Introduction to Jmol

Jmol is the usual standard for visualization of molecular models. It recognizes, among others, structural files in the **.pdb** (*Protein Data Bank*) format. Jmol is a Java application, and its use requires the last Java updates to be installed in your system. Jmol is a freeware program and can be freely distributed.

A complete description of Jmol utilities can be found in the following web addresses:

http://www.jmol.org/

http://jmol.sourceforge.net/

http://biomodel.uah.es/Jmol/ (*in Spanish*)

http://wiki.jmol.org:81/index.php/Main_Page

Elementary use of Jmol

1. Opening the program

The Java executable file for Jmol is

Jmol_files_English\Jmol_files**Jmol.jar**

Upon opening, there appears a black window with the usual format of Windows applications. There are menus called **Archivo (*File*), Editar (*Edit*), Estilo (*Style*), Vista (*View*), Herramientas (*Tools*), Macros**. This black window is the Display Window.

First, we change the language. With the pointer on the black window, display with the mouse right button the Main Menu. Choose **Idioma** and then **Inglés (en_US)**. The language then changes to English. Note, however, that the names of folders and files referred to appear in Spanish.

There also appears another window (white background). This is the Command Window. If it does not appear, display with the mouse right button the Main Menu over the Display Window. Choose the option **Console**. The Command Window should then appear.

2. Opening a file

In the display window, go the **File** menu and go the option **Open**. There appears a dialog box. Go the folder **Demo_Jmol/Aminoacidos/estructuras**. In this folder are the structural files for the 20 protein aminoacids (identified with lower-case one-letter

code), among others. Open the file called **m.pdb** that contains the structure of the aminoacid Methionine (Met, M). In the display window will appear a ball-and-stick model of the aminoacid.

.pdb files are the standard form of storing structural information for molecules. They consist of text files that specify the X, Y and Z coordinates for all the atoms present in the molecule.

Note that the amino group appears as its protonated form ($-NH_3^+$) and the carboxyl group as its dissociated form ($-COO^-$). This the usual way of representing aminoacids (that is, its *zwitterionic* form)

3. Display forms

In the display window, open the main menu with the secondary button of the mouse. Go to submenu **Style**, and then to the option **Scheme**. By default, the molecule appears as **Ball and Stick**. Test other display forms as follows:

CPK Spacefill: (**Style > Scheme > CPK Spacefill)** The molecule appears with the atoms represented at their Van der Waals radii. This kind of displaying is the closest to the real form of the molecule. Atoms appear with the color code called CPK code (Corey-Pauling-Kultun): Carbon in grey, Hydrogen in white, Oxygen in red, Nitrogen in blue, Sulfur in yellow and so on.

Ball and Stick: It is the default option of the program. It uses the same color code, and we can visualize the bonds between atoms. Atoms do not appear at its real size (as in the **CPK Spacefill** option) but it is the most convenient form of looking at the structure of low molecular weight molecules.

Sticks: We only see the interatomic bonds as sticks and conserving the CPK color code.

Wireframe: The interatomic bonds appear as wires (as in the Dreiding type models used in Organic Chemistry) and conserving the CPK color code.

Options **Cartoon** and **Trace** only function in macromolecules (Proteins and Nucleic acids)

Go the **Ball and Stick** option to continue with the rest of the demo.

4. Rotation

The molecule can rotate over the X (horizontal) and the Y (vertical) axes by **dragging the molecule over the display window with the left button of the mouse**.

Rotation over the Z axis can be done **dragging with the right button of the mouse while maintaining down the Shift key**.

Precise rotations can be introduced in the command window following the prompt **$**:

$ rotate x 32.5 [Enter]

This command rotates the molecule 32.5 degrees clockwise over the X axis. The command

$ rotate z -67 [Enter]

will rotate the molecule 67 degrees counterclockwise over the Z axis.

5. Zoom

We can zoom on the molecule **dragging with the mouse left button while maintaining down the Shift key**.

