

Lab 2

Molecules and Molecular Orbitals

2.1 Overview

In this lab we will investigate how the ADFINPUT program can be used to define the structures of molecules. The two approaches that will be discussed are importing structures defined elsewhere and building from scratch. Once we built some molecules we will look at a useful piece of information that can be extracted from our molecular calculations: molecular orbitals.

2.2 Defining Molecules

In the previous lab we ran a number of calculations on atomic systems. If we want to investigate problems that are more interesting chemically we will have to study molecules. Telling the program to study a particular atom was straightforward. How can we tell ADF exactly what molecule we are interested in? There are two main approaches, **Importing** and **Building**. We shall consider each in turn.

2.3 Importing Molecules

2.3.1 Chemical Structure Formats

Many molecules have had their structures studied already, either experimentally or theoretically. The results of these studies may well be published somewhere. If a recorded structure is written in a format that ADFINPUT can understand then ADFINPUT can read a file containing the structure and immediately produce the required molecule.

Aside There are many, many, many different ways that a chemical structure may be stored. So many, that a special computer program called Babel (http://openbabel.sourceforge.net/wiki/Main_page) exists to translate between the various possible formats. A large number of these formats exist for commercial reasons and have no advantage over the alternatives except that they may be used efficiently with a particular piece of software. Other formats, such as internal coordinates and the PDB format are useful because they do a good job of describing aspects of molecular structure that are go beyond the simple placement of atoms. Examples of such extra information include bond lengths and

angles, locations and types of chemical bonds and the information required to describe a biological molecule in ribbon form.

The two chemical structure formats that we will be using are XYZ and PDB.

XYZ is what you might expect. The locations of the atoms in Cartesian space (usually in Ångstrom) are recorded. The only other information included in an XYZ file is the identity of each atom and maybe the number of atoms and a title. This format is the simplest possible way to describe the locations of atoms in a molecule. As such it is very easy to read and transfer between programs but contains no additional information.

The PDB format is more involved. PDB stands for Protein Data Bank. More information on the PDB file type can be found at <http://www.rcsb.org/pdb/Welcome.do> and http://www.okino.com/conv/imp_pdb.htm. This format is obviously designed for describing biological molecules. It is considerably more complicated than the XYZ format but also can include information about the secondary and tertiary structure of a protein that an XYZ format knows nothing about.

2.3.2 Importing XYZ Molecular Structures

Importing a file in the XYZ format is achieved by using the *File:Import Coordinates...* command. If you are importing an XYZ file make sure that you select the *Any XYZ-like file* file type.

Exercise 2.1 Import the file

`/home/seth/575/2_2_Mo_comp.xyz`. Save a picture of the molecule.

You can look at what the XYZ file looks like by typing

```
cat /home/seth/575/2_2_Mo_comp.xyz
```

into your xterm.

2.3.3 Viewing a Molecule

An imported molecule can be rotated, moved, magnified and shrunk.

- Rotation is achieved by pressing the left mouse button and moving the mouse.
- Movement is achieved by pressing the middle mouse button and moving the mouse.
- The molecule can be magnified by pressing the right mouse button and moving the mouse away from you corresponding to moving the cursor up on the screen.
- The molecule can be shrunk by pressing the right mouse button and moving the mouse towards you (down).
- The view can be reset by the *View : Reset View* command.

Exercise 2.2 Load the Mo dimer complex structure. Rotate the molecule so that the view looks along the Mo-Mo axis. Save a picture of the molecule. Rotate and enlarge the structure so that the two Mo atoms touch the left and right sides of the viewing box. Save a picture of the molecule.

