

# Gaining a Structural Perspective on the Molecular Basis of Inherited Diseases

Isabelle H. Barrette-Ng<sup>1</sup> and Don MacMillan<sup>2</sup>

<sup>1</sup> University of Calgary, Department of Biological Sciences, 2500 University Drive N.W., Calgary Alberta T2N 1N4 CAN

<sup>2</sup> University of Calgary, Taylor Family Digital Library, 2500 University Drive N.W., Calgary Alberta T2N 1N4 CAN

([mibarret@ucalgary.ca](mailto:mibarret@ucalgary.ca); [macmild@ucalgary.ca](mailto:macmild@ucalgary.ca))

Many undergraduate students in introductory biochemistry courses find it challenging to understand how different levels of protein structure relate to each other. To address this problem, we introduced an inquiry-based laboratory exercise in which students are challenged to explain how the effects of mutations on different levels of protein structure lead to changes in protein function and ultimately to genetically-inheritable diseases. The implementation of this exercise in a large, second-year undergraduate, introductory biochemistry course led to a high level of student satisfaction and a more integrated view of biochemistry and genetics.

**Keywords:** biochemistry, bioinformatics, inquiry-based learning, genetics, information literacy, levels of protein structure

## Introduction

Most undergraduate biology programs require students to master basic concepts in genetics and biochemistry, because they are essential for understanding many aspects of living things. Despite the central importance of both genetics and biochemistry to modern biology, these two topics are still unfortunately often taught from very different historical and ideological perspectives. As a result, students often fail to appreciate the connections between genetics and biochemistry.

Over the last five years, as part of an effort to eliminate the artificial barriers between the teaching of introductory genetics and biochemistry at the University of Calgary, we have introduced novel laboratory exercises in both courses that help students understand more of the key concepts common to both disciplines. One of these exercises builds upon a classic example of a very striking molecular connection between genetics and biochemistry. As first demonstrated in a classic paper by Linus Pauling and co-workers (1949), mutations in hemoglobin lead to changes in protein structure, which in turn lead to a molecular explanation for the development of an important human disease, sickle-cell anemia. Since this classic study, many other papers have described examples of mutations that lead to changes in protein structures, and which in turn lead to the development of diseases (Steward et al., 2003).

Previously published laboratory exercises for undergraduate students have described different approaches to linking

concepts in genetics to biochemistry using genetically-inheritable diseases. Bednarski et al (2003) described an exercise in which students were introduced to a wide range of bioinformatics software and databases to study the effects of a mutation on protein structure and metabolic pathways. The central learning objectives of this exercise focused on the use of a variety of different bioinformatics tools to illustrate various aspects of genetics and protein structure. Schneider and Linton (2008) described quite a different type of exercise in a senior undergraduate Molecular Medicine course designed to engage students in using interactive molecular graphics software. The primary goal of their exercise was to learn more about the use of molecular graphics software to explore the three-dimensional structure of proteins.

When we considered introducing a similar exercise in our large-enrollment introductory biochemistry course, we realized that this type of exercise could help achieve an additional learning objective that we have found to be very challenging for most students. Our exercise focuses on understanding the meaning of different levels of protein structure, and how different tools can be used to analyze each level of structure. The reason we chose to emphasize the relationships between different levels of protein structure is that students seem to compartmentalize the information about sequence, secondary structure, three-dimensional structure and quaternary structure. As a result, they fail to see how the different levels of protein structure are intimately tied together. We feel that

it is critical for them to have this knowledge before they begin to learn about more advanced concepts in biochemistry that are dependent on a solid understanding of the basics of how protein structure relates to function.

In our exercise, students start by seeing how a change in gene sequence leads to a change in the amino acid sequence or primary structure level of the protein. Then, through the use of sequence alignment programs, students can compare how primary structure varies in homologous proteins from mouse, chicken and zebrafish. This analysis provides an important level of context to evaluate the effects of their disease-related mutation against the level of amino acid sequence variation seen in homologous proteins. At the secondary structure level, students use a graphical representation of secondary structure mapped onto the protein's primary structure to see how segments of their protein form helices, strands or non-regular secondary structure, and how this level of structure again relates to the variation of sequence in homologous organisms and the conservation of sequence at the site of their disease-causing mutation. At the tertiary structure level, students use a sophisticated molecular graphics program (PyMOL) to access specific views through a simple and intuitive button-access interface to see how the primary and secondary levels of protein structure relate to the tertiary structure in three dimensions. Students focus on both the covalent and noncovalent interactions of the disease-related mutant amino acid and explore rotamers using a simple but powerful interface available in PyMOL. Finally, the students also view the overall quaternary structure of the protein to put the other levels of protein structure into perspective, and to finally relate the mutation to the overall molecular structure and function of the protein.

Students work in pairs and choose a topic to investigate from a list of eleven different diseases. Over the course of eight weeks, each pair of students follows a detailed series of exercises tailored to guide their investigations to achieve the specific learning objectives described above. The overall design provides students with enough direction and instruction to allow them to discover and explore the specifics of their disease and protein structure primarily using their own initiative. Support and guidance are also provided through three workshop sessions led by graduate teaching assistants, librarians and instructors. After they complete a series of investigative exercises focusing on exploring the effects of a specific mutation on different levels of protein structure, each student pair organizes the information gained through their investigations to prepare a capstone Power Point presentation. The 528 students in the course are divided into 24 laboratory sections, and the eleven student pairs in each laboratory section present their findings to each other, as well as to a graduate teaching assistant. A key aspect of the oral presentation is an emphasis on the discussion and peer evaluation of each presentation by all students in the laboratory section. The final grade for each student is determined by the teaching assistant after they examine the results from peer evaluation.

## Student Outline

Over the next few weeks, you will have the opportunity to gain, using various bioinformatics tools, 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 can be found in the coding region of genes, these mutations often affect 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 is or are replaced by another. This mutation results in a new codon, which causes a different amino acid to be inserted into the growing polypeptide chain during translation.

For this laboratory exercise, you will be asked to work with your laboratory partner. You and your laboratory partner will be guided in the use of various bioinformatics tools to study the effects of disease-causing mutations on protein structure and function. We will specifically focus on different levels of protein structure and how they are intimately related to one another in the formation of the final, fully-folded protein. At the end of this exercise, you and your laboratory partner will be asked to orally present your results to the other members of your laboratory session via a 10-minute Power Point presentation.

### Part A: Selecting Your Topic

The first part of this project involves selecting your topic. 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. All eleven available topics are listed in Appendix A. Also included in Appendix A are the protein structure coordinates for the wild-type protein and a file with a “.pse” file name extension. You will need this file for your work with the protein visualization software PyMOL. Appendix A also contains one seed reference for each disease, to help you get started in locating background information on your topic as well as structural information and the disease-causing mutation.

### Part B: Studying the Protein’s Structure and Physicochemical Properties

To help you complete your project, you will be guided through all of the steps using the K-Ras protein, which has been implicated in lung cancer. To make it easier for you, screenshots using the K-Ras example have been inserted in the text below.

#### *Finding 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 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.