Precise values of zoom can be achieved with the command window. The default value of zoom is 100. Test several values of zoom, for example

$ zoom 50 [Enter]

$ zoom 200 [Enter]

Go back to zoom 100.

6. Translation

We can translate the molecule **dragging with the right mouse button while maintaining down the Ctrl key**.

7. Measurements

Jmol allows to make some measurements to the model, which give an idea on the real dimensions of the molecule. Go the display **Ball and Stick** at zoom 100.

7.1 Distances

Let us measure the distance between the sulfur atom and the nitrogen atom. On the main menu (mouse right button over the display window) choose **Measurement > Click for distance measurement**. Click on the sulfur atom (yellow) and then click on the nitrogen atom (blue). A straight line appears joining both atoms and labelled **0.458 nm** (4.58 angstrom).

Practice with distance measurements.

In the Main Menu choose **Measurement > Delete measurements**

7.2 Angles

To measure the angle formed by the sulfur atom and its neighboring carbon atoms, choose **Measurement > Click for angle measurement.** Click on the terminal methyl carbon, on the sulfur atom and then on the next carbon. The value of the angle appears in the display window, **105.0º**.

Practice with other angle measurements on the same molecule.

In the Main Menu choose **Measurement > Delete measurements**

7.3 Torsion Angles

If we have a chain of four consecutive atoms, a Torsion Angle is the dihedral angle between the plane determined by the first three atoms and the plane determined by the last three atoms. In other words, the torsion angle is the torsion around the bond linking the second and the third atoms in the chain. Torsion angles are important in determining Secondary Structure of proteins.

To measure torsion angles in the Main Menu choose **Measurement > Click for torsion (dihedral) measurement**. Click on four consecutive atoms of the chain to get the measurement.

8. Atom recognition

In the Main Menu choose **Set picking > Identity**

Taking the pointer on any atom of the model, it will appear a small label indicating the nature of the atom pointed to. If you click on the sulfur atom, in the command window will appear the information

S #4 -2.093 0.89599997 0.049999997

Saying that is a sulfur atom (S), that is atom #4 in the structure file and its three spatial coordinates X (-2.093), Y (0.89599997) and Z (0.049999997). We'll see that in protein models we can get much more information.

10. Other molecules

Practice with molecules other than methionine. These are the folders in which structural files are stored (Remember that their names are in Spanish):

/Demo_Jmol/Estructuras_organicas/estructuras	Basic organic compounds
/Demo_Jmol/Hidratos_de_Carbono/estructuras	Carbohydrates
/Demo_Jmol/Lipidos/estructuras	Lipids
/Demo_Jmol/Aminoacidos/estructuras	Aminoacids and Oligopeptides
/Demo_Jmol/Proteinas/estructuras	Proteins
/Demo_Jmol/Acidos_Nucleicos/estructuras	Nucleic Acids
/Demo_Jmol/varios	Coenzymes, drugs, hormones, etc.

Study of proteins with Jmol

Jmol was designed to study macromolecules. The program has many more possibilities when applied to proteins and nucleic acids. We'll start with a very small protein, **Rubredoxin** (53 residues).

Rubredoxin is an electron-transport protein that belongs to a family of proteins containing iron and sulfur. They are also known as Ferredoxins or NHI proteins (Non-Heme Iron proteins). Iron appears coordinate to four sulfur atoms of cysteine residues.

Rubredoxin is the simplest ferredoxin, and it is found in Bacteria. All living organisms have ferredoxins in both mitochondrial and photosynthetic electron transport systems.

1. Opening a file

The Java executable file for Jmol is

Jmol_files_English\Jmol_files**Jmol.jar**

In the menu **File**, choose **Open**. Go to the folder **Demo_Jmol/Proteinas/estructuras** and select **rubredoxina.pdb**. The protein (Rubredoxin, a redox transport protein) appears in the display window in the default mode of Ball and Sticks. Observe that it is usual in proteins not to represent hydrogen atoms.