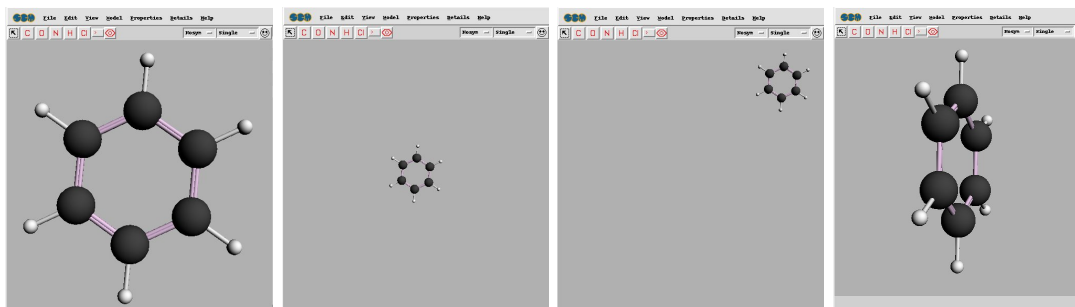


Figure 2.1: Change the view of the molecule

2.3.4 Importing PDB Molecular Structures

A file containing a molecular structure in the PDB format is imported using the *File: Open...* command. The two possible file types are *adf* and *pdb*. Opening an file of type *adf* is useful for opening calculations previously created by ADFINPUT. We are interested in the *pdb* file type here.

Exercise 2.3 Import one of the following files: `/home/seth/575/Cytochrom_Lab2.pdb`, `/home/seth/575/ERCC1_Lab2.pdb` or `/home/seth/575/MT_DSBF_Lab2.pdb`. Save a picture of your molecule.

Note that PDB files often take a long time to load because they can be rather large.

Some of the extra information included in a PDB file concerns which atoms are part of a protein's backbone and which atoms are part of sidechains. Once a PDB file is loaded, it is initially presented without the sidechains and with the backbone in ribbon form. The panel to the right of the molecule window is also changed to give some controls for visualizing the molecule (figure ??). The two drop menus indicated in figure ?? control how the molecular backbone and sidechain are represented while the window lists all molecules and amino acids in the file. Selecting one of the items in the list will cause the subject to be highlighted.

Exercise 2.4 Use the new menus on the right hand side to change how the backbone and sidechains are represented. Save a picture of the molecule with the sidechains visible and another picture with the backbone represented as a chain of C.

2.4 Building Molecules

If you don't have an appropriate file containing the structure of the molecule that you are interested in then you will have to build the molecule yourself from scratch. Alternatively, you may have a predefined structure that is related to your molecule that you would like to modify. In this section we will go over how to build molecules using the following procedures:

