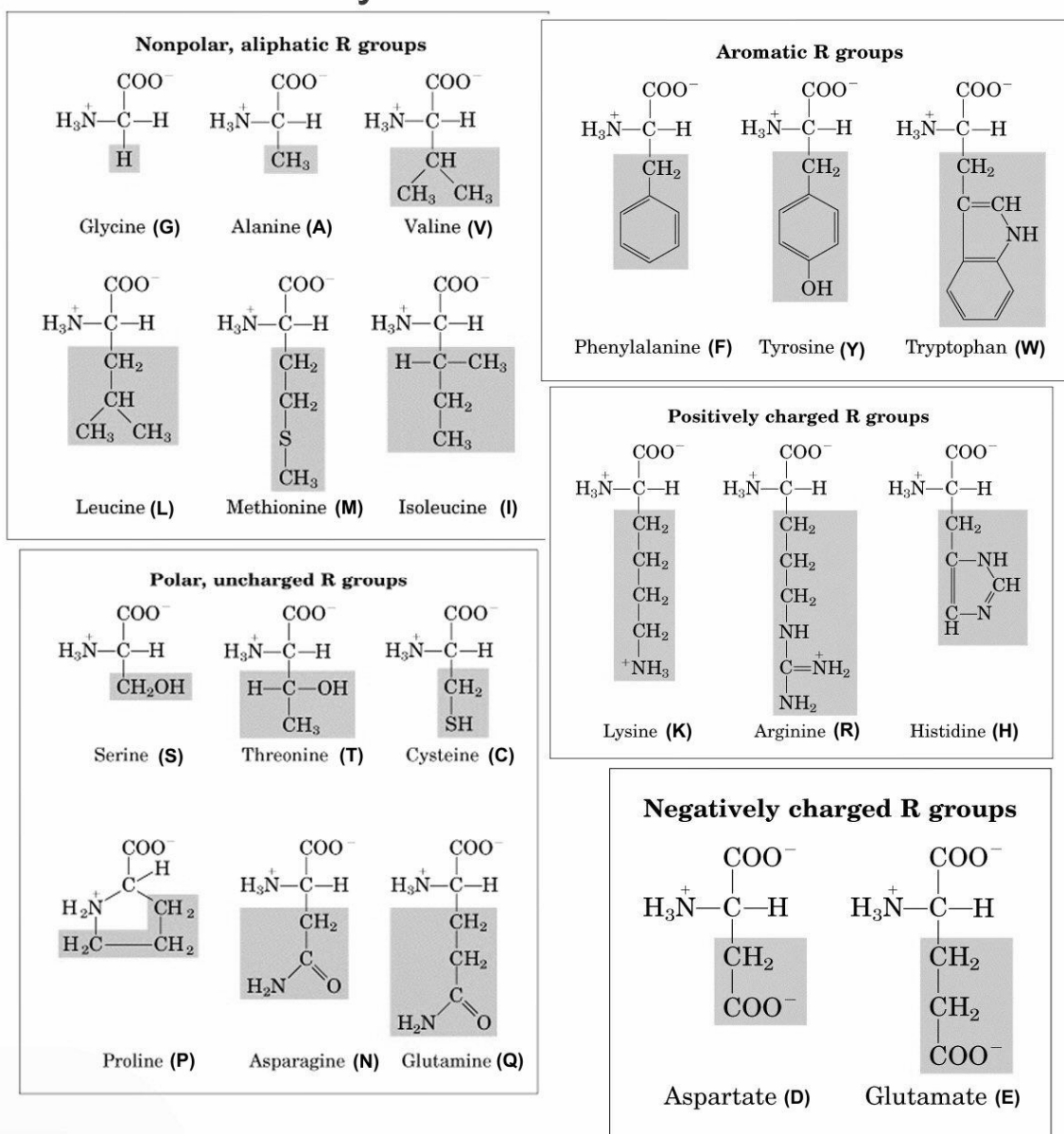


## LAB 4: LEARNING TO VIEW PROTEIN STRUCTURES.

### PROTEIN FUNDAMENTALS

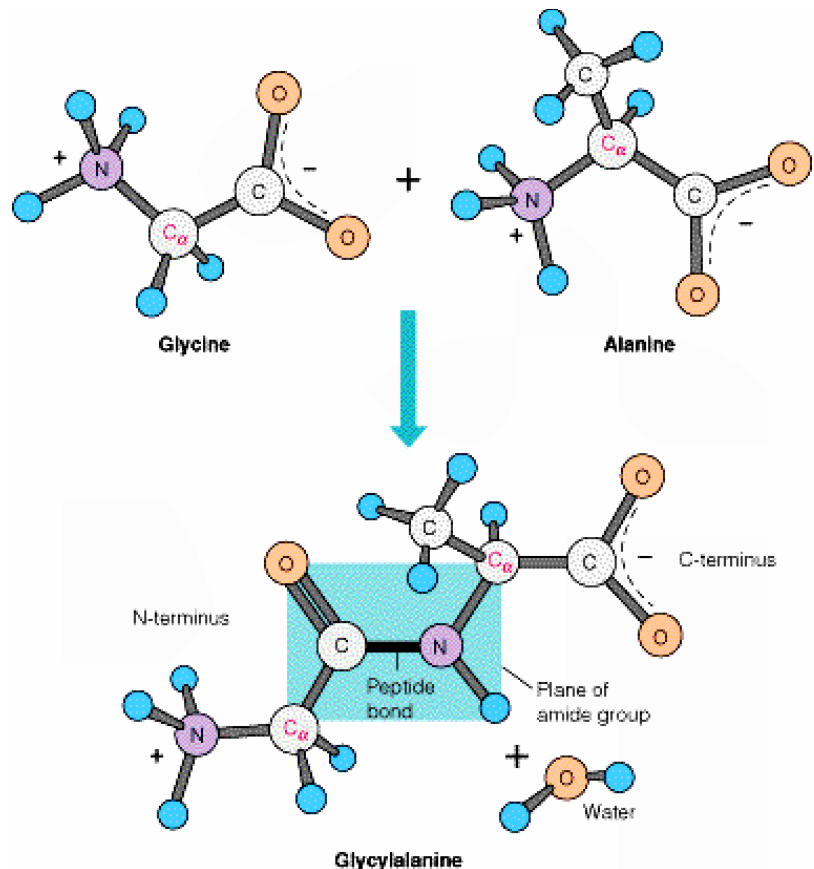
Proteins are polymers of amino acids. There are 20 amino acids which have differing properties. They all have a central carbon, called  $C_{\alpha}$ , to which is attached a carboxylic acid, an amine, a hydrogen, and some other group called the "side chain". That is a general structure like this  $H_2N-CHR-COOH$ . The 20 are shown here.

## Twenty standard Amino Acids



The amino acids all have one letter codes, which are shown. It is useful to memorize them for work with proteins. As you know from work with acids and bases, the carboxylic acids, like D and E, will be ionized at physiologic