In your literature searches, you should identify the following:

- a general description of the genetically-inheritable disease (symptoms, therapies, etc.)
- the inheritance pattern of the genetically-inheritable disease (i.e., autosomal, sex-linked, dominant, recessive);
- the chromosome on which the gene encoding the enzyme is located;
- the class to which your enzyme belongs;
- the function of the enzyme and the reaction it catalyzes; and
- the amino acids that make up the catalytic site of the enzyme of your chosen topic (these residues are the ones that allow the enzyme to catalyze a specific reaction).

### A First Look at Protein Structure Coordinate Files (PDB Files)

As we have studied in class, a protein's function is intricately linked to its structure. The loss of a protein's three-dimensional structure usually results in a complete loss of function of the protein. Protein structural information is very important to gaining a deep understanding of the function of a protein. In many cases, the way in which a protein functions remains unknown until structural information is available. 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 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. Protein crystals are very tiny, often measuring no more than 0.5 mm in diameter.

A protein crystal 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 black spot. 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 level of detail that you can see in the map depends on the quality of the crystal, and data quality is known as "resolution". 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 the atoms in the polypeptide chain.

By studying the electron density, you can determine the three-dimensional coordinates of all amino acids that comprise a protein. These three-dimensional coordinates 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.

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. Let's look at an example: the file 3GFT.pdb, which contains the three-dimensional coordinates of the human K-Ras protein in complex with a GTP analogue.

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 Fig. 1). 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).

Once the record for the PDB file appears, use the information shown on the page to answer the following questions:

- (1) What is the resolution? You can obtain this information in the "Experimental Details" box located at the bottom right of the page (see Fig. 2).

Figure 1. Screenshot of PDB website, with the search box highlighted in red.

Experimental Details		Hide
<b>Method:</b> X-RAY DIFFRACTION		
<b>Exp. Data:</b>		
Structure Factors		
<b>Resolution[Å]:</b>	2.27	
<b>R-Value:</b>	0.215 (obs.)	
<b>R-Free:</b>	0.267	
<b>Space Group:</b>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
<b>Unit Cell:</b>		
<u>Length [Å]</u>	<u>Angles [°]</u>	
a = 63.53	α = 90.00	
b = 118.28	β = 90.00	
c = 156.84	γ = 90.00	

Figure 2. Screenshot of the Experimental Details box, with the resolution highlighted in red.

(2) Which expression system was used to overexpress the protein for purification purposes? You can obtain this information in the “Source” box (see Fig. 3).

Source		Hide
<b>Polymer: 1</b>		
<b>Scientific Name:</b>	Homo sapiens	<b>Taxonomy</b>
<b>Common Name:</b>	Human	<b>Expression System:</b> Escherichia coli

Figure 3. Screenshot of the Source box, with the expression system used highlighted in red.

(3) Are there any ligands bound to your protein? These could be cofactors and/or substrates. To find the answer to this question, look in the “Ligand Chemical Component” box (Fig. 4). Use the “seed” article you were given in Appendix A if ligands are present to determine if they are important for function and/or structure.

Ligand Chemical Component				Hide
Identifier	Formula	Name	Interactions	
<b>CIT</b> Search Download	 C <sub>6</sub> H <sub>8</sub> O <sub>7</sub>	CITRIC ACID	Ligand Explorer	
<b>GNP</b> Search Download	 C <sub>10</sub> H <sub>17</sub> N <sub>6</sub> O <sub>13</sub> P <sub>3</sub>	PHOSPHOAMINOPHOSPHONIC ...	Ligand Explorer	
<b>MG</b> Search Download	Mg <sup>+2</sup> Mg	MAGNESIUM ION	Ligand Explorer	
<b>UNX</b> Search Download	UNK X	UNKNOWN ATOM OR ION	Ligand Explorer	

Figure 4. Screenshot showing the “Ligand Chemical Component” box.

### A Look at Primary and Secondary Structure

To obtain information on the primary and secondary structure of your protein, click on the “Sequence” tab at the top of the record (Fig. 5).

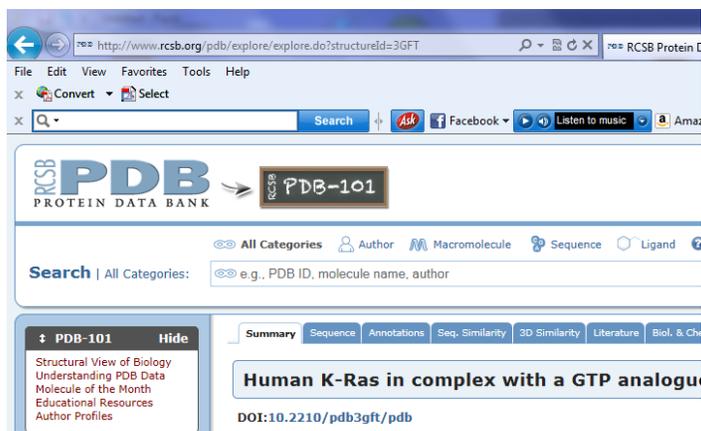


Figure 5. Screenshot highlighting in red the Sequence tab.

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

- (1) Download the secondary structure diagram. You will need to include this diagram in your Power Point presentation. See Fig. 6 for an example.

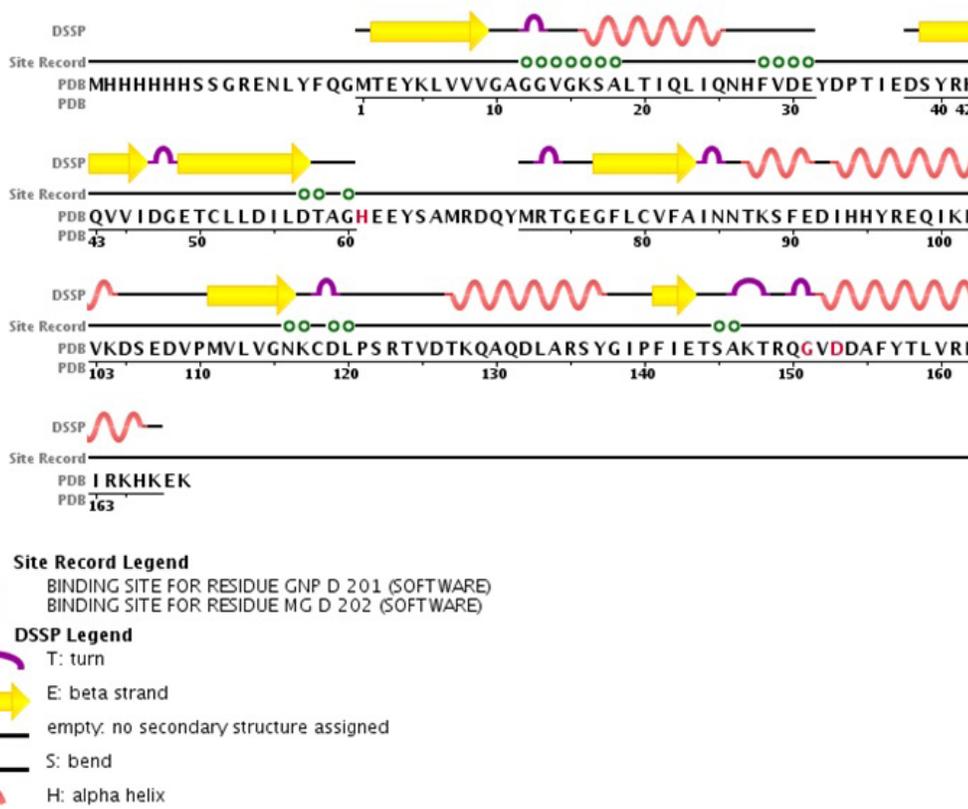


Figure 6. Screenshot of the secondary structure diagram for K-Ras.

