

# Exploring Proteins and Nucleic Acids with Jmol

Revised March 30, 2023

Jeffrey A. Cohlberg

Department of Chemistry  
and Biochemistry

California State University, Long Beach

Copyright 2010, Jeffrey A. Cohlberg  
Copyright 2012, W. H. Freeman and Company

## Contents

A Note on This Edition . . . . .	2
Preface . . . . .	3
Guide for Instructors: Which Exercises Should You Select? . . . . .	5
To the Student . . . . .	8
Downloading Jmol . . . . .	10
Exercise 1: Some Basics of Secondary and Tertiary Structure . . . . .	11
Exercise 2: Lysozyme, a Globular Protein with Disulfide Bonds . . . . .	23
Exercise 3: Beta Turns and Pi Helices . . . . .	28
Exercise 4: Supersecondary Structure . . . . .	31
Exercise 5: Lysozyme with Bound Inhibitor: Working with Ligands . . . . .	35
Exercise 6: Fibrous Proteins . . . . .	39
Exercise 7: Searching for PDB Files . . . . .	45
Exercise 8: What's in a File? . . . . .	48
Exercise 9: Domains . . . . .	51
Exercise 10: Subunits – Asymmetric Units and Biological Assemblies . . . . .	57
Exercise 11: Hemoglobin . . . . .	61
Exercise 12: The Mechanism of Chymotrypsin . . . . .	65
Exercise 13: Aspartate Transcarbamoylase . . . . .	68
Exercise 14: Membrane Proteins . . . . .	71
Exercise 15: Complexes of the Electron Transport System . . . . .	75
Exercise 16: Photosynthetic Reaction Centers . . . . .	83
Exercise 17: DNA . . . . .	90
Exercise 18: Transfer RNA . . . . .	91
Exercise 19: Ribosomes . . . . .	102
Reference section on using Jmol:	
Summary of Jmol Script Commands . . . . .	105
Printing and Saving an Image . . . . .	109
Saving Scripts . . . . .	110
Creating Labels . . . . .	110
Adding Hydrogens to a Structure. . . . .	111
Working with NMR Structure Files . . . . .	112
JSmol, a Web Browser Based Version of Jmol . . . . .	114
Jmol and Chime Web Pages . . . . .	115
Term Paper Assignment . . . . .	116

## A Note on This Edition

*Exploring Protein and Nucleic Acid Structure with Jmol* is a manual containing self-guided exercises that enable the user to learn how to download structures from the Protein DataBase (PDB) and display them and manipulate them using the program Jmol. The user gradually acquires the skills to use Jmol while proceeding to learn the basics of protein structure in a manner that parallels the way they will be learning the material in an introductory biochemistry course.

A manual similar to this one has been used in biochemistry classes at California State University, Long Beach, since 2001.

From 2011 to 2017 an earlier version of this manual was part of the Launchpad online package that accompanied the texts *Lehninger Principles of Biochemistry*, by Nelson and Cox, and *Biochemistry*, by Berg, Tymoczko and Stryer, both published by the WH Freeman division of Macmillan and Company. Separate versions were tailored to the Lehninger and Stryer books. It was available, in both a PDF version and an interactive version, to instructors using either text and to students who purchased access to Launchpad. An E-book version of the manual could also be purchased separately. The manual was revised once or twice a year to keep up with changes in Jmol and in the PDB web portal.

The present manual is freely available to both instructors and students. I will update it as needed so that the instructions accurately reflect what students will see when they access the PDB web site and use newer versions of Jmol. If anyone using the manual discovers any errors, ambiguous or unclear instructions, or anything else that needs to be altered, please let me know, and I'll see that the necessary changes are made promptly and the corrected version posted at [www.csulb.edu/~cohlberg/Jmolmanual.pdf](http://www.csulb.edu/~cohlberg/Jmolmanual.pdf).

## PREFACE

Accessing the Protein Data Base (PDB) and downloading structure files and displaying them with molecular modeling programs have become indispensable skills for a biochemist. A researcher studying a protein often needs to examine the three-dimensional structure of the protein in order to plan experiments and interpret data. Other experiments may rely on information about the domain structure of the protein and the presence of conserved sequence motifs. Pharmaceutical companies use protein 3-D structures to help design drugs that will bind to sites on the protein. A student getting a degree in biochemistry ought to have some ability to use the PDB and to visualize and manipulate 3-D molecular structures on a computer.

In addition, interactive molecular graphics can be of great value in helping beginning students understand the structures of proteins and nucleic acids. Static figures in textbooks can be useful if they are well-designed, but only to a point. Viewing animations gives a student more insight than text figures, but not as much as interactive exercises in which the students controls the way molecules are displayed.

This manual is designed to be a learning aid to supplement a biochemistry course in which students are learning about protein and nucleic acid structure. The self-guided exercises in this booklet enable the user to learn how to understand the information on PDB web pages, and how to download structures from the Protein DataBase (PDB) and display them and manipulate them using the program Jmol. The user gradually acquires the skills to use Jmol while proceeding to learn the basics of protein structure in a manner that parallels the way they will be learning the material in an introductory biochemistry course.

In addition to use by beginning biochemistry students, this manual will be useful to many others who want to learn the basics of Jmol.

A manual similar to this one has been used in biochemistry courses at California State University, Long Beach, since 2001. Since then all biochemistry majors, and most chemistry and biology majors, graduating from CSULB have learned the basic skills of using the PDB and doing basic molecular modeling.

Until 2010, the exercises were based on the use of a different program, Rasmol, which is a precursor of Jmol. Rasmol cannot be used with Mac operating system 10. In addition, many educators had used Chime, a sort of web version of Rasmol designed to be used as a "helper application" for web browsers, to compose biochemistry education sites with animated interactive images. But Chime is not compatible with many, if not most, current web browsers, and is used most effectively with an older version of Netscape.

As a consequence, there was a need to have a new program that retained the simplicity of Rasmol but did not have these compatibility problems. Jmol has now replaced Rasmol as the program referenced by leading biochemistry textbooks. The simple scripting language that was one of the best features of Rasmol has been retained in almost unchanged form in Jmol. Also, many of the web sites containing Chime images have been rewritten to use Jmol, which operates as a Java applet.

Jmol is capable of much more than what a user will learn in this manual. Interatomic distances and electrostatic potentials can be measured. Educators can use Jmol to produce animations of both small organic molecules and macromolecules. In addition to the standalone version of Jmol used in this manual, there is now JSmol, a web-based version of Jmol that can be run in a browser window. I have included instructions for using JSmol in a reference section at the end of the manual. All the exercises can be done in either Jmol or JSmol.) I have not attempted to cover everything that Jmol can do and have restricted myself to what students need to know to manipulate static molecular images.

We have found that students can generally follow the instructions in the manual without any assistance, with only a few needing a little help in getting started. Instructors who are not themselves proficient in Jmol or in molecular modeling in general can still assign these exercises to their students.

I have had positive feedback from a number of biochemistry professors who have used the manual in their teaching, and I would appreciate feedback from others. If you find any errors or problems with these exercises, or if you have any suggestions for improving them, please let me know!

Jeff Cohlberg, cohlberg@csulb.edu

I would like to acknowledge the support of a grant from the California State University Program in Education and Research in Biotechnology (CSUPERB) in the initial development of this curriculum, as well as the valuable editorial assistance of Andrea Gawrylewski at W. H. Freeman for help in producing the first versions used in their Launchpad online resources.

## GUIDE FOR INSTRUCTORS: WHICH EXERCISES SHOULD YOU SELECT?

The exercises are written so that a student who goes through them in sequence will gradually become proficient with Jmol. However, the instructions are written so that all the essential information on how to use Jmol is fully contained in just a few exercises.

Exercise #1 teaches the student how to download and display a structure file in Jmol and provides examples of most of the Jmol commands needed to manipulate the display of a molecule. Students must complete Exercise #1 before attempting other exercises.

Exercise #2 explores globular protein structure and teaches the students how to work with disulfide bonds. This is a required followup to Exercise #1 and is a prerequisite for some of the following exercises.

Exercise #5 shows the students how to construct Jmol commands for working with ligands that bind to proteins. This skill is needed for a number of the other exercises.

Exercise #10 teaches the students how to work with files for multi-chain proteins in which the Asymmetric Unit is different from the Biological Assembly. Students learn about this distinction and learn how to download Biological Assembly files and properly display the structures. These skills are required for many of the later exercises, generally all those that deal with multi-chain structures. However, in the lesson on hemoglobin (Exercise #11), the files used are ones in which the Asymmetric Unit and Biological Assembly are the same; consequently this exercise can be done by someone who has not completed exercise #10.

Finally, Exercise #17 introduces the terminology used by Jmol to select parts of nucleic acid structures, and these skills are required for the exercises on RNA and ribosomes.

Exercise	Prerequisites	Exercise	Prerequisites
1	none	11	1 and 5A
2	1	12	1 and 5A
3	1 and 2	13	1, 5A and 10
4	1, 2 and 3	14	1, 5A and 10
5	1 (2 recommended)	15	1, 5A and 10
6	1	16	1, 5A and 10
7	none	17	1 and 10
8	none	18	1 (17 recommended)
9	1 and 5A	19	1 and 10 (17 recommended)
10	1		

The exercises chosen will depend on the goals of the instructor:

First, the exercises can be used to give students a better understanding of macromolecular structure by allowing them to display and manipulate these structures:

If your goal is to help students understand protein three-dimensional structure, you can assign exercises #1 and #2. These will help students to appreciate the nature of secondary and tertiary structure and the way globular proteins are organized. Exercise #9 will show students the way larger polypeptide chains are organized into domains.

Other exercises can be selected to help students acquire a better understanding of specific topics in biochemistry. For those instructors who go into detail on fibrous proteins, exercise #6 will help students appreciate the nature of the coiled coil and the collagen triple helix. Exercise #11 allows the students to examine and compare the R and T states of hemoglobin, including zooming in on the heme binding sites, and Exercise #13 allows a similar comparison for the two conformations of aspartate transcarbamoylase. Exercise #12 allows students to explore the orientation of the side chains that are important in the catalytic mechanism of chymotrypsin.

A number of other exercises can be assigned once the student has done Exercise #10 and learned to work with Biological Assembly files. Exercise #14 guides the students through explorations of a few different integral membrane proteins. Exercises #15 and 16 allow the students to display the structures of the electron transfer complexes of both mitochondrial respiration and photosynthesis; in both cases displaying the redox centers and looking at their orientation helps the students to understand how these complexes operate.

The last few exercises are concerned with DNA, RNA and ribosomes. Many biochemistry courses present these topics in a manner that treats the structure in much less detail than is used for proteins. Students who are presented with figures of the three types of DNA helix accompanied by brief descriptions will gain a much better understanding of these structures by performing Exercise #17. The same is true for transfer RNA, where the nature of the L-shaped structure and its relationship to the cloverleaf structure, as well as the nature of the tertiary interactions, will be much clearer to a student who has performed Exercise #18. Exercise #19 allows the students to observe and manipulate images of ribosomes with three tRNAs and mRNA bound.

Second, the exercises can be used to give students the skills to download protein and nucleic acid structures from the Protein Database and to explore them on their own. Nearly every student who pursues a career in the biomedical sciences, especially those going on to graduate study, will find these skills important. A student working on a research project involving a protein can use molecular modeling software to locate critical residues and plan experiments intelligently. Those who work in the pharmaceutical industry can use molecular modeling to examine drug interactions. A student who learns Jmol will find it easy to learn other molecular modeling programs subsequently. No matter which software the student uses, an ability to use the PDB search engine to locate appropriate structure files, to read and use the information on the PDB Structure Summary page, as well as the header of the PDB file itself, in a knowledgeable fashion, and to download and correctly display both Asymmetric Unit and Biological Assembly files, will make it easier to investigate new structures and to interpret the images. For this purpose, assigning Exercises #6, 7, and 10 are important.

In my own first-semester biochemistry classes at California State University, Long Beach, I gave the students five assignments, spaced a week or two apart, in which

students were required to complete specific exercises and hand in the answers. They were: I (Exercises 1 and 2), II (Exercises 3 and 6), III (Exercises 9, 10 (including one of the Further Explorations) and 11), IV (Exercises 5, 12 and 13) and V (Exercises 7 and 8). In the second semester, I assigned Exercises 16 through 19 (we covered DNA and RNA structure in our second semester).

One way to get students to apply these skills is to assign a term paper in which each student writes about a protein of their choosing which they illustrate with figures composed in Jmol. About one month before the end of the semester, the students chose their proteins, and they then had about three weeks to work on a paper (1-1/2 to 3 pages of double-spaced text, plus at least three Jmol figures, each with a title and legend). Most of the students were able to do a satisfactory job, and the best results were of publication quality.

Note that each exercise includes at least one question in which the students are directed to produce and hand in a figure constructed with Jmol. This makes it harder for a student to copy the answers from a friend and hand it in as their own work. If a more rigorous test that the student is not handing in copied work is desired, then one could require that the figure be authenticated as described on the last page of the manual for the figures in the term paper assignment, or additionally one could require that the student submit a script file (see the reference section on "Saving Scripts") that will produce the image when opened.



## TO THE STUDENT

Gaining an understanding of protein and nucleic acid structure is an important part of learning biochemistry. Your textbook has excellent illustrations, but they are limited by being two-dimensional representations of three-dimensional structures. A number of computer programs are available which display molecular structures that you can interact with by moving, rotating, zooming in and out, and displaying in different ways. Using these programs makes it much easier to gain an appreciation of three-dimensional structure.

Biochemists determine the three-dimensional structure of proteins and nucleic acids by X-ray crystallography, by NMR (nuclear magnetic resonance) spectroscopy, or by cryo-electron microscopy. The result is a map of the structure in which it is possible to specify the location of every atom by its x, y, and z coordinates. A list of these coordinates along with other information is deposited as a structure file in the Protein Data Bank (PDB), which is maintained on a server at Rutgers University in New Jersey. Using a variety of programs, both freely available software and commercial programs, anyone can easily download the coordinates of a macromolecule and display it on a personal computer. Biochemists use such programs to explore protein and nucleic acid structure, to plan experiments, and to interpret experimental findings. The use of these programs has become one of the indispensable tools of every biochemist.

One of these programs, **Jmol**, forms the basis of the lessons in this booklet. **Jmol** is used widely by biochemists. It requires the use of typed instructions in addition to menus, but most people find it easier to use than other completely menu-operated programs. It is among the easiest to learn of the programs that are available. You can easily install Jmol on your computer by following the instructions in this manual. Alternatively, you can download Jmol onto a flash drive and use a computer in a campus computer lab to do these exercises.

Other molecular modeling programs are also used widely by research biochemists. But once you learn Jmol you will find it easy to learn any of the other programs.

When students started to use this manual, they found a number of errors – hopefully these have now been corrected. But Jmol is constantly being revised and new versions are appearing, so many of the instructions have had to be modified to keep up with changes in the software. Also the PDB web site and others referenced in the manual are constantly making changes that make the instructions in this manual out-of-date. On top of that, many of the older files have been recently "remediated" in such a way that the commands in the manual have had to be changed to work with them. **So if you come across anything that seems to be incorrect or instructions**

**that are difficult to understand, please let me know!** I'll try to pass the word along to your fellow students this semester and make corrections for the next edition.

Hopefully your experience with these lessons will give you a better appreciation of protein and nucleic acid structure and will equip you with tools that you can use as you pursue a career in science or in fields related to medicine.

Dr. Jeff Cohlberg  
cohlberg@csulb.edu

## HOW TO DOWNLOAD JMOL

To download Jmol, go to <http://jmol.sourceforge.net/>. In the upper right corner click on **Download**. Under **Quick links** click on **Download Jmol\_JSmol**. On the next page click on **Download Latest Version**. The current version of Jmol will be downloaded to your computer.

[Note: Most of the web sites cited in this manual are also available from the California State University, Long Beach, Biochemistry Links page. Go to [www.csulb.edu/~cohlberg/biochemlinks.html](http://www.csulb.edu/~cohlberg/biochemlinks.html). Click on **Molecular Modeling**. There is a section on **Free Software Downloads**, one on **Structure Browsers** that includes a link to the PDB site, and sections with links to software tutorials, biochemistry tutorials and other sites related to molecular modeling.]

Locate the downloaded file on your computer, in your Downloads folder or wherever downloaded files are usually found. The file is compressed in a .zip format and must be expanded. If the file doesn't expand automatically, try double-clicking on the icon of the compressed version. If you have the appropriate software to expand the file, a new folder entitled Jmol-16.1.9 (or a later version) will appear.

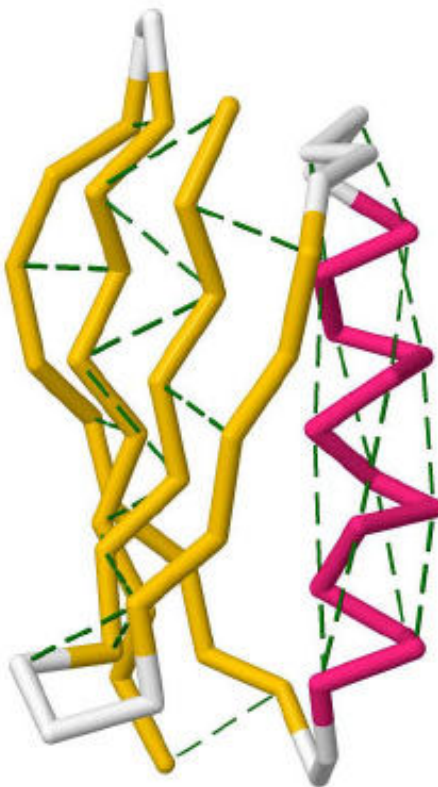
You must have Java installed on your computer in order to run Jmol. If your computer does not already have Java installed, you must download it from [www.java.com](http://www.java.com).

Since Jmol is a java applet, it doesn't have to be installed. To run the program, open the Jmol folder and double-click on the executable java applet. On a Mac this file will be named Jmol.jar. On a PC there will be several files named Jmol, and you can determine which one is the executable jar file by moving your mouse over each icon. (Or choose Details from the View menu to see a list of files with the file type identified – double-click on the file whose type is Executable Jar File.)

On a Mac running a recent version of OS X, you may get a message that Jmol cannot be run since it is an application downloaded from an unknown source. It will then be necessary to launch System Preferences and adjust your Security preferences to allow Jmol to be run. You may get similar security messages with Windows.

You can also create a shortcut (alias on a Mac) to launch Jmol.jar from your desktop. Also, the Jmol folder is small enough to copy it onto a flash drive --so you can run it from the flash drive on any computer.

## EXERCISE 1: SOME BASICS OF SECONDARY AND TERTIARY STRUCTURE



In this lesson you will learn the following Jmol skills:

- How to download a protein structure file from the Protein Data Bank (PDB) web site
- How to read the information on the PDB page for a protein structure
- How to display an interactive image of the protein in Jmol
- How to manipulate images of a protein with your mouse
- How to use the Jmol menus to change the way the protein is displayed
- How to use commands in the Script Console to change the way the protein is displayed
- How to change the display style of selected parts of a protein
- How to display hydrogen bonds

You will use these skills to:

- Examine a short  $\alpha$ -helical peptide and a small protein with both an  $\alpha$ -helix and a  $\beta$ -pleated sheet
- Gain a better understanding of different types of protein secondary structure
- Gain a better understanding of the principles of the tertiary structure of globular proteins

LAUNCHING Jmol: Download Jmol to your computer as described in "Downloading Jmol." Then launch your web browser (e.g. Internet Explorer, Firefox, Netscape, Safari, etc.).

Go to the web page of the Protein Data Bank at **www.rcsb.org**. In the search box at the top of the page, type **1AL1**. Click on the search icon (magnifying glass) or hit **Return**. You will get a page with information about the molecule; this is the **Structure Summary** tab. The Structure Summary page has information about who deposited the data and the date the data was deposited, the method of structure determination (X-ray diffraction in this case) and the resolution (2.7 Å). The R-value is a measure of how well the 3-D model fits the data. The **wwPDB Validation** has information of interest to crystallographers about the quality of the structure – including how well it fits the X-ray crystallography data, how well the bond lengths and angles conform to standard values and how well it avoids steric clashes.

Next comes the citation and abstract of the journal article in which the structure was first published.

There is a description of the molecule in the **Macromolecules** box; in this case the molecule is a 12-amino acid residue peptide: **ELLKKLLEELKG**. (The **Sequence Length** is shown as 13, not 12, and the sequence shows residue 1 as X. It turns out that X is an acetyl (CH<sub>3</sub>CO-) group. This is indicated by the green circle under X in the **Modified Monomer** row.)

The **Small Molecules** box tells us about non-amino acid compounds or functional groups present in the structure. The **Ligands** box tells us that in addition to the peptide there is a sulfate ion.

At the top left is a cartoon image of the molecule, consisting of a helix colored blue at one end (the N-terminus) going to red at the other end (the C-terminus). Shown in ball-and-stick are the N-terminal residue (N-acetylglutamate) and the sulfate ion.

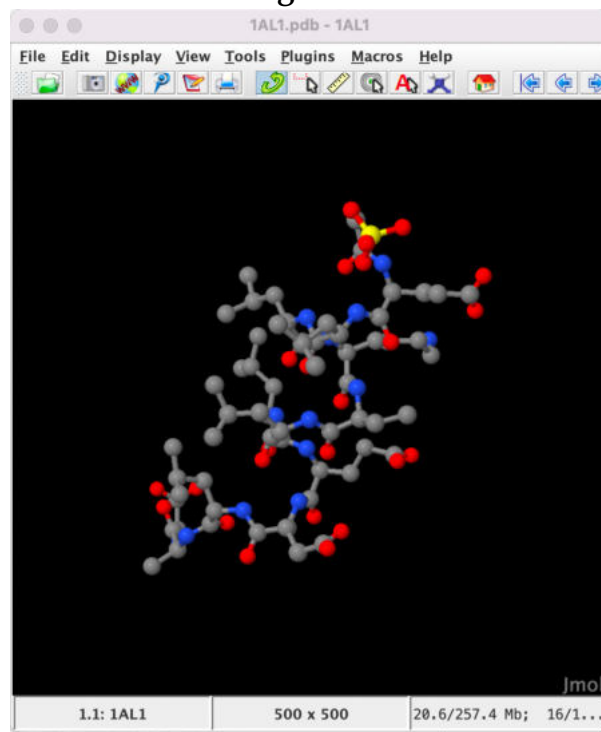
Before you go further, you must download Jmol by following the instructions in "Downloading Jmol."

Launch Jmol by double-clicking on the file **jmol.jar**.

Choose the command **Get PDB** from the **File** menu and type **1AL1** in the **Input** window.

In order to proceed further you must have the program Jmol running with a ball-and-stick image of the molecule in a Jmol window.

Figure 1



If the image of the molecule appears in a browser window instead of a Jmol window, you will not be able to perform the exercises that follow. Go back and follow the instructions again.

DOWNLOADING STRUCTURE FILES: Another way to open a structure file in Jmol is to download the file to your computer and then open it in Jmol. There are some situations that we will encounter later where it's necessary to do this.

Go back to the PDB web page for 1AL1. At the upper right, click on the **Download Files** and then on **PDB Format (gz)** in the drop-down menu. Save the file **1al1.pdb** to your computer, and remember where you saved it.

There are two ways to open the downloaded file in Jmol:

(1) (works with Windows but currently not with Mac) From the **File** menu choose **Open**, locate the file 1AL1.pdb.gz and select it.

(2) Locate the icon of the downloaded file and drag it into the Jmol window.

In either case you should see the ball-and-stick image shown in Figure 1.

(For an alternative way of viewing a molecule, see "JSmol, the Web Browser-Based Version of Jmol" in the reference section at the end of this manual.)

Before leaving the PDB page, click on each of the tabs across the top of the page to see what's there. Note that the **3D View** tab enables you to view the 3D structure in the web page with a program called Mol\*.

### ALPHA HELICAL PEPTIDE 1AL1:

GETTING FAMILIAR WITH THE JMWOL WINDOW:

Practice manipulating the molecule with your mouse:

-- You can rotate the molecule in the X and Y directions by left-clicking on the molecule and dragging it. To rotate in the Z direction (in the plane of the screen), hold down the shift key and move the mouse to the left or right.

-- You can move (translate) the molecule by holding down the shift key, left-double-clicking on the molecule, holding the mouse button down on the second click, and dragging the molecule.

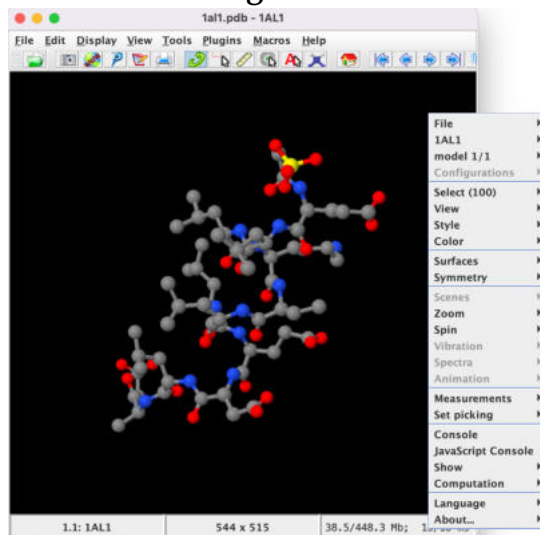
-- To zoom in or out, hold down the shift key, left-click on the molecule, and move the mouse up to zoom out or down to zoom in. If your mouse has a scroll wheel, you can use it to zoom in and out on the molecule.

-- See [http://wiki.jmol.org/index.php/Mouse\\_Manual](http://wiki.jmol.org/index.php/Mouse_Manual) for further instructions.

Initially the image of the molecule will be a ball-and-stick representation with atoms shown as spheres colored using the Corey-Pauling-Koltun scheme (carbon colored gray, oxygen colored red, nitrogen colored blue, hydrogen colored white, sulfur or phosphorus colored yellow). Hydrogen atoms are generally not detected in X-ray crystallography and are not present in PDB structures.

We can display the molecule in many different ways. One way to do this is by using the Jmol pop-up menu. On a PC, right-click anywhere in the window (on a Mac click while holding down the Control key) to bring up the pop-up menu.

**Figure 2**



- Choosing **Style → Scheme → CPK Spacefill** displays each of the atoms as a sphere with a radius equal to the van der Waals radius of the atom and colored using the Corey-Pauling-Koltun scheme (carbon colored gray, oxygen colored red, nitrogen colored blue, hydrogen colored white, sulfur or phosphorus colored yellow). Hydrogen atoms are generally not detected in X-ray crystallography and are not present in PDB structures.
- **Style → Scheme → Ball and Stick** restores the original view. Atoms are displayed as spheres with a radius that is 20% of the van der Waals radius; the bonds are displayed as lines 0.15 Å in diameter. You can control these parameters: Choose **Style → Bonds → 0.25 Å** to make the bonds thicker or **Style → Atoms → 50% van der Waals** to make the atoms larger. **Style → Atoms → Off** gives a representation with bonds only. A second way to control these parameters is through the **Display** menu at the top of the Jmol window.
- **Style → Scheme → Stick** displays the bonds but not the atoms.
- **Style → Scheme → Cartoon** shows the alpha helix as a twisted ribbon colored magenta. The sulfate ion is not displayed in this display scheme. **Style → Scheme → Wireframe** is similar, but the bonds are thinner.
- **Style → Scheme → Trace** shows a trace of the N and C atoms of the polypeptide backbone.

What else can we do with the pop-up menu? Let's look at a few more ways to manipulate the molecule. First restore the Ball-and-Sticks display.

- Choose **Spin → On**, then click on the molecule to cause the molecule to rotate and **Spin → Off** to stop it. Now try **Surfaces → Dot Surface**. Next **Surfaces → Van der Waals Surface**. Both of these show the outer surface of the protein obtained when all the atoms are solid spheres with radii equal to the van der Waals radii. Next try **Surfaces → Make Opaque**. To return to the original view, choose **Surfaces → Off**.

