Name:

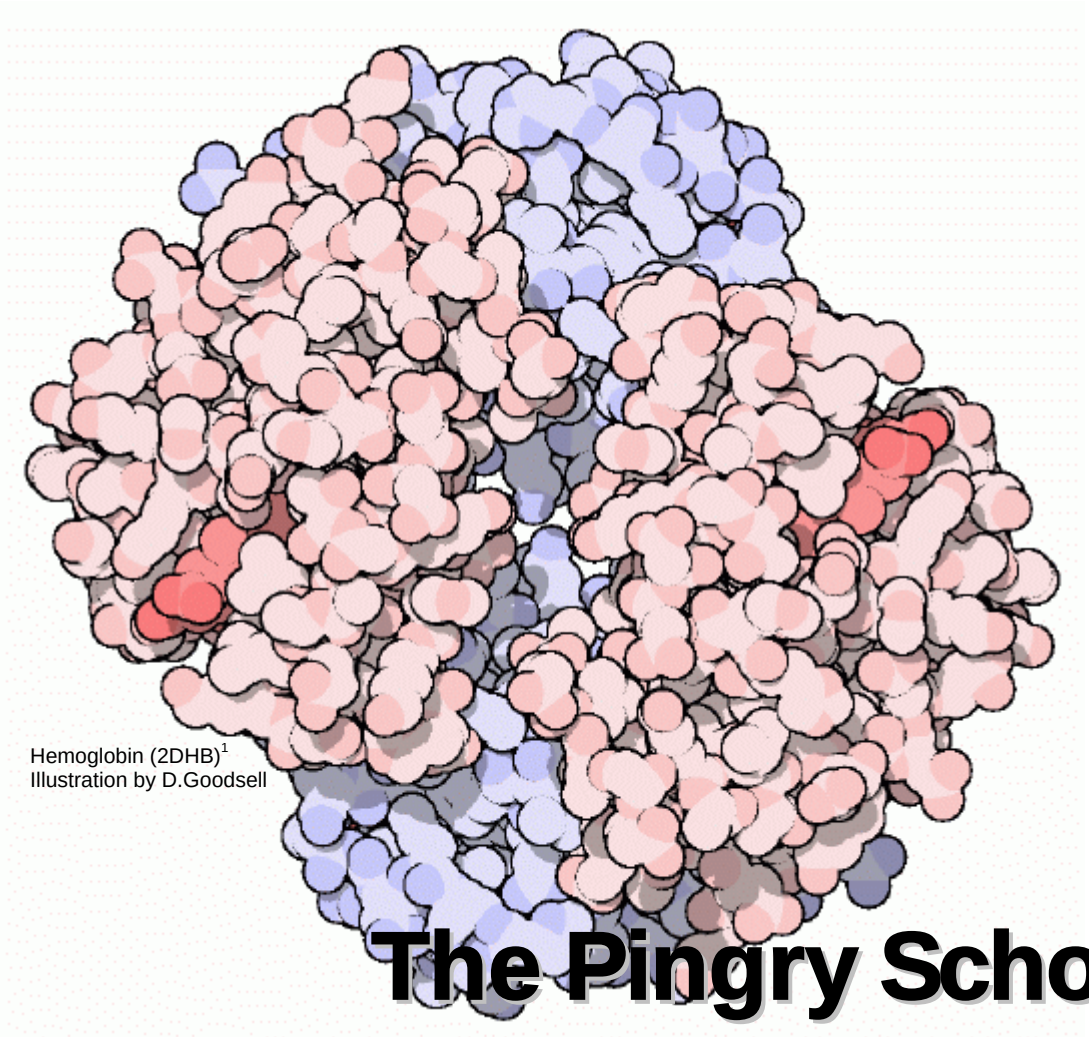
Date Distributed: September 13, 2004
Due Date: September 28, 2004

Project

1

MACROMOLECULES, ENZYMES AND CELLS

Computer Visualization of Molecular Structures: Using MDL Chime and the Protein Data Bank (PDB)



Hemoglobin (2DHB)[1]
Illustration by D.Goodsell

The Pingry School
Biology Honors Projects

http://www.mybiology.com/honors_projects.htm

Introduction

Description of the honors projects and Project #1

The Honors Projects in Biology are designed to further explore the topics discussed in class while providing you with opportunities to apply these to additional topics and concepts. As outlined in the Honors Designation Letter, you are expected to maintain a grade of B or above in the course while you progress through this material on your own. Honors designees will also meet during additional lectures and will complete individual projects that will be assessed by the biology teachers.

