$$\boxed{\textbf{LABORATORY 6}}$$

Gaining a structural perspective on the molecular basis of inherited diseases

Objectives: By the end of this laboratory exercise, you should be able to:

- obtain a protein's three-dimensional coordinates from the Protein Data Bank;
- use PyMOL to examine the effects of a mutation on the non-covalent interactions present in a protein's three-dimensional structure;
- work with your team to develop a hypothesis that relates clinical data to the effects of a mutation on a protein's structure and function; and
- prepare a short movie or infographic in which you justify your team's hypothesis to your peers.

Structure of this laboratory exercise:

This laboratory exercise has been designed to give each team the opportunity to apply the concepts we will study together this semester on protein structure and function to real-world medical examples. This exercise focuses on the effect of mutation on the structure and activity of enzymes involved in important pathways in human metabolism. Working with your team, you will be asked to complete the following three parts of this exercise at different times during the semester:

- **Part A: Use bioinformatics and structural biology tools to gain a structural perspective on the molecular basis of inherited diseases**

Your team will choose a human genetically-inheritable disease of interest from Appendix A and attend a tutorial where you will learn how to use various bioinformatics and structural biology tools to study the study the effects of mutations on protein structure and function. These tools are now used on an almost daily basis by most researchers in biochemistry, genetics and structural biology.

- **Part B: Relate clinical data with the effects of a mutation on a protein's structure and function**

This semester, we will explore various metabolic pathways that are critical for meeting the energy needs of the human body. As we spend time together in class studying these pathways, your team will be asked to use the KEGG database to learn more about the specific role the enzyme relating to your chosen topic plays in metabolism. You will then be asked to analyze clinical data to help you develop a hypothesis that relates these data to the effects of the mutation on your chosen enzyme's structure and function

- **Part C: Prepare a short movie or infographic to present your hypothesis**

Your team will be asked to prepare a short movie or infographic in which you justify the hypothesis that you developed to your peers. This will give you a great opportunity to gain experience in disseminating scientific information and defending a hypothesis.

Overall goal of this laboratory exercise:

The overall goal of this laboratory exercise is to help you gain a deep understanding of the relationship between protein structure and function. Through this project, you will use various bioinformatics tools that will allow you to explore the effects of mutations at the atomic level on protein structure. From the observations you will make in combination with the information you will obtain from the peer-reviewed literature, metabolic databases and clinical data, you will be asked to describe how the mutation affects protein structure and function, and how these effects explain the occurrence of a specific genetically-inheritable disease and its symptoms.

Part A: Using bioinformatics and structural biology tools to gain a structural perspective on the molecular basis of inherited diseases

Over the next few weeks, you will have the opportunity to gain a structural perspective on the molecular basis of genetically-inherited diseases. As you saw in your introductory genetics course, human genetically-inherited diseases are caused by DNA sequence variations. Although disease-causing DNA sequence variations can occur in both non-coding and coding regions of the genome, the majority of characterized mutations occur in the coding region of genes since they most often result in easily observable effects on the structure and function of proteins. For this laboratory exercise, we will focus on genetically-inheritable diseases that are caused by this type of mutation. More specifically, we will focus on genetically-inheritable diseases that result from a missense mutation. Recall that a missense mutation is a change in the nucleotide sequence of a gene, where one or more nucleotides are replaced by others. This mutation results in a new codon, which causes a different amino acid to be inserted into the growing polypeptide chain during translation.

(A)   Selecting your topic

Select your topic from the list of human genetically-inheritable diseases available in Appendix A. There are eleven topics from which to select, and only one pair per laboratory section can work on each topic. So, topic selection is first come, first served.

Also included in Appendix A are the name for the file that contains the protein structure coordinates for the **wild-type** protein (*i.e.*, it is a file with a ".pse" file name extension). You will need this file for your work with the protein visualization software PyMOL. You will also see from the appendix that the mutation you will study is listed next to each topic. **Although there may be many different missense mutations associated with the topic you choose, you are asked to please only work on the mutation listed in Appendix A.**

Appendix A also contains one seed reference for each disease, to help you get started in locating background information on your topic. Please note, however, that the seed reference does NOT discuss

the mutation you are asked to study. Most mutations listed in Appendix A are not yet well characterized, and as such, there is very little published literature available on these. The choice of these mutations was intentional – we want you to gain experience formulating hypotheses and defending them with all currently available data to your peers. The ability to formulate hypotheses and properly support them is important to many different career paths. Please use this laboratory exercise as an opportunity to keep improving this important skill!

To help you select your topic, you can find more information on the following website:

- "Genes and Diseases" list on the NCBI website (http://www.ncbi.nlm.nih.gov/books/NBK22183); and
- "Genetics Home Reference" website (http://ghr.nlm.nih.gov/).

Now that you have selected your topic, we will focus on learning how to use various bioinformatics and structural biology tools to study the effects of the mutation on protein structure and function.

(B)   Studying the protein's structure and physicochemical properties

To help you complete your project, you will be guided through the use of all of the tools required for completion of this exercise using the K-Ras protein, which has been implicated in lung cancer. To make it easier for you to use the tools, screenshots using the K-Ras example have been inserted at appropriate locations in the text below.

(1)   Finding peer-reviewed background information on your topic

In Appendix A, you were provided with one "seed" reference article to help get you started on your project. However, you will need to find further articles to gain sufficient background information to answer all of the questions in this laboratory exercise and present these data to your peers. To gain further background information on your chosen topic, please make use of the following two databases, to which you should have already been introduced in your introductory genetics course:

- PubMed – This is a database that contains millions of citations from MEDLINE and other life science indexes for biomedical journals that date back to the 1950s. This database is extremely useful as it easily and quickly allows you to locate published information on your selected topic. Additionally, PubMed includes links to full-text articles and other related resources. The University of Calgary has electronic subscriptions to many journals, which allows you to quickly download a copy of an article. You can access PubMed through the following address: http://www.ncbi.nlm.nih.gov/pubmed.

- OMIM – OMIM stands for Online Mendelian Inheritance in Man and is available from the NCBI website using the following link: http://www.ncbi.nlm.nih.gov/omim. OMIM is a database containing annotated records of human genes and genetic diseases that is authored and edited by Dr. Victor A. McKusick and his colleagues at John Hopkins University. Each record provides textual information and references about the gene (*i.e.*, location, mapping, history, etc.) or disease

(*i.e.*, clinical features of the disease, inheritance pattern, etc.), as well as links to many references available through PubMed.

Your task

Using peer-reviewed literature, please complete Table 6-1 below.

Table 6-1: Background information on inheritance pattern of chosen topic

Question	Answer
What is the inheritance pattern of the disease? (*i.e.,*)	CIRCLE your answer below: dominant OR recessive autosomal OR sex-linked
What are the symptoms associated with the disease?	
Are there any therapies currently used to treat the disease? If so, please describe at most 2.	
On which chromosome is the gene responsible for the disease located? Make sure the gene you are looking up is the one that results in the expression of the protein listed in Appendix A for the topic you chose.	
To which class does the enzyme listed in Appendix A for your topic belong?	
What is the function of the enzyme? What is the reaction it catalyzes?	
What are the amino acids that make up the **catalytic site** of the enzyme for your chosen topic? These residues are the ones that allow the enzyme to catalyze a specific reaction.	

(2)   Analyzing protein structure coordinate files (PDB files)

Background

A protein's function is intricately linked to its structure, and the loss of a protein's three-dimensional structure usually results in a complete loss of function of the protein. As such, protein structural information is very important to gaining a deep understanding of the function of a protein. This is particularly true for enzymes, which are involved in the catalysis of various reactions. Due to the importance of structure in defining function, pharmaceutical companies invest many research dollars towards protein structure determination each year.