- (2) Immediately above the secondary structure diagram is given the percentage of each secondary structure element in the protein as well as the length of the protein construct used in the expression system. What is the length of the protein used in the expression system? What is the percentage of  $\alpha$ -helix and  $\beta$ -sheet in your protein? How many  $\alpha$ -helices and  $\beta$ -strands are present in your protein? See Fig. 7 for an example.



Figure 7. Screenshot of the percentage of  $\alpha$ -helix and  $\beta$ -sheet in K-Ras.

(3) Are the structural coordinates of all amino acids in the protein contained in the PDB file? See Fig. 8 for an example.

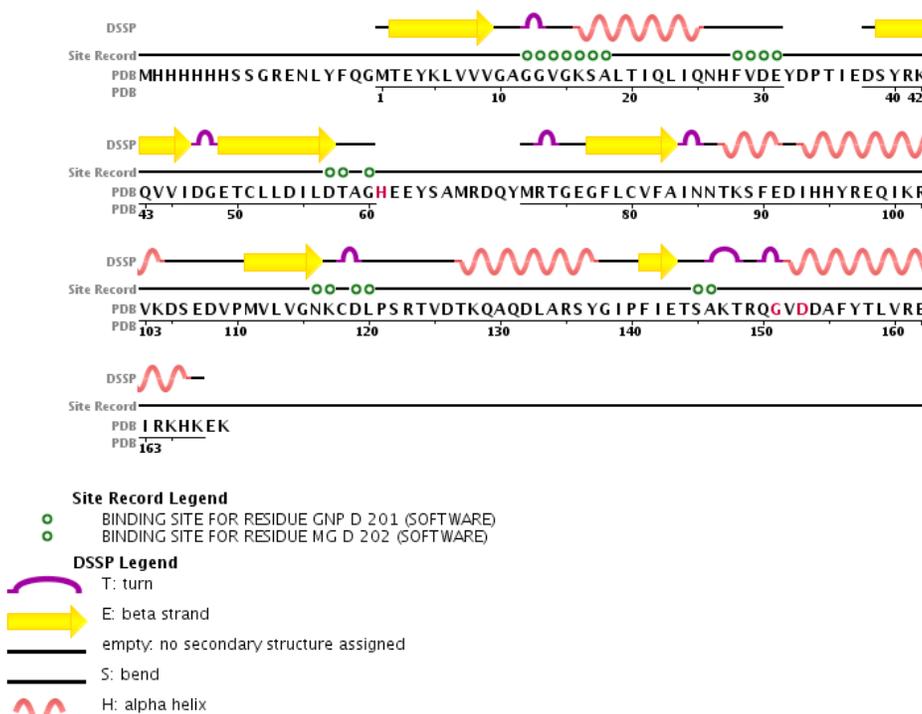
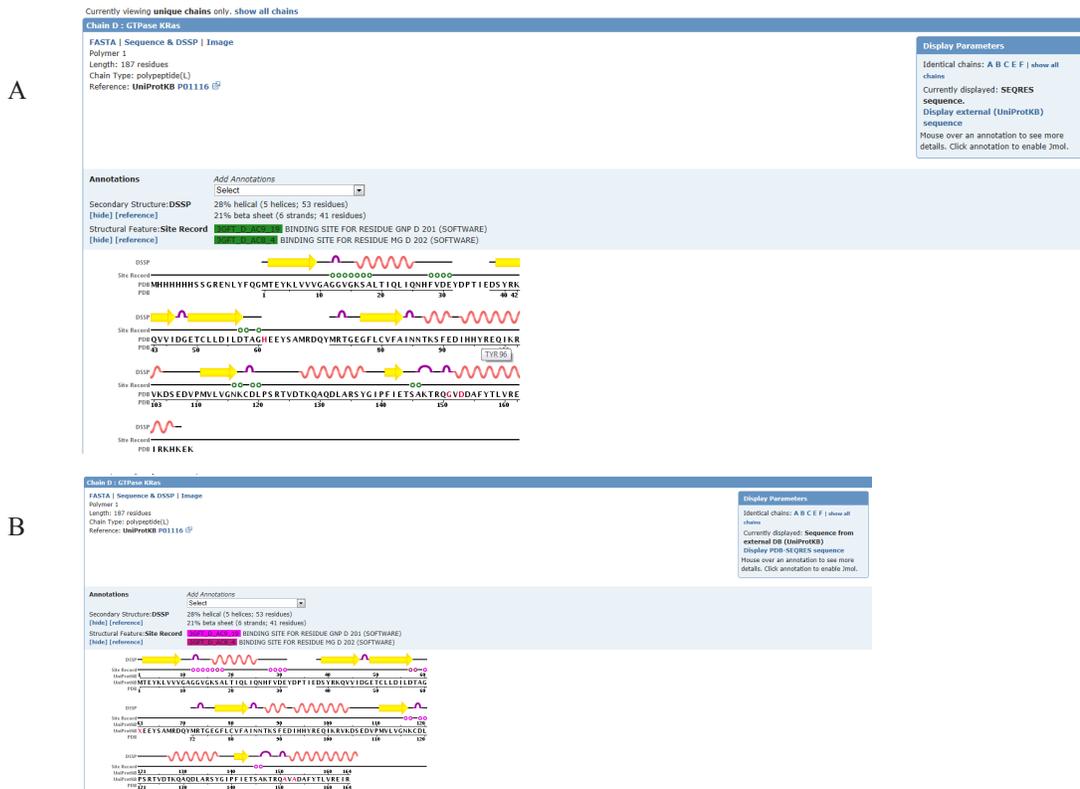


Figure 8. Screenshot of the secondary structure diagram, with the amino acids whose structural coordinates are not included in the PDB file circled in red.

(4) In what type of secondary structure element is the disease-causing mutation located in your protein? 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  $\beta$ -strand and the second  $\alpha$ -helix.

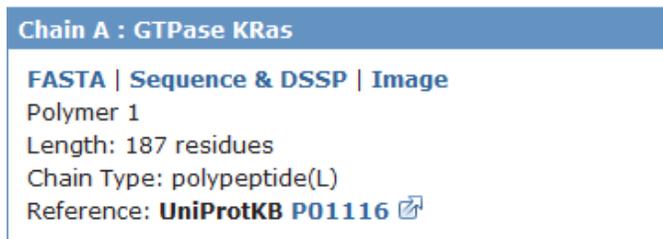
Note that the numbering of residues in the PDB file may not correspond directly with the numbering used in the medical genetic literature describing your mutation (such as in the “seed” reference given in Appendix A for your chosen topic). The reason for this is that the form of the protein that is crystallized 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

9(A) below. This causes the primary/secondary structure diagram to be redrawn with the external/UniProt numbering ABOVE the primary sequence and the PDB numbering BELOW the primary sequence. Use the UniProt numbering when referring to the mutation in your presentation.



