BIOC0006 Tutorial practicals on protein structure

Name:

In these tutorials you will use:
1. Interactive Kinemage exercises from Branden and Tooze *Introduction to Protein Structure*
2. Structural bioinformatics resources on the web
3. The pymol molecular graphics software

The ability to visualise protein structures in 3D is critical to many aspects of pharmacology, physiology, and biochemistry. Molecular graphics packages are used to view protein structures in a number of different ways. They allow scientists to investigate:
- the 3D arrangement of secondary structure elements and individual amino acids in 3D space;
- the biochemical function of the protein;
- the effects of changing parts of the protein;
- the interactions between two proteins or between a protein and a ligand;
- the cause of disease as a result of protein modification, and the design of drugs to target the protein.

Bioinformatics can be defined as the application of computers to the analysis of biological data (in particular, sequences and structures). This enables scientists to compare their discoveries with what is previously known from the work of others.

In this tutorial (split into two parts), you will look at the following aspects of protein structure:
1. The backbone phi and psi dihedral angles. You will use an interactive graphics program to look at a short stretch of protein structure and try modifying these backbone dihedral angles to see their effect on structure.
2. You will then go on to look at how the combinations of allowed phi/psi angles are restricted and the Ramachandran plot which describes the favoured regions of conformational space.
3. You will look at proteins from 3 conformational classes and see what effects (if any) the class has on the Ramachandran plot
4. You will investigate a particular protein structure and how it interacts with an inhibitor (a substrate analogue) using the molecular graphics program Pymol.
5. You will then use the CATH classification of protein structures to look at the characteristics of other proteins which adopt the same protein fold.

> **You must read the printed information in this booklet to understand what you are doing.**
>
> Instructions for using the software and databases are provided on web pages so that we can make last minute changes if the resources you are going to use change.

Preliminaries

The software and data files that you need should be installed on the teaching lab computers – it is easiest if you can run the pre-installed versions, so try that first.

If the software or data files are unavailable or not working on the teaching lab computers, you can download them from:

> http://www.bioinf.org.uk/teaching/bioc0006/

The KineMage software can be saved to the desktop and run from there (right click on the links and choose the option to save), or can be run directly from the web site by left-clicking the link. Simply say 'yes' to the warnings about security.

If you have to run KineMage this way then you will have to download its datafile to your desk top. When you load it into KineMage, you may need to click the option to see all files rather than just ones with the known extension (.kin).

Part 1 – The General Properties of Protein Structure

1.1. Protein conformation - Phi and psi angles

The two conformational dihedral (or 'torsion') angles **phi** and **psi** largely determine the conformation of the whole protein backbone. The third backbone dihedral angle (**omega**) is fixed close to 180 (or occasionally 0) degrees because of the delocalized nature of the peptide bond.

Visual exploration of variable dihedral angles with Kinemage

Follow the instructions on the web at:
http://www.bioinf.org.uk/teaching/bioc0006/
and click **Ex1**

Question 1.1: Only certain combinations of phi and psi angles for an amino acid are favoured and certain combinations are not permitted. Give one reason for these restrictions.

1.2. The Ramachandran plot

G.N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, made calculations and identified favoured and dis-favoured combinations of phi and psi angles in L-form amino acids. The plot they derived is known as a Ramachandran plot.

When **observed** phi/psi angle pairs from known proteins were plotted, it was found that they fell into the predicted favourable regions, with one region corresponding to phi/psi angle pairs for residues in α-helices and the other region corresponding to phi/psi angle pairs found in β-strands. **Note that these angles must be repeated for each residue along the peptide**

chain in order to obtain a helix or strand – just because a residue has the phi/psi angles of an α-helix does not mean it is actually part of an α-helix! See the lecture notes.