Students are expected to work without aid unless specifically mentioned in the project. Please turn a final draft of the project with your name on each page, along with your teacher, and each question fully typed out with a complete answer underneath. Please see your teacher if you have format questions. Make every effort to have your paper in on time. The department will not accept papers late unless arrangements are made ahead of time. You may email your teacher a copy if you absent but turn in a hardcopy when you return to school. Since this is an independent project you need to budget your time well ahead of the deadline. Please include the honor pledge on the last page.

The Topic

This first project will cover the topics of macromolecules, enzyme structure and activity, and cellular structure and function. This project will also introduce you to a number of computer resources that you will explore further in subsequent projects. Learning to identify and use these computer resources is crucial to successfully completing the honors projects.

This project is broken up into a number of smaller modules. Read the directions for each module carefully and complete the accompanying readings. Readings will be pulled from additional texts and resources on the internet. Each module is accompanied with questions which will be assessed by the biology teachers at the end of the project.

Keeping up with changes in Biology

Biology and related fields are experiencing rapid changes as improved technology allows us to better understand what goes on in the living system. Arguably the most significant accomplishment in recent years was the announcement of the working draft of the Human Genome in 2000 and its completion this past April, 2003. The Human Genome Project was a world-wide effort to "reading" and "mapping" the 6.4 billion A, T, G, C nucleotides (or 3.2 billion pairs) contained in the nucleus of a human cell. Recall from your readings that these letters in DNA are the "instructions" for building organisms. These instructions are organized into functional units known as genes. Currently there are about 30,000 genes. Among DNA's functions (which we are still trying to fully understand) is coding for the primary structure of proteins. The resulting proteins(over 100,000 in humans) and enzymes are responsible for carrying out all functions in the cell; including the synthesis and hydrolysis of macromolecules, catalyzing biological reactions, transport and storage of molecules and ions, coordinating motion and structural support, signaling, and control of growth and other processes. According to the National Center for Biotechnology Information, the Human Genome Project has lead to identification of the genetic causes of diseases including cystic fibrosis, breast cancer, hereditary deafness, hereditary skeletal disorders, and a form of diabetes[2].

Scientists are also completing genomes of additional organisms, which allow the comparison of organisms at the genetic level. Integration of computer programs to understand, "read" and compare these genomes has opened up a brand new field collectively called *bioinformatics* and *computational biology*. Through this project, you will use some of the tools used in bioinformatics to explore the structure and functions of a number of enzymes and proteins. We will use this understanding of protein function to discuss additional topics in cellular structure and processes. Through the rest of the Honors Projects, you will continue to build upon your understanding of biology and extend the use of different tools and skills through topics in energetics, genetics, biotechnology, and evolution.

Topics

The following are topics and themes discussed in class and through the course readings. Make sure that you are comfortable with these basic concepts before beginning appropriate modules in the Honors Projects. The readings in the Honors Projects assume your familiarity with these concepts.

1. Macromolecules
2. Protein Structure
    -Amino acids and primary structure
    -Properties of individual amino acids and its result on protein structure
    -Role of "R" groups in tertiary structure and protein function
    -Relationship of structure to function
3. Enzymes
    -Enzyme specificity and reaction catalysis
    -Enzyme "active site"
4. Cells
    -Types of movement across cell membranes
    -Structures and function of organelles

Honors Projects 2004-2005

Project 1 Outline

Protein Structure and Function

September 2004

1. Introduction to MDL Chime and computer visualization of proteins

    A closer look at protein structure: Introduction to the Protein Data Bank (PDB).

    -myoglobin

    -hemoglobin

    More on "R" groups and protein function

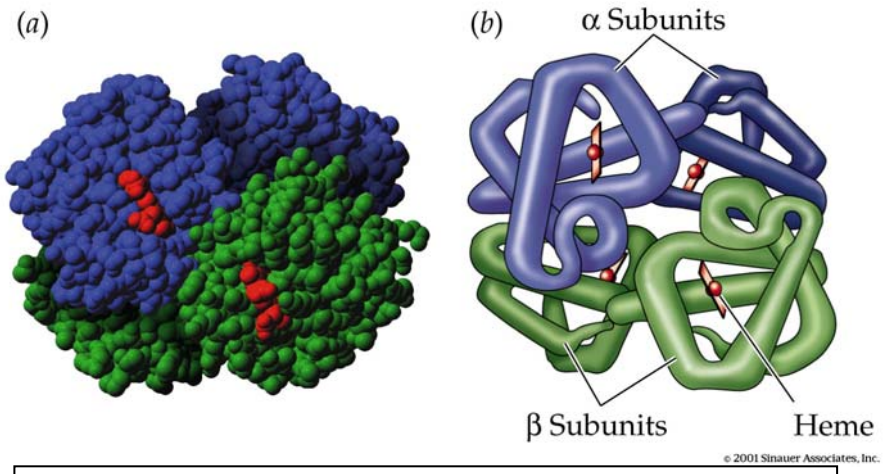
    -Human serum albumin (HSA) (PDB:1AO6)

