expressed and purified independently of the rest of the protein, and they would still show some characteristic activity, like ligand binding, metal binding or interaction with other proteins or even with other domains of the same protein. Some proteins consist of one single domain while others may contain several domains. A protein domain is assigned a certain type of fold. Domains with the same fold may or may not be related to each other functionally or evolutionary. This is because Nature appears to have re-used some protein folds multiple times in different contexts. The currently known protein three- dimensional structures have been classified into more than 1000 different unique folds. In the following chapters we will discuss some examples of these folds, to illustrate the basic principles used for their definition.

The fourth structural level, the **quaternary structure**, is an oligomeric structure and usually involves several polypeptide chains (called **subunits**). It may be the same protein molecule (**homo-oligomer**) or different protein molecules (**hetero-oligomer**). An oligomer is stabilized by subunit interactions and may involve hydrophobic interactions, hydrogen bonds, salt bridges, etc. The different molecules within an oligomeric structure may contribute to an active site (or sites), contribute to the dynamics of the complex and may interact with some target proteins outside the complex.

Since large variations in the sequence may result in the same type of three-dimensional structure, we say that **structure has a higher degree of conservation than sequence**. This can be understood if we take into account function – for example binding of a certain ligand, specificity of interactions with other proteins, dynamic behavior of a structure – all depend on the type of the structure. This is why you may hear that the determination of the structure of a protein with unknown function may help in revealing the function. An interesting example was provided by the

degree.
The vertical axis shows the fraction of highly buried residues, while the horizontal axis shows the amino acid names in one- letter code.

Torsion angles and the Ramachandran plot

Two **torsion angles** in the polypeptide chain, also called **Ramachandran angles** (after the Indian physicist who worked on modeling the interactions in polypeptide chains, Ramachandran), describe the rotations of the polypeptide backbone around the bonds between N-Cα (called Phi, φ) and Cα-C (called Psi, ψ, see below for the graphics view of the angles).

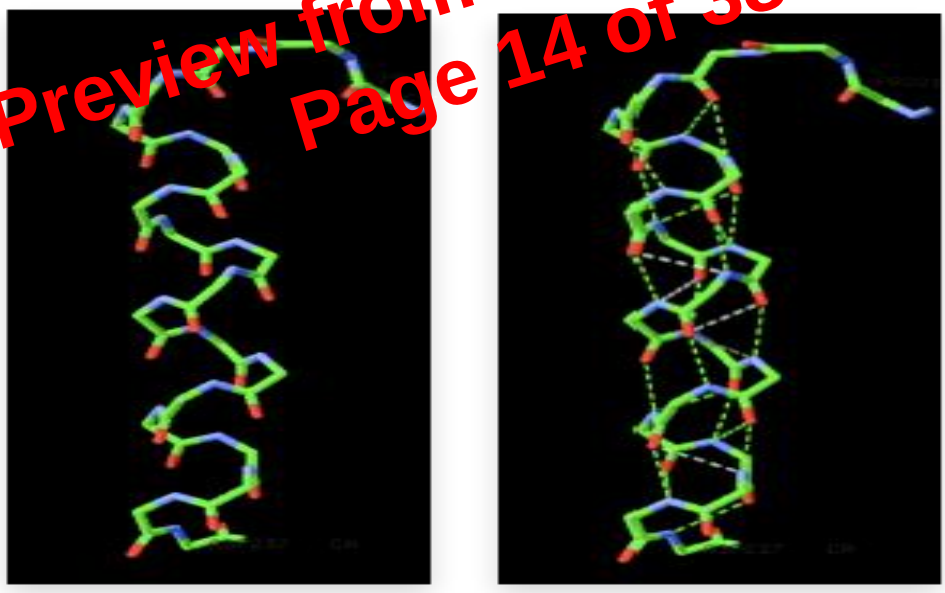
A special way for plotting protein torsion angles was also introduced by Ramachandran and co-authors, and was subsequently named the **Ramachandran plot.** The Ramachandran plot provides an easy way to view the distribution of torsion angles in a protein structure.

It also provides an overview of excluded regions that show which rotations of the polypeptide are not allowed due to steric hindrance(collisions between atoms).The Ramachandran plot of a particular protein may also serve as an important indicator of the quality of its three-dimensional structure.

Torsion angles are among the most important local structural parameters that control protein folding - essentially, if we would have a way to predict the Ramachandran angles for a particular protein, we would be able to predict its fold.