Around the structure appear many isolated red atoms (oxygen). These are water molecules trapped into the crystal used to determine the structure of rubredoxin by X-ray diffraction. To delete them, type in the command window the following command:

$ restrict not solvent [Enter]

Note that the command window only understands English!

We can get a lot of information about the structural file in the following way: In the main menu, choose **Show > File header.** In the command window will appear the first lines of the file (a protein file may contain thousands of lines) telling the protein name, the biological source, the aminoacid sequence and basic crystallographic data.

COMPND	Compound; a description of the model
SOURCE	Biological source; in this case, the archaeon *Pyrococcus furiosus*
AUTHOR	Author(s) of the structural file
REVDAT	Date(s) of model revision by the authors
JRNL	Journal of publication
REMARK	General data on structure determination, mainly crystallographic

SEQRES	Aminoacid sequence of the protein
HET	Hetero (non-protein) molecules in the structural file
HELIX	Residues forming α-helix secondary structure
SHEET	Residues forming β pleated sheets
TURN	Residues forming β turns

2. Display modes

In the main menu the option (**Style > Scheme > CPK Spacefill**) gives the spacefill representation with atoms at their van der Waals radius, as in small molecules. The same can be said about other display forms: **Ball and Sticks, Sticks** and **Wireframe**.

In macromolecules, there are two further options:

Style > Scheme > Cartoon: This is the ribbon model of protein structure. We are looking at the backbone of the protein (side chains are not represented). Ribbons represent secondary-structured zones. Arrows point to the C-terminus. α-Helices appear in red, β-sheets in yellow and β-turns in blue. Zones without secondary structure appear in white or grey.

Style > Scheme > Trace: Like the Cartoon representation but the secondary-structured zones do not appear as ribbons. The whole protein backbone appears as a thread extending from the N-terminus to the C-terminus.

3. Colors

Go to the command window (Main menu **Console**)

- Type the command

$ color group [Enter]

It appears a color pattern from blue to red going through the rainbow sequence (blue – cyan – green – yellow – orange – red). This sequence indicates the relative position with respect to the termini: N-terminus is blue, and C-terminus red.

- The command

$ color chain [Enter]

Gives a different color for each subunit in the protein. This does not function in rubredoxin because this protein has no quaternary structure.

- The command

$ color temperature [Enter]

Gives the values of temperature anysotropic factors; that means that the “coldest” zones of the molecule appear in blue and the “hottest” in red. Cold atoms are those that are not

subject to thermal agitation. Red atoms are those whose position is more uncertain due to thermal agitation.

- The command

$ color structure [Enter]

Displays the usual colors of structure code (red for α-helices, yellow for β-sheets and blue for β-turns)

Go back to the Ball and Stick display with the main menu **Style** > **Scheme** > **Ball and Stick** and to the default color

$ color cpk [Enter]

4. Atoms and Residues recognition

As in small molecules, Jmol can recognize atoms and aminoacid residues.

First we reset the protein to its initial position with the command

$ reset [Enter]

Click the red atom appearing in the lowest part of the molecule. In the command window will appear the following information:

[GLU]31.OE1 #238 9.875999 -10.771 5.058

That means the following:

- The atom belongs to the residue Glu 31

- The atom is in the ε1 (E1) position of the aminoacid. Remember that in organic acids we call α to the carbon immediate to carboxyl group, β to the next, and the others with the successive letters of the Greek alphabet: γ, δ, ε, ζ, η, θ, etc. We use here the Latin equivalent to the greek letter; thus, CA stands for carbon α, CB for carbon β, CG for carbon γ, CD for carbon δ, and so on.

- Is atom #238 in the structure file **rubredoxina.pdb**.

- Its spatial coordinates are (**X** *9.875999* **Y** *-10.771* **Z** *5.058*)

Note that pointing merely with the mouse to the atom in the display window will appear a label with the same information (except for the atom coordinates)

Practice in recognition on other atoms in the structure. Remember that hydrogen atoms are not represented.

5. Atoms and structures selection

Jmol allows to select atoms or sets of atoms with some common property. Once selected, all commands entered in the command window will affect only to the selected atom(s). Let's see some of the selection commands.

