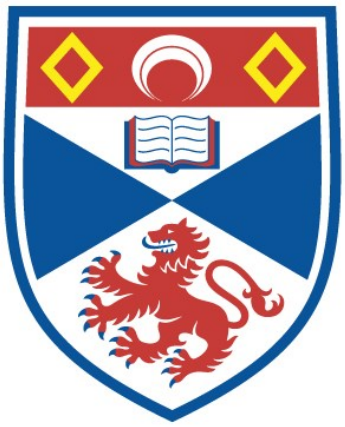
THEORETICAL STUDIES OF STEROID HORMONES AND RELATED COMPOUNDS

Michael Hugh Charlton

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1992

Theoretical Studies of Steroid Hormones and Related Compounds.

A Thesis Presented for the Degree of
Doctor of Philosophy
in the Faculty of Science of the
University of St. Andrews.

by Michael Hugh Charlton, B.Sc. (Hons).

St. Leonard's College,
September 1991.

Th
B47

Abstract.

A theoretical study of steroidal inhibitors of the enzymes Glucose-6-Phosphate Dehydrogenase and Aromatase is presented. Both enzyme systems are of interest in the study of cancer, the latter being the final step in the biosynthesis of oestrogens which are involved in certain types of breast cancer. Two levels of theory are employed in the study, namely, Ab Initio and Semi Empirical methods.

Structures and charges have been calculated using the MOPAC and GAUSSIAN programs and these have been used to model the efficacy of various inhibitors. The major tool in comparing these steroids has been the molecular electrostatic potential (MEP). Maps of the MEP and an analysis of the similarity between the MEP s of different molecules have led both to a method of assessing the activities of steroids as enzyme inhibitors and requirements for the electronic structure of the steroid binding sites within these enzymes.

A molecular graphics display program has been developed to facilitate this work. It has been designed to make full use of the facilities available. The quality of the resulting display has improved greatly on what was previously available and has been of value in studies of large molecular systems. The program is written in VAX FORTRAN and uses the Graphics Kernel System (GKS) to produce graphical output and should be reasonably easy to transfer to other systems.

Finally, to determine whether PM3 really is a significant advance on AM1, a comparison of the two semi empirical methods is presented. The calculated properties of steroid hormones are compared to those of both Ab Initio calculations and experimental determinations, allowing the quality of the semi empirical predictions to be assessed.

Declarations.

I, Michael Hugh Charlton, hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

19-Sep-91

Michael Charlton.

I was admitted to the Faculty of Science at the University of St. Andrews under Ordinance General No. 12 on 1st October 1988 and as a candidate for the degree of PhD on the 1st October 1989.

19-Sep-91

Michael Charlton.

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the degree of PhD.

30 Sept 1991

Colin Thomson.

Copyright.

Acknowledgements.

I would like to thank Colin Thomson for the opportunity to carry out this project and for his advice, the S.E.R.C. for financial support and the N.F.C.R., A.I.C.R. and Digital for their support of the group. I have also had useful discussions with other members of staff, and would like to thank David Calvert, Chris Glidewell, Gordon Harris, Kenneth Harris, and Gordon Woolley. Thanks are also due to John Walton, whose question sparked off the idea for the Geometry Optimisation study. Other members of the department deserving mention are Andrew Leech for a helpful discussion, and Jim Bews for his skill in keeping the Macintoshes under control.

The past and present members of the group have been invaluable, and I would like to mention Paola, Piera, Derek, Donatella and John, but particularly my great friends Andrew and Julie. Audrey and Catherine deserve a mention, just for listening to my ideas and pretending that I made sense.

The encouragement and financial assistance of my parents has kept me going both mentally and physically, and I should like to thank them both.

Finally, the greatest thanks must go to Sarah, whose help with typing, references, spelling and grammar have been invaluable. She has had to become a minor expert in Theoretical Chemistry to understand me, and a disaster relief worker to clear up the trail of destruction following in the wake of this thesis.

Contents

Appendices.

Chapter One

Introduction.

The topic of this thesis is a study, using the methods of theoretical chemistry, of some of the processes involved in cancer. This chapter gives a short outline of the disease, and briefly indicates how steroid hormones (a field of particular interest in this work) are involved. It also describes how computational chemistry can be used to study biological systems, and gives an outline of how it will be applied to the study of the mechanism of action of steroids in the chapters of this thesis.

(A) The Occurrence of Cancer.

Cancer occurs in all multi - cellular life-forms, and its manifestation has been recorded since the era of the ancient Egyptians.[1,2]. After cardio - vascular disease, it is the second largest cause of death in the Western World. In one year (1986) in the UK, various forms of cancer were responsible for the 25.3% of male and 22.6% of female deaths, whilst in 1987, more than half a million Americans died from cancer and one million contracted the disease [3].

Cancer is considered not to be one single disease. It consists of a variety of disorders with many different causes, although this has been a subject of debate. The types of cancer leading to the greatest number of deaths in the West are cancer of the lung, large intestine, breast, prostate, pancreas and stomach [4]. However, this ranking varies in different parts of the world. For example, skin cancer is more than 200 times as prevalent in white Australians as it is in Bombay, and breast cancer is approximately four times more prevalent in the USA as it is in Senegal [5].

Breast cancer also serves to illustrate the fact of a hereditary influence in cancer, in that the daughters of women who have had breast cancer are twice as likely to develop the disease as other patients. This may be partly due to environmental factors because both generations may have been brought up in a similar manner, although, in general, many changes in lifestyle occur between one generation and the next.

One further fact of interest is that the normal incidence of cancer in migrant populations changes within a few generations to follow the pattern of cancers of the host country to which they move.

These factors imply that the causes that contribute to cancer include dietary, hereditary or environmental, and the control of carcinogens (compounds that cause cancer) in the diet, via adopting more healthy eating habits, might be an effective way to reduce the total incidence of the disease.

(B) Cell Growth and Cancer.

Cancer is the uncontrolled growth of a group of cells. This group, which is the progeny of a single cell, has somehow developed the ability to ignore the usual inter - cellular messages that control cell growth. For example, cultures of cancerous cells spread throughout the culture medium in an uncontrolled manner, whilst normal cells stop growing when they touch one another, a process known as contact inhibition. This growth, which is normally rapid in cancerous cells, eventually invades the surrounding tissue and organs, which is the reason for the tissue damage caused by cancer. Most cancers also undergo a process known as metastasis, which involves the spread of cells to remote sites in the body, where they grow into secondary tumours.

This is the difference between malignant and non-malignant tumours. The latter are non-invasive, encapsulated [2] and exhibit slow growth. Nevertheless, it should be noted that the accelerated rate of growth in malignant tumours can sometimes be no faster than that of normally rapidly dividing cells in various parts of the body, such as in hair follicles and in the gut.

(C) Carcinogenesis.

(C1) The Multi-stage Process of Carcinogenesis.

One of the most commonly observed features of carcinogenesis is the delay between the exposure to the carcinogen and the appearance of tumours. This latency has been demonstrated in numerous of cases [6,1], and was rationalised by the introduction of the concept that carcinogenesis can be divided into more than one stage. Initially, the terms 'initiation' and 'promotion' were used to describe a two stage process. To this, the concept of tumour progression can be added as a third stage.

Certain chemical compounds are found to have only initiating or promoting properties. In such cases, the production of tumours only occurs from the application of an initiator, followed by repeated exposure to a promoter. Neither the promoter nor the initiator alone are capable of producing tumours, and generally the reversal of the order of application is similarly ineffective. The application of the promoter following a long delay after initiation is still capable of causing cancer. This implies that once a cell is initiated, the process is irreversible. This effect can also explain the latency period of tumour growth, in that the appearance of cancer after initiation has to be followed by the promotion, which may take many years.

Carcinogenesis is the process that causes normal cells to become cancerous. It effects the genetic material (DNA) of a cell in such a way as to lead to uncontrolled growth of that cell and its progeny.

The environmental factors that lead to carcinogenesis are many, and include exposure to chemical carcinogens, ionising and ultraviolet (UV) radiation and viruses. Viruses can cause cancer because their genetic material can become integrated with that of the host [6], for example, as is the case with mouse mammary tumour virus, although the evidence for viral involvement in carcinogenesis in humans is not

conclusive. The effects of UV and shorter wavelength (as well as particulate) radiation are due to mutations and the destruction of sections of DNA leading to improper gene expression and reproduction. The effect depends upon the dose received and the ability of the cell to repair the damage to the DNA.

Like radiation, carcinogenic chemicals also damage the DNA. The classic case of the determination of a chemical cause of cancer was that of Percival Pott in 1775 [1], which reported the high occurrence of scrotal cancer in chimney sweeps due to the exposure to Polycyclic Aromatic Hydrocarbons (PAH s) in coal tar. It has since been found that such compounds cause cancer by inducing errors in DNA transcription, because the planar nature of these molecules enables them to lie between the successive spirals of the DNA helices. Some compounds cause damage by reacting with the DNA base pairs, as is the case with nitrogen mustards, which cause alkylation of the genetic material [2]. Other highly potent carcinogens include aromatic amines, vinyl chloride, nitrosamines, aflatoxins and certain metals and their compounds. The mode of action of many of these molecules is often unknown, and they do not all act in a similar manner.

(C2) Pro - Carcinogens.

One feature of chemical carcinogens that is particularly important is that not all such compounds are carcinogenic in themselves, but must be metabolised to an active form to have any effect. Again, PAH s are an example of pro - carcinogens, which must be both hydoxylated and require the formation of an epoxy - bridge before they can intercalate with the DNA. This reaction actually occurs as part of the process of detoxification of these compounds, in which the hydroxy groups are added to make the substance more water soluble to allow excretion [6]. This feature is of interest in the study of anti - carcinogenic effects of the steroid Dehydroepiandrosterone (DHEA),

which has been shown to have a protective effect against PAH s, and is discussed later in this thesis.

(D) Steroid Hormones and Cancer.

(D1) Introduction to Steroids.

Steroids play an important role in a wide range of physiological processes, such as sexual development, cell growth, metabolism of proteins, carbohydrates, lipids and nucleic acids, as well as maintaining cell water and electrolyte balance.

The basic structure of a steroid is a four - ringed system, one of these being a five - membered ring, whilst the others are six - membered (Figure 1.1).

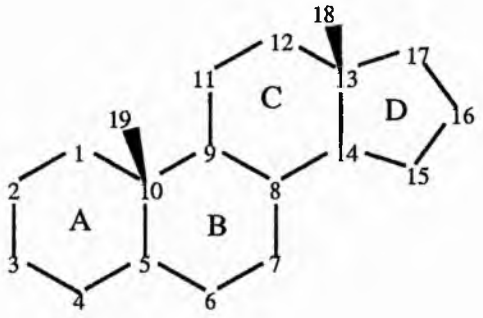


Figure 1.1

The steroid backbone, showing the ring nomenclature and numbering system.

The back bone shown in Figure 1.1 is that of Androstane, and the nomenclature of many of the steroids studied in this work is based upon this name. Substituents are named using the usual chemical terms, and designated α if they are below the ring system (as viewed in Figure 1.1) or β if they are above. Double bonds are indicated either with a Δ symbol or simply with a number. For example a double bond between C_4 and C_5 in Androstane could give Δ^4-Androstene or Androst-4-ene.

Another basic ring system studied is that of pregnane, which has a C_2H_5 chain attached at position 17.

Exceptions to the rules are the molecules derived from Testosterone, which is 17β-hydroxyAndrost-4-ene-3-one. The importance of these steroids merits a separate system of naming based upon Testosterone itself.

Oestrogens, which are responsible for the development and maintenance of female sexual characteristics, lack the angular C_{19} methyl group. They form another basic group of steroids, and are characterised by an aromatic A-ring (Figure 1.2).

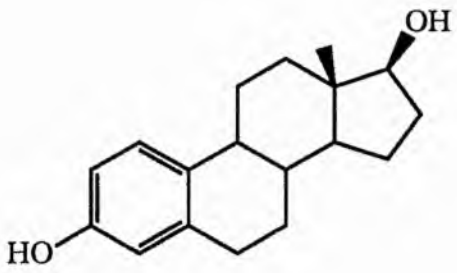


Figure 1.2

Oestradiol, an example of an oestrogen.

(D2) Steroid Action.

The mechanism of action of steroids is mediated by various macro - molecular receptors which were traditionally thought to be located in the cytoplasm of the cell. The binding to a receptor is extremely steroid specific, and minor structural modifications to a hormone can make it totally inactive [7].

Once the steroid has bound to the receptor, the complex undergoes a conformational change and moves into the cell nucleus, where it bonds to the DNA and induces its biological response. More recently data have been published contradicting

this traditional two-stage mechanism with the discovery that steroid hormone receptors have been found in the nucleus without the associated steroid ligand. Thus the need for the process of translocation of the steroid-receptor complex to the nucleus is redundant. [8].

It should be noted that this physiological response is only observed in cells that contain the appropriate receptor and neither steroid nor receptor alone are sufficient to cause such an effect [7].

(D3) Involvement of Steroids in Cancer.

Different steroids are involved in both the therapy and the genesis of cancer. The presence of certain steroids that are not within the normal biological range of concentrations, can lead to the production of tumours, whilst others inhibit certain enzymes necessary (often indirectly) for tumour growth, and lead to tumour regression.

(D4) Steroid as a Cause of Cancer.

It has also been known since the end of the last century, when Beatson found a link between breast cancer and the ovary [9,10], that steroids are involved in the genesis of this cancer. It is also known that androgens can induce the production of prostatic tumours [5]. Another relevant fact is that the risk of contracting breast cancer is increased by an early onset of menarche or late onset of menopause [5]. This effect is mediated though the greater exposure to oestrogens during the extended reproductive period of life associated with these two effects.

One possible mechanism by which steroids cause cancer is that if an organ is under hormonal control, and excessive steroidal stimulation occurs, then neoplasia can

arise [11]. This is because the steroid often induces a higher rate of cell division, and hence an increased chance of an error in DNA copying, or the reproduction of a cell that has already been initiated but was previously lying dormant. [1]

(D5) Steroids and Cancer Therapy.

Steroids are used extensively in cancer therapy. Some tumours are hormone dependent, including those of the breast and female genital tract, which means that the presence of the steroid is necessary for tumour growth, and remission can occur if the source of the steroid is removed [7]. For example, about one third of breast cancers (which caused the deaths of 570,000 people in 1980 [3]) are oestrogen dependent [9] and are responsive to the appropriate endocrine manipulations. The other two thirds must be treated by other means, which are not within the scope of this work.

Hence, molecules can be designed to antagonise the normal steroid - receptor binding process by introducing modifications to the structure of the natural substrate of an enzyme. Thus, they bind to the enzyme active site, but do not induce the usual biological effect. In a similar manner, drugs can be designed to interfere with the normal steroid - receptor binding process by blocking the access of the steroid to the binding site.

(D6) Aromatase Inhibitors.

Another method of control of steroid dependent cancers, particularly effective if these tumours have acquired the ability to synthesise the hormones themselves, is the inhibition of the production of the steroids. The rate limiting step in oestrogen synthesis is performed by an enzyme called Aromatase or oestrogen synthetase [9] and this is a suitable system for cancer therapy via inhibition of steroid production. This is

the subject of Chapter Six which attempts to rationalise the design of Aromatase inhibitors.

(D7) Dehydroepiandrosterone.

Dehydro epiandrosterone (DHEA) Figure 1.3 is a particularly interesting steroid in that it seems not to have any particular direct role in human physiology, but nevertheless it has many beneficial effects, including anti - carcinogenicity and anti - obesity action, and it is also active in preventing viral infection.

These factors will be described in Chapter Six, in which DHEA and a range of related steroids are discussed, and an attempt to correlate their activity as inhibitors of the enzyme Glucose-6-Phosphate Dehydrogenase (G6PDH) with various calculated properties will be presented.

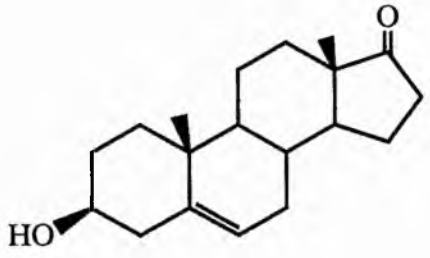


Figure 1.3

DHEA

(E) The use of Theoretical Chemistry in Biological Studies.

Theoretical chemistry has been used to assess the biological activity of a wide range of problems by many workers. It enables the prediction of molecular structures and properties which can be correlated with various activities of a range of molecules.

These properties are often not available through any other means, because the experiment has not been performed, or the result is difficult to measure accurately. They can also be calculated without the molecule ever having been synthesised, and hence theoretical chemistry can be used to pre - select a number of compounds from a list of potential drugs without the need for expensive synthesis and testing.

Some examples of the use of theory relevant to biological situations include the study of nucleic acids and their constituent compounds by the Pullmans [12], the study of hydrogen bonding by Kollman [13] and the investigation of the sweet - taste receptor [14]. Examples of the use of theoretical chemistry in the study of carcinogenesis include examinations of the activities of PAH s [15, 16, 17]. A description of the theory used to make these predictions is presented in Chapter Two.

(E1) Limitations of Different Theoretical Methods.

The time taken to perform a theoretical calculation is related to the quality of the results we wish to obtain. This relationship is often not linear in nature, as will be discussed in Chapter Two, with the most accurate results taking many times as long to calculate. It is therefore necessary to balance the time (and hence the monetary cost) of the calculation against the accuracy of the results we require. In general, we wish to select the theoretical model that will give consistently reliable results in the shortest time.

Although the rapidly increasing power of modern computers makes ever more sophisticated calculations possible, approximations still have to be made. In particular, the use of semi - empirical Quantum Mechanical calculations (Chapter Two) is particularly fast. If the results are suitably reliable, such methods would save a lot of time over Ab Initio work (see Chapter Two for a discussion of these terms). Two such semi - empirical methods have been used to perform much of the work presented

in this thesis, (Chapters Five and Six) and Chapter Seven presents a comparison of these two.

(F) Molecular Graphics.

An equally important aspect of assessing biological activity is the property of molecular shape. A large part of drug design involves the use of enzyme inhibitors and receptor agonists and antagonists. Enzymes and receptors are highly specific in shape and will only admit molecules of a certain shape in the same manner as a lock will only accept a particular key.

The use of molecular graphics has revolutionised the design of such drugs because it allows easy visualisation of the 3D shape of a molecule. If the enzyme / receptor structure is known, then it is simple to tell which molecules are sterically unable to fit into the active site or receptor pocket using molecular graphics.

Graphics also allow the display of molecular properties in an easily interpretable way, such as in the form of contour maps and surface plots. Thus it is a simple matter to compare these properties for a range of compounds.

Chapter Three gives a description of a molecular modelling program written to take advantage of the graphics facilities that have been available during the course of most of this work. Other members of the research group have used this program extensively and it has been indispensable in the modelling of large molecular systems such as proteins [18] and cell membranes [19].

(F1) Geometry Comparison.

An additional feature of this program that has been used in the discussion in Chapter Seven is that it enables the comparison of structures of a molecule obtained from many different theoretical and experimental sources. This is done by plotting various graphs of the differences between the structures and should enable the determination of which of the different computationally cheap methods best model more accurate calculations and experimental geometries. This geometry comparison program, being incorporated into the molecular modelling program, is discussed in Chapter Four.

(G) Equipment, Program Details and Methods.

Theoretical chemistry calculations are very computationally intensive and require sophisticated software and hardware. Chapter Two includes a brief description of the programs used, whilst the computing facilities used in the work described in this thesis are described below. Chapter Four discusses a particularly useful tool used in evaluating biological activity called the Molecular Electrostatic Potential (MEP). This chapter also gives an analysis of the geometry optimisation process used in predicting molecular structure, and discusses the level to which the optimisation needs to be taken to give reliable results.

(G1) Computer Hardware Within the Research Group.

The Computational Chemistry Group at the Chemistry department of the University of St. Andrews has access to a number of machines. Firstly, there is a MicroVAX II / GPX colour workstation, which was donated to the group by DEC, running the VMS operating system, version 5.3-1. It has 13 MBytes of main memory (or RAM) and 370 MBytes of disk space, divided across 3 drives of 90, 140 and

140 MBytes each and these are used for swap space, user files and temporary (or scratch) space used to store the intermediate results of the Quantum Mechanical calculations. Graphics display is performed upon the workstation terminal, which is able to display up to 32768 colours at once, from a selection of over 16 million. Graphics can also be displayed on the Tektronix 4109 and Digital VT340 16 colour terminals, as described in Chapter Three. Hard copy is available either by a screen dump using the Tektronix companion printer, or upon a Digital LJ250 colour printer.

The majority of the calculations performed in this research have been performed upon an FPS-500 mini - supercomputer. This runs the Unix operating system, and is a dual processor machine, with one Scalar and one Vector processor. The latter allows a number of equivalent calculations upon array elements to be performed simultaneously, rather than sequentially and hence greatly improves upon computation speed for such calculations. The FPS has 128 MBytes of RAM and over 900 MBytes of disk space, although due to the necessity of setting aside a large space for scratch work, the available user space is quite restricted.

The scalar processor on this machine generally performs at an average rate of about 15 MFLOPS (million floating point operations per second), but can peak at over 20, whilst the vector processor generally runs at 4 or 5 MFLOPS. This gives the machine an overall performance of over 40 times faster than the MicroVAX for certain codes.

(G2) Computers Facilities of the University of St. Andrews.

For about ten years, the University has provided a service based upon two VAX 11/785 s, which have now been decommissioned, but upon which some of the preliminary work was performed. This system has been superseded by a network of both black and white and colour Sun workstations, of which eight are based in the

Chemistry department. These have been in place since the Autumn of 1990, and there are several similar groups situated at a number of sites about the University. The Chemistry cluster (like the others), also runs the Unix operating system, and consists of one colour and seven black and white machines connected to a common fileserver. This consists of two disk drives providing about 600 MBytes each. The more powerful workstations have 16 MBytes of RAM, whilst the smaller ones have 8 MBytes.

These machines are less powerful than the FPS (the fastest runs at about one sixth of the speed), but the large number of workstations makes them an attractive option for the concurrent execution of a large number of smaller jobs with lower memory requirements, whilst leaving the FPS free to perform the more sophisticated calculations.

Chapter Two

Theoretical Methods And Computational Chemistry Software.

(A) Classical Mechanics.

At the beginning of the twentieth century, classical Newtonian mechanics was found to fail to account for black body radiation, the Photo - Electric effect, Compton scattering of radiation, Atomic Spectra and the diffraction of particles. Planck solved the problem of black body radiation by introducing the concept of quantisation, which restricted the values of the energy of a particle to discrete values, rather than a range of continuous ones. This notion of energy occurring in these discrete packets, or quanta, also led to solutions to these other problems, and represents one of the major breakthroughs of modern physics.

(B) The Schrodinger Wave Equation.

In 1926 Schrodinger provided a unifying solution to the shortcomings of classical theory, introducing the Schrodinger Wave Equation. Because all the systems studied in this thesis involve stationary states, only the time independent equation is used [20] :

$$\mathbf{H}\Psi(\underline{r})=E\Psi(\underline{r}) \qquad (2.1)$$

where :

Ψ = The exact non - relativistic wavefunction describing the system.

$\underline{r}$ = Co-ordinates of the particles in the system.

H = The Hamiltonian Operator.

E = The energy of the system, which is the sum of kinetic and potential terms.

In quantum mechanics, it is usual to make use of atomic units, which lead to a simplification of the equations because quantities are expressed in terms of physical properties of elementary particles. In this system of units, lengths are measured in terms of a_0 (= 0.529 Angstroms), known as the Bohr. Mass is given in multiples of the electronic mass, m_e (= 9.110×10^{-31} Kg), charge as multiples of the electronic

charge, e (= 1.602 x 10^{-19} C). Energies are measured in units of Hartrees (= 4.360 x 10^{-18} J, or 27.21 eV) where one Hartree is equivalent to twice the ground state energy of a Hydrogen atom.

Thus the Hamiltonian operator for a system of N nuclei of charge Z and n electrons [21] is :

$$\mathbf{H} = \sum_{A>}^{N}\sum_{B}^{N}\frac{Z_A Z_B}{r_{AB}} + \sum_{i>}^{n}\sum_{j}^{n}\frac{1}{r_{ij}} - \sum_{A}^{N}\frac{\nabla_A^2}{2M_A} - \sum_{i}^{n}\frac{\nabla_i^2}{2} - \sum_{A}^{N}\sum_{i}^{n}\frac{Z_A}{r_{Ai}} \qquad (2.2)$$

Lowercase subscripts refer to individual electrons, whilst uppercase ones refer to nuclei. This convention will be used throughout this chapter. Hence, r_{ij} are the inter-electronic separations, whilst r_{AB} are the inter-nuclear separations. M_A are the nuclear masses. The five terms are as follows :

1) Inter-nuclear repulsion.

2) Inter-electron repulsion.

3) Nuclear kinetic energy.

4) Electronic kinetic energy.

5) Nuclear-electronic attraction.

This equation neglects smaller magnetic interactions, such as Spin - Orbit and Spin - Spin Couplings, which are extremely small in magnitude in comparison to the terms in the Hamiltonian, and have little effect upon chemical bonding. For this reason, such effects are not pertinent to the work described in this thesis.

(B1) The Born - Oppenheimer Approximation

From the nature of the Hamiltonian, it can be seen that Ψ for a system of N particles (electrons and nuclei) is a function of the 3N co-ordinates of the system. The problem can be reduced in size to an electronic one, since the mass of the nuclei is

many times greater than that of the electrons, and hence their motions are much slower. This means that they are essentially clamped, and the fast moving electrons can be thought of as interacting with the fixed potential field provided by the nuclei [21, 22, 23].

Thus we have a simplified electronic problem, in which the nuclear Kinetic Energy is neglected, and the inter-nuclear repulsions are constant for a given nuclear configuration :

$$\mathbf{H}_{el}\Psi_{el}(\underline{r}) = E_{el}(\underline{r})\Psi_{el}(\underline{r}) \qquad (2.3)$$

where $\underline{r}$ are now the electronic co-ordinates, $\mathbf{H}_{el}$ is the electronic Hamiltonian, and E_{el} is the electronic energy. This is the Born - Oppenheimer approximation. From now on the electronic wavefunction, Ψ_{el}, will be refered to simply as Ψ.

The total energy is the sum of the electronic energy and the nuclear repulsion term :

$$E_{TOT} = E_{el} + \sum_{A>B}^{N}\sum^{N}\frac{Z_A Z_B}{r_{AB}} \qquad (2.4)$$

The electronic energy, and hence the total energy depend parametrically upon the nuclear co-ordinates. This dependence defines a potential energy hyper - surface for the molecule. For an N-atom molecule, the energy is a parametric function of 3N - 6 (or 3N - 5 for a linear molecule) atomic co-ordinates. In the case of a diatomic molecule, this surface is a simple Morse curve, being a plot of energy versus nuclear separation. For larger systems, the surface is impossible to visualise, and must be examined by studying the variation of energy with only one or two atomic positions at a time.

This approximation greatly simplifies the problem, and holds for most chemical situations, exceptions being where the motion of the nuclei is important, such as the Renner effect [20], in which deviations in spectra occur due to the interaction of vibrations with electronic excitations.

(C) Properties of the Wavefunction.

(C1) Electron Spin.

An electron not only has an angular momentum due to its motion about the nuclei, but it has been experimentally observed that it has an intrinsic angular momentum of its own. This concept is required to account for facts such as the fine splitting of spectral lines. It requires the electronic wavefunction to be a function of both spatial and spin co-ordinates, represented by $\Psi(r_1...r_n,s)$ or $\Psi(\underline{r},s)$. From now on, the spin will be treated as another co-ordinate, and the wavefunction will be written $\Psi(\underline{q})$, with $\underline{q}$ representing both spatial and spin co-ordinates.

In an external magnetic field, the component of electron spin in the direction of the field is found to take up one of two values, either aligned with or against the field. We therefore introduce two spin functions to account for the two possible values of the z-component of the spin quantum number. These are refered to as α (spin up) and β (spin down).

(C2) Antisymmetry.

The electronic wavefunction must be antisymmetric with respect to the interchange of two electrons, i.e. the wavefunction for the system will change sign after such an interchange : $\Psi(q_1q_2 ... q_iq_j ... q_n) = - \Psi(q_1q_2 ... q_jq_i ... q_n)$ for a system of n electrons where the co-ordinates of electrons i and j have been swapped. This is a general statement of the Pauli Exclusion Principle, which does not allow two electrons to have the same set of spatial and spin co-ordinates.

(D) Hartree-Fock Molecular Orbital Theory.

(D1) Molecular Orbitals.

In Molecular Orbital (MO) theory, we extend the orbital picture of atoms to molecules. Electrons may move anywhere within the molecule, but can only have certain discrete values for their energy, rather than a continuous range of values. The problem can be considerably simplified if each electron is described by a one electron function, or molecular orbital, ψ_i, which is delocalised over the entire extent of the molecule. The molecular wavefunction, Ψ, in equation (2.1) is expressed as a product of these one electron orbitals [23]. To satisfy the antisymmetry principle, this product is written in the form of a Slater Determinant, which is a linear combination of products of orbitals.

For the special case of a closed shell system (where all the electron spins are paired), which is the situation in of all the studies in this thesis, this is written :

$$\Psi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(1)\,\psi_2(1) \ldots \psi_n(1) \\ \psi_1(2)\,\psi_2(2) \ldots \psi_n(2) \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \psi_1(n)\,\psi_2(n) \ldots \psi_n(n) \end{vmatrix} \qquad (2.5)$$

for an n-electron system (n is even). The expression is often abbreviated to show only the diagonal terms :

$$\Psi = \frac{1}{\sqrt{n!}} \left| \psi_1(1)\,\psi_2(2) \ldots \psi_n(n) \right| \qquad (2.6)$$

(D2) Calculating the Energy.

Any observable (i.e. a property of a system that can be measured), may be calculated from the correct quantum mechanical operator, **O** using the following equation :

$$\langle o \rangle = \frac{\int \Psi^* \mathbf{O} \Psi \, .d\tau}{\int \Psi^* \Psi \, .d\tau} \qquad (2.7)$$

where <o> is the expectation value of the operator, which is equivalent to the average of a large number of experimental determinations. The integral is performed over all space and spin co-ordinates, which is usually denoted $d\tau$. If the wavefunction is real (rather than imaginary) and normalised, then this equation becomes :

$$\langle o \rangle = \int \Psi \mathbf{O} \Psi \, .d\tau \qquad (2.8)$$

The operator used to evaluate the energy is the Hamiltonian, and thus :

$$E_{el} = \int \Psi \mathbf{H} \Psi \, .d\tau \qquad (2.9)$$

This expression can be written using a useful notation introduced by Dirac and adapted to the Schrodinger formulation of quantum mechanics [21] :

$$E_{el} = \langle \Psi \mid \mathbf{H} \mid \Psi \rangle \qquad (2.10)$$

If we use the wavefunction in the form of a single Slater Determinant, then the energy can be written as :

$$E_{el} = 2\sum_{i}^{n} H_i + \sum_{i}^{n}\sum_{j}^{n} (2J_{ij} - K_{ij}) \qquad (2.11)$$

where

$$H_i = \langle \psi_i | \mathbf{H}_1 | \psi_i \rangle$$
$$J_{ij} = \langle \psi_j | \mathbf{J}_i | \psi_j \rangle \qquad (2.12)$$
$$K_{ij} = \langle \psi_j | \mathbf{K}_i | \psi_j \rangle$$

The core operator, $\mathbf{H}_1$, the Coulomb operator, **J**, and the exchange operator, **K**, are defined as :

$$\mathbf{H}_1 = -\frac{1}{2}\nabla_i^2 - \sum_A^N \frac{Z_A}{r_{Ai}} \qquad (2.13)$$

which is the sum of the kinetic energy of the electrons and the nuclear - electronic attraction terms

$$\mathbf{J}_j\psi_i(1) = \left[\int \psi_j(2)\frac{1}{r_{12}}\psi_j(2)\,.d\tau_2\right]\psi_i(1) \qquad (2.14)$$

which describes the Coulombic repulsions between electrons i and j.

$$\mathbf{K}_j\,\psi_i(1) = \left[\int \psi_j(2)\frac{1}{r_{12}}\psi_i(2)\,.d\tau_2\right]\psi_j(1) \qquad (2.15)$$

This last term is the exchange repulsion of two electrons of the same spin, and has no classical analogue.

(D3) The Variation Theorem.

In determining the energy and electron distribution of a system, the variation theorem [21, 22] is an invaluable tool. It states that the energy calculated as the expectation value of the Hamiltonian using any approximate wavefunction is always higher (i.e. more positive) than the exact energy, and is hence an upper - bound to the exact ground state energy, which can only be obtained using the exact wavefunction.

(D4) Minimisation of the Energy.

Fock applied the variation principle to MO theory to minimise the energy of a system in which the wavefunction was written in the form of a single Slater Determinant, but with the constraint that the MO s are orthonormal, i.e.

$$\int \psi_i \psi_j \, . d\tau = \langle i \mid j \rangle = \delta_{ij} \qquad (2.16)$$

where δ_{ij} is the Kronecker delta symbol. The result was a set of differential equations known as the Hartree Fock equations :

$$\mathbf{F}\psi_i = \varepsilon_i \psi_i \qquad (2.17)$$

The eigenvalues, ε_i are the orbital energies of the one - electron MO s and **F** is the Fock operator :

$$\mathbf{F} = \mathbf{H}_1 + \sum_i \left[2\,\mathbf{J}_i - \mathbf{K}_i \right] \qquad (2.18)$$

(D5) The Self Consistent Field Method.

The form of **F** depends upon the solution of equation (2.17), through the above relationship with **J** and **K**, and hence the equations can not be solved analytically, and must be subjected to an iterative procedure known as the Self Consistent Field (SCF) method. This involves the following steps (for a particular molecular geometry) :

1) Choose the initial MO s, ψ_i.

2) Calculate **J** and **K**.

3) Solve $\mathbf{F}\psi_i = \varepsilon_i \psi_i$, and hence

4) Obtain a new set of ψ_i.

5) Repeat this cycle from step 2), until there are no changes between successive wavefunctions, Ψ, or these changes are small enough to be neglected.

At this point, the energy has converged to a minimum value, and the orbitals are said to have reached self - consistency.

(E) The Basis Set Expansion : The Linear Combination of Atomic Orbitals Method.

The one-electron Hartree-Fock wavefunctions, ψ_i, are difficult to calculate for many systems with more than one electron, and in practice, they are expanded in terms of a linear combination of m simpler functions, known as Basis Functions [24] :

$$\psi_i = \sum_{\nu}^{m} c_{\nu i} \chi_i \qquad (2.19)$$

Because atomic orbitals were originally used, this became known as the Linear Combination of Atomic Orbitals (LCAO) method. Today, although basis functions are not necessarily atomic orbitals functions, the term is still used.

The variable, m, above indicates how many basis functions are used, and to give full variational flexibility to an MO, this expansion has to form a complete set, and will yield an energy calculated with Hartree-Fock theory that is said to be at the Hartree-Fock limit. In practice, the number of electron repulsion integrals necessary for such calculations prohibits their use for all but the smallest systems, and a truncated set of basis functions has to be used, leading to reduced flexibility in the wavefunction, and hence a higher energy as dictated by the variation theorem. It is important to choose the basis set very carefully so that the MO s are represented as accurately as possible within the bounds of available computational time and space. This is done by selecting a basis that yields the lowest electronic energy for given molecular configuration, but will use the smallest amount of computational resources.

Once we have chosen the basis set, the Fock equations become :

$$\mathbf{F}\sum_{\nu} c_{\nu i}\chi_i = \varepsilon \sum_{\nu} c_{\nu i}\chi_i \qquad (2.20)$$

and these are solved by varying the basis set co-efficients so that :

$$\frac{\partial E}{\partial c_{\nu i}} = 0 \qquad (2.21)$$

To solve the relevant matrix equations, we define the overlap matrix, $\underline{S}$, where :

$$S_{\mu\nu} = \int \chi_\mu \chi_\nu \, .d\tau \qquad (2.22)$$

(The diagonal elements of this matrix represent the overlap of a basis function with itself, and are hence unity). We also explicitly define the Fock matrix elements, $F_{\mu\nu}$, where :

$$F_{\mu\nu} = \int \chi_\mu \, \mathbf{F} \, \chi_\nu \, .d\tau \qquad (2.23)$$

Hence, the Fock equations become :

$$\sum_{\nu} F_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu} c_{\nu i} \qquad (2.24)$$

for i = 1 ... m, the number of basis functions. If we define a diagonal matrix, $\underline{\varepsilon}$, with elements comprising the individual orbital energies, ε_i, we can reduce the above equations to the matrix form :

$$\underline{F}\,\underline{C} = \underline{S}\,\underline{C}\,\underline{\varepsilon} \qquad (2.25)$$

We must solve this series of equations, known as the Roothaan - Hall equations, to obtain the solutions of the Fock equations, which, as previously stated, involves knowledge of the form of the Fock matrix. This is derived in terms of the density matrix, $\underline{P}$, such that

$$P_{\mu\nu} = 2\sum_{a}^{N/2} c_{\mu a} c_{\nu a} \qquad (2.26)$$

The Fock matrix can be defined in terms of the density matrix :

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda}\sum_{\sigma} P_{\lambda\sigma}\left[(\mu\nu \mid \sigma\lambda) - \tfrac{1}{2}(\mu\lambda \mid \sigma\nu)\right] \tag{2.27}$$

$\underline{H}$ is the matrix representation of the core Hamiltonian, and its elements are defined by :

$$H_{\mu\nu} = \int \chi_\mu \, \mathbf{H}_1 \, \chi_\nu \, .d\tau \tag{2.28}$$

and,

$$(\mu\nu \mid \sigma\lambda) = \int \chi_\mu(1)\, \chi_\nu(2)\, \frac{1}{r_{12}}\, \chi_\sigma(1)\, \chi_\lambda(2)\, .d\tau_1\, .d\tau_2 \tag{2.29}$$

To solve the equations, we transform the basis coefficients so that the basis functions become orthonormal. This eliminates the overlap terms, $\underline{S}$, and allows us to solve the simpler eigenvalue equation :

$$\underline{F}\,\underline{C} = \underline{\varepsilon}\,\underline{C} \tag{2.30}$$

which is done by diagonalisation of the Fock matrix. We now obtain a new $\underline{C}$ matrix, from which a new $\underline{P}$ can be calculated, and repeat until self-consistency is obtained.

The total energy is then given by :

$$E_{TOT} = \sum_{A>B}^{N}\sum^{N} \frac{Z_A Z_B}{r_{AB}} + \sum_{\mu}\sum_{\nu} P_{\mu\nu} H_{\mu\nu} + \frac{1}{2}\sum_{\mu}\sum_{\nu}\sum_{\lambda}\sum_{\sigma} P_{\mu\nu} P_{\lambda\sigma}\,(\mu\lambda \mid \nu\sigma) \tag{2.31}$$

or, in terms of the Fock matrix,

$$E_{TOT} = \sum_{A>B}^{N}\sum^{N} \frac{Z_A Z_B}{r_{AB}} + \sum_{\mu}\sum_{\nu} P_{\mu\nu}\,(H_{\mu\nu} + F_{\mu\nu}) \tag{2.32}$$

(F) Types of Basis Set.

One possible set of basis functions are Slater Type Orbitals (STO s), which represent atomic orbitals and have the form :

$$\chi_{(r)}^{STO} \propto r^{n-1} \times e^{-\zeta r} Y_{lm}(\theta,\phi) \qquad (2.33)$$

where Y is a spherical harmonic term and describes the angular dependence of the basis function. ζ is the Slater orbital exponent, and n is the principle quantum number.

These functions represent the atomic electron distribution well, but the exponential term makes calculation of three and four - centre integrals using STO s difficult and time consuming. A dramatic improvement in integral evaluation was introduced by Boys [25], which was the use of Gaussian functions (or simply Gaussians) in the basis set. These are of the form :

$$\chi_{(r)}^{GTO} \propto x^l y^m z^n \times e^{-\alpha r^2} \qquad (2.34)$$

In which α is the Gaussian orbital exponent, and x, y and z are the co-ordinates. It should be noted that this has an exponential term in r^2, compared to the exponential term in r for the Slater functions.

The advantage of such functions is that the product of two Gaussians is a third Gaussian, centred on a point on the line joining the two original functions. This makes the calculation of the four - centred integrals much simpler because they can all be reduced to a two-centre problem.

There are two fundamental differences between STO s and GTO s [21], occurring at $r = 0$ and as $r \rightarrow \infty$. At $r = 0$, for s-type functions, an STO has a finite gradient, whilst for a GTO, the gradient is zero. The latter also tail off more quickly than their STO counterparts, leading to a poor description of the inner and outer regions

of the orbitals. Thus, to obtain Hartree-Fock energies of similar quality to those derived from Slater Orbital calculations, many more Gaussian functions have to be used than Slater functions. In general, it has been found that a minimum of three GTO s have to be used for every STO. This leads to a large number of functions that have to be optimised to obtain the wavefunction, which in turn means that it can be difficult to achieve convergence of the Hartree-Fock wavefunction. The use of GTO s also leads to an increased number of integrals that need to be calculated (the number varies with order of the fourth power of the number of basis functions), the effect is more than compensated for by the improved ease with which they can be calculated. The speed up from using this method is about two or three orders of magnitude for integral evaluation, which has made these basis functions extremely popular.

(F1) Contraction.

A way to circumvent the problem of an excessive number of basis functions is to use the process known as Basis Set Contraction[1] proposed by Clementi. This involves fixing the ratio of the orbital co-efficients in the form of a linear combination :

$$\chi_{\mu}^{CGF}(r) = \sum_{n=1}^{L} d_{n\mu}\chi^{GTO}(\alpha_{n\mu}, r) \qquad (2.35)$$

χ^{CGF} is a contracted Gaussian Function, L is the length of the contraction and $d_{n\mu}$ are the contraction co-efficients. A careful selection of these terms can give a contracted function that fits any functional form we might require, and it can be performed so that the contracted function fits the form of the STO s, as in the popular STO-nG [26] series of basis sets. Here, n is the length of the contraction, i.e. the number of Gaussian functions used to represent the STO s. The minimum value of n to yield results of similar accuracy to an equivalent STO basis is 3, i.e. the STO-3G basis set of Pople *et.*

---

[1] A more detailed account of contraction may be found in references [19] and [21].

al.[26]. In general, values of n from 3 to 6 are used, giving an increasingly superior description of the STO s, and hence lower energies, for larger n.

Alternatively, the contraction can be determined so that it minimises the energy of the relevant atomic system, and multiplied by an appropriate scale factor to account for the effects of bonding in molecules.

(F2) Minimal Basis Sets.

STO-3G [26] is an example of what is called a Minimal Basis, i.e. it contains the minimum number of contracted functions necessary to describe the system. This is slightly more than the number of doubly occupied orbitals because the number of orbitals from each atom is the minimum number to describe a complete shell. In the atomic situation, this shell may not be completely filled. STO-3G may actually be described as a sub-minimal basis, due to the fact that it gives energies that are higher (and thus a worse evaluation of the energy of the system) than a minimal basis STO calculation.

(F3) Extended Basis Sets.

An Extended Basis is one that has more than the minimum number of functions to describe each atomic orbital. The two main types are Split Valence and multiple Zeta basis sets, which allow the MO s to expand or contract to suit various molecular environments [23], and hence confer greater variational flexibility to the wavefunction. They also compensate for the deficiency in describing anisotropic electron distributions inherent in a minimal basis calculation.

(F4) Split Valence Basis Sets.

To give more flexibility to the orbitals, a split valence basis, such as 3-21G and 6-31G [27], can be used, which are the two most frequently used basis sets of this type within the scope of this work. This uses a minimal basis to describe the core electrons, but divides the valence space into two regions which are varied independently. Although there is obviously some loss of accuracy in this different treatment of the core and valence electrons, as opposed to the use of the same number of functions to describe both regions, it may be justified in that the former do not contribute much to bonding or to many physical properties, and are similar to the functions in the unperturbed atoms, thus less functions are needed to describe them accurately.

(F5) Double and Triple Zeta Basis Sets.

Double Zeta is the logical extension of split valence basis sets, and uses two (contracted Gaussian) functions for the core orbitals as well. Triple Zeta is similar, but uses three functions to describe all the orbitals. Examples of such basis sets are Clementi's Geomedium and Geolarge sets [28].

(F6) Polarisation Functions.

The increase in the number of functions used to describe each atomic orbital could continue upwards, but leads to two problems. Firstly, the number of integrals, and hence computation time, increases dramatically with the number of basis functions. Secondly, the use of such basis sets tends to lead to incorrect theoretical predictions. For example, a large basis constructed only of functions of s and p symmetry, predicts a planar structure for ammonia. [21]

Greater flexibility can be imparted to the wavefunction by the use of Polarisation Functions. These are functions with an angular momentum quantum number which is higher than the minimum necessary to describe the atoms making up the molecule, allowing polarisation, or a shift of electron density, away from the nuclear centres. This accounts for the distortion of the symmetrical charge distribution of an atom, when it is present in the non-uniform electric field of a molecule.

Polarisation functions are denoted by adding a star (*) to the basis set name, and are normally added in two stages. Both give an increase in flexibility to the wavefunction, but the first does not involve the use of extra functions on the Hydrogen atoms. Hence in, for example, the 6-31G* basis set [29], d-orbitals are added to the first row atoms. In 6-31G** [29] (an example of the second category) the basis set for hydrogen is also supplemented by the addition of p-functions.

(G) Direct SCF.

A useful tool in Ab Initio calculations is the Direct SCF method [30]. This has been an option available in the Gaussian series of programs since the 1988 release, and has been the basis of nearly all of the Ab Initio calculations performed in this work.

As previously stated, the number of integrals evaluated in each SCF calculation rises rapidly with the number of basis functions, m. The number of one-electron integrals is given by :

$$N_{1e} = m(m + 1) / 2 \qquad (2.36)$$

whilst the number of two-electron integrals is :

$$N_{2e} = N_{1e}(N_{1e} + 1) / 2 \qquad (2.37)$$

Thus the total number of integrals is of order $m^4 / 8$. Although some of these can be neglected due to distance criteria, most have to be calculated and stored. This means that the largest calculation one can perform is limited by the disk space available, which

on the current configuration of the FPS-500 (see Chapter One) gives a maximum of about 130 basis functions. This is not enough space even for a 3-21G calculation on a typical bicyclic molecule.

Direct calculations give us a means to go beyond this limit in that the integrals are calculated as and when they are needed, and do not need to be stored. Although this requires individual integrals to be evaluated several times during the iterative SCF procedure, and hence slows it down, it also means that we are no longer limited by the extent of the available disk space in the calculations we choose to perform.

(H) Limitations of the Hartree-Fock Method.

The main assumption of this theory is that the electronic forces (due to electron - electron repulsions and nuclear - electron attractions) acting on an individual electron are independent of the instantaneous positions of all the other electrons. Despite this approximation, such calculations yield over 99 % of the non - relativistic energy. Nevertheless, the motions of the electrons do affect one another, manifested in a tendency to avoid each other, a phenomenon known as electron correlation, which is not accounted for by Hartree-Fock theory for electrons of different spins.

The correlation energy, which is the difference between Hartree - Fock theory and the exact, non - relativistic energy can be calculated, but this process is extremely time consuming. Hence, such work can only be performed upon relatively small molecules, and is beyond the scope of this work. The most common methods used for such calculations are known as Configurations Interaction (CI) and Moller - Plesset perturbation theory[2].

---

[2] Discussions of these methods are presented in references [21] and [23].

One further approximation in Hartree - Fock calculations is the treatment of relativistic effects, which are totally neglected. Relativistic corrections are mainly associated with the kinetic energies of the core electrons in atoms of high atomic number, which are not generally of importance in most biological systems.

Inclusion of these two factors in a calculation leads to an expression for the exact ground state energy of a molecule which is the sum of Hartree - Fock, correlation and relativistic terms :

$$E_{Exact} = E_{HF} + E_{Corr} + E_{Rel} \qquad (2.38)$$

(I) Semi Empirical Methods.

The solution of the SCF problem is extremely time consuming for large molecules, or where extended basis sets are used. Although Ab Initio single point calculations (i.e. without geometry optimisation) may be performed with these basis sets for small molecules, high quality work upon large molecules, or the repeated SCF calculations needed in a geometry optimisation are beyond today's resources for the study of large molecular systems . Thus if we wish to have some knowledge of the electronic structure or molecular geometry of large systems, we use Semi Empirical methods.

Semi empirical calculations are based upon quantum mechanical theory, but many of the integrals are approximated by the use of optimised parameters derived from experimental data. Hence, such methods can represent basic electronic bonding and properties, but only require a fraction of the time needed for a full Ab Initio calculation.

The parameters and the types of experimental data used vary from method to method, and depend upon what the method is used for. For example, the methods of Pople's group, such as CNDO [31, 32] and INDO [32, 33] were designed to reproduce

Ab Initio charge densities and bond orders from a minimal basis STO calculation with the minimum effort [34]. Zerner's work, in methods such as ZINDO [35] was designed to reproduce atomic and molecular spectra. Finally, the efforts of Dewar et al. were aimed at reproducing experimental geometries and heats of formation (ΔH_F). It is the last of these methods, based upon the Neglect of Diatomic Differential Overlap (NDDO) [31, 32] approximation that has been the basis of most of the geometry optimisations performed in this work.

Another advantage of the use of experimentally obtained parameters is that it is possible to produce geometrical predictions that are more accurate than those obtained from single determinant Ab Initio calculations (i.e. those that do not account for electron correlation), since the effects of electron correlation are introduced into the parameterisation because they are an integral part of the experimental determination.

The increase in speed of semi empirical calculations is dramatic, and is approximately proportional to the size of the molecule under consideration [34] ; the relationship of time with the fourth power of the number of atoms for Ab Initio work is replaced by a cubic function implying that semi - empirical calculations are an order of magnitude quicker than Ab Initio work. An example of this increase in speed can be shown for the steroid hormones studied in this thesis ; a complete geometry optimisation for such a molecule (with about 50 atoms) can be performed using the semi empirical methods of MOPAC [36] in around one hour on the FPS-500[3]. A single SCF calculation with the minimal STO-3G [26] basis (using direct SCF) would take about three to four hours using Gaussian88 [37] on the same machine, and many such steps could be required during the course of the optimisation.

---

[3] Computer programs used in this thesis are described in sections L, M, N, and O of this chapter.

(I1) Semi Empirical Approximations.

As previously stated, methods employing the NDDO approximations have been selected. These have been well tested, and been shown to give reliable scientific results. The approximations used in the NDDO method are [32] :

1) The method only treats valence electrons explicitly. Electrons from inner shells which take no part in bonding are treated as part of a nuclear core of charge ($Z - n_{inner}$). This is known as the core approximation.

2) The basis set comprises atomic orbitals of the same principle quantum number as for the electronic configuration of the isolated atoms making up the molecule.

3) As its name suggests, diatomic differential overlap is neglected. This means that the overlap between any orbitals not based upon the same atoms is zero, and that all three - and four - centred, and some two-centred integrals are neglected.

(I2) NDDO Methods Employed in the Calculations.

There are three NDDO methods available within MOPAC 5 [36], which has been used to determine the structures of most of the molecules of interest in this thesis. These are MNDO [38], AM1[39], and PM3 [40], which stand for Modified Neglect of Diatomic Overlap, Austin Model 1 and Parametric Method 3. respectively.

Of these three, MNDO results have been largely been superseded by AM1 and PM3 [39, 40], and hence it has not been used in this work. Nevertheless, it is the basis for the later methods and the theory employed is similar for all three methods. The major weaknesses of MNDO [39] are the failure to account for hydrogen bonding, excessively positive energies for sterically crowded systems, energies that are too low for four - membered rings and over estimation of activation barriers.

The source of all these problems was found to be the over - estimation of the repulsion between atoms. This problem was solved in AM1 and PM3 by the addition of Gaussian functions to the Core Repulsion Function (CRF). The number of Gaussians added is two for PM3 [41], whilst the AM1 term is supplemented with between two and four such functions, depending upon the atom type.

Initially, AM1 was only parameterised for Carbon, Hydrogen, Nitrogen and Oxygen, and the additions to the CRF for the first three of these elements involve both attractive and repulsive Gaussians centred upon the region exhibiting the excessive repulsion, and repulsive functions centred at smaller separations. Conversely, only repulsive functions were added to Oxygen CRF.

The diagonal Fock matrix elements under the NDDO approximation are :

$$F_{\mu\mu} = H_{\mu\mu} + \sum_{\nu}^{A} (P_{\nu\nu} < \mu\mu \mid \nu\nu > - P_{\nu\nu} < \mu\nu \mid \mu\nu >) + \sum_{B} \sum_{\lambda}^{B} \sum_{\sigma}^{B} P_{\sigma\lambda} < \mu\mu \mid \lambda\sigma > \qquad (2.39)$$

The off-diagonal elements are given by :

$$F_{\mu\nu} = H_{\mu\nu} + 2P_{\mu\nu} < \mu\nu \mid \mu\nu > - P_{\mu\nu} (< \mu\nu \mid \mu\nu > + < \mu\mu \mid \nu\nu >) \qquad (2.40)$$

(I3) Parameters.

Apart from the difference in the CRF, the other main difference between AM1 and PM3 is the way in which the parameters were obtained. After the initial parameterisation of C, H, O and N in AM1, other elements were added to the list without altering the previously optimised values. For PM3 a new method of determining parameters was used, relying upon the fact that calculated properties vary linearly with the values of the parameters over a sufficiently small range [40]. Because

this enables simple evaluation of those properties and their derivatives, it has allowed the simultaneous optimisation of all the parameters for a much wider range of elements. Hence the values obtained represent "the global minimum in parameter space" [40], which was not attainable using the fixed values for the C, H, O and N parameters in the AM1 method.

(I4) Applicability of the methods.

Semi empirical methods, in general, have been shown to give calculated ΔH_F s that are of similar quality to moderately large basis set calculations [42], although Richards' group took the view that AM1 should be compared to minimal basis calculations when examining the electrostatic potential derived from the AM1 wavefunction [43].

AM1 has been extensively tested and has been shown to reproduce certain kinds of experimental results to a good degree of accuracy. It has been used to predict experimental properties in a number of cases [19, 44, 45,46, 47]. Although PM3 became available in the version 5 release of MOPAC [36], and the original papers [40] were published early in 1989, there is less data published on its performance Results published using both methods are discussed in Chapter Seven. Although the author of the latter method demonstrated a general improvement over AM1 properties, doubt has been cast over its ability to model particular situations, especially the geometries and properties of nitrogen [19]. In addition to this, John Wilkie's work also reveals closer correlation of observed properties with AM1 than with PM3 [19].

(J) Molecular Geometries

The atomic co-ordinates for a molecule can be obtained from a variety of experimental methods, including X-ray crystallography, neutron diffraction and

Nuclear Magnetic Resonance (NMR). Nevertheless, such information can not always be obtained. If, for example, the molecule may not crystalise, or large enough crystals cannot be obtained, in which case, X-ray crystallography is of no use. Such structures can also be subject to packing forces due to the proximity of many other molecules, which can cause structural deviations from the unperturbed energy minimum.

Quantum Mechanical studies are normally carried out upon isolated, unsolvated molecules, i.e. those in the gas phase. Although this is a drastic simplification of a biological system, it has been proposed that the interactions of molecules with protein receptors and active sites involves 'naked' substrates, stripped of their coat of water (47).

Theoretical chemistry enables us to examine any system and to determine its structure, which is done by evaluating the gradients of the energy with respect to nuclear co-ordinates and optimisation of the geometry. It is particularly useful for the study of molecules for which there is no experimental data available and the study of transition states, rotational isomers, short lived species and molecules that have not yet been synthesised and tested.

(J1) Geometry Optimisation.

Using geometry optimistion methods, and the most sophisticated quantum mechanical methods, it is often possible to calculated bond lengths to within ±0.02 Angstroms and angles with an error of ±5° [49]. This structural information is extremely useful.

(J2) Energy Minimisation.

Once the energy of a molecule has been calculated, the energy derivative with respect to change in atomic co-ordinates can be calculated as first described by Pulay [50]. These values tell us how the energy alters when the geometry is changed by a small amount, i.e. what forces act on the individual atoms. Thus we are able to predict the necessary alterations to the bond lengths and angles that will lead to the minimisation of the energy. This is a state which occurs when all bond lengths, angles and dihedrals are at their optimum values, i.e. the forces upon the atoms are zero, hence the molecular structure is said to be optimised. Nevertheless, the optimisation routine actually searches for a stationary point upon the potential energy hyper - surface rather than just a minimum, and hence it is possible for the program to locate a transition point instead of a minimum. This can occur if the initial geometry used to start the calculation is not very good, and means that it is necessary to check the final geometry by performing a force calculation (MOPAC keyword *Force*.), which calculates the vibrational frequencies for the molecule. Analysis of these results will intimate the fact the the geometry is a transition state by revealing the presence of one vibration with a negative frequency.

The optimisation process can be performed by altering each variable individually, but this method can lead slow convergence to an energy minimum and has to be repeated several times to yield an energy minimum. Conversely, all parameters can be altered simultaneously, leading to rapid geometry optimisation.

Two programs have been used to obtain molecular geometries (MOPAC 5 [34] and the Gaussian series [37, 51]), both of which have a range of available geometry optimisation routines, although it has only been necessary to use the default method in each for the molecules of interest in this research.

They rely upon the fact that the energy of a molecular configuration, $\underline{x}'$, can be expressed as a Taylor series expansion of the energy for the same molecule with geometry $\underline{x}$, where the differences between $\underline{x}'$ and $\underline{x}$ are small. Hence, we obtain :

$$E_{(\underline{x}')} = E_{(\underline{x})} + \underline{g}(\underline{x}' - \underline{x}) + {}^1/_2\, \underline{H}\, (\underline{x}' - \underline{x})^2 + \ldots \qquad (2.41)$$

Here, $\underline{g}$ is a matrix of the first derivatives with respect to energy :

$$g_i = \frac{\partial E(\underline{x})}{\partial x_i} \qquad (2.42)$$

and $\underline{H}$ is the second derivative, or Hessian matrix :

$$H_{ij} = \frac{\partial^2 E(\underline{x})}{\partial x_i \partial x_j} \qquad (2.43)$$

If we assume that near a stationary point, the energy surface can be approximated to a parabola, then we can generate the changes that need to be made to the geometry, $\Delta\underline{x}$, using [49] :

$$\Delta\underline{x} = -\underline{H}^{-1}\underline{g} \qquad (2.44)$$

and modifying this solution to eliminate the three rotational and three translational degrees of freedom for the molecule.

The Gaussian series of programs use Schlegal's "Berny" method of optimisation [52]. This uses an approximation to the Hessian matrix itself, as opposed to its inverse. This is so that the initial Hessian can be constructed to incorporate any information that can be derived from our chemical knowledge of the system under study. Thus, the inverse is generated explicitly, when it is required.

The Hessian is updated as the optimisation proceeds, so as to contain information relevant to the current geometry. This update procedure is not performed

when the gradients are large, so as to avoid the necessity of compensating for errors introduced by bad optimisation steps made early on in the calculation.

The geometry changes are then predicted as described above, although polynomials of degrees 3, 4 or 5 are used, depending upon which function has a minimum within the desired range [53].

The Berny optimisation terminates when four conditions are satisfied. These require that the maximum force, the RMS force, the maximum predicted geometry change and the RMS predicted geometry change all fall below their respective preset thresholds.

MOPAC 5 uses a modified form of the Broyden - Fletcher - Goldfarb - Shanno (BFGS) [54] method. This uses the inverse Hessian, but not the second derivatives themselves. An initial guess is made at this matrix, which has to be performed carefully in order to preclude any extraordinary geometry changes [41]. To obtain a realistic inverse Hessian, the off-diagonal elements are set to zero, whilst the on-diagonal terms are calculated by making small perturbations to the geometry and recalculating the energy.

New geometries are initially predicted using a line search method, until successive energy changes become suitably small (about 0.5 kcal mol^{-1}). Geometries are modified by adding or subtracting a scaled product of the inverse Hessian and first derivative matrices :

$$\underline{x}' = \underline{x} + \alpha \underline{H}^{-1} \underline{g} \qquad (2.44)$$

where α is the step size.

Once the cutoff threshold has been reached, the accuracy of this method decreases and it is replaced by a binary search. This involves initially setting the step

size to unity, and then continuously halving it until the predicted change in energy yields a minimum value.

More precise details of this method are described in reference [41], including cutoff values for the change from the line minimisation to the binary search.

There are a number of tests taken within MOPAC to determine the end point of the optimisation [55]. These are in the table below, with their default cutoff values :

Test	Criteria To Be Satisfied.
Herbert's Test	Projected decrease in energy < 0.001 kcal mol^{-1}
Test on X	Projected change in geometry < 0.0001 Angstroms
Test on gradient	Gradient norm < 1.0 * √no. of co-ordinates to be optimised
ΔH_F test	Two successive ΔH_F values differ by < 0.002 kcal mol^{-1}

Table 2.1

(K) Properties Calculated from the Hartree-Fock SCF Wavefunction.

As previously described, the calculation of any property requires knowledge of the wave function and the appropriate operator. The observable is evaluated as the expectation value of the operator. For one - electron properties (i.e. those that depend upon the co-ordinates of only one electron at once), the value of the observable can be calculated using the density matrix and the relevant one electron integrals :

$$\langle o \rangle = \sum_j \sum_k P_{jk} \int \chi_j \, O \, \chi_k \, .d\tau \qquad (2.46)$$

(K1) Atomic Charges

The charges on atoms are not physically observable, but nevertheless they are a useful concept because they indicate likely positions of attack by other chemical species, in particular nucleophiles and electrophiles. They can also be used as a computationally cheap method of approximating the electrostatic potential (see later section). The charges can be calculated using the methods of Population Analysis or via a fitting process designed to reproduce the electrostatic potential of a molecule.

(K2) Population Analysis

The scheme used extensively in this thesis is a method of Population Analysis due to Mulliken [56]. There is no unique method of assigning electrons to particular atoms, and hence other methods do exist, such as Lowdin Population Analysis [21]. The Mulliken system is the most commonly used, and it is this method that is the basis for most of the charges used in this thesis.

In Ab Initio theory the atomic charges are given in terms of the density and overlap matrices, $\underline{P}$ and $\underline{S}$:

$$q_A = Z_A - \sum_{\mu \in A} (\underline{P}\,\underline{S})_{\mu\mu} \qquad (2.47)$$

The sum is over all occupied orbitals belonging to atom A.

The approximations used in semi empirical theory mean that the Mulliken analysis is simpler because there is no overlap matrix. The number of electrons associated with a particular atom can be calculated from the diagonal elements of the the

density matrix corresponding to basis functions centred upon that atom. Hence the atomic charge is given by

$$q_A = Z_A - \sum_{\mu \in A} P_{\mu\mu} \qquad (2.48)$$

Mulliken charges (and population analysis results in general) are particularly sensitive to the choice of basis set, which means that results taken from calculations using different basis sets should not be compared with one another. Despite this, charges calculated with a particular basis can be compared with one another, and useful results obtained.

(K3) Electrostatic Potential.

One of the most useful properties in predicting biological activities is the Molecular Electrostatic Potential (MEP). This gives the first order interaction of a molecule with a positive test charge of unit magnitude. This means that it does not take into account any perturbation in the electronic structure of the molecule caused by the presence of the test charge. The uses and implications of MEP are discussed in a later section.

The potential at a point is defined as the energy needed to move the test charge from infinity to that point, and the MEP has to take into consideration the potential due to several fixed nucleic point charges of magnitude Z_A and an electron density distributed smoothly throughout the molecule :

$$V_{(\underline{r})} = \sum_{A}^{N} \frac{Z_A}{|\underline{r}_A - \underline{r}|} - \int_{-\infty}^{\infty} \frac{\rho_{(\underline{R})}}{|\underline{R} - \underline{r}|} . d\underline{R} \qquad (2.49)$$

where the integration is over all space and involving the electron density $\rho_{(\underline{R})}$ at point $\underline{R}$.

The first denominator represents the distance between the point $\underline{r}$, at which the potential is being calculated, and the individual nuclei. The second is the distance between $\underline{r}$ and all points in the integration at which the electron density is calculated.

The electron density can, of course, be calculated by reference to the Born interpretation of the wave function, i.e. that $|\Psi^*_{(\underline{r})}\Psi_{(\underline{r})}|$ is a probability density at point $\underline{r}$.

A particularly useful aspect of the MEP is its use in describing the activities of a series of molecules, and extremely useful information can be derived by calculating the MEP at a large number of points upon the Connolly surface [57] of a molecule. This surface is constructed from the points upon the Van der Waal's surfaces of the individual atoms, but neglects any points that are within the surface of a neighbouring atom. With this view in mind it is obvious that Ab Initio calculations (even with a small basis set) could take a long time for all the molecules in the series. An approximation that has been found to be particularly useful at St. Andrews is the Point Charge Approximation [58]. This uses the atomic charges (calculated by such schemes as the Mulliken Population Analysis) to represent the potential due to the nuclei and their associated electrons. The potential can now be calculated with the following equation :

$$V_{(\underline{r})} = \sum_{A}^{N} \frac{q_A}{|\underline{r} - \underline{R}_A|} \qquad (2.50)$$

where q_A are the net atomic charges resulting from the population analysis.

This equation obviously leads to a potential which is dependent upon the source of the calculated charges, and it is well known that Mulliken charges have an arbitrary nature and are dependent upon the basis set with which they are calculated. Nevertheless it is extremely useful in the comparison of a series of molecules, where calculations have been performed with the same method. It is also very fast, especially if combined with a semi empirical program such as MOPAC, which can easily produce both geometries or charges.

The approximation has been incorporated into a locally written program called 3D2 [59], described in section M2 of this chapter. The results such calculations have been used extensively [19, 60, 61] and have correlated well with more precise determinations.

(K4) Potential Fitted Charges.

The basis set dependence of Mulliken charges has lead to some concern over their use, particularly in Molecular Mechanics calculations, which use atomic charges to reproduce the MEP of large systems in order to evaluate the energy. Another method has been suggested to calculate these charges, which relies upon fitting the charges so that they model the MEP of a molecule. This involves a procedure that is the reverse of the point charge approximation in equation (2.50). A least-squares fit is performed using the MEP calculated directly from the AM1 wavefunction [43], although any wavefunction can be used. The charges obtained are sensitive to the wavefunction used, and to the points upon which the potential is calculated, but to a much smaller degree than the variability exhibited by charges derived from population analysis calculations. Such charges have been studied to some extent in this thesis, and can be calculated easily as described in section M3.

(K5) Other Properties

There are also other properties that can be calculated from the wave function, which can be used both in a predictive sense, and as a means to assess the accuracy of the theoretical methods. These include dipole and higher moments [21], and the electron density.

The dipole moment has been used in later chapters of this thesis, and has been calculated from the Population Analysis charges :

$$\mu = \sum_{A}^{N} q_A r_A \qquad 2.51$$

where q_A is the charge on atom A, whilst r_A is the distance of this atom away from the origin.

A further property of interest is the Molecular Similarity Index [62] which was originally calculated using the electron density. It has been found that the Similarity Index calculated using the Electrostatic Potential (EPSI) is a better guide to biological activity. This property has been used in several sections of this thesis, and either of two indices can be calculated. The Carbo Index [62] for molecules A and B is given by

$$C_{AB} = \frac{\int V_A V_B \, .d\tau}{(\int V_A^2 \, .d\tau)^{1/2} (\int V_B^2 \, .d\tau)^{1/2}} \qquad 2.52$$

whilst the Hodgkin Index [63] also accounts for the magnitude of the Potential, and is given by

$$H_{AB} = \frac{2 \int V_A V_B \, .d\tau}{\int V_A^2 \, .d\tau + \int V_B^2 \, .d\tau} \qquad 2.53$$

The latter has been found to give a better correlation with activity.

Molecular ionisation energy and individual orbital energies are also of interest. To calculate these, Koopman's theorem [21] is used. It assumes that there is no electronic re-arrangement upon removal of an electron. Thus the ionisation energy is the negative of the energy of the Highest Occupied Molecular Orbital (HOMO). A molecule would normally be able to re-arrange its electronic structure upon removal of an electron, Koopman's theorem ionisation potentials tend to give values that are too positive [21]. (Szabo and Ostland also state that this relaxation energy is partially cancelled by the change in Correlation Energy associated in going from an n to an n-1 electron system).

The energies of other orbitals can also be used as a guide to further ionisation energies of a molecule.

(L) Software.

Most of the programs used have been written in FORTRAN, and compiled on the VAX using the standard Digital FORTRAN compiler. The main quantum mechanical packages have been transferred to the FPS and modified to enable execution and optimum performance. Vectorisation is achieved automatically on the latter machine by using the VAST-II pre-processor, which selects all the array calculations suitable for execution upon the vector processor, and generates the appropriate code.

(L1) Semi Empirical Calculations.

MOPAC 5 [36] is the main tool used for structure determinations, and was chosen because of the accuracy of structures and properties calculated with it and because it has been well tested. As previously described, it includes the AM1 and PM3 Hamiltonians, which have been used in the calculations described in this thesis. The program can normally be obtained from the Quantum Chemistry Program Exchange

(QCPE), although the MicroVAX version used in this work was obtained directly from the original author, and this was adapted for the FPS by Keith Fielding of Floating Point Systems. The resultant code was easy to port onto the Sun network, and a working version was achieved with only minor modifications to the source code, as part of this work.

Input to MOPAC is in the form of a Z - matrix, (which is described in Chapter Three), and the type of calculation performed is determined by a range of simple keywords. The program can perform geometry optimisations within a reasonable time (only a couple of hours are required for systems with over fifty atoms upon the FPS), and also calculate a number of molecular properties such as dipole moments and ionisation potentials and can also carry out the Mulliken Population Analysis to determine atomic charges. MOPAC can also perform force calculations to determine vibrational frequencies for molecular spectra as well as reaction co-ordinate calculations to evaluate the effect of varying one or two geometrical parameters upon the heat of formation of the system.

(L2) Ab Initio Calculations.

The Gaussian series of programs has been used to perform Ab Initio calculations. The Gaussian88 [37] and Gaussian90 [51] programs have been used in this thesis, the latter being substantially faster in performing SCF calculations than its predecessor. These programs have a range of basis sets built in, and have the ability to perform direct SCF calculations [30]. Input to Gaussian is via a slightly different Z - matrix format, and as with MOPAC, the program options are controlled by a series of keywords.

The Gaussian programs are both available on the MicroVAX, although the complexity of the calculations involved has meant that most of the Ab Initio work has

been performed with Gaussian88 on the FPS because the 1990 version for this machine only became available in the last few months of this research.

A small number of calculations were also performed on the FPS with the Quest module of the AMBER [64] package, which is a modified version of Gaussian80 and was the only Ab Initio code present on the machine for the first year of its use. This program was unsuitable for the study of large systems because it lacks the direct SCF procedure and has less sophisticated optimisation algorithms than Gaussians 88 and 90.

Some Ab Initio work has also been performed with Clementi's KGNMOL [65] program, but again, the lack of the direct SCF procedure and the limited disk space available for integral storage have limited its use, despite the fact that it is more efficient than Gaussian, especially when the Geometric basis sets [28] are used.

(M) Electrostatic Potential Software.

(M1) ASP.

ASP [66] is an automated similarity package produce by Oxford Molecular Ltd.. It uses the point charge approximation to calculate the Molecular Electrostatic Potential (MEP) for a series of molecules and to evaluate the Carbo [62] and Hodgkin [63] Electrostatic Potential Similarity Indices (EPSI s). The program also has the ability to optimise the fit of the similarity indices by rotating and translating the test molecules with respect to the lead molecule, which is the system that the other molecules are being compare to.

