Informatica Biomedica

lezione17

Alberto Paoluzzi Mauro Ceccanti
www.dia.uniroma3.it/ paoluzzi/web/did/biomed/

Informatica e Automazione, "Roma Tre" — Medicina Clinica, "La Sapienza"

May 17, 2010

Fonte essenziale: J- Gu amp; P.E- Bourne, *Structural Bioinformatics*, Wiley (2009)

Informatica Biomedica: Lezione 17

Fundamentals of Protein Structure
Primary Structure - Amino Acid Sequence
The Peptide Bond
Secondary Structure is Local 3D Structure
α Helices
β Sheets
Other Secondary Structure
Representations of protein structure
Tertiary Structure - Global 3D Structure
Side Chains and Tertiary Structure
Domains and Motifs
Biochemical Classification of Folds
Structural Classification of Folds
Quaternary Structure - Associations of Multiple Polypeptide Chains
Functional Relevance of Quaternary Structure

The Importance of Protein Structure

- Most of the essential structure and function of cells is mediated by proteins.
- These large, complex molecules exhibit a remarkable versatility that allows them to perform a myriad of activities that are fundamental to life.
- No other type of biological macromolecule could possibly assume all of the functions that proteins have amassed over billions of years of evolution.

Protein structure leads to protein function

Fundamental principle:

- The distinctive structures of proteins allow for the placement of particular chemical groups in specific places in three-dimensional space.
- It is this precision that allows proteins to act as catalysts (enzymes) for an impressive variety of chemical reactions.
- Precise placement of chemical groups also allows proteins to play important structural, transport, and regulatory functions in organisms.
- Further, the functional diversity of proteins is expanded through the interaction of proteins with small molecules, as well as other proteins.

After discovery of structure of *myoglobin*

Lack of regularities

> *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 predicated by any theory of protein structure.*

Discovery of myoglobin (1958)
– Kendrew et al.

Emerging regularities in protein structures

Despite these initial frustrations, subsequent studies of the myoglobin structure based on higher-quality data revealed that the protein did have some regularities; these regularities were also observed in other protein structures.

Decades of research have now yielded a coherent set of principles about the nature of protein structure and the way in which this structure is utilized to effect function-

Four-tiered hierarchy

These principles have been organized into a four-tiered hierarchy that facilitates description and understanding of proteins:

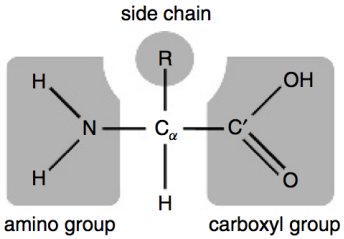
- primary structure
- secondary structure
- tertiary structure
- quaternary structure-

This hierarchy does not seek to describe precisely the physical laws that produce protein structure, but rather is an abstraction to make protein structural studies more tractable.

The structure of a prototypical amino acid

Proteins are linear polymers of amino acids

- it is the distinct sequence of component amino acids that determines the ultimate three-dimensional structure of the protein.
- The chemical groups bound to the central α-carbon are highlighted in gray
- The R-group represents any of the possible 20 amino acid side chains.

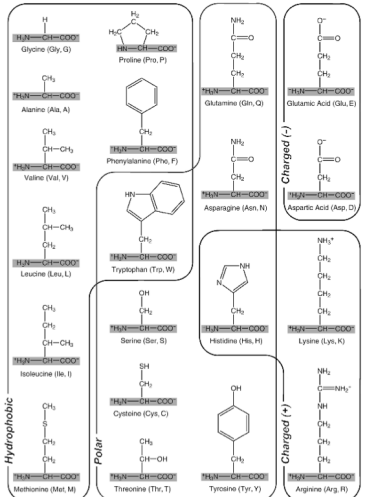


Amino acids classification

The 20 standard amino acids can be loosely grouped into classes based on the chemical properties conferred by their side chains

Three classes are commonly accepted:

- hydrophobic
- polar
- charged



Amino acids classification

- hydrophobic
- polar
- charged

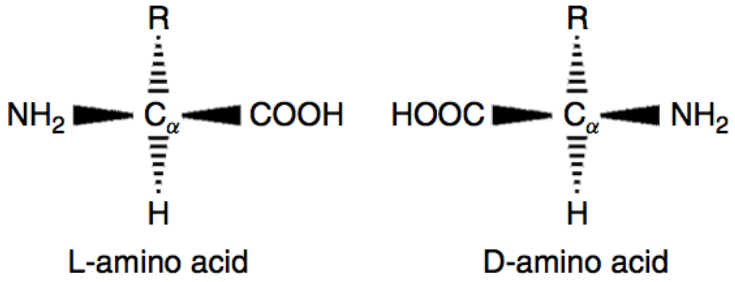
This classification groups amino acids based on the form that predominates at physiological conditions (note that their amino and carboxyl groups are charged under these conditions)-

This classification is useful as a guideline, but does not convey the full complexity of side chain properties.