The three-dimensional structure of a protein can be experimentally determined using two main techniques: X-ray crystallography and NMR. All protein structures that will be used in this laboratory exercise were determined using X-ray crystallography.

To determine a protein's structure through X-ray crystallography, a pure sample of the protein must be induced to form crystals (see Figure 6-1). Protein crystals are very tiny, often measuring no more than 0.5 mm in diameter.
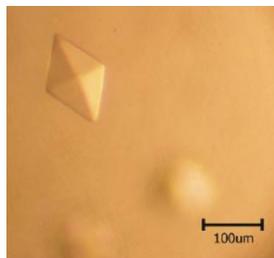
Figure 6-1: Crystals of RNA-dependent RNA polymerase from feline calicivirus.

Once a crystal is sufficiently large, it is placed in a narrow beam of X-rays, and the atoms within the crystal scatter the X-rays. The scattered X-rays can be recorded electronically on an image plate or on a piece of X-ray film. These form what is known as an **X-ray diffraction pattern**, where each diffracted X-ray forms a spot (see Figure 6-2). Analysis of these X-ray diffraction patterns, which involves measuring the intensity and position of each spot, yields a three-dimensional map of electron density for the crystallized protein. The electron density reveals the location of atomic nuclei; this information is used to slowly piece together the protein's structure. The example shown in Figure 6-2 shows how this process was used to determine the structure of the heme prosthetic group in myoglobin.

The level of detail that you can see in the map depends on the quality of the crystal, and data quality is known as "**resolution**". As shown in Figure 6-3, the greater the resolution of the crystal (*i.e.*, resolution of 1.5 Å as opposed to 5 Å), the more detailed the map and the better you can see the location of all of the amino acids and their atoms in the polypeptide chain. "High resolution data" from good-quality crystals provide much more reliable three-dimensional structures than those obtained from "low resolution data".

(a)　　　　**(b)**

(c)　　　　**(d)**
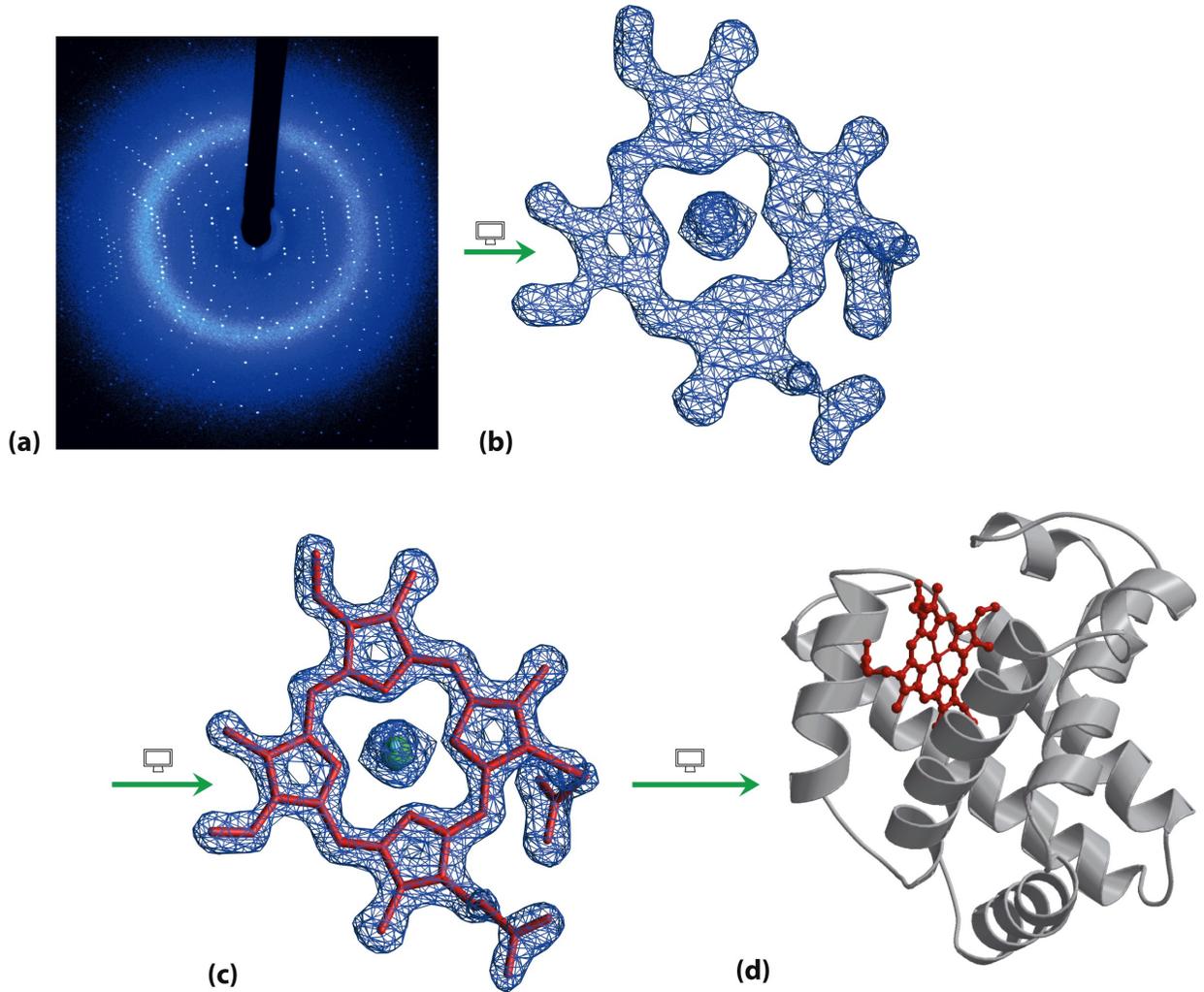
Box 4-5 figure 1
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Figure 6-2: An X-ray diffraction pattern. Reproduced from Nelson and Cox (2013).

5.0 Å　　　3.0 Å　　　1.5 Å
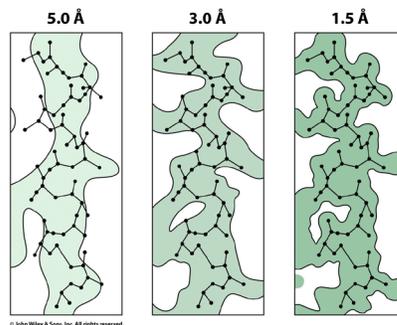
Figure 6-3: Protein structure at different resolutions. Reproduced from Pratt and Cornely (2011).

Once the three-dimensional coordinates of all amino acids that comprise a protein are determined based on electron density, they are stored in a file called a "**PDB**" file. PDB files are usually named as xxxx.pdb, where the four x's correspond to a combination of numbers and letters that is unique.

PDB files containing the three-dimensional coordinates of proteins are deposited into a database called the Protein Data Bank ("PDB"), which is accessible from the Research Collaboration for Structural Bioinformatics website (RCSB, available at http://www.rcsb.org/pdb/home/home.do). All PDB files contained in this database are carefully checked and reviewed to make sure that there are no errors before being made available to the public.

The PDB is available for free and you can search the database by protein name, by keyword or by PDB file name (referred to as "PDB ID" on the website). For this assignment, you will be asked to search the PDB using the PDB ID available from Appendix A for your selected topic.

As you will see below, the PDB file for your protein also contains a lot of information about the various levels of protein structure, the quality of the data used to generate the three-dimensional coordinates and other information relating to the experimental conditions used to generate the data.