The program was originally supplied for the MicroVAX, but proved to be too slow to be useful (a typical optimisation of the EPSI for a steroid hormone takes about seven and a half hours). Thus the program was adapted, as part of this work, for use

with the FPS and Sun computers. An equivalent opimisation to the one described above only takes ten or eleven minutes on the FPS.

(M2) 3D2.

3D2 [67] is a program developed by Derek Higgins whilst he was in this research group, based upon the work of previous members of the group [68]. It calculates and displays MEP s calculated using the point charge approximation, and is also capable of reading in and displaying Ab Initio potentials.

It works by constructing the molecular surface of the system under study using Connolly's MS program [57], and calculating the potential at the points upon this surface. An octagon is then drawn at this point, and is coloured to represent the value of the potential. The orientation of the polygon is altered according to the angle between the perpendicular to the molecular surface and the line of sight. When coupled with a shading routine, this provides an excellent 3-D representation of the potential on the MicroVAX workstation screen.

(M3) Potential Fitted Charges.

Some work has been performed using atomic charges calculated to reproduce the MEP calculated directly from the AM1 wavefunction [69]. This is done with a series of programs obtained from the group at Oxford University, and includes an adapted version of MOPAC, which works with a second program to calculate the potential. A program called ESPFIT was also supplied, which us part of AMBER [64]. This performs a least squares fit of the charges to reproduce the calculated potential.

An automated system to perform such a calculation using the MEP calculated upon the Van der Waal's surface of the molecule was developed by Andrew Phin, and further modified as part of this work, to allow greater flexibility and larger systems to be processed.

The resultant charges can be used in ASP or to regenerate and display the potential in 3D2. The potential can also be calculated upon the molecular surface directly from the wavefunction and the results can be read directly into 3D2 and displayed, allowing a visual assessment of the efficacy of the fitting procedure.

(N) Molecular Graphics.

In addition to the graphics program described in the previous chapter, the research group has also been able to use Chem-X [70]. Although this is a sophisticated program, the screen output produced upon the terminals used by the group is inferior to the in - house written software which was designed to take advantage of the facilities available. Nevertheless, some of the features, such as the least squares fitting of molecules have proved useful.

(O) Other Programs Written as Part of this Research.

(O1) Gradients.

This is a program designed to analyse a MOPAC logfile (from a calculation run with the *Gradients* keyword). It extracts the forces on the atoms from the logfile, and divides them into ten groups, reporting how many gradients there are of greater magnitude than each of the ten values. The user can then enter a value, above which all variables (i.e. bond lengths, angles and dihedrals) will be optimised, whilst those remaining are kept frozen. The program then produces a new file suitable for input to MOPAC.

This program has been particularly useful in the optimisation of molecular structures using MOPAC when difficulties have arisen in obtaining a minimum in energy, due to the shape of the potential energy hyper - surface. The user is able to select the parameters from the Z - matrix which are responsible for the largest forces on the atoms, and optimise these alone.

(O2) Rotate.

Rotate is a program written using the Ghost graphics library on the VAX to draw a reaction profile upon the Tektronix 4109 terminal. It automatically scans the output file from a MOPAC reaction co-ordinate calculation and extracts the final heat of formation (ΔH_F) for each value of the reaction co-ordinate, and plots the former against the latter. Thus, the user is able to visualise the energy as a function of a certain dihedral angle in the molecule using MOPAC, and easily ascertain where the minima occur as a group is rotated.

This program was written for the Tektronix because of the ease of obtaining hard copy output from this terminal. It also produces an output file containing a summary of the MOPAC calculation, which can be used to enter the information into other plotting packages, if so desired.

(O3) Connect.

This program was designed to supplement the main graphics program described in Chapter Three, and is a method of increasing the processing speed of protein (PDB) files. The production of a stick display of a molecule requires the program to have a knowledge of which atoms are joined, and this information is not present in a PDB file, such as that generated in AMBER Molecular Mechanics calculations. The

connectivities have to be calculated, and for a system with many atoms, the process can be extremely time - consuming upon the MicroVAX. A program has been written on the FPS to convert the PDB filetype produced by the AMBER molecular mechanics program, to a CHG filetype, which contains the necessary connectivity information.

The input to the program is a typical PDB file, which can be entered "free format" if required. The data is read in by dividing each line of the file into separate words, the program then reads in the correct data by selecting the required word from the line. For example, the atom name is the third word upon the line, whilst the co-ordinates are words six, seven and eight.

After the file has been read in, the program calculates the connectivities by evaluating the distance between each pair of atoms, and comparing it to the sum of their Van der Waal's radii. Two atoms are considered to be bonded if they are separated by a distance less than this value, although in practice, it is necessary to add an extra 5 % to the sum of the radii to successfully evaluate the connectivities correctly.

The program was written to take advantage of the vector processor on the FPS-500, by performing the distance calculations in a vectorisable loop, and hence gives a remarkable speed up over the same calculation on the VAX, although it is limited in that it will only deal with atoms commonly found in proteins, and to perform a full connectivity calculation involving more exotic atoms, the original graphics program on the MicroVAX must be used.

Chapter Three

Design and Implementation of a Molecular Graphics Display Program.

(A) Introduction.

When work began on this project, the Computer Graphics laboratory at the University of St. Andrews had the following equipment : Tektronix 4109 16 - colour terminal plus printer, a Digital VT340 16 - colour terminal, a MicroVAX GPX/II colour workstation, and an LJ250 colour printer.

The colour workstation has an extremely large palette of colours available, and it can display many more of these at once than is possible on the other two VDU s. Production of the graphics is also much quicker because the workstation is connected directly to the computer, and driven by special software. A recent acquisition is a Tektronix CAChe system, which is a Macintosh based colour graphics system, which can produce 3-D images using crossed polaroid spectacles and filters over the screen to make the appropriate image visible to only one eye at once. Although the program described in this chapter is not compatible with the new system, it will be possible to modify it so that molecular structure files can be interchanged, and converted into CAChe format.

The Chem-X [70] package was also running on the VAX, and this was the main program available for the manipulation of molecular graphics, and it has three major drawbacks. The first of these is that the program will only drive the VT340 and Tektronix terminals, which are the least powerful of the available display equipment, and the second is that the display of large structures is painfully slow. This is because the picture has to be transmitted down a 9600 baud communications line to these terminals. The final drawback is that the program is only supplied as an executable module, and we are therefore unable to modify the program to perform any specific tasks related to the work at St. Andrews. Other disadvantages of Chem-X are its difficulty of use, and the particularly long time between starting up the program and being able to display a structure. Although the latter is not actually a problem, it is

annoying to the user, especially if the program is only being used to check a single structure, for example to ensure that it is the correct geometrical isomer. Therefore, it was decided that a new program should be written, which we would be able to modify at will, and which would use the more powerful facilities of the MicroVAX workstation.

(B) Details of the Program.

The program, MHCDraw is written in the FORTRAN programming language on the VAX, and uses the Graphics Kernel System (GKS) for device independent use of many different terminals, this also means that individual device drivers do not have to be written for use with other terminals. GKS provides many standard routines to perform graphics output, including line drawing and polygon filling, display area scaling and the ability to provide input via the mouse, and hence allows easy access to these features, without the programmer needing to worry about the details. Most of the program is written in standard FORTRAN-77, so the task of moving it to another machine should be fairly easy. It consists of between 8000 and 9000 lines of code, including comments and blank lines. It also has a small on-line help facility, which gives the user a list of all the commands, and a brief description of what they do.

The main uses of the program can be divided into three types :

1) Conversion between different molecular graphics file formats.
2) Graphics display.
3) Comparison of structures of molecules from different sources.

(C) Data type conversion and file manipulation.

(C1) Output File Formats.

The storage and manipulation of molecular structures is best done in Cartesian Co-ordinates. This enables the use of simple transformations to perform rotations and translations to obtain different views of the molecule under study. In addition to the co-ordinates, the machine stores a list of connectivities, which tells it which atom is joined to which, allowing the representation of the molecule as a stick picture, with bonded atoms being joined by lines. This method of representation has the advantage that it is extremely fast to draw.

It is not particularly easy for the user of a program to obtain a mental picture of a molecule when it is represented in Cartesian co-ordinates, and the generally excepted method for the input of data into Theoretical Chemistry programs is in internal co-ordinates in the form of a Z - matrix. For most atoms, this involves providing the atom type, bond length, angle and dihedral angle, plus atom numbers with which these parameters are refered to. In MOPAC [36], these are called N_a, N_b and N_c. The first atom is always at the co-ordinate origin, and requires no other parameters. The second is along the x-axis, and only requires a bond length, whilst the third requires both an angle and a bond length. An example of a possible MOPAC Z - matrix for methanol is shown in Figure 3.1.

Atom	Length	Flag 1	Angle	Flag 2	Dihedral	Flag 3	N_a	N_b	N_c
C									
H	1.1	1					1		
H	1.1	1	109.5	1			1	2	
O	1.3	1	109.5	1	120.0	1	1	2	3
H	0.9	1	109.5	1	90.0	1	4	1	2
H	1.1	1	109.5	1	240.0	1	1	2	3

Figure 3.1

MOPAC Z - matrix for Methanol

All angles are entered in degrees and bond lengths in Angstroms. The flags tell the program whether or not the geometry of a particular variable is to be optimised, whilst N_a, N_b and N_c for the penultimate hydrogen atom (number 5), show that the bond length is defined as the distance to O_4 (since $N_a = 4$), the angle is that between H_5, O_4 and C_1 (i.e. $N_b = 1$). The dihedral angle is the twist angle between H_5, O_4, C_1 and H_2 (i.e. $N_c = 2$). More specifically, in this example, if atoms O_4 and C_1 are superimposed, with O_4 at the front, then the dihedral is the apparent angle between H_5 and H_2 (see Figure 3.2). Values of τ in the diagram are positive if we rotate clockwise when moving from front (H_5)to back (H_2).

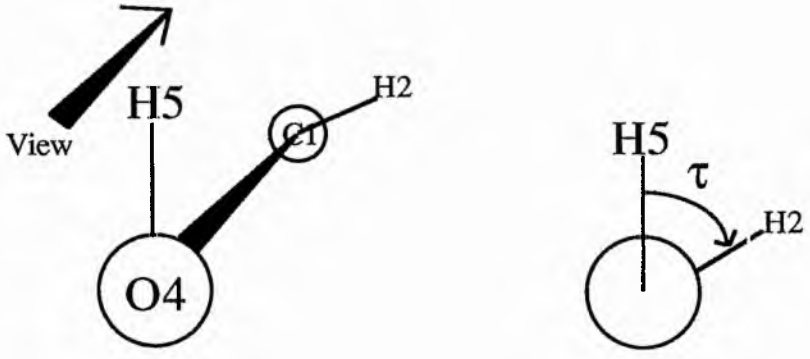


Figure 3.2

Definition of Dihedral H_5-O_4-C_1-H_2

This method of representation makes it much easier for the user to picture the molecule, and to build up a data file for it. Hence the graphics display program has to be able to convert between the two different data types.

It should be noted that input to programs such as MOPAC is possible using Cartesian co-ordinates, but the use of an internal co-ordinate system makes it much easier to follow the course of the optimisations, and to analyse results, such as the forces on the atoms, which can be reported in terms of bond stretching and angle deformation, rather than a change in X, Y and Z co-ordinates. It follows that the ability to convert cartesian co-ordinates to internal co-ordinates is also a useful feature.

The various programs used to calculate structures and display properties all use different formats and orders for the data. Some also use atomic symbols, whilst others use the atomic number. Hence, there was a potential use for a program that converts between these different formats to save time for the users, and also to save space on the computer disks, because the ease of inter - conversion means that only one co-ordinate file needs to be stored on disk.

Since routines to calculate dihedrals and angles from Cartesian co-ordinates and the Cartesian system from a Z - matrix were already available in the code for MOPAC [36], it was not necessary to rewrite such algorithms. Therefore, the source of these subroutines is acknowledged.

MHCDraw will produce the input atomic co-ordinates for MOPAC (using the command *mopac*), the Gaussian series [37, 51] (command *gaussian*), the Quest module of AMBER [64] (command *quest*), and it will write out the modified Crystal Structure Cartesian format required by 3D2. This file format also includes the atomic charges for the molecule, and also can be read by Chem-X. A further possible file format, is output using the form of the Protein Data Bank (PDB). This format can

also be used with the AMBER program. This contains the atom type, co-ordinates, amino acid residue type and number. The latter information can also be stored in an extended crystal structure-type file (again, filetype .CHG) designed especially for this use. This speeds up the study of protein molecules because it stores both the amino acid residue information from the PDB file and the connectivity information found in the CHG file format, and hence the connectivities do not need to be calculated every time the protein is read in.

(C2) Data File Formats used for Storage of Molecular Structures.

For the results of a MOPAC calculation, the archive files (filetype .ARC) are probably most useful to keep, because they contain both the structure and results of the calculation. For other calculations, the logical filetype to keep are the CHG files. Not only can this be used to store a particular view of the molecule, but it also has the atomic charges and connectivities, which are quite time consuming to calculate. For very large molecules this can also be written out in binary file format (filetype .PAC) to save disk space. Although these files are not transferable to other computers, they can save about 25% of the space used by a normal CHG file, and are quicker to read because less disk access is needed.

To produce the above output, the program stores the following information : Atom type, atomic number, 'read-in' co-ordinates, transformed co-ordinates (i.e. those corresponding to the current view of the molecule), connectivities, plus amino acid residues and numbers from protein structure files, as well as the atomic charges. The use of two sets of co-ordinates enables a chosen view of the molecule, such as one that highlights the active site of an enzyme, to be stored whilst maintaining the original orientation of the molecule. Internal co-ordinates are also stored in the program, but only if they have been read in from a file or they have been calculated because they are explicitly needed. They are not calculated automatically because the process would be

very time consuming for a protein or other large system. Thus it is necessary for the program to store a series of flags to indicate if the internal co-ordinates data is present, or if they need to be calculated before writing out a MOPAC input file, for example.

(C3) Construction of the Z - matrix.

If the Z - matrix is constructed in the order that the atoms were read in, and atoms are 'referenced' with respect to previous ones, the values of N_a, N_b, and N_c must be less than the number of the atom currently being considered. This system works if the atomic ordering has been done sensibly, but data from sources such as the Crystal Structure database often deviate from a numbering system that is logical as far as data file generation is concerned. If the ordering is bad, then this process leads to bad referencing and meaningless bond lengths. This can be seen in the following example :

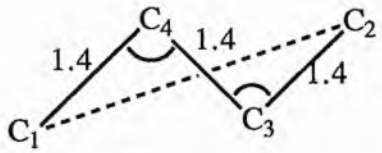


Figure 3.3

Atom 2 can only be referenced back to atom 1, giving an extremely long bond length, and small angle C_1-C_2-C_3. This could lead to trouble if the data file produced is used for a geometry optimisation using internal co-ordinates.

If the source of the co-ordinates for a molecule is in the form of internal co-ordinates (i.e. bond lengths, bond angles and dihedrals) then the z-matrix does not need to be constructed. If this is not the case, and only the co-ordinates and connectivities are available, the program has to decide which atoms to use as reference atoms, i.e. the values of N_a, N_b and N_c. Thus the algorithm used by MHCDraw

searches the connectivity table starting from atom one, for an atom that is chemically bound to the one under consideration. This is done in two passes : initially, hydrogen atoms are ignored because it leads to a more useful ordering if the referencing is done with atoms along a carbon chain. If the program cannot find a suitable atom, it searches again for possible atoms, this time including hydrogen atoms. If this second search also fails, it means that the atom is part of a separate molecule, and it is referenced with respect to the preceding atom in the list.

(C4) Input File Formats.

The program can obtain geometrical data using the *read* command, from the Crystal Structure database, Protein Structures, output from AMBER [64], Chem-X [70], MOPAC archive files, Gaussian [37, 51] log files and QUEST [64] log files. It will also look for the atomic charges from the last two formats, and extract them if they are present. Charges are also read from the preceding two formats. The extraction of atomic charges is particularly useful because it allows for easy input to two electrostatic potential programs used by the group. These are ASP [71], which compares the potential, and 3D2 [67], which is used to display the maps. This automatic analysis of the output files from programs such as Gaussian and QUEST is extremely useful because these are painful tasks to do by hand. A recent addition to the program is a routine to analyse the output of the ESPFIT program. This is used to fit charges to reproduce electrostatic potentials, which gives a simple method of comparing potentials calculated in a rigourous manner. The potential can be simulated with an array of point charges using Espfit, the geometry and charges can then be read (and displayed) using the graphics program, and the written out in the CHG format used for normal structure storage.

(D) Graphical Display.

Molecules can be displayed as stick pictures, in which chemically bonded atoms are joined with a line. This line is split into two colours, indicating the types of atom that make up the bond. The colouring system normally depends upon atom type, and is generally consistent with that used by 3D2, although all halogens are coloured light blue, and the default colour (for atoms which do not have a specified colour) is orange. These colours are as follows : Green (carbon), red (oxygen) dark blue (nitrogen), white (hydrogen), yellow (sulphur), purple (phosphorous) and as mentioned above, light blue for halogens. The display background is black to make the display easier to view, which is why this colour has not been used for carbon, as is sometimes the case. A second method of colouring atoms is possible, which is used if more than one molecule is being studied at once. This colours all the atoms in a particular molecule one colour, and colours different molecules (or segments, as they are termed in MHCDraw) with different hues.

Alternatively, each atom can be displayed as a shaded sphere with a radius equal to the covalent bonding radius of the atom. This radius was chosen for reasons of speed, because it allow the picture to be constructed from a series of five concentric shaded circles, as opposed to constructing a molecular surface using individual points on the surface as with Connolly's MS [57] program. At this radius, there is only a small amount of overlap between adjacent atoms, so intersections on the surface do not need to be calculated, but the display still gives a reasonable idea of the space occupied by the atoms, and of the cavities between them. The atoms are sorted according to z - co-ordinate, and those at the back are drawn first, so that they do not obscure those at the front. The sorting is performed upon an array of pointers to the atoms, rather than by swapping the actual co-ordinates and connectivity information, so that the ordering of the molecule is preserved.

(D1) Three - Dimensional Aspects of the Display.

Most molecules are inherently three - dimensional, and as described in Chapter One, their 3D shape is a vital factor in determining their use as a drug, but it is impossible to represent this information truly on most graphical displays. Nevertheless, some indication of the 3D structure of a molecule can play a crucial role in studies such as inhibitor docking (i.e. visualising the active site of a protein, and positioning the inhibitor within it) and checking structural isomers to make sure that the correct molecule is being studied.

The program attempts to enhance the users perception of 3D in two ways. Firstly, the extensive colour map available on the GPX workstation allows depth - cueing of the atoms. This involves chosing a shade of the colour used for each atom that depends upon the z - co-ordinate. Atoms at the back of the molecules are coloured dimly, whilst at the front, they are bright. This makes the far side of the molecule fade into the distance, and hence makes those atoms which are nearer look closer. The range of shades can be adjusted by the user, so that the effect can either be slight, or conversely, so that the atoms at the back almost disappear.

The second method of enhancing the three - dimensionality of a molecule relies upon the fact that GKS allows the easy modification of line width, and so thicker lines are draw for bonds at the front of the molecule, and thinner ones at the back. This function is obtained with the *auto-width on* command, and should be followed by a second command (*width* followed by an integer) to tell the program the largest line width it should use. Normally, only four different widths are needed to obtain a realistic effect, especially when used used in conjunction with shading. This method is also of some use with the less versatile terminals in the laboratory, such as the Tektronix, from which it is simple to obtain hard copies.

(D2) Multiple Segments.

MHCDraw allows the user to read in and manipulate up to twenty molecules at once using the *read / append* command. Hence molecules can be superimposed and compared or added together and written out in a single file. Individual molecules are refered to by the program as segments, and it will give the user a list of the current segments with the *showseg* command.

(D3) Rotation of Molecules.

The molecular view can be changed by rotating the molecule with the *rotate* (or *rot*) command. With this, three angles have to be specified, these are the rotations about the x, y and z - axes respectively. Angles may be specified to be positive or negative (i.e. 315° is the same as - 45°) or greater than 360° if so desired.

The rotation is always done with respect to the initial view (i.e. the orientation read in from the file). Hence, two commands of *rot 0 0 30* will produce the same picture, rather than one of rotation by 60° about the z - axis. If cumulative rotations are required, the *keep* command can be used. This uses the current display co-ordinates as the new base.

The centre of rotation is the origin of the Cartesian co-ordinates, i.e. (0, 0, 0), which is often the first atom in a molecule. The viewing window is automatically re-scaled and translated upon rotation, so that the molecule can always be seen completely.

If more than one segment has been read in, then it is possible to rotate one molecule with respect to the other using the *rot / seg* command. In this case, using (0, 0, 0) as the centre can also produce a translation of the molecule if it is not centred

upon the origin. This problem is solved by rotating about the molecular centroid (the average co-ordinate) of the molecule in question.

(D4) Moving a Molecule about another.

It is possible to move one molecule whilst keeping the other still. Coupled with individual segment rotation, this allows orientation of two molecules in any way. For example it simplifies the process of inhibitor docking. Shifting is done with the *move* command, and by specifying */x=dist*, */y=dist* or */z=dist* or any combination of the three to describe a desired movement. X and Y are positive to the right and upwards respectively, whilst positive Z is out of the screen.

(D5) Identifying Atoms and Calculating Molecular Geometries.

The *identify (ident)* command allows the use of the mouse to identify atoms. Clicking on an particular atom will make the program print its number and name. If more than one segment is present, the it will also print the molecule name, which it extracted from the input file.

To determine the geometrical parameters of a molecule, the *geometry (geom)* command is used. Again, this uses the mouse, this time to pick up to four atoms. The program draws two boxes on the screen (red and green) signifying "stop" and "calculate" respectively. After selecting the desired atoms, the user then clicks on the "calc." box, which signals the program to print the result. If two atoms have been selected, then the program assumes that the user wishes to know the distance between the two atoms, whilst three and four selected atoms signify that the evaluation of an angle or a dihedral respectively, is required. Hence the user is able to obtain information not present in the output matrix or Cartesian co-ordinate file, and could, for example, tell if two distant bonds are parallel.

(E) Lists.

MHCDraw uses three structures, held in logical arrays with one element per atom, which it refers to as lists. Lists can be defined for atom labelling, enhancement and ignoring, and can be specified by defining atom type (e.g. *label C,N,O*), by atom number (e.g. *ignore (51-99)*) or a combination of the two (e.g. *emph (12-78)C,O* which will emphasise all carbon and oxygen atoms in the range 12 to 78). Lists can also be added to, so *emph add N* would add all nitrogen atoms to the current emphasis list. It is also possible to specify protein amino acid residue names and number instead of atom names, and these are recognised as being strings of three characters.

(E1) Atom Emphasis.

In large molecules such as proteins, it may be difficult to visualise particular atoms or groups, but the program facilitates their identification using the *emphasise (emph)* command followed by a list specification. There are also three different emphasis modes, giving three ways in which the selected atoms can be displayed. The default display mode colours the chosen atoms orange, whilst the other two are used with the *draw sphere* command. The first is selected with the *emph hatch* command. This draws solid spheres for all the selected atoms, and hatched ones for the others. The *emph ball-stick* command draws the chosen atoms as solid spheres and the others as a normal stick drawing. These two modes are de-selected with the *emph nohatch* and *emph noball-stick* commands.

(E2) Hiding Atoms from the Display.

It is possible to temporarily hide atoms or molecules from the picture using the *ignore* command followed by an atom list. Atoms specified in this list are not drawn, and cannot be selected with the mouse using the *ident* or *geom* commands, and hence are totally invisible until the list is cancelled. This feature is useful to obtain a clearer view of a cluttered group of molecules, for example with the *ignore H* command, which will hide all hydrogen atoms from the display when the picture is next redrawn. It is also a step in removing a group of atoms permanently from those stored by the program, when used in conjunction with the *chop / ignore* command.

This command also has an opposite, which is the *only* command. It also uses the ignore list to store the numbers of atoms, but causes the program to display only the atoms specified. Its use is the same as that for the *ignore* command.

(E3) Labelling Atoms.

Labelling is a useful way of identifying a large number of atoms, and can be performed with the *label* command, again in association with a list. Initially, all atoms are in this list, but the command can be used to label only selected atoms, such as only carbon, nitrogen and oxygen without cluttering the display by labelling all the hydrogens as well.

The label name consists of the atom type (e.g. C, H, S etc.) and either a number that was part of the name when the structure was read in from a file, or if this is not present, then an equivalent number is added by the program. This latter feature is useful to show the numbering order in a molecule.

An additional feature of the *label* command is the ability to label atoms with charges calculated by the various Quantum Mechanical packages used. This type of labelling is obtained with the *label / charge* command. Other qualifiers for the *label* command are */length* and */angle*, which write out bond lengths over the centre of the specified bonds, and the angles between two bonds respectively.

(F) Chopping Out Sections of Molecules.

The *chop* command, as mentioned previously, is used to remove selected atoms or molecules permanently. An example of such a use could be the removal of water molecules from a hydration study, so that the structure of the solute can be compared directly to that from a gas - phase study. Chopping can be done in four ways, using the *chop / distance, chop / ignore , chop / box and chop / coord* commands. The first of these requires the user to select a central atom, and then specify a distance or a second atom. Any atoms beyond this distance are removed. As stated above, *chop / ignore,* removes all the atoms specified in the ignore list, whilst *chop / box* allows the user to draw a box on the screen using the mouse, all atoms outside of which will be removed. When using the *chop / coord* command, the user can either supply the co-ordinates of a cuboid, outside of which all atoms are removed, or the program will prompt for them.

This command was originally intended as part of a system to design molecules and build them up with the mouse, but because this feature has been superseded by the purchase of the CAChe system, it has not been necessary to write the code to perform this function.

(G) Geometry Comparison.

(G1) Reasons For the Design of the Comparison Program.

This section of MHCDraw was written to enable the user to automatically compare two different geometries of the same molecule. It was designed with a view to helping to solve two problems.

Firstly, as described in Chapter Two, two semi empirical quantum mechanical methods have been used to obtain molecular structures, namely AM1 and PM3. There are some reported comparisons in the literature between the two [19, 40, 72, 73], but these are analyses of the differences between calculated properties (such as ionisation potentials [74]). Some examples of the comparison of computed structures of small molecules are given in the original papers describing the methods [40, 41]. Thus there is little information on which of the two methods will yield the more accurate work for the larger (i.e. > 50 atoms) molecular species studied in this thesis. Thus a study was embarked upon to compare the semi empirical geometries with X-ray crystal structures for steroid hormones, and some method of comparing the results was required. A second part of this study also required such a tool to compare the semi empirical structures of sections of steroids with the results of Ab Initio optimisations. A further description of these results is presented in Chapter Seven.

The second use for the geometry comparison subroutine was to answer the question of what level of geometry optimisation is required to obtain an accurate picture of the MEP (see Chapter Two) with the minimum expenditure of CPU time. Although the facilities available at St. Andrews during the latter part of this work have allowed optimisations to be performed to an extremely high level if necessary, the answer to this question will be of interest to anyone doing molecular modelling with a more restricted computer budget. The comparison program was used to highlight the

differences between structures taken from various stages during the optimisation procedure, and the results are presented in Chapter Four.

(G2) Program Design.

The idea behind the program was to produce a graphical representation of the differences between two structures of the same molecule in the form of graphs. These would be plots of the differences between bond lengths, angles and dihedrals along the x - axis, and frequency of such differences in a certain range up the y - axis. The results are displayed as a bar graph, which is coloured (or shaded) into separate sections to highlight the differences due to non - hydrogen (i.e. "heavy") atoms and hydrogen atoms. This distinction was deemed necessary because the co-ordinates of hydrogen atoms cannot be determined accurately from X-ray crystallography due to the low electron density associated with hydrogen atoms. Hence, it was useful to divide the results as previously described.

A fourth type of comparison is also available that tries to produce an overall picture of the differences by plotting the difference in the distance between the molecular centroid and each atom in the molecule, and this is termed the difference in co-ordinates. The centroid was chosen as a reference point, rather the the co-ordinate origin, because its position in relation to the atoms in the molecule does not depend upon the actual co-ordinates of the molecule.

To perform a calculation, the user has to read in two or more structures using the *read / append* command, and then issue the *compare* command. If only two segments have been read in, or the */auto* qualifier (followed by the numbers of two segments) has been used, then the program automatically produces the four types of plot described above for the differences between the two structures selected.

If */auto* is not used and more than two structures are present, then the program prompts (for each of up to four graphs) for the numbers of the two segments to be compared, the function to plot (*length, angle, dihed* or *coord*) plus an optional individual title. (A main title, which is written at the top of the picture, can be added using the */title* qualifier). The desired functions are then plotted for the selected geometries.

It is possible to obtain plots of the differences over standard ranges (± 1Å for lengths, ± 180° for angles and dihedrals and ± 5Å for co-ordinates) for easy comparison but poor resolution of the differences of individual graphs. Conversely, the / *scale* qualifier automatically scales the plot over the minimum range of differences necessary to include all the data.

The comparison of bond lengths, angles and dihedrals is performed for every possible variable present, rather than the minimum number necessary to specify the structure. For example, a molecule can be defined by N - 3 dihedral angles (plus the corresponding bond lengths and angles) but there are many more possible combinations of four joined atoms in, for example, the four ring system of a steroid hormone and the differences between these are all presented in the graph. To restrict the variables used in each graph to the minimum necessary to describe the geometry of the molecule, the / *unique* qualifier can be used.

It is possible to obtain a report of the differences over a certain value by using the / *report* qualifier. This prompts the user for a maximum difference between the two molecules (for each graph to be plotted), and lists the appropriate bond lengths, angles etc. that correspond to differences over this value.

A further qualifier (*/laser*) produces a file (called "COMPARE.LAS") of PostScript commands, suitable for transfer to a Macintosh and for down - loading to a

LaserWriter using the appropriate software for the production of publication quality output. ("SendPS", which is part of the Cricket Draw suite is an example of a suitable program for this task).

(H) Other Functions of the Graphics Program.

The main functions of MHCDraw have been described in the previous sections of this chapter. Since the program has an online help facility, the rest of its features are described by way of this help file, which is presented as Appendix Two. As described in the section named HINTS, the files uses the protocol that any item in a command written between square brackets, i.e. [], is optional, whilst one of the objects written between round brackets, i.e. (), must be selected. If a number must be entered with the command, this is denoted as <int> for an integer and <real> for a floating point number.

(I) Conclusion.

A molecular graphics display program, MHCDraw, has been presented, which was written to take advantage of the facilities available at St. Andrews at the time.

The program was written in Fortran and uses the Graphics Kernel System (GKS) library calls to produce the display. The program allows various sections of a molecule to be highlighted and shown in different display formats. It also gives a good representation of the 3D structure of a molecule. Simple conversion from one filetype to another is also available, giving a flexible system that has been of great use in the laboratory.

A further section of the program allows the comparison of different structures of the same molecules, giving a method of assessing the suitability of the structural predictions of different theoretical methods.

A future addition to the program would be a facility allowing the generation of trial structures for the input to the Quantum Mechanical programs with an easy - to - use, mouse driven system of geometry input, giving a 3D equivalent of programs such as "ChemDraw" [75] in the Macintosh. Such a system would lead to rapid generation of input structures, and make work in the group more productive.

(J) Description of Photographs.

Several photographs are presented at the end of this chapter, which serve to illustrate the effectiveness of the MHCDraw program at portraying the three - dimensional nature of molecules. On the first page, two views of some of the steroid hormones studied in later chapters of this thesis are shown in the stick display format. Green represents the carbon atoms, red the oxygen, white the hydrogen, dark blue the nitrogen atoms, whilst light blue is used for the halogen atoms (in this case bromine). Both pictures are drawn using the shading and automatic line width facilities of the program to enhance the perception of depth.

The upper photograph shows four molecules studied in the Glucose-6-Phosphate Dehydrogenase inhibition work, discussed in Chapter Five. At the top left is the molecule Dehydroepiandrosterone, top right is 16α-Bromoepiandrosterone. The other two steroids, from left to right, are Androstanediol and 3β-Hydroxy-7α-methylandrost-5-ene-17-one. The structures have been read into the program as separate segments and rotated and translated into their respective positions using commands available within the program.

The lower picture is one of the inhibitors of the enzyme Aromatase, which are discussed in Chapter Six. In this case, the *TITLE* facility of the program has been used to produce the large lettering, whilst the smaller lettering was generated using the *LABLE* command. It shows the atomic symbols and the numbers of the atoms in the MOPAC optimisation for all but the hydrogen atoms.

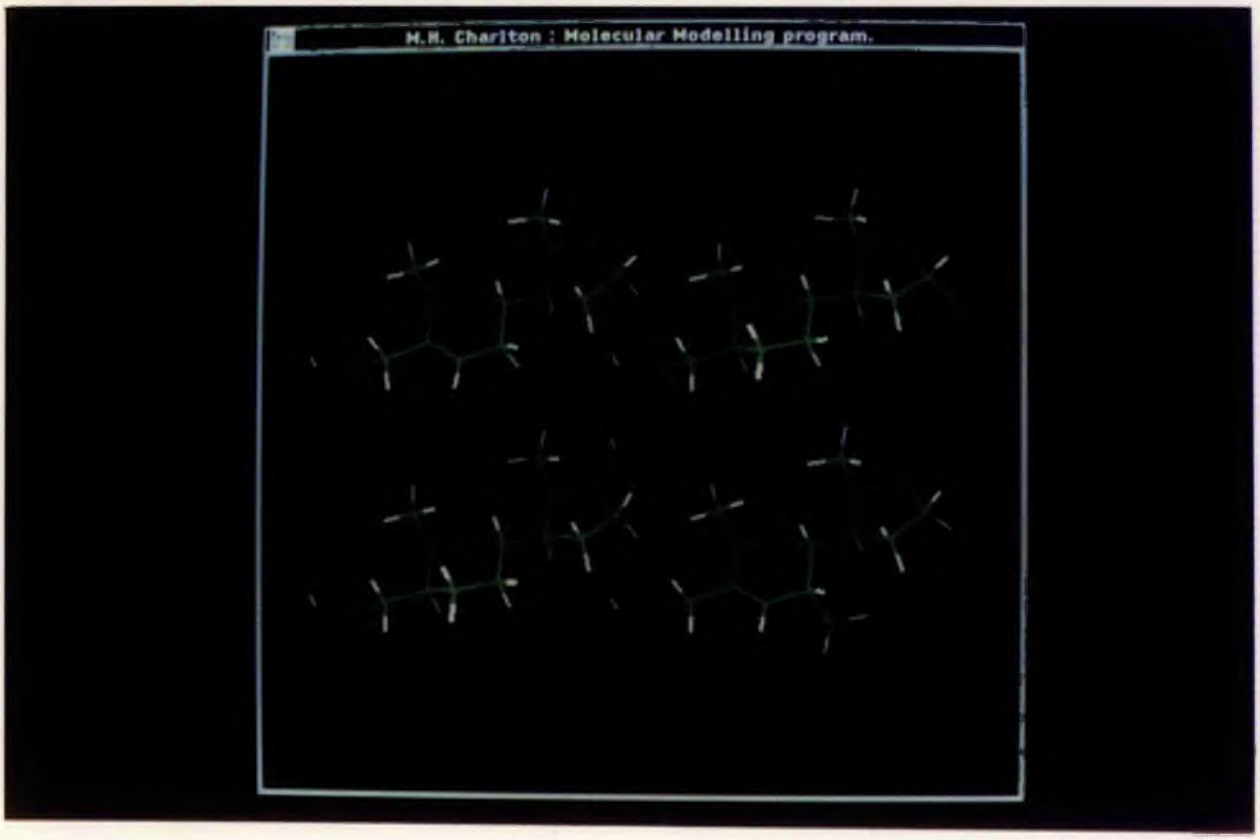
The photographs on the second page are of two larger systems. The upper picture is of a small section of a DNA double helix, in which each atom has been represented as a sphere. The twists of the helices can be seen quite clearly in the picture due to the use of shading. The lower picture comes from the studies John Wilkie [19]. It shows two lipid monolayers sandwiched between two layers of water molecules. The MHC Draw program has been particularly useful in examining the results of these studies, which involve several thousand atoms.

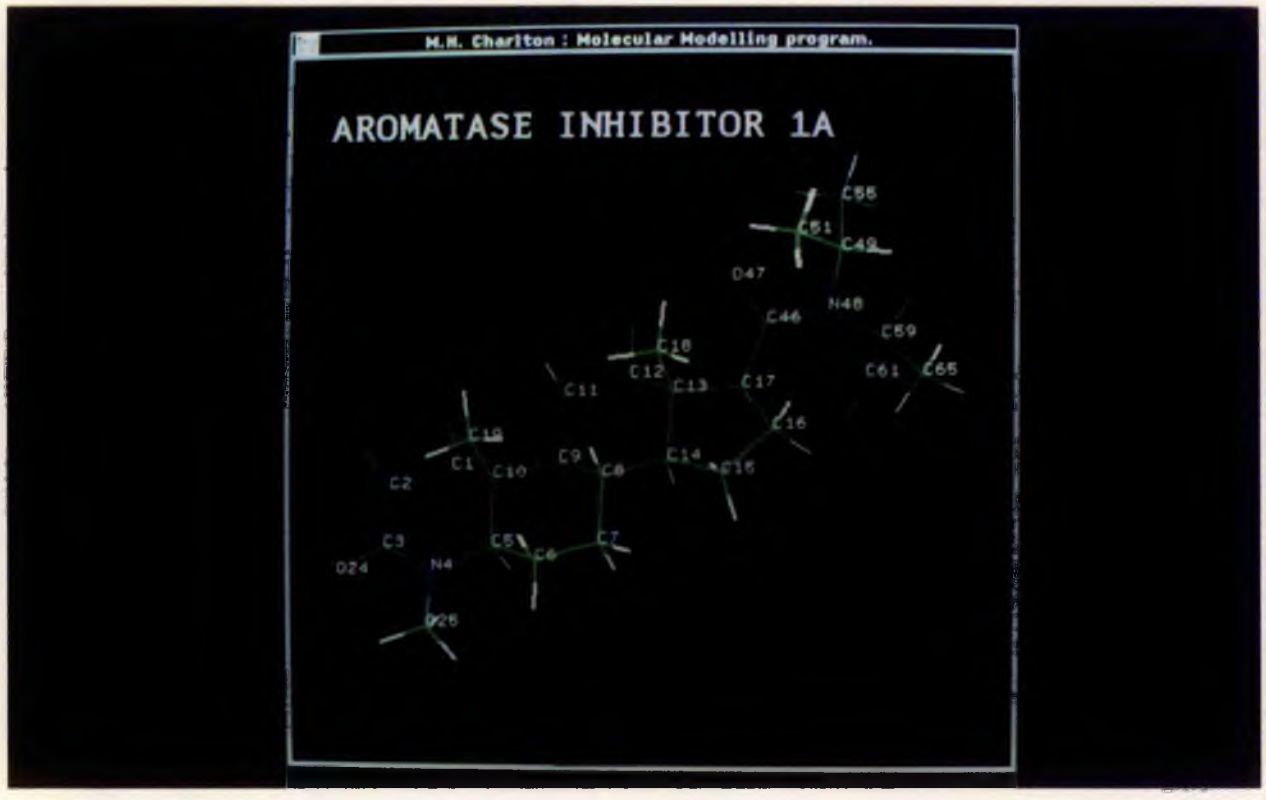
The third page depicts the enzyme Phospholipase A2, which has over two thousand atoms. The light blue circle in the picture is the calcium ion present in the active site of the molecule, and proximal to this is an inhibitor of the enzyme, which is being positioned for docking studies between the two.

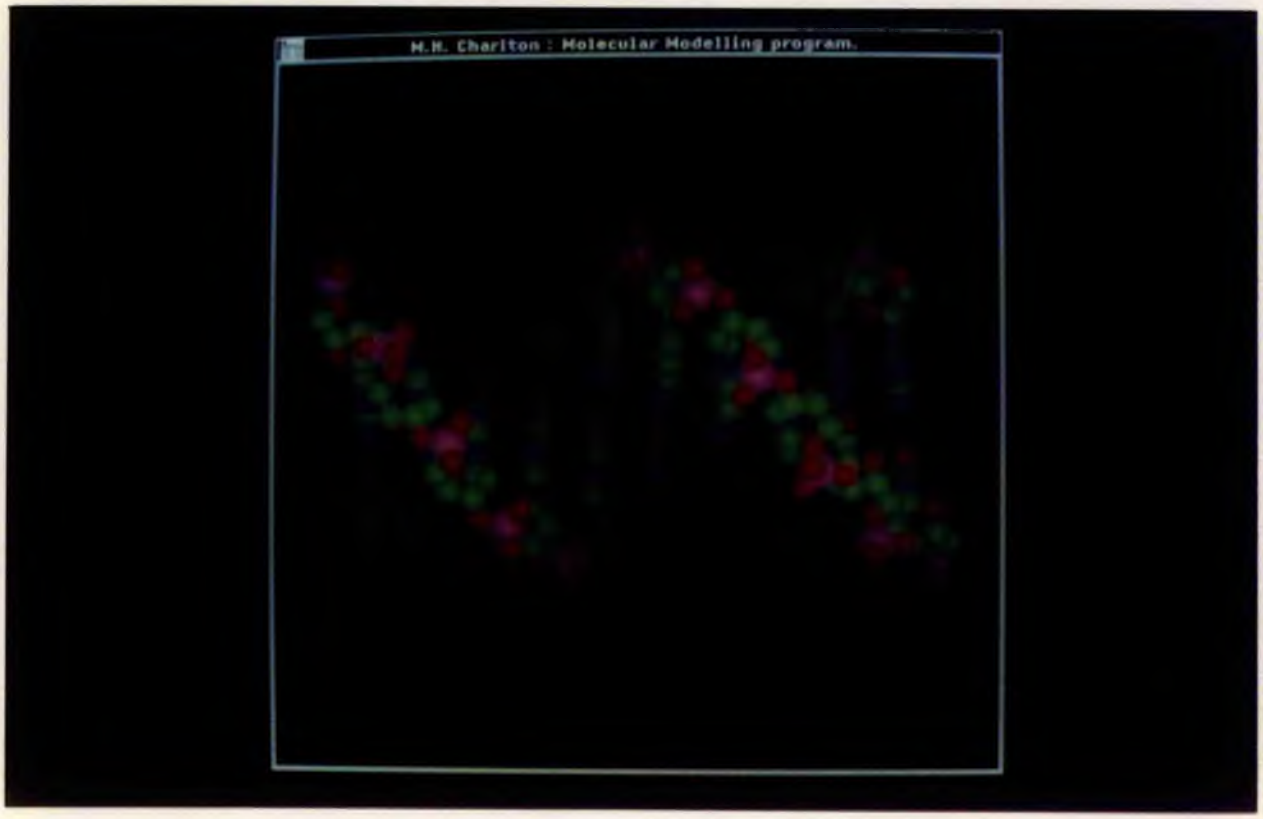
The active site of the enzyme can be located relatively easily using this display. It is the cavity at the centre of the molecule, near the calcium ion, and is highlighted by the 3D aspects of the display.

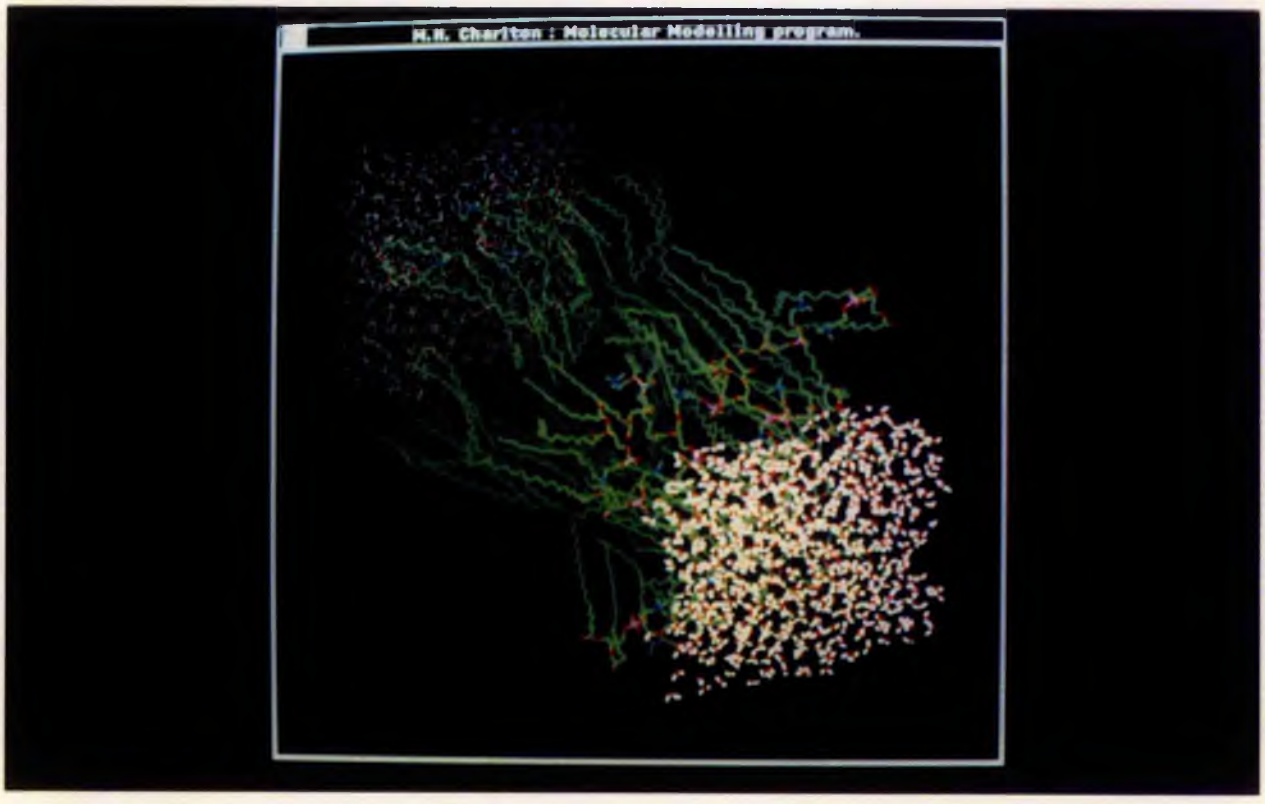
The fourth page of photographs shows a close up of the active site of the same enzyme. This was generated using the *CHOP* function of the program. The upper picture shows the section displayed using MHCDraw. The lower picture was drawn using Chem-X [70], which is the next best available picture quality. The upper picture shows the active site cavity much more clearly due to the 3D effects of the display. In addition. the upper picture shows the Histidine 48 amino acid residue of the enzyme,

which is vital for the catalytic process, emphasised in orange using the *EMPH* command.


M.H. Charlton : Molecular Modelling program.


M.H. Charlton : Molecular Modelling program.
AROMATASE INHIBITOR 1A
C55
C51
C49
O47
N48
C46
C59
C61
C65
C18
C12
C11
C13
C17
C16
C19
C1
C10
C9
C8
C14
C15
C2
C3
C5
C7
O24
N4
C6
O25


M.N. Charlton : Molecular Modelling program.


M.N. Charlton : Molecular Modelling program.


M.H. Charlton : Molecular Modelling program.


M.M. Charlton : Molecular Modelling program.


VT200 Series Terminal

Chapter Four

A Discussion of Electrostatic Potential and Geometry Optimisation.

(A) Introduction.

The MEP has been of extreme importance in this work because of the ease with which it can be used to evaluate the biological activity in a wide range of compounds as described in Chapter One [11, 12, 15, 16, 17]. It is used with preference over the electron density as an indicator of reactivity because the latter is more susceptible to inaccuracies in the wavefunction, which are not so noticeable in the MEP because of the integration performed in its calculation. As described in Chapter Two, the MEP is relatively easy to calculate from the SCF wavefunction, although the process may be time - consuming if a large basis set is used. The use of the point charge approximation has also been described as one way of simplifying the calculation of the MEP by expressing it as the sum of partial atomic charges positioned at the nuclei of the molecule. Since it is possible to measure the potential experimentally using scattering techniques [76], we are also able to evaluate the accuracy of the potential calculations and to use this as a measure of the accuracy of the calculated wavefunction.

(B) Rationale Behind the use of the MEP in Biological Studies.

To understand the interaction between two molecules, we break the total interaction energy, ΔE, down into a series of terms. One method of doing this was developed by Morokuma [77], known as Energy Decomposition Analysis. Thus, the interaction energy is given by [78] :

$$\Delta E = E_{es} + E_{pol} + E_{ex} + E_{ct} + E_{disp} \qquad (4.1)$$

E_{es} is the electrostatic energy, resulting from the Coulombic forces between charges. E_{pol} is the polarisation interaction, which involves the polarisation of the electron distribution in one molecule caused by charges and multipoles in the other. The E_{ex} term corresponds to the exchange repulsion, which is a quantum mechanical effect of the exchange of electrons between the two molecules, introduced to satisfy the Pauli Exclusion Principle. E_{ct} is a charge transfer term, whilst E_{disp} is the dispersion energy,

which is a second order term corresponding to the instantaneous positions of the individual electrons in one molecule inducing instantaneous dipoles in the second, leading to a mutual attraction. This last term, also known as the van der Waals' force, is responsible for the attraction between non - polar molecules, and is, for example, the force that is involved in holding together molecules in a liquid hydrocarbon, where there are virtually no dipole or Coulomb interactions between the individual molecules.

In biological complexes, the electrostatic interaction is generally the most important, which means that variations in the total interaction energy are similar to those in the electrostatic binding energy. E_{es} may not be the largest term in evaluating the interactions at short range, but at medium to large separations of the molecules forming a complex, it is the dominant term, and although it may not have the largest magnitude, variations in E_{es} are responsible for the greatest effect upon the total binding energy, ΔE. Hence, it will determine how the substituents of the complex will orient with respect to one another. The remaining terms in the expansion normally only come into play at very close range in most biological studies, by which time the geometry of the complex will already be established. The tight - fitting nature of enzyme active sites and protein receptors means that at small separations of the two molecules, there can be little re-orientation of the complex.

(C) Uses of the MEP.

The first use can be seen directly from the definition of the MEP given in Chapter Two that it is the same as the force acting upon a proton (i.e. a single positive point charge) due to the molecule under consideration, if the second order effects of the approaching proton upon the electron distribution are neglected. Hence the immediate application of the MEP is the prediction of protonation sites in a molecule. For example, Tomasi has shown a correlation between protonation energies and the electrostatic energy, although to properly reproduce the trends of protonation of simple

amines, polarisation terms also have to be included in the calculation [79]. Edwards and Weinstein have used the MEP to predict the relative affinities of various sites in nitrous acid towards protonation using the 4-31G basis set. They calculated the depths of the potential minima at four such possible sites, and were able to account for the experimentally preferred ordering of protonation [79].

The studies of the interaction of a molecule with a proton have been extended to cover the interactions with more complex systems, especially macro - molecules. In these cases, we can assume that complementary regions of electrostatic potential exist between a protein receptor (or enzyme active site), and compounds such as drug molecules interact strongly with this region and are hence able to initiate a biological response. In the case of enzyme inhibitors, the inhibitor must be able to mimic the MEP of the natural substrate to bind to the active site. The degree with which a compound binds to the site, and hence its biological activity, depends upon how well the substance is able to reproduce the potential of the substrate. Thus the MEP can be used to evaluate the biological activity of a series of molecules without any knowledge of the electronic structure of the macro - molecule, because it can be assumed that this complimentary structure exists.

It is this idea which has been used extensively in the studies of steroid hormones presented in this work, and attempts have been made to break the potential down into those regions that are likely to be involved in the binding process, and those regions that are not.

Examples of the use of the MEP to study biological situations have been given in Chapter One [12,13,14,15,17]. Further examples include the work of this research group [19,60,61] and that of Weinstein upon 5-hydroxytryptamine [80] in which the relative orientation of potential minima in the plane of the molecules has been used to determine the efficacy of the binding to the receptor.

As already described in Chapter Two and earlier in this chapter, another use for the MEP is the accurate evaluation of atomic charges, which are chosen to reproduce the potential. These charges are also useful in macro - molecular studies where the system is too large to be modelled using quantum mechanical techniques. In such cases, molecular mechanics programs are used; these involve the evaluation of the electrostatic interaction between atoms using a point charge model, for which the potential fitted charges are reported to give good results [69].

(D) Calculation of the Potential Maps.

In planar molecules, it is possible to calculate the MEP on a plane through the molecule, or slightly above or below it. The results can then be displayed as a contour map in which points on an isopotential are joined by a line. This gives a useful picture that is easy to visualise, of the active regions of the molecule.

In non - planar molecules, this is not possible because there may be many planes of interest, and it is not possible to view all of them concurrently. Hence, the 3D2 program, is used (see Chapter Two). This constructs a Molecular Surface, and the points upon this surface are used as the basis for the potential calculation. A commonly used Molecular Surface is that constructed using the van der Waals' radii of the atoms in the molecule, known as the Connolly Surface. This consists of the van der Waals' surface points of the atoms that make up the molecule, but excludes any points that are within the surface of any other atom in the molecule. Such a surface is a useful point at which to study inter - molecular interactions because molecules tend to position themselves approximately where the individual surfaces meet. Nevertheless, it is at this sort of separation that variations in the other terms in Equation (4.1) start to become important, and thus the results should be treated with care in some circumstances.

(E) Dependency of the MEP upon the Accuracy of the Wavefunction.

Despite the fact that the potential is less sensitive to the quality of the wavefunction than is the electron density, there is still quite a variation of the MEP when it is calculated using different basis sets and methods. For example, the positions and depths of potential minima are affected by the quality of the wavefunction, with less sophisticated methods (i.e. semi - empirical or small basis set calculations) giving minima which are, in general, deeper and closer to the nuclei than their more rigorous counterparts [81]. The behaviour of the nodal plane (the region at which no electrostatic forces act upon the positive test charge) also depends upon the method used to calculate the MEP [82]. In a minimal basis calculation for ethylene, the nodal plane crosses the molecular plane, implying the presence of a region of negative potential in this plane, which is not the case in larger basis calculations [82]. In more sophisticated calculations it was also found to be necessary to use the 4-31G basis set [83] to accurately predict the relative affinities of various sites towards protonation [79], implying that the use of a minimal basis calculation is not always sufficient. Despite this fact, MEP s calculated directly from semi - empirical wave functions, which are based upon a minimal basis set, have agreed well with various Ab Initio calculations [81], and most methods give a good indication of the relative values of the minima energies calculated with the 6-31G* basis [29].

One method used in this thesis is that of Ferenczy, Reynolds and Richards [69] to calculate the AM1 MEP , and to evaluate potential fitted charges. They obtained good agreement between STO-3G [26] and AM1 MEP s, and good correlation with the charges calculated with this method and those using the 6-31G* basis. Hence an attempt has been made in Chapter Five to correlate biological activity using these charges as well as those derived from the Mulliken analysis. Other methods of obtaining semi empirical MEP s and potential fitted charges are also available, including

that of Luque et al [81] who evaluated the MNDO MEP, and Besler and co-workers [84], who calculate charges by fitting them to both AM1 and MNDO potentials.

(F) Calculation of Potential Fitted Charges.

The AM1 Electrostatic Potential is evaluated at points upon the Connolly Surface [57] and the ESPFIT [86] program described in Chapter Two is then used to calculate the charges, located at the nuclei, that will reproduce this potential. This is done using a least - squares fitting process.

The calculation time and results obtained depend both upon the point density and van der Waals' scaling factor used in the determination of the molecular surface. The former factor denotes the number of surface points per square Angstrom ($Å^2$), whilst the latter defines the multiple of the atomic van der Waals' radii used to calculate the surface.

Ferenczy, Reynolds and Richards [69] state that either experimental or calculated geometries may be used to evaluate the charges, and suggest a van der Waals' scaling factor of 1.2 as a compromise between accuracy and computation time. Singh and Kollman [85] report the use of a point density of between one and five per $Å^2$, whilst Besler et al use the value of 1.0 in their study [84]. They also state that the location of the points at which the MEP is calculated has little effect upon the charges as long as there are enough points, and that they are located outside of a single van der Waals' radius (i.e. using a scaling factor of ≥ 1.0).

The effects of varying the two parameters has also been briefly investigated as part of this work, using the AM1 optimised structure of ethanol as a test case.

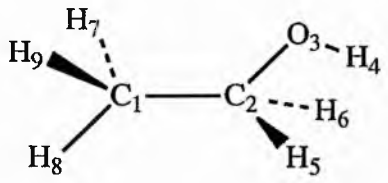


Figure 4.1

Numbering System and Atomic Positions used in Ethanol Study.

(F1) Results.

The results for the variation of the scaling factor are shown in Table 4.1. These calculations have been performed with a point density of 10.0. The column headed 'AM1' is the charges resulting from the population analysis upon the AM1 wavefunction

	Co-ordinates			Scaling Factors and Charges			
Atom	X	Y	Z	1.0	1.5	2.0	AM1
C_1	0.000	0.000	0.000	-0.159	-0.254	-0.268	-0.254
C_2	1.511	0.000	0.000	0.319	0.167	0.131	-0.025
O_3	2.046	1.311	0.000	-0.581	-0.529	-0.521	-0.327
H_4	1.672	1.773	0.759	0.318	0.314	0.314	0.195
H_5	-0.381	-1.046	-0.069	0.022	0.062	0.069	0.084
H_6	-0.394	0.462	0.937	0.046	0.075	0.060	0.073
H_7	-0.388	0.584	-0.869	0.053	0.087	0.093	0.088
H_8	1.916	-0.463	-0.940	0.007	0.055	0.067	0.105
H_9	1.903	-0.569	0.886	-0.026	0.023	0.035	0.061

Table 4.1.

Co-ordinates for Ethanol, with ESPFIT and AM1 Charges.

(F2) Discussion.

The results of this study are plotted in Figure 4.2, and in general, the fitted charges are of larger magnitude than those obtained from the Mulliken analysis process, implying a more radical distribution of charge and potential than can be obtained using Mulliken charges.

The results for scaling factors 1.5 and 2.0 are very similar, whilst those using the single van der Waals' surface (Scaling factor = 1.0) are not consistent with the other two. The latter charges cannot be obtained by a simple scaling of the results from the other calculations.

Correlation of the various charges with those calculated using the double surface (Scaling factor = 2.0) yields correlation co-efficients of 0.966, 0.890 and 0.893 for scaling factors 1.5, 1.0 and AM1 Mulliken charges respectively. This implies that the AM1 charges are of similar quality to those evaluated upon the single van der Waals' surface.

One point of note is the disagreement between the methods over the sign of the charge upon Carbon C_2, which the standard AM1 method predicts to be negative, despite the fact that it is attached to an electronegative oxygen atom. The ESPFIT charges all agree as to the positive charge upon this atom.

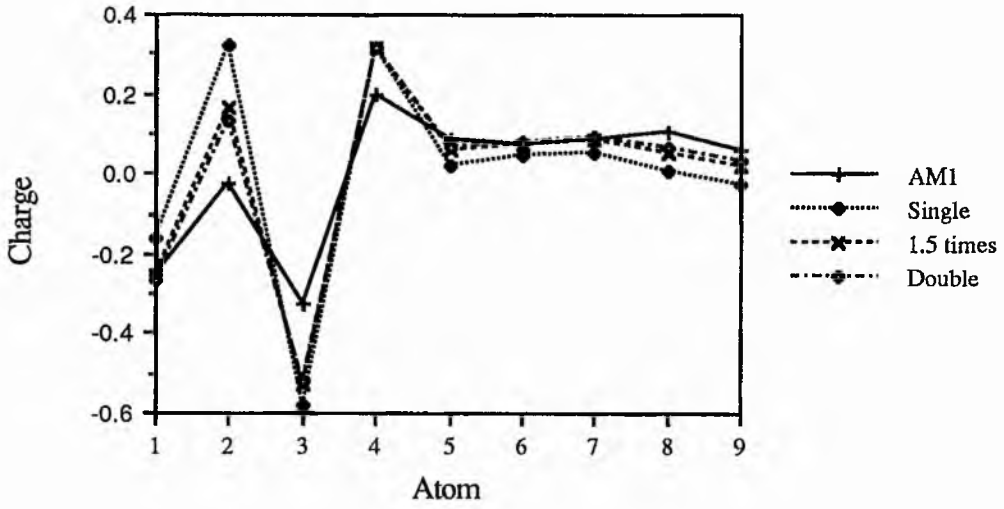


Figure 4.2

(F3) Conclusions Upon the Choice of Scaling Factor.

Assuming that to obtain the best atomic charges, a large scaling factor should be used, but that the minimisation of computational time is required, then the selection of a scaling factor of 1.5 should be made.

(F4) Dependence of ESPFIT Charges Upon Point Density.

A similar study was undertaken, using a van der Waals' scaling factor of 2.0 and point densities of 1, 2, 3, 4, 7 and 10 points / $Å^2$, and the results are shown in Table 4.2 below.

	Point Densities and Charges					
Atom	PD 10	PD 7	PD 4	PD 3	PD 2	PD 1
C1	-0.256	-0.254	-0.250	-0.239	-0.232	-0.229
C2	0.167	0.167	0.176	0.165	0.193	0.226
O3	-0.529	-0.529	-0.529	-0.524	-0.528	-0.529
H4	0.314	0.313	0.312	0.309	0.308	0.303
H5	0.024	0.023	0.020	0.021	0.012	0.000
H6	0.056	0.056	0.052	0.055	0.047	0.037
H7	0.088	0.087	0.085	0.083	0.078	0.074
H8	0.062	0.062	0.060	0.058	0.053	0.050
H9	0.075	0.075	0.073	0.072	0.069	0.066

Table 4.2

(F5) Discussion.

The graphs shown below (Figure 4.3) show plots of charges against point density, and they show a smooth variation, except for the case where the point density = 3.0. The minimum variation is for the oxygen atom, for which the charge only changes by about 1 %.

The charges upon the hydrogen atoms increase as the point density gets large, and although the magnitude of the changes is quite small, in one case the charge drops to 0.0, leading to a 100 % change, although in general, this change is in the range of about 10 or 20 %.

Nevertheless, a graph of charge calculated using a point density of 10.0 against that for 1.0 (Figure 4.4) gives a straight line of gradient 0.998 and a correlation co-efficient of 0.988.

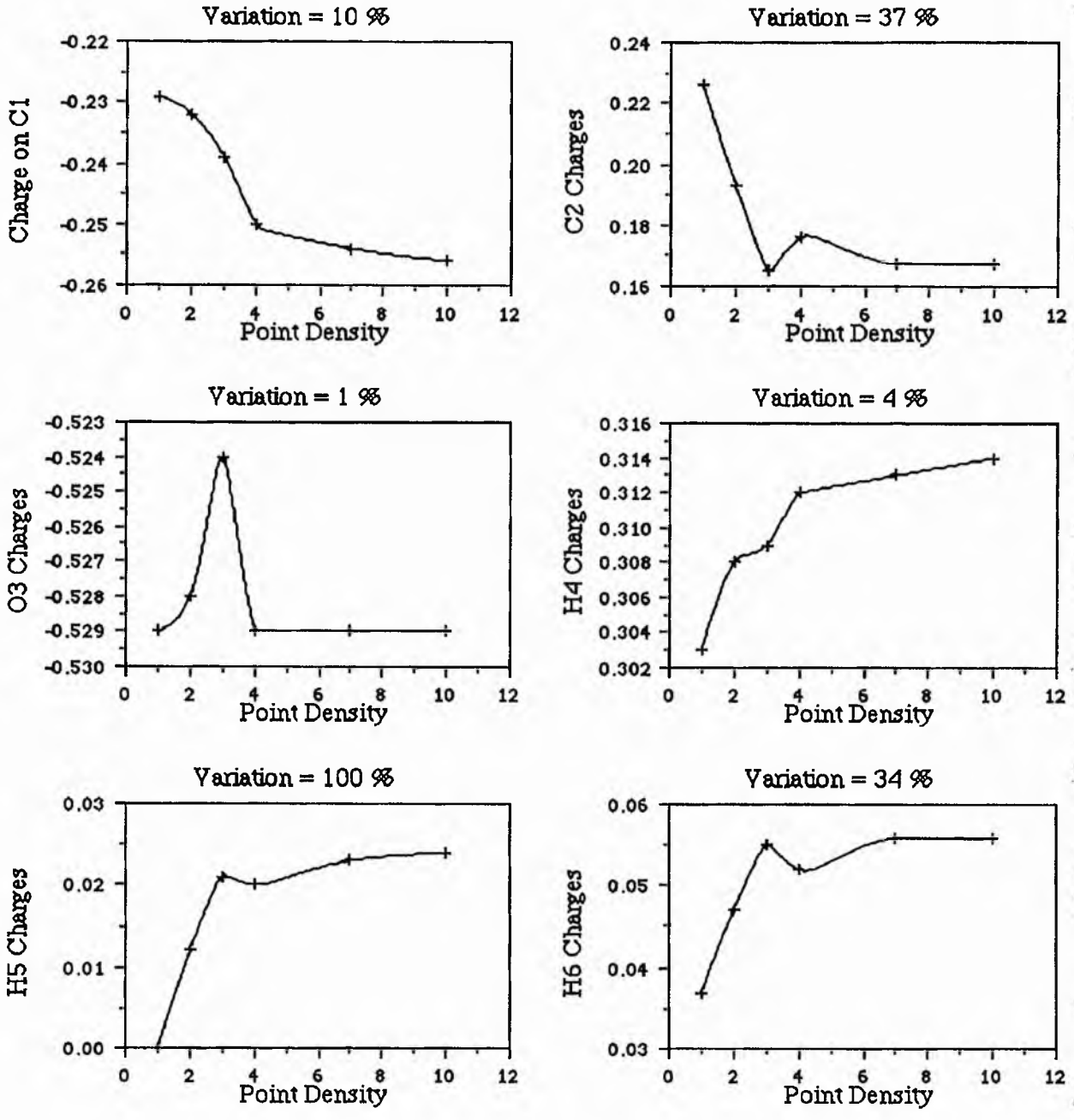


Figure 4.3

Graphs of the Charges on Ethanol Calculated at Various Point Densities.

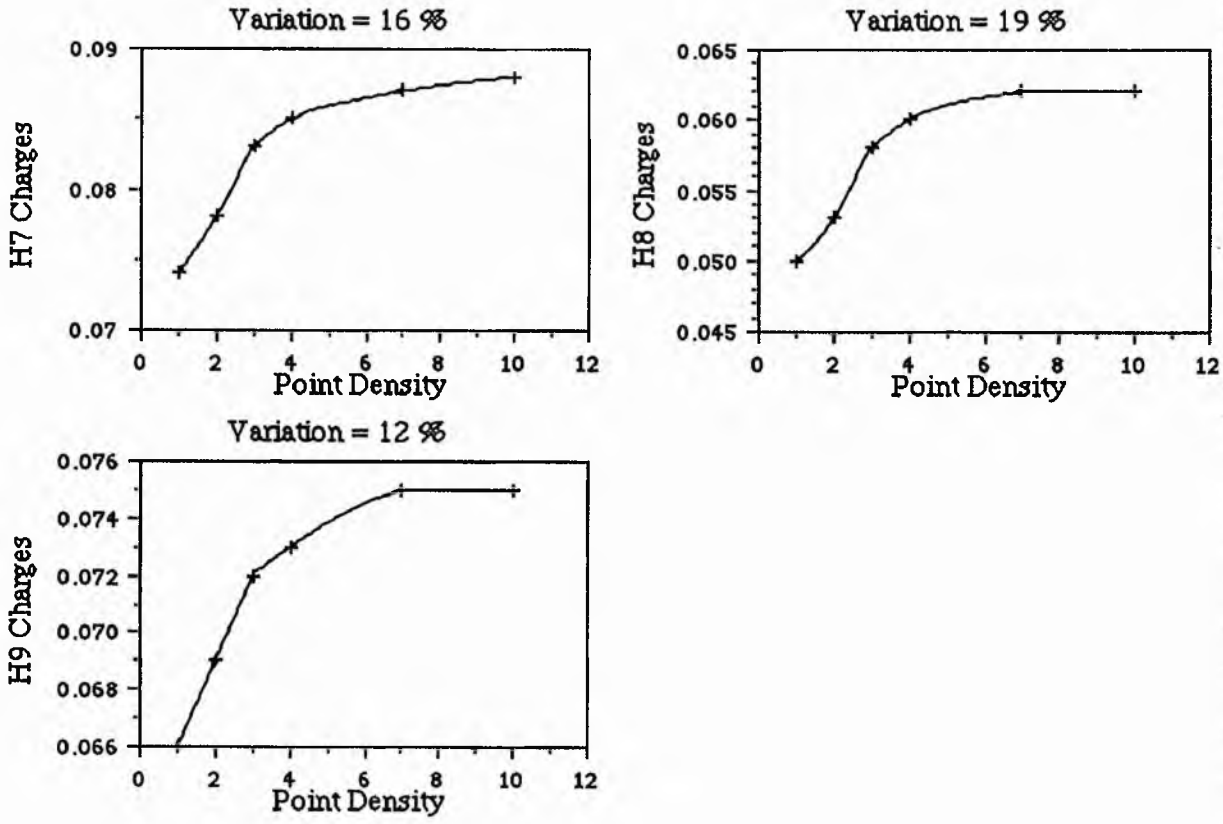


Figure 4.3 (Continued)

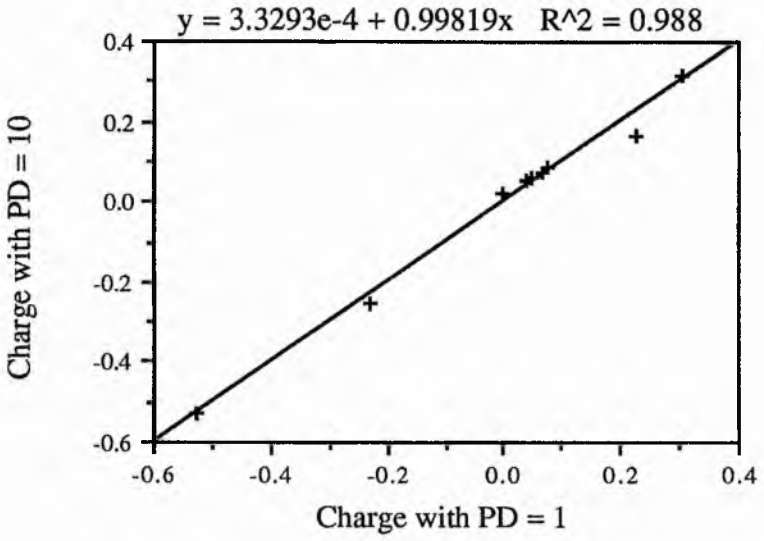


Figure 4.4

Graph Showing Differences due to Altered Point Densities.

(F6) Conclusions.

Ferenczy and co-workers have shown that the potential fitted charges are less dependent upon the quality of the wavefunction than are population analysis charges [69]. Nevertheless, there are other variables involved in the use of the ESPFIT method that affect calculated charges, namely the point density and the scaling factor used to define the molecule surface.

This work has shown that the effects of the point density become less marked above a density of 5.0 points / $Å^2$. To obtain consistent results, a van der Waals' scaling factor of 1.5 should also be used, although the Ferenczy work indicates that a value of 1.2 is sufficient.

(G) Study of Geometry Optimisation Using MOPAC.