The Peptide Bond

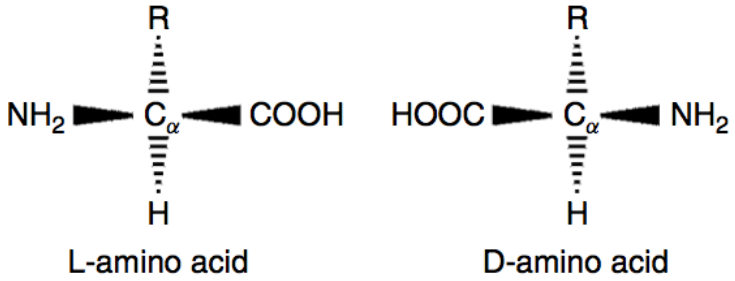
- Amino acids can form bonds with each other through a reaction of their respective carboxyl and amino groups-
- The resulting bond is called the peptide bond, and two or more amino acids linked by such a bond are referred to as a peptide
- The atoms involved in the peptide bond are referred to as the peptide backbone

Stereoisomers of a prototypical amino acid (1/2)



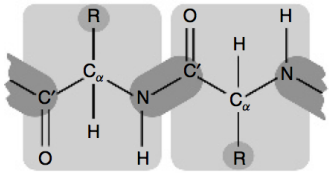
- These structures are mirror images of each other
- The L-form is the only type incorporated into proteins via the genetic machinery.

Stereoisomers of a prototypical amino acid (2/2)



- The R-group represents any of the possible 20 amino acid side chains

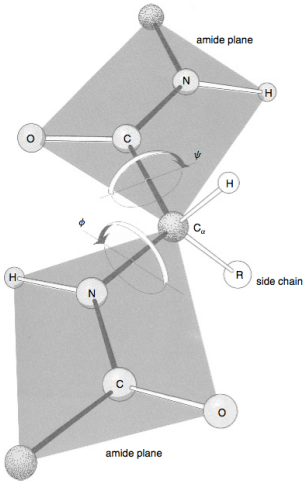
The peptide bond



- Two peptide units (amino acid residues) are shown shaded in light gray
- The peptide bond between them is shaded in dark gray

- The specific characteristics of the peptide bond have important implications for the three-dimensional structures that can be formed by polypeptides-
- The peptide bond is planar and quite rigid-
- Therefore, the polypeptide chain has rotational freedom only about the bonds formed by the α-carbons-

Rotation of the peptide backbone about the Cα atom.



- Rotation is only possible about the (Cα–N) and (Cα–C') angles.
- Arrows about the two angles show the positive rotation.

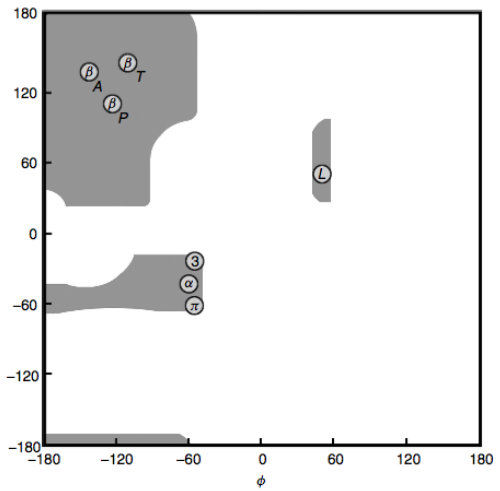
Secondary Structure is local 3D Structure

The secondary structure of a protein can be thought of as the local conformation of the polypeptide chain, independent of the rest of the protein.

