

6.02 Molecular Modelling (5 points)

Introduction

The need for computer modelling

Traditionally chemists have synthesized molecules and then tested their properties by experiment. In the pharmaceutical industry, for example, it was in the past common for a company to produce hundreds of compounds for every one worth pursuing into the clinic. This is naturally very expensive. If instead a compound with the desired properties can be designed merely by executing an appropriate computer program, there are obvious benefits both intellectually and commercially.

This hope has become in large part reality. All major pharmaceutical and agrochemical companies now employ specialists in computer-aided molecular design and the synthetic chemists are increasingly using these facilities for themselves.

Long before the advent of computers, chemists used models to aid their understanding of molecules. Basically these were of two types: wire models such as the Dreiding sort which represent bonds by fine metal tubes giving an idea of the molecular skeleton, and space-filling models like the CPK (Corey, Pauling, Kolthun) type which represent each atom by a sphere of the appropriate radius and give an idea of the electronic "flesh".

These types of representation have been carried over into the computer graphic molecular modelling programs, as you will see during this experiment. There are, however, major advantages in the computer displays over mechanical models. Molecular geometries can be drawn from crystallographic databases so that they are realistic rather than average values of bond lengths and angles; the display is created in moments rather than weeks for large molecules; the pictures are not subject to gravity (!); the images may be manipulated interactively and parts may be removed or added. It is possible, for example,

to 'observe' the active site of an enzyme from inside the protein. In fact the most sophisticated modelling programs rival flight simulators, which is not surprising since the hardware and many aspects of the software have much in common.

Most drugs and agrochemicals are small molecules of less than fifty atoms. Almost always the molecules are flexible and evoke their important biological response by binding to a specific site on a macromolecule, generally a protein or nucleic acid. Although it is possible to learn about the shape or conformation of the small binding partner from X-ray crystallography of the compound, or by nuclear magnetic resonance spectroscopy of solutions, these data are insufficient to understand the shape of molecules when they bind into the receptor slots of the macromolecule. The small molecules, and indeed the protein, may adapt their shapes on binding. Hence it becomes important to calculate the energy of the drug as a function of its shape so as to be able to answer the question "into what shapes could this molecule be distorted on binding without expending more than a given energy threshold?" It is this latter question which is answered by energy calculations. The computer performs the calculation and a graphics terminal can display the results.

Calculating molecular properties

Two quite distinct approaches are used in the computation of the energy of a molecule as a function of its shape. The purer of the possibilities is the use of molecular quantum mechanics to solve the Schrödinger equation in an approximate Hartree-Fock manner. These computations may be of the so-called *ab initio* variety in which all the many millions of integrals implied in the theory are rigorously evaluated. The alternative is a semi-empirical calculation, in which time is saved by replacing computed values of integrals by experimentally determined parameters. Quantum mechanical calculations are the favoured approach if, in addition to a value of the energy or stability of a molecule, we also require some other molecular properties which can be derived from the molecular wavefunction (such as dipole moment, electrostatic potential or energies of individual molecular orbitals).

A different strategy is to use the so-called molecular mechanics methods. These are purely empirical. The methods treat a molecule as a collection of balls (the atoms) and springs (the chemical bonds). Potentials are introduced for all atom-atom interactions; for the bending of bond angles and for torsional changes. The number of disposable parameters fixed by forcing agreement with experimental quantities, such as heats of formation, may be as large as a hundred, but over the years these potentials have been refined to the point where molecular calculations are reliable for many types of molecule even when they contain heteroatoms like sulphur or rings of atoms.

Once the energy is calculated, the range of shapes which a molecule may adopt may be compared for molecules which fit a receptor site of interest with the hope of learning something about the demands of the binding process.

Structure-activity relationships

In addition to computing the energy of an isolated small molecule, other calculable properties may be of utility, especially in the search for a relationship between molecular structure and biological activity.

From the molecular wavefunction which is, along with molecular energy, the output from a quantum mechanical calculation, a large variety of molecular properties may be calculated. The square of the function at any point in the space surrounding a molecule is, of course, a measure of electron density. We may also compute at the same point the interaction energy of a unit charge (the electrostatic potential) or its gradient, the electrostatic field, which is a measure of the interaction energy a dipole would have at that position. Computer graphics permit representations of these important quantities to be made using colour coding and stereoscopic views. Such graphical displays permit qualitative connections to be made between properties and activity.

For quantitative structure activity relationships ("QSARs") the numerical values of computed quantities are used in regression analyses to provide statistically supported predictions of the activity of hypothesized new compounds. The parameters used in this way include the charges on specific atoms; the electron density in bonds; the availability of electrons as judged by orbital energies; the capacity to accept electrons and perhaps most effectively the properties of the so-called frontier orbitals - those electrons which are least tightly bound by the molecule.

Protein-drug calculations

The use of computers described so far has concentrated on calculating and displaying properties of the small molecules which act biologically by binding to receptor sites on macromolecules. Computational techniques have been developed to the extent that the more logical approach of including the protein or nucleic acid macromolecule is now also possible. This adds a whole new dimension to molecular design.

The first requirement is knowledge of the molecular structure of the large molecule binding partner. In ideal circumstances this comes from X-ray crystallographic studies of crystals of the macromolecule. High resolution refined crystal structures are now available for large numbers of proteins, mostly enzymes and for some oligonucleotides. Based on these crystal structures the architecture of similar proteins can be derived by a combination of computer techniques including recognizing similar structural elements,

theoretical calculation of relative energies and computer-aided model building. This type of work is becoming increasingly important as gene sequences (and hence the amino acid secondary structure of proteins) become more readily available. Even artificial intelligence techniques are being used to predict the structures of the binding sites of receptors.

If we have a reasonable knowledge of both the macromolecular binding site and the small molecular partner then the calculation of their binding energy is a most important quantity which can come from theoretical computation. If we want to design an inhibitor of an enzyme then that inhibitor must bind to the active site much more tightly than the natural substrate. If that inhibitor is to act as a drug then the stronger it binds, the smaller the dose which will be required medicinally.

The actual calculation of binding energy can be done using the same types of computer program which are used to calculate the energetic properties of isolated small molecules. In quantum mechanical calculations, inclusion of every atom of receptor protein is too demanding in terms of computer time, so approximations such as treating atoms as point partial charges are introduced. The molecular mechanics calculations by contrast are rapid enough to cope with the tens of thousands of enzyme atoms. Both approaches have now reached a level of sophistication where the calculations of binding energy are in good agreement with experimental binding enthalpies so that novel inhibitors can now be designed.

Introduction of molecular motion

Thus far molecular design has been treated almost like engineering design: molecules must fit together and bind tightly but they have been considered to have definite static shapes. In reality, of course, molecules are in dynamic motion. When a small molecule binds to an enzyme, it is not a key fitting a lock but a throbbing vibrating pair of species which interact. These dynamic effects need no longer be ignored. Since the molecular mechanics potentials can be computed readily, the gradients of the potential and dynamic behaviour may be predicted using Newton's laws of motion. This use of so-called molecular dynamics has become particularly important now that experimental dynamic properties are emerging from nuclear magnetic resonance studies. It has added an extra level of realism to computer-aided molecular design and again is much clarified by using computer graphic displays to illustrate the motions involved.

Procedure

This experiment is run on a PC in the lower teaching laboratory. No special knowledge of computing is required.

The experiment is in two parts. In the first, you will use the molecular modelling package Spartan to complete a number of straightforward tasks, whose aim is to allow you to become familiar with the software. In the second, you will devise an exercise which uses Spartan to solve some appropriate chemical problem. You may find it easiest to propose a suitable project after you have spent some time becoming familiar with the program and have an clear understanding of what it can do, but it is essential that you discuss your ideas with a demonstrator before you start work on any project.

The project forms the bulk of this experiment, and may require several hours work with the modelling program and possibly some time spent investigating the relevant literature. Once your exercise is approved, you will:

- (i) use the molecular modelling package to investigate your topic;
- (ii) prepare a suitable write-up, discussing your results, comparing them to any found in the literature and considering their reliability.

A copy of your project will be added to the file beside the experiment, so that later students can learn from the approach you have taken, and any extra features of Spartan which you have used but are not described in these notes.

Starting Spartan

Spartan is a powerful program which can be run on a PC or Unix workstation. It requires the presence of a hardware key (a "dongle") connected to the rear of the computer. If Spartan will not run, check with the technician that the dongle is present.

To start Spartan, double click on the Spartan icon. The principle aim of this experiment is to use Spartan to solve a real chemical problem. In order to do this, you will need to become familiar with various aspects of the program, and the best way to do this is through the completion of tasks which illustrate the capabilities of the software. Print out or record in your data book any graphics or data which seem interesting, but do not print everything, otherwise you will collect huge volumes of paper!

Task 1. Building and manipulating a small molecule

To get you started, you will first build and display a small molecule, bromomethoxymethane, ($\text{CBrH}_2\text{-O-CH}_3$). Once the molecule is built, you will determine its equilibrium geometry and dipole moment. New molecules are built by selecting atomic fragments from a menu and joining them as required. As you will discover later, it is possible to specify geometry, stereochemistry, charge, multiplicity and other characteristics of the molecule.

a) Click the left hand mouse button (lmb) on the **File** menu and select **New**. (Alternatively, you can click on the **New** icon, which is the leftmost icon.) A panel of common molecular fragments will appear, with tetrahedral carbon already highlighted. Move the cursor to the middle of the green display area and click the lmb to place a tetrahedral carbon as the starting point for your structure.

b) Click on the **-Br** box in the fragment panel and then click the lmb over the end of one of the carbon single bonds to add the bromine.

c) Select from the fragment panel the singly-bonded oxygen atom and add that to the structure.

d) Finally add a second tetrahedral carbon to the free oxygen bond. All remaining free valences will automatically be completed with hydrogen atoms.

You can rotate the molecule by pressing and holding down the lmb in the display area and moving the mouse. You can move the molecule using the rmb. Try both now.

e) You can determine the molecular mechanics strain energy by selecting **Minimise** from the **Build** menu. (Alternatively, you can use the icon for this, which is the E with a downward-pointing arrow). The geometry will adjust rapidly as the calculation is in the progress, and the final strain energy will be displayed at the bottom right, together with the molecular point group. Record both of these in your data book.

The molecule can be displayed on screen in several different formats.

f) Choose **View** from the **Build** menu and note the effect of selecting different types of display from the **Model** menu. The molecule can be rotated and moved in each display type. Print out one of the structures if you wish.

g) Return to the ball and wire model. In this format (and in the wire model) it is possible to display atom and bond labels. Select **Configure Labels...** from the **Model** menu and make sure that **Labels** in the **Model** menu is ticked. Click **OK**, and then select **Labels** from the **Model** menu. You can remove labels by again selecting **Labels** from the **Model** menu.

Task 2. Determining molecular geometry

It is simple to make measurements of the geometric parameters for your molecule.

a) Return to the ball-and-spoke model. Select **Measure Distance** from the **Geometry** menu. Click the mouse on two atoms in succession, or (if you wish to measure the distance between bonded atoms) on the bond which joins them to display the distance between the atoms. Measure the length of each of the C-O bonds and record the distances in your data book. Comment on the values.

b) Determine the C-Br bond length and compare this to the average value quoted in any standard chemistry textbook.

c) Choose **Measure Angle** from the **Geometry** menu. Click three atoms in succession (or on two bonds sharing a common atom) to determine the angle between the two bonds. Determine the angles around the bromine-bonded carbon atom and comment on their values.

Task 3. MO Calculations

A variety of parameters can be found through a MO calculation. The first step required is to choose the type of calculation.

a) Select **Calculations...** from the **Setup** menu. In the dialogue box which opens up select **Equilibrium geometry** from the leftmost pull-down menu. Select **Hartree-Fock** and **3-21G** from the two rightmost menus to specify a Hartree-Fock calculation using the 3-21G split-valence basis set. Verify that the **Total Charge** is **Neutral** and the **Multiplicity** is **singlet**. Click on **OK** to remove the dialogue.

b) Select **Submit** from the **Setup** menu. A file browser appears. Choose the default name (probably "Spartan1" or something similar) and click on **Save**. This step will start your calculation off-line. Click **OK** to remove the message.

c) Once the calculation has completed (between ten seconds and several minutes will be required depending upon the complexity of the calculation and speed of the machine) you will be notified. You can, if you wish, inspect output from the job using **Output** in the **Display** menu. Of particular interest is the dipole moment. To find this, select **Properties** from the **Display** menu. Record the dipole moment and compare it to the literature value for this molecule of 2.06 Debye.

Task 4. Displaying electron density and potential energy surfaces

a) Select **Surfaces** from the **Setup** menu. Click on **Add...** Now select **density** from the **Surface** menu and **none** from the **Property** menu. Click a second time on **Add...** and this time select **density** from the **Surfaces** menu and **potential** from the **Property** menu. This specifies the calculation of an electron density surface onto which the value of the electrostatic potential has been mapped. Click on **OK**.

b) Submit the job (**Submit** from the **Setup** menu); you do not need to remove the dialogue window to do this. When the calculation is complete again select **Surfaces**. Click on the line **density Completed 0.002**. Click on the density surface. A menu will appear at the bottom right of the screen. Note the effect of changing the display style and print out one of the figures.

Select **Close** from the **File** menu to remove the molecule from the screen.

Task 5. Specifying molecular conformations; calculating and displaying Molecular Orbitals

Depending upon the project you propose, you may need to specify the conformation of a molecule, or calculate and display molecular orbitals. This section explains how those tasks can be completed.

a) Construct trans-acrolein ($\text{CH}_2=\text{CHCHO}$) as follows: Ensure any previous molecule has been removed from the screen. Select **New** from the **File** menu and click on trigonal planar sp^2 hybridised carbon. Click on the screen, rotate the molecule if necessary to identify the free double bond and click again on that to make ethylene. Click on any of the free valences to add the third carbon, then finally add a doubly-bonded oxygen.

b) In the trans molecule the two double bonds point in the same direction. If the molecule you have made is in the cis configuration, change the conformation as follows: Click on the $\backslash\? \backslash$ icon in the icon bar. Click in order on the oxygen and the three carbons. The value of the dihedral angle which these atoms define will appear in a box at the bottom right. Replace 0 by 180 and press the return key; the correct conformation will be created.

Acrolein and other α,β -unsaturated carbonyl compounds are known to undergo both nucleophilic addition at the carbonyl carbon and Michael addition at the β olefin position. We can determine which is more likely in acrolein by calculating and displaying the LUMO, into which it is likely an incoming pair of electrons in a nucleophile will be directed. (The carbon on which the LUMO is most localised will be the most susceptible to attack by a nucleophile).

a) Enter the **Surfaces** dialogue. Click on **Add...** and select **LUMO** from the **Surface** menu. Click on **OK** and submit the job.

b) When the job has completed, click on the line **LUMO Completed 0.032** inside the **Surfaces** dialogue. Print out the resulting surface and comment on whether this is consistent with experimental evidence for reactivity of α,β -unsaturated carbonyl compounds.

Task 6. Calculation and display of vibrational frequencies

It is possible to calculate and display the vibrational motions of molecules using Spartan. In this step you will do this for acetylene, which is one of the species whose IR spectrum is recorded in experiment 7.05.

a) Remove any molecules from the screen, then construct acetylene (ethyne).

b) Select **Calculations...** from the **Setup** menu. Select **Equilibrium geometry** from the leftmost menu and **Hartree-Fock** and **3-21G** from the two rightmost menus. Click on the **Frequencies** box, then click on **OK** and submit the job.

c) Once the job has completed, display the vibrational frequencies by selecting **Vibrations** under the **Display** menu. Make a note of the predicted vibrational frequencies with their symmetry, and compare with those measured experimentally for acetylene (in the low resolution spectrum of acetylene, bands appear centred around 750, 1250, 1630 and 3200 cm^{-1}). Calculated frequencies are typically high by around 10% and it is common practice to scale frequencies (by 0.9 in the case of the 3-21G basis set) to bring them more into line with experimental data. You should also bear in mind that combination and difference bands, though not calculated by this program, may appear in the spectrum.

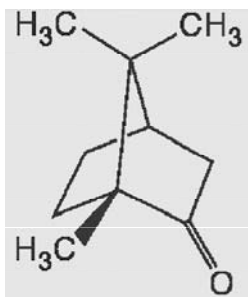
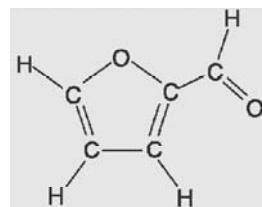
d) To animate a vibration, click on its frequency in the vibrations dialogue. You can stop the animation by clicking a second time on the frequency. You may adjust the display style to give the clearest motion. Check that the form of each vibration is consistent with what you would expect from the symmetry shown in the vibrations table. Sketch the form of each vibration, indicate which frequency is associated with which vibration, and identify those which should be Infrared active.

Task 7. Building more complex molecules

Spartan contains a small library of pre-formed molecules and fragments, which simplify the building of complex molecules. A few examples will illustrate how you can construct large molecules.

a) **Pydriene**: Click on **New** in the **File** menu. Click on **Rings** in the model kit and select **benzene**. Place it on the screen. Select aromatic nitrogen (nitrogen with two single bonds and a dotted third bond) from the atomic fragments, position the cursor over a ring carbon and double click to make pyridine.

b) **Furufral**: Click on **Rings** in the model kit and select a cyclopentane ring. Choose an Oxygen atom with two single bonds and double click on a ring carbon. To make the double bonds, select **Make Bond** from the **Build** menu. Click on one of the free valences of each of the two carbons you wish to join to make the double bond. Click on the minimise energy icon, then add the carbonyl group from the **Groups** menu.



c) **Camphor**: Remove any molecules from the screen and click on **New** in the **File** menu. Place a cyclohexane ring on the screen. Place an sp^2 hybridised carbon in the ring, then add the carbonyl oxygen. Click on sp^3 hybridised carbon and click on an axial free valence on one of the carbons adjacent to the carbonyl group. Click on the make-a-bond icon (8th from the left). Click on one of the free valences of the atom you have just added and then on the equatorial free valence on the carbon on the opposite side of the ring. Create a refined structure by clicking on the minimise icon. then add three methyl groups to complete the structure.

d) **Sulfur tetrafluoride:** SF_4 is a simple molecule to construct, but cannot be done from the standard model kit since sulfur is not in its normal bent dicoordinate geometry, but in a trigonal bipyramidal form.

Bring up the expert model kit by clicking on the expert tab at the top of the model kit. Click on **S** in the periodic table, check in the window above the table that it is 5-coordinate; if it is not, select trigonal pyramidal co-ordination from the buttons below the table. Place the five co-ordinate sulfur anywhere on screen. Select **F** from the table and the one-coordinate entry -J from the list of atomic hybrids. Click on both of the axial free valences of sulfur and two of the equatorial free valences. You need to delete the remaining free valence on sulfur, otherwise it will be assumed to be attached to a hydrogen atom. Click on the delete symbol (the 7th icon from the left) and then on the remaining free valence. Run the molecular mechanics minimisation, and check that the molecule has C_{2v} symmetry. If the symmetry is T_d instead, exit the program and restart.

Other facilities

Spartan also allows the calculation of

- low-energy conformations using Monte Carlo methods,
- enthalpies of reaction,
- barriers of rotation,
- electron distribution in LUMOs and HOMOs,
- properties of intermolecular complexes,
- the location of transition states,
- the variation of energy as reaction occurs, and
- the properties of molecules with unpaired electrons.

Details of how to accomplish these tasks may be given in the past projects which are available by the PC used in the experiment. If not, and you wish to use one or more of these features of the software, ask the technician for the Spartan Pro tutorial; note that this must be signed for and returned to the technician at the end of the experiment.

Project

Devise a suitable project, and discuss your ideas with a demonstrator before starting. After approval, use the program to carry out your project, and try to find in the literature whether or not published data are available so you can compare experiment with prediction.

Once your project is complete, and written up, provide a copy for the files to act as a resource for later students doing this experiment.

A list of past projects will be made available on the web as experience with the software grows.

References

1. W.G.Richards, Quantum Pharmacology, Butterworths, London, 1983, [Hooke Re25].
2. W.G.Richards, Computer-aided molecular design, *Sci. Progr., Oxf.* (1988), **72**, 481-492.

Further Reading

- M.J.S.Dewar and W.Thiel, *J.Am.Chem.Soc.* **99**, (1977),4899
D.M.Hayes and P.A.Kollman, *J.Am.Chem.Soc* **98**, (1976), 3335
C.A.Reynolds *et al*, Anti-cancer drug design 1, (1987) 291, *J. Chem. Soc.Perkin II* (1988), 1434, *Nature* 334 (1988), 80, *J. Chem. Soc. Commun.* (1988), 1434
P.Cieplac, U.C.Singh and P.A.Kollman, *Int. J. Quant. Chem. Obs* **14** (1987), 65
P.A.Bash, U.C.Singh and P.A.Kollman, *Science* 235 (1987), 65, *Science* 236 (1987), 564, *Nature* 328 (1987), 551
K.Jug, *Theoret.Chim.Acta* **54** (1980), 263
J.D.Head *et al*, *Int.J.Quantum Chem.* **23** (1988),177