MODELLING MOLECULAR STRUCTURES

Desktop Molecular Modeller - a:\sucrose.mol

File Style View Calculate Modify Utilities Annotation Peptides Help

New...
Open...
Save
Save As...
Delete...
Print...
Restart program...
Preferences
Exit   Alt + F4

ENERGY

Alan Hinchliffe

MODELLING MOLECULAR STRUCTURES

Alan Hinchliffe

UMIST, Manchester, M60 1D, UK

email Alan.Hinchliffe @umist.ac.uk

Series Preface

Theoretical chemistry is one of the most rapidly advancing and exciting fields in the natural sciences today. This series is designed to show how the results of theoretical chemistry permeate and enlighten the whole of chemistry together with the multifarious applications of chemistry in modern technology. This is a series designed for those who are engaged in practical research, in teaching and for those who wish to learn about the role of theory in chemistry today. It will provide the foundation for all subjects which have their roots in the field of theoretical chemistry.

How does the materials scientist interpret the properties of a novel doped-fullerene superconductor or a solid-state semiconductor? How do we model a peptide and understand how it docks? How does an astrophysicist explain the components of the interstellar medium? Where does the industrial chemist turn when he wants to understand the catalytic properties of a zeolite or a surface layer? What is the meaning of 'far-from-equilibrium' and what is its significance in chemistry and in natural systems? How can we design the reaction pathway leading to the synthesis of a pharmaceutical compound? How does our modelling of intermolecular forces and potential energy surfaces yield a powerful understanding of natural systems at the molecular and ionic level? All these questions will be answered within our series which covers the broad range of endeavour referred to as 'theoretical chemistry'.

The aim of the series is to present the latest fundamental material for research chemists, lecturers and students across the breadth of the subject, reaching into the various applications of theoretical techniques and modelling. The series concentrates on teaching the fundamentals of chemical structure, symmetry, bonding, reactivity, reaction mechanism, solid-state chemistry and applications in molecular modelling. It will emphasise the transfer of theoretical ideas and results to practical situations so as to demonstrate the role of theory in the solution of chemical problems in the laboratory and in industry.

D. Clary, A. Hinchliffe and D.S. Urch
June 1994

Preface

In the beginning, quantum chemists had pencils, paper, slide rules and log tables. It is amazing that so much could have been done by so few, with so little.

My little book '**Computational Quantum Chemistry**' was published in 1988. In the Preface, I wrote the following:

> As a chemistry undergraduate in the 1960s . . . I learned quantum chemistry as a very 'theoretical' subject. In order to get to grips with the colour of carrots, I knew that I had to somehow understand
>
> $$\left| \int \Psi^*_k \sum \underline{r}_i \Psi_0 d\tau \right|^2$$
>
> but I really didn't know how to calculate the quantity, or have the slightest idea as to what the answer ought to be . . .

and I also drew attention to the new confidence of the late 1980s by quoting

> Today we live in a world where everything from the chairs we sit in to the cars we drive are firstly designed by computer simulation and then built. There is no reason why chemistry should not be part of such a world, and why it should not be seen to be part of such a world by chemistry undergraduates.

The book seemed to capture the spirit of the 1980s, and it became accepted as a teaching text in many universities throughout the world. In those days, computing was done on mainframes, scientific programs were written in FORTRAN, and the phrase 'Graphical User Interface' (GUI) was unknown.

Personal computing had already begun in the 1980s with those tiny boxes called (for example) Commodore PETs, Apples, Apricots, Acorns, Dragons and so on. Most of my friends ignored the fact that PET was an acronym, and took one home in the belief that it would somehow change their life for the better and also become a family friend. Very few of them could have written a 1024 word essay describing the uses of a home computer. They probably still can't.

What they got was an 'entry level' machine with a simple operating system and the manufacturer's own version of BASIC. There were no application packages to speak of, and there was no industry standard in software. Anyone who wrote software in those days would have nightmares about printers and disk files.
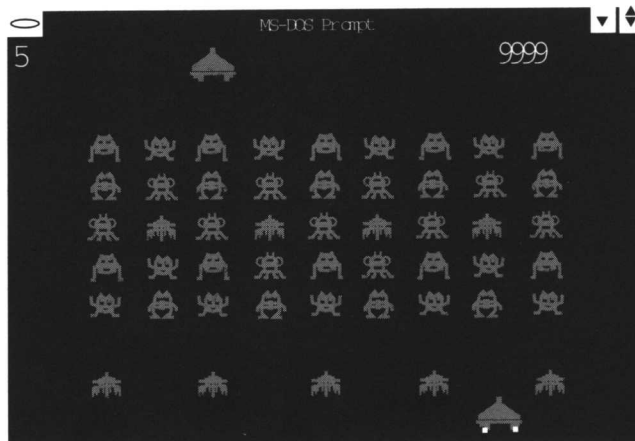
IBM (the big blue giant) slowly woke up to the world of personal computing, and gave us the following famous screen, in collaboration with MICROSOFT.

