

## Protein modelling

---

Computer based protein 3d modelling is an essential component of understanding/demonstrating the relationships between proteins and their substrates. In this class we will use a free (for academic purposes) pdb viewer called Deepview, to understand how mutations could affect the function of our protein and what outcome that may have for the protein both experimentally and in a living cell.

*\* The Deepview software has been provided on your computers along with a .pdb file of the wild type of your protein.*

## Generating models

Prior to computer modelling possible protein structures were cumbersome determined using physical models and drawing possible structures (of which there are many) on paper. Now digital, this process is still essentially the same (though vastly easier). These models must be confirmed using a process called x-ray crystallography. This is based on the scatter of x-rays by a pure crystal of the protein of interest. Unfortunately the process of generating x-ray crystallographic images of a molecule is very difficult. Some proteins have too many possible conformations (shapes) for them to crystallize properly. Even with a good crystal, proteins often contain unstable/ flexible regions which appear in multiple conformations. Such regions are often excluded from the protein data base entries. These problems can sometimes be dealt with through providing the protein with its substrate or other proteins it may interact with during crystallisation which can ensure it stays in a mostly stable confirmation.

*\* Disordered AKA flexible regions, are the reason many crystallographic structures of proteins will be smaller than the total known sequence of a protein.*

## Computational modelling

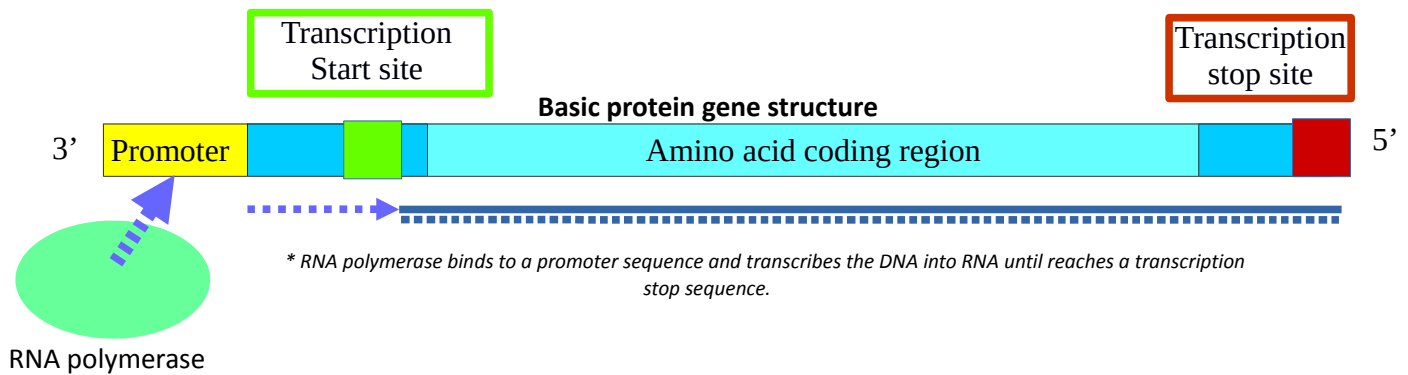
Many years of research have allowed us to draw from a vast wealth of crystallography data and as proteins often share important sequences (called conserved sequences), we can use this to model a proteins potential tertiary structures with software. This requires a great deal of computing power to accurately model and must be done using large servers with massive data bases. You can do this through the submission of a sequence (DNA/AA) as a fasta file to services such as PHYRE2, which will generate a .pdb file for 3d modelling software.

*\* Thankfully there does exist x-ray crystallography information regarding hSSB1 and a larger structure it interacts with (consider this while investigating your mutation).*

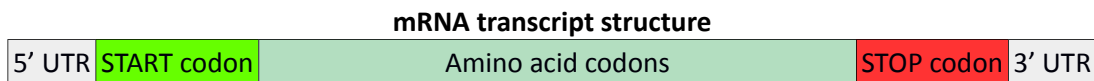
## Mutagenesis

Amino acids, called residues when bound together in a sequence, are the building blocks of proteins. They form a contiguous sequence that folds (and is often edited by other proteins) to create a stable tertiary structure. This structure is what imparts a proteins function, changes to important regions called domains can cause a protein to miss-fold, retarding the proteins function and sometimes inhibiting entire cellular pathways.