During the course of protein structure research, two types of secondary structure have emerged as the dominant local conformations of polypeptide chains:

- alpha (α) helix
- beta (β) sheets

Ramachandran plot(ϕ vs ψ angles)



- Gray regions denote the allowed configurations of the polypeptide backbone
- Circles denote the paired angle values of the secondary structures

Definitions of symbols

Types of secondary structures:

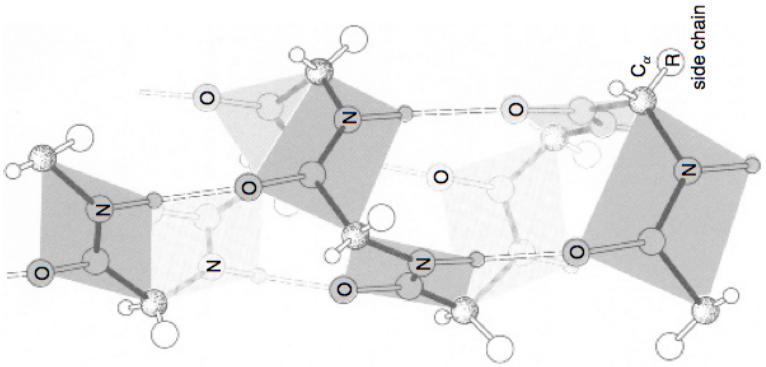
- βA, antiparallel β sheet;
- βP, parallel β sheet;
- βT , twisted β sheet (parallel or antiparallel);
- α, right-handed α helix;
- L, left-handed helix;
- 3, 3.10 helix;
- π, π helix.

Helix structures

A helix is created by a curving of the polypeptide backbone

- Because the polypeptide backbone can be coiled in two directions (left or right), helices exhibit handedness
- A helix with a rightward coil is known as a right-handed helix
- Almost all helices observed in proteins are right-handed, as steric restrictions limit the ability of left-handed helices to form
- Among the right-handed helices, the α helix is by far the most prevalent.

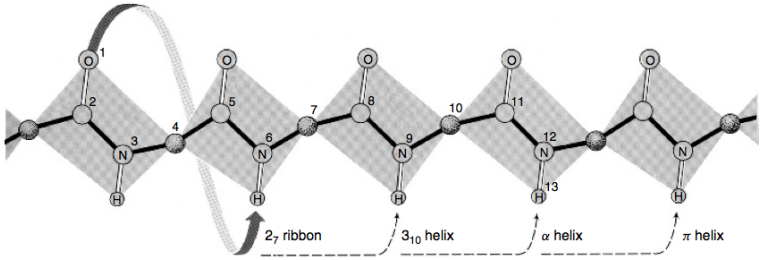
Helix stabilization



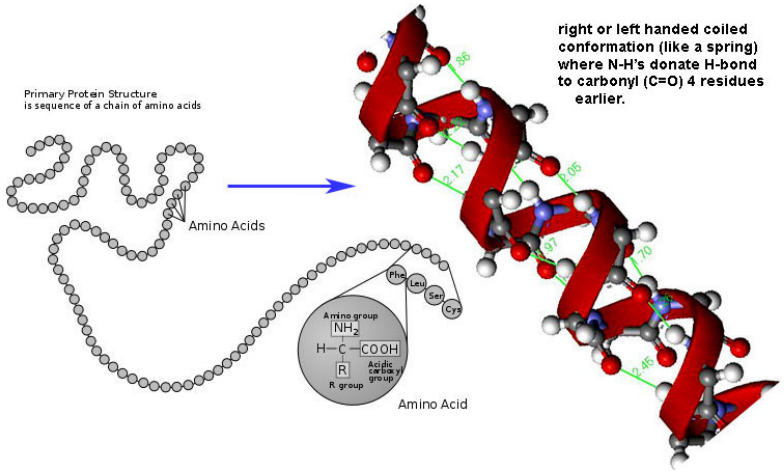
The α-helix is stabilized by internal hydrogen bonds formed between the carbonyl oxygen of each residue and the amide proton of the residue 4 residues ahead in the helix, shown here as dashed lines.

Hydrogen bonds

The hydrogen bonding patterns of different helical secondary structures-



Hydrogen bonds of α-helix



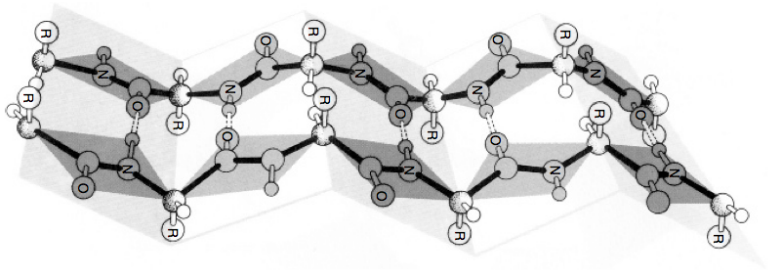
β sheets bonding patterns

Unlike helices, β sheets are formed by hydrogen bonds between adjacent polypeptide chains rather than within a single chain

- Sections of the polypeptide chain participating in the sheet are known as β strands-
- β strands represent an extended conformation of the polypeptide chain, where the and angles are rotated (approximately) 180° with respect to each other
- This arrangement produces a sheet that is pleated, with the residue side chains alternating positions on opposite sides of the sheet.

β sheet configurations

Figure: Diagram of an antiparallel β sheet using a ball-and-stick model-



Two configurations of β sheet are possible: parallel and antiparallel.

Irregular structure regions

- α helices and β sheets account for the majority of secondary structure seen in proteins
- However, these regular structures are interspersed with regions of irregular structure that are referred to as loop or coil-
- Loop regions are usually present at the surface of the protein
- These regions are often simply transitions between regular structures, but they also can possess structural significance, and can be the location of the functional portion, or active site, of the protein.

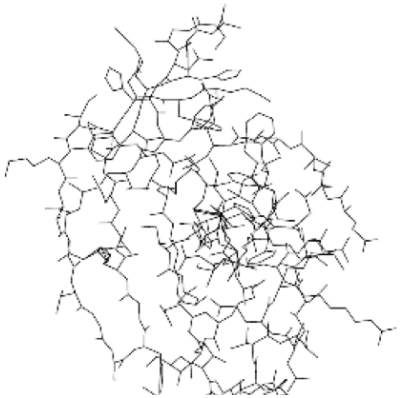
ACTIVE SITE

All-atom line representation

The N-terminal domain of eukaryotic protein kinase A (PKA; PDB-id 1APM) is shown using different representations.

This section of PKA contains a five-stranded antiparallel β sheet and three helices.

Figure: All-atoms diagram of 1APM

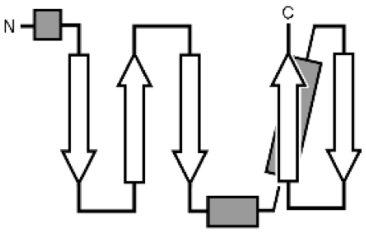


Topology diagram representation

It is difficult to determine the overall structural characteristics of a protein using the all-atom line representation

- Because proteins are often large and complex structures, views at the atomic level tend to obfuscate the important features.
- Simple topology diagrams are two-dimensional projections of the protein structure

Figure: Topology diagram of 1APM



Cartoon diagram representation

These diagrams clearly illustrate the topology (connectivity) between the secondary structural elements and parallel or antiparallel nature of β sheets

Figure: cartoon diagram of 1APM

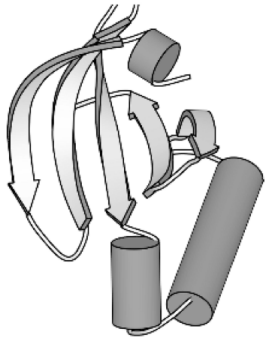
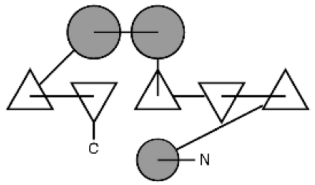


figure generated by MolScript package (Kraulis, 1991)

TOPS diagram representation

Figure: TOPS diagram of 1APM

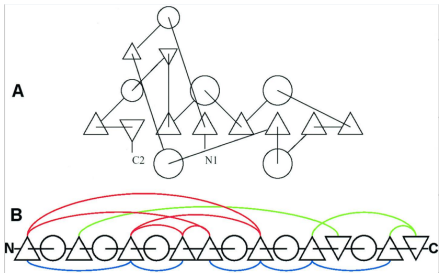


TOPS: an enhanced database of protein structural topology Ioannis Michalopoulos et al., *Nucleic Acids Res.* 2004 January 1; 32(Database issue): D251–D254.

Abstract topological representation (1/2)

Consider a sequence of SSEs, i.e helices (circles) or strands (triangles), together with relationships like spatial adjacency within the fold and approximate orientation, neglecting details like the lengths of SSEs and loops

Figure: Abstract topology diagram: 1ra9



Abstract topological representation (2/2)

(A) 2D TOPS cartoon for of 1ra9 (dihydrofolate reductase)

- TOPS cartoons are pseudo-2D schematic abstractions, where the third dimension is implied, since SSEs are considered to have an approximate direction of 'up' or 'down' (connecting lines drawn to the centre of the symbol indicate connection to the top, and those drawn to the edge indicate connection to the base)
- Direction information for strands is duplicated, upward pointing triangles indicating 'up' strands and vice versa
- Adjacent strand pairs are connected by H-bonds, being parallel or anti-parallel
- Chiralities between parallel strands are also implicit

(B) TOPS diagram of 1ra9

- Hydrogen bonds and supersecondary chiralities are shown explicitly (parallel in red, anti-parallel in green, right-handed chiralities in blue).

Tertiary structure

- The tertiary structure of a protein is defined as the global 3D structure of its polypeptide chain.
- Tertiary structure describes the folding of the polypeptide chain to assemble the different secondary structure elements in a particular arrangement
- As helices and sheets are units of secondary structure, so the domain is the unit of tertiary structure

Side Chains and Tertiary Structure

- At the level of tertiary structure, the side chains play a much more active role in creating the final structure
- In contrast, backbone interactions are primarily responsible for the generation of secondary structure (particularly in the case of helices and sheets)
- The 3D tertiary structure of a protein is commonly referred to as its fold

Domains and Motifs

Tertiary Protein Structure and folds

Within the protein fold, domains and motifs can be recognized

Domains

are compact sections of the protein that represent structurally (and usually functionally) independent regions-

Motifs

(also referred to as supersecondary structure) are small substructures that are not necessarily structurally independent: generally, they consist of only a few secondary structural elements-

Specific motifs

Specific motifs are seen repeatedly in many different protein structures; they are integral elements of protein folds

See Tertiary Protein Structure and folds

- Further, motifs often have a functional significance, and in these cases represent a minimal functional unit within a protein-
- Several motifs can combine to form specific domains.

Protein classification

One method of protein classification partially sidesteps the issue of structural organization in favor of biochemical properties

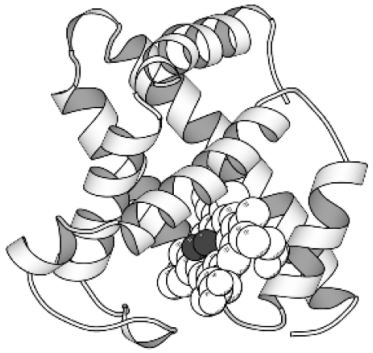
Here, proteins are classified into three major groups:

- globular (see Enzymes amp; other Globular Proteins)
- membrane (see Membrane Proteins)
- fibrous (see Fibrous and Structural Proteins)

Globular protein

Globular proteins Myoglobin

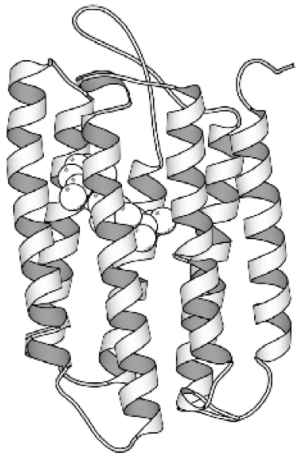
(PDB id 1A6M)



Membrane protein

Membrane proteins Rhodopsin

(PDB id 1AT9)



Fibrous protein

Fibrous proteins Collagen

(PDB id 1QSU)

Structural classification of Folds

- As more and more protein structures have been determined, development of increasingly specific fold classifications has become possible.
- Cyrus Chothia and Michael Levitt derived one of the first such classifications, which grouped proteins based on their predominant secondary structural element (Levitt and Chothia, 1976)-
- This classification consisted of four groups: all α, all β, α/β, and $\alpha+\beta$

Predominant secondary structural element

all α proteins

as the name suggests, are based almost entirely on α helical structure

all β structures

are based almost entirely on β sheet

α/β structure

is based on a mixture of α helix and β sheet, often organized as parallel β strands connected by α helices

$\alpha + \beta$ structures

consist of discrete α helix and β sheet motifs that are not interwoven (as they are in α/β structure).