C:\ >

The DOS prompt.

Not very user-friendly!

Then came the games, and most older readers will recognise the Space Invaders screen shown below . . .

The Graphical User Interface was then born, courtesy of the the 'A' manufacturers such as Apple, Apricot, Amiga and Atari. Perhaps that is why so many of them went down the Games path. But they certainly left IBM behind.

These days we have Lemmings, Theme Park and SimCity. Many of them are modelling packages dressed up as games.

In the world of serious software, we soon saw the introduction of packages reflecting the three legs of the information technology trilogy

- word processing
- databases
- spreadsheets

and IT is now a well-established part of secondary education.

I don't want to bore you. As time went on, molecular modelling packages began to appear. Many ran under DOS (with the famous prompt screen above), but the more popular ran with GUIs on Apple Macs. Well, what happened is that MICROSOFT introduced WINDOWS, the famous graphical interface designed to protect users from DOS. There are now said to be more users of WINDOWS on IBM compatible PC's worldwide than all the other operating systems combined. But has all this actually changed our ability to understand molecules? Cynics will still argue that there have been no new major discoveries about molecular electronic structure theory since the heady days of the 1920's when Schrödinger, Pauli, Heisenberg and Dirac were active. Dirac said it all, in his oft-quoted statement

Dirac's famous statement

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that exact application of these laws leads to equations much too complicated to be soluble

But computers and computing have moved along apace. This is especially true for personal computing, whereby powerful modelling packages are now available for everyday use. Most of these packages use molecular mechanics, and these come with brilliant graphics and excellent user-interfaces. Conformational problems involving protein strands that have been tackled using these packages are becoming common-place in the primary literature.

It is interesting to note that all the simple theories (such as Hückel π-electron theory) have now reappeared as options in these very same packages! Thus, very many scientists now routinely use computational quantum chemistry as a futuristic tool for modelling the properties of pharmaceutical molecules, dyestuffs and biopolymers. I wrote the original 'Computational Quantum Chemistry' Text as an Introduction for senior undergraduates and beginning postgraduates. True, the original edition had some flaws; reviewers pointed out that there was no need for a revision of the principles of quantum mechanics, however 'brief and breakneck' (and I quote).

It seemed to me that the time was ripe for a new text that would focus on recent applications, especially those reflected in current modelling packages for PCs. Hence this book!

I have also made available a disk containing the following teaching material

- BASIC programs to do the following:
  calculations for the H_2 problem; LCAO, VB and CI treatments
  Matrix diagonalization (for simple Hückel problems, etc.)
  a PPP π-electron program
  a rudimentary CNDO/2 all-valence electron program.
- End-of-Chapter problems.
- Solutions to all the End-of-Chapter problems.

In keeping with the spirit of the 1990s, I should mention that I have made use of the following PC packages in preparing this text.

ChemDraw Pro for Windows, by mass times velocity squared limited (mc^2), 46 Solent Road, London NM6 1TX (United Kingdom).

Chemintosh/ChemWindow, by Cherwell Scientific Publishing Limited, The Magdalen Centre, Oxford Science Park, OX4 4GA (United Kingdom).

MOBY Molecular Modelling on the PC, Version 1.5, by U. Höweler, Münster, FRG. Springer-Verlag GmbH & Co. KG Heidelberger Platz 3 D-14197 Berlin (Germany).

Mathcad 5.0 for Windows, by MathSoft, Inc. 101 Main Street, Cambridge, MA 02142.

DTMM3.0 for Windows, by M. James, C. Crabbe, John R. Appleyard and Catherine Rees Lay, Oxford University Press, Walton Street, Oxford OX2 6DP (United Kingdom).

I have also made use of the GAUSSIAN92 and GAUSSIAN94 packages. The literature citations are shown with the computer outputs.

Alan Hinchliffe
UMIST,
Manchester, 1995

Contents

0 Prerequisites

0.1 WHAT ON EARTH IS A CHAPTER 0?

Let me tell you how things were in the heady days of the 1960s, when scientists (like me) and engineers first got our hands on computers. Computers were large beasts, and they consumed very many kilojoules (kilocalories in those days, or if you are a North American reader) per unit time. For this reason, they usually had a refrigeration plant, where the three resident engineers kept the milk for their coffee.