**Figure 9.** (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.

(5) Click on the UniProtKB reference number for your protein. See Figure 10 for an example with K-Ras.



**Figure 10.** Screenshot of the UniProtKB reference number for K-Ras.

UniProtKB is a database of protein structure and function information that is freely available on the internet.

Click on the “Sequences” tab at the top of the page (Fig. 11).

There may be different isoforms of the protein that are shown on the resulting page. Select the isoform that most closely resembles the sequence you examined in the PDB file. In the case of K-Ras, isoform 2B is the most similar. In the pull-down menu to the right of the isoform you selected, select “ProParam” and press “Go” (Fig. 12).

UniProtKB

Search in: Protein Knowledgebase (UniProtKB) [Search] [Advanced Search] [Clear]

**P01116 (RASK\_HUMAN)** Reviewed UniProtKB/Swiss-Prot  
Last modified November 28, 2012. Version 160. [History]

Clusters with 100%, 90%, 50% identity | Documents (6) | Third-party data

Names · Attributes · General annotation · Ontologies · Interactions · Alt products · Sequence annotation · **Sequences** · References · Web links · Cross-refs · Entry info · Documents (Customize order)

**Names and origin**

Protein names	Recommended name: <b>GTPase KRas</b> Alternative name(s): K-Ras 2 Ki-Ras c-K-ras c-Ki-ras Cleaved into the following chain: 1. <i>GTPase KRas, N-terminally processed</i>
Gene names	Name: <b>KRAS</b> Synonyms: KRAS2, RASK2
Organism	<b>Homo sapiens (Human)</b> [Reference proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorhini › Catarrhini › Hominoidea › Homo

**Protein attributes**

Sequence length	189 AA
Sequence status	Complete.
Sequence processing	The displayed sequence is further processed into a mature form.

Figure 11. UniProtKB webpage for K-Ras, with the “Sequences” tab highlighted in red.

**Isoform 2B [UniParc] [UniParc].** FASTA 188 21,425 **ProtParam** go

Checksum: B1B6D189BB259861

```

10      20      30      40      50      60
MTEYKLVVVG AGGVGKSALT IQLIQNHFVD EYDPTIEDSY RKQVVIDGEI CLLDILDTAG

70      80      90      100     110     120
QEEYSAMRDQ YMRTEGEGFLC VFAINNTKSF EDIHHYREQI KRVKDSEDPV MVLVGNKCDL

130     140     150     160     170     180
FSRTVDTKQA QDLARSYGIP FIETSAKTRQ GVDDAFYTLV REIRKHKERM SKDGKRRKKK

SKTKCVIM

```

« Hide

Figure 12. Screenshot of the selected isoform for K-Ras, with the “ProtParam” link highlighted in red.

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

- What is the molecular weight of the protein?
- What is the pI of the protein?

(6) Return to the UniProtKB record for your protein. In the pull-down menu to the right of the isoform you selected, select “Blast” instead of “ProtParam” and press “Go”.

BLAST is a program used to find sequences (nucleotide or amino acid) in databases that are similar to your query sequence. It also provides a score based on the degree of similarity between the sequences, allowing you a measure of the extent of similarity.

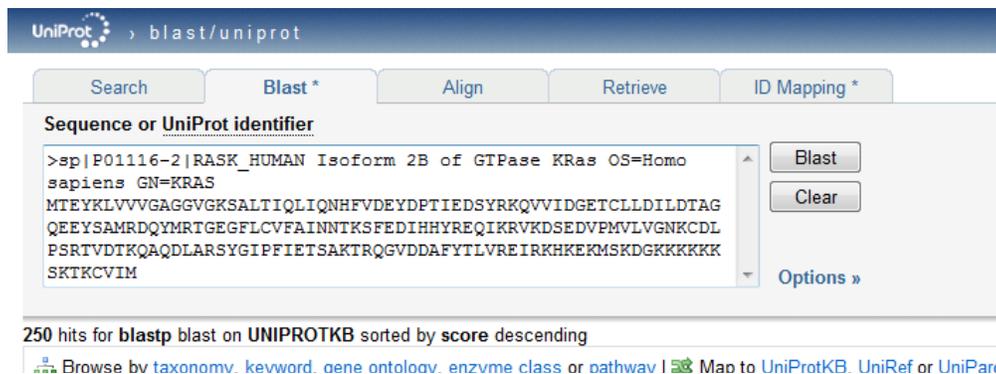
There are different types of BLAST searches that vary primarily on two factors:

- Is the query sequence nucleotide or protein?
- Do you want to search databases that contain nucleotide sequences or ones that contain protein sequences?

We will focus on a Protein blast (blastp). This type of BLAST search is done when you have a protein sequence and want to search a protein database to find similar sequences. blastp searches can be very useful to determine the function of a protein. Similarity between protein sequences may imply similar roles. Another advantage of these types of searches is the possibility to determine whether a certain protein is conserved throughout very different organisms. This may point to an important function of the protein. Our use of BLAST for this project will focus on the latter use of this software.

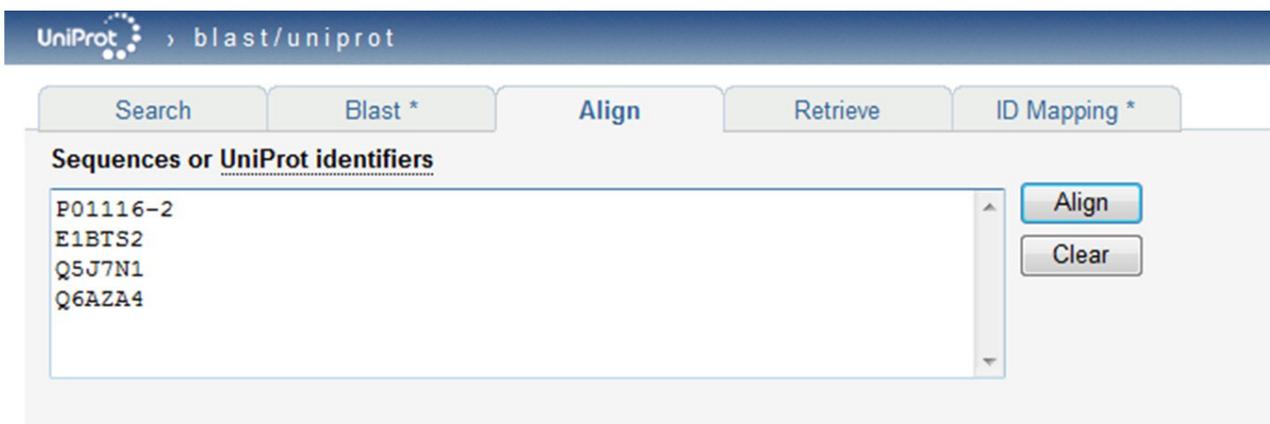
Let's get started!