    -Retinol-binding protein (RBP)

Module 1.1

A closer look at protein structure: Introduction to the Protein Data Bank

You have been introduced to the basics of protein structure through our discussions in class. You know that proteins are made of one or more polypeptides, which in turn are long polymers of amino acids. Countless numbers of proteins are created in our cells every day. By now, you have seen a number of "pretty" pictures displaying the complex and elaborate structures of proteins. You have seen these pictures in your textbook, your teachers have shown them to you in class, and you might have even seen them in other media such as magazines. You may have even used computer based visualization programs to rotate and manipulate different molecules and proteins.



(a) Spacefill display of human hemoglobin. Hemoglobin is made up of four polypeptides (two alpha- globins and two beta-globins). The shaded regions in the center are the heme groups, a non-protein molecule that is responsible for binding oxygen and other gases. (b) Cartoon representation of the same file. The structure of the backbone is emphasized without displaying the sidechains of the amino acids. Notice the position of the heme groups.

But how do we know what proteins look like? Proteins are so small that it is impossible to see them with the naked eyes even with the best microscopes available. So how do we "solve" the structure of proteins and other molecules? How has this technology allowed us to better understand the functions of these molecules? How can understanding protein structure also help scientists study and understand genetic diseases, poisons and toxins, and other biological functions?

Through this module, you will learn about the Protein Data Bank (PDB), an international effort and collaboration to collect and categorize the "solved" structures of proteins and other molecules. You will learn to navigate through the PDB and view the structure of molecules of interest. You will also learn to manipulate MDL Chime, a simple, free program that allows you to manipulate the files available through the PDB.

Downloading and Installing MDL Chime to view molecules on your computer:

The first thing you must do is to download MDL Chime onto a computer you are using or to use any of the PC's at The Pingry School (Chime is installed on all laptops and desktops in the computer lab). The file is available for download through the following link:

http://www.mybiology.com/chime/download.htm

When you click on the link to download the program, you will be prompted for a login and password. Login is "pingry" and password is "biology", both without the quotations marks.

Download the program onto your computer first, then install by following the directions. MDL Chime installs as a plug-in to your web browser. After installation, you should be able to use the chime tutorials that were shown in class along with many others that are available on the web. You will be required to navigate through a few tutorials to get acquainted with the use of the program.

Learning to use Chime:

Go to the website www.mybiology.com/chime and click on "View Molecules" on the left menu. There are a number of separate Chime modules on this page. Work through the following: Glucose, Catalase, Hemoglobin, and Immunoglobin. There are questions along with the tutorial that can be answered below but your final draft will require you to prepare your answers on a separate paper. Please see the description section on page one.

I. Catalase:

3.

4.

6.

9.

II. Hemoglobin:

3.

4.

5.

III. Immunoglobin:

2.

3.

9.

A closer look at protein structure: Introduction to the Protein Data Bank (PDB)

All of the molecules that you viewed in the preceding activity were crystallized by scientists in order to "solve" its structure. As mentioned earlier, individual proteins are too small to see through a microscope. Through a technique called X-ray crystallography, scientists are able to determine the three-dimensional (x,y,z) coordinates of each atom that makes up a protein. This data is then processed by a computer to create the files we can view through programs such as MDL Chime.

The Protein Data Bank (PDB) is an international depository of protein and molecule structures that have been solved by scientists across the world. When a team of scientists "solve" the structure of a particular molecule, that information is catalogued by those at PDB. The Protein Data Bank's main office is located at Rutgers University in New Brunswick, NJ. There are also a number of other mirror sites around the world.

In this part of the module, you will learn the basic skills to navigate through the PDB website to access their resources. You will then review a number of learning modules that are posted on their website to further explore protein structure and function.

Access the PDB website:
www.pdb.org

In the PDB, each structure is assigned a four alpha-numeric ID number. For example, the hemoglobin shown on the cover of this packet has ID number 2DHB. Every time you come across such ID, you can enter it into the search field to access information regarding that structure. For this part of the module, you will not be asked to research through the database; you will be given specific files to look up and explore.