In those days, there were no packages such as GAUSSIAN92. All we had were rudimentary libraries containing matrix diagonalization 'subroutines' etc. The first step was to write your own code. We wrote programs in PIG (Programming Input General) or in an **autocode**, such as MERCURY AUTOCODE. By 'we', I mean the budding electrical engineers, the crystallographers and the very select group of quantum chemists who spent their nights and weekends running calculations on the Ferranti MERCURY. This machine occupied a large portion of the top floor of the Hicks Building in the University of Sheffield (UK).

What you would recognize as random access memory (RAM) was extremely limited on the Ferranti MERCURY, and so programs had to be segmented into 'CHAPTERS'. The first 'segment' of code was CHAPTER 0. CHAPTER 0 **had** to be executed before any other, and I hope that you will regard this chapter of my book in the same way. That is to say, you really must read this chapter before you begin the book in earnest.

The reason is quite simple. I am going to have to assume some prerequisite knowledge of the techniques of quantum chemistry. In particular, I think that you should know the following.

- The basic ideas of quantization (black body radiation, the photoelectron effect . . . that kind of thing).
- The basic postulates of quantum mechanics (stationary states, the Born interpretation of quantum mechanics which has $\psi^*\psi$ representing a probability and so on).

- Schrödinger's treatment of the hydrogen atom, and a familiarity with his famous equation for the stationary states

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{8\pi^2 m}{h^2}(\epsilon - U)\psi = 0$$

where ϵ is the electron energy and U the mutual potential energy of the electron and the nucleus.

- A knowledge of how to write the Schrödinger equation as an **eigenvalue equation**, involving the **Hamiltonian operator** H

$$H\psi = \epsilon\psi$$

- An appreciation that wavefunctions for molecules depend on the coordinates of **all** the particles present. I am going to write **capital** Ψ when dealing with many-particle systems, and lower case symbols such as ψ above when dealing with single particles such as the electron in a hydrogen atom or H_2^+ molecule-ion. I will also use χ to stand for atomic orbitals in a molecular treatment (e.g., the LCAO method).

- A knowledge of the **variation principle**. In particular, you should be familiar with variational integrals such as

$$\frac{\int \Psi^* H\Psi d\tau}{\int \Psi^* \Psi d\tau}$$

I will normally treat all wavefunctions as real quantities, so you can forget about the complex conjugate * in equations of this kind.

- An appreciation that this variational integral has dimensions of energy, and that the variational energy is always greater or equal to the true energy, depending on whether the wavefunction is an approximation to the state of a given symmetry having the lowest energy, or exactly equal to it.

- A familiarity with the variation principle in its **linear** form, where we seek an approximate wavefunction as a linear combination

$$\Psi = c_1\Psi_1 + c_2\Psi_2 + \ldots + c_n\Psi_n$$

and you should know that this leads to the so-called **secular equations**

$$\begin{pmatrix} H_{1,1} & H_{1,2} & \ldots & H_{1,n} \\ H_{2,1} & H_{2,2} & \ldots & H_{2,n} \\ \ldots\ldots \\ H_{n,1} & H_{n2} & \ldots & H_{n,n} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \ldots\ldots \\ c_n \end{pmatrix} = \epsilon \begin{pmatrix} S_{1,1} & S_{1,2} & \ldots & S_{1,n} \\ S_{2,1} & S_{2,2} & \ldots & S_{2,n} \\ \ldots\ldots \\ S_{n,1} & S_{n,2} & \ldots & S_{n,n} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \ldots\ldots \\ c_n \end{pmatrix}$$

- The concept of electron spin and the ideas of spin multiplicity.
- A familiarity with the idea of indistinguishability.
- A knowledge of the LCAO (linear combination of atomic orbitals) methodology.
- The **aufbau** principle, and the ability to write an electronic configuration for a first row diatomic such as O_2.

Also, you will need the concept of a **vector**. As you should know, vectors are quantities having magnitude and direction, and I am going to represent them by

symbols such as $\underline{r}$ and $\underline{R}$. The point is that we need to describe the positions of electrons and nuclei in space, and the language of vectors turns out to be the most succinct.