The purpose of this study is to determine the degree of geometry optimisation required to obtain an accurate picture of the MEP using the point charge approximation and Mulliken charges. The study was performed using MOPAC, because this is the method used to calculate most of the structures used in this thesis.

The degree of optimisation can be assessed using the Gradient Norm, or Gnorm. This is the scalar value of the vector of derivatives, with respect to the co-ordinates, of the energy. It indicates the average forces upon the nuclei at a particular configuration, and should be zero at a stationary points upon the potential energy hyper - surface of the molecule. Such a Gnorm is not attainable within MOPAC because of rounding errors due to the way in which the data are stored in the computer and the difficulty of actually locating the minimum. Hence, a value of greater than zero must suffice. The default gradient norm cutoff in MOPAC is a value of 1.0, but the accuracy can be increased one hundred - fold using the *Precise* keyword, and structures obtained

using this keyword are refered to as "Precise" structures. (It should be noted that using Precise also increases the accuracy of the convergence of the SCF procedure and this lengthens the computational time dramatically). This level of precision has been used for all the geometry optimisations preformed in this work, although it has on occasions been necessary to settle for higher Gnorms (up to about 5.0) in certain cases due to the large size of the molecules under study, or the low curvature of the potential energy hyper - surface, which leads to problems in the optimisation.

One method of lowering the Gnorm further is to selectively optimise a few parameters, using, for example, the "Gradients" program described in Section O1 of Chapter Two.

The optimisation process works by systematically reducing the forces, generally until the Gnorm drops below the preset value. As noted, getting the gradient norm down to small values ($<<1.0$) can be difficult, if not impossible and the MOPAC manual [55] recommends values between 0.1 and 0.5, although this is for smaller molecules than the ones of interest here. It is stated in the manual that the six decimal place accuracy used in the input data file only allows an accuracy of three decimal places in the gradient norm, thus setting a minimum value below which changes in the Gnorm are meaningless.

This study was designed to find a suitable cut-off point at which the optimisation process could be terminated without compromising the quality of the MEP.

(G1) Method used in the Study.

The starting structures chosen for the study were constructed using a table of standard bond lengths with the tetrahedral or trigonal angles. The dihedral angles were guessed by using an appropriate model.

The MOPAC keyword *Gnorm* = *n* and *T* = *m* were used to stop the optimisation process, either when the gradient norm dropped below the value n or the time exceeded m.

The calculations were then run to various level of optimisation, and the resultant calculation times, gradient norms, number of optimisation steps (cycles) and heats of formation were recorded. The structures and charges were also used to visualise the MEP and to evaluate the similarity between the various molecules, whilst the geometries were used to evaluate the differences in bond lengths and angles using the geometry comparison feature of the program described in Chapter Three.

The study was performed using the AM1 method upon two molecules that inhibit the enzyme Glucose-6-Phosphate Dehydrogenase. The first is the steroid DHEA (see Figure 1.3 of Chapter One) and contains the usual steroid back - bone and includes a double bond, a keto - group and a hydroxy - group, and so should adequately represent this class of molecule and aliphatic compounds in general. The second molecule is 1,8-dihydroxy Anthrone or Dithranol (Figure 4.5). This system is mainly aromatic, and thus gives a good representation of aromatic molecules. Hence the conclusions drawn from the two molecules studied should be applicable to a large range of molecules.

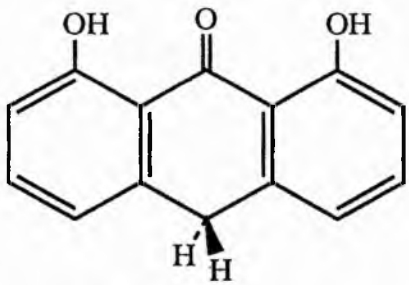


Figure 4.5 Dithranol.

(G2) Results

The results for Dithranol, including the Electrostatic Potential Similarity Index (EPSI) are tabulated below. These are followed by those for DHEA. In the Dithranol results, the heat of formation is also presented. This is purely for interest to show how this property depends upon the Gnorm. Similarly, the column headed "File" in the DHEA results has also been presented for interest. The file names indicate the method used to terminate the optimisation. For example, "39cyc" indicates that the process was terminated after 39 cycles, whilst "Gnorm400" indicates that the gnorm = 400 keyword was used. "Opt" is the results of allowing the optimisation to run to completion, whilst "Precise" is the precise equivalent.

The results can be informatively displayed in the form of graphs of EPSI against gradient norm, whilst graphs of EPSI versus number of cycles and total time are also of interest.

ΔH_f^{Θ}	Gnorm	Time	Cycles	EPSI
270.400	2449.430	42	0	0.705
36.676	779.309	249	1	0.808
15.575	407.780	327	3	0.832
-61.329	93.726	581	9	0.872
-28.296	278.271	625	5	0.894
-66.489	60.857	1173	15	0.945
-75.292	43.178	2166	36	0.967
-76.014	23.789	2435	45	0.985
-76.670	4.559	3135	59	0.987
-76.712	0.900	3992	81	0.993
-76.714	1.891	10306	148	0.994
-76.713	0.391	6212	93	0.997
-76.623	12.219	2497	48	0.997
-76.691	1.139	4490	86	1.000

Table 4.3

Results of Geometry Optimisation Study for Dithranol.

File	Gradient	Time	Cycles	EPSI
1scf	1532.3	114	0	0.728
6cyc	410.9	273	6	0.753
gnorm400	379.9	310	7	0.762
1cyc	1532.3	114	1	0.803
gnorm800	733.3	154	2	0.814
gnorm100	80.0	801	22	0.865
39cyc	16.3	1188	39	0.933
gnorm200	151.2	612	15	0.952
gnorm50	47.3	1015	30	0.957
gnorm10	9.9	1400	46	0.973
opt	9.8	1797	63	0.987
gnorm5	3.6	1967	71	0.996
gnorm1	1.0	3052	123	0.999
gnorm01	0.3	4192	185	1.000
precise	2.3	3944	128	1.000

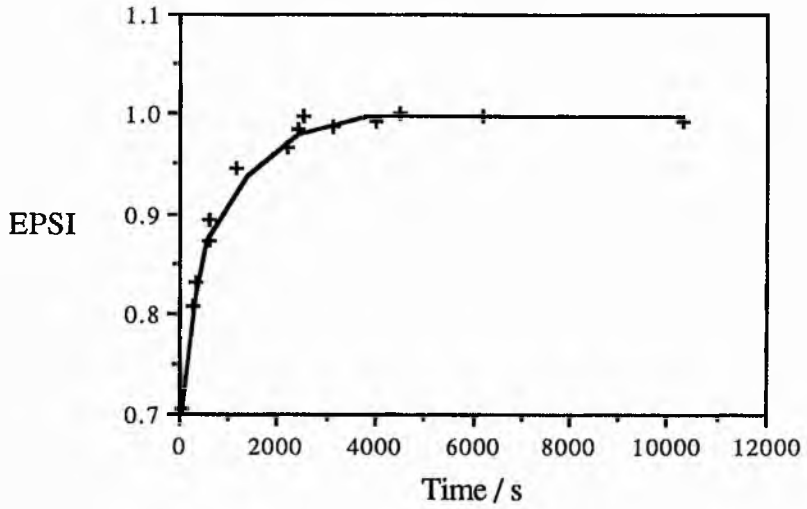
Table 4.4

Results of Geometry Optimisation Study for DHEA.

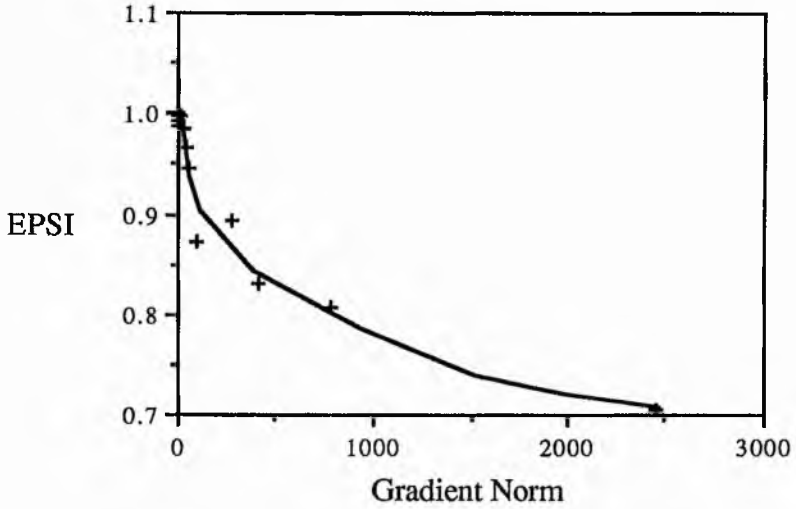
(G3) Discussion.

In both cases, the tables have been presented with the data sorted according to the value of the Electrostatic Potential Similarity Index, the index used being the optimised Hodgkin Index. It can be seen from the ordering of the gradient norms in the tables, that the optimisation process is not a smooth one, and as the gradient norm decreases, it is not necessarily the case that the EPSI increases. This is highlighted by the case of DHEA, in which the geometry after one cycle is more similar to the precise structure than that with a Gnorm of 400. It should be noted that in both cases, the precise structure was used as the basis for the ASP calculations, to which all the other structures were compared to. This geometry does not necessarily have the lowest gradient norm, but took the longest time to obtain because the *PRECISE* keyword also effects the convergence of the SCF procedure by lowering the termination criteria. The performance of the optimisation can be seen in the graphs shown below.

Dithranol : Hodgkin Index vs Time



Dithranol : Hodgkin Index vs Gradient Norm



Dithranol : Hodgkin Index vs Number of Cycles

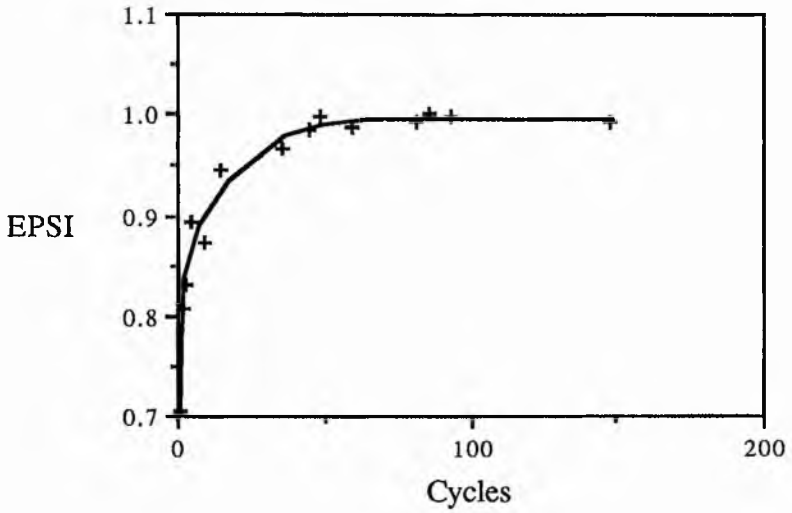
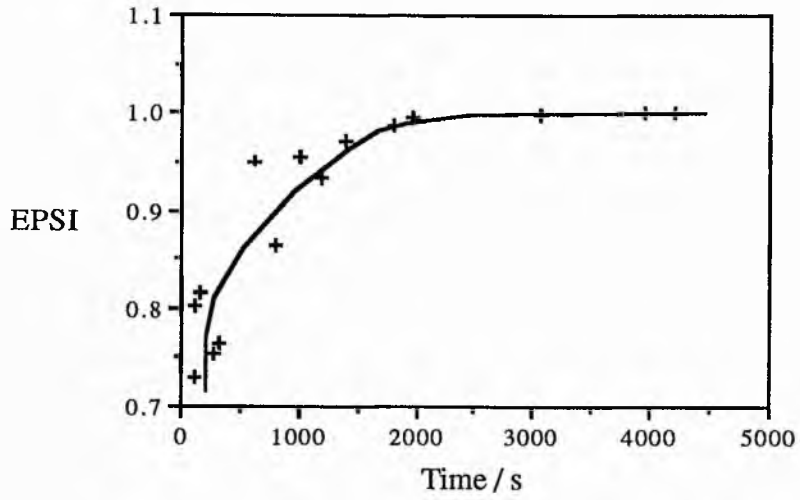
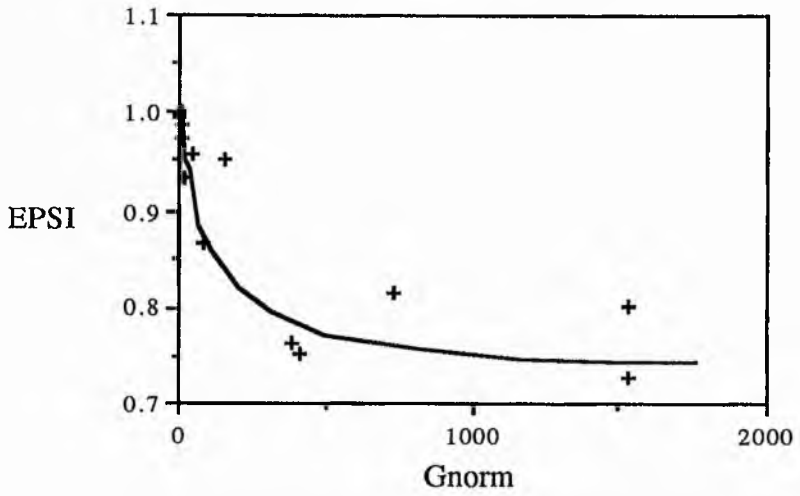


Figure 4.6

DHEA : Hodgkin Index vs Time



DHEA : Hodgkin Index vs Gradient Norm



DHEA : Hodgkin Index vs Number of Cycles

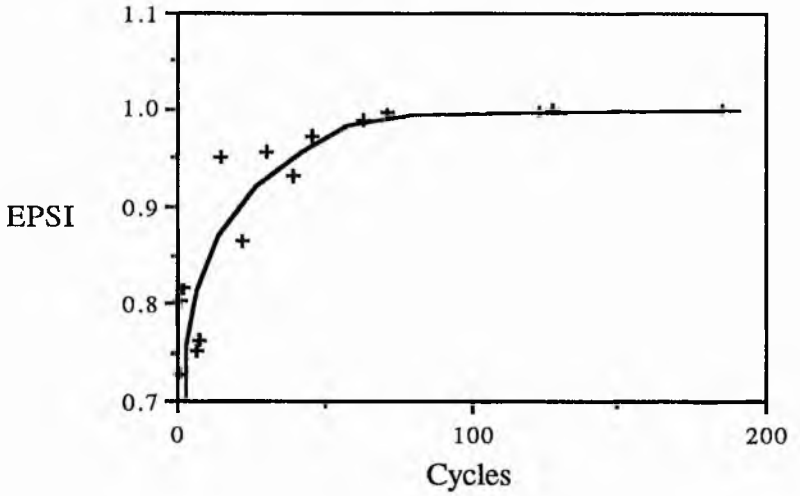


Figure 4.7

In both cases, the structure attained an EPSI close to 1.0 (the maximum value possible due to the normalisation constraint in the equations used to calculate the Similarity Index) after about half of the time needed to obtain the lowest gradient norm, and this takes about one third of the maximum number of cycles.

Since the purpose of this study is to locate a possible point at which the optimisation can be terminated, the latter sections of the Similarity Index against Gradient Norm graphs have been expanded. In the case of DHEA (Figure 4.8), the plot seems to be quite linear, and implies that even if a gradient norm of 10 is selected as the minimum, then the MEP will be accurate to within about 1 %.

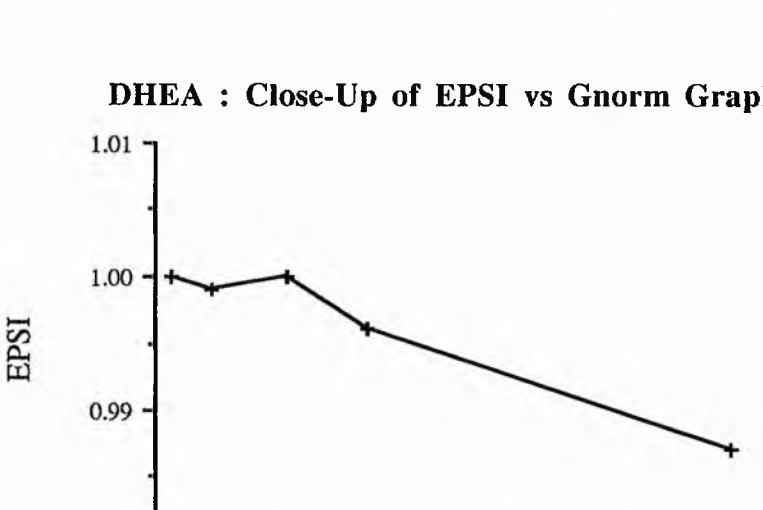


Figure 4.8

The case for Dithranol (Figure 4.9) is not so clear, but again, optimisation to a gradient norm in the region of 10 will yield an MEP that is accurate to within 1 or 2 %

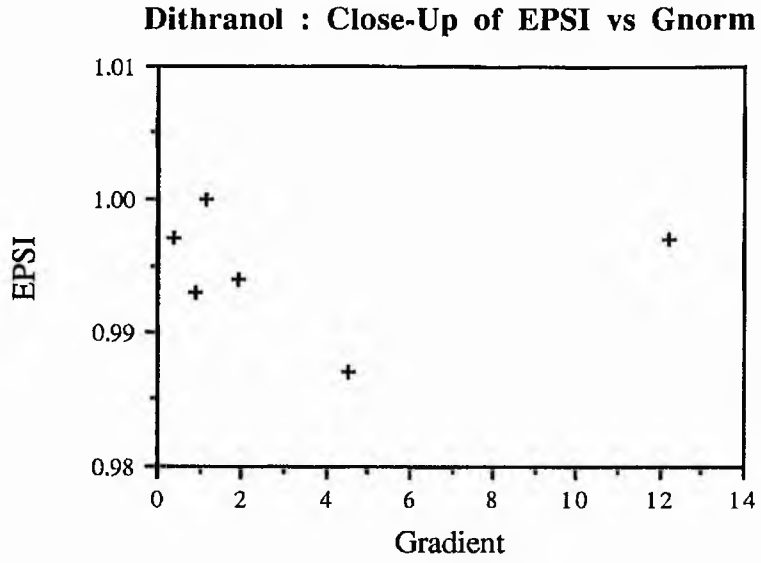


Figure 4.9

The maximum time for the Dithranol calculations is longer than that for DHEA, despite the latter being a larger system because the DHEA calculations were performed upon the FPS-500, whilst those for Dithranol were run upon the Sun workstations.

(G4) Comparison of the Geometries of the Structures Generated in the Study.

The EPSI is extremely useful in that it assess the differences in both the geometry and the charges of the different conformations. The algorithm used in the ASP program was designed to exclude from the similarity calculation any region within the van der Waals' surface of either molecule. Hence small differences in the geometry probably have little effect upon the EPSI because points upon the grid used in ASP that are near the changes will be excluded.

In this section, the effects of the optimisation process upon the bond lengths, angles and dihedral angles of the molecule are investigated. To examine these changes, the values of the differences in these parameters between some of the intermediate structures and the precise structure are evaluated. As described in the introductory

discussion of this study, this process is performed automatically using the molecular modelling program, MHCDraw, presented in Chapter Three. The resulting graphical output for selected molecules is presented in Figures 4.10 and 4.11, whilst the unsigned average of the differences in the structures are detailed in Tables 4.5 and 4.6.

The following graphs indicate the progress of the optimisation and show that range of differences in the geometric variables. As two geometries become more similar, the graphs should no longer depict a spread of values, but a single peak, centred at a difference of 0.0.

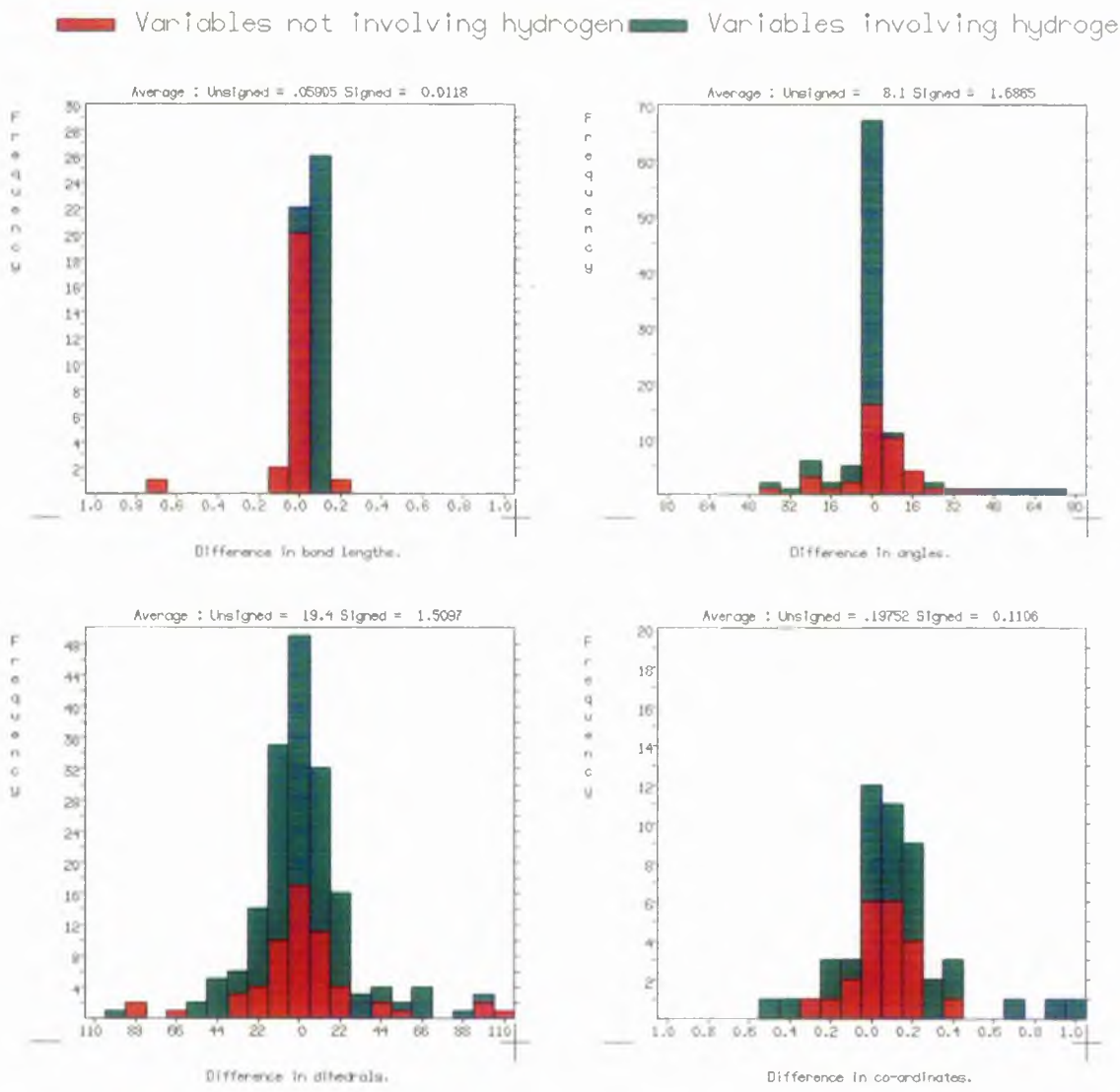


Figure 4.10a.
DHEA : Differences between Precise and Initial Sturctures.

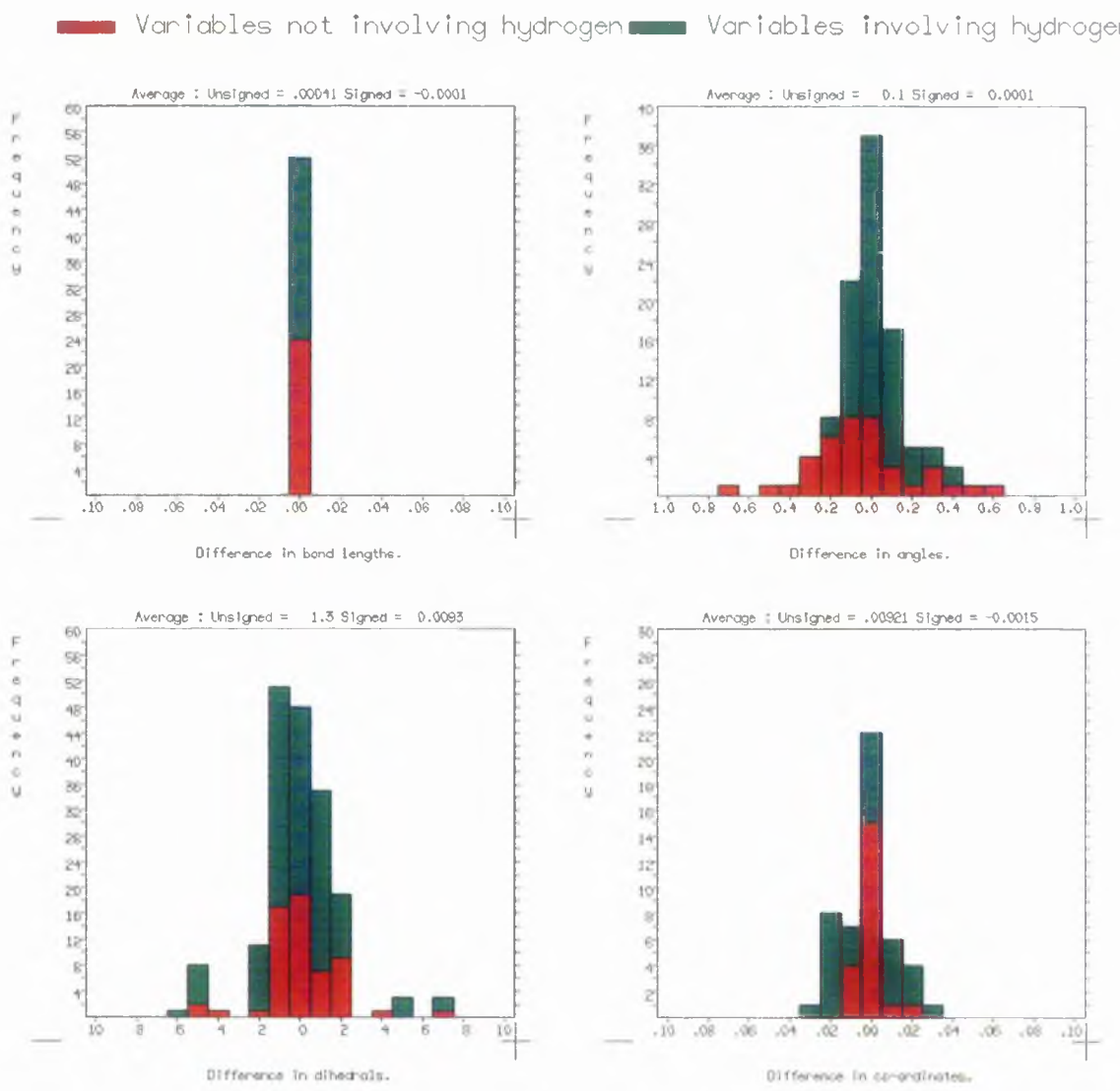


Figure 4.10b.
DHEA : Differences between Precise and Optimised Sturctures.

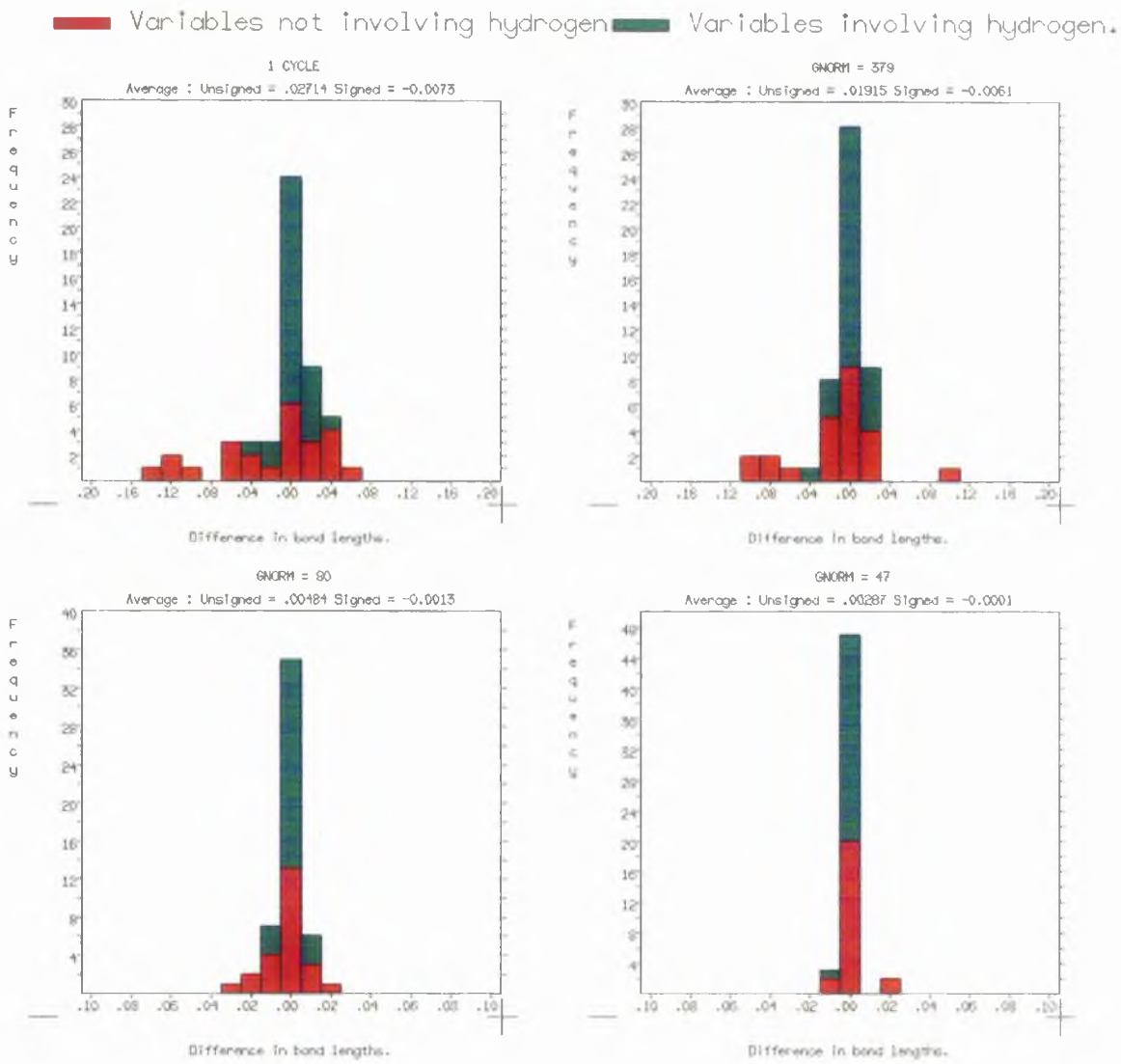


Figure 4.10c.
DHEA : Changes in Bond Lengths During Optimisation.

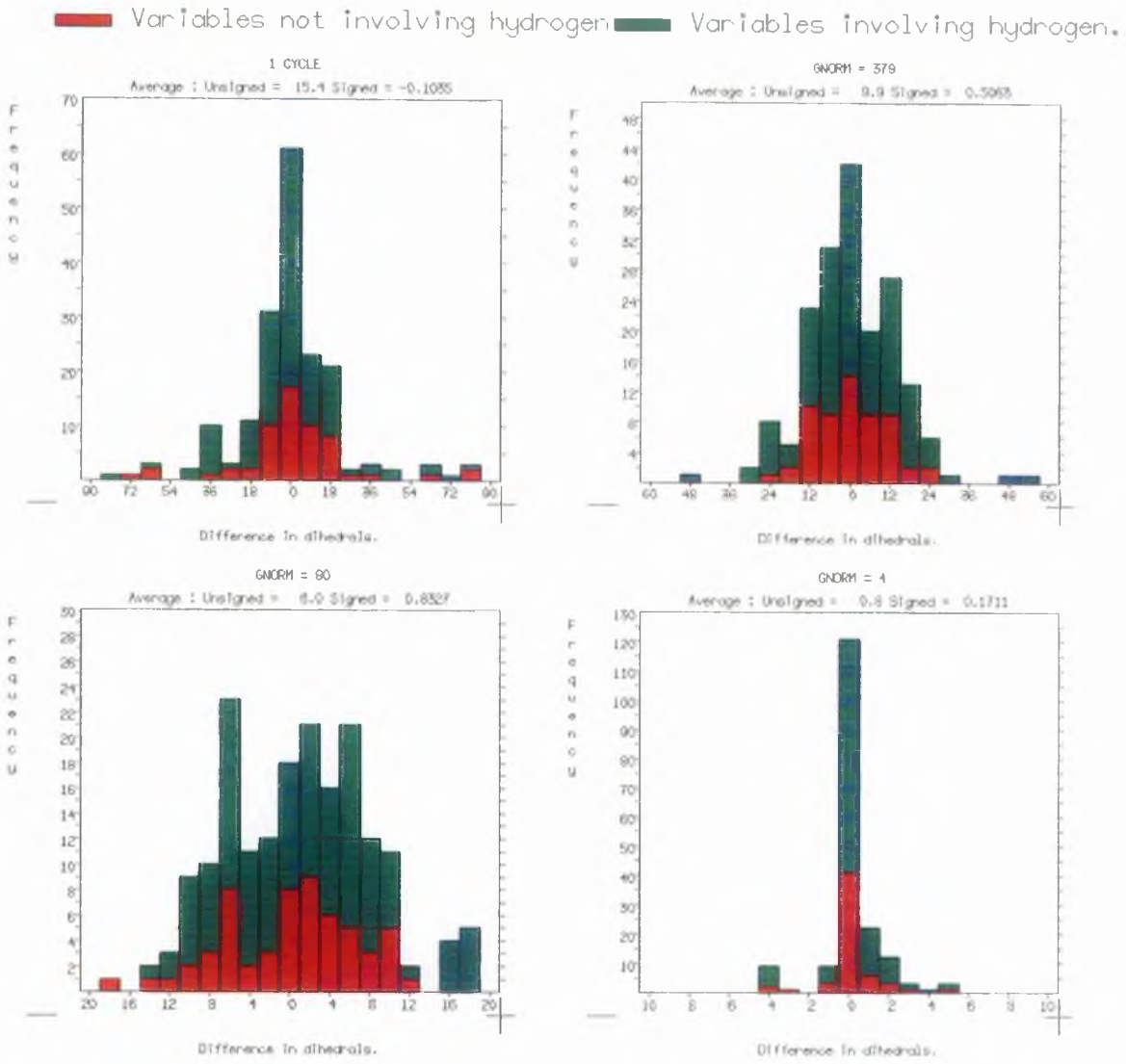


Figure 4.10d.
DHEA : Changes in Dihedrals During Optimisation.

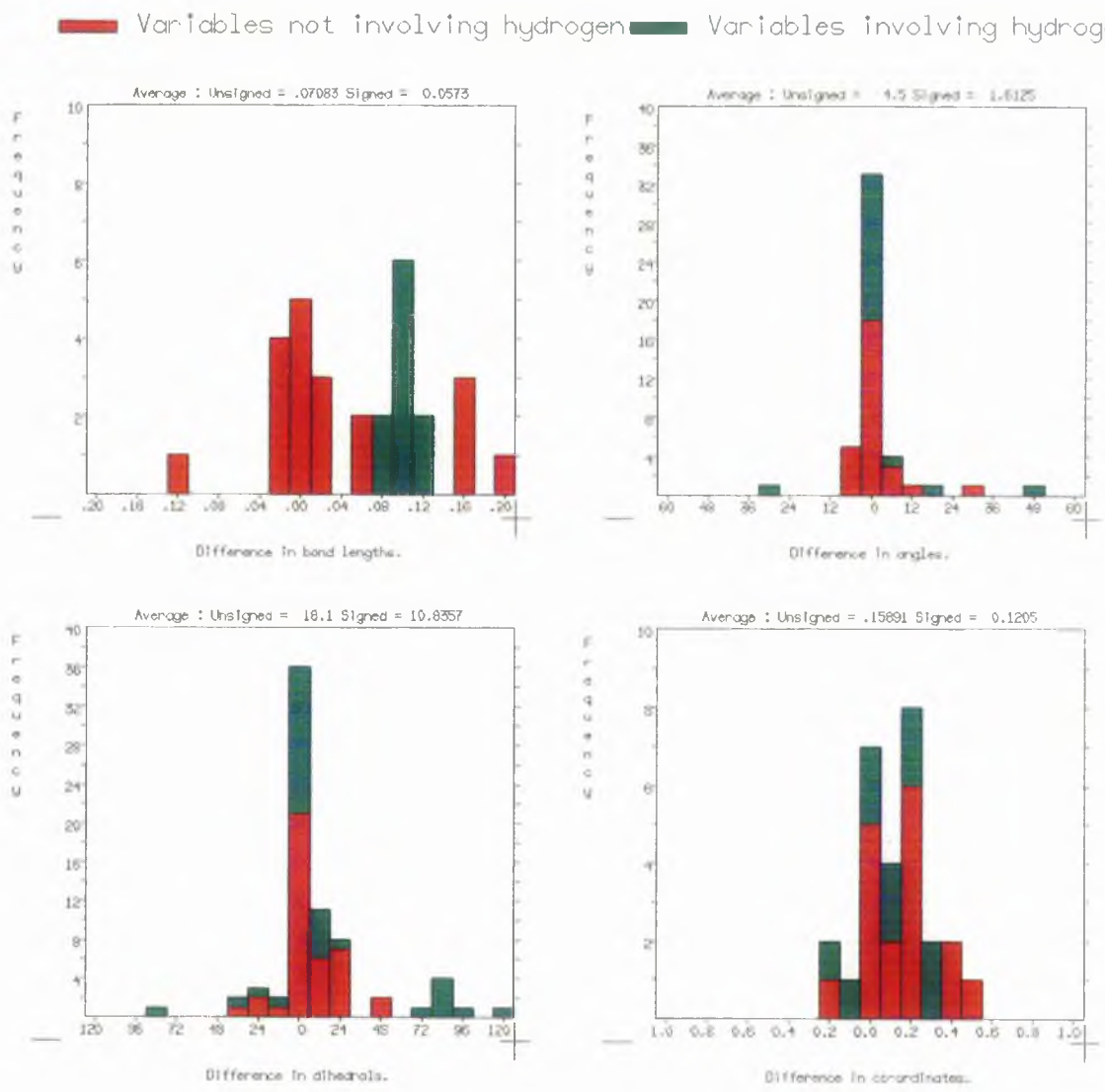


Figure 4.11a.
Dithranol : Differences between Precise and Initial Sturctures.

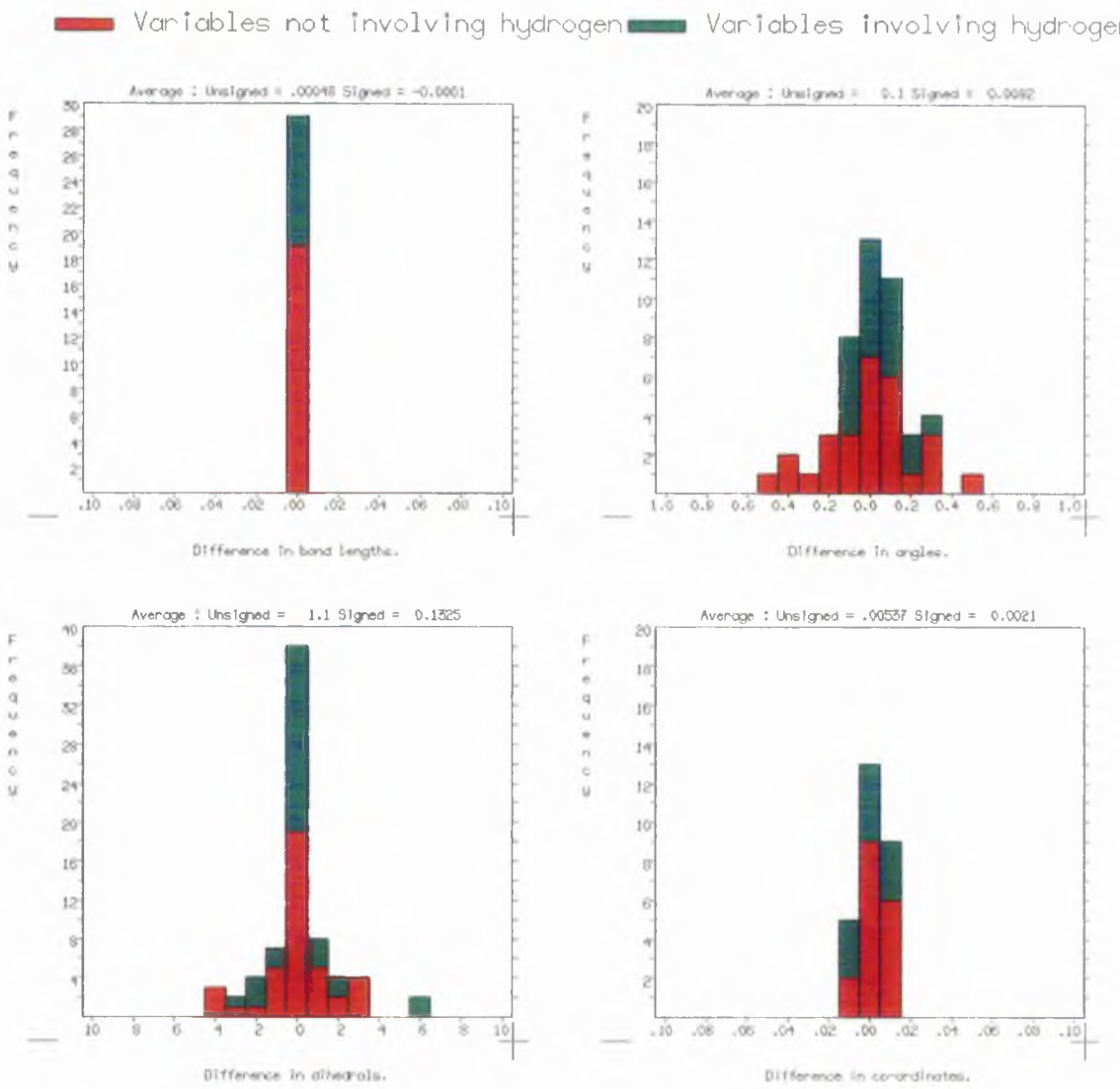


Figure 4.11b.
Dithranol : Differences between Precise and Optimised Sturctures.

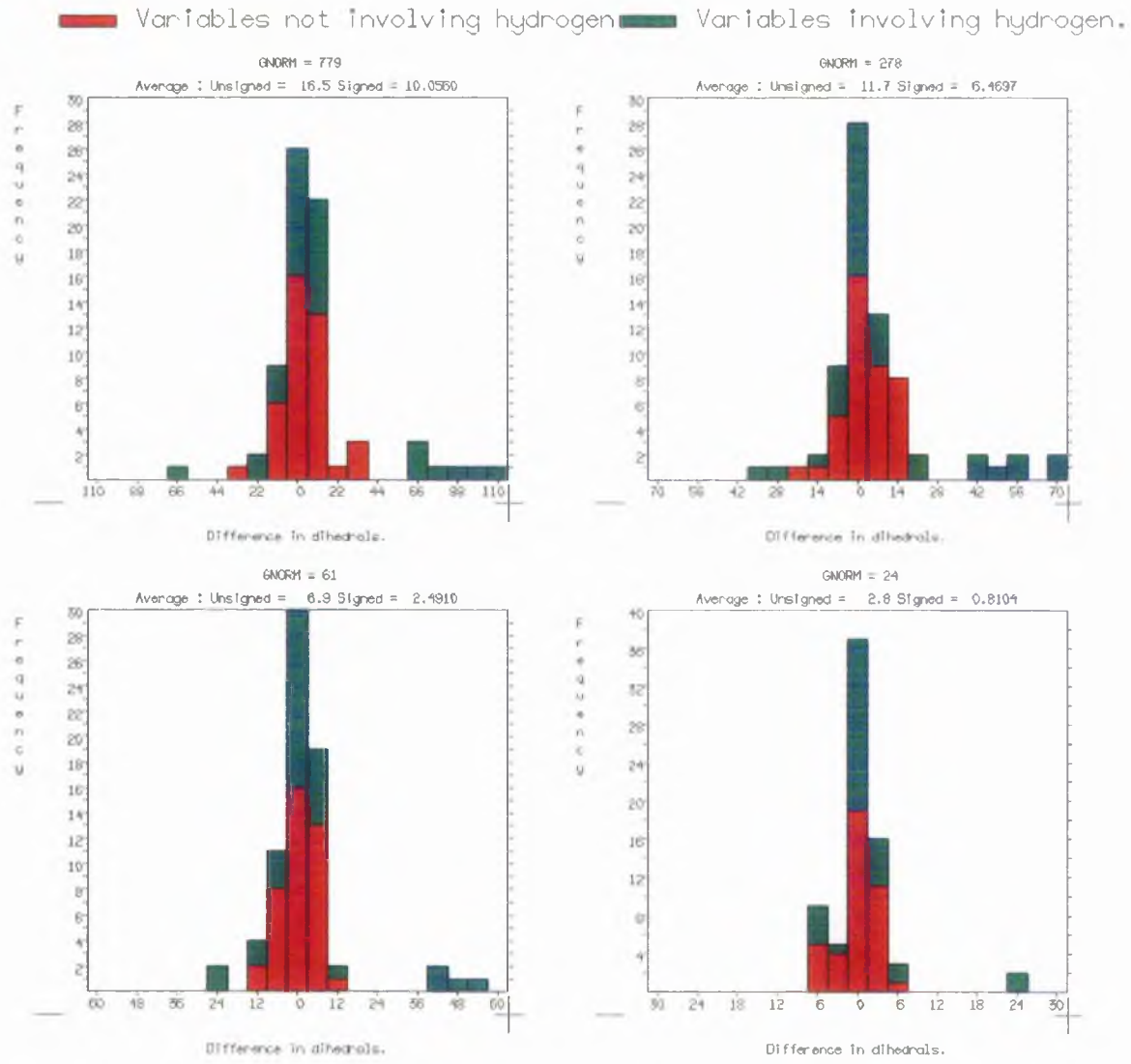


Figure 4.11c.
Dithranol : Changes in Dihedrals During Optimisation.

(G5) Discussion.

Figure 4.10a represents the differences between the initial and precise structures of DHEA. These structures are illustrated in Figure 4.12. It can be seen that the use of standard bond lengths has been extremely successful for all except one bond. The average difference is 0.0590 Å, whilst the unsigned average is 0.012 Å. Since the graphs represent (Precise Value) - (Intermediate Value), the latter average indicates that the initial estimate of the bond lengths was slightly too short. Close examination of the graph reveals that this deviation is due to the positioning of the hydrogen atoms in the molecule.

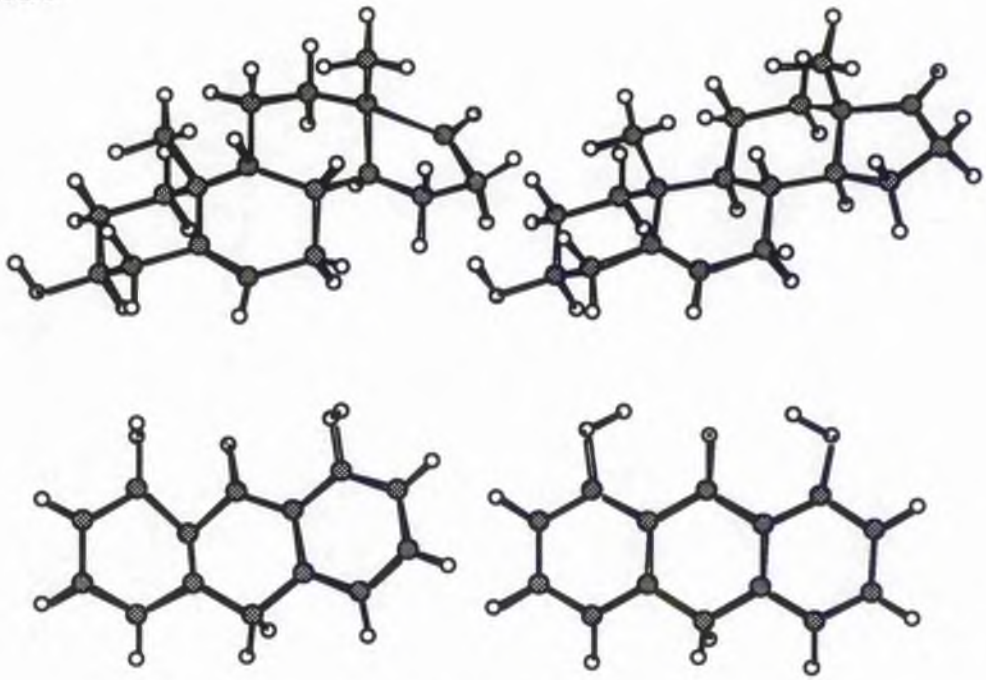


Figure 4.12

Initial and Final Structures of DHEA (above) and Dithranol (below).

The assignment of the angles was less successful, with a range of values differing by up to 80°, although the average error is only 8.1°. Dihedral angles were also similarly poorly estimated, with an average error of 19.4°. The difference in co-ordinates plot indicates how far the individual atoms are from their positions in the precise structure, and the average of this is quite small, being 0.1975 Å.

Figure 4.10b represents the differences between the results of the precise and standard optimisations, the latter being refered to as the optimised structure. It indicates that the bond lengths are almost exactly the same and that the deviation in angles is small (0.1°). Nevertheless, some of the dihedrals differ by up to six degrees.

Figure 4.10c illustrates how the bond lengths change during the optimisation, showing the differences to the precise structure after one cycle, and with gradient norms of 379, 80 and 47. It should be noted that the rogue bond depicted in Figure 4.10a was quickly corrected and that even with a Gnorm as high as 47, the bond lengths are extremely accurate.

The last of the DHEA plots (Figure 4.10d) depicts the spread of the dihedral angles at four stages during the study. It can be seen that these parameters do not attain an average difference of less than one until the gradient norm reaches single figures.

Figures 4.11a to c represent a similar study upon Dithranol, with (a) being a comparison of the initial and precise structures. In this, the bond lengths are less successfully estimated, especially for the hydrogen atoms, which are all significantly underestimated. This is due to the fact that the same value was used for this aromatic compound as for the aliphatic case of DHEA.

Figure 4.11b depicts a comparison of the precise and optimised structures, which again exhibits a small spread in the dihedrals.

The last graph, Figure 4.11c, shows how the differences in the dihedral angles decrease during the optimisation process.

The average values of the differences in bond lengths ($\overline{\Delta L}$ / Å), bond angles ($\overline{\Delta\theta}$ / °), dihedral angles ($\overline{\Delta\tau}$ / °) and co-ordinates ($\overline{\Delta C}$ / Å) are tabulated below for a selection of intermediate structures for DHEA (Table 4.5) and Dithranol (Table 4.6), along with the gradient norms associated with each geometry.

File	$\overline{\Delta L}$ / Å	$\overline{\Delta\theta}$ / °	$\overline{\Delta\tau}$ / °	$\overline{\Delta C}$ / Å	Gnorm
1 SCF	0.0591	8.1	19.4	0.198	4676.722
1 CYC	0.0271	6.8	15.4	0.791	1532.294
GNORM 800	0.0259	5.8	14.1	0.178	733.261
6 CYC	0.0214	4.1	10.9	0.148	410.956
GNORM 400	0.0192	3.7	9.9	0.125	379.309
GNORM 200	0.0079	2.0	7.1	0.093	151.210
GNORM 100	0.0048	1.5	6.0	0.066	79.991
GNORM 50	0.0029	0.7	5.6	0.036	47.293
GNORM 10	0.0011	0.3	2.9	0.029	9.933
OPT	0.0000	0.1	1.3	0.009	9.773
GNORM 1	0.0000	0.0	0.2	0.002	0.960

Table 4.5

File	$\overline{\Delta L}$ / Å	$\overline{\Delta\theta}$ / °	$\overline{\Delta\tau}$ / °	$\overline{\Delta C}$ / Å	Gnorm
1 SCF	0.0708	4.5	18.1	0.159	2449.427
GNORM 2000	0.0289	3.8	16.5	0.122	779.040
T=650	0.0172	3.7	11.0	0.067	414.889
GNORM 500	0.0244	3.8	15.5	0.094	407.780
GNORM 300	0.0167	4.0	11.7	0.080	278.271
GNORM 100	0.0044	1.7	6.9	0.042	60.857
GNORM 50	0.0026	0.6	3.4	0.022	43.178
GNORM 25	0.0019	0.4	2.8	0.034	23.789
OPT	0.0005	0.1	1.1	0.005	12.219
GNORM 5	0.0005	0.2	1.4	0.008	4.559
GNORM 1	0.0002	0.1	1.1	0.006	0.900

Table 4.6

(G6) Conclusions.

Despite fairly poor initial geometries, both molecules started with EPSI s of > 0.7, and this value rose slowly until the gradient norm reached about 500, at which point is started to rise rapidly. The graphs depicting the expanded low gradient region show that the EPSI is accurate to within 1 or 2 % once the gradient norm has dropped below about 10.0, which is normally about the level at which the optimisation (without *PRECISE*) terminates.

The study of how the geometry varies during the optimisation shows that errors in bond lengths are quickly ironed out, and that the final stages of the optimisation involve the "fine - tuning" of the dihedrals, and to a lesser extent, the bond angles.

Hence, if the purpose of the study is to produce structures and charges to calculate the MEP, it is not necessary to force the optimisation below these levels, leading to a possible saving of about 50 % of the time needed for the most precise calculations performed. Nevertheless, greater accuracy may be required for other types of work, such as force calculations to evaluate vibrational frequencies.

Chapter Five

A Theoretical Investigation of DHEA and Other G6PDH Inhibitors.

(A) Introduction.

Dehydroepiandrosterone (DHEA, Figure 1.3), is a steroid secreted by the adrenal cortex, and is produced in one of the highest concentrations of all steroids [87]. It has been shown to be a particularly interesting compound, having cancer preventative and anti - carcinogenic action [88-93], anti - obesity effects [92,94], it is anti - atherogenic [95], has ability to inhibit viral transformation of human lymphocytes [91], and can inhibit DNA synthesis [96].

Concentrations of DHEA in humans remains high until after the second decade of life when they fall off dramatically, finally reaching only 10 - 20 % of the peak levels [97]. It is interesting to note that cancer is sometimes referred to as a 'disease of ageing' because the incidence of most cancers greatly increases with age, and thus there could be some link with DHEA levels and cancer incidence. A study of the excreted metabolites of DHEA indicate a low level of the steroid is associated with a high risk of breast cancer in women [98].

The cancer preventative action of DHEA cited above includes the inhibition of spontaneous mammary tumours in certain breeds of mice [99,100], whilst examples of its inhibition of carcinogenesis include giving protection against carcinogenesis by 7,12-dimethylbenz[a]anthracene (a PAH) [92], urethan [92], dihydroxy-di-n-propylnitrosamine [88,93] and 1,2-dimethylhydrazine [90].

The anti - obesity effects of DHEA are also of interest because it has been noted that restriction of calorie intake has a general anti - carcinogenic effect in laboratory animals [101,102]. Nevertheless, Schwartz and Tanne concluded that although the action of DHEA may be partly through the suppression of weight gain, the steroid may also act directly as an anti - tumour promoter [92].

(B) DHEA as an Inhibitor of G6PDH.

(B1) Structure and Function of G6PDH.

It is thought that the mode of action of DHEA is through its inhibition of the enzyme Glucose-6-Phosphate Dehydrogenase (G6PDH) [103].

This is a large enzyme, the active form of which is either dimeric or tetrameric, with each of the subunits having a molecular weight of about 50,000.[1] The ratio of dimer and tetramer depends upon such factors as pH and ionic strength. The structure of the active form also involves molecules of Nicotinamide - Adenine Dinucleotide Phosphate (NADP), which has both an active role and a structural role, as depicted in the hypothetical schematic diagram in Figure 5.1, adapted from Luzzato and Testa [104].

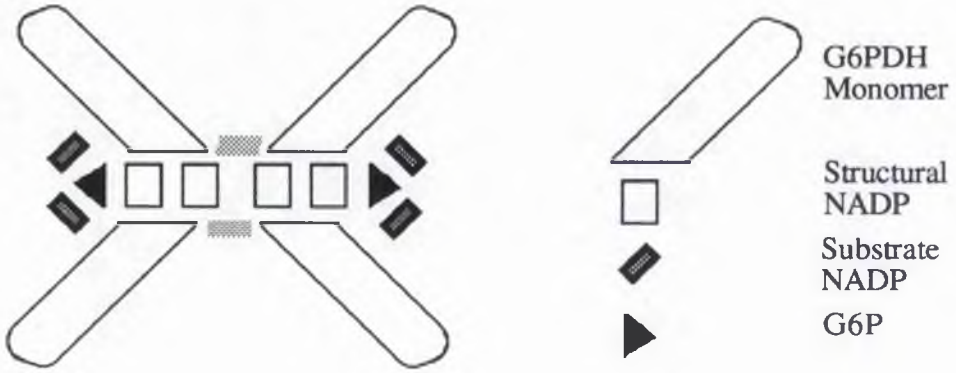


Figure 5.1

Schematic Diagram of G6PDH Tetramer

Four NADP molecules hold the tetramer together, whilst it is hypothesised that for each dimer, there are two molecules of Glucose-6-phosphate (G6P) substrate and two of substrate NADP.

---

[1]G6PDH structure and function is reviewed in 104.

The reaction catalysed by the enzyme is an important step in the system of energy producing reactions known as the pentose - phosphate shunt:

$$G6P + NADP \rightleftharpoons 6PG + NADPH$$

where 6PG is 6-Phosphoglucono lactone.

Thus G6PDH is a major extra - mitochondrial source of NADPH, which is an extremely important biological compound utilised as an energy source to drive biological reactions. The other product (6PG) is a precursor of the five - membered ring (pentose) sugars which are essential precursors for the synthesis of both DNA and RNA.

Thus there are two ways in which G6PDH inhibition can have anti - cancer effects. Firstly, reduction of NADPH levels leads to a decreased ability of cells to metabolise pro-carcinogens to their active forms. Secondly, cancer is usually associated with high rates of cell division, which requires high rates of DNA and RNA synthesis, which requires the presence of precursor sugars.

(B2) Inhibition of G6PDH by Steroids.

DHEA is a potent inhibitor of mammalian G6PDH [105-108]. The fact that the inhibition is non-competitive implies that the action is not mediated by competition between the substrates and DHEA for the active site of the enzyme, and hence the latter must bind at a separate (possibly remote) site. It is likely that the anti - carcinogenic activities of DHEA are exerted through this inhibition because certain Mediterranean variants of the enzyme have a low catalytic activity, and associated with this is a reduced ability to metabolise pro - carcinogens. It is interesting that this effect is similar

to that produced by DHEA administration and also that the effects can be removed by administration of exogenous G6PDH [109]. DHEA also inhibits DNA synthesis [96], as would be expected of an inhibitor of this enzyme, indicating the importance of G6PDH in the study of cancer.

(C) Quantum Mechanical Study of Steroidal Inhibitors of G6PDH.

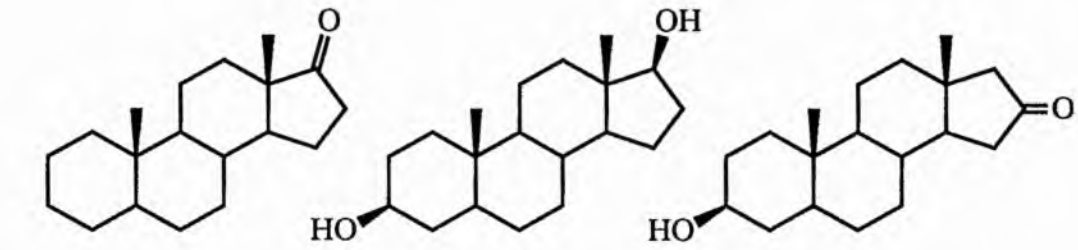
G6PDH is also inhibited by a range of other steroids and it was decided to study some of these inhibitors in an attempt to rationalise the activities of the various steroid hormones. The first data obtained for a large number of compounds were those of Raineri and Levy [106]. This tabulates the enzyme activity in the presence of a number of steroids, and is expressed as enzyme activity remaining upon addition of the inhibitors.

(C1) Methods.

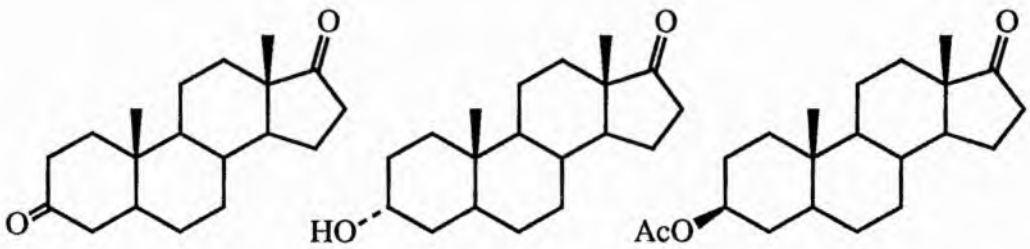
The structures of the various steroids were optimised using the AM1 and PM3 methods available in MOPAC and the resultant structures and charges were extracted and used to draw MEP maps using 3D2 and to calculate Electrostatic Potential Similarity Indices (EPSI s) using the ASP program, details of which are given in Chapter Two.

Optimisations were performed at the precise level for most of the molecules for which experimental activity data were available the Androstane series and the Pregnane series. The exceptions were the molecules for which the activity was measured at a different concentration due to solubility difficulties or for the 5-β steroids, in which steric differences are liable to be the major factors in determining activity.

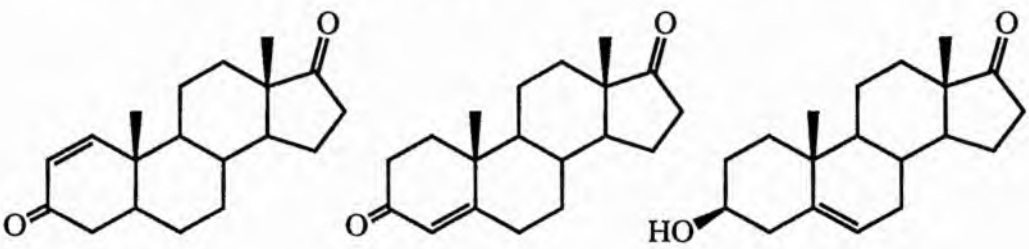
The molecules studied in the Androstane series are shown below in Figure 5.2, with the abreviated names used in the text. Also shown are the abbreviated names used in the text, which are derived from the number of the molecule in the Raineri and Levy paper. The molecules from the Pregnane series follow in Figure 5.3.



And-02 And-05 And-07

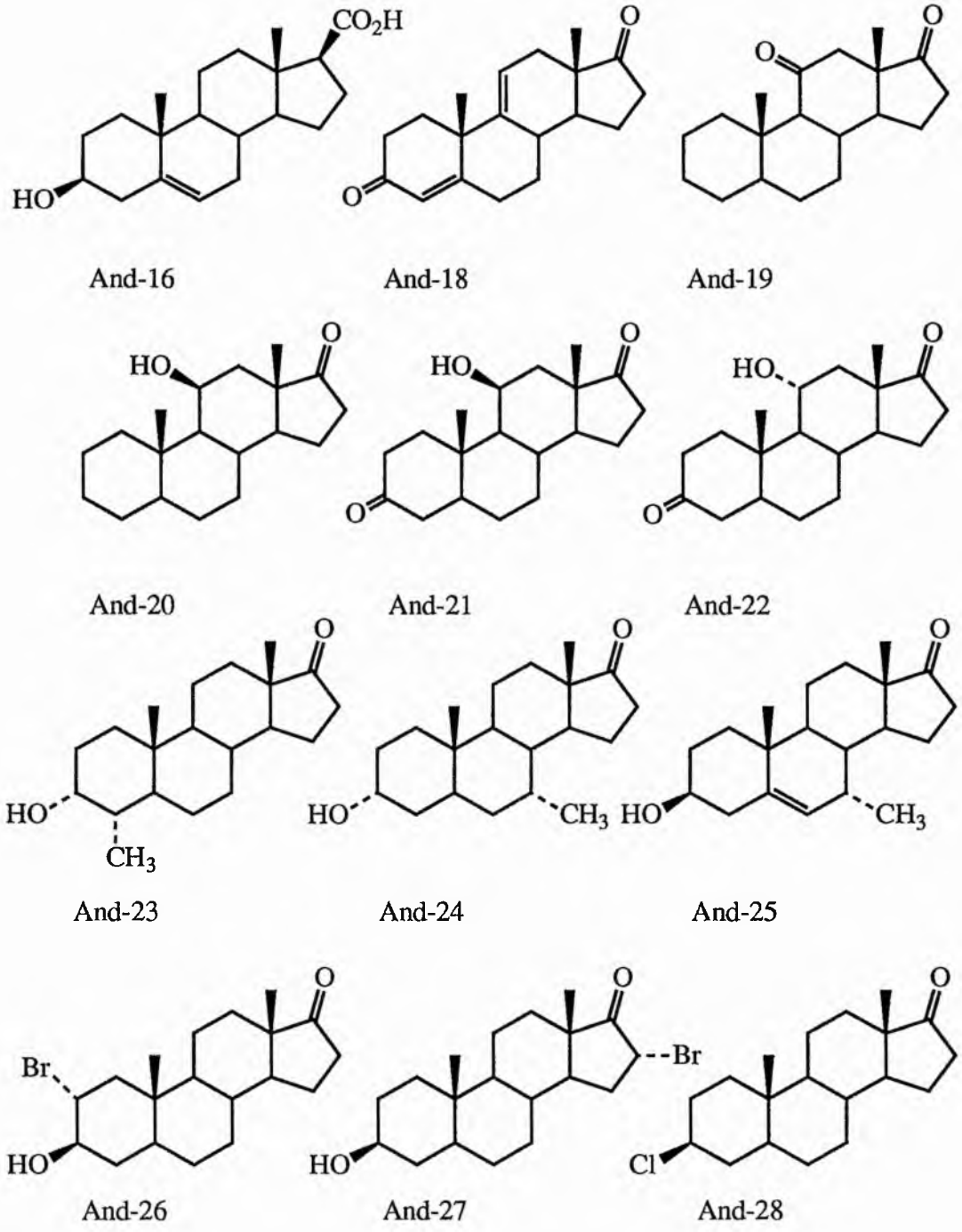


And-09 And-11 And-12



And-13 And-14 And-15

Figure 5.2



And-16 And-18 And-19

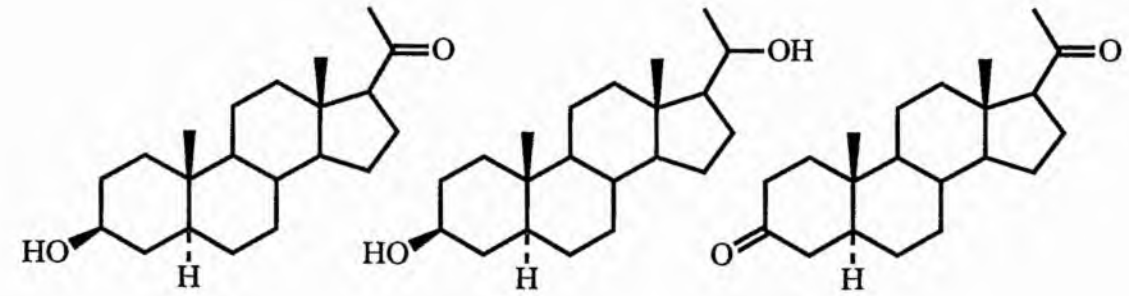
And-20 And-21 And-22

And-23 And-24 And-25

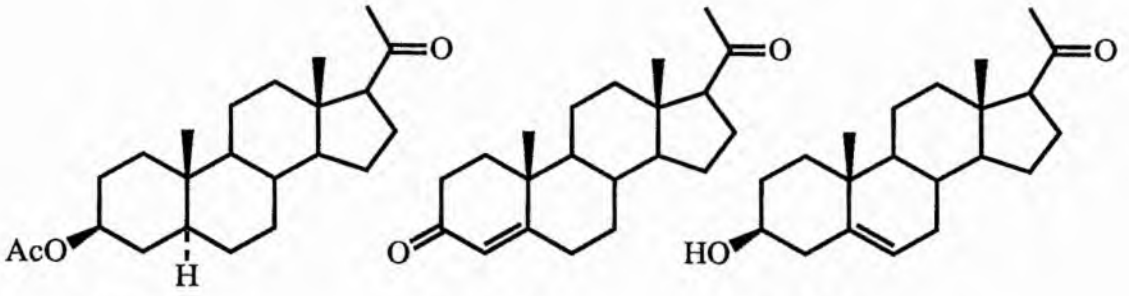
And-26 And-27 And-28

Figure 5.2 (Cont.)

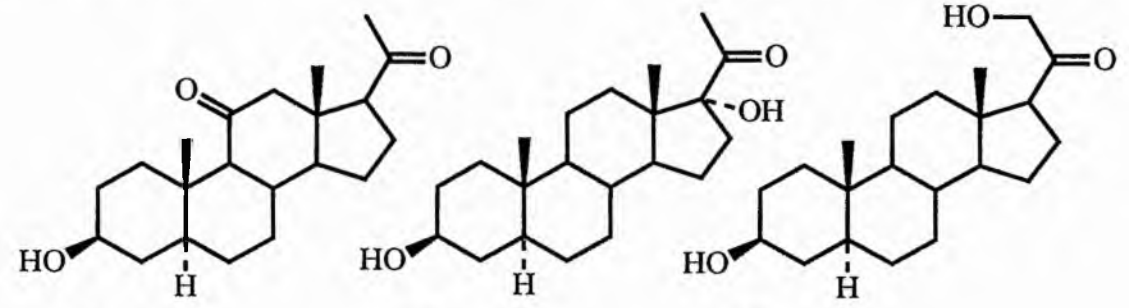
Molecules in the Androstane Series.



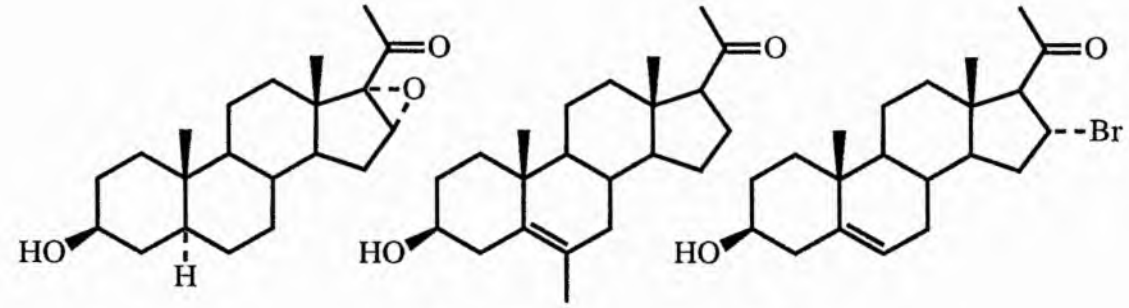
Preg-02 Preg-03 Preg-05



Preg-06 Preg-07 Preg-08



Preg-09 Preg-10 Preg-11



Preg-12 Preg-13 Preg-14

Figure 5.3.

Molecules in the Pregnane Series.