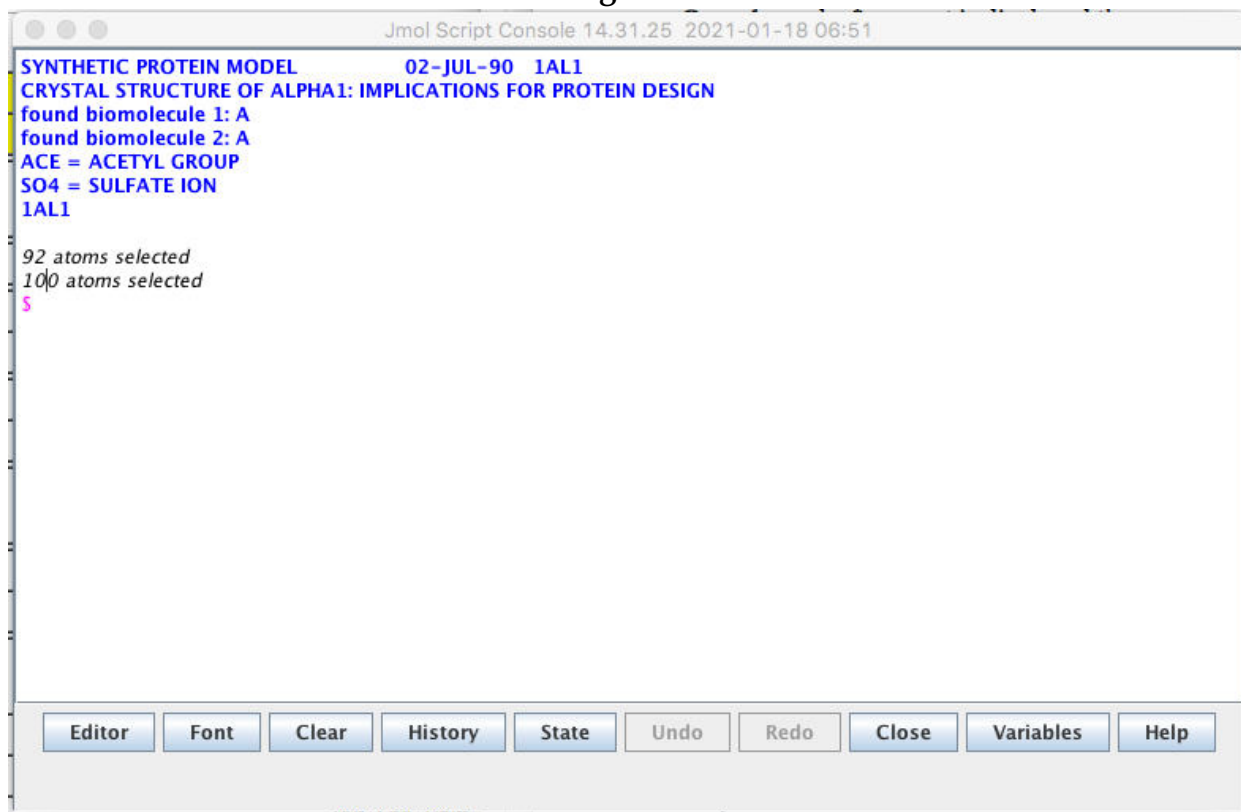
USING THE SCRIPT CONSOLE:

Another way to manipulate the molecule is through the Jmol Script Console. You may already see a second window open on your desktop titled Jmol Script Console. If not, from the **File** menu choose **Console...** A new window opens called **Jmol Script Console**, and a \$ prompt is displayed there.



(See the reference section on JSmol to learn how to open the Console when using the web browser-based version of Jmol.)

**Figure 3**



The blue lines in the Script Console show the PDB ID and filename. Note also **ACE = ACETYL GROUP**. This is the covalent modification that was designated **X** on the Structure Summary page.

We can type commands into this window using a simple language that is easy to learn. Type **show sequence**. The printout shows that the acetyl group occupies position 0 in the sequence (a covalent attachment to the N-terminal amino group of the peptide).

The Script Console can also be used to manipulate the display of the molecule. Throughout this manual we will do most of our manipulations using the Script Console. At first it may seem as if there is a lot to learn, but most Jmol users find that for most manipulations of the molecules, it is easier and faster to use Script commands than to use either the main Jmol menu or the popup menus. As you go through these exercises, you will learn the commands slowly and gradually become competent at using them.

We can put the entire molecule in spacefill representation by typing the word **spacefill**. To return to the original display, type **undo**. (Often you can do this by clicking on the **Undo** button of the Script Console.) You can also type **spacefill off**.

Let's start over with the molecule in **Style →Scheme → Ball and Stick**. Now type **ribbon**. To see just the ribbon and not the balls and sticks, type **ribbon only**. For a narrower ribbon type **ribbon 200**.

Now type **ribbon off** and then **backbone** (or just type **backbone only**). The backbone trace shows straight lines connecting the alpha carbon atoms. Type **backbone 100**. Rotate the molecule so that you are looking down the helix axis, which is a line running lengthwise down the center of the helix. You should see that there are between three and four amino acid residues per turn of the helix.

Now use the pop-up menu to return the molecule to **Style →Scheme → Ball and Stick**. If you are still looking down the helix axis, you can now see the side chains projecting outward from the helical backbone.

We can also use the Script Console to control the colors of the atoms and bonds.

Type **select hydrophobic** (hit return after each line)

**color green**

**select positive**

**color yellow**

**select negative**

**color magenta**

You will now be able to see the relative locations of the different types of amino acid residues in the molecule. (Note that you can also change atom colors by using the **Select** and **Color** commands in the pop-up menu -- but using the script console is a little easier.)

Let's focus just on the chemical nature of the sidechains. Type **select backbone** and then **color cpk**. (You can also type the two commands on a single line: **select backbone;color cpk**. Now only the sidechains are coded by chemical type. It might be interesting to see what the molecule looks like in spacefill display. Type **spacefill**. Oops! What happened? Since only the backbone was selected, only the backbone atoms were changed into sticks.

Choose **Select →All** from the Jmol **Display** menu. (You can also type "select all" on the command line.) Now type **spacefill** again. That's much better.

(Note: Instead of typing **spacefill** in the script console, you could have chosen **Style→Scheme→CPK Spacefill** in the pop-up menu, but then you would have lost the color coding and displayed all the atoms in CPK colors. Choosing any of the style

schemes automatically also chooses an associated color scheme. For this reason it is probably a good idea to use **Style** in the pop-up menu only for the very initial stages of displaying a molecule and to use script commands for subsequent changes of display style.)

Type **show info**. The script console displays the name of the molecule, along with information about its sequence, chain nomenclature, size, and secondary structure.

Double click on an atom. A message appears in the Jmol script console which gives the type of amino acid residue followed by the residue number (e.g. [LEU]6), the polypeptide chain (:A – in this case there is only one chain, the A chain), the type of atom (N = NH nitrogen, O = CO oxygen, C = CO carbon, CA = alpha carbon, CB = beta carbon, etc.), the atom number (each atom in the molecule is assigned a unique number), and the x, y, and z coordinates of the atom that locate the atom in three-dimensional space in the model.

Now type **spacefill off**, then **wireframe 100**. Click on the various atoms and answer the following:

Question 1: Which amino acid is in position 8?

Now type **show sequence**. See whether your answer was correct.

There's one complication that needs to be pointed out. You should see three yellow amino acid residues, all of them lysine residues. However, only one of them, lysine 4, has a complete side chain. According to the crystallographers who determined this structure, for some reason the side chains of lysines 5 and 11 could not be clearly located in the electron density, so only partial side chains were included in the structure.

Question 2: What do you observe about the distribution of the various types of side chains? (Examine the molecule in the wireframe 100 display. Then go to the spacefill display by typing **spacefill**. To return, either type **spacefill off** or click on the **Undo** button. (If your Script Console doesn't have an Undo button, try expanding the console by dragging downward on the lower right corner.) In both display modes, examine the molecule in both side and end views.) How would this helix be oriented if it were on the surface of a globular protein?

## IMMUNOGLOBULIN BINDING PROTEIN (1PGB):

### OVERALL STRUCTURAL ORGANIZATION:

Go back to the PDB web site and obtain the information page for the file 1PGB. This is a 56-amino acid residue portion of a larger protein, namely the B1 immunoglobulin-binding domain of streptococcal protein G. We will use this small structure for our first exploration of tertiary structure.

Open the structure 1PGB in Jmol. You should see a Ball-and-Stick image of a protein molecule. There are also small red spheres that seem to be floating unconnected to the protein. They represent the oxygen atoms of water molecules that are part of the crystal structure. To make them disappear, type **delete water**. The water molecules are removed from the working PDB file and cannot be restored without reopening the original file in a new Jmol session.

Alternatively you can type **restrict not water**. (A "restrict" command causes everything not selected to disappear from the display, but the selected atoms can be restored later.)

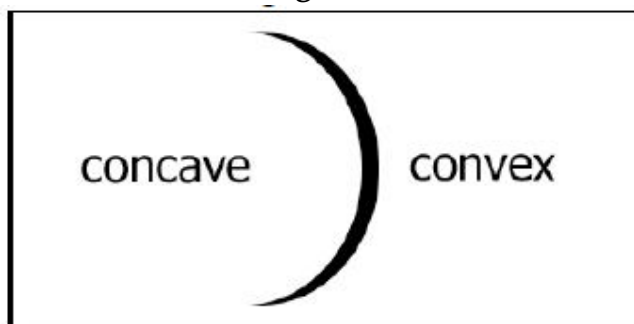
The instructions for the exercises in this manual will assume that you have removed the water molecules with a **delete** command.

Choose **Style → Scheme → Cartoon**. You will see a Cartoon view of a protein colored according to structure. In this color scheme beta sheets are yellow, alpha helices are magenta, and regions with none of these secondary structures -- sometimes called "coil"-- are white.

Type the command **show structure** and see what information is displayed.

Rotate the molecule to get a sense of the relative orientations of the helix and beta sheet. Orient the molecule so that the helix axis is pointing straight at you. Notice that the beta sheet is curved so that it has a convex surface and a concave surface.

**Figure 4**



Question 3: Describe the structural organization of this protein. Use the word “convex” or “concave” in your description.

Put the molecule into backbone display, in which the polypeptide is represented as a series of lines connecting the alpha-carbon atoms. To do this type **backbone 100** and then **cartoon off** (or **backbone only; backbone 100**). Click on the two ends of the protein to identify the N- and C-termini. Another way to do this is by the command **color group**. In this scheme the colors change from cooler to warmer (blue to red) as we go from the N-terminus to the C-terminus. Restore the molecule to **color structure** before proceeding.

Now start at the N-terminus of the molecule and trace the course of the polypeptide backbone, rotating the molecule as necessary to see where you are going.

Question 4: Describe the order of the various secondary structure elements as they occur in the primary sequence, starting as follows: “Residues 1 to 8 are in a beta sheet conformation. Residues 9 and 10...”

#### A CLOSER LOOK AT THE BETA SHEET:

Question 5: Number the four regions of beta sheet 1 through 4 to indicate their position within the amino acid sequence. Now start at one edge of the beta sheet and look at the order of the strands as you go to the other edge of the sheet. Indicate the order in which the four beta strands occur. (Hint: Is it 1-2-3-4? 1-3-2-4? Something else?)

Question 6: Beta sheets can be either parallel (with strands running from their N-termini to their C-termini in the same direction) or antiparallel (with strands running in opposite directions). Examine each pair of adjacent strands in the beta sheet (there will be three pairs). For each pair of strands, indicate whether they are parallel or antiparallel.

To see the hydrogen bonds (H-bonds), type **hbonds calculate**. This command directs the program to identify each pair of NH hydrogens and CO oxygens of the backbone (for the currently selected atoms) that are close enough to allow H-bond formation. The H-bonds involving amino acid residues from the sheet will appear as dash lines. But since the backbone display shows only the alpha carbon atoms of the backbone, the H-bonds connecting amide nitrogens to carbonyl oxygens seem to be floating in space. To fix this type **set hbonds backbone**. You can make the H-bonds thicker by typing **hbonds 20**. If you want to make them a different color, type **color hbonds green**. They're a little hard to see against the black background, so type **background white**. (If you're in a playful mood, try **background magenta**.) Notice how the H-bonds run

perpendicular to the strands of the beta sheet and lie in the plane of the sheet, while the H-bonds in the helix are parallel to the helix axis.

Question 7: Print an image of the molecule. (See "Printing and Saving an Image" in the Appendix for instructions on printing.)

Let's display the other atoms as well. Type **select all; wireframe only; wireframe 100**. Now type **set hbonds sidechain** to make the H-bonds connect the amide N and carbonyl O atoms. (Note that in this command language, "sidechain" does not refer to an amino acid R-group.) Typing **color hbonds cpk** will color each H-bond blue at the amide end and red at the carbonyl end. To see how the H-bonds connect the N and O atoms more clearly, we'll look first at the beta sheet and then at the helix. Type **select sheet; color cpk**, then **restrict sheet**. Finally type **select not sheet; hbonds off**. Zoom in on the sheet. You should now be able to see how each H-bond connects an NH group to a CO group in an adjacent strand. (The hydrogens are not displayed, so the NH hydrogens are missing.) Note how within a given strand the red oxygen atoms alternate up and down, as do the blue nitrogens.

#### A CLOSER LOOK AT THE ALPHA HELIX:

Now let's restore the helix and examine it. Type **select all; wireframe 100; color cpk**, then **restrict helix**. Now type **select helix; hbonds 20; select not helix; hbonds off**.

Question 8: Choose two H-bonds. Click on the atoms at each end to identify the amino acid residue numbers. List the pairs of residues for two H-bonds. In an alpha helix, how far apart is each pair of amino acid residues?

(Is amino acid residue X hydrogen-bonded to residue X+3? X+4? X+5?)

Type **select all; hbonds 20; wireframe off**. Only the H-bonds remain. Can you tell where the helix is and where the sheet is?

Restore the cartoon view (**select all; cartoon; color structure**) and, just for fun, type **rotate**. To stop the rotation type **rotate off**. (You can also use the commands **spin** and **spin off**.)

**\*\*An Important Note About Changing Display Styles\*\***

When you switch from wireframe to spacefill display, you can just type **spacefill**. However, when you switch from spacefill to wireframe display, you have to type **wireframe; spacefill off** or **wireframe only**; otherwise the spacefill will block your view of the wireframe inside.

If you're in wireframe and you type **cartoon**, you will see the cartoon image superimposed on the wireframe image. If you just want to see the cartoon image, type **cartoon; wireframe off** or **cartoon only**.

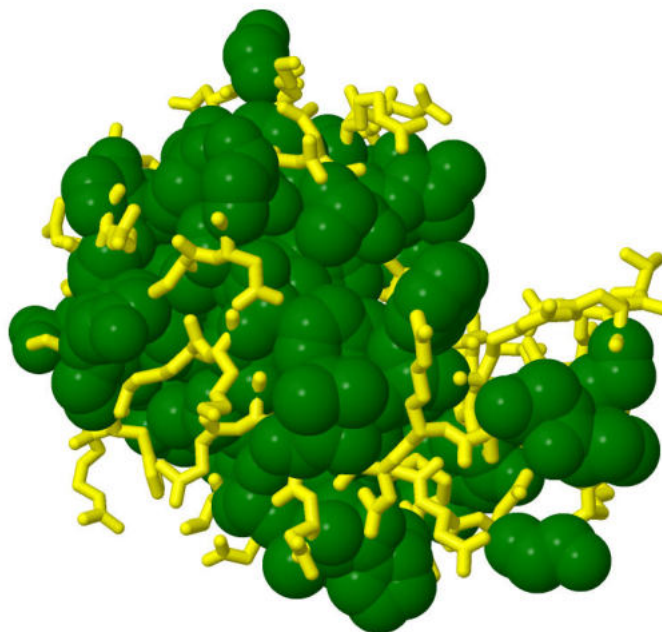
## FURTHER EXPLORATIONS

**FE 1:** Display the structure 1GZI. Which protein is this? Trace the polypeptide chain and describe the sequence of secondary structure elements within the protein. Analyze the beta sheet in terms of parallel and antiparallel orientation. Look at the region from residues 41 to 56. Describe the secondary structure of this region.

**FE 2:** Display the structure 2KT2. Which protein is this? Trace the polypeptide chain and describe the sequence of secondary structure elements within the protein. Analyze the beta sheet in terms of parallel and antiparallel orientation.

**FE 3:** On the basis of your observations in exercises FE1 and FE2, describe the role of beta sheets in bringing different parts of the polypeptide chain together. Cite some specific examples.

## EXERCISE 2: LYSOZYME, A GLOBULAR PROTEIN WITH DISULFIDE BONDS



Jmol

In this lesson you will learn the following Jmol skills:

- How to display disulfide bonds
- How to use slab and surface views in displaying protein structures

You will use your Jmol skills to:

- Explore the structure of the enzyme lysozyme.
- Study lysozyme as an example of a protein containing disulfide bonds.
- Look in detail at alpha helices and beta sheets that are part of an intact globular protein.
- Learn the difference between an alpha helix and a  $3_{10}$  helix.
- Examine the distribution of polar and nonpolar amino acids within globular proteins, using lysozyme as an example.

Lysozyme is an enzyme that cleaves the polysaccharides present in the peptidoglycan of bacterial cell walls. The structure of lysozyme, first published in 1965, is of great historical interest – it was the first structure of a protein containing a beta-pleated sheet and also the first structure of an enzyme, which gave biochemists their first direct glimpse of an active site. Like many other extracellular proteins, lysozyme's structure is stabilized by disulfide bonds.

Prerequisite: You must have completed Exercise 1 before attempting this exercise.



### LYSOZYME: ITS ACTIVE SITE CLEFT AND DISULFIDE BONDS

Go to the Structure Summary page for 2LYZ. In the **Macromolecules** section, under **Protein Feature View**, click on **Expand** in the upper right corner (if you don't see it, try making your window wider) and look at the information displayed there.

Display the structure in Jmol. Type **show info** and see what information is displayed.

Type **delete water** to get rid of the water molecules. Type **cartoon only; color structure**. Notice the structure color scheme. The beta sheet is colored yellow and the alpha helices are magenta. There is also a short helix colored a shade of purple. This is a  $3_{10}$  helix, which we will learn more about below.

Set the **Style Scheme** to **CPK Spacefill**. Zoom in until the molecule fills the screen. Lysozyme is an enzyme that cleaves bacterial cell wall polysaccharides. If you rotate the molecule to the proper orientation, you can see the cleft where the polysaccharide binds -- this is the active site of the enzyme. The active site cleft of lysozyme can bind a hexasaccharide, and the inhibitor tri-NAG (a trisaccharide of N-acetylglucosamine residues) binds to the first three of the six monosaccharide binding pockets.

Now reset the **Style Scheme** to **wireframe**. Make sure that the background is black. You should be able to see the disulfide bonds -- yellow lines connecting yellow spheres (yellow is the CPK color for sulfur). To make the SS bonds thicker type **ssbonds 60**. To see things more clearly display the molecule in the backbone mode (**backbone only**). To make the backbone thicker, type **backbone 80**. Typing **backbone only** made the disulfide bonds disappear; restore them by typing **ssbonds on; ssbonds 60**.

The disulfide bonds seem to be floating in space. To correct this, type **set ssbonds backbone**. (This command draws the SS bonds connecting the alpha carbons of the amino acid residues instead of the sulfur atoms, which are not visible in the Backbone display. The alternative is **set ssbonds sidechain**.) If the disulfide bonds are gray, you can make them yellow by typing **color ssbonds yellow**.

Question 1: How many disulfide bonds are present in lysozyme? Double click on both ends of each disulfide bond. List each pair of cysteines involved in disulfide formation. (Hint: One bond links cysteines 6 and 127)

Put the molecule in cartoon display (**cartoon only**) and color by secondary structure (**color structure**). In addition to the magenta alpha helices and yellow beta sheets, you will now see purple  $3_{10}$  helices. A  $3_{10}$  helix is slightly different from an alpha helix; we will explore the difference later in this exercise.

Question 2: How many of each type of helix are there in lysozyme? How many sheets? Describe the location of the sheet relative to the active cleft.

To see the cleft more clearly, use the Spacefill display mode (**spacefill only**). See if this view affects your answer to Question 2.

#### HYDROPHOBIC INTERACTIONS IN LYSOZYME

Go back to the **CPK** color scheme (N blue, O red, C gray, S yellow) by typing **color cpk**. Type **select hydrophobic; color green; select polar; color yellow**. Notice the relative location of the polar and nonpolar amino acid residues. To make this even more obvious, type **wireframe only; wireframe 100**. Since only the polar amino acids are selected, they will change to wireframe display, while the other amino acids will remain in the Spacefill representation.

Now change the image so that the hydrophobic amino acid residues are Wireframe and the polar amino acids are Spacefill. See if you can figure out how to do this without looking ahead.

Having trouble? OK, here goes: As long as the polar amino acid residues are still selected, **spacefill** will return them to Spacefill representation. Then type **select hydrophobic; wireframe only; wireframe 100**. You should see a marked difference between the relative locations of these two types of amino acid residues. Try going back and forth between the two views, using the **Undo** and **Redo** buttons (clicking each button twice in a row), to see the difference.

Finally, restore the spacefill display (**select protein; spacefill only**). Another way to visualize this is to use the **slab** command. Type **slab on; slab 50**. This allows you to see a slice through the molecule 50% of the way through it. (Note: Typing **slab 80** will change the display so that the slice is 80% of the way (from back to front) through the molecule rather than the default value of 50%. You can also gradually change the depth of the slab -- hold down the Control and Shift keys and move the mouse vertically.) Rotate the molecule, and notice which type of amino acid residue is internal and which type is external.

When you are finished, type **slab off**.

Question 3: Print an image of lysozyme with the slab function turned on. Describe the location of the polar and nonpolar amino acid residues in lysozyme.

Type **select positive; color red; select negative; color blue**. Notice how the positive and negative amino acid residues on the surface of the protein are interspersed so that they tend to neutralize each other.

How would you select the uncharged polar amino acid residues? Type **select polar and not charged; color magenta** and see what happens. (Note: For a list of available colors, go to <http://jmol.sourceforge.net/jscolors/> and click on **JavaScript colors**. They include things like "hotpink" and "dodgerblue"!)

### BETA SHEETS

Type **select protein; wireframe only; color cpk; set ssbonds sidechain**. Now type **select sheet; cartoon; wireframe 50**. You will see a three-stranded  $\beta$ -sheet of lysozyme lining the surface of the active cleft, along with a short two-stranded  $\beta$  sheet on the surface of the protein away from the cleft. Rotate the molecule so that you can clearly see the three strands of polypeptide chain in the principal  $\beta$  sheet. To make this easier, look at the  $\beta$  sheet by itself by typing **restrict sheet**.

To see the hydrogen bonds, type **hbonds calculate** and **background white**. To make the bonds thicker, type **hbonds 40**. Type **cartoon off**. Then try **color hbonds orange**. Zoom in on the  $\beta$  sheet and center it on your screen (remember: shift double-click to translate). Type **restrict backbone**. You can now turn the sidechains on and off by clicking on the **Undo** and **Redo** buttons.

Question 4: How many H-bonds stabilize the main  $\beta$  sheet in lysozyme? List them. (Hint: Click on each NH and CO group to identify them. For example, one of the H-bonds is NH52-CO44.)

Question 5: Describe the orientation of the side chains with respect to the plane of the  $\beta$  sheet.

### ALPHA HELICES

Type **select 5-14; cartoon; wireframe 100**. You will see an image of an  $\alpha$ -helix sitting on the surface of the protein. Turn on the H-bonds as you did for the  $\beta$  sheet, but choose a different color for them. Turn the cartoon display off. Notice how the H-bonds connect NH and CO groups. Click on any pair of NH and CO groups bonded to each other.

Question 6: List any two pairs of NH and CO's joined by an H-bond. How many amino acid residues apart are the two residues in each pair?

Rotate the molecule so that you are looking down the axis of the helix. Select the hydrophobic residues and color them green; color the polar residues your favorite color.

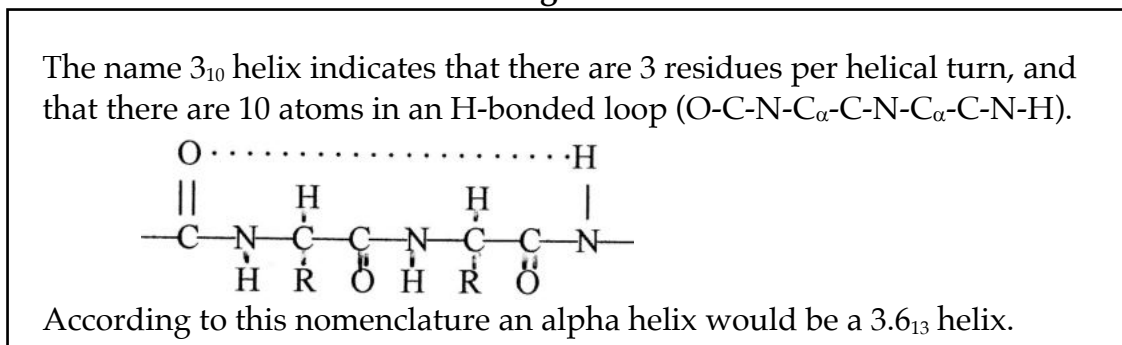
Question 7: Describe the distribution of the hydrophobic and polar sidechains of the helix. Why should they be arranged in this fashion?

### 3<sub>10</sub> HELICES

Now let's look at the 3<sub>10</sub> helix. Let's go back to the original image of the molecule by typing **select protein**, setting the Style Scheme to **Cartoon** and typing **hbonds off**. Clicking on the ends of the purple helix shows that it extends from residue 80 to residue 85. Display only this helix (**restrict 80-85**). Calculate and display the H-bonds as you did earlier. Click on each pair of NH and CO groups joined by an H-bond.

Question 8: How many residues apart are each pair of NH and CO groups joined by an H-bond in a 3<sub>10</sub> helix? Is this a tighter and narrower helix than an alpha helix, or is it a looser and wider helix?

**Figure 5**



### IRREGULAR SECONDARY STRUCTURE

Restore the entire molecule and put it in backbone display (**select protein; backbone only; backbone 40; hbonds off**). Now **select not sheet and not helix**. Calculate and display the H-bonds, and select the backbone option (**set hbonds backbone**).

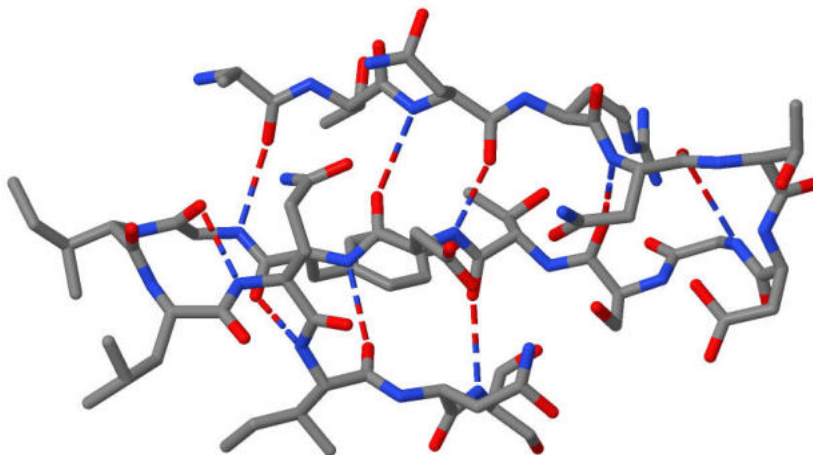
Question 9: Are hydrogen bonds restricted to the helices and sheets? Describe the role of H-bonds in stabilizing the tertiary structure of lysozyme, providing some specific examples.

Type **select all; dots; background black**. This will generate a van der Waals dot surface around the molecule. Rotating the molecule will enable you to see the active site cleft clearly. To get rid of the dots, type **dots off**.

## FURTHER EXPLORATIONS

**FE 1:** Display the structure 3I3Z, human insulin. Explore the molecule. Note that there are two polypeptide chains, which you can see separately by typing **color chain**. (Multi-chain proteins will be explored in more detail in Exercises 6 and 10.) Visualize the disulfide bonds and list them, distinguishing the A and B chains. One or more disulfide bonds connects cysteine residues in the two different chains—which one(s)? Look at the hydrogen bonds both within and outside the helices. Describe the role of H-bonds in stabilizing the secondary and tertiary structure.

## EXERCISE 3: BETA TURNS AND PI HELICES



In this lesson you will learn the following Jmol skills:

- How to display beta turns and other types of secondary structure in proteins by doing a secondary structure calculation with the program DSSP

You will use your Jmol skills to:

- Explore the structure of beta turns in lysozyme
- Look at examples of proteins that contain pi helices

In addition to alpha helices and beta sheets, the third principal type of secondary structure found in globular proteins is the beta turn. In this exercise you will look at some of the beta turns in lysozyme, determine their H-bonding patterns, and examine their roles in allowing the polypeptide chain to change directions and in connecting helices and beta strands. You will also examine a structure called a pi helix.

Prerequisite: You must have completed Exercises 1 and 2 before attempting this exercise.

### CALCULATING THE SECONDARY STRUCTURE

Display structure 2LYZ, lysozyme. Delete the water. Set the Style Scheme to Cartoon.

Type **calculate structure**. The protein disappears, and a long message is printed out in the Script Console. Jmol has applied a program called DSSP (Define Secondary Structure of Proteins) to determine the location of various types of secondary structure in a protein, including beta turns, as well as helices and sheets. The final portion of the Script Console printout is a table listing the elements of secondary structure according to the following scheme:

H = alpha helix

E = extended strand that participates in a beta sheet

T = hydrogen bonded turn, including beta turns

Less common are the following:

B = residue in an isolated beta-bridge (similar to a beta sheet, but containing only two H-bonds)

G =  $3_{10}$  helix

I = pi helix (more about this below)

S = bend (a region of high curvature, where the chain changes direction by an angle greater than  $70^\circ$ )

Without doing the DSSP structure calculation, you can only distinguish alpha helices, beta sheets and  $3_{10}$  helices. Doing the calculation allows you to see the other types of secondary structure as well.

### BETA TURNS AND BETA HAIRPINS

Restore the image of the molecule by typing **cartoon; color structure**. Now in addition to magenta or purple helices and yellow sheets there are blue H-bonded turns.

Let's take a closer look at the beta turns within the beta sheet. Type **restrict 42-60; wireframe only; wireframe 50; color cpk**. Now calculate the H-bonds and turn them on. You will see that in addition to the H-bonds within the beta sheet, there is one H-bond within each of the turns.

Question 1: For each of the two turns click on the H-bonded amino acid residues to identify them. How many residues apart are these residues? Is the H-bonding pattern similar to an alpha helix or to a  $3_{10}$  helix?

Question 2: Describe the organization of the beta sheet in terms of beta strands and beta turns.

A structure comprising two antiparallel beta strands H-bonded to each other and separated by a small loop is called a **beta hairpin**. Lysozyme has two beta hairpins, one comprising the strand of residues 42-46 and the strand of residues 50-54, and a second hairpin comprising strand 50-54 and strand 57-60. A beta hairpin is a type of **motif** or **supersecondary structure**. Supersecondary structure will be explored in greater detail in the next exercise.