pH. They will have negative charges which interact strongly with water. Similarly, basic amino acids, like K and R, will be positively charged and interact with water. Non polar side chains, like L or F, are “hydrophobic” (water fearing).

The amino acids are linked together like beads on a string. A water molecule is removed at the intersection; the bond joining two amino acids is called a peptide bond; an example is shown of G and A forming a dipeptide.



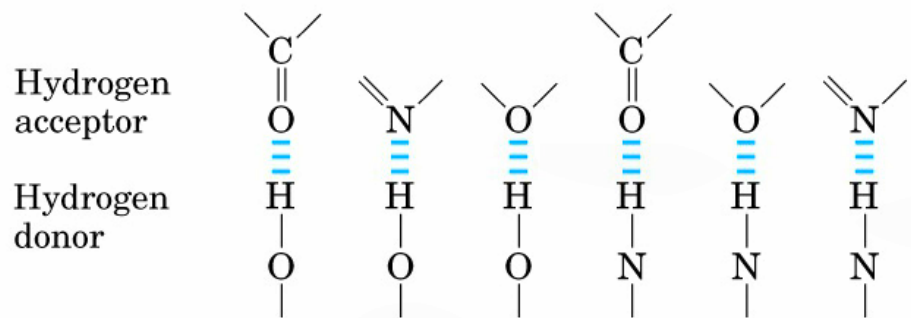
Although the peptide bond is shown as a single bond, in fact it is partially aromatic and so does not rotate. When you look at protein structures you will note that they are flat. Also the peptide is neither acidic nor basic so it is always uncharged.