5.1 Aminoacid selection

Let's suppose that we wish to visualize all the isoleucines present in the molecule. We enter the command

$ select ile [Enter]

The program informs that 32 atoms have been selected. From now on, all the commands will affect only to this set of atoms (the isoleucines). The command

$ color white [Enter]

Will color in white all the atoms in the isoleucines present in rubredoxin

To go to the initial situation, we type

$ select not solvent [Enter] (or **$ select all** [Enter])

$ color cpk [Enter]

The rubredoxin molecule has one iron atom coordinated to the sulfur atom of four cysteines. To visualize this structure, we type at the command window

$ color green [Enter] (the whole molecule appears in green)

$ select cys [Enter] (Cysteines selected)

$ color cpk [Enter] (Cysteines colored in cpk code)

$ select fe [Enter] (Iron atom selected)

$ **color cpk** [Enter] (Iron atom colored in cpk code)

Sequences of commands can be entered in a single line, with individual commands separated by semicolons (;). Thus, the above sequence can be entered as

$ color green; select cys; color cpk; select fe; color cpk [Enter]

We only have selected the iron atom. To select the whole molecule (except solvent) we enter in the command window

$ select not solvent [Enter] (or **$ select all** [Enter])

And back to cpk color code:

$ **color cpk** [Enter]

Practice selecting other aminoacids (they must be entered with the 3-letter code)

We can also select aminoacid according their position in the peptide chain. For example, to select the aminoacid at the N-terminus the commands would be

$ select 1; color white [Enter]

In this way we could select every residue in the protein (from 1 to 53)

Back to the initial view of the molecule:

$ select all; color cpk [Enter]

5.2 Heteroatoms

We know as *heteroatoms* those atoms that do not belong to the peptide structure, although they are forming part of the protein (prosthetic groups and metal atoms) and also the molecules of solvent trapped in the crystal structure. To locate the heteroatoms in rubredoxin we color in green the molecule and the heteroatoms in cpk code:

$ color green; select hetero; color cpk [Enter]

In this case we select only one atom, the iron atom (we have previously excluded the crystallization water with the command **$ restrict not solvent**)

We select back the molecule with the commands

$ select not solvent; color cpk [Enter]

5.3 Selection of secondary structures

Jmol allows the selection of zones with a certain type of secondary structure. We use the keywords **helix** (for α-helices), **sheet** (for β-sheets) and **turn** (for β-turns). The command sequence

$ color green; select helix; color cpk [Enter]

Select those parts of the molecule forming α-helices.

Back to the initial view:

$ select all; color cpk [Enter]

5.4 Predefined sets in Jmol

The command **$ restrict** works like **$ select**, but it erases from vision all those atoms not covered by the instruction.

We can use **$ select** and **$ restrict** to select atoms or atom sets with some properties in common. Jmol has three groups of predefined sets:

5.4.1 Predefined sets of atoms

alpha: all the α carbons in the protein backbone
amino: the set of all atoms in a protein structure
backbone: atoms of the protein backbone N - Cα - C for all the aminoacids.
hetero: Non-peptide atoms in the structure
ligand: Non solvent heteroatoms; molecules that bind to the protein.
nucleic: the set of all atoms in a nucleic acid structure
protein: the same as **amino**
solvent: the atoms of the solvent; usually crystallization water.
water: atoms belonging to water molecules.