On the left margin of the PDB website, you should see a feature titled "Molecule of the Month." Below this link, there is another link for the current month's molecule. The molecule for November is Simian Virus 40. It may be good to note here that there are two separate links in this area; "Molecule of the Month" is separate from "Simian Virus 40." We want to access "Molecule of the Month" to browse molecules that were featured in the past.

If you click on this link, you should see a brief explanation of the “Molecule of the Month” feature along with links to past molecules. Be sure to read the explanation.

Read the following two modules (keep in mind that there are multiple pages per module). You can access them through the appropriate links from the main page or alternately use the links provided below.

Myoglobin (2000): http://www.rcsb.org/pdb/molecules/mb1.html
Hemoglobin (2003): http://www.rcsb.org/pdb/molecules/pdb41_1.html

While you read through these modules, you should notice that the author often refers to specific PDB files using the four letter code. You will later learn to access more detailed information regarding each file using this code.

After reading the two modules, answer the following questions:

IV. What is the function of myoglobin and hemoglobin? How are they similar? How do they differ?

V. How does the structure of myoglobin differ from hemoglobin? How are they similar?

VI. The “heme” group is a prosthetic group (a non-protein component of a large protein complex) containing iron, nitrogen, oxygen, and carbon. On the fourth page of the hemoglobin module, you see that the amino acid histidine is responsible for interacting with the heme group in each polypeptide. This heme group is then responsible for interacting with oxygen. Based on the reading, explain in your own words the mechanism through which hemoglobin transports oxygen. Relate the importance of structure to its function.

VII. The module explains that there could be slight changes in the amino acid sequence of hemoglobin chains. Describe one way that scientists manipulated the sequence to improve the quality of hemoglobin. Describe one way that a change in the sequence can lead to a potentially debilitating disease.

VIII. What makes red meats (beef, lamb) red? (Hint: it’s not the blood) Why do you think some fish have red flesh, like large tuna, while most have white flesh? What determines this?

More on "R" groups and protein function

-Human serum albumin (HSA) (PDB:1A06)

Read the module on serum albumin in the "Molecules of the Month" feature (2003).

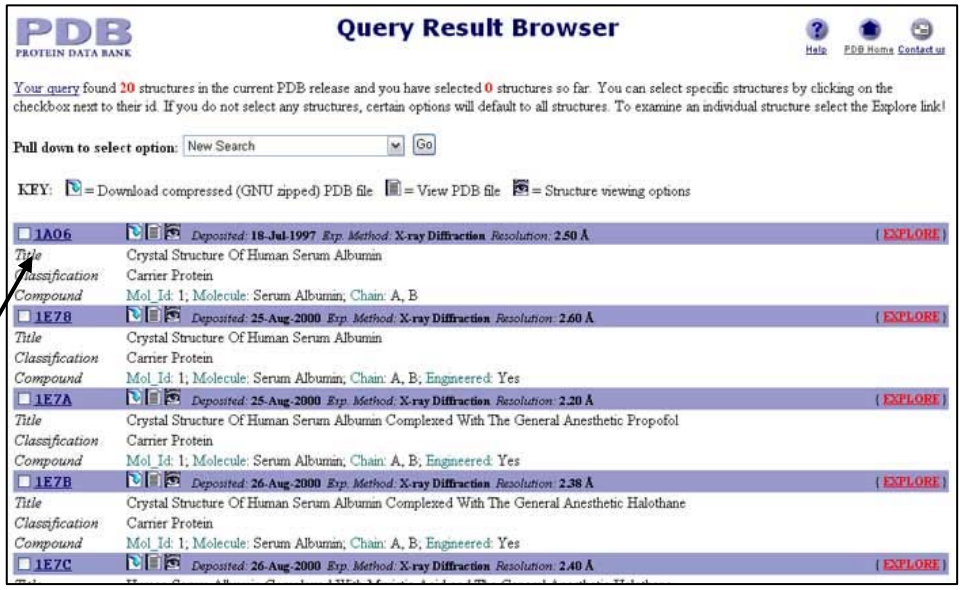
IX. Describe the functions of albumin along with its molecular structure that is significant for this function.

