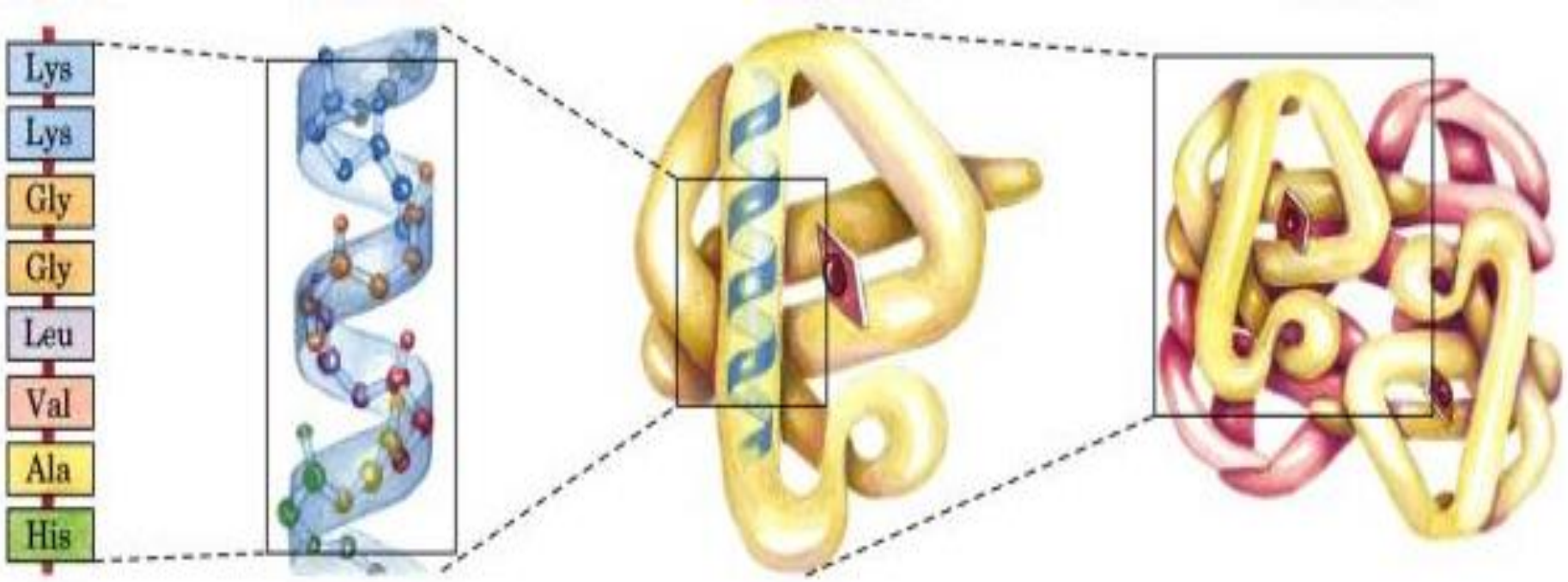
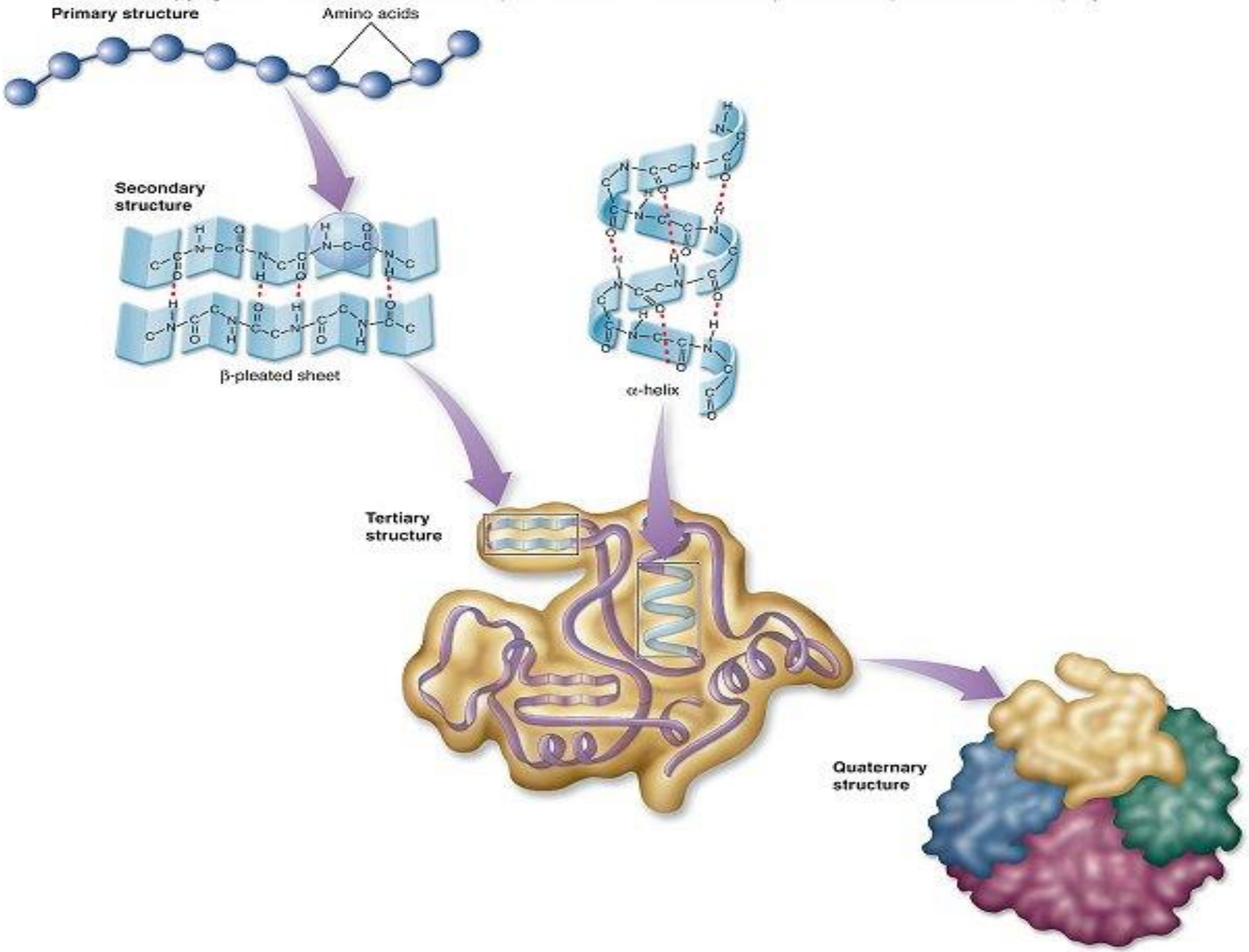
STRUCTURE and FUNCTIONS of PROTEINS

Lecture II

There Are Several Levels of Protein Structure



1) The ***primary structure*** consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds.
2) The resulting polypeptide can be coiled into units of ***secondary structure***, such as a helix.
3) The helix is a part of the ***tertiary structure*** of the folded polypeptide.
4) folded polypeptides make up the ***quaternary structure*** of the multisubunit protein


Primary structure
Amino acids
Secondary structure
β-pleated sheet
α-helix
Tertiary structure
Quaternary structure

Primary structure

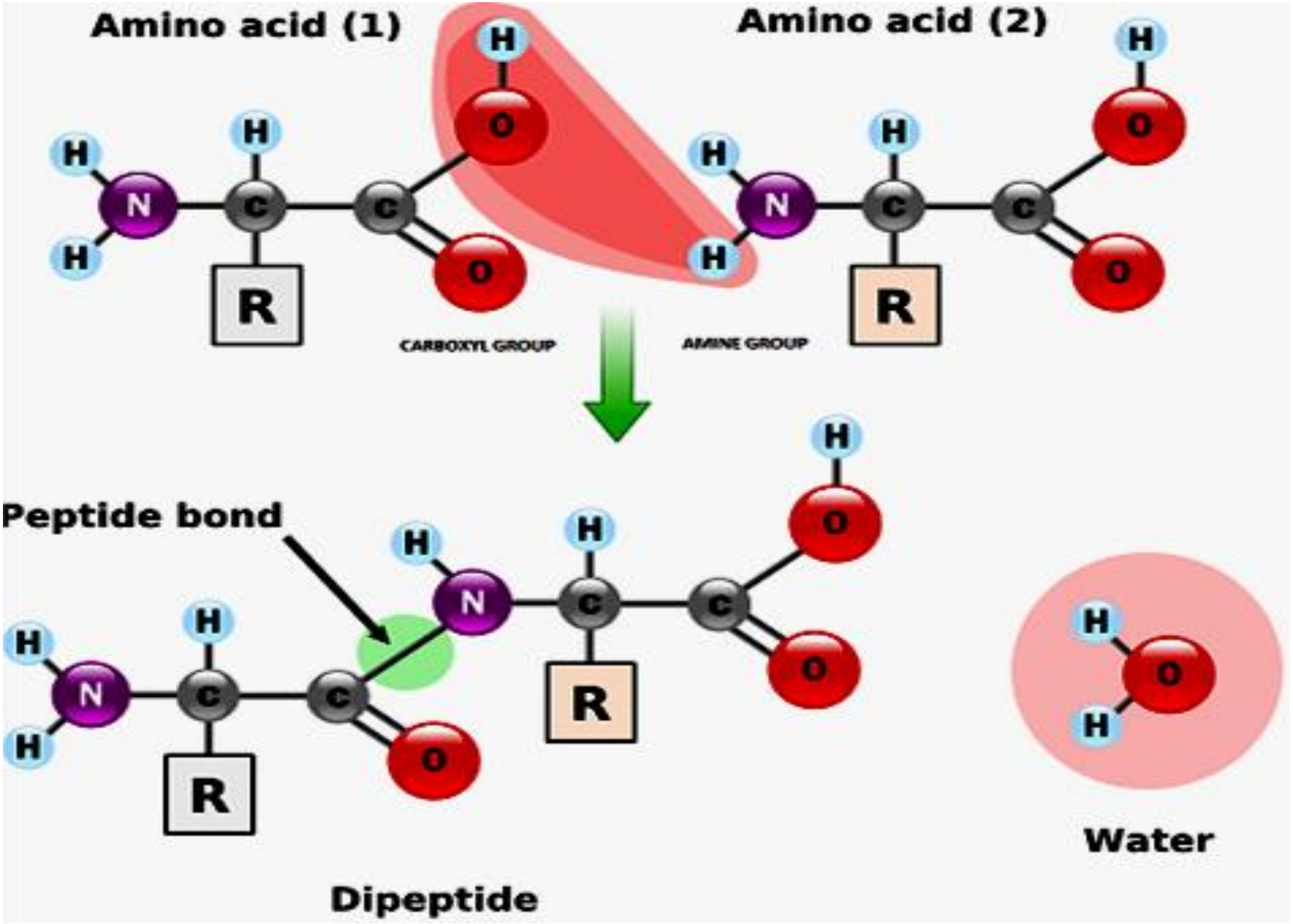
the *sequence* of amino acid residues in the polypeptide chaine.