(C2) Conformational Analysis of Molecules in the Pregnane Series.

All the molecules in this group of G6PDH inhibitors have a β-side chain at the 17 position on the D-ring. This group contains two carbon atoms, and presents the only major possibility for the molecule to attain different conformations, in what is, in general, a rigid series of molecules. Figure 5.4 represents these rotations, which are about C_{17}-C_{20} and C_{20}-C_{21} respectively. In most cases, C_{21} is the centre of a methyl group, and hence the rotation about the latter of these two bonds can be ignored because the group will be able to rotate freely. The conformation of this group is only of interest in the molecule Preg-11, which has a 21-hydroxy group.

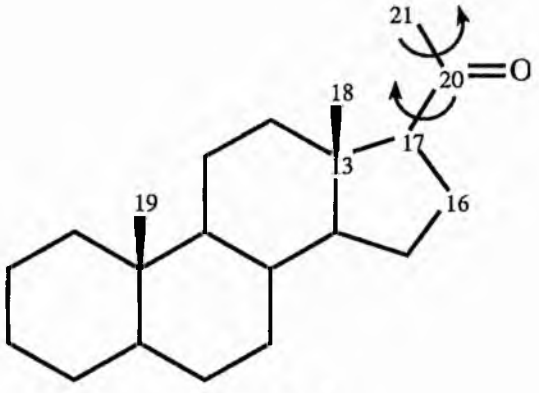


Figure 5.4

Possible Bond Rotations in Pregnane Side Chain.

The following molecules were studied to determine the minimum energy conformation to use in the MEP work, whilst the others had the same functionality in the D - ring as one of the molecules below.

Molecule	Rotate About	Groups of Interest
Preg-03	C_{17}-C_{20}	20-hydroxy
Preg-05	C_{17}-C_{20}	20-keto

Preg-10	C_{17}-C_{20}	20-keto, 17α-hydroxy
Preg-11	C_{17}-C_{20}	20-keto, 21-hydroxy
Preg-12	C_{17}-C_{20}	20-keto and 16α, 17α–epoxy
Preg-14	C_{17}-C_{20}	20-keto, 16α-bromo

In addition, Preg-11 was studied by rotating about C_{17}-C_{20} and C_{20}-C_{21} simultaneously.

The aim of this study was to investigate the possibility that some of the molecules may be inactive because they are unable to attain a specific conformation necessary to give enzyme inhibition, due to steric constraints within the molecule. This question is particularly important for molecule Preg-14, which is totally inactive, despite the fact that the Androstane equivalent (And-27) is substantially more active an inhibitor than is DHEA.

(C3) Method of Conformational Analysis.

It is possible to perform a 1-Dimensional and 2-Dimensional conformational analysis automatically using MOPAC. The calculated heats of formation can then be plotted against the conformational variable to give a reaction profile for the rotation. In the 1D case, the user enters specific values for the geometric variable, which in these cases is generally the dihedral angle C_{21}-C_{20}-C_{17}-C_{13}. For the 2D, the program defines a grid of eleven by eleven conformations and evaluates the energy at each of these. The results for the latter method can be displayed either as a contour map, or as a 3D surface graph from which the minimum energy conformations can be located.

In both cases, the specified dihedrals are fixed at the required values, and the rest of the molecule can either be fixed, and a single point calculation performed, or a full or partial optimisation can be performed. In the case of large steroid molecules, full

optimisation of 121 different conformations can be extremely time consuming, so a partial optimisation is necessitated. Such a process will usually yield sensible results in this sort of molecule, where interactions of the 17-side chain and the A- and B-rings are extremely small. Calculations without any optimisation are inadvisable because of the possibility of very high energies due to steric crowding which could be relieved during the optimisation process, but can be used for a rapid assessment of the positions of the potential minima.

A further approximation that can be used to reduce computational time in a conformational search is the use of model compounds. This technique has been used in the next chapter in the study of inhibitors of Aromatase.

The 1D searches have generally been made using the final point from the preliminary MOPAC optimisation as a starting point. The other points were taken in 20° steps, starting at the nearest multiple of 20 above the starting point and wrapping around from 360° to 0°. Full optimisation was allowed.

The 2D search was made using steps of 36°, and spanning the full 360° of conformational space for both variables. Again the geometry of the preliminary MOPAC optimisation was taken as the starting point.

(C4) Results of Conformational Searches.

The results for these calculations are presented in the form of reaction-profile graphs in Figures 5.5 to 5.8 and 5.10 to 5.11, with the reaction co-ordinate being the dihedral C_{21}-C_{20}-C_{17}-C_{13}.

(C5) Discussion.

Preg-03.

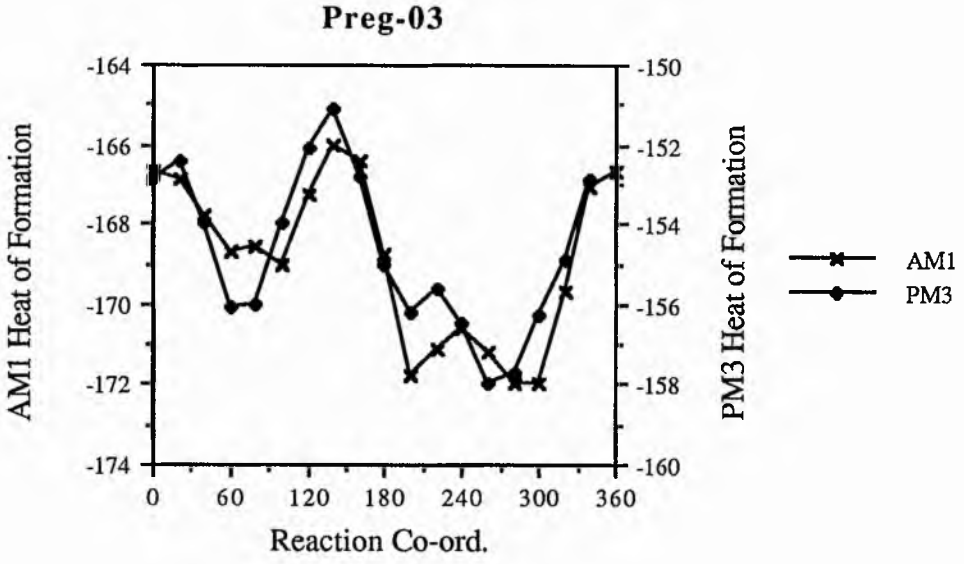


Figure 5.5.

Both the AM1 and PM3 curves are fairly similar, and both indicate an approximate barrier to rotation (given by the difference in maximum and minimum energies) of about 5.9 kcal mol^{-1} for AM1 and 6.9 kcal mol^{-1} for PM3. The peaks in the graphs are due to the interactions shown in Table 5.1 where τ represents the value of C_{21}-C_{20}-C_{17}-C_{13}.

τ / °	Interaction
0	C_{21} Methyl and C_{18} Methyl
130	OH and C_{18} Methyl
250	H on C_{20} and C_{18} Methyl

Table 5.1

There are two fairly equivalent minima at $\tau = 200°$ and $\tau = 290°$, and full optimisations at these geometries yield ΔH_F values of -171.84 kcal mol^{-1} and -172.28

kcal mol^{-1} for AM1 and -156.90 kcal mol^{-1} and -157.98 kcal mol^{-1} for PM3 respectively. These values indicate that the two potential wells are of similar depth and are likely to be similarly populated, with an easy transition between the two due to the low barrier height. The minimum at $\tau = 80°$ is less likely to be important because it is about 3 kcal mol^{-1} higher in energy and separated by large barriers.

It is interesting to note that the $\tau = 130°$ peak in the PM3 case is larger than the $\tau = 0°$ peak, indicating a greater repulsion indicated by the C_{20}-hydroxy group than the C_{21}-methyl group, the reverse of the AM1 situation

Preg-05.

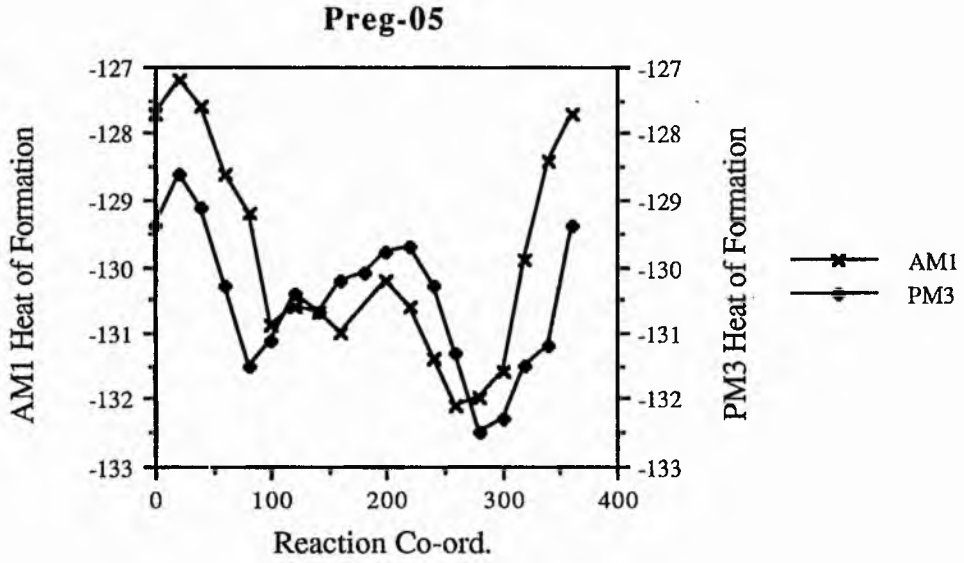


Figure 5.6.

This molecule has a keto - group attached to C_{20}, and the two rotation profiles indicate differences in the way the two methods handle the interactions of this group with others.

Both methods predict the lowest minimum at $\tau = 290°$, with other minima at about $\tau = 100°$ and $\tau = 160°$, but the barriers between these differ : the total barrier to rotation is 4.9 kcal mol^{-1} for AM1 and 3.9 kcal mol^{-1} for PM3, whilst the latter method indicates a barrier that is about 1 kcal mol^{-1} higher for the peak at $\tau = 220°$. The peaks are described in Table 5.2.

τ / °	Interaction
0	C_{21} Methyl and C_{18} Methyl
120	C_{21} passes close to C_{16}
220	Carbonyl O and C_{18} Methyl

Table 5.2

The minimum configuration is $\tau = 268.48°$ for AM1 and $\tau = 286.75°$ for PM3, although the former predicts a greater likelihood of a less restricted rotation, with possible values of τ between 100° and 300°.

Preg-10.

The rotation profile for this molecule is shown over the page in Figure 5.7. In addition to a 20-keto group, this molecule also has a 17α-hydroxy group, which will interact with the former during the rotation. The two groups show a major difference in that the overall barrier height is again larger for AM1, but the carbonyl - C_{18} methyl interaction is larger for PM3. The minimum energy configuration is clearly at about $\tau = 290°$. In the AM1 case, it would seem that this is the only minimum. This minimum is fairly broad, and the possibility of τ varying between 100° and 300° cannot be dismissed if the AM1 profile is the more accurate. In the PM3 case, the larger barrier height makes this less probable.

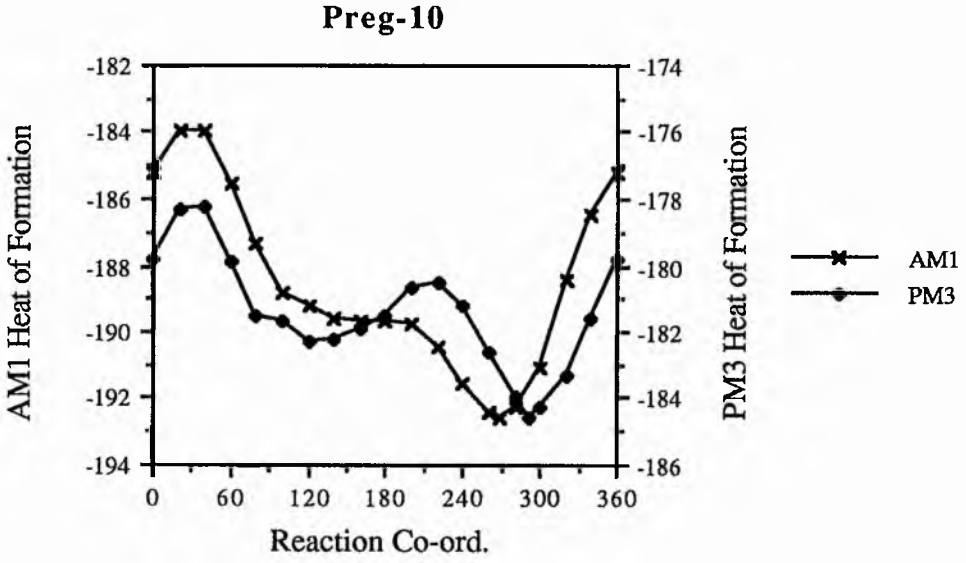


Figure 5.7.

Preg-11.

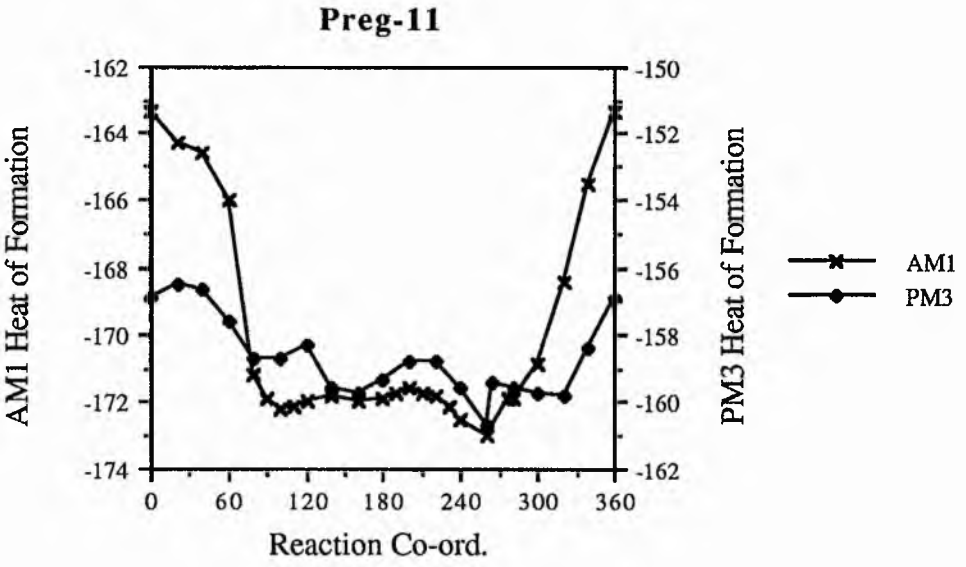


Figure 5.8.

In this study, the dihedral specifying the position of the OH - group was allowed to vary, and the rotation profiles for $\tau = C_{21}\text{-}C_{20}\text{-}C_{17}\text{-}C_{13}$,are completely

dissimilar. AM1 depicts a large barrier (9.7 kcal mol^{-1}) but a very broad minimum running from 100° to 300° with virtually free rotation in this range. The global minimum is at $\tau = 277°$. In the PM3 case, there are a number of intermediate peaks in this range, but the overall barrier is less than half of the AM1 value at 4.2 kcal mol^{-1}. The global minimum is the same as that for AM1, at $\tau = 265°$. The PM3 peaks are described in Table 5.3.

τ / °	Interaction
20	CH_2OH and C_{18} Methyl
130	CH_2OH and $C_{16}H_2$
220	C=O and C_{18}
270	C=O and C_{17}-Hα

Table 5.3

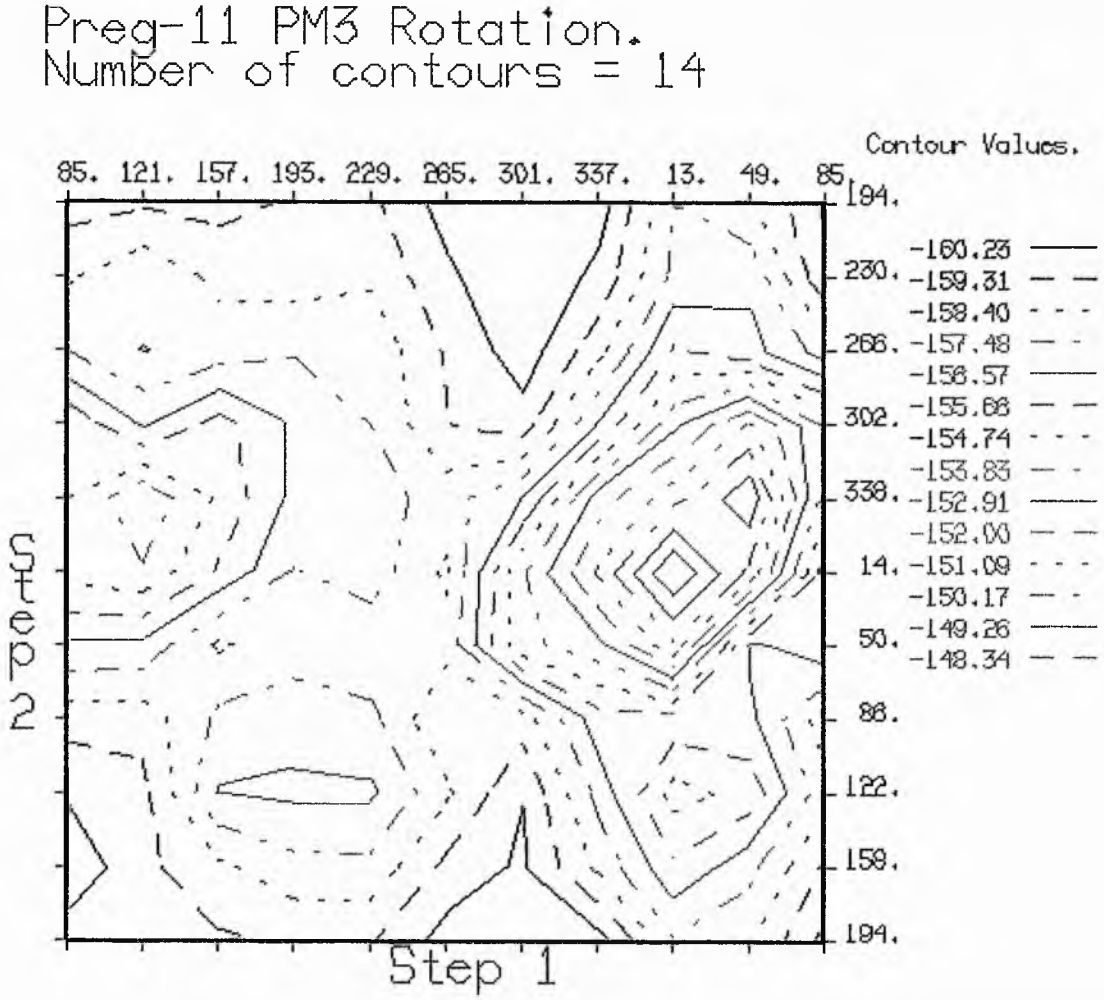

Preg-11 PM3 Rotation.
Number of contours = 14
Contour Values.
85. 121. 157. 193. 229. 265. 301. 337. 13. 49. 85
194.
230.
266.
302.
338.
14.
50.
86.
122.
158.
194.
-160.23
-159.31
-158.40
-157.48
-156.57
-155.66
-154.74
-153.83
-152.91
-152.00
-151.09
-150.17
-149.26
-148.34
Step 2
Step 1

Figure 5.9.

Figure 5.9, above, shows a contour map depicting the concurrent rotation, in 36° steps of dihedrals C_{21}-C_{20}-C_{17}-C_{13} (Step 1), which defines the postion of the carbonyl oxygen, and O_{21}-C_{21}-C_{20}-C_{17} (Step 2), which defines the position of the hydroxy group. The centre of the plot represents the starting geometry obtained from the initial optimisation, and it can be seen that this is not the overall minimum. Several minima can be located upon this map, and the AM1 equivalent, which has not been presented. The main PM3 minima are at represented in Table 5.4, along with the approximate heats of formation for these conformations. The first two of these represent the same minimum at opposite sides of the map.

τ_1 /°	τ_2 /°	Approx. ΔH_F (kcal mol^{-1})
301	194	-161.1
301	230	-161.1
85	158	-160.9

Table 5.4

Minimisations started at these points revealed that the global minima occur at (262.2°, 162.9°) for AM1 and (286.1°, 161.5°) for PM3, with heats of formation of -173.6 and -162.2 kcal mol^{-1}. This conformation places the carbonyl oxygen over the D - ring (as with the other molecules of this type), and the hydroxy group about 20° out of an eclipsing conformation with the carbonyl oxygen.

Since the ΔH_F for the result of the initial PM3 optimisation was -159.5 kcal mol^{-1}, the search has located points which are superior to the initial optimised geometry. Full optimisations were performed, starting with these structures, using both AM1 and PM3. The results are indicated in Table 5.5, with heats of formation in kcal mol^{-1}.

Method	τ_1 /°	τ_2 /°	τ_1 /°	τ_2 /°	ΔH_F
AM1	301	230	262.2	162.9	-173.6
PM3	301	230	318.6	211.3	-161.6
AM1	301	194	269.1	163.1	-173.6
PM3	301	194	286.1	161.5	-162.2
AM1	85	158	145.3	155.0	-172.5
PM3	85	158	84.1	164.9	-161.3

Table 5.5

Preg-12.

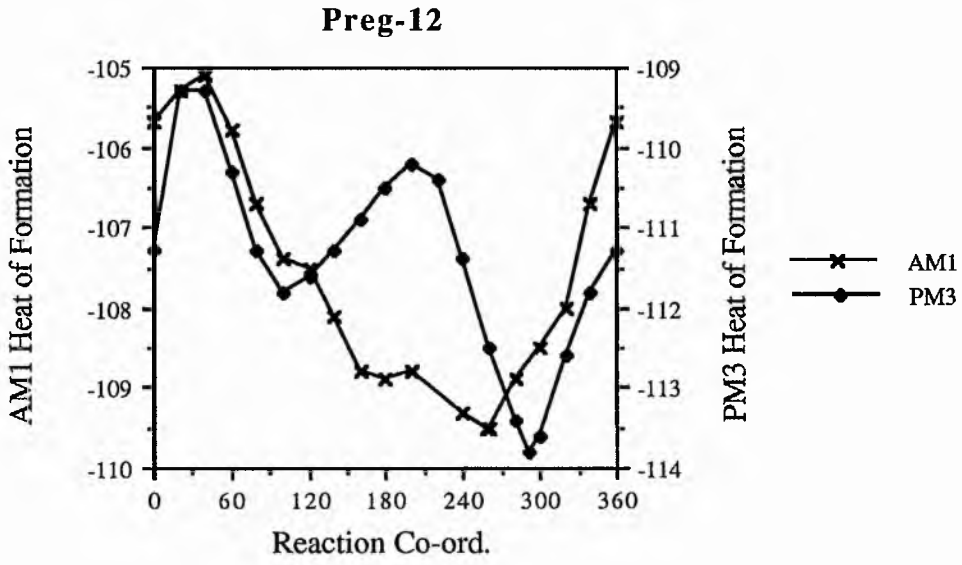


Figure 5.10.

The interaction of interest here is that of the C_{20}-keto group and the 16α, 17α-epoxy group, and again there is an obvious difference in the way in which AM1 and PM3 handle this. Both groups predict a similar overall barrier to rotation of 4.4 kcal mol^{-1}, but whereas AM1 predicts only one major repulsion (that of C_{21}-methyl and C_{18}-methyl groups), PM3 indicates an interaction of similar magnitude between

the C_{20}-keto and C_{18}-methyl groups at $\tau = 220°$. Any effect of the epoxy-group is dwarfed by this.

Thus AM1 predicts virtual freedom of τ between about 100° and 300°, whilst the 220° region of the profile is a high energy conformation for PM3. Both methods predict the overall minimum to be around 280° with exact optimised values of 277° and 265° for AM1 and PM3 respectively.

Preg-14.

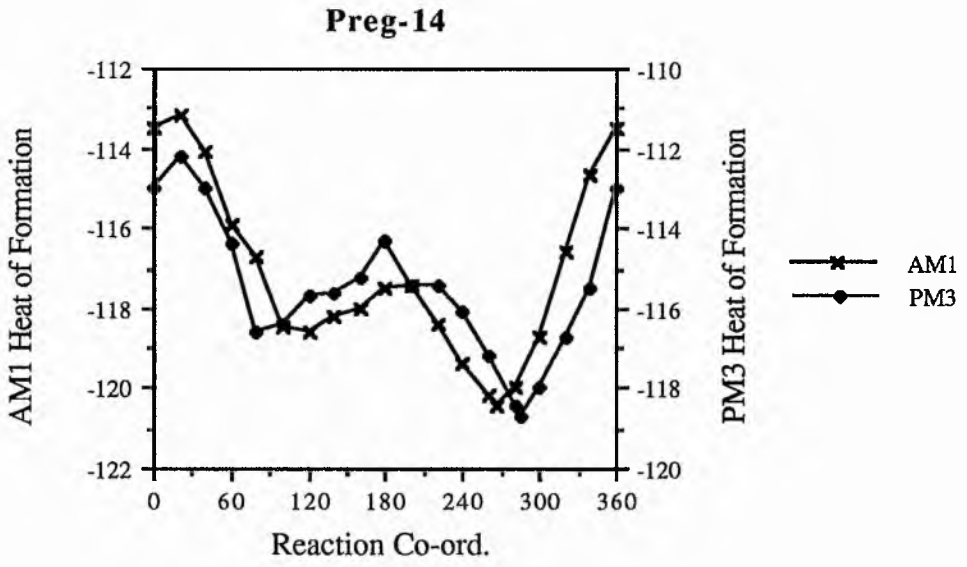


Figure 5.11.

As with the highly active And-27, this molecule contains a 16α-bromo group, but is totally inactive as an inhibitor of G6PDH. The rotation profiles in Figure 5.11 indicate that there are no steric reasons for this, and the molecule is quite able to take up the same conformations as the active Preg-13, which has a keto - group at C_{20} but no bromo - group. Both rotation profiles indicate similar overall approximate barriers of 7.2 and 6.4 kcal mol^{-1} for AM1 and PM3 respectively. The interaction between C_{20}-keto and C_{18}-methyl groups is again predicted to be greater by PM3 and any effects

due to the 16α-bromo group are small. The global minimum conformation for both methods is similar, with $\tau = 267°$ for AM1 and 286° for PM3, which is very similar to Preg-13.

(C5) Conclusions.

In general, it is likely that the rotation of the Pregnane C_{17} side chain is restricted, although the barriers to rotation vary between about 4 and 10 kcal mol^{-1}.

Both semi - empirical methods agree to the fact that the global minimum value for τ is about 270°, indicating that C_{21} points towards the back of the molecule, C_{20} = O or C_{20}-OH points forwards (Figure 5.12).

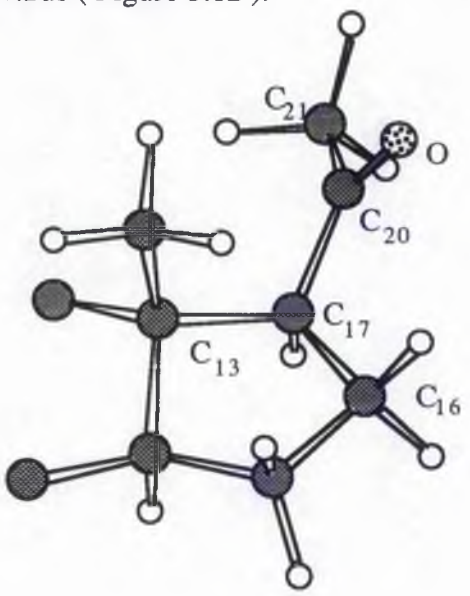


Figure 5.12

AM1 Optimised Structure of Preg-05 (D-Ring Only)

The two methods do not agree over exact barrier heights, with AM1 generally predicting a larger overall barrier, whilst PM3 gives a greater repulsion between C_{20}-keto and C_{18}-methyl groups.

It is interesting to note that X-ray structures for Preg-07 indicate a value for τ of 285.3°, whilst for pregnenolone τ = 299.1°, which is in fairly good agreement with the theoretical predictions. The predictions for Preg-12 are less favourable, with the x-ray value of τ being 180.6°. The geometry of this molecule is examined further in Chapter Seven. The x-ray structure for 20-hydroxy Pregn-4-ene-3-one, which is similar to Preg-03 except for the configuration of C_3, has a value of τ of 303.5°. Again this is reasonably similar to the theoretically predicted dihedrals of about 300°.

(D) Rationalisation of the Activities of the Steroids as Inhibitors of G6PDH.

(D1) Method.

The Electrostatic Potential Similarity Index (EPSI) calculations performed with ASP involves the selection of a lead compound, which is generally the most active in the series, as a basis for the comparison. The other drug molecules in the series are then compared to this, and the EPSI is evaluated.

As described in Chapter Two, ASP calculates two different EPSI s: Hodgkin's and Carbo's. It also allows the fitting of the EPSI s by a series of rotations and translations of the test molecule with respect to the lead in order to maximise the value of the Similarity Index. This is referred to as an Optimised Similarity Index, whereas if fitting is not performed, the result is an Unoptimised EPSI.

In general, optimised EPSI s have been calculated but the final results are not always suitable to use because the optimised position and orientation of the test molecule compared to the lead would be unlikely to be able to fit in the steroid binding site of the enzyme due to tight geometrical constraints. Figure 5.13 shows an example of the results of such an optimisation using ASP.

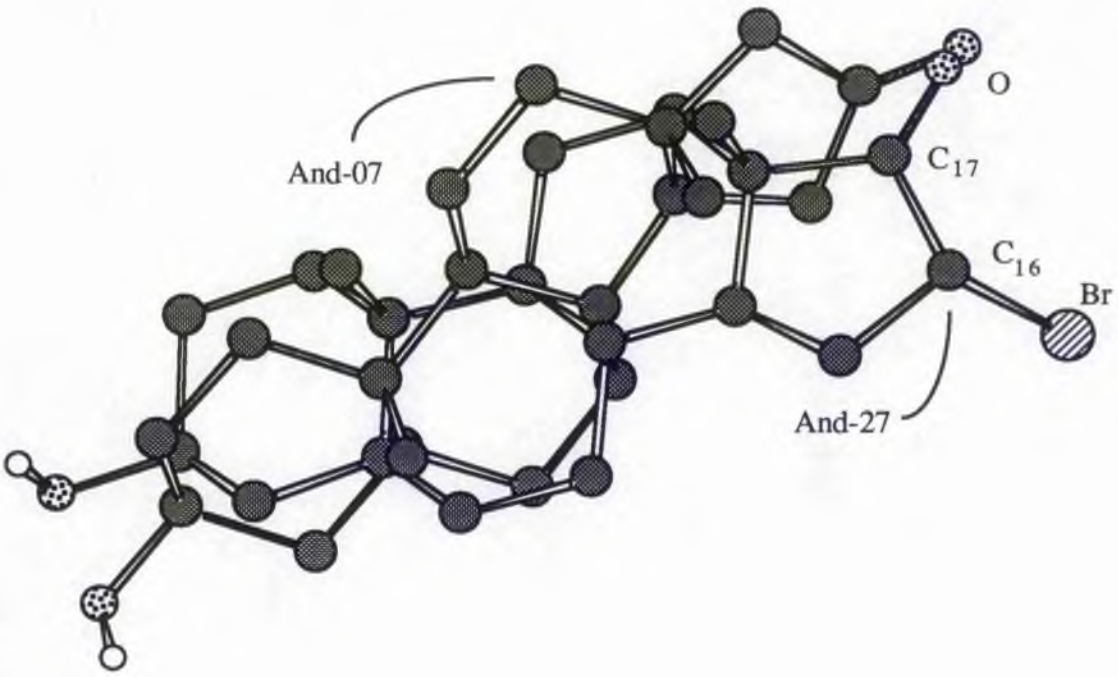


Figure 5.13

When such a result is obtained, it was found that the calculated EPSI was unsuitable, and hence the unoptimised value was used. This was obtained by performing a least squares fit of the steroid backbone to that of the lead molecule, and using the resultant geometry in the EPSI calculation. The Electrostatic Potential Maps (EPM s) and the ASP work were performed using both the AM1 and PM3 geometries and charges. In addition the ESPFIT program was used to calculate the AM1 potential fitted charges for the AM1 structure, and these were also used in the ASP and 3D2 work.

(D2) Choice of Lead Compound.

Raineri and Levy [106] quote two molecules as having the highest activity. These are 16α-bromoepiandosterone (And-27) and Androstan-17-one (And-02), the activities of which are 10% and 11% respectively, although the latter had to be measured at lower concentrations due to having a low solubility, which implies it is

possibly the more active. Nevertheless, And-27 has been chosen as the lead compound, both because the activity was measured at the same concentration as for the other inhibitors, and because it has been quoted frequently as being active against carcinogenesis [91,96]. Thus all the EPSI s are evaluated in comparison to this compound.

Once the EPSI s were obtained, they were plotted on a graph of Activity against Similarity Index for the molecules. The calculated dipoles for the molecules were also used in an attempt to predict activity, as these are less time consuming to evaluate than the EPSI, but are also dependent upon the electronic structure of the molecules.

(D3) Results : Dipole Moments.

Table 5.6 shows the calculated dipole moments for the Androstane series of molecules. These are quoted in Debye.

Molecule	AM1 Dipole	PM3 Dipole	Activity (%)
And-02	2.862	2.764	11
And-05	1.773	1.913	100
And-06	2.368	2.341	20
And-07	2.376	2.382	103
And-09	2.526	2.274	60
And-11	2.866	2.588	90
And-12	3.470	3.205	73
And-13	2.864	2.727	77
And-14	3.477	3.156	82
And-15	2.169	2.328	52
And-16	0.531	0.536	97
And-18	4.023	3.591	92
And-19	4.066	3.916	23
And-20	1.777	1.615	58
And-21	4.344	4.099	95
And-22	4.375	3.561	101
And-23	2.864	2.594	70
And-24	2.838	2.544	91
And-25	2.052	2.205	77
And-26	3.159	2.918	77
And-27	3.159	3.196	10
And-28	1.786	1.743	35

Table 5.6

(D4) Discussion.

These results have been plotted as a graph, with the inhibitory activity of the molecules. The resultant graphs are depicted in Figures 5.14 (AM1) and 5.15 (PM3).

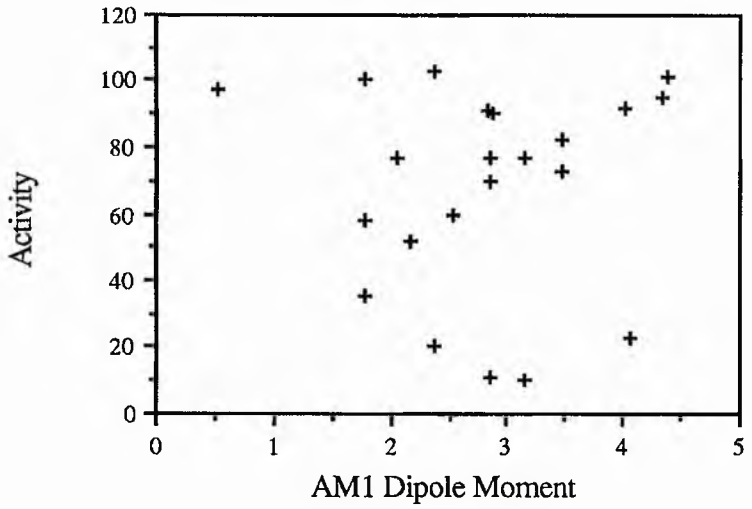


Figure 5.14.

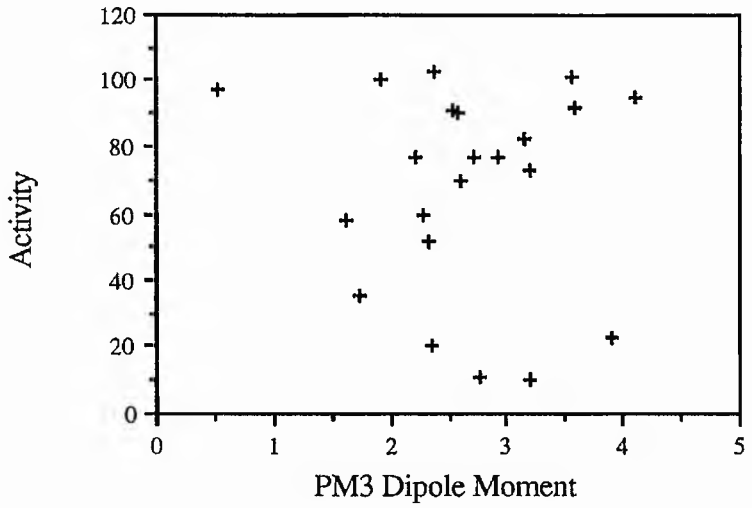


Figure 5.15.

At first sight, there seems to be little correlation between the calculated dipole and the activity, although closer examination reveals an approximately Z-shaped distribution in the graph. This is highlighted in the graph in Figure 5.16, below.

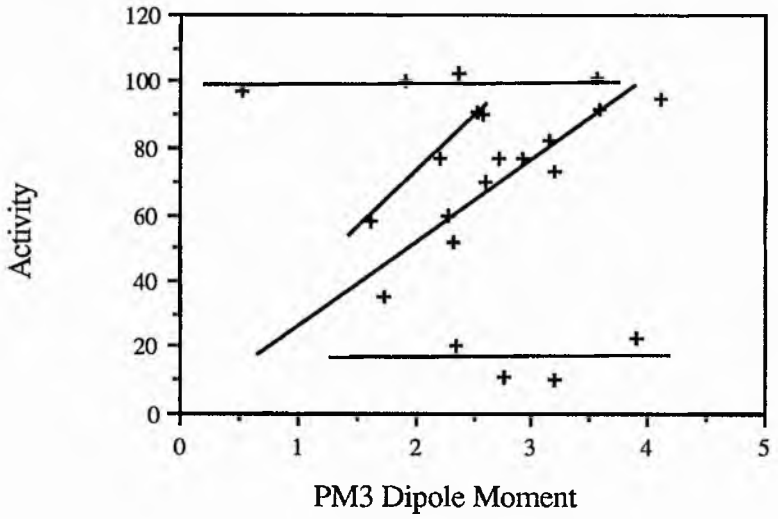


Figure 5.16

There seems to be little in common between the molecules upon the various arms of the graph, and there also seems to be a second, almost parallel, diagonal section to the distribution involving molecules 11, 24, 25, and 20. Nevertheless, the fact that some pattern has emerged implies that the distribution is not a random effect.

(D5) Conclusion.

Raineri and Levy [106] have attempted to explain the different activities of the steroids in terms of perturbation of the electron distribution of the basically hydrophobic Androstan-17-one molecule. Their discussion indicates that the addition of other functional groups to this molecule diminishes its activity, but this seems to fail to account for the high activity of Bromoepiandosterone (And-27).

The above graphs indicate that the dipole in the molecule is an important factor, although obviously not the only one, in determining the activities of the steroids. This implies that the electron distribution of the molecule is also more important than implied by the authors of the activity data.

(D6) Electrostatic Potential Maps.

Figure 5.33, at the end of this chapter, shows the electrostatic potential maps for the Androstane series of G6PDH inhibitors calculated using the AM1 structures and charges upon the single Van der Waal's surface using 3D2. Figures 5.34 and 5.35 are the PM3 and AM1-ESPFIT maps for the molecules And-15 (DHEA) and And-27 and And-05. The complete sets have not been included for reasons of space and because the following discussion is applicable to all three because of the similarities between the maps, although the magnitude of the potential reproduced by the EPSFIT charges is significantly larger than that of the other maps.

The maps on the left hand side are those drawn in the standard orientation for viewing steroids (front view) as they are depicted in Figure 5.2, with the A - ring to the left, D - ring to the right and the methyl groups pointing upwards. The maps on the right hand side are the view from behind, with the molecule rotated through 180° about the Y-axis (rear view), i.e. the α - face of the molecules.

The contours used in the maps are described in Table 5.7. The AM1 and PM3 calculations were performed using slightly different contour levels because the PM3 charge distribution tends to be less polarised than that for AM1. Thus PM3 potential maps drawn with the AM1 contours had few regions of interest. Maps similar to those calculated using the AM1 results can be generated by reducing the contour levels by about 10%. The ESPFIT maps are displayed using the same contours as for the AM1 maps.

Colour	AM1 Contour Level / AU	PM3 Contour Level / AU
Red	-0.0225	-0.0200
Buff	-0.0150	-0.0130
Yellow	-0.0075	-0.0066
Light Green	0.0000	0.0000
Dark Green	+0.0075	+0.0066
Light Blue	+0.0150	+0.0130
Dark Blue	+0.0225	+0.0200

Table 5.7

(D7) Discussion of the Androstane Series Electrostatic Potential Maps.

To describe the features of the steroids that are necessary to produce a good inhibitor of G6PDH, the molecules have been divided up into a number of active regions, which are depicted in Figure 5.17.

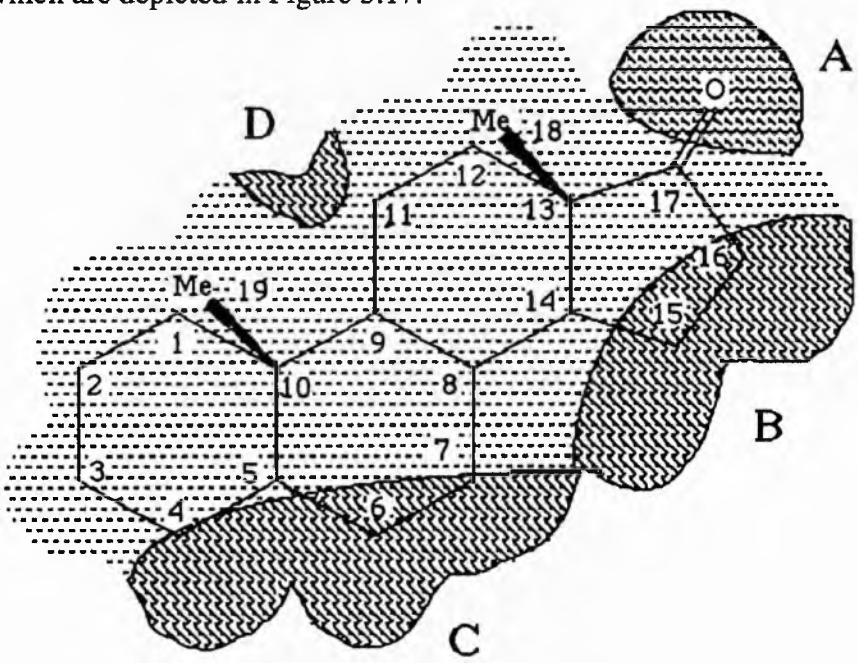


Figure 5.17

Regions Involved in G6PDH Inhibitory Activity.

The following rules can be drawn up from a careful study of the EPM s and the activities for the molecules.

1). A bright red (strongly positive) region is always required at Region A. Raineri and Levy state that this must in fact be a carbonyl group at position 17 if activity is to be observed, because both And-05 and And-16 which have a hydroxy and acetic acid group, respectively, are inactive.

2). Region B is light blue in the best inhibitors, although it may be dark blue. Both colours indicate a positive region of potential.

3). Region C seems to be a hydrophobic binding region, and hence is green in colour. Deviations from this reduce inhibitory activity.

4). A very small region of light blue at Region D (which may extend around the back of the molecule) is present on some of the best inhibitors. This again indicates a region of positive MEP.

(D8) Analysis of Individual EPM s using the Rules.

A visual comparison of the inhibitors of And-27 is presented for the AM1 maps.

And-02 5α-Androstan-17-one

This molecule is very similar to And-27, except at the 3 position. It also has a slight blue spot in the centre of the map. The activity should be similar to And-27.

Activity :11

And-05 Androstane-3β-17β-diol

Most of the molecule is green (near zero potential). Region B is green rather than light blue, as is region D. Region A has a red - blue dipole due to the hydroxy group. The activity should be low.

Activity : 100

And-06 3β-Hydoxyandrostan-17-one

This active molecule has the 17-keto red area, the light blue regions B and D indicate the standard profile of a good inhibitor, whilst region C is green as expected.

Activity : 20

And-07 3β-Hydroxyandrostan-16-one

The red of Region A has been moved towards where the blue of Region B (which has totally disappeared) is normally located. The map also has a small positive region in the centre. We might predict a moderate to low activity, whereas the actual activity is very low.

Activity : 103

And-09 5α-Androstan-3,17-dione

The centre of the map is more polarised with a small dark blue region (very positive potential). Region C is also distorted by light blue areas near the the 3- and 4-positions. Regions A and B are the same as for And-27. The predicted activity would be in the mid to high range, but the disruption of the hydrophobicity of Region C has a more marked effect than might first be thought as the activity is in the mid to low region.

Activity : 60

And-11 5α-Hydroxy-5α–androstan-17-one

This lacks the light blue of regions B and D, although regions A and C are as normal. It is interesting to note the differences in activity between this and And-06,

which has a 3β-hydroxy group. The effects at a quite distant site (Region B) are quite noticeable, and this may account for the large difference in activity between the two, which was explained by Raineri and Levy in terms of steric repulsions between the receptor sit and the α-OH group.

Activity : 90

And-12 17-Oxo-5α–androstan-3β-acetate

This molecule has radical charges at the 3 position in comparison to And-27. This has altered Region C slightly making it slightly yellow. Region B has no light blue colouring at all. Again, there is a small light blue region at the centre of the map.

The predicted activity would probably be moderate to low for this molecule.

Activity : 77

And-13 5α-Androst-1-ene-3, 17-dione

The addition of the Δ^1 double bond has markedly altered the hydrophobicity of Region C, but Regions A and B are unchanged. Region D is positive in potential, especially when viewed from behind. We might predict a moderate to high medium high activity, but again the disruption of the green Region C has a quite large effect, giving a moderately low activity.

Activity : 77

And-14 4-Androstene-3,17-dione

A large portion of the front face of this molecule is light or dark blue, modifying binding Region C. The Δ^4 double bond also has an effect detrimental to binding on this region. The predicted activity would be lower than And-13.

Activity : 82

And-15 3β-Hydroxy-5-androsten-17-one (DHEA)

This has a reduced blue area at Region B, and a red area in Region C, due to the Δ^5 double bond. Region D has a small positive area.

The predicted activity for this molecule may be lower than experimental, but the double bond produces a region that obviously has less effect upon Region C than might be expected.

Activity : 52

And-16 3β-Hydroxy-5-androsten-17β-carboxylic acid

This map is not markedly different from that for And-15, although Regions B and D show more blue areas. Thus we might predict a slightly reduced activity from the above. The low experimental activity highlights the importance of the 17-keto group in this series of molecules.

Activity : 97

And-18 4,9-Androstadiene-3,17-dione

Regions A and B are as expected for active molecules, but the large positive region on the underside of the molecule and the negative region due to the Δ^4 double bond totally remove the hydrophobic binding site C, whilst the Δ^9 double bond produces negative potential at Region D. Thus this molecule would be predicted to have a very low activity.

Activity : 92

And-19 5α-Androstane-11,17-dione

This molecule has normal Regions A, B and C, but Region D is totally disrupted by the 11-keto group. The high experimental activity of this could indicate that Region D is not involved in the binding of the molecule and may just be an artifact that just happens to be present in the better inhibitors. This molecule also seems to disprove Oertel and Benes theory [108] that ring position 3 is involved in the binding

since there is no functional group at this position. Despite the red area due to the 11-keto group, there is a nearby region of positive potential at the centre of the molecule. This is on the hydrogen atom at C_7. Thus, if this region is involved in the binding, it implies that it binds to a negatively charged region of the enzyme that has some degree of conformational flexibility. Such a group would be able to interact with the region of positive potential whether it is situated closer to the C_{11} or the C_7 hydrogen atoms, hence facilitating the binding of molecules such as And-27 and And-19.

Activity : 23

And-20 11β-Hydroxy-5α-androstan-17-one

Region B has no region of markedly positive potential, and Region C has more yellow potential. Again the electron distribution at Region D is hard to explain, although there is a region of blue nearby due to the hydroxy hydrogen.The predicted activity might be moderate to low.

Activity : 58

And-21 11β-Hydroxy-4-androsten-3,17-dione

Although this map shows the necessary potential at Regions A and B to make a good inhibitor, the extended light blue and red at position C indicate a reduced activity. The oxygen at the 11β-hydroxy group causes a red area at Region D, although there is also a dark blue section due to the hydrogen of this group. The predicted activity for this molecule could be medium - low.

Activity : 91

And-22 11α-Hydroxy-4-androstan-3,17-dione

The map for this molecule indicates a more polar charge distribution than for the 11β isomer, And-21, with more red and blue regions. Again, Regions A and B have the required colouring for active molecules but the C Region does not have the

appropriate hydrophobic (green) colour. The disturbance of this region is greater than for And-21, and hence the predicted activity would be lower.

Activity : 101

And-23 3α-Hydroxy-4α-methyl-5α-androstan-17-one

This molecule has the required 17-keto group, but the B - Region has no blue colouring, as is the case for the other 3α-hydroxy molecule, And-11, to which this map has a strong resemblance. The activity of this molecule is greater than one would predict from comparison to the And-11 map, probably due to the effects of the 4α-methyl group extending the hydrophobic C - Region.

Activity : 70

And-24 3α-Hydroxy-7α-methyl-5α-androstan-17-one

This map is similar to that for And-23 and And-11, both being 3α-hydroxy steroids, and in a similar manner, has no blue at Region B. Predicted activity would be close to that for And-11, since the methyl group is not as close to Region C as for And-24.

Activity : 91

And-25 3β-Hydroxy-7α-methyl-5-androsten-17-one

This map looks similar to that for And-15, with the characteristic A and B Regions of an active molecule. A predicted activity would be similar to that for And-15, but would be an overestimate. It seems likely that the lower activity for this molecule is due to steric repulsions due to the methyl group being in a conformation dissimilar to that attained by the saturated And-24 case.

Activity : 77

And-26 2α-Bromo-5α-androstane-3,17-dione

The 2α-Bromo group disrupts the hydrophobicity of Region C to a much greater extent than does the unsubstituted 3,17-dione compound, And-09. Thus this molecule would have a lower predicted activity than the latter.

Activity : 77

And-27 3β-Hydroxy-16α-bromo-5α-androstan-17-one

This is the lead molecule in this study, with the greatest activity. It has the standard 17-keto group, leading to a red A - Region, and a light blue B - Region. The latter region is likely to be only on the upper (or β) face of the steroid, because the 16α-Bromo group does not interfere with this. The C - Region is green, and there is a very small touch of blue at Region D.

Activity : 10

And-28 3β-Chloro-5α-androstan-17-one

This map is typical of other good inhibitors, with all four regions showing the normal colours of active molecules. The fact that this molecule has a lower activity than And-27 could possibly be due to the chlorine atom having a slightly greater effect than a hydroxy group upon Region C

Activity : 35

(D9) Conclusions.

The first of the three regions depicted in Figure 5.17 definitely seem to be involved in the binding. Region D does seem to be light blue in many of the most potent inhibitors, indicating a region of positive potential. Despite this, the location of a keto - group at position 11, causes a large negative region of potential at this location. Thus, as this molecule (And-19) is a very active inhibitor of G6PDH, it seems that region D may not be involved.

Examination of the EPM for And-19 does reveal a region of positive potential near Region D, situated on the hydrogen atom attached to C_7. It may be possible that conformational flexibility in the proposed binding region of G6PDH complementary to Region D, allows the two regions of opposite potential to interact positively. Hence this region may still be involved in the interactions between the steroids and the enzyme.

From this study, we can deduce the presence of three binding sites within the steroid binding pocket, at least one of which (Region B) could not be deduced from a study of the functional groups of the inhibitors alone. The presence of a fourth binding site, with a less rigorous demand upon the location of the complimentary region of potential can also possibly be inferred. This is juxtaposed with Region D of Figure 5.17

(E) ASP Study of the Androstane Series.

As previously stated, it is not sufficient to use solely either optimised or unoptimised Electrostatic Potential Similarity Indices (EPSI s), but a mixture of the two, depending upon the quality of the overlap between the two molecules being compared.

The final similarity index was selected using the quality of the fit as a guide, and graphs of Activity against both Carbo and Hodgkin EPSI were plotted in an attempt to find a correlation between the two.

(E1) Results and Discussion.

The calculated similarity indices are presented in Tables 5.8 and 5.9, in which the column headings have the following meanings :

C Carbo EPSI, no optimisation.

CO Carbo EPSI, optimised.

H Hodgkin EPSI, no optimisation.

HO Hodgkin EPSI, optimised.

Fit Indicates the quality of the fitting after optimisation, and is divided into four categories :

A Good.

B Reasonably Good.

C Poor.

D Very Poor.

Molecule	C	CO	H	HO	Fit	Activity
And-02	0.820	0.832	0.813	0.819	A	11*
And-05	0.739	0.784	0.672	0.710	A	100
And-06	0.938	0.944	0.937	0.941	A	20
And-07	0.442	0.829	0.442	0.829	D	103
And-09	0.415	0.754	0.409	0.720	A	60
And-11	0.764	0.783	0.764	0.785	C	90
And-12	0.514	0.533	0.512	0.527	A	73
And-13	0.705	0.730	0.689	0.707	A	77
And-14	0.648	0.703	0.626	0.527	A	73
And-15	0.859	0.896	0.856	0.906	A	52
And-16	0.538	0.638	0.535	0.631	A	97
And-18	0.607	0.651	0.582	0.630	A	92
And-19	0.475	0.709	0.474	0.708	D	23
And-20	0.745	0.779	0.732	0.771	A	58
And-21	0.571	0.662	0.553	0.649	A	95
And-22	0.629	0.645	0.596	0.609	A	101
And-23	0.736	0.816	0.735	0.814	A	70
And-24	0.731	0.797	0.731	0.799	C	91
And-25	0.802	0.873	0.800	0.881	C	77
And-26	0.704	0.741	0.696	0.745	A	77
And-27	1.000	1.000	1.000	1.000	-	10
And-28	0.845	0.860	0.845	0.864	A	35

Table 5.8

ASP results using AM1 Structures and Charges.

* Indicates Activity measured at different concentration of Steroid.

Molecule	C	H	CO	HO	Fit	Activity
And-02	0.787	0.771	0.810	0.790	A	11*
And-05	0.777	0.674	0.781	0.680	B	100
And-06	0.894	0.888	0.901	0.905	A	20
And-07	0.444	0.443	0.793	0.795	B	103
And-09	0.637	0.633	0.768	0.774	A	60
And-11	0.765	0.761	0.779	0.770	C	90
And-12	0.440	0.440	0.586	0.588	D	73
And-13	0.724	0.714	0.749	0.744	A	77
And-14	0.322	0.317	0.714	0.695	A	82
And-15	0.853	0.845	0.878	0.864	A	52
And-16	0.289	0.287	0.656	0.646	A	97
And-18	0.621	0.611	0.691	0.674	A	92
And-19	0.444	0.444	0.651	0.651	D	23
And-20	0.699	0.680	0.780	0.742	A	58
And-21	0.550	0.540	0.625	0.631	A	95
And-22	0.709	0.687	0.735	0.717	C	101
And-23	0.795	0.790	0.802	0.797	A	70
And-24	0.724	0.720	0.790	0.783	C	91
And-25	0.828	0.821	0.879	0.864	C	77
And-26	0.703	0.695	0.773	0.763	A	77
And-27	1.000	1.000	1.000	1.000	-	10
And-28	0.828	0.815	0.852	0.840	A	35

Table 5.9

ASP results using PM3 Structures and Charges.

The graphs of Activity against ESPI are shown in Figures 5.18 to 5.21. The regression lines refer to the first set of points, denoted by plus signs (+), and include the bulk of the data points. The second set of points lie away from this line, and are denoted by crosses (x). This accounts for four molecules, which have been divided into two groups. The first group consists of the molecules And-02, And-12 and And-19. All of these are below the line, indicating a higher activity than predicted using the EPSI. They also have radical changes in the functional group at the 3 - position of the steroid. This is reflected by a low Similarity Index because of the differences between the molecules and the lead compound. These structural differences are highlighted in Figure 5.2. The fact that changes in this position produce such a marked alteration in the EPSI, without dramatically altering the activity implies that this position of the steroid backbone is not involved in binding. This refutes the conclusion of Oertel and Benes [108], who site this as an important binding site, despite the earlier work of Marks and Banks [105], who concluded that this position is not involved in the binding process. The Raineri and Levy data were also published prior to the Oertel and Benes work, but the latter seem to have ignored the high activity of molecules such as And-19 in making their conclusions.

The second disparate group consists of And-07, which has a keto - group at C_{16}, rather than at C_{17}. The optimised EPSI for this molecule would be quite a distance above the lines on the graphs, indicating a lower activity than predicted for this molecule. A predicted activity for this molecule, based upon its optimised AM1 Hodgkin Index would be in the region of 60 %, but the unoptimised EPSI for this molecule (using the same methods) is 0.442, which is equally far from the line. Thus neither the optimised nor the unoptimised Similarity Indices for this molecule give a good indication of its activity, and an EPSI of about 0.6 would be necessary to make this point close to the least - squares fitted line.

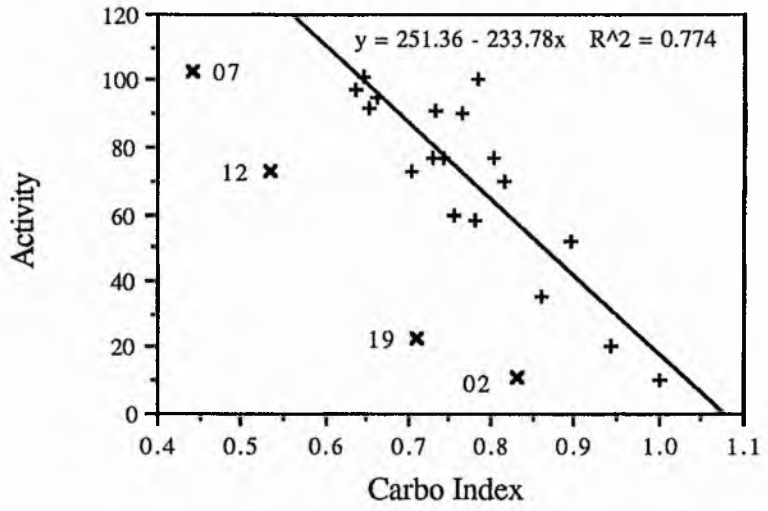


Figure 5.18.

Results for AM1 Structures and Charges using the Carbo EPSI.

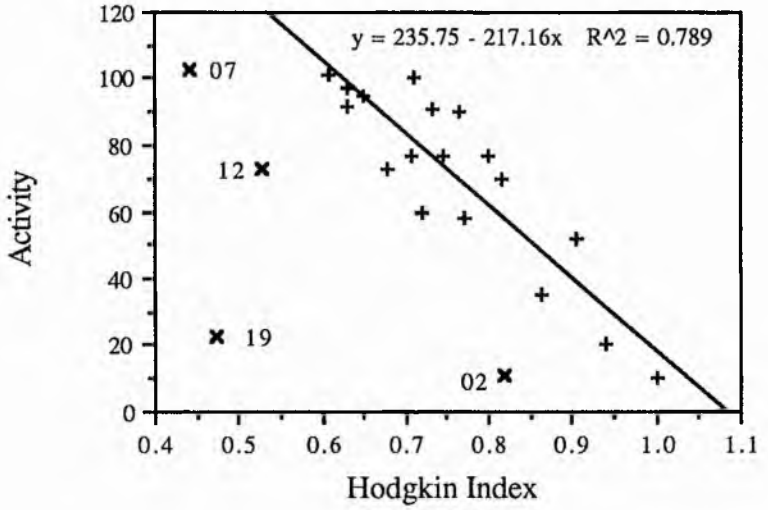


Figure 5.19.

Results for AM1 Structures and Charges using the Hodgkin EPSI.

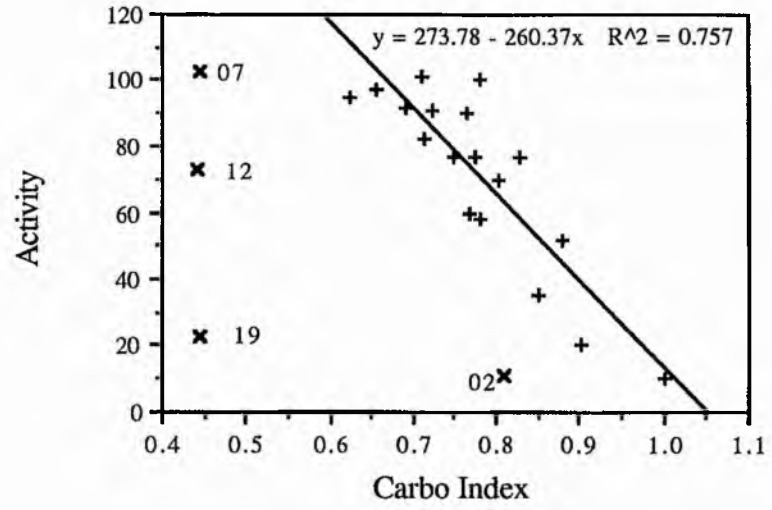


Figure 5.20.

Results for PM3 Structures and Charges using the Carbo EPSI.

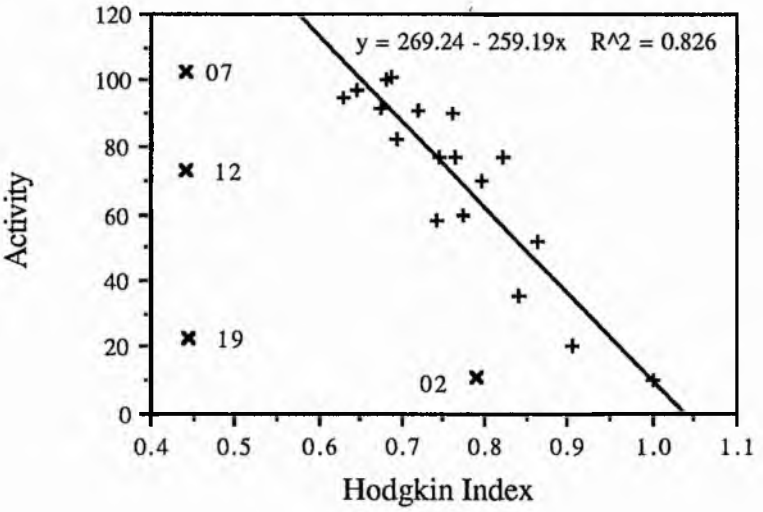


Figure 5.21.

Results for PM3 Structures and Charges using the Hodgkin EPSI.

(E2) Conclusions.

The Similarity Index method has shown that the MEP has an important role to play in assessing the activity of the G6PDH inhibitors, and that the visual analysis of

the potential maps has a definite basis behind it. The method fails to account for the higher than predicted activities of the molecules with radically altered groups at position 3, or the value of the activity for And-07, which has an EPSI too low in its unoptimised form, but is too high after optimisation. Despite the first of these two failures, the low EPSI s for And-02, And-12 and And-19 serve to indicate that the suggestion of Oertel and Benes [108], stating that the 3 - position is important in the binding of the steroid to the enzyme, is probably false.

(F) Other Similarity Work on the Androstane Series.

(F1) Effects of van der Waals' Scaling Factor on the Similarity.

One of the parameters in the ASP calculation that can be varied is the van der Waals' Scaling Factor. This is normally set at 1.0, but can be increased. This factor determines at what multiple of the van der Waals' surface grid points are excluded from the calculation of the EPSI. Thus of a value larger than 1.0 is selected, then points close to the single van der Waals' surface are not used. Since the evaluation of the MEP using the point charge approximation used in ASP (See Chapter Two) becomes more accurate at points further from the atomic nuclei, one might expect the results to give a better correlation at, say, the double van der Waals' surface, which is achieved by setting the scaling factor to two. Such a study has been performed using the Androstane series of inhibitors of G6PDH.

(F2) Results.

The results of this study, after adjusting for poor fitting after optimisation, are displayed in Figures 5.22 and 5.23.

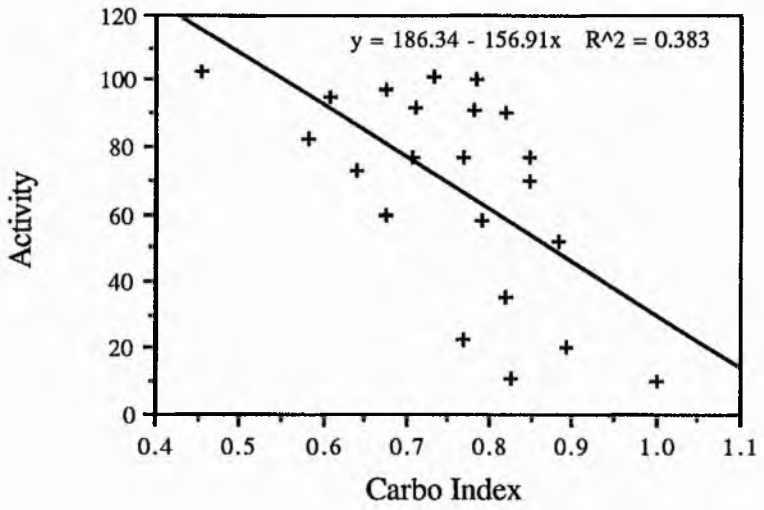


Figure 5.22.

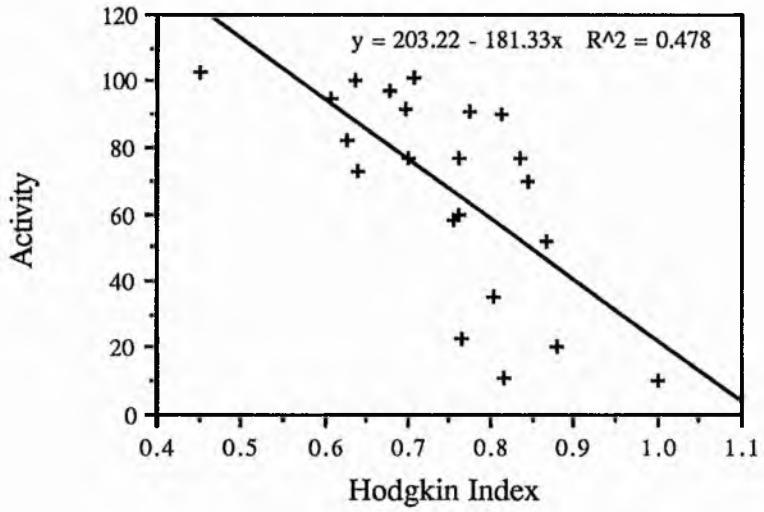


Figure 5.23.

(F3) Discussion.

These graphs show a poor correlation between activity and Electrostatic Potential Similarity Index, with correlation co-efficients of 0.383 and 0.478 for the Carbo and Hodgkin indices, respectively.

One interesting feature of these graphs concerns the three molecules that deviated below the line to a large degree in the preliminary ASP study using a single van der Waals' scaling factor, namely And-02, And-12 and And-19. These are not only closer to the line, but possibly form a separate line of their own, although they still have higher than predicted activities. A possible reason for this could be that the differences in the EPSI due to modifying the functional groups at position 3 are less noticeable when points close to these groups are not considered. Thus the effects of a greater distance between the point at which the potential is calculated and the nuclei in question reduces the effect of the substitution upon the EPSI.

An explanation for the poor performance of the EPSI at predicting activity may be found from the study of the concluding remarks of Raineri and Levy [106]. They state that on binding, the whole steroid could be located inside a tight fitting pocket within the enzyme. Thus the electrostatic interactions are only of interest at a short range, and the variation in the potential out with the double van der Waals' surface are less important than the short range interactions, despite the fact that the latter may be less well modelled because of the breakdown of the Point Charge Approximation at short distances from the nuclei.

(F4) ESPFIT Charges.

The AM1 structures and ESPFIT potential fitted charges have also been use as input to the ASP program. This was done because these charges should give a more accurate picture of the MEP, and hence yield results that correlate to a higher degree with the activity.

(F5) Results and Discussion.

The results for these calculations are presented in the graphs of Activity versus Hodgkin EPSI in Figures 5.24 and 5.25, the latter being the result of the calculation using a double van der Waals' scaling factor in the ASP work. Both Similarity Indices fail to indicate any correlation between the two, although only the Hodgkin results are shown here for reasons of space.

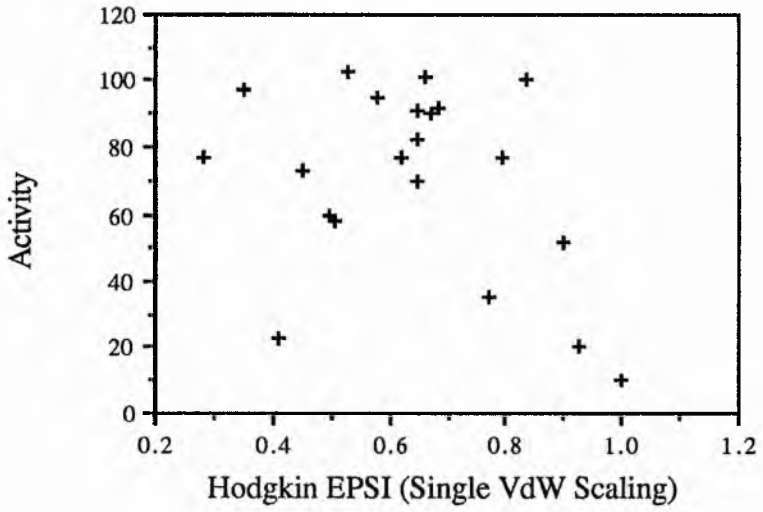


Figure 5.24.

Graph of Activity against EPSI using ESPFIT Charges.

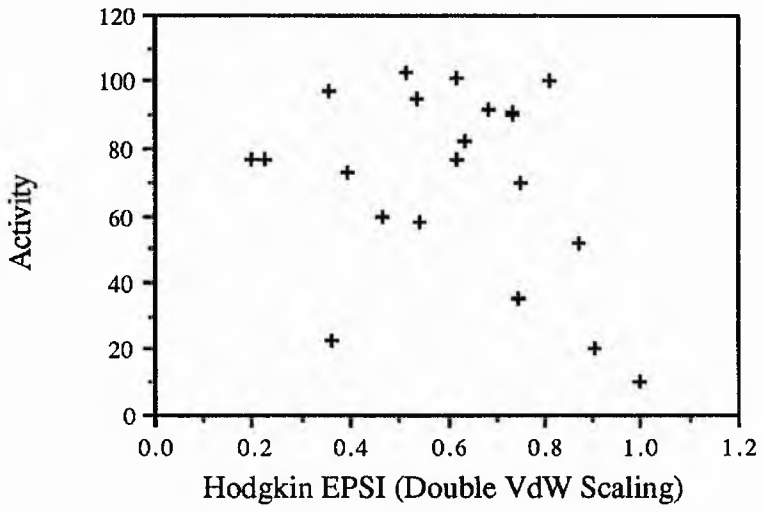


Figure 5.25.

Graph of Activity against EPSI using ESPFIT Charges.

(F6) Conclusions.

Despite the fact that the use of ESPFIT charges in ASP should yield a more accurate representation of the MEP, this method has failed to find any correlation between activity and the EPSI. No explanation can be found for this fact, except perhaps that the activity must depend upon factors other than the MEP, which are represented by the Mulliken, but not the ESPFIT charges. There may also be an effect due to the fact that the charges are fitted to reflect the MEP upon the double Van der Waals' surface, but are being used in a study involving points within this surface, although the use of a scaling factor of 2.0 in ASP also failed to show any correlation.

