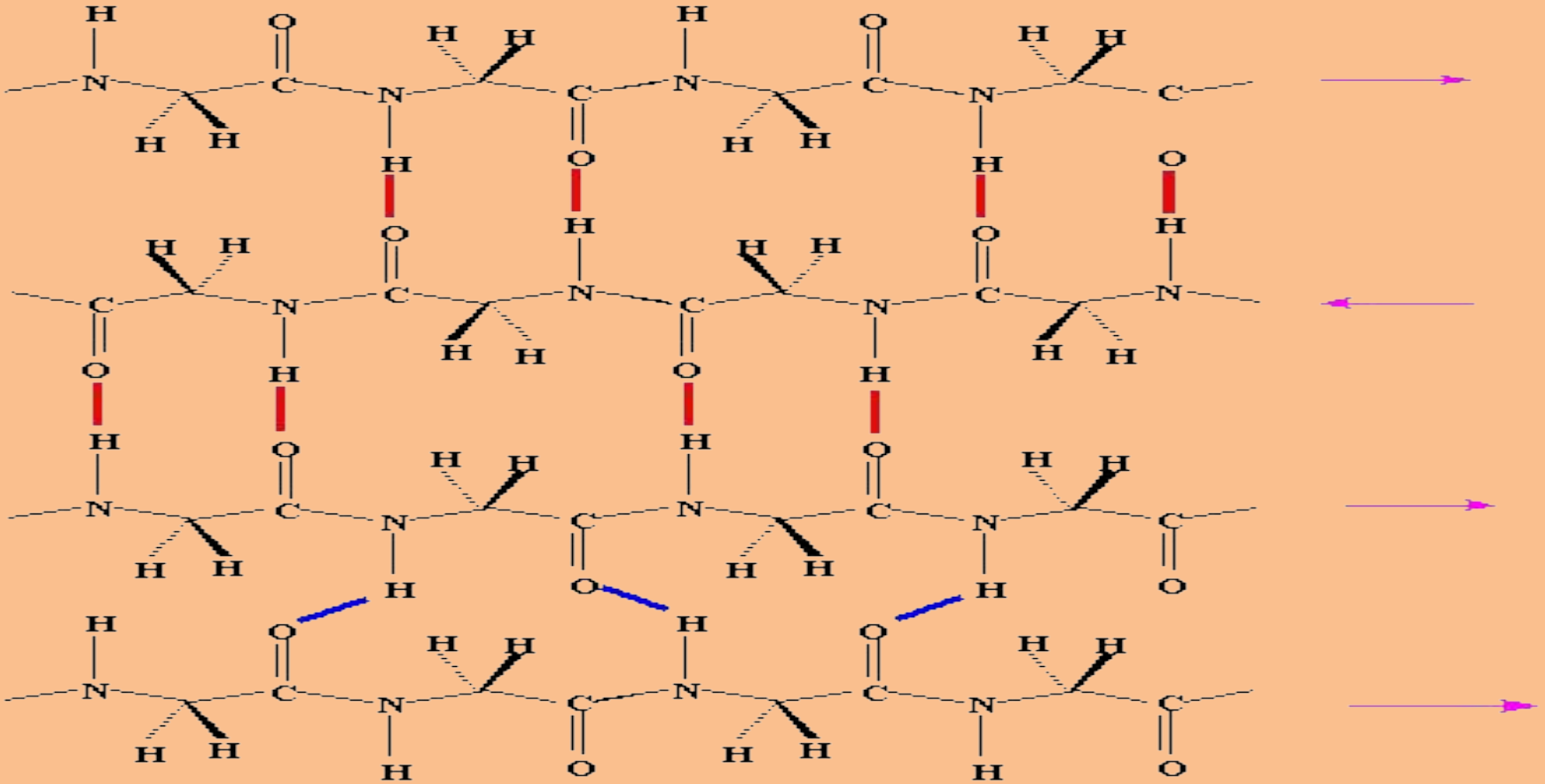
- Helices and sheets –commonly present in many proteins. They satisfy the hydrogen bonding potential of the mainchain N–H and C=O groups, while keeping the main chain in an unstrained conformation.
- A stretch of 6-20 consecutive residues generates an α Helix. Repeating the β conformations generate the β sheets. Two or more β strands can interact laterally to form β sheets
- Helices are a **local structure of the polypeptide chain, i.e.** they form h-bonds between consecutive residues.
- Sheets form by lateral interactions of several independent sets of residues to create a hydrogen-bonded network that is often nearly flat, but sometimes cylindrical (forming a **barrel structure). Unlike helices,** sheets bring together sections of the chain separated widely in the sequence.

Beta strands

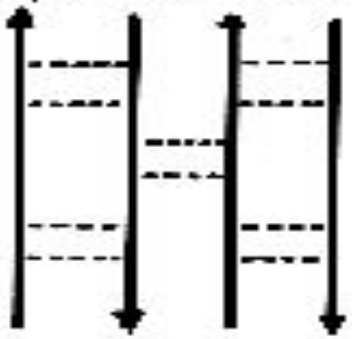
- Occurs in many proteins as two or more parallel or anti-parallel adjacent polypeptide chains arranged in such a way that hydrogen bonds can form between them
- very common in globular proteins.
- The number of strands in a sheet ranges from 2 to 10, with 5-6 being most common
- Width of 6 stranded β-sheet is appx 25 A°.
- Average length of β-strands is 5-7 residues although they can be 2 to 11 residues in length
- Mixed sheets with some β-strands pairs as parallel and some as anti-parallel also occur sometimes.