Your task

To help you use the PDB website and its associated tools, we will work with the file 3GFT.pdb, which contains the three-dimensional coordinates of the human K-Ras protein in complex with a GTP analogue.

Read through each step carefully and use the PDB file code that contains the three-dimensional coordinates of the protein for your chosen topic.

Working as a team, record the answers to each of the following questions in the space provided.

> **a. Retrieving the three-dimensional coordinates of your protein in the PDB**

Go to the RCSB website (http://www.rcsb.org/pdb/home/home.do) and enter your PDB ID in the search box at the top of the page (see Figure 6-4). Recall that these coordinates are the coordinates for the **wild-type protein**. The PDB file that we will use for K-Ras is 3GFT.pdb (Note: only enter 3GFT in the search box, and NOT 3GFT.pdb).
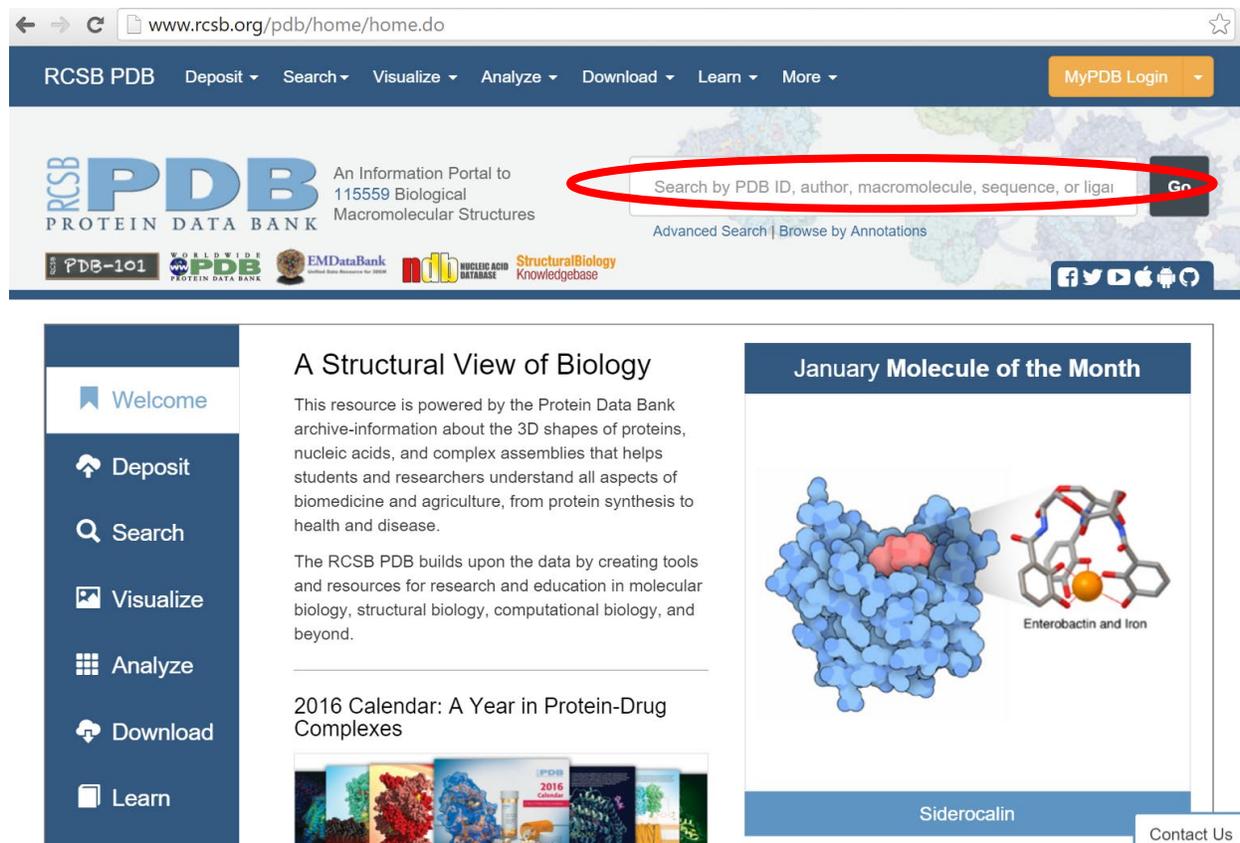
Figure 6-4: Screenshot of PDB website, with the search box highlighted in red.

b.   What is the resolution of the data?

This information can be found in the main page you will obtain after you input your PDB code. You can find this information in the "Experimental Data Snapshot" section located near the top of the page (see Figure 6-5). Recall from above that this can be used as a measure of confidence in the data. Data that are of 1-2.5 Å resolution are considered quite reliable.

Figure 6-5: Screenshot of the Experimental Data Snapshot, with the resolution highlighted in red.

The resolution for the crystal data for our topic is _____.

        c.   **Which expression system was used to overexpress the protein that was used for crystallization?**

You can obtain this information immediately above the "Experimental Data Snapshot" portion of the website, where you found the resolution (see Figure 6-6). Knowing the source of the protein used for crystallization can help determine whether any modifications to the protein's primary structure or post-translational modifications needed to be made. These modifications may or may not affect function.

Figure 6-6: Screenshot of the Source box, with the expression system used highlighted in red.

The expression system used to produce our protein of interest is _____.

Does your seed reference indicate whether any modifications had to be made to the protein for its expression?

_____

_____

     d.   **Are there any ligands bound to your protein?**

As shown in Figure 6-7, the PDB record includes a table entitled "Small Molecules". Scroll down a little from your current position to find it. These ligands could be cofactors and/or substrates.

Note of caution: Not all molecules listed in this table will be "true" ligands. Some of the molecules are "crystallization artefacts". This means that the molecules were added ONLY to help promote crystallization, but are not true, biological ligands (*i.e.*, the fact that they bind to the protein does not have a biological role). To determine whether the molecule is a true ligand, you will to refer to the "seed" article you were given in Appendix A. Please pay particular attention to the section of the article that describes how the protein was crystallized. The article will contain information letting you know whether the molecule is important for function and/or structure.

Small Molecules

Ligands (4 Unique)

ID	Chains	Name / Formula / InChI Key	2D Diagram & Interactions	3D Interactions
CIT Query on CIT Download SDF File ⊕ Download CCD File ⊕	F	CITRIC ACID $C_6 H_8 O_7$ KRKNYBCHXYNGOX- UHFFFAOYSA-N		Ligand Explorer JSmol
GNP Query on GNP Download SDF File ⊕ Download CCD File ⊕	A, B, C, D, E, F	PHOSPHOAMINOPHOSPHONIC ACID-GUANYLATE ESTER $C_{10} H_{17} N_6 O_{13} P_3$ UQABYHGXWYXDTK- UUOKFMHZSA-N		Ligand Explorer JSmol
MG Query on MG Download SDF File ⊕ Download CCD File ⊕	A, B, C, D, E, F	MAGNESIUM ION Mg JLVVSXFLKOJNIY-UHFFFAOYSA-N	Mg^{2+}	Ligand Explorer JSmol
UNX Query on UNX Download SDF File ⊕ Download CCD File ⊕	A	UNKNOWN ATOM OR ION X null	X	Ligand Explorer JSmol

Figure 6-7: Screenshot showing the "Small Molecules" box.

Do you have any biologically important ligands? CIRCLE your answer below. If you circled "YES", please indicate which are biologically relevant.

<p style="text-align:center">YES   NO</p>

e.   What elements of secondary structure does your protein have?