- Adding atoms
- Selecting and modifying atoms

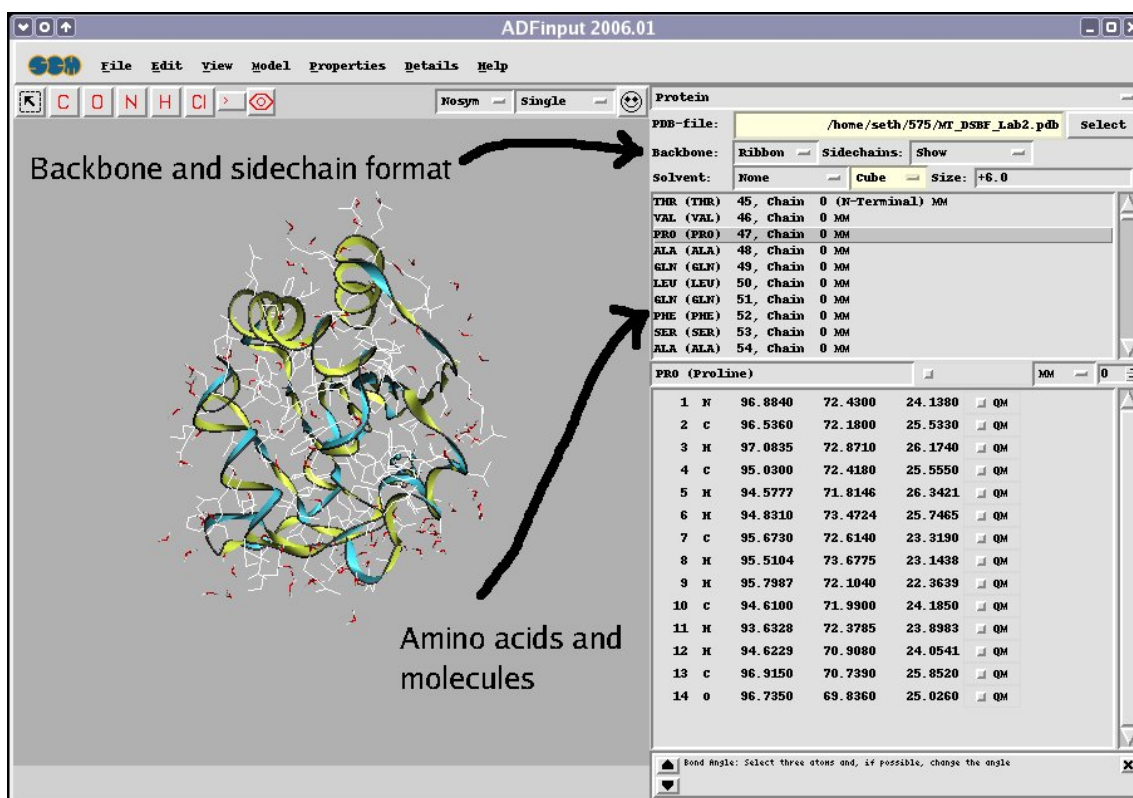


Figure 2.2: An imported molecule from a PDB file

- Modifying bonds
- Adding Groups and hydrogens
- Cleaning up your structure

2.4.1 Adding Atoms

Exercise 2.5 In the previous lab we made use of the “H” button to convert an atom into a hydrogen atom. The “C”, “O”, “N” and “Cl” buttons next to the “H” have the expected effect. Other elements can be selected using the > button. Select an element. Press the left mouse button inside the build window, move the mouse and click again. Repeat several times placing atoms wherever you want. Make a chain of 8 atoms. A chain can be terminated by clicking on the terminal highlighted atom. Add a chain of a different element attached somewhere to your first chain. Save a picture of your “molecule.”

2.4.2 Selecting and Modifying Atoms

An atom that is selected may be modified. The simplest way to select an atom is to click on it with the left mouse button as has already been done. Multiple atoms can be selected by holding down the shift key while clicking on the atoms of interest. All of the atoms in a region of space can be selected by holding down the shift key, pressing the left mouse button while pointing at empty space and moving the mouse to select a rectangle. All atoms can be unselected by clicking on empty space. Groups of atoms can also be selected by the *Edit: Select...* commands.

Once an atom or atoms are selected, they can be modified. The selected atoms can be rotated, moved etc in a similar way to how the whole molecule was moved. If you desire to rotate or move the selected atoms only then the pointer arrow must be over a selected atom when the mouse button is pressed (figure ??).

Selected atoms can be deleted with the backspace key or the *Edit: Clear* command.

Exercise 2.6 Delete four of the atoms in your molecule. Save a picture.

To change an atom from one element to another, select the element that you want to change it to from the periodic table tool >, select the atom to be changed and click on the atom again.

Exercise 2.7 Change one of the atoms of your molecule to Po atom. Change one to an I atom. Save a picture.

2.4.3 Modifying Bonds

By default, all bonds created are single bonds. A bond itself can be selected. A bond can be modified by selecting it then choosing the bond type from the drop-down menu just to the left of the smiley face (figure ??). A new bond can be created by selecting two atoms and choosing the *Edit: Add Bond* command.

Exercise 2.8 Change two of the bonds in your molecule to double bonds and two to triple bonds. Add two new bonds. Save a picture.

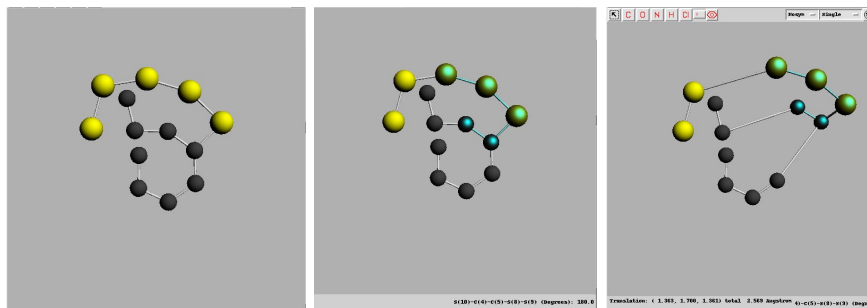


Figure 2.3: Molecule manipulations. Select atoms (centre). Modify selected atoms (right).