Once your BLAST results appear on the webpage (it may take a minute or two!), click on the "Align" tab at the top of the page (Fig. 13).



**Figure 13.** Screenshot of the blast/uniprot page, with the "Align" tab highlighted in red.

From the list of proteins returned by the BLAST search as having some similarity to your human protein, select a protein from the following organisms: zebrafish (*Danio rerio*), mice (*Mus musculus*), and chicken (*Gallus gallus*) if available. Be sure to look through all pages of results, not just the first page to find these! Go to the bottom right-hand corner to look through the other pages of results. In the "Sequences or UniProt identifiers" box, type in the accession numbers for the human sequence, the zebrafish sequence, the mouse sequence and the chicken sequence (Fig. 14).



**Figure 14.** Screenshot, showing the accession numbers for the human sequence, the zebrafish sequence, the mouse sequence and the chicken sequence typed into the "Sequences or UniProt identifiers" box.

The alignment of sequences is made possible through the use of a program called ClustalW. ClustalW is a multiple sequence alignment program. It allows you to enter a series of protein sequences that you believe are similar in sequence and it will align the sequences so that the lowest number of gaps is introduced, and that most similar residues are aligned with one another. To perform the alignment, ClustalW uses a scoring matrix similar to BLOSUM-62, where different scores are assigned for aligning identical or similar amino acids with one another, and scores are removed when a gap must be introduced or different/dissimilar amino acids are aligned with one another.

Using your sequence alignment, please answer the following questions:

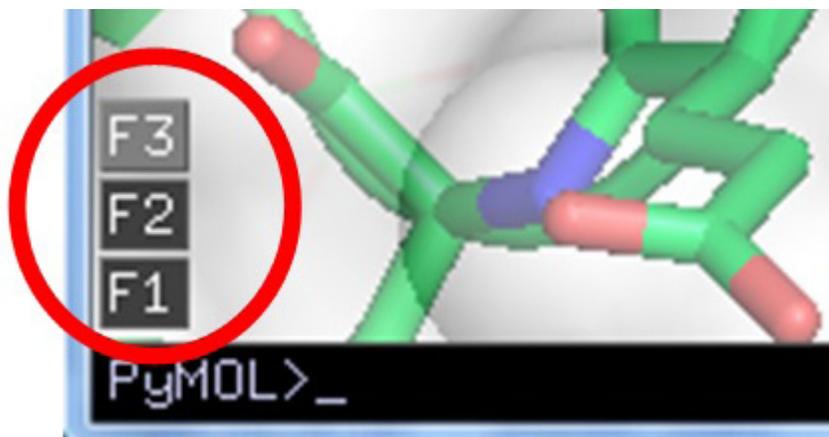
- Examine the residues identified from your literature searches to be important for activity (e.g., residues that form the enzyme's catalytic site). Are they conserved amongst the four sequences? If they are not conserved, is the substitution conservative (i.e., an amino acid with similar properties occupies the same location)? What does this tell you about the role/importance of these amino acids for the function of the protein?
- Using the secondary structure diagram you obtained previously, pick a residue that does not lie in a regular secondary structure element (i.e., is not in an  $\alpha$ -helix or in a  $\beta$ -sheet) and is not conserved in all 4 species. Make a note of this residue for the next portion of the exercise, when you will examine the three-dimensional structure of your protein using PyMOL.
- Focus on the mutated residue that leads to the disease. What is the identity of this residue in the wild-type form of the protein in each of the four species you used to create your alignment? Is this residue conserved? If not, is the substitution conservative or non-conservative? What does this tell you about the role of this particular amino acid?

### *A Look at Tertiary and Quaternary Structure*

In this section of the exercise, we will study the effects of a single amino acid substitution on protein structure and function. 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.

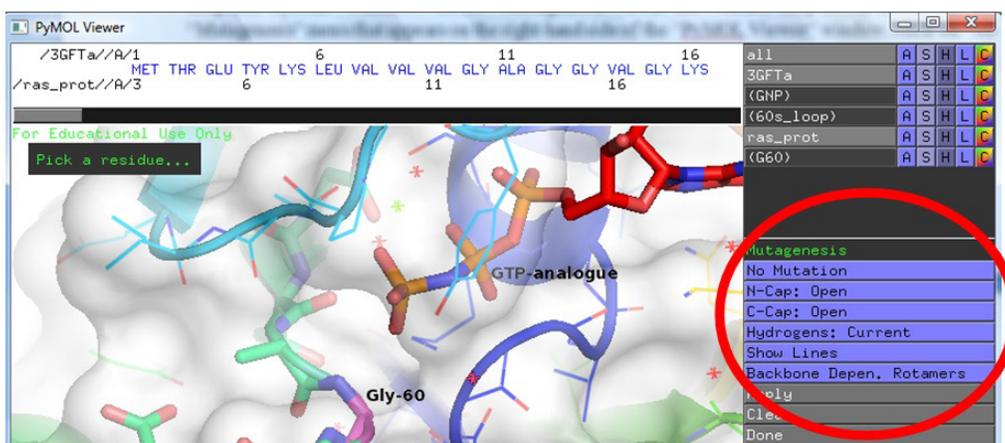
Open PyMOL and load the PyMOL file for your chosen topic (see Appendix A for file name; note that a PyMOL file ends with a ".pse" file extension). You will note that two windows open when the PyMOL program is loaded: a window called "The PyMOL Molecular Graphics System" and a window called the "PyMOL Viewer". The first window contains many different menus; one of these menus will allow you to 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 15). 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.



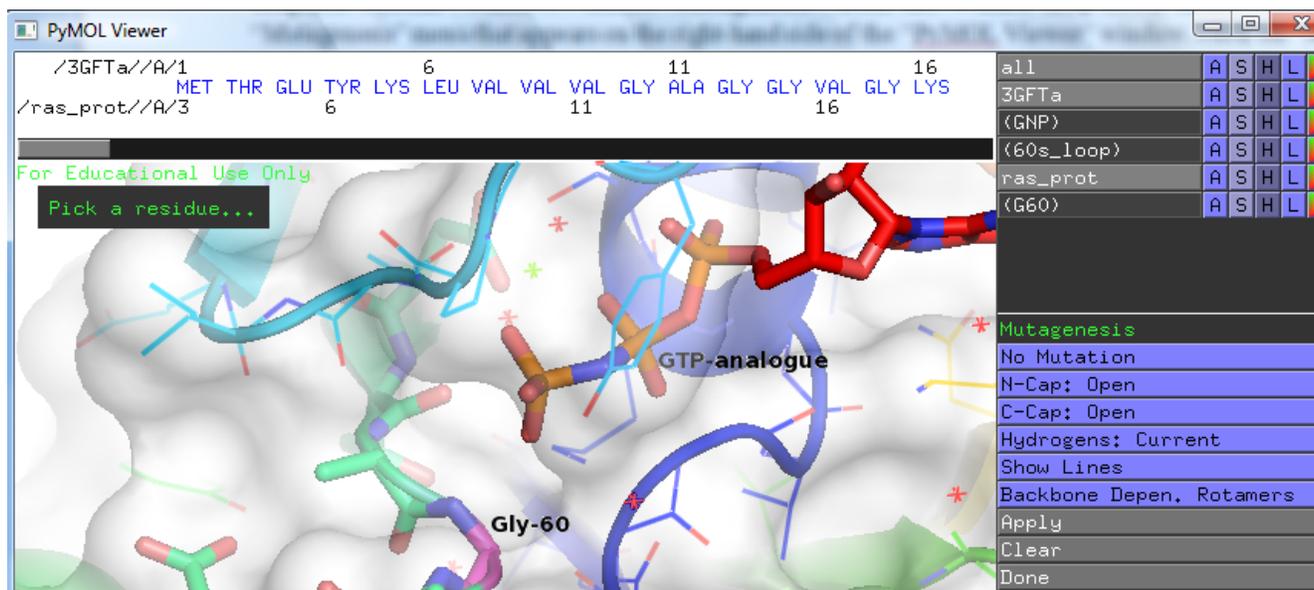
**Figure 15.** Screenshot of the "PyMOL Viewer" window, with the toggle buttons highlighted in red. In this example, three different views (available by pressing the F1, F2 and F3 toggle buttons) are available.