There are two types of secondary structure elements: alpha-helices and beta-sheets. Different amounts of each are found in different proteins. Click on the "Sequence" tab at the top of the PDB record (Figure 6-8).

Figure 6-8: Screenshot highlighting in red the "Sequence" tab.

Using the information on this webpage, please answer the following questions:

(i)   Download the secondary structure diagram. You will likely need it when formulating your hypothesis. See Figure 6-9 for an example.

Sequence Chain View
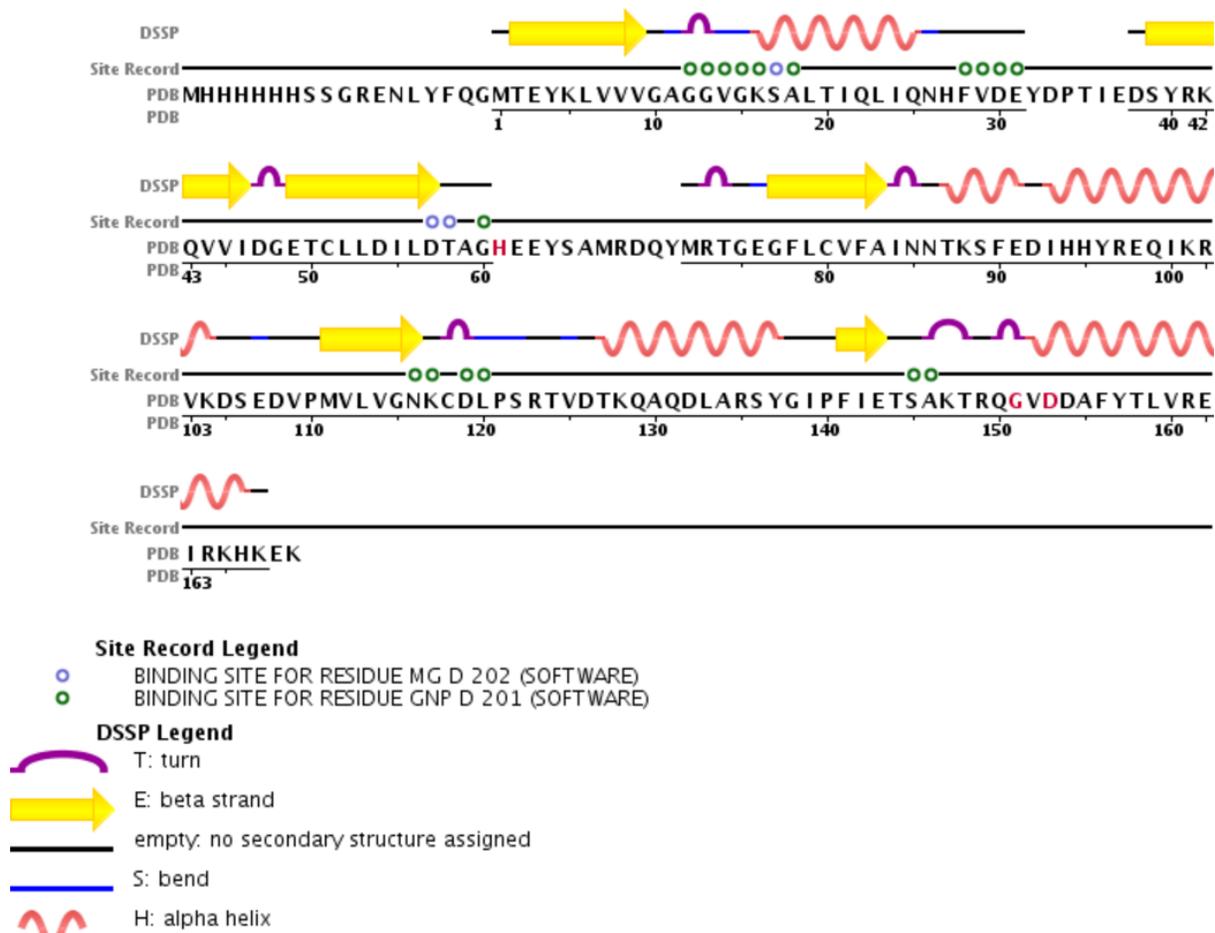
Figure 6-9: Screenshot of the secondary structure diagram for K-Ras.

(ii)   Immediately above the secondary structure diagram, you will see listed the percentage of each secondary structure element in the protein as well as the length of the protein expressed in the expression system used to overexpress the protein for crystallization. See Figure 6-10 for an example.

Chain Downloadable Files

Download FASTA File

View Sequence & DSSP Image

Download Sequence Chain Image

Chain Info

Polymer: 1
Length: 187 residues
Chain Type: polypeptide(L)
Reference: **UniProtKB** (P01116)
Up-to-date UniProt Ids are provided by the SIFTS project

Display Parameters

Identical Chains

Show A chain | Show B chain | Show C chain
Show E chain | Show F chain | Show all chains

Currently displayed
SEQRES sequence.
Display external (UniProtKB) sequence

Mouse over an annotation to see more details. Click on any annotation to enable Jmol.

Annotations	Details
Secondary Structure:**DSSP** [hide] [reference]	28% helical (5 helices; 53 residues) 21% beta sheet (6 strands; 41 residues)
Structural Feature:**Site Record** [hide] [reference]	3GFT_D_AC9_19 BINDING SITE FOR RESIDUE GNP D 201 (SOFTWARE) 3GFT_D_AC8_4 BINDING SITE FOR RESIDUE MG D 202 (SOFTWARE)

Figure 6-10: Screenshot of the percentage of α-helix and β-sheet in K-Ras.

What is the length of the protein studied and reported in the PDB file? _____

What is the percentage of α-helix and β-sheet in your protein? _____

How many α-helices are present in your protein? _____

How many β-strands are present in your protein? _____

(iii)   In many cases, crystallographers cannot see the residues at the N- and C-termini of a protein. This is due to the fact that these portions of the polypeptide often have increased mobility and this makes gathering data through X-ray crystallography impossible. As you will see in Figure 6-11,

the amino acids for which structural data are available for K-Ras are underlined. The amino acids that are not underlined are still part of the protein, but their 3-D structural coordinates cannot be determined. In the structure contained in 3GFT.pdb, the first 18 amino acids and the last 2 amino acids cannot be seen in the final 3-D structure of the protein. There are also 17 residues that form loops that are mobile and cannot be seen in the PDB file.
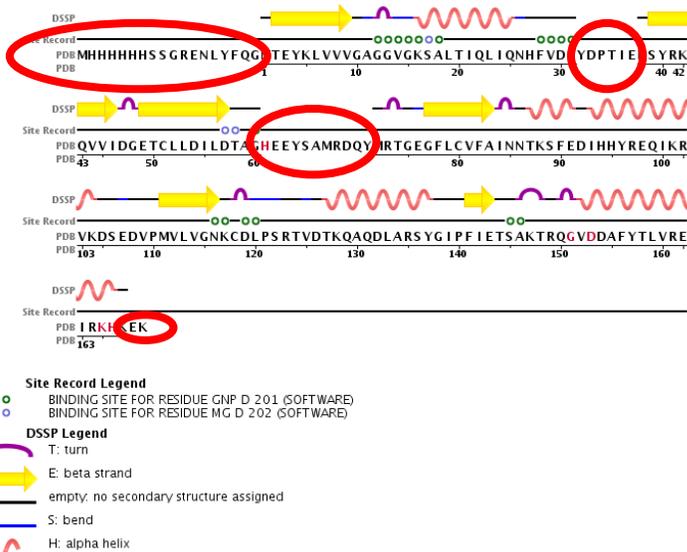
Figure 6-11: Screenshot of the secondary structure diagram, with the amino acids whose structural coordinates are not included in the PDB file circled in red.