PI HELIX:

**Pi helices** are much less common than other types of secondary structure and occur only rarely. There are no pi helices in lysozyme, so we'll have to turn to other proteins to see examples of this structure.

Download and display the file 3EE4. This is a Mn/Fe oxidase from the organism that causes tuberculosis. Delete the water, calculate the structure, restore the Cartoon display and color by Structure, as you did above for lysozyme.

Note that the report in the Script Console indicates that residues 163-167 form a pi helix. These residues are colored a deep violet. Let's have a look at this portion of the protein. Type **restrict 159-168**. Display these residues in **wireframe 50** mode and color by the CPK scheme. Calculate and display the H-bonds.

Question 3: Click on each pair of residues connected by an H-bond and identify them. How many residues apart are the residues connected by H-bonds?

(Note: The structure calculation output identified only residues 163-167 as pi helix, but we can see that the pi helix actually comprises residues 159 to 168. This is because the pi helix is the C-terminal portion of a helix that begins as an alpha helix. So, while the CO group of residue 159 is H-bonded to the NH group of residue 164, its NH group is H-bonded to the CO of residue 155. So residue 159 is simultaneously part of an alpha helix and a pi helix. In the DSSP program calculation, the classification is based on the H-bonding pattern of the NH group.)

Question 4: Summarize the H-bonding patterns seen in alpha helices,  $3_{10}$  helices, and pi helices.

#### FURTHER EXPLORATIONS

**FE 1:** Other proteins with pi helices include 3GS6, 3FAY, and 3H9V. These include an isolated pi helix, a pi helix at the end of an alpha helix, and a pi helix within an alpha helix. Explore the three proteins, determine the length of each pi helix and its location, and report your findings.

**FE 2:** One of the longest pi helices occurs in human ferrochetalase (1HRK). Display the structure and characterize the pi helix. Note: This protein contains two identical polypeptide chains. Typing **restrict :A** will allow you to view a single chain. (You will learn more about working with multichain proteins in Exercise 10.)



## Exercise 4: SUPERSECONDARY STRUCTURE



In this exercise you will learn the following skill:

- How to use a database called PDBSum to get information about protein supersecondary structure.

You will use this skill and the skills learned previously to:

- Examine the way elements of secondary structure are connected to each other in globular proteins.
- Examine a number of protein motifs that occur widely in globular proteins.

The term "supersecondary structure" refers to the way that successive alpha helices and beta strands are organized to form larger units called protein motifs. In globular proteins the polypeptide chain consists of a series of straight and relatively rigid alpha helices and beta sheets interspersed with turns or loops. The turns or loops sometimes themselves stabilized by hydrogen bonds, enable the polypeptide chain to change directions. In this lesson you will examine some globular proteins and visualize some of the protein motifs that they contain.

Prerequisites: You must have completed Exercises 1, 2, and 3 before doing this exercise.

## MOTIFS IN CARBOXYPEPTIDASE, BETA-ALPHA-BETA, AND ALPHA-ALPHA UNITS

We noted in exercise 3 that lysozyme has two beta hairpins. This is one type of supersecondary structure -- a level of structure built by combining helices and beta strands with connecting loops. In this exercise we will examine some additional examples of supersecondary structure.

Display the structure of carboxypeptidase (1YME) from the PDB in Cartoon style. Examine the beta sheet.

Question 1: How many strands comprise the beta sheet? Describe the organization of the sheet in terms of parallel and antiparallel strand interfaces.

Calculate and display the H-bonds. Type **set hbonds backbone**.

Question 2: Are there any beta hairpins in carboxypeptidase? If so, which strands are involved?

Beta hairpins are the connections between two antiparallel beta strands. One way of connecting two parallel strands is to join them by a loop containing one or more alpha helices, forming a **beta-alpha-beta unit**. Carboxypeptidase has two beta-alpha-beta units. To see one of the beta-alpha-beta units, type **restrict 201-242**. To see the second type **undo** and then **restrict 61-108**.

Question 3: Describe each of the two beta-alpha-beta units, indicating the residues comprising the strands and helices.

Now **restrict 73-102** to see an example of an  **$\alpha\alpha$  motif**. An  $\alpha\alpha$  motif is sometimes called an  **$\alpha\alpha$  hairpin** if the helices are antiparallel to each other or an  **$\alpha\alpha$  corner** if they are roughly perpendicular.

Sometimes adjacent alpha helices interact with one another through van der Waals interactions resulting from the interdigitation (interlocking like fingers) of their side chains. One helix in carboxypeptidase, residues 286-305, forms such interactions with three other helices, those comprising residues 73-89, 94-102 and 216-231. Display these portions of carboxypeptidase in both cartoon and spacefill display.

To see a listing of the various types of secondary structures and motifs in carboxypeptidase we can use the **PDBSum** database, which provides a summary of various structural features for each structure in the PDB. ([www.ebi.ac.uk/pdbsum/](http://www.ebi.ac.uk/pdbsum/)). Enter the **PDB code** 1YME and click **Find**. Click on the **Protein** tab. The next page displays a secondary structure "wiring diagram" of the protein. Under **ProMotif** in the left menu you will find a list of many of the various secondary structure elements and

motifs. Clicking on each motif individually brings up a page with more details about that motif, including a link to a **Motif Description**, which provides a formal definition of each type of motif.

Click on **1 beta hairpin** and see whether you correctly identified the beta hairpin in your answer to Question 2.

### BETA BARREL

Display the structure of bovine retinol-binding protein (1KT7). Delete the water. Display in Cartoon mode and color by group. Trace the chain from the blue N-terminus to the red C-terminus.

Question 4: Describe the organization of the protein in terms of secondary structure elements. How many beta strands are there? Are they parallel or antiparallel? Describe the structure in terms of beta hairpins. If the strands are numbered according to their position in the sequence, which strands form H-bonds with strand #1?

Note the bound retinol molecule in the center of the beta barrel. Render the structure in spacefill display. Then **restrict protein** to make the retinol disappear. You should be able to see that the beta barrel is hollow with empty (water-filled) space in the center; this central cavity is accessible to external ligands.

### ALPHA/BETA BARREL

Display the structure of triose phosphate isomerase (1YPI) as a Cartoon. Restrict your view to a single chain of the dimeric enzyme by typing **restrict :A**. Color by group. Trace the chain as you did for the retinol-binding protein.

Question 5: Describe the organization of the protein in terms of secondary structure elements. How many beta strands are there? Are they parallel or antiparallel? Describe the structure in terms of and bet-alpha-beta units.

Render the molecule in spacefill display and rotate it until you can see the central cavity.

### PARALLEL BETA HELIX

Display the structure of bovine pectate lyase (2PEC). Color by group and trace the polypeptide chain.

Question 6: Why is this structure referred to as a parallel beta helix? How many beta strands comprise each turn of the helix? Are the loops between adjacent beta strands stabilized by H-bonds?

#### HELIX-TURN-HELIX, A DNA-BINDING MOTIF

Display the structure of the DNA-binding domain of the *lac* repressor bound to operator DNA (1LCC). Since this is an NMR structure, the hydrogens are visible, including the hydrogens of water molecules. For a display that looks more like what you are used to, go to the **Display** menu of the Jmol window and uncheck the box **Hydrogens**.

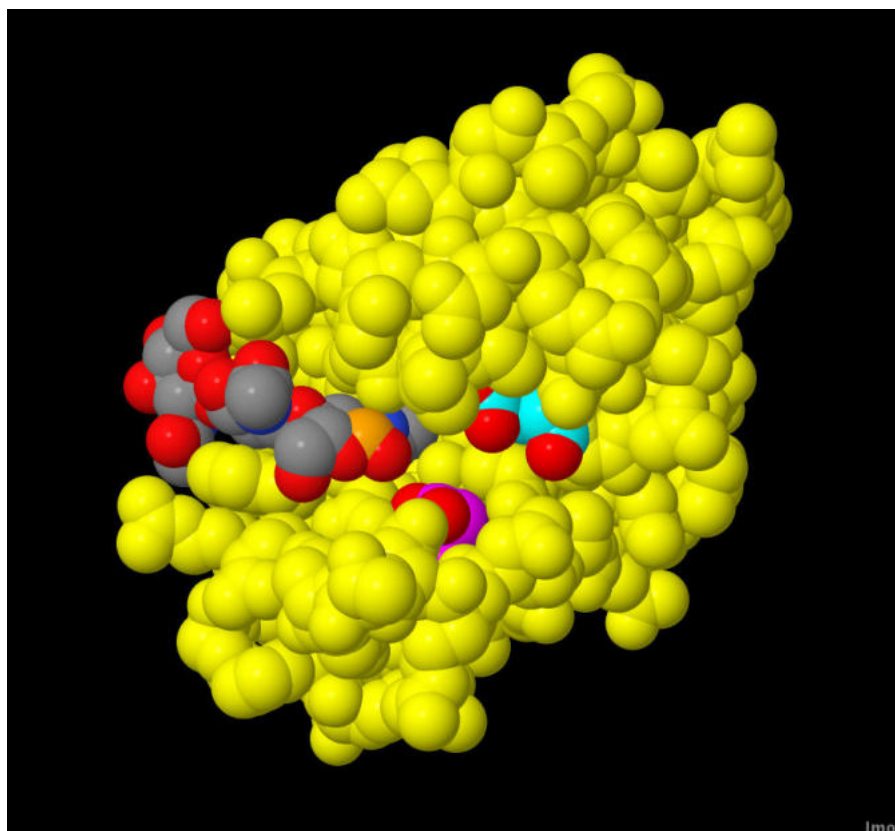
Residues 5-26 form the helix-turn-helix motif. To see only that motif bound to the DNA, type **restrict dna or 5-26**. Examine the structure in different orientations. Then display in wireframe 50 to see the orientation of the sidechains.

Question 7: Which of the two helices interacts with the DNA? Does that helix lie in the major groove of the DNA (the side facing the bases) or the minor groove (the side facing the backbone)? Describe the orientation of the helix axis with respect to the plane of the base pairs, as well as the orientation of the sidechains. Print an image of the structure that illustrates your answer.

#### FURTHER EXPLORATIONS

**FE 1:** Locate the discussions in your biochemistry text of other motifs found in DNA-binding proteins: zinc finger, homeodomain, leucine zipper and basic helix-loop-helix motifs. For each motif look up the PDB file identified in the text, and display the structure in Jmol. Explore the structure of the motif, its orientation with respect to the DNA, and the potential for interactions between the sidechains and the DNA bases. Report your observations.

## EXERCISE 5: LYSOZYME WITH BOUND INHIBITOR WORKING WITH LIGANDS



In this lesson you will learn the following Jmol skills:

- How to get information from the PDB about the ligands bound to a protein and the abbreviations used to select those ligands
- How to work with protein structures that contain bound ligands

You will use your Jmol skills to:

- Examine the active site of lysozyme and see how it binds a trisaccharide that fills part of the site normally occupied by the polysaccharide substrate.
- Examine the orientation of the catalytic side chains of lysozyme in relation to the mechanism of the enzyme.

The enzyme lysozyme catalyzes the hydrolysis of bonds linking sugars in the carbohydrate chains of the peptidoglycan of bacterial cell walls. Having a three-dimensional structure for lysozyme in 1965 enabled biochemists to see for the first time the spatial relationships between an enzyme and its substrate, including the orientation of functional groups at the active site that are involved in substrate binding and catalysis. This allowed David Phillips, then a researcher at the Royal Institution in London, to identify the amino acid residues at the active site of lysozyme that are

important for catalyzing the hydrolysis reaction and to propose a catalytic mechanism based on both structural and chemical evidence. More than 35 years later, new chemical evidence led to a revision of the "Phillips mechanism" in which the enzyme forms a covalent complex with its substrate during catalysis. This history is discussed in many biochemistry textbooks.

Prerequisite: You must have completed Exercise 1 before attempting this exercise. Exercise 2 is recommended.

#### EXERCISE 5A: THE INHIBITOR BINDING SITE OF LYSOZYME

In the PDB, search for 1HEW -- lysozyme with the inhibitor tri-N-acetylchitotriose bound to the active site. The search results page contains three files, all of which have **1HEW** in the filename. One of these is the actual file 1HEW.

Click on **1HEW** to get the Structure Summary page. Scroll down to the **Oligosaccharides** section. You can see that in addition to protein the structure contains the trisaccharide triacetyl-beta-chitotriose. The information in the box tells us that it contains three N-acetylglucosamine (2-acetamido-2-deoxy-beta-D-glucopyranose) connected in  $\beta(1\rightarrow4)$  glycosidic linkages.

Open the file in Jmol. Look at the heading in the console. Note the line **NAG = N-ACETAMIDO-2-DEOXY-BETA-D-GLUCOPYRANOSE** (another name for N-acetyl-D-glucosamine), which tells us that we can select the trisaccharide using the abbreviation **NAG**.

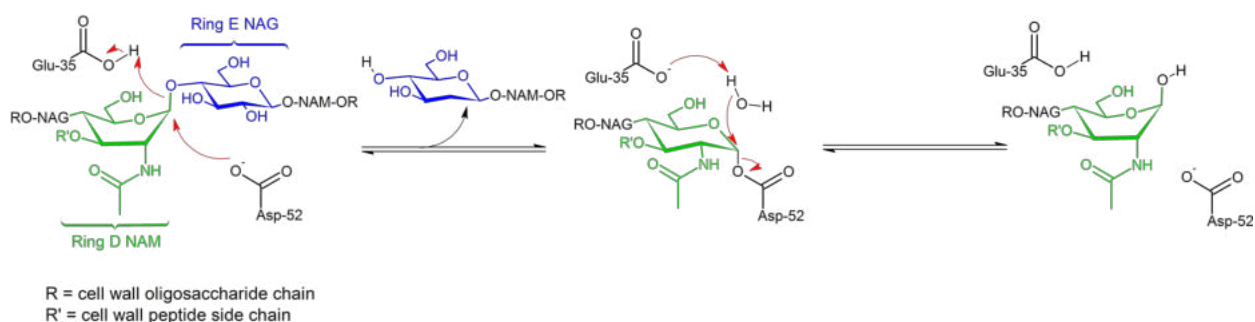
Get rid of the water (**delete water**). You can see the lysozyme molecule with the inhibitor bound to the active site cleft. The active site cleft of lysozyme can bind a hexasaccharide, and the inhibitor tri-NAG binds to the first three of the six monosaccharide binding pockets (designated subsites A, B and C). In the default view, the protein is in Cartoon display and the inhibitor is in ball-and-stick display.

Put both the protein and inhibitor (**select protein or nag**) in **spacefill** display. You should now see the inhibitor (in CPK colors) bound tightly to half of the active site cleft, while the other half of the cleft is vacant. If you like, you can make the inhibitor easier to see by putting the protein in CPK colors (**select protein; color cpk**) and coloring the inhibitor (NAG) green (**select nag; color green**). You can also try selecting nag and putting it in wireframe 100 display (don't forget to turn spacefill off) for perhaps a better view of the active site cleft.

Question 1: Describe what you see. Where in the active site cleft does the inhibitor bind? Describe the orientation of each of the three sugar rings with respect to the binding site. Roughly what fraction of the active site cleft is occupied by the inhibitor?

### EXERCISE 5B: CATALYTIC MECHANISM OF LYSOZYME

The currently accepted mechanism of lysozyme is shown below. A hexasaccharide portion of the cell wall polysaccharide binds to the active site cleft of the enzyme. The cleavage occurs between the fourth and fifth monosaccharide residues of the hexasaccharide bound to the cleft. The mechanism involves cleavage of the glycosidic bond with Glu35 acting as a general acid to donate a hydrogen to the fifth monosaccharide residue that comprises the leaving group. The resultant oxonium ion of the fourth monosaccharide residue undergoes nucleophilic attack by Asp 52, resulting in formation of a covalent complex. In the second phase of the reaction, Glu35 acts as a general base to accept a hydrogen from a water molecule as the water molecule acts as a nucleophile to attack the C1 carbon and displace Asp52.



Source:

[https://commons.wikimedia.org/wiki/File%3AMechanism\\_of\\_Lysozyme\\_action.png](https://commons.wikimedia.org/wiki/File%3AMechanism_of_Lysozyme_action.png)  
By LHchem (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons

**Question 2:** Color the protein yellow and the inhibitor in CPK colors. Select aspartate 52 (**select 52**) and color it magenta, but color its oxygen atoms red (**select 52 and oxygen; color red**). Select glutamate 35 and color it cyan but color its oxygen atoms red. Print an image of lysozyme showing the locations of the inhibitor and these two amino acid residues.

Examine the inhibitor in both spacefill and wireframe displays.

**Question 3:** Describe the location of residues 35 and 52 with respect to the active site and the bound inhibitor. Is the location and orientation of the atoms of the side chains of residues 35 and 52 consistent with the currently accepted mechanism of lysozyme?


## EXAMINING SIDECHAINS INVOLVED IN BINDING

Return to the PDB page for 1HEW. In the **Oligosaccharides** box, under **3D Interactions**, click on **Oligosaccharides Interaction**.

A new page opens with an image of the molecule in the program Mol\*, with the individual monosaccharide units represented as cubes. Letting the mouse cursor rest on each cube outlines the cube and displays a message identifying each as **NAG1** (nonreducing end), **NAG2** or **NAG3** (reducing end). For now, don't try to manipulate the structure with your mouse.

(Note: The mouse controls in Mol\* are different from those in Jmol and are summarized at [www.rcsb.org/3d-view/molstar/help/getting-started](http://www.rcsb.org/3d-view/molstar/help/getting-started) as follows:

- **Rotate:** click the left mouse button and move. Alternatively, use the Shift button + left mouse button and drag to rotate the canvas.
- **Translate:** click the right mouse button and move. Alternatively, use the Control button + the left mouse button and move. On a touchscreen device, use a two-finger drag.
- **Zoom:** use the mouse wheel. On a touchpad, use a two-finger drag. On a touchscreen device, pinch two fingers.
- **Center and zoom:** use the right mouse button to click onto the part of the structure you wish to focus on.
- **Clip:** use the Shift button + the mouse wheel to change the clipping planes. On a touchpad, use the Shift button + a two-finger drag.
- If an action is available for the right mouse button, it can also be completed by using the Control button + left mouse button.)

On the right, under **Components**, click on the  icon, then on **Non-covalent interactions**. We see that several types of interactions are turned **On**. Let's focus on the H-bonds by toggling everything else to **Off**.

There is a dashed line connected to NAG3, the monosaccharide on the left. Find it by clicking on the molecule and rotating it. (You may have to click on NAG3 once or twice when you're finished.) This represents an H-bond linking the sugar to the protein. Let your mouse cursor hover over the H-bond until it is highlighted. The message **A ASN103 OD1 – B NAG3 O6** (or **B NAG3 O6 – A ASN103 OD1**, depending on which end of the H-bond is selected) is displayed. This indicates the presence of an H-bond with the O6 hydroxyl group of NAG as the H-bond donor and the  $\delta$ -O atom (the sidechain carbonyl oxygen) of asparagine 103 as the H-bond acceptor.

Explore the H-bonds involving the other two monosaccharide units by clicking on each of the other two cubes. In some cases the interaction involves a bridging water molecule. For example, the O1 (ring oxygen) of NAG1 H-bonds to water molecule 278, which is in turn H-bonded to the carbonyl O of alanine 107.



Now turn **Hydrogen-bonds** off and turn on **Hydrophobic**. You should be able to determine that there is a hydrophobic contact between the acetyl group of NAG3 and the indole ring of tryptophan 62.

Applying a similar procedure to other ligand-bound proteins will enable you to detect ionic interactions, pi-stacking interactions, etc. Such interactions are not present at the lysozyme active site.

Question 4: Which of the three sugar rings interacts most extensively with the active site? Create and print a Jmol image that shows the inhibitor along with all of the amino acid side chains that interact with it. Use a display style that you think shows these interactions most effectively.

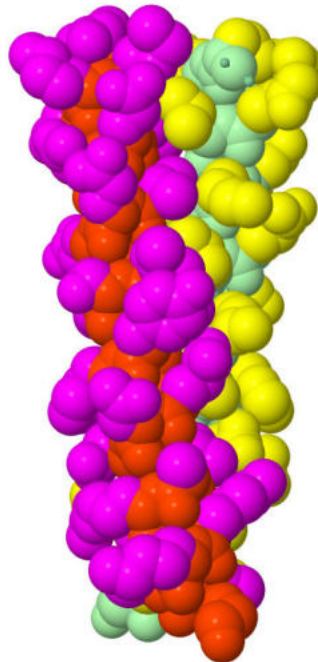
#### FURTHER EXPLORATIONS

**FE 1:** Display the file 3PGK, the structure of yeast phosphoglycerate kinase, one of the enzymes of glycolysis. Find out from the Jmol header which ligands are bound to the protein and what their abbreviations are. Explore the protein and describe the binding sites of the two substrates and the magnesium ion, and their spatial relationship to each other. Is the relative orientation of the two substrates and their reactive functional groups consistent with the reaction catalyzed by this enzyme? (NOTE: The command to select 3-P-glycerate is **select [3pg]**. The brackets are necessary whenever the name of the ligand begins with a number.)

Use the Ligand Interaction tool to investigate the binding interactions of the two substrates, and use Jmol to help you visualize these interactions. Note that in this protein electrostatic interactions are also important in ligand binding. What type of amino acid residues would you predict should be present in the vicinity of the phosphate groups of the two substrates? Use Jmol to verify your prediction. (Hint: Try displaying most of the protein in backbone display and just the sidechains of interest in spacefill.) Explain your observations and identify specific amino acid residues involved in electrostatic interactions.

**FE 2:** Display the file 1U0F, the structure of mouse phosphoglucose isomerase, another enzyme of glycolysis. Note that the substrate, glucose-6-phosphate, is bound in both the straight chain and ring forms, and both forms are shown bound to the same site, so you need to select them one at a time. Notice how deep the active site is. Putting the enzyme in a light color, like white, will help you to see the active site depression in a spacefill display. Examine the binding site using a slab setting of **slab 70**. Describe the binding site and compare the orientations of the bound straight-chain and ring forms of glucose-6-phosphate.

## EXERCISE 6: FIBROUS PROTEINS



In this lesson you will learn the following Jmol skill:

- How to select and separately manipulate individual polypeptide chains of a multi-chain structure

You will use your Jmol skills to:

- Examine examples of fibrous proteins
- Explore the two-stranded coiled coil structure, with two alpha-helices wrapped around each other to form a two-stranded rope. This forms the basis of the structure of keratin
- Explore the silk fibroin structure which is based on a stack of beta-pleated sheets; the structure is made possible by the unusual amino acid composition of silk
- Explore the collagen triple-helix, a three-stranded rope of extended helices (not alpha-helices) held together by a unique pattern of H-bonding
- Gain a better understanding of why each of these structures is favored by a specific type of amino acid sequence

Proteins can be either globular proteins, in which the polypeptide chains fold on itself to form a compact, roughly spherical structure, or fibrous proteins, in which the polypeptide chains adopt extended orientations with a regularly repeating secondary structure, allowing them to form fibers. In this exercise you will explore examples of some of the major types of fibrous proteins.

Prerequisite: You must have completed Exercise 1 before attempting this exercise.

TWO-STRANDED COILED COIL:

Display the file 2ZTA in Jmol. This is a leucine zipper domain from a yeast transcription factor called GCN4, which binds to DNA. The transcription factor is composed of two polypeptide chains that associate to form a dimer, in which the C-terminal  $\alpha$ -helical portions of the two polypeptide chains wrap around each other to form a two-stranded rope, called a coiled coil. A similar structure is found in the coiled coil region of  $\alpha$ -keratin.

Get rid of the water and display as a cartoon. You can see how this protein contains two alpha-helices that wrap around each other. Rotate the structure and observe different orientations to get a better sense of this wrapping.

Set the **Style Scheme** to **Ball and Stick**. Type **color chain**. Now type **backbone 100** to see the backbone superimposed on the ball-and-stick representation. Set the background to white (**background white**) to visualize the molecule better. Turn on the H-bonds and make them thick enough to see. Rotate the molecule with your mouse to get a sense of how the two helices wind about each other. Notice that the two right-handed helices form a left-handed superhelix.

Type **select backbone; spacefill**. Notice how the side chains interdigitate (interlock like fingers) with one another and come into close van der Waals contact. Then **select protein** or **select amino** and display the entire protein in spacefill to see these interactions more clearly.

To color the sidechains and backbone differently, try **select sidechain and :A; color magenta; select sidechain and :B; color yellow** (Note: The PDB page for 2ZTA tells us that the structure contains two polypeptide chains called A and B. The respective designations for these two chains in Jmol script are **:A** and **:B**.)

Type **show sequence** and find the amino acid sequence. The display **[ACE]0** indicates that the alpha-amino group is acetylated. The other abbreviations are the standard three-letter abbreviations for amino acids. Notice that the sequence exhibits a heptad repeat, where the first and fourth amino acid out of every seven (*a* and *d* positions) is nonpolar. See if you can figure out how to divide the sequence into heptads, so that nonpolar amino acids occur at these positions. (You will have to make an exception for the asparagine residues at position 16; asparagine is the one polar amino acid residue often found at the *a* and *d* positions of coiled coils.)

Question 1: Rewrite the sequence by hand in rows of seven amino acids, where each row corresponds to one heptad. Draw a box around the columns corresponding to the *a* and *d* positions.

Now **select sidechain; wireframe only; wireframe 100** to return the sidechains to wireframe display. Then **select 2 or 5 or 9 or 12 or 16 or 19 or 23 or 26 or 30; spacefill** [or **select (2,5,9,12,16,19,23,26,30); spacefill**] to show how the sidechains in the *a* and *d* positions form a hydrophobic strip that winds around the coiled coil.

Question 2: Print an image of the molecule in spacefill display color-coded as indicated above. Describe the noncovalent interactions between the two chains that stabilize the coiled coil.

Some proteins are known to form three- or four-stranded coiled coils as well. The type of coiled coil depends on the particular amino acids present at the *a* and *d* positions – substitution of isoleucine and/or valine for leucine alters the geometry for optimal van der Waals interaction, and GCN4 mutants containing these substitutions form three- or four-stranded coiled coils. You can see examples of each in 1GCM (3-stranded) and 1GCL (4-stranded).

#### SILK FIBROIN:

Type **1SLK** in the search window of the PDB site. Silk forms fibers, not crystals, so low-resolution X-ray fiber diffraction data is available, but it cannot be analyzed by X-ray crystallography. The available structure file for silk fibroin is a theoretical model based on constraints from the fiber diffraction data and knowledge about bond lengths and bond angles in peptides. It used to be part of the PDB but has now been removed to a separate Model Archive. For more information click on [10.2210/pdb1SLK/pdb](https://www.rcsb.org/structure/10.2210/pdb1SLK/pdb).

Display the structure in Jmol. The structure shown is a portion of a lattice that extends long distances in all dimensions. The portion shown contains fifteen polypeptide chains in a 5 x 3 array. It consists of three beta pleated sheets, each with five polypeptide chains hydrogen-bonded to each other. Type **show sequence** to see that this is an alternating copolymer of alanine and glycine residues, with each chain an (ala-gly)<sub>3</sub> hexapeptide with an acetylated N-terminus and with each chain's C-terminus linked in an amide bond to a molecule of methylamine. (You may have to scroll up in the Script Console window to see the complete sequence.) The acetyl and methylamine groups are shown in ball-and-stick display.

Again, most of the following views will be more easily seen with a white background. Note that this structure contains hydrogen atoms.

Using the pop-up menu, set the style scheme to wireframe. Select the C chain (**select :C**), which is the central chain in the structure and set it to wireframe 100. Now rotate the array and examine it from different directions. If you orient the molecule so that the axes of the chains are perpendicular to the screen, you will be looking at each chain from the end, and you will see the 15 individual chains.

Rotate the molecule and view it in different orientations. In one orientation you will see three sets of chains, each consisting of five chains superimposed on each other. You will be looking at the three pleated sheets from the end. In this orientation all the methyl sidechains of the alanine residues of the c chain point to one side and all the H sidechains of glycine point the other way. In this orientation it is easy to see the pleat in the polypeptide chain, with alternate peptide planes oriented in different directions. Orient the structure so that the methyl groups of the c chain are pointing toward you. Notice how successive CO groups (distinguishable from the red oxygens) point in alternate directions, as do the NH groups.

Let's examine a single pleated sheet. Type **restrict (:A,:B,:C,:D,:E); wireframe 100**. Rotate the molecule and explore different orientations. In one orientation you will see all five chains superimposed on each other, and you can see how the pleats in successive chains lie in the same orientation. In another orientation the five chains run parallel to each other. Display the hydrogen bonds as you have done previously (Exercise 2).

Question 3: Describe the orientation of the H-bonds within the sheet. Rotate the sheet so that you are looking at the ends of the five chains. Are all the H-bonds lined up with each other – what do you observe?

#### COLLAGEN TRIPLE HELIX:

Display 1CGD in Jmol. This is the structure of a peptide with repeats of the sequence proline-hydroxyproline-glycine, a model for collagen structure. Set the Style Scheme to CPK Spacefill. Type **delete (water,acy)** to get rid of the water and the acetic acid molecules that are part of the structure. Color by chain (**color chain**). Examine the structure in wireframe 100, spacefill, and backbone 100 displays.

Question 4: Is the triple helix right-handed or left-handed?

Now examine only one polypeptide chain (e.g. **restrict :C**) in Spacefill display.

Question 5: Is each individual polypeptide chain a right-handed or left-handed helix? (Hint: Look at how the five-membered rings wind around the helix axis.)

Restore all three chains to wireframe display (**select protein; wireframe only; wireframe 100**). (Don't forget to turn **spacefill off**.) Of the three amino acids in this molecule, which one will have all the NH groups? Why?