(F7) Using other Molecules as Lead Compounds in the ASP Work.

One drawback of similarity studies is that a lead molecule has to be selected, and this has an EPSI of the maximum value of 1.0, meaning total similarity. Thus, for any other molecule studied, the EPSI has to be less or equal to that of the lead. Hence, it may be possible to miss potential new drugs because that have a lower EPSI.

It was considered that it may be possible to spot a trend in the Activity against EPSI graphs for molecules more active than the lead by selecting other molecules from the Androstane series as a lead. Two studies have been performed, with two different lead molecules. The first of these is And-15 (DHEA itself), which was selected because it has an activity of approximately 50 %. The second molecule is And-26, which was chosen because it is roughly half way through the list of inhibitors, when sorted by activity.

(F8) Results.

The graph below (Figure 5.26) indicates the results of the study using And-15 as the lead compound. The similarity index used was that of Hodgkin for the AM1 geometries and charges. As usual, the optimised values were used, unless the atomic fit after optimisation was poor.

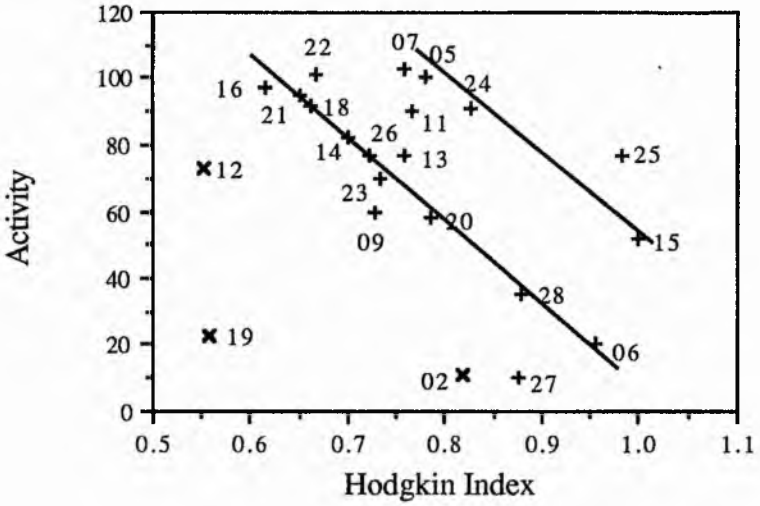


Figure 5.26.

ASP Results using DHEA as the Lead. AM1 Structures and Charges.

When the inhibitors were compared to And-26 as the lead, the graph in Figure 5.27 was obtained, again using the AM1 structures and charges and the Hodgkin Similarity Index.

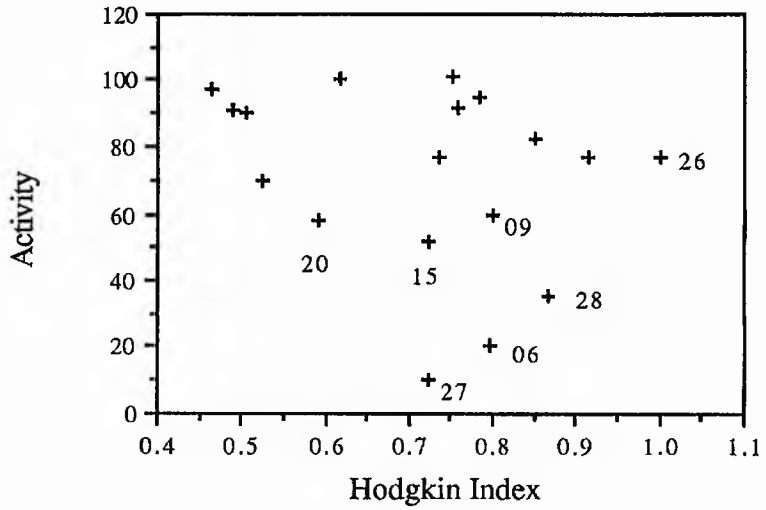


Figure 5.27.

ASP Results using And-26 as the Lead. AM1 Structures and Charges.

(F9) Discussion.

In the And-26 graph, in which some of the more active compounds have been labelled, there is no way to distinguish between the active and inactive molecule using the Similarity Index, with poor correlation between the results.

For And-15, the graph has a more interesting form, with three different groups of points. The first consists of most of the molecules forming a straight line with a good correlation. The second group consists of six molecules that are above this line, and possibly form a line of their own. This groups includes DHEA itself. The last group consists of the three molecules with the radically modified 3 - position, and And-27, the previous lead molecule. All of these molecules have corresponding points upon the graphs which are below the primary line.

The points in the second group represent the following molecules : 3β-hydroxy androstan-16-one (7), androstane-3β,17β-diol (5), 3α-hydroxy-7α-methyl androstan-

17-one (24), 3α–hydroxy androstan-17-one (11), 3β-hydroxy-7α-methyl androst-5-ene-17-one (25) and DHEA i.e. 3β-hydroxy androst-5-ene-17-one (15). Apart from the penultimate molecule, they have little in common with And-15 itself, suggesting a possible fortuitous alignment of points on the graph, although the group does include both of the 3α-hydroxy molecules, both molecules without a 17 - keto group and both 7α-methyl substituted groups.

(F10) Conclusions.

Using either of these two molecules as a lead in the similarity calculation fails to provide a method of clearly identifying molecules with a greater activity than the lead, and hence molecules that are more active than And-27 may not be spotted for use as a more powerful inhibitor.

There is some interesting structure in the graph using And-15 as a lead, but there is no obvious correlation in the And-26 case.

(G) Final Conclusions on the Androstane Electrostatic Potential Studies.

This study has analysed the electrostatic potential maps for inhibitors of G6PDH in the Androstane series, and highlighted three regions that are involved in the bonding and a fourth region of interest that may be located close to a region of negative potential on the enzyme that has some degree of conformational flexibility. From this work, assumptions can be made about associated areas within the enzyme that have a complementary region of electrostatic potential. This analysis has also provided a possible explaination for the large differences in activity often caused by small changes in the inhibitor, such as the 3α-OH / 3β-OH case of And-06 and And-11.

Work with the similarity index has provided a quantitative guide to assessing the activity of the molecules with a reasonable level of accuracy. The highest correlation between activity and the EPSI was found with the Hodgkin Index, using PM3 structures and charges.

The Similarity Index studies have also indicated the failure of either the AM1 ESPFIT charges or the use of a double van der Waals' scaling factor to provide viable results in the ASP calculation. The the quality of the ESPFIT charges such as those that have been utilised in some of these calculations is also analysed in Chapter Seven, in which various types of charge are used to predict ^{13}C chemical shifts. The ESPFIT charges give a lower correlation with the shifts than do those calculated using population analysis. When viewed in conjunction with this study the quality of such charges may be questionable.

(H) Electrostatic Potential Work on the Pregnane Series.

The AM1 Electrostatic Potential Maps (EPM s) for the molecules studied from this series are displayed in Figure 5.36, which can be found at the end of the chapter. These maps correspond to the minimum energy conformation for the C_{17} side chain of these molecules, and indicate that the C_{20} oxygen present in most of them is in a similar position to the C_{17} - keto group in the Androstane series. This is not the same conformation as used by Raineri and Levy [106] in their analysis of these molecules, which is based upon the C_{21} - methyl group pointing forwards and the C_{20} - keto group pointing backwards. Although AM1 and PM3 predict the former to be the overall minimum energy conformation, many molecules exhibit a considerable flexibility of the dihedral $\tau = C_{21}\text{-}C_{20}\text{-}C_{17}\text{-}C_{13}$. In many cases, it has been demonstrated that τ may vary between 100° and 300° with very little change in energy.

(H1) Analysis of the EPM s with respect to the Rules for the Androstane Series.

A study of the maps in Figure 5.36 shows immediately that the four regions denoted A to D in Figure 5.17 have a less clear cut role in determining the activity of the Pregnane series.

Raineri and Levy state that a 20 - keto group is necessary for activity, again leading to a red (or negative) area of potential at Region A. Region B has no light blue colouring in the active inhibitors, such as Preg-02, Preg-10 and Preg-13, although some blue in this regions is visible in the moderately active Preg-09. Preg-12, on the other hand, exhibits a strongly positive potential at Region B, and is totally inactive, although this is likely to be due to the effects of the 16α,17α-epoxy group.

The hydrophobic region C is green for all molecules except for the inactive Preg-07, for which this ares is largely light blue, and Preg-05, which is also inactive, and has some blue areas near C_4. Preg-06, which is slightly active, is also quite yellow around this position.

A very small light blue area around Region D is visible in some of the molecules, both active and inactive. For example, Preg-02, Preg-10 and Preg-13 exhibit this feature, whilst so do the inactive Preg-05, Preg-07 and Preg-12. It should be noted that this feature often extends some way onto the α face of the molecules, as can be seen on the view from behind.

One further region of interest, not present upon the maps for the Androstane series is the potential of the C_{21}-methyl group. This is illustrated in Figure 5.28, which is a schematic representation of the steroid from the rear (i.e. α) face.

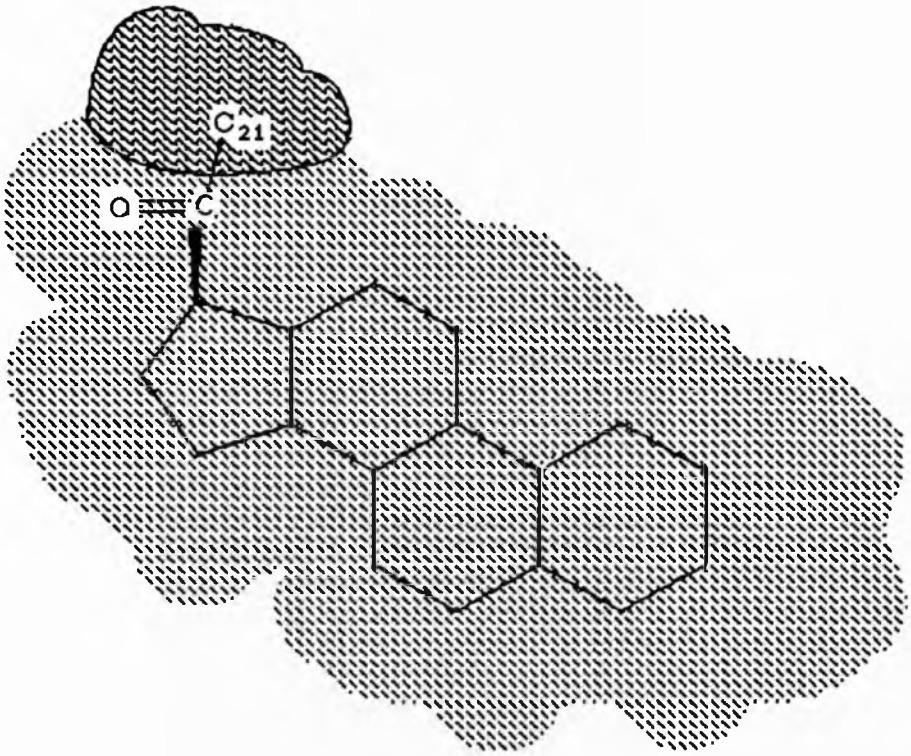


Figure 5.28.

A study of the maps for the Pregnane series indicates a difference in the level of positive potential in the indicated region, between active and inactive molecules.

The active inhibitors, such as Preg-02, Preg-08, Preg-10 and Preg-13 exhibit a potential that is less positive that that of inactive molecules such as Preg-03, Preg-05, Preg-07 and Preg-14. These have more regions of dark blue at this location, compared to the light blue or green of the active molecules.

An obvious exception to this rule is Preg-12, which has the 16α,17α - epoxy group, which produces a potential in the area of interest that is colour - coded green and light blue. A further exception is Preg-09, which is slightly active, but has one of the largest regions of dark blue of all the molecules. This effect is probably due to the proximity of the 11 - keto group causing electron withdrawal from the C_{21} - methyl group.

(H2) Conclusions.

All the active molecules have a 20 - keto group, giving a red area equivalent to Region A in Figure 5.17. Region B does not exhibit the characteristic blue area of the active molecules from the Androstane series. This could be due to the closeness of the carbonyl oxygen to the Region. For example, the O-$H_{16\beta}$ distance is 2.838Å in DHEA, but only 2.381Å in Preg-02. The proximity of the oxygen could lead to any small positive regions of potential being dwarfed by the negative oxygen.

Region C must still be hydrophobic if the steroid is to make a good inhibitor, and molecules in which this pattern is disturbed, such as Preg-05 and Preg-07, are inactive. A very tiny light blue area at Region D is present in a number of active and inactive molecules.

A further region of interest has also been postulated, which can be observed when the EPM is viewed from behind. It is located on the C_{21} - methyl group, and features an area of less positive potential in the most active molecules.

Thus the rules for Regions A and C still hold for the Pregnane series of molecules, whilst the case for Region B may be dwarfed by other factors. As is the case for the Androstane series, the involvement of Region D is inconclusive.

(I) ASP work for the Pregnane Series.

The Similarity Indices for the Pregnane series of molecules were calculated using the same method as used previously, with both And-27 and Preg-02 as lead compounds. The activity for the latter cannot be plotted on the graphs because it was measured at a different concentration, although it is still a very potent inhibitor of G6PDH.

(I1) Results and Discussion.

The results using both Carbo and Hodgkin Indices are depicted in Figure 5.29 using And-27 as the lead compound. The Preg-02 lead study results are shown in Figure 5.30.

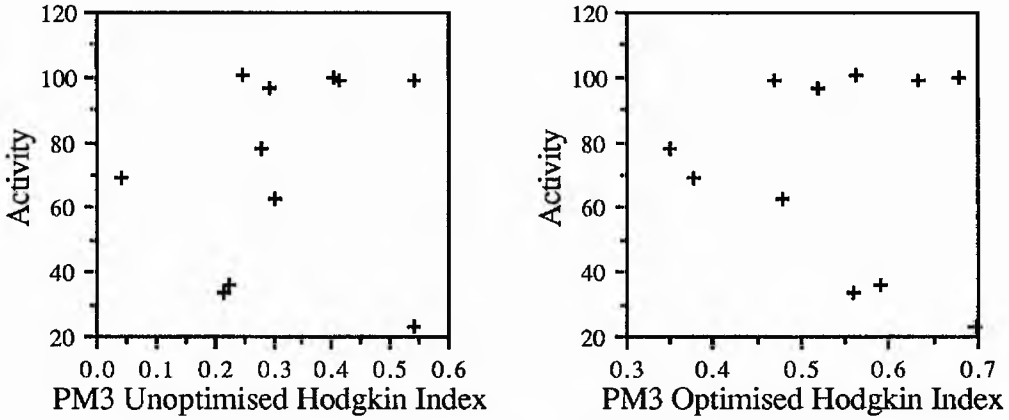


Figure 5.29.

Optimised EPSI s for the Pregnane series ASP work with And-27 Lead.

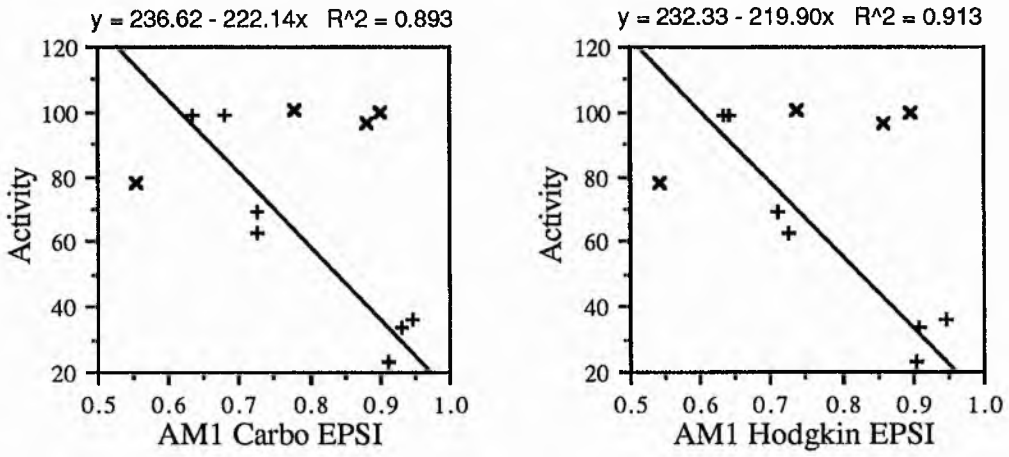


Figure 5.30.

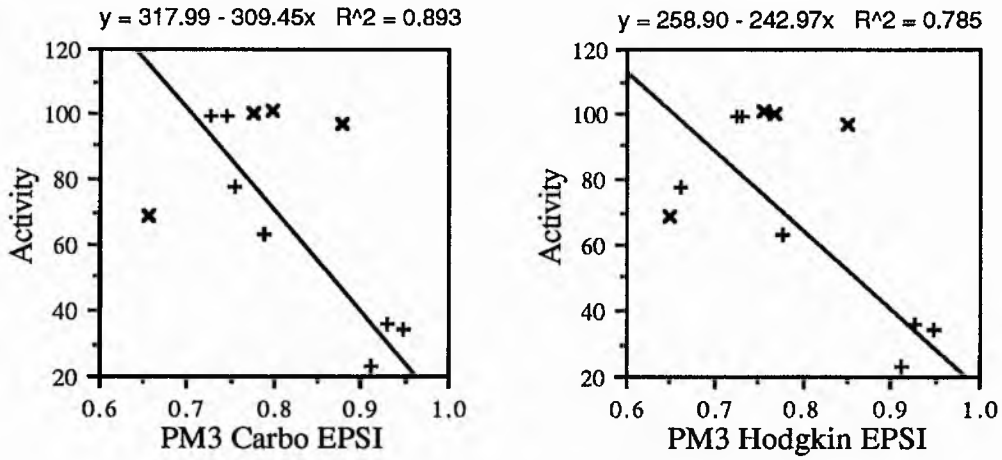


Figure 5.30. (Cont.)

Pregnane series ASP work with Preg-02 Lead.

The graphs for the And-27 comparison represent the unoptimised and optimised Hodgkin index results. In the unoptimised case, there seems to be no correlation with activity, whilst examination of the optimised graph reveals the presence of two lines, although the quality of the optimised fit for these molecules to the orientation of And-27 was very poor in almost every case. One of these lines is a diagonal, similar to that obtained for the Androstane series, whilst the second is a horizontal line containing the least active molecules. Some of these have an EPSI which is as high as that of the most active molecules.

Thus, these graphs indicate that the EPSI alone is incapable of predicting the biological activity of the Pregnane series when And-27 is used as the lead compound.

When Preg-02 is used as the lead, the graphs shown in Figure 5.30 are obtained. These are for both Similarity Indices, and are adjusted so that the optimised index is only used for molecules in which a good final fit to the lead compound is obtained.

rest, except for Preg-06, making a diagonal line. The points on the graphs have again been divided into two groups. Those marked with plus signs are on the diagonal line, to which the regression information on the graphs refers. The second set of points are those for Preg-05, Preg-06, Preg-07 and Preg-14, which are marked with crosses, and are excluded from the calculation of the least squares line. In this latter group, Preg-06 has a lower than predicted activity and like some of the molecule in the Androstane series has dramatic changes in the functional group at the 3 - position. The other points in this selection correspond to three of the molecules which exhibit dark blue colouring at the C_{21} - methyl group.

This analysis leads to the conclusion that this region also plays an important part in the binding process, but the Similarity Index is not sensitive enough to highlight what seems only to be a small change in the potential. This is because the EPSI is calculated over the complete molecule and hence the changes in the MEP over a small region are dwarfed by the similarity of the molecules as a whole.

(I2) Conclusions.

The electrostatic potential work on the Pregnane series of G6PDH inhibitors is a less useful guide to the biological activity of the molecules than it is for the Androstane series. Although the correlation for seven of the points is good, the Similarity Index is, in general, a very poor predictor of activity. If it is used in conjunction with the knowledge that a dark blue colouring around C_{21} on the EPM s leads to a low activity in most cases (Preg-09 being an exception), then a reasonable good indication of the potency of the molecules as inhibitors can be obtained.

The gradients and intercepts of the correlation lines, especially for the AM1 results, agree reasonably well with those from the Androstane study, which implies a common explaination for the activities of the molecules and the fact that this is not a

fortuitous alignment of the data points. It may also be concluded that although AM1 gave lower correlation co-efficients in the Androstane study, it is superior to PM3 in assessing the activities of the wider range of compounds, in that the predictions for the Androstane series of molecules are more suitable for use with the Pregnane series.

(J) Comment on the Accuracy of the Experimental Data.

One point of note in these studies is that some of the poorer predictions may actually be due to inaccuracies in the experimental data. This fact may be deduced from the comparison of sets of activity values from different authors for the same molecules. Graphs are presented in Figures 5.31 and 5.32 below, of a comparison of the calculated Inhibitor Constants (K_i) from the Raineri and Levy work and those of Oertel and Benes [108], and also of the activity, as measured by Marks and Banks [105].

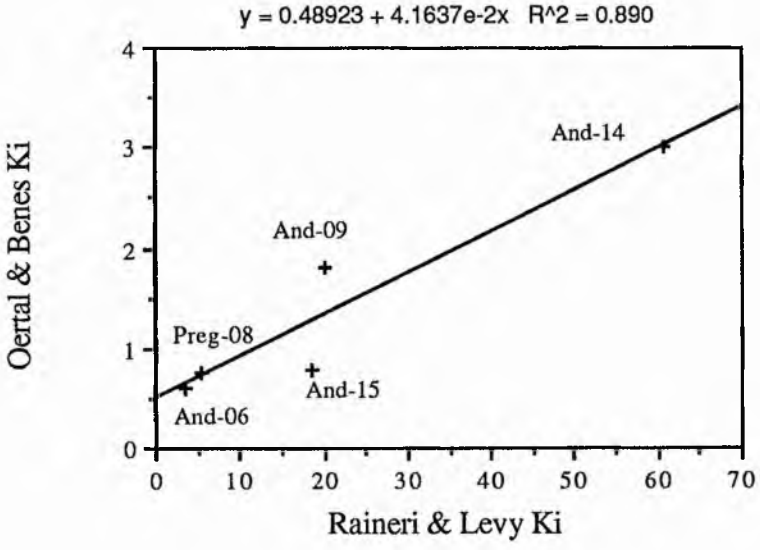


Figure 5.31

Graph Comparing Raineri and Levy Activity Data with that of Oertel and Benes.

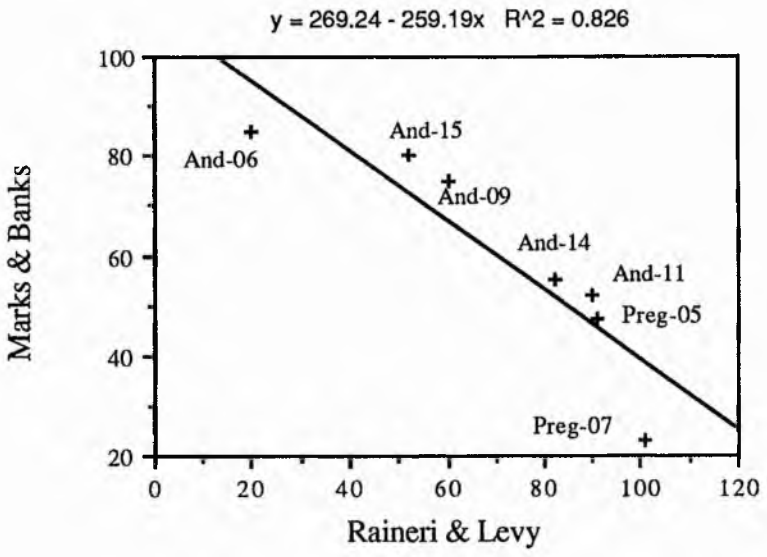


Figure 5.32

Graph Comparing Raineri and Levy Activity Data with that of Marks and Banks.

Both correlation co-efficients (denoted by R^2 in the graphs) are quite high, but there are some deviations from the lines, indicating possible experimental error in the data sets. Thus it can be concluded that some of the poorer theoretical predictions may in fact be due to inadequate accuracy in the experimental data.

(K) Summary of the Work on Steroidal Inhibitors of G6PDH.

The results in this Chapter have provided an insight into the importance of the MEP in the steroid - enzyme binding process, as well as a moderately accurate guide to qualitatively assessing the activities of many of the molecules studied using purely theoretical techniques.

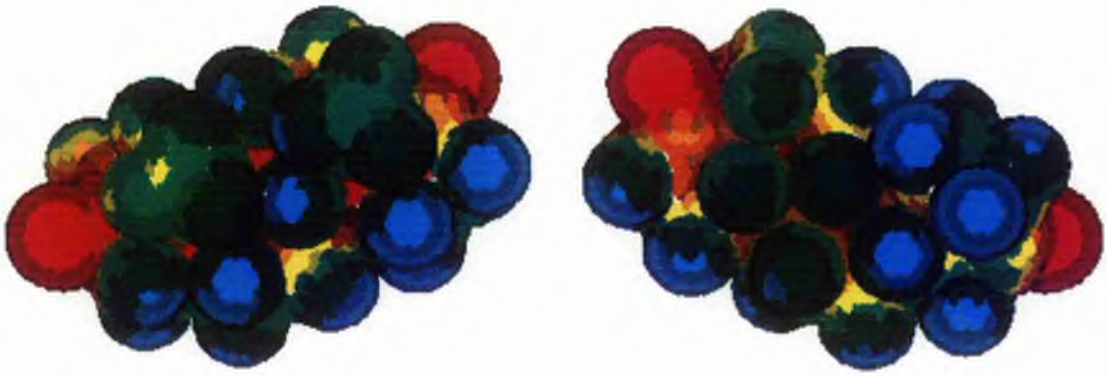
And-02

And-05

And-06

Figure 5.33

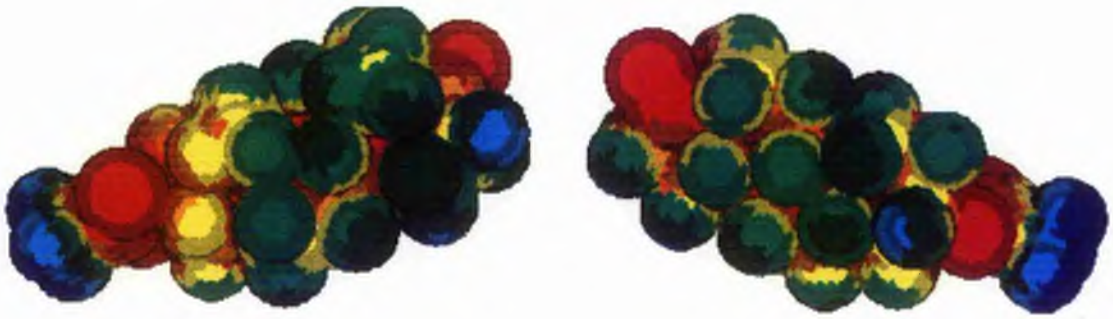
And-07



And-09

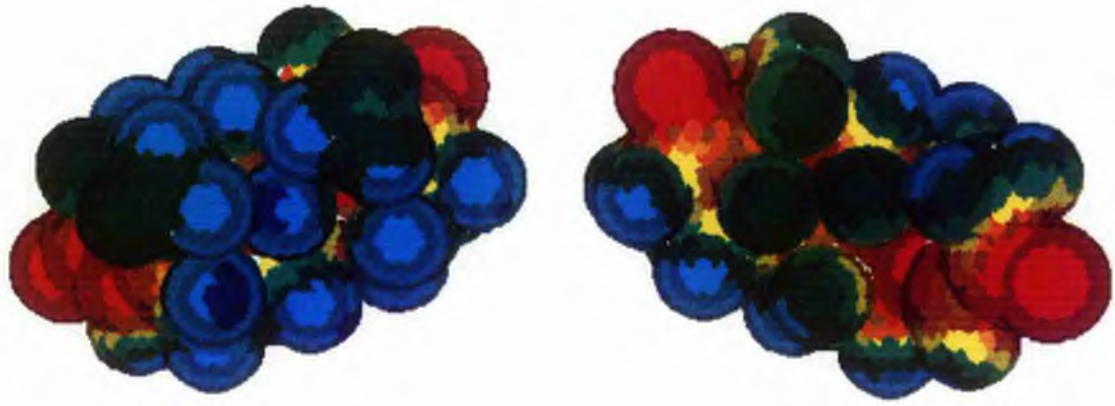
And-11

Figure 5.33 (Cont.)



And-12

And-13



And-14

Figure 5.33 (Cont.)

And-15

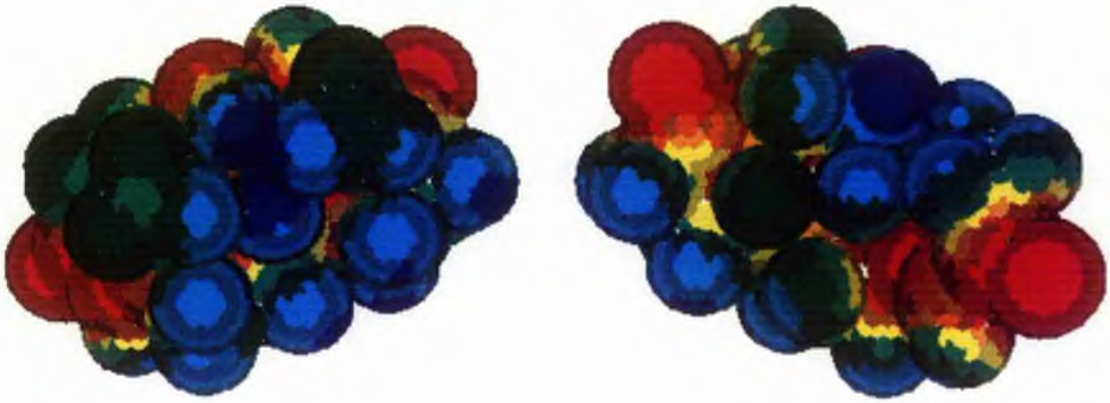
And-16

And-18

Figure 5.33 (Cont.)

And-19

And-20



And-21

Figure 5.33 (Cont.)

And-22

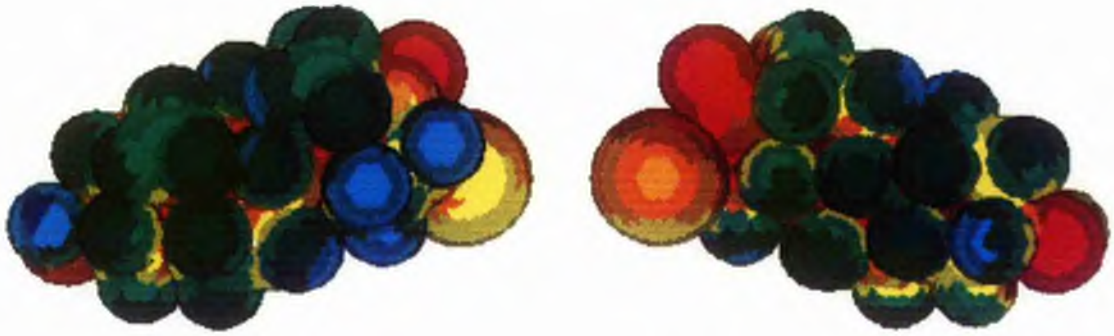
And-23

And-24

Figure 5.33 (Cont.)

And-25

And-26



And-27

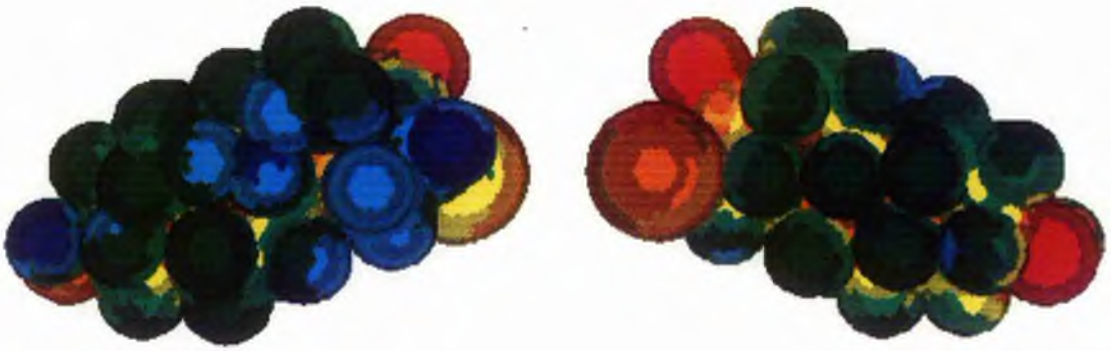
Figure 5.33 (Cont.)

And-28

Figure 5.33 (Cont.)

And-05

And-15



And-27

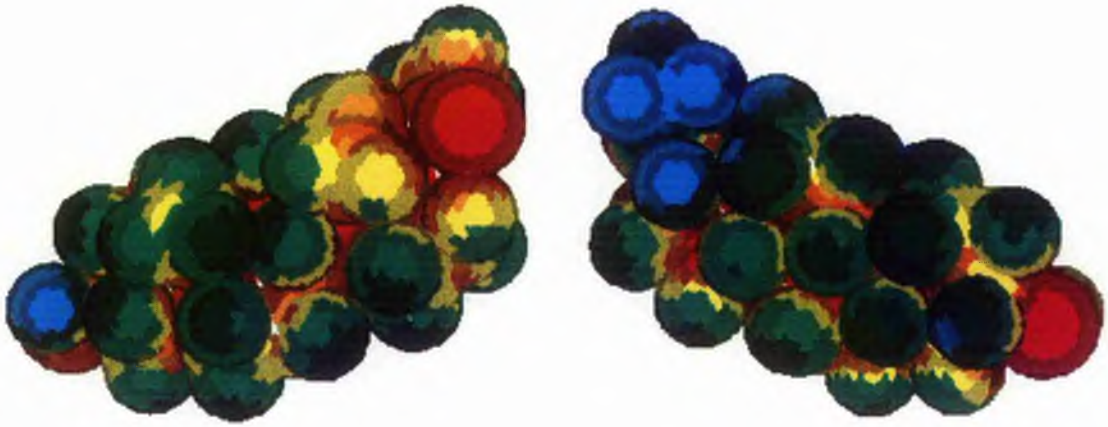
Figure 5.34

And-05

And-15

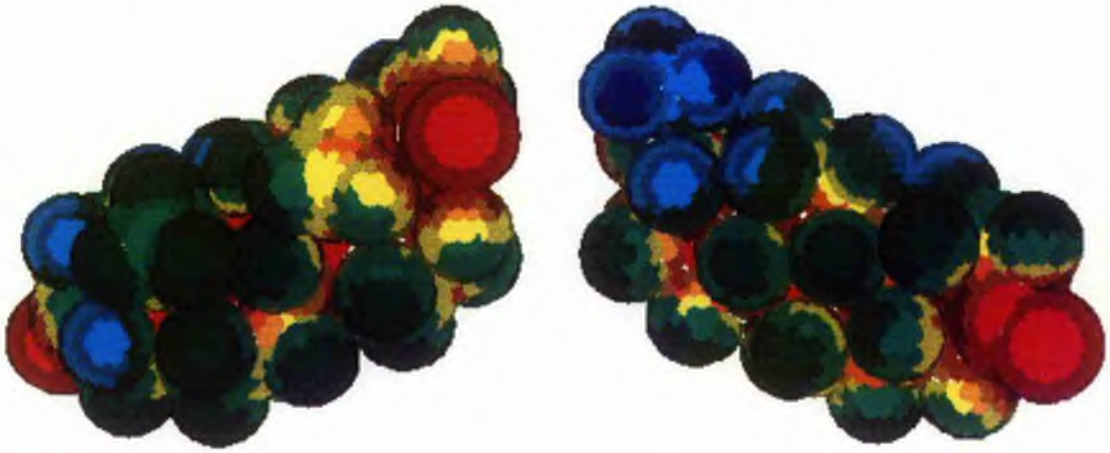
And-27

Figure 5.35



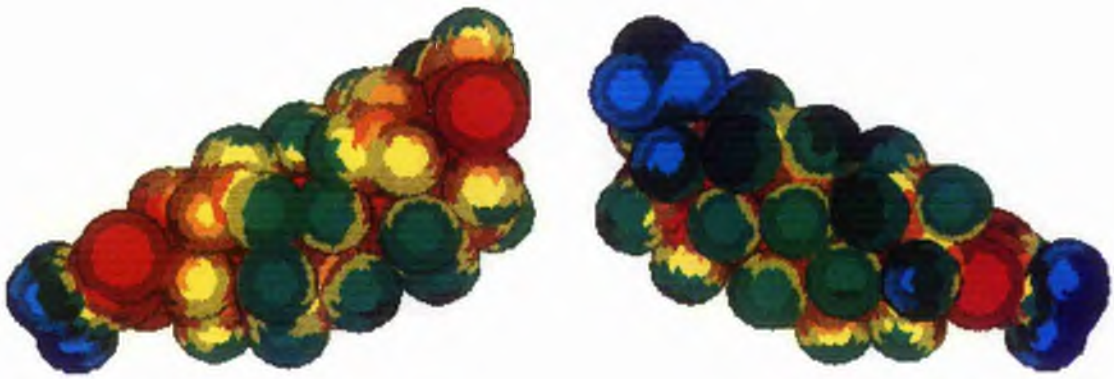
Preg-02

Preg-03



Preg-05

Figure 5.36

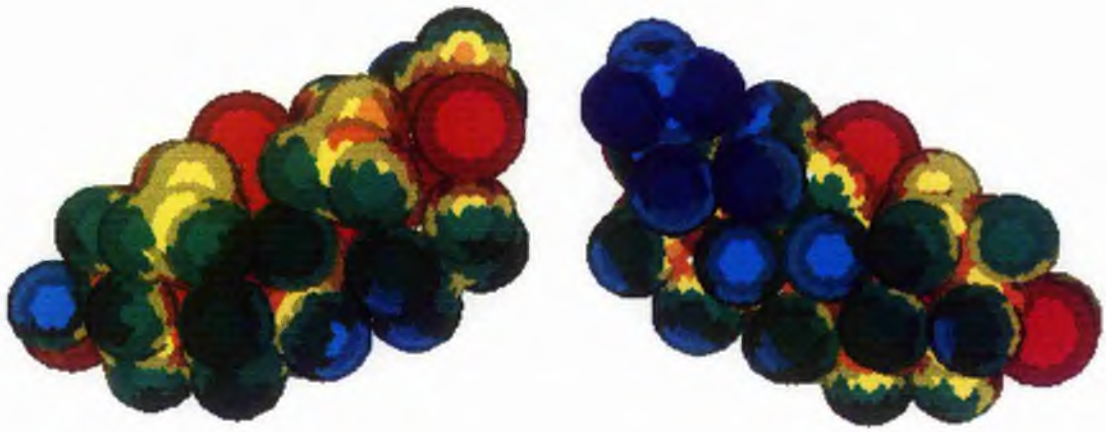


Preg-06

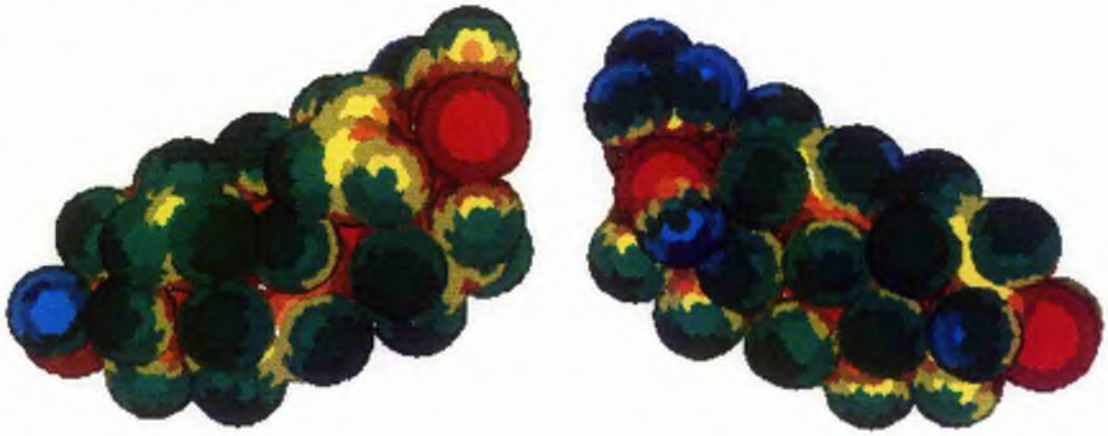
Preg-07

Preg-08

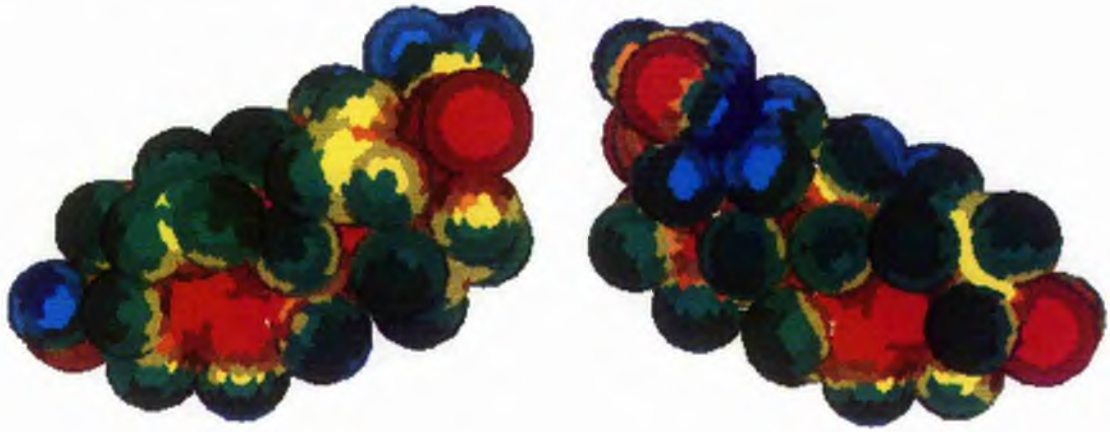
Figure 5.36 (Cont.)



Preg-09



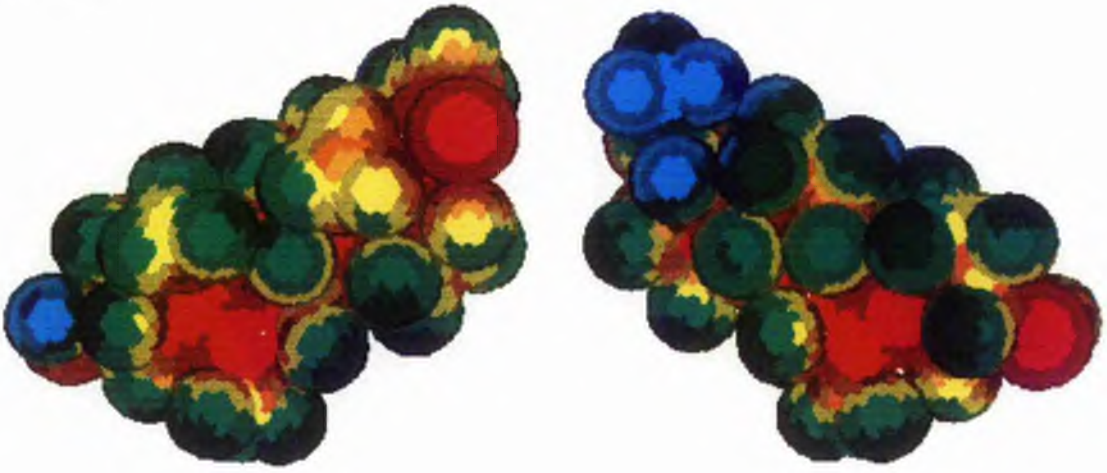
Preg-10



Preg-11

Figure 5.36 (Cont.)

Preg-12



Preg-13

Preg-14

Figure 5.36 (Cont.)

Chapter Six

Studies of Inhibitors of Oestrogen Biosynthesis.

(A) Introduction.

Breast cancer has the highest mortality rate in women in the western world [5], with a death rate of nearly 80 per 100,000 population. It is interesting to note that the rate is much lower in certain parts of the world, where the incidence may be as low as one quarter of this value. The occurrence of the disease rises steadily with age, and is associated with a number of risk factors. These include age at first birth, body weight and height, with decreased values of all these factors leading to a protecting effect.

About one third of breast cancers can be termed Oestrogen Dependent, and require the presence of Oestrogens to survive. There are also examples of excess oestrogen leading to higher rates of cancer, for example, with extended use of certain types of oral contraceptive, or after treatment with an artificial oestrogen, diethyl stilbesterol. [5]

(A1) Treatment of Breast Cancer.

Breast cancer may be treated either by surgical removal of the ovaries, which are the main source of oestrogens, or by interfering with oestrogen action with drugs such as the important anti - oestrogen, Tamoxifen (Figure 6.1). Finally treatment may be carried out with compounds that block the biosynthesis of oestrogens.

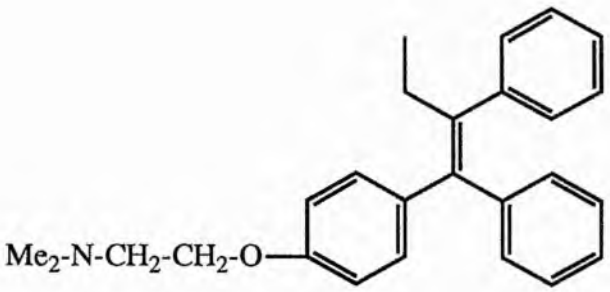
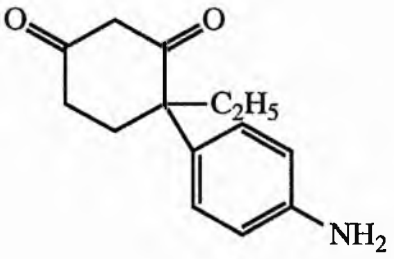


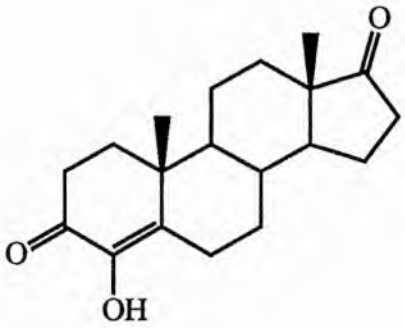
Figure 6.1.

Tamoxifen.

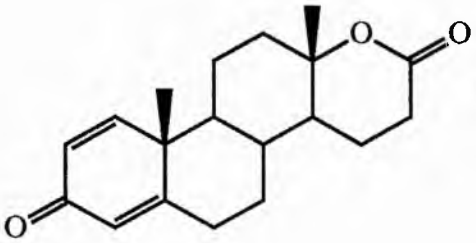
Examples of this are inhibitors of the enzyme Aromatase, and include compounds such as Aminoglutethimide, 4-hydroxy Androstenedione, Testolactone and CGS 16949A [110], which are shown in Figure 6.2.



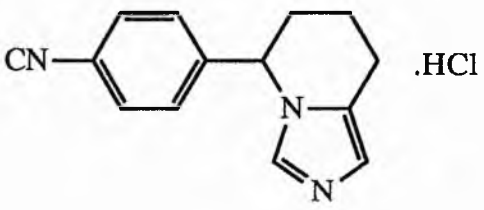
Aminoglutethimide



4-HydroxyAndrostenedione



Testolactone



CGS 16949A

Figure 6.2

The last of these compounds has an extremely high potency for the inhibition of Aromatase. It has little effect upon the synthesis of other steroids and few other side effects and is hence a very interesting drug.

Although surgical treatment has been very successful, it has been found that some tumours are capable of synthesising oestrogens themselves [111], and hence inhibition of the synthesis would be preferable.

(A2) Action of Aromatase.

Aromatase is a cytochrome P-450 enzyme, and is the last, and rate limiting step in the synthesis of oestrogens such as oestradiol and oestrone, as shown in Figure 6.3.

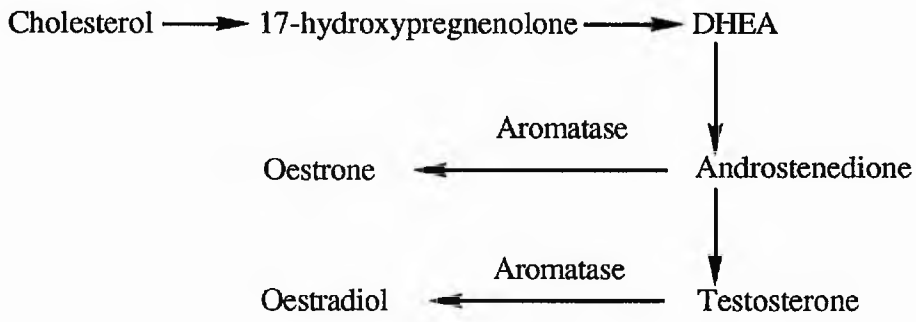
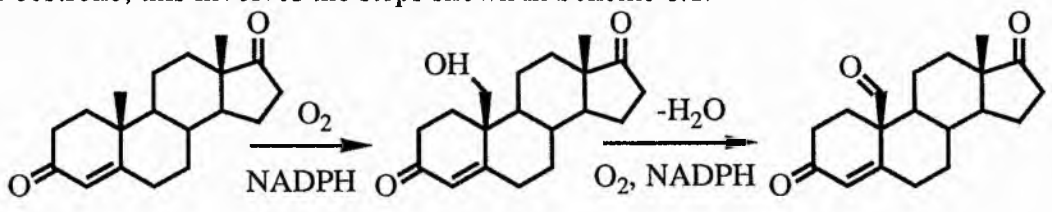


Figure 6.3

Pathway of Oestradiol Synthesis

The mechanisms of the Aromatase reaction has been under study for many years, and is known to involve three successive steps, each requiring one mole each of NADPH and O_2. [9] The mechanism of the first two of these steps has been elucidated using labelling techniques. In the case of the aromatisation of Androstenedione, the precursor of oestrone, this involves the steps shown in Scheme 6.1.

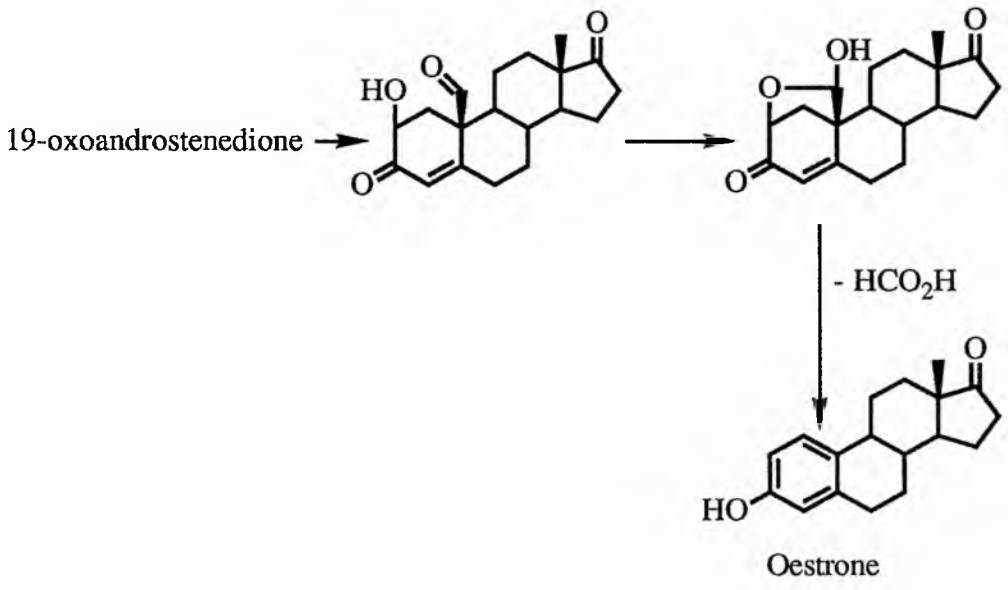


Androstenedione 19-oxoandrostenedione

Scheme 6.1

This process proceeds via two hydroxylations of C_{19} and the removal of water.

Initially, the final step was thought to proceed via a 2β-hydroxylation and deformylation via a cyclic hemiacetal [112,113], as in Scheme 6.2.



Scheme 6.2

Caspi et al called the intermediacy of the 2β-hydroxy compound into question [114], and more recently, a radical mechanism and a peroxo - iron (III) intermediate have been invoked to explain the reaction [115,116,117]. This involves a peroxo linkage between the iron - porphyrin complex of the enzyme, and C_{19} on the steroid (Figure 6.4).

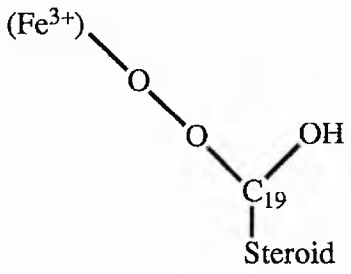


Figure 6.4

(A3) Inhibitors of Aromatase.

There has been great interest in the inhibition of this enzyme[1], and the various inhibitors can be divided into two kinds, depending upon their mode of action. Competitive inhibitors, as their name suggests, compete with the natural substrates (i.e. Testosterone and Androstenedione) for occupancy of the active site without undergoing any biological reaction. Suicide, or mechanism - based inhibitors are activated by the enzyme and bind to it permanently, thus rendering it inactive. There has been much interest in the latter group of inhibitors, because competitive inhibitors can often be metabolised quickly, whereas in the case of suicide inhibitors, more of the enzyme must be produced before activity can resume. Thus, less frequent treatment is required to maintain the blockage of oestrogen synthesis.

There have been many examples in the literature of both types of inhibitor, and their use to treat breast cancer [118-128]. Many of these compounds of interest are thiol containing androgens, substituted at C_{19}, C_4, and C_7.

(B) Aromatase Inhibitors Studied in this Chapter.

The molecules of interest in this chapter are those studied be Greway and Levy [129], which are a series of 3-oxo-17β-carboxamido steroids, of the type shown in Figure 6.5. The molecules are described in Table 6.1, below.

---

[1]For a review see ref 9.

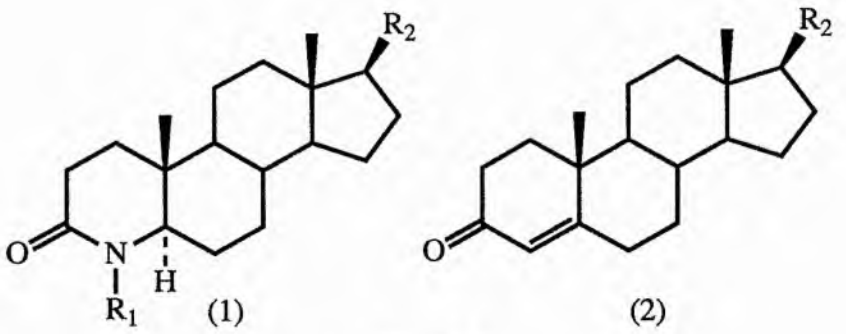


Figure 6.5

Steroid	R_1	R_2	Other Functionality	K_i	$t_{1/2}$
1a	Me	$C(O)N(CH(CH_3)_2)_2$	-	3.3±0.6	15
1b	Me	$C(O)N(CH(CH_3)_2)_2$	10β-H (19 Nor)	9.0±0.8	28
1c	Me	$C(O)N(CH(CH_3)_2)_2$	10β-CH_2CH_3	4.8±0.5	14
1d	Me	$C(O)N(CH(CH_3)_2)_2$	2β, 10β-Ethano Bridge	3.8±0.4	40
1e	Me	$C(O)N(CH_2CH_3)_2$	-	6.2±0.4	20
1f	H	$C(O)N(H)C(CH_3)_3$	$\Delta^{1,2}$	2.5±0.3	19
1g	Me	=O	-	5.8±0.3	NI
1h	Me	=O	2β, 10β-Ethano Bridge	0.8±0.1	NI
1i	Me	$CH(CH_3)CH_2OH$	-	NI	NI
2a	-	$C(O)N(CH(CH_3)_2)_2$	5α-H	7.6±0.9	NI
2b	-	$C(O)N(CH(CH_3)_2)_2$	-	1.1±0.1	36
2c	-	=O	5α-H	0.03±0.002	NI
2d	-	=O	2β,10β-Ethano Bridge	0.02±0.002	NI

Table 6.1

$t_{1/2}$ in minutes K_i in μMol.

NI: No Inhibition, or No Time Dependent Inactivation

Some of these compounds exhibit time dependent inhibition of Aromatase, meaning that if the enzyme is incubated with the steroid in the presence of an NADPH generating system, then suicide inhibition, which is dependent upon the length of

incubation, occurs. Most of the molecules also act as competitive inhibitors. The full list of compounds, shown in Table 6.1, also shows the K_i values (in μMol) and half life for the inactivation, $t_{1/2}$ (in minutes).

(B1) Reasons for the Study.

It was decided to study these molecules in particular because both competitive and inactivation data were available and it would be of interest to try to rationalise the differences between the active and inactive molecules in each case. A second reason for the interest in these molecules is because of the amide linkage shown in Figure 6.6, which occurs both in the 17β - side chains and in the A-rings of the compounds in Group 1.

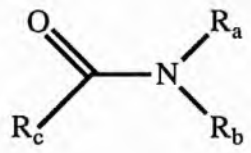


Figure 6.6

In the A-ring, R_c and R_b represent the rest of the ring, C=O is carbon 3 and the nitrogen is at position 4. In the side chain, R_c is the main body of the steroid, joined at position 17, whilst R_a and R_b are isopropyl or some other alkyl groups.

The group is shown in Figure 6.6 because it is similar to the peptide linkage in proteins, in which R_b is a hydrogen atom. The poor ability of the methods parameterised within MOPAC to reproduce exoerimental properties of this group is a known failing [55]. The results predict that the nitrogen atom exhibits a pyramidal structure despite experimental evidence as to the planarity of the group. This is because a delocalised system occurs involving the oxygen, carbon and nitrogen atoms. The

nitrogen lone pair becomes delocalised into the CO π - system [130], and hence the nitrogen becomes sp^2 hybridised and hence is planar (Figure 6.7).

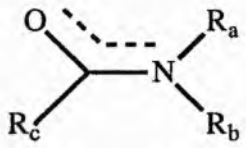


Figure 6.7

MOPAC automatically searches the molecule for peptide linkages and if any are detected, it requires the user to select one of the keywords *MMOK* or *NOMM*. The former adds a molecular mechanics correction to the calculation, which forces the barrier of rotation for N-Methyl Acetamide to be 14.0 kcal mol^{-1} [55], although Lim and Francl quote a value of 20 kcal mol^{-1} to be typical [130]. This is a wholly artificial correction to attempt to make the calculated results agree with the experimental ones. The *NOMM* keyword ignores the peptide linkage, and continues with the calculation as normal.

A successful search for the peptide linkage requires the presence of the hydrogen atom corresponding to R_b, and hence the "peptide - like" group in Figure 6.6 will not be found in the steroids in question. Thus part of this chapter is concerned with the comparison of MOPAC structures with those calculated with Ab Initio methods, which should not be subject to the same failings.

(C) Semi Empirical Calculations.

MOPAC calculations were run at the precise level of geometry optimisation using a modified version of the program dimensioned to be large enough to cope with these molecules, which contain about 70 atoms, since the default version is too small.

In addition, the conformation of the heterocyclic A-ring was examined for steroid 1g. Optimisations were performed starting with the methyl group at N_4 either in an axial or an equatorial position.

Finally, a conformational search was performed upon the carboxamido side chain of molecule 1a by twisting the bond with C_{17}, and evaluating the heat of formation of the various rotamers at 20° intervals.

It was considered that the methyl groups making up the four isopropyl sections of the side chain would be able to freely rotate, and hence no study of the conformation of this region of the molecule has been made. This conclusion was also valid for the ethyl group at C_{10} in molecule 1c.

(D) Structure of 17β-Side Chain.

(D1) N, N-Dimethyl Acetamide.

Ab Initio work has been performed to examine the structure of the group shown in Figure 6.6, involving optimisations using various basis sets and model systems.

The first molecule studied was N,N-dimethyl Acetamide (DMA), Figure 6.8, illustrating the four conformations initially studied.

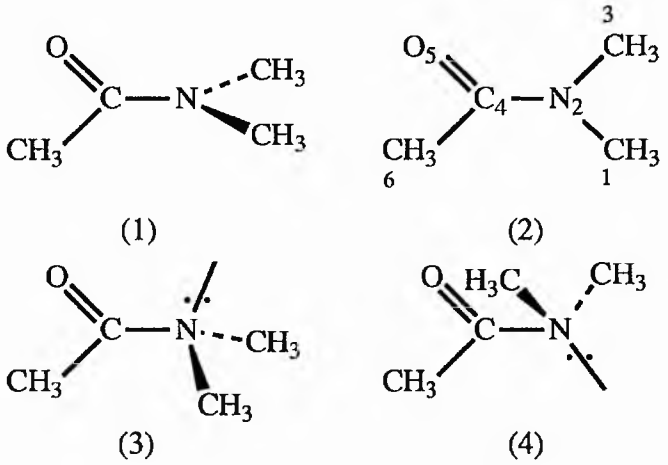


Figure 6.8

Attempts to find the transition states for internal rotations within this type of compound have been made. For example, Sawary and Yadav found that Ab Initio calculations upon N-Methyl and N-Ethyl Acetamide suggest that the nitrogen is non - pyramidal unless a non - planar geometry is used. If the molecule is twisted about the C-N bond, overlap between the nitrogen lone pair and the C=O system is reduced, leading to a pyramidal N atom [131]. They were also able to find the minima for the non - planar peptide unit, which were higher in energy. These calculations were performed without full optimisation, and many geometrical parameters were fixed. The basis sets used were STO-3G, 4-31G and 6-31G. Sapse, Fugler and Cowburn performed STO-6G and 6-31G* optimisations upon N-Methyl Acetamide, although they imposed a planarity constraint upon the system, and hence the results are of less importance in this situation [132].

Initially, four different geometries were set up, using optimised bond lengths and angles, but keeping the dihedrals fixed in the required conformation. The AM1 heats of formation, in kcal mol^{-1}, and the 3-21G energies, in AU, of the molecules were calculated, and the results are shown in Table 6.2.

Model	N Conformation	Me Conformation	ΔH_F	$E_{3\text{-}21G}$
1	Planar	Staggered	-28.65	-284.39
2	Planar	Eclipsed	-40.99	-284.43
3	Pyramidal	All Me down	+5.67	-284.32
4	Pyramidal	2 up, one down	-32.80	-284.40

Table 6.2

This table suggests that the planar, eclipsed structure, 2, is the most stable. Subsequent to this work, full Ab Initio optimisations using the 3-21G basis set were performed, and all four converged to the same point, equivalent to structure 2. Full AM1 and PM3 optimisations have also been performed. The results of these calculations, using the atom numbering system in molecule 2 of Figure 6.8 are shown in Table 6.3.

Parameter	AM1	PM3	3-21G
C_1-N_2	1.432	1.480	1.457
C_1-N_2-C_3	117.6	112.8	115.1
O_5-C_4	1.248	1.221	1.222
C_6-C_4	1.508	1.507	1.519
C_6-C_4-O_5	120.9	122.7	120.6
C_1-N_2-C_4-O_5	180.0	154.4	180.1
C_3-N_2-C_4-C_6	180.0	189.5	180.2

Table 6.3

The methods agree well about the calculated bond lengths, and the values of the reported angles are reasonably close to those of the 3-21G structure. The last two rows in the table represent how flat the molecule is, and both 3-21G and AM1 predict a very planar system. The case for PM3 is very different, with the oxygen atom being

substantially below the plane and C_3 being slightly below. The figures above also agree with the comparable sections of the N-Methyl Acetamide calculations of Sapse et al. [132].

These calculations agree with the results shown in the MOPAC manual, which state that N-Methyl Acetamide is a planar molecule, and this is predicted by AM1. PM3 produces a pyramidal structure for this molecule.

A second interesting feature of these molecules, is the calculated Mulliken Analysis charges, shown in Table 6.4. Also shown are the charges calculated for N-Methyl Acetamide by Supse, Fugler and Cowburn [132] using the minimal STO-6G, and the extended 6-31G basis sets. The C_1 charge is taken from the cis configuration results, whilst that for C_3 is taken from the trans results. The has been done to allow partially for the differences between the two molecules in question, and for the proximity of the oxygen atom.

	N,N-Dimethyl Acetamide			N-Methyl Acetamide*	
Atom	AM1 Charge	PM3 Charge	3-21G Charge	STO-6G	6-31G
C_1	-0.071	-0.086	-0.328	-0.081	-0.199
N_2	-0.343	-0.072	-0.885	-0.382	-0.894
C_3	-0.076	-0.074	-0.347	-0.076	-0.210
C_4	+0.303	+0.255	+0.866	+0.325	+0.734
O_5	-0.370	-0.362	-0.639	-0.295	-0.617
C_6	-0.237	-0.142	-0.695	-0.215	-0.512

Table 6.4

* NMA results from [132].

It should be noted that the Ab Initio and Semi Empirical charges should not be compared directly because they are calculated using different basis sets, which radically affect the Mulliken Charges. Nevertheless, for AM1 and 3-21G, it can be seen that the charges on N_2 and C_4, although of opposite sign, are of similar magnitude to one another. In the PM3 case the charges on these two atoms are totally dissimilar, indicating a different electronic structure for PM3.

(D2) Larger Systems.

It is possible that as the size of R_a, R_b and R_c in Figure 6.6 is increased, the molecule could become more planar (a possibility in the PM3 case) due to greater steric interactions between the groups. Thus, using methyl groups still for R_a and R_b, a molecule was designed in which R_c modelled part of the steroid ring system. In particular, the geometry of atoms C_{12}, C_{13}, C_{18}, C_{17}, and C_{16} of a steroid was set up, and the angles and dihedrals were set and fixed to a geometry corresponding to part of the C and D rings. The geometry of the rest of the molecule, including the hydrogen atoms in the steroid model was optimised.

The optimised PM3 structure for this molecule is shown in Figure 6.9, which also indicates the numbering system used in these molecules.

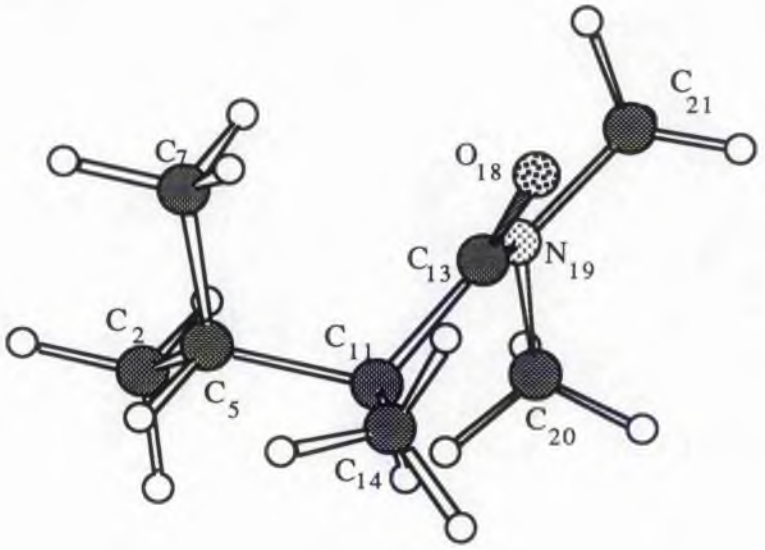


Figure 6.9

This picture shows the non - planar nature of N_{19} in the PM3 structure, and that geometry is still roughly in an eclipsed form. The important torsion angles are shown in Table 6.5.

Dihedral	AM1	PM3
C_{21}-N_{19}-C_{13}-O_{18}	5.6	5.5
C_{20}-N_{19}-C_{13}-O_{18}	189.8	145.5

Table 6.5

The first of these figures shows how close both methods predict the structure to being eclipsed, whilst the second indicates the planarity. AM1 predicts a structure that is about 10° off being planar, whereas for PM3, the result is over 30° out.

A similar study has been performed, using full isopropyl group, for R_a and R_b. The resultant dihedrals, using the same numbering system, are shown in Table 6.6.

Dihedral	AM1	PM3
C_{21}-N_{19}-C_{13}-O_{18}	345.4	329.3
C_{20}-N_{19}-C_{13}-O_{18}	136.6	289.9

Table 6.6

(D3) Rotation about the $C_{17}\beta$ Bond.

The structure of molecule 1a was examined in more detail, and a rotation about the bond joining the side chain to C_{17} was performed by varying the dihedral O=C-C_{17}-C_{13} (Figure 6.10).

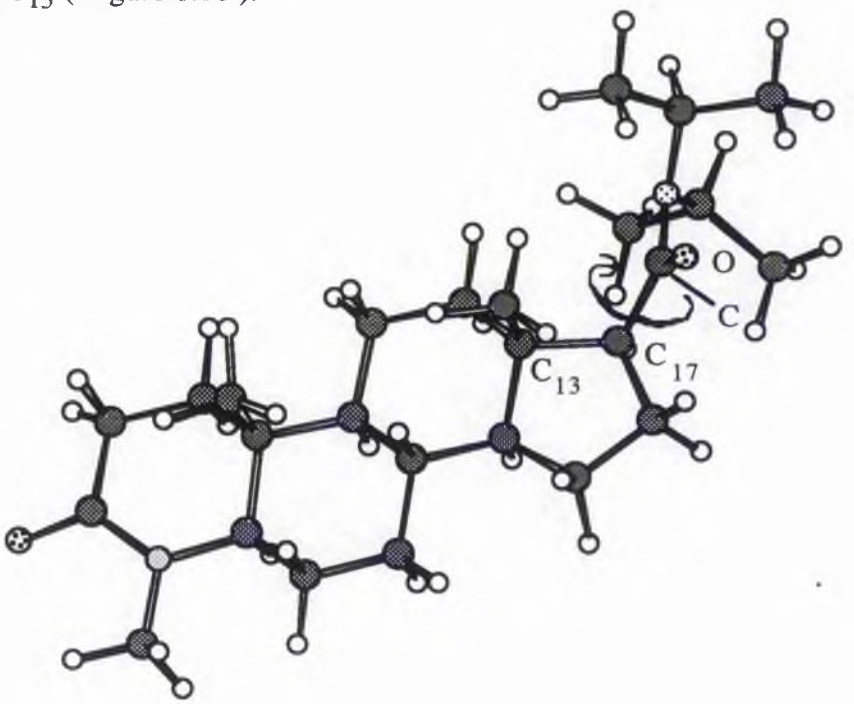


Figure 6.10

Molecule 1a.

As usual, the heat of formation was evaluated as the bond was rotated. The step size in the rotation was 20°.

