

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 Structure](#).⁴ Don't proceed until you're at this introduction. (If you have trouble running the program, try this [possible fix](#)).

Parts 1 and 2 have the purpose of introducing you to the graphical conventions that will become second nature as you use the protein visualization tools. To this end, it uses simple molecules, e.g. glycine, but don't use your possible lack of interest in glycine to skip over these

preliminaries. 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. Parts 1 and 2 show different representations, e.g. spacefilling and stick. What are the advantages and disadvantages of each? When would you prefer one representation over another.

SQ4. From the stick representation (Part 1, View 3) of glycine its the colors, draw the chemical structure of glycine.

SQ5. What is haloed in Part 2, View 4?

Part 3 gives a useful introduction to, hemoglobin, one of the two featured proteins of the article. This is an excellent time to get yourself introduced.

SQ6. What are the red blobs in Part 3, View 2? What do the blue, blue-green, green, and yellow squiggles represent? Why aren't the red blobs represented as squiggles?

SQ7. Part 3, View 10, shows two hollow frying pans extended towards red and orange blobs. Explain what these shapes represent. What's special about the frying pan molecule that makes it suitable for the purpose, compared to other amino acids?

SQ8. While in the area, rotate the protein until you're looking down the axis of O₂, i.e. the two atoms are one in front of the other. When you look at the protein this way, you'll see four fused frying pans surrounding the spheres. What are they? How are these frying pans chemically related to the two referred to in SQ7?

Part 4 is perhaps the main event. How do amino acids interact locally with one another to form secondary structures?

SQ9. Rotate Part 4, 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).

SQ10. The text associated with Part 4, View 4 describes the red helix as lying at the protein-water interface". What does that mean? Where's the water?

SQ11. 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?

SQ12. Each hydrogen bond shown in Part 4, View 8, stabilizes the helix by connecting what to what? (Look at the colors) If a hydrogen bond begins at amino acid #1, which amino acid will it end at?

Part 5 introduces a concept essential for understanding how secondary structure relates to tertiary structure.

SQ13. What is the specific meanings of the colors shown in Part 5, View 3?

SQ14. The text associated with Part 5, View 4 refers to "the differences between sides". What does this mean? What is its significance? If two helices such as the one shown were floating freely in water, how might the two interact?

Part 6 extends the lessons of Part 5 to a complete protein, hemoglobin. The seven views represent successive slices through hemoglobin, from front to back, revealing different internal cutaway views of the protein. The images are sometimes difficult to absorb, but considering that they are made on the fly using general purpose software, they're pretty amazing!

SQ15. What general lesson can you take away from the slices?

The last part doesn't relate much to Perutz et al, so I'll leave that one for you.

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. The tools are provided by the resource [Protein Explorer](#), available from both the Protein unit web page and the Links page on the course web site. We'll be looking at whale myoglobin and horse hemoglobin.

SQ16. Open up Protein Explorer. What kind of information to 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.

SQ17. 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.

SQ18. 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!

SQ19. How does this Protein Data Base figure compare to Fig. 1 of Perutz et al?

SQ20. Using the two figures, what is the 3rd amino acid of myoglobin in Helix A?

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

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 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.

SQ21. What fraction of globin amino acids would you expect to be found in all instances of the protein?

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 the same places. Why not globin?

SQ22. Comparing the results shown in Table 2 of Perutz and the sequence shown on the PDB page, what fraction of myoglobin is invariant?

If you're like Perutz et al, then you're pretty surprised by this result! This small number of amino acids are probably pretty important 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).

SQ23. Using the nomenclature of Perutz et al, what five amino acids are they?

SQ24. Use the sequence shown on the PDB page to translate the abbreviations of these five amino acids to standard 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 we need to get that three-dimensional structure. Go to Protein Explorer (not FirstGlance), type in the PDB ID of myoglobin, and press Enter. In several seconds you should see a representation of myoglobin spinning on your screen. Click the **Toggle Spinning** button to put the molecule to sleep. Then click the **Quick** tab at the top left of the page, to bring you to the **QuickViews** page.

I presume the representation on the right looks very familiar, but to make it even more familiar, do this:

- On the **SELECT** line, change **what?** to **Protein**
- On the **SHOW AS** line, change **what?** to **Cartoon**
- On the **COLOR BY** line, change **COLOR** to **N→C Rainbow**

You should now be looking at a figure equivalent to Part 4, View 3 of the Hemoglobin tutorial.

You can identify the first amino acid by doing the following:

- In the Commands box of Protein Explorer, type **select 1** and press Enter. You should see a message in the Messages box that 7 atoms have been selected. These are the 7 atoms of amino acid #1.
- In the Commands box, type **color gray** and press Enter. You should see a message in the Messages box that color of the selected amino acid is now gray, and more importantly, you should see a gray segment at one end of the protein.
- Test to see if indeed that segment is the first amino acid by clicking the beginning of the segment. You should see a message in the Messages box identifying the group as **VAL 1**.
- Return to the protein sequence on the PDB page... is the first amino acid valine?

SQ25. If that worked out, then try it with the interesting amino acids. Select the five amino acid coordinates you identified in SQ24 (numbers, not amino acid abbreviations). You can select them all at once, separating the coordinates with commas. Then change their color to purple. Where are the purple regions relative to the heme?

The cartoon view is good for rendering secondary structure, but it does not accurately render distances. To get a better picture of the relationship between the selected amino acids and heme, type in the Commands box **spacefill** and press Enter.

SQ26. Which amino acids are very close to the heme? (Click to identify them). Any that are not?

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.

SQ27. What is the PDB ID of horse hemoglobin as described in an article one of whose authors was Max Perutz? Bring up that protein in Protein Explorer.

SQ28. How many polypeptide chains are shown? Is this what you expected? Explanation?

SQ29. Render the protein as a cartoon, showing the secondary structure, and color the helices N→C rainbow. Then select all the prolines of hemoglobin (select pro**), color them something, and render them according to space fill. Where do the proline residues sit relative to the helices?**

SQ30. Why does proline have this property and not some other amino acid? What's so special about proline?

Positions of hydrophobic and polar residues

The most far-reaching result obtained by Perutz et al from their comparisons of globin is the conservation of hydrophobic residues within the protein. Just as plastic beads aggregate together

to minimize contact with water, so should hydrophobic amino acids, and this could be the dominant force governing protein tertiary structure.

...if it's true. Is it?

SQ31. Render the same hemoglobin protein in Protein Explorer as ball+stick, colored blue (scroll down the color list). Then:

- On the **SELECT** line, chose **Polar**
- On the **SHOW AS** line, choose **Spacefill**
- On the **COLOR BY** line, choose **Red**

(the residues that are not changed, i.e. the hydrophobic residues, remain blue). Describe how the hydrophobic and polar residues are situated in the three dimensional structure.

SQ32. Maybe it's just an optical illusion. As a control, reverse the selection, making hydrophobic residues red and large while polar residues are small and blue. Are your conclusions affected?

VI. 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. Hemoglobin Structure. <http://www.umass.edu/molvis/tutorials/hemoglobin/>