0.2 BASIC ELECTROSTATICS

At various points throughout the text, I will need to refer to some of the basic concepts and results of classical electrostatics. This is a field of human endeavour that deals with the forces between electric charges at rest, the fields and electrostatic potentials produced by such charges, and the mutual electrostatic potential energy produced by a pair of charge distributions. To get us started, consider two point charges Q_A and Q_B, as shown in Figure 0.1.
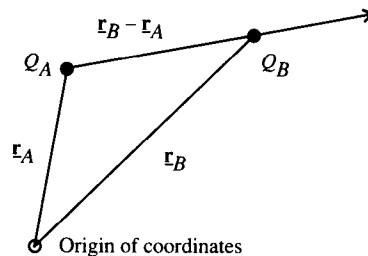

Figure 0.1 Construct needed to discuss the force between the point charges Q_A and Q_B

Point charge Q_A is located at vector position $\underline{r}_A$ and point charge Q_B is at vector position $\underline{r}_B$. The vector joining the positions of Q_A and Q_B is also shown; $\underline{r}_B - \underline{r}_A$ points in the direction **from A to B** as shown.

I am going to quote a number of results without proof, and I want you to bear them in mind:

- Point charge Q_A exerts a force $\underline{F}(Q_A$ on $Q_B)$ on Q_B given by

$$\underline{F}(Q_A \text{ on } Q_B) = \frac{Q_A Q_B}{4\pi\epsilon_0} \frac{(\underline{r}_B - \underline{r}_A)}{|\underline{r}_B - \underline{r}_A|^3}$$

- Point charge Q_A generates a field $\underline{E}$ at points in space, and in particular it generates a field at the position in space $\underline{r}_B$ where I originally placed Q_B. The field generated by Q_A exists irrespective of the presence or absence of Q_B, and I normally omit all mention of Q_B (Figure 0.2), and just focus attention on that particular point in space. The field at point $\underline{r}$ is written $\underline{E}(\underline{r})$

$$\underline{E}(\underline{r}) = \frac{Q_A}{4\pi\epsilon_0} \frac{(\underline{r} - \underline{r}_A)}{|\underline{r} - \underline{r}_A|^3}$$
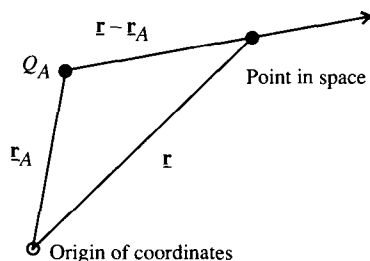
Figure 0.2   *Construct needed to discuss the field generated by Q_A at a point in space*

- Point charge Q_A generates an electrostatic potential at points $\underline{r}$ in space, given by

$$\varphi(\underline{r}) = \frac{Q_A}{4\pi\epsilon_0}\frac{1}{|\underline{r} - \underline{r}_A|}$$

The potential at $\underline{r}$ is written $\varphi(\underline{r})$.

- The **work done** in building up the charge distribution shown in the Figure 0.1, where the two point charges are initially at infinity, is called the **mutual electrostatic potential energy** $U(Q_A, Q_B)$ of Q_A and Q_B. For the pair of charges in my diagram it is

$$U(Q_A, Q_B) = \frac{Q_A Q_B}{4\pi\epsilon_0}\frac{1}{|\underline{r}_B - \underline{r}_A|}$$

When we deal with charge distributions rather than point charges, the definitions have to be generalized. What we do is to divide these charge distributions into differential charge elements $\rho(\underline{r})d\tau$, work out the contribution to the force, field, potential . . . made by this differential charge element, and then sum over them.

For example, if we were dealing with a charge distribution $\rho_A(\underline{r})$ instead of point charge Q_A (Figure 0.3) and we wanted to calculate the electrostatic field generated by ρ_A at the point $\underline{r}$ in space, then we would divide up the charge distribution ρ_A into differential charge elements $\rho_A d\tau$, and apply the basic formula for the electrostatic field. Here, $d\tau$ is a volume element. Finally, we would have to integrate over the coordinates of the charge distribution, in order to find the total field. This might not be easy!

$$\underline{E}(\underline{r}) = \int \frac{\rho_A(\underline{r}_A)}{4\pi\epsilon_0}\frac{(\underline{r} - \underline{r}_A)}{|\underline{r} - \underline{r}_A|^3}d\tau_A$$

Likewise to find the mutual potential energy of a pair of charge distributions $\rho_A(\underline{r}_A), \rho B(\underline{r}_B)$ we would have to calculate

$$U(\rho_A, \rho_B) = \int\int \frac{\rho_A(\underline{r}_A)\rho_B(\underline{r}_B)}{4\pi\epsilon_0}\frac{1}{|\underline{r}_B - \underline{r}_A|}d\tau_A d\tau_B$$