Polypeptides are longer, LINEAR, unbranched strings of amino acids, which are often referred to in the trade as “residues”. Proteins are polypeptides covering a wide range of lengths, from, say, 50 to over 1000. A typical protein has about 300 residues. On average each residue has a MW ~110 so a 300 residue protein will have a molecular mass of ~33,000 daltons, or 33 KD.

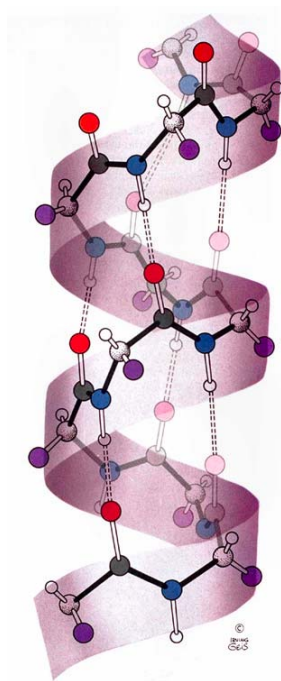
## Secondary Structure

Proteins fold up in a complicated way governed by thermodynamics. Most globular proteins are like “oil drops”. They fold with many non-polar side chains inside and the charged one tend to be outside, interacting with water.

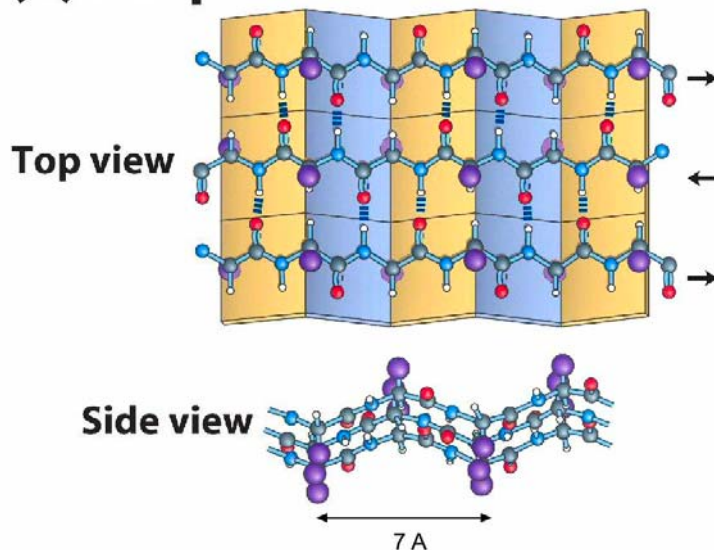
The polypeptide backbone often folds in a locally repetitive way called secondary structure. These are secured by **hydrogen bonds**. These are interactions between two polar atoms, O or N, one of which has a hydrogen (this is the hydrogen bond donor) and one with a lone pair of electrons, the acceptor atom. Some examples are shown.



Secondary structure of proteins is of two main types,  $\alpha$ -helix and  $\beta$ -sheet. The latter can have parallel, antiparallel, or mixed sheets. Below are cartoon representations of the helix and one type of sheet.