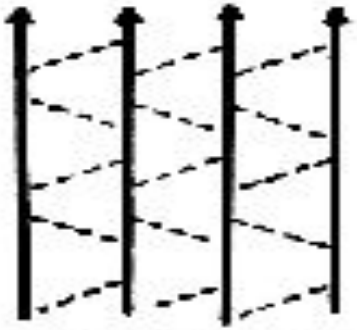
- No preference for parallel or anti-parallel β-sheets
- β-sheets with less than 4 strands are rare, indicating low stability
- φ , Ψ and ω are -120°, 120° and 180° respectively.
- Beta sheets have a right handed twist that may fold back upon itself leading to a barrel shape (a beta barrel)
- Beta bulge is also a variant; residue on one strand forms two hydrogen bonds with residue on other – causes one strand to bulge – occurs most frequently in parallel sheets

β-sheet representation



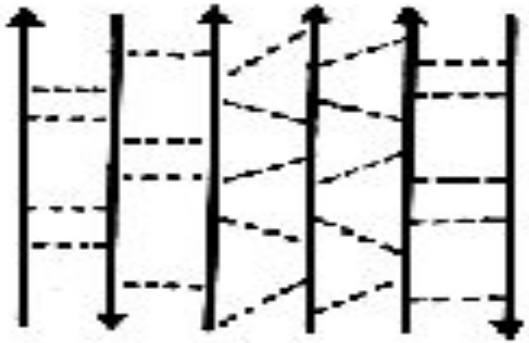
Antiparallel beta-sheet





Parallel beta-sheet

The different types of beta-sheet. Dashed lines indicate main chain hydrogen bonds.



Mixed beta-sheet

Types of β-strands

- **Parallel strand-** in parallel β- sheets, the evenly spaced hydrogen-bonded chains extend in the same direction. Parallel sheets are less twisted than the antiparallel ones and are always found buried
- **Antiparallel strand** – in this, the neighbouring narrowly spaced hydrogen bonded polypetide chains run in opposite directions. This type of sheets are more stable than parallel ones, which is consistent with the hydrogen bond geometry and the fact that small parallel sheets rarely occur

Turns

- Composed of 3-4 residues, turns are compact, U-shaped secondary structures stabilized by a hydrogen bond between their end residues.
- located on the surface of a protein, forming a sharp bend : redirects the polypeptide backbone back toward the interior.
- Glycine and proline are commonly present in turns. The lack of a large side chain in the case of glycine and the presence of a built-in bend in the case of proline allow the polypeptide backbone to fold into a tight U-shaped structure.
- Without turns, a protein would be large, extended, and loosely packed.
- A polypeptide backbone also may contain long bends, or *loops*.

Other secondary structure elements

- Loops
- Reverse turns
- β- hairpins
- Coiled coils

Loops

- Occur at the surface of the protein molecule
- These are regions of a protein chain that are between α- helices and β-sheets
- Varying length and 3D conformations
- The mainchain CO & NH groups of the loop regions, which generally do not form H-bonds with each other are exposed to the solvent and can form H- bonds with the water molecules
- Loops exposed to solvent have large quantities of charged and polar hydrophilic residues
- Loops interact with surrounding aqueous environment and other proteins
- It is possible to predict loop regions with higher accuracy than α- helices and β-sheets
- In homologous amino acid sequences, insertions or deletions of a few residues are found almost only in the loop regions because during evolution protein cores are much more stable than the loops

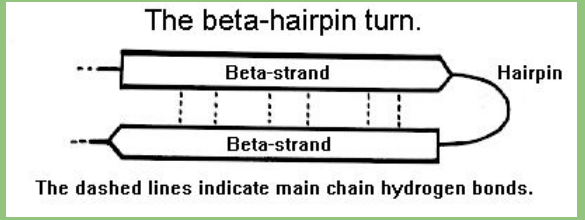
Reverse turns

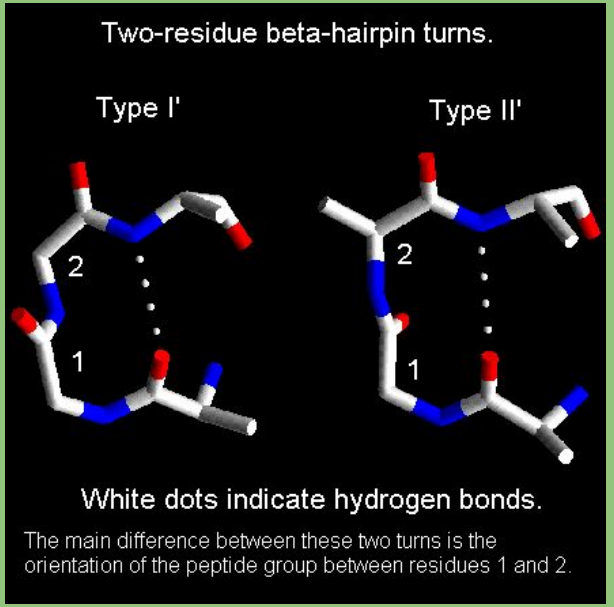
abundant in globular proteins and generally occur at the surfaces of the molecule

- These regions of the polypeptide chain have a H-bond from one mainchain carbonyl oxygen to the mainchain NH group 3 residues along the chain (i.e Oi to N(i+3))
- These regions act as nucleation centres during protein folding

β-hairpins

- One of the simplest supersecondary structures
- Widespread in globular proteins
- Consist of short loop regions/turns between antiparallel H-bonded β-strands.
- Length is less than 8 residues, 2 residue loops being most common





Coiled coil

- Is a ubiquitous protein structural motif that can mediate protein interactions
- 5-7% eukaryotic proteins have a coiled coil regions
- Are associated with several cellular functions like transcription, oncogenesis, cell structure, membrane fusions etc.
- Consist of 2 or more right handed α- helices wrapped around each other with a slight left-handed super helical twist