Notice how the chains are staggered so that at each level in the triple helix there is one gly residue, one pro, and one hypro. Is there a possibility of forming any interchain hydrogen bonds among these three residues?

Make the H-bonds visible as you have done previously. Click on two residues connected by an H-bond to identify them. Try this at different levels of the structure to convince yourself that it is always the same.

Question 6: Describe the H-bonds in collagen. Which residues are involved and which functional groups on those residues are involved?

#### FURTHER EXPLORATIONS

**FE1: Globular Proteins with Two-Stranded Coiled Coils:** This exercise allows you to explore some globular proteins that contain two-stranded coiled coils, both parallel and antiparallel, as part of their structure. Display 3COQ in Jmol. Look at the PDB page to see which letters are assigned to the two protein chains. Color by chain. Describe the secondary structure organization of each chain. Locate the portion of the protein involved in coiled coil formation, display the sequence and verify the presence of heptad repeats. How many heptad repeats are observed? Is the coiled coil parallel or antiparallel? For other example see 1G6N, 1GRJ and 1SER. Explore one or more of these proteins and report your observations on their structures.

**FE2: Triple-Stranded Coiled Coils:** This exercise allows you to explore some examples of triple-helical coiled coils, also known as 3-helix bundles. Display the protein 1RTM, color by chain, and locate the coiled coil. Display the amino acid sequence. Does it contain heptad repeats? Are the amino acid side chains in positions *a* and *d* facing inward as they do in two-stranded coiled coils? For other examples of both parallel and antiparallel three-stranded coiled coils see 3F57 and 3GHG. For the antiparallel coiled coil, determine which of the three chains is antiparallel to the other two. Explore one or more of these proteins and report your observations on their structures.

**FE3: Four-Stranded Coiled Coils:** As an example of a four-stranded coiled coil (four-helix bundle) see the *lac* repressor, 3TLB. Note that the four polypeptide chains are organized into two dimers. For the four helices that make up the four-stranded coiled coil, determine the relative orientation (parallel or antiparallel) of the two helices within each dimer as well as the relative orientations of each pair of adjacent helices from different dimers. Verify the existence of heptad repeats and examine the orientations of the amino acid side chains in positions *a* and *d*. Report your observations.

**FE4: Polyproline II Helix:** Why are the individual collagen helices extended left-handed helices with three residues per turn? This is because such a helix, called a polyproline II helix, is a favored conformation for a sequence containing multiple proline residues in a *trans* configuration. Display the protein 2IOC and color by chain. Examine the amino acid sequence and conformation of residues 54-62 and compare them to collagen. For another example see 3LII. Display the amino acid sequence and locate the region rich in prolines. Restrict the display to this region, and specifically examine the portion that is neither helix nor sheet. Locate the portion of the protein that exists as a polyproline II helix, and report your findings.

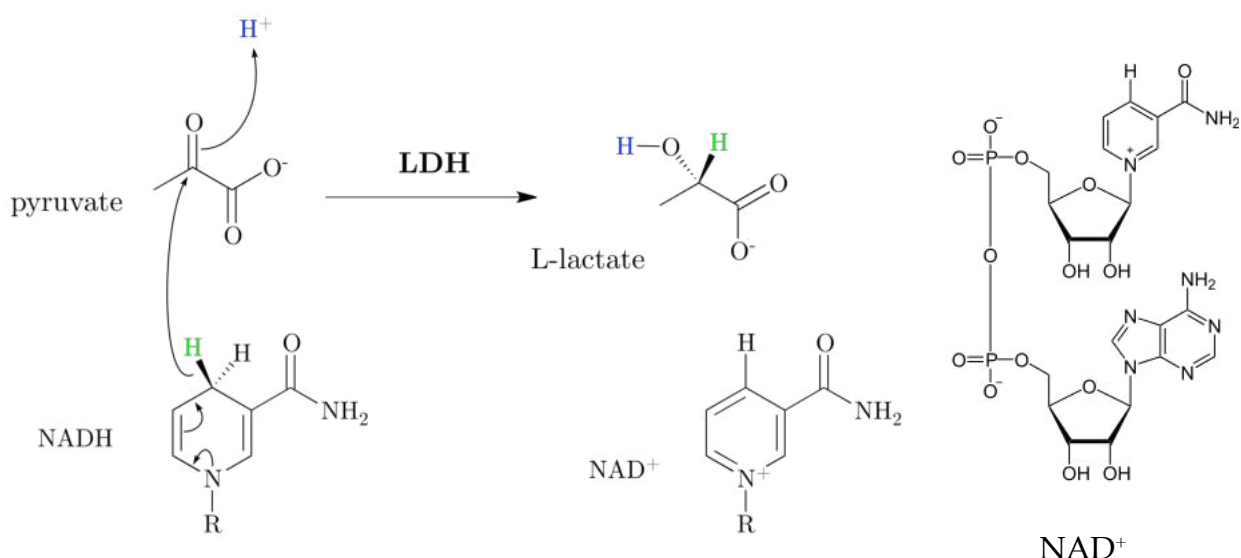
## EXERCISE 7: SEARCHING FOR PDB FILES

In this lesson you will learn the following skill:

- How to locate the structure file for a given protein in the PDB.

Up till now, you have been given the PDB ID numbers of the structures to download for viewing in Jmol. What if you want to locate the PDB file for your favorite protein? How can you search for that file? If your search returns a long list of structure files, how will you know which one to use? In this exercise you will learn how to conduct a search and select the proper file for use with Jmol.

We will search for a structure file for the enzyme lactate dehydrogenase (LDH), which participates in the pathway of glycolysis. The enzyme catalyzes the reaction:



From: [https://en.wikipedia.org/wiki/Lactate\\_dehydrogenase](https://en.wikipedia.org/wiki/Lactate_dehydrogenase),  
[https://en.wikipedia.org/wiki/Nicotinamide\\_adenine\\_dinucleotide](https://en.wikipedia.org/wiki/Nicotinamide_adenine_dinucleotide)

The keto group of pyruvate is reduced to a hydroxyl group. The hydrogen added during the reduction is removed from the nicotinamide ring of NADH, converting it to NAD<sup>+</sup>. The complete structure of NAD<sup>+</sup> is shown at the right. The enzyme has binding sites for each of the substrates.

Go to the PDB web page. Instead of entering a PDB ID, type **lactate dehydrogenase**. Click on **Go**. As of March 29, 2023, this brings up a list of 46,589 structures where the words "lactate" and "dehydrogenase" appear somewhere in the PDB file.



Do a new search, typing “**lactate dehydrogenase**”, including the quotation marks. Now we’re searching for the phrase “lactate dehydrogenase.” The search returns 3409 structures.

To get a more restricted list, scroll up to **Advanced Search Query Builder**. Click next to **Structure Attributes** to display a drop-down menu. Next to the words **Type to filter and/or select an attribute**, click on the down arrow to display the dropdown menu. Click on **Structure Title** and then type **lactate dehydrogenase** in the window and click **Search**. We now have 160 structures.

Let’s refine the search and try to find the human enzyme. In the list of **Refinements** on the left, notice the category **SCIENTIFIC NAME OF SOURCE ORGANISM**. Click on **More...** to see an even longer list. Then click on **Homo sapiens (34)** and then the arrow to the right of the word **Refinements** to bring up the 34 human structures. To make things easier for now, on the horizontal bar above the results, click on the icon with the horizontal bars to see a list of the structures.

Of these 34 structures, two (1T2F and 1I0Z) are for the heart isozyme (H isozyme) of LDH, also called B LDH, and the other 32 are for the muscle isozyme (M isozyme), also called A LDH.

You can also use the **Sort by** pulldown menu to help find the file you are looking for. Selecting **Resolution: best to worst** will bring the structures with the best resolution to the top of the list.

On the banner above the results, click on **Tabular Report** and then on **Ligand** in the pulldown menu. The files are listed with information about bound ligands, including the 3-letter abbreviation of each ligand (e.g. NAI is NADH).

Go back to the **Advanced Search Query Builder** at the top. In the **Chemical Component** section click on **Chemical ID(s)**. Enter **NAI** in the search window and click on the search icon to see the structures that contain bound NADH.

Question 1: How many lactate dehydrogenase structures contain bound NAD<sup>+</sup>? Bound NADH? Write a Jmol command to select the NADH. (Hint: The answer is not **select nadh.**)

Question 2: How many structures of mouse (*Mus musculus*) lactate dehydrogenase are available in the PDB?

Question 3: Find the highest-resolution structure for human enolase with bound 2-phosphoglyceric acid. State the PDB ID number, the resolution, the number of polypeptide chains, and the identity of the bound ligands.

Question 4: Locate a structure for rabbit (*Oryctolagus cuniculus*) actin that comprises only actin rather than a complex of actin with one or more other polypeptides. Describe the structure (number of chains, ligands, experimental method, resolution).

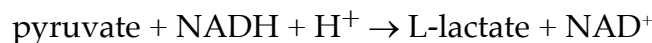
## EXERCISE 8: WHAT'S IN A FILE?

In this exercise you will examine the text of a PDB structure file and learn what is contained in it.

You will learn the following skill:

- How to display and read the header of a PDB file and how to get useful information that will help you to manipulate the image of the protein in Jmol and to understand what you are seeing.

We will examine a PDB file for the heart isozyme of human lactate dehydrogenase (LDH). The enzyme catalyzes the reaction



The reaction is shown on the first page of Exercise 7, including structures of all the substrates and products. In Exercise 7 we determined that there are two files for this enzyme, one of which is 1I0Z (the third character is a zero).

Search for the PDB page for 1I0Z. The information on the page is similar to what was noted in Exercises 1 and 2. There is an abstract of the journal article in which the structure was originally published, along with information about when the structure was deposited and later modified, as well as information about the X-ray crystallographic data and links to the original X-ray data. The **Macromolecules** section tells us that there are chains designated A and B.

The **Ligands** section tells us that we also have 1,4-dihydronicotinamide adenine dinucleotide (NADH) abbreviated NAI, and oxamic acid, an analog of pyruvate, abbreviated OXM. Oxamic acid is simply pyruvic acid with the methyl group replaced by an amino group. It is an inhibitor that occupies the site where pyruvate binds.

Let's now take a look at the actual PDB file. To the right of the PDB ID number click on **Display Files** and then **PDB Format**. We see a browser tab displaying the text of the file. This file contains all the information that Jmol and other programs need to locate every atom of a protein in space, to identify each amino acid residue and hetero group by name and number, to locate the various elements of secondary structure, and to distinguish the different polypeptide chains. It also contains information that may be important to you if you want to use this structure to explore the protein, including information about which are the more reliable and less reliable regions of the model. Let's take a closer look at the contents of the file:

The initial information provides the name and source of the enzyme, the submission history and the title of the paper where the structure was published. There is information about the crystal preparation and the X-ray diffraction results. There is a remark (REMARK 300) that "the biological assembly is a tetramer generated from the dimer in the asymmetric unit." We will explain what that means in Exercise 10.

REMARK 465 tells us that leucine 333 was not located in the experiment. This is the C-terminal residue. The C-terminal residue probably had some mobility within the crystals, causing it to lack a well-defined position. Occasionally residues from the middle of a polypeptide chain are also missing from the PDB model. It is important to know this when you are using the model.

There is information about atoms in close contact and dihedral angles that are outside the allowed regions of the Ramachandran plots. These may be areas where the protein is slightly strained or possibly areas where the model is less reliable.

Some abbreviations for the NADH and oxamic acid binding sites are defined.

The amino acid sequence is given.

Then the HET (hetero = non-amino acid) molecules are identified: note that there is one NADH and one OXM for each chain.

The locations of the helices and sheets are specified as well as the positions of three *cis* peptide bonds and the amino acids comprising the different named sites.

After this there is a list of every atom in the structure along with its *x*, *y*, and *z* coordinates, occupancy and B-factor (temperature factor). For example, atom number 26 is the gamma carbon atom of lysine 4 in the A chain. Its *x*, *y*, and *z* coordinates with respect to the origin are 11.808, 72.154, and 35.838. The occupancy is 1.00, as it is with all the atoms in this file, indicating that all the molecules in the crystal have exactly the same conformation. (In some structures the molecules are found in multiple conformations, and a given atom may have more than one position, each with a designated fraction of occupancy.)

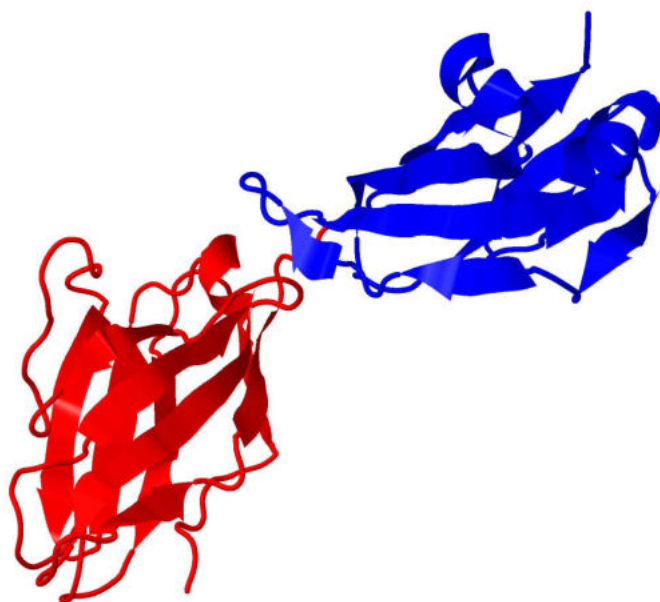
The B-factor (temperature factor) provides information about how much the position of an atom varies about its specified position. A high value of the B-factor indicates either that the atom experiences a large degree of freedom of movement due to thermal motion, or that the protein is disordered at this position and that the atom occupies a range of different positions in different molecules within the crystal. The B-factor is equal to  $8\pi^2$  (which is equal to 79) times the average value of the square of the displacement of the atom due to thermal vibration. So a B-factor of  $79\text{\AA}^2$  means that the average displacement is 1  $\text{\AA}$ ; a B-value of 30 indicates a displacement of 0.62  $\text{\AA}$ . In this

model the B-factors of the A chain are 30-56 Å<sup>2</sup> for the first 20 residues, the last 6 residues, and for a stretch of 17 residues (212-228) in the middle of the chain, and lower (as low as 7Å<sup>2</sup>) everywhere else.

The CONECT information at the end tells the program which atoms in the HET molecules are connected by bonds.

Question: Examine the PDB file 1TAQ. What protein is this? How many different polypeptide chains are present, and what is the length of each? Are any residues missing from the structure, and if so, which ones? Are there any *cis* peptide bonds, and if so, which one(s)?

## EXERCISE 9: DOMAINS



In this exercise you will learn the following Jmol skills:

- How to find information about the boundaries between the domains of a polypeptide chain and how to visualize them.
- How to select and separately manipulate individual polypeptide chains of a multi-chain structure
- You will get additional practice working with ligands bound to proteins

You will use these skills to:

- Examine the domain organization of an antibody light chain and lactate dehydrogenase, each of them a protein with two continuous domains. You will then examine D-glycerate dehydrogenase, a protein where one of the domains is discontinuous.

Most globular proteins (all except for the very smallest proteins) contain polypeptide chains organized into two or more structural **domains** – compact, stably folded regions of the chain, which may be connected to each other through flexible hinges, or which may dock against each other to form a compact structure for the entire chain. These structural units often correspond to different functional units as well with, for example, binding sites for different ligands.

Prerequisites: You must have completed Exercises 1 and 5A before attempting this exercise.

### DOMAIN ORGANIZATION OF AN ANTIBODY LIGHT CHAIN

Display the structure 3MUH. This is the crystal structure of a light chain derived from an antibody that binds to the envelope glycoprotein of HIV-1 virus. You can see immediately that the protein is comprised of two distinct regions, each containing a sandwich of two beta sheets facing each other. These are referred to as the two domains of each polypeptide chain. A structural domain is a compact unit of a protein that contains a hydrophobic core and can fold independently. Domains are generally separated spatially from each other, and often each domain has a distinct function.

Immunoglobulin light chains contain a variable domain whose sequence and structure vary according to the antigen that the antibody recognizes, and a constant domain that is the same in all antibody molecules. Then click on the two ends of the molecule to locate the N- and C-termini and to identify their residue numbers. Consult your biochemistry text to determine which is the variable domain and which the constant domain.

Question 1: Describe the secondary and tertiary structural organization of each of the two domains.

Locate the approximate position that divides the chain into two domains. Select the N-terminal domain (**select 1-xx**, where xx is the residue number of the amino acid at the division between the two domains) and the C-terminal domain in turn and assign different colors to them. Set the display to spacefill.

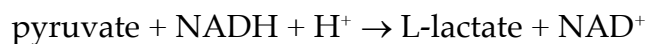
Question 2: Print an image of the molecule in spacefill display with the domains colored differently.

Note the ligand, which is identified on the PDB page.

Question 3: Click on the amino acid to which the ligand is attached and identify it. How many monosaccharide units are present in the ligand? Is it an N-linked or O-linked sugar?

### DOMAIN ORGANIZATION OF LACTATE DEHYDROGENASE:

The enzyme lactate dehydrogenase (LDH) catalyzes the reaction:



The reaction is shown on the first page of Exercise 7, including structures of all the substrates and products.

Go to the PDB page for 1I0Z (the third character in the ID is a zero). (This is the file used as an example in the exercise "What's in a File?" and is also one of the two structures for the heart isozyme of human LDH found in the exercise on "Searching for PDB Files.") Display the structure in Jmol, and delete the water. Color according to chain. The PDB page tells us that there are two chains, labeled A and B. To examine just one chain, type **restrict :A**.

It is not at all obvious that this chain contains multiple domains, as it was for the antibody light chain. Type **color group**. You can now see that the N-terminal (cooler colors) and C-terminal (warmer colors) portions are spatially distinct.

Go to the PDBSum site at <https://www.ebi.ac.uk/pdbsum/>. Enter 1I0Z in the search window. On the next page select the **Protein** tab. We see that the CATH database recognizes two domains, an Alpha Beta 3-Layer(aba) Sandwich, colored red, and an Alpha Beta Alpha-Beta Complex, colored blue. Examining the sequence below shows that domain 1 comprises residues 15-162 and domain 2 comprises residues 163-330, with residues at the N- and C-termini not part of either domain.

Let's color-code the two domains. Select amino acids 15-162 (**select 15-162**) and color them orange. Color residues 163-330 cyan. (Note: If you want to find the N-terminus, type **select 1** and color it blue. It should be apparent where the C-terminus is.)

Finally select the entire A chain and render it in spacefill display. There is an extensive buried interface between the two domains, and the spacefill image gives no indication of the presence of multiple domains.

#### LIGANDS OF LACTATE DEHYDROGENASE:

LDH has binding sites for each of the substrates, and the Ligands section of the PDB page tells us that the structure contains ligands corresponding to both binding sites – 1,4-dihyronicotinamide adenine dinucleotide (NADH, designated NAI) and oxamate (OXM). The structure shown for oxamate is identical to pyruvate except that the methyl group is replaced by an amino group. Oxamate is an inhibitor that competes with pyruvate for binding at the active site.

Let's take a closer look at the ligands. With everything displayed as wireframe 100, type **restrict not protein** and color in CPK style. Zoom in on one of the NADH's. In this reaction the NADH donates a hydrogen from the top of the nicotinamide ring to the carbonyl group of pyruvate, converting it to a hydroxyl group. (See the figure at the beginning of Exercise 7.) Notice that the NADH and oxamate are oriented so that the carbonyl of oxamate is close to the hydrogen of NADH that will be donated. Examine the ligands in both wireframe and spacefill display.

Select the ligands separately, put them in spacefill display and color them different colors. Now restore the protein (by selecting **:A and amino** and rendering it in cartoon display).

Question 4: Which domain binds NADH and which binds oxamate?

Note: Dehydrogenases that use NAD<sup>+</sup> as an oxidizing agent often have this type of domain organization; one domain binds NAD<sup>+</sup> or NADH and the other binds the substrates to be oxidized or reduced. So, structural domains of proteins often correspond to functional domains. The domain that binds NAD<sup>+</sup> adopts an organization referred to as a Rossmann fold. Display everything as a cartoon. If possible find a figure in your textbook showing a Rossmann fold or NAD-binding domain and compare that image with your Jmol image.

If you like, you can use the Ligand Interaction button to investigate the substrate binding sites as you did for lysozyme in Exercise 5.

#### DETERMINING DOMAIN BOUNDARIES:

There are other databases that detect domains in proteins. Earlier in this exercise we used the information from the CATH database to identify domain 1 as residues 15-162 and domain 2 as residues 163-330. It is possible to display other domain assignments from different databases as well. Links to the home pages of several databases are available in the **Annotations** tab of the PDB page for 1I0Z.

The SCOP (Structural Classification of Proteins) database is available at [scop.mrc-lmb.cam.ac.uk/](http://scop.mrc-lmb.cam.ac.uk/). Type 1I0Z in the search window at the upper right and hit Return. The next page tells us that there is a single domain comprising residues 1-332. Interestingly, if you return to the SCOP home page and click on **SCOP 1.75** to access an older version of SCOP, then under **Access methods** click on **Enter SCOP at the top of the hierarchy**, search for 1I0Z and click on each line of the search results, you'll see that the first domain comprises residues 1-160 and the second domain residues 161-332.

The ECOD (Evolutionary Classification of Protein Domains) database is available at <http://prodata.swmed.edu/ecod/>. Click on **Search by PDB ID** and then enter 1I0Z in the search window. The next page shows that domains are assigned as 1-160 and 161-332.

The Pfam database is available at [pfam.xfam.org](http://pfam.xfam.org). Enter 1I0Z in the search window next to the words **Jump To** and click **Go**. On the next page click on the fifth green square (with the image of a beta strand connected to a helix). A new browser tab will open. Scroll to the bottom and look at the line labeled **Pfam**. Clicking on the dark and lighter



purple rectangles in sequence reveals that domain 1 is assigned to residues 21-160 and domain 2 to residues 163-333.

Note that the four databases assign the domains in slightly different ways. There is no universally accepted definition of a domain, and different servers apply different criteria in their assignments.

#### DISCONTINUOUS DOMAINS IN D-GLYCERATE DEHYDROGENASE:

The enzyme D-glycerate dehydrogenase participates in one pathway of serine metabolism. In this pathway the amino group of serine is replaced by a carbonyl group in a transamination reaction, forming hydroxypyruvate. The carbonyl group of hydroxypyruvate is then reduced to D-glycerate by NADH in a reaction catalyzed by D-glycerate dehydrogenase. A kinase then converts D-glycerate to 3-phospho-D-glycerate, an intermediate of glycolysis.

Display the structure 1GDH. Eliminate the water and color by chain (**color chain**). Note the domain structure of the individual chains and the way the two chains associate to form a dimer by using one of the two domains to form the dimer interface. Type **restrict :A** to view a single chain. In order to distinguish the N- and C-termini, color by group.

Question 5: Describe the domain organization of this protein.

You can confirm your answer by going to the domain databases mentioned above. The assignments are:

CATH	domain 1 = 2-101, 291-318	domain 2 = 102-285
SCOP 2.0	single domain = 2-321	
SCOP 1.75	domain 1 = 101-291	domain 2 = 2-100, 292-321
ECOD	domain 1 = 2-100, 292-321	domain 2 = 100-291
PFAM	single domain = 6-322	

Note the assignments of the CATH, SCOP 1.75 and ECOD databases to two domains. One of the domains is formed by the folding of the middle of the protein and the other by the folding of the portions of both ends of the protein into a single compact unit -- a discontinuous domain. Select individually the middle domain and each of the two regions of the discontinuous domain and color them differently. Finally, select the entire A chain and display it in spacefill mode.

## FURTHER EXPLORATIONS

**FE1:** Display each of the following proteins, and do the following: Determine the domain boundaries, using one or more of the databases (one is enough), and observe the domains in Jmol; describe the domain organization of each protein in terms of the primary sequence, the relative positions of the domains within the tertiary structure, and, when relevant, the locations of binding sites for ligands: troponin C (4TNC), phosphoglycerate kinase (2ZGV), aspartate transcarbamoylase regulatory chain (8ATC B chain), aspartate transcarbamoylase catalytic chain (8ATC A chain), NADH peroxidase (12NPX), hexokinase (1QHA), pyruvate kinase (1PKM), E. coli DNA polymerase I Klenow fragment (1D8Y) and alcohol dehydrogenase (5ADH).

**FE2: DOMAIN SWAPPING:** In Exercise 1 you looked at file 1PGB, the B1 immunoglobulin-binding domain of Streptococcal Protein G. Display that structure in Jmol to refresh your memory. Now display the structure 1Q10 and color by chain. This is a mutant version of the protein that forms dimers. Notice how the dimer is formed – examine the conformation of each polypeptide chain and compare it to the conformation of the chain in 1PGB. Two beta strands from each chain are swapped with two beta strands from the same position of the other chain. Each four-stranded beta sheet contains two strands from the green chain and two strands from the blue chain. This is referred to as a "domain-swapped dimer."

For a second example, display the structure 1R5I, a complex that contains a protein named MAM (*Mycoplasma arthridis*-derived mitogen). Color by chain. The D chain is the MAM protein; restrict the display to the D chain, and note the two-domain structure of MAM. (Save an image of this structure for the next part of the exercise.) Now display 3KPH, the structure of the isolated MAM protein. What do you observe? Compare this structure to that of the MAM protein in 1R5I. Describe the 3KPH structure with attention to the ways the two chains associate to form a dimer. Why is this type of structure referred to as a domain-swapped dimer? Is it really a full domain that is "swapped?"

Examine one or more of the following files to see other examples of domain swapping, and report your observations: prephenate dehydrogenase (3DZB), cystatin C (1G96), apo-calmodulin (1QX5), glucocorticoid receptor (3H52), thioredoxin mutant (3DIE) and antithrombin-III (2ZNH).

## EXERCISE 10: SUBUNITS – ASYMMETRIC UNITS AND BIOLOGICAL ASSEMBLIES

In this exercise you will learn:

- The difference between an Asymmetric Unit and a Biological Assembly and how the number of polypeptide chains may differ between these two

You will learn the following Jmol skills:

- How to download and properly display the Biological Assembly of a protein using the Biological Assembly file from the PDB
- How to select and separately manipulate individual polypeptide chains of a multi-chain structure represented by multiple “models” in a Biological Assembly file
- How to download and properly display the Biological Assembly of a protein using the Protein Surfaces, Interfaces and Assemblies (PISA) database

**Subunits** are the individual polypeptide chains that make up an oligomeric structure. An important issue often comes up in working with proteins containing multiple subunits: a PDB file may describe either the **Asymmetric Unit** of a structure or the **Biological Assembly**. In this exercise, you will learn what these terms mean and learn how to display a structure that corresponds to the form of a protein that actually occurs in solution and in cells.

Prerequisite: You must have completed Exercise 1 before attempting this exercise.

An important aspect of examining multichain proteins has to do with the distinction between the Biological Assembly and the Asymmetric Unit. For this lesson we will use the enzyme lactate dehydrogenase (1I0Z) as an example, as we did in Exercises 7, 8, and 9. The **Macromolecules** section on the PDB page tells us that there are two chains, labeled A and B. Display the structure in Jmol. Type **color chain**. Go back to the PDB page. Note that the image of the molecule at the upper right has the heading **Biological Assembly 1** with an arrows to the left and right. Click on one of the arrows to see an image of the **Asymmetric Unit**.

Question 1: Which of the two images resembles the structure that you see in the Jmol window? When you display an ordinary PDB File, is it the Asymmetric Unit or the Biological Assembly?

(Note: In some cases, there is more than one possible way that polypeptide chains can be arranged in an oligomer, so there may be a Biological Assembly 2, Biological Assembly 3, etc.)

In this case the Asymmetric Unit is a dimer, while the Biological Assembly is a tetramer. What is an Asymmetric unit? Clicking on the question mark next to the words

"Biological Assembly" brings up a page that tells us, "The asymmetric unit is the smallest portion of a crystal structure to which symmetry operations can be applied in order to generate the complete unit cell (the crystal repeating unit)." In other words, even though this enzyme is a tetramer, the coordinates of the Asymmetric Unit (a dimer) are sufficient information to generate the complete structure, as long as the file tells you which symmetry operations to apply. So the Asymmetric Unit coordinates convey the same amount of information that the complete coordinates of the tetramer would convey, but in a file half the size.

How can we visualize the tetramer? There are two ways to do this:

First method: On the Structure Summary page for 1I0Z, expand the **Download Files** dropdown menu. Click on **Biological Assembly 1 (PDB-gz)**. Save the file to your computer, and remember where you saved it. Now open the file in Jmol BY USING ONE OF THE METHODS DESCRIBED AT THE BOTTOM OF PAGE 13. DO NOT USE THE GET PDB COMMAND.

Color by chain (**color chain**). Type **rotate x 90**. Then type **model all**.

Question 2: What do you observe when you type **model all**? Print a copy of this image.

What is happening here? Jmol has loaded two "models," each consisting of one of the two symmetry-related dimers. When you first open the file, only model 1 is displayed. You can return to that view by typing **model 1**. To see the second dimer, type **model 2**. The command **model all** superimposes all the models. (You can also use the command **model 0**).

At this point you have two blue A chains and two green B chains. Let's change the chains in model 2 to different colors. You can do this by typing **select :A/2; color red; select :B/2; color yellow**. Notice the tetrahedral arrangement of the four chains.