Are the structural coordinates of all amino acids in the protein contained in the PDB file for your topic? CIRCLE YOUR ANSWER BELOW. If you circle "NO", please indicate below which residues are missing.

<div align="center">YES   NO</div>

(iv)   Is the disease-causing mutation located in an element of secondary structure? If so, in which type is it? Use the secondary structure diagram to answer this question. In the case of K-Ras, the disease-causing mutation is G60R and this mutation is located in a loop between the third β-strand and the second α-helix.

Important note: The numbering of residues in the PDB file may not correspond directly with the numbering used in the peer-reviewed literature describing your mutation (such as in the "seed" reference given in Appendix A for your chosen topic). The reason for this is that the form of the protein that is crystallized is sometimes truncated or altered in some way to allow it to crystallize. To avoid confusion, click on the "Display external (UniProtKB) sequence" button circled in Figure 6-12(A) below; the external (UniProtKB) sequence numbering will match what is usually used in the peer-reviewed literature. When you click on this option, the primary/secondary structure diagram will be redrawn with the external/UniProt numbering ABOVE the

primary structure and the PDB numbering BELOW it. If the two sets of numbers match, the numbering used in the PDB file is identical to what is used in the peer-reviewed literature. **Please use the external (UniProtKB) numbering when referring to the mutation in your presentation.**

(A)

Chain Downloadable Files

Download FASTA File

View Sequence & DSSP Image

Download Sequence Chain Image

Chain Info

Polymer: 1
Length: 187 residues
Chain Type: polypeptide(L)
Reference: **UniProtKB** (P01116)
Up-to-date UniProt Ids are provided by the SIFTS
project

Display Parameters

Identical Chains

Show A chain | Show B chain | Show C chain
Show E chain | Show F chain | Show all chains

Currently displayed
SEQRES sequence.
Display external (UniProtKB) sequence

Mouse over an annotation to see more details. Click on any annotation to enable Jmol.

Annotations	Details
Secondary Structure:**DSSP** [hide] [reference]	28% helical (5 helices; 53 residues) 21% beta sheet (6 strands; 41 residues)
Structural Feature:**Site Record** [hide] [reference]	3GFT_D_AC9_19 BINDING SITE FOR RESIDUE GNP D 201 (SOFTWARE) 3GFT_D_AC8_4 BINDING SITE FOR RESIDUE MG D 202 (SOFTWARE)

(B) Sequence Chain View

Figure 6-12: (A) Screenshot highlighting the "Display external (UniProtKB) sequence" button in red. (B) Screenshot showing the external/UniProt numbering ABOVE the primary sequence and the PDB

numbering BELOW the primary sequence.

(3) ***Examining the effects of a mutation on protein tertiary and quaternary structure***

Background

In this section of the exercise, we will study the effects of a single amino acid substitution on protein structure and function. Amino acid substitutions generally affect one or more of the following:

- active site interactions with substrates, which can render the enzyme inactive;

- protein folding, which can reduce enzyme activity, alter cellular trafficking of the protein, or lead to protein degradation; or

- interactions between separate protein chains in oligomeric proteins, which reduce enzyme activity or weaken structures needed for function.

In studying the protein structures of the wild-type protein and that of the mutant protein, you will be asked to determine which of the above effects are caused by the mutation identified in your chosen topic.

To help you with completing this portion of the exercise, we will study together the effect of the mutation G60R in K-Ras, where Gly-60 has been substituted by Arg-60. We will make use of the program PyMOL to visualize protein structures in three-dimensions.

Your task

First, install the program PyMOL on your computer. This program is available for download from the "Laboratory exercise 2" folder on D2L. Executable files are available for PC, Mac and Linux platforms.

Once PyMOL is installed on your computer, open the program and load the PyMOL file for your chosen topic (see Appendix A for file name; note that a PyMOL file ends with a ".pse" file extension). You will note that two windows open when the PyMOL program is loaded: a window called "The PyMOL Molecular Graphics System" and a window called the "PyMOL Viewer". The first window contains many different menus that will allow you to, amongst many other options, open and save files. The other window is where the structural coordinates of the protein will appear. To load a PyMOL file in the program, click "File" in the "The PyMOL Molecular Graphics System" window and then "Open" to select your file. We will use 3GFT.pse for the K-Ras example.

In the PyMOL file for each topic, pre-selected views of the protein structure have been saved to help you focus on certain portions of the structure. You can toggle between the pre-selected views by pressing on the buttons available on the lower left of the "PyMOL Viewer" window (Figure 6-13). To rotate the protein in the "PyMOL Viewer", click anywhere in the window with the left mouse button, and, with the left mouse button still depressed, move the mouse to move the protein structure. To zoom in and out on a portion of the structure, click anywhere in the window with the right mouse button, and, with the right mouse button still depressed, move the mouse to zoom in and out on the protein structure.

Figure 6-13: Screenshot of the "PyMOL Viewer" window, with the toggle buttons highlighted in red. In this example, three different views (available by clicking on the F1, F2 and F3 toggle buttons) are available.

Working as a team, carefully follow each of the following steps using the PyMOL file corresponding to the structural coordinates for the enzyme associated with your chosen topic.

> **a.** **How can we insert the disease-causing mutation in the wild-type three-dimensional coordinates of the wild-type protein? Does the introduction of this mutation cause steric clashes?**

To mutate glycine-60 to arginine-60 in the three-dimensional structural coordinates of K-Ras, first select the best view that shows the residue to be mutated and the nearby region of the protein structure (usually view "F7"). Next, select the "Wizard" menu from "The PyMOL Molecular Graphics System" window. Select "Mutagenesis" from the Wizard drop-down menu. In the "Mutagenesis" menu that appears on the right-hand side of the "PyMOL Viewer" window, click on "No mutation" (Figure 6-14).
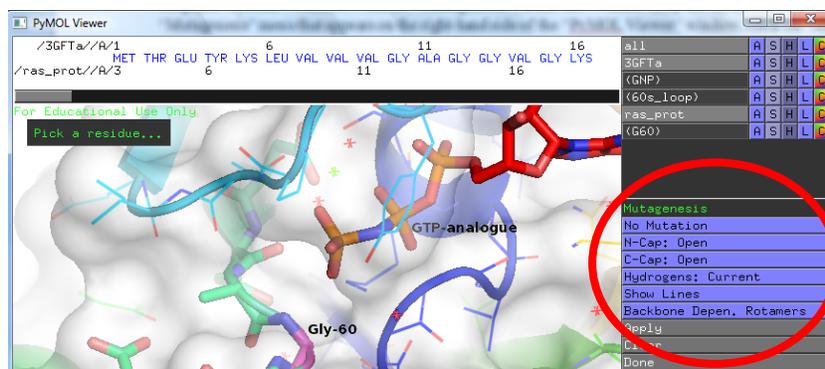
Figure 6-14: Screenshot of the "PyMOL Viewer" window, with the "Mutagenesis" menu and the "No Mutation" button highlighted in red.

