

Name:

Date:

# WP2 – Using the PDBe website

First, access the PDBe databank at [www.ebi.ac.uk/pdbe/](http://www.ebi.ac.uk/pdbe/). From here you can search for over 100,000 structures.

1. Search for 'Hemoglobin'. How many different entries are found?  
**Over 700**, more will have been added since this worksheet was produced.
2. Why are there so many different entries found, and not just one universal structure of hemoglobin?  
Hemoglobin is a type of molecule, so there are many **different variations** of it. These include variations between **different species**. The hemoglobin commonly found in humans is very different to that found in an earthworm. There are also lots of entries in which the hemoglobin is **interacting with other molecules** and small compounds. A hemoglobin bound to oxygen is different to one bound to carbon monoxide etc.
3. Along the left-hand pane of the page, you will be able to refine the number of results down depending on which type of molecule you want to look at. One of these allows you to refine by 'Experimental methods'. If you click on this section, you will be able to see some of the different ways that we can determine the structure of a protein. Why do you think that it is important to be able to search for structures which have been determined by different methods?  
**Each experimental method has positives and negatives.** Some methods can more accurately locate the position of hydrogen atoms, but give a less precise overall picture for example. It is important to see the method used so that you can **know how accurate your data is**. Some methods require the protein to be crystallised, whereas some require it to be in solution, and these can affect the final structure that you see. It is also useful to be able to **compare results of different methods**, to check that different methods can both give the same result. If two methods give completely different results, then at least one of them must be wrong!
4. Two of the most important ways to refine a search are using 'Interacting compounds' and 'Species name'. Using these, refine your search to samples of human hemoglobin bound to O<sub>2</sub>. How many entries are there?  
37 at the time of publication, however more may have been added!

You should be looking at a page similar to this:



The screenshot shows a search results page for PDB. The first entry is **2dn1**, titled "1.25Å resolution crystal structure of human hemoglobin in the oxy form". It was released on 09 May 2006. The source organism is *Homo sapiens*. The assembly composition is "protein/protein complex". Interacting compounds are listed as MBN, HEM, and OXY. The second entry is **4x0l**, titled "Human haptoglobin-haemoglobin complex". It was released on 24 Dec 2014. The source organism is *Homo sapiens*. The assembly composition is "polysaccharide/protein complex". Interacting compounds are listed as OXY, GOL, and HEM. Both entries include 3D molecular models and quality metrics for X-ray diffraction.

It is worth noting that the database will naturally order the results by their quality, so that the highest quality result is first. This is done by looking at the resolution of the data, and the accuracy of it.

Click on the first entry, 2dn1. This should open the following page:



The screenshot shows the PDB entry page for **2dn1**. The title is "1.25Å resolution crystal structure of human hemoglobin in the oxy form". The source organism is *Homo sapiens*. The primary publication is "1.25 Å resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms" by Park SY, Yokoyama T, Shibayama N, Shiro Y, Tame JR, published in *J. Mol. Biol.* 360:690-701 (2006). The page includes a "Quick links" tab with options like "2dn1 overview", "Citations", "Structure analysis", "Function and Biology", "Ligands and Environments", and "Experiments and Validation". The "Ligands and Environments" section shows "3 bound ligands" with chemical structures for heme, oxygen, and toluene. The "Citations" section shows "10 review citations" with the title "The role of alpha-hemoglobin".

This will show all the information about this entry. You can access different sections using the 'Quick links' tab on the right-hand side. One of the more useful parts on the main page is the 'Ligands and Environments' section. This shows the molecules which interact with the main protein.