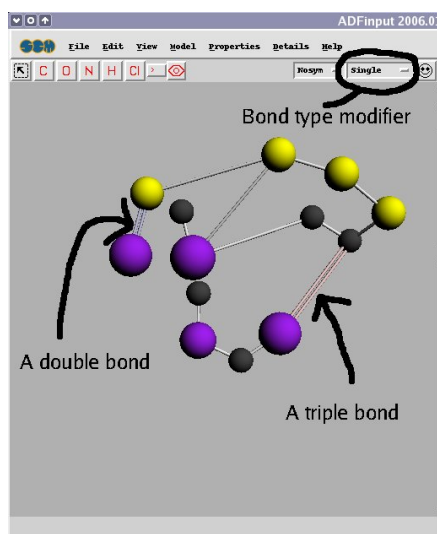


Figure 2.4: Modifying bond types.

Aside: The choices of bond type that you make here will have no influence on the later quantum mechanical calculation. They are for cosmetic purposes and for the structure cleanup procedure we will investigate soon

2.4.4 Adding Groups of Atoms, Hydrogens

It is often the case that you will repeatedly need to add certain groups of atoms, e.g. a butyl group, or that you want a common type of metal complex structure. These can be created from scratch using the procedures just outlined or, in many cases, predefined groups can be used. Just to the right of the > button is a button with a phenyl ring on it. After pressing this button, you will find many options to choose from of predefined groups. Many of these options (Alkyl chains, Amino acids, aromatics, cyclic hydrocarbons, DNA, Ligands) act like the command to add an atom except the whole group is added. The Metal Complexes and Polyhedra commands are used as starting points for making molecules of particular symmetry.

Do any two of these four exercises.

Exercise 2.9 Build any molecule with an Asp-Gly-Lys chain in it. Save a picture.

Exercise 2.10 Build a molecule with three pyridine groups in it. Save a picture.

Exercise 2.11 Build a trigonal bipyramidal metal complex with at least three different ligands. Save a picture.

Exercise 2.12 Build a molecule of white phosphorus. Save a picture.

Often a molecule will be built that has a lot of hydrogen atoms. Rather than adding all of them by hand, ADFINPUT can do it automatically through the *Edit: Add Hydrogen*

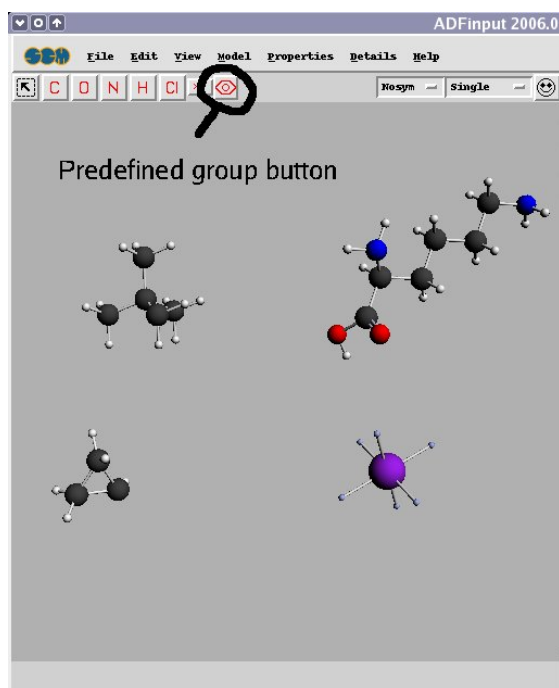


Figure 2.5: Some predefined groups.

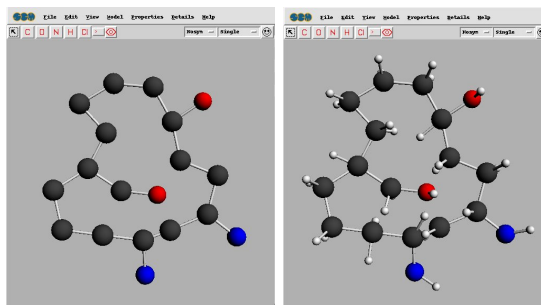


Figure 2.6: Hydrogens can be added easily in one command.

command (figure ??). Note that this command gets confused by multivalent atoms like transition metals.