The image of the Ramachandran plot shown below indicates shaded regions which are the favoured combinations of phi/psi angles. The black regions are the most highly favoured, while the grey regions are somewhat less favourable.
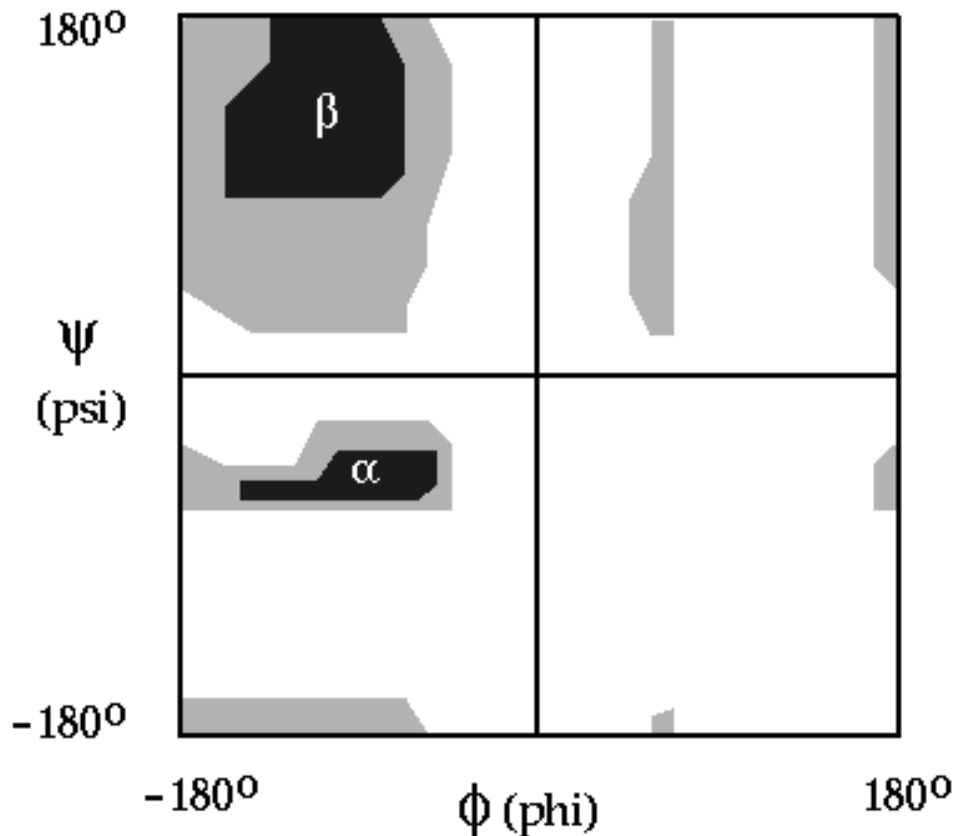
These shaded regions come from an analysis of several hundred high quality protein structures.

When you look at an individual protein, the white regions will not normally be populated by amino acids. Most of those that do fall into the white regions will be Glycines.

Question 1.2: Given your understanding of why certain combinations of phi/psi angles are not allowed, why is Glycine special in allowing additional phi/psi combinations?

Comparison of Ramachandran plots for example proteins belonging to the 'mainly α', 'mainly β' and 'few secondary structures' classes.

The PDBSum website provides brightly coloured Ramachandran plots. As with the grey one above, the shaded areas represent the results of analyzing several hundred high quality structures and consequently will be identical in all the plots you examine. The most populated (and therefore most energetically favourable) regions are shown in red, slightly less favoured regions are in brown and borderline-favoured regions are in yellow. The pale yellow background is disfavoured for all amino acids other than glycine. The favoured 'L' region represents a hypothetical left-handed alpha helix.

The amino acids of your protein will be indicated with pale blue squares (or triangles for glycines). If a non-glycine residue is in a disfavoured region, it will be indicated in red.

Question 1.3: What features(s) do you see <u>in the three plots</u>? How do the plots compare? Are the plots <u>very</u> different in the distribution of phi/psi combinations or not? Note that 1wvk is a small protein and consequently has fewer points on the plot – this is not relevant to the <u>distribution</u> of those points!

1sqc

2bbk

1wvk

Comparison

Part 2 - Enzyme-ligand interactions

Orotidine 5'-monophosphate decarboxylase catalyzes the conversion of orotidine 5'-monophosphate to uridine 5'-monophosphate, the last step in biosynthesis of pyrimidine nucleotides. A high resolution crystal structure bound to the inhibitor 6-azauridine 5′-monophosphate was determined and the 3D co-ordinates of the complex are deposited in the Protein Data Bank (PDB). The protein adopts a **TIM-barrel** fold with one side of the barrel (made up of β strands) closed off and the other side binding the inhibitor. See the lecture notes.

2.1 Exploring a protein structure with pymol

Definition

Proteins are defined as having the same fold if they have the same major secondary structures in the same arrangement and with the same topological connections.

> Follow the instructions on the web at:
> http://www.bioinf.org.uk/teaching/bioc0006/
> and click **Ex3**

Question 2.1a: Identify the C-terminus and the N-terminus of the protein chain and give residues and their sequence numbers. To do this, click on both ends of the Ribbons model in pymol. The residue and number will be shown in the Command Line window of pymol.

Question 2.1b: Your display should show an inhibitor (a substrate analogue) spacefilled interacting with the cartoon protein. Where in the protein is the active site located?

2.2 Detailed view of Protein/Ligand interactions with LIGPLOT

LIGPLOT is a program which draws a flattened view of the interactions between a ligand and a protein.

> Follow the instructions on the web at:
> http://www.bioinf.org.uk/teaching/bioc0006/
> and click **Ex4**

Question 2.2a: Indicate the type and residue number (e.g. Arg203) of each of the amino acids involved in formal charge electrostatic interactions with the inhibitor molecule *UP6 5001*. Note that only the phosphate group of the ligand will have a formal charge. (Electrostatic interactions are indicated by green broken lines, together with the distance of the interaction. Remember that only Arg, Lys, His, Asp and Glu have formal charges.)

Question 2.2b: The remaining interactions shown by green broken lines will be hydrogen bonds. Indicate the type and residue number of each of the amino acids making hydrogen bond (rather than electrostatic interactions) with the inhibitor molecule *UP6 5001*.

2.3. Structural classification of TIM barrel proteins

CATH ('**C**lass **A**rchitecture **T**opology **H**omology') is a hierarchical classification of protein structures.

- **'Class'** indicates whether the protein is mostly-alpha, mostly-beta, mixed, or low-secondary structure (you looked at examples from 3 of these classes last week).
- **'Architecture'** refers to the arrangement of secondary structure elements in three dimensions
- **'Topology'** refers to the connectivity between the secondary structure elements as they are arranged in this architecture. This is also commonly referred to as the **protein fold**.
- **'Homology'** indicates whether proteins of a given protein fold are evolutionarily related.
- Further levels of classification categorize proteins into sequence families.

Follow the instructions on the web at:
http://www.bioinf.org.uk/teaching/bioc0006/
and click **Ex5**

Question 2.3: What are the typical biochemical functions of proteins with TIM barrels? Look at the protein names in CATH (most of them have an ending that is characteristic of a certain type of protein) and try searching Wikipedia for clues.

Before leaving, you must discuss your answers with a demonstrator who will check your written answers.

Demonstrator's signature:

In your own time...

Pymol and Kinemage are open-source programs which you can download and install on your own computer. To explore protein structure more thoroughly, download a copy of pymol and download protein structures from the Protein Databank (PDB): http://www.pdb.org/

Kinemage provides a further useful introductory exercise. You can download a copy of Kinemage onto your own computer and obtain a copy of the **c1basics.kin** file from http://www.bioinf.org.uk/teaching/bioc0006/ (or do a Google search for c1basics.kin).

Proceed as in the first tutorial, but click on ***{Kin 3}***. A protein structure will be displayed. In the right-hand panel, you will find a button labelled 'Animate', as you click this button, you will see different subsets of amino acids being highlighted.

This shows you how hydrophobic amino acids tend to pack on the inside while hydrophilic amino acids are on the outside. Charged amino acids, when buried, always form charge pairs. The 'special' amino acids (glycine which has no sidechain and proline which has a cyclic sidechain) are also highlighted.

Useful Websites

Data and programs for this practical	http://www.bioinf.org.uk/teaching/bioc0006/
PDBsum	http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/
CATH	http://www.cathdb.info/
RasMol	http://www.openrasmol.org/
Kinemage	http://kinemage.biochem.duke.edu/