Once you click on "No Mutation", you will now see buttons for all of the different types of amino acids. In the case of K-Ras, select "ARG", because we want to see what happens when this residue is mutated

from Gly to Arg, as has been seen with some cancer patients (Figure 6-15). For your specific topic, refer to Appendix A to see the name of the residue to be mutated and the mutation to be made.
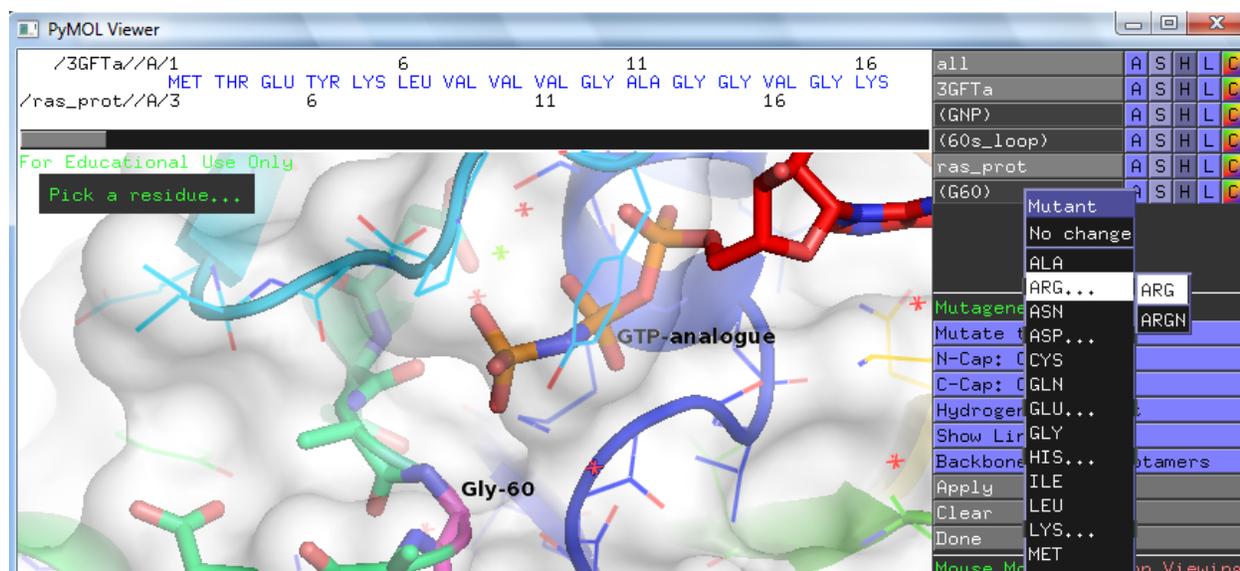
Figure 6-15: Screenshot of amino acid selection. "ARG" has been selected in this case.

Once "ARG" has been selected, click on the C-alpha atom of Gly-60 (coloured magenta) in the "PyMOL Viewer" window. Note that a grey Arg side chain has now been modelled into the structure, with red disk symbols showing steric clashes between the Arg side chain and the other atoms in the protein (Figure 6-16). You may want to rotate the view (by holding down the left mouse button and dragging) to have a better view of the steric clashes introduced by the mutated side chain.

Click on the 5th grey button (counting from the left) on the bottom, right portion of the "PyMOL Viewer" window to try a different conformation for the Arg side chain. You should see the "state" number advance each time you click on this grey button. You should click on this button repeatedly to try all of the different **rotamers** (*i.e.*, commonly seen 3-D conformations) of the Arg side chain (Figure 6-16).

Why do you think that different rotamers exist?

_____

_____

_____
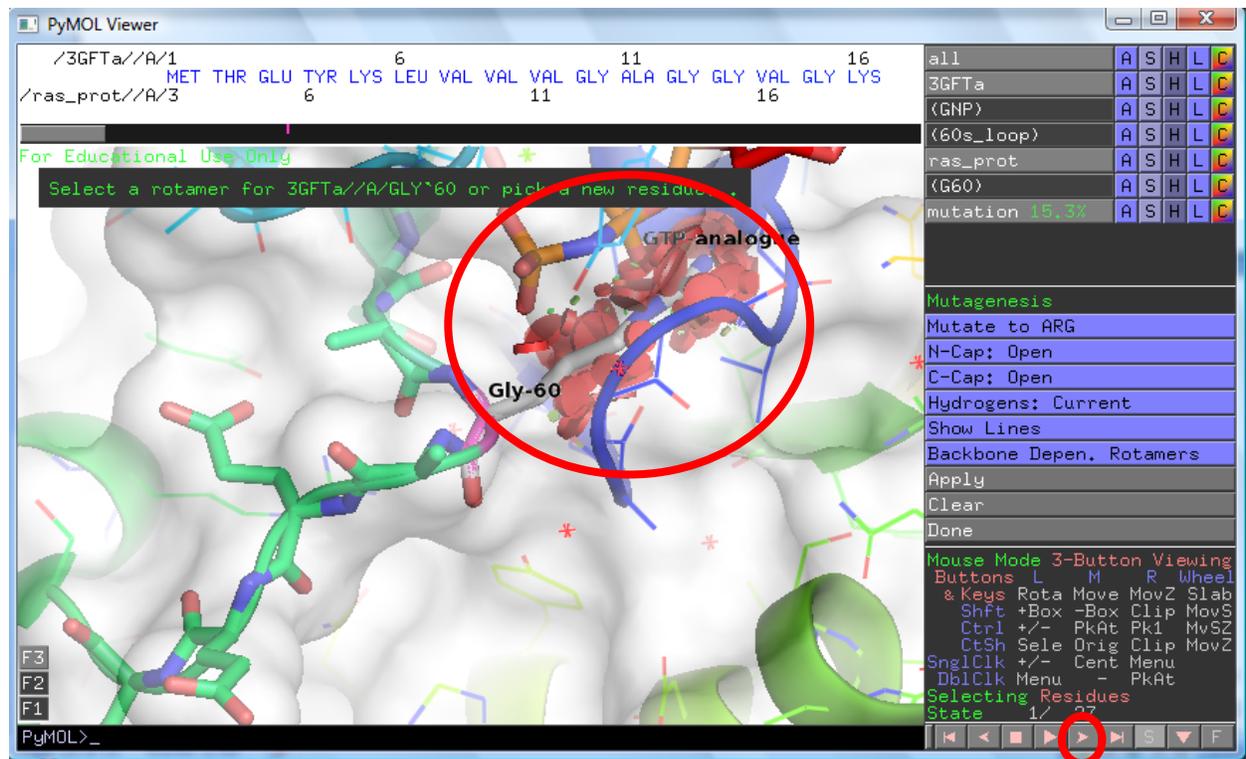
Figure 6-16: Screenshot of the "PyMOL Viewer" window after the Gly-60 has been replaced with Arg-60. The grey button on the bottom right-hand side of the window that allows you to change between side chain rotamers has been highlighted in red. The red disks showing steric clashes have also been highlighted in red.

In the case of K-Ras, you should see that each of the rotamers gives many steric clashes with other atoms in the active site near the GTP substrate-binding site. These clashes may indicate why the mutation of Gly-60 to Arg disrupts the structure of K-Ras and disrupts the normal function of the protein, which in turn leads to the development of cancer. Take a few screenshots of a few rotamers of the mutated side chain; you may find these quite helpful to support your hypothesis.

After you have tried all 27 rotamers of Arg, click on the grey "Clear" button, and then click on the "Done" button. For other mutations you will have fewer rotamers to examine, because shorter side chains have fewer allowed conformations.

> **b.** **Which non-covalent interactions are disrupted in the three-dimensional structure when the disease-causing mutation is present?**

Return to the view that best shows Gly-60 in the wild-type protein. **Take note of a few of the types of non-covalent interactions in which this residue is involved.** Which of these, if any, would be destroyed when this amino acid is replaced with the mutant residue? You may wish to save a few pictures of the wild-type amino acid involved in different types of non-covalent interactions in order to

support your hypothesis. To save pictures from PyMOL, click on the "File" menu under the "The PyMOL Molecular Graphics System" window. Click on "Save Image As", and then click on "PNG ...". A pop-up window will appear asking you to give the figure a name.