2.4.5 *Cleaning up Structures*

Using the commands described in the previous sections it is generally straightforward to build a molecule with the desired formula and bonds. It is much more difficult to build a molecule with bond lengths, bond angles and dihedral angles that are all reasonable guesses for what should be present in a molecule. As we shall see in lab 5, one of the uses of quantum mechanical molecular calculations is to predict molecular structures. It is therefore possible to take

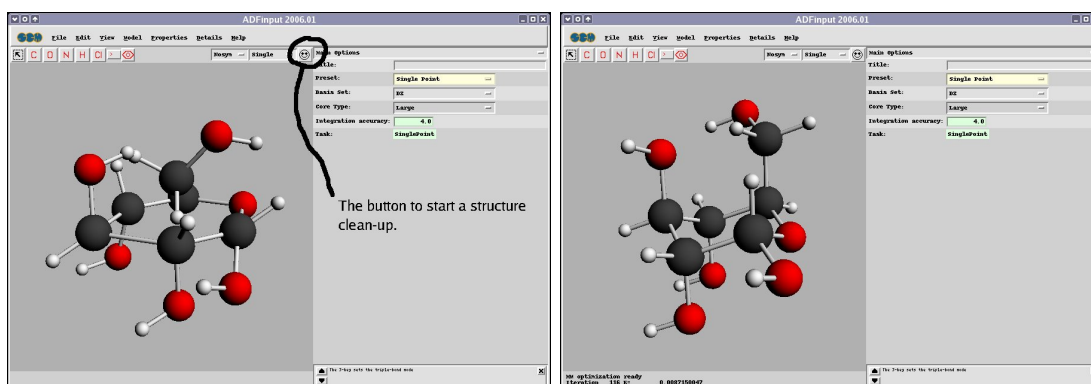


Figure 2.7: Cleaning up the structure of glucose. Left is before. Right is after.

any guess at a molecular structure and convert it into a reasonable structure with a quantum mechanical calculation. There is a serious problem with this approach. Quantum mechanical calculations can take a long time. A better methodology is to first improve or “clean up” the structure of the molecule with an approximate *and fast* calculation and then, if necessary, further improve the structure with a slow quantum mechanical calculation. The fast calculation is usually some kind of classical mechanics calculation utilizing empirical parameters to describe the interactions between the atoms. This type of calculation, often called a molecular mechanics calculation, is a huge field by itself but is outside the scope of this lab course. More information on molecular mechanics calculations can be found in the book by Leach (Molecular Modelling 2nd edition by Andrew Leach, Prentice and Hall, 2001) or at http://cmm.cit.nih.gov/modeling/guide_documents/molecular_mechanics_document.html, <http://www.accelrys.com/technologies/modeling/materials/atomistic/>, <http://chemistry.gsu.edu/glactone/modeling/MMintro.html> and <http://www.chem.swin.edu.au/modules/mod6/>.

In ADFINPUT, a simple molecular mechanics optimization of the structure is started by pressing the smiley-face button (figure ??). The optimization runs until it is suitably converged, runs out of iterations or the button is pressed again.

Aside MM calculations are not usually suitable for molecules that include transition metals or *f*-elements. The clean up button may not work well for such molecules.

Exercise 2.13 Build cyclohexanone atom by atom and maybe with the *Edit: Add Hydrogen* command. Clean up its structure. Save a picture.