### (a) Antiparallel

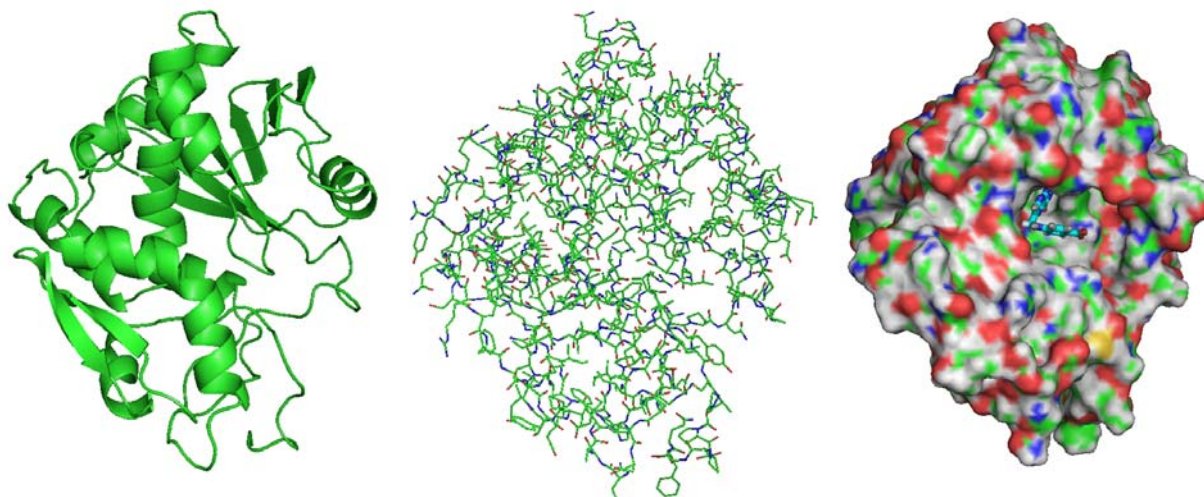


### Tertiary Structure of Proteins.

The overall folding of a protein in space is called the tertiary structure. It may contain elements of secondary structure as well as seemingly random coil. This 3D structure can be experimentally observed by two methods, **X-ray crystallography** and nuclear magnetic resonance (NMR). Each has its advantages, but X-rays can solve much larger proteins and account for the vast majority of known structures.

A 3D model of a protein can be represented as a list of atoms, and their coordinates in a 3D space, that is x, y, and z. In addition they usually have a measure of how much the atom moves around called a temperature factor, B. The most common format for these lists is called a protein data base or **.pdb** file. We will use these in our project. There are other formats for describing structures including **.sdf** and **.mol2** files; these will often be used to display small molecules or “drug candidates” for virtual screening by computer.

Protein structures are complicated and are often displayed in different ways. Below are three views of the same protein.



The left image is a cartoon of the backbone (no side chains); it is fairly easy to see the secondary structure. The middle view shows all of the atoms, except the hydrogens. It is pretty busy. The right hand image is a space-filling model in which all atoms are surrounded by their Van der Waal radii. (Note in the center of this image is a stick image of an inhibitor. This sets the stage for the kind of work we want to do later, using computers to find small molecules that fit neatly into protein crevices and pockets.)

## USING MOLECULAR VIEWING SOFTWARE

The goal of this lab is to load software onto your computer that will allow you to display protein structures (read pdb files). You will be able to rotate, translate, zoom, and display them in a variety of ways. Your goal is to become familiar with protein structures and folds; you want to look at how different residues interact to form cores, align into secondary structures, and form “active sites” that bind other molecules to help the protein do its biological job. Many of the proteins we are interested in are enzymes, that is catalysts that speed up key biochemical reactions. This chemistry is always carried out in the active site. This is nearly always the target area we want to test for drug binding in virtual screening efforts. For example, the protein illustrated above is an enzyme called ricin A chain. It is the business end of a very toxic protein. There is

interest in finding small molecule antidotes to it, that will fit and bind so tightly to the active site, that ricin A chain can no longer catalyze its reaction (which happens to be cleaving a key adenine on ribosomes).

There are a number of programs available for molecular graphics. The one we will use is called **PyMol**. PyMol is not the easiest graphics program available, even among the free ones, but it is very powerful, and worth the investment in time to learn about. Most professional structural biology labs use it. There are several tutorials online to help you get comfortable with the program and its many, many facets. To be sure you get started though, let's make our own mini-tutorial to get on track.