Second method: Another way to get a PDB file containing the coordinates of the biological assembly is to go to the **Protein Surfaces, Interfaces and Assemblies (PISA)** server ([www.ebi.ac.uk/msd-srv/prot\\_int/pistart.html](http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)). Click the button to **Launch PDBePisa**. Enter 1I0Z and click on **Assemblies**. On the next page, make sure the tetramer is selected and click **Download**. The contents of a PDB file will be displayed. Save the file, renaming it 1I0Z Assembly, and notice where you saved it. Now open the file in Jmol BY USING ONE OF THE METHODS DESCRIBED AT THE BOTTOM OF PAGE 13. DO NOT USE THE GET PDB COMMAND. You can see that when this file is opened, a tetramer is immediately displayed. Each chain is assigned a unique letter, and **color chain** results in an image with chains in four colors. Type **restrict not protein**. You can now see that each NADH and oxamate ligand is given its own chain designations, and each has a unique color.

By double-clicking in turn on each chain, you can determine that the four protein chains are designated A, D, G and J. (Alternatively, display the entire model in wireframe. Then **select :a; spacefill** will reveal which chain is the A chain, **select :b; spacefill** will do the same for the B chain, etc.) Everything is in a single model, so you don't have to be concerned about models when you work with the file obtained by this second method.

We can observe different relationships between the Asymmetric Unit and Biological Assembly with other lactate dehydrogenase structures. In a previous version of this manual (December 2017), when there were fewer structures in the PDB, there were only 20 structures for human lactate dehydrogenase (LDH), not the 34 found in 2023. Let's examine those 20 files – 5IXS, 5IXY, 4ZVV, 4R68, 4R69, 4OJN, 4OKN, 4QSM, 4QT0, 4RLS, 4QO7, 4QO8, 4L4R, 4L4S, 4M49, 4JNK, 1T2F, 1I0Z, 4AJP and 1I10.

LDH exists in solution and in cells as a tetramer. If we look at the Macromolecules section of the summary page for each of the files listed above, we discover the following: As we have just seen, for 1I0Z the Asymmetric Unit is a dimer, while the Biological Assembly is a tetramer. The same is true for 4L4R and 4L4S. However, for 5IXS, 5IXY, 4ZVV, 4R68, 4R69, 4RLS, 4QO7, 4QO8, 4M49, 4JNK, 1T2F and 4AJP the Asymmetric Unit is the same as the Biological Assembly, a tetramer. For 4OJN, 4OKN, 4QSM, 4QT0 and 1I10, the Asymmetric Unit contains eight chains comprising two tetramers whose 3-D structures are very close but not identical -- and for each structure there are two Biological Assemblies, one corresponding to each of the two tetramers.

Situations like these are very common. If you click on the question mark next to the words Biological Assembly above the figure of the protein on the PDB page, you will see another example, hemoglobin, of a tetrameric protein with structures containing dimers, tetramers, or octamers as the Asymmetric Unit. We will see in Exercise 14 that the Asymmetric Unit of the membrane protein (porin) is a monomer, while the Biological Assembly is a trimer. Also some of the PDB files for double-stranded DNA molecules have a single polynucleotide chain as the Asymmetric Unit (Exercise 17).

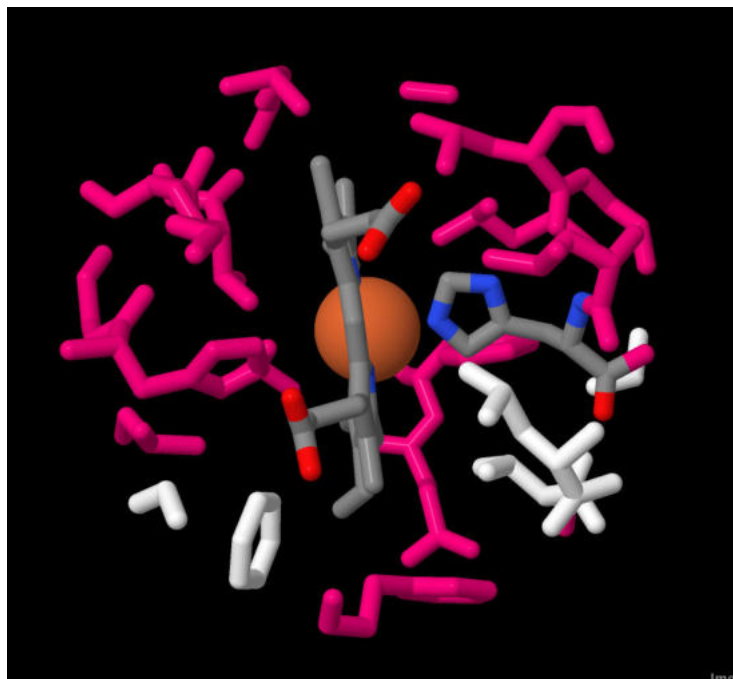
It's **very important** to notice when the Asymmetric Unit and Biological Unit are different. In Exercise 9, working with the Asymmetric Unit of LDH was sufficient to teach us some things about domains, but in general you should always work with the Biological Assembly.

## FURTHER EXPLORATIONS

**FE1:** For the following proteins, determine the number of chains in the Asymmetric Unit and in the Biological Assembly, and print an image of the Biological Assembly, with each chain a different color: liver alcohol dehydrogenase (5ADH), bacteriorhodopsin (1KGB), ribulose-1,5-bisphosphate carboxylase (RuBisCo, 8RUC), insulin (3JSD), bacterial alcohol dehydrogenase (1ZK3), and glutamine synthetase (2UU7).

**FE2:** For the following proteins, determine the number of chains in the Asymmetric Unit and in the Biological Assembly, and print an image of the Biological Assembly, making each chain a different color: glycogen phosphorylase (8GPB), phosphofructokinase (4PFK), glutamine synthetase (2OJW), tryptophan synthase (1TTQ), aldehyde dehydrogenase (2VLE), and fatty acid binding protein (1FDQ).

## EXERCISE 11: HEMOGLOBIN



In this lesson you will learn the following Jmol skills:

- How to select and separately manipulate individual polypeptide chains of a multi-chain structure
- How to visualize the region of a protein in the vicinity of a bound cofactor
- Additional experience working with ligands bound to proteins

You will gain the following Jmol experience:

- Comparing two structures that represent alternative conformations of a single protein
- Exploring the interconversion of the R and T states of hemoglobin and its effect on the geometry of the heme group and its ligands

Hemoglobin, the oxygen carrier of blood, exists in two different conformations, the T state with a low affinity for oxygen, and the R state with a high affinity for oxygen. The existence of an equilibrium between these two conformations leads to cooperative oxygen binding. The mechanism of this effect can be traced to a movement in the position of the iron atom of the heme group when oxygen binds. Here you will examine and compare the structures of T state hemoglobin and R state hemoglobin. Working with hemoglobin may provide you with a better understanding of how this complex protein functions.

Prerequisites: You must have completed Exercises 1 and 5A before attempting this exercise.

Display the structure 1THB, deoxyhemoglobin, in Jmol. Set the background to white, delete the water and color by chain (**color chain**). This is actually hemoglobin with oxygen bound to two of the four heme groups, but it crystallizes in the T state, the same conformation as deoxyhemoglobin. There are two alpha chains, labeled A (purple) and C (pink), and two beta chains, labeled B (green) and D (yellow). (You can determine the color assignment by clicking on each chain and verifying its chain designation in the Script Console.) A molecule of inositol hexaphosphate (IHP) is bound to the molecule. In the file it is assigned to the B chain and colored green (although it is a darker green than the B chain itself). IHP is an analog of 2,3-bisphosphoglycerate (BPG) that is used in place of BPG by some non-mammalian hemoglobins to regulate the oxygen affinity.

Set the Style Scheme to Spacefill and color by Chain. Notice the two  $\alpha\beta$  dimers, alpha chain A and beta chain B (purple and green) and alpha chain C and beta chain D (pink and yellow). There are extensive contacts between the  $\alpha$  and  $\beta$  chains in each dimer, less contact between  $\alpha$  and  $\beta$  chains in different dimers (AD and BC), and no contact between the two  $\alpha$  chains or the two  $\beta$  chains.

Question 1: Which subunits serve as the binding site for BPG?

Sandwiched within each subunit you can also see a flat heme group. To see these more clearly **select protein** and display in backbone 100 or cartoon mode (don't forget to turn spacefill off), while leaving the hemes in spacefill.

With the protein displayed as a cartoon type **restrict :B** to examine one polypeptide chain. (The Jmol designation for the B chain is **:B**.) Select the heme (**select hem and :B**) and color it cpk. Click on the two ends of the polypeptide chain to figure out which is the N-terminus and which is the C-terminus. Trace the polypeptide chain from N- to C-terminus, going through the eight helices A through H.

Now zoom in on the heme. Select the porphyrin ring (**select hem and not iron and :b**) and display in wireframe 80 mode (don't forget – **wireframe only** or **spacefill off**). Make sure that the iron atom is still in spacefill display – color it red. Select all the sidechains of the B chain (**select sidechain and :B**), turn off cartoon display and put them in wireframe 80 display. Can you tell what type of amino acid residue serves as the axial ligand for the iron atom? If you're not sure, click on the wireframe image to identify it.

(Note: You may be having a problem with rotating the molecule, such that the center of rotation is off the center of the screen and the object of interest moves off the screen when you try to rotate the molecule. If so, choose **Define Center** from the **View** menu. The problem should be fixed. Alternatively you can click on a residue in the center of the image to identify it (e.g. [GLY]64:B), then type **center 64:B**..



Question 2: Which amino acid residue (type and number) is the axial ligand for the iron?

Select this residue, display it in spacefill mode and color it cpk. Look at the intimate contact between this sidechain and the iron atom. Notice how this sidechain is attached to one of the helices. Now restore the sidechain to wireframe 80 display. Rotate the porphyrin ring until you are looking at it directly along one of the edges and the plane of the ring is perpendicular to the computer screen.

Question 3: Is the center of the iron atom in or out of the plane of the porphyrin ring? Describe the relationship of the iron, the axial ligand, and the porphyrin plane. Print a figure that illustrates your answer.

Let's look at the vicinity of the iron atom. Type **restrict within (10.0,iron) and :B** to eliminate everything more than 10.0 Å from the iron atom. Notice that the sixth coordination site is empty so that an oxygen molecule can bind. Notice the distal histidine just adjacent to the empty oxygen binding site.

Let's take a look at the inositol hexaphosphate binding site. Restore the entire molecule to wireframe 20 display. Now **select positive** and **color magenta**. Then **select ihp** and render in spacefill display with CPK colors. Finally **restrict within (10.0,ihp)**. Click on each of the magenta groups interacting with the negatively charged phosphates of IHP.

Question 4: List the positively charged amino acid residues and the specific polypeptide chains that interact with the bound inositol hexaphosphate. (Consult the second paragraph of this exercise to help you identify the chains.)

Close this file and load 1CBM, which will serve as our structure for the oxy conformation (R state) of hemoglobin. For technical reasons it is difficult to obtain good crystals of oxyhemoglobin, and the 1CBM structure is a  $\beta_4$  tetramer of hemoglobin with carbon monoxide bound to the oxygen binding sites. A variety of evidence has shown that this molecule adopts the oxy conformation.

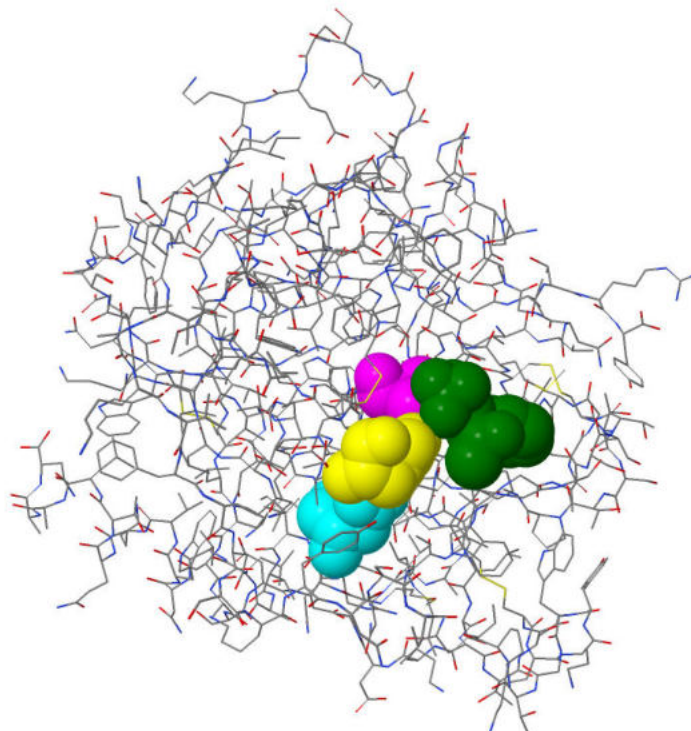
Set the Style Scheme to CPK Spacefill and color by chain. Notice that no IHP is bound, and there is no space between the chains large enough to accommodate an IHP molecule.

As before, zoom in on one of the hemes, putting the protein in cartoon display, the sidechains in wireframe, the heme in spacefill and CPK colors, as you did above. See if you can do all this without looking at the instructions above – if not, read the steps above and repeat them with this structure.

Notice that there is a molecule of CO bound to the sixth coordination site of the iron. As before, **select hem and not iron** and display in wireframe 80, keeping the iron in spacefill.

Question 5: Describe the position of the center of the iron relative to the plane of the porphyrin ring, and compare it to what you observed for deoxyhemoglobin.

## EXERCISE 12: THE MECHANISM OF CHYMOTRYPSIN



In this exercise you will gain the following Jmol skill:

- Zooming in on the active site of an enzyme and exploring it in close detail.

Chymotrypsin cleaves proteins on the C-terminal side of aromatic amino acid residues. Chymotrypsin is a member of the serine protease family, whose mechanism of action is one of the most intensively studied and best understood of all enzyme mechanisms. By closely examining the spatial relationships between the protein, its substrate binding pocket, and the amino acid sidechains present in the vicinity of the active site, you will gain a better understanding of the mechanism of chymotrypsin than you can get from looking at static two-dimensional figures.

Prerequisites: You must have completed Exercises 1 and 5A before attempting this exercise.

Load the file 6GCH into Jmol. Note that the **Macromolecules** section tells us that the protein is gamma chymotrypsin.  $\gamma$ -Chymotrypsin is the same as  $\alpha$ -chymotrypsin except that residues 11 to 13 have been cleaved from the end of the A chain. This structure is that of a covalent complex of  $\gamma$ -chymotrypsin with the inhibitor N-acetylphenylalanine-trifluoromethylketone (APF). APF is a derivative of phenylalanine with an acetyl group ( $\text{CH}_3\text{CO}$ ) attached to the alpha-amino group and a trifluoromethyl

group (CF<sub>3</sub>) replacing the OH of the carboxyl group, giving -COCF<sub>3</sub>. APF binds to the active site of chymotrypsin where a peptide substrate would bind. The phenylalanine sidechain of APF binds to the specificity pocket, and the CO group binds at exactly the same place as the CO of the peptide bond that will undergo cleavage. After APF binds, it reacts with the active site serine (ser 195), with the side chain oxygen of the serine forming a covalent bond with the CO carbon of APF, just as it does with the CO of a peptide substrate.

Delete the water. Set the Style Scheme to CPK Spacefill. Color the APF green and display it as wireframe 100 (don't forget to turn the spacefill off). You should be able to see how the phenylalanine sidechain fits snugly into the specificity pocket of the enzyme. You can put the APF in spacefill as well. Turn spacefill on and off repeatedly to see the shape of the pocket and how the APF fills it. To make this more vivid, you could try **select amino** and then choosing **Surfaces/van der Waals Surface** or **Surfaces/Solvent-Accessible Surface** from the popup menu.

Type **select amino** and put the enzyme in wireframe 20 display. Now **select (195,57,102)** and put them in Spacefill. This puts the catalytic triad in spacefill in addition to the inhibitor. In order to tell the residues apart, select each in turn and color 195 magenta, 57 yellow, and 102 cyan. It should be clear how the catalytic triad residues are all in a line, with ser195 poised to attack the inhibitor CO. Displaying these residues in wireframe 80 display may enable you to better distinguish the various functional groups involved in the charge relay system.

Return the three catalytic triad residues to Spacefill. Now **select amino and not (57,102,195)** and put the rest of the protein in wireframe 50. Rotate and/or zoom until you locate the green inhibitor. You will now get a better sense of the orientation of the active site.

Question 1: Print an image of the molecule with the catalytic triad in Spacefill.

Question 2: One of the members of the catalytic triad is buried, one is on the surface, and one is on the surface in the free enzyme but buried when the inhibitor binds. Which residue corresponds to each description?

Remember that it is the burial of asp102 that makes its carboxyl group an unusually good proton acceptor (abnormally high pK<sub>a</sub>), an essential feature of the charge relay mechanism.

The amide NH groups of Gly193 and Ser195 form the "oxyanion hole" in which both NH groups form H-bonds to the negatively charged oxygen of the tetrahedral intermediate. Restrict the view to the inhibitor plus these two residues, and color everything in CPK colors. Note the proximity of the NH groups to the carbonyl oxygen

of the inhibitor. (The hydrogens are not displayed, but you should be able to figure out where they would be located.)

Select residues 16 and 194 and put them in Spacefill display. Notice how the positively charged  $\alpha$ -amino group of ile16 forms a salt bridge with the carboxyl sidechain of asp194. This  $\alpha$ -amino group is produced by the cleavage of the peptide bond between residues 15 and 16 during the conversion of chymotrypsinogen to chymotrypsin, and the formation of the salt bridge involving this amino group is necessary to maintain the specificity pocket in an open state so that aromatic side chains can bind. Notice that the specificity pocket is not far from this ile16-asp194 salt bridge.

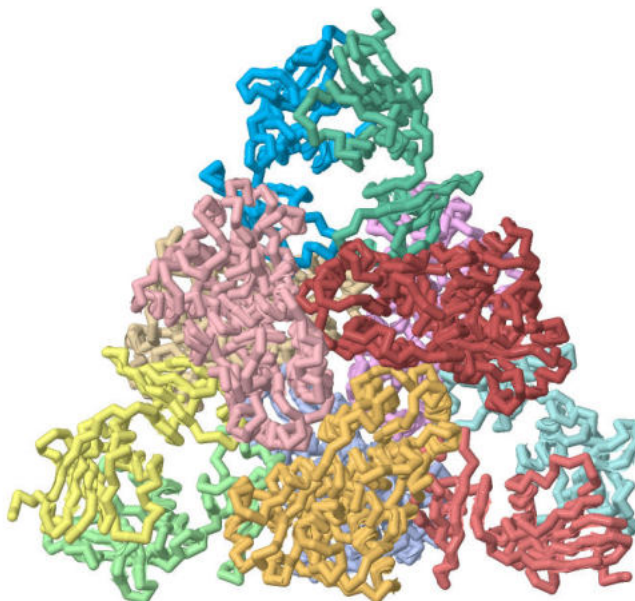
Question 3: Describe the spatial relationship of residues 16 and 194 to the substrate binding site.

Examine the structure of chymotrypsinogen (1CHG). Locate ile16 and asp194.

Question 4: What do you observe? How can you explain the difference between the locations of these amino acid residues in chymotrypsinogen and chymotrypsin?

For more on chymotrypsin, see the appendix on "Adding Hydrogens to a Structure."

## EXERCISE 13: ASPARTATE TRANSCARBAMOYLASE



In this lesson you will learn the following Jmol skill:

- How to visualize the region of a protein in the vicinity of a bound ligand

You will gain the following experience:

- Looking at the alternative conformations of an allosteric enzyme

Aspartate transcarbamoylase (ATCase) catalyzes the reaction of carbamyl phosphate with aspartate, the first reaction in the biosynthesis of pyrimidine nucleotides. The reaction is subject to inhibition by CTP, which binds to regulatory sites on the enzyme distant from the active sites. Like hemoglobin, ATCase exhibits cooperative substrate binding, and like hemoglobin it exists in both a low affinity T state and a high-affinity R state. You will be provided with instructions on what to look for but not with details of all the Jmol commands needed to do so. But if you have done the previous exercises, you should be able to do this one without much difficulty.

Prerequisites: You must have completed Exercises 1, 5A and 10 before attempting this exercise.

1RAA is the PDB file for the T state of the enzyme bound to the allosteric inhibitor CTP; 1AT1 is the file for the R state bound to the bisubstrate analog phosphonacetamide (PCT). We know that the enzyme contains six catalytic and six regulatory polypeptide chains, for a total of 12 chains, but the PDB files have only information for four of the 12 chains, since the Asymmetric Unit contains two catalytic chains (C chains), one from each trimer, along with the two regulatory chains (R chains) that link them.

As before (Exercise 10) there are two ways to download the Biological Assembly files. YOU MUST USE ONE OF THE DOWNLOAD METHODS DESCRIBED IN EXERCISE 10; if you simply download the PDB Format (gz) file or use the Get PDB command, the images will not be correct.

In this case the two methods yield slightly different results:

1) Download the Biological Assembly files from the PDB (Exercise 10, page 59, First Method) – **Don't forget to include the command "model all" (see Exercise 10).** In this case all three catalytic chains in one dimer will be designated A and all three catalytic chains in the other dimer will be designated C. (The regulatory chains in the top half will be designated B and the others D). When you color by chain, all three catalytic chains in a single trimer will be the same color, etc. If you want to select a single catalytic chain, the command format is **select :A/1** (compare Exercise 10, page 59).

2) Download the Biological Assembly files from PISA (recommended): From the results choose the structure in set #1 that has 12 subunits. In this case the protein chains and bound ligands all have separate chain designations (and separate colors). However, there doesn't seem to be a listing of the assignments that you can consult. In this case, since we know how the enzyme is organized, we can figure it out by using the method described at the top of page 60.

For 1RAA, chains A through C and D through F are the catalytic chains in the two trimers. G and H are one regulatory chain and its bound CTP, while I and J are the other regulatory chain and its bound CTP in the same regulatory dimer. (K-L-M-N are the second dimer and O-P-Q-R the third dimer.)

For 1AT1, A is a catalytic chain, B its bound CTP, and C the adjacent regulatory chain. The other such combinations are DEF and GHI on the same catalytic trimer and PQR, STU, and VWX on the opposite trimer.

Download the Biological Unit files of the R and T states of ATCase by one of these methods and explore the structures. Useful images can be obtained by displaying the molecules in backbone 300 mode. Spacefill is also very useful to help you see the general shape of the enzyme in the two conformational states.

You might want to use different display modes for the two types of chain to help distinguish them as you explore the molecule.

Question 1: Print a structure for either the R state or the T state that clearly shows the spatial relationship among the different polypeptide chains. Be sure that the structure you see in the Jmol window is the complete enzyme with twelve polypeptide chains.

Question 2: Examine the side views of the two structures and describe any differences that you see. Describe a rotational movement that could allow one structure to convert to the other one.

Note that both structures contain zinc ions. Also, as noted above, 1RAA has a bound CTP, while 1AT1 contains the bisubstrate analog phosphonacetamide (PCT). So you should be able to locate the zinc ions, the active sites, and the allosteric regulatory sites.

Locate the zinc ions. Display everything as wireframe 80. Then **restrict within (8.0,zinc)** eliminates everything more than 8.0 Å from the zinc atom. Finally render the zincs in Spacefill. Zoom in.

Question 3: What are the ligands for the zinc ions? What type of geometry (tetrahedral, cubic, octahedral, etc.) is observed for the zincs? Is the zinc bound to the catalytic or the regulatory chains?

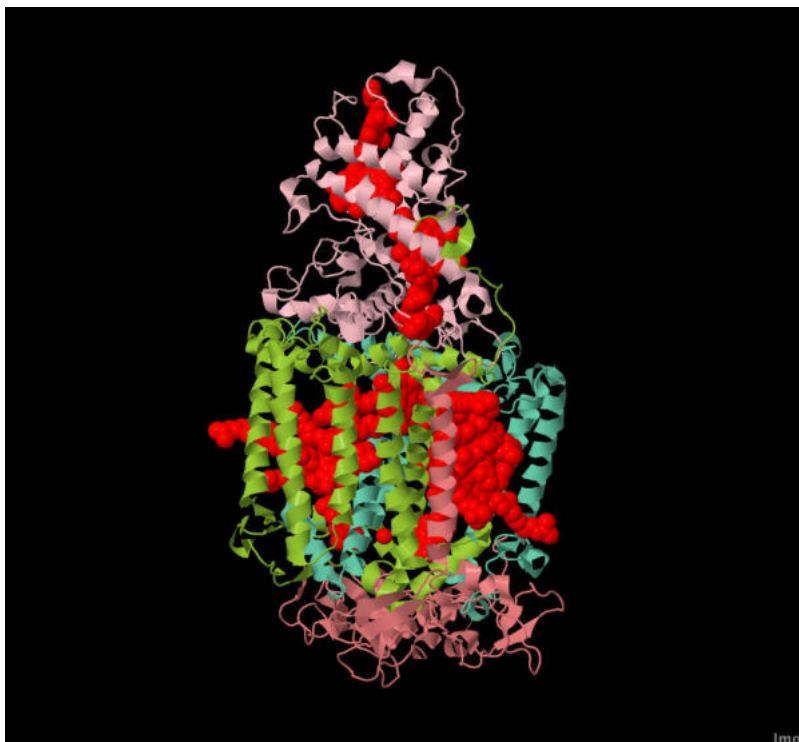
Question 4: Examine one of the regulatory chains closely. Describe the domain structure of the regulatory chains and the location of each domain within the enzyme, its interactions with other chains and its role in zinc binding.

Question 5: Describe the location of the CTP molecules in the T state enzyme.

In the R state enzyme, set the polypeptide to wireframe 80 display and chain color. Now set the substrate molecules in spacefill and color them magenta. By rotating the molecule you should be able to convince yourself that the active sites of the enzyme face the central cavity.



## EXERCISE 14: MEMBRANE PROTEINS



In this exercise you will gain the following Jmol experience:

- Examining the structural arrangement of several examples of membrane proteins, including some with multiple polypeptide chains

The folding principles for integral membrane proteins are quite different from those for globular proteins. In this exercise you will explore a few examples of membrane proteins, paying close attention to the distinctive distribution of polar and nonpolar amino acids and the relation of the protein structure to its location with respect to the lipid bilayer.

Prerequisites: You must have completed Exercises 1, 5A and 10 before attempting this exercise.

#### BACTERIORHODOPSIN:

Display the asymmetric unit of bacteriorhodopsin (1QHJ) in Jmol. Delete the water. The protein is surrounded by several molecules of 1,2-[di-2,6,10,14-tetramethyl-hexadecan-16-oxy]-propane (PH1) that were used in the crystallization. Get rid of them by typing **delete ph1**. There is also one molecule of retinal.

Color by Group. Bacteriorhodopsin is a transmembrane protein with seven transmembrane helices. Visually trace the chain from the N-terminus to the C-terminus. Also select the retinal and put it in Spacefill display.

Now **select amino**, color gray and display the molecule as wireframe 80. Select the hydrophobic sidechains and color them yellow. Color the polar sidechains green. Put the hydrophobic sidechains in spacefill. Now **undo** and put the polar amino acids in spacefill.

Question 1: What did you observe? Is the distribution of polar and nonpolar amino acids similar to or different from that of globular proteins? Why should the distribution be different for membrane proteins?

Question 2: Print an image of the molecule to support your answer to Question 1.

Bacteriorhodopsin exists in membranes as an oligomeric structure. You can display this structure and select the individual chains by following the instructions in Exercise 10 for downloading the Biological Assembly file.

Question 3: Describe the quaternary structure of bacteriorhodopsin.

PORIN:

A different type of membrane protein structure is exemplified by the bacterial porin, OMP.

Question 4: How is the secondary structure of porins different from that of bacteriorhodopsin?

Examine the location of the polar and nonpolar amino acids in porin.

Question 5: Which type of amino acid is found lining the central pore?

Porin exists in membranes as an oligomeric structure. You can display this structure and select the individual chains by following the instructions in Exercise 10 for downloading the Biological Assembly file.

Question 6: Describe the quaternary structure of porin.

PHOTOSYNTHETIC REACTION CENTER:

This is the first membrane protein to have its 3-D structure determined (Deisenhofer et al. (1995) *J. Mol. Biol.* **246**, 429-457). It is a bacterial photosynthetic reaction center. We will study the function of this molecule in the exercise on photosynthetic reaction centers. We examine it here because it is an example of a membrane protein containing more than one type of polypeptide chain.

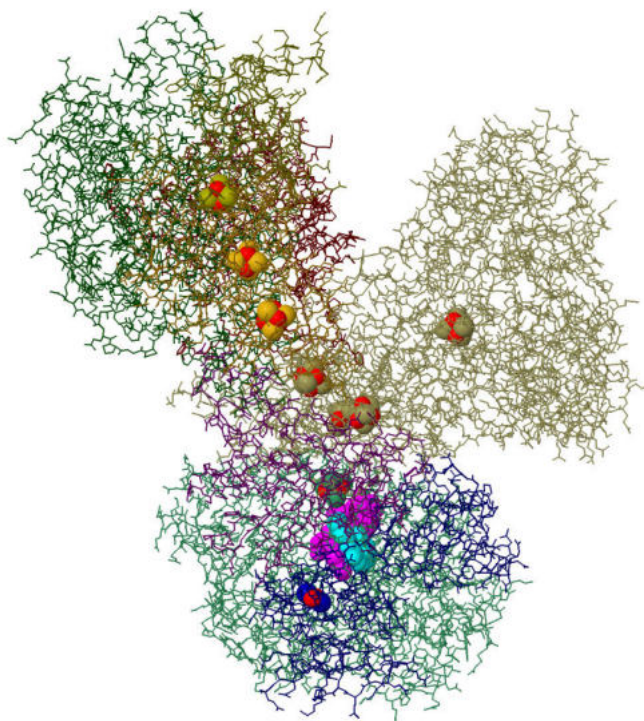
Display the structure 1PRC in Jmol, and delete the water. It contains four polypeptide chains, labeled C (cytochrome), L (light), M (medium), and H (heavy). This complex structure also includes a number of molecules that transfer electrons during photosynthesis -- two chlorophyll b's (BCB), two pheophytins (BPB), an iron ion (FE), a heme (HEM), two quinones (MQ7 and UQ1) – as well as other ions and detergents used in the crystallization (LDA, SO4 and NS1), plus some water.