To mutate glycine-60 to arginine-60, 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” (Fig. 16).



**Figure 16.** Screenshot of the “PyMOL Viewer” window, with the “Mutagenesis” menu and the “No Mutation” button highlighted in red.

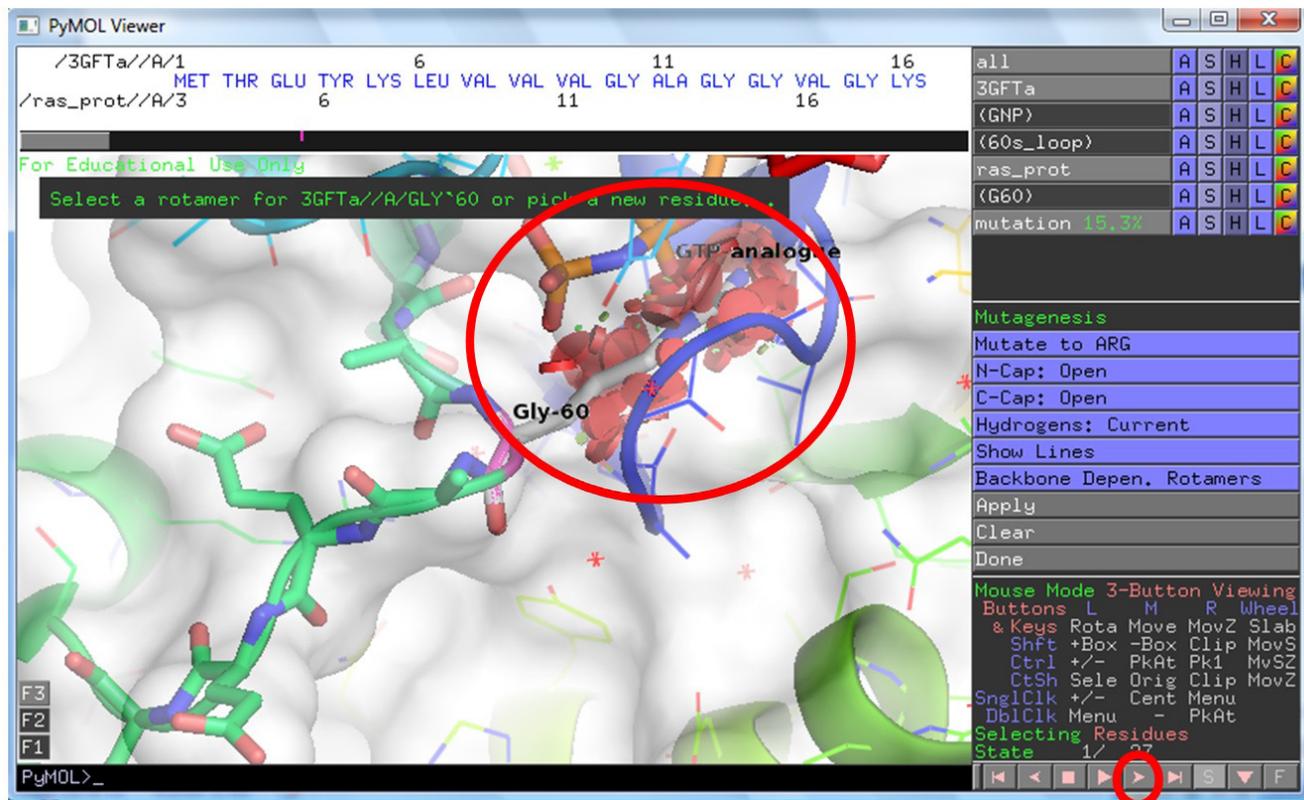
Once you click on “No Mutation”, you will now see buttons for all of the different types of amino acids. In the case of K-Ras, select “ARG”, because we want to see what happens when this residue is mutated from Gly to Arg, as has been seen with some cancer patients (Fig. 17). For your specific topic, refer to Appendix A to see the name of the residue to be mutated and the type of mutation to be made.



**Figure 17.** 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 (Fig. 18). 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 part 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 conformations) of the Arg side chain (Fig. 18).



**Figure 18.** 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 indicate why the mutation of Gly-60 to Arg disrupts the structure of K-Ras and leads to an inactive protein, which in turn leads to the development of cancer. Take a few screenshots of a few rotamers of the mutated side chain to include in your Power Point presentation.

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.

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. You should include a few pictures of the wild-type amino acid involved in different types of non-covalent interactions in your Power Point presentation. 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. You can then insert this image in your presentation.

In looking at your protein structure in PyMOL, also determine the highest level of protein structure your protein contains (tertiary or quaternary).

Go back to the residue you selected earlier when looking at the protein alignment, where you selected a non-conserved residue that did not fall within a segment of secondary structure. Where does this residue lie in the protein's three-dimensional structure (i.e., does it point inside the structure, toward the substrate/cofactor/subunit or does it point to the outside without any interaction with other biomolecules?) Are you surprised with its location in the three-dimensional structure?

### **Part C: Preparing and Presenting Your Power Point Presentation**

With your laboratory partner(s), you will present a Power Point presentation in front of your peers in your laboratory section. Your presentation should include the following:

- Background information you gathered on your disease of interest from the OMIM website, as well as from your selected review articles.
- All of the information you gathered on your disease and protein using the PDB, the UniProtKB record, the ProtParam record, the ClustalW alignment, and PyMOL, paying special attention to describing 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 inheritance pattern of your chosen disease. More specifically, you must be able to describe how the mutation in the protein affects its structure and how these structural perturbations cause the disease to be either recessive or dominant.
- Are there any therapies that can be used to combat your chosen genetic disease, and how do they relate to the protein structure?

Your Power Point presentation should contain no more than 10 slides. Please remember that Power Point presentations are VISUAL summaries and are meant to present data in an appealing fashion. Your slides should not consist only of writing. Please include as many figures and drawings as you feel are necessary. Also, be careful about font choice and font size. A good rule of thumb is to use a font size no smaller than 24 points. The use of color is always recommended to make your presentation more attractive to your audience.