If you have your own computer you can load PyMol 0.99) onto it. The link to the free version is: <http://delsci.com/rel/099/>. If you want to install PyMol on your machine, but have problems, contact the RE or TA. Alternatively, **PyMol is available on the computers of the Welch Hall Computer Lab in Wel 2.306**. You will need your UTID to log into these machines. From the Start menu, go to the Chemistry folder and PyMol can be opened by the appropriate icon.

For now, certain molecule pdb files can be downloaded from a chemistry web site. The link is: <http://courses.cm.utexas.edu/jrobertus/ch204/>

For the first lab download the file called Ahelix.pdb to your machine (desk top or where ever you like).

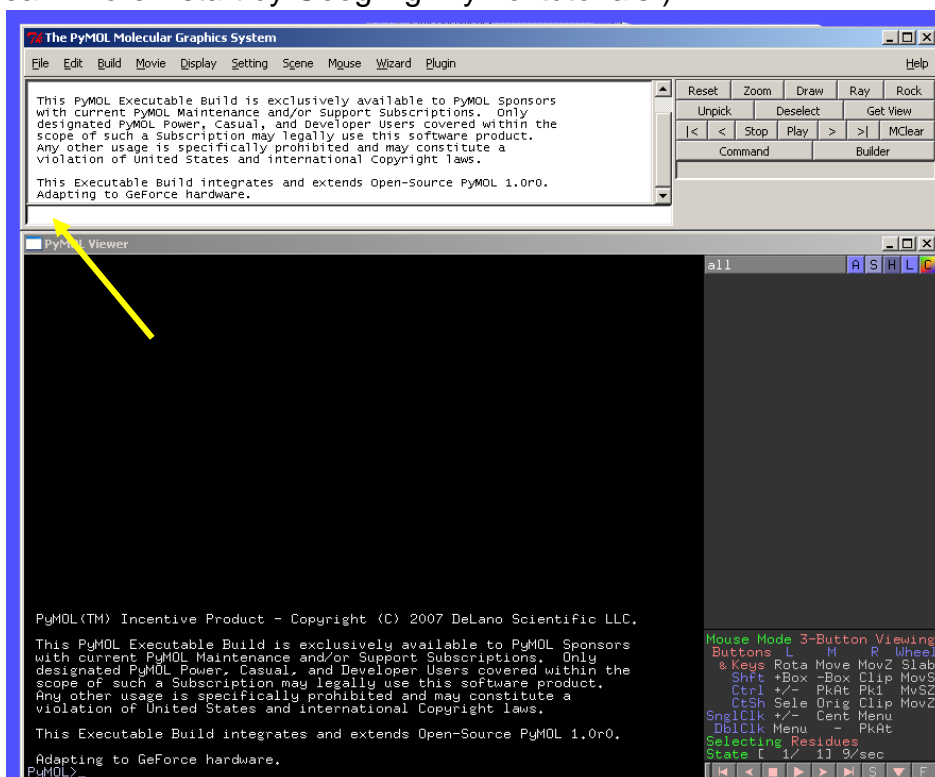
There are so many things to do with PyMol it is hopeless to discuss them all. For now we will get you started with some basic, but very important, fundamentals. (You can go to some of the online tutorials to learn more—start by Googling “PyMol tutorials”)

Here we go.

1. Bring up PyMol - it will look like this:

The upper panel has a row of pull-down menus that can handle most of your needs. It also has a box section (yellow arrow) where you can enter line commands. More on that later.

The larger lower panel is where the molecules will be displayed, on its left hand side. On the right is an administrative section that we will see soon.