*\* The process whereby DNA is transcribed to RNA and translated into an amino acid sequence is called **"The Central Dogma of Molecular Biology"**. This process is essential to understanding the formation of most mutations.*



When a given protein is needed RNA polymerase (an enzyme that attaches nucleotide bases together) creates a RNA transcript using the bases of the required gene.



*\* UTR= Untranslated region.*

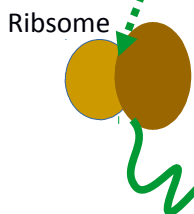
*\* DNA is written 5' to 3' by most DNA processing enzymes.*

The amino acid coding region may appear as

### Codon sequence from mRNA



*\* Codons are triplets of bases that are interpreted by a ribosome*



Which would have a final sequence in a protein as

### Translated amino acid sequence



An mRNA codon to amino acid table is provided below. This table contains the necessary information for making changes to a DNA codon and understanding how it can change the translation of an amino acid. This may also demonstrate how many mutations are degenerate (redundant) as a codon may still sometimes code for the same or similar amino acid.

### mRNA codon table

mRNA codon table									
Nucleotide positions in codon									
First	Second								Third
	U		C		A		G		
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	(Trp/W) Tryptophan	G
C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	
	AUG	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Translation START
nonpolar
polar
basic
acidic
Translation STOP

*\* Your mutations will have been provided in the format **X#Y** whereby X is the original amino acid at residue # and has been changed to amino acid Y. For example, K92F would mean at residue 92 Lysine (Basic) has been changed to Phenylalanine (nonpolar).*

## Mutation 3D Modelling

---

\* If you have any trouble during this session, raise your hand for assistance.

For this project 3d modelling will be done with software called Deepview (AKA swiss-pdbviewer) from the Swiss Institute of Bioinformatics. Such software allows for the viewing and altering of .pdb files. When modelling a protein, crystallographic data should be used where available for an accurate representation of your protein.

*\*.pdb/protein data bank file, a standardized file-type for compatibility across protein modeling software, contains the position and characteristics of each atom in a protein or given molecule.*

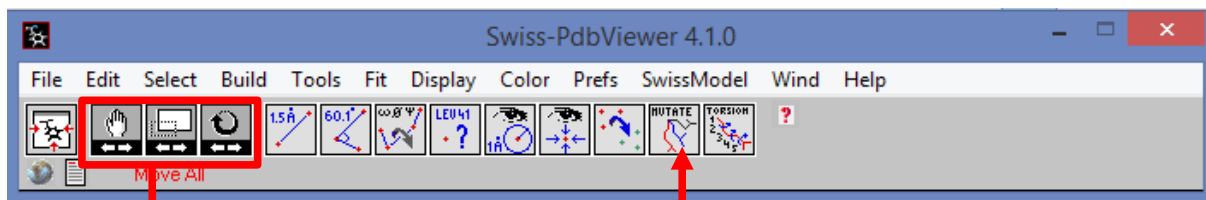
### Getting started

1) Do a google search in your browser. Type in: hSSB1 crystallography. Open up the third link- **Structural basis of SOSS1 Complex Assembly** – this is a crystallography paper. What we want from this paper is the model. Scroll down and find the **accession numbers** heading. Click on 4OWW and it will take you to a protein data bank entry. Click on the Download file > PDB format and save it to the desktop. **#Don't download the (gz) file.**



2) Start by clicking the Deepview icon on your desktop.

3) An about window should appear, close it and the program toolbar should appear as below.



Visual manipulation tools

Mutation tool

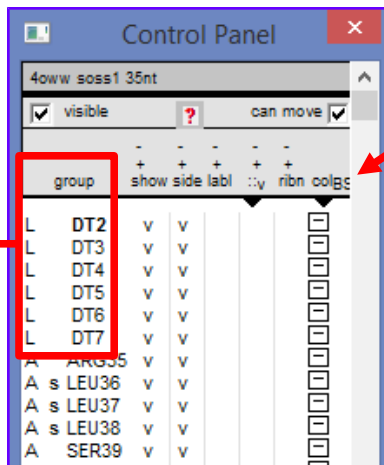
4) In the toolbar click on File and go to Open PDB file and navigate to the provided .pdb file of your protein. (Toolbar > File > Open PDB file ....)

5) A 3D backbone + side chains model of your protein should appear along with a log-file which you should close. Play with the 3 visual manipulation tools in the toolbar to understand what they do.

6) Click Wind in the toolbar and select control panel, you will make most of your changes to your protein from here. (Toolbar > Wind > Control Panel)

The left most character indicates the molecule, while the characters to its right indicate the specific residue.

Various components can be switched on and off by selecting them and pressing the - or + symbols above the label of the feature you wish to change. E.g. A selected group can be hidden by clicking the - symbol above the "show" label.



Colours can be attributed to a selected group by clicking the colBS label and selecting the desired colour.

Large groups can be selected by left clicking + holding shift on a desired residue then doing the same on the final residue.

## Visuals (cleaning it up)

- 1) Click Display in the toolbar and deselect Show Backbone Oxygens.
- 2) In the control panel, hide unwanted proteins, residues and side chains. Ensure you don't hide the DNAs side chains (L).  
  
*\* INTS3 (A) is a protein hSSB1 (B) associates with as part of the DNA repair process. hSSB1 is provided as residues 5-111 despite its database entry being much larger, this is because only these residues could be imaged accurately. C is a portion of another protein in this complex, as it isn't directly associated with hSSB1 we can hide this as well.*
- 3) Attribute different colours to the DNA (L), hSSB1 (B) and INTS3 (A) groups to help differentiate them.
- 4) To help you see your mutation find the residue and change its colour so it's different to the rest of group B and show the side chain.
- 5) Use the 3 visual manipulation tools in the toolbar to orient into a good view of the your residue.
- 6) Select the relevent substrate or residue and turn on '::v' in the control panel, this will show a dot sphere of its Van der Waal forces. If necessary select 'Slab' from the Display menu in the toolbar to hide atoms not in the current plane of view.
- 7) Take a picture by selecting File > Save > Image so you can use images of the normal residue in your presentation.

*\* Though poorly calculated in software some relevant hydrogen bonds can be demonstrated by selecting "Compute H-Bonds" in the "Tools" menu on the toolbar.*

## Mutations

1) With a good view of the interactions of your residue, click the mutate tool button in the toolbar and select the residue in the visual window. A drop down menu of possible amino acids will appear, using the code for your mutant, select the amino acid for your mutation. Click on the mutation tool button again and you will be asked to accept the change, click Yes.

*\* E.g Y98A means Y has been replaced with A on the 98<sup>th</sup> residue/amino acid. This will be demonstrated for you.*

2) Take another image and continue to examine your protein as you desire.

3) Repeat this process with other mutations. When you're done attempt the questions below.

## Questions

---

### Record the mutations you investigated here

- \_\_\_\_\_
- \_\_\_\_\_

Working with your group and using the mRNA codon table above answer the following questions.

What amino acid change occurred and did the type of amino acid change (e.g. polar)?

---

---

---

What region of the protein is your mutation in and how might this affect its function?

---

---

---

---

Suggest what DNA codon changes in the hSSB1 gene mutant could cause this change (e.g. CAG to CCA)?

What effect might this have on a cell?

---

---

---

---

How might this affect a carrier of this defect?

---

---

---

---

## Further reading

---

\* **Central Dogma of Molecular Biology**

[https://en.wikipedia.org/wiki/Central\\_dogma\\_of\\_molecular\\_biology](https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology)

\* **Principles of x-ray crystallography**

[http://chemwiki.ucdavis.edu/Core/Analytical\\_Chemistry/Instrumental\\_Analysis/Diffraction/X-ray\\_Crystallography](http://chemwiki.ucdavis.edu/Core/Analytical_Chemistry/Instrumental_Analysis/Diffraction/X-ray_Crystallography)

\* **hSSB1 x-ray crystallography paper**

<http://www.sciencedirect.com/science/article/pii/S221112471400120X>

\* **swisspdbviewer AKA Deepview website**

<http://spdbv.vital-it.ch/>

\* **Computational protein modeling services**

[http://molbiol-tools.ca/Protein\\_tertiary\\_structure.htm](http://molbiol-tools.ca/Protein_tertiary_structure.htm)

\* **Mutation nomenclature**

<http://www.hgvs.org/mutnomen/>

\* **Protein structure levels**

<https://www.ebi.ac.uk/training/online/course/biomacromolecular-structures-introduction-ebi-reso/proteins/levels-protein-structure>