> c.   **What is the highest level of protein structure in your protein? If your protein has quaternary structure, does the disease-causing mutation interfere with its formation?**

In looking at your protein structure in PyMOL, determine the highest level of protein structure your protein contains (tertiary or quaternary). To do so, determine whether the protein is made up of a single chain or whether it has several different chains. CIRCLE YOUR ANSWER BELOW.

> protein has tertiary structure only because it only has 1 chain

> protein has quaternary structure because it has more than 1 chain

Part B: Relate clinical data with the effects of a mutation on a protein's structure and function

The goal of the previous portion of this exercise was to propose a hypothesis to explain the effects of the disease-causing mutation on the enzyme's structure and function. We will now focus on extending the hypothesis you have been developing, by using another bioinformatics tool (*i.e.*, the KEGG database) to best understand how the enzyme fits into the broader metabolic landscape in the body.

The KEGG database, which stands for "Kyoto Encyclopedia of Genes and Genomes", contains useful maps of metabolic pathways, as well as description of diseases and drug targets, among others. We will be making use of the KEGG database in this project to uncover in which metabolic pathway(s) the enzyme you have been studying is involved, and more importantly, which reaction(s) it catalyzes. This knowledge will help you extend your hypothesis so that you can relate the effects of the mutation on your protein's structure and function to clinical data.

(1) Using the KEGG database

To get to the KEGG database, please go to the following website:

http://www.genome.jp/kegg/disease/

In the search box at the top of the page, type in the **name** of the disease you selected earlier this term. Press on "Go".

A table with one or more entries will be returned to you. Click on the hyperlinked entry number that best relates to your chosen topic.

You will be taken to a table containing data on the disease. In the table, you will see an entry called "Pathway". The metabolic pathway(s) associated with the disease will be listed. To the left of each pathway name is a hyperlinked reference number. Click on this hyperlinked reference number. You will

be taken to a schematic diagram of the metabolic pathway. Defective enzymes in the pathway associated with disease will be shown in red numbers. **BE CAREFUL – SOME DISEASES ARE CAUSED BY MULTIPLE DIFFERENT ENZYMES! IF MULTIPLE BOXES WITH RED NUMBERS APPEAR IN THE PATHWAY, THIS SIGNIFIES THAT MULTIPLE ENZYMES CAN BE INVOLVED. YOU NEED TO FIND THE BOX WITH THE RED NUMBERS THAT CORRESPONDS TO THE ENZYME YOU HAVE BEEN STUDYING.**

Take note of the metabolic pathway(s) and use the map(s) to explain the clinical symptoms that are described in the table below.

(2) Clinical data for each possible topic

Using the knowledge you have gathered above about the metabolic pathway(s) in which your enzyme is involved, extend your original hypothesis to explain the clinical data that are described below.

Table 6-2: Clinical data for each available genetically-inheritable disease

Genetically-inheritable disease	Clinical data
Fructose intolerance	An 18-year old patient presents himself to the emergency department with hypoglycemia. Initial tests confirm he has an A150P mutation in the gene for liver fructose-1,6-biphosphate aldolase (a.k.a. fructose-1,6-bisphosphate aldolase B). Propose a hypothesis to explain the presence of hypoglycemia.
Galactosemia type III	A woman presents herself to the emergency department with her two-week old infant. She is breastfeeding her infant, but the infant is starting to show diarrhea and an enlarged liver. Further tests show that there are high levels of galactose-1-phosphate in the red blood cells of the infant. The infant has the G90E mutation in the gene that encodes UDP-galactose 4-epimerase. Propose a hypothesis to explain these observations.
Gaucher disease	A 1-year-old child is found to have an accumulation of glucosylceramide in macrophages. The liver is showing signs of enlargement as well. He has the V437F mutation in the gene that encodes acid-beta-glucosidase. Propose a hypothesis to explain these observations.
Glutathione synthetase deficiency	A 2-year-old infant presents with high levels of 5-oxoproline and metabolic acidosis. He has the R267W mutation in the gene that encodes glutathione synthetase deficiency. Propose a hypothesis to explain these observations.
OTC deficiency	A 14-year-old boy presents with high levels of ammonia in the blood. The synthesis of orotic acid is also increased. He has the I172F mutation in the gene that encodes ornithine transcarbamylase. Propose a hypothesis to explain these observations.
Lesch-Nyhan syndrome	Uric acid (a.k.a. urates) levels are high in both plasma and urine in a 22-year-old patient. She has the S162R mutation in the gene that encodes HGPRTase. Propose a hypothesis to explain these observations.

Lou Gehrig's disease	A 52-year-old man presents with difficulty controlling voluntary muscle. He has the L85F mutation in the gene that encodes Cu,Zn superoxide dismutase and shows signs that some neurons are dying. Propose a hypothesis to explain these observations.
Phenylketonuria	A 1-week-old infant presents with urine that has a "mousy" odor due to the presence of phenylpyruvate. He has the R408W mutation in the gene for phenylalanine hydroxylase. Propose a hypothesis to explain these observations.
PNP deficiency	A 2-year-old child presents with a gradual decrease in T-cell immunity. She also has a high level of guanosine in the blood. She has the Y192C mutation in the gene that encodes purine nucleoside phosphorylase. Propose a hypothesis to explain these observations.
Porphyria (hepatic)	A 25-year-old women presents with high levels of uroporphyrin in her urine. She has the L282R mutation in the gene that encodes uroporphyrinogen decarboxylase. Propose a hypothesis to explain these observations.
Tay-Sachs disease	A 1-week-old infant has an accumulation of GM2 gangliosides, which is leading to premature death of nerve cells. He has the S210F mutation in the gene that encodes beta-hexosaminidase A. Propose a hypothesis to explain these observations.

Part C: Synthesizing your findings and sharing your hypothesis

In your presentation, you are asked to **propose a hypothesis** that describes how the disease-causing mutation affects protein structure and function, and how these effects on protein structure and function explain the metabolic clinical data presented in Part B.

To support your hypothesis and persuade your peers to accept it, your presentation must include evidence! You must include the following:

- Background information you gathered on your disease of interest from the OMIM website, as well as from your selected review articles. Be critical with your selection here! Your presentation should NOT entirely be made up of background information. Only include what you believe your peers need to know to understand your hypothesis.

Select pieces of data you gathered on your disease and protein using the PDB, PyMOL and the KEGG database – be picky! You only want to include the pieces of data that support your hypothesis and are convincing. Remember that your hypothesis must describe the effects of the mutation on your enzyme (both structurally and functionally). Your description must be sufficiently detailed to clearly illustrate the link between the effects of the mutation on the protein structure and the clinical data presented for your topic.

You have two options to share your work:

- Creating a short movie using Powtoon; or

- Creating an infographic using Piktochart.

It is entirely your choice as to which option you select.

Your presentation should be no longer than 2 minutes and no shorter than 1 minute. Remember to speak slowly and clearly.

Below, you can find more information about the options from which you can select:

(1) Creating a short movie using Powtoon

You can find Powtoon at the following address: https://www.powtoon.com/

You can create a free account, which will let you make a movie that is up to 3 minutes in length.

There are multiple templates to choose from, sample Powtoon videos to view, and a lot of how-to videos freely available to watch through YouTube.

The following how-to video is particularly useful for using a free account on Powtoon: https://www.youtube.com/watch?v=lEQiZQi-aGY

(2) Creating an infographic using Piktochart

An infographic is information arranged visually. It minimizes text by combining text and images such as drawings, charts, arrows, with use of color and 'white space', to help the audience grasp essential ideas and content.

