

Molecular Biology Through Discovery
Companion to Perutz et al (1965)
Structure and function of hemoglobin

II. Some relations between polypeptide chain configuration and amino acid sequence
Perutz MF, Kendrew JC, Watson HC (1965). J Mol Biol 13:669-678

I. Introduction

There were three great centers of molecular biology in the early years when the field was undergoing its birth. Cal Tech was the home of Linus Pauling, whose work on the secondary structure we have silently passed over thus far this semester (to our shame). We will hear a great deal in the coming weeks from the group centered at Cal Tech and led by Max Delbrück. A second center was Institut Pasteur in Paris – the groups of André Lwoff and Jacques Monod – that focused on what we would now call gene regulation in bacteriophage and bacteria. The third center was the closely related Cavendish Laboratory, led by Lawrence Bragg (developer of X-ray crystallography) and the Laboratory of Molecular Biology at Cambridge, UK.

Everyone knows of the announcement that came from Cambridge in 1953 – Watson and Crick's double helical structure of DNA. Fewer are aware of the second epochal announcement, from the same group, the same year. For decades many had attempted to use X-ray crystallography to elucidate the structure of proteins, in the hope that the structure might shed light on how proteins are able to perform the variety of roles required for the functioning of a cell. Unfortunately, while the technique led to great advances in the understanding of small molecules, proteins were too big. We now know that a typical protein has thousands of atoms. They produce too many X-ray reflections to allow an intelligible structural interpretation.

The story has often been told¹ about how a graduate student at the Cavendish speaking at his first scientific meeting gave a talk entitled "What Mad Pursuit", in which he argued from first principles that those attempting to find the structure of proteins through X-ray crystallography were wasting their time. It says something about the value placed on communication over hierarchy that a graduate student could speak thus to an audience that included the head of the institute, Lawrence Bragg, who had devoted his life to the technique. It perhaps doesn't matter that the student happened to be Francis Crick and that Bragg was reported to be not amused. Also in the audience was Crick's mentor, Max Perutz, who later directed the lab and wrote of the need for barrier-free communication.²

Perutz embraced the sentiments of Crick's talk and looked for a way around the theoretical obstacle Crick had presented.² In 1953, Perutz found a way, showing that certain derivatives of hemoglobin that contained a heavy atom, mercury, maintained their natural structure, and the heavy atom grossly simplified the problem of interpretation.³ Perutz' group used this trick to solve the structure of hemoglobin. This structure provided first detailed insights into the mechanism underlying hemoglobin function and thus validated the initial motivation to pursue structure determination.



Figure 1: Max Perutz and John Kendrew, with models of hemoglobin and myoglobin, respectively (1962; http://www.nobelprize.org/nobel_prizes/medicine/articles/perutz/)

Meanwhile, John Kendrew, also at the Cavendish, used the same techniques to solve the structure of myoglobin, which differs from hemoglobin in having one polypeptide chain instead of four.

In the article that is the focus of these notes, Max Perutz, John Kendrew, and H.C. Watson compared the amino acid sequences and structures of different versions of myoglobin and hemoglobin, in the hopes of finding common features that define the proteins. Have the article on hand as you go through the notes. The purpose this time out isn't to extract nuggets from the article but rather to use it as inspiration for your own process of discovery, using protein visualization tools.

II. Initial survey of article

The first paragraph of the article states clearly what the authors are aiming for.

SQ1. From the first paragraph of the Introduction of Perutz et al (1965), what question is central to the work reported in this article? What motivated the question?

This is clearly a big, important question, but it isn't one that is easy for you to deal with, given the tools you have available to you. Instead, I propose that we go through the article together, using it to point out important features of interest in two specific globins: horse hemoglobin (the protein whose structure was determined by Perutz) and whale myoglobin (the protein whose structure was determined by Kendrew).

Now to the body of the article.

SQ2. What kind of results are presented?