Your presentation should be no shorter than 5 minutes, and no longer than 10 minutes. Remember to speak slowly and clearly. Your goal is to teach your peers the aspects of your chosen disease that you feel are most important for others to understand.

Your presentation will be evaluated using the attached rubrics (see APPENDIX B and APPENDIX C). The rubric in Appendix B will be used by your instructor to evaluate your presentation. Your instructor's evaluation of your oral presentation is worth 3% of your final grade. Note that your instructor will evaluate each member of the group separately. Please note that the rubric in Appendix B awards marks for asking 2 questions during the entire laboratory session. To be awarded these marks, you must ask 2 thoughtful questions during the session (i.e., the questions you ask must show that you carefully listened to the presentation). The 2 questions must be asked on different presentations.

The rubric in Appendix C will be used by your peers to evaluate your presentation. Their evaluation of your presentation is worth 0.5% of your final grade. Your peers will only fill out one rubric per presentation (i.e., each presenter will NOT be evaluated individually by their peers). Gaining experience in performing peer evaluation is important, and is something that you will most likely be asked to do in many different professional settings. We ask that you please take the process of peer evaluation seriously when completing the rubric. Through peer evaluation, we wish to provide feedback on both areas of strength and areas of growth, so that we all can help one another deliver great oral presentations. Please make sure to read this rubric CAREFULLY before attending your laboratory session. You should find that it also helps you prepare for your own presentation!

## Materials

A computer with Internet access and the PyMOL program (educational version freely available for download from <http://pymol.org/educational/>) is required for each pair students. LCD projector and computer are required for student presentations.

## Notes for the Instructor

One of the major challenges that we faced in implementing an inquiry-based exercise in a large class of over 500 students was to organize the exercise in a way that maximized the inquiry experience of each student without placing excessive demands on the limited time and resources of a small team of graduate teaching assistants, librarians and instructors. Several design elements of the exercise were specifically chosen to meet this significant challenge.

First, an introductory computer-based workshop session is conducted during a regularly scheduled, weekly laboratory section of the introductory biochemistry course. The relatively small groups of students in individual laboratory sections (approximately 22 students in each of 24 laboratory sections) facilitated the interactive nature of the computer-based exercises by providing opportunities for one-on-one interactions with teaching assistants and librarians, as well as peer-to-peer learning. Following this introductory session, students are given six weeks to complete the remaining self-guided exercises and to prepare their Power Point presentation, before the final student presentations.

A second key design element was to encourage a student-driven, inquiry-based learning experience. The students enjoyed making the initial choice of their research topic from a list of diseases suitable for student-directed investigations. In addition, it was particularly important to provide an open and flexible learning environment during the library and bioinformatics session to encourage student-initiated inquiry-based learning. The library session was geared towards explaining the technical aspects of bioinformatics tools and databases through the use of K-Ras as an example. K-Ras was specifically excluded from the list in APPENDIX A to allow for its use as an example during the introductory help session. The computer laboratory should be equipped with an LCD projector to allow the librarian to easily demonstrate the use of the bioinformatics tools and databases. Students are required to apply their basic understanding of genetics and biochemistry to effectively use these tools to find the information necessary for their chosen disease. The combination of librarians and graduate teaching assistants leading small groups in a loosely structured setting was particularly effective for guiding students through the technical and conceptual aspects of this inquiry-based discovery process.

Over the last four years, it has been our experience that the success of the inquiry-based discovery process depends on a few critical factors. First, bioinformatics and other non-bibliographic resources are complex and generally contain

more information than students need in an introductory genetics course. During the help session, students are shown how to focus on the most useful aspects of the tools and databases, thereby greatly reducing frustration and encouraging later exploration of other aspects of these resources. Second, the exercise should try to emulate the way in which an expert would use these tools. Using these resources to ask “real” questions lends credibility to the exercise. Third, ensure that sufficient time is provided for hands-on practice with readily available assistance from the librarian, instructor, or teaching assistant. Questions often arise regarding searching and interpreting data that require expertise.

Fourth, PyMOL is a particularly powerful and attractive molecular visualization tool, but its wide range of capabilities can introduce an intimidating level of complexity to students and instructors. To reduce the learning curve for using PyMOL, we have created eleven .PSE session files, one for each disease. Each session file can be downloaded by students and instructors to assist with their investigations of protein structure (session files can be downloaded from <http://people.ucalgary.ca/~mibarret/bcem/molmed.html>). In each session file, we have saved specifically designed views of protein structure, which students can access by pressing on a series of buttons. The views provide a common starting point for student-directed investigations. Students can rotate molecules, change the zoom level, change molecular representations, measure distances, etc. If they make a mistake and lose the original view, all they have to do is press on the view button to restore the original view. The use of these views in PyMOL greatly increases student satisfaction and reduces the complexity of instruction needed to deal with the technical issues of using the software. Instructors who want to modify the existing session files or create new session files for different proteins can consult the extensive online documentation freely available for PyMOL (<http://pymolwiki.org>).

The final design element essential for the success of the exercise was the use of an oral presentation as a final evaluation tool. It was particularly effective to use the final laboratory session for student presentations and discussion. Most students viewed the final presentation of their results as a meaningful capstone learning experience. Through this part of the exercise, students had to organize and present their work in a way that demanded a deeper level of understanding than normally required in a typical exam or laboratory report. Peer evaluation through student questioning and marking also increased the level of engagement and satisfaction with the evaluation process.

One of the most difficult challenges facing this project was to devise a way to evaluate how students performed in the inquiry-based exercises. Since the oral presentation was designed to be the culmination of the student-initiated inquiry-based learning process, the overall performance of the students in this exercise was evaluated by marking the quality of the oral presentations for each pair of students. To standardize the evaluation of students in a large number of

separate laboratory sessions, we developed a detailed marking rubric that provided specific guidance to the graduate teaching assistants regarding the grading of the final student presentations (Appendix B).

The marking rubric was carefully designed to emphasize the importance of creativity and inquiry, as opposed to a non-selective listing of information. Students were informed well in advance of their presentations that they would be marked for their creativity and the quality of their presentation, as well as for the scientific accuracy and completeness of information. As a result, students needed to master basic concepts and apply them in a meaningful way to prepare a successful presentation. To prepare their final presentation, each pair of students synthesized their abstract knowledge and applied it towards explaining a specific disease. This was one of the first opportunities for most of the students to present scientific information in a formal setting, and it proved to be a challenging but rewarding experience for most of them. As an added incentive, we informed the students at the beginning of the exercise that the students who prepared the best three posters would be selected to participate in the annual departmental research symposium. The high quality of most of the presentations attests to the success of this method of evaluation, as well as the success of the project as a whole.

### Literature Cited

- 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.
- Pauling, L., H.A. Itano, S.J. Singer, and I.C. Wells. 1949. Sickle cell anemia: a molecular disease. *Science*, 110:543-548.
- 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.

### Acknowledgments

Thank you very much to all of the BCEM 393 students and teaching assistants who have helped improve this laboratory exercise over the last four years.

