# Molecular Basis for Sickle Cell Disease

**Adapted from Nicholas' Story** by Didem Vardar-Ulu for CH373-S20

**NAME:** **Time Spent:**

**PRE-CLASS WORK:** Please complete this worksheet and print out to bring to your Case Study In-Class Activity on 03/04/2020 based on your group assignment. Please make sure to include your name and the time it took you to complete this worksheet above.

**PART I: COMPILING BACKGROUND INFORMATION FOR THE CASE STUDY:**

1. Introduction to the case: Watch the video titled "Managing Sickle Cell Disease as a Teenager" (https://www.youtube.com/watch?v=iKQmQHh4E2w) and answer the following questions.

   A. What is Nicholas suffering from? Please include in your answer what his diagnosis is as well as what symptoms he has.

   Nicholas is suffering from a blood disorder called Sickle Cell Anemia. His primary symptoms are pain in his body that feels like "pinching" and the pain is often so excruciating he must be hospitalized until the pain subsides. Other symptoms were swelling of the hands and feet as well as fever.

   B. Do Nicholas's parents have the same condition? Are their lives impacted the same way as Nicholas? Why or why not?

   Nicholas's parents do not have the same condition as him. However, they are both carriers of the disorder. His father has the sickle cell trait, but his mother has thalassemia minor. Their lives were definitely impacted by their son having Sickle Cell Anemia as they were often in the hospital and taking extra steps to ease the pain Nicholas was experiencing. On the other hand, their own health stste was not impacted the same way because sickle cell anemia is a recessive genetic disease, being carrier doesn't affect the individual.

2. Introduction to Sickle Cell Disease (SCD):
   a. WATCH: The video titled"Sickle Cell Disease) https://www.youtube.com/watch?v=Y66B7PWrE00&feature=youtu.be
   b. READ: Section 9.6 First "Clinical Insight" in your textbook
   c. READ: The Fact Sheet document at https://ghr.nlm.nih.gov/condition/sickle-cell-disease#genes **paying special attention to the figures embedded in the text** and answer the following questions.

   A. What causes Sickle Cell Disease (SCD)?

Sickle Cell disease is a genetic disorder where at least one of the two Beta-globin subunits of Hbb is mutated at position six (6), causing a Glutamate amino acid to be changed into a Valine amino acid. Sickle Cell Anemia, a type of sickle cell disease, is when BOTH beta-globin subunits are mutated.

   B. How does the sickle cell mutation in hemoglobin (HbS) cause red blood cells to sickle?

When the hydrophilic amino acid Glu is mutated to the hydrophobic amino acid Val, this new hydrophobic amino acid associates with an existing hydrophobic patch on a different Hbb subunit causing a self-assembly of large fibers which distort the shape of the blood cell.

   C. What are some current medications/strategies for curing or treating sickle cell disease?

There is still no cure for SCD, however there are treatment options for those who are suffering from the disease. Staying hydrated is important in decreasing pain associated with HbS. There are medications to alleviate pain as well such as Hydroxyurea to prevent that sickling of the red blood cells (RBCs). Blood transfusions also provide some temporary relief by replacing the sickled cells with wild-type RBCs. This treatment option does not last as RBCs are continuously being produced by the body anyway.

3. Introduction to Hemoglobin Structure:
   READ: Section 9.2 & 9.3 in your textbook and produce a reading log to refer to during class activity and answer the following questions. **Make sure to spend extra time looking at the corresponding figures in detail.**

   A. What is hemoglobin? Where is it present in our body? Why is it so important for us?

It is the iron-containing oxygen-transport metalloprotein in the red blood cells (erythrocytes) of almost all vertebrates as well as the tissues of some invertebrates. Hemoglobin in blood carries oxygen from the lungs or gills to the rest of the body (i.e. the tissues) where it releases the oxygen to permit aerobic respiration to provide energy to power the functions of the organism.

   B. What is the overall composition and structural organization of the hemoglobin molecule?

Hemoglobin has four subunits. Hb is comprised of 2 alpha subunits and 2 beta subunits but they are paired together such that an alpha and a beta operate together as do another pair. Each subunit of hemoglobin can bind one oxygen molecule. Each molecule of hemoglobin can bind four oxygen molecules.