Hmmm. This poses somewhat of a problem. Table 1 is just a listing of sequences they used – not too interesting unless we want to look up some sequences. It's the other tables that present the core of the article... and they all use a strange language! What is meant by "B6" in the first line of Table 2? The authors presume familiarity with their earlier work, where the nomenclature is defined, but you can figure it out from Fig. 1. There you see what is described as the sequence of globin (graphically represented), broken up into different segments, labeled A through H, plus regions in between two adjacent segments (e.g. AB). Each label lies above a wavy line, which represents a helix – a region of a particular secondary structure. The symbols in Table 2, e.g. "B6", means the 6th amino acid in Helix B. Unfortunately, no one uses this nomenclature anymore. We'll have to translate it into standard amino acid coordinates to make the protein visualization tools understand what we're talking about.

It's clear that we're going to have to have some prior knowledge of helices and how they fit into protein structures before diving into the article.

III. Introduction to helices (and graphical representations)

It's a good thing I don't have to introduce you to this topic myself! Fortunately, there's an excellent on-line introduction, entitled [Hemoglobin Molecular Structure](#).⁴ Don't proceed until you're at this introduction.

Part 1 (Introduction to Molecular Visualization) tells you about graphical conventions that will become second nature as you use the protein visualization tools. It's not so important now, so give it a quick look to see what's there. That's probably enough. For this and future parts, I

suggest that you click the **Toggle spinning** button to stop the molecule in place and then use your mouse to rotate it (try it out!).

SQ3. Part 1 shows different representations, e.g. spacefilling and stick. What are the advantages and disadvantages of each? When might you prefer one representation over another.

Part 2 (Hemoglobin & Heme) gives a useful introduction to hemoglobin, one of the two featured proteins of Perutz et al (1965). This is an excellent way to get yourself introduced.

SQ4. What are the red blobs in Part 3, View 2? What do the blue, magenta, green, and yellow squiggles represent? Why aren't the red blobs represented as squiggles?

Part 3 (Hemoglobin Secondary Structure) leads you gently to View 9, which shows the mechanistic basis for alpha helices, a common element protein structures.

SQ5. Part 3, View 9, shows that the alpha helix is stabilized by hydrogen bonds between atoms of the backbone – N from amines and O from carboxylates. But all amino acids of a protein are part of the backbone. Why isn't every protein a gigantic alpha helix?

SQ6. How many amino acids are contained in the helix shown in Part 4, View 7? How many amino acids are there in one turn of the helix?

Part 4 (Amphipathic Alpha Helices) relate the hydrophobicity of certain amino acids with the three-dimensional structure of proteins. View 5 presents a fascinating view of slices through the structure of one of the chains of hemoglobin.

SQ7. Why do most alpha helices in hemoglobin have predominantly hydrophilic amino acids on one side and hydrophobic amino acids on the other?

Part 5 (hydrophobicity, Polarity & Charge) continues the theme of the previous part, connecting hydrophobicity with protein structure. Views 1 and 2 are particularly interesting.

SQ8. What general lesson can you take away from the slices?

Parts 6 (Regulation of Oxygen Binding) and 7 (Sickle Hemoglobin) may be too much for now, but keep them in mind if you encounter the need to understand these matters.

SQ9. Rotate Part 3, View 3, so that you can count how many helices there are, presuming that each one has its own distinct color. If you labeled each helix consecutively, with a letter, starting with "A", where would you start, where would you end, and what would be the last letter? Compare your answer to Fig. 1 in Perutz et al (1965).

IV. Visualization of specific observations from Perutz et al (1965) - preliminaries

The rest of our time with this article will be spent taking observations by the authors as the starting point and seeing for ourselves, using protein visualization tools. For this purpose, use the protein visualization tool, [Chimera](#).⁵ This program does not run online. You need to download it to your own computer and install it (which is easy, at least when I did it). Like Perutz et al (1965), we'll be looking at whale myoglobin and horse hemoglobin.

SQ10. Open up Chimera on your own computer. Click the *File* menu and select *Fetch by ID*. What kind of information do you have to feed this tool in order to display your favorite protein?

"PDB" stands for "Protein Data Base". To get a PDB ID, go to [Protein Data Base](#) (see link on the Protein unit web page or the Links page). We'll start with the simpler protein, whale myoglobin.