The integration would have to be over the volume of charge distribution A and charge distribution B.
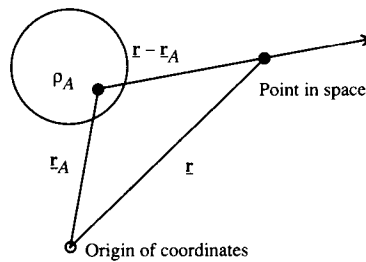
Figure 0.3 Construct needed to discuss the electric field generated by a charge distribution ρ_A

0.3 SYSTEMS OF UNITS

0.3.1 The Système International

It is usual these days to express all physical quantities in the system of units referred to as Système International–SI for short. SI units, which are recommended by the International Union of Pure and Applied Physics, are based on the metre, kilogram and the second as the fundamental units of length, mass and time. The SI electrical units are determined from these, via the fundamental constants ϵ_0 and μ_0, the permittivity and permeability of free space respectively. The ampere is defined in terms of the force between two straight parallel conductors placed a metre apart, and once this has been defined the coulomb must be such that one coulomb per second passes along a conductor if it is carrying a current of one ampere.

It turns out that the speed of light c_0, ϵ_0 and μ_0 are interrelated

$$c_0 = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

and since 1983 the speed of light has been **defined** in terms of the distance that light travels in one second. It has the exact value

$$c_0 = 2.99792458 \times 10^8 \, \mathrm{m \, s^{-1}}$$

We **are** going to be concerned with electrical and magnetic properties, so I had better put on record the fundamental force laws for stationary charges and steady currents. These are as follows.

$$\underline{F}(Q_A \text{ on } Q_B) = \frac{Q_A Q_B}{4\pi\epsilon_0} \frac{(\mathbf{r}_B - \mathbf{r}_A)}{|\mathbf{r}_B - \mathbf{r}_A|^3}$$

is the electrostatic force exerted **by** point charges Q_A **on** point charge Q_B, where $\mathbf{r}_A$ is the position vector of Q_A and $\mathbf{r}_B$ the position vector of Q_B. I discussed this above, and you should be aware that this force is exactly equal and opposite to the force exerted by Q_B on Q_A.

The corresponding force between two complete electrical circuits A and B is

$$\underline{F}(A \text{ on } B) = I_A I_B \frac{\mu_0}{4\pi} \oint \oint d\underline{l}_B \times (d\underline{l}_A \times \frac{\underline{r}_B - \underline{r}_A}{|\underline{r}_B - \underline{r}_A|^3})$$

which is much more complicated, because the integrations have to be done around the complete electrical circuits. The details do not matter, the point being this. Because ϵ_0 and μ_0 are interrelated, we are free to give one of them an arbitrary value, and in SI we choose arbitrarily to make

$$\mu_0 = 4\pi \times 10^{-7} \, \text{H m}^{-1}$$

0.3.2 Gaussian units

The most commonly used system apart from SI is the Gaussian system, sometimes called 'cgs' because it is based on the centimetre, gram and second. The unit of force is the dyne, and the unit of energy the erg.

There is usually no problem in converting between SI and Gaussian units, until we have to consider electrical phenomena. In cgs we take the proportionality constant in Coulomb's law to be unity (a number)

$$\underline{F}_{\text{cgs}}(Q_A \text{ on } Q_B) = Q_A Q_B \frac{(\underline{r}_B - \underline{r}_A)}{|\underline{r}_B - \underline{r}_A|^3}$$

and this means that derived equations have a different form. The unit of charge is called the **electrostatic unit** (esu). When two charges each of magnitude 1 esu are separated by a distance of 1 cm, each experiences a force of 1 dyne. The electric field is measured in statvolts cm^{-1}.

As a rule of thumb, be wary of equations which have an $(-e)^2$ but no $4\pi\epsilon_0$, and of equations that relate to highly symmetrical charge distributions but seem to have a 4π too many.

It gets worse with magnetic properties, and the Lorentz force

$$\underline{F} = Q(\underline{E} + \underline{v} \times \underline{r})$$

is written in such a way as to make the magnetic field $\underline{B}$ have the same dimensions as the electric field $\underline{E}$, namely force per unit length. In cgs units, the Lorentz force law becomes

$$\underline{F} = Q(\underline{E} + \frac{1}{c_0} \underline{v} \times \underline{r})$$

so as a final *aide-mémoire*, beware of magnetic equations that have a c_0 in them.

A quick conversion table is given in Table 0.1. It isn't comprehensive, but you should find it useful.