The results are shown below in a graph of ΔH_F against reaction co-ordinate. (Figure 6.11)

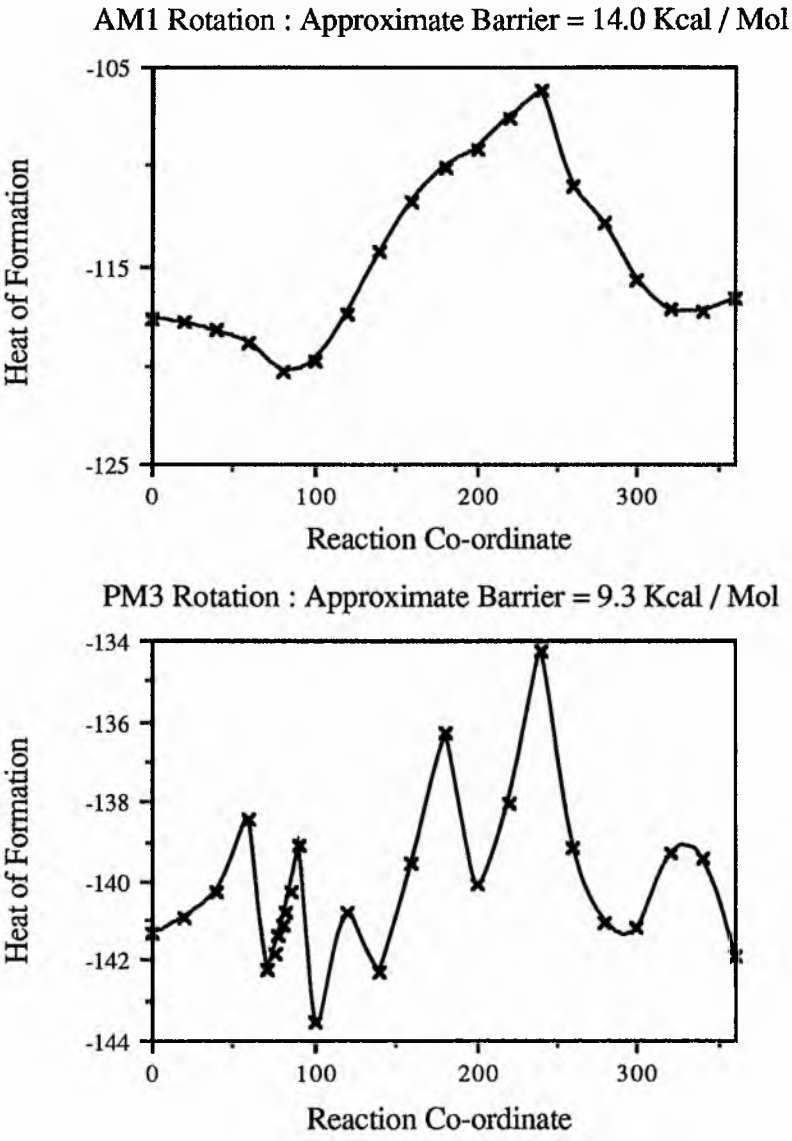


Figure 6.11

The graphs for the two methods seem totally different, with five peaks in the PM3 case and only two for AM1. The global minimum is around 90 - 100° according to both methods. The major AM1 peak at τ=260° corresponds to the isopropyl groups approaching C_{16}. There is a similar peak on the PM3 profile, which shows many more maxima and minima than does the AM1 graph, although the general trends in the graphs are not totally dissimilar.

Full optimisations were performed, starting at each of the minimum conformations, and the results are reported in Table 6.7. The heats of formation are in kcal mol^{-1}.

Method	τ init / °	ΔH_F	τ final / °
AM1	90	-120.36	84.6
AM1	320	-117.44	330.63
PM3	100	-144.36	85.9
PM3	140	-143.66	113.2
PM3	200	-140.64	189.6
PM3	300	-141.33	291.6
PM3	360	-142.16	61.54

Table 6.7

Minima Located in the Rotation About the $C_{17\beta}$ Bond.

The results in this table confirm that the minimum energy conformation is around τ=90° for both methods, which agrees with the results of the rotation in the Pregnane series of molecules in Chapter Five. This conformation was used as the starting point for all the optimisations.

(E) Conformation of the Steroid A-ring.

(E1) Boat - Chair Isomerism.

As mentioned in Section C of this chapter, the structure of molecule 1g was optimised using both methods starting from two different geometries for the heterocyclic nitrogen atom, resulting in distorted boat and chair conformations for the A-ring. The final heats of formation are reported in Table 6.8.

Conformation	AM1 ΔH_F	PM3 ΔH_F
Boat	-105.82	-115.40
Chair	-106.89	-118.71

Table 6.8

Both methods confirm the expected conclusion that the chair conformation has the lower energy.

(E2) Geometry of the Peptide - Like Amide Linkage.

The A-ring of this series of inhibitors of Aromatase also contains the peptide - like group depicted in Figure 6.6, and the calculations in this section attempt to examine the geometry of this linkage and the effects of steric interactions upon it caused by being part of the A-ring.

A model compound was selected to represent the steroid A-ring, as shown in Figure 6.12.

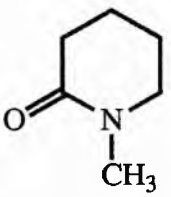


Figure 6.12

Full semi - empirical optimisations of this molecule were performed at the precise level with MOPAC. The Ab Initio calculations using the Quest module of Amber [64] were limited by the available disk space and by the poor performance of the optimisation routines. Once the Gaussian programs had been obtained for the FPS-500, work with larger basis sets was possible, allowing optimisation with both 3-21G

and 6-31G* bases. Some of the resulting torsion angles for these calculations are shown in Table 6.9.

Angle	Dihedral	AM1 /°	PM3 /°	3-21G /°	6-31G* /°
1	N_4-C_3-C_2-C_1	16.5	38.9	23.5	24.5
2	O_2-C_3-C_2-C_1	194.1	213.7	201.5	201.3
3	C_8-N_4-C_3-C_2	185.8	185.4	180.7	182.1
4	C_8-N_4-C_5-C_6	193.3	184.0	192.1	191.7
5	O_7-C_3-N_4-C_5	174.3	149.9	174.8	168.9
6	O_7-C_3-N_4-C_8	8.2	10.4	2.8	5.4

Table 6.9

In addition, the average difference between the 6-31G* and the AM1 dihedrals is 4.0°, whilst for PM3, this value is almost twice as big, at 7.3°.

The Ab Initio results in the table indicate that C_8, N_4, C_3, O_7 and C_2 are virtually co - planar, with both semi empirical methods reproducing this feature well (dihedrals 3 and 6).

In several other cases, it seems that AM1 predicts dihedrals that are more planar than the PM3 structure, and the Ab Initio results are between the two extremes (dihedrals 1, 2, 4, 5 and 6).

The only major deviation between the 3-21G and 6-31G* dihedrals reported above is for the fifth value, in which the larger basis set predicts a structure that is less planar by about 6°. The rest of the dihedrals are within 2 or 3 degrees.

The conclusions that can be drawn from these calculations are that AM1 produces a structure that is much more planar than the results of PM3 calculation. The AM1 results are, on average, closer to the Ab Initio results, which predict a structure in between those of the semi empirical methods. On the whole, the structure of the peptide - like linkage is more planar when present within the A-ring than when part of the 17β - side chain.

(F) Rationalisation of the Competitive Inhibition Using Theoretical Techniques.

(F1) Conformation of the Molecules Used in this Study.

Full optimisation at the precise level of the molecules 1a - 1i and 2a - 2i were performed starting with the dihedral τ (O=C-C_{17}-C_{13}) at 90° (see Figure 6.10). The optimised AM1 structures converged to values of τ between 84.5° and 87.6°, except for molecule 1f, for which the minimum was at 105.1°. This molecule has a single tertiary - butyl group attached to the nitrogen atom, as opposed to the two iso - propyl groups prevalent in most of the other molecules, which explains the different conformation taken up by this molecule.

Starting the PM3 optimisations at this value of τ gave a much wider range of final values for the dihedral, ranging from 82.4° up to 117.0°.

It should be noted that no attempt has been made to search the complete conformational space of these 17β side - chains because of the large number of variables involved. The structures used in any further calculations are the results of the optimisations discussed above, with the structures of the alkyl - groups attached to the nitrogen atom in the conformation obtained in these optimisations. These may or may not be the global minima.

A second notable point is that the barriers between possible minimum energy side - chain conformations are likely to be reasonably small and the molecules will be able to move over small barriers by virtue of their thermal energy. Thus, although there will be a broad range of conformations present in a sample of molecules, it is necessary to select one static conformation for further calculations. This has O=C-C_{17}-C_{13} in the global minimum energy position, but all other dihedrals in the side - chain could be in a local minimum.

One experimental geometry for this type of compound has been obtained, for a molecule in which the D-ring is 6-membered, and the side - chain is in the equivalent of the 17α - position. Analysis of the crystal structure shows that, as expected, the amide groups have a geometry that is very nearly planar. Some interesting geometrical parameters are show in Table 6.10, in which the first two dihedrals refer to the side - chain. Angles and dihedrals are quoted in degrees, and bond lengths in Å.

Parameter	X-ray	AM1	PM3
$O_{26}C_{20}N_{21}C_{24}$	175.4	167.5	248.8
$O_{26}C_{20}N_{21}C_{22}$	356.7	3.0	24.1
$O_{28}N_4C_3C_2$	185.9	181.0	187.4
$C_{28}N_4C_5C_{10}$	201.2	199.0	182.0
$C_{28}N_4C_3O_{27}$	6.1	4.0	12.2
$C_5N_4C_3C_2$	355.5	346.0	325.6
$C_3N_4C_5C_{10}$	31.5	33.8	43.1
$C_3C_2C_1C_{10}$	324.0	318.0	306.6
$O_{27}C_3N_4$	121.5	120.4	118.6
$C_{28}N_4$	1.4642	1.4374	1.4827
$O_{27}C_3$	1.2648	1.2499	1.2220

Table 6.10.

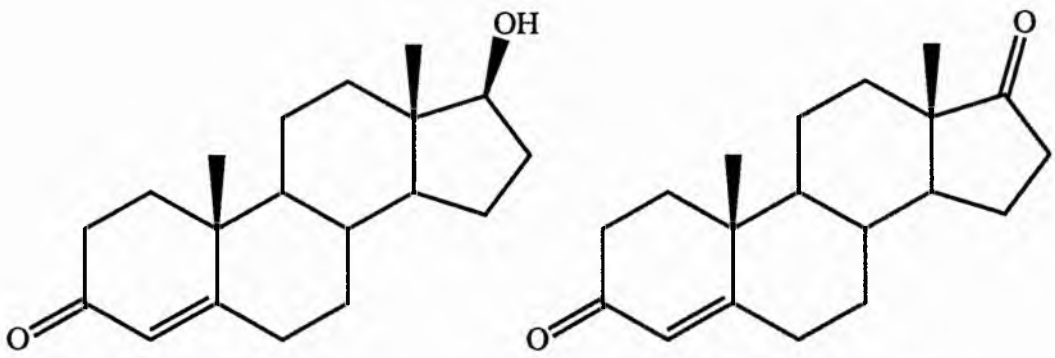
The atom numbers in the above table refer to the numbering system for the X-ray structure, and are compared to the calculated geometries for molecule 1e, which, like the experimental structure, has ethyl groups attached to the nitrogen atom in the side chain.

It can be seen that, in general, AM1 gives a better representation of the dihedrals in this type of molecule than does PM3, whilst the PM3 bond lengths are also not as good because of the differences in the electronic structure predicted by the two methods for the amide group.

(F2) Electrostatic Potential Similarity Work.

The optimised structures discussed in the previous section were thus formatted for use with the ASP program to evaluate the Electrostatic Potential Similarity Index (EPSI). This work was done using both AM1 and PM3 structures and charges, although less emphasis can be placed upon the results using the latter due to its poor description of the charge distribution and geometry of the "peptide - like" amide group.

This section is interested in the Competitive Inhibition of Aromatase by these molecules (see Table 6.1 for the activities), which means that the inhibitors produce their activity by blocking the entry of the substrates to the active site. To do this, they must be able to bind strongly to the active site without undergoing aromatisation, and hence must display a Molecular Electrostatic Potential (MEP) which is similar to that of Testosterone and Androst-4-ene-3,17-dione (Androstenedione), which are the two possible substrates (Figure 6.13). Full semi empirical optimisations of these two molecules have been performed, the latter being molecule And-14, one of the inhibitors of G6PDH studied in Chapter Five.



Testosterone Androstenedione

Figure 6.13

Since the substrates exhibit either a 17β-hydroxy or a 17-keto group, which produce markedly different regions of potential, it would seem that direct binding between the enzyme and the C_{17}-substituents is unlikely occur. Thus large charges at this position may affect the EPSI without significantly altering the affinity of the molecule for the receptor. Nevertheless, similarity work has been attempted, using Androstenedione as the lead compound.

(F3) Details of the ASP Calculations.

The inhibitors were overlaid on the structure of Androstenedione using the least squares fitting option within Chem-X [70]. For most of the molecules, carbons 1, 17, 11 and 17 were superimposed each with equal weighting. For the molecules containing a 10β-2β - ethano bridge, the positions of carbon 1 are likely to be different, and thus carbons 10, 3, 7, 11 and 17 were superimposed. The numbering system from Figure 1.1 is repeated below in Figure 6.14, for convenience.

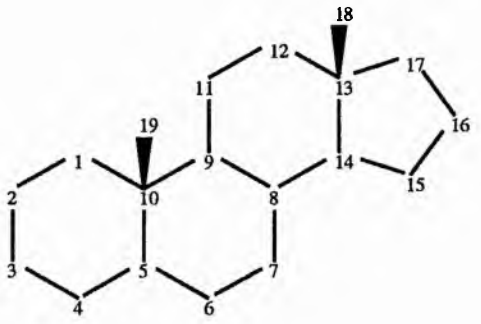


Figure 6.14

The overlaid structures were then used as input to ASP, and as usual, the Carbo and Hodgkin Indices were calculated, and the program also attempted to maximise the values by altering the orientation of the inhibitors with reference to the lead molecule.

(F4) Results.

It was found that, on the whole, the optimised fit produced by ASP gave a poor overlap of the inhibitors with Androstenedione. A qualitative measure of this fit, as described in Chapter Five, along with the optimised and unoptimised EPSI s for the AM1 structures is shown in Table 6.11. The PM3 unoptimised EPSI s are also included.

Molecule	K_i	AM1 C	AM1 H	AM1 CO	AM1 HO	PM3 C	PM3 H	AM1 Fit
1a	3.3	0.388	0.388	0.485	0.480	0.572	0.571	C
1b	9.0	0.308	0.307	0.566	0.619	0.537	0.536	D
1c	4.8	0.421	0.420	0.503	0.501	0.554	0.553	B
1d	3.8	0.375	0.375	0.745	0.745	0.546	0.545	D
1e	6.2	0.347	0.344	0.459	0.459	0.549	0.549	C
1f	2.5	0.489	0.485	0.600	0.810	0.613	0.609	C
1g	5.8	0.815	0.814	0.888	0.893	0.950	0.950	C
1h	0.8	0.791	0.790	0.894	0.896	0.900	0.899	C
1i	-	0.575	0.574	0.672	0.671	0.743	0.742	C
2a	7.6	0.455	0.451	0.484	0.481	0.563	0.562	B
2b	1.1	0.585	0.585	0.634	0.613	0.587	0.587	B
2c	0.03	0.951	0.947	0.953	0.949	0.968	0.965	A
2d	0.02	0.885	0.885	0.927	0.921	0.929	0.929	B

Table 6.11.

K_i in μM. C : Carbo Index. H : Hodgkin Index.

CO : Optimised Carbo. HO : Optimised Hodgkin.

(F5) Discussion.

The K_i values for the inhibitors shown in Table 6.11 refer to the reaction

$$E + I \rightleftharpoons EI$$

where E is the enzyme, I the inhibitor and EI the enzyme - inhibitor complex. K_i is the equilibrium constant for the dissociation reaction, and is calculated by

$$K_i = \frac{[E]\,[I]}{[EI]}$$

it depends upon the concentrations of the complex, the enzyme and the inhibitor. Small values of K_i indicate that a large amount of the enzyme is present in the form of the complex, meaning that very little of the enzyme is available to catalyse the reaction.

The equilibrium constant for the reverse reaction is associated with the free energy of dissociation by the equation

$$\Delta G = -RT \ln K_i$$

where T is the temperature and R is the Ideal Gas Constant. ΔG tells us the driving force behind the dissociation, and if the interaction between the enzyme and inhibitor is largely based upon the electrostatic attractions, there should be a linear relationship between ln K_i and the electrostatic binding energy. By assuming electrostatic complementarity in the enzyme - inhibitor complex, the relative electrostatic binding energies of various inhibitors can be estimated from their similarity to the substrates, which can be assumed to bind strongly. Thus a graph of ln K_i against EPSI may give a straight line.

It should be noted in this analysis that it is possible to obtain inhibitors with a greater affinity for the enzyme than the substrates. In particular the transition state of the reaction can have an affinity several orders of magnitude greater than those of the substrates. Hence any substance that partially resembles this transition state will have a very high affinity. Such a molecule will have a Similarity Index less than 1.000, and the straight line relationship between ln K_i and the EPSI will not hold.

The graphs of ln K_i (in μM) against unoptimised EPSI are presented in Figure 6.15, below, which portrays the EPSI s calculated using the AM1 structures and charges for the molecules. Some correlation was obtained using the PM3 results, but this was inferior to that obtained with AM1.

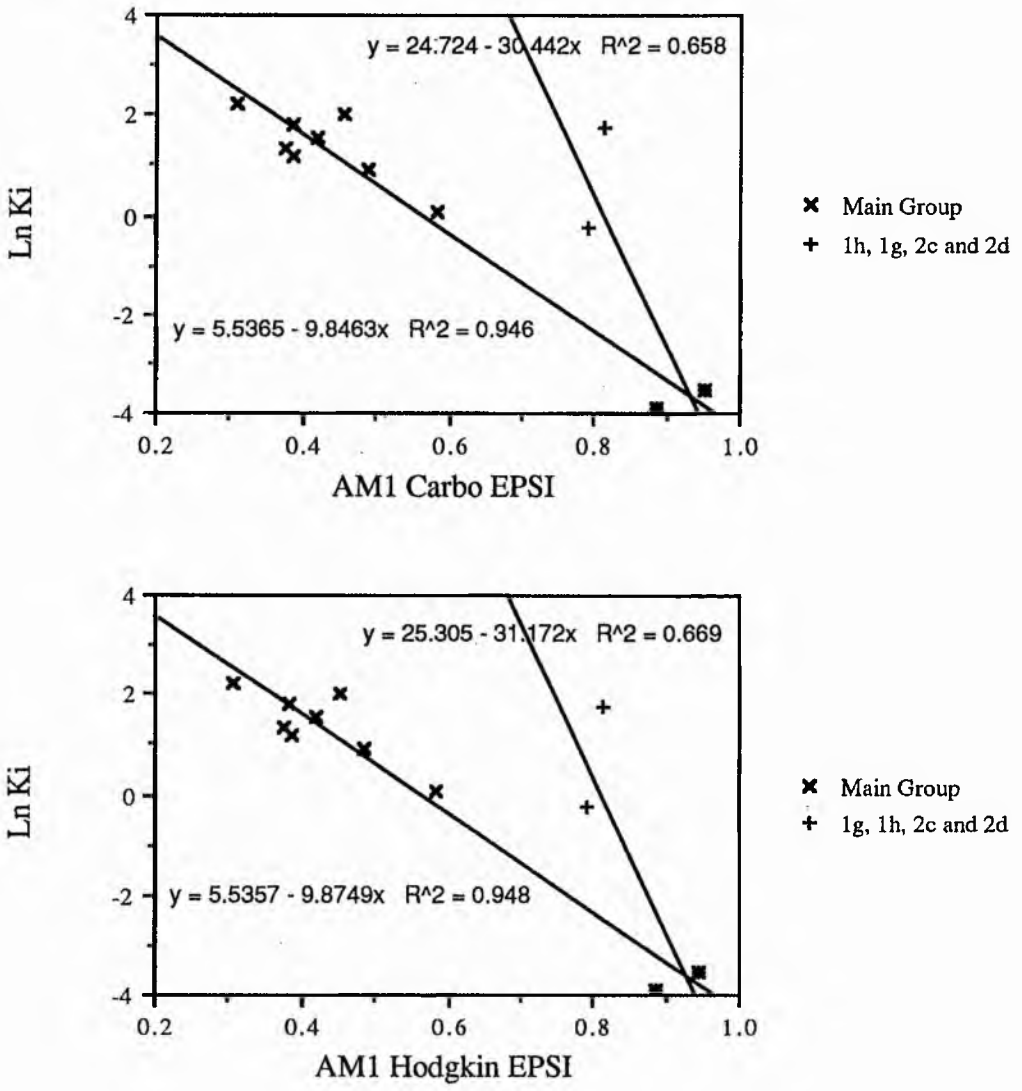


Figure 6.15

Two points in the above graph have been omitted from the calculation of the correlation,line and these correspond to the molecules 1h and 1g, which are plotted as plus signs (+) on the graph.

The crosses (x) indicate the main data points for molecules 1a - 1f and 2a - 2d. 2c and 2d also form a separate line including 1h and 1g with a moderate correlation co-efficient. The former points lie on both lines, and are represented by stars (i.e a

plus on top of a cross). There is no K_i value for 1i, and hence this cannot be included on the graph.

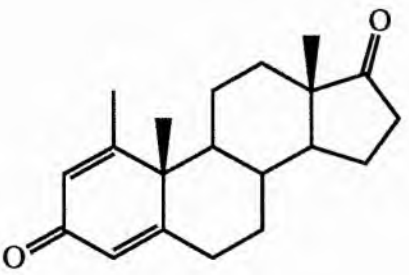
The justification for putting 1h and 1g on a separate line is that they are the two molecules in group 1 with a 17-keto group. This is also present in molecules 2c and 2d. It is not known whether these four molecules truly form a separate line, or if it it a chance alignment of 1g and 1h. If the latter were the case, then the EPSI is indicating large changes in the MEP in a region that is not involved in the binding. The fact that the line has a large gradient implies a reduced dependence of the binding process on the overall electrostatics, as would be the case if this region were not involved.

An argument against the presence of two different lines in the graph is the fact that 2c and 2d would not be expected to lie upon the main line but on the 17-keto line because they are 17-keto derivatives.

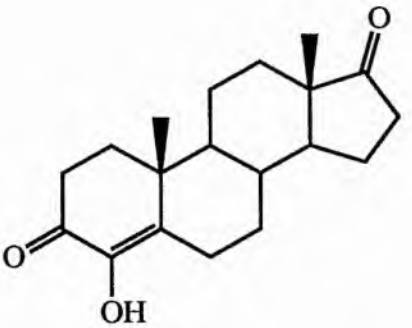
(F6) Other ASP Work.

As with the Androstane series of inhibitors in Chapter Five, an attempt has been made to assess the EPSI using a double Van der Waal's scaling factor. The result of such calculations is that the correlation of the main points is improved to 0.960 using the Hodgkin Index, whilst the Carbo Index correlation co - efficient decreases to 0.879. The correlation for the second line involving molecules 1h, 1g, 2c and 2d improves over the single Van der Waal's scaling in both cases. The co-efficients are 0.795 and 0.782 for the Hodgkin and Carbo Indices respectively.

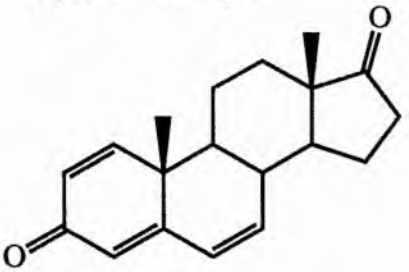
One further study has involved the calculation of the EPSI s for alternative inhibitors, which do not contain the amide group. The molecules studied are shown in Figure 6.16.



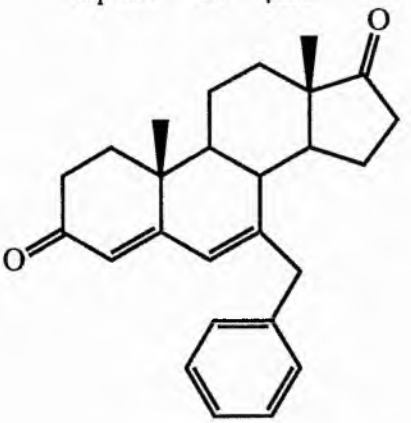
1-MethylAndrostan-1,4-diene-3,17-dione
K_i=8.2 * 10 $^{-2}$ μM



4-HydroxyAndrost-4-ene-dione
K_i=3.7 * 10 $^{-2}$ μM



1,4,6-Androstatriene-3,17-dione
K_i=1.0 * 10 $^{-1}$ μM



7-BenzylAndrost-4,6-diene-3,17-dione
K_i=61 * 10 $^{-3}$ μM

Figure 6.16.

Other Aromatase Inhibitors[1]

The results from these calculations, which were performed in the same manner as those presented earlier in the chapter, have been added to the graphs in Figure 6.15, and are shown in Figure 6.17. The Carbo EPSI s are 0.881, 0.911, 0.936 and 0.860 respectively. The Hodgkin values are 0.874, 0.906, 0.934 and 0.859.

---

[1]The experimental data for these molecules is reported in ref 133, except for the last molecule, which is studied in refs 134 and 135. The latter reference also discusses Molecular Modelling of this, and related compounds.

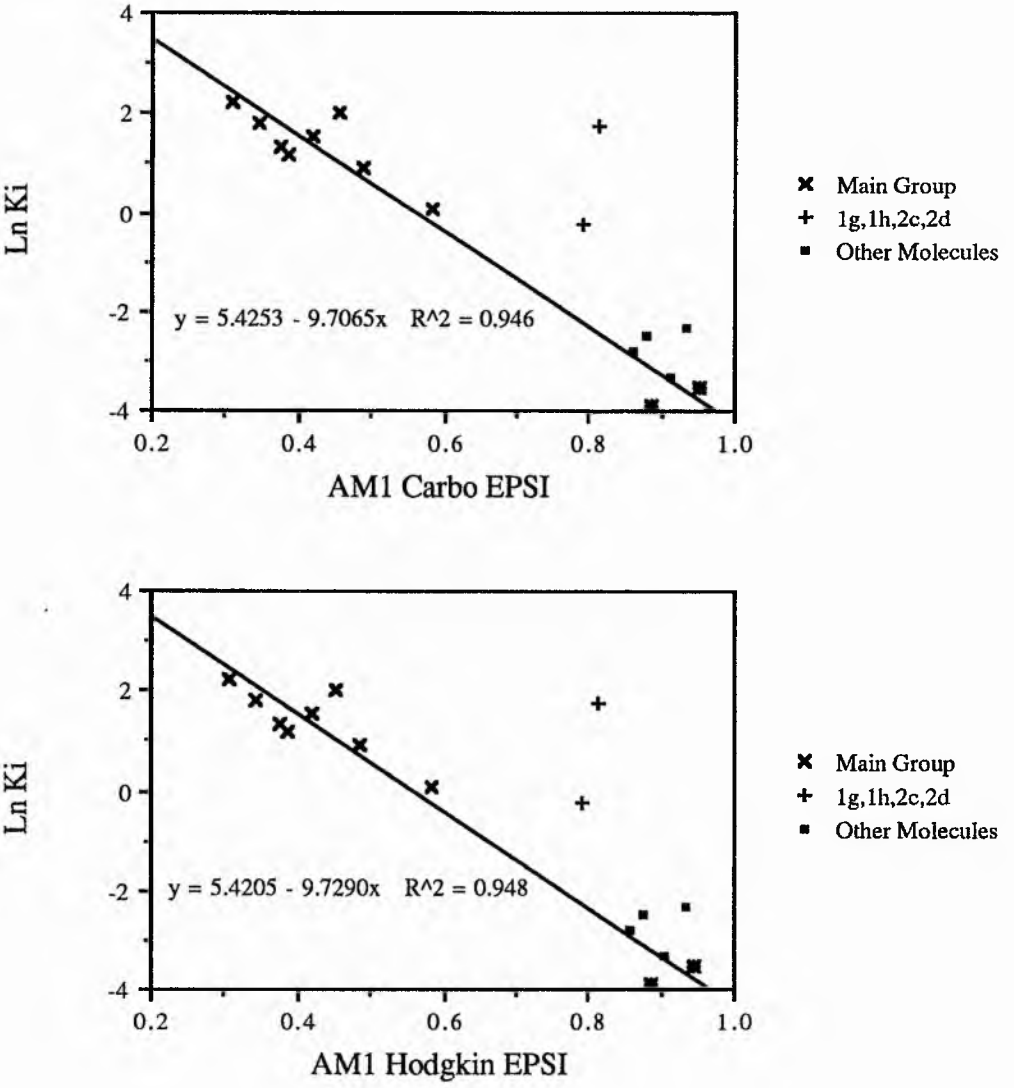


Figure 6.17

The additional points on the graphs neither confirm nor deny the possibility of the second line, but show predictive power of using this method to assess inhibitor activity using the primary line.

(G) Analysis of the Electrostatic Potential Maps for the Aromatase Inhibitors.

The potential maps in Figure 6.21 at the end of this chapter are those for a selection of inhibitors, the two substrate molecules and the inactive molecule, 1i.

A study of Testosterone and Androstenedione, and the differences between the two, indicates that the D - ring of the molecules is probably unimportant in determining the binding affinity for the two.

All the molecules exhibit a 4-ene-3-one configuration, which obviously plays a vital role in the binding process, probably via hydrogen bonding to the enzyme. Adjacent to this region is an area that is commonly light blue (positive) in most but not all of the molecules, not being present in the least active. It is conspicuously absent in molecule 1i, which has no affinity for the active site. This region involves hydrogens attached to C_6, C_{19} and C_{11}. The high affinity of the 4,6-Androstadienes [134,135], such as the molecule 7-Benzyl ADD studied in this chapter indicates that the central part of this region is not involved in the binding because this is where the conjugated double bond (Δ^6) would be located. The delocalised system would have a region of negative (red) potential associated with it.

The rear (α) face of the steroids, as with the G6PDH inhibitors of Chapter Five, seems to have little involvement in the binding, other than via possible hydrophobic interactions. Thus the important regions in determining the affinity of steroids for the aromatase active site are shown in Figure 6.18.

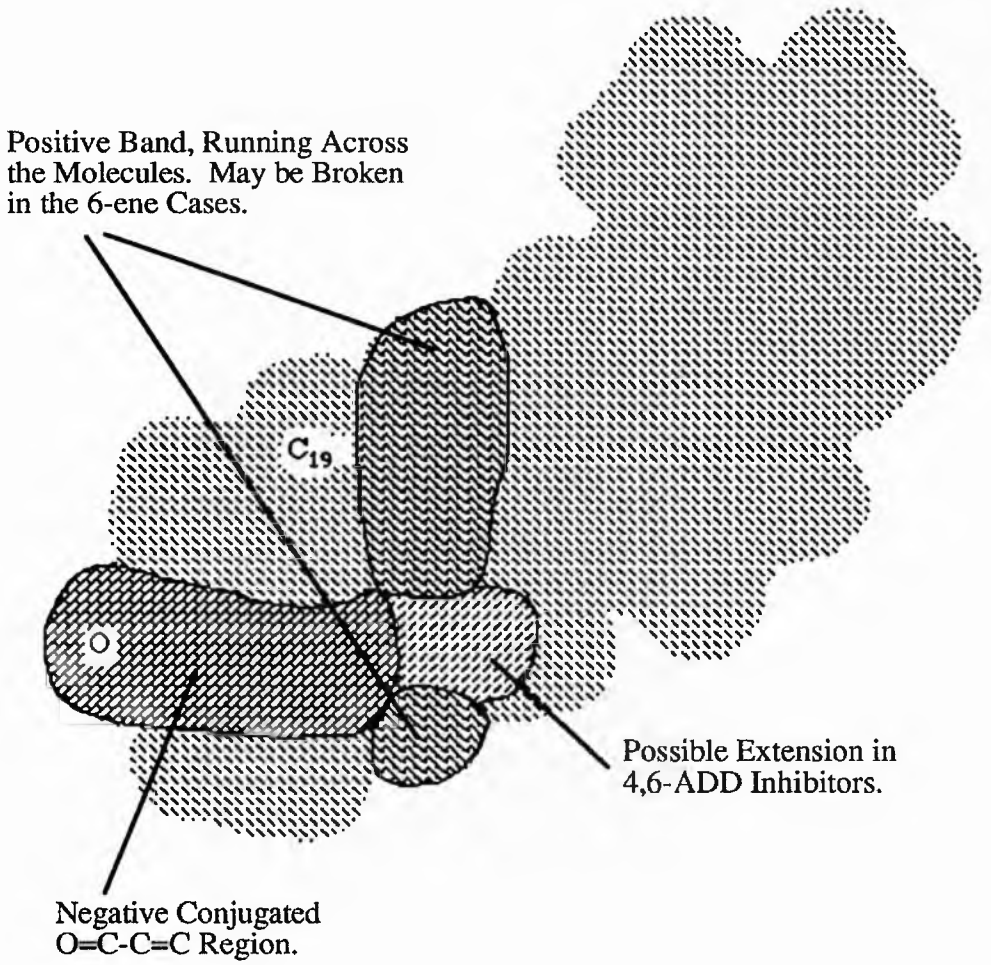


Figure 6.18.

Regions Involved in Binding to the Aromatase Active Site.

(H) Theoretical Model of the Enzyme Inactivation.

There is no correlation between the half - lines shown in Table 6.1 and the Similarity Indices calculated in the previous sections. The EPSI s for the inactive molecules cover the same range of values as do the results for the active compounds. Thus, the inactivation is not a virtue of a high affinity of the steroids for the active site of the enzyme.

The experimental data show that the molecules that lead to time dependant inactivation require the 17β-N,N-dialkylcarbamoyl group, and that neither the keto nor the alcoholic group (molecule 1i) result in active molecules. Not all of the steroids which have the amide group inactivate the enzyme, with the exception being molecule 2a. The difference between this molecule and the active 2b is a Δ^4 double bond. Changes in the 17β-side chain are also important, as exemplified by the difference in enzyme half - life between 1a and 1e, which have N,N-diisopropyl and N,N-diethyl groups respectively.

Greway and Levy [129] suggest the possibility of an alternative mode of binding of the active molecules, such that the 17β-side chain is juxtapositioned with the active site of the enzyme, thus allowing an enzyme activated oxygen species to cause the inactivation. They cite the relatively low affinity of molecule 2a for the enzyme as the reason for its failure to cause inactivation.

Inspection of the experimental data reveals that 1b has a K_i value greater than that of 2a, yet it is still able to inactivate the enzyme, which calls into question the theory of the authors of the experimental data as to the inactivity of molecule 2a.

(H1) Study of the Ionisation Potential.

The aromatisation reaction involves oxidation of the steroid A-ring, which corresponds to a loss of electrons. These electrons are likely to be removed from the HOMO of the steroids, and thus if the inactivators participate part way in the reaction, there may be some correlation with the ease of oxidation, as predicted by the energy of the HOMO, and the degree of inactivation. Thus an attempt has been made to correlate the Ionisation Potential (which is the negative of the HOMO energy) and the half life. These values are shown in Table 6.12, below, there was no difference between the

calculated IP s of the active and inactive molecules. This table also shows the AM1 calculated dipole moments in Debye, which are discussed in Section H4.

Molecule	AM1 IP	PM3 IP	AM1 Dipole
1a	9.398	9.424	4.41
1b	9.425	9.436	4.80
1c	9.390	9.405	4.17
1d	9.398	9.363	4.70
1e	9.419	9.457	4.24
1f	9.835	9.726	3.28
1g	9.481	9.498	2.79
1h	9.485	9.434	2.85
1i	9.411	9.441	2.77
2a	9.572	9.491	2.84
2b	9.606	9.496	3.38
2c	10.174	10.411	2.71
2d	10.013	10.140	2.89
Testosterone	10.052	10.187	3.54
Androstenedione	10.101	10.242	3.32

Table 6.12.

It should be noted that although the theoretical methods employed in these calculations are able to give and IP that approximates the experimental value, as shown in Chapter Seven, the actual correlation between experimental and predicted values is poor. The study in Chapter Seven also applies to steroids that do not contain the amide moiety, which as previously mentioned, is poorly reproduced by the semi empirical methods, particularly PM3.

(H2) Location of the Frontier Orbitals.

One possible difference between the two groups of molecules could be the actual position of the HOMO or near - lying orbitals. For example the HOMO could possibly be located upon the 17β-side chain of the active molecules, but elsewhere upon the others and the substrates. Thus the enzymic oxidation would be directed to the side chain in the former group, leading to the inactivation.

This idea has been explored by obtaining the normalised eigenvectors for the Molecular Orbitals using the *VECTORS* keyword in MOPAC, and extracting those corresponding to the HOMO and the orbital lying one below the HOMO. Any particularly large eigenvectors would indicate that the Molecular Orbital (MO) resembles, to some degree, the Atomic Orbital to which that eigenvector corresponds, implying localisation of the MO at the atom in question.

The eigenvectors for these large molecules are too large to report in this thesis, since there are four for each non - hydrogen atom and one for each hydrogen atom. Thus a qualitative picture of the orbital location will be presented for certain molecules, including the two intermediates in the aromatisation shown in scheme 6.1.

Substrates.

The HOMO for both Androstenedione and Testosterone is largely located around the A / B-ring junction, especially O, C_3, C_4, C_5, C_6, C_7, C_{10}, C_{19}, and C_9. The involvement of C_{19} is as would be expected as this is the site of the two hydroxylation reactions occurring during the aromatisation. The "HOMO-1" orbital is located on the A-ring of Testosterone, but mainly upon the upper section of the C and D rings of Androstenedione. This difference is due to the C=O double bond in the latter molecule.

Intermediates.

The location of the HOMO in both the 19-Hydroxy and the 19-Keto Androstenedione derivatives is the same as for the substrate molecules. In the 19-Hydroxy molecule, the HOMO-1 is mainly based upon the A-ring as for Testosterone. The second intermediate is more similar to Androstenedione, with the HOMO-1 on the C and D rings.

Inhibitors.

The HOMO of all the type 1 inhibitors is located around the lower part of the A and B rings, in particular, the oxygen atom, N_4, C_5, C_6 and the methyl group attached to the nitrogen. C_3 has small eigenvectors. The HOMO s of 2a and 2b are located on the 17β-side chain, the most marked contributors being from the O and N atoms and the two carbons attached to the nitrogen.

The HOMO s of 2c and 2d are located on the upper regions of the C and D rings and the A / B-ring junction respectively.

The HOMO-1 s of 1a - 1f are located on the side chain, similar to the HOMO s of 2a and 2b. For 1h and 1g it is on the C and D-rings, whilst for 2a and 2b it is in the A and B-rings. The HOMO-1 for 2c is localised on the A-ring, but there is no particular positioning for 2d.

The location of the two orbitals on the type 2 inhibitors is thus the reverse of that of the type 1 molecules. This and the lack of a visible difference between 2a and 2b indicates that the location of the orbitals is not of vital importance in determining the ability of a compound to cause inactivation.

(H3) Inversion of the Inactivators Within the Active Site.

As already mentioned, the authors of the experimental data suggested a second orientation of the inactivators within the active site of the enzyme, leading to an interaction between the 17β-side chain and the cytochrome P450. For this to occur, the molecules must have a high affinity for the second binding mode, which would be indicated by a high degree of similarity between the inactivators in their reversed conformation, and the substrates.

Assuming various regions of close contact between the enzyme and the substrate, it would seem likely that there is little scope for the accommodation of bulky side groups near the active site. Exceptions to this are the regions around the 4 and 7α positions of the steroids reported in the literature [120,126,135]. One possible inverted alignment of the 17-substituted inactivators would be a complete rotation about an axis running into the page as shown in Figure 6.19. Thus the 17β-side chain would be able to fit into the cavity in the enzyme adjacent to carbon 4 of the substrate.

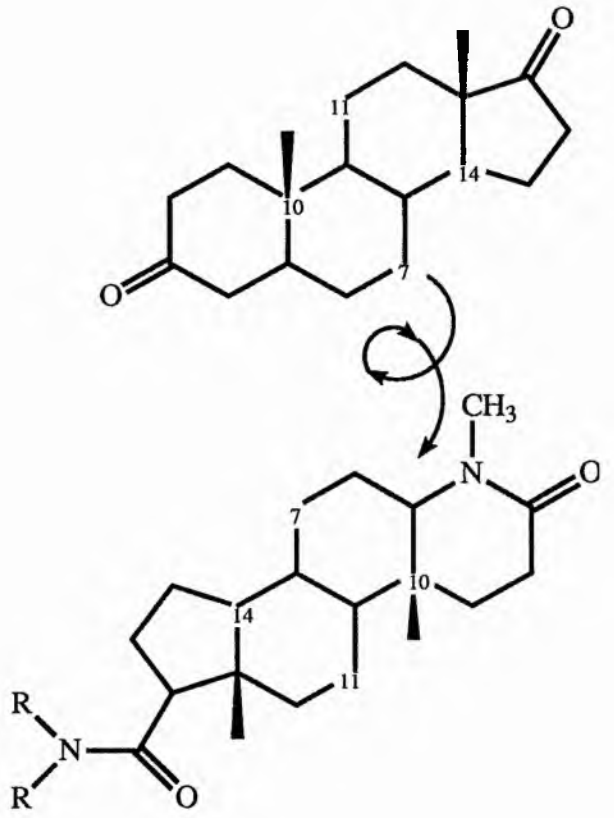


Figure 6.19.

Possible Inverted Binding Mode of Active Molecules.

To test this hypothesis, all of the inhibitors were aligned with the substrate, using the fitting option within Chem-X, such that equal weighting was given to the overlap of the following atom pairs : C_{10}-C_{14}, C_7-C_{11}, C_{11}-C_7 and C_{14}-C_{10}. Using these alignments as input to the ASP program, similarity calculations were performed. Optimised and unoptimised EPSI s were calculated and recorded, although optimisation was only carried out for the Carbo Index. The results are reported in Table 6.13, below. These calculations were performed upon both the active and the inactive molecules, in an attempt to discover a difference in affinity for the enzyme between the two groups.

Molecule	C	H	CO	$t_{1/2}$
1a	-0.031	-0.031	0.536	15
1b	0.021	0.021	0.569	28
1c	0.057	0.057	0.525	14
1d	0.011	0.011	0.655	40
1e	0.093	0.093	0.578	20
1f	0.229	0.228	0.490	19
1g	0.457	0.457	0.690	-
1h	0.405	0.405	0.688	-
1i	0.262	0.262	0.420	-
2a	0.090	0.089	0.631	-
2b	0.154	0.154	0.566	36
2c	0.531	0.528	0.805	-
2d	0.499	0.498	0.706	-

Table 6.13.

AM1 EPSI s for Rotated Conformations.

Analysis of the results in the table shows neither a correlation between the $t_{1/2}$ values and the EPSI, nor a clear indication of the difference between active and inactive molecules. The active molecules also have the smallest EPSI s, indicating that the affinity of the inactivators in this mode of binding is low. Thus it is unlikely that a second mode of binding occurs.

This conclusion is supported by the high correlation between the EPSI and the ln K_i values. An alternative binding mode would be unlikely to have a similar affinity for the active site as would the primary mode, which would cause a deviation from the line in Figure 6.15.

(H4) Dipole Moments.

As previously mentioned, the dipole moment can be used as a guide to biological activity, and thus the calculated dipoles have been plotted upon a graph with the $t_{1/2}$ values. This has only been performed with the results of the AM1 calculations, due to the poor modelling of the electron distribution in the amide - group by PM3. The results are shown in Figure 6.20, and the calculated values in Table 6.12, earlier in the chapter. The inactive molecules do not have an associated half - life, and thus they are represented upon the graphs as vertical lines, as are the dipole moments of the substrate molecule. All dipoles are measured in Debye.

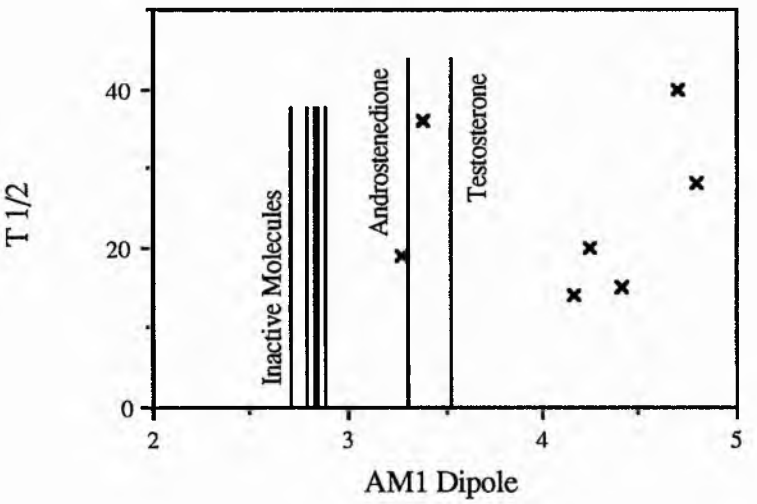


Figure 6.20.

The graph indicates a possible method of distinguishing between the active and inactive molecules, although it can not be used to predict the half - life values. It can be seen that the dipole moments of the inactive molecules are all substantially lower than those of the substrates, whereas the inactivators of the enzyme have dipoles that are approximately equal or greater.

It should be noted that there is a strong possibility that this is not a cause and effect relationship, but merely a result of the differences between the substitution of carbon 17. The dipole moment is also extremely dependent upon the orientation of the 17β - side chain. If the molecule is able to take up a number of conformations, then an average value of the possible dipoles will be observed, whereas in this type of theoretical consideration, only one configuration (usually the lowest energy one) can be considered at once. Thus any dependence upon the dipole moment could be an artificial effect of this limitation. Nevertheless, the analysis does separate molecules 2a and 2b, which is a necessary attribute of any method used to assess the ability of the molecules to cause Time Dependent Inactivation. To thoroughly confirm the practicality of this method, more experimental data for the 17-substituted molecules, both active and inactive, is required.

(I) Conclusions.

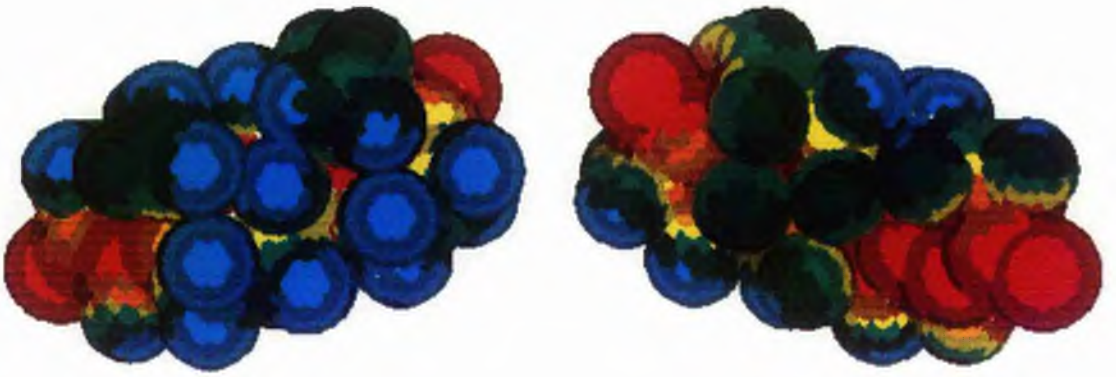
This chapter has presented a method of assessing the activities of competitive inhibitors of the enzyme Aromatase using purely theoretical methods. The AM1 semi empirical structures and charges have been used to calculate the Electrostatic Potential Similarity Indices, which correlate with the natural logarithm of the K_i values quoted in the literature. The correlation co-efficients give an accuracy of prediction of about 95 %, although the errors will be increase due to the process of taking logs. The calculations fail to predict accurately the activity of molecules which exhibit a high similarity to Testosterone at the 17 - position, such as 1h and 1g. This indicates that this position is probably not involved in the binding. These calculations also can not account for the inactivity of molecule 1i as a competitive inhibitor, since its calculated EPSI is reasonably high.

The calculations have also been extended to several other molecules, and although the numerical predictions may deviate slightly from those predicted using the EPSI, the order of magnitude of the calculated results is extremely accurate.

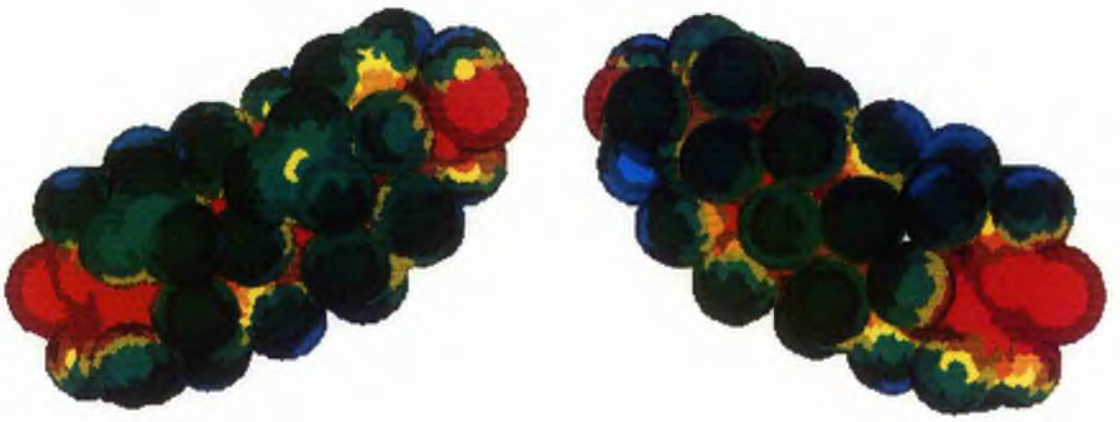
Very little information has been found in the theoretical calculations which allows the prediction of the activity of the 17-substituted molecules studied in this chapter as Time Dependent Inactivators of the enzyme. One possible guide to separating the active from the inactive molecules is the AM1 Dipole Moment, for which the calculated dipoles of the inactive molecules are substantially lower than those of the active molecules and enzyme substrates.

Finally, this chapter has also highlighted the failure of the PM3 method to reproduce charge distributions and geometries for the amide groups present within these molecules. The AM1 results correlate to a better degree with those obtained from Ab Initio work, but are not totally satisfactory for this type of chemical system.

Testosterone

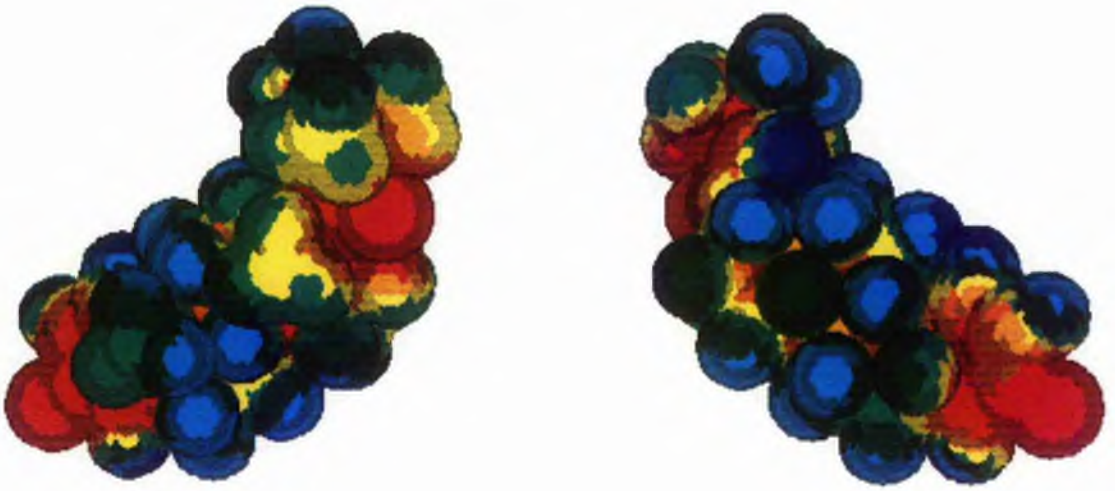


Androst-4-ene-3,17-dione

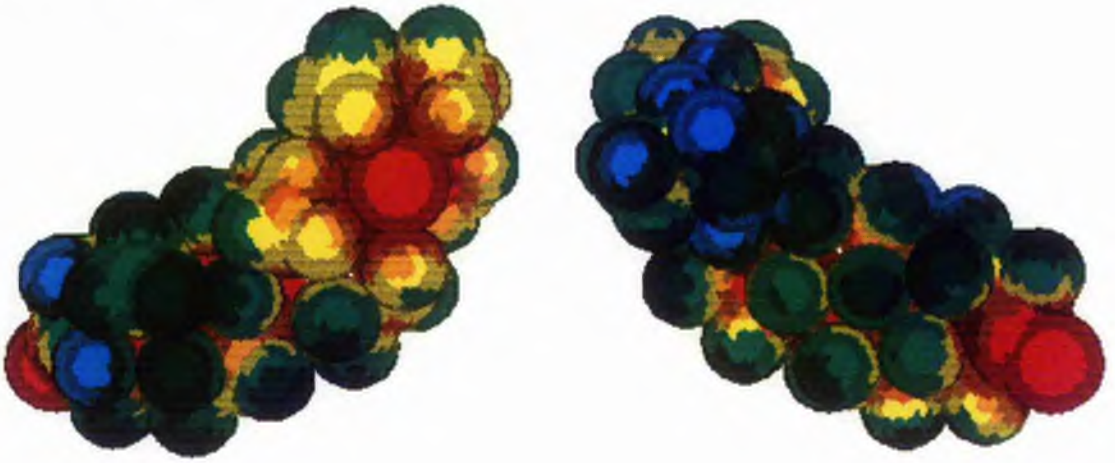


Arom-1i

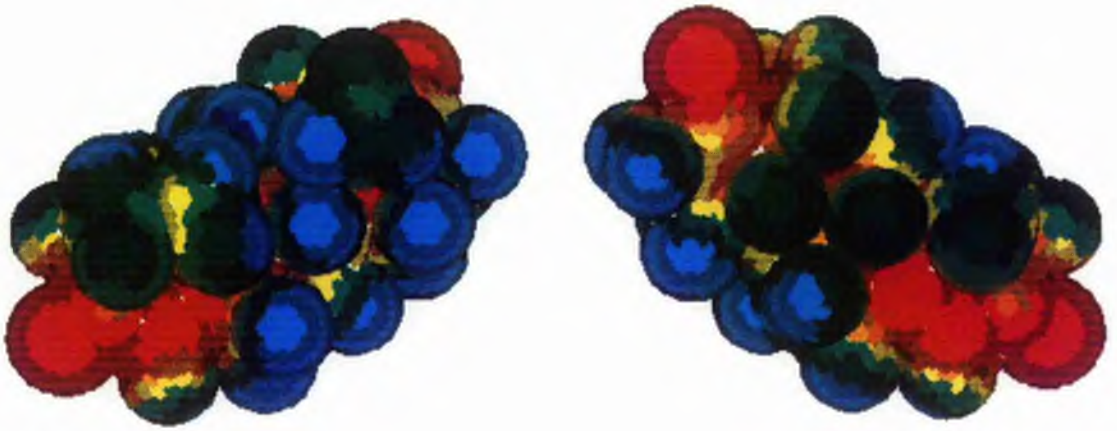
Figure 6.21.



Arom-1f



Arom-2a



Arom-2d

Figure 6.21 (Cont.)

Chapter Seven

A Comparison of AM1 and PM3 with Ab Initio and Experimental Results.

(A) Introduction.

Ideally, we might wish to perform a full geometry optimisation at the Ab Initio level using a large basis set, such as 6-31G* [29] on all the molecules we wish to study. This is not generally possible for the steroid molecules of interest in this work because of the prohibitive time needed for such calculations. Ab Initio calculations have been done for this class of molecule [136], but for studies of the large number of molecules presented here such a procedure would be impractical, and the AM1 and PM3 semi empirical methods have been used to obtain molecular geometries.

Both AM1 and PM3 have been well tested, and have been found to give results in agreement with experiment in many cases, as already mentioned in Chapter Two. Various comparisons of the performance of the two methods have also been presented in a number of cases : Dewar and co-workers showed that PM3 gave better results than AM1 in reproducing the properties of the set of molecules used in the parameterisation [40]. John Wilkie from this research group found AM1 to give Infrared vibrational frequencies in better agreement with standard experimental values than did PM3 and that PM3 failed to give accurate predictions for nitrogen - containing compounds [19]. Paola Scano, also from St. Andrews, found that, on average, PM3 was better than AM1 at reproducing heats of formation for a range of large molecules [72].

Other comparisons include the comparison of AM1 and PM3 structures with the X-ray crystallography geometry by Dega - Szafran and co-workers [73] for the molecule 1-benzyloxy-4-methoxypyridinium perchlorate. Although both methods failed to reproduce the trans conformation of the molecule, PM3 reproduced bond lengths and angles to a slightly better degree than AM1.

Since the various reports have given conflicting conclusions, it was decided to evaluate the superiority of one or other method when applied to the steroid hormones

studied in this work, which in general are larger than those used in previous studies. This study was aimed at determining which of the methods reflect both geometries and calculated properties most accurately.

(A1) Methods used in the comparison.

This analysis has been divided into two sections. Firstly, a study of how well the semi empirical methods reproduce Ab Initio calculations is presented. To analyse the differences in geometry between AM1 and PM3 two methods have been used. Since geometries of molecules as large as steroids cannot be optimised using Ab Initio methods, some representative model systems have been chosen. These are two - ringed sections of steroid hormones, and highlight various chemical systems, including 5- and 6-membered rings, double bonds, hydroxy and carbonyl groups. Such systems containing about a dozen heavy and about twenty light atoms are within the scope of optimisations using extended bases. The final structures obtained from the Gaussian programs [37,51] have been compared, using the geometry comparison section of the molecular graphics program presented in Chapter Four.

A study of complete steroids has also been made, but this time single point calculations (i.e. with no geometry optimisation) were performed upon the AM1 and PM3 structures for the hormones. The Variation Theorem described in Chapter Two allows us to determine which is the better conformation because it will have the lowest energy. In the same vein, a comparison has been made by Dewar and co-workers of the difference between AM1 and 3-21G optimised geometries, using large basis set and correlated Ab Initio methods . This study found that AM1 was superior in many cases [137]. As a prelude to this part of the study, it was necessary to select a basis set with which to perform the Ab Initio calculations. This choice was based upon studies performed upon two model compounds, selected to represent various structural features of the steroids. Single point calculations were performed for the two molecules with

various bases, to determine the basis set dependence of the energy difference between the AM1 and PM3 structures.

The second comparison is that of AM1 and PM3 with experimental results (including geometries obtained from the Daresbury Crystal structure Database). Measurements of dipole moments for a few steroids have been reported [138,139], and these have been compared with calculated values.

Since the evaluation of the electrostatic potential calculated using Mulliken Analysis charges has played a key role in the rationalisation of biological activity in this work, an analysis of these charges, and those calculated from the wave function would be useful.

Although the dipole moment is very sensitive to variations in charge distribution and a comparison of calculated and experimental dipoles is presented, individual atomic charges can be assessed using ^{13}C NMR chemical shifts. Such a study has already been made by Repmann for nine steroids [140] using the Extended Huckel MO (EHMO) method. This analysis relies upon the fact that the ^{13}C NMR chemical shift depends upon the amount of electronic screening the nuclei experience due to the local electron density. The latter is correlated with atomic charge, and hence a correlation between charges and chemical shifts arises.

The chemical shift, δ, at a fixed frequency is given by

$$\delta = \frac{H_A - H_0}{H_0} \qquad 7.1$$

Ho is the field strength at which resonance occurs for some standard, and Ha is the corresponding field for a given nucleus, A. The total field experienced by nucleus A, is given by [141]

$$H_A = H_0 (1 - \sigma_A) \qquad 7.2$$

Where σ_A is the magnetic screening constant, which is the sum of several terms:

$$\sigma_A = \sigma_d^{AA} + \sigma_p^{AA} + \sum_{B \neq A} (\sigma^{AB} + \sigma^{A,\,deloc}) \qquad 7.3$$

σ_d^{AA} results from the diamagnetic shielding by electrons of the inner shells of atom A, σ_p^{AA} corrects for the deviation from spherical symmetry of these electrons [142]. σ^{AB} is the effect of electrons localised upon other atoms, whilst $\sigma^{A,\,deloc}$ represents the effect of the motion of the electrons delocalised in π - systems.

The contribution of these last two terms is fairly independent of the environment of atom A, normally accounting for 10% or less of σ_A [142]. The local diamagnetic term, σ_d^{AA} is fairly constant for C, N, O and F atoms [143], and thus changes in σ_p^{AA} are almost totally responsible for the differences in ^{13}C chemical shifts due to differing chemical environments.

This factor is proportional to the local electron density around the atom in question, as is the charge as calculated using the method of Mulliken analysis, and hence we can deduce a linear relationship between the calculated charges and the ^{13}C NMR shifts.

It should be noted that electron density is not the only factor involved in determining the chemical shift, but Repmann states that it should be the most important (140), and obtained good correlation between EHMO charges and observed shifts, with an average correlation coefficient of 0.81 for aliphatic carbon atoms, 0.98 for carbonyl and double bonded systems and 0.95 for the aromatic system of estrone.

A similar analysis was performed using AM1 and PM3, as well as a smaller study of AM1 - ESPFIT and Ab Initio Mulliken charges.

(B) Comparison of AM1 and PM3 by evaluation of the Ab Initio energy for the optimised geometry.

The object of this study is to produce AM1 and PM3 optimised geometries for a number of steroids, and to calculate the Hartree - Fock energy (E_{HF}) for both. If one method is consistently able to give a lower value (more negative) of E_{HF}, the results will indicate that the method is best able to reproduce Ab Initio structures. Such a study requires the effects of electron correlation to be equivalent in the two semi empirical structures, and thus may not truly represent the real physical situation.

(B1) Choice of Basis Set

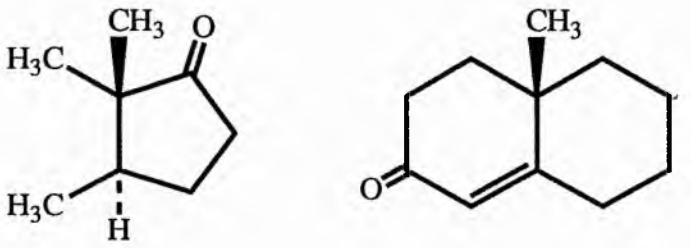
In the Dewar study [137], AM1 and 3-21G geometries were compared by evaluating the energies of the two structures at the 6-31G, and 6-31G* levels, using both Hartree - Fock theory and Moller - Plesset perturbation theory [21] to assess the contribution of correlation effects.

The resultant energies implied that AM1 geometries were at least as good, if not better, than those using the 3-21G basis set, especially if correlation is taken into account.

In this study, we want to know which is the smallest basis set that will give a valid representation of large basis set calculations in order to minimise the computational time needed to evaluate the Ab Initio energies for the steroids. These systems have about 50 atoms (including around 20 non-hydrogen atoms), and hence large basis calculations are time consuming to perform.

(B2) Method of selection.

Two model systems based upon sections of the backbone of steroids, which were small enough to allow extensive Ab Initio calculations were selected. Full AM1 and PM3 optimisations were performed upon these molecules, whose structures are shown below in Figure 7.1.



Model 1 : $C_8H_{14}O$ Model 2 : $C_{11}H_{16}O$

Figure 7.1.

Model compounds used in Ab Initio energy evaluations of AM1 and PM3 structures.

Model one was chosen to represent the D-ring system of the steroids (see Figure 1.1). The keto group above, at the position equivalent to Carbon-17 (in the standard steroid numbering system) was selected because it is present in all the most active inhibitors of G6PDH. Two methyl equatorial groups were also included in the system to simulate the steric effects of the C-ring system.

Model two highlights the A- and B-ring system of steroids, and is especially relevant to the structures of the Aromatase inhibitors and substrates studied in Chapter Six. Again, this has a keto system, and it also exhibits a carbon-carbon double bond, examples of which can be seen in the steroidal inhibitors of both enzymes mentioned above.

The basis set used in the calculations are some of those readily available within the Gaussian and KGNMOL programs, and cover a broad spectrum of qualities.

(B3) Results.

The Ab Initio energies calculated using various bases are shown in Tables 7.1 and 7.2, along with the differences between the two energies corresponding to the AM1 and PM3 geometries. The former are quoted in Atomic Units (AU), as these are the results taken directly from the Gaussian or KGNMOL calculations. The energy differences are given in kcal mol^{-1} to put the results in perspective with the calculated values of ΔH_F obtained from MOPAC. This calculation is performed by subtracting the AM1 value from that for PM3. This number is then multiplied by 627.47237, which is the conversion factor between Hartrees and kcal mol^{-1}. Hence ΔE = 627.47237 (E_{PM3} - E_{AM1}), and negative values imply that the PM3 structure has a lower energy than that calculated using AM1.

Basis Set	E_{AM1}/ AU	E_{PM3}/ AU	ΔE / kcal mol^{-1}
STO-3G	-381.266	-381.279	-8.18
3-21G	-383.816	-383.826	-6.76
Geosmall	-384.822	-384.804	+11.38
Geosptv	-385.378	-385.389	-6.88
Geomedium	-385.754	-385.763	-5.69
6-31G	-385.788	-385.798	-6.86
D95	-385.836	-385.847	-7.11
Geolarge	-385.890	-385.902	-7.16
6-31G**	-385.974	-385.986	-7.66
D95**	-386.038	-386.050	-7.54

ΔH_F (AM1) = -65.88 ΔH_F (PM3) = -69.78

Table 7.1

Ab Initio Energies for Model One.

Basis Set	E_{AM1}/ AU	E_{PM3}/ AU	ΔE / kcal mol^{-1}
STO-3G	-494.644	-494.654	-6.08
3-21G	-497.937	-497.944	-4.33
Geosmall	-499.267	-499.249	+11.30
6-31G	-500.506	-500.512	-3.92
Geolarge	-500.633	-500.639	-4.23
6-31G*	-500.716	-500.724	-4.73
D95*	-500.791	-500.798	-4.64

ΔH_F (AM1) = -53.30 ΔH_F (PM3) = -52.67

Table 7.2

Ab Initio Energies for Model Two

(B4) Discussion.

Most of the basis sets used in the study of Model One yield negative values for ΔE in the region of six or seven kcal mol^{-1}, implying that the PM3 structure has a lower energy than AM1. An exception to this is the Geosmall [28] basis, with a values of ΔE of +11.38. This is a minimal basis set, although it is of quite good quality, yielding a lower energy than the 3-21G [29] split valence calculation.. It is interesting to note the STO-3G minimal basis set [26] agrees with the bulk of the results in giving a lower value for the energy of the PM3 conformer, although the size of ΔE in this calculation is larger than for the others.

For Model Two the range of basis sets used is smaller as this study is to confirm the general conclusion for Model One. It can be seen that most of the basis sets show that the PM3 structure is of lower energy, this time being about 4 kcal mol^{-1} lower. Again the Geosmall basis predicts the opposite result from the general consensus, and STO-3G yields a value of ΔE that is more negative than do the others.

(B5) Conclusion.

The results in Tables 7.1 and 7.2 demonstrate that the STO-3G basis set, although yielding values of ΔE slightly greater than the modal average, is quite capable of reproducing the trends of the larger, more accurate calculations for the molecules in question. Thus, in the interests of saving computer time, it was this basis set that was used in the comparison of the complete steroid molecules.

(B6) Ab Initio Study of Complete Steroid Molecules.

A number of steroid geometries generated using the AM1 and PM3 methods were then used as the basis for calculations with the Gaussian programs using the

STO-3G basis set. As above, the energies of the two structures were calculated, and the differences between the two values were evaluated. In addition, to test the result stating that the STO-3G basis set is adequate for this study, one of the steroids (DHEA) was used as the basis for a number of calculations with a few larger bases.

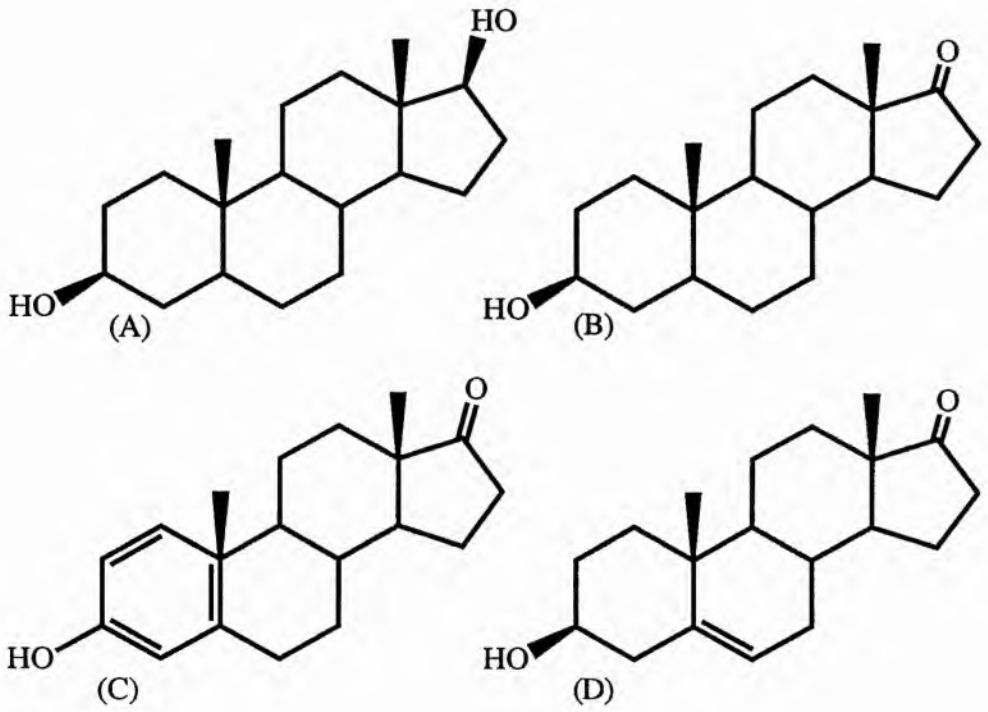


Figure 7.2.

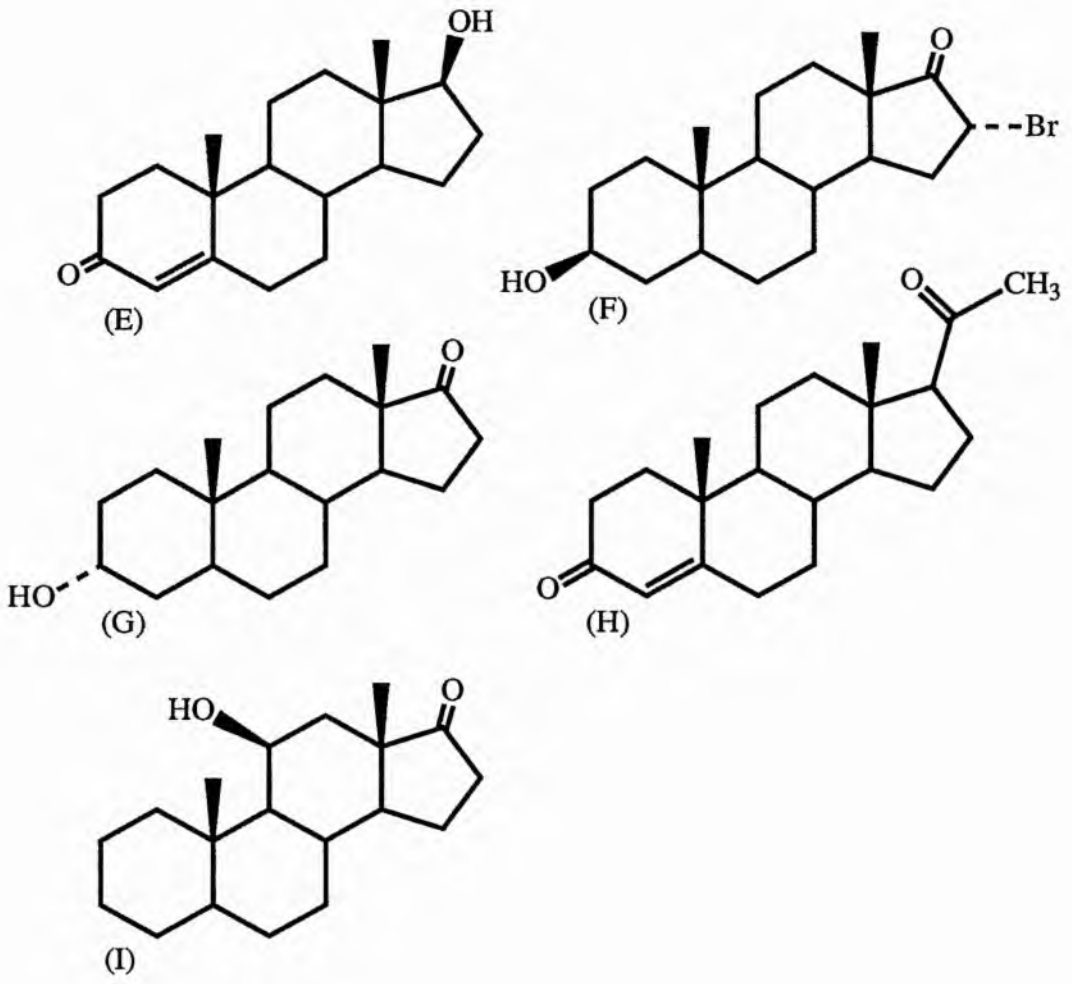


Figure 7.2 (Cont.).