Make the background white. Color by chain. Select all the small molecules (**select not amino**); put them in spacefill display and color them red. Get rid of both the water and some cosolutes present in the crystal by typing **delete (water,lda,ns1,so4)**. Select the amino acids (**select amino**) and display as **cartoon** only. Orient the complex so that the pink subunit is pointing up and the copper-colored subunit down. Turn on the rotation by typing **rotate** (type **rotate off** to stop). Look at the complex with the protein in both spacefill and cartoon display. The dark green and light green subunits are the L and M subunits, which consist mostly of transmembrane helices. The pink subunit is a cytochrome, a peripheral membrane protein that does not extend into the bilayer but just interacts with the L and M subunits. The copper-colored H subunit lies mostly outside the lipid bilayer but has one helix in the bilayer along with the L and M subunits.

To convince yourself of the location of the lipid bilayer, do the following: With the protein in spacefill display, type **select amino; color gray**, making the polypeptide chain gray but leaving the ligands red. Now color the hydrophobic amino acids green and the polar amino acids magenta.

Question 7: What do you observe about the distribution of hydrophobic and polar amino acids on the surface of the protein? What can you conclude about the location of the lipid bilayer? (After you figure this out, you can check your answer by examining the Transmembrane View of the protein.)

## EXERCISE 15: COMPLEXES OF THE ELECTRON TRANSPORT SYSTEM



In this exercise you will gain the following Jmol experience:

- Examining the structures of the respiratory complexes of the electron transport system
- Examining the spatial arrangement of the redox carriers within those complexes and learn the path of electrons as they pass through each complex

The electron transport system transfers electrons from NADH or succinate to molecular oxygen and uses the energy released in highly exergonic redox reactions to pump protons across the inner mitochondrial membrane. It is organized as a series of transmembrane complexes. The complete structures of all the complexes of the electron transport system have been elucidated -- complexes I (NADH dehydrogenase), II (succinate dehydrogenase), III (cytochrome  $bc_1$  or ubiquinol cytochrome c oxidoreductase), IV (cytochrome oxidase), and V (ATP synthase). For complexes I through IV, examining the spatial arrangement of the redox carriers will help you understand the pathway of electrons as they pass through each complex, the mechanisms by which the complexes pump protons, and other details of the electron transport process.

Prerequisites: You must have completed Exercises 1, 5A and 10 before attempting this exercise.

### COMPLEX I (NADH DEHYDROGENASE):

This huge molecule was the last of the respiratory complexes to have its structure determined. The first complete structure of the complex I (4HEA) was the complex from the bacterium *Thermus thermophilus* (Baradaran *et al.* (2013) *Nature* **494**, 443-450). It contains six transmembrane polypeptides, as well as a hydrophilic domain consisting of seven polypeptide chains that extends into the mitochondrial matrix. The hydrophilic domain contains the redox carriers, while the transmembrane domain is responsible for proton pumping.

This 16-subunit complex is a simpler version of the 45-subunit complex found in mitochondria, but the 14 "core" subunits are highly conserved in organisms ranging from bacteria to humans, and the three-dimensional structures of the cores are similar in bacteria and eucaryotes. See Further Exploration FE1 for a look at one of these larger complexes.

The structure of the hydrophilic domain of the bacterial complex was determined in 2009 (Berrisford and Sazanov (2009) *J Biol Chem* **284**, 29773-83). The structure was solved for both the reduced form, which has NADH bound and the oxidized form, which has no bound dinucleotide coenzyme. (NAD<sup>+</sup> is bound very weakly, so that when NADH is oxidized to NAD<sup>+</sup>, it immediately dissociates.) The PDB ID of the reduced form is 3IAM.

We'll look at the reduced hydrophilic domain with bound NADH first; download the coordinates of Biological Assembly 1 (see Exercise 10). Open it in Jmol. Set the Style Scheme to Wireframe and color by chain. Select the redox centers (**select ligand**) and put them in spacefill. Put the protein (**select protein**) in spacefill to get an idea of the arrangement of the polypeptide chains, then **undo** to return to thin wireframe.

There is one NADH, one FMN, a number of Fe-S centers, several magnesium ions, and one calcium. Get rid of the ions (**delete (calcium,magnesium)**). Electrons pass from NADH to FMN and then through the chain of Fe-S centers and eventually to ubiquinone bound in a cleft between chains 4 and 6. Click on the molecule to locate chains 4 and 6, and orient the complex so that you can see the cleft. You should be able to see the electron pathway clearly.

Question 1: How many iron-sulfur centers are present? How many Fe<sub>2</sub>S<sub>2</sub> and how many Fe<sub>4</sub>S<sub>4</sub>?

Look at the PDB page to find out the abbreviations for the various types of electron carriers – NADH (1,4-dihyronicotinamide adenine dinucleotide), FMN and the two types of iron-sulfur centers. Select the FMN and NADH separately and color them different colors. Now select NADH and color it by the CPK scheme. Locate the

nicotinamide ring. Now put the NADH back in one color and the FMN in CPK so you can locate the isoalloxazine ring of FMN.

Question 2: Describe the orientation of the nicotinamide ring with respect to the FMN molecule.

Finally **select amino** and set to Backbone display (with Wireframe off) to get a better view of how the redox carriers are situated within the protein.

Question 3: Describe the sequence of carries leading from FMN to the bound ubiquinone.

Now let's take a look at the entire complex. Download the Biological Assembly 1 file for the entry 4HEA, open it in Jmol and color by chain. Note the orientation of the transmembrane and hydrophilic domains with respect to each other. The three largest transmembrane subunits, the blue-green, yellow-green, and magenta subunits (starting from the end farthest from the hydrophilic domain) have sequences that allow their identification as members of a family of membrane antiporters and are presumably the proteins involved in proton translocation.

The authors of the structure paper proposed a mechanism for the coupling of proton translocation to the passage of electrons from NADH to ubiquinone. Reduction of ubiquinone at its binding site at the top of the copper-colored subunit (the transmembrane subunit adjacent to the hydrophilic domain) results in the formation of an anionic  $Q^{2-}$  species. This results in changes in the ionization states of nearby sidechains, triggering a chain of ionization changes along a network of H-bonded residues that form a hydrophilic channel that eventually runs through the transmembrane domain. These ionization changes are accompanied by conformational changes. Note that the carboxyl end of the final blue-green membrane subunit has a very long amphipathic helical extension that extends along the surface of the protein and terminates in a transmembrane helix that abuts the magenta subunit. It is possible that a conformational change in the magenta subunit causes a shift in the interface with this transmembrane helix, causing a piston-like motion of the long helical extension that helps to transmit this conformational change to the other transmembrane subunits.

With quinone in the oxidized form, each transmembrane subunit has an open half-channel that takes up a proton from the matrix. Reduction of the quinone and the subsequent conformational change closes this half-channel and opens a half-channel that releases a proton to the cytosol. As a result, passage of a pair of electrons through the complex results in translocation of three protons. Further experiments will determine whether this model is correct. The more recently reported structures of eukaryotic complex I are consistent with this model.

COMPLEX II (SUCCINATE DEHYDROGENASE):

Display the structure 1YQ3 in Jmol. Color by Chain. Delete the water, as well as the cosolutes BHG and GOL. In addition to bound water, the crystal structure contains many molecules of an "unknown ligand," so get rid of that too (**delete unl** (that's lower case UNL)). Then **select ligand** and set to spacefill display.

The enzyme contains four polypeptide chains, organized into a transmembrane region with membrane-spanning helices and a globular region that extends into the matrix. Orient the molecule so that the membrane-spanning helices are at the bottom and the globular matrix portion is at the top. Select the protein and color by chain, then click on the chains to identify them.

Question 4: Which polypeptide chains comprise the transmembrane portion and which chains comprise the matrix portion?

To see the sequence of electron transfers, let's hide the polypeptide chain (**restrict not amino**) and just look at the ligands.

The transmembrane region contains a phospholipid molecule, phosphatidyl ethanolamine. Can you identify it by looking at the structures? It is designated PEE (this is listed on the PDB page). Get rid of it by typing **restrict not pee**.

Select the malate-like intermediate (TEO) and color it green. This compound competitively inhibits the enzyme by occupying the succinate binding site. Which non-protein molecule is adjacent to this inhibitor? (Click on it to identify it.) The structure also contains several redox carriers that participate in the reaction. Identify the various ligands by clicking on them. (If necessary, consult the list of ligand identifiers on the PDB page.) Color each one a different color to help you distinguish them. Now **select amino** and display in **cartoon 50** mode to help you see the location of the redox carriers in the membrane.

Question 5:

- a) Print an image of your color-coded ligands and identify them in a figure legend.
- b) Propose a pathway for the sequence of carriers in the transfer of electrons from succinate to ubiquinone. Explain briefly how the location of the various carriers on matrix or transmembrane subunits is consistent with the organization of the electron transport system.

Before you finish, **select amino** and restore to spacefill display to see once again how the electron carriers fit into the polypeptide complex.

COMPLEX III (UBIQUINOL CYTOCHROME C OXIDOREDUCTASE):

We will be looking at the structure 3H1I, complex III from chicken. It is not the highest resolution structure available, but it contains two different antibiotics bound, making it easy for us to distinguish the two quinone binding sites. Before you get started, review the Q cycle in your biochemistry text. If you find this difficult to understand, don't be discouraged -- it's not an easy concept.

Display the structure 3H1I. Color by chain and put the ligands in spacefill.

The complex has a dimeric organization. Each "monomer" has 10 different polypeptide chains, labeled A through J in one "monomer" and N through W in the second "monomer." You should be able to recognize the transmembrane helices that are in contact with the lipid bilayer. The large globular section is the portion in the matrix, and the opposite end extends into the intermembrane space.

It will be easier to explore this protein if we look at just one monomeric unit. Type **define monomer :A,:B,:C,:D,:E,:F,:G,:H,:I,:J**), then **restrict monomer**. During the course of this exercise, if you accidentally display all 20 chains, you can correct the situation by typing **restrict monomer**.

You can visualize the A chain individually by the command **restrict :A**, then **undo** to restore the complete monomer.

Question 6: Examine each chain individually. List the polypeptide chains, and indicate whether each is in the bilayer, the matrix, the intermembrane space, or a combination of the above, and list the number of transmembrane helices. (Notice that several of the chains contain a single transmembrane helix.)

Now we'll look at the location of the electron carriers. There are several lipid and solvent molecules in the crystal structure that we need to get rid of – they are listed on the PDB page. Type **delete water,cdl,gol,pee,plc,unl**. Make sure that the ligands are still displayed in spacefill.

Note the following: The C chain is the cytochrome b subunit. It contains the antibiotics antimycin (ANY), which binds to the Q<sub>N</sub> site (negative side of the membrane), and stigmatellin (SMA), which binds to the Q<sub>P</sub> site (positive side of the membrane). [Note that some textbooks refer to the Q<sub>N</sub> and Q<sub>P</sub> sites as Q<sub>i</sub> and Q<sub>o</sub> (inner and outer).] So to locate the two quinone binding sites, just color the antimycin magenta and the stigmatellin purple. The complex also contains two cytochrome-like hemes (HEM) which relay the electrons from the Q<sub>P</sub> site to the Q<sub>N</sub> site. These are cytochrome b<sub>562</sub> or b<sub>H</sub> (closer to the matrix side) and b<sub>566</sub> or b<sub>L</sub> (closer to the intermembrane space). Select the hemes (**select hem**) and color them red to make them easier to distinguish. The D chain



is the cytochrome  $c_1$  subunit, which carries the  $c$  type heme (HEC). The E chain binds an iron-sulfur center.

Does this arrangement correspond to the model for the Q scheme? Consult the discussion of the Q scheme in your biochemistry text. You should see that Q bound to the  $Q_P$  site near the intermembrane space passes an electron to cytochrome  $b_L$  which then goes to  $b_H$  and then to the  $Q_N$  site near the matrix. You should be able to trace this pathway easily on the structure.

Alternatively, Q bound to the  $Q_P$  site passes an electron to the iron-sulfur center and from there to cytochrome  $c_1$ . Select the FeS center and the heme  $c$  and color them colors of your choice. Note that the FeS center is very far from the  $Q_o$  site, much too far to accept an electron! However, in the dimeric complex, the other FeS center from the second monomer is actually positioned between the  $Q_N$  site and the  $c_1$  heme. To see this, display the other monomeric unit. Type **select ligand and not monomer** and display in spacefill. Color all the ligands from the second monomer green. You should be able to see the FeS center from the second monomer positioned between the  $Q_P$  site and the  $c_1$  heme on the first monomer.

#### COMPLEX IV (CYTOCHROME C OXIDASE):

Download the Biological Assembly 1 file for 2OCC, using one of the methods described in Exercise 10 (DO NOT USE THE GET PDB COMMAND), and display the structure in Jmol. Display the protein as a cartoon colored by polypeptide chains and everything else (**not amino** – there are no water molecules in this particular structure) in spacefill and colored red. This huge complex has 13 polypeptide chains. If you display the polypeptides in spacefill, you can see that it is a flat, densely packed structure. Orient it so that the transmembrane helices are roughly vertical.

Restrict the view to the ligands. Choose a different color for each (suggestion: NA yellow, CU green, MG magenta, ZN orange, PER (peroxide) cyan, heme a (**select hea515 and not iron**) blue, heme  $a_3$  (**select hea516 and not iron**) purple, and IRON red). Zoom in on one half of the structure. Click on an atom in the center of the image and type **center selected**. Electrons pass from cytochrome  $c$  to the pair of green copper atoms known as  $Cu_A$  to the blue heme a to the purple heme  $a_3$  to  $Cu_B$ .

The molecule is shown in its fully oxidized state with peroxide bound.

Question 7: What are the two ligands that directly coordinate the peroxide substrate?

Now put the polypeptide chains (**select amino**) in Backbone display. How do protons reach the active site to reduce oxygen? And how are protons pumped through the complex? It has been proposed that hydrogen-bonded networks of amino acid

sidechains are involved, with different networks for oxygen reduction and for proton pumping. For example, type **define Hchannel :A and (51,441,440,371,38,382,386,390)**, then **select Hchannel**, put it in spacefill and color CPK. Also select heme a (see the last paragraph) and put it in spacefill and color CPK. You will see a network that actually traverses the protein, starting from the bottom of an aqueous channel on the matrix side and leading through a series of amino acid residues as well as side chains of the porphyrin ring to the cytoplasmic side. There is evidence that an H<sup>+</sup> binds at one end of this network, causing H-bonds to shift by a relay mechanism like the charge relay in chymotrypsin, resulting in release of an H<sup>+</sup> on the other side. Two other H-bonded networks (the D and K channels) seem to be involved in relaying protons from the matrix to the active site. Mechanisms have been proposed by which changes in the redox state of the oxygens bound at the active site cause motions of critical amino acid side chains. This movement of the side chains then causes the opening or closing of one or more of the proton channels. This results in a gating mechanism that allows the pumping of a proton only when an electron reaches the active site. For more details see Brzezinski and Gennis (2008) *Bioenerg Biomembr* **40**, 521-531.

#### COMPLEX V (ATP SYNTHASE):

The PDB ID is 1BMF. The ligands are MG, ADP, and ANP (a nonhydrolyzable analog of ATP). By now you should know how to explore the molecule. Try Spacefill, Backbone and Cartoon display. Look at both top and side views. One particularly informative view is with chains A through F in Backbone display (better on a black background) and chain G (the gamma subunit) in Cartoon or Spacefill. The asymmetrical gamma subunit within the six-fold symmetrical alpha and beta subunits is clearly seen. Another interesting view is obtained from the side view by putting the entire molecule in Spacefill and typing **slab on; slab 55; rotate**. You get a cross-section of the molecule sliced just above the middle. For more information on the Slab command, click on the **Help** button, then choose **slab** in the window that opens.

### FURTHER EXPLORATIONS

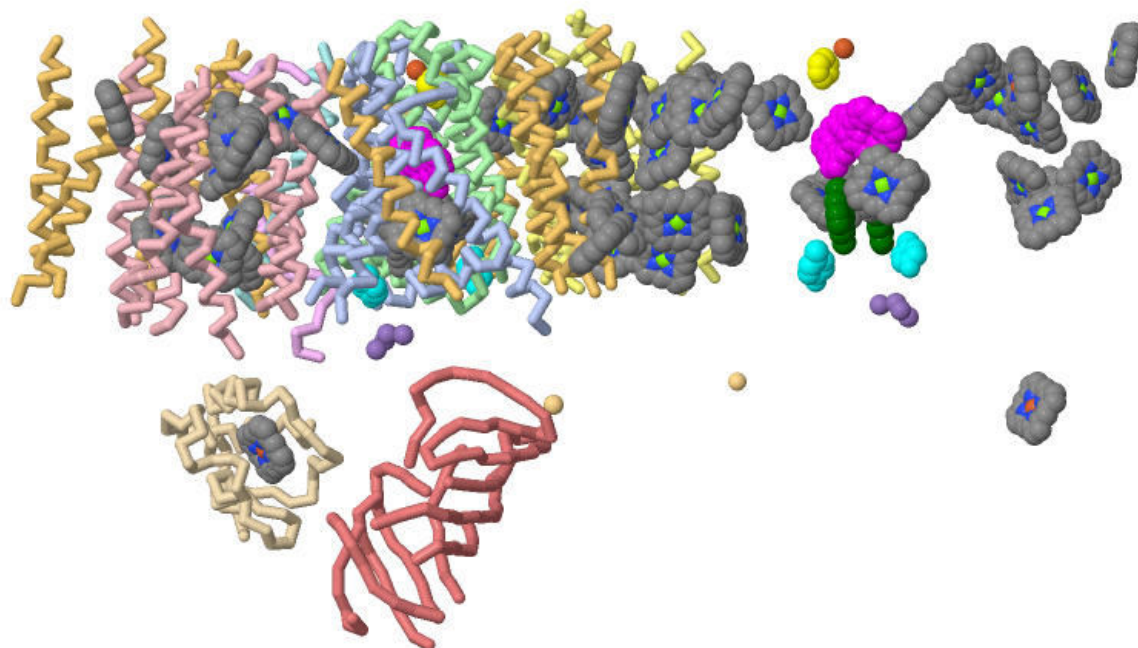
#### **FE1: CRYO-ELECTRON MICROSCOPY OF BOVINE COMPLEX I**

Bovine complex I consists of the 14 core subunits and 31 “supernumerary subunits” that surround and stabilize the core. Cryo-electron microscopy identified two types of particles with different conformations, which are thought to be related to the resting state and the activated state in which reduction of ubiquinone has facilitated proton transfer. You can examine the active conformation in the structure 5LC5. By comparing it to the bacterial complex, it should be easy to see the added mass of the supernumerary subunits.

**FE2: THE RESPIRASOME**

There is evidence that complexes I, III and IV occur in mitochondria as a supercomplex called the respirasome. Download one or more of these structures and view them in Jmol. You should be able to identify the parts of the respirasome that correspond to each of the three respiratory complexes. Try to visualize the path of an electron as it goes from NADH through the complex to oxygen.

## EXERCISE 16: PHOTOSYNTHETIC REACTION CENTERS



In this exercise you will gain the following Jmol experience:

- Examining the structures of photosystems representative of those found in photosynthesis of bacteria and green plants
- Examining the structural arrangement of the special chlorophyll pairs at the photosystem reaction centers
- Examining the spatial arrangement of the redox carriers within the photosystems and learn the path of electrons as they pass through each complex

The photosynthetic photosystems trap light energy and use that energy to drive highly endergonic redox reactions. In each photosystem absorption of a light photon by a special pair of reaction center chlorophylls causes the chlorophylls to become a powerful electron donor. As in Exercise 15 on the mitochondrial electron transport complexes, examining the structural organization of these photosystems and the spatial organization of the electron carriers within them will help you to understand the path of electron transfer and the role of the "special pair," as well as the arrangement and function of the accessory light-harvesting pigments. You will examine the structure of the photosystem of a purple bacterium and also photosystems I and II of cyanobacteria, which are similar to those found in green plants.

Prerequisites: You must have completed Exercises 1, 5A and 10 before attempting this exercise.

BACTERIAL PHOTOSYSTEM (1PRC):

This is the first membrane protein to have its 3-D structure determined (Deisenhofer et al. (1995) *J. Mol. Biol.* **246**, 429-457). It is an example of a membrane protein containing more than one type of polypeptide chain. The first few paragraphs of the text that follows are the same as the last section of Exercise 12.

Display the structure 1PRC in Jmol, and delete the water. 1PRC contains four polypeptide chains, labeled C (cytochrome), L (light), M (medium), and H (heavy). This complex structure also includes a number of molecules that transfer electrons during photosynthesis -- two chlorophyll b's (BCB), two pheophytins (BPB), an iron ion (FE), a heme (HEM), two quinones (MQ7 and UQ1) – as well as other ions and detergents used in the crystallization (LDA, SO4 and NS1), plus some water.

Make the background white. Color by chain. Select the ligands, put them in spacefill display and color them red. Get rid of some cosolutes present in the crystal by typing **delete (lda,ns1,so4)**. Orient the complex so that the pink subunit is pointing up and the copper-colored subunit down. Turn on the rotation by typing **rotate** (type **rotate off** to stop). Look at the complex with the protein in both Spacefill and Cartoon display. The dark green and light green subunits are the L and M subunits, which consist mostly of transmembrane helices. The pink subunit is a cytochrome, a peripheral membrane protein that does not extend into the bilayer but just interacts with the L and M subunits. The copper-colored H subunit lies mostly outside the lipid bilayer but has one helix in the bilayer along with the L and M subunits.

To convince yourself of the location of the lipid bilayer, do the following: Type **select amino; color gray**, making the polypeptide chain gray but leaving the ligands red. Now color the hydrophobic amino acids green and the polar amino acids magenta.

Question 1: What do you observe about the distribution of hydrophobic and polar amino acids? What can you conclude about the location of the lipid bilayer?

Now let's look at the organization of the photosynthetic electron carriers. Type **restrict ligand** to eliminate the polypeptide chains, and make sure that they are still in **Spacefill** representation.

To distinguish the different pigments, make each a different color:

**Select hem; color red** puts the cytochrome hemes in red.

There are four closely spaced bacteriochlorophylls, numbered as residues 601 to 604. Select them one by one (e.g. **select 601**) and color them magenta, purple, orange, and pink respectively.

Select the two bacteriopheophytins (**select bpb**) and color them yellow.

Select the menaquinone (**select mq7**) and color it blue.

Select the ubiquinone (**select uq1**) and color it green.

Question 2: Print an image of the molecule showing the color-coded electron carriers in spacefill display.

Rotate the figure in various orientations to see the arrangement of the two special pair chlorophylls (602 in purple and 603 in orange).

Question 3: According to current theory, the  $\pi$  electrons of the two special pair chlorophylls form a single combined molecular orbital. Describe the spatial arrangement of these two chlorophylls. Is it consistent with this theory?

You can also see how the four hemes in the cytochrome are aligned so as to deliver an electron to the special pair. Finally you should be able to trace the path of the electrons as they go from the special pair to the bacteriopheophytin to the menaquinone to the ubiquinone.

For an interesting view of the molecule, type **select ligand and not iron** and put them in **wireframe** display.

#### PHOTOSYSTEM II (4V62 AND 1FE1):

The first crystal structure for photosystem II (1FE1) was reported in 2001 (Zouni *et al.* (2001) *Nature* **409**, 739-743). It is from the cyanobacterium *Synechococcus elongatus* (also called *Thermosynechococcus elongatus*) and is similar to photosystem II from green plants. The structure was determined to only 3.8Å resolution. The PDB file includes only the alpha carbons in helices and beta sheets, mostly in the transmembrane region, as well as the various electron carriers and metal atoms. More complete structures with all atoms have been subsequently determined to higher resolution. We'll take a look at one of the all-atom high-resolution structures first, then use the earlier file for detailed exploration of the redox carriers.

Download the Biological Assembly 1 file for 4V62 and open it in Jmol (see Exercise 10; DO NOT USE THE GET PDB COMMAND). This is a complete structure, determined in 2008 to a resolution of 2.9 Å. The filename is 4v62-assembly1.cif.

(Higher resolution structures (e.g. 5B66 from 2017, 1.85 Å) are available, but they show dimeric structures and are more difficult to work with. The file is in the mmCIF format, since for some large structures PDB-format files are not available. Information about the structure in PDB format is available by selecting **PDB-format like files (tar.gz)** from the **Download Files** dropdown menu.)

Display as **cartoon only** and color by Chain. Rotate the structure so that the transmembrane helices are in a vertical orientation. The structure contains a P680

reaction center and is composed of 20 subunits, all but three of them integral membrane proteins.

The non-protein groups are listed on the Structure Summary page and in the HET section of the PDB-format file. (There are 35 chlorophylls (CLA); two chlorophylls form the special pair and two or four others may be involved in electron transfer. The others are part of the antenna system, along with 12 molecules of beta-carotene (BCR). There are two pheophytins (PHO) and one non-heme iron atom (FE2). There is no bound quinone, but a molecule of 2,3-dimethyl-5-(3,7,11,15,19,23,27,31,35-nonamethyl-2,6,10,14,18,22,26,30,34-hexatriacantanonaenyl-2,5-cyclohexadiene-1,4-dione-2,3-dimethyl-5-solaneyl-1,4 benzoquinone (PL9) probably occupies the quinone site. In addition, the structure shows the Mn cluster of the oxygen-evolving complex (OEC). There is a cytochrome  $c_{559}$ , which is part of the transmembrane portion, and an extramembrane cytochrome  $c_{550}$ . Of course each cytochrome contains a heme (HEM). The roles of the cytochromes are not well understood.

Change the display to wireframe so we can see the pigments more easily. Let's **delete bct,dgd,lhg,img,lmt,sqd** – these are all the detergents and cosolutes present in the crystal. Select the ligands, display in Spacefill and color magenta. You can now see the antenna chlorophyll and beta-carotene molecules in addition to the heme groups in the luminal portion of the complex. Select the manganese complex (**select oec**) and color it red. You should be able to see them in the luminal region just adjacent to the bilayer. The P680 special pair and the other electron carriers are buried in the middle of the complex. If you like, you can select each of the ligands and color-code them. We'll look at them in greater detail with the 1FE1 structure. You can also turn on rotation (**rotate**) to get a better feeling for the complex.

Close 4V62 and load 1FE1 in Jmol. This structure is a dimeric complex with two copies of the photosystem and the associated proteins. Since the PDB file contains coordinates only for the alpha carbons of the protein, the protein appears as a series of dots representing the alpha carbons. To see the protein, **select protein** and display as **backbone 250**. Color by chain. Then **select not protein; spacefill**. Once again, rotate the structure in different orientations, and make sure you can distinguish the two halves of the dimer, the transmembrane and luminal portions, and the two luminal polypeptides in each dimer.

The non-protein groups in 1FE1 are slightly different from those in 4V62. Beta-carotene is absent. There are 32 chlorophylls, rather than 35. The molecule bound at the quinone site, 2-[(3-hydroxy-2-methyl-5-phosphonooxymethyl-pyridin-4-ylmethyl)-amino]-2-methyl-succinic acid, is designated PLA instead of PL9, and the manganese cluster is labeled MN instead of OEC. Also, the structure includes the redox-active tyrosines (TYR) which accept electrons from the manganese cluster. (The reason we are using

this structure is that it's easier to find the redox-active tyrosines than in the all-atom structure.)

It will be instructive to hide the polypeptide chains in one half of the complex. To do this type **restrict not (:J,:K,:L,:M,:N,:O,:P,:Q,:R)**. You should now see the polypeptide chains in only one half. If you now retype **select not protein;spacefill**, you will see the polypeptide chains plus pigments in one half of the dimeric complex with naked pigments in the other half.

That was a lot of work. Just in case you mess up, save this view as a script. (See the instructions in the appendix "Saving Scripts.")

Remember that electrons from water pass from the Mn center to a tyrosine radical and then to the special pair. Select the tyrosines (TYR), color them cyan and display them in Spacefill Using the "naked" half of the complex, zoom in and identify the residue number of the tyrosine that transfers electrons from the manganese complex by clicking on it.

Question 4: What is the residue number of the redox-active tyrosine? On which chain is it located?

Now locate the special pair of chlorophylls that accepts electrons from the tyrosine. (Look for the same spatial arrangement that you observed in the bacterial photosystem.) Select these chlorophylls by residue number (**select 175;j,177:k**) and color them green. Select the pheophytin, quinone, and iron atom (see abbreviations on the last page) and give them distinct colors.

Question 5: Look at the spatial arrangement of the redox carriers. Propose a pathway for the transfer of electrons based on the spatial arrangement of the electron carriers and your knowledge of the bacterial system. Do you think that other chlorophylls are involved in carrying electrons between P680 and the pheophytins? Why or why not?

Question 6: Describe the orientation of the other chlorophylls with respect to the photosystem itself. How is this arrangement consistent with an antenna function for these chlorophylls?

Now let's look at the various polypeptide chains. We can get information about the chains from the header of the PDB file. They are:

A and J	PsbA (D1) subunits of the reaction center
B and K	PsbD (D2) subunits of the reaction center
C,D and L,M	antenna subunits
E,F and N,O	$\alpha$ and $\beta$ subunits of cytochrome $b_{559}$
G and P	each comprises 12 small polypeptide chains



H and Q                      protein containing the Mn center (lumen)  
I and R                      cytochrome  $c_{550}$  (lumen)

We have hidden chains J through R, so only A through I are visible. A third luminal polypeptide chain is not shown in the 1FE1 structure.

Type **select :A or :B; backbone 100**. This will make the chains thinner so you can see where they are in the structure. Restore them to **backbone 250**. Now do the same thing with the C and D subunits, then with the H subunit.

Question 7: Describe the location of the A, B, C, D, and H chains with respect to the membrane and the pigments.

#### PHOTOSYSTEM I (2O01):

Load Biological Assembly 1 of 2O01, photosystem I from peas. Color by **Chain**. Orient the molecule with the transmembrane helices oriented vertically. Then **select ligand; spacefill**. Set **background white** to make everything easier to see. You can see that there are a huge number of pigment molecules in the plane of the bilayer along with the transmembrane helices of the proteins. A smaller amount of protein extends into the stroma and just a very small amount of protein is the luminal surface of the complex. Look at both side and top views to get a feeling for how the pigments and transmembrane helices are dispersed. Finally **select all** and render in **Spacefill**.

Now let's have a closer look at the pigment molecules. There are five  $\beta$ -carotene molecules (BCR), 172 chlorophyll *a* (CLA), two phylloquinones (PQN), and three iron-sulfur centers (SF4). Put the protein in **Wireframe 80** display and color by the CPK scheme. Color the CL1's green, the iron-sulfur centers blue, and the phylloquinones magenta.

The special pair chlorophylls are residues 9010 and 9011. Color one of them purple and the other cyan.

Question 8: What do you think is the function of the other 170 chlorophylls?

Now **restrict (9010, 9011, pqn, sf4)**.

Four other chlorophylls are part of the electron transport chain. Select residues 9012 and 9022, display them in **spacefill**, and color them yellow. Do the same with residues 9013 and 9023 but color them red. You should now be able to see how the electron transfers from the special pair to the ferredoxin binding site on the stromal side of the membrane.

Question 9: What path does an electron take from the reaction center to ferredoxin?

Finally **select protein**, render in **wireframe**, and color by CPK to remind yourself of how the pigments involved in electron transfer are situated in the protein complex.

### FURTHER EXPLORATIONS

**FE1:** In bacteria photosystem I has been shown to exist as an oligomeric complex. Download the Biological Assembly of 1JB0, photosystem I from *Synechococcus elongatus*. Using the instructions from Exercise 10, display the complete structure. Put the phylloquinone and iron-sulfur centers in spacefill to locate the electron transfer centers. Describe what you see.

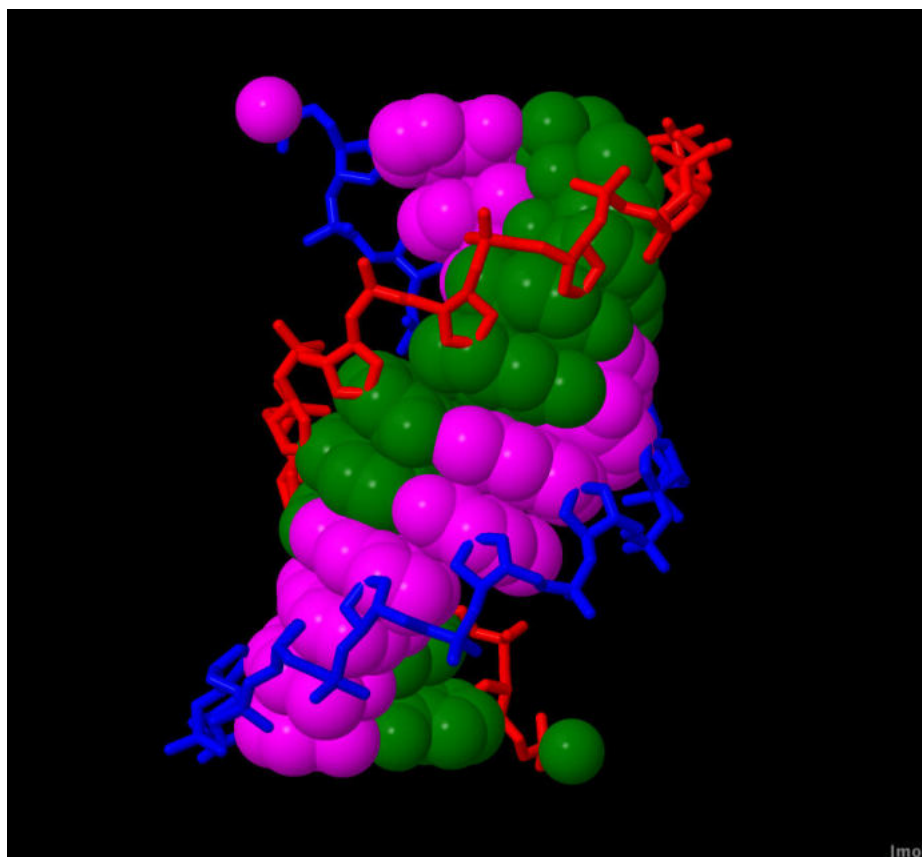
**FE2:** Structures of photosystem II from plants have recently become available. 4YUU, deposited in 2016, is the X-ray structure of photosystem II from a red alga, and 3JCU, also reported in 2016, is the cryo-EM structure of photosystem II from spinach. Download and display one of these structures. Locate and display the electron carriers.

**FE3:** Plant photosystem II exists as a supercomplex comprising two “core” photosystem assemblies associated with four copies of the light-harvesting complex LHCII. A cryo-EM structure of this complex has recently been published (Su *et al.* (2017) *Science* **357**, 615-20). The main PDB accession number is 5XNL. Examine this structure with the aid of the description in the journal article, and report what you see.

**FE4:** A 1.9 Å resolution structure of photosystem II (3WU2) revealed for the first time the detailed geometry of the Mn and O atoms in the Mn complex (Umena *et al.* (2011) *Nature* **473**, 55-60). Download the file and display it in Jmol. Restrict the view to the manganese complex. Produce a figure that shows the geometry of the Mn and O atoms and describe what you see. (Note: Recent publications have captured the Mn complex in different oxidation states. See Suga *et al.* (2017) *Nature* **543** 131-135.)

**FE5:** Crystal structures of plant cytochrome b<sub>6</sub>f complex (2ZT9), and the plant major light-harvesting complex (1RWT, 2BHW) are also available. Download and display one or more of these structures. Investigate the arrangement of the electron carriers in the cytochrome b<sub>6</sub>f complex and the arrangement of the pigments in the light-harvesting complex, and report your observations.

## EXERCISE 17: DNA



In this exercise you will learn the following Jmol skills:

- How to write Jmol commands to select the various structural elements of nucleic acids, including strands, bases, and base pairs
- How to locate, download, and display structures from the Nucleic Acids Database

You will use these skills to:

- Examine the structures of the three helical forms of DNA and learn the differences among them
- Examine the structure of DNA complexed with an intercalating dye and learn about the geometry of the complex
- Examine the structure of a nucleosome core particle and learn about the arrangement of DNA and proteins with the particle

The structure files of the PDB describe not only proteins but many nucleic acids as well, including DNA and RNA. If you know how to use Jmol with protein structures, you can easily use it with nucleic acid structures as well, as long as you learn just a few new commands that will allow you to select various parts of nucleic acid molecules.

For double-stranded DNA there are PDB structures of short pieces of DNA of defined sequence that have enabled biochemists to study the different helical forms of DNA and to understand the influence of the sequence of DNA on its local structure. In this exercise you will examine examples of the A, B, and Z forms of DNA in detail. Exploring these molecules will help you understand how these types of helix differ by such properties as their overall shapes, the appearance of the grooves, and the inclination of the base pairs. It will help you to see how the alternating structure of Z-DNA gives rise to a left-handed double helix.

In addition to the PDB itself there is a Nucleic Acids Database that has many nucleic acid structure files in PDB format. You will learn how to access the Nucleic Acids Database so that you can examine the structure of DNA with a bound molecule of the intercalating dye ethidium bromide.

Prerequisites: You must have completed Exercises 1 and 10 before attempting this exercise.

Before we get started, let's learn how to use Select commands to select parts of a nucleic acid molecule in Jmol. You can select:

nucleic or DNA	the entire molecule
backbone	the sugar-phosphate backbone
sidechain	just the bases
a	adenosine residues (base+sugar+phosphate)
c,g,t	cytidine, guanosine, thymidine residues
a and sidechain	just the adenine bases
purine [or pyrimidine]	purine residues
at [or gc]	at base pairs (with sugar+phosphate)
:B	strand B (like chains in proteins)
a and :B	all adenosine residues in strand B
a and :B and sidechain	all adenine bases in strand B
6 and :B [or 6:B]	residue 6 in strand B

**B DNA (423D):** This is the crystal structure of a dodecanucleotide. Delete both the water and magnesium ions (MG). Type **show sequence** to see the sequences of the two complementary strands. Set the Style Scheme to **Sticks**.

Rotate the molecule in the Z plane (hold down the shift key and drag horizontally) so that the axis is vertical.

Now **select backbone** and put in **spacefill**. To distinguish the two chains **select :A** and color it red, then color the B chain blue. Rotate the molecule about its axis. Now rotate

it so that you are looking at an end view. You should gain an appreciation of the relative locations of the backbone and base pairs.

After performing each of the following operations, stop and examine the molecule, rotating it to different orientations:

First **select dna; spacefill**. To distinguish the two chains, **select sidechain and :A; color green**. Then **select sidechain and :B; color magenta**.

Then **select backbone** and put in wireframe 80. This last view is especially instructive, showing how the inside of the double helix consists of a series of base pairs stacked on top of each other and in close van der Waals contact. Note that the planes of the base pairs are roughly horizontal, perpendicular to the helix axis.

Now let's look at a single base pair. Type **restrict 8 or 17**. Display in both **spacefill** and **wireframe 40** to see what a base pair looks like.

Now zoom in, keeping everything in **wireframe 40** display. Focus on one of the two bases. Rotate the structure so that you are looking at the base exactly edge on, so that the plane of the base is exactly perpendicular to the plane of the screen. Do the two bases lie in exactly the same plane? Now rotate the second base so that it is perpendicular to the screen and examine the first base. You should see that a base pair has a slight propeller twist. This was not a feature of the original Watson-Crick model. We now know that this propeller twist has important effects on DNA structure and its dependence on base sequence.

Question 1:

a) Is the propeller twist clockwise or counterclockwise? (If you look down each glycosidic bond from the sugar toward the base, is each base rotated clockwise or counterclockwise relative to a completely coplanar alignment (in which both bases lie in the same plane)?)

b) Print an image of the base pair that illustrates your answer.

Now let's take a closer look at the sugar rings. **Select backbone; color CPK**.

Question 2: What is the orientation of the plane of the sugars relative to the plane of the base pair? Relative to the helix axis? (If you have trouble answering this last question, **select backbone; wireframe 40** to see where the helix axis is.)

Now zoom in on one of the sugars. Look at the sugar ring in different orientations.

Question 3: Is the sugar-ring planar? Orient the ring so that the red oxygen is at the rear and C5 is to the left. Which atom in the furanose ring lies above the plane of the ring and which one below the plane of the ring?

Now put the entire molecule in **wireframe 40** display. Notice that the molecule has a wide groove and a narrow groove.

Question 4: Look at the wide groove. Which functional groups of an AT base pair line the surface of this groove? Which functional groups line the surface of the narrow groove? (To help you answer, type **restrict 8 or 17**, then **undo**, then **redo**, etc., to allow you to focus on one base pair while keeping track of which side is the major groove and which side the minor groove. Use CPK colors to help identify the functional groups.)

Question 5: Which functional groups of a GC base pair line the surface of the wide groove and which line the surface of the narrow groove?

The grooves are referred to as the major and minor grooves. Other forms of DNA lack the wide and narrow grooves seen in the B form. But the major groove is always defined as the groove on the side of the base pairs and the minor groove as the groove on the side of the sugars.

A DNA (1QPH): Download the Biological Assembly file for 1QPH (See Exercise 10; DO NOT USE THE GET PDB COMMAND), open it in Jmol, and type **model all**. For this structure the Asymmetric Unit is a single chain, and the Biological Assembly is a double helix.

We'll begin by color-coding the molecule. Delete the water. Set the Style Scheme to **Sticks**. Using commands similar to those used for B DNA, color the backbones red and blue and the bases green and magenta. (Note that the two chains in this structure are designated :A/1 and :A/2 instead of :A and :B.)

Now let's make sure the molecule is oriented with the helix axis vertical (it's not nearly as obvious how to do this as it was with B DNA). Rotate in the z plane (shift-click and drag horizontally) until the red and blue riboses are at the same vertical level at each end of the molecule.

One thing that should be immediately obvious is that the base pairs are strongly tilted with respect to the helix axis. Now look at a top view of the molecule (either rotate with your mouse or type **rotate x 90**; to restore the original orientation type **rotate x -90**.) Notice the channel through the middle of the molecule, which was absent in B DNA.

Rotate the molecule in different orientations with your mouse. Display in spacefill (after selecting **nucleic**) and examine again. Then put the backbone in spacefill and the

bases in wireframe 80. Then reverse, putting just the bases in spacefill. You should get the impression of a molecule shaped like a ribbon wound about a hollow core.

Using the definitions of major and minor grooves just given, locate the major and minor grooves in A DNA.

Question 6: What do you notice about the depths of the two grooves? Are their widths very different?

Click on two bases opposite each other to identify them. Restrict the display to this base pair. (Note: In this structure, each strand is labeled from 1 to 12, so your command will be of the form **restrict x:A/1 or y:A/2**, where x and y are the appropriate residue numbers.) Calculate and display the H-bonds between the bases.

Question 7: The bases in B DNA had a clockwise propeller twist. Is this also true of A DNA?

Now **select backbone** and **color cpk**. Orient the sugar as you did for B DNA.

Question 8: Describe the sugar conformation as you did in Question 3 for B DNA.

Z DNA (1DJ6): Download the Biological Assembly file for 1DJ6 and open it in Jmol. **Restrict dna**. Type **show sequence**. We are looking at the structure of the hexanucleotide (GC)<sub>3</sub> forming a double helix with itself.

Set the Style Scheme to Sticks, then **wireframe 40**. Color the backbones and bases red, blue, green, and magenta as you did with B and A DNA. Examine the molecule. Note the left-handed helix. Also, there are some strange orientations of the sugars and the bases which we will be taking a closer look at shortly.

Question 9: Locate the major and minor grooves. Which groove is deep and which one shallow?

Now color the molecule in the CPK scheme. Click on the strands to identify the A and B strands. Orient the molecule so that the 5' end of the A strand is pointing down and the 3' end pointing up. (The 5' side of each sugar has C and O between the sugar and the phosphate and the 3' side just an O atom between the sugar and the phosphate. Orient the molecule so that the 5' sides of the A strand are pointing down.)

When you're finished type **select g:A; color red; select c:A; color yellow**. This will code the A strand as alternating red cytidines and yellow guanosines. In a similar fashion, render the B strand as alternating blue cytidines and cyan guanosines. Finally highlight the A backbone by typing **select backbone and :A; wireframe 80**.

Question 10: Print an image of the molecule in this format.

Starting at the bottom, notice that from the yellow ribose the phosphodiester bond goes left to the phosphorus atom, then up to the red ribose. This is true of each phosphodiester bond – to the left and then up. So the backbone has a zigzag pattern, giving this structure the name Z DNA.

Now look at the orientation of the riboses. The red riboses have their 5' carbons pointing down and their 3' carbons pointing up. The yellow riboses are actually flipped over nearly 180 degrees, so that the 5' carbon is above the 3' carbon.

The orientation of the bases about the glycosidic bond is also different for C and G residues. Zoom in on a yellow C residue and notice that the hydrogen bonding groups of the cytosine are pointing away from the sugar, just like the bases in B and A DNA. But for the red G residues, the guanine base is rotated 180 degrees so that the 6-membered ring is over the ribose, with the hydrogen bonding groups pointing back over the sugar. The C residues are said to be in the *anti* conformation and the G residues in the *syn* conformation.

Now **select nucleic** and return to **wireframe 80** display. Type **restrict 3,10** to look at a single base pair. Now do this thought exercise: Flip the yellow ribose over so that the phosphodiester backbone is straight rather than kinked. As you're doing this, imagine that the C base turns as the ribose turns. Imagine further that the GC base pair remains intact but that the G base is allowed to rotate about the glycosidic bond connecting it to the blue ribose. What would you end up with? If you did the thought experiment correctly, you should realize that you'll end up with two riboses in the normal orientation, with both bases of the GC base pair in the *anti* orientation. In other words, you'd end up with a conformation characteristic of a right-handed double helix. So it is possible to go from left-handed to right-handed DNA without disrupting the base pairs! In fact this conversion occurs both in the test tube and in living cells.

COMPLEX BETWEEN DOUBLE-STRANDED RNA AND ETHIDIUM (DRB018): Go to the Nucleic Acids Database (<http://ndbserver.rutgers.edu>). Then enter DRB018 in the search window. At the bottom left of the next page, click on **Asymmetric Unit Coordinates (pdb format, Unix compressed (.gz))**, save the file to your hard disk and open in it Jmol.

Delete the water. Set the Style Scheme to Sticks. Color the ethidium (**select ligand**) green. Examine the complex from the side and top. Now **select nucleic** and put in **Spacefill** display. **Restrict nucleic** will make the ethidium disappear. Typing **undo** will make it appear again. Put the ethidium in spacefill and make it disappear and reappear.

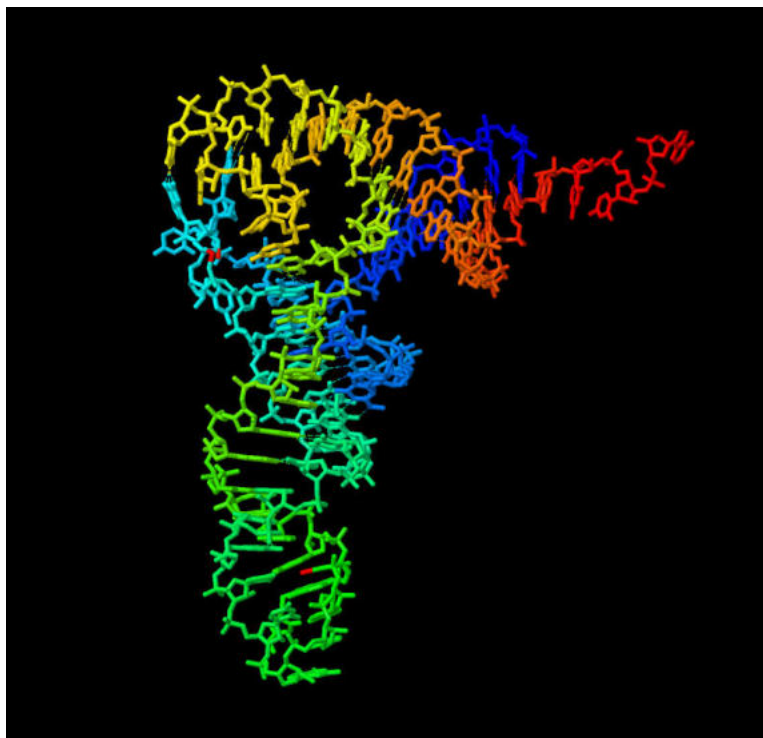


Question 11: Summarize your observations and describe how ethidium binds to nucleic acids.

#### FURTHER EXPLORATION

**FE1**: The file 1AOI is the structure of a nucleosome core particle. It consists of a histone octamer wrapped with two left-handed superhelical turns of double-stranded DNA. The chain designations of the individual histone molecules (two each of H2a, H2b, H3, and H4) are given in the PDB page. You can separately **select protein** or **select dna** and display them in different formats (try **Cartoon** for the protein) to visualize the organization of the complex. You can also select the individual histone molecules by their chain designations and examine their location in the complex. Explore the complex, report your observations and support them with Jmol images. What type of conformation does the DNA adopt? How are the eight histones arranged within the octamer? Describe the contacts of the histone molecules with each other and the contacts between the histones and the DNA, pointing out the functional groups involved in the interactions.

## EXERCISE 18: TRANSFER RNA



In this exercise you will gain the following Jmol experience:

- You will get additional experience in manipulating nucleic acids
- You will explore the structure of a transfer RNA (tRNA) molecule and gain an understanding of the relationship between the primary, secondary, and tertiary structure.
- You will visualize the various types of base pairs in tRNA and gain an understanding of how they contribute to its stability.

In 1972, biochemists took their first look at the three-dimensional structure of an RNA molecule when they determined the X-ray crystallographic structure of yeast phenylalanyl-tRNA<sup>Phe</sup>. The structure revealed for the first time that RNA can adopt a complex tertiary structure. When you encounter this structure for the first time, it is not easy to see how the L-shaped structure can form and how it is related to the cloverleaf model showing the location of the double-helical regions within the primary structure. By going through this exercise you will gain an understanding of the organization of the L-shaped structure and an understanding of how tertiary H-bonds between bases distant in the primary sequence help to stabilize this structure.

Prerequisite: You must have completed Exercise 1 before attempting this exercise. Exercise 17 is recommended.

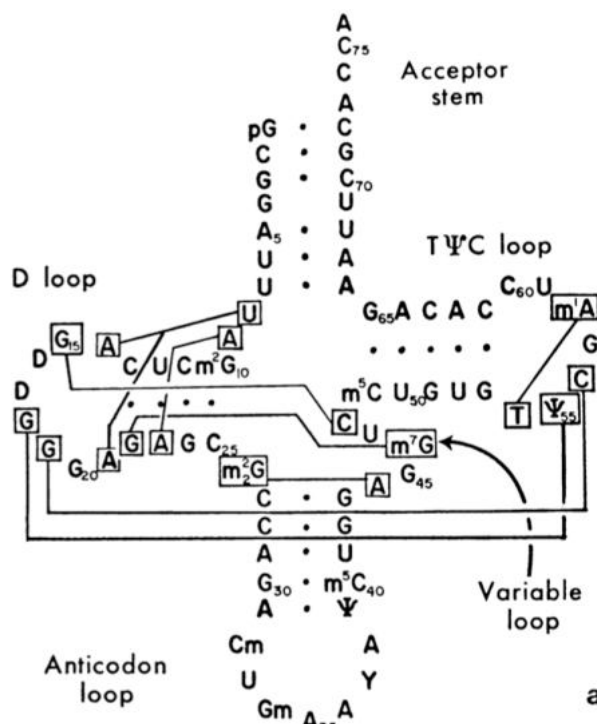
Before starting this exercise, review the rules for writing Select commands for nucleic acids at the beginning of Exercise 17.

Display the file 6TNA in Jmol. Before we get started we have to move the center of rotation to the middle of the molecule. Rotate the structure with your mouse to see the L shape. Now select **Define Center** from the view menu and click on the center of the image.

Zoom in and center the molecule in the window. Delete the water. Set the Style Scheme to **Trace** and color by **Group**. This makes the 5' end blue, with the color gradually changing to green at the 3' end. Orient the molecule so that the 5' and 3' ends are at the upper right. With a diagram of the cloverleaf secondary structure as a reference, trace the chain as it traverses the acceptor stem, D stem and loop, anticodon stem and loop, T stem and loop, and acceptor stem. Try out **Strand** and **Cartoon** also. Rotate the molecule so that you are looking down perpendicular to each of the arms of the L.

If you are having trouble seeing how different parts of the L correspond to different parts of the cloverleaf, you can select various regions, using the figure to the right as a guide, and color code them. For example, to see where the variable loop is, type **select 44-48; color red**. To see where the D stem is, type **select 10-13 or 22-25; color yellow**. Do this until you can identify the different stems and loops.

(The figure is taken from Kim et al. (1973) *Science* **185**, 435-440, by permission. The tertiary base pairs are shown by solid lines and Watson-Crick helical base pairs by dots.)



For another interesting image, put the molecule in **wireframe 80** display, then **select backbone** and display in **wireframe** display.

**Question 1:** Examine the orientation of the base pairs in the different parts of the molecule. What do you observe?

**Question 2:** What type of helix (A, B, or Z) is contained in tRNA? How can you tell?

Select the three anticodon residues (36-38) and color them red.

Question 3: How is the mRNA oriented with respect to bound tRNA?

Now let's look at some of the tertiary base pairs. Display the hydrogen bonds. The structure of tRNA is maintained not only by Watson-Crick base pairs within the helical stems, but also by hydrogen bonds involving bases from the loops as well. Watson-Crick base pairing rules are not observed in the tertiary base pairs, and unusual H-bonding patterns are observed.

We'll start with a base pair between G19 and C56. This is a Watson-Crick base pair, but the G and C are in non-helical portions of the tRNA molecule. Restore the **Group** color scheme to the tRNA. Now **select 19 or 56; color red**.

Question 4: Describe the location of the 19-56 base pair and its role in stabilizing tRNA tertiary structure.

Other tertiary base pairs present in the corner of the L include the following:

G18 forms a single hydrogen bond with  $\psi$ 55. This H-bond also connects the D loop to the T loop – you should be able to locate this base pair by selecting it and assigning a color.

T54 forms a base pair with A58. To see this most clearly **restrict 54 or 58**. Notice that the bases are oriented so that the adenine will form hydrogen bonds involving the “top” of the purine ring, including the nitrogen of the five-membered ring, rather than the “side” of the ring as in Watson-Crick hydrogen bonding. This arrangement is called a Hoogsteen base pair. Note that Jmol does not display the hydrogen bonds between bases in a Hoogsteen base pair.

A complete list of the nine tertiary base pairs in this tRNA can be compiled by noting the solid lines in the figure on the last page. They are: 18-55, 19-56, 54-58, 22-46, 26-44, 15-48, 9-23, 8-14, and 8-21.

To see the role of all the tertiary base pairs in tRNA structure, do the following: Put the molecule in **wireframe 80** display, then select the backbone and put it in **Backbone** display. Now select each pair or triplet in turn and give it a different color, *e.g.* **select 18 or 55; color orange**. Use warm colors so you can see them easily against the green-to-blue colors of the rest of the molecule. Save this view as a script so you can restore it later.

Click on the individual bases to help identify them. Notice how base pairs G19-C56, G18-ψ55, and m<sub>1</sub>A58-T54 all connect the T loop to the D loop at the corner of the L. Notice also how the single base G57 or the T loop is inserted between G18 and G19 of the D loop. The stacking interactions of these bases also help to stabilize the corner of the tRNA.

The other region where tertiary base pairs are important is in the upper part of the vertical arm of the molecule, where the D and anticodon stems come together. Bases 8 to 10 between the anticodon stem and the beginning of the D stem are involved here, as well as bases 44 to 46 in the variable loop that intervenes between the anticodon stem and the T stem.

Here you will find a series of four tertiary base pairs and triplets along with one Watson-Crick base pair, all stacked on top of each other.

The lowest of the four is the 2-dimethylG26-A44 base pair, between the two bases just above the top of the D stem in the cloverleaf structure. This base pair extends the vertical helix of the anticodon stem.

The vertical helix continues with Watson-Crick base pairs m<sup>2</sup>G10-C25, C11-G24, and U12-A23, and C13-G22 of the D stem. For three of these four helical base pairs, one of the bases also is involved in a tertiary base pair, resulting in the formation of a base triplet:

m<sup>2</sup>G10 forms a second base pair with G45 of the variable loop

A23 forms a second base pair with A9

G22 forms a second base pair with m<sup>7</sup>G46

This is a region of the molecule where four sections of the backbone are running parallel to each other, and the base triplets help to keep these sections in the proper orientation. To get a better idea of the importance of these interactions, **select nucleic** and put the molecule in **spacefill**. If you colored all the tertiary basepairs with warm colors, you should see a blue-and-green molecule with warm colors in the two regions of the molecule with the tertiary base pairs.

Let's look at one of the base triplets in detail. Type **select not (9,13,22,46)** and color by **Group**. If you have not already colored the base triplet a warm color, select these four residues and color them orange. You can see that we are looking at a planar section through the molecule in the upper part of the vertical helix.

Now **restrict (9,13,22,46); color cpk**. Zoom in on the structure. View it in both spacefill and wireframe display. You can see the GGC base triplet, but now you can also see that the amino substituents of m<sup>7</sup>G46 and C13 are in position to form hydrogen bonds to one of the phosphate oxygens of residue 9. (There are no hydrogens in the structure,

but you should realize that each blue nitrogen sticking out from the ring has a pair of hydrogens in the plane of the ring at a 120-degree angle, one of which would be pointing at the phosphate oxygen. So, all four sections of the backbone at this level of the molecule are connected by hydrogen bonds.)

Question 5: Using the figure from the first section of this exercise, select any one of the tertiary base pairs or base triplets other than 19-56. Create and print a Jmol image that shows the location of this base pair or triplet within the 3-D structure of tRNA.

Hydrogen bonds involving the sugars are also important in tRNA structure. Place the entire molecule (**select nucleic**) in **wireframe 80** display and **color green**. Then **restrict backbone**. Finally **select (6-8,48-50)** and **color cpk**. Notice that this is a portion of the molecule where the ends of the acceptor stem and T stem come together, and both sections of backbone make a sharp turn and then continue downward. If you zoom in, you can see that the 2'-OH of ribose 7 forms a hydrogen bond with the ring oxygen of sugar 49. This hydrogen bond acts as a sort of paper clip to hold these two portions of the backbone together at this point.

Hydrogen bonds important in tRNA tertiary structure include base-base, base-sugar, base-phosphate, sugar-sugar, and sugar-phosphate bonds.