X. Serum albumin is an extracellular protein (found outside of a cell) that is a component of our blood. As you can imagine, the chemical environment of our blood can be harsh relative to that inside of a cell. Recall that proteins easily denature with changes in pH, salt concentration, and temperature (all of which fluctuate a lot more in the blood than inside the cell). Without looking at the structure of serum albumin, think about what type of tertiary interactions are favorable for a protein that functions in a harsh environment? In other words, for serum albumin to remain functional in the blood, without getting denatured, what types of tertiary interactions would it depend upon? What specific amino acid would you expect to find more frequently than in other proteins?

In order to verify your prediction, you will access the pdb file for serum albumin, learn how to download it onto your computer so that you can view it using MDL Chime, and apply the menu commands you learned earlier to identify individual amino acids and types of interactions.

Enter the code 1ao6 into the search field. You should get the following screen:

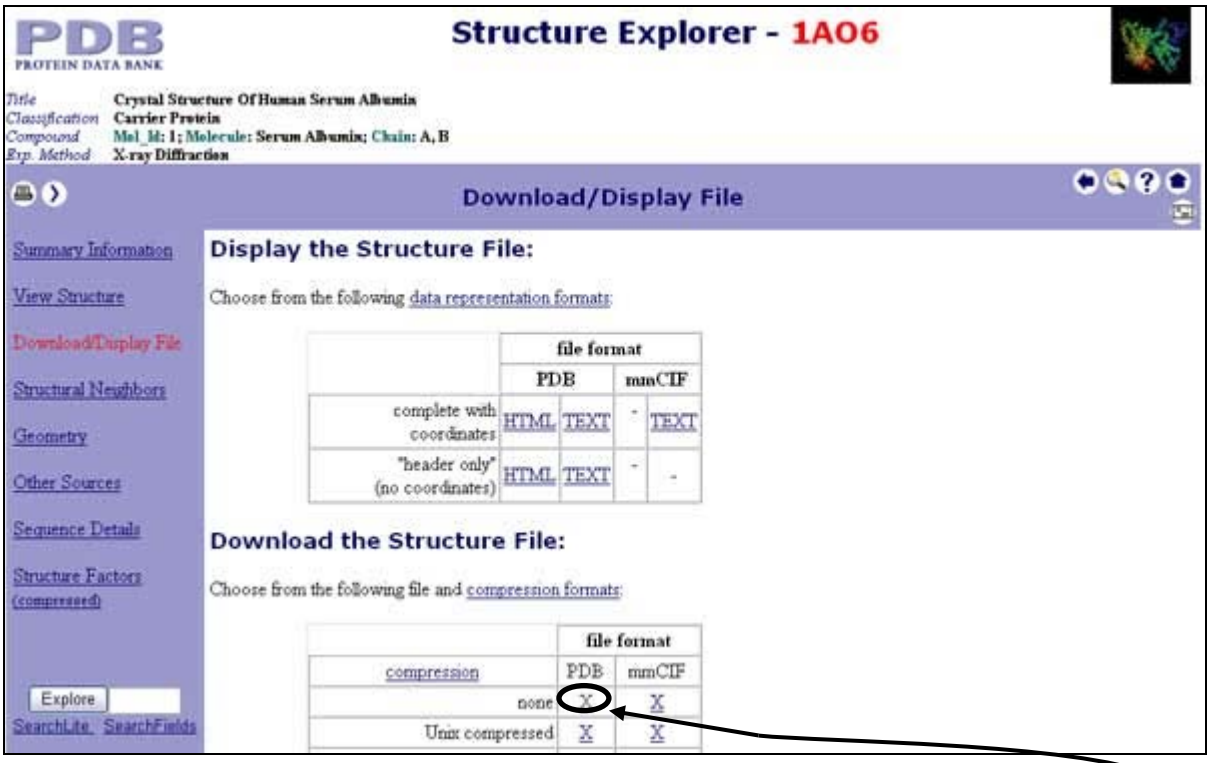
Select the file that we want to view by clicking on the appropriate code:



In the next page, you will see a summary of the file 1ao6. See that the title of the structure is "Crystal structure of human serum albumin." It was deposited into the PDB on July of 1997. In later projects, we will also learn to access the primary literature that has been published in association with the deposited structure file. The literature will provide you with specific information regarding the functions and significance of the molecule.

Downloading the structure file (.pdb file)

In order to download the actual PDB file so that we can manipulate it through Chime, first go to "Download/Display File" on the left menu. If you scroll down half a page, you should come to the following page:



We want to "Download the Structure File" in the PDB format. Read through the next few steps before you continue since these steps could be confusing.