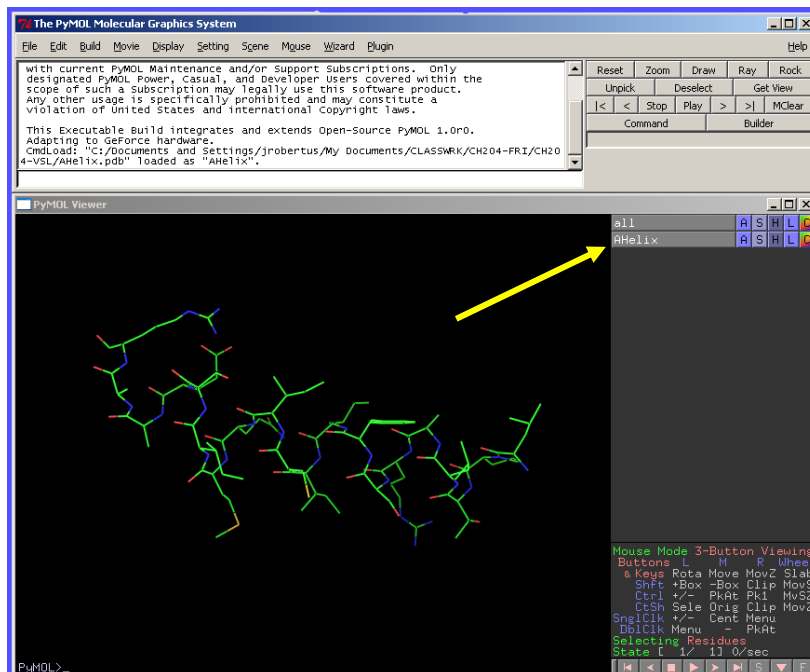




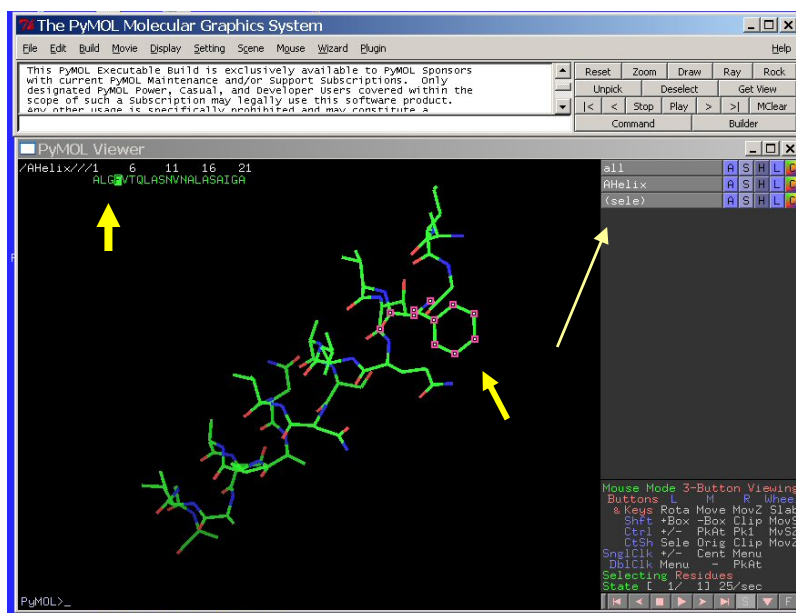
2. There are many ways to load your molecule, but a simple one is to go to the top panel File pull down, and navigate to where you put Ahelix.pdb.

When you bring it up it will look like this:

Notice in the right hand side of the lower panel that a line is now present that says ALL, and Ahelix (yellow arrow). Next to the title are boxes labeled A (actions), S (show), H (hide), L (label), and C (color scheme). If you click in the AHelix name box itself, the molecule becomes inactive and disappears as the box goes gray. Click it again to activate. Now go to H and hide everything. The screen goes blank. Now go to show, and hit lines – the model returns. You can also show sticks, ribbons, cartoons, surface or whatever. Try them all. You can also hide or show side chains separately from main chain and so on. (You can also open the C tab and change colors as you like; just fool around with it.)



3. The next thing is to move the model. The left mouse button will rotate the structure, the middle button or wheel, if held down, will translate the structure, and the right button will zoom it in or out. Be advised that the wheel when rolled will change the “slab”, that is the amount of the model that is in view in a direction perpendicular to the page. If things seem to disappear, change the slab to bring them back. IF you don't have a 3-button (wheel) mouse, go to the top panel pull down menus, go to mouse, and change to a 2-button mouse; you may need to use the pull down menus to carry out some of our operations.