2.5 Real Molecule/Model Molecule

It is a well-known “feature” of quantum chemical calculations that they consume an enormous amount of computer time. Many factors contribute to how long a given calculation will take. One important factor is the size of the molecule of interest. It seems intuitively obvious that a larger molecule would take longer to calculate than a smaller molecule but how much

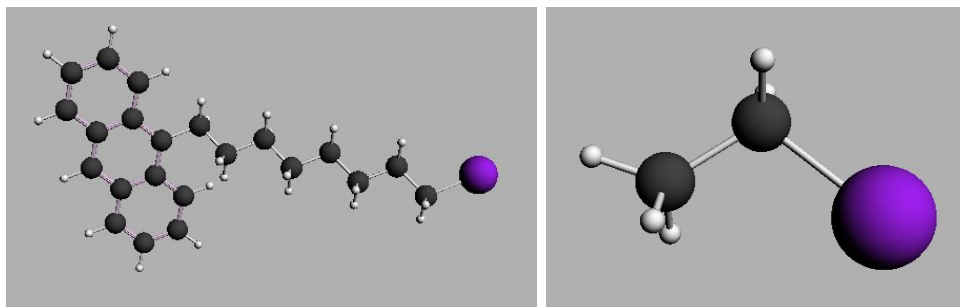


Figure 2.8: A molecule (left) and its model (right).

longer? DFT methods generally scale about as N^3 where N is the number of atoms in the molecule. So, if one molecule is twice as large as another then, with all other things being equal, a calculation of the larger molecule should take eight times as long as the calculation of the smaller molecule. This level of scaling is actually very good for a quantum mechanical calculations and is one of the major reasons why DFT methods have become so popular. Many programs (including ADF) include tricks that make calculations scale proportionately to N^1 (“linearly”) once the molecule under consideration is large enough. The research area of “linear scaling” is still active at the present time.

Whatever the exact scaling behaviour of a given calculation it is still true that quantum mechanical require a lot of computer time and that calculations involving bigger molecules will generally take longer than calculations involving smaller molecules. It is therefore desirable to perform calculations on molecules that are as small as possible.

On the other hand, many molecules of interest are very large.

There are numerous ways of dealing with this problem. One option is to actually perform the calculations on a molecule that is smaller than the real molecule but has enough of the properties of the big molecule that such a calculation would be expected to give something useful. The smaller molecule is a model of the large molecule.

Choosing a suitable model system is not always a trivial matter. In general, a good model should be able to express all of the properties and reactions of interest of the bigger molecule. This generally means that all of the atoms directly involved in a property or reaction should be retained and all other atoms that have a strong electronic effect on the reaction should also be retained. Another challenging question is what the deleted atoms should be replaced with. They must be replaced with something otherwise there will be a dangling bond. Usually the deleted atoms are replaced with a single atom that has the appropriate electronegativity.

Example Consider the molecule in figure ?? a) (the purple atom is I). If we are considering the substitution of the I atom by a Cl^- ion in a S_N2 reaction then structure shown in figure ?? b) could be a good model. On the other hand, if we are interested in a substitution reaction on the anthracene molecule then this is obviously a hopeless model.

Example Import structure `/home/seth/575/Bigmol2-Lab2.xyz`. If we are considering the electrophilic substitution at the position ortho to the ether group then

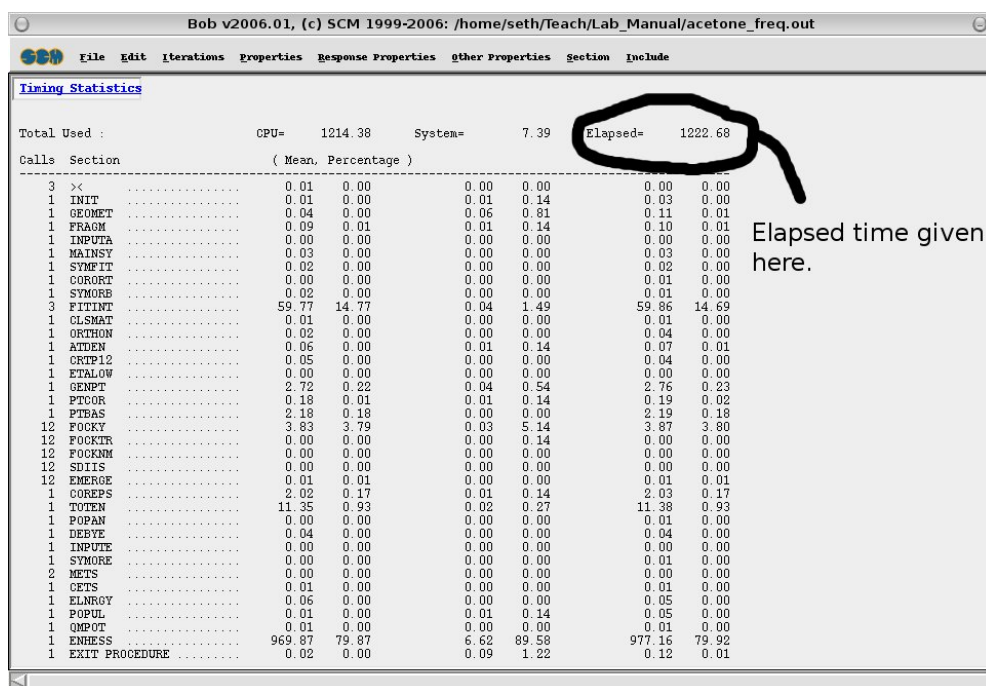


Figure 2.9: The timing analysis.

/home/seth/575/Modelmol2_Lab2.xyz could be a good model. /home/seth/575/Modelmol3_Lab2.xyz would not.

Exercise 2.14 Propose a model for the molecule /home/seth/575/Bigmol4_Lab2.xyz by editing the imported structure. Save a picture of your model. The property that we are interested in is the substitution reactions at the carbons of the benzene ring that are ortho to the Cl atom.

2.5.1 Calculation Time

In this section the fact that calculations involving larger molecules take longer than calculations on smaller molecules was introduced. This begs the question: how do we know how long our calculations take? There are several ways to determine how long a calculation took. The ADF program explicitly prints out an analysis of the time a calculation takes. This analysis can be found in the detailed output. You may recall that this output can be accessed with the *Output* option from the *SCM* menu. Once the window containing the detailed output opens the timing analysis can be accessed through the *Section: Timing* menu item. The analysis includes lots of interesting information but the most important in the total time which is called “Elapsed” (see figure ??). Make sure that you are looking at the correct section by choosing the **** (NO TITLE) **** option from the *include* menu.

Exercise 2.15 Choose one of the calculations that you ran in the previous lab or that you will run in this lab. Specify which calculation you have chosen. Report the elapsed

time of this calculation.

2.6 Initial Molecular Calculations and Molecular Orbitals

Now that we can build molecules, we are going to try some calculations on molecules using some of the techniques we learnt in the previous lab. You will need to refer to that lab for additional instructions.

2.6.1 Molecular Orbitals

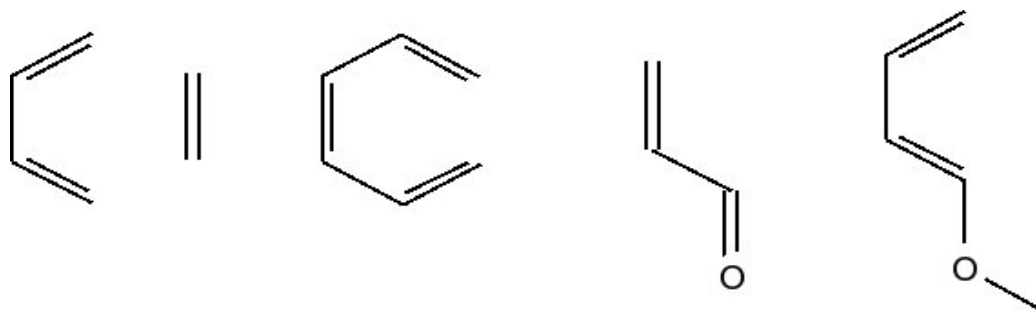
Once you have run a calculation on a molecule you immediately have access to something useful: the molecular orbitals of that molecule. These orbitals can be analyzed with the LEVELS and VIEW programs in the same way that we looked at the orbitals of atoms in the last lab.