   C. What are the differences in oxygen binding and releasing behavior for the two molecules myoglobin and hemoglobin?

The primary difference is that myoglobin can only bind one oxygen at once whereas hemoglobin can bind four at once. Their binding curves also differ and for an important reason. Myoglobin needs to be able to bind oxygen better than the hemoglobin when the hemoglobin is moving through the tissue capillaries. It is important that myoglobin can bind oxygen better at lower partial O2 pressure because it needs to be able to take O2 from transient hemoglobin. We see these differences in binding curve shapes because Hemoglobin has four binding sites whereas myoglobin only has one. With hemoglobin, the binding of one oxygen molecule increases the likelihood of another oxygen molecule binding to the other three site s because of allosteric effects. Myoglobin does not have this advantage because of its single binding site and thus we don't see a sigmoidal curve

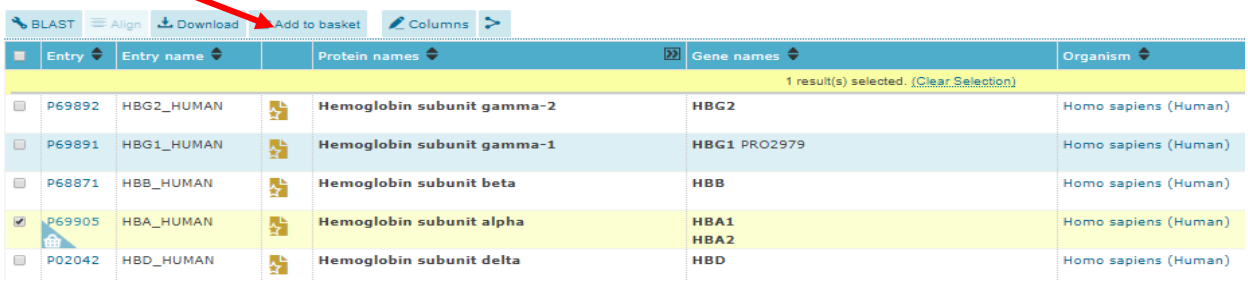
**PART II: COMPILING BIOINFORMATICS DATA ON HEMOGLOBIN**:

This part of your pre-class work is designed to introduce you to several useful bioinformatics tools in order to better understand characteristics of the protein hemoglobin whose structure you will be studying in detail during the in-class portion of this case-study.

You will be using The Universal Protein Resource (UniProt) that provides the scientific community with a comprehensive, high-quality, and freely accessible resource of protein sequence (attributed with a unique identification number) and functional information.

**PLEASE FOLLOW THE STEPS OF THIS GUIDED TUTORIAL AND RECORD YOUR ANSWERS BELOW EACH REQUESTED STEP INDICATED IN BLUE:**

1. Go to UniPro Home Page (https://www.uniprot.org/) and enter “human hemoglobin” in the text box and click “search”
2. Find the entry for “Hemoglobin subunit alpha” (Entry ID: P69905).
3. Click on the small square on the left of the entry number to select the item and the click on the “Add to basket” tab on the top of the list.




4. Click on the Entry ID P699050 that will take you to a page with extensive information about for human hemoglobin alpha subunit.
5. Take some time to familiarize yourself with the kinds of information this file contains.
6. Scroll down to the “Sequence” section of the file and click the on the blue tab titled “FASTA” which will open a page that contains the single letter amino acid sequence for human hemoglobin alpha subunit. In this format the first line starts with the character “>” followed by some informational text, indicating that, that line is for informational content only and will be ignored by other programs running their own algorithms. This line is followed by the single letter amino acid sequence of the protein.
7. **Copy this sequence below.**
   **Amino Acid Sequence of for human hemoglobin alpha subunit (in one-letter code):**
   MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHV
   DDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR

8. Go back to the previous page (human hemoglobin alpha subunit homepage). To the right of the Sequence box, there is a scroll down menu (showing BLAST as default).