4. It is important to be able to **select** residues or longer segments for action. There are

many ways to do this. Here are a couple:

a) go to the upper panel menu, click “display” and then “sequence”. A bar giving the one letter codes of your molecule appears. You can use the cursor to select one or more (control or shift click) of them; the bar is shown, with Phe 4 selected, in the picture as indicated by a short, fat yellow arrow. When you make a selection, the appropriate part of the molecule is “marked” by highlighting the atoms of the selection, as shown by the medium yellow arrow. The selection also appears as an “object” in the molecule list as indicated by the long thin arrow. If you want to maintain this selection as a recurring object, you can go to the A tab of the sele, and **rename** it. In either case, you can work with the selection like other molecules, changing its display or color etc.

If you **click your mouse in an empty part of the display window**, the selection highlighting disappears, although the selection remains in the object list.

b. right clicking on an atom will select that residue as well, but you may not know its name.

c) another way to select a specific group is to use the **line commands**. There are 2 places to enter line commands: in the box in the upper panel, or at the PyMol> prompt at the bottom of the graphics area. We use the “resi” mode which selects residues by **number**.

To select a residue, say Val 12, you can simply type: select resi 12.

To select a group, say residues 4-6 type: select resi 4-6. OR select resi 4+5+6

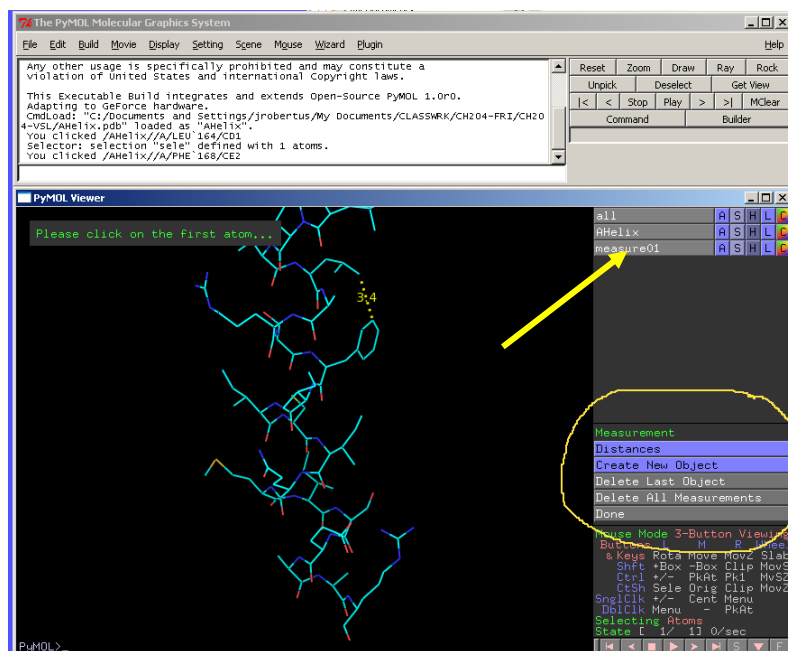
In either case the appropriate atoms light up, and the selection is displayed in the molecule list (thin yellow arrow).

If you want to use a selection for future work, give the selection a name. It could be “Nterm”, “tom”, or whatever, but let’s call this one v12 because that reminds us of what we will select. The command to select it then is: select v12, resi 12. That command highlights Val12 in the displayed model, and it puts a line in the molecule list called v12. Go to the H tab of v12 and click “unselected”. Everything except V12 disappears. Now go to the L tab of v12 and scroll down to “atom name”. Clicking it gives the name of the atoms. To return rest of the molecule click on S then lines for the Ahelix tabs.

d) If we may want to select residues by chemical class **type**. For example we may want to select all the Leucines; to do that enter the line command: select leu, resn leu. The term “resn”, instead of “resi”, is now used - it signifies residue name. (There are many other line commands in PyMol that you can explore online). As usual, the selection opens a new item in the molecule list and you can operate on it.

If you wanted to look at all the hydrophobic residues you might want to include several types of residues. Try this line command: select hyfob, resn leu+val+phe. The selection adds to the molecule list, and can be operated on. Go to the S button and show these side chains as sticks, or spheres.