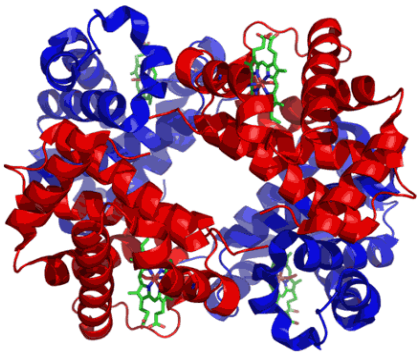
The torsion angles provide the flexibility required for the polypeptide backbone to adopt a certain fold, since the third possible torsion angle within the protein backbone (called omega, ω) is essentially flat and fixed to 180 degrees.

structure is called sticks representation. To give you a better impression of how a helix looks like, only the main chain of the polypeptide is show in the figure, no side chains. There are 3.6 residues/turn in an α-helix, which means that there is one residue every 100 degrees of rotation (360/3.6). Each residue is translated 1.5 Å along the helix axis, which gives a vertical distance of 5.4 Å between structurally equivalent atoms in a turn (pitch of a turn). The repeating structural pattern in helices is a result of repeating φ values and ψ values, observed as mentioned earlier in the text, as clustering of the corresponding torsion angles within the helical region of the Ramachandran plot. The α-helix is the major structural element in proteins. When looking at the helix in the figure below, we notice how the carbonyl oxygen atoms C=O (shown in red) point in one direction, towards the amide NH groups 4 residues away (*i*, *i*+4). Together these groups form a hydrogen bond, one of the main forces in the stabilization of secondary structure in proteins. The hydrogen bonds are shown on the right figure as dashed lines.



The α-helix is not the only helical structure in proteins. Other helical structures include the 3_10 helix, which is

The functional units of both proteins consist of 4 subunits, by other words they are arranged into a **quaternary structure**. In the case of hemoglobin this will make 4 domains, while for pyruvate kinase there will be 12 protein domains in the functional unit. The domains in pyruvate kinase are well separated from each other. The top domain on the figure below is built up by β-sheets, while the other two domains contain a mixture of helices and strands. For illustration, the



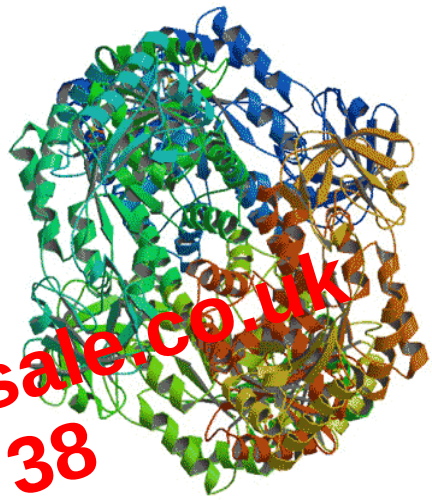


figure below shows the quaternary structure of hemoglobin (left) and pyruvate kinase (right):

In pyruvate kinase the domains are well separated from each other, but in many cases it may be difficult to separate them visually without prior knowledge. As an example, performing a search with PDB ID 1E0T would return the following result for the 3 domains.

1e0tA01 corresponds chain A (there are 4 chains – one for each subunit) and domain number.

As mentioned above, there are in total 3 domains in each chain: 01, 02 and 03. If we click on one of the IDs, for example the first one for domain 1, we get information about its classification - **Class: Alpha Beta, Architecture: 2-Layer Sandwich, Topology: Pyruvate**

geometry, secondary structure content, regions missing in the structure, etc). The R-factor is an essential parameter for the assessment of high well the structure fits the X-ray data. The lower the value of the R-factor, the better the fit. Well-refined protein structures have R-factor values below 20%:

First of all, notice that this structure starts from amino acid Arg 18! No amino acids from 1 to 17. The reason is that there was no electron density for these residues (see for example the discussion on structure quality in homology modeling). This is normally a result of a high flexibility of that particular region of the structure. It is essentially impossible to find the correct positions for amino acids without the guiding electron density. We need to be aware that many structures in the PDB have missing parts, sometimes in loop regions, sometimes just a side chain, and in the worse cases a whole domain may be missing.

The numbers after the first record in the file, ATOM, are just sequential numbers of the atoms in the structure. This is followed by the atom type - for example, CA means C-α, the carbon atom to which the side chain of the amino acid is attached. The next carbon atom is C- β, and following atoms are named after the Greek alphabet, gamma, delta, etc. Except C-α, main chain atoms do not have any Greek letters attached to them. They are just C, O and N.

This is the so-called chain identifier. In cases when the structure consists of several polypeptide chains (a **multi-subunit protein**), each chain will get its own identifier, like A, B, C, etc (as in the case of Pyruvate kinase discussed earlier). Without chain identifiers graphics programs will get confused having the same amino acids names and numbers for different chains (in cases of homo-multimeric proteins). The 3 numbers which follow (e.g., 14.699, 61.369, 62.050 for the very first atom) are the x,y,z coordinates of the atom. They describe the position of each atom in an orthogonal coordinate system. If we can describe the position of each atom in the protein, we will obviously be able to draw the whole tertiary structure.