## EXERCISE 19: RIBOSOMES



In this exercise you will gain the following Jmol skill:

- You will learn how to display and manipulate a structure whose coordinates are contained in a split entry involving more than one PDB file.

You will gain the following Jmol experience:

- You will examine the structure of a 70S bacterial ribosome with mRNA and tRNA's bound.
- You will gain a better appreciation of the spatial organization of the RNA and protein molecules of the ribosome
- You will gain a better appreciation of the locations of the various binding sites and the peptidyl transferase active center and the orientation of the three bound tRNAs as they participate in peptide bond formation.

The ribosomes of *E. coli* are the most complex structures determined thus far. Determining the structures of the individual ribosomal subunits in 2000 was followed in 2001 by the solution of the structure of the 70 S ribosome with bound mRNA and tRNAs.

Prerequisites: You must have completed Exercises 1 and 10 before attempting this exercise. Exercise 17 is recommended.

Before starting this exercise, review the rules for writing Select commands for nucleic acids at the beginning of Exercise 17.

#### 30S Subunit (1J5E):

Display the structure in Jmol. Set the Style Scheme to cartoon. Color the RNA green and the protein blue.

Question 1: Is the structure mostly RNA or mostly protein?

Question 2: The 30S subunit has a slab-like appearance (a fairly flat pancake with an irregular shape). Set the display to spacefill. Now rotate the ribosome about its flat axis and look at the distribution of RNA and protein on the two sides. What do you observe?

Color the RNA by the CPK scheme and the protein by Chains. Note that the proteins are all rather small. The largest ribosomal protein, protein S1, is missing from this structure.

Type **restrict rna**. Notice that the overall shape doesn't change very much. So the 16S RNA forms the central core of the 30S ribosomal subunit and determines its overall shape, while the proteins bind to this central core and help to stabilize its conformation.

#### 50S Subunit (1JJ2):

This structure has more than 90,000 atoms! Display it in Jmol and set the Style Scheme to Cartoon.

Color the 23S RNA (**select :0**) green, the 5S RNA (**select :9**) yellow, and the protein blue. Set the display to Spacefill.

Question 3: Look at the distribution of RNA and protein as you did for the 30S subunit. What do you observe?

Question 4: Describe the location of the 5S RNA.

Use your command of Jmol to explore the large subunit, as you did above with the small subunit.

70S Ribosome (4V42): Download and display the **PDBx/mmCIF Format (gz)**. Use one of the methods described on page 13.

This structure contains the 30S and 50S subunit of the ribosome from the bacterium *Thermus thermophilus* bound to a 6-nucleotide segment of mRNA and three tRNAs



bound to the A, P and E sites. It contains all the atoms of the tRNAs and mRNA, but just the alpha carbons of the proteins and the phosphorus atoms of the rRNAs.

Since this file was originally split into two separate files, one for the 30S subunit and one for the 50S subunit, in the combined PDBx file the chain designations have two letters instead of one, with those for the 30S subunit RNA and proteins and the mRNA and tRNAs beginning with the letter A and those for the 50S subunit RNAs and proteins beginning with B. They are listed on the Summary tab and can be summarized as follows:

16 S RNA	chain AA
23 S RNA	chain BA
5 S RNA	chain BB
mRNA	chain A1
A site tRNA	chain AB
P site tRNA	chain AC
E site tRNA	chain AD
small subunit proteins	AE through AX
large subunit proteins	BC through BX

Not all the ribosomal proteins are resolved in the structure.

When you open the file you will see the rRNAs and proteins in backbone display and the tRNAs and the mRNA in ball and stick display. Select the RNAs (**select rna**) and color by Chain. Select the three tRNAs and the mRNA and display them in Spacefill. Select the proteins, display them as Cartoons and color them red. You should be able to see clearly the binding of the tRNA anticodons to the mRNA within the 30S subunit and the acceptor stems where the aminoacyl groups would be surrounded by the 50S subunit. You should be able to see that the ends of the tRNAs in the A and P sites are very close to each other. This defines the peptidyl transferase site of the ribosome.

Question 5: Examine this site to see which molecules of the ribosome are present at the peptidyl transferase site. Does it contain 16S RNA? 23S RNA? 5S RNA? Ribosomal proteins? You may have to rotate the structure and look at it from different angles to be sure. Are you convinced that the ribosome is a ribozyme? Print an image that you used to arrive at your answer.

Restrict your view to the three tRNAs and the mRNA. Turn **spacefill off** to return to the wireframe display of the tRNAs. Then look at just the following combinations: A site tRNA and mRNA; P site tRNA and mRNA. View each image in different orientations.

Question 6: Which of the tRNAs interact with the mRNA? Describe as precisely as you can the orientation of the mRNA with respect to the L-shaped tRNAs.

## SUMMARY OF JMOL SCRIPT COMMANDS

A **select** command specifies a set of atoms, and then the following directions apply only to the selected set. A **restrict** command specifies a set of atoms, and then only the atoms specified are displayed. A **delete** command irreversibly removes the selected set; they can be restored only by reopening the file in a new Jmol session.

Select and restrict commands are described at the web page [www.umass.edu/microbio/rasmol/seleccmd.htm](http://www.umass.edu/microbio/rasmol/seleccmd.htm). The Select commands in the programs Chime and Rasmol are the same ones used in Jmol. You can select

everything	all
elements	e.g. carbon, sulfur, iron
types of atoms	n = NH nitrogen, o = CO oxygen, c = CO carbon, ca = alpha carbon, cb = beta carbon, etc.
specific atoms	atomno=<number>
(Both the atom type and atom number are displayed when you click on an atom.)	
residue numbers	e.g. 5, 278-283
types of amino acids	e.g. ala, ile
hetero groups	we had the example nag in lysozyme, with the abbreviations defined on the structure info page
polypeptide chains	e.g. :A, :B

alpha (meaning alpha carbons)  
amino or protein (protein only, not solvent or hetero groups)  
backbone  
sidechain

cyclic (aromatic amino acids and proline)  
acyclic  
aliphatic  
aromatic

hydrophobic or nonpolar  
polar  
charged  
neutral  
basic (or positive)  
acidic (or negative)

cystine (cysteines linked by a disulfide bond)

surface

buried

helix

sheet

turn (after doing a DSSP structure calculation)

hetero (non amino acid)

water

ions (e.g. phosphate or sulfate)

solvent (water or ions)

ligands (hetero and not solvent)

within (x.x,species) within x.x Ångstrom units of the designated species

e. g. select within (5.0,hem) within 5.0 Å of the heme

select within (4.0,:A) within 4.0 Å of chain A)

To select a named chemical component type **select** followed by the identifier, e.g. **select hem** (Exercise 11). If the name of the chemical identifier begins with a number within it, you must put it in brackets, e.g. **select [3pg]** (Exercise 5, FE1).

You can also create “Boolean expressions” combining more than one term.

not solvent

:A and lys

lys and 23

neutral and not hydrophobic

:B and 43-50

3 or 7 or 11 (can also be written 3,7,11)

(3,7,11) and :A

1-40 and hydrophobic

lys and nz and :B (sidechain nitrogens of lysines in chain b)

If you need to select the same set of atoms over and over again, you can create a name for that set of atoms, for example:

define myatoms helix and backbone and (3,13,53,186,187,204) and polar

select myatoms; spacefill

Once atoms are selected you can color them whatever color you like. Examples:

color red

color structure (different colors for different secondary structures)

color cpk

color group (cool colors at the N-terminus going to warm colors at the C-terminus)

color chain (each polypeptide chain a different color)

You can also select the color of the background by typing **background <color>**.

You can change the display mode of the selected atoms by typing **spacefill only**, **wireframe only**, **backbone only**, **ribbons only**, **strands only** or **cartoon only** or by typing simply **spacefill**, **wireframe**, **backbone**, etc. and turning the previous display off (e.g. **spacefill off**).

You can also determine the width of the bonds or the diameter of the spheres by specifying a number. For example, **wireframe 80**, **spacefill 200**, **cartoon 500**, etc.

You can do a DSSP structure calculation by typing **calculate structure**. After this is done, coloring by structure will separately code turns and pi helices, in addition to alpha helices,  $3_{10}$  helices and beta sheets.

You can turn hbonds and ssbonds on and off (**hbonds calculate** (or **hbonds on** once they have been calculated), **hbonds off**, **ssbonds on**, **ssbonds off**).  
**set hbonds backbone** and **set hbonds sidechain** determine whether H-bonds connect alpha carbons or amide and carbonyl groups.  
**set ssbonds backbone** and **set ssbonds sidechain** determine whether disulfide bonds connect alpha carbon atoms or cysteine sulfur atoms  
**ssbonds 20** or **hbonds 50** controls the thickness of the bonds.

You can manipulate the molecule with the mouse (a summary of mouse commands is given in Exercise 1), but you can also do this with a typed command, such as:

**zoom 100** (a number between 10 and 1000)

**translate x 20** (a number between -100 and 100) and **translate y -10** move the molecule in the x and y directions

**rotate x 90** rotates the molecule 90 degrees about the x axis (**rotate y** and **rotate z** commands can also be used)

**show info** shows the information for the molecule

**show sequence** displays the amino acid sequence

**show selected** lists the atoms that are currently selected

**show structure** displays information about the secondary structure of the protein

**undo** reverses the last command; **redo** reverses the undo command

#### SOME ADVANCED COMMANDS

To measure the distance between two atoms, type **set picking distance**. Double-click successively on each of the atoms. The identity of the atoms and the distance between

them will be displayed in the Command window and on the structure. When you are finished, type **set picking on**.

To measure the angle between three atoms, type **set picking angle**. Double-click on the first atom, click on the second (central) atom, and double-click on the third atom. When you are finished, type **set picking on**.

Typing **slab on; slab 50** will slice the molecule through the middle and display the surface. You can vary the depth at which the slice is taken from 0 to 100.

For instructions on use of a command, click on the Help button on the scripts console and scroll down to that command.

For more information on Jmol commands, you can consult the Jmol scripting documentation at <http://jmol.sourceforge.net/docs>, the Jmol scripting documentation at <http://chemapps.stolaf.edu/jmol/docs/>, or the RasMol manual at [www.bernstein-plus-sons.com/software/rasmol/doc/rasmol.html](http://www.bernstein-plus-sons.com/software/rasmol/doc/rasmol.html). There is also a Jmol wiki at <http://wiki.jmol.org>, that has links to lots of sites with information about how to use Jmol.

#### COPYING AND PASTING

To copy and paste text within the Script Console use the keystrokes Control-C and Control-V. Note that these keystrokes are used on Macs as well as PCs, even though the usual keystrokes for Copy and Paste on a Mac are Command-C and Command-V.

## PRINTING AND SAVING AN IMAGE

There are a few different ways of printing Jmol images.

1. Direct printing: Choose **Print** from the Jmol **File** menu. The disadvantage of this method is that the image is not saved and is lost as soon as you quit Jmol.
2. Copying the image and pasting it into another application: Choose **Copy Image** from the **Edit** menu. Then open another application (like Microsoft Word) and choose **Paste**. (Alternatively you can save the file and insert it later – see paragraph 3.)
3. Exporting a file and saving it as a graphics file: From the **File** menu choose **Export/Export Image**, assign a filename, and save as a .jpg file. The image can then be opened in a graphics program and printed. Alternatively, the image can be inserted into a Microsoft Word or PowerPoint document by using the Insert/Picture/From File... command from the Insert menu. Finally the image can be attached to an e-mail or uploaded to a web site that accepts student assignments.
4. Screen capture: On a PC hold down the Shift key and press the Print Screen button (to the right of the function keys) to copy a screen image onto the clipboard which can then be pasted into an application like MS Word, or saved as a file that can be attached to an e-mail or uploaded to a web site. In Mac System X, launch the application Grab (from Applications/Utilities) and click on Capture/Screen.

Note: Generally you will want to use a white background for a printed image, because printing an image with a black or colored background uses LOTS of ink.

## SAVING SCRIPTS

Want to save a state of the protein (including both view formats and labels) to open it later in Jmol and manipulate it further? From the Jmol **File** menu, choose **Export/Write State**. A **Save** window will open. Give the file a name and save it as an .spt (script) file (by setting **Files of Type:** to **SPT (\*.spt)**). The suffix .spt will not be part of the filename.) Make sure that you save it to your Jmol folder (the folder that contains the jmol.jar file).

How can you restore that state of the protein? Let's assume that you named the file **myview**. Launch Jmol and type **script myview**. If you are in the middle of a complicated project and don't have time to finish, save your work as a script file, and pick up later where you left off.

You should be able to copy the script file to a flash drive and open it on a different computer. First copy the file into the folder on the computer that contains the Jmol executable jar file. Then launch Jmol, select the **Open** command from the **File** menu, and open the script file. This way you can generate the saved state of your protein on the other computer. This is handy if you are working on a computer that is not connected to a printer, or if you don't have a color printer at home and want to print out an image on a color printer somewhere else.

## CREATING LABELS

In order to label an amino acid residue, select an atom where you want the label to be, and note the atom number. For example, if clicking on the atom brings up the line:  
[PHE]137:G CE2 #1043 -34.397 28.993 20.914, then the atom number is 1043, and you would type **select atomno=1043**.

To turn on the label, type **label %n %r** (%n means 3-letter residue name and %r means residue number – to see other options, click the Help button and go to **label**.)

The default color for the label is white. To get different colors use the command **color label <color>**. To turn the label off, type **label off**.

If you want to change the text of the label, just type **label <text>**.

The font size of the lettering can be controlled with the command **set fontsize <number>** (e.g. **set fontsize 30**).

## ADDING HYDROGENS TO A STRUCTURE

Here's one final trick that may come in handy. PDB files generally do not contain coordinates for hydrogen atoms. Sometimes it's useful to be able to visualize hydrogen atoms, for example when looking at a closeup of the active site of an enzyme. Jmol is capable of adding hydrogens to a structure.

In order to do this the following command must be typed before you open a structure file: **set pdbAddHydrogens**. The Script Console will respond with **pdb AddHydrogens = true**.

As an example, you can try taking another look at the active site of chymotrypsin. Type the above command and then download and open the file 1YPF. Now **restrict (57,102,195)**. Since the structure contains two chymotrypsin molecules, type **restrict :C,:E** to display only one of the two. Set the style scheme to Sticks. You will now have a view of the charge relay system that includes the hydrogens. The program has added hydrogen atoms to the N-H nitrogens, the histidine imidazole and the serine hydroxyl.

Note that Jmol adds hydrogens to both of the imidazole nitrogens of histidines to produce the protonated form of histidine that exists at acidic pH. (At higher pH, only one or the other of the imidazole nitrogens is protonated.) A hydrogen has been added to the serine side chain oxygen in a conformation in which the new hydrogen is trans to the alpha carbon of the serine. This is not necessarily the actual conformation, as there is free rotation about the C-O bond. Nevertheless, you can see how the three residues form a hydrogen-bonded network that facilitates the charge relay that is important in catalysis.

Now load the file 6GCH. Type **restrict (57,102,193,195)**. Set the style scheme to Sticks. You will now have a view of the substrate, the charge relay system and the oxyanion hole that includes the hydrogens. Select the inhibitor (**select apf**) and put it in Spacefill. In this structure the oxygen of the serine hydroxyl has reacted with the carbonyl oxygen of the inhibitor APF, forming a covalent bond. Consequently, the hydrogen previously attached to that oxygen is not present. You can still see the protonated histidine. In addition you can now clearly see the role of the NH groups of residues 193 and 195 in forming H-bonds to the carbonyl oxygen of the inhibitor. This illustrates how the oxyanion hole is involved in binding the substrate and the transition state.

If you now want to load a different structure file in the normal way, without adding hydrogens, first type the command **set pdbAddHydrogens false**.



## WORKING WITH NMR STRUCTURE FILES

Nearly all of the structures we have examined were determined by X-ray crystallography. Another technique used by biochemists to determine the three-dimensional structure of macromolecules is NMR (nuclear magnetic resonance). The information contained in an NMR structure file is different from that contained in a file derived from X-ray crystallography, and there are some important things to know if you are to use these files correctly.

The basic concepts underlying the determination of a protein structure from NMR data are simple. The signals corresponding to individual amide hydrogens in a protein can be identified and information can be obtained about which amino acid residues are close in space, as well as some information about the values of the various dihedral angles in the protein.

This information forms a set of constraints for a molecular dynamics analysis. A computer starts out with a model of the protein in a completely random conformation. The computer then subjects the protein model to random fluctuations in conformation and calculates the energy of the structure after each fluctuation. When the protein settles into a conformation that satisfies the distance constraints from the NMR data and also appears to be at a minimum potential energy, the computer produces a three-dimensional structure. This is repeated 20 to 30 times. Since each molecular dynamics run starts from a slightly different random conformation, a slightly different structure is produced each time. We end up with an ensemble of structures.

This information can be deposited in the PDB in three forms. One type of file simply contains all the NMR structures. Another type of file contains something called the minimized average structure, which is produced by obtaining an average position for each atom and then adjusting the structure to minimize the potential energy (e.g. by avoiding unfavorable steric interactions). A third type of file is a representative structure, usually the single structure that is closest to the average. (Many biochemists think that a representative structure is more likely to be an accurate representation of the protein than an average structure, and newer NMR structure files are often representative structures.)

As an example, examine the structure of 1FO7. This is an ensemble of 30 NMR structure models for a fragment of the bovine prion protein. When you open the file in Jmol, it automatically displays model number 1. You can view other models by typing **model 2** or **model 3**, etc. To see all the models displayed, type **model all**. Note that in some regions the different models are almost perfectly superimposed, while in some of the nonhelical regions the superimposition is not as good. 1FKC is the minimized average structure of the same protein. However, go to the EMBL-EBI site [www.ebi.ac.uk/pdbe-apps/nmr/main.html](http://www.ebi.ac.uk/pdbe-apps/nmr/main.html), and click on **Olderado**. On the next page,

enter 1FO7 in the window and click on **View Report**. The window that opens next, under **NMR Resources**, tells us that structure 14 is the most representative structure. In the opinion of many if not most biochemists, model 14 is the best one to work with. You can view this structure in Jmol by typing **model 14**. Note that PDB file 1FKC is the minimized average structure of the same protein. [Note, March 2023: The URL shown above for the Olderado database is not working, and I have not been able to locate a replacement.]

Another thing to keep in mind is that an NMR structure, unlike an X-ray structure, contains all the hydrogen atoms. For a display that looks more like what you are used to, go to the **Display** menu of the Jmol window and uncheck the box **Hydrogens** (see the portion of Exercise 4 dealing with the helix-turn-helix motif). If you want to restore the hydrogens, simply uncheck the box. To irreversibly remove the hydrogens, you can type **delete hydrogen**.

## JSMOL, A WEB BROWSER BASED VERSION OF JMOL

The lessons in this manual use the standalone version of Jmol. They can also be done with JSmol, a version of Jmol that runs in a browser window.

When you download Jmol as described in the first section of this manual, the downloaded folder contains JSmol as well as Jmol.

One way to access JSmol is through the site [chemapps.stolaf.edu/jmol/jsmol/simple.htm](http://chemapps.stolaf.edu/jmol/jsmol/simple.htm). Click on **Load PDB by ID**. Enter the PDB ID in the window that open, and click **OK**. An image of the molecule (asymmetric unit) will appear. If you want to display the biological assembly, check the box next to **biomolecule 1** before clicking on **Load PDB by ID**. (See Exercise 10 for an explanation of asymmetric units and biological assemblies.) You can type script commands into the window next to the **console** button.

## JMOL WEB PAGES

There are many web pages that have prescribed animated lessons that use Jmol. For example, go to

<http://bioinformatics.org/jmol-tutorials/jtat/hemoglobin/contents/contents.htm>.

You can go through this lesson by clicking on the links and buttons in the text. Also, for any image you can bring up the Jmol pop-up menu by right-clicking (Control-clicking), and you can change the representation of the molecule using the menu commands.

There are also many Jmol lessons available on the internet. There are several excellent places to find these web sites. The site molviZ.org provides links to excellent Jmol lessons on protein structure, hemoglobin, antibodies, and lipid bilayers. The Online Macromolecular Museum

([https://earth.callutheran.edu/Academic\\_Programs/Departments/BioDev/omm/exhibits.htm](https://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/exhibits.htm)) has some excellent tutorials on basic protein and nucleic acid structure, as well as tutorials on enzymes involved in replication and transcription and many other biochemical processes. There are also links from the CSULB Biochemistry Links web page ([www.csulb.edu/~cohlberg/biochemlinks.html](http://www.csulb.edu/~cohlberg/biochemlinks.html)) under Molecular Modeling/Images. The Jmol wiki (<http://wiki.jmol.org>) also has a link to **Websites Using J(S)mol**.

## Term Paper Assignment

Summary of the Assignment: Pick any protein other than the ones that were the subjects of the lessons in this manual, and write a short paper on the protein illustrated with figures which you prepared by using Jmol. The paper should have as a major focus the relationship between the structure of the protein and its function. The main body of the paper should be 1-1/2 to 3 pages long (double-spaced, 12 point font, 1 inch margins). In addition you should have a list of references, plus at least three Jmol figures, each with a title and a brief legend, which illustrate the points in your text.

This is a suggested format for the assignment. Your own instructor may choose to modify these instructions or those that follow.

How can I pick a protein? Try looking ahead in your text to see what interesting proteins we'll be studying (for example, some of the enzymes of carbohydrate metabolism). Maybe you've heard about some interesting proteins in a cell biology course (a polymerase, a viral protein, a growth factor, etc.). Look them up in the PDB and see if their structures are known. Or look in your cell biology textbook. Or consult the journals *BioEssays* or *Trends in Biochemical Sciences*, both of which have short review articles. Or follow the link to **Molecule of the Month** from the PDB web site. Or browse the PDB-101 site (<http://pdb101.rcsb.org/>).

How can I locate the right PDB file for my protein? Review Exercise 7, "Searching for PDB Files." After doing an initial search, don't just choose the first file in the list. Try to obtain a file that contains the entire structure, not just a fragment or a single domain. Use the instructions in Exercise 7 to locate a file that contains bound ligands that are relevant to the function of the protein.

After you have chosen a protein and verified that a structure is available, submit your choice to your instructor. There are thousands of known protein structures, so it should be possible for everyone to write about a different protein. If two people want to do the same protein, your instructor may try to suggest alternatives.

What should I write about? You can choose various things to write about. You can say something about the general secondary structure organization. If you have a multi-domain protein, you can talk about the different domains and their different functions. If it's an enzyme, you can talk about how the amino acid sidechains at the active site participate in the binding and catalysis. If it's not an enzyme but a protein that binds something else, you can talk about the amino acid sidechains involved in the binding reaction. If the protein undergoes a conformational change, you can say something about the nature of the change at a molecular level. You don't have to cover everything that's known about the structure and function of the protein in this short paper, as long as you include material on at least some aspect of the structural basis of the protein's

function at an atomic level. Consider this as something with a level of detail somewhere between an illustrated show-and-tell exercise and a more traditional term paper.

How should I organize the paper? Different types of organization are possible. You will probably want to start with an explanation of what the protein is, what it does, and its significance in the cell or the body. Then you might describe the overall structure of the protein, finally zeroing in on a particular aspect of the protein which you will illustrate with your Jmol figures. Your discussion should refer to the figures as you go along. e.g. "As you can see in Figure 2, lysine 85 (shown in red) forms a salt bridge with...." Every figure in your paper should be referenced at some point in your text.

Where can I get more information about the protein? Consult the PDB page for a citation of the original publication describing the protein structure. If your protein is an enzyme, try the BRENDA database (look under the Metabolism heading in CSULB [www.brenda-enzymes.org](http://www.brenda-enzymes.org)) or go to Pubmed ([pubmed.gov](http://pubmed.gov)). Put your protein name in the search field. When the search results come back, look at the Reviews tab. This will list only review articles on your protein and will prevent you from getting long lists of primary research articles.

Read the PDB page and the header for the pdb file – both have important information. If you haven't already gone through Exercise 8, "What's in a File?" do it now. It has valuable instructions on how to get information about your protein from these documents.

Examine the diagram in the Molecular Description box of the PDB page. It often contains valuable information about the primary structure of the protein and its posttranscriptional processing, the location of modified amino acids, the secondary structure and the locations of active site residues. (See the section of Exercise 1 on the immunoglobulin binding domain.)

Is the structure file derived from X-ray crystallography or NMR, or is it a theoretical model? If it's an NMR structure, you MUST read the appendix "Working With NMR Structure Files."

What should be in the figure titles and legends? The title of the figure should be a short phrase or sentence indicating what is shown or what the main point of the figure is. The legend, which goes into this in greater detail, could be as short as one sentence or longer if necessary. It should point out the features that are being illustrated and explain any color schemes or labels whose meaning is not obvious. At least one figure should identify the accession number of the protein. If you did anything unusual with Jmol or with another computer program to help prepare the figure, this should also be

explained in the legend. The legend may be printed on the same page as the figure that it refers to, or you may have your legends printed on separate pages.

How should I cite my sources? If you use only one source, it's adequate to list it in your reference list. If you use multiple sources, you should identify which material came from which reference. The easiest way to do this is to number your references and then put a number in parentheses in the text at the end of a sentence to indicate the source of the information in the preceding sentences.

How should I list the references? Please use the following formats. (Don't change these formats; we want you to demonstrate that you are capable of following directions.)

journal article -- Mucklefutz, Q., and Gesundheit, A. C. (1997) "The secret of life as revealed in the molecular structure of cookie dough," *J. Biol. Chem.* 458, 134-142

book – Darwin, C. (1999) "Evolution for Dummies: A Guide for the Kansas Board of Education," Neanderthal Press, Wichita, pages 115-164

article from a book -- Winken, R., Blinken, L. M., and Nod, U. U. (1992) "Structure of lactate dehydrogenase from the roadrunner *Beepbeep californicus*," in Simon, P., and Garfunkel, A. (editors) "Collected Essays in Comparative Biochemistry", Wiley Coyote Press, Phoenix, 1998, pages 59-87

web site -- [www.crazyguys.csulb.edu/toocool/proteins](http://www.crazyguys.csulb.edu/toocool/proteins), "Famous Proteins We Have Known" (give the web address and the title of the web page)

Can I copy material word-for-word from a reference? NO. Copying word-for-word (or copying entire passages almost word-for-word with only slight paraphrasing) is plagiarism. Please put your ideas in your own words. If you must quote from a paper, put the quote in quotation marks. If you simply quote verbatim from a reference, we have no way of knowing whether you have any real understanding of the stuff you've copied. If you don't think that you're capable of putting things in your own words, see your instructor for help in getting started. If we suspect that you've plagiarized, we will take time to look at your original references and will penalize you heavily if we find that you have copied the original language.

This is my first biochemistry course and I haven't taken any cell biology. I'm not very good with computers. I really don't know what to do and I'm frankly very intimidated by this assignment. What can I do to keep from totally freaking? Hey, chill out!

First of all, if you need help with any aspect of this assignment, please come and see your instructor. If you've done some looking but just can't figure out what protein to choose, the instructor can make some suggestions. If you've chosen a protein but don't know where to find the information you need, the instructor can point you in the right direction. If you found a discussion of your protein that you can't understand, bring it in and look at it together with your instructor. If you've done your research but can't figure out what to focus on, let your instructor help you. If you've gone through the

Jmol exercises but can't figure out how to manipulate the images the way you want to, you may be able to sit down at the computer with your instructor and figure it out together. (Of course this means that you can't start the paper two days before it is due.)

Second, we don't expect everyone to hand in a masterpiece. We understand that this will be difficult for some of you, but if you demonstrate that you put in a reasonable effort and were able to make some nice pictures and explain some things reasonably well, you'll get a good grade. You'll have to earn your A, but there should be a lot of B's on this assignment, and no one will get a D or F who puts forth any reasonable effort. You can write as little as two pages, so you are not expected to do the amount of work for this project that you'd do for most term papers.

Common mistakes that students make when doing this assignment that prevent them from getting a favorable grade:

- Some PDB files represent an individual domain of a protein rather than the entire protein. Read the PDB page to make sure you know what it represents.
- In some X-ray structures there are missing amino acid residues that are not resolved in the X-ray data and are therefore omitted from the structure. You need to be aware of this to understand what you are looking at. This information is contained in the header of the PDB file.
- If your structure has bound ligands, be sure to include them in your figures and comment on their significance.
- In many cases you have a choice of a structure with bound ligands and a structure with the protein alone. Choose one with bound ligands and comment on them. Don't take the first structure that pops up on the page of search results.
- Find out whether the asymmetric unit and the biological assembly are the same. If they are different, use the biological assembly for your report (See Exercise 10 for instructions).
- If your structure has lots of water molecules, eliminate them from your figures.
- If a file is based on an NMR structure, see if an X-ray structure exists. A file based on an X-ray structure will be easier to work with. If only an NMR structure is available, you MUST follow the instructions in "Working with NMR Files" in the Appendix and select a single model for your analysis, and you must explain how you selected the model to work with.



-- The text of your discussion should focus on the structure of your protein as illustrated by your figures. Don't just generate some figures and then simply discuss the biological function of the molecule without referring to the figures.

-- Use the **label** commands (see "Creating Labels" in the Appendix) rather than adding handwritten labels.

Authenticating your paper: In order to authenticate your paper, do the following:

1. Form an image in the Jmol main window similar to one of your figures.
2. In the script console window, type your name and the date.
3. Capture an image of the screen and include a printout with your paper. On a PC hold down the Shift key and press the Print Screen key (the key just to the right of the function keys) to copy a screen image onto the clipboard which can then be pasted into an application like MS Word. In Mac OS X, launch the application Grab (from Applications/Utilities) and click on Capture/Screen.