# **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:

- Introduction to the case: Watch the video titled "Managing Sickle Cell Disease as a Teenager" (<u>https://www.youtube.com/watch?v=iKQmQHh4E2w</u>) 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.
  - B. Do Nicholas's parents have the same condition? Are their lives impacted the same way as Nicholas? Why or why not?
- 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 <u>https://ghr.nlm.nih.gov/condition/sickle-cell-disease#genes</u> paying special attention to the figures embedded in the text and answer the following questions.
  - A. What causes Sickle Cell Disease (SCD)?
  - B. How does the sickle cell mutation in hemoglobin (HbS) cause red blood cells to sickle?
  - C. What are some current medications/strategies for curing or treating sickle cell disease?
- 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?
- B. What is the overall composition and structural organization of the hemoglobin molecule?
- C. What are the differences in oxygen binding and releasing behavior for the two molecules myoglobin and hemoglobin?

#### 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 (<u>https://www.uniprot.org/</u>) 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.

	ЪВ		Nign 土 Download	Add to	o basket 🖉 Columns 🗲			
		Entry 🗢	Entry name 🗢		Protein names 🗢	>>	Gene names 🗢	Organism 🗘
	1 result(s) selected. ( <u>Clear Selection</u> )							
1		P69892	HBG2_HUMAN	<mark>∱</mark>	Hemoglobin subunit gam	nma-2	HBG2	Homo sapiens (Human)
1		P69891	HBG1_HUMAN	삹	Hemoglobin subunit gam	nma-1	HBG1 PRO2979	Homo sapiens (Human)
I		P68871	HBB_HUMAN	Å	Hemoglobin subunit beta	a	нвв	Homo sapiens (Human)
1	•	P69905	HBA_HUMAN	삹	Hemoglobin subunit alph		HBA1 HBA2	Homo sapiens (Human)
1		P02042	HBD_HUMAN	<b>ئ</b> د	Hemoglobin subunit delt	ta	HBD	Homo sapiens (Human)

- 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):
- 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).

Sequence (1+) <sup>i</sup>		
Sequence status <sup>i</sup> : Complete. Sequence processing <sup>i</sup> : The displayed sequence is fu This entry has 1 described isoform and 2 potential i	rther processed into a mature form. oforms that are computationally mapped.  Show all	
P69905-1 [UniParc] J FASTA M Add to basket « Hide 10 20 30 MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSI 60 70 80 HGSAQVKGG KKVADALTNA VAHVDOMPNA LSALSDI 11 12 0 130 LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVI	90 100 HAH KLRVDPVNFK 140	Length: 142 Hass (Da): 15.258 Last modified: January 19.2007 - v2 Checksum: <sup>1</sup> 1554-0665738BBAE BLAST ProtScale Compute pi/NW PeptideMass PeptideCutter

- 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):

- 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 #	Name of	# of	MW	Calcula	# neg	# pos	Aro	matic		Aliphatic
(Protein	Protein	amino	g/mol (Da)	ted pl	(asp & glu)	(arg & lys)	ami	no aci	ds	index
Accession #)		acids					tyr	phe	trp	
P699050										
P68871										

- 23. Based on the table you filled out above list which properties are similar and which ones are different between the two subunits.
- 24. Go back to to UniPro Home Page (<u>https://www.uniprot.org/</u>) 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):

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.

UniProtKB (3)		UniRef (0) Uni	Parc (0) (n	(max 400 entrie 🗙		
	Entry	Entry name	Organism	Remove		
	P69905	HBA_HUMAN	Homo sapiens (Human)	Ē		
	P68871	HBB_HUMAN	Homo sapiens (Human)	亩		
	P02144	MYG_HUMAN	Homo sapiens (Human)	亩		

- 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.

	None	± Download < Edit and resubmit
Alignment		Alignment
Tree		How to print an alignment in color
Result info		Job status: COMPLETED
		JOD Status. COMPLETED
Highlight		P69905 HBA_HUMAN 1 -MVLSPADKINVKAAWGKVGA

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 strongly similar properties. A . (period) indicates symbols is \* (identical) > : (colon) > . (period).

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

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/

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 4<sup>th</sup> 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?
  - b. What are the residue ranges for each helix (i.e. list the first and last residue number for each helix)?

## 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) (<u>https://www.rcsb.org/</u>) 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 Bank (<u>https://www.rcsb.org/</u>) 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 s	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 (und	der the title) – when the structure was deposited to PDB and who
solved the structure	
4. Literature (under its own banner)	<ul> <li>for access the article that describes the structure.</li> </ul>
5. Macromolecules (under its own b	anner) – All proteins and nucleic acids present in the structure are
listed here. Each unique type of mac	romolecule or molecular chains is listed as a separate entity. There may
be multiple copies of a molecule in the	ne structure.
6. Small molecules (under its own ba	anner) – All ligands, ions, cofactors, inhibitors that are present in the
structure are listed here.	
7. Experimental Data and Validation	n (under its own banner) – describe details about the experiments that
led to the structure determination ar	nd 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/gu	ide-to-understanding-pdb-data/introduction for details

2. Use information from the structure summary page for 1MBD to complete the following table (Table 2). TABLE 2:

Structure Title	
Authors of entry	
Macromolecules (#, Name, and chain ID)	
Small molecule (# and Name)	

- **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).

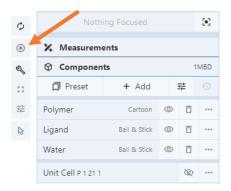
As you go through each representation note down what kind of molecular information about the protein or its ligand is best in emphasized by each representation.

- 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  $\rightarrow$  Residue property  $\rightarrow$  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.



Component     Componen	its		1	MBD					
미 Preset	+ Add		輩	Ð					
Polymer	2 reprs	0	Ō						
✓ Add Representation									
Cartoon									
Ball & Stick									
Gaussian Su	rface								
Label									
Molecular Su	urface								
Orientation									
Point									
Putty				_					
Spacefill				K					
	nt Interactions	5							
Validation Cl	ashes								
▼ Cr toon Repr	resentation		Ô	0					
Ту	pe Cartoor	n							
Color Ther	ne Chain lo	ł							
Size Ther	ne Uniform	ı							
G	Noth	ing t	o Up	date					
Ball & Stick F	Representation		Ō	٢					
Ligand	Ball & Stick	0	Ô						

- 6. Insert the snapshot below and include a figure legend describing the different elements it highlights about the structure.
- **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	
Non-standard residues (identified with entries that start with 'HET' (#, symbol, name)	
# of chains and chain ID	
# and type of secondary structures identified , and residue range for each secondary structural element	

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?)	

**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?