(A) Androstanediol

(B) Epiandrosterone

(C) Oestrone

(D) DHEA

(E) Testosterone

(F) 16α-Bromoepiandrosterone

(G) 3α-hydroxyandrostane-17-one

(H) Progesterone

(I) 11β-hydroxyandrostane-17-one

(B7) Results of the Ab Initio Work.

The results of this study are presented in Table 7.3, below, whilst the calculations upon DHEA using various basis sets are detailed in Table 7.4. The units used are the same as in Tables 7.1 and 7.2, with the absolute values of E in AU and the values of ΔE in kcal mol^{-1}.

Steroid	E_{AM1}	E_{PM3}	ΔE
11β-hydroxyandrostane-17-one	-875.997	-876.009	-7.38
16α-Bromoepiandrosterone	-3420.088	-3420.094	-3.25
3α-Hydroxyandrostane-17-one	-876.001	-876.012	-6.91
Androstanediol	-877.176	-877.186	-6.28
DHEA	-874.787	-874.798	-6.74
Epiandrosterone	-876.000	-876.012	-7.16
Oestrone	-833.869	-833.877	-5.03
Progesterone	-950.762	-950.774	-7.51
Testosterone	-874.789	-874.797	-4.99

Table 7.3

STO-3G Energies for AM1 and PM3 Steroid Structures.

Basis Set	E_{AM1}	E_{PM3}	ΔE	Approx Time
STO-3G	-874.009	-874.798	-6.74	6 hours
3-21G	-880.593	-880.600	-4.28	10 hours
Geosmall	-883.040	-883.009	19.81	2 days
6-31G	-885.118	-885.125	-4.30	1 day

Table 7.4

Ab Initio Energies for DHEA using a Small Selection of Basis Sets.

(B8) Discussion.

All the values of ΔE in Table 7.3 are negative. This demonstrates the superiority of PM3 geometries over those generated using AM1.

The values of ΔE in Table 7.4 confirm that the selection of the STO-3G basis set was an adequate choice for this study, although as before, the results calculated using the Geosmall basis is in disagreement with the others.

(B9) Conclusion.

The approximate CPU times for the calculations are also shown in Table 7.4, and refer to Gaussian88 calculations performed upon the FPS-500. The more recently acquired Gaussian90 gives a considerable improvement upon these times. These emphasize the need for powerful computing resources when undertaking a study of this class of molecule, and also demonstrates the benefits of using as small a basis set as possible.

This study has clearly shown the superiority of the PM3 calculations in giving lower energies in Ab Initio Hartree - Fock calculations. Nevertheless, it is not known whether PM3 would be better than AM1 at predicting the results of calculations involving the treatment of electron correlation, and further work would be needed to discover to confirm this.

(C) Comparison of Semi Empirical and Ab Initio Structures.

This is the second method of evaluating which of the two semi empirical methods used gives the more reliable results using Ab Initio techniques. It involves the selection of model structures based upon steroid molecules and their optimisation using AM1 and PM3 in MOPAC and Ab Initio methods in Gaussian88 and 90.

The model compounds chosen were selected to give a good representation of the variety of important regions within the steroids. They highlight such features as two six - membered rings, keto - groups, hydroxy - groups, carbon - carbon double

bonds, and the interaction of double bonds with the oxygen - containing groups. Other functionalities include the study of a double ring system containing six and five members, and the effects of keto and hydroxy - groups upon these. Also considered is an aromatic - saturated double ring system, as found in oestrogens and the O=C-C=C-C=C system found in some of the Aromatase inhibitors discussed in Chapter Six.

A further interaction of interest is that found between a Bromo moiety and an adjacent keto - group, as found in the most potent steroidal inhibitor of G6PDH discussed in Chapter Five, 16α-Bromo epiandrosterone. It is important to obtain a reliable structure for this molecule because it is used as a lead compound in the ASP calculations performed on the G6PDH inhibitors, and hence is used as a standard in the similarity study. Limitations of the basis sets available as standard within the Gaussian programs and problems with the section handling the core - potential approximation made the study of bromine - containing molecules with anything except the STO-3G and STO-3G* bases difficult. Hence, use was made of the general basis option within Gaussian, which allows the user to specify any basis set they wish. The bases chosen to study this model compound were Clementi's Geometric Basis sets [28], because they are present as standard within the KGNMOL program, with which the preliminary studies were made. Limitations of disk space and the lack of an optimisation routine in the program made the use of the Gaussian programs necessary. These were not designed to handle this type of basis set efficiently, and hence these calculations are extremely slow to perform.

A second, smaller study involves calculations upon the molecule 4,4-dimethyl Androstan-3-one (4,4-DMA). A theoretical structure for this steroid has been presented in the literature [136], which was generated using a complete geometry optimisation using the 4-21G basis set, and a method known as the Multiplicative Integral Approximation (MIA).

The MIA is used in evaluating two - electron integrals, in which the Gaussians in the integrals are expanded in terms of secondary functions. The errors involved in the process are estimated and corrected for if they are greater than a cut - off level.

The details of this calculation are given in the aforementioned paper, but the approximation is said to yield results that are equivalent to the full SCF calculation, but in a much shorter time, hence enabling a full optimisation upon this large molecule.

The AM1 and PM3 geometries have been calculated for this molecule using MOPAC, at the Precise level of accuracy, and these have been compared using the compare routine of the MHCDraw graphics program presented in Chapter Three.

(C1) Model Systems Used in the Study.

The molecules studied in this section of the work are shown in Figure 7.3, which is followed by a brief description of the important regions in these systems.

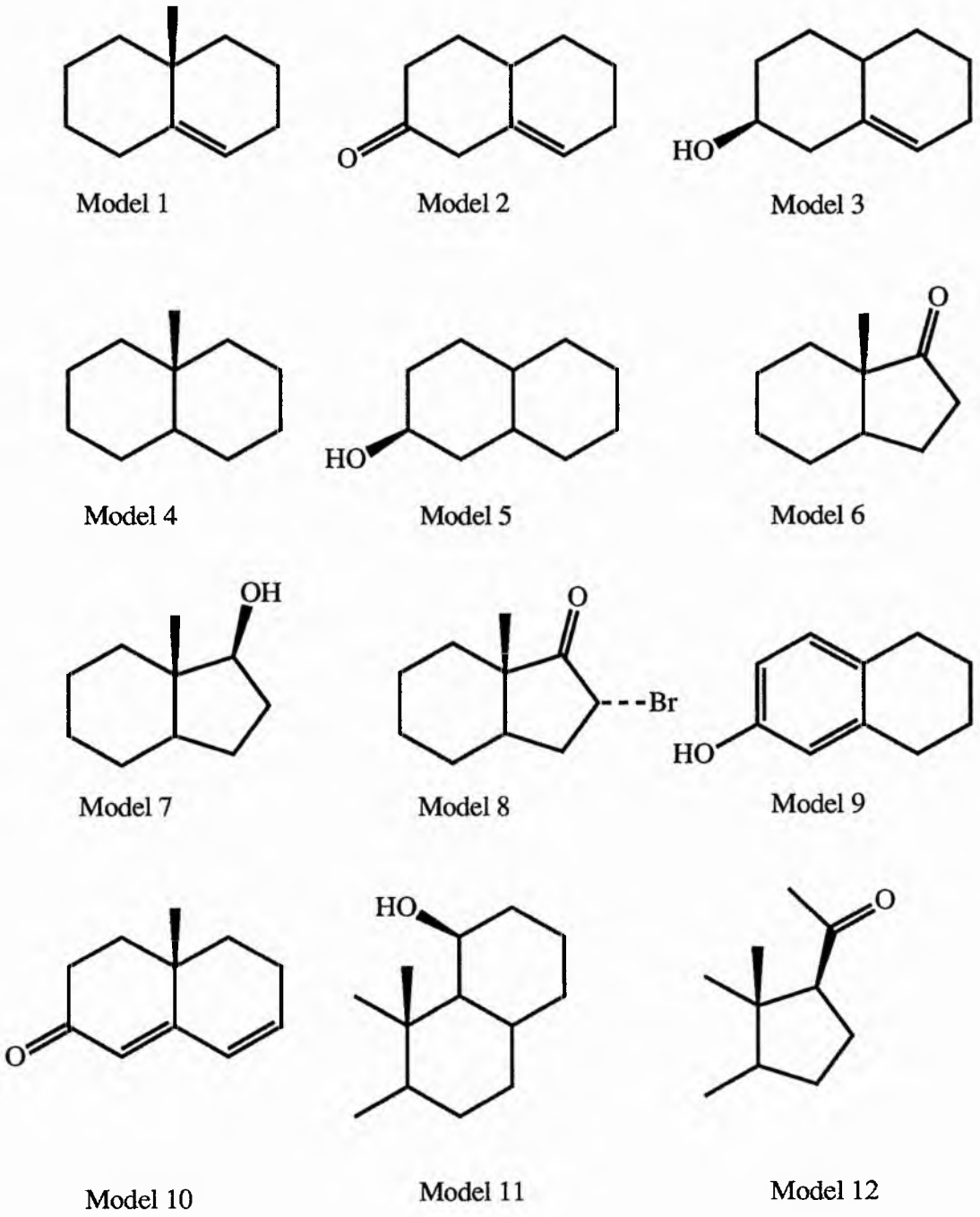


Figure 7.3

Steroid Hormone Model Compounds used in the Ab Initio Geometry Study.

Model	Formula	Comments
1	$C_{11}H_{18}$	Same as A and B ring system of DHEA.
2	$C_{10}H_{14}O$	Δ^5 Double bond and keto - system.
3	$C_{10}H_{16}O$	Beta hydroxy group and Δ^5 double bond.
4	$C_{11}H_{20}$	Saturated A and B rings.
5	$C_{10}H_{18}O$	Hydroxy group without double bond.
6	$C_{10}H_{16}O$	Same as C and D rings of steroids.
7	$C_{10}H_{18}O$	Androstanediol model (poor inhibitor of G6PDH).
8	$C_{10}H_{15}Br$	Bromo epiandrosterone model.
9	$C_{10}H_{12}O$	Oestrogen A and B ring model.
10	$C_{11}H_{24}O$	7-substituted aromatase inhibitor model.
11	$C_{13}H_{24}O$	11β-hydroxy steroid model.
12	$C_{10}H_{18}O$	Model for Pregnane - like inhibitors of G6PDH.

(C2) Methods Used in the Study.

Each of the model compounds listed in the table was optimised using both the AM1 and PM3 methods to the "PRECISE" level. The AM1 structure was then taken as the starting point for the Ab Initio study, and all the compounds were optimised at the 3-21G level. The work of Dewar and co-workers [137] indicates that according to higher quality calculations, AM1 structures often yielded lower energies in single point calculations than did their 3-21G counterparts, some of these compounds (numbers 1, 3, 6 and 9) have also been optimised at the higher 6-31G level.

Once all the structures had been computed, the results of the MOPAC semi empirical calculations were compared to those obtain from the Gaussian programs. This was done in a similar manner to the Geometry Optimisation study presented in Chapter Four, and uses the molecular graphics program, MHCDraw.

(C3) Results : Steroid Model Compounds.

Table 7.5, below, indicates the results of the comparison of 3-21G geometries with those from AM1 and PM3 calculations. The tabulated values are the average of the results of the evaluation of semi empirical geometry parameter minus the 3-21G geometry parameter for each of the model compounds. The column headings are as follows : $\overline{\Delta L}$(AM1) and $\overline{\Delta L}$(PM3) are the average differences in bond length for AM1 and PM3 calculations respectively. $\overline{\Delta\theta}$ represent the average bond angle differences, $\overline{\Delta\tau}$ dihedral and $\overline{\Delta c}$ co-ordinate differences respectively.

	$\overline{\Delta L}$		$\overline{\Delta\theta}$		$\overline{\Delta\tau}$		$\overline{\Delta c}$	
Molecule	AM1	PM3	AM1	PM3	AM1	PM3	AM1	PM3
1	0.031	0.020	0.5	0.9	2.2	3.2	0.020	0.029
2	0.030	0.020	0.8	1.0	3.8	2.5	0.024	0.019
3	0.028	0.021	0.7	0.9	209	2.0	0.027	0.022
4	0.032	0.032	0.5	0.5	2.1	2.1	0.02	0.020
5	0.030	0.022	0.6	0.8	4.7	4.4	0.037	0.034
6	0.029	0.019	0.6	0.9	3.5	3.3	0.027	0.031
7	0.029	0.021	0.7	0.9	4.1	5.1	0.029	0.035
8	—	—	—	—	—	—	—	—
9	0.026	0.019	0.7	0.9	1.6	1.2	0.02	0.020
10	0.027	0.017	0.7	0.8	3.2	3.2	0.02	0.016
11	0.029	0.019	0.7	0.9	5.5	5.6	0.056	0.037
12	0.029	0.017	0.8	1.0	7.3	5.9	0.064	0.079

Table 7.5

The basis set limitation on calculations involving Bromine have necessitated the exclusion of Model Compound 8 from the above table. The results of the comparison

to STO-3G and STO-3G* geometries are indicated in Table 7.6, below, in which the symbols used have the same meaning as those in Table 7.5

	$\overline{\Delta L}$		$\overline{\Delta\theta}$		$\overline{\Delta\tau}$		$\overline{\Delta c}$	
Basis	AM1	PM3	AM1	PM3	AM1	PM3	AM1	PM3
STO-3G	0.029	0.021	0.5	1.2	3.6	5.8	0.031	0.055
STO-3G*	0.031	0.022	0.5	1.2	3.5	5.6	0.028	0.052

Table 7.6

The results of the 6-31G comparison are shown in Table 7.7

	$\overline{\Delta L}$		$\overline{\Delta\theta}$		$\overline{\Delta\tau}$		$\overline{\Delta c}$	
Molecule	AM1	PM3	AM1	PM3	AM1	PM3	AM1	PM3
1	0.028	0.018	0.6	0.7	1.1	3.1	0.020	0.029
3	0.026	0.018	0.7	0.8	1.6	1.4	0.024	0.020
6	0.027	0.017	0.7	0.8	3.1	3.7	0.027	0.031
9	0.023	0.016	0.6	0.7	0.7	0.4	0.016	0.016

Table 7.7

(C4) Discussion : 3-21G and Model 8 Comparison.

The overall average of the differences in bond lengths values for are 0.029Å and 0.021Å for AM1 and PM3 respectively, and the PM3 values are always smaller than the corresponding AM1 values.

The situation for bond angles is reversed, with averages of 0.65° and 0.85° respectively, whilst for dihedral angles, the case is not so clear cut, with both methods showing superiority in some cases. The average indicates a slight improvement in the

PM3 results over those calculated with AM1, with values of 3.36° and 3.26°. The overall average, indictaed by the average of the differences in co-ordinates, for AM1 is 0.0274Å and 0.0263Å for PM3.

The 3-21G optimisation of Model 12 was particularly time consuming to perform, and the results indicate a greater similarity with AM1. In particular the important dihedral angle in the molecule, defining the conformation of the side chain corresponding to the 17 position of a complete steroid shows a closer fit to the AM1 value. The Ab Initio value of this variable (equivalent to O=C-C_{17}-C_{13}) is 80.5°, whilst the semi empirical results are 88.2° and 100.7° for AM1 and PM3 respectively.

For Model Compound 8, the STO-3G averages for AM1 are 0.0291Å, 0.5°, 3.6° and 0.0306Å for bond lengths, angles, dihedrals and co-ordinates respectively. The equivalent PM3 values are 0.0208Å, 1.2°, 5.8° and 0.0552Å, indicating that the AM1 structure shows a greater similarity the the STO-3G geometry than does that calculated with PM3.

Using STO-3G*, the equivalent values for AM1 are 0.0306Å, 0.5°, 3.5°, 0.0281Å, whilst for PM3 they are 0.0222Å, 1.2°, 5.6° and 0.0524Å. This again indicates the greater similarity of the AM1 geometry to the Ab Initio structure.

A Geosmall optimisation of this compound has also been performed, and a comparison with the semi empirical results gives the following for AM1 and PM3 respectively : $\overline{\Delta L}$ = 0.057Å, 0.063Å ; $\overline{\Delta\theta}$ = 0.7°, 1.2° ; $\overline{\Delta\tau}$ = 4.8°, 5.7° ; $\overline{\Delta c}$ = 0.090Å, 0.101Å. Hence, the AM1 structure is significantly closer to the Geosmall geometry than is the PM3 structure. In both cases, the Ab Initio bond lengths are longer than those calculated using the semi empirical methods, indicating that this basis set is perhaps less suited for the calculation of structures than it is of energies, since the more common basis sets used do not disagree to such an extent over the bond lengths.

The Br-C_{16}-C_{17}-O dihedral angle, which is also of interest, is much closer to the AM1 value than the PM3 value, which are 64.6° and 78.1° respectively, and the Geosmall value is 60.7°.

(C5) Discussion : 6-31G Comparison.

The 6-31G results indicate that the PM3 method produces bond lengths that are superior to the AM1 ones. The average value for $\overline{\Delta L}$ for the AM1 calculations is 0.026Å, whilst for PM3, it is 0.017Å for the compounds studied at this level. This situation is reversed for the bond angles, with the respective AM1 and PM3 averages being 0.65° and 0.75°. The case for the dihedral angles shows no clear superiority for either method, although the averages are 1.625° and 2.15°, implying that the AM1 dihedrals are closer, on average, to the 6-31G ones than are those calculated by PM3. Overall, the average values for $\overline{\Delta c}$ are 0.021575Å for AM1 and 0.02238Å for PM3, showing a slight superiority for the AM1 method.

(C6) Results : 4,4-DMA Comparison.

The results of the comparison between the AM1, PM3 and published geometries are presented in the graphs in Figures 7.4 (AM1) and 7.5 (PM3). In each case, the graphs represent the value of the published 4-21G geometrical parameter minus that of the semi empirical method. Thus negative differences indicate that the semi empirical values of the variables are the larger.

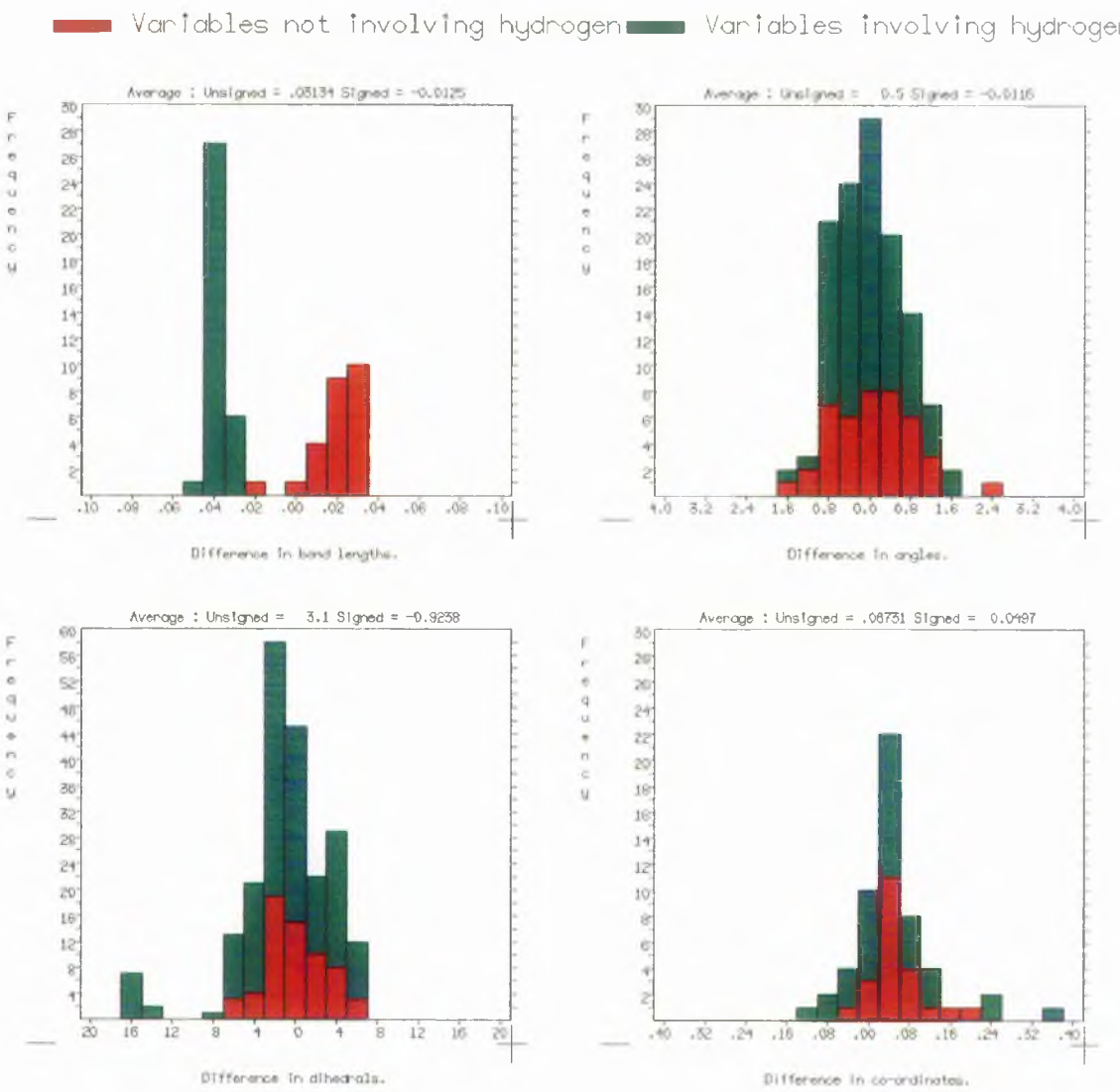


Figure 7.4.
Differences between 4-21G and AM1 Structures.

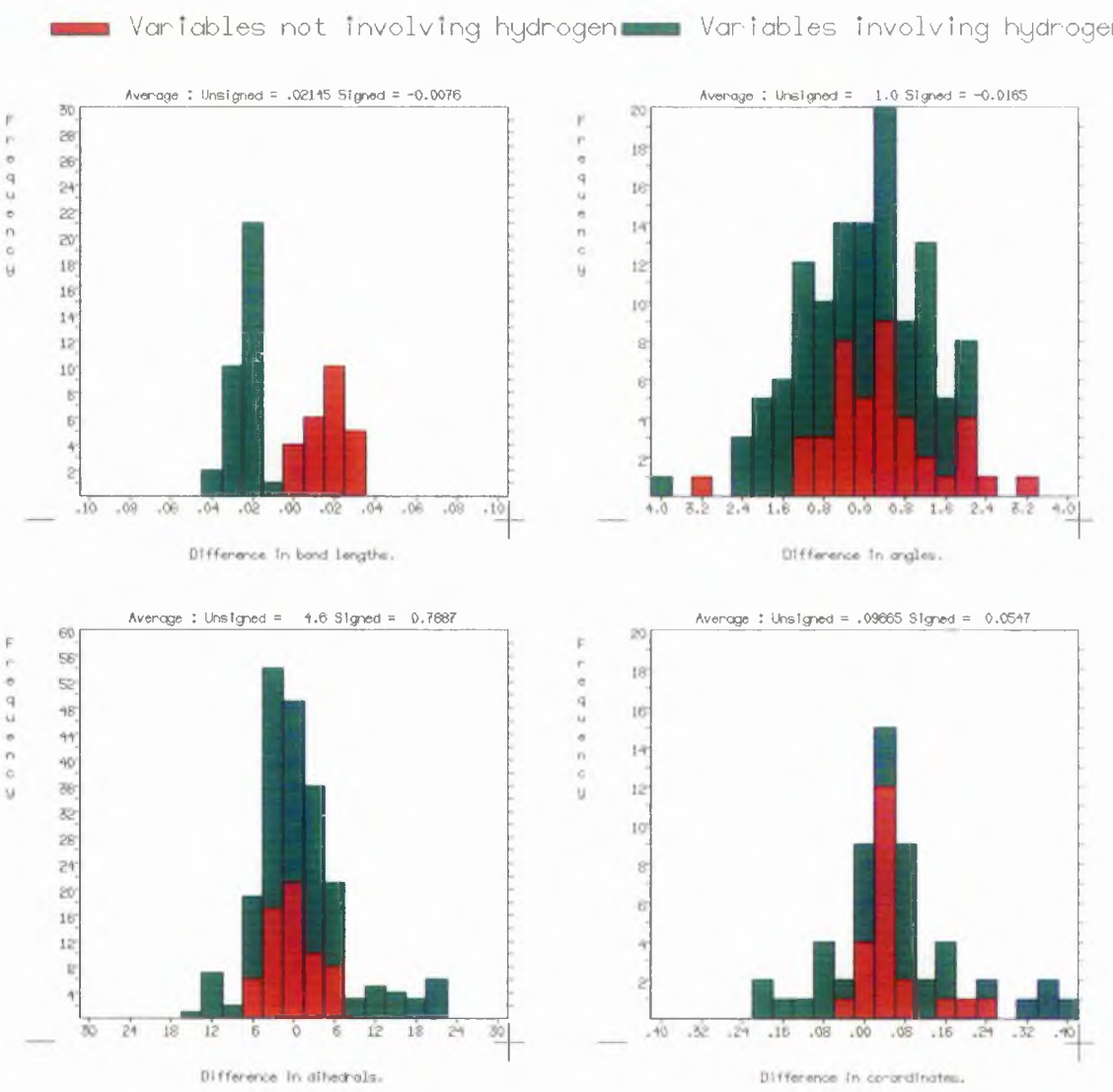


Figure 7.5.
Differences between 4-21G and PM3 Structures.

(C7) Discussion : 4,4-DMA.

The graphs depicted above indicate that the bond lengths for the calculations are fairly close (i.e. within ~0.04Å) to the 4-21G values, with the PM3 average being significantly lower than that for AM1. It is interesting to note that for both cases, the semi empirical methods both predict bonds to hydrogen atoms that are longer than those from the paper, whilst the bond lengths for the non - hydrogen atoms are, on average, shorter.

AM1 is visibly superior at reproducing the 4-21G bond angles, with an average difference that is half the value of the equivalent PM3 result. The AM1 dihedral angles are again superior, which again has a significantly superior average, although in both cases, the distribution of the differences in non - hydrogen dihedrals is fairly similar. The poorly predicted dihedrals are mainly due to the hydrogen atoms upon the β-methyl group attached at position 4, which account for all the differences in dihedrals for the AM1 case which are > 12°.

The difference in co-ordinates distribution shows that overall, the calculations are in favour of AM1, indicating that, in general, atoms in the AM1 structure are closer to the positions predicted by the 4-21G calculation than are those in the PM3 structure.

(D) Comparison of Experimental and Calculated Dipole Moments.

There are a number of experimental values for steroid dipole moments available in the literature, and there have been attempts to calculate these using theoretical techniques [138,139]. Because the dipole moment is derived from the wavefunction and in particular, the charge distribution, the ability of a theoretical method to reproduce experimental results depends upon the accuracy of the wavefunction, and thus the correlation between calculated and experimental dipoles gives an indication of the

accuracy of the calculated wavefunctions. Dipole moments can also be used as a method of predicting biological activity because the interaction energy between two dipoles can be an important part of the overall attraction between two molecules. Hence, it is important that the calculated dipoles used in such a study are reliable.

(D1) Method.

Attempts have been made to reproduce two sets of experimental values for the dipole moments of steroids using the AM1 and PM3 methods. As with other studies, these calculations have been performed with full geometry optimisation at the precise level using the MOPAC program. The dipole moments are calculated automatically by MOPAC, so no further work was necessary to obtain the results.

The molecules in the first study are shown in Figure 7.6, and those from the second are illustrated in Figure 7.7.

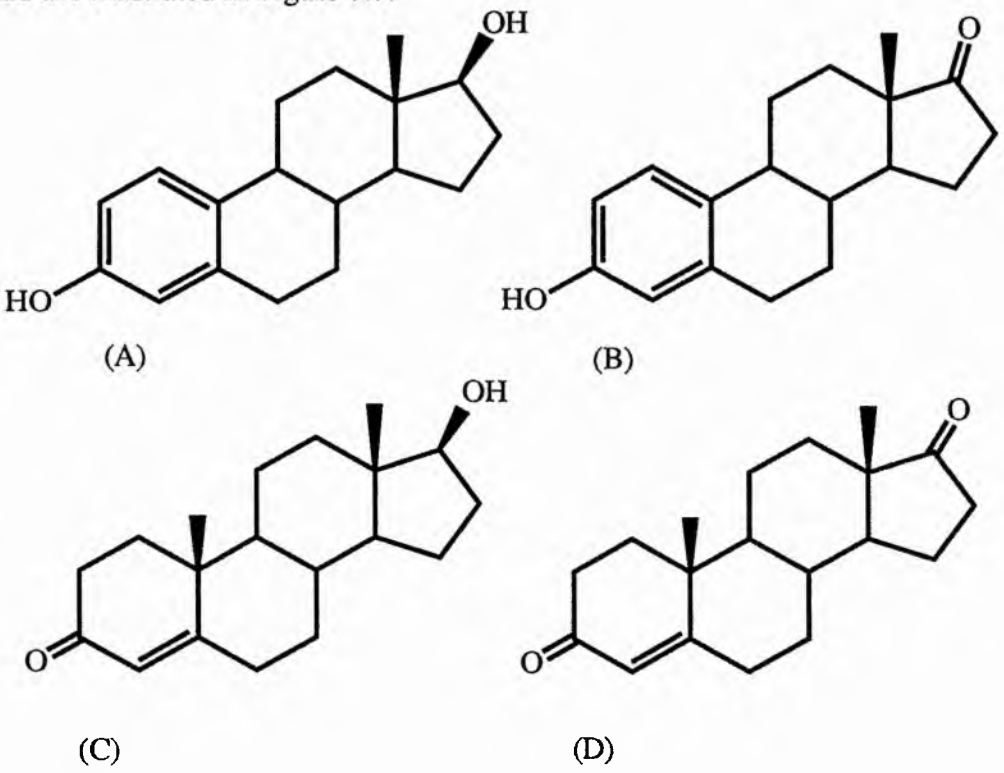
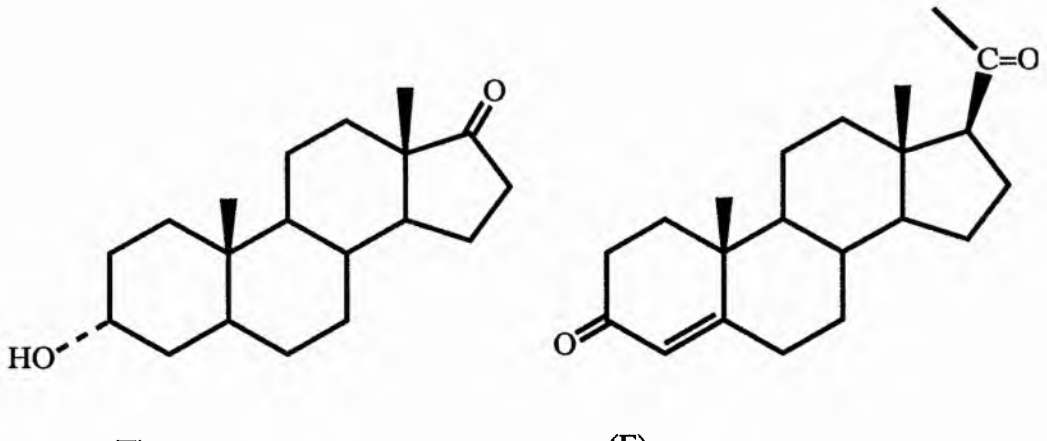


Figure 7.6.



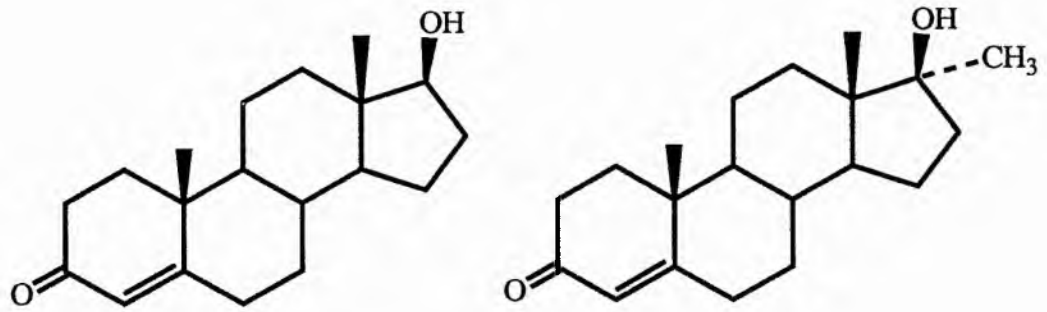
(E) (F)

Figure 7.6 (Cont.).

(A) Oestradiol (B) Oestrone

(C) Testosterone (D) Androst-4-enedione

(E) Androsterone (F) Progesterone



(A) (B)

Figure 7.7.

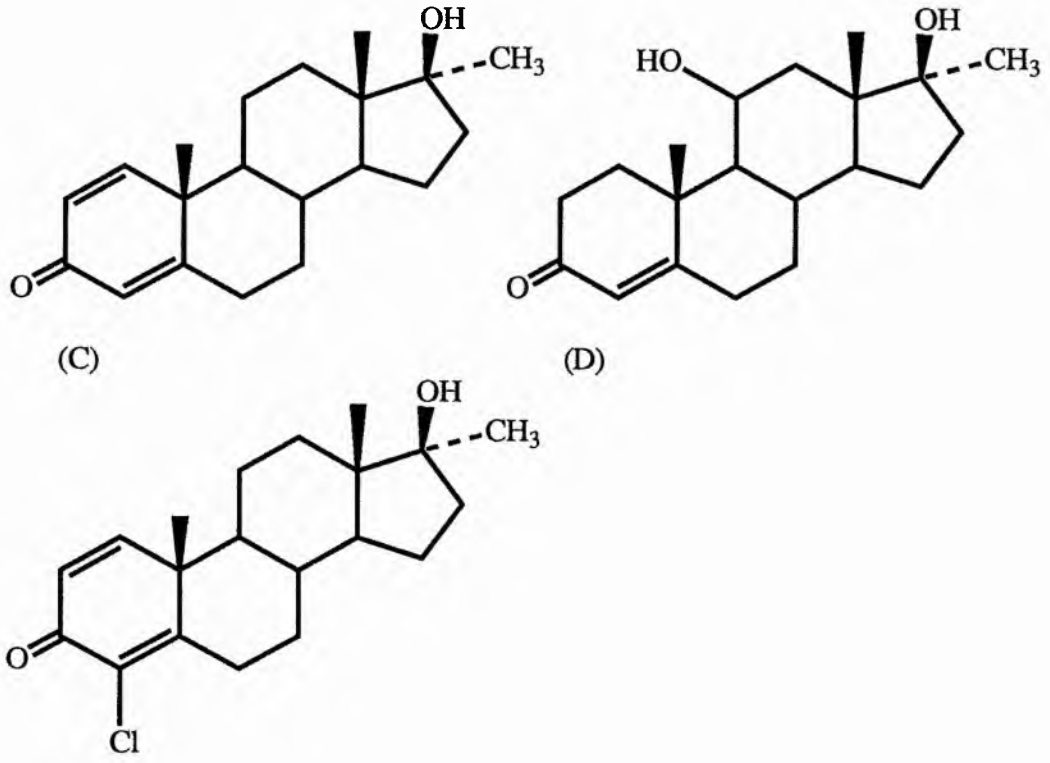


Figure 7.7 (Cont.).

(A) Testosterone

(B) 17α-methyltestosterone

(C) 1-Dehydro-17α-methyltestosterone

(D) 11β-hydroxy-17-α-methyltestosterone

(E) 1-dehydro-4-chloro-17α-methyltestosterone

(D2) Results.

The calculated and experimental results for the first group of molecules are shown in Table 7.8 below.

Molecule	μ (Exp)	μ (AM1)	μ (PM3)
Oestradiol	2.3	0.46	0.97
Oestrone	3.1	1.61	1.68
Testosterone	4.1	3.54	3.34
Androstenedione	3.5	3.48	3.16
Androsterone	3.7	2.87	2.59
Progesterone	2.7	2.27	1.79

Table 7.8

Experimental and Calculated Dipole Moments in Debye.

The results for the second set of calculations are shown in Table 7.9. The molecules are the same as those refered to by letter in Figure 7.7.

Molecule	μ (Exp)[1]	μ (Exp)[2]	μ (AM1)	μ (PM3)
(A)	3.13	4.04	3.54	3.34
(B)	4.14	3.99	3.39	3.06
(C)	4.93	4.84	4.03	3.55
(D)	4.31	4.46	4.20	4.18
(E)	5.69	5.40	4.45	3.84

Table 7.9

Experimental and Calculated Dipole Moments in Debye.

[1] Measured in Dioxane solvent.

[2] Measured in Benzene solvent.

(D3) Discussion.

The results have been plotted upon graphs of Experimental Dipole against Calculated Dipole for the various sets of data, and the graphs are shown in Figures 7.8 to 7.10, below :

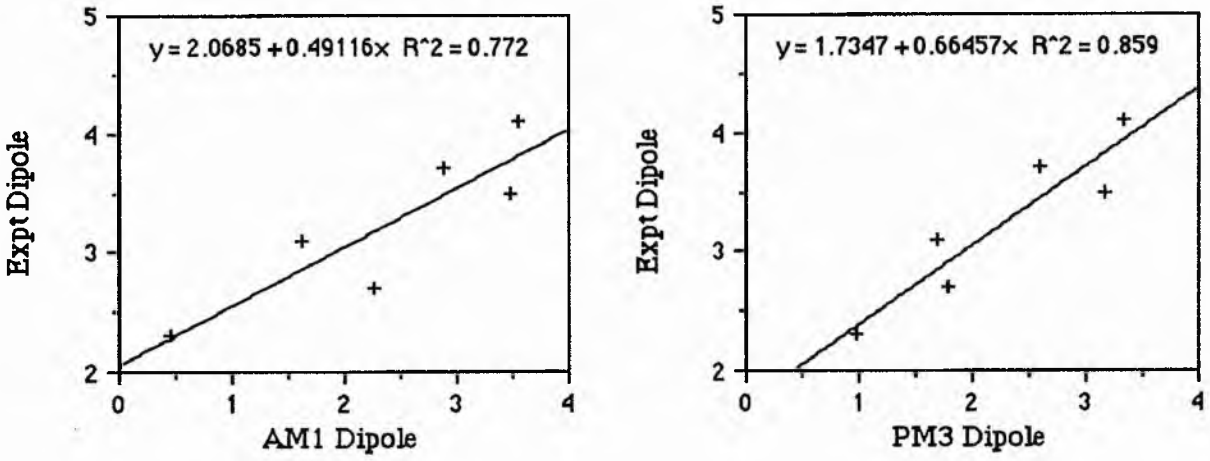


Figure 7.8

Calculated and Experimental Dipoles from First Study.

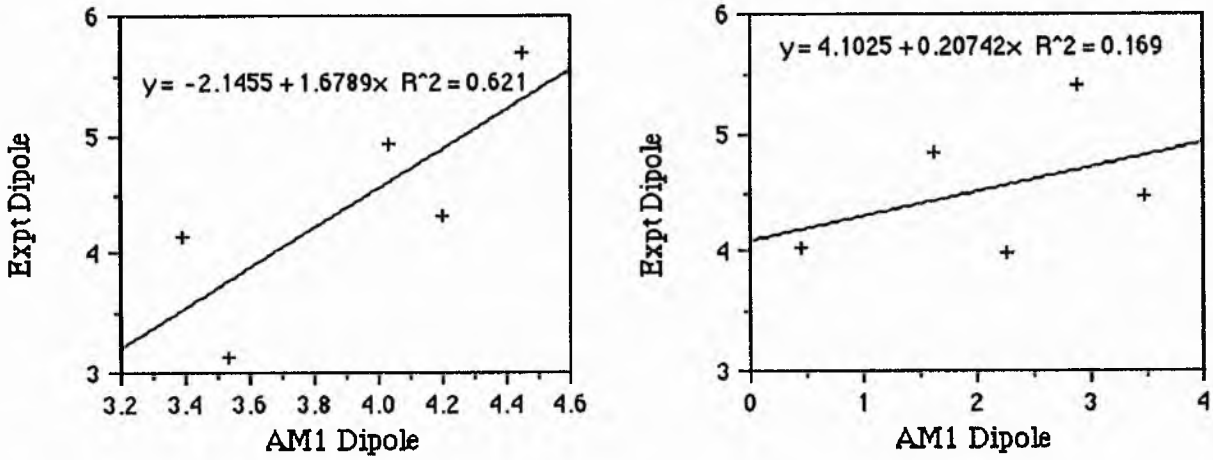


Figure 7.9

AM1 and Experimental Dipoles Measured in Dioxane and Benzene from Second Study.

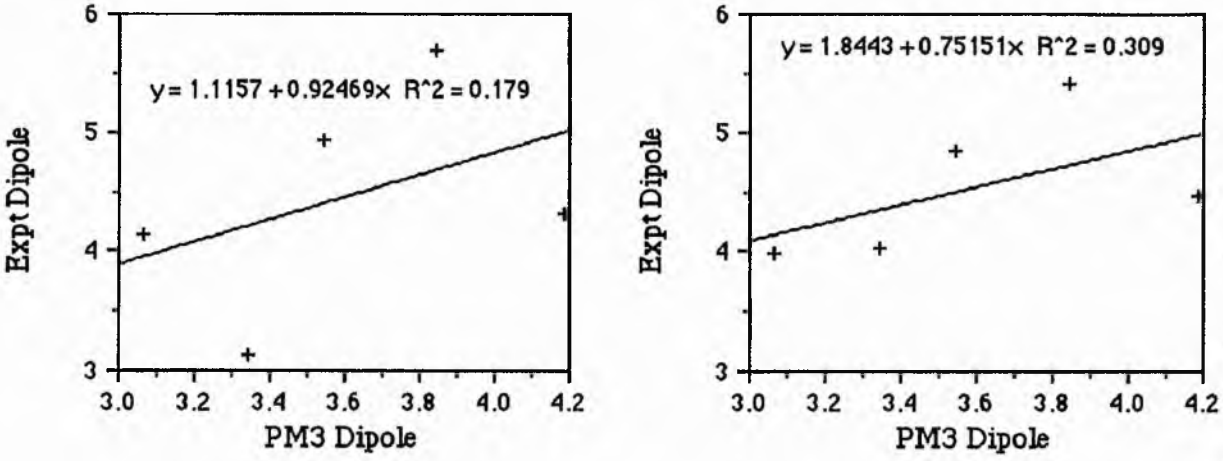


Figure 7.10

PM3 and Experimental Dipoles Measured in Dioxane and Benzene from Second Study

The results from the first study are quite encouraging, with quite high correlation co-efficients, with values of R^2 reasonably close to 1.0 for both methods[1], and the PM3 method is the superior of the two. Nevertheless, the absolute values of the Dipole Moments are not particularly accurate, as evinced by non - zero intercepts on the graphs, and gradients of around about a half. Despite this, the straight - line form of the graphs means that these calculated values can be used in predicting biological activity as has been done in Chapter Five

The results from the second study are less encouraging, with correlation co-efficients that are either moderate or extremely poor. It seems that the AM1 method is better at reproducing the results as measured in Dioxane (i.e. a polar solvent), whilst the results obtained using non - polar Benzene as a solvent are slightly better predicted using PM3, although these results, on the whole, are disappointing.

---

[1]The correlation co-efficient is denoted R^2 in the graphs.

(D4) Conclusions.

The first set of results indicated that both methods are extremely successful at reproducing dipole moments, although the results can not be used directly, and require some mathematical processing to obtain experimental values.

The results attempting to reproduce the second set of experimental data fail to give any significant results, indicating the important role of the solvent in such a study..

(E) Quality of Calculated Ionisation Potentials.

The Ionisation Potential (IP) of a molecule is automatically evaluated during a MOPAC calculation. As described in Chapter Two, it is given by the negative of the energy of the HOMO, and can be obtained experimentally from Photo Electron Spectroscopy (PES) studies. The results of PES studies upon steroids are available in the literature [144], and as with the Dipole Moment, the quality of the calculated results is dependent upon the quality of the wavefunction, and hence can be used as a test of the theoretical methods.

(E1) Methods used in the Study.

As with the Dipole Moment work, all molecules under study were optimised at the Precise level within MOPAC using both the AM1 and PM3 methods, and the calculated IP s were recorded from the output data from the program.

The study was performed upon the molecules shown in Figure 7.11.

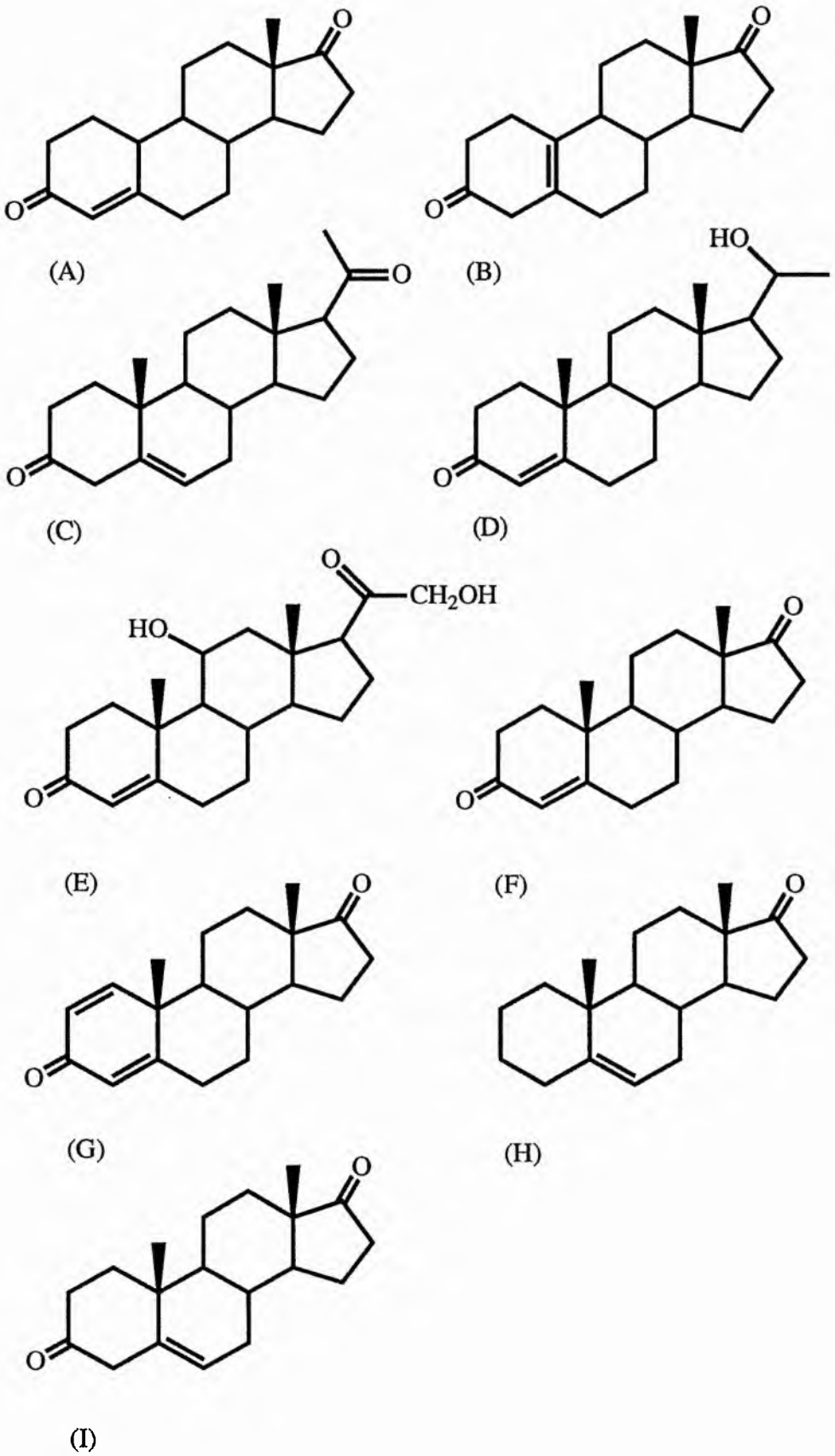


Figure 7.11.

(E2) Results.

The lowest energy peaks from the PES spectrum were used to obtain the energy of the HOMO, and these are shown, along with the calculated values, in Table 7.10, below. These are quoted in Electron Volts (eV), because these are the energy units used by MOPAC for the IP, and in the reference from which the experimental values are taken [144].

Molecule	IP (Exp)	IP (AM1)	IP (PM3)
(A)	8.98	10.08	10.24
(B)	8.28	9.42	9.55
(C)	8.84	9.55	9.71
(D)	8.94	10.04	10.17
(E)	8.78	9.59	9.77
(F)	8.9	10.07	10.25
(G)	8.81	10.10	10.23
(H)	8.6	9.36	9.52
(I)	8.9	9.61	9.75

Table 7.10

(E3) Discussion.

These results are certainly of the right order of magnitude, but when the experimental values are plotted against the calculated values, the correlation between the two is poor, as shown in Figure 7.12, below. In general, the calculated IP s are larger than the experimental ones, and they also cover a larger range of values.

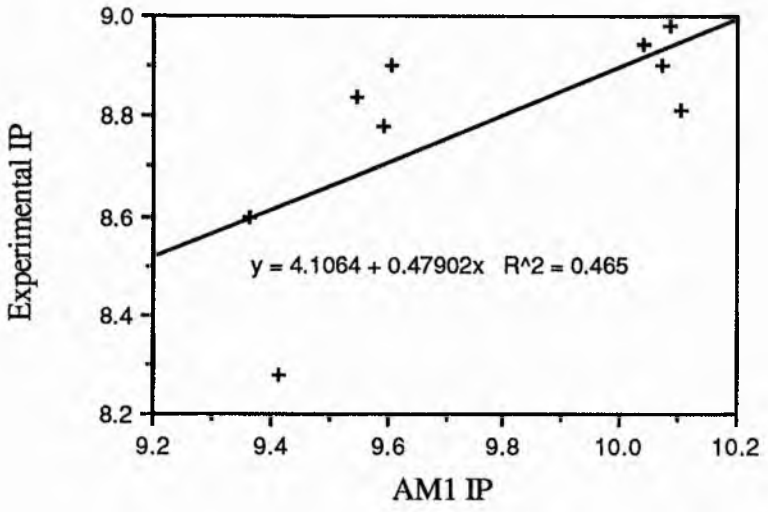


Figure 7.12

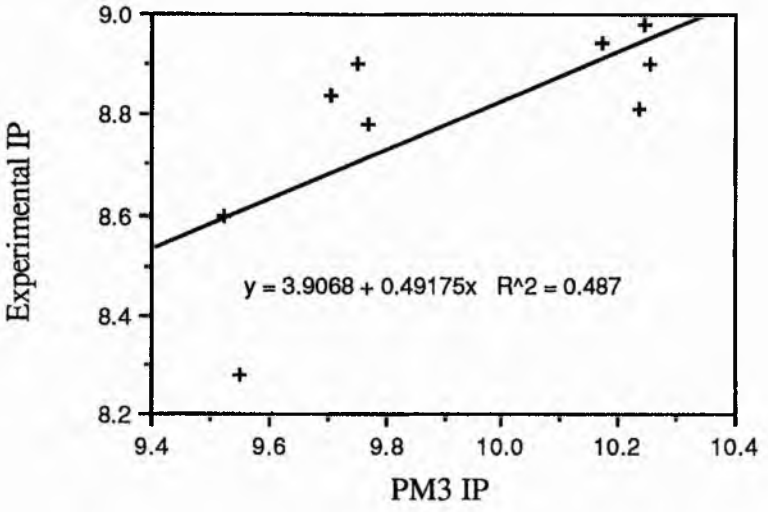


Figure 7.12 (Cont.)

The graphs indicate a very slight superiority of the PM3 method, although both correlation co-efficients are under 0.5. The results are particularly poor for molecule B (the lowest point on the graphs) , and in general, the calculated Ionisation Potential cannot be used to differentiate between the experimental values.

(E4) Conclusions.

On the whole, although the calculated IP s are reasonably close to the experimental values, there is not a simple relationship between the two sets of values, and the correlation is poor. The correlation co-efficient, R^2, is slightly higher for PM3, but the difference is not very significant.

(F) Assessment of Calculated Atomic Charges and Chemical Shifts.

In this thesis, the atomic charges, both calculated using the method of Mulliken Analysis and by fitting the charges to reproduce the electrostatic potential (ESPFIT), have been of paramount importance in the modelling of biological systems. Hence it is necessary that the methods used yield results suitable for use in potential calculations.

One method to assess these charges for the carbon atoms in a molecule, is by comparing them to ^{13}C NMR chemical shifts. A previous study has been made in this area upon steroid hormones [140], as mentioned in the introduction to this chapter.

(F1) Methods.

The nine molecules studied in this section of the research are shown below in Figure 7.13. Each of these was optimised using the AM1 and PM3 methods to the Precise level, and the charges calculated during the Mulliken Population Analysis were extracted and plotted on graphs, along with the chemical shifts. ESPFIT calculations using the AM1 wavefunction were also performed upon a selection of molecules to investigate if potential fitted charges were an improvement over the Mulliken Charges. Finally, the AM1 and PM3 geometries for some of the molecules were used as a basis for Ab Initio calculations using the STO-3G and 3-21G basis sets, in which the charges were again evaluated using the Mulliken method.

for Ab Initio calculations using the STO-3G and 3-21G basis sets, in which the charges were again evaluated using the Mulliken method.

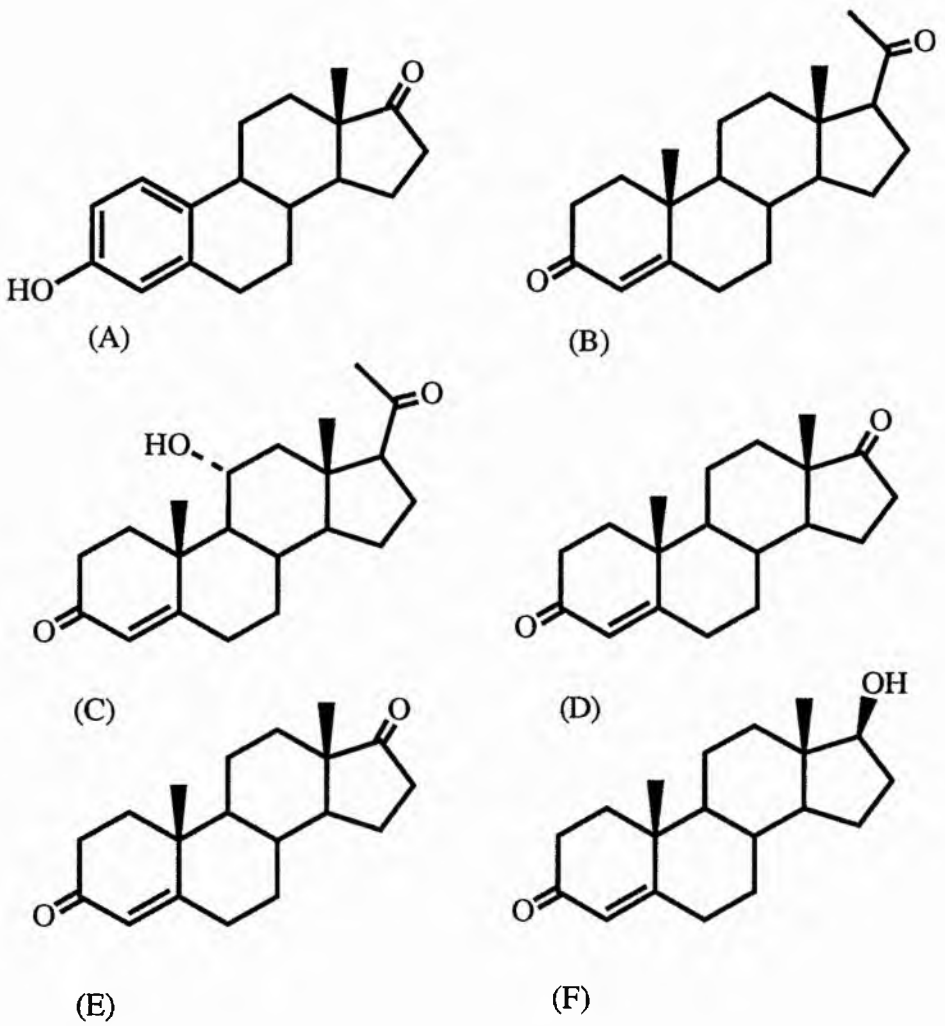
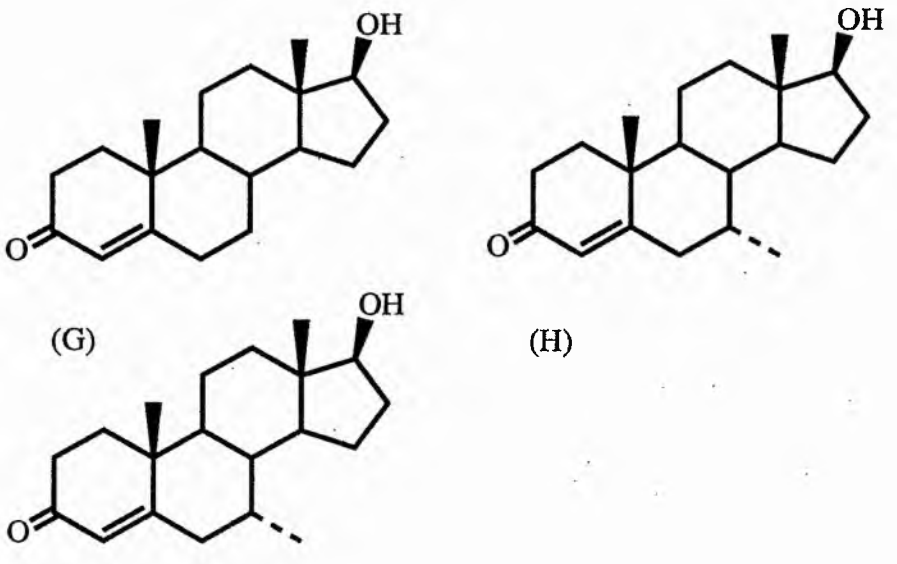


Figure 7.13.



(I)

Figure 7.13 (Cont.).

(A) Oestrone

(B) Progesterone

(C) 11α-hydroxyprogesterone

(D) Androst-4-ene-3,17-dione

(E) 19-norandrost-4-ene-3,17-dione

(F) 19-nortestosterone

(G) Testosterone

(H) 7α-methyltestosterone

(I) 19-nor-7α-methyltestosterone

(F2) Results.

The results of this study are presented in the form of graphs of Chemical Shift (in PPM, up field of CS_2) against calculated charge for the nine molecules in the figures below. The actual charges are tabulated in Appendix Three.

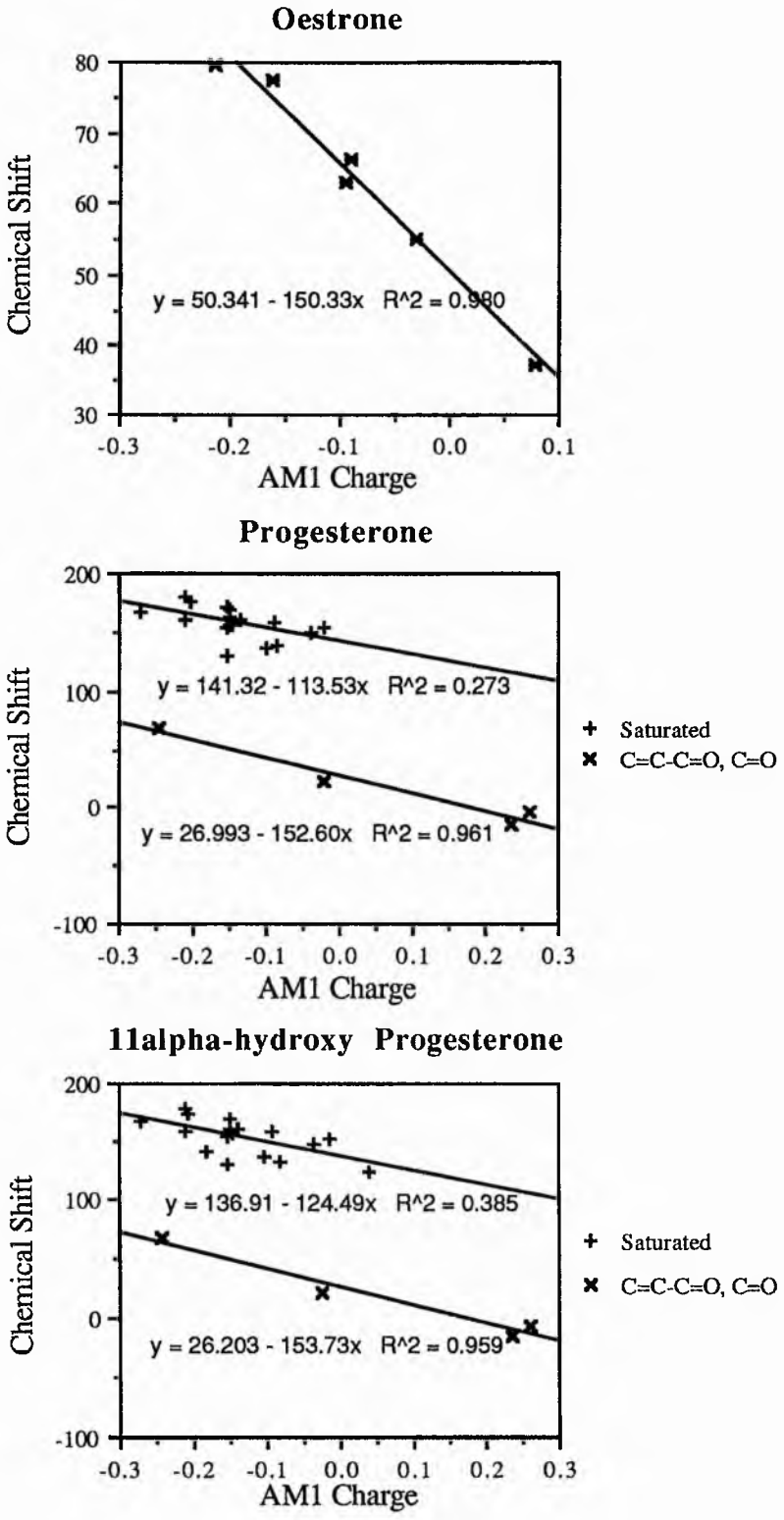

Oestrone
Chemical Shift
AM1 Charge
y = 50.341 - 150.33x R^2 = 0.980
Progesterone
y = 141.32 - 113.53x R^2 = 0.273
y = 26.993 - 152.60x R^2 = 0.961
Saturated
C=C-C=O, C=O
11alpha-hydroxy Progesterone
y = 136.91 - 124.49x R^2 = 0.385
y = 26.203 - 153.73x R^2 = 0.959

Figure 7.14

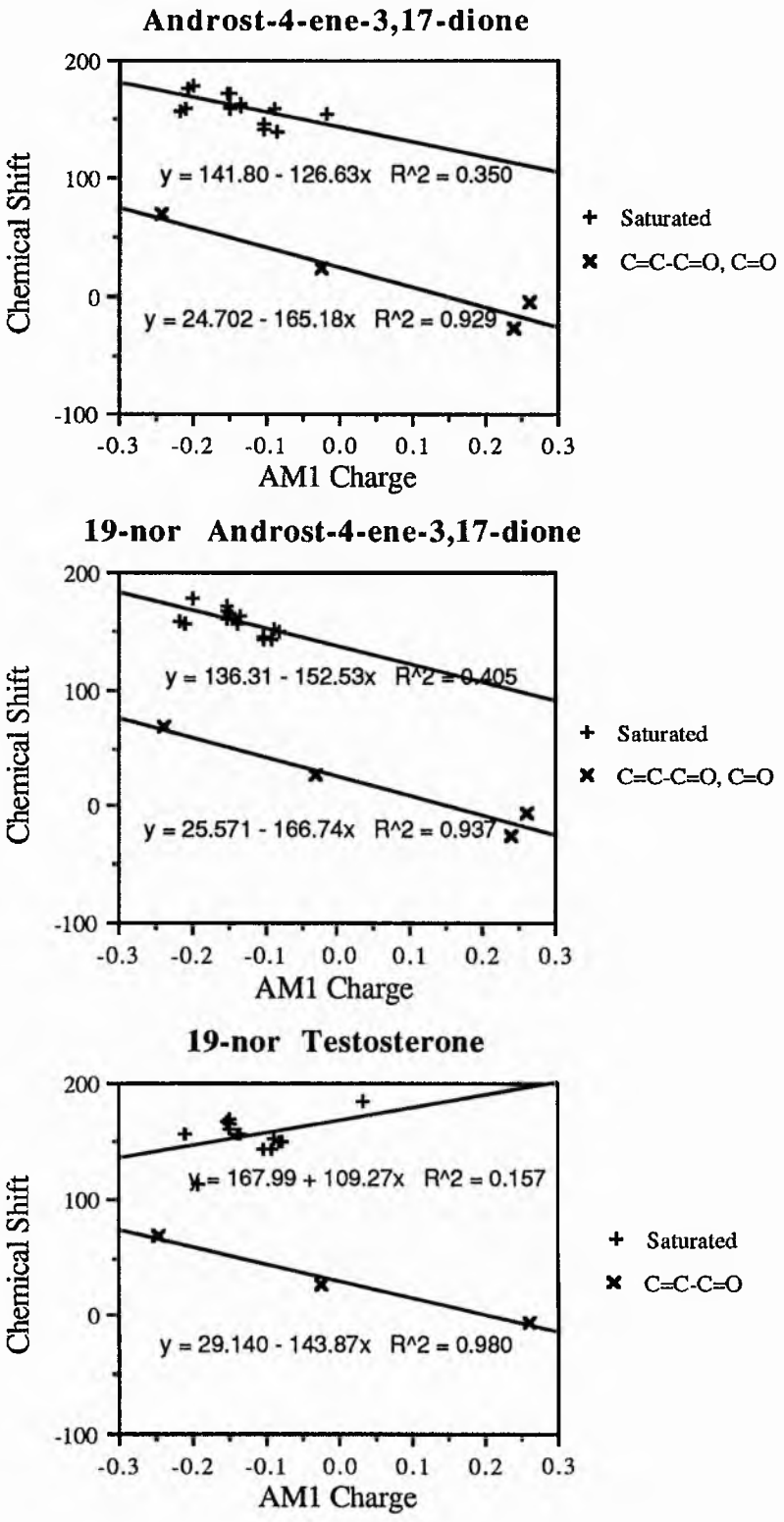


Figure 7.14 (Cont.)

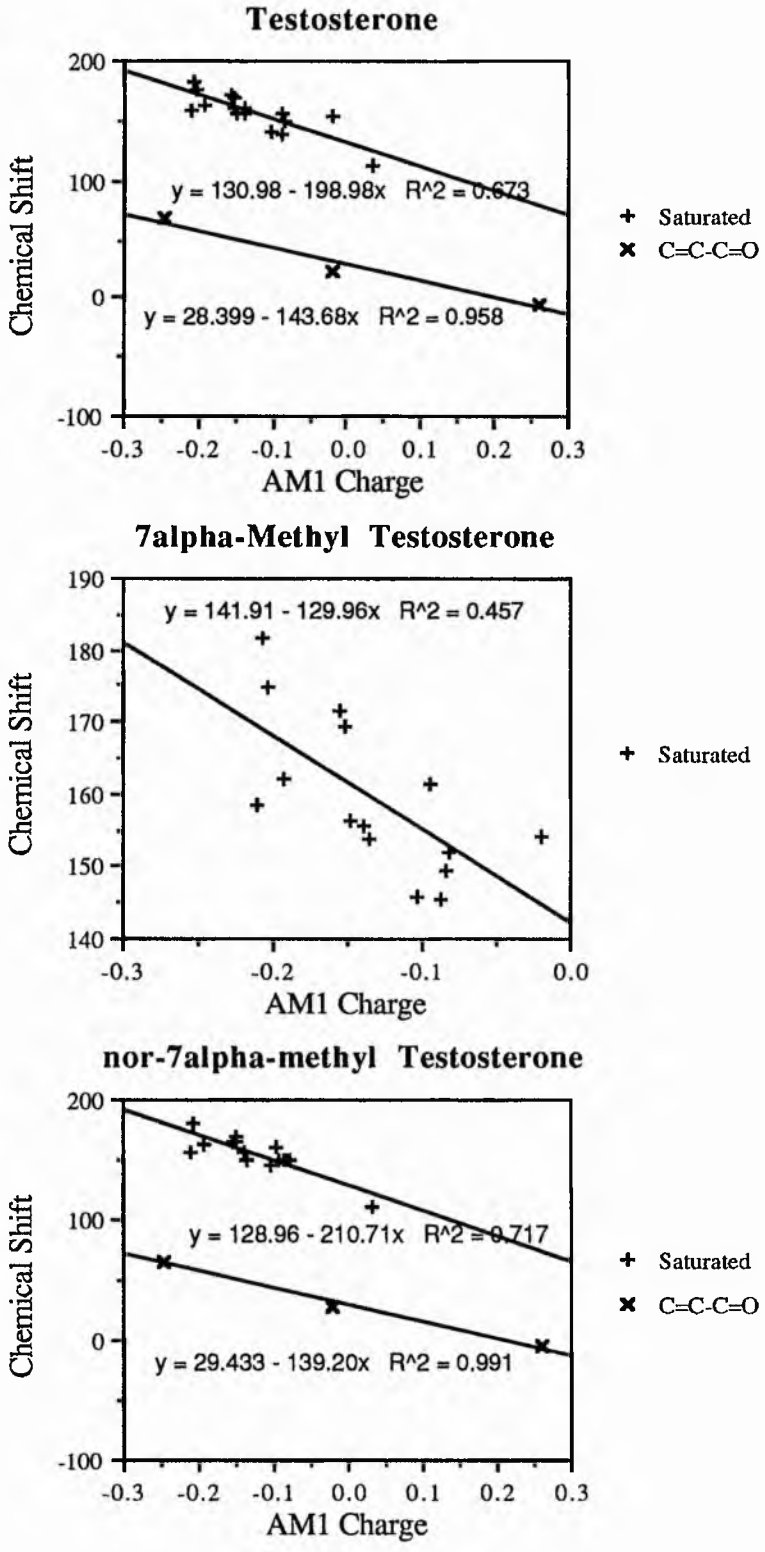


Figure 7.14 (Cont.)

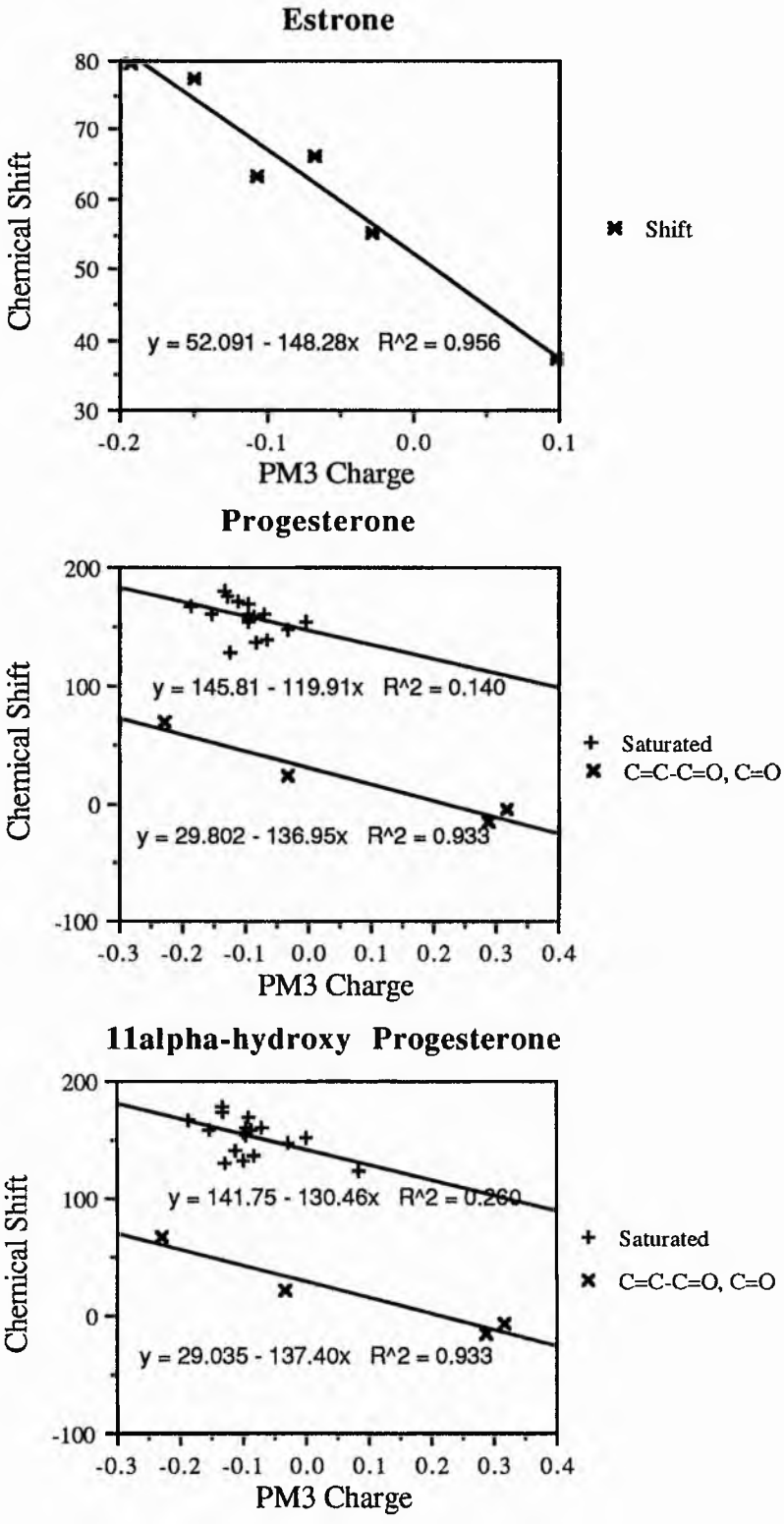


Figure 7.14 (Cont.)

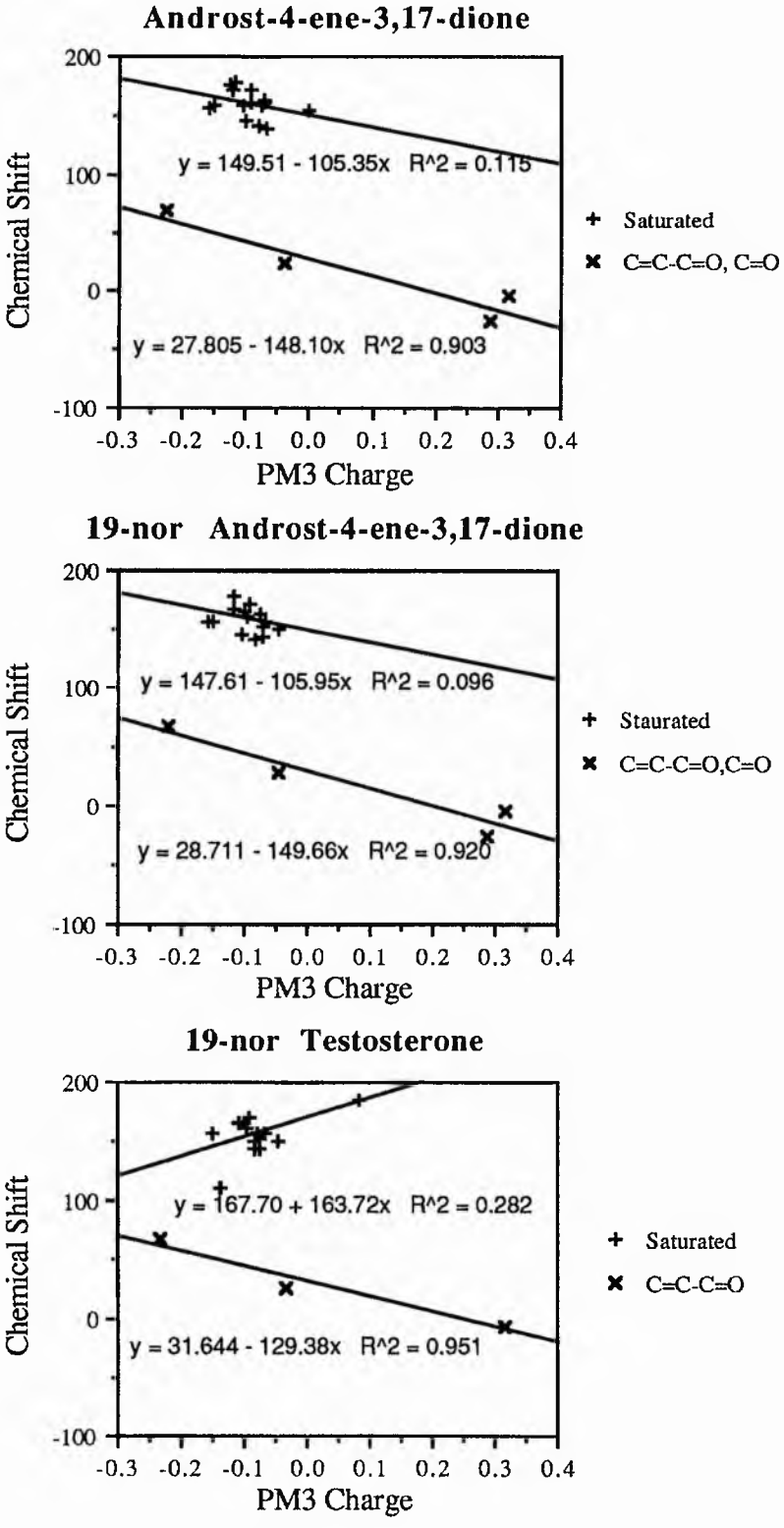


Figure 7.14 (Cont.)

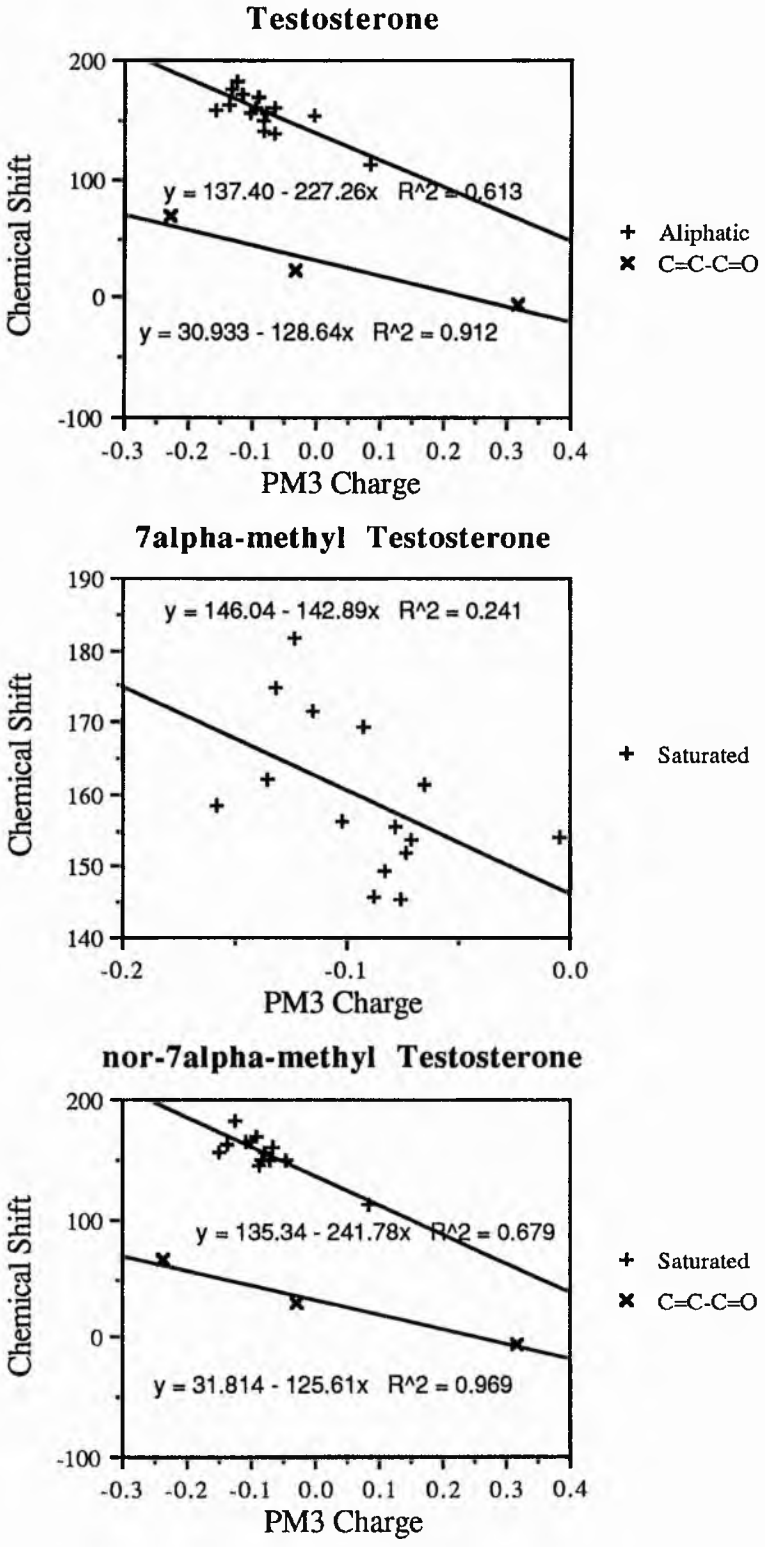


Figure 7.14 (Cont.)

The intercepts, gradients and correlation co-efficients for the lines in the above graphs are tabulated below in Tables 7.10 and 7.11.

Molecule	Method	Intercept	Gradient	R^2
(A)	AM1	50.3	-150.3	0.980
(A)	PM3	52.1	-148.3	0.956
(B)	AM1	27.0	-152.6	0.961
(B)	PM3	29.8	-137.0	0.933
(C)	AM1	26.2	-153.7	0.959
(C)	PM3	29.0	-137.4	0.933
(D)	AM1	24.7	-165.2	0.929
(D)	PM3	27.8	-149.1	0.903
(E)	AM1	25.6	-166.7	0.937
(E)	PM3	28.7	-149.7	0.920
(F)	AM1	29.1	-143.9	0.980
(F)	PM3	31.6	-129.4	0.951
(G)	AM1	28.4	-143.7	0.958
(G)	PM3	30.9	-128.6	0.912
(H)	AM1	---	---	---
(H)	PM3	---	---	---
(I)	AM1	29.4	-139.2	0.991
(I)	PM3	31.8	-125.6	0.969
Average	AM1	28.6*	-151.9	0.962
Average	PM3	29.9*	-138.1	0.935

Table 7.10

Gradients and Intercepts for Unsaturated / Aromatic Atoms.

* These averages do not include the values for Oestrone (A).

Molecule	Method	Intercept	Gradient	R^2
(A)	AM1	---	---	---
(A)	PM3	---	---	---
(B)	AM1	141.3	-113.5	0.273
(B)	PM3	145.8	-119.9	0.140
(C)	AM1	136.9	-124.5	0.385
(C)	PM3	141.8	-130.5	0.260
(D)	AM1	141.8	-126.6	0.350
(D)	PM3	149.5	-105.4	0.115
(E)	AM1	136.3	-152.5	0.405
(E)	PM3	147.6	-106.0	0.096
(F)	AM1	168.0	+109.3	0.157
(F)	PM3	167.7	+163.7	0.282
(G)	AM1	131.0	-199.0	0.673
(G)	PM3	137.4	-227.3	0.613
(H)	AM1	141.9	-130.0	0.457
(H)	PM3	146.0	-142.9	0.241
(I)	AM1	129.0	-210.7	0.717
(I)	PM3	135.3	-241.8	0.679
Average	AM1	140.8	-118.4	0.427
Average	PM3	146.4	-113.8	0.303

Table 7.11

Gradients and Intercepts for Saturated Atoms.

(F3) Correlation of Chemical Shifts and AM1 ESPFIT Charges.

The AM1 optimised geometry used in the above study was also taken as the basis for calculations using the electrostatic potential fitted charge method using the AM1 wavefunction for the molecule Progesterone (B), Testosterone (G) and Oestrone (A). The calculated charges were again plotted against the ^{13}C chemical shifts for the appropriate atoms. The results of these calculations are summarised below in the form of a table of intercepts, gradients and correlation co-efficients.

Molecule	Method	Intercept	Gradient	R^2
(A)	ESPFIT	60.1	-60.9	0.982
(B)	ESPFIT	34.9	-70.6	0.958
(G)	ESPFIT	31.7	-65.0	0.989

Table 7.12

Correlation between AM1 ESPFIT charges and chemical shifts for unsaturated atoms.

Molecule	Method	Intercept	Gradient	R^2
(A)	ESPFIT	---	---	---
(B)	ESPFIT	150.1	-62.5	0.261
(G)	ESPFIT	147.6	-94.6	0.569

Table 7.13

Correlation between AM1 ESPFIT charges and chemical shifts for aliphatic atoms.

(F4) Correlation of Chemical Shifts and Ab Initio Mulliken Charges.

Both the AM1 and the PM3 optimised geometries for the three molecules studied in the previous section were used to generate input for Gaussian, and single point Ab Initio calculations were performed at the STO-3G and 3-21G levels. This

study was performed to investigate the possibility of a defect in the semi empirical methods implied by the fact that they are unable to give a good correlation for the non - delocalised atoms, when the less sophisticated EHMO could. A good correlation for the Ab Initio charges for the saturated atoms would indicate such a flaw in these methods. The results are summarised in the form of tables of correlation co-efficients for the least squares line drawn through the points upon the Chemical Shift against Calculated Charge plot.

Molecule	STO-3G//AM1	3-21G//AM1	STO-3G//PM3	3-21G//PM3
(A)	0.980	0.924	0.981	0.925
(B)	0.975	0.986	0.969	0.983
(G)	0.960	0.982	0.953	0.979

Table 7.14

Correlation co-efficients using Ab Initio Charges for Aromatic / Unsaturated atoms.

Molecule	STO-3G//AM1	3-21G//AM1	STO-3G//PM3	3-21G//PM3
(A)	---	---	---	---
(B)	0.307	0.384	0.313	0.392
(G)	0.747	0.893	0.734	0.890

Table 7.15

Correlation co-efficients using Ab Initio Charges for Saturated atoms.

(F5) Discussion.

The results for the unsaturated and aromatic systems are extremely good for all the methods, with average correlation co-efficients of greater than 0.9 for both semi empirical methods. The fact that most of the molecules produce a graph on which the

least squares fitted lines have similar gradients and intercepts also implies that this is a useful technique for assessing the quality of the calculated charges.

The results for the aliphatic carbon atoms are less encouraging, with values of R^2 that are much less than 0.5 for both AM1 and PM3. This is quite surprising, because the EHMO study [140] mentioned above (which also uses the Mulliken Analysis method) yields results that are much more useful for the prediction of chemical shifts for this type of carbon atom. The gradients and intercepts of the lines corresponding to these atoms show a greater variation than do those for the double bonded and aromatic carbons, which indicates that this study is less useful in assessing the charges for the saturated systems.

The AM1 ESPFIT results for these atoms are also not particularly good, although the values for Testosterone (G) are markedly better than those for Progesterone (B) (this is also true for the Mulliken charges for the same molecules). The values of R^2 are smaller than those using Mulliken Analysis, except for the Testosterone PM3 results.

The Ab Initio results for the unsaturated and aromatic carbon atoms are, in general, similar to the ESPFIT values, although it is interesting to note that the STO-3G basis set yields correlation co-efficients that are greater than those for the Spilt Valence 3-21G basis for the aromatic Oestrone molecule. This situation is reversed for the cases of the C=C and C=O double bonded cases. Neither of the semi empirical methods produced geometries that yielded results which are significantly superior when the charges were compared at the Ab Initio level.

For the aliphatic carbon atoms, the results are somewhat better than the semi empirical and ESPFIT charges, although the results for Progesterone are still poor. For

both molecules, the 3-21G basis set yields a superior value of R^2 to STO-3G, and there is only a slight difference between the use of the two semi empirical geometries.

(F7) Conclusion.

For both the saturated and the unsaturated cases, it seems that AM1 is able to produce Mulliken charges that are more suitable to modelling ^{13}C chemical shifts than is PM3, as shown by the larger value of R^2 in both cases. Contrary to the literature [43], the AM1 ESPFIT charges seem to be inferior to the Mulliken results, although this could of course be due to a breakdown of the theory behind this study, indicating that other factors are important in determining the chemical shift.

Although the predictions for delocalised systems seems to be very successful, the treatment of other carbon atoms, even using Ab Initio methods, is extremely variable in the quality of the results.

(G) Comparison of Calculated Structures and X-ray Geometries.

The structures for many steroid compounds are available on the Daresbury Crystal Structure database [143], and a comparison of some of these with the AM1 and PM3 calculated structures. In most cases, comparisons have been performed on molecules for which the calculations had already been performed, and in some cases a comparison has been made between some of the parameters of the calculated geometry and the corresponding ones of a similar molecule taken from the database.

(G1) Method.

In the cases in which molecules are the same as those extracted from the database, the comparison has been done in the usual way, using the MHCDraw

program, and the average differences are reported. If the molecules are only similar, and not exactly the same, then the values of certain important geometrical parameters are quoted.

(G2) Results : MHCDraw Comparison.

The results of the complete comparisons are shown in Table 7.16, below. In some cases, the results are also reported for the comparison between the molecule in which the hydrogen atoms have not been considered. The group of molecules consists of a selection of G6PDH inhibitors, the structure of which are described in Chapter Five.

Molecule	Method	$\overline{\Delta L}$/Å	$\overline{\Delta\theta}$/°	$\overline{\Delta\tau}$/°	$\overline{\Delta c}$/Å
And-11	AM1	0.081	2.0	6.5	0.077
And-11	PM3	0.074	2.1	6.8	0.096
And-06	AM1	0.057	3.1	5.8	0.084
And-06	PM3	0.051	3.3	7.1	0.090
And-06*	AM1	0.017	1.2	3.5	0.041
And-06*	PM3	0.012	1.3	3.9	0.053
And-15	AM1	0.034	1.8	6.6	0.056
And-15	PM3	0.028	2.2	6.8	0.081
And-15*	AM1	0.027	1.9	4.6	0.029
And-15*	PM3	0.027	2.1	5.8	0.040
Preg-12	AM1	0.065	3.2	6.7	0.137
Preg-12	PM3	0.060	3.4	8.6	0.217
Preg-12*	AM1	0.014	1.0	8.1	0.073
Preg-12*	PM3	0.012	1.2	12.1	0.112

Table 7.16

Differences Between Experimental and Calculated Geometries.

* Indicates Results of Comparison Excluding Hydrogen Atoms.

(G3) Discussion.

It is interesting to note that in molecule And-11, the calculated bond lengths of the hydrogen atoms were generally longer than those from the X-ray determination, a feature common to a number of molecules. The worst dihedral predictions in this molecule concern the definition of the hydroxy - group at ring - position 3 on the molecule. This is probably due to intermolecular hydrogen bonding between adjacent molecules in the crystal lattice causing a distortion from the gas - phase geometry predicted by the theoretical methods.

In And-06, it is again the description of the 3β-hydroxy group that is particularly poor, along with the description of the hydrogen atoms in general. The latter observation can be deduced from the fact that the results of the comparison which ignored the hydrogen atoms are greatly superior to those that include these atoms, as is the case in all the comparisons in which hydrogens have been ignored. This is due to the poor accuracy of this method in determining the locations of these atoms because they have a low electron density localised around the nucleus, and thus have little effect upon the X-ray beam used in determining the crystal structure.

PM3 predicts bond lengths that are superior to the equivalent AM1 values, although both AM1 angles and dihedrals are closer to those from the X-ray determination than are the PM3 results. In every case, the average value of the difference in co-ordinates indicated that the AM1 structure is the closer of the two semi empirical methods to the X-ray structure, this include the cases in which the hydrogen atoms have been ignored.

(G4) Results : Critical Angles and Lengths.

The results of the determination of some of the angles and bond lengths of interest in a few molecules are outlined below. Angles and dihedrals are quoted in degrees, and bond lengths in Å.

And-28

This molecule has a 3β-chloro group, for which a model compound was found, although the latter differs in the conformation of the carbon atom C_{17}.

Parameter	X-ray	AM1	PM3
$Cl\text{-}C_3\text{-}C_2\text{-}C_1$	178.8	177.6	173.6
$Cl\text{-}C_3\text{-}C_2$	106.1	110.1	108.2
$Cl\text{-}C_3$	1.8502	1.7721	1.7900

Table 7.17

Selected Geometrical Parameters for And-28.

And-25

This molecule has a methyl group attached at C_7, the geometry of which is examined below.

Parameter	X-ray	AM1	PM3
$C_{Me}\text{-}C_7\text{-}C_8\text{-}C_9$	77.7	76.2	78.3
$C_{Me}\text{-}C_7\text{-}C_6\text{-}C_5$	249.9	251.3	249.2
$C_{Me}\text{-}C_7$	1.5271	1.5188	1.5303

Table 7.18

Selected Geometrical Parameters for And-25.

Preg-09

This is an example of a the Pregnane - type inhibitors of G6PDH, studied in Chapter Five. A similar X-ray structure has been found, which differs only in the fact that the hydroxy - group at position 3 is α to the ring in the crystal structure. The region of interest in this molecule is the conformation of the side chain attached to C_{17}, which has been examined extensively in Chapter Five.

Parameter	X-ray	AM1	PM3
C_{21}-C_{20}-C_{17}-C_{13}	278.6	269.3	287.2
O=C_{20}-C_{17}	120.2	123.0	123.4
O=C_{20}	1.2144	1.2156	1.2350
O=C_{11}-C_9-C_{10}	359.3	356.2	354.3
O=C_{11}-C9	124.1	124.7	124.3

Table 7.19

Selected Geometrical Parameters for Preg-09.

Preg-12

This molecule also has the usual pregnane - type side chain, with a keto - group at C_{20}, but includes a 16α,17α-epoxy group, and the interactions between the two deserve special mention, because of the conformational aspects of this type of molecule.

Parameter	X-ray	AM1	PM3
C_{21}-C_{20}-C_{17}-C_{13}	180.6	257.7	290.8
O=C_{20}-C_{17}-C_{13}	0.4	76.8	110.8
O=C_{20}-C_{17}	120.6	122.0	121.2
O_{Epoxy}-C_{17}-C_{13}-C_{14}	39.3	38.1	38.9
O_{Epoxy}-C_{17}	1.4470	1.4410	1.4401
C_{16}-O_{Epoxy}-C_{17}	60.8	63.1	62.9
C_{15}-C_{16}-C_{17}-C_{13}	3.2	3.8	3.6
O-C_3-C_4-C_5	186.7	183.0	179.3

Table 7.20

Selected Geometrical Parameters for Preg-12.

(G5) Discussion.

The PM3 prediction of the carbon - chlorine bond length was more accurate on the And-28 model, although the AM1 dihedral for this atom is clearly superior. PM3 also yielded better results for the structure of the methyl group in the And-25 model compound, with both the C-C bond length and the dihedral angle being closer to the X-ray geometry.

In the case of Preg-09, the two methods gave similar results, with notable exceptions being the superiority of AM1 at treating the O=C double bond at C_{20} and the twist angle for the oxygen at C_{11}. The predictions of the twist angle defining the pregnane side chain are both approximately 10° out, but it is interesting to note that the two methods predict the minimum energy conformation to be on opposite sides of that depicted by the X-ray geometry.

For Preg-12, the prediction of the conformation of the side chain is totally wrong (although AM1 is considerably closer to the X-ray value than is PM3). Both semi empirical methods tend to favour the same conformation predicted for most of the other pregnane - type molecules studied, with the keto - group pointing in front of the molecule when viewed in the orientation depicted in Figure 1.1 of Chapter One. In the X-ray structure, this group points out to the left of the molecule, towards the angular C_{18} methyl group as portrayed schematically in Figure 7.15. It should be noted that this difference could well be the result of packing and intermolecular hydrogen - bonding forces occurring in the crystal lattice, since the hydroxy group at C_3 in this molecule would make an ideal partner for hydrogen bonding. Nevertheless, the rotation profile calculations for this molecule (see Chapter Five) seems to indicate very little interaction between the keto - and epoxy - groups, which could be due to a failing of the semi empirical methods.

(G6) Conclusions.

It seems that overall, AM1 is able to give a better average description of the X-ray results, although PM3 bond lengths have been shown to be superior. It also seems that the methods cannot always reliably predict the X-ray values of critical torsion angles, such as those occurring in the pregnane side chain, but in the study presented here, PM3 seems more able to give more values that are closer to the experimental structures.

(H) Final Comments.

Although PM3 yields geometries that have superior bond lengths and lower energies when compared using Ab Initio or experimental results as a standard, the AM1 structures have angles and dihedrals that are, on average, the closer of the two. The molecules of interest in Chapter Six provide an example of the poor results calculated using PM3 for nitrogen - containing compounds. Not only does PM3 fail to reproduce the planarity of the amide systems, but the calculated charges are also in poor agreement with those obtained using other theoretical methods.

When calculated properties are considered, this chapter has shown that calculated dipoles and ionisation potentials are poor, but the correlation of PM3 with the experimental values is very slightly better. This difference between the two methods is not a particularly useful method of judging the semi empirical methods, considering their poor performance.

The AM1 calculated charges have proved to be a superior method of predicting ^{13}C chemical shifts, although both methods faired significantly worse than the less sophisticated EHMO method. The charges calculated by fitting to the calculated AM1 MEP faired no better in this comparison, and thus would seem to be of less use than the

authors of the method suggest [43], especially when considered in conjunction with their poor performance in the prediction of biological properties in Chapter Five.

In summary, it seems that although PM3 gives a more accurate prediction of bond lengths, overall, structures calculated using AM1 are superior, and there is very little to choose between the two as far as calculated properties for many of the molecules of interest in this thesis are concerned, the 3-keto-4-aza steroids of Chapter Six being an exception.

Chapter Eight

Conclusions.

This work has achieved a number of goals, concerned both with predicting and accounting for the biological activity of steroid hormones using theoretical methods, and with the assessment of these methods themselves. One further achievement was the design and writing of a Molecular Graphics display program.

The graphics program, MHCDraw, has been use to various extents by most people in the theoretical research group at St. Andrews over the past three years. It has provided the facility to depict large molecular systems with a display quality far superior to anything previously available within the group. In particular, MHCDraw gives a perception of depth and 3D shape within molecules, which has made possible the study of biological systems with several thousand atoms, such as the enzyme and the lipid bilayers depicted in the photographs in Chapter Three.

A secondary feature of the program is its ability to portray the differences between two structures of a molecule in a graphical form. Thus it is possible to compare calculated and experimental geometries in a simple manner. This technique has been utilised in several sections of the thesis.

Following the acquisition of a Tektronix CAChe system in the latter part of 1990, many features of the program have been superseded. In particular, the real 3D effect of the CAChe yields an excellent representation of a molecule. Thus future work upon the MHCDraw program seems to be of low priority. Nevertheless, the fact that the FORTRAN source - code for the program is available means that specialist requirements can be incorporated into the program. Possibilities include subroutines to allow the comparison of equivalent sections of different molecules and the addition of commands to improve the manipulation of molecules such as building up systems and altering geometric parameters. A least - squares fitting procedure, similar to the one available in Chem-X would also be useful.

The studies of steroidal inhibitors of the G6PDH and Aromatase inhibitors have been very successful, particularly when all the parameters within the biological interactions are considered. Firstly, obtaining accurate experimental data can be difficult, and differences in reported properties can arise, as demonstrated at the end of Chapter Five. Such problems can markedly effect any correlation obtained.

Secondly, the use of the Molecular Electrostatic Potential as the sole tool to predict biological activities, particularly when calculated using semi empirical charges, neglects important intermolecular interactions, as described in Chapter Four and introduces possible sources of error into the calculations. The use of the semi empirical methods assumes that the individual parameterisations are capable of reproducing structures, energies and charge distributions accurately upon quite large systems. Additionally, in the use of MEP s, the electronic effects of both solvent and the enzyme itself have been neglected, and these can have a marked effect upon the electronic structure of the inhibitors.

Despite these possible sources of error, the calculations have yielded interesting results that could not have been obtained in other ways. In particular, the study of the MEP s of the G6PDH inhibitors has located a region which is likely to be involved in the binding to the enzyme and accounts for the difference in activity for molecules with a very small change in structure. The work upon the Aromatase inhibitors, although it failed to account for the Time Dependent Inactivation, provided an excellent prediction of activity for most of the competitive inhibitors. The correlation in this case was approximately 95 %.

Future work on these systems would involve the examinations of how other chemical groups, such as halogens, can alter the MEP, and thus predictions can be made of molecules which will be better inhibitors of the enzymes, or of molecules with equivalent activity, but without undesirable side - effects.

The study of the Aromatase inhibitors could also benefit from a more thorough conformational search of the 17β-side chain. At present, the minimal energy configuration of the N-alkyl groups have not yet been investigated. A useful tool in this search, which involves several independent variables, is Molecular Dynamics (MD). MD assigns both kinetic and potential energy to the atoms, and allows their positions to vary with time. The energy and the time spent in the conformations generated can be calculated, and this can yield information regarding the stability of the different structures. An additional advantage of MD is that because Quantum Mechanics is not used to evaluate the energy, a large number of atoms can be considered at once, meaning that solvent effects can also be studied.

The final chapter of computational work attempts to assess the quality of the results of the AM1 and PM3 semi empirical methods, and their ability to reproduce both experimental and Ab Initio calculated properties for steroids.

The semi empirical methods were compared extensively for their ability to predict energies, geometries and charge distributions, and in addition, charges calculated by fitting to the MEP were evaluated. Although PM3 results seemed, on the whole, to be slightly superior to those of AM1, the former method's total failure to model the amide bond means that the results for nitrogen - containing compounds should be regarded with some scepticism. The ESPFIT charges tested in this work have, contrary to expectation, proved less suitable than those calculated from the semi empirical density matrix, for modelling experimental properties.

A large amount of work in the studies of the inhibitors has been performed using the ASP program as a tool for the comparison of molecules, and the results have, on the whole, been extremely successful. Nevertheless, it should be noted that the Similarity Index method does have some drawbacks. In particular, it would be impossible to recognise a molecule that has a greater activity than that of the compound

chosen as the lead using similarity calculations alone. The EPSI is also less useful in cases in which the structure of the molecule changes in a region that is not involved in the binding process. Such conditions cause a large change in the EPSI, often without altering the activity, and thus molecules no longer appear upon the calculated line.

A possible cure for this problem would be the introduction of regional Similarity Index calculations, in which the differences between certain areas of molecules can be ignored in the evaluation of the EPSI. The areas could be selected using the actual potential maps as a guide.

Finally, it should be noted that the optimisation of the EPSI within ASP is not a fool proof method, and often leads to a poor overlap of the test molecule with the lead. Thus the orientation of the molecule should be checked before the optimised EPSI is used.

In summary, the calculated properties have been more than sufficient to provide a useful model of the steroid - enzyme interactions, and have been obtained from calculations that have been reasonably quick to perform.

References.

1) H. Pitot. "Fundamentals of Oncology.", Marcel Dekker Inc., New York. 2nd Edition (1980).

2) Ed. R.E. LaFond. "Cancer : The Outlaw Cell.", .American Chemical Society, Washington D. C. 1st Edition (1978).

3) Ed. D.L. Davis and D. Hoel. "Trends in Cancer Mortality in Industrial Countries", Ann. NY Acad. Sci., 609 (1990).

4) R. Doll and R. Peto. "The Causes of Cancer.", Oxford University Press, New York. 1st Edition (1981).

5) Ed. I.F. Tannock and R.P. Hill. "The Basic Science of Oncology.", Pergamon Press, New York. 1st Edition (1978).

6) J. Cairns. "Cancer : Science and Society.", W. H. Freeman and Company, San Francisco. 1st Edition (1978).

7) E. Gurpide and M. Lippmann. "Steroid Hormone Receptors in Breast and Endometrial Cancer.", Upjohn, Michigan. 1st Edition (1977).

8) J - M Gasc and E - E Baulieu. Biology of the Cell, 56, 1-6 (1986).

9) J. O'Neal Johnston and B.W. Metcalf. "Aromatase : A Target Enzyme in Breast Cancer Chemotherapy.", 307-328 in "Novel Approaches to Cancer Chemotherapy." Academic Press, New York. 1st Edition (1984).

10) A.P.M. Forrest. 1-10 in "Oestrogen and the Human Breast." Proc. Roy. Soc. Edinburgh ,95 (1989).

11) B.E. Henderson, R. Ross and L. Berstein.Rosenthal Lecture, Cancer Res., 48, 246-253 (1988).

12) A. Pullman and B. Pullman. "Molecular Electrostatic Potential of Nucleic Acids." in Quarterly Rev. Biophys., 14, 289-380 (1981).

13) P.A. Kollman."The Role of the Electrostatic Potential in modelling Hydrogen Bonding and other non-covalent interactions." in "Chemical Applications of

Atomic and Molecular Electrostatic Potentials." Ed. P. Politzer and D.G. Truhler Plenum Press, New York. 1st Edition (1981).

14) T. Venanzi and C. Venanzi."Sweet - Taste Receptor Recognition of the Conformationally Flexible Aldoxime Molecule', 321-324 in QSAR: Quantitative Structure-Activity Relationships in Drug Design, ed. J. L. Fauchiere ©Alan R. Liss, New York. 1st Edition(1989).

15) A. Kumar and P.Mishra.Int. J. Quant. Chem., 38, 11-23 (1990).

16) I. Smith and P. Seybold.Int. J. Quant. Chem. Biol. Symp., 5, 311-320 (1978).

17) P. Politzer and K. Daiker.Int. J. Quant. Chem. Biol. Symp., 4, 317-325 (1977).

18) A. Phin. Unpublished Work

19) J. Wilkie. Ph.D. Thesis. University of St. Andrews.(1990).

20) R. McWeeny and B. Sutcliffe. "Methods of Molecular Quantum Mechanics." Academic Press, London. 1st Edition (1969).

21) A.Szabo and N.S. Ostlund. "Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory.", Macmillan, New York. 2nd Edition (1989).

22) R. McWeeny. "Coulson's Valence.", Oxford University Press, Oxford. 3rd Edition (1979).

23) W.J. Hehre, L. Radom, P. Schleger and J.A. Pople. "Ab Initio Molecular Orbital Theory.", J. Wiley and Sons, New York.1st Edition (1986).

24) C.C.J. Roothaan. Rev. Mod. Phys., 23, 69-89 (1950).

25) S.F. Boys. Proc. Roy. Soc. (London) ,A200, 542 (1950).

26) W.J. Hehre, R.F. Stewart and J.A. Pople. J. Chem. Phys., 51, 2657-2664 (1969).

27) W.J. Hehre, R. Ditchfield and J.A. Pople. J. Chem. Phys., 56, 2257-2261 (1972).

28) E. Clementi and G. Corongiu. "Geometrical Basis Sets for Molecular Computations". IBM Corporation.

29) P.C. Hariharan and J.A. Pople. Theor. Chim. Acta., 28,213-222 (1973).

30) J. Almioff, K. Faegri and K. Korsel. J. Comp. Chem., 3, 385-399 (1982).

31) J.A. Pople, D.P. Santry and G.A. Segal. J. Chem. Phys., 43, 5129-5135 (1965) and
J.A. Pople and G.A. Segal. J. Chem. Phys., 43, 5316-5151 (1965).

32) G. Klopman and B. O'Leary. "All-Valence Electrons SCF Calculations" Springer-Verlag, Berlin. 1st Edition (1970).

33) J.A. Pople, D.L. Beveridge and P.A. Dobosh. J. Chem. Phys., 47, 2026 (1967).

34) M.C. Flanigan, A. Komornicki and J.W. McIver. in "Semi Empirical Methods of Electronic Structure Calculation" Part B ed. G.A. Segal. Plenum Press, New York and London. 1st Edition (1977).

35) M. Zerner. J. Chem. Phys., 62, 2788-2799 (1975).

36) MOPAC : A General Molecular Orbital Package. Version 5.0. QCPE 455

37) Gaussian88 M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegal, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Khan, J.J.P. Stewart, S. Topiol, J.A. Pople Gaussian Inc. Pittsburg PA., U.S.A (1988)

38) M.J.S. Dewar, W. Thiel. J. Am. Chem. Soc., 99, 4899-4907 (1977).

39) M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart. J. Am. Chem. Soc., 107, 3902-3909 (1985).

40) J.J.P. Stewart. J. Comp. Chem., 10, 209-220 and 221-264 (1989).

41) J.J.P. Stewart. J. Computer Aided Molecular Design, 4 (1990).

42) M.J.S. Dewar and D.M. Storch. J. Am. Chem. Soc., 107, 3898-3902 (1985).

43) G.G. Ferenczy, C.A. Reynolds and W.G. Richards. J. Comp. Chem., 11, 259-169 (1990).

44) C. Thomson. Int. J. Quant. Chem. Biol. Symp., 16, 219-234 (1989).

45) J.M. Chance, B. Kahr, A.B. Buda and J.S. Siegal. J. Am. Chem. Soc., 111, 5940-5944 (1989).

46) G.M. Anstead and J.A. Katzenellenbogen. J. Phys. Chem., 94, 1328-1334 (1990).

47) R.J. Woods W.A. Szarek V.H. Smith. J. Am. Chem. Soc., 112, 4732-4741 (1990).

48) M.J.S Dewar and D.M. Storch. Proc. Natl. Acad. Sci. USA., 82, 2225-2229 (1985).

49) M. Zerner. Appendix C in "Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory." by A. Szabo and N.S. Ostlund. Macmillan, New York. 2nd Edition (1989).

50) P. Pulay. Mol. Phys., 17, 197-204 (1969).

51) Gaussian90, Revison I. M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B.Foresman, H.B. Schlegal, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J.Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin,L.R. Khan, J.J.P. Stewart, S. Topiol, J.A. Pople (1990) Gaussian Inc. Pittsburg PA. U.S.A

52) H.B. Schlegel. J. Comp. Chem., 3, 214-218 (1982).

53) M. Frisch "Gaussian88 User's Guide and Programmer's Reference." Appendix 4 (1988).

54) R. Fletcher and M. J. D. Powell. Comp. J., 6 163 (1966); R. Fletcher, Comp. J. 8 33 (1965); W.C. Davidson.Comp. J., 10 406 (1968).

55) MOPAC 5 Manual, QCPE 429 455.(1983).

56) R.S. Mulliken. J. Chem. Phys., 23, 1833-1840 (1955).

57) M. Connolly Q.C.P.E. (M.S.Program).

58) D.E. Williams and J. M. Yan. Adv. Atomic Mol. Phys., 23, 87 (1988).

59) C. Thomson, D. Higgins and C.M. Edge. J. Mol. Graph., 6, 171-177 (1988).

60) D. Higgins. Ph.D. Thesis. University of St. Andrews (1988).

61) J. Barker. Ph.D. Thesis. University of St. Andrews (1991).

62) R. Carbo, L. Leyda and M. Arnau. Int. J. Quant. Chem., 17, 1185 (1980).

63) E.E. Hodgkin and W.G. Richards. Int. J. Quant. Chem., 14, 105-110 (1987).

64) S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner. J. Am. Chem. Soc., 106, 765-784 (1984).

65) KGNMOL QCPE 538 University of Indiana, Bloomington, Indiana, U.S.A.

66) C. Burt and W.G. Richards. J. Comp. Chem., 11, 1139-1146 (1990).

67) C. Thomson, D. Higgins and C.M. Edge. J. Mol. Graphics., 6, 171-177 (1988).

68) C.M. Edge. Ph.D. Thesis, St. Andrews. (1987)

69) G.G. Ferenczy, C.A. Reynolds and W.G. Richards. J. Comp. Chem., 11, 159-169 (1990).

70) Chem-X. Developed and Distributed by Chemical Design Ltd., Oxford, U.K.

71) C. Burt. ASP: Automated Similarity Index Program, Oxford Molecular Limited (1990).

72) P. Scano, C. Thomson. J.Comp. Chem., 12, 172-174 (1991).

73) Z. Dega-Szafran, M. Jaskolski, Z. Kosturkiewicz, M. Szafran and E. Tykarska. J. Chem. Soc. Perkin Transactions, 2, 69-72 (1991).

74) J. E. Gano, E. J. Jacob and R. Roesner. J.Comp. Chem., 12, 126-134 (1991).

75) ChemDraw 2.1.2 © Cambridge Scientific Computing, Inc.

76) D.G. Truhlar in "Chemical Applications of Atomic and Molecular Electrostatic Potentials". Eds P. Politzer and D.G. Truhlar. Plenum Press, New York. 1st Edition (1981).

77) K. Morokuma and K. Kitaura in "Chemical Applications of Atomic and Molecular Electrostatic Potentials". Eds P. Politzer and D.G. Truhlar. Plenum Press, New York. 1st Edition (1981).

78) C. Hansch, P.G. Sammes and J.B. Taylor. (Chief Eds.). "Comprehensive Medicinal Chemistry" Vols I - VI.Chief Eds. Pergamon Press, Oxford, U.K.1st Edition.(1990).

79) W.D. Edwards and H. Weinstein. Chem. Phys. Lett., 56, 582-584 (1978).

80) H. Weinstein, R. Osman, J. Green and S. Topiol in "Chemical Applications of Atomic and Molecular Electrostatic Potentials". Eds P. Politzer and D.G. Truhlar. Plenum Press, New York. 1st Edition (1981).

81) F.J. Luque, F. Illas and M. Orozo.J. Comp. Chem., 11, 416-430 (1990).

82) E. Scrocco and J. Tomasi in "Advances in Quantum Chemistry". Vol II. Ed. P. Lowdin. Academic Press Inc., London.1st Edition (1978).

83) R.Ditchfield, W.W. Hehre and J.A. Pople. J. Chem. Phys., 51, 2651-2657 (1969).

84) B.H. Besler, K.M. Merz and P.A. Kollman.J. Comp. Chem., 11, 431-439 (1990).

85 U.C. Singh and P.A. Kollman. QCPE Bull. 2, 17 (1982).

86) ESPFIT, Program within AMBER, see ref [64].

87) C.J. Migeon, A.R. Keller, B. Lawrence, and T.H. Shepard. J. Clin. Endocrinol. and Metab., 17, 1051-1062 (1957)

88) M.A. Moore, W. Thamavit, H. Tsuda, S.K. Sato A. Ichihara and N. Ito. Carcinogenesis, 7, 311-316 (1986).

89) T.A. Ratko, C.J.Detrisac, R.G. Mehta, G.J.Kellogg and R.C. Moon. Cancer Res., 51,481-486 (1991).

90) J.W. Nyce, P.N. Magee, G.C. Hard and A.G. Schwartz. Carcinogenesis, 5, 57-62 (1984).

91) E. Henderson, A. Schwartz, L.L. Pashko, M. Abou - Gharbia and D. Swern. Carcinogenesis, 2, 683-686 (1981).

92) A.G. Schwartz and R.H. Tannen. Carcinogenesis, 2, 1335-1337 (1981).

93) M.A. Moore, E.Weber, M. Thornton and P. Bannasch. Carcinogenesis, 9, 1507-1509 (1988).

94) T.T. Yen, J.V. Allen, D.V. Pearson, J.M. Acton and M. Greenberg. Lipids, 12, 409-413 (1977).

95) G.B. Gordon, D.E. Bush and H.F. Weisman. J. Clin. Invest., 82, 721-720 (1988).

96) L.L. Pashko, A.G. Schwartz, M. Abou - Gharbia and D. Swern. Carcinogenesis, 2, 717-721 (1981).

97) N. Orentreich, J.L. Brind, R.L. Rizer and J.H. Vogelman. J. Clin. Endocrinol. Metab., 59, 551-555 (1984).

98) R.D. Bullbrook, J.L. Hayward, C.C. Spicer and B.S. Thomas. Lancet II, 1238-1240 (1962).

99) A.G. Schwartz. Cancer Research., 39, 1129-1132 (1979)

100) A.G Schwartz, G.C. Hard, L.L. Pashko, M. Abou - Gharbia and D. Swern. Nutr. and Cancer, 3, 46-53 (1981).

101) M.H. Ross and G. Bras. J. Natl. Cancer Inst., 47, 1095-1113 (1971).

102) A. Tannenbaum and H. Silverstone. Adv. Cancer Res., 1, 451-501 (1953).

103) A.G Schwartz, L. Pashko and J.M. Whitcomb. Toxicol. Pathol., 14, 357-362 (1982).

104) L. Luzzatto and U. Testa. Current Topics In Haematology, 1, 1-70 Alan R. Liss Inc. 1st Edition New York (1978).

105) P.A. Marks and J. Banks. Proc. Natl. Acad. Sci. (U.S.A), 46 447-452 (1960).

106) R. Raineri and H.R. Levy. Biochemistry, 9, 2233-2243 (1970).

107) G.W. Oertel and I. Rebelein. Biochim. Biphys. Acta, 184, 459-460 (1969).

108) G.W. Oertel and P. Benes. J. Steroid Biochem., 3, 493-496 (1972).

109) F. Feo, L. Pirisi, R. Pascale, L. Daino, S. Frasseto, S. Zanetti and R. Garcea. Toxicol. Pathol., 12, 262 (1984).

110) S. Bhatnager, K. Schieweck and A. Hausler. 293-303 in "Oestrogen and the Human Breast." Proc. Roy. Soc. Edinburgh, 95 (1989).

111) W.R. Miller and A.P.M. Forrest. Br. J. Cancer, 33, 116-118 (1976).

112) J. Fishman and J. Goto. J. Biol. Chem., 256, 4466-4471 (1981)

113) M. Akhtar, M.R. Calder, D.L. Corina and J.N. Wright. J. Chem. Soc. Chem. Commun., 129-130 (1981).

114) E. Caspi, J. Wicka, T. Arunachalam, P. Nelson and G. Spiteller. J. Am. Chem. Soc., 106, 7282-7283 (1984).

115) P.A. Cole and C.H. Robinson. J. Am. Chem. Soc., 110, 1284-1285 (1988).

116) D.E. Stevenson, J.N. Wright and M. Akhtar. J. Chem. Soc. Perkin Trans I, 2043-2052 (1989).

117) Y. Watanabe, Y. Ishimura. J. Am. Chem. Soc., 111, 8047-8049 (1989).

118) A.M.H. Brodie, D. Marsh and H.J. Brodie. J. Steroid Biochem., 10, 423-429 (1979).

119) A. Seago, M.H. Baker, J. Houghton, M. Jarman, C-S Leung and M.G. Rowlands. Biochemical Pharmacology, 37, 2167-2172 (1988).

120) Y.J. Abul-Hajj. J.Steroid Biochem., 35, 139-143 (1990).

121) A.M.H. Brodie, W.C. Schwarzel, A.A. Shaikh and H.J. Brodie. Endocrinology, 100, 1684-1695 (1977).

122) W.C. Schwarzel, W.G. Kruggel and H.J. Brodie. Endocrinology, 92, 866-880 (1973).

123) K. Schieweck, A.S. Bhatnagar and A. Matter. Cancer Res., 48,834-838 (1988).

124) R.W. Brueggemeier, E.E. Floyd and R.E. Counsell. J. Med. Chem., 21, 1007-1011 (1978).

125) M.V. Darby, J.A. Lovett, R.W. Brueggemeier, M.P. Groziak, and R.E. Counsell. J.Med. Chem., 28, 803-807 (1985).

126) Y.J. Abul-Hajj. J. Med. Chem., 29, 582-584 (1986).

127) G.A. Flynn, J.O. Johnston, C.L. Wright and B.W. Metcalf. Biochem. Biophys. Res. Commun., 103, 913-918 (1981).

128) D.F. Covey, W.F. Hood and V.D. Parikh. J. Biol. Chem., 256, 1076-1079 (1981).

129) A.T. Greway and M.A. Levy. J. Steroid Biochem., 33, 5573-579 (1989).

130) K-T Lim and M.M. Francl. J. Phys. Chem., 91, 2716-2721 (1987).

131) A. Sawaryn and J.S. Yadav. Int. J. Quant. Chem., 29, 1241-1251 (1986).

132) A.M. Sapse, L.M. Fugler and D. Cowburn. Int. J. Quant. Chem., 29, 1241-1251 (1986).

133) D. Henderson, U-F. Habenicht, Y. Nishino, U. Kerb and M.F. El Etreby.J. Steroid Biochem., 25, 867-876 (1986).

134) P-K. Li and R.W. Brueggemeier. J. Med. Chem., 33, 101-105 (1990).

135) P-K. Li and R.W. Brueggemeier. J. Enzyme Inhibition, 4, 113-120 (1990).

136) C. Van Alsenoy, K. Wolinski, K. Siam and L. Schafer. J. Molecular Structure (Theochem), 180, 343-352 (1988).

137) M.J.S. Dewar, A.J. Holder, E.F. Healy and S. Olivella. J. Chem. Soc. Chem. Commun., 1452-1454 (1989).

138) R. Huttenrauch and K. Matthey. Archiv der Pharmazie, 300, 1007-1016 (1967).

139) H. Repmann. Theoret. Chim. Acta (Berl.), 17, 396-407 (1970).

140) H. Repmann. Z. Naturforsch, 29, 1172-1178 (1974).

141) J.A. Pople. Discuss. Farady Soc., 34, 7-14 (1962).

142) D.G. Farnum. Adv. Phys. Org. Chem., 11, 123-175 (1975).

143) K.A.K. Ebraheem, G.A. Webb and M. Witanowski. Org. Magn. Resonance, 8, 317-320 (1976).

144) L. Klasinc, L.J. Pasa-Tolic, B. Kovac and S.P. McGlynn. Int. J. Quant. Chem. Quant. Biol. Symp., 331-341 (1989).

145) Cambridge Structural Database, Daresbury Laboratory, Daresbury, Warrington.

Appendices.

Appendix One

Abbreviations.

AM1	Austin Method 1
CI	Configuration Interaction
CNDO	Complete Neglect of Differential Overlap
CRF	Core Repulsion Function
DHEA	Dehydroepiandrosterone
4,4-DMA	4,4-Dimethylandrostan-3-one
DNA	Deoxyribo Nucleic Acid
EHMO	Extended Huckle Molecular Orbital Theory
EPSI	Electrostatic Potential Similarity Index
G6PDH	Glucose-6-Phosphate Dehydrogenase
HOMO	Highest Occupied Molecular Orbital
INDO	Intermediate Neglect of Differential Overlap
IP	Ionisation Potential
LCAO	Linear Combination of Atomic Orbitals
LUMO	Lowest Unoccupied Molecular Orbital
MEP	Molecular Electrostatic Potential
MIA	Multiplicative Integral Approximation
MNDO	Modified Neglect of Diatomic Overlap
MO	Molecular Orbital
NDDO	Neglect of Diatomic Differential Overlap
NMR	Nuclear Magnetic Resonance
PES	Photo Electron Spectroscopy
PM3	Parametric Method 3
RNA	Ribose Nucleic Acid
SCF	Self Consistent Field

Appendix Two

Help file for the Molecular Graphics Program

HINTS

Type HELP followed by the command you want help on. Anything in square brackets is optional, whilst you must use one of the objects specified in round brackets.

LISTS

Lists are used to define a group of atoms that can be ignored, labelled etc.

The following rules should help with the definition of a list :

Examples of possible ranges - things in [] are optional, () are all compulsory.

1. (a-b)[(c-d)....]	all atoms in the range a->b [and c->d ...].
2. C	all carbon atoms.
3. C,N,O[...]	all carbon, nitrogen and oxygen atoms.
4. C(a-b)[(c-d)...]	all carbons in the range a->b [and c->d].
5. C(a-b)[,]N(c-d)[,]O	carbons a->b, nitrogens c->d and all oxygens.
6. (a-b)C,N	all carbons and nitrogens in the range a->b.
7. C(a)[...]	carbon a only [...]
8. ALL	all atoms.
9. NONE	clears the list.

Ranges can also be defined as (a-) meaning all specified atoms starting from atom a to the end of the atom list. Putting the word ADD before the list adds the existing list to the one being defined. e.g. LABEL ADD C(5-10)

Note : specifying C(8) will label carbon atom number eight (if such an atom exists), and not the carbon whose name is 'C8'. i.e. the program uses the atom numbers from the order that they are read in, and not those from the atom names.

ALPHA

Format : alpha

Function : Joins the alpha carbons.

AUTO-WIDTH

Format : auto-width (on)(off)

Function : Draws thick lines at the front and thin ones at the back. The thickest are drawn with the width specified in the WIDTH command.

Uses : Helps to show depth in the molecule.

CHOP

Format : chop(/dist)(/ignore)(/box)(/coord)

Function : Permanently removes portions of a molecule. /dist : Removes all atom greater than a certain distance from a central one.

Format : chop/dist[/atom=<int>][/radius=<real>]. The two variables will be prompted for entry by mouse if not supplied. /ignore : Removes all atoms currently ignored from the display. (See IGNORE).

/box : Click on 2 corners of a rectangle. All atoms outside of this are removed. /coord : Chops off all atoms outside a specified co-ordinate range.

Format : chop/coord [xmin ymin zmin xmax ymax zmax] if these real values are not supplied, then they are prompted for.

CLEAR

Format : clear

Function : Clears the display.

COLOUR

Format : colour by (atom)(segment)

Function : Colour by atom colours each atom according to its atom type (red for oxygen, green for carbon etc.), whereas colour by segment is used to colour different molecule in the same display differently.

Uses : This is useful when manipulating separate segments in docking studies.

COMPARE

There are two possible ways to use this command :

Format 1 : compare/auto[/scale][/report][/title][/laser][/unique] [seg1] [seg2]

Format 2 : compare[/scale][/report][/title][/laser][/unique]

Also : compare/clear to remove the compare window.

Function : This compares two different geometries of a molecule. The two geometries should be read in as separate segments (using read/append). The program will then give a plot of number of differences in, for example, bond lengths, vs the differences. The spread of these differences shows how alike the two are.

Format1 : Automatically compares bond lengths, angles dihedrals and co-ordinates for two of the segments. If seg1 and seg2 are specified then they should be segment numbers, otherwise, the program will prompt for them.

Format2 : For up to 4 graphs it prompts for function to plot and segment numbers (press return to end the list). /report will prompt the user for a maximum value and report all differences greater than this value. Plot types are length, angle, dihed(ral) and coord(inate).

/laser produces a postscript command file called COMPARE.LAS.

/scale scales the differences over a reasonable range

DRAW

Format : draw [sphere][ball][stick][stereo]

Function : Draws a molecule on screen. Sphere represents the atoms as spheres, whilst ball draws quicker flat circles. DRAW STICK, or just DRAW, draws a plain

stick picture, and DRAW STEREO produces a red-green stereo picture (using the read in co-ordinates only) To do this at any other angle, use the KEEP command).

EMPH

Format : emph <list>

Function : Emph is used to emphasize certain atoms in the display. There are a number of emphasis modes, which can also be accessed through this command. The default is to draw the emphasized atoms orange. The other commands are : EMPH HATCH / EMPH NOHATCH. Draws ball and sphere pictures using solid and hatched circles. EMPH BALL-STICK / EMPH NOBALL-STICK. Draws emphasized atoms as circles, and the rest as sticks when used in conjunction with DRAW BALL or DRAW SPHERE.

See notes on lists.

END

Format : end

Function : Exits from the program. This is the same as EXIT, QUIT and cntrl-Z.

EXIT

Format : exit

Function : Exits from the program. This is the same as EXIT, QUIT and cntrl-Z.

GAUSSIAN

Format : gaussian [filename]

Function : Writes out the geometry in a format suitable for use with the Gaussian programs. If the geometry has been read in from a MOPAC data or archive file, then the MOPAC optimisation flags are used to determine wether or not the same variable is to be optimised in Gaussian. If the input has come from any other source, then all parameters are set to be optimised.

GEOM

Format : geom [/file[=filename]]

Function : Uses the mouse to determine bond lengths, angles and dihedrals. If /file is specified, then output is sent to a file and the screen. The default file name is <molecule_name>.geo. Click on the atoms defining the geometry, and then on the GREEN box to calculate.

GRID

Format : grid [<real>]

Function : Draws a grid [of the specified density in Angstroms] on the screen. The default grid size is 1 Angstrom.

Uses : This is helpful when manipulating separate segments with MOVE.

IDENT

Format : ident

Function : Uses the mouse to identify atoms.

IGNORE

Format : ignore <list>

Function : Temporarily hides the atoms specified in the list from the display. It has the opposite effect of ONLY, and both use the same logical array to store their data, and so affect one-another. This command can be used in conjunction with CHOP and WRITE.

See notes on lists.

KEEP

Format : keep

Function : Atomic co-ordinates are kept in two arrays. One of these is those that have been read in, whilst the other is the result of manipulations using ROT, and are calculated from the former upon each rotation. KEEP transfers the latter to the former.

Uses : This helps with additive rotations as they are normally defined w.r.t. the initial co-ordinates. This is also necessary to view stereo pictures from any angle other than from where they have been read in.

LABEL

Format : label [(/charge)(/length)(/angle)] <list>

Function : Labels the atoms defined in the list. If no list is specified, then it uses the previous labelling list. Note : initially, no list is defined so no atoms will be labelled unless a list is given.

/charge labels the atoms with atomic charges.

/length labels the bonds with lengths.

/angle labels between the atoms with the bond angle. Otherwise, it labels the atoms with their names.

See notes on lists.

MODE

Format : mode (1)(2)

Function : This is used only on the TEK and VT340 terminals. It switches from multi-colour mode (mode 1) to 3 colours by 5 shades (mode 2).

Uses : It can be used to produce (limited) depth pictures on these terminals.

MOPAC

Format : mopac [filename]

Function : Writes out the MOPAC z-matrix for a molecule. Atoms are referenced w.r.t the lowest numbered (if possible) non-hydrogen atom that they are bonded to. It cannot currently handle disconnected systems.

MOVE

Format : move [seg=<int>](/x=<real>)(/y=<real>)(/z=<real>)

Function : Moves the designated segment by the amounts specified. If no segment is chosen, then the machine prompts to pick one with the mouse.

NOTE : MOVE affects the permanent co-ordinate set, and will therefore not have the desired effect if the viewing angle is not 0 0 0. Use KEEP before MOVE to avoid this.

ONLY

Format : only <list>

Function : Used to temporarily hide atoms from the display, and only the atoms specified are drawn. This has the opposite effect of IGNORE.

See notes on lists.

QUEST

Format : quest [filename]

Function : Writes out the geometry in a format suitable for Quest.

READ

Format : read [/noconn][/append][/react] filename

Function : Reads in structure and wavefunction data from disk. The file format is always determined by the file type. It can read :

1) .CHG graphics structure (formatted)

2) .CSR as above

3) .XR as above (mutually perpendicular axes only)

4) .PDB protein data base / Amber output (free format)

5) .DAT Mopac input data file (free format)

6) .ARC Mopac archive file (free format)

7) .CHK Mopac checkpoint file (formatted)

8) .QLOG Quest logfile (formatted)(this must have converged)

9) .PAC binary graphics structure file

10) .POUT output from potential fitted charges program

11) .LOG output from Gaussian - final z-matrix from logfile.

Read/append is used to manipulate more than one molecule at once, whilst read/noconn disables connectivity calculations for when stick pictures are not needed (from PDB files). Read/react reads MOPAC reaction co-ordinate geometries into separate segments. The filetype specified must be .ARC

RENUMBER

Format : renumber

Function : Changes the atom ordering of a specified segment. The segment chosen MUST be drawn on the screen, because the new order is entered by clicking on each atom in turn with the mouse. For this reason, it is limited to 99 atoms, as any more than this would not be feasible. This function can be useful if, for example, you wish to compare two molecules, which have been calculated in a different atomic order.

ROT

Format : rot [seg[ment][=<int>]] <real> <real> <real>

Function : Rotate the angle of view of the display by the specified amounts w.r.t the read in (or last kept (see KEEP)) co-ordinates. ROT SEGMENT is used to rotate one segment w.r.t the others

SEGMENT

Format : segment (first)(last) <int>

Function : Takes ONE molecule and divides it into two segments using the residue information. i.e. The separation is made after e.g. the first <int> residues of the molecule.

Uses : Helpful in docking studies if the enzyme and substrate have been written out in one file.

SETCOL

Format : setcol <int> <real> <real> <real>

Function : Changes one of the default colour definitions. The integer is the colour to be changed, and the reals are the R,G & B values in the range 0->1. e.g. setcol 0 1 1 1 changes the background colour to white.

Uses : Not many !

SHADE

Format : shade <real>

Function : Changes the degree of shading shown in the display. The real should be in the range 0->1. Workstation or mode 2 only. (See MODE).

Uses : Can be used to enhance the perception of depth.

SHOWSEG

Format : showseg

Function : Produces a list of segment information

SLICE

Format : slice (<int>)(step)

Function : Uses shading to emphasize a slice through a large protein. Slice <int> (1->16) does one particular slice, whilst SLICE STEP steps through, using the mouse.

SPAWN

Format : spawn

Function : Creates a VAX/VMS sub-process.

TITLE

Format : title [auto]

Function : Puts text anywhere on the screen, using the mouse to locate the required position. Using AUTO defines and prints a set of labels for each segment.

Uses : Helpful when producing labelled photographs.

TEK

Format : tek

Function : Changes the normal shades of grey (for hydrogen) to shades of yellow for use with the Tek printer. Only if mode = 2 (see MODE).

Uses : This avoids the black/white interchange problems with the printer.

WIDTH

Format : width <int>

Function : Sets the line width. The width range should be 1 to about 6.

WINDOW

Format : window <int>

Function : Sets the size of the display window on the workstation. The size range is 1 to 100, and the default is 43.

WRITE

Format : write[/ignore] <filetype> [filename]

Function : Stores structures or electron density maps on disk. The filename is optional, otherwise the input name is used. <Filetype> can be :

1) CHEMX : graphics structure in Chemx format (no residue info.)

2) PDB : Protein database.

3) CHG : graphics structure + residue info. if available.

4) PAC : same as CHG, except binary file to save space.

The program actually writes the current viewing co-ordinates onto screen, rather than those that were read in. Write/ignore does not write out any atoms that have been ignored.

Appendix Three.

Calculated Charges and Reported Chemical Shifts from ref. 140.

Oestrone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.090	-0.068	66.3
C2	-0.215	-0.193	79.7
C3	0.079	0.099	37.4
C4	-0.162	-0.150	77.4
C5	-0.030	-0.029	55.1
C6	-0.123	-0.052	
C7	-0.152	-0.095	
C8	-0.087	-0.077	
C9	-0.055	-0.022	
C10	-0.094	-0.107	63.1
C11	-0.151	-0.107	
C12	-0.135	-0.073	
C13	-0.102	-0.103	
C14	-0.102	-0.083	
C15	-0.151	-0.091	
C16	-0.217	-0.158	
C17	0.238	0.289	
C18	-0.202	-0.114	

Progesterone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.148	-0.096	156.8
C2	-0.211	-0.153	159.9
C3	0.261	0.316	-4.6
C4	-0.246	-0.230	68.5
C5	-0.020	-0.031	23.2
C6	-0.137	-0.070	160.2
C7	-0.151	-0.094	161.8
C8	-0.090	-0.087	158.6
C9	-0.087	-0.067	138.3
C10	-0.020	-0.004	153.9
C11	-0.153	-0.113	171.3
C12	-0.146	-0.094	156.5
C13	-0.038	-0.033	148.8
C14	-0.101	-0.083	136.3
C15	-0.151	-0.094	169.5
C16	-0.154	-0.094	153.7
C17	-0.154	-0.127	129.2
C18	-0.209	-0.134	179.5
C19	-0.204	-0.130	175.5
C20	0.237	0.287	-14.7
C21	-0.272	-0.186	168.1

11α-Hydroxyprogesterone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.147	-0.098	157.4
C2	-0.210	-0.155	159.0
C3	0.261	0.316	-6.0
C4	-0.242	-0.229	68.1
C5	-0.025	-0.033	22.1
C6	-0.138	-0.072	160.5
C7	-0.151	-0.098	161.9
C8	-0.092	-0.086	158.3
C9	-0.082	-0.101	133.3
C10	-0.013	-0.001	152.4
C11	0.038	0.082	124.3
C12	-0.184	-0.112	142.1
C13	-0.035	-0.031	148.5
C14	-0.103	-0.085	136.8
C15	-0.151	-0.093	169.5
C16	-0.154	-0.094	154.8
C17	-0.154	-0.127	129.4
C18	-0.211	-0.135	178.3
C19	-0.206	-0.135	174.5
C20	0.236	0.287	-14.9
C21	-0.271	-0.187	168.2

Androst-4-ene-3,17-dione

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.151	-0.104	157.8
C2	-0.211	-0.152	159.4
C3	0.261	0.316	-5.1
C4	-0.241	-0.224	68.5
C5	-0.026	-0.040	23.1
C6	-0.137	-0.071	160.8
C7	-0.150	-0.095	161.6
C8	-0.089	-0.078	157.9
C9	-0.085	-0.070	138.4
C10	-0.019	-0.001	154.5
C11	-0.154	-0.120	172.1
C12	-0.135	-0.073	162.2
C13	-0.103	-0.102	145.3
C14	-0.104	-0.081	141.6
C15	-0.151	-0.092	170.8
C16	-0.217	-0.157	157.4
C17	0.238	0.289	-25.9
C18	-0.201	-0.117	179.1
C19	-0.205	-0.126	175.5

19-norandrost-4-ene-3,17-dione

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.150	-0.099	165.7
C2	-0.210	-0.151	156.1
C3	0.260	0.316	-5.2
C4	-0.238	-0.222	67.8
C5	-0.032	-0.045	27.4
C6	-0.137	-0.068	157.2
C7	-0.151	-0.096	160.8
C8	-0.090	-0.071	152.5
C9	-0.090	-0.073	142.7
C10	-0.080	-0.044	150.1
C11	-0.152	-0.116	166.7
C12	-0.135	-0.074	162.3
C13	-0.103	-0.103	145.1
C14	-0.103	-0.083	142.3
C15	-0.151	-0.092	170.9
C16	-0.217	-0.157	157.4
C17	0.238	0.289	-25.8
C18	-0.201	-0.117	179.0

19-nortestosterone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.150	-0.101	165.8
C2	-0.211	-0.149	156.2
C3	0.261	0.316	-5.8
C4	-0.247	-0.235	68.0
C5	-0.024	-0.033	26.6
C6	-0.137	-0.066	157.2
C7	-0.152	-0.095	161.5
C8	-0.089	-0.077	151.8
C9	-0.092	-0.074	143.6
C10	-0.078	-0.045	149.9
C11	-0.154	-0.110	166.2
C12	-0.139	-0.081	155.7
C13	-0.084	-0.084	149.4
C14	-0.102	-0.085	143.9
C15	-0.152	-0.092	169.2
C16	-0.193	-0.136	111.4
C17	0.034	0.083	184.4
C18	-0.208	-0.124	

Testosterone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.148	-0.103	156.7
C2	-0.211	-0.158	158.7
C3	0.261	0.317	-5.2
C4	-0.246	-0.230	68.6
C5	-0.020	-0.032	22.4
C6	-0.137	-0.068	160.0
C7	-0.151	-0.094	160.6
C8	-0.089	-0.077	156.7
C9	-0.087	-0.068	138.2
C10	-0.020	-0.006	153.8
C11	-0.155	-0.118	171.6
C12	-0.140	-0.079	155.7
C13	-0.084	-0.084	149.6
C14	-0.103	-0.085	141.7
C15	-0.152	-0.092	169.0
C16	-0.193	-0.136	162.1
C17	0.034	0.083	111.5
C18	-0.208	-0.124	181.5
C19	-0.204	-0.132	175.5

7α-methyltestosterone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.148	-0.102	156.3
C2	-0.211	-0.158	158.4
C3	0.261	0.317	
C4	-0.247	-0.232	
C5	-0.017	-0.029	
C6	-0.136	-0.071	153.6
C7	-0.094	-0.066	161.2
C8	-0.083	-0.074	151.8
C9	-0.087	-0.076	145.3
C10	-0.020	-0.005	154.0
C11	-0.155	-0.115	171.5
C12	-0.140	-0.079	155.7
C13	-0.084	-0.083	149.4
C14	-0.103	-0.088	145.9
C15	-0.152	-0.092	169.5
C16	-0.193	-0.135	162.0
C17	0.034	0.083	
C18	-0.208	-0.124	181.5
C19	-0.204	-0.132	174.8

nor-7α-methyltestosterone

Atom	AM1 Charge	PM3 Charge	^{13}C Shift
C1	-0.150	-0.102	165.7
C2	-0.211	-0.149	156.0
C3	0.261	0.316	-5.2
C4	-0.248	-0.237	66.1
C5	-0.021	-0.030	28.5
C6	-0.136	-0.069	149.5
C7	-0.095	-0.066	161.6
C8	-0.083	-0.073	149.4
C9	-0.091	-0.082	149.3
C10	-0.079	-0.045	149.9
C11	-0.154	-0.107	165.7
C12	-0.139	-0.081	155.7
C13	-0.084	-0.083	149.3
C14	-0.102	-0.088	145.7
C15	-0.152	-0.092	169.9
C16	-0.193	-0.136	162.2
C17	0.034	0.083	111.5
C18	-0.208	-0.123	181.5