9. Click on the arrow to activate the menu and select **ProtParam** and click “GO”
10. Click submit at the bottom of the page that opens. *Note that by default the complete sequence of human* hemoglobin alpha subunit *will be used for calculations. However, you also have the option to analyze only a certain piece of the protein sequence by selecting the fragment of interest from the displayed options.*
11. **Using the Results of the ProtParam, fill out the first row of the table 1 below**
12. Go back to your “human hemoglobin” search results (after Step 1)
13. Find the entry for “Hemoglobin subunit beta” (Entry ID: P68871)

14. Click on the small square on the left of the entry number to select the item and the click on the "Add to basket" tab on the top of the list.
15. Click on the Entry ID P68871 that will take you to a page with extensive information about for human hemoglobin beta subunit.
16. Take some time to familiarize yourself with the kinds of information this file contains.
17. Scroll down to the "Sequence" section of the file and click the on the blue tab titled "FASTA" which will open a page that contains the single letter amino acid sequence for human hemoglobin beta subunit.
18. **Copy this sequence below.**
    **Amino Acid Sequence of for human hemoglobin beta subunit (in one-letter code):**
    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSD
    GLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

19. Go back to the previous page (human hemoglobin beta subunit Homepage). To the right of the Sequence box, there is a scroll down menu (showing BLAST as default).
20. Click on the arrow to activate the menu and select ProtParam and click "GO"
21. Click submit at the bottom of the page that opens.
22. **Using the Results of the ProtParam, fill out the first row of the table below**

**TABLE 1:**

| Entry ID # (Protein Accession #) | Name of Protein | # of amino acids | MW g/mol (Da) | Calculated pI | # neg (asp & glu) | # pos (arg & lys) | Aromatic amino acids | | | Aliphatic index |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | tyr | phe | trp | |
| P699050 | Alpha Hemoglobin Human | 142 | 15257.55 | 8.72 | 12 | 14 | 3 | 7 | 1 | 90.77 |
| P68871 | Beta Hemoglobin Human | 147 | 15998.41 | 6.74 | 15 | 14 | 3 | 8 | 2 | 93.47 |

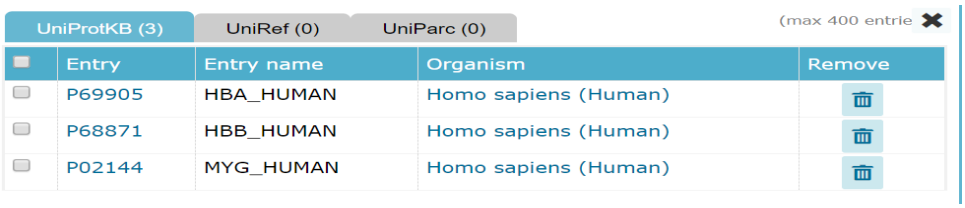
23. **Based on the table you filled out above list which properties are similar and which ones are different between the two subunits.**

The beta chain is slightly longer (5 more amino acids) and 70 Da heavier than the alpha chain. Both seem to have the same number of positive aas but different number of negative amino acids. The calculated pIs are different such that alpha has an alkaline pI whereas beta has an acidic pI which is consistent with the increased # of acidic amino acids in the beta chain.

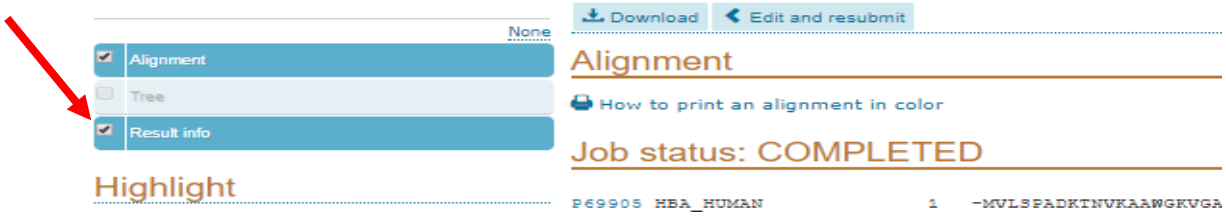
24. Go back to to UniPro Home Page (https://www.uniprot.org/) and enter "human myoglobin" in the text box and click "search"
25. Find the entry for "Human myoglobin" (Entry ID: P02144).
26. Click on the small square on the left of the entry number to select the item and the click on the "Add to basket" tab on the top of the list.
27. Click on the Entry ID P02144 that will take you to a page with extensive information about for human myoglobin.
28. Scroll down to the "Sequence" section of the file and click the on the blue tab titled "FASTA".
29. **Copy this sequence below.**

**Amino Acid Sequence of for human myoglobin (in one-letter code):**
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILK
KKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

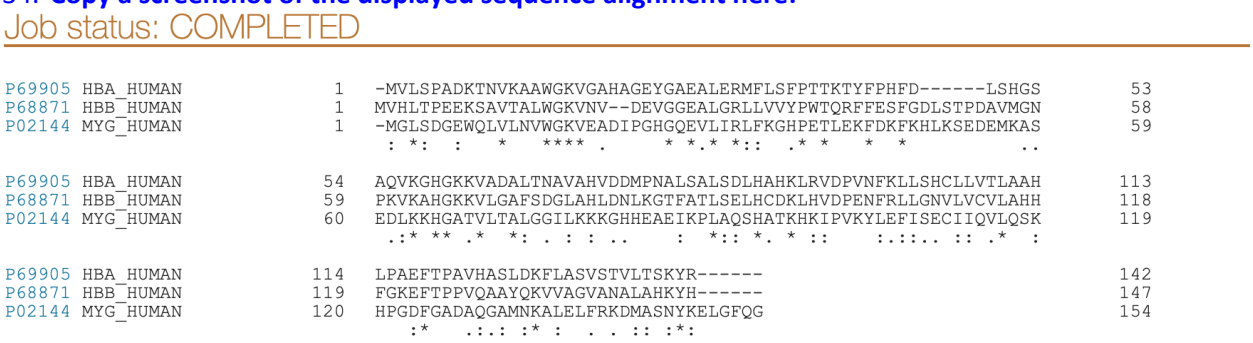
30. Go up to the top right corner of the page and click on "Basket." This opens a small window that lists all the entries in the Basket.




31. Select/check all three entries and click on the "Align" button at the bottom left of the window.
32. When the alignment procedure is complete, a page displaying the arrangement of sequences from all of the entries will appear. (Be patient it might take a few minutes).
33. To ensure that all the necessary information is displayed, make sure that the "Result Info" optionson the top left side of the page is checked.




34. **Copy a screenshot of the displayed sequence alignment here:**




Sequence alignments can provide valuable insights into the evolution of a particular protein. Protein sequences are typically aligned by comparing amino acid identities, amino acid types, amino acid similarities, and protein structural motifs or domains. A set of symbols are typically used to identify identical and similar amino acids in aligned sequences; the UniProt tools use the "*", ":", and "." symbols. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties. Therefore the hierarchy of conservation using these symbols is * (identical) > : (colon) > . (period).

35. **How many positions are identical in the three sequences? (Look at the report section for the answer)**

26 identical positions

36. **How many positions are similar in the three sequences? (Look at the report section for the answer)**

48 similar positions

37. Use the "Annotation" tools on the left side of the page (under the "Highlight" heading) to selectively highlight amino acids with specific properties (e.g., metal binding, aromatic, etc.).

38. **Write down one similarity and one difference between the three sequences from the "annotation" category and one similarity and one difference from the "Amino Acid property" category.**

**This answer will vary from student to student.**

**Now that you have compiled a significant amount of information for the human hemoglobin subunits, from their primary structures, let's use another online program to predict secondary structures for these subunits based on sequence information.**

39. **Go to [http://www.compbio.dundee.ac.uk/jpred/](http://www.compbio.dundee.ac.uk/jpred/)**

40. Paste the human myoglobin sequence from step 29 into the appropriate field and click "Make a Prediction"

You will get a response saying "Match found in PDB" meaning that there is an experimentally determined structure for the protein sequence you submitted that is already deposited in the protein data bank (PDB) that gives far more accurate structural information than the prediction tool.

41. For demonstration purposes select the option to carry out the prediction by clicking on "continue". Be patient. It will take a few minutes for the computer to do the computation and display it even after it says 100% completed.

42. **Insert a screenshot of the graphical results below:**


43. **Using the information included in the 4th line "jnetpred" answer the following questions. Please note that a green arrow is used to represent a beta strand and a red cylinder is used to represent an alpha helix when predicted by the jnetpred algorithm.**
    a. **How many alpha helices are predicted for the human myoglobin?**
       **6**
    b. **What are the residue ranges for each helix (i.e. list the first and last residue number for each helix)?**
       First == 5-18 Second == 22-37 Third == 60-76 Fourth == 83-96 Fifth == 105-119 Sixth == 126-150

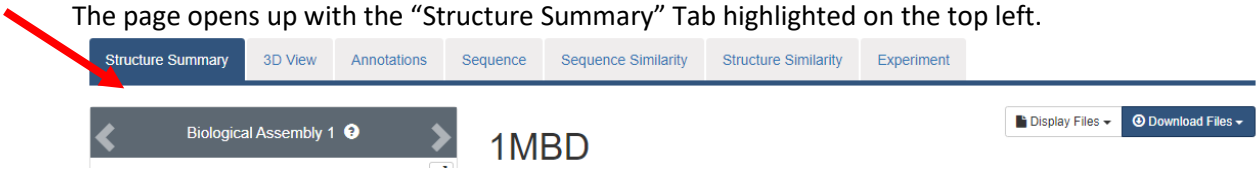
**PART III: BECOMING FAMILIAR WITH MOLECULAR STRUCTURE DATA FILES**:

This part of your pre-class work is designed to introduce you to the Protein Data Bank (RCSB PDB) ([https://www.rcsb.org/](https://www.rcsb.org/)) that hosts experimentally determined 3D structure information of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease. During your in-class activity

you will be using this resource extensively, so please spend some time to familiarize yourself with the what information is archived in this data bank and how it is presented to the users by completing the following steps.

1. **Go to the RCSB Protein Data Ba**nk (https://www.rcsb.org/) input 1MBD into the PDB ID or Text search bars in the top menu and click "GO".

The page opens up with the "Structure Summary" Tab highlighted on the top left.




***What can you find on the structure summary page?***

1. **Title (under the displayed PDB ID)** - that tells you what the structure is about
2. **Snapshot of the molecule (on the upper left)** - The structure of the molecule/complex in cartoon representation.
3. **Deposition Date and Authors (under the title)** – when the structure was deposited to PDB and who solved the structure
4. **Literature (under its own banner)** – for access the article that describes the structure.
5. **Macromolecules (under its own banner)** – All proteins and nucleic acids present in the structure are listed here. Each unique type of macromolecule or molecular chains is listed as a separate entity. There may be multiple copies of a molecule in the structure.
6. **Small molecules (under its own banner)** – All ligands, ions, cofactors, inhibitors that are present in the structure are listed here.
7. **Experimental Data and Validation (under its own banner)** – describe details about the experiments that led to the structure determination and a slider that provides insights about the quality of the structure and its agreement with the experimental data and geometric standards.

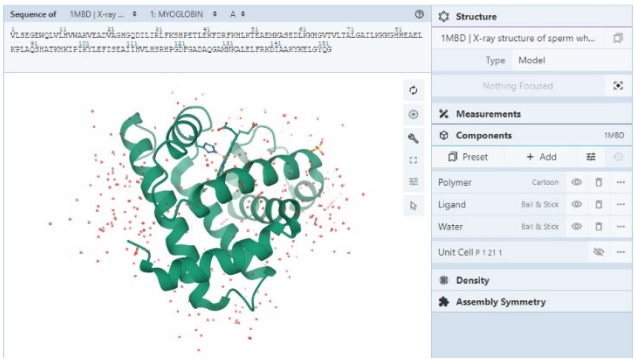
See http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction for details

**2. Use information from the structure summary page for 1MBD to complete the following table.**

**TABLE 2:**

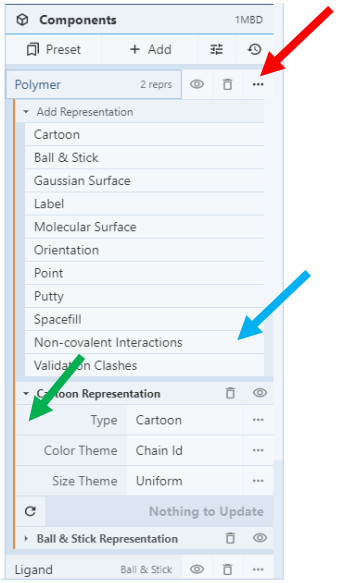
| | |
|---|---|
| **Structure Title** | X-ray structure of sperm whale deoxymoglobin refined at 1.4A resolution |
| **Authors of entry** | Phillips, S.E.V. |
| **Macromolecules (#, Name, and chain ID)** | 1, Myoglobin, A |
| **Small molecule (# and Name)** | 2, $SO_4$, HEM |

3. **Click on the 3D view** tab on the top of the RCSB PDB structure summary page or the hyperlinked word "Structure", below the snapshot. This should open an interactive view that shows the macromolecule and the associated small molecules based on the actual coordinates submitted to the PDB. Above the molecule you will also see its sequence. The various pull-down menus on the right hand side of the page allow you to



explore the structure interactively by viewing and/or hiding each element (polymer, ligand, and associated water molecules), moving, rotating, changing the colors, representations etc. You can get a brief explanation for the associated function of each button when you rest your mouse over each button. Spend some time to familiarize yourself with this interface.

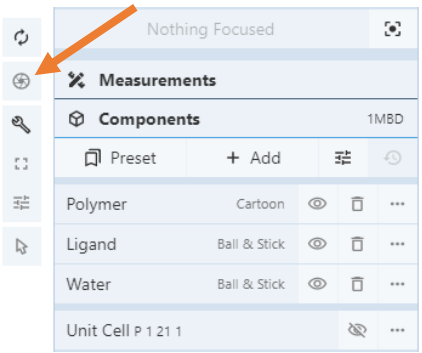
**4.** **Explore** the molecular structure of myoglobin molecule trying out different representations Different representations are activated by clicking on the "…" (red arrow) which allows you to "Add representation". To remove an existing representation, use the garbage can symbol (blue arrow) next to the corresponding representation from the growing list below. For each representation you can access further options by clicking on the small arrow on the left of the representation name (green arrow).




As you go through each representation note down what kind of molecular information about the protein or its ligand is best in emphasized by each representation.

A. View the "polymer" (in this case your myoglobin molecule" using the different "style" options: Cartoon, Ball & Stick, Molecular Surface, Spacefill, etc.)
B. Repeat the same exercise with the ligand
C. Keeping the polymer "representation" as "cartoon" and "ligand" as "ball & stick" try the different color options from the "set coloring" pull down activated by clicking on the "…" (red arrow). You can further narrow your selection using the small arrows on the left of each new menu that opens up. Ex: Set coloring→ Residue property → Residue name
Bring your cursor over each colored region and read the residue information displayed. What color corresponds to what class of amino acids?

**5.** Take a screenshot (orange arrow) of your favorite view of the structure that you believe would be most effective in communicating information about the overall shape and the predominant secondary structure of myoglobin as well as where the heme group is located with respect to the protein.



6. **Insert the snapshot below and include a figure legend describing the different elements it highlights about the structure.**

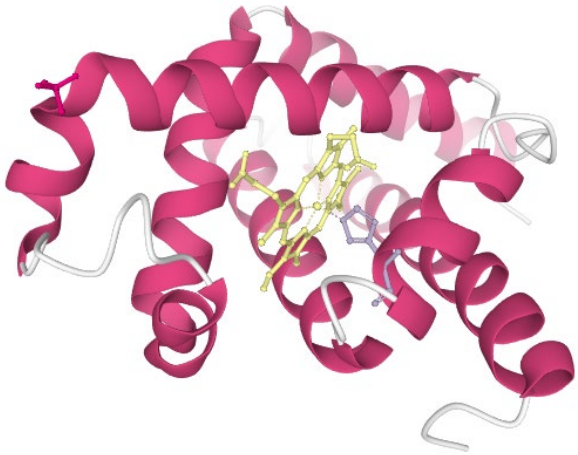
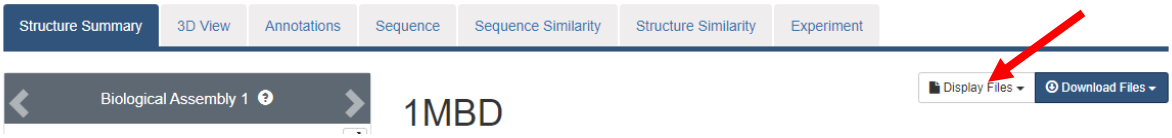


**Figure legend:** Three dimensional structure of sperm whale myoglobin (PDB ID 1MBD) shown in fuchsia cartoon representation. The structure is composed of eight alpha helices The heme group associated with the protein is shown in yellow and histidine side chain that coordinates the iron within the center of the heme group is shown in gray ball and stick representation.

7. Go back to the 1MBD Structure Summary Tab on the initial 1MBD View page and click on "Display Files" tab.




8. Select "PDB" Format which contains the experimental data information used to create the structural views you have been investigating in an easily readable text file. Please note that the current format that is used for data transactions within PDB is the mmCIF is machine readable and does not have some of the limitations of the PDB file format.

***What can you find in the PDB file?***

**1. Introduction** & **Title Information** - that tells you background information about the molecule whose structure is included in the file and ends with 'remark' entries.

**2. Primary Structure Section** – includes cross-ref sequence information between the PDB and sequences deposited in other databases.

**3. Heterogen Section** – that describe non-standard residues, such as prosthetic groups, inhibitors, solvent molecules, and ions for which coordinates are supplied.

**4. Secondary Structure Section** – that identifies the positions of secondary structures such as helix and sheet.

**5. Connectivity Annotation Section** – that identifies disulfide bonds, specify connectivity between residues that is not implied by the primary structure, as well as positions for cis conformations.

**6. Crystallographic and Coordinate Transformation Section-** that describes the geometry of the crystallographic experiment and the coordinate system transformations

**7. Coordinate section** – that contains the collection of atomic coordinates and models used.

**8. Connectivity Section**- that provides information on additional atomic connectivity

See http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html for details

**9. Use information from the PDB file for 1MBD to complete the following table (Table 3).**

**TABLE 3:**

| | |
|---|---|
| **Name of the molecule and its function** | Deoxymyoglobin in whale sperm |
| **Non-standard residues (identified with entries that start with 'HET' (#, symbol, name)** | $SO_4$ (Sulfate ion)<br>HEM (Protoporphyrin IX containing Fe) |
| **# of chains and chain ID** | 1 Chain, A |
| **# and type of secondary structures identified , and residue range for each secondary structural element** | Helix 1, SER 3 - GLU 18<br>Helix 2, ASP 20 - SER 35<br>Helix 3, HIS 36 - LYS 42<br>Helix 4, THR 51 - ALA 57<br>Helix 5, SER 58 - LYS 77<br>Helix 6, LEU 86 - THR 95<br>Helix 7, PRO 100 - ARG 118<br>Helix 8, GLY 124 - LEU 149 |
| **All the structural information you can gather for ATOM #100 (What is the atom, which residue it belongs to, what chain it belongs to, what are its Cartesian coordinates?)** | Nitrogen, Valine residue, Chain A. Residue number 13 with coordinates 18.828, 12.610, 1.00, 13.20. |

**10. How do the identified secondary structural elements in the PDB file compare to the elements identified by the "Jpred" program (steps 39-43)? What might be a plausible explanation for any discrepancy?**