5.4.2 Predefined sets of aminoacids

acidic: The dicarboxylic aminoacids Asp and Glu
acyclic: Aminoacids not having a cyclic structure.
aliphatic: The set of aliphatic aminoacids Gly, Ala, Val, Leu, Ile.
aromatic: The set of aromatic aminoacids Phe, Tyr, Trp, His.
basic: The set of dibasic aminoacids Lys, Arg, His
charged: The set of aminoacids with electric charge
cyclic: Aminoacids with a cyclic structure: Pro, His, Phe, Tyr, Trp
cystine: The cysteines linked by a -S-S- bond
large: Aminoacids with a large side chain
hydrophobic: Aminoacids with a hydrophobic side chain
medium: Medium size side chain aminoacids
neutral: Neutral aminoacids
polar: Polar (hydrophilic) aminoacids
small: Small side chain aminoacids

5.4.3 Predefined sets of structures

buried: Atoms that are in the interior of the molecule, apart from solvent
helix: residues forming α-helices
sheet: residues forming β-sheets
surface: atoms in the surface of the molecule, in contact with solvent.
turn: residues forming β-turns

Practice with selections of these sets.

6. Hemoglobin

It is interesting to practice with molecules bigger than rubredoxin. Thus, in the same folder **Demo_Jmol > Proteinas > estructuras** you can find files for many other protein molecules.

Open the file **Demo_Jmol** > **Proteinas** > **estructuras** > **1rvw.pdb**. This is a file containing a model of Hemoglobin in the oxy- (R) form (actually, carbomonoxy-hemoglobin, that is, hemoglobin loaded with carbon monoxide (CO) instead of oxygen.

First, suppress solvent atoms with the command **$ restrict not solvent**. Change the display mode to **Style** > **Scheme** > **Cartoon**. Note that α-helix secondary structure is predominant.

6.1 Subunits and hem group

The command **$ color chain** will assign a different color to each of the subunits in the protein. Subunit **a** (α1) in blue, subunit **b** (β1) in green, subunit **c** (α2) in pink and subunit **d** (β2) in yellow.

In the structure file, atoms in the four hem groups appear as HETATM (heteroatoms). To see the hem groups enter the command **$ select ligand** and change the style to **Style** > **Scheme** > **Ball and Stick**. The four hem groups appear, one in each subunit.

6.2 An isolated subunit

We can restrict the view to the α1 subunit with the command

$ restrict *a

(or ***b** the β1, ***c** the α2 and ***d** the β2 subunits). It is convenient to center the view in the subunit *a. In the Main Menu choose **Set picking > Center**. Then locate the iron atom in the center of the hem group and click on it. The model will center around this atom. Change immediately to **Set picking > Identity** (otherwise the model will keep centering on each click of the mouse)

7. Immunoglobulins

Open the file **Demo_Jmol** > **Proteinas** > **Estructuras** > **1igt.pdb**. This is a structure file for a model of Immunoglobulin G.

Note that hydrogen atoms are represented in this file. Change the display to **Cartoon** and the color to **chain**. Identify the light and heavy chain, and the fragments Fab (2) and Fc (1). At the center of the model, a region appear with no secondary structure: this is the *hinge* region, the point at which the enzyme *papain* acts cleaving the molecule into two Fab fragments and one Fc fragment.

7.1 The Light chain

the command **$ restrict *a** will restrict the display to the L1 light chain. It is clearly formed by two domains, the variable domain (VL) at the N-terminal half and the constant domain (CL) at the C-terminal half. Center the molecule at the region between both domains. To distinguish the domains, change the color to **$ color group**. The VL domain contain the N-terminus (blue) and the CL domain contains the C-terminus (red) Note the characteristic secondary structure of both domains: a *double antiparallel pleated sheet*.

Select the whole molecule with the command **$ select all**

7.2 The Heavy chain

Restrict the view to one of the heavy chains with the command **$ restrict *b**. We can clearly view the four domains of the heavy chain. These are the variable domain (VH) and the three constant domains, CH1, CH2 and CH3. Center the molecule at the region between CH1 and CH2. The variable domain contains the N-terminus and the CH3 domain the C-terminus. They can be distinguish by the command **$ color group**.

The domain CH2 is linked to a complex oligosaccharide. To display this, enter the command

$ select ligand and *b

and change the style to **Ball and Stick**.

Jmol in web pages

Jmol can be inserted in web pages. In the web page for this course (PHS 2201) you can find a demonstration called

Molecular models demo

designed to help you in the structural study of biomolecules. You are supposed to study it along this course.