5. A useful tool with PyMol is to **measure** the distance between two atoms. To do this, go to the top panel pull-down list and choose “wizard” and when it pulls down choose “measurement”. New menu items appear in the right side of the lower panel, they are circled in yellow on the picture. Many options are now available including deleting measurements you make, and closing the window with “done”. PyMol also prompts you to click on the first atom and then click on the second and the distance appears as a yellow dashed line with the distance measured in Angstroms ( $10^{-10}$  m). Note that the act of measuring puts another item in the molecule list, it is shown with the yellow arrow. (A lot of PyMol actions are like this.) You can use the associated tabs to manipulate the measurement, including to delete the dash line etc. but this can also be done in the measurement options (circled).



6. Another important operation is to **identify, or label**, various residues. There are several ways to do this.

a) put the arrow on an atom and RIGHT CLICK. A window will open and tell you the residue name, and open other fly out window. You should play with these options. One thing to do is scroll down to residue and another fly out opens from which you can label it, color it, center it in the screen etc.

b) put the cursor arrow on an atom and LEFT click. The residue is selected and a new item opens in the molecule list called “sele”. It has its own A,S,H,L,C options and these can be used to change its presentation (to say sticks), color, or to add a label so you know what it is.

7. Making Pictures with PyMol

PyMol can make very fancy pictures (as you will see later). Unfortunately it only makes a format called .png, but these are readily read by contemporary viewers and can be made into .jpg or other formats if you like. There are several ways to make pictures, lets check out the easiest one.

To make a handsome picture, convert your helix model to sticks (using the S tab). To create a picture go to the top panel pull-down menu, click file, click “save image as” and choose png. A “save as” window opens and you can name the picture and put it where you want. The picture is essentially a screen dump – that is, it is the size of the screen



you are using. If you enlarge your display screen by dragging on the corners, and redo the picture it will now store as a larger image.

**Now that you have some idea of how PyMol works, here are some things to do.**

1. Display the helix as sticks, and then show the cartoon to map the polypeptide backbone. Label the amino terminal N and the carboxy terminal C. Rotate the structure, looking down and at the helix axis. Make a picture.
2. Count the number of residues in a turn of helix; a good estimate is to count the number of turns and number of residues over this entire helix and divide residues by turns.
3. Look down the helix axis. Note that the side chains splay out away from the backbone. In projection the backbone looks like a circle. Estimate its width in angstroms, Å; this is easier if you **hide the side chains**.
4. The pitch of a helix is the distance the chain rises, along its imaginary central axis, for each turn of  $360^\circ$ . You can measure this along the outer edge of the helix by choosing atoms nearly directly above one another (along the axis direction). A better way might be to measure the distance spanned by 3 or 4 turns and divide by that number.
5. Look at the pattern of hydrogen bonds along the  $\alpha$ -helix. Select resi 1, go to the A tab, and select "center"; the view changes to put this residue in mid screen. The carbonyl oxygen of resi 1 points toward the N of resi 5. Measure that distance. A hydrogen bond will generally be formed if these residues are  $<3 \text{ \AA}$  apart or so. Continue to measure other pairs like O2 to N6, O3 to N7 etc. Note the characteristic pattern that a given  $O_i$  accepts an hydrogen bond from the nitrogen of the residue  $i+4$ . (In more advance work little pieces of other helices, like the  $\pi$  or 4-10, are seen, but they have different patterns.)
6. This particular helix has "amphipathic nature". It is non-polar on one side and polar on the other; in a larger protein the hydrophobic side would probably stick to other like structures and the polar side would be on the protein surface. Display the non polar side chains and color them green. Display the charged side chains (D,E,K, R) as blue. Display the non-charged, but polar ones (S,T,N,Q) as red. Confirm the amphipaticity of the helix. Make a picture or two.
7. Display a single residue, say V12. Spin it around, look at the atom names and be sure are familiar with the amino acid structure. Find the  $C_\alpha$  (the central carbon atom), amino nitrogen, carbonyl group and the side chain. Bring up another residue and check it out.