SQ11. What does the search box at the top of the Protein Data Base main web page want? Do you have it?

Well, of course you don't have the PDB ID of whale myoglobin – that's what we're trying to find! But you do have the molecular name and at least one author.* That should do.

SQ12. What is the PDB ID of whale myoglobin as described in an article one of whose authors was John Kendrew?

Click the PDB ID, which will be four letters/numbers, to bring you to the whale myoglobin structure page. We want two things from it: the PDB ID (which you now have) and the sequence of the protein. You'll find a **Sequence** tab near the top of the page. Click it.

If you scroll down a bit on the resulting page, you should get to an amino acid sequence of myoglobin (it will begin VLS...) with a figure above it, and... well, strike me blind!

SQ13. How does this Protein Data Base figure compare to Fig. 1 of Perutz et al?

SQ14. Using the two figures, what is the 3rd amino acid of myoglobin in Helix A? Note that counting from the beginning of Helix A is different from counting from the beginning of the protein.

There remain two slight problems. First, the PDB figure uses one-letter amino acid abbreviations, while Perutz et al use three-letter abbreviations (e.g. in Table 2). There are many resources that translate between the two. One is accessible through the Links page as [Abbreviations](#). The second problem is not so straightforward to solve: Where precisely do the helices begin? There may be a difference of opinion on this point, so you should be prepared to be flexible.

Now we need to get that three-dimensional structure. Go back to Chimera (remember that?), type in the PDB ID of whale myoglobin, and click *Fetch*. As a result, you should see a representation of myoglobin on your screen.

I presume the representation on the right looks sort of familiar (Hemoglobin tutorial, Part 3, View 3?), but to make it even more familiar, do this:

- From the **Presets** menu, choose **Interactive 1 (ribbons)**

This sets the view to Ribbon (which it already was) and the color to Rainbow (for 1-chain proteins: beginning of protein is blue, end is red, in between is in between).

* Of course the author is not the author of the article we're reading but the author of the study that produced the structure of whale myoglobin, who was mentioned twice earlier in these notes.

SQ15. Mouse over the tip of the blue string (the N-terminus of myoglobin). You should see “Val 1.A” appear, meaning valine is the first amino acid of chain A. Moving your mouse, find out the names of the first three amino acids.

The whole point of myoglobin is to carry oxygen on the bound heme. Identify the heme in the structure of hemoglobin.

SQ16. Where is the heme?

You probably have a guess, but let's see what Chimera has to say on the matter:

- From the **Select** menu, choose **Residue / HEM** (for heme). You should see a green line around the heme.
- From the **Actions** menu, choose **Atoms/bonds / sphere**. You should see the heme blow up so that each atom in it takes on its full volume
- From the **Actions** menu, choose **Color / hot pink** (so the heme can be readily distinguished later).

Now you're ready to interpret the data from Perutz et al and identify amino acids of interest in a protein graphic that you can manipulate.

V. Visualization of specific observations from Perutz et al (1965)

Invariant amino acids

The first question addressed by Perutz et al was which amino acids of globins are found in all globin proteins in the same positions.

SQ17. What fraction of the amino acids in globin would you expect to be the same in all instances of the protein (e.g. whale myoglobin, human hemoglobin, lama hemoglobin...)?

If you're like Perutz et al, you'd figure that all globins do about the same thing – carry oxygen – and they all have about the same structure. So they should have *lots* of amino acids in common. Mammals all have similar functions, and almost all of us have the same sort of organs in more or less the same places. Why not globin?

SQ18. Comparing the results shown in Table 2 of Perutz and the number of amino acids in the sequence shown on the PDB page, what fraction of myoglobin is invariant?

If you're still like Perutz et al, then you're pretty surprised by this result! The number is very small. Those invariant amino acids must have important functions if they never change. Table 2 identifies five amino acids as having a related function (at least they all have the same word in their functional description).

SQ19. Using the nomenclature of Perutz et al, Table 2, what five amino acids are invariant and have the same word in their function description? Why is this word not surprising?