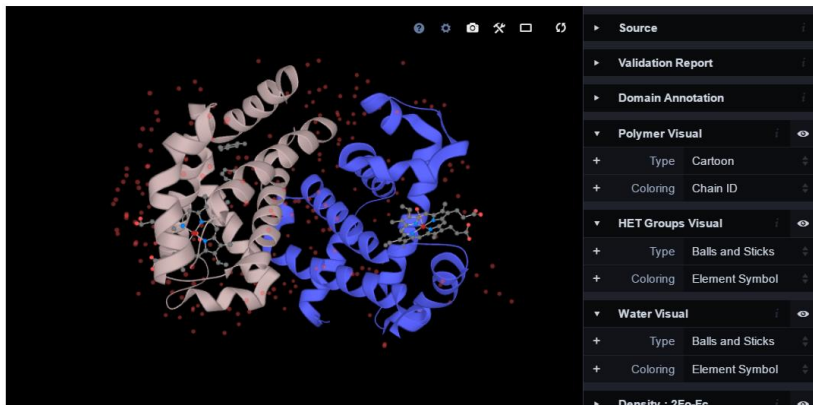
5. What are the three ligands which interact in this protein? And why do you think the final ligand has been detected in the sample?

**Prophyrin, molecular oxygen, and toluene**

Prophyrin and oxygen are expected in oxyhemoglobin. **Toluene** is most likely used as a **solvent** in the structure determination.

### 3D VISUALISATION

To view the structure in 3D, open up the '3D Visualisation tool on the right-hand side. This should open a viewer like this:



Heterogroups – these are groups attached to the protein such as ligands.

Shows the water groups around a protein. This is important for hydrogen bonding, but for the most part can be ignored. You can hide these by clicking on the eye icon.

This is a fully 3D viewing tool which you can now use to investigate proteins. There are a few things to note:

Firstly, each structure is shown as an 'Asymmetric unit', this means that it shows you the smallest part of the structure which you cannot break down by symmetry. Most of these structures are analysed as crystals of repeated units. The asymmetric unit just means that this is considered. In this case, hemoglobin actually exists as a tetramer – it has 4 chains. But this model only shows two of them (pink and blue), as the tetramer can be seen by repeating the unit shown.

Along the right-hand side there are few different options to help you view the structure. One of the most useful is the Visual Type option, which allows you to select how the molecules are displayed.

1. Try out some of the different types of Polymer Visuals, giving a brief comment on why each may be useful.

**Cartoon:** shows common structures easily. It is **simple** and easy to spot **key features**.

**C- $\alpha$  Trace:** shows the **general shape** of the polypeptide to see how it has folded.

**Balls and Sticks:** shows each **individual atom** and **covalent bond**

**VDW Balls:** shows a **more accurate representation of the size of each atom** to see the **3D space** and investigate active sites

**Surface:** Shows the **surface of each chain**, to see the **overall 3D shape** and investigate active sites

As a 3D viewer, you can move this model around using your mouse. It is possible to focus on a specific section by clicking on it. When you do this, you will see blobs of blue, green and red appear around your selection. These represent the electron density of the structure and can be ignored by hiding the three density sections on the right-hand pane.

## PROTEIN OVERVIEW PAGE

Load up the entry 4ins. This will be a structure of insulin from a pig. The Protein Overview page that you see when you click onto an entry is full of information. By using the Overview of 4ins find out:

1. The scientific name of the pig species which this sample is from  
***Sus scrofa***
2. The year that this structure was first published. Not that this is not the same as the year the structure was released. You should look for its journal article which will look like:  
*J. Mol. Biol.* **136** 103-28 (1980)  
Why is it important to know when this structure was first published?  
**1988**. It is important because this structure was calculated 30 years ago. This is a **relatively long time** and **experimental methods have improved** much since then. This means that there may be **more accurate** and detailed structures of pig insulin **available**. Or if they aren't, then we may be able to determine the structure again using current, improved methods, to give a better final structure.
3. The biochemical function of the protein  
**Hormone activity**
4. The cellular component – is it found on the inside or outside of the cell?  
**Extracellular** – on the outside of the cell
5. The type and number of bound ligands  
**2x Zn<sup>2+</sup>**
6. The number of distinct macromolecules (large molecules making up the overall structure).  
**2**