Exercise 2.16 Build an H₂ molecule. Clean up its structure. Save your work with *File: Save As..* and then run a quantum chemical calculation on H₂ with the *File: Run* menu option. **a)** View the orbital energy diagram of H₂ with LEVELS and save a picture of it. (section 1.3.1 **b)** What are the energies of its HOMO and LUMO (section 1.5.2)? Visualize the HOMO and LUMO with VIEW (section 1.3.1). **c)** Describe these orbitals in terms of σ , π etc and bonding/antibonding. Save pictures of these orbitals.

2.7 Projects

Do project 2.7.1 and one of 2.7.2, 2.7.3 or 2.7.4.

2.7.1 Stereochemistry and Regioselectivity in Cycloadditions



Build cis-1,3-butadiene, ethene, 1,3,5-hexatriene, acrolein (CH₂=CH-CH=O) and cis-1-methoxy-1,3-butadiene. Run a calculation on each. Save a picture of each molecule.

In a cycloaddition reaction the reactants can approach each other in a suprafacial or an antarafacial manner. In a suprafacial reaction bonding occurs between one side of the π -system on both reactants. In an antarafacial reaction bonding occurs between one side of the π -system of one reactant and opposite sides of the π -system of the other reactant. Molecular orbitals can be used to decide whether a reaction will be suprafacial or antarafacial. In a

cycloaddition reaction consider the LUMO of the diene and the HOMO of the dienophile. A bond will be formed between lobes of MOs with the same sign.

For more information on cycloadditions and pericyclic reactions consult <http://www.cem.msu.edu/~reusch/VirtualText/pericycl.htm>, http://www.meta-synthesis.com/webbook/49_pericyclic/pericyclic.html <http://www.ch.ic.ac.uk/wiki/index.php/Organic:pericyclic:examples> and http://www.ch.ic.ac.uk/local/organic/pericyclic/p1_theory.html

Predict whether the cycloaddition reactions between a) cis-1,3-butadiene and ethene, b) 1,3,5-hexatriene and ethene and c) acrolein and cis-1-methoxy-1,3-butadiene will be antarafacial or suprafacial. Support your predictions with reference to the MOs from your calculations. Note that the sign of an MO is given by its colour in the VIEW program. Save pictures of the MOs that you use to confirm your prediction.

If both the diene and dienophile are substituted then a cycloaddition can produce different regioisomers. The preferred isomer can be determined with reference to MOs. If we again consider the HOMO of the diene and the LUMO of the dienophile the size of these orbitals will not be the same at both ends of the molecule if symmetry is broken by a substituent. If this is the case, then the preferred orientation of the molecules with respect to each other is such that the larger ends of the MOs are close and the smaller ends of the MO are close.

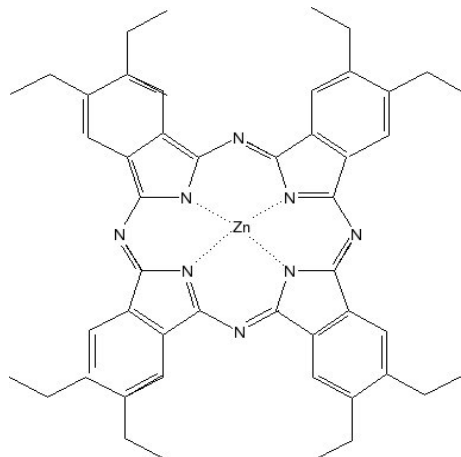
Predict the structure of the product formed from the cycloaddition reaction of acrolein with 1-methoxy-1,3-butadiene. Justify your prediction with reference to your calculated MOs of these molecules. Save pictures of the important MOs.

2.7.2 Glucose

Build β -D-glucose in its ring form and clean up its structure. Save a picture.

2.7.3 Azurin

Import the PDB file `/home/seth/575/Azurin_Lab2.pdb`. The protein Azurin has four copper atoms in it. Find the third one (according to the numbering of the PDB file as shown by the amino acid molecule window), expand the view to show only this copper atom and its immediate environment and change the display to show bonds and sidechains. Save a picture of the molecule from this perspective.

2.7.4 Octaethyl-Zn Phthalocyanine

Build Octaethyl-Zn-phthalocyanine. Save a picture of it. A calculation on this molecule would take several hours.

Build a smaller model of this system that may be suitable for calculations involving reactions at the Zn atom.