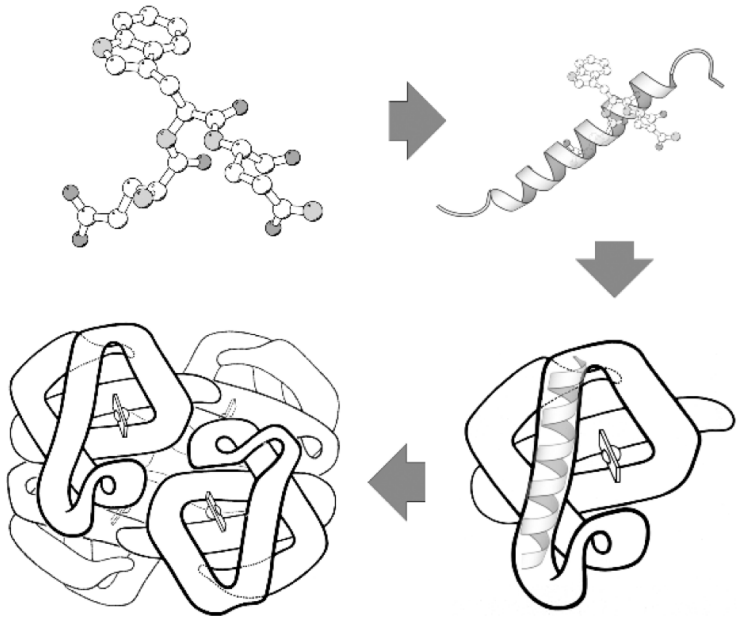
Multimeric proteins

Tertiary structure

describes the structural organization of a single polypeptide chain

- *However, many proteins do not function as a single chain, or monomer*
- *Rather, they exist as a *noncovalent association of two or more independently folded polypeptides**
- *These proteins are referred to as multisubunit, or multimeric proteins and are said to have a quaternary structure*

The four-tiered hierarchy of protein structure (hemoglobin)



Functional Relevance of Quaternary Structure

Interestingly, most proteins are folded such that aggregation with other polypeptides is avoided (Richardson, 1992);

The formation of multisubunit proteins is therefore a very specific interaction.

The quaternary structure is the interaction between several chains of peptide bonds

- The individual chains are called subunits (or domains)
- Complexes of two or more polypeptides (i.e. multiple subunits) are called multimers
- Not all proteins have quaternary structure, since they might be functional as monomers
- The quaternary structure is stabilized by the same range of interactions as the tertiary structure
- The individual subunits are usually not covalently connected, but might be connected by a disulfide bond

Molecular chaperones that assist protein folding

A proportion of all newly-made proteins require assistance to convert from a linear chain of amino acids to a functional three-dimensional entity (folding).

- Chaperonins are protein complexes that assist the folding of these nascent, non-native polypeptides into their native, functional state.
- These proteins belong to a large class of molecules that assist protein folding, called molecular chaperones.
- These molecular machines use chemical energy, in the form of adenosine triphosphate (ATP), to promote protein folding in all cells.

see Chaperonin Structure – The Large Multi-Subunit Protein Complex

Cooperativity

Cooperativity
The association of subunits that bind the same substrate is often able to enhance binding capabilities of the multimer beyond what is possible with individual subunits. This cooperativity is realized through the ability of the subunits to influence each other based on their close proximity.

Hemoglobin

(PDB id 1A3N)

Co-localization of Function

Co-localization of Function

Different subunits can associate in order to confer multiple functions on a single protein. Often these functions involve distinct steps in the processing of a single substrate. Thus, the co-localization of function provided by a multisubunit complex can further enhance the abilities of a protein.

Tryptophan Synthase

(PDB id 1QOP)

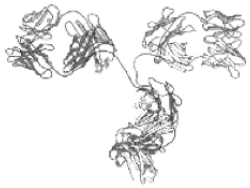
Combinations of Subunits

Combinations of Subunits

Combinatorial shifts in quaternary structure are able to bestow impressive versatility to protein function and regulation. Protein function can be altered by subunit swapping, and protein regulation can be achieved via interactions with different subunits.

Immunoglobulin

(PDB id 12E8)



Structural Assembly

Structural Assembly

Very large structural proteins are made possible by the association of a large number of small subunits. This component-based assembly simplifies the construction of such structures and allows the information required to code these proteins to be more concise.

Actin

(PDB id 1ALM)