The practice used by Perutz et al of identifying amino acids in terms of the helix in which it appears (e.g. B 6 Gly = glycine, the 6th amino acid in Helix B) makes sense for comparing

structures. The proteins being compared may have slightly different number of amino acids, but they all have the same helices. However, the practice disappeared from the literature – it was so much easier to find amino acids in a sequence when they were numbered simply 1, 2, 3,..., as they are in the PDB file. So a translation from the numbering system of Perutz et al to standard numbering is necessary.

SQ20. Use the sequence shown on the PDB page to translate the abbreviations of these five amino acids to standard purely numerical coordinates.

Write these five coordinates down somewhere. You'll be using them in a moment to identify the position of the amino acids in the three-dimensional structure of myoglobin.

Now to find those amino acids.

- Bring up the sequence of myoglobin by clicking **sequence** from the **Tools / Sequence** menu.
- Select the first amino acid in your list of five, using the coordinate to guide you to the proper letter. Drag the mouse over that letter.
- Select the second through fifth amino acids by dragging the mouse over the appropriate letter (again using the coordinate to get you there) *with the shift key depressed*. Unfortunately, if you miss, you'll have to start over by **Select / Clear Selection**. I don't know how to unselect a single amino acid.
- Remove the ribbon format by clicking **hide** from the **Actions / Ribbon** menu. The selected amino acids will disappear. Not to worry.
- Show the atoms of the selected amino acids by clicking **show** from the **Actions / Atoms/Bonds** menu.
- Make the selected atoms spheres by clicking **sphere** from the **Actions / Atoms/Bonds** menu.

SQ21. What can you say about the five conserved amino acids conserved in all globins? Why do you think they're conserved?

Anti-helical residues

Why isn't the entire protein one big helix? Perutz et al propose a partial answer to this question, finding a correlation between the positions of prolines and the ends of helices (see p.672 and p.677). Let's see if that works.

SQ22. What is the PDB ID of whale hemoglobin as described in an article one of whose authors was Max Perutz? Bring up that protein in Protein Explorer. If you already have myoglobin on the screen, first get rid of it by closing the session (File / Close Session).

SQ23. How many polypeptide chains are shown? (to see them more clearly, use Presets / Interactive 1 (ribbon)) Is this what you expected? Explanation?

SQ24. Select all the prolines of hemoglobin (Select / Residue / Pro). Color them yellow (Actions / Color / yellow). Where do the proline residues sit relative to the helices?

SQ25. Why does proline have this property and not some other amino acid? What's so special about proline?

V. Parting thoughts

- The most important lesson to take away from this exercise is that understanding is much deeper if you do something than if you just read about it. It is a rare molecular biology article that can be replicated without a lot of work, but this is quite common in articles focusing on *information*.
- Bear in mind that this article examined only globin sequences. For all you know, the conclusions apply only to globin. You might ask yourself, which conclusions appear to rest on characteristics that should hold in all proteins? Which are merely observations and might be very local in reach?
- From the principles set forth by Perutz et al, do you think it should be possible to predict the three dimensional structure of a protein given only the amino acid sequence of the protein and a knowledge of the properties of amino acids?

References

1. Anonymous. Cracking the Phase Problem. The Nobel Prize in Chemistry 1962, Perspectives. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1962/perspectives.html
2. Perutz MF (1997). The Medical Research Council Laboratory of Molecular Biology. Nobelprize.org. http://www.nobelprize.org/nobel_prizes/medicine/articles/perutz/
3. Green DW, Ingram VM, Perutz MF (1954). The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. Proc Royal Soc London A 225:287-307.
4. Martz E, Reichsman F (2018). Hemoglobin Molecular Structure. <http://bioinformatics.org/jmol-tutorials/jtat/hemoglobin/contents/contents.htm>
5. Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004). UCSF Chimera — A visualization system for exploratory research and analysis. J Comput Chem 25:1605-1612. DOI:[10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084), <https://www.cgl.ucsf.edu/chimera/>