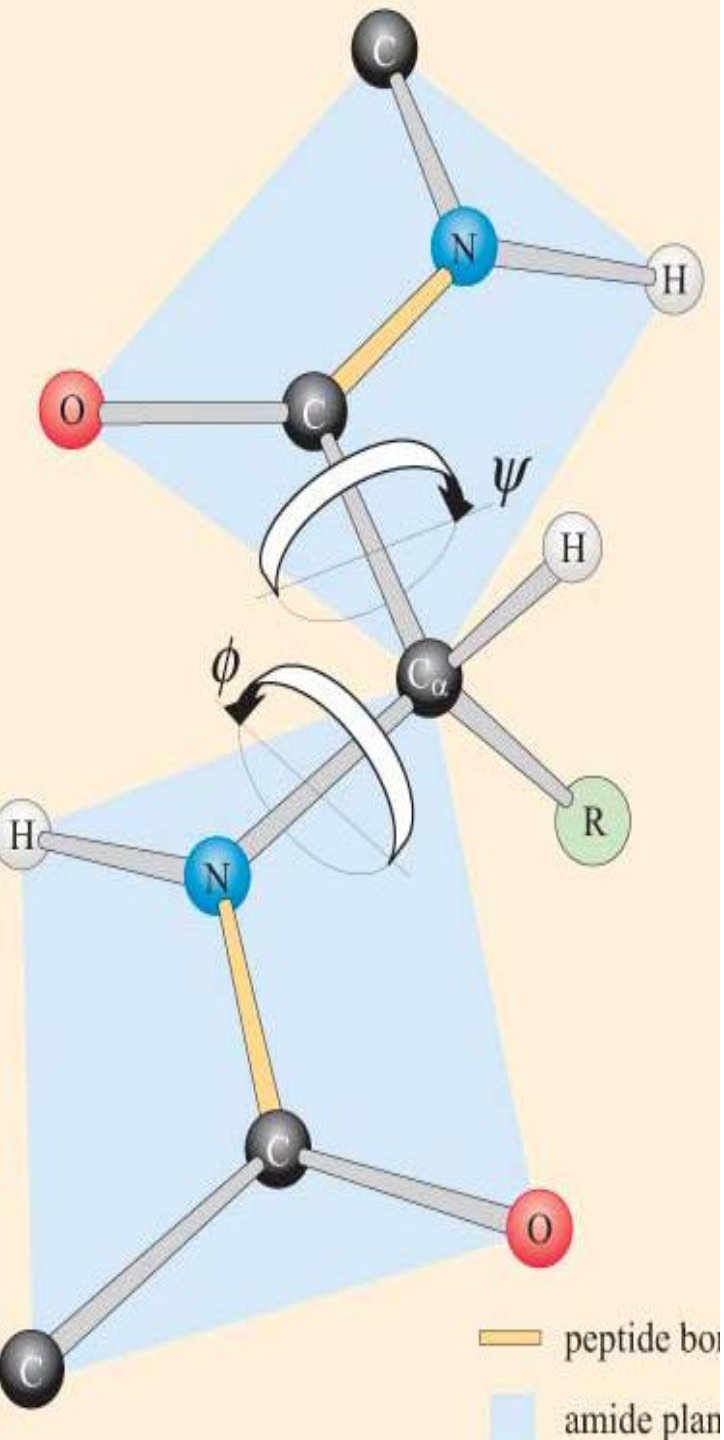
Each type of protein has a unique amino acid sequence. Amino acid sequence play a fundamental role in determining the three-dimensional structure of the protein, and its function.

Peptide Bond

the special name given to the amide bond between α-carboxyl group of one amino acid and α-amino group of another amino acid.

Peptide Bond



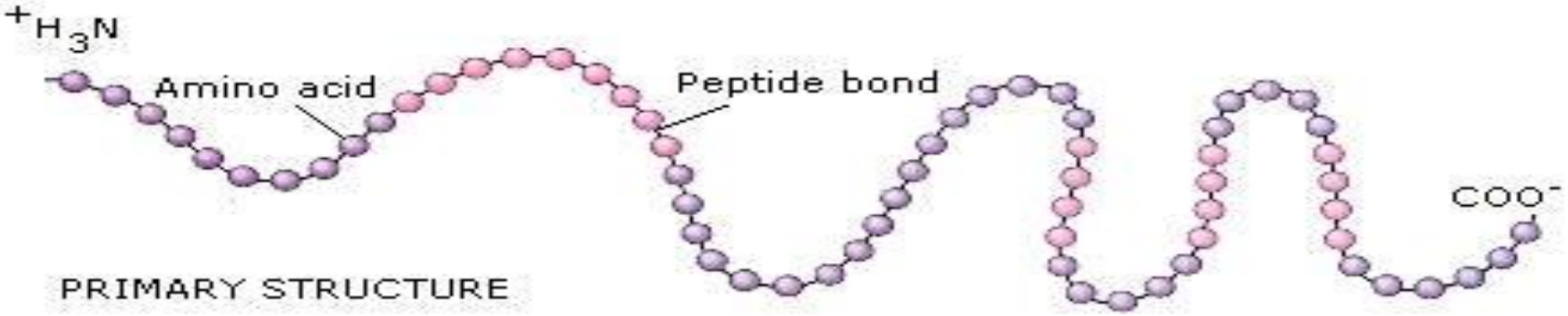


Characteristics of Peptide Bond

- **always in trans configuration**
- **some double nature of peptide bond caused planar geometry**
- **bond is rigid and *do not rotate***
- the other backbone bonds **Cα–N and Cα–C, are theoretically *free to rotate***
- if the R group attached to the α carbon alone is large enough, it will prevent complete rotation around the Cα–N and Cα–C bonds
- if these bonds are rotated in relation to each other, ***angles will be found where the two H (O) atoms of the peptide bonds would overlap each other and obstruct free rotation.***

The free -NH_2 group of the terminal amino acid is called as N-terminal end and the free -COOH end is called as C-terminal end.

It is a tradition to number the amino acids from N-terminal end as No.1 towards the C-terminal end. Presence of specific amino acids at a specific number is very significant for a particular function of a protein. Any change in the sequence is abnormal and may affect the function and properties of protein.



Relationship between amino acid sequence and biological function:

Proteins with different functions always have different amino acid sequences.

Functionally similar proteins from different species, often have similar amino acid sequences.

An extreme case is ubiquitin, a 76-residue protein involved in regulating the degradation of other proteins. The amino acid sequence of ubiquitin is identical in species as disparate as fruit flies and humans.

Relationship between amino acid sequence and biological function:

If the primary structure is altered, the function of the protein may also be changed.

Thousands of human genetic diseases have been traced to the production of defective proteins. One-third of these proteins are defective because of a single change in their amino acid sequence.

Some flexibility in amino acid sequence is possible.

An estimated 20% to 30% of the proteins in humans are

polymorphic,

having amino acid sequence variants in the human population. Many of these variations in sequence have little or no effect on the function of the protein.

Main steps in determination of primary structure

1) **Determination of amino acid composition.**

2) **Identyfying of N- and C-terminal amino acids.**

3) **The protein is cleaved into a set of specific fragments by chemical or enzymatic methods.**

   It is necessary to generate several sets of peptides using more than one method of cleavage.

Main steps in determination of primary structure

4) If any disulfide bonds are present, they must be broken.

5) Each fragment is purified, then sequenced by the Edman procedure.

6) Finally, the order in which the fragments appear in the original protein is determined and disulfide bonds (if any) are located.

Cleaving the Polypeptide Chain

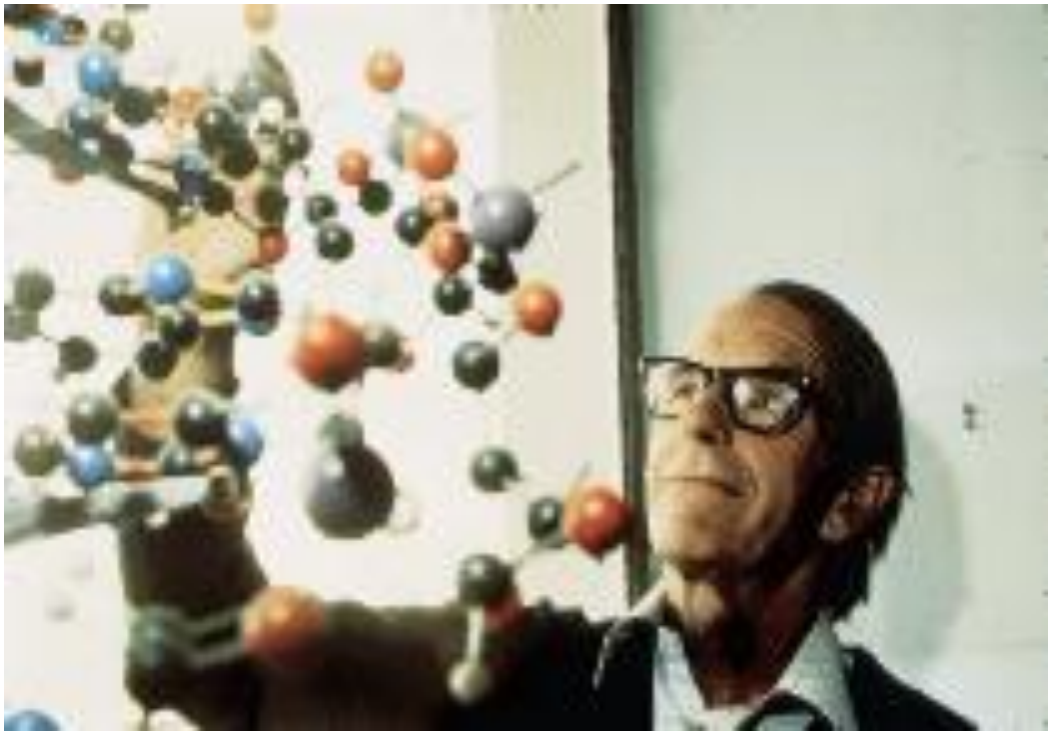
Several methods can be used for fragmenting the polypeptide chain:

- **Enzymes called proteases catalyze the hydrolytic cleavage of peptide bonds. Some proteases cleave only the peptide bond adjacent to particular amino acid residues.**

- **A number of chemical reagents also cleave the peptide bond adjacent to specific residues.**

Methods for determination of N-terminal amino acids

- **Sanger's method.**
- **Edman degradation procedure.**
- **Reaction with dansyl chloride.**
- **Use of aminopeptidase.**



The amino acid sequences of proteins have been determined using principles first developed by F.Sanger.

Sanger was first to sequence a polypeptide. He determined the complete primary structure of the hormone insulin.

Identification of N-terminal Residue

N-terminal residue can be identified by using a reagent that binds covalently with its α-NH_2 group. Because the bond is stable to hot acid hydrolysis, the derivative of the N-terminal residue can be identified by chromatographic procedures after the protein has been hydrolysed.

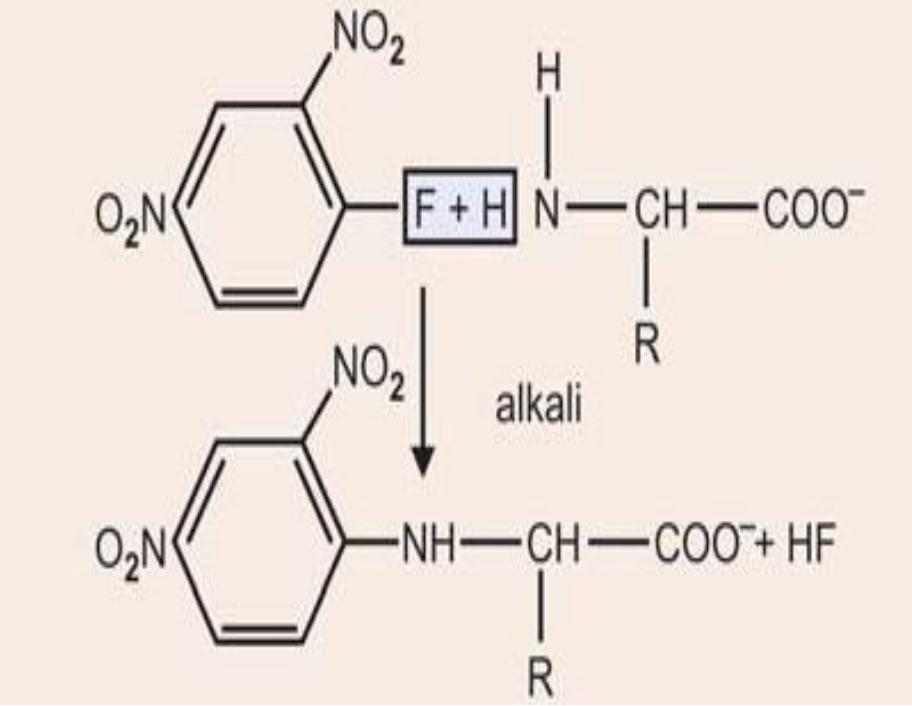
Identification of N-terminal Residue

Two reagents are commonly used

Sanger's reagent:

***1-fluoro-2,*4-dinitrobenzene (FDNB)**

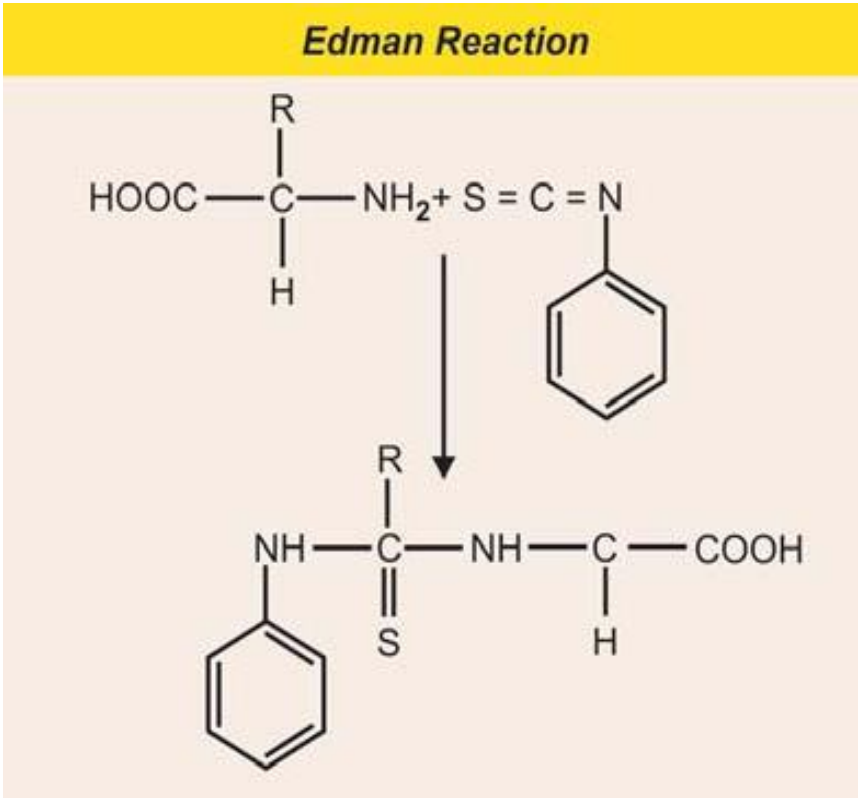
It reacts with free –NH_2 group in an alkaline medium



The reaction can also take place with the N-terminal –NH_2 group of the polypeptide chain. The compound so formed can be isolated after protein hydrolysis and identified.

In practice, a chemical method devised by P.Edman is usually employed.

The Edman degradation procedure (with *phenylisothiocyanate*) labels and removes only the amino-terminal residue from a peptide, leaving all other peptide bonds intact



The phenyl thiohydantoin amino acid liberated is identified by high performance liquid chromatography (HPLC).

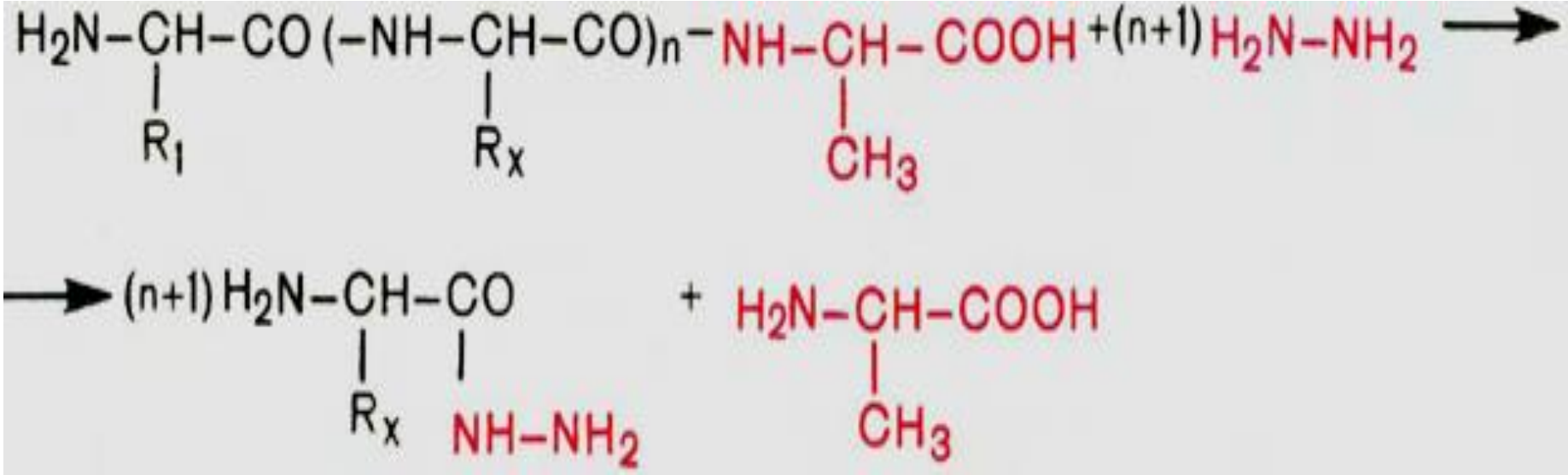
The Edman degradation is carried out in a machine, called a sequenator.

Automated amino acid sequencers now widely used, which permit very rapid determination of the amino acid sequences of polypeptides up to 100 amino acid approximately.

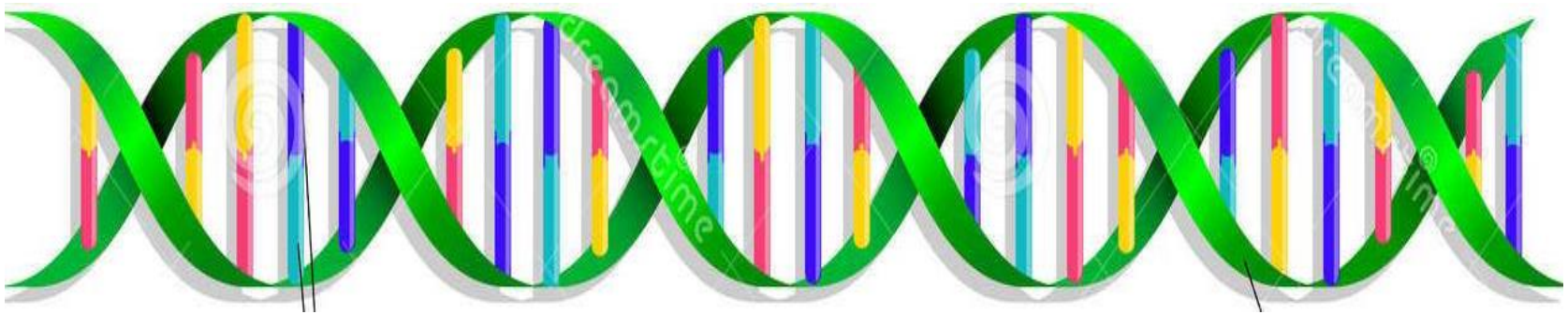
Methods for determination of C-terminal amino acids

- **Akabori's method.**
- **Reaction with cyanogen bromide (CNBr).**
- **Use of carboxypeptidase.**

Akabori's method
(based on hydrazinolysis of polypeptides)



A protein sequence can also be deduced from the nucleotide sequence of its corresponding gene in DNA



Secondary structure

local conformation of protein backbone

or

regular arrangement of amino acid residues in a segment of a polypeptide chain, in which each residue is spatially related to its neighbors in the same way

or

local arrangement of adjacent amino acids formed as the polypeptide chain folds

SECONDARY STRUCTURE

Regular

α-helix

β-sheet

Irregular

tight turns

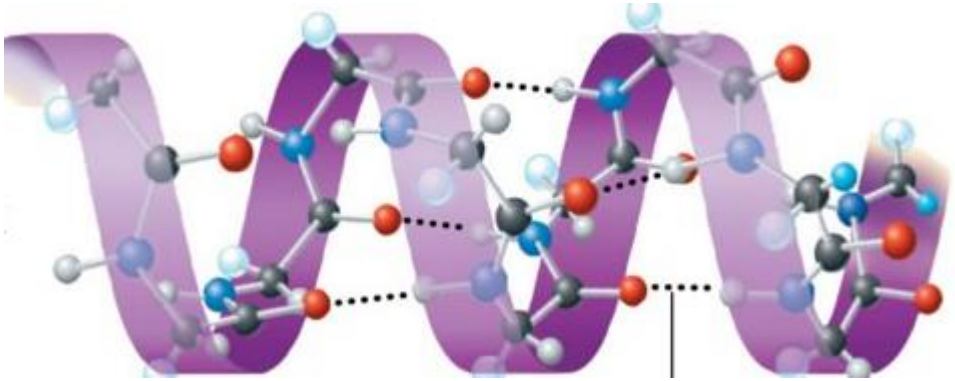
random coils

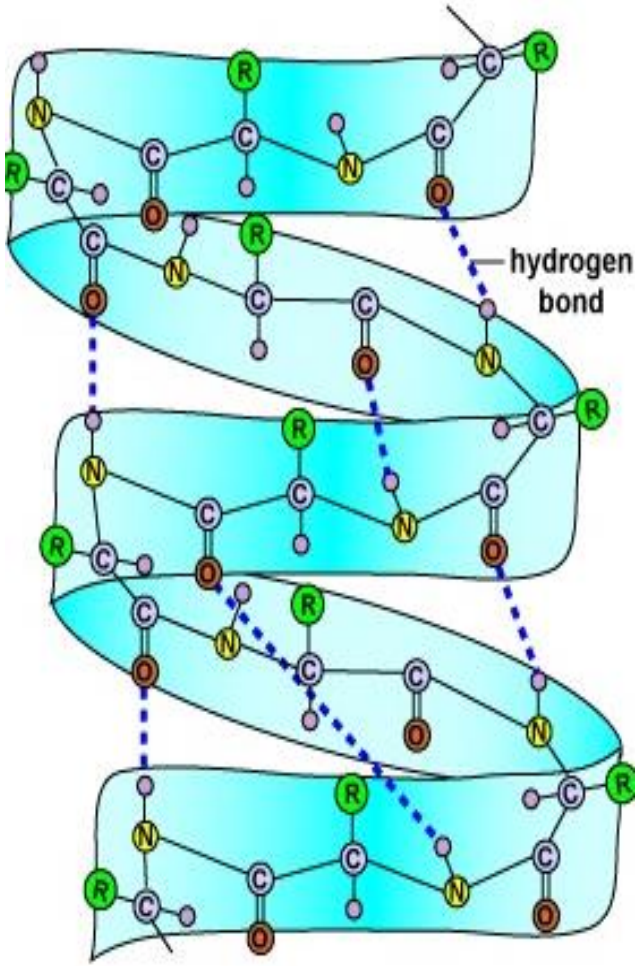
bulges

Secondary structure

The simplest arrangement of the polypeptide chain is a helical structure, which Pauling and Corey called the

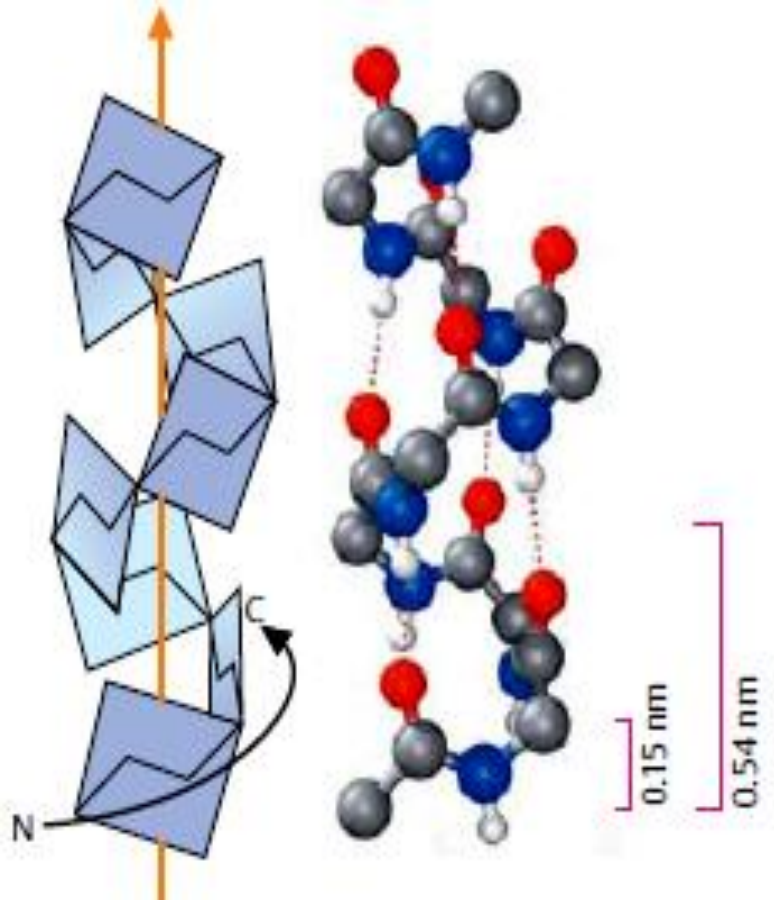
α-helix.





In this structure the polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues protrude outward from the helical backbone.

(Harper's Illustrated Biochemistry)



Only the right handed α-helix has been found in protein structure.

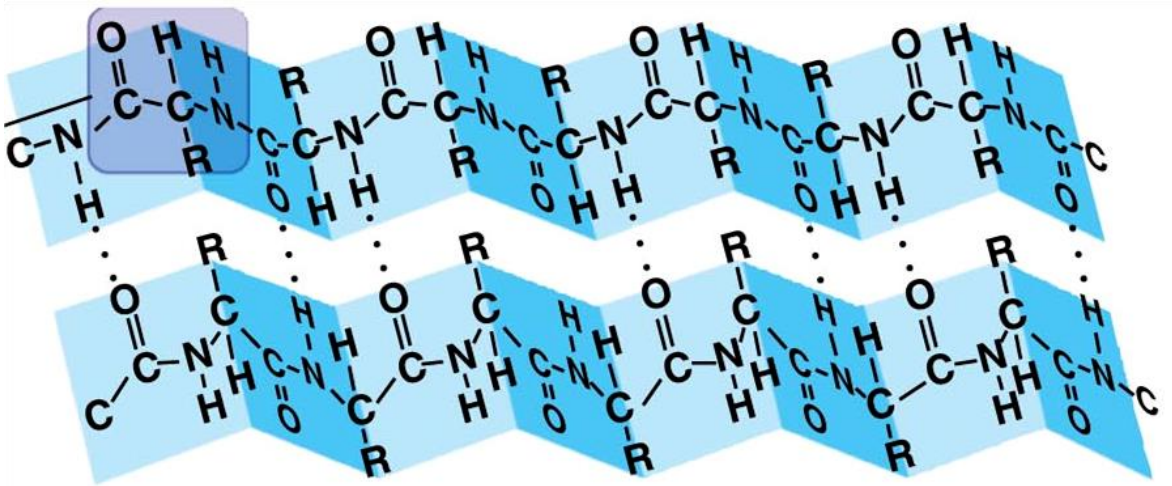
Each amino acid residue advances by 0.15 nm along the helix, and ***3.6 amino acid residues are present in one complete turn.***

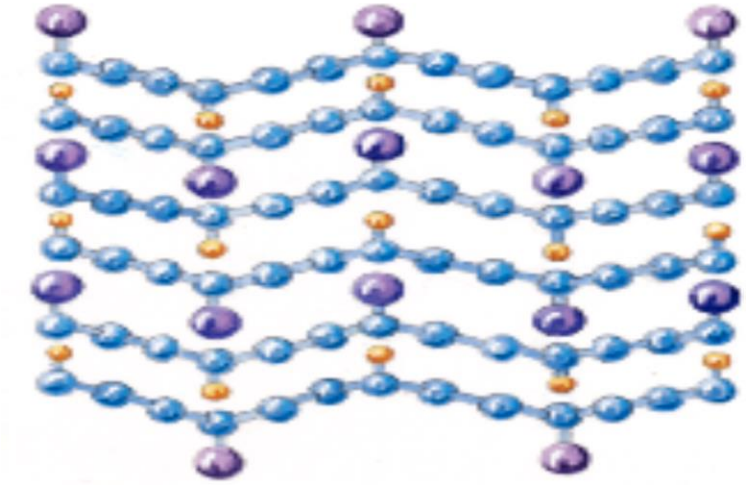
The distance between two equivalent points on turn is 0.54 nm and is called ***a pitch.***

Secondary structure

The second recognizable regular secondary structure in proteins is

β sheet





Secondary structure: stabilizing bonds

- **hydrogen bonds**

(bonds sharing a single hydrogen by two electronegative atoms such as O and N)

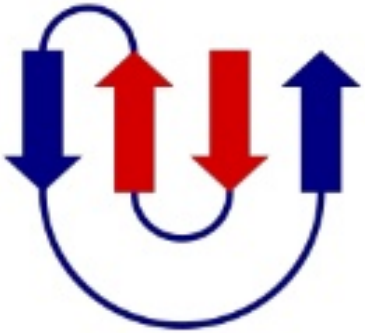
- **disulfide bonds**

(are formed between two cysteine residues. They are strong, high energy covalent bonds)

Supersecondary structures,

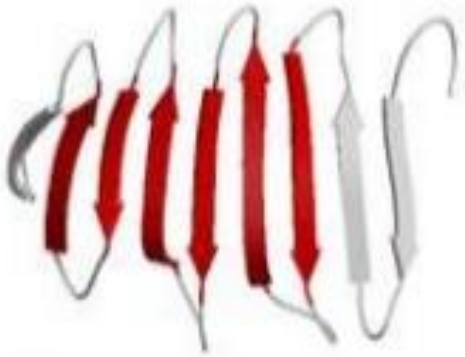
also called motifs or simply folds,

are particularly stable arrangements of several elements of secondary structure and the connections between them.



Greek key motif

beta-alpha-beta motif

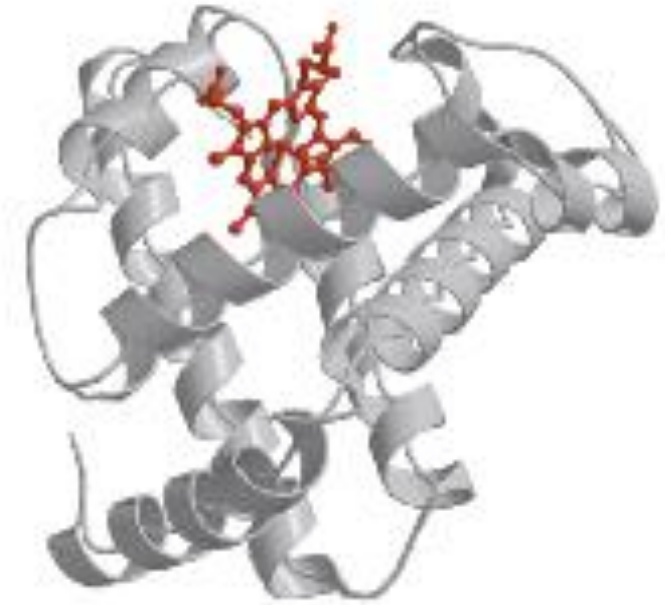


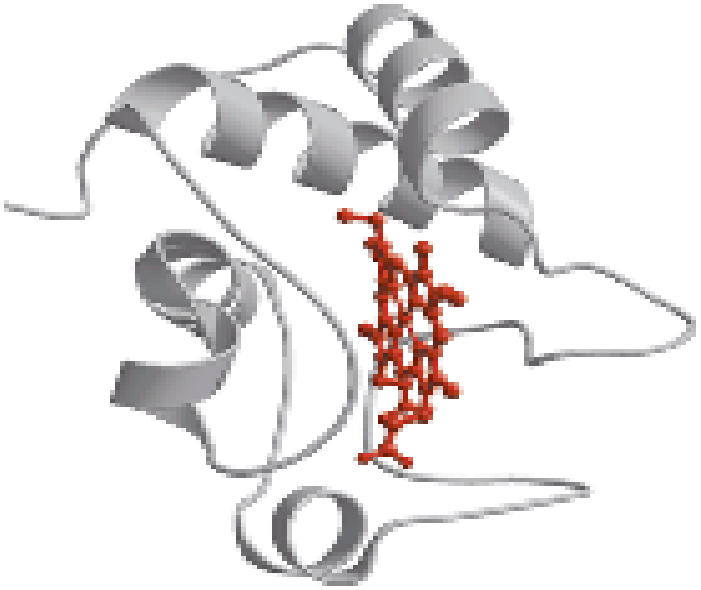
β-meander motif

polypeptide chain with secondary structure may be further folded

TERTIARY STRUCTURE

The overall three-dimensional arrangement of all atoms in a protein





Tertiary structure*: stabilizing bonds*

- **hydrophobic interactions** (normally occur between nonpolar side chains of amino acids such as ***alanine, leucine, methionine, isoleucine and phenylalanine.*** They constitute the major stabilising forces for tertiary structure forming a compact three-dimensional structure).

- **ionic or electrostatic interactions.** (These are formed between oppositely charged polar side chains of amino acids, such as basic and acidic amino acids).

Tertiary structure*: stabilizing bonds*

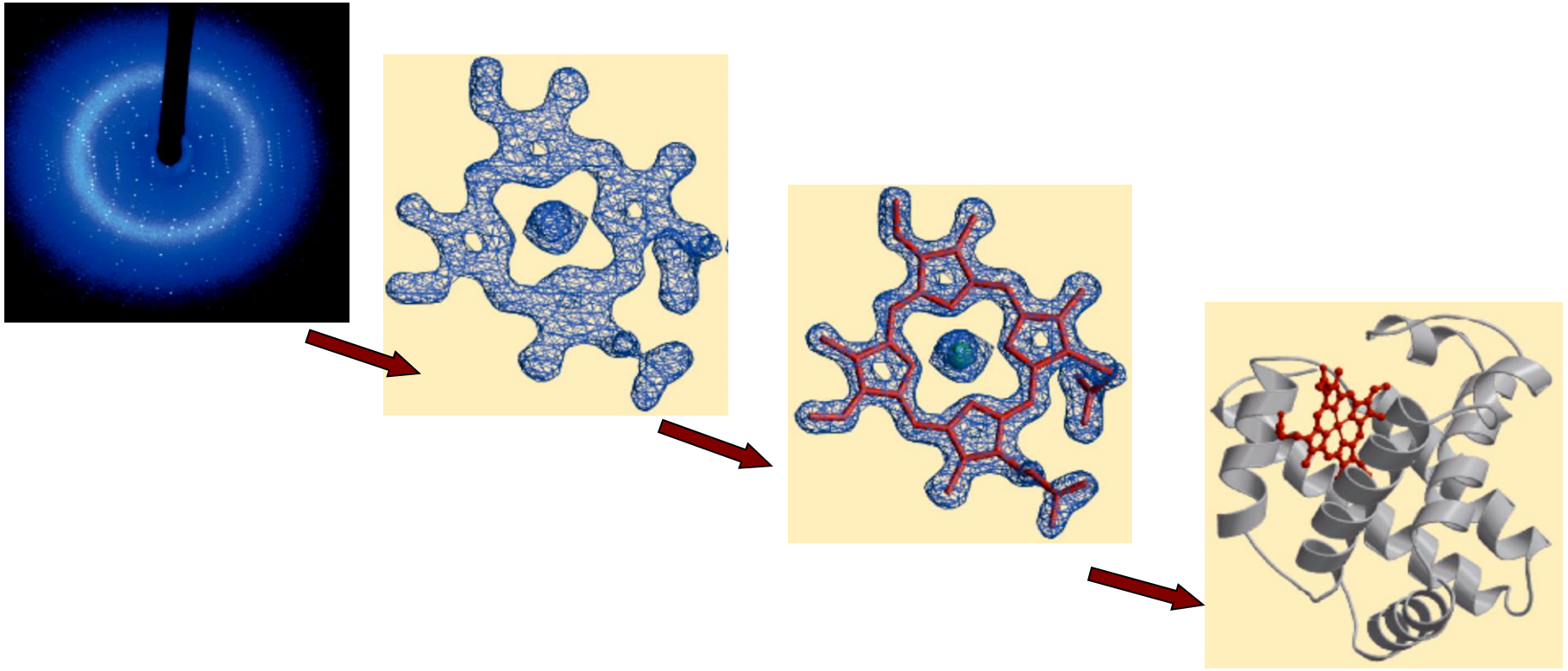
- **Van der Waal Forces.** (Occur between nonpolar side chains.

- **hydrogen bonds** (normally formed by the polar side chains of the amino acids).

- **disulfide bonds (**These are the S–S bonds between -SH groups of distant cysteine residues).

Methods for Determining the Three-Dimensional Structure of a Protein

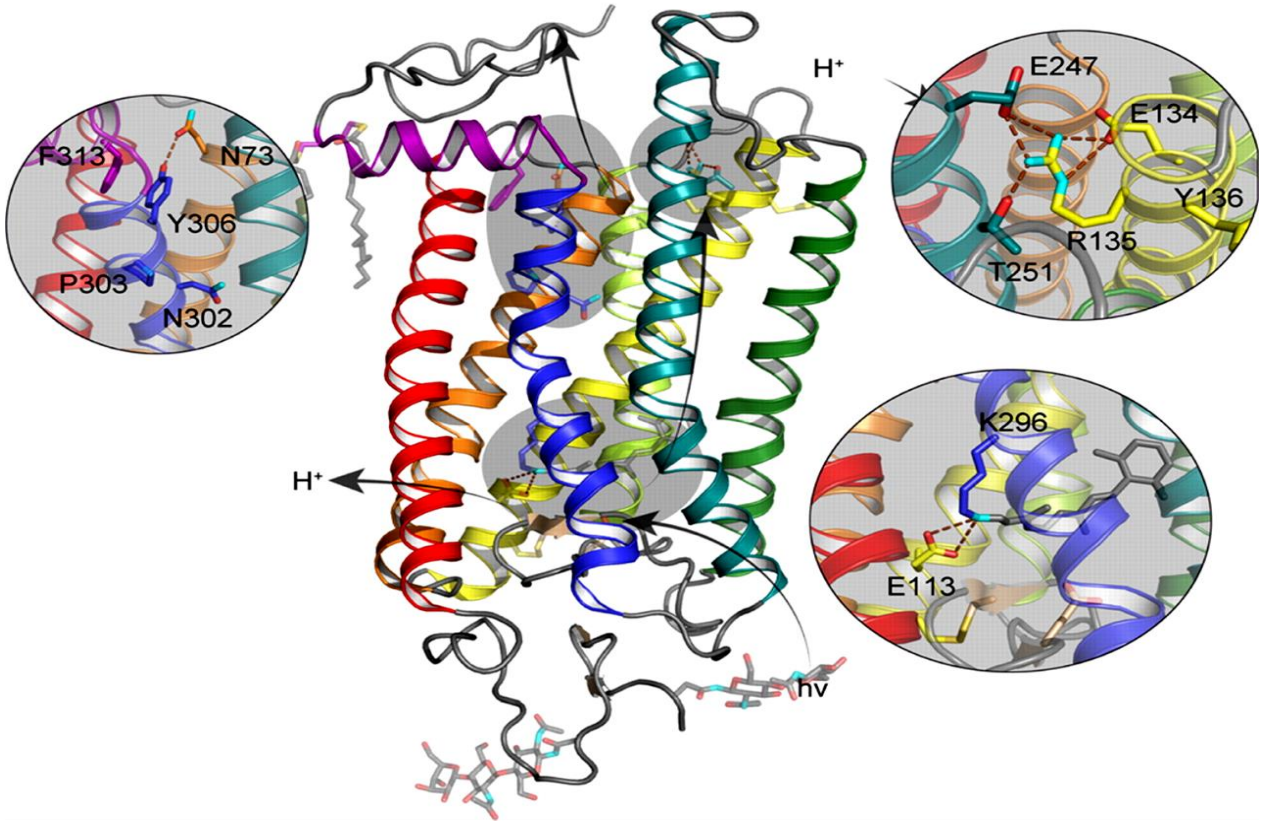
1. X-Ray Diffraction

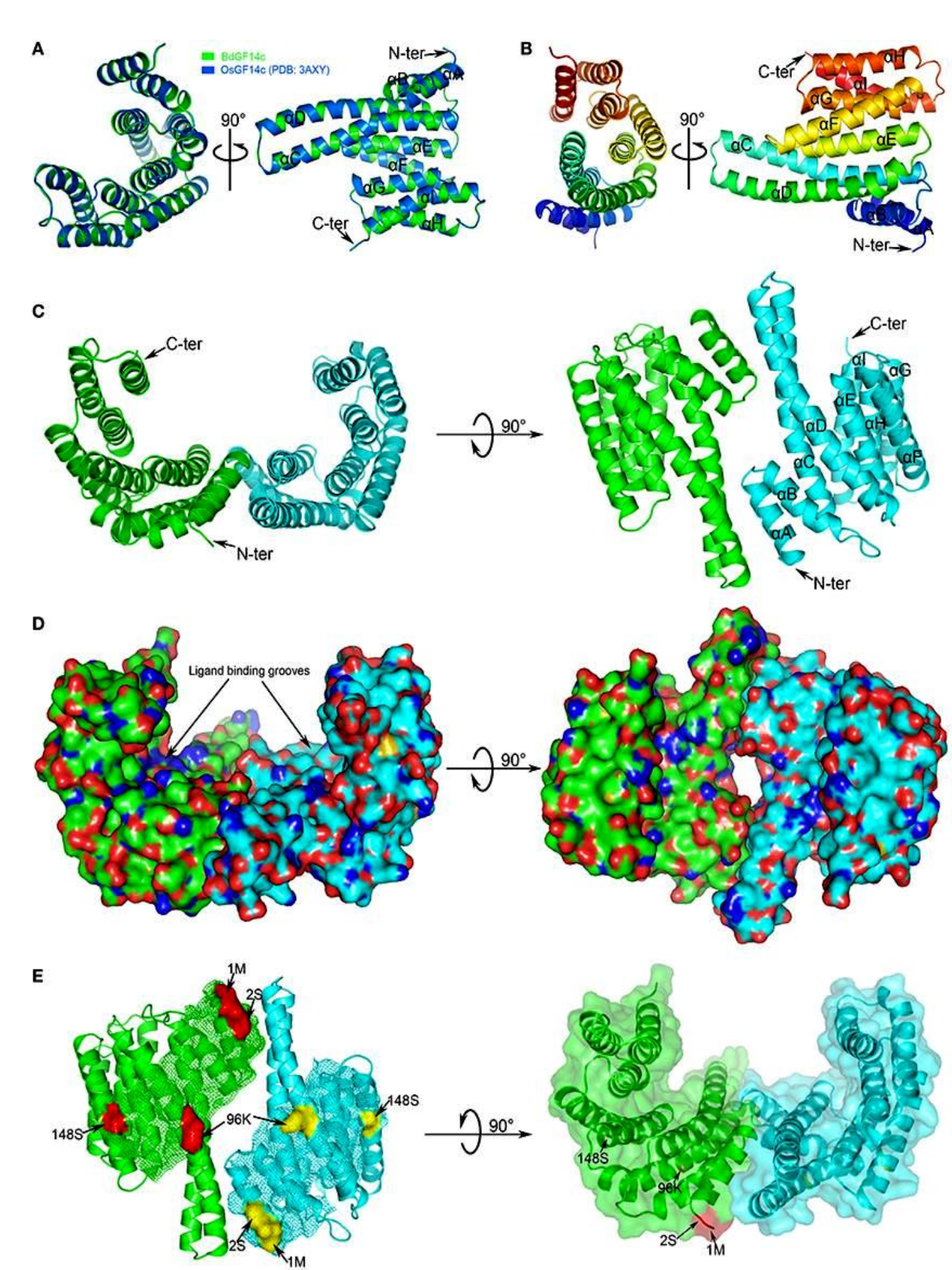


2 . Nuclear Magnetic Resonance

Scientists who determine the structure of a proteins deposit their data into a database such as Protein Data Bank (PDB).

A structure record shows the three-dimensional coordinates of every atom in the molecule.




A
BdGF14c
OsGF14c (PDB: 3AXY)
N-ter
αB
αA
αD
αE
αC
αF
αG
αI
C-ter
αH
90°
B
C-ter
αH
αI
αG
αF
αE
αC
αD
αB
αA
N-ter
90°
C
C-ter
N-ter
90°
C-ter
αI
αG
αE
αD
αH
αC
αF
αB
αA
N-ter
D
Ligand binding grooves
90°
E
1M
2S
148S
96K
148S
2S
1M
90°
148S
96K
2S
1M

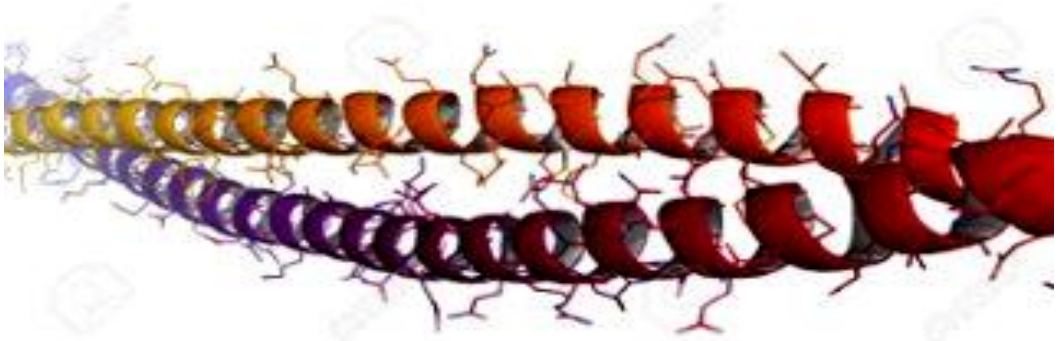
Tertiary structure

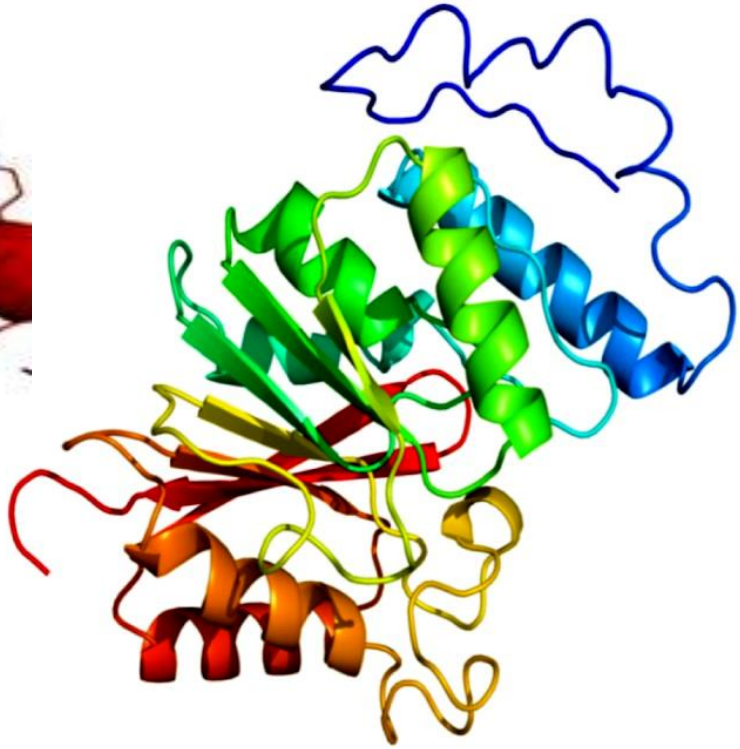
There are two general classes of proteins based on tertiary structure: fibrous and globular.

Fibrous proteins, which serve mainly structural roles, have simple repeating elements of secondary structure.

Globular proteins have more complicated tertiary structures, often containing several types of secondary structure in the same polypeptide chain.

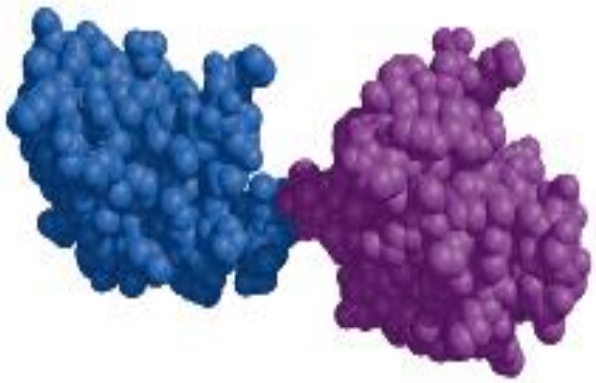
The thousands of known protein structures are generally assembled from a few hundred motifs.





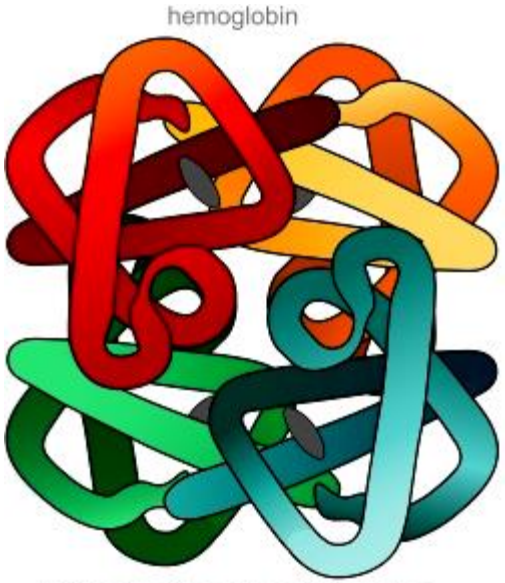
Polypeptides with more than a few hundred amino acid residues often fold into two or more stable, globular units called

domains.

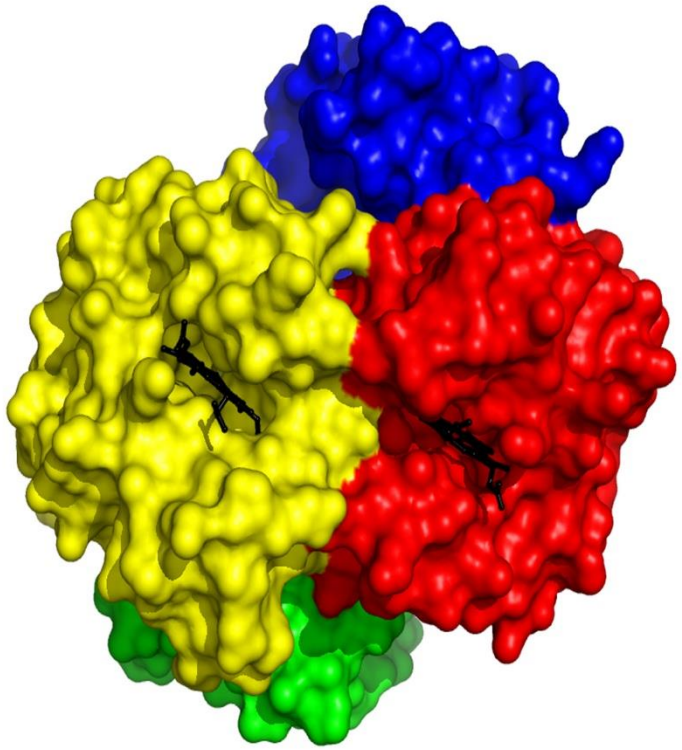


Quaternary structure

When a protein has two or more polypeptide subunits, their arrangement in space is referred to as quaternary structure.



Quaternary structure results from interactions between the subunits of multisubunit (multimeric) proteins .



A multisubunit protein is also referred to as a multimer. Multimeric proteins can have from two to hundreds of subunits.

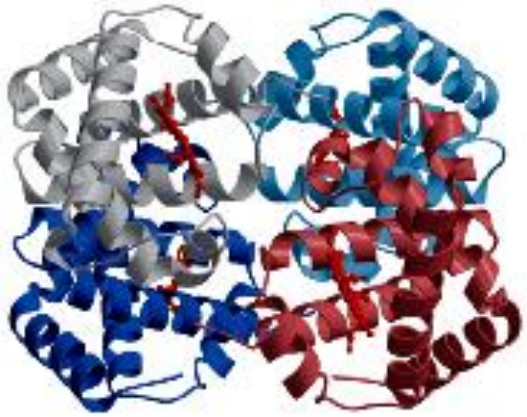
A multimer with just a few subunits is often called an oligomer.

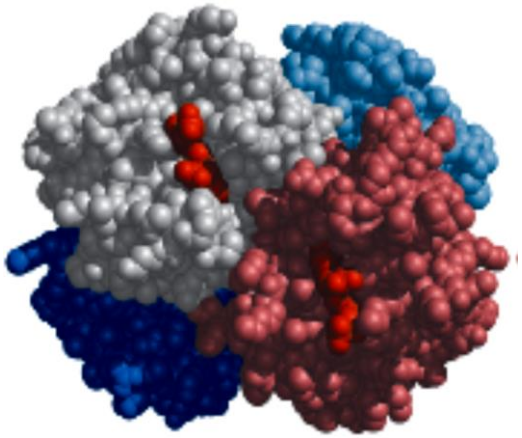
Most multimers have identical subunits or repeating groups of nonidentical subunits, usually in symmetric arrangements.

The repeating structural unit in such a multimeric protein, whether it is a single subunit or a group of subunits, is called a protomer.

The first oligomeric protein for which the 3D structure was determined was hemoglobin, which contains four polypeptide chains and four heme prosthetic groups.

The subunits of hemoglobin are arranged in symmetric pairs, each pair having one α-and one β-subunit. Hemoglobin can therefore be described as a tetramer





The association of polypeptide chains can serve a variety of functions:

- **Many multisubunit proteins have regulatory roles; the binding of small molecules may affect the interaction between subunits, causing large changes in the protein's activity in response to small changes in the concentration of substrate or regulatory molecules.**

- **In other cases, separate subunits can take on separate but related functions, such as catalysis and regulation.**

- **Some associations, such as the fibrous proteins and the coat proteins of viruses, serve primarily structural roles.**

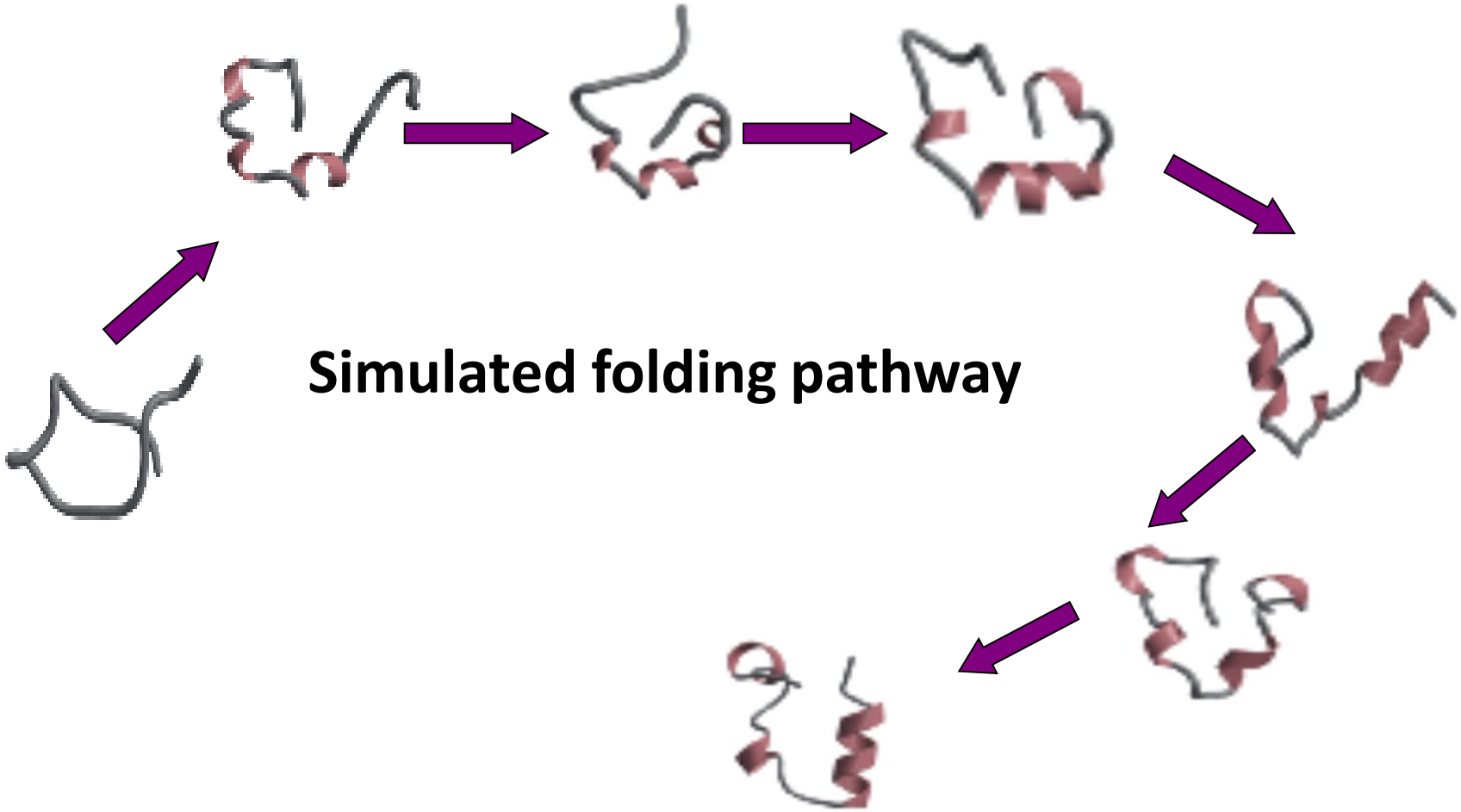
- **Some very large protein assemblies are the site of complex, multistep reactions.**

Protein Folding

All proteins begin their existence on a ribosome as a linear sequence of amino acid residues.

This polypeptide must fold during and following synthesis to take up its native conformation.

The folding pathway of a large polypeptide chain is complicated

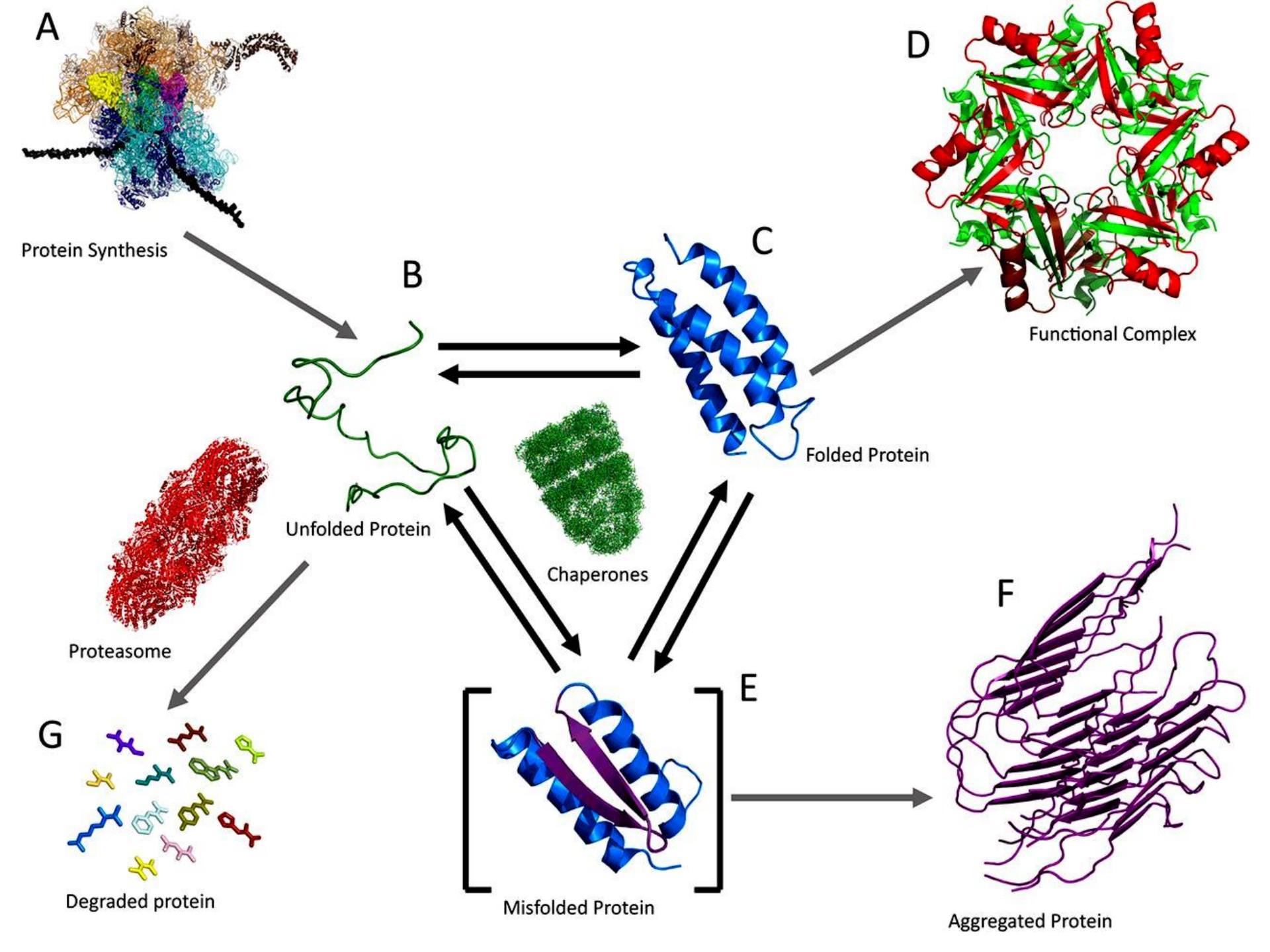

Simulated folding pathway

Not all proteins fold spontaneously as they are synthesized in the cell.

Folding for many proteins is facilitated by the action of specialized proteins called

chaperones

Chaperones interact with partially folded or improperly folded polypeptides, facilitating correct folding pathways or providing microenvironments in which folding can occur.


A
Protein Synthesis
B
Unfolded Protein
C
Folded Protein
D
Functional Complex
Chaperones
Proteasome
E
Misfolded Protein
F
Aggregated Protein
G
Degraded protein

A misfolded protein appears to be the causative agent of a number of rare degenerative brain diseases in mammals.

Perhaps the best known of these is mad cow disease (bovine spongiform encephalopathy, BSE).

The infectious agent has been traced to a single protein, which dubbed

prion protein (PrP).

(from *pr*oteinaceous *i*nfectious *on*ly)

Prion protein is a normal constituent of brain tissue in all mammals. Its role in the mammalian brain is not known in detail, but it appears to have a molecular signaling function.

Illness occurs only when the normal cellular PrP, or PrPC, occurs in an altered conformation called PrPSc (Sc denotes scrapie).

The interaction of PrPSc with PrPC converts the latter to PrPSc, initiating a domino effect in which more and more of the brain protein converts to the disease-causing form.

In inherited forms of prion diseases, a mutation in the gene encoding PrP produces a change in one amino acid residue that is believed to make the conversion of PrPC to PrPSc more likely.