### About the Authors

Isabelle Barrette-Ng has been an Instructor at the University of Calgary since 2006, where she teaches large courses in genetics and biochemistry, primarily at the second-year undergraduate level. She graduated from Queen's University in 1998 with a BSc (Hons.) in Biochemistry, received a MSc in Bioinformatics/Biochemistry from the University of Montreal in 2001 and completed her PhD studies in Biochemistry at the University of Alberta in 2003. Her interests include the development and integration of inquiry-based learning approaches in large undergraduate courses. Isabelle was awarded the Student's Union Award for Teaching Excellence for the Faculty of Science at the University of Calgary in 2008 and 2009. In 2010, she was inducted into the Student's Union Teaching Hall of Fame. In 2010, she also won the Teaching Excellence Award for the Faculty of Science at the University of Calgary. In 2013, Isabelle was co-host of the ABLE Annual Meeting held at the University of Calgary.

Don MacMillan has been the Liaison Librarian for Biological Sciences, Math, Physics and Astronomy at the University of Calgary Library since 2003. He provides program-integrated information literacy instruction and advanced reference and training services to students and faculty in those disciplines and carries out research on student learning, information literacy and the incorporation of tools and technology in information literacy instruction. Prior to his present position, Don was EMBA/MBA librarian at the Haskayne School of Business, University of Calgary and held several positions at the Calgary Public Library. Don received his Bachelor of Science and MLS degrees from Dalhousie University, Halifax, Nova Scotia.

## Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit <http://www.ableweb.org/>.

Papers published in *Tested Studies for Laboratory Teaching: Peer-Reviewed Proceedings of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

## Citing This Article

Barrette-Ng, I.H. and D. MacMillan. 2014. Gaining a Structural Perspective on the Molecular Basis of Inherited Diseases. Pages 5-24, in *Tested Studies for Laboratory Teaching*, Volume 35 (K. McMahon, Editor). Proceedings of the 35th Conference of the Association for Biology Laboratory Education (ABLE), 477 pages.

<http://www.ableweb.org/volumes/vol-35/?art=1>

Compilation © 2014 by the Association for Biology Laboratory Education, ISBN 1-890444-17-0. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner.

ABLE strongly encourages individuals to use the exercises in this proceedings volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.

## 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. <i>Acta Crystallogr. Sect. D</i> , 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. <i>J. Biol. Chem.</i> , 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. <i>J. Biol. Chem.</i> , 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. <i>EMBO J.</i> , 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 Å resolution. <i>Proteins</i> , 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 HGPRase, PRPP, Mg <sup>2+</sup> , and the inhibitor HPP reveals the involvement of the flexible loop in substrate binding. <i>Protein Sci.</i> , 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. <i>J. Mol. Biol.</i> , 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. <i>J. Biol. Chem.</i> , 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. <i>Acta Crystallogr., Sect. D</i> , 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. <i>EMBO J.</i> , 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. <i>J. Mol. Biol.</i> , 359:913-929.

## Appendix B

### Instructor Marking Rubric for Presentations

Name: \_\_\_\_\_

Topic: \_\_\_\_\_

#### Content (6 marks):

Give 2 marks if that aspect is well done, 0.5 marks for evidence of effort, and 0 for no effort.

- \_\_\_\_\_ appropriate amount of background information is presented, with selected protein and mutation clearly identified
- \_\_\_\_\_ links between the inheritance pattern and the effects of the mutation on protein structure are clearly explained (i.e., why is the disease recessive or dominant based on the mutation's effect on protein structure and function)
- \_\_\_\_\_ the effects of the mutation on the different levels of protein structure are clearly explained through a detailed description of the disruption of any non-covalent interactions and the introduction of any steric clashes

#### Presentation (6 marks):

Give 1 mark if that aspect is well done, 0.5 marks for evidence of effort, and 0 for no effort.

- \_\_\_\_\_ demonstrates good speaking skills (tone, volume, pace, avoiding “ums”, “you know”, etc.)
- \_\_\_\_\_ logical and confident delivery of material
- \_\_\_\_\_ presenter makes eye contact, engages the audience and is enthusiastic
- \_\_\_\_\_ presentation is clear and visually appealing
- \_\_\_\_\_ equal participation of group members
- \_\_\_\_\_ references are properly cited and presented in the presentation

#### Answering questions (2 marks):

- 2 marks will be given if all questions are answered thoroughly
- 1 mark will be given if the questions are only partially answered
- 0 marks will be given if the questions are not answered satisfactorily

#### Asking questions to others (2 marks):

- 2 marks will be given if 2 thoughtful questions are asked
- 1 mark will be given if only 1 thoughtful question is asked or if the questions are not thoughtful
- 0 marks will be given if no questions are asked

**Deduct 1 mark for every minute that the talk goes over 10 minutes or for every minute that is missing to reach a minimum of 5 minutes.**

Presentation mark = \_\_\_\_\_ /16 (Worth 3.0% of your final grade)

#### Comments:

(a) Areas of strength:

(b) Areas of growth:

## Appendix C

### Peer-Review Marking Rubric for Presentations

	4 – Excellent	3 – Good	2 – Fair	1 – Area of growth	Grade (out of 4)
Delivery	<p>Hold attention of entire audience with the use of eye contact, seldom looking at notes</p> <p>Speak with fluctuation in volume and inflection to maintain audience interest and emphasize key points</p>	<p>Consistent use of direct eye contact with audience, but still returns to notes</p> <p>Speak with satisfactory variation of volume and inflection</p>	<p>Display minimal eye contact with audience, while reading mostly from the notes</p> <p>Speak in uneven volume with little or no inflection</p>	<p>Hold no eye contact with audience, as entire report is read from notes</p> <p>Speak in low volume and/or monotonous tone, which causes audience to disengage</p>	_____/4
Content/ organization	<p>Demonstrate full knowledge by answering all class questions with explanations and elaboration</p> <p>Provide clear explanation of effects of mutation on protein structure and how this affects the inheritance pattern of the disease</p>	<p>Are at ease with expected answers to all questions, without elaboration</p> <p>Provide somewhat clear explanation of effects of mutation on protein structure and how this affects the inheritance pattern of the disease</p>	<p>Are uncomfortable with information and are able to answer only rudimentary questions</p> <p>Attempt to explain the effects of the mutation on protein structure, but do not adequately show why the disease is dominant or recessive</p>	<p>Do not have grasp of information and cannot answer questions about subject</p> <p>Do not clearly explain the effects of the mutation on protein structure nor how these effects cause the disease's inheritance pattern</p>	_____/4
Enthusiasm/ audience awareness	<p>Demonstrate strong enthusiasm about topic during presentation</p> <p>Significantly increase audience understanding and knowledge of topic; convince an audience to recognize the validity and importance of the subject</p>	<p>Show some enthusiastic feelings about topic</p> <p>Raise audience understanding and awareness of most points</p>	<p>Show little or mixed feeling about the topic being presented</p> <p>Raise audience understanding and knowledge of some points</p>	<p>Fail to increase audience understanding, knowledge of topic</p> <p>Show no interest in topic presented</p>	_____/4

Total: \_\_\_\_/12

**Comments:**

Areas of Strength:

Areas of Growth: