

Struktura a funkce proteinů

Protein Structure and Function

C7920

doc. RNDr. Břetislav Brzobohatý, CSc.
Tel. 541517184, E-mail: brzoboha@ibp.cz
www.ibp.cz/labs/LMAPD

INTRODUCTION

The ultimate goal of molecular biology is to understand biological processes in terms of the chemistry and physics of the macromolecules that participate in them.

The importance of molecular structure for an understanding of function is best exemplified by DNA.

The simple and beautiful double-helical, base-paired structure of DNA immediately made genetics intelligible in chemical terms.

Genes, the previously mysterious factors that controlled inheritance of particular traits, were segments of DNA molecules that could be spooled out of solution at the end of rotating glass rod, like cotton candy on a stick.

INTRODUCTION

Understanding DNA structure explained the two cardinal properties of genes:

- their ability to replicate
- and
- their ability to determine structure of proteins.

Molecular genetics came into being and flourished as the genetic code was deciphered and patterns of gene organization and expression were elucidated.

INTRODUCTION

By contrast, the determination of structures of myoglobin and hemoglobin, less than a decade after the structure of DNA, and many more proteins since, has not yielded a simple and all-embracing explanation of protein structure and function.

- protein structures are much more complex than that of DNA
- proteins are built up from 20 different amino acids compared to 4 nucleotides of DNA

Proteins fulfill a much wider range of biological functions than does DNA, and functional diversity has dictated structural diversity.

By comparison with molecular genetics, progress in research on protein structure has been painfully slow due to

- great diversity that protein structures have proven to have
- the simple technical problems of obtaining protein crystals that are large enough for crystallographic analysis to atomic resolution.

INTRODUCTION

Recent technical developments that minimized the technical problems include:

- recombinant proteins can be produced and purified in high degree of purity with high yields, thus, even rare proteins are amenable to x-ray analysis
- very sensitive x-ray detectors were developed to replace x-ray films as the means of collecting x-ray diffraction data
- power of computers increased dramatically
- NMR developed into a tool for rapid determination of three-dimensional structures of small molecules

Molecular geneticists using site directed mutagenesis and other techniques of reverse genetics are now redesigning proteins both to investigate their biological function and to tailor enzymes for biotechnological processes.

The realization of full potential of protein engineering will depend critically on having a thorough grasp of the three-dimensional structure of the protein that is to be redesigned.

PART 1

Basic Structural Principles

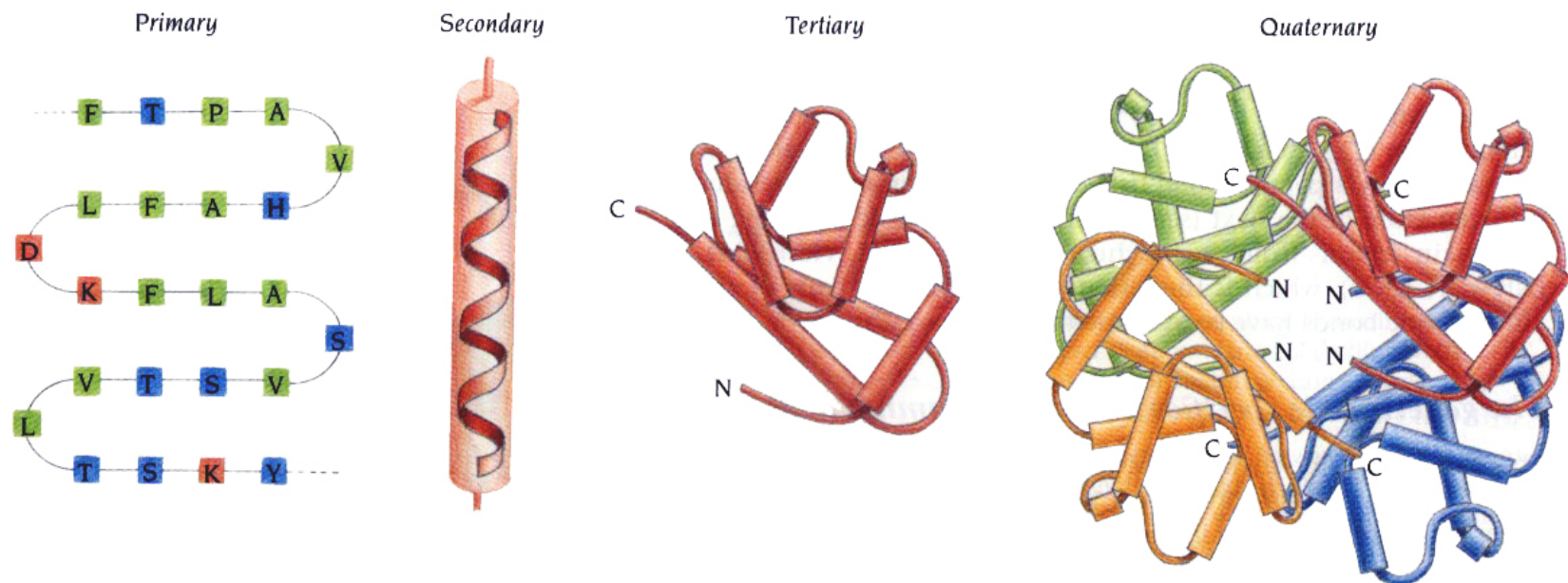
PART 2

Structure, Function, and Engineering

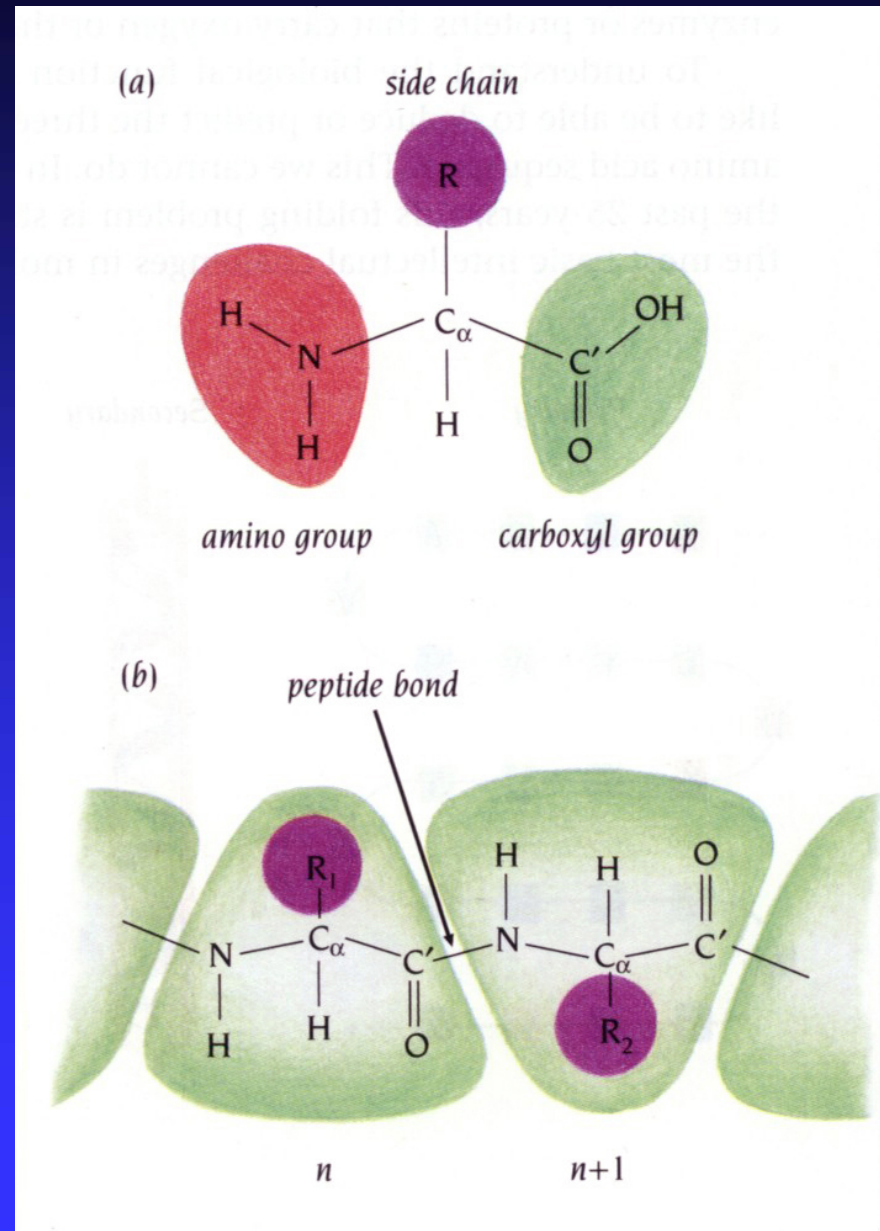
PART 1

Basic Structural Principles

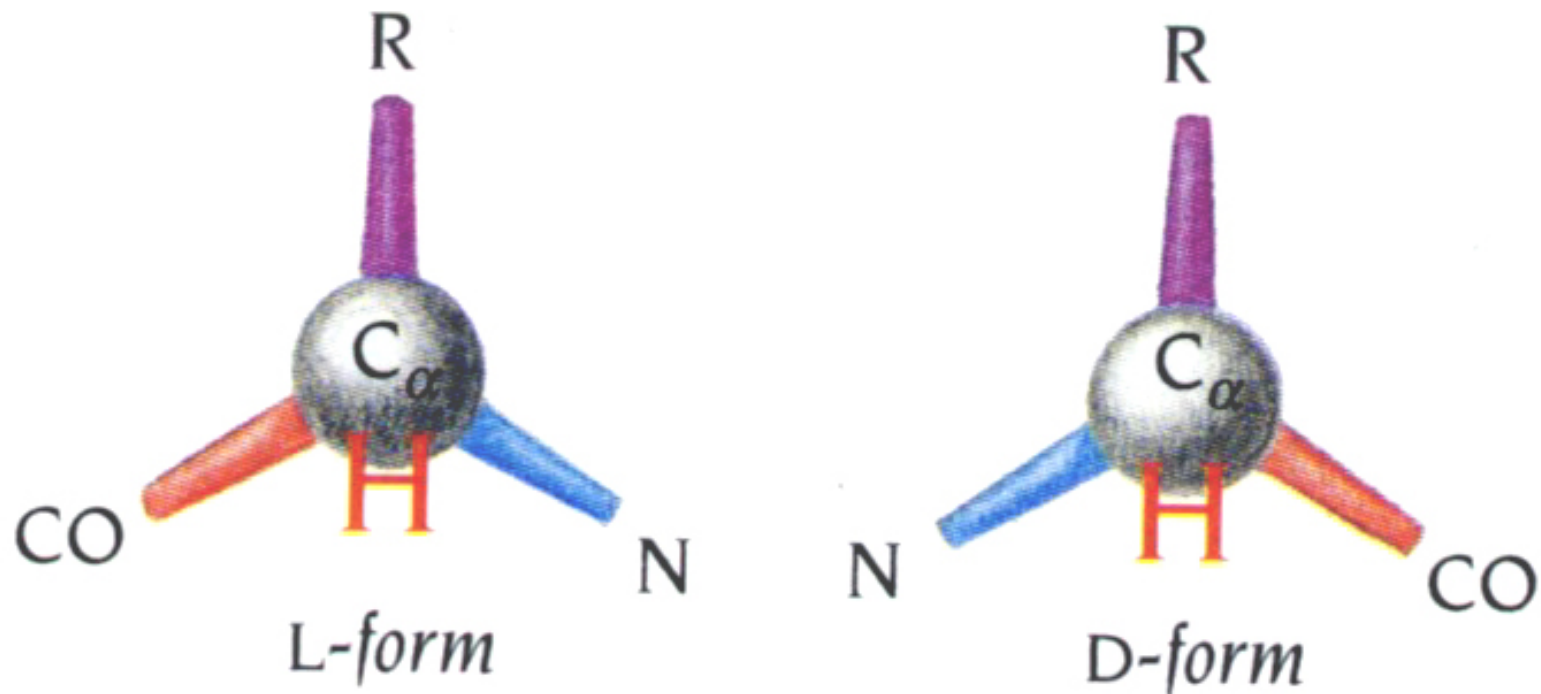
Protein structure



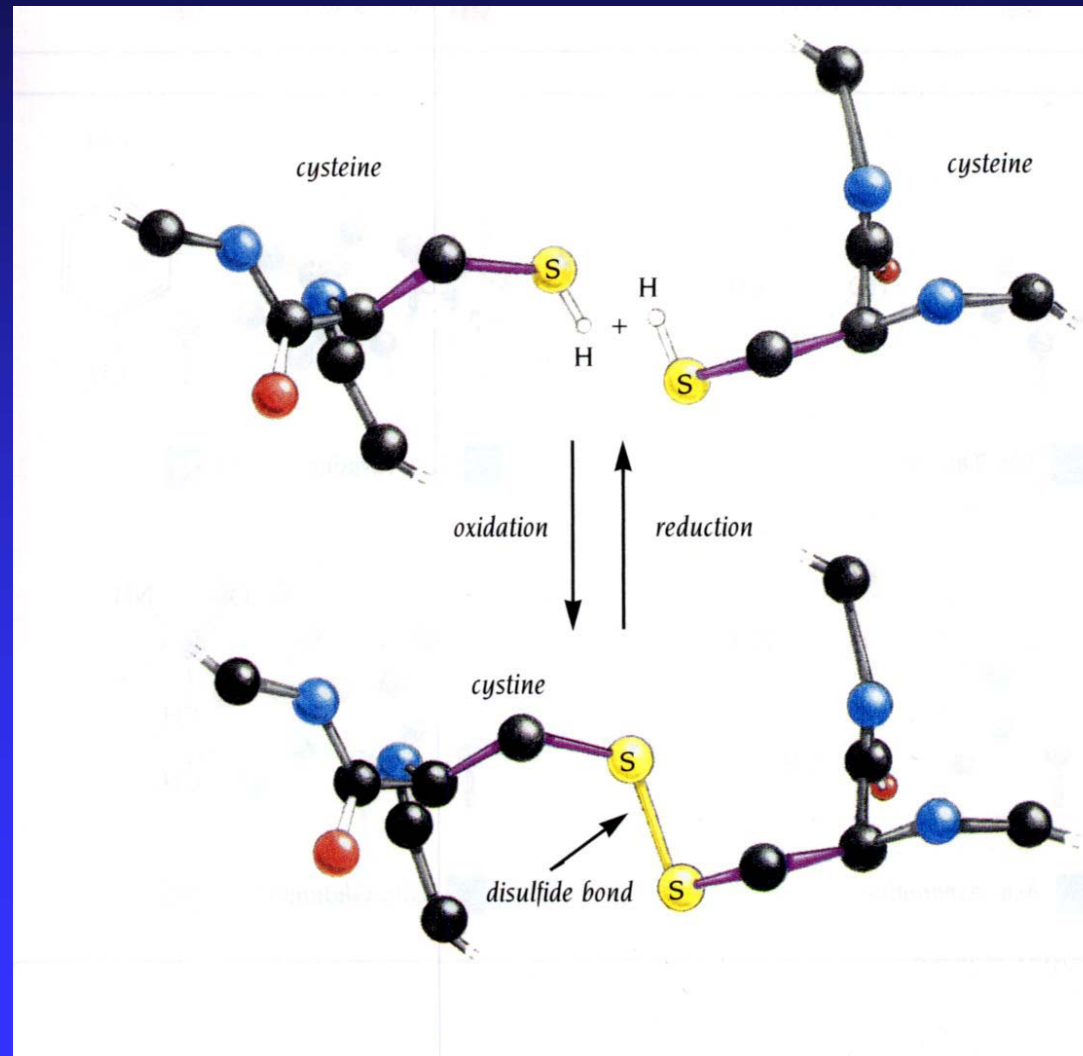
Peptide bond formation



L-forms of amino acids are utilized by translation machinery for protein synthesis

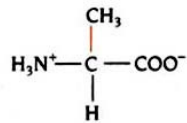
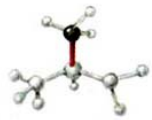


Two adjacent cysteine residues can form a disulfide bond in proteins



Hydrophobic amino acids

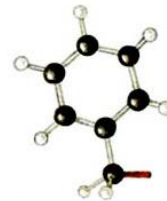
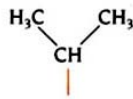
(a) Hydrophobic amino acids



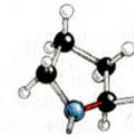
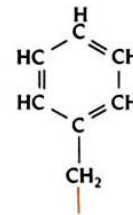
A Ala, Alanine



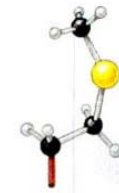
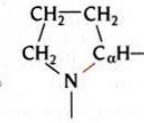
V Val, Valine



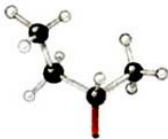
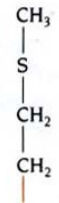
F Phe, Phenylalanine



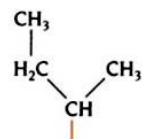
P Pro, Proline



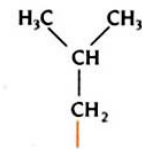
M Met, Methionine



I Ile, Isoleucine



L Leu, Leucine

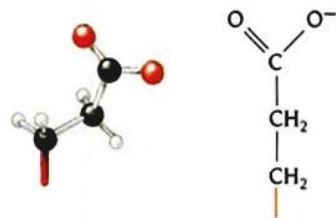


Charged amino acids

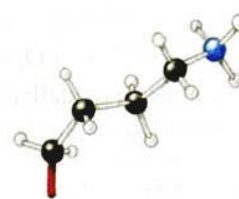
(b) Charged amino acids



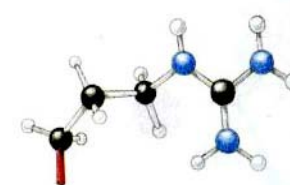
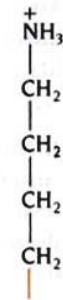
D Asp, Aspartic acid



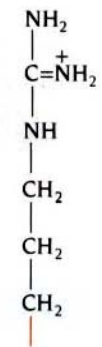
E Glu, Glutamic acid



K Lys, Lysine



R Arg, Arginine



Polar amino acids

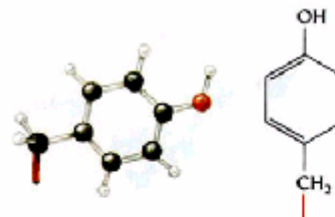
(c) Polar amino acids



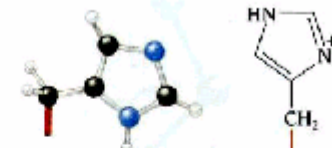
S Ser, Serine



T Thr, Threonine



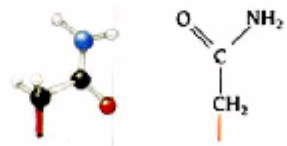
Y Tyr, Tyrosine



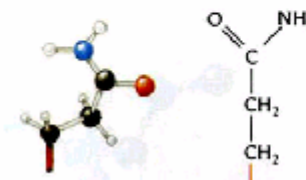
H His, Histidine



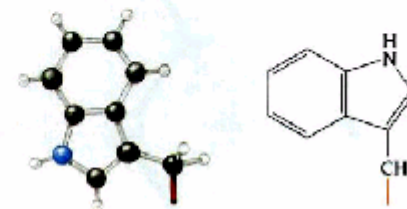
C Cys, Cysteine



N Asn, Asparagine



Q Gln, Glutamine



W Trp, Tryptophan

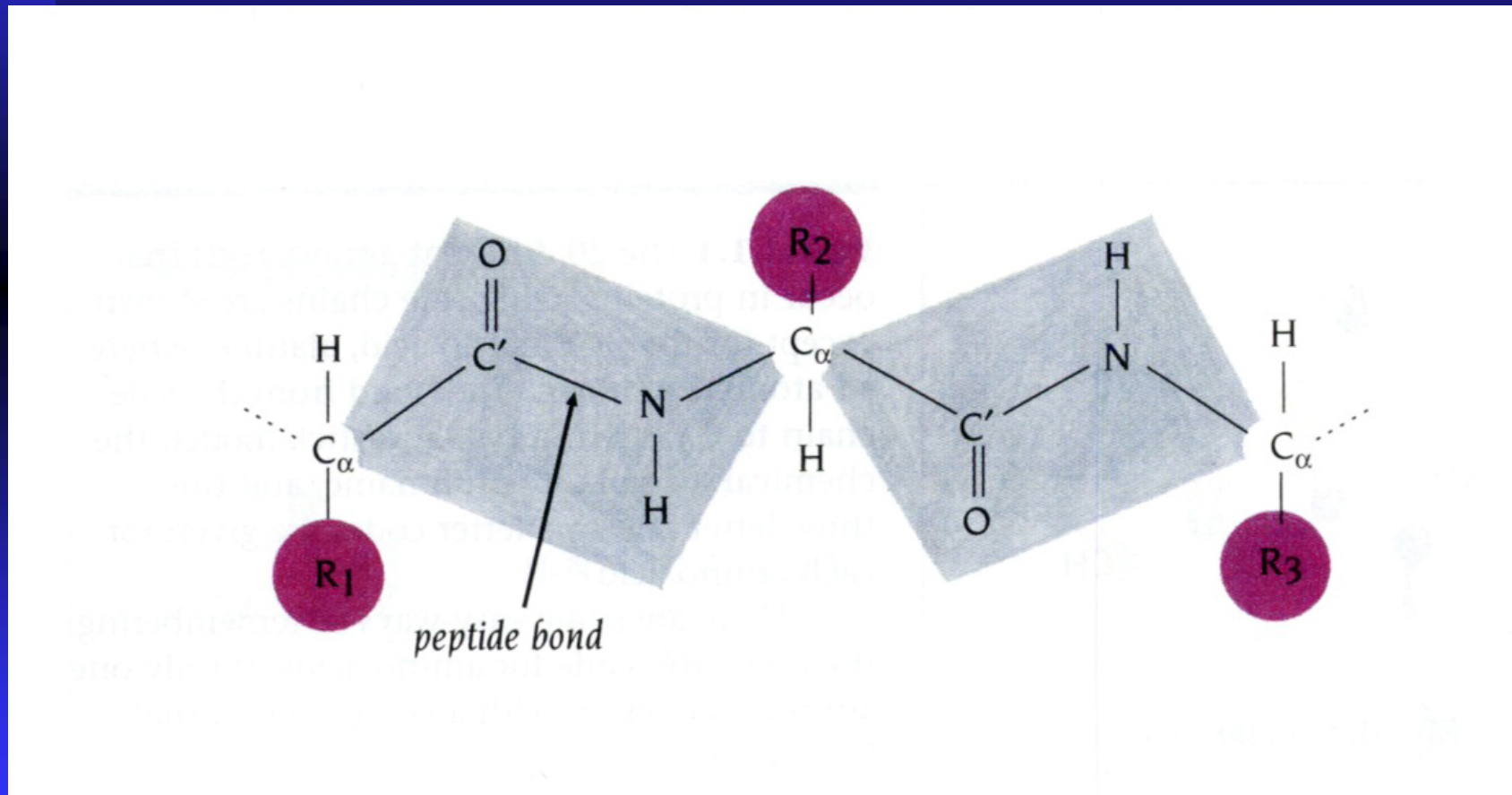
Glycine

(d) Glycine

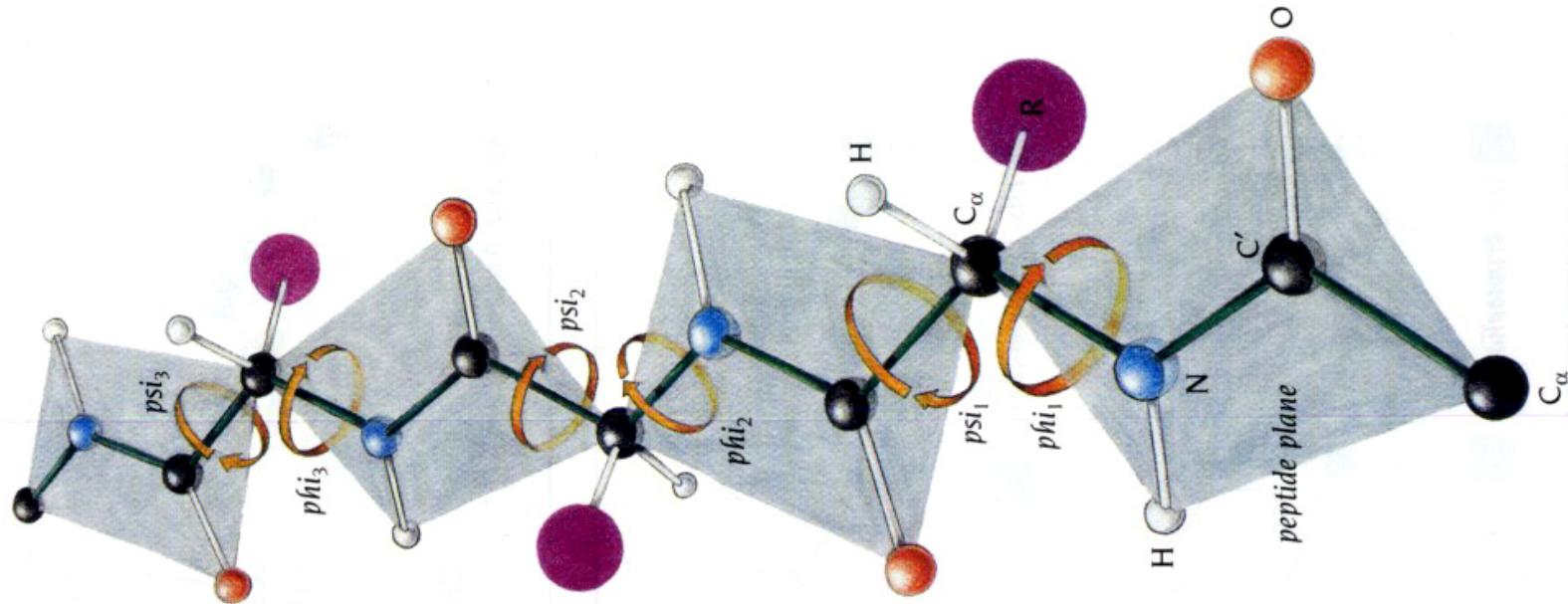


G Gly, Glycine

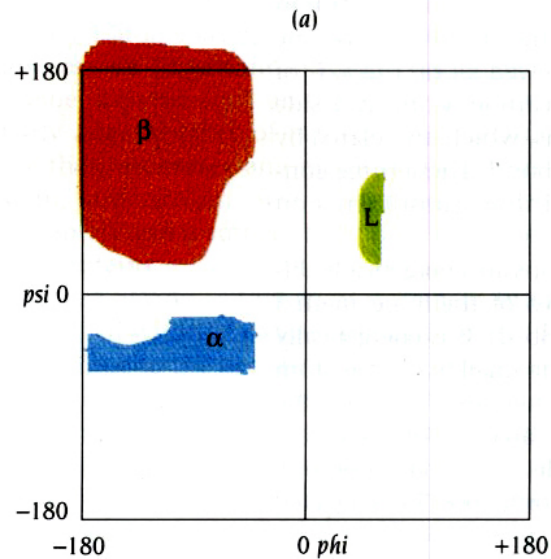
Bond lengths and bond angles of a peptide bond are nearly the same in all peptide bonds in all proteins, and the peptide group is planar



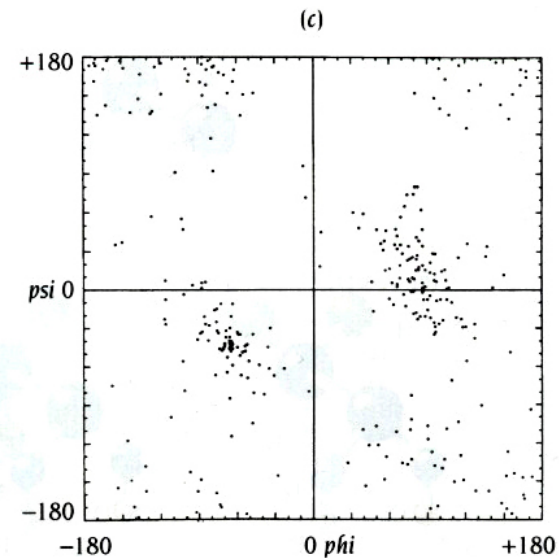
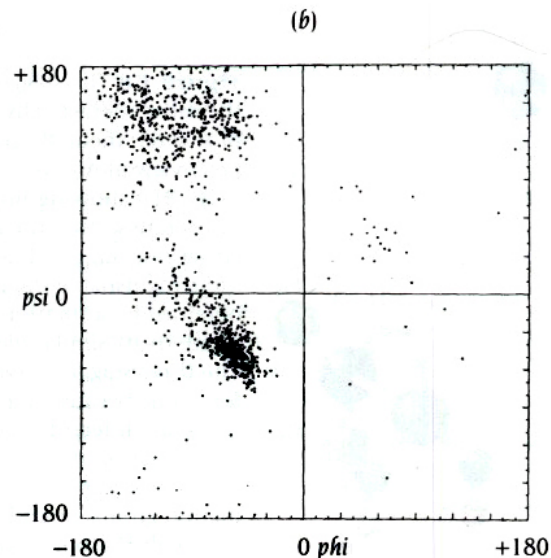
The conformation of the whole main chain of the polypeptide is completely determined when the ϕ and ψ angles for each amino acid are defined



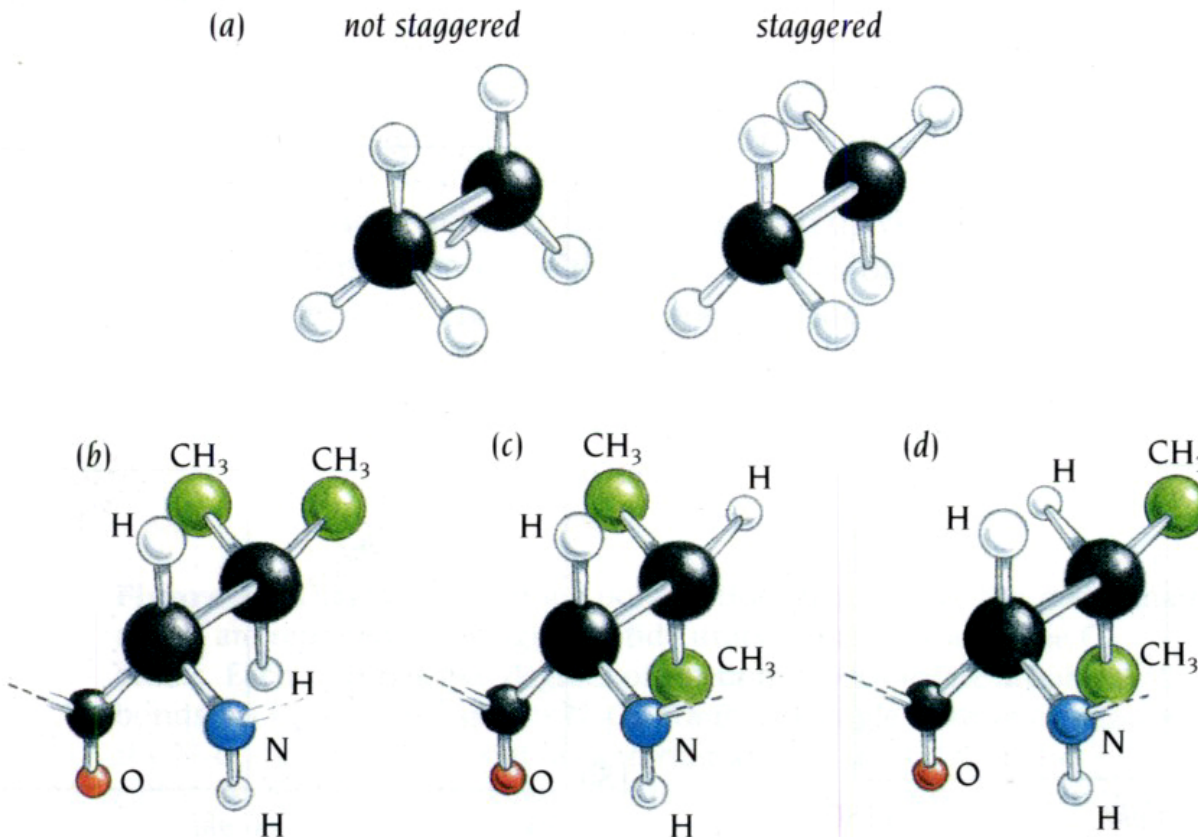
Ramachandran plot



- a) Colored areas show sterically allowed combinations of the ϕ and ψ angles.
- b) Observed values for all residue types except for glycine. Each point represents ϕ and ψ values for an amino acid residue in a well-refined x-ray structure.
- c) Observed values for glycine



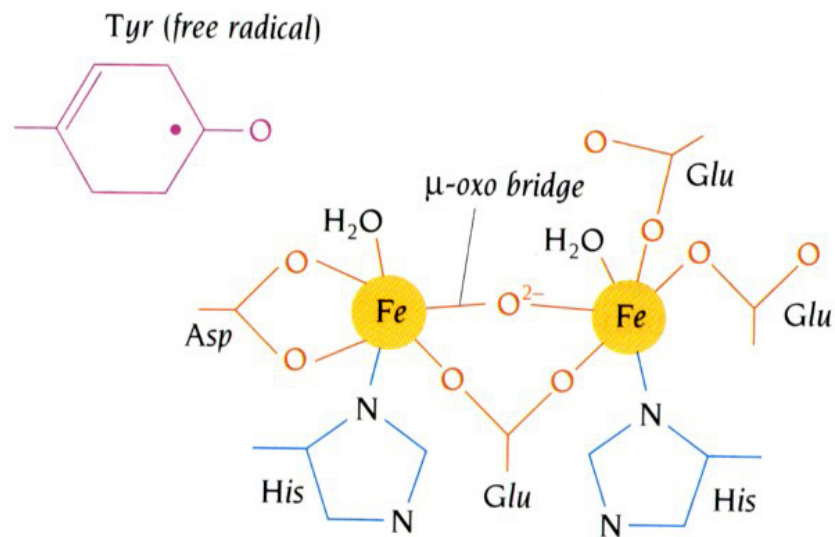
The staggered conformations are the most energetically favored conformations of two tetrahedrally coordinated carbon atoms



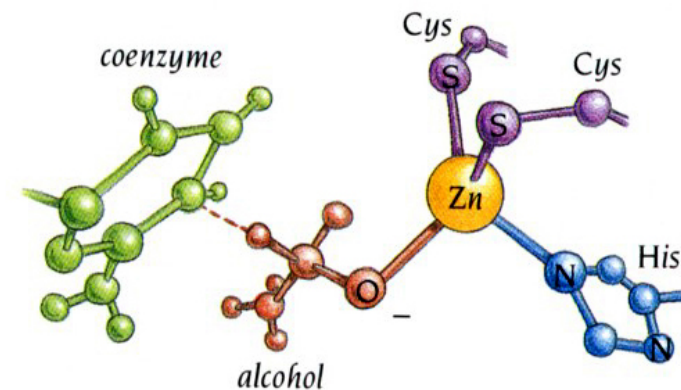
Large number of proteins have recruited metal atoms as intrinsic parts of their structure

- a) the di-iron center of the enzyme ribonucleotide reductase
- b) the catalytically active zinc atom in the enzyme alcohol dehydrogenase

(a)



(b)



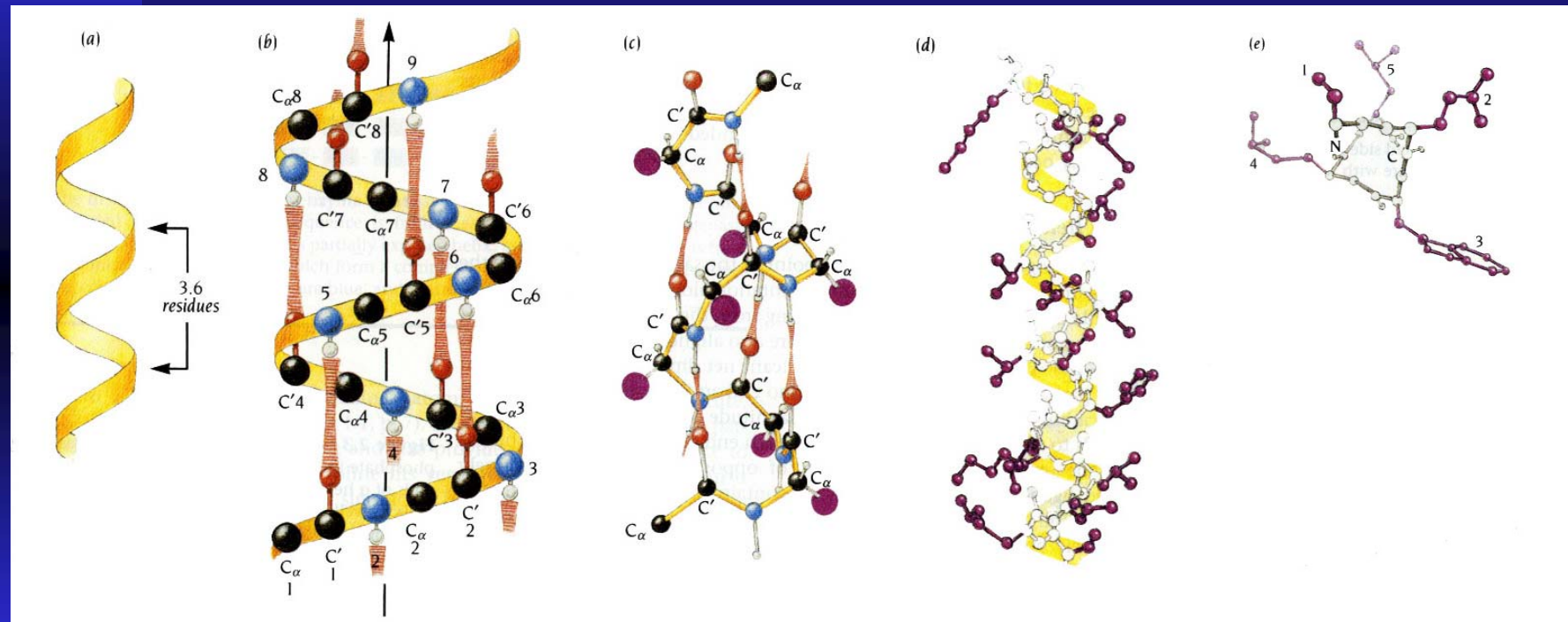
Model of the low-resolution structure of myoglobin



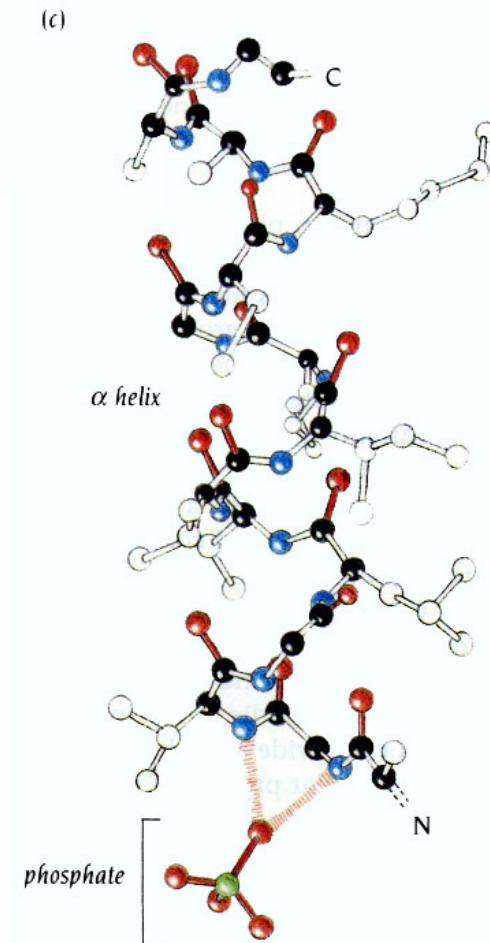
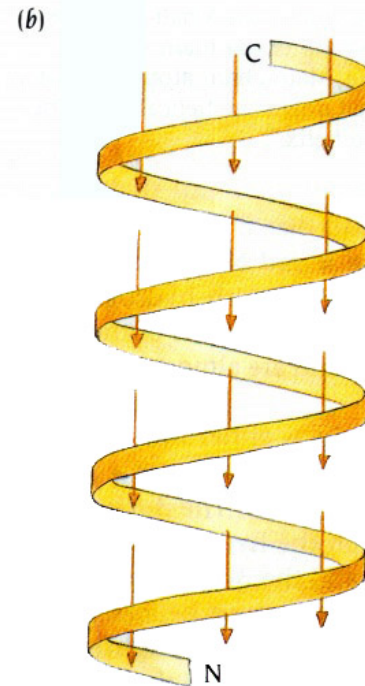
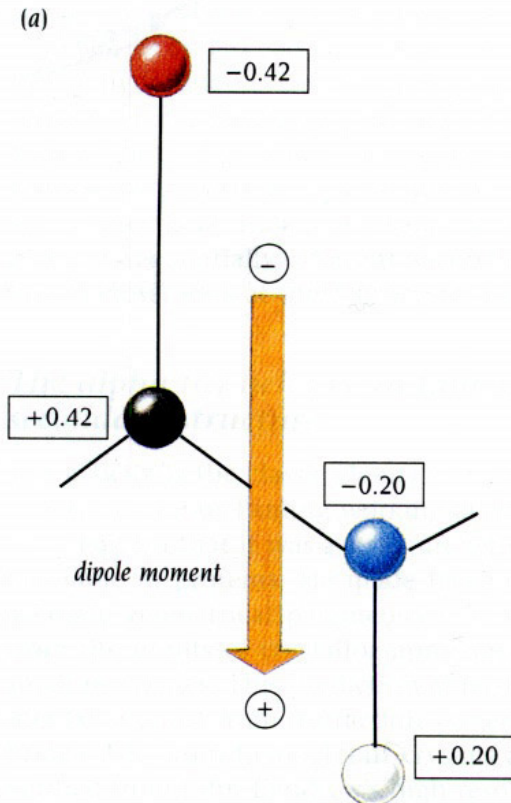
John Kendrew, Cambridge, 1958:

„Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure.“

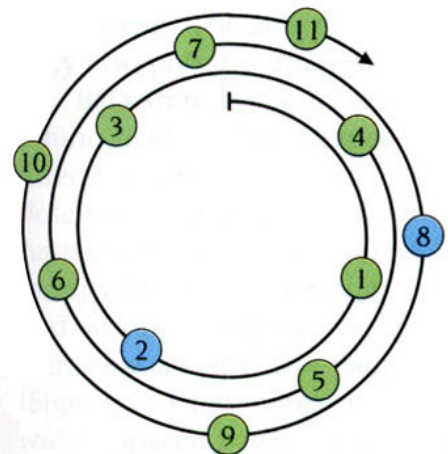
The α -helix – one of the major elements of secondary structure in proteins



The α -helix has a dipole moment

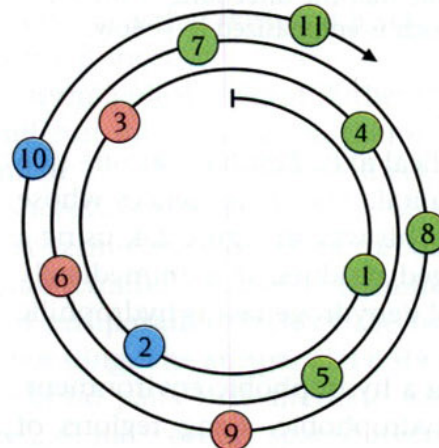


The helical wheel or spiral – a convenient way to illustrate amino acid sequences in helices



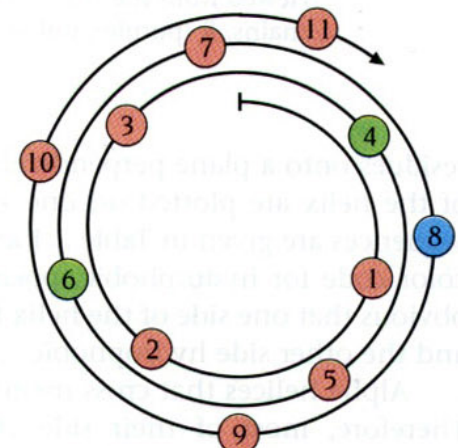
citrate synthase

1 2 3 4 5 6 7 8 9 10 11
L S F A A A M N G L A



alcohol dehydrogenase

1 2 3 4 5 6 7 8 9 10 11
I N E G F D L L R S G



troponin-C

1 2 3 4 5 6 7 8 9 10 11
K E D A K G K S E E E

Examples of amino acid sequences of a totally buried, partially buried, and a completely exposed α helix

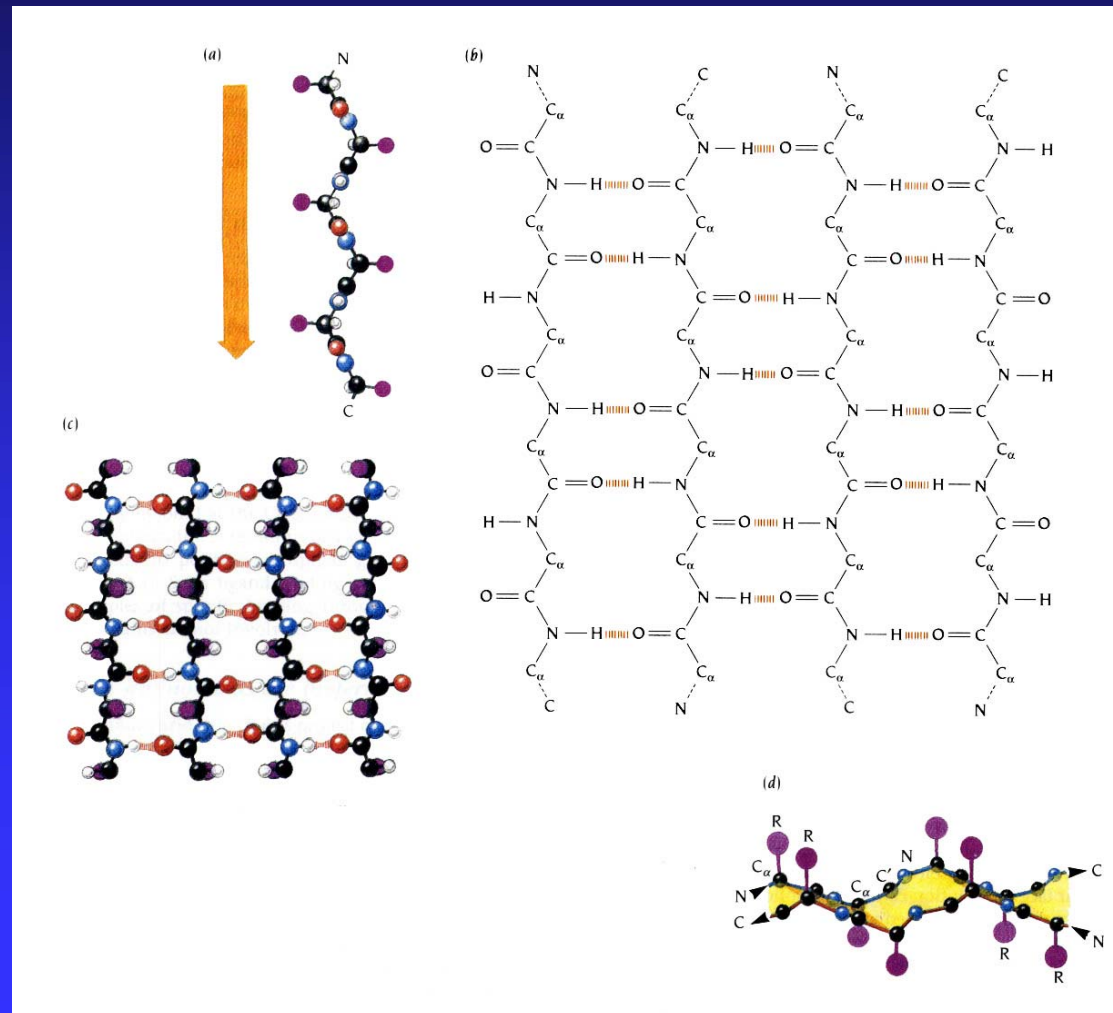
Table 2.1 Amino acid sequences of three α helices

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 -

The first sequence is from the enzyme citrate synthase, residues 260–270, which form a buried helix; the second sequence is from the enzyme alcohol dehydrogenase, residues 355–365, which form a partially exposed helix; and the third sequence is from troponin-C, residues 87–97, which form a completely exposed helix. Charged residues are colored red, polar residues are blue, and hydrophobic residues are green.

The β -sheet – the second major structural element found in globular proteins

a) antiparallel β -sheet



The β -sheet – the second major structural element found in globular proteins

b) parallel β -sheet

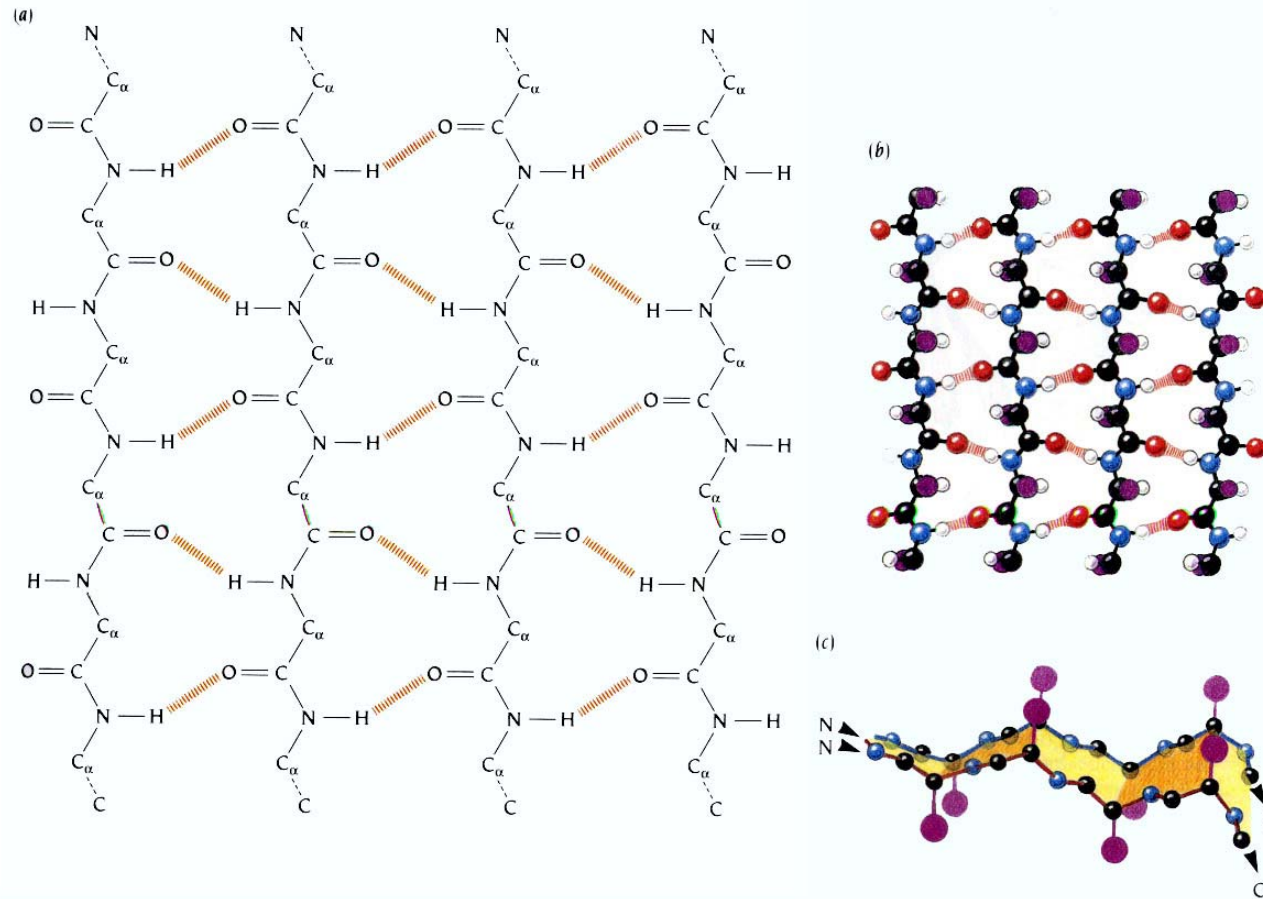
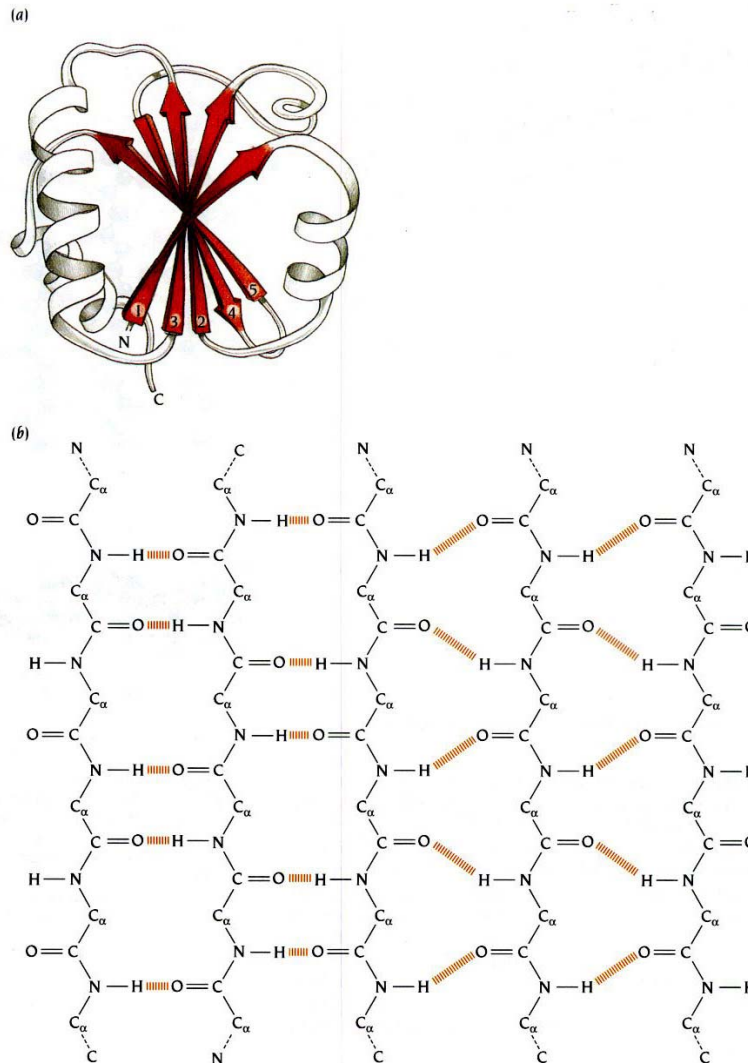
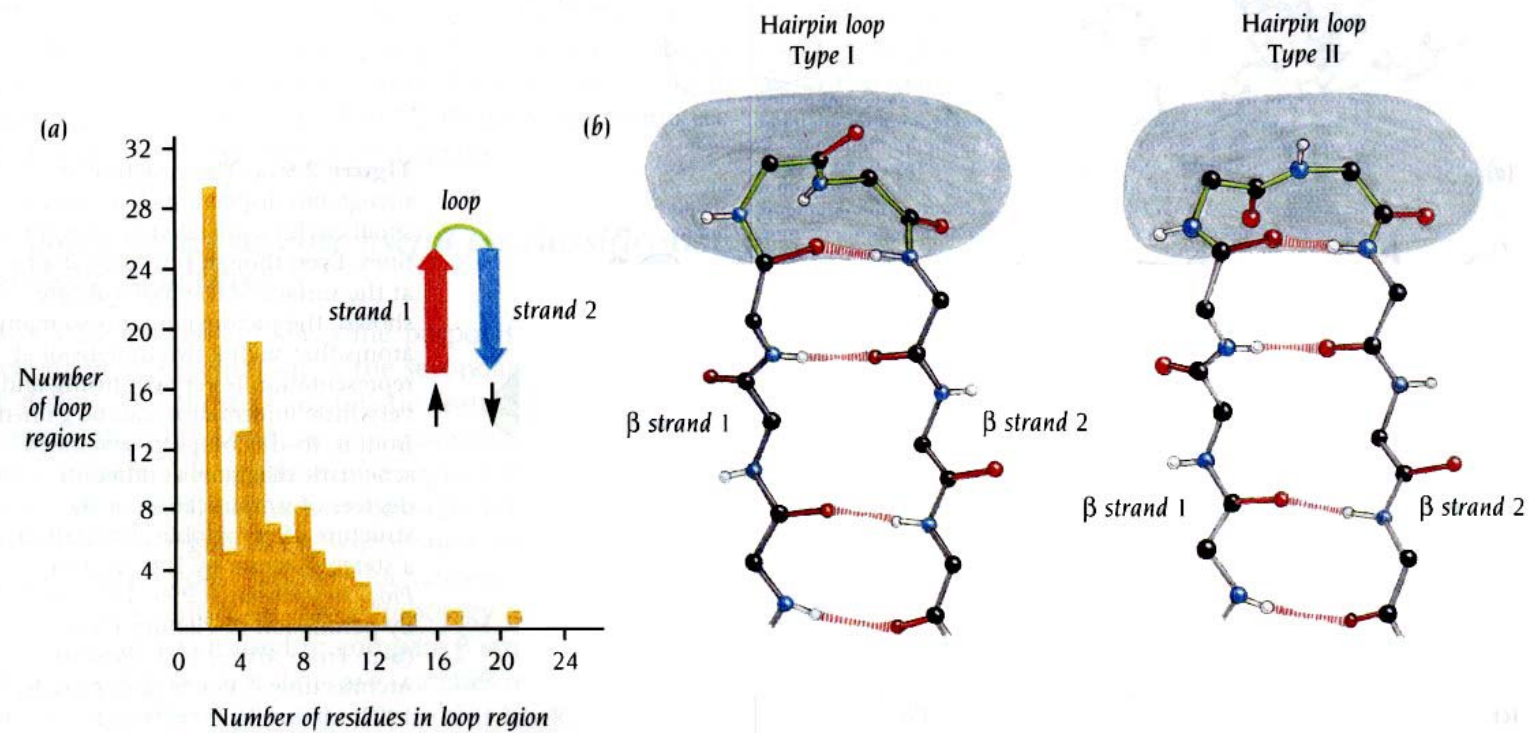


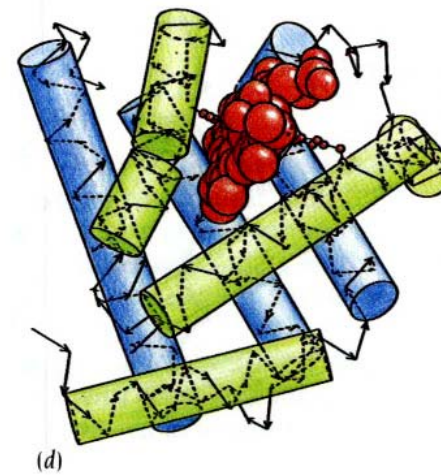
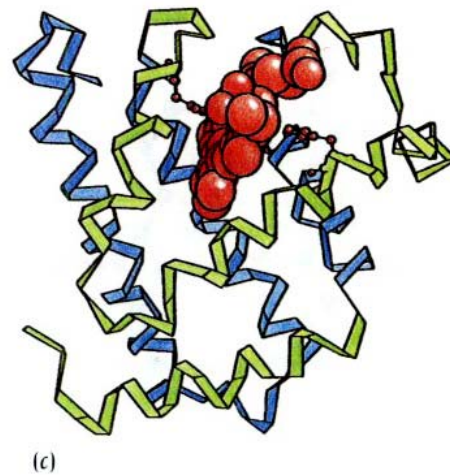
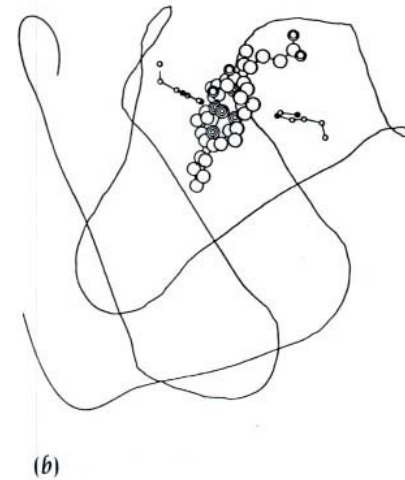
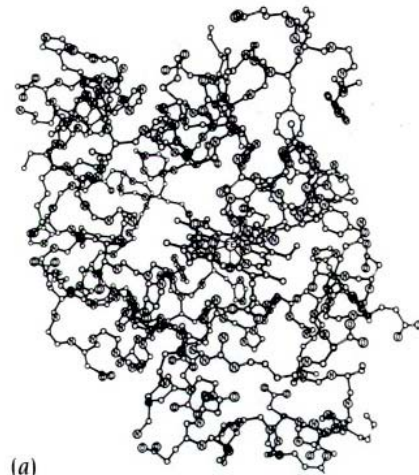
Illustration of the twist of mixed β -sheets found in thioredoxin from *E. coli*



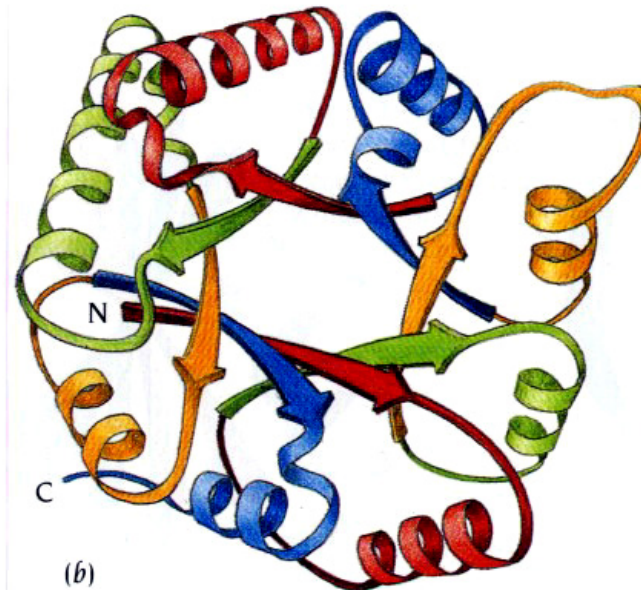
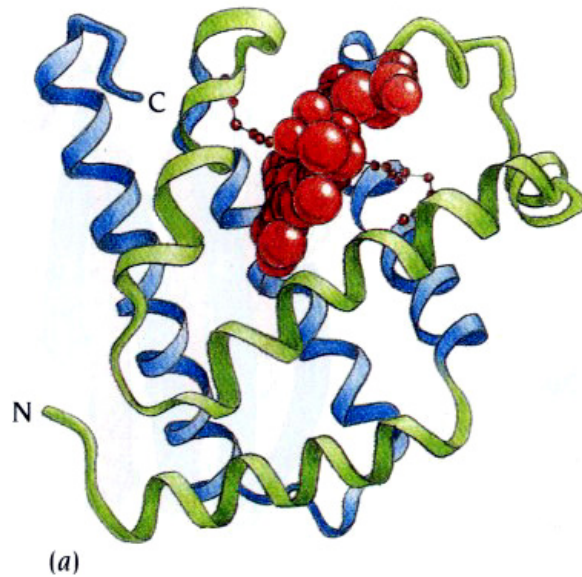
Adjacent antiparallel β -strands are joined by hairpin loops



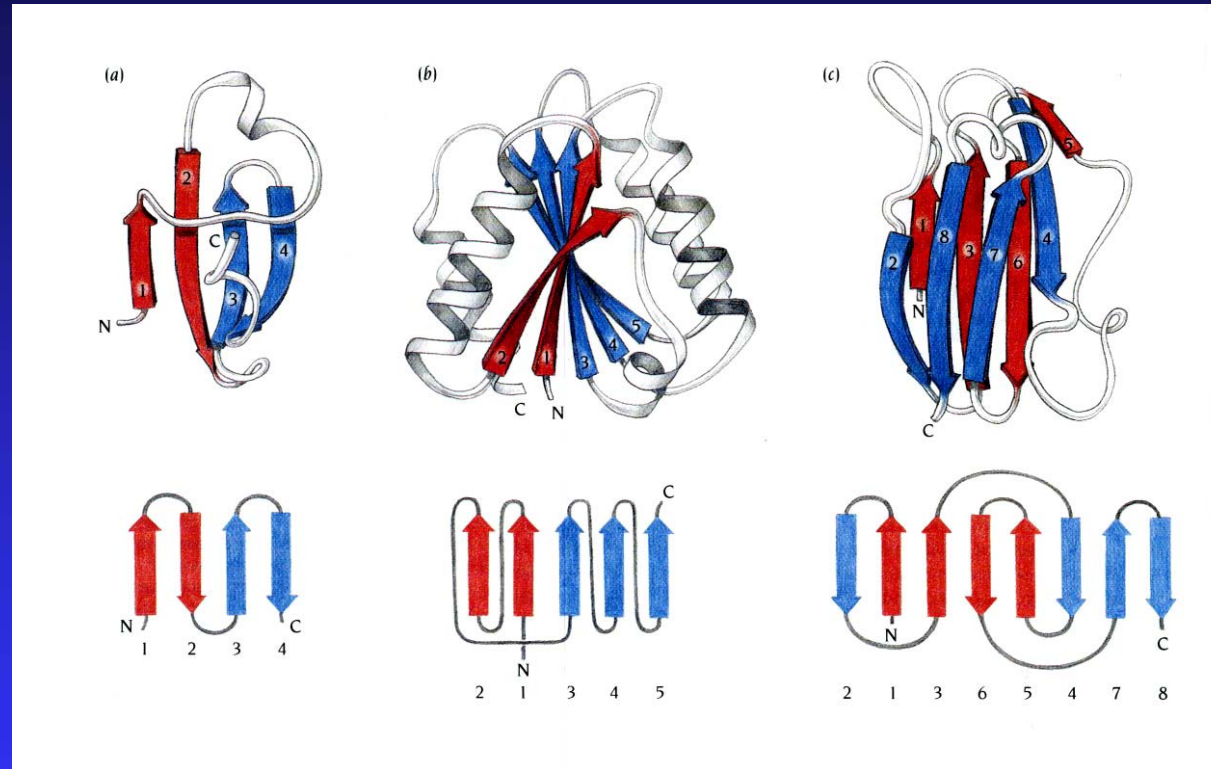
Schematic pictures of proteins highlight secondary structure (b-d – by Arthur Lesk)



Schematic pictures of proteins highlight secondary structure (by Jane Richardson)

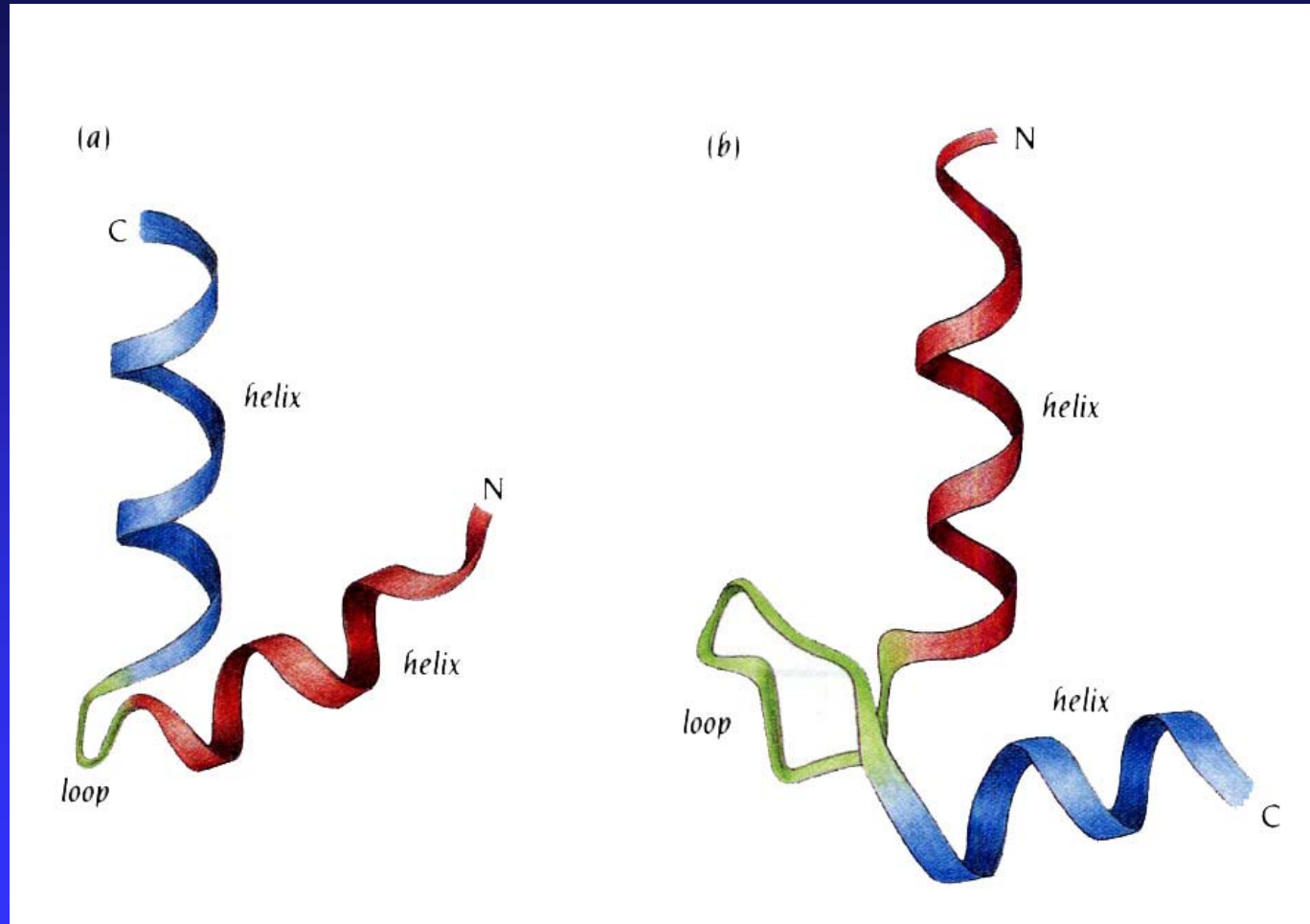


Topology diagrams are often used for presentation and classification of protein structures



- a) Four strands. Antiparallel β sheet in one domain of the enzyme transcarbamoylase.
- b) Five strands. Parallel β sheet in the redox protein flavodoxin.
- c) Eight strands. Antiparallel barrel in the electron carrier plastocyanin.

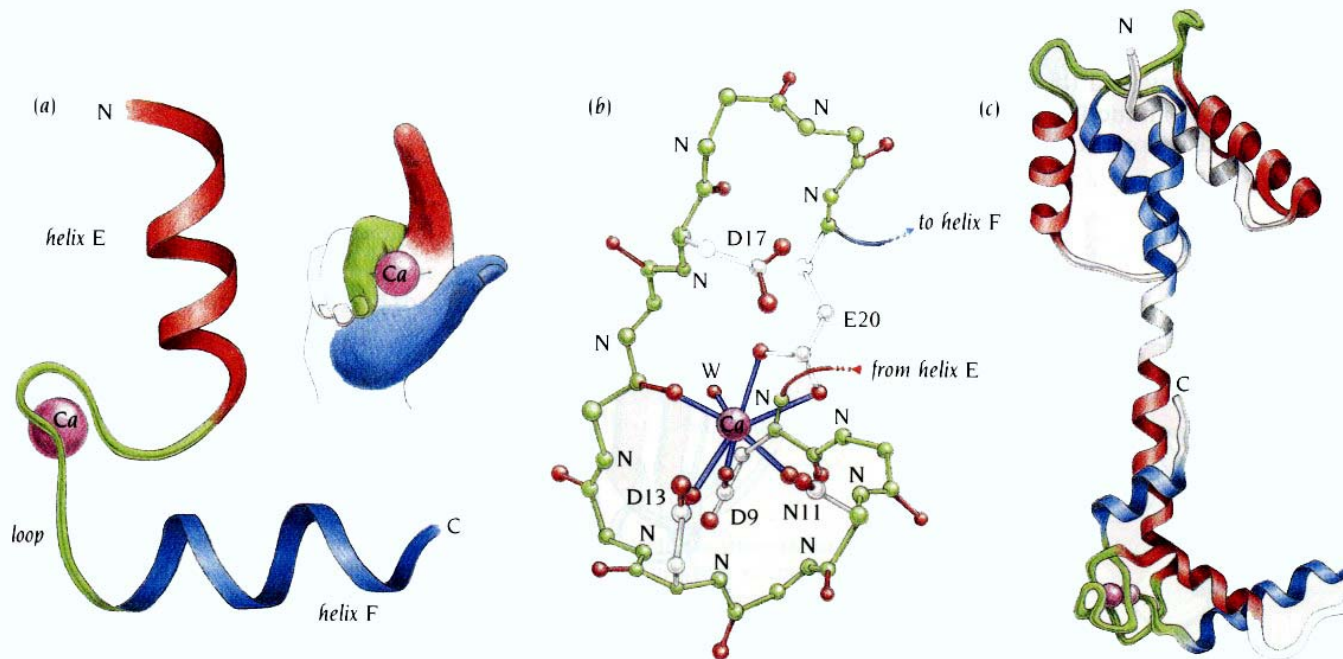
Secondary structure elements are connected to form
simple motifs
Helix-turn-helix motif



a) The DNA-binding motif

b) The calcium-binding motif

Schematic diagrams of the calcium-binding motif



- a) Calcium-binding motif was first characterized as the EF hand motif in parvalbumin.
- b) The calcium atom is bound to one of the motifs in the muscle protein troponin-C through six oxygen atoms.
- c) The structure of troponin-C is built up from four EF motifs.

Set of constraints that an amino acid sequence must conform to in order to form the calcium-binding motif

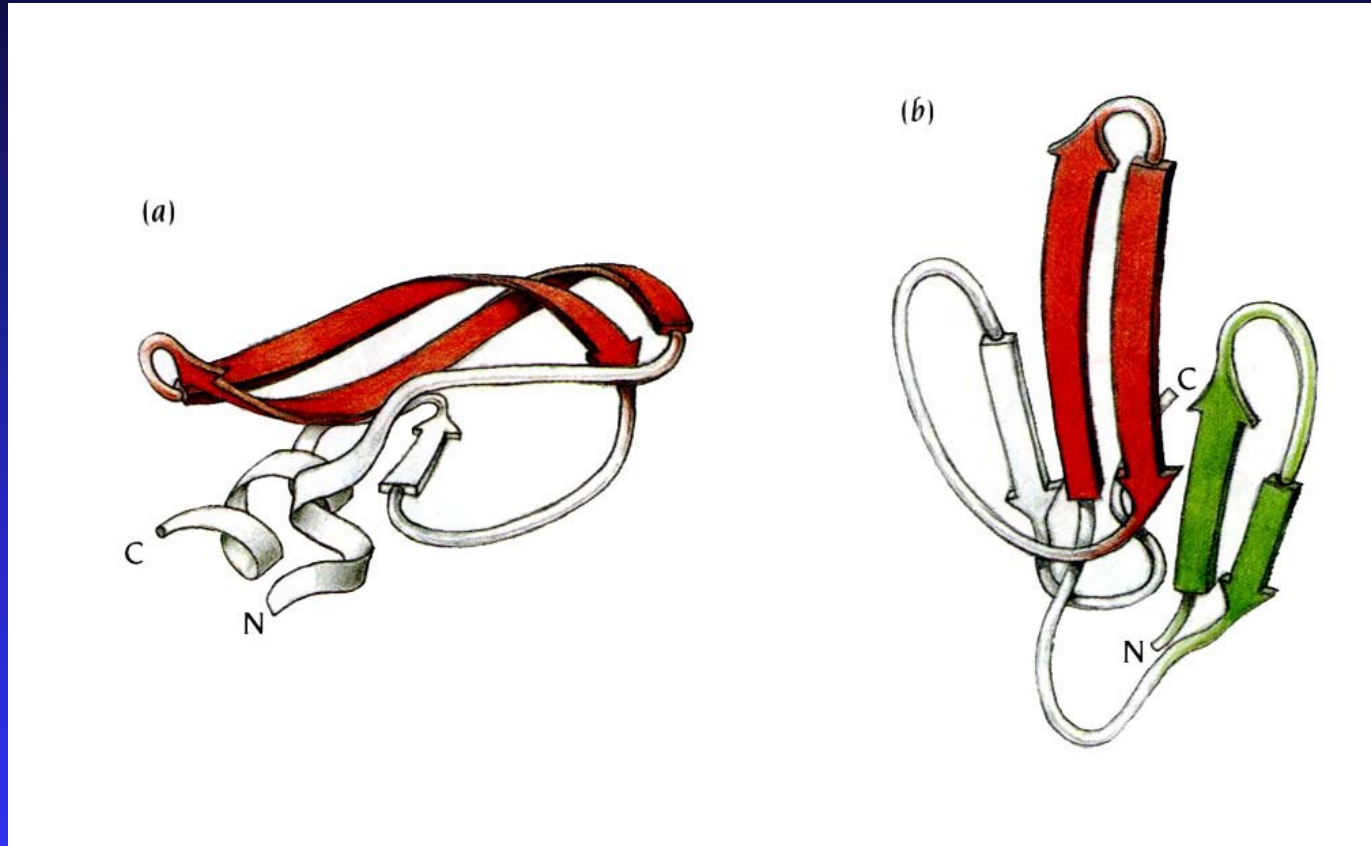
Table 2.2 Amino acid sequences of calcium-binding EF motifs in three different proteins

Parvalbumin	V	K	K	A	F	A	I	I	D	Q	D	K	S	G	F	I	E	E	D	E	L	K	L	F	L	Q	N	F
Calmodulin	F	K	E	A	F	S	L	F	D	K	D	G	D	G	T	I	T	T	K	E	L	G	T	V	M	R	S	L
Troponin-C	L	A	D	C	F	R	I	F	D	K	N	A	D	G	F	I	D	I	E	E	L	G	E	I	L	R	A	T



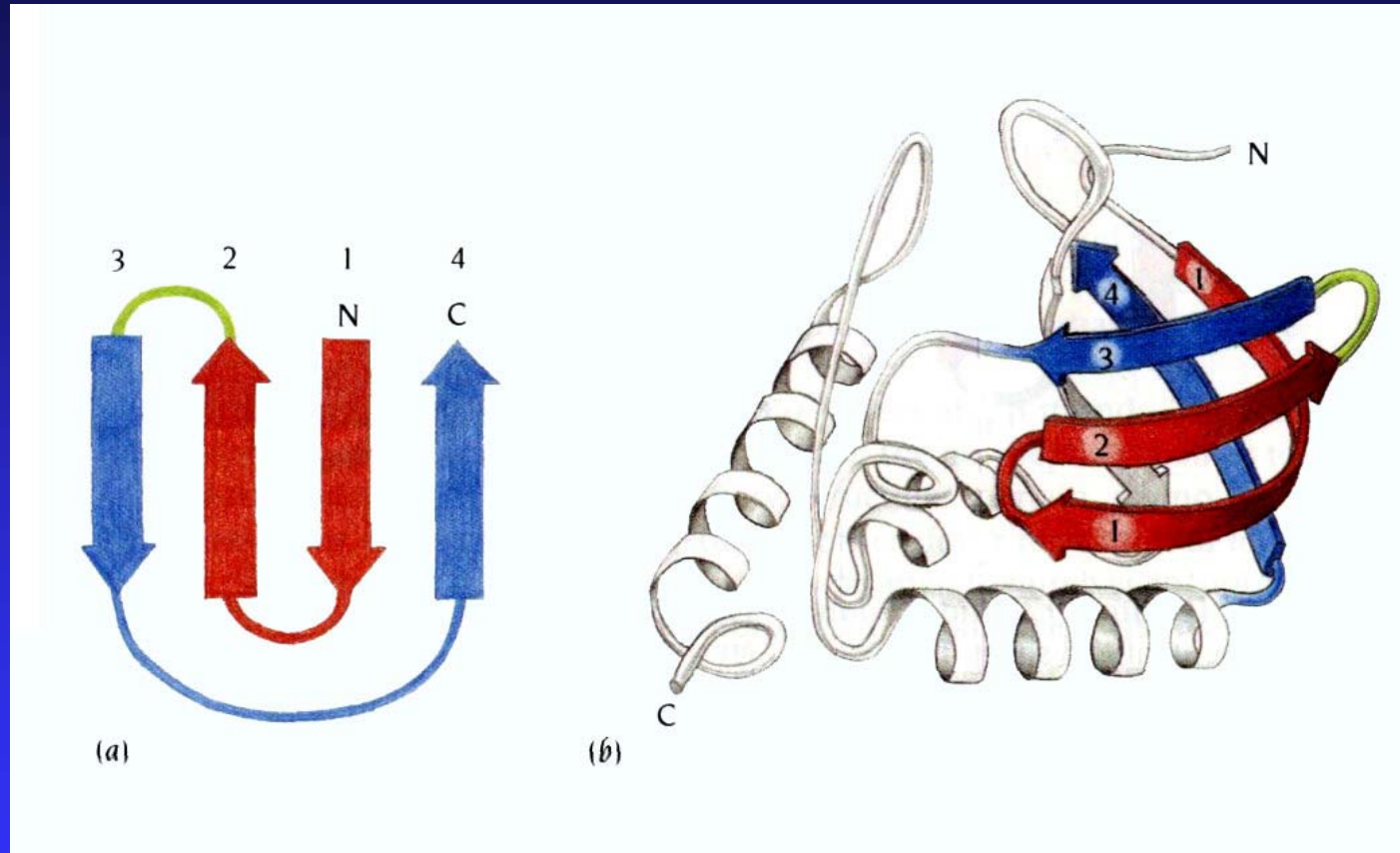
Calcium-binding residues are orange, and residues that form the hydrophobic core of the motif are light green. The helix-loop-helix region shown underneath is colored as in Figure 2.13.

The hairpin β motif occurs frequently in protein structures



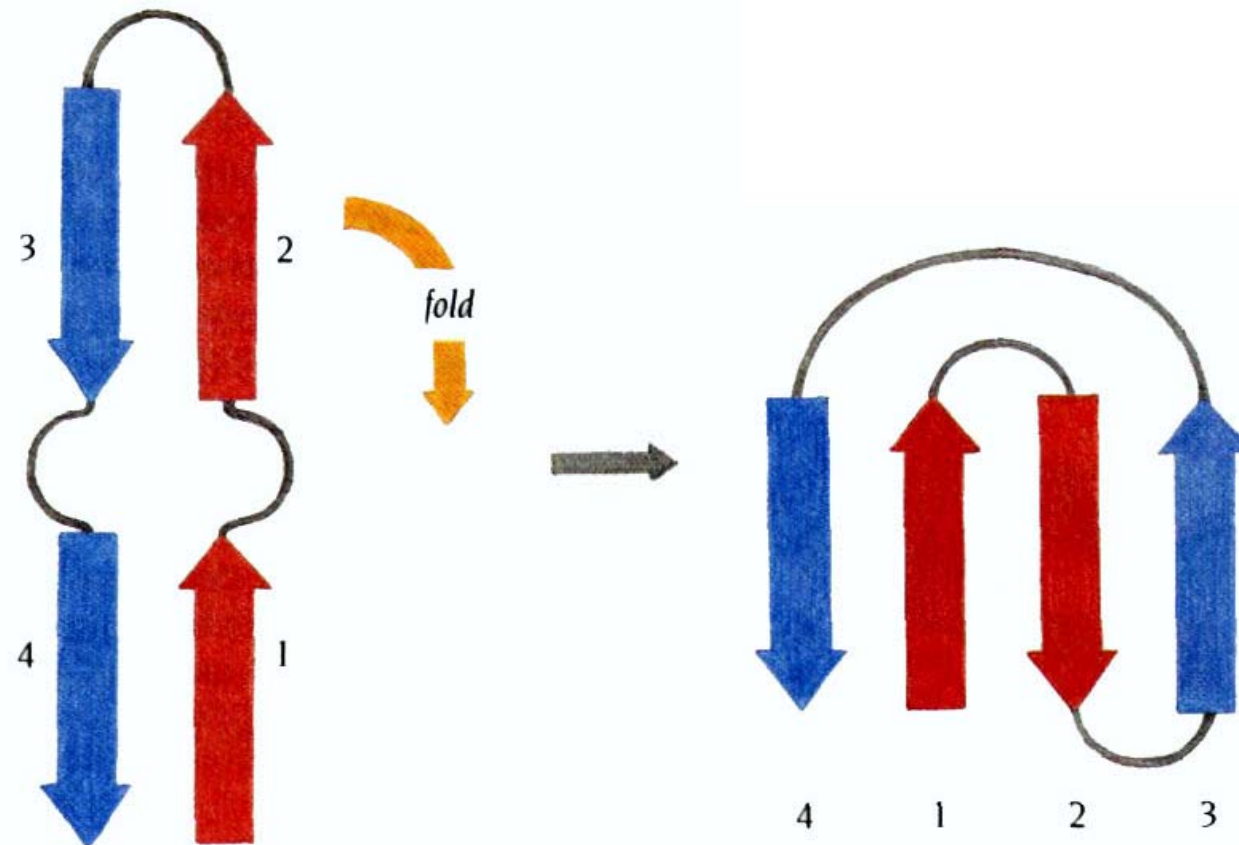
- a) Schematic diagram of the structure of bovine trypsin inhibitor. The hairpin motif is colored red.
- b) Schematic diagram of the structure of the snake venom erabutoxin. The two hairpin motifs within the β sheet are colored red and green.

The Greek key motif is found in antiparallel β sheets when four adjacent β strands are arranged in the pattern shown as a topology diagram in (a)

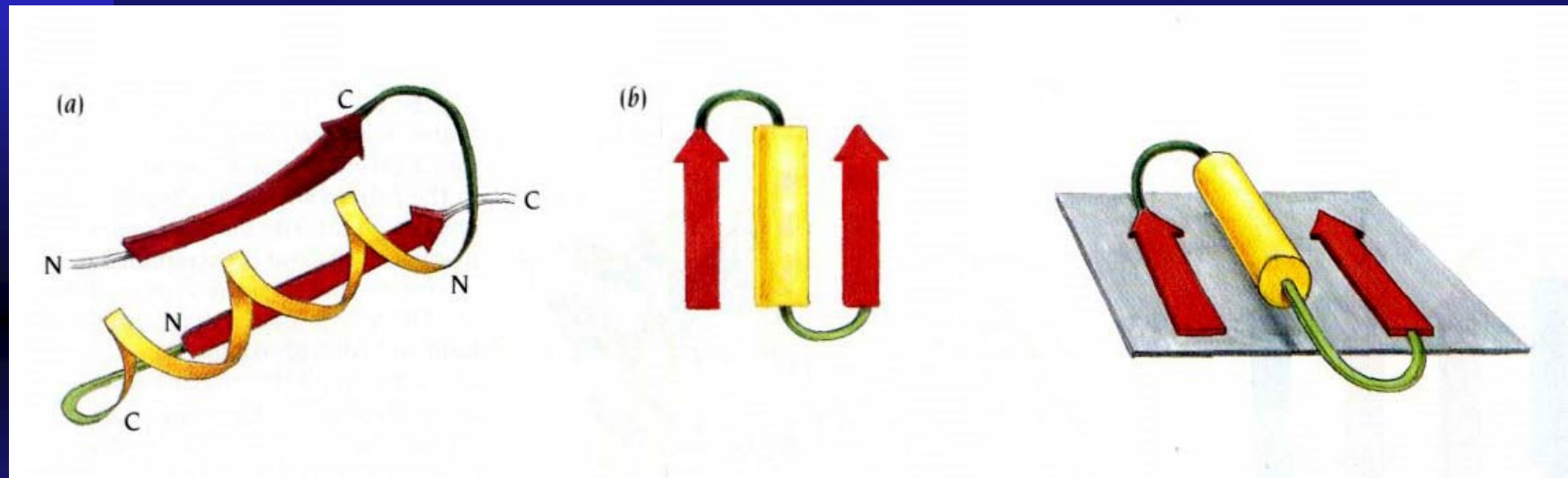


- a) The motif occurs in many β sheets and is exemplified here by the enzyme *Staphylococcus* nuclease.
- b) The four β strands that form this motif are colored red and blue.

Suggested folding pathway from a hairpin-like structure to the Greek key motif

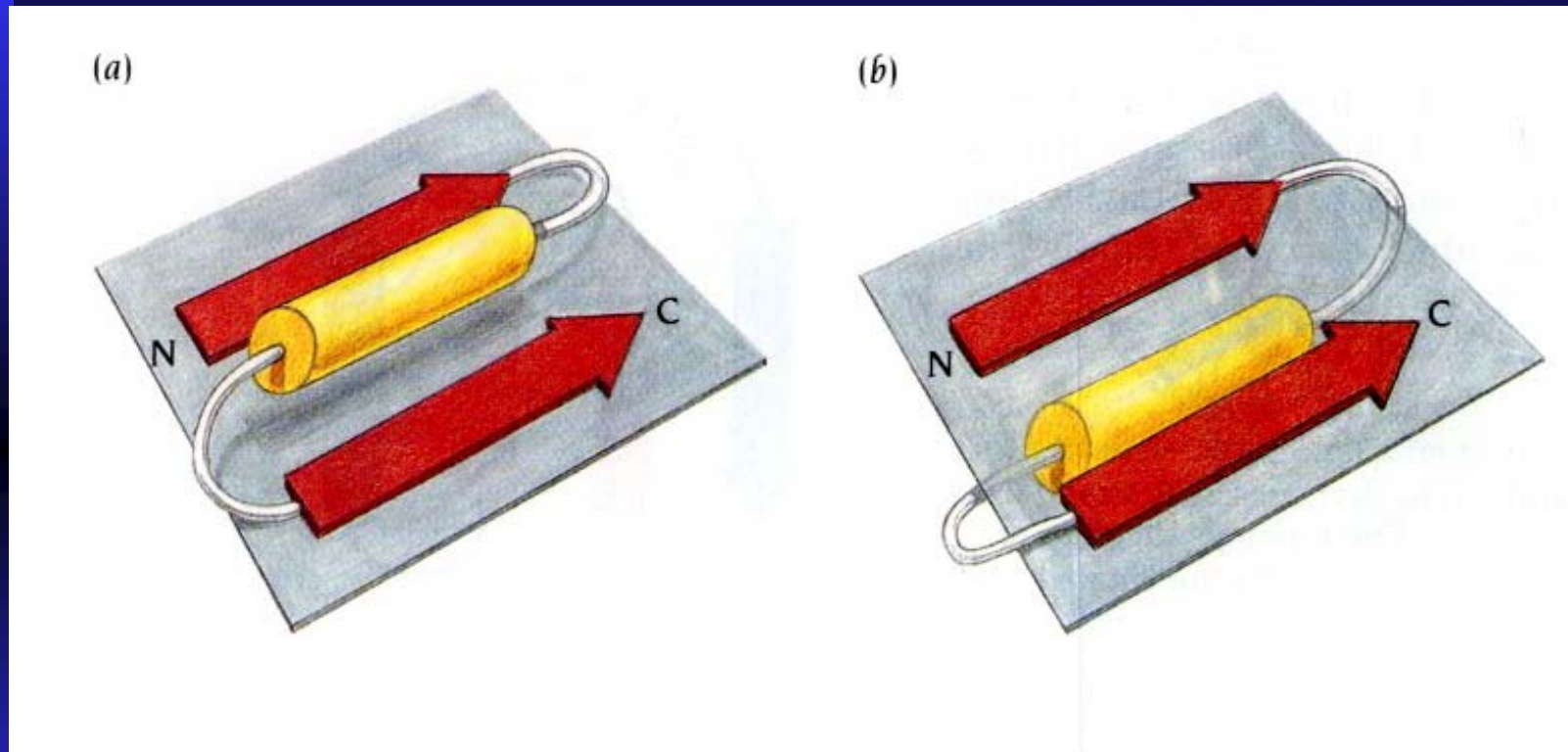


The β - α - β motif contains two parallel β strands



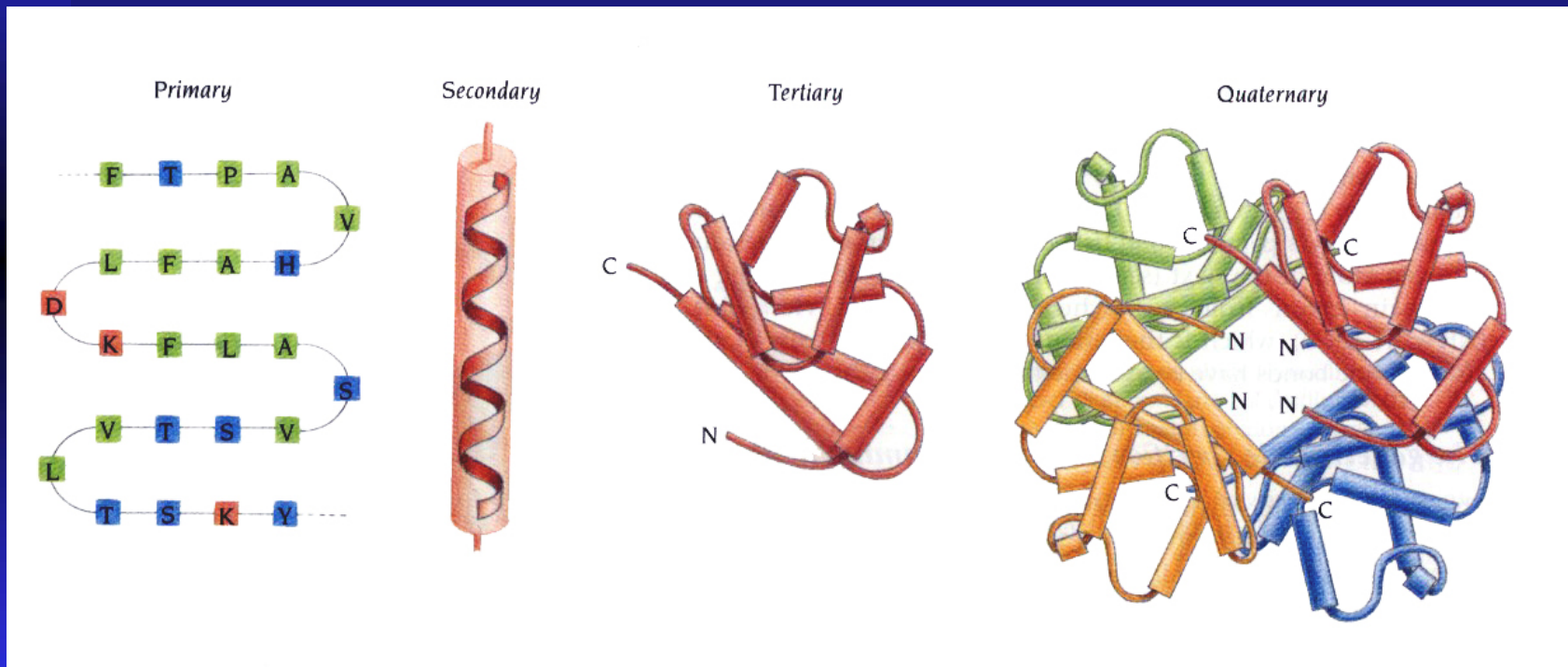
Two adjacent parallel β strands are usually connected by an α helix from the C-terminus of strand 1 to the N-terminus of strand 2. Most protein structures that contain parallel β sheets are built up from combinations of such β - α - β motifs.

The β - α - β motif can in principle have two „hands“

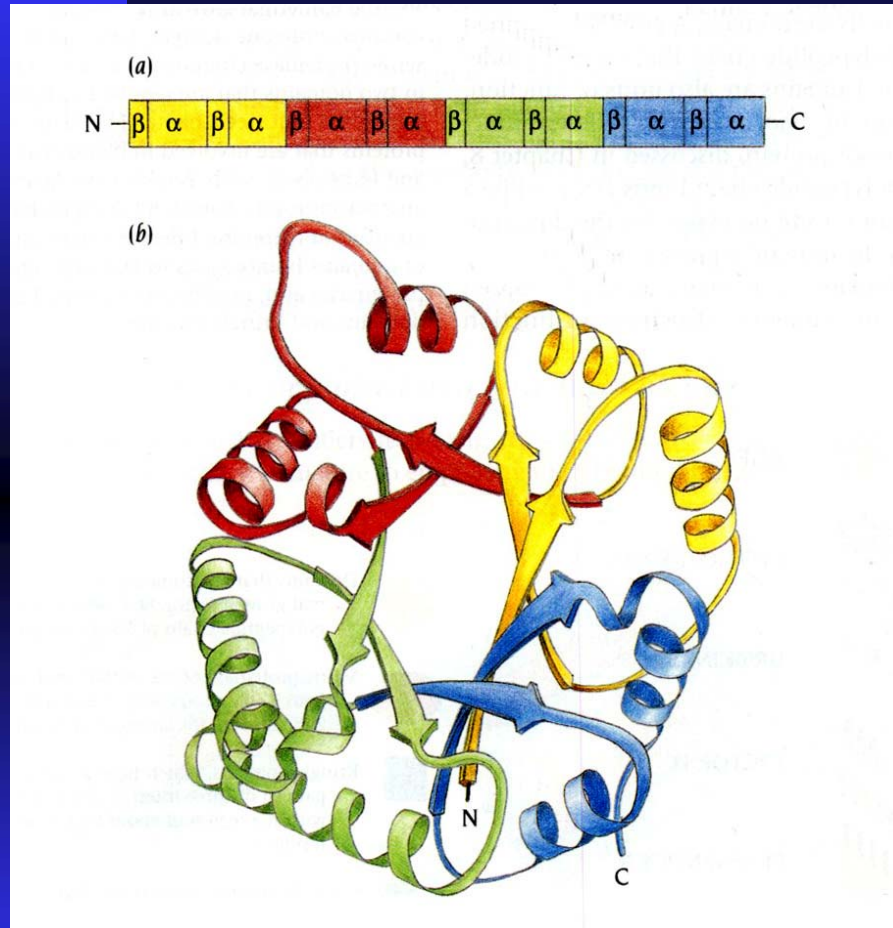


- a) This connection with the helix above the sheet is found in almost all proteins and is called right-handed because it has the same hand as a right-handed α helix.
- b) The left-handed connection with the helix below the sheet.

Protein molecules are organized in a structural hierarchy



Domains are build from structural motifs



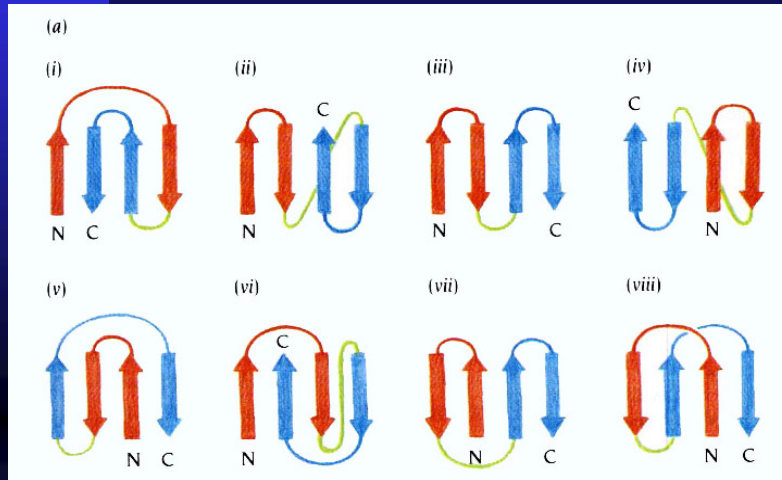
To a first approximation a polypeptide chain can be considered as a sequential arrangement of simple motifs that are formed from consecutive regions of the primary structure. For example, triosephosphate isomerase is built up from four β - α - β - α motifs that are consecutive both in the amino acid sequence (a) and in the three-dimensional structure (b).

The number of such combinations found in proteins is limited, and some combinations seem to be structurally favored.

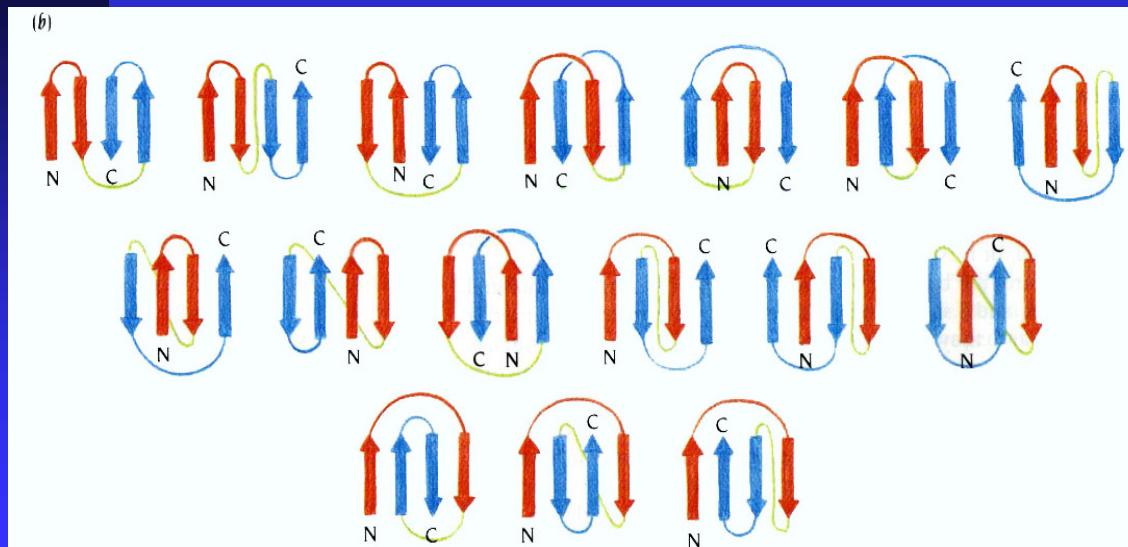
Thus similar domain structures frequently occur in different proteins with different functions and with completely different amino acid sequences.

Simple motifs combine to form complex motifs

Two sequentially adjacent hairpin motifs can be arranged in 24 different ways into a β sheet of four strands.

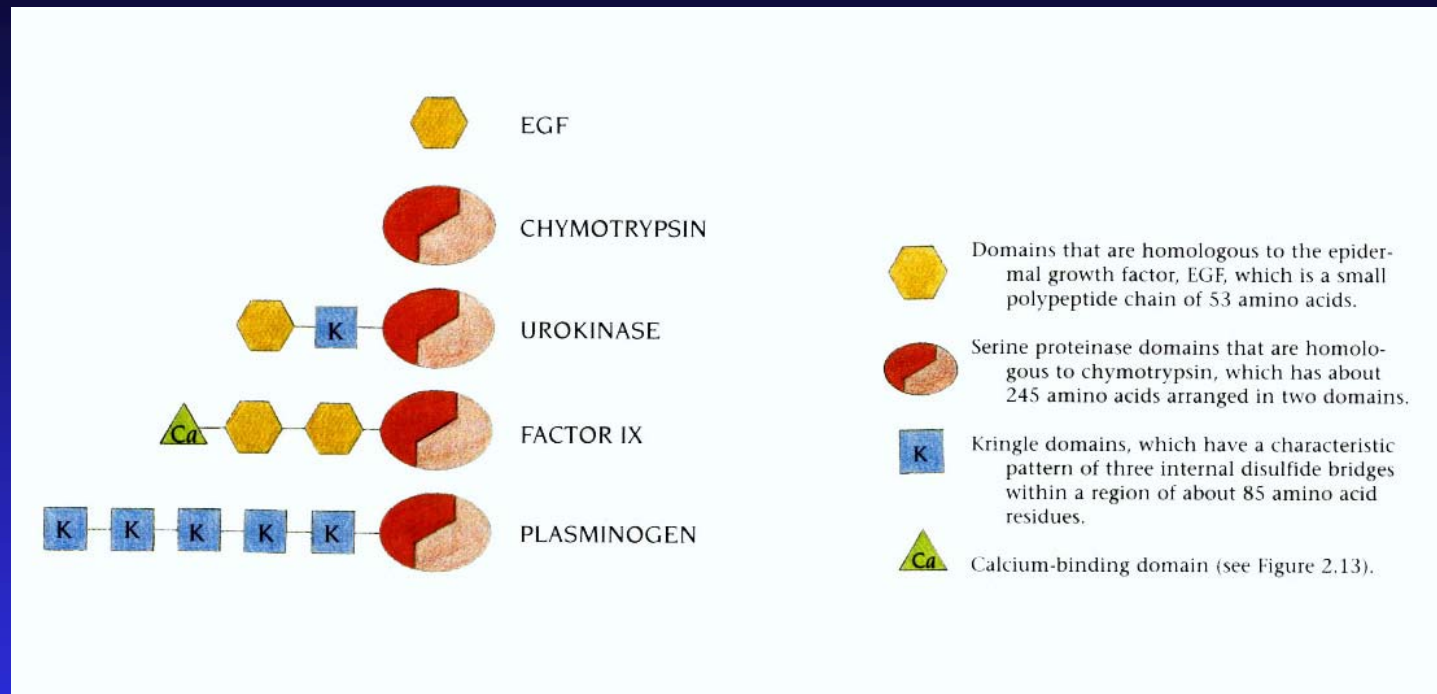


a) Topology diagrams for those arrangements that were found in a survey of all known structures in 1991. The Greek key motifs in (i) and (v) occurred 74 times, whereas the arrangements shown in (vii) occurred only once.



b) Topology diagrams for those 16 arrangements that did not occur in any structure known at that time. Most of these arrangements contain a pair of adjacent parallel β strands.

Large polypeptide chains fold into several domains



Small protein molecules like the epidermal growth factor, EGF, comprise of only one domain. Others, like the serine proteinase chymotrypsin, are arranged in two domains that are required to form a functional unit. Many of the proteins that are involved in blood coagulation and fibrinolysis, such as urokinase, factor IX, and plasminogen, have long polypeptide chains that comprise different combinations of domains homologous to EGF and serine proteinases and, in addition, calcium-binding domains and Kringle domains.

Protein structures can be divided into three main classes

α structures

the core is built up exclusively from α helices

β structures

the core comprises antiparallel β sheets and are usually two β sheets packed against each other

α/β structures

are made from combinations of β - α - β motifs that form a predominantly parallel β sheet surrounded by α helices

Minor groups

- proteins built up from a combination of discrete α and β motifs and usually form one small antiparallel β sheet in one part of the domain packed against a number of α helices
- small proteins rich in disulfide bonds or metal atoms; the structures of these proteins seem to be strongly influenced by the presence of these metals or disulfides and often look like distorted versions of more regular proteins.