There is no link through this page that allows you to directly display the PDB file on your web browser. You must first download the file onto your computer then open it by selecting the file.

a. Right click and "save target as..." the link marked "X" marked above:

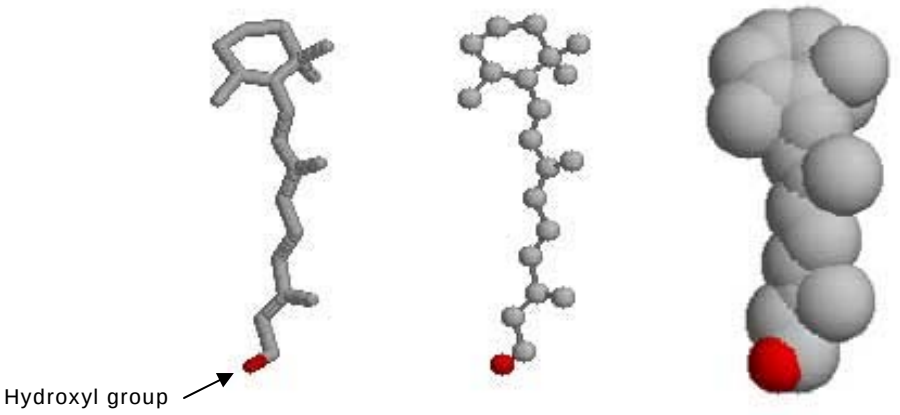
b. Save the file somewhere on your computer that you can access easily.

c. Open the file by double clicking through your desktop. If you get the message telling you that there is no program associated with the file type, manually select to open through MS Internet Explorer (your web browser).

XI. Spend time to explore the structure of serum albumin. Were your earlier predictions accurate? What characteristics can you now identify? What bonds/interactions/residues (if any) are abundant in this structure?

-Retinol-binding protein (RBP)

Below is the molecule retinol (vitamin A).



Retinol:
Display in (1) sticks, (2) ball and stick, and (3) spacefill.

Retinol is an amphipathic molecule. Which region do you think is hydrophobic? Which region do you think is hydrophilic? Label these properties on one of the models above.

XII. So far you have seen examples of proteins that act as transporters in our body: hemoglobin transports oxygen and albumin transports fatty acids. Using your understanding of the nature of interactions in these two transporter proteins, think about the possible structure of the protein responsible for transporting retinol in our bodies. In a paragraph, describe the structure of your hypothetical protein and your reasoning for its structure (i.e. why did you choose your particular structure?). Supplement your answer with a diagram displaying the interaction between your hypothetical protein and a retinol molecule. Although your diagram will be 2-dimentional, do your best to describe the 3D nature and structure of your protein.

Notes and Bibliography

1. PDB ID: 2DHB
Bolton, W., Perutz, M. F.
Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 Angstrom units resolution.
Nature 228 pp. 551 (1970)

2. Research Collaboratory for Structural Bioinformatics
H.M.Berman, J.Westbrook, Z.Feng, G.Gilliland, T.N.Bhat, H.Weissig, I.N.Shindyalov, P.E.Bourne
The Protein Data Bank
Nucleic Acids Research, 28 pp. 235-242 (2000)

3. National Center for Biotechnology Information. <http://www.ncbi.nih.gov>

4. MDL Chime is a registered product of MDL Information Systems, Inc. The distribution and use of the product through www.mybiology.com is limited to educational use only.

5. All "Molecule of the Month" features are illustrated and written by David S. Goodsell of the Scripps Research Institute.

6. PDB ID: 1MBN
H. C. Watson, J. C. Kendrew
The Stereochemistry of the Protein Myoglobin
Prog.Stereochem. 4 *pp.* 299 (1969)

7. PDB ID: 1AO6
S. Sugio, S. Mochizuki, M. Noda, A. Kashima
Crystal Structure of Human Serum Albumin
Protein Eng. 12 *pp.* 439 (1999)

8. PDB ID: 1AQB
G. Zanotti, M. Panzalorto, A. Marcato, G. Malpeli, C. Folli, R. Berni
Retinol-Binding Protein (Rbp) From Pig Plasma
Acta Crystallogr D Biol Crystallogr 54 *pp.* 1049 (1998)

The Pingry School Biology Honors Projects have been developed and written by Tommie S. Hata during the 2003-2004 school year. Thanks to those who have contributed by offering suggestions and providing help with content. A special thanks to Dr. Tim Herman and Dr. Mike Patrick of the Center for BioMolecular Modeling for their inspiration through the summer modeling institute and those at the Protein Data Bank for their work with the PDB website.

Project edits and revisions:
September 2004: Deirdre O'Mara