Infographics can be used when you want to get across a big idea, or make a point to your audience. Concepts that are tricky for audiences might lend themselves well to an infographic. Or, if you have facts that are hard to understand, you might investigate how they might be turned into an infographic.

You can find Piktochart at the following address: https://piktochart.com/

You can create a free account, which will let you create an infographic.

There are multiple templates to choose from in Piktochart.

The following how-to video is particularly useful for creating an infographic on Piktochart using a free account: https://www.youtube.com/watch?v=Eq-85gzw3GI

You are encouraged to google "infographic examples" for lots of great examples.

Your presentation will be evaluated using the rubric in APPENDIX B by your G.T.A.

References

Bednarski, A.E., S.C.R. Elgin, and H.B. Pakrasi. 2003. An inquiry into protein structure and genetic disease: introducing undergraduates to bioinformatics in a large introductory course. Cell Biology Education, 4:207-220.

Pratt, C.W., and K. Cornely. Essential Biochemistry. Second edition. Wiley, Hoboken, NJ, 704 pages.

Nelson, D.L. and M.M. Cox. Lehninger Principles of Biochemistry. Sixth edition. W.H. Freeman, New York, NY, 1198 pages.

Schneider, T.L., and B.R. Linton. 2008. Introduction to Protein Structure through Genetic Diseases. Journal of Chemical Education, 85(5):662-665.

Steward, R.E., M.W. MacArthur, R.A. Laskowski, and J.M. Thornton. 2003. Molecular basis of inherited diseases: a structural perspective. Trends in Genetics, 19(9):505-513.

Appendix A: List of available genetically-inheritable diseases for selection

Genetically-inheritable disease	Protein structure coordinates of wild-type protein	PyMOL file	Disease-causing mutation	Seed reference
Fructose intolerance	1QO5.pdb	1QO5.pse	A150P	Dalby, A.R., D.R. Tolan, and J.A. Littlechild. 2001. The structure of human liver fructose-1,6-bisphosphate aldolase. Acta Crystallogr. Sect. D, 57:1526-1533.
Galactosemia	1I3K.pdb	1I3K.pse	G90E	Thoden, J.B., T.M. Wohlers, J.L. Fridovich-Keil, and H.M. Holden. 2001. Molecular basis for severe epimerase deficiency galactosemia. X-ray structure of the human V94M-substituted UDP-galactose 4-epimerase. J. Biol. Chem., 276:20617-20623.
Gaucher's disease	1Y7V.pdb	1Y7V.pse	V437F	Premkumar, L., A.R. Sawkar, S. Boldin-Adamsky, L. Toker, I. Silman, J.W. Kelly, A.H. Futerman, and J.L. Sussman. 2005. X-ray structure of human acid-beta-glucosidase covalently bound to conduritol-B-epoxide. Implications for Gaucher disease. J. Biol. Chem., 280:23815-23819.
Glutathione synthetase deficiency	2HGS.pdb	2HGS.pse	R267W	Polekhina, G., P.G. Board, R.R. Gali, J. Rossjohn, and M.W. Parker. 1999. Molecular basis of glutathione synthetase deficiency and a rare gene permutation event. EMBO J., 18:3204-3213.
OTC deficiency	1C9Y.pdb	1C9Y.pse	I172F	Shi, D., H. Morizono, M. Aoyagi, M. Tuchman, and N.M. Allewell. 2000. Crystal structure of human ornithine transcarbamylase complexed with carbamoyl phosphate and L-norvaline at 1.9 A resolution. Proteins, 39:271-277.
Lesch-Nyhan syndrome	1D6N.pdb	1D6N.pse	S162R	Balendiran, G.K., J.A. Molina, Y. Xu, J. Torres-Martinez, R. Stevens, P.J. Focia, A.E. Eakin, J.C. Sacchettini, and S.P. Craig 3rd. 1999. Ternary complex structure of human HGPRTase, PRPP, Mg^{2+}, and the inhibitor HPP reveals the involvement of the flexible loop in substrate binding. Protein Sci., 8:1023-1031.
Lou Gehrig's disease	1N18.pdb	1N18.pse	L85F	Cardoso, R.M.F., M.M. Thayer, M. DiDonato, T.P. Lo, C.K. Bruns, E.D. Getzoff, and J.A. Tainer. 2002. Insights into Lou Gehrig's disease from the structure and instability of the A4V

				mutant of human Cu,Zn superoxide dismutase. J. Mol. Biol., 324: 247-256.
Phenylketonuria	2PAH.pdb	2PAH.pse	R408W	Fusetti, F., H. Erlandsen, T. Flatmark, and R.C. Stevens. 1998. Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria. J. Biol. Chem., 273: 16962-16967.
PNP deficiency	1RFG.pdb	1RFG.pse	Y192C	Canduri, F., R.G. Silva, D.M. dos Santos, M.S. Palma, L.A. Basso, D.S. Santos and W.F. de Azevedo. 2005. Structure of human PNP complexed with ligands. Acta Crystallogr., Sect. D, 61:856-862.
Porphyria	1R3Q.pdb	1R3Q.pse	L282R	Phillips, J.D., F.G. Whitby, J.P. Kushner, and C.P. Hill. 2003. Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase. EMBO J., 22:6225-6233.
Tay-Sachs disease	2GK1.pdb	2GK1.pse	S210F	Lemieux, M.J., B.L. Mark, M.M. Cherney, S.G. Withers, D.J. Mahuran and M.N.G. James 2006. Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. J. Mol. Biol., 359:913-929.

Appendix B: Rubric used by your G.T.A. to evaluate your presentation

Criteria	4 – Excellent	3 – Good	2 – Fair	1 – Area of growth	Grade (out of 4)
Quality of background information	Appropriate amount of background information is presented, with protein and mutation clearly identified and described.	Too much background information is presented. It is not succinctly tied to the hypothesis.	Consists mostly of background information. There are no ties to the hypothesis and/or very little reference to data and/or literature.	Only consists of background information. There is no hypothesis.	____ / 4
Quality of hypothesis and justification for proposed mechanism	The hypothesis provides a relationship between the clinical data and the effects of the mutation on protein structure and function. The effects of the mutation on protein structure are clearly explained through a detailed description of the disruption of any non-covalent interactions and the introduction of any steric clashes. The hypothesis is well supported by data.	Although the hypothesis is presented, insufficient details are included and/or it is not well supported by data.	The hypothesis is stated, but no support or description is included. There are very few references to the data collected.	There is no hypothesis that attempts to describe the links between the mutation, the effects on protein structure and the molecular basis for the disease.	____ / 4
Quality of presentation material	All tables and figures have an appropriate caption and title (*i.e.*, Figure 1, Table 1, etc.). Data presented are easy to read and understand. There are screenshots of the mutated protein structure showing the impact of the mutation. Materials are well prepared, are easy to read and are visually appealing. There are limited grammatical and spelling errors.	Table and figure captions do not contain all required information and/or titles. More information is required to fully understand what is presented. More relevant data should have been presented to support the hypothesis. Materials are good, but some are difficult to read. There are some grammatical and spelling errors.	Tables and figures only have a title. Data are hard to read. Few relevant data/figures are presented in support of the hypothesis. Materials are hard to read. There are many grammatical and spelling errors.	Captions are absent. Data are hard to read and understand. No data/figures are included. The materials are disorganized and difficult to follow. The grammatical and spelling errors are very distracting.	____ / 4

Areas of strength:

Areas of growth: