THEORETICAL CALCULATIONS ON LARGE MOLECULES: DEVELOPMENT OF NEW TECHNIQUES AND THE STUDY OF ENZYME INHIBITORS

A Thesis

presented for the degree of

Doctor of Philosophy

in the Faculty of Science of the

University of St Andrews

by

Derek R Higgins, BSc

St Leonard's College 1988

Declaration

I, Derek Higgins, 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.

Signed .

Date ...4th..August...1988

I was admitted to the Faculty of Science of the University of St Andrews under Ordinance General No 12 on 1st October 1985 and as a candidate for the degree of PhD on 9th October 1986.

Signed **Date** 4th August 1988

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

Signature of Supervisor . Date 4/8/88

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Signed ..... Date 4th August 1988

ABSTRACT

This thesis is in two parts: Part I includes chapters 2 to 8, part II, chapters 9 to 12. Part I describes the development of a program which allows the calculation, display and comparison of Molecular Electrostatic Potentials of molecules using a simple point charge model.

Test calculations are reported which compare the results obtained using this method and AM1 optimised geometries and charges with more accurate ab-initio methods.

The program described above has been used to study the MEP's of the inhibitors of the enzyme poly(ADP-ribose) transferase (ADPRT), in conjunction with theoretical calculations performed using the AM1 method.

Some compounds are suggested which may be powerful inhibitors of ADPRT but at the time of writing are untested.

Part II of this thesis reports the results of frequency calculations on the rotamers of nitrosomethanol.

Preliminary work on the development of a semi-empirical method for theoretical calculations on large systems using geminal type wavefunctions is described.

CONTENTS

PART I

1 INTRODUCTION

The introduction of the fundamental equation of quantum mechanics by Schrödinger [1] in 1926 brought with it a mathematical model for describing the structure of individual atoms or molecules. It also provided a method by which the outcome of a chemical experiment could, in principle, be predicted without that experiment ever being performed. The implications of such a method must have seemed quite foreign to the traditional view of chemistry as an experimental science concerned with the study and analysis of chemical reactions and with the measurement of the physical properties of matter. In addition the postulates involved were difficult to understand since they introduced concepts which were far from everyday experience.

The Schrödinger equation is an equation whose solution represents the motion of electrons, in the potential field of each other, and of their associated nuclei. The way in which the equation is formulated means that its solution involves solving a partial differential equation in $3n$ dimensions, where n is the number of electrons present in the system. It is not generally possible to perform any further analytical reduction to give an equation of smaller dimensions.

The main difficulty in the application of the Schrödinger equation lies in the calculation of the interactions between electrons. This is not a problem for one-electron systems, where there are no electron repulsion terms, and so the resulting Schrödinger equation can be solved easily. This type of equation can be solved by analytical methods for one-electron systems with one or two atoms, and by numerical methods for systems with 3 or more atoms. In the case of many electron systems the exact analytic solution of the Schrödinger equation is impossible and so many methods have been developed to obtain approximate solutions to

the equation. Examples of such approximations include the neglect of relativistic effects, the introduction of the Born-Oppenheimer approximation, the total or partial neglect of electron correlation and the neglect of inner-core electrons in semi-empirical methods.

The use of approximations does not detract from the usefulness of the Schrödinger equation in any way, since the use of modern digital computers has led to the development of many sophisticated approximate quantum mechanical methods. The larger the chemical system to be studied, the cruder the approximations which must be employed. Such approximate methods, however, should not be judged on the crudeness of the approximations employed, but rather on their ability to predict experimental facts and by their general applicability.

Of all methods in common use the ab-initio Hartree-Fock approximation first formalised for molecules by Roothaan [2] is of particular importance. This method is the zeroth-order approximation of more accurate many-electron theories. The uses of this method have been well documented [3] and the approximations used in its formulation can be gradually eliminated. It is possible to carry out reasonably accurate ab-initio calculations on molecules containing around 20 atoms using this method although the size of molecule which can be handled satisfactorily, on any given computer system, will depend upon the resources available. The use of more accurate methods for molecules of this size is still prohibitive.

Ab-initio Hartree Fock calculations can, of course, be carried out on larger systems by using smaller basis sets or by restricting the calculation in some other way. There are many examples in the literature of basis sets which have been specifically designed for such work [4].

In the study of complex chemical systems, such as those of

biological importance, there is often a need to use theoretical methods as an aid to the interpretation of available experimental data. Such systems need not necessarily be large in which case it would be appropriate to use one of the ab-initio methods for such a study. For larger systems, however, a purely ab-initio study, although desirable, will generally be out of the question because of prohibitive computing costs. In such cases it will be necessary to use one of the semi-empirical methods which are currently available always providing that the results are checked against more reliable ab-initio methods wherever possible.

The most popular of the semi-empirical methods in common use are those based upon the NDDO approximation such as MNDO [5] and the more recent AM1 [6] method. Both are examples of semi-empirical molecular orbital methods which have been parameterized against the properties of a large range of organic molecules. The success of each method in predicting molecular properties is due, in part, to the quality of the parameterized atomic functions which are used. In practice, MNDO and AM1 are generally found to give results which compare well with those obtained from ab-initio calculations using a split-valence basis set. There has, however, been some controversy over such claims [7]. The advantage gained by using such methods is that they only require a fraction of the time required for a similar ab-initio calculation. The MNDO method, in particular, has been well tested and has been used successfully in the study of many biological systems [8-11]. Such widespread use has, however, shown that MNDO suffers from several weaknesses [12-14], the most important of which is its tendency to overestimate the repulsive interactions between atoms. This results in a failure to reproduce hydrogen bonding effects for example, which has limited the use of MNDO in biological applications where hydrogen

bonding is involved.

The recently introduced AM1 method is a modified version of MNDO which has been specifically designed to overcome some of the weaknesses inherent in the MNDO model. The modifications made to the core-repulsion function, (CRF), and the re-parameterization of the model have shown that AM1 performs more reliably than MNDO in most cases [6]. The shortcoming of MNDO to reproduce hydrogen bonding has been improved upon in AM1. The original version which was only parameterized for the elements C, H, N and O, has recently been extended to include halogen parameters [15]. As yet there have been few studies reported using AM1.

Details of the AM1 method and the results of calculations comparing AM1 with both MNDO and ab-initio methods are reported in chapter 5. These calculations clearly show that the use of AM1 for the determination of the conformations and properties of molecules is preferred over MNDO. The majority of the semi-empirical calculations reported in this thesis have therefore been performed using AM1.

Another area where the use of modern computer technology has been applied to chemical problems is in the field of computer, or molecular, graphics. Whereas computer graphics is a well recognized and established field in computer science, applications in chemistry have, until recently, not been very widespread. The reasons for such limited use, in the past, are many and will not be discussed here. The potential applications of computer graphics to chemistry are so great, however, that graphics systems are now being used in the examination of structural geometries, molecular shapes, electronic properties etc.

Most chemists will agree that the use of mechanical models, such as the Drieding, Minit, or CPK models, can greatly facilitate the visualisation

of the three-dimensional structure of molecules. For non-planar and complex 3-D molecules such as enzymes, and many other biological compounds, the use of models is essential.

The earliest applications of computer graphics to chemistry were mainly in the study of macromolecular structures such as proteins and nucleic acids. Much use was also made for the display of information obtained from X-ray crystallography [16]. More recently use has been made of computer graphics in the modelling and design of drugs [17-19]. Many of the graphics programs available today allow chemists with a limited knowledge of quantum mechanics to build molecular structures directly on the screen of a graphics terminal, and then perform molecular mechanics, semi-empirical or even ab-initio calculations on the final structure. The results from such a calculation can then be read in to the same program to enable the display of conformations, properties etc.

The choice of molecular geometry used to start a calculation can either be guessed, using chemical knowledge to estimate bond lengths, dihedral angles etc, or can be based upon the results of previous calculations. An alternative approach is to obtain the molecular geometry, or that of a similar compound from a database such as the Cambridge crystallographic data base [20]. This particular data base contains crystallographic data on some 65,000 compounds and allows for on-line search and retrieval of coordinate data. A similar database exists for protein structures, namely the Brookhaven Protein Data Base (PDB) [21], which can be used to obtain protein structure information. Some use has been made in this work of the databases mentioned above, and of the graphics package CHEM-X [22].

Recently, a graphics program has been developed in this

laboratory for displaying the electrostatic potential of molecules in three dimensions. The MEP is depicted by a series of contour levels, colour coded on the van der Waals surface of the molecule. The original version of this program was developed on a small Tektronix 4170 microcomputer with the graphics output being displayed on a Tektronix 4109 colour terminal. A VAX based version of the program has also been produced. The recent addition to the laboratory of a Microvax II/GPX workstation [23] has significantly increased the resources available for both the calculation and display of properties such as the electrostatic potential. This thesis describes the development and extension of the program described above to make use of the improved graphics facilities now available.

The basic structure of this thesis is in two parts. Part I deals with the development of the graphics program described above and also reports the application of this program to the study of the inhibitors of the enzyme poly(ADP-ribose) transferase (ADPRT). This enzyme, described in detail in chapter 2, is required for efficient repair of DNA after certain kinds of damage. The involvement of the enzyme in DNA repair is thought to occur because DNA ligation is regulated by ADP-ribosylation. The importance of ADP-ribosylation in cancer research lies in its involvement in DNA recombination, sister chromatid exchanges, chromosome aberrations and certain categories of cell differentiation. The potential also exists for the design of more powerful inhibitors of the enzyme for use in the treatment of cancer, since inhibitors of the enzyme have been found to increase the cytotoxicity of various DNA damaging agents.

Part II reports the results of other work carried out during the last three years. Chapter 9 reports the results of ab-initio and

semi-empirical studies on the nitrosomethanol system. Chapter 10 gives an account of the progress made in the development of a program system for the treatment of large molecules using a geminal type wavefunction.

2 Poly(ADP-ribose) transferase

Poly(ADP-ribose) transferase (EC 2.4.2.30) is a chromatin bound enzyme, located in the nucleus of eukaryotic cells, which catalyzes the polymerisation of ADP-ribose to form poly(ADP-ribose). This enzyme is also called poly(ADP-ribose) synthetase or poly(ADP-ribose) polymerase but in the remainder of this thesis the term ADRPT will be used to refer to the enzyme.

ADPRT is capable of catalyzing the three types of reaction shown in figure 2.1 below.

i) **Initiation**

NAD + acceptor $\longrightarrow$ ADP-ribose-acceptor + Nm

(carboxylic ester)

ii) **Elongation**

n NAD + ADP-ribose-acceptor $\longrightarrow$

$(ADP\text{-ribose})_n$-ADP-ribose-acceptor + nNm

(1"- 2' glycoside)

iii) **Branching**

NAD + $(ADP\text{-ribose})_n$-ADP-ribose-acceptor $\longrightarrow$

ADP-ribose-$(ADP\text{-ribose})_n$-acceptor + Nm

ADP-ribose

(1"' $\longrightarrow$ 2" glycoside)

where Nm = nicotinamide

Figure 2.1 Three types of reaction which are catalyzed by poly(ADP-ribose) transferase.

2.1 Distribution of ADPRT

ADPRT activity has been found in almost all eukaryotic cells which have been examined for its activity [24, 25] including animal tissues, plants, and lower organisms such as slime moulds, plasmodium (malaria parasites) [26] and yeast, although the existence of the enzyme in yeast is controversial. The enzyme has been purified extensively from many tissues including rat, cow, calf, pig and lamb thymuses, Ehrlich ascites tumour cells [27], and HeLa cells [28]. The first tissues which were found to lack ADPRT activity were mature rat and human granulocytes [29].

Early studies on rat liver cells by Nishizuka *et al*, [30] and Oikawa *et al*, [31] showed that ADPRT activity was mainly localized in the cell nucleus. More recently work by Ikai et al [29, 32] confirmed that poly(ADP-ribose) synthesis and ADPRT activity is localized in the nucleus of rat, bovine and human tissues. This is also supported by the work of Kanai *et al*, [33] on HeLa cells.

2.2 Physicochemical properties of ADPRT

Extensive physicochemical and chemical studies of calf thymus ADPRT, carried out by Ito *et al* [34], show that the enzyme is a globular protein with molecular weight in the range 110,000 to 120,000. Ito found that the enzyme was a Lysine-rich basic protein. Similar properties have been reported by other groups using enzyme preparations obtained from various sources [35, 36].

Work by Kameshita *et al*, [37] has shown that the enzyme consists of three proteolytic fragments which have been identified as the substrate-binding domain, the DNA-binding domain, and the automodification domain. The automodification domain is so called since

the enzyme itself is found to be capable of accepting the poly(ADP-ribose) polymer: the site at which this takes place is called the automodification domain. Cleavage of the enzyme by limited proteolysis with α-chymotrypsin results in the formation of two fragments of MW 54,000 and 66,000 respectively. Kameshita has shown that the 54K fragment contains the substrate-binding site. Further treatment of the 66K fragment with papain results in cleavage into two fragments of MW 46,000 and 22,000. The 46K fragment is found to retain the DNA-binding domain and the 22K fragment the site(s) for accepting poly(ADP-ribose) (the automodification domain).

More recently Kurosaki *et al* [38] have predicted the complete amino acid sequence of ADPRT and have deduced the structural characteristics of the three domains described above from cDNA sequencing. From the coding sequence they predict that the enzyme consists of 1,013 amino acid residues with a molecular weight of 113,203. The amino acid sequence compares well with that obtained from a previous determination by Ushiro *et al*, [39].

Based upon the amino acid composition predicted for each domain, Kurosaki found the DNA-binding domain to be rich in lysine residues and to be quite basic, with a net charge of +15. This suggests that DNA binds to this domain by ionic interactions. The automodification domain was also found to be lysine-rich with a net charge of +7. The NAD-binding domain, on the other hand, was found to be neutral overall.

2.3 Enzymology of ADPRT

Although the exact physiological function of ADPRT is not fully understood the overwhelming evidence suggests that it is involved in biologically important processes such as DNA repair and replication, RNA

synthesis, and cell differentiation. The role which ADPRT plays in the above processes and how the gene for ADPRT is regulated in eukaryotic cells is still unknown. The list given below summarises some of the most important enzymological properties of ADPRT.

(a) ADPRT is almost entirely dependent on DNA

(b) Only double-stranded DNA is effective with single stranded DNA being inhibitory [40]

(c) The enzyme is activated by nicks in the DNA [40, 41]

(d) β-NAD functions as the donor of the ADP-ribose unit

(e) Histones may function as acceptors of poly(ADP-ribose) in *vitro* [42] and in *vivo* [43]

(f) Evidence suggests that the enzyme itself is the major acceptor of the polymer. This occurs at the automodification domain described earlier [44]

(g) ADP-ribosylation activates DNA ligase activity [45]

(h) The enzyme can be selectively inhibited by various small molecules

2.4 ADPRT inhibitors

Four chemical classes of ADPRT inhibitors are known; nicotinamides (such as 5-methylnicotinamide), pyrazinamides, purine analogues (such as theophylline and theobromine), and aromatic amides such as benzamides. Figure 2.2 shows the structures of some of the compounds, from the 4 classes of compounds mentioned above which have been tested for their ability to inhibit ADPRT. A selection of these inhibitors together with details of their inhibitory power are listed in Table 2.1. It is clear from Figure 2.2 that most of the inhibitors shown are similar in structure to the nicotinamide moiety of the substrate NAD. Most of the inhibitors

which have been tested seem to function as analogues of NAD and compete with NAD for binding to the receptor. The analogy between NAD and the purine inhibitors is not immediately obvious.

The following sections describe the main features of the four different types of inhibitors.

2.4.1 Nicotinamide inhibitors

Of the nicotinamide analogues which have been tested nicotinamide (compound 1, Figure 2.2) and picolinamide (compound 2) are the most effective ADPRT inhibitors. Some modifications to the structure of these molecules reduces their potency as ADPRT inhibitors while other modifications completely abolish their ADPRT activity. Movement of the carboxamide group from position 3 to position 4 in the ring to give isonicotinamide (compound 6) reduces inhibitory capacity by 20% from 89% to 69% (Table 2.1). Alteration of the carboxamide group itself produces more drastic changes in potency as in 3-acetylpyridine (compound 5) where the inhibitory capacity is only 27%.

If the carboxyamide group is replaced with a carboxylic acid group as in nicotinic acid (compound 4), then the ADPRT activity is completely lost. This is also the case with isonicotinic acid which does not cause any inhibition. Substitution at the ring nitrogen does however reduce potency.

The observations described above suggest that a carboxamide group at position 3 or 4 on the pyridine ring is necessary for good inhibition. An aromatic ring is also essential since inhibition is completely abolished when the pyridine ring system is completely saturated.

2.4.2 Benzamide analogues

The benzamide based compounds are the most potent inhibitors of ADPRT. Table 2.1 shows that benzamide, 3-aminobenzamide and 3-methoxybenzamide all cause inhibition of ADPRT activity by 96%, when present at a concentration of 1mM. When the concentration is decreased to 50 μM, however, it is possible to rank these inhibitors in order of potency. In the case of 3-aminobenzoic acid, a decrease in concentration from 1mM to 50 μM results in 10% increase in activity.

The results with benzamide analogues show that a ring nitrogen is not necessary for ADPRT inhibition. The trends observed for the benzamide analogues also follows those observed for the pyridine analogues. The carboxamide group is also essential for activity. Substitution in the ring at positions 2 or 6 with $-NH_2$ groups also cause a slight loss of potency, which is attributed to internal hydrogen bonding between the $-NH_2$ group of hydrogens and the carboxamide carbonyl oxygen.

Chapters 6 and 7 present a more detailed study of the benzamide based ADPRT inhibitors and offers some explanation of their relative potency.

2.4.3 Pyrazinamide analogues

The main observations from studies of the pyrazinamide analogues show that the presence of a pair of nitrogen atoms at positions 1 and 4 in the ring does not interfere with ADPRT inhibition. Studies also suggest that a carboxamide group is essential for inhibition.

In general the pyrazinamide analogues are less potent inhibitors than either nicotinamide or benzamide based inhibitors.

2.4.4 Purine analogues

Figure 2.2 shows three purine type analogues which have been tested for ADPRT inhibition. Since these compounds are structurally different from the benzamide or nicotinamide inhibitors, they are thought to compete for a different part of the NAD binding site on the enzyme. The figures given in table 2.1 show that both theophylline (compound 17) and theobromine (compund 16) are relatively potent ADPRT inhibitors. A study of both theophylline and theobromine is also presented in Chapter 6.

2.5 ADPRT and DNA damage

As mentioned previously, all available evidence suggests that poly(ADP-ribose) is involved in the regulation of cell proliferation, differentiation and DNA repair. Many studies have been done in an attempt to determine the ADPRT activity, and concentrations of poly(ADP-ribose) and poly(ADP-ribosylated) proteins present in various types of cells.

Several groups have reported a dramatic increase in ADPRT activity after DNA damage has occurred [46]. The involvement of ADPRT in DNA repair was first shown by Durkacz *et al* [47] when they exposed L1210 mouse leukemic lymphoblast cells to dimethyl sulphate causing breaks to occur in the DNA. Five hours after removal of the dimethyl sulphate, the DNA was found to be of similar length as the undamaged DNA and thus had presumably been effectively repaired. When the ADPRT inhibitor 3-aminobenzamide was included in the reaction mixture, the rejoining of the DNA was almost completely inhibited. These results were consistent with the idea that ADPRT is involved in DNA excision repair, and is strongly activated by DNA containing single-strand

breaks.

Despite the above findings and all evidence gathered since this discovery, it is still not possible to assign a specific role to poly(ADP-ribosylation) in DNA repair. A common theory used to explain the role played by ADPRT in DNA repair postulates an activation of DNA ligase II by the process of ADP-ribosylation [48]. Recently Cleaver *et al* proposed an alternative hypothesis which suggest that ADPRT regulates an intracellular endonuclease [49].

More recently Cantoni *et al* [50] have suggested that DNA fragmentation is not the only factor which controls ADPRT activity. Cantoni has demonstrated that ADPRT activity can be induced by chilling Chinese Hamster Ovary (CHO) cells to ice temperature for 60 minutes followed by a return to normal culture conditions. The effect of this cold shock is to decrease the cellular NAD^+ content which can be prevented by the ADPRT inhibitor 3-aminobenzamide. This suggests that ADPRT activity is responsible for the drop in the NAD^+ content although other mechanisms such as decreased synthesis of NAD^+ cannot be discounted.

The apparent induction of ADPRT activity in the absence of DNA-damage has also been reported by Wallace *et al* [51] in polyamine depleted cells and by Jackowski *et al* [52].

An interesting hypothesis has been proposed by Wintersberger *et al* [53] who suggest that the activation of ADPRT by DNA damage serves to decrease rapidly and transiently the cellular level of NAD by synthesis of poly(ADP-ribose). This results in a slow-down of energy requiring reactions giving cells more time to repair DNA damage.

2.6 ADPRT and cancer research

Since inhibitors of ADPRT retard DNA repair, then this has two important biological consequences: inhibitors can potentiate the cytotoxicity of small alkylating agents, and bring about an enhancement of sister chromatid exchanges (SCE) and chromosomal aberrations. Many mutagens and carcinogens are known to increase SCEs but the inhibitors of ADPRT are not known to be mutagenic or carcinogenic.

As mentioned previously, Durkacz *et al* reported that 3-aminobenzamide potentiates the killing of L1210 cells *in vitro* by dimethyl sulphate. This property of ADPRT inhibitors to potentiate cell killing by DNA damaging agents could be applied in treatment of cancers. As an example, benzamide has been found to enhance the killing of HeLa cells by bleomycin [54] and to potentiate the anti-tumour effect of bleomycin on Ehrlich ascites cancer cells.

This suggests that the ability to predict compounds which are more potent inhibitors of ADPRT might be extremely useful in the field of cancer chemotherapy.

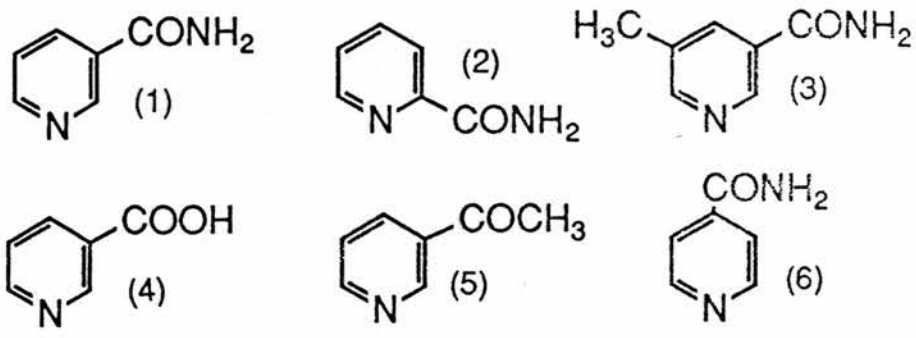
2.7 Theoretical study of ADPRT inhibitors

We are not aware of any previous studies made on ADPRT inhibitors in connection with their ability to inhibit the enzyme activity. We have undertaken a detailed theoretical study of the inhibitors of ADPRT using modern theoretical methods and computer graphic techniques. The calculations reported in Chapter 6 are based upon the relative potency of the inhibitors already mentioned in this chapter and upon the work of Cantoni [55]. Chapter 7 of this thesis extends the calculations to possible compounds which may be more potent ADPRT inhibitors but as yet are untested.

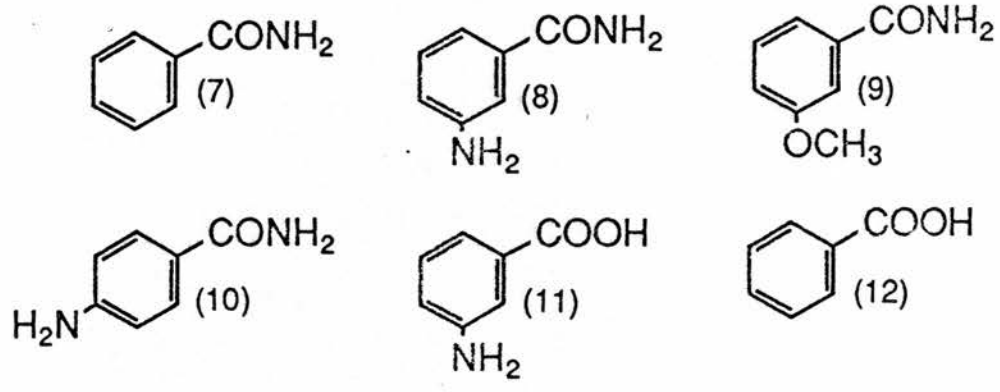
2.8 Reviews of ADPRT

It is appropriate to mention here some excellent and comprehensive reviews which have appeared on various topics concerned with ADPRT. The review by Shall [56] gives an overview of the relationships involved between ADP-ribosylation, DNA repair, cell differentiation and Cancer. The review by Gaal *et al* [57] gives an exellent account of the cellular roles played by ADP-ribosylation reactions. Finally the review by Ueda and Hayaishi [58] sumarises the important findings in the wider aspects of both mono and poly(ADP-ribosylation) processes.

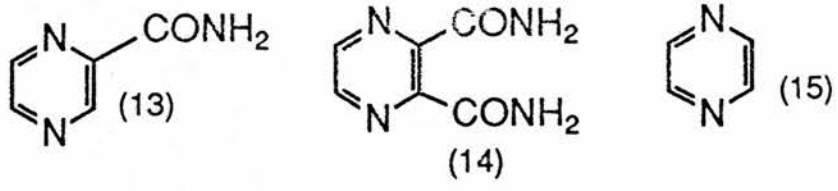
Nicotinamide analogues



Benzamide analogues



Pyrazinamide analogues



Purine analogues

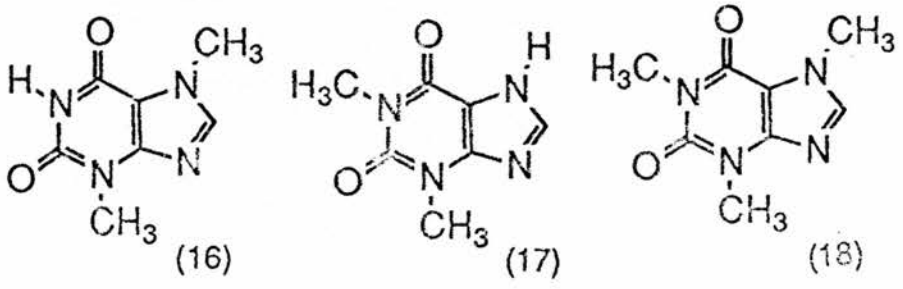


Figure 2.2 Structures of the ADPRT inhibitors listed in Table 2.1.

Table 2.1 Inhibitors of poly(ADP-ribose) transferase

Number	Compound	% inhibition (1mM)[a]	%inhibition (50 μM)[b]
Nicotinamide analogues			
1	nicotinamide	89	63
2	picolinamide	86	
3	5-methylnicotinamide	83	
4	nicotinic acid	0	0
5	3-acetylpyridine	27	0
6	isonicotinamide	69	
Benzamide analogues			
7	benzamide	96	96
8	3-aminobenzamide	96	90
9	3-methoxybenzamide	96	98
10	4-aminobenzamide	71	
11	3-aminobenzoic acid	0	10
12	benzoic acid	0	
Pyrazinamide analogues			
13	pyrazinamide	78	
14	pyrazinedicarboxamide	15	
15	pyazine	0	
Purine analogues			
16	theobromine	81	
17	theophylline	89	
18	caffeine	35	

[a] data taken from reference [59]
[b] data taken from reference [60]

3 Theoretical methods and properties

This chapter gives a brief outline of the ab-initio and semi-empirical methods used in this work and describes the molecular properties which have been calculated using these methods. It is not intended to provide rigorous proofs of the theory behind such calculations; instead a qualitative description of their application is provided.

3.1 Semi-empirical methods

The semi-empirical methods used were limited to those available as options in the MOPAC program package [61]. As mentioned previously the majority of the semi-empirical calculations reported here were carried out using AM1. Full details of this method are given in the original papers [6]. Details of test calculations using AM1 and the reasons for choice of this method are discussed in Chapter 5.

Briefly, AM1 has been designed to overcome the main deficiencies of the MNDO method while still retaining the same basic approximations. The major changes are in the modification of the MNDO core repulsion function (CRF), to compensate for excessive interatomic repulsions at large separations, and a reparameterisation. The modifications to the CRF involved supplementing the original CRF with additional attractive and repulsive Gaussian terms resulting in a method which represents a considerable improvement over MNDO. In particular, the ability to reproduce hydrogen bonding satisfactorily should make AM1 especially useful in biological applications.

3.2 Ab-initio methods

In most cases the ab-initio calculations reported here were performed using the Gaussian 82 [62] or more recent Gaussian 86 [63] system of programs. In cases where problems were encountered with convergence in the SCF links, then the GAMESS program [64] was used since this program allows more control over convergence using techniques such as level shifting and damping [65].

For the ab-initio calculation of the electrostatic potential a modified version of the Gaussian 80 [66] program was used. This program uses an algorithm which is around 5 times faster than the equivalent routine in Gaussian 82, hence the reason for its use. Input to the program consists of two data files, the first of which contains control cards for the program options along with the molecular geometry. The second file contains the coordinates of each point at which the MEP is to be calculated. Output consists of a file containing the coordinates and the electrostatic potential at each point. This file can then be used as input to subsequent processing programs to give a file which is suitable for display.

3.3 Atomic charges

One of the most important areas where the use of ab-initio or semi-empirical calculations may lead to a better understanding of the underlying processes is in the study of the reactive properties of molecules. A concept which can be useful in such studies is that of atomic charge, where the charge associated with each atom can be used as an indication of the susceptibility of a molecule to electrophilic or nucleophilic attack at that position. The drawback with this approach is that atomic charge is not a physically observable property and cannot be

rigorously calculated.

There are, however, various schemes which have been devised to calculate atomic charges, although the concept upon which they are based may be an arbitrary one since the charges obtained are dependent upon the scheme used. The most common of such schemes is that due to Mulliken [67] and is the scheme which is used in most ab-initio programs including the Gaussian programs. Atomic charges obtained using this, and many other, methods are dependent upon the basis set used for the calculation. Although it would seem attractive to use the largest basis set possible, the effect of simply adding more basis functions does not guarantee that a better charge distribution will be obtained. In each case the choice of basis set must be made carefully to ensure that a balanced basis set is obtained which does not favour any specific part of the molecule.

As a result of the weaknesses inherent in the calculation of atomic charges, the values obtained from such calculations are not normally used as the sole predictor of molecular reactivity. Instead, it is more usual to use atomic charges as an aid to interpreting reactivity in conjunction with other properties or to use the charges in the calculation of some more reliable index of reactivity. One such index where charges may be used is in the calculation of the molecular electrostatic potential as described below.

3.4 Dipole moment

The dipole moment is a one-electron property which may be used to give an indication of the overall charge distribution in a molecule. It can easily be calculated for a closed system molecule providing a means by which the calculated and measured charge distributions can be

compared. Experimental information on dipole moments for large molecules is scarce and so such comparisons can only be made in a few cases. Where comparisons are possible, however, some trends are observed.

In the case of ab-initio dipole moments, those calculated using the STO-3G basis set are generally smaller than the corresponding experimental values. The use of larger basis sets does not necessarily guarantee better results, although the addition of polarisation functions to small basis sets, such as 3-21G, often gives improvements. The poorest agreement is usually found in molecules which contain strongly electronegative or electropositive atoms [6].

In the case of semi-empirical dipole moments calculated using AM1 or MNDO, the agreement with experiment is generally satisfactory. The AM1 method has been found to give smaller average errors than MNDO [6].

Overall, the calculated dipole moments within a series of related molecules are generally found to follow the experimental trends. This is true for both ab-initio and semi-empirical methods.

3.5 Molecular electrostatic potential (MEP)

The molecular electrostatic potential (MEP) at any point **r** (around a molecule), which is generated by its electrons and nuclei, is defined as

follows:

$$V(\mathbf{r}) = \sum_{A} \frac{Z_A}{|R_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|}$$

Z_A = the charge on the nucleus A.

R_A = coordinates of nucleus A.

$\rho(\mathbf{r}')$ = electronic density function for the molecule.

This equation is the difference between two terms:

$$V(\mathbf{r}) = \text{nuclear term} - \text{electronic term}$$

Since both terms have opposite signs then the MEP, $V(\mathbf{r})$, is the net electrostatic effect produced at point $\mathbf{r}$ by the total charge distribution of the molecule.

The MEP is a real physical property which can be measured using electron scattering techniques [68], unlike atomic charge which cannot be physically measured. The main use of the MEP is in obtaining a reliable indication of molecular interactions such as those sites around a molecule which may be attacked by electrophiles. In such cases its use may be qualitative rather than quantitative. The use of the calculated molecular electrostatic potential as a guide to the reactive sites of molecules was pioneered by Bonaccorsi and co-workers [69, 70].

One area where the use of MEP's is potentially most informative is the study of biological processes including hydrogen bonding and the interaction of drugs with receptors. There are many examples in the

literature where the MEP has been used in an attempt to correlate biological activity with electrostatic features, including nucleic acid bases [71, 72], enzyme active sites [73], the opiate receptor [74], neuroleptic drugs [75] and DNA [19, 76]. Weinstein, in particular, has explored the extents and the limitations of the method [77-79].

A major drawback to the use of the MEP in biological applications is that its calculation is time consuming and restricted where ab-initio methods are used. The length of time taken to calculate V(r) depends upon the number of two and three-centre nuclear-electronic attraction integrals which must be calculated at each point. For a wavefunction described by **n** basis functions there are typically **½n(n + 1)** of such integrals. Table 3.1 gives an indication of the amount of CPU time required in calculating the MEP for a range of molecules, using various basis sets. The times given in this table represent the time taken for the whole calculation which is the sum of the time taken in both SCF and MEP parts of the program.

It is clear from these figures that the time involved to calculate the MEP for molecules with more than about 10 atoms is considerable, especially where large basis sets are used. For the largest molecules in this table, the STO-3G basis set is the smallest which is practicable yet still involves large computation times. This fact, more than any other, limits the use of ab-initio MEP calculations for large systems unless unlimited computing resources are available.

One approach which enables the MEP to be calculated for large molecules is to introduce various levels of approximation into the basic equation for V(**r**). The simplest possible approximation is the point charge model [80] where the MEP is approximated by replacing the molecular charge distribution with a set of point charges. The accuracy

of such a model ultimately depends upon the source and location of the point charges. This method has been used throughout this work to calculate the MEP unless otherwise stated. The equation used to calculate the MEP is given below:

$$V(\mathbf{r}) = \sum_{A} \frac{q_A}{|\mathbf{r}' - \mathbf{r}|}$$

q_A = atomic charge on nucleus A centred at r'

$\mathbf{r}$ = point where MEP is to be evaluated

Although this model may seem to oversimplify the calculation of the MEP, it is applicable to any size of molecule as long as the atomic charges can be calculated or estimated. It also produces surprisingly good results for what would appear to be a drastic simplification of the original equation. In all cases the MEP's reported were calculated using AM1 molecular geometries and charges. Test calculations comparing this method with ab-initio MEP's are reported in Chapter 5.

Another approach which can be used is to approximate the integral evaluation in some way. The aim of such approximations is usually to decrease the number of integrals which must be calculated thereby reducing the time necessary for calculation.

3.6 Molecular Geometries

The molecular geometries reported in this work have been obtained using the optimisation routines available in the specific programs. For the ab-initio optimisations using Gaussian 82, the Berny routine [81] was used except in some cases where this routine failed, in which case the

Murtagh-Sargeant [82] routine was used. One other routine which has been used is that developed by Baker [83] for use in Gaussian 82. This routine is designed for locating transitions states using an eigenvector following technique and may also be used to locate minima. Part II of this thesis reports the results of calculations where this routine has been used with great success to locate transition states in the nitrosomethanol system.

The semi-empirical geometries have been obtained using the Fletcher-Powell (FLEPO) routine [84] in the MOPAC package with the PRECISE option also being used. This routine gave reliable results with the gradient norm falling below 1.0 in most cases. For some structures, however, the gradient norm could not be reduced to a low enough value using the FLEPO routine or any of the other optimisation routines available in the MOPAC package. This situation was not satisfactory since the energy difference between some conformations of a molecule were often less than a few tenths of a kilocalorie which meant that the gradient norm in these molecules had to be as low as possible to ensure that the energy difference between conformations could be properly calculated.

This problem was overcome by using the Gaussian 86 program instead of MOPAC in such cases. As mentioned earlier Gaussian 86 has the facility to perform semi-empirical calculations which has the advantage that the more reliable Berny and MS optimisation routines available in this program may be used to obtain properly optimised geometries.

3.7 Vibrational Frequencies and stationary points

Part II of this thesis reports the results of vibrational frequency calculations on the nitrosomethanol molecule. This area of computational chemistry is interesting from a fundamental viewpoint, since the effectiveness of such calculations can easily be judged by how well the results agree with experimental infrared and Raman spectra.

It is often the case, that calculated vibrational frequencies and assignments can be used as an aid to the correct identification of fundamentals in experimental vibrational spectra. Other areas where calculated frequencies are important are in cases where other techniques cannot readily be applied, such as short lived intermediates, and in identification of fundamentals in interstellar spectra [85, 86].

One less exotic, but arguably more important, use for calculated force constants and frequencies is in the characterisation of stationary points on potential energy surfaces. For a molecule with M atoms the gradient of the potential energy is zero at a stationary point:

$$\partial g_i = \frac{\partial E}{\partial X_i} = 0 \text{ for } i = 1, 3M$$

X_1, X_2 ... X_{3M} are the Cartesian coordinates for the M atoms.

The nature of the stationary point can be characterised by the Cartesian force constant matrix.

$$F_{ij} = \frac{\partial^2 E}{\partial X_i \partial X_j} = \frac{\partial g_i}{\partial X_j}$$

where the F_{ij} are the individual force constants. If this matrix is positive definite then the stationary point is a true minimum on the surface. If the force constant matrix has one negative eigenvalue then the stationary point represents a saddle point or transition structure. Stationary points may also exist where the force constant matrix has more than one negative eigenvalue or where zero eigenvalues arise. Such points represent special features on the PE surface and will not be discussed here.

The negative eigenvalues of the force constant matrix give a measure of the negative curvature along the axes where the stationary point is located. The corresponding eigenvectors may be used to obtain useful information about the atoms involved in, and the paths leading away from, the transition structure. Use of this property is made in Baker's eigenvector following optimisation routine mentioned earlier.

A vibrational analysis can be carried out which requires mass weighting of the Cartesian force constant matrix and transformation of the coordinates to give 3M equations each of which depends upon a single coordinate. For each so called "normal coordinate" a harmonic oscillator equation can be written and solved to give a force constant λ_i. For non linear molecules only 3M-6 of the force constants (λ_i) are useful.

The method by which a vibrational analysis is usually carried out is the GF matrix formulation of Wilson [87]. This method is used in most of the popular ab-initio and semi-empirical programs which allow frequency calculations to be carried out and involves transformation to a coordinate system where the 3M-6 degrees of freedom of the molecule are specified in terms of internal coordinates such as bond lengths and bond angles. In this way the force constants and therefore the vibrational frequencies obtained are easier to interpret in chemical terms. The basis

of this method and the theory of molecular vibrations is described in the standard text by Wilson *et al* [87].

All ab-initio frequencies reported in this work have been calculated using the programs described in section 3.2 above. Where possible the frequencies were calculated analytically, failing which numerical methods were used. Semi-empirical frequenices were calculated using the FORCE option in the MOPAC program.

In some of the calculations reported in part II of this thesis an analysis of the vibrational modes contributing to each fundamental frequency is given. This information has been obtained using a program written specifically for this purpose [88] which gives a simple description of the vibrational components of each frequency in terms of the internal coordinates of the molecule.

Table 3.1 **Typical times taken for ab-initio electrostatic potential calculations on various sized molecules.**

Basis set	No of basis functions	No of points	Time[a]
Imidazole (9 atoms)			
STO-3G	29	752	00:46
3-21G	53	743	02:08
6-31G	53	743	02:25
6-31G*	83	743	08:50
Benzamide (16 atoms)			
3-21G//AM1	95	708	10:50
5-hydroxytryptamine (17 atoms)			
3-21G//AM1	104	706	13:55
8AV1[b] (30 atoms)			
STO-3G//AM1	94	1160	20:30
4AV1[b] (35 atoms)			
STO-3G//AM1	103	1361	26:12

[a] times quoted are in hours and minutes. Calculations performed on a Microvax II using program in reference [66].

[b] 8AV1 and 4AV1 are codes given to derivatives of benzamide.

4 3D2 Graphics Program

This chapter describes the development of the 3D2 graphics program and gives an account of its capabilities. As mentioned earlier 3D2 is based upon the 3D1 program previously developed in our laboratory [89]. A brief description of the 3D1 program is given below, a more detailed description may be found elsewhere [90]. Unless otherwise specified all MEP values reported in this chapter are in atomic units.

4.1 3D1

The 3D1 program was originally developed on a Tektronix 4170 microcomputer, running under CP/M-86, using a Tektronix 4109 colour graphics terminal as the display device. The program was written in Fortran-86 using a set of Direct Terminal Interface (DTI) routines [91] to handle all the graphics output to the 4109. The 4109 terminal is a 4 bit-plane device capable of displaying 16 colours on the screen, at one time, from a pallette of 256 colours. Hardcopy output was available by dumping the screen display directly to a Tektronix 4695 colour plotter.

The basic idea of the 3D1 program was to depict the molecular electrostatic potential on the van der Waals surface of a molecule using coloured polygons. Each polygon was drawn at the point on the surface where the MEP was calculated and coloured according to the value of the MEP. Since the 4109 terminal was limited to displaying a maximum of 16 colours it was necessary to partition the MEP values into a number of ranges, each range being assigned a specific colour. In practice only 8 colours were used for display purposes since this number was found to give a balance between the legibility of the display and its usefulness as a predictive tool. Table 4.1 shows the ranges used and the colour

associated with each range. Figure 4.1 shows a typical display from the 3D1 program, using the colours and ranges given in table 4.1, for the STO-3G MEP of benzamide.

From figure 4.1 it is clear that the size of the polygons is such that they overlap slightly forming a polygonal mesh over the surface of the molecule. Each polygon is also tilted according to a vector perpendicular to its surface point which has the effect of drawing the polygon tangential to the surface. This creates a surface which looks more 3-dimensional than if the polygons were simply drawn flat. The polygons are also drawn from the back of the molecule to the front so that the forward facing polygons obscure those behind. This eliminates the need for the use of hidden surface routines.

Input to the 3D1 program was through a formatted disk file which contained an entry for each point on the surface. Each entry consisted of the coordinates of the surface point, the coordinates of the perpendicular unit vector and an integer corresponding to the colour of the polygon to be drawn at that point. Figure 4.2 shows the steps involved in producing a 3D1 input file.

The first step involves calculation of the molecular surface points using the algorithm due to Connolly [92] and implemented in the MS program [93]. A modified version of Gaussian 80 [66] was then used to calculate the electrostatic potential at each point. The resulting MS and Gaussian 80 output files were then combined and processed using a few utility programs to produce a data file for the 3D1 program.

4.2 Reasons for development of 3D2

Despite the number of steps involved, this method worked well and was adequate for the needs of the laboratory at that time (mainly

ab-initio MEP calculations on small molecules of up to 30 atoms). Our increasing interest in large biological systems and the availability of semi-empirical methods such as AM1, which can handle systems of up to 100 atoms, meant that it was desirable to have a method available for calculating MEP's using data from an AM1 calculation, for example. In this way it would be possible to deal with larger systems instead of being limited by use of ab-initio methods. Although the 3D1 program was capable of displaying the MEP from any source, since it only required a formatted input file, the major difficulty lay in the methods available for calculating the MEP for these large systems.

One method which was tried was to use the CHEM-X program [22] to calculate the MEP. CHEM-X can calculate MEP's using either ab-initio or semi-empirical methods and can display the results in an number of formats. Ab-initio MEP's can be calculated for small molecules using the DENPOT program [94] in conjunction with the GAUSSIAN 80 program already mentioned [66]. Semi-empirical MEP's can be calculated using the VSS program [95]. In all cases the MEP can be calculated, using a simple point charge model wherever charges are available. Once CHEM-X has calculated the MEP it can be displayed in the form of a "dot surface", or a contour map, colour coded in terms of potential. This approach was unsatisfactory, however, since the CHEM-X displays were difficult to interpret and lacked the clarity of the 3D1 type of display. The answer to this problem seemed to be either to modify 3D1 to enable MEP calculations to be carried out directly or to develop a program to calculate the MEP externally and to use 3D1 for display purposes only.

At this time, however, the computing facilities available in our laboratory were greatly improved by the addition of a Microvax II/GPX workstation [23]. It was thus decided to re-write the 3D1 program to

make full use of the powerful graphics capabilities of the GPX system and to incorporate facilities for direct calculation of the MEP within the program.

4.3 Program design

The main function of the 3D2 program is to allow the user to display and manipulate MEP maps which may be calculated by the program itself or which have been obtained in a similar way to the 3D1 program. Facilities are also provided to allow the MEP to be displayed in a variety of formats and to allow comparisons to be made between the MEP maps of different molecules. Input to 3D2 is in the form of unformatted data files for MEP maps and formatted files for all structure and geometry input. The main input file has been modelled upon the CSSR format as used by CHEM-X. This file contains atomic coordinates, charges, and connectivity information and may be used to display structures, or as input for a MEP calculation when charges are present. The input files can be obtained directly from our modified version of MOPAC [96], by addition of a single keyword, making it a simple matter to produce a MEP map from the results of an MNDO or AM1 calculation.

The GPX workstation is based around a Microvax II minicomputer driving a 19 inch VR290 colour monitor which has a screen resolution of 1024 by 864 pixels. The powerful graphics capabilities of this system are provided by the GPX graphics coprocessor and a VCB02 video subsystem providing eight planes of bitmap memory capable of displaying 256 colours from a pallette of 16.7 million. This offers scope for a vast improvement of display quality when compared with the 16 colour Tektronix based system.

The 3D2 program is written in VAX FORTRAN and is built around several core routines, which are responsible for the basic operation of the program, and a set of graphics and utility routines. The graphics routines have been adapted from the 3D1 program and have been modified to include the use of shading and local printing. In its present form the 3D2 program consists of 800 K of Fortran source which produces an executable image of around 230 K after compilation and linking. The program is dimensioned for a maximum of 500 atoms and 15,000 surface points although these limits may be altered easily if necessary. Much care has been taken to ensure that maintaining and modifying the program is relatively straight forward due to its modular design.

4.4 User Interface

The interface between a program and a potential user is of particular importance since it can make its use pleasant and straight forward or an unpleasant and unfriendly experience. A great deal of effort has been spent in making the interface between the 3D2 program and the user as informative and easy to use as possible. Like 3D1, the 3D2 program has been designed around a series of screen displays and menus which are used to display information such as filenames, display parameters and textual output. All of the displays and menus have been constructed using the VMS Screen Management (SMG) routines [97]. Although this restricts use to terminals which are compatible with, or can emulate, the VT52, VT100 or VT200 series terminals, it has allowed a screen based interactive interface to be produced. In any case most of the terminals in our laboratory are compatible with the above.

Figure 4.3 shows the layout of the screen displays of both the 3D1

and 3D2 programs. The 3D1 display is included only for comparison: the descriptions which follow refer to the 3D2 screen display in figure 4.3a. The uppermost display is the status display which holds information such as the current filename and the values of the rotation matrix. Beneath this lies the main display area where most of the textual output and messages generated by the program are displayed. To the right of the main display is a menu which is used to display the current values of the graphics variables and MEP information. This menu is automatically updated when any changes are made to the variable list. The display at the bottom of the screen is the command line which is used to enter all of the program commands. Beneath the command line lies the error line which is used to display most of the error messages generated by the program. This display does not have a border and is cleared whenever a valid command has been entered.

All of the user input is through a "virtual keyboard" on the command line. The virtual keyboard is a logical structure used for input operations and is created and maintained using the SMG routines. The advantage of using a virtual keyboard is that it is device independent and is associated with the physical keyboard on the terminal being used: any commands typed on the terminal will therefore be echoed on the command line.

All 3D2 commands take the form of simple English words which are processed by a number of parsing routines to determine which action or function is required. If the command is valid then the appropriate routines will be called to perform the desired function if possible. In this way the program is easy to use and since the commands are simple words they should be easy to remember. In any case "on-line" help is available at any time which gives a brief description of the commands

available and their appropriate action. The help facility follows the same format as the standard VMS help library and should therefore be familiar to any VAX user.

4.5 Graphical Output

All graphical output from 3D2 is displayed on the VR290 workstation monitor. The code which performs this output has been written using the UIS [98] workstation software routines, rather than GKS routines, since the UIS routines are specifically designed for use on the GPX workstation. In speed of operation, the UIS routines are many times faster than the equivalent GKS routines. The UIS software also allows a number of windows to be created on the workstation each of which can contain a separate image. This facility is useful where it is required to compare the MEP maps for a series of molecules since each map can be drawn in a separate window and the windows manipulated on screen at random.

As well as using the same basic graphics routines as 3D1, the 3D2 program allows surfaces to be shaded. For each of the basic 8 colours used, a series of 14 shades can be obtained. Each shade is calculated according to the value of the perpendicular unit vector. When the vector is in the same direction as the viewpoint the basic colour is used. As the angle between the viewpoint and the perpendicular vector increases then the amount of shading increases until it reaches a maximum where the two are perpendicular. This method, among others, is described in detail elsewhere [99]. The difference between the maximum and minimum shades for each value can be controlled in the 3D2 program by the value of the shading parameter which can lie between 0 and 100; 0 gives no shading and 100 maximum shading. Figure 4.4a

shows the MEP of benzamide calculated at the STO3G//AM1 level and displayed using 3D2.

By default, 3D2 calculates the MEP on the van der Waals surface of a molecule using standard values of the van der Waals radii [100] for the atoms. In some cases, however, it may be desirable to calculate the MEP on the surface created using some multiple of the van der Waals radii. This can be achieved using 3D2 by specifying a van der Waals scale factor before calculating the MEP. In this way all van der Waals radii used in the surface calculation will be scaled by this factor. Figure 4.4b shows the MEP of benzamide on the "double" van der Waals surface. This map was calculated externally using an STO-3G basis set at the AM1 optimised geometry.

The 3D2 program can also display the MEP of a molecule in the form of a "profile map". Figure 4.5 shows the profile map for the benzamide MEP of Figure 4.4a where the MEP value is plotted on the Y-axis against each surface point on the X-axis. The surface points along the X-axis are sub-divided into individual atom contributions so that it is easy to identify which part of the profile corresponds to any individual atom. This is possible from the way in which the MS program calculates the surface points, since it associates each point with a particular atom.

4.6 Additional features in 3D2

This section describes some of the features included in the 3D2 program which were not present in 3D1. Many new features have been added to 3D2, the most important of which are described below.

4.6.1 Structure display

One of the shortcomings of 3D1 was the difficulty involved in visualising the MEP map in terms of the underlying molecular structure. Although this was a trivial matter for small molecules, it often became difficult for larger systems where it was not often clear which parts of the display corresponded to a particular part of the molecule for example. To overcome this problem a few routines were added to 3D2 which draw a stick representation of the molecular structure on top of the MEP map. This can be clearly seen in Figure 4.4c again shows the MEP of the benzamide molecule calculated by 3D2 using AM1 charges and geometry.

The format of this stick drawing can easily be changed to allow thicker bonds to be drawn, or to set the atom colours for example. The default is to draw the molecular structure using a "split-bond" display where the bond between each pair of atoms is split in two with each half being coloured according to the atom type (O = red, N = blue) etc. This allows for easy identification of the molecular structure.

The same routines are also used to display the molecular structure on its own using the "Display/stick" command. This is useful when trying to orientate a molecule to give a desired view since it takes less time to draw the stick structure than the MEP map. When the required view is found the MEP map can be displayed using the "Display/surface" command. Alternatively the "Display/stick" command can be used to display the molecular structure in any CSSR file.

4.6.2 Calculation of MEP

As mentioned earlier in this chapter, and in chapter 3, the 3D2 program can also be used to calculate MEP maps. All that is required

for such a calculation is a structure file containing the molecular geometry and atomic charges. Once this data has been read in, the MEP can be calculated by using the "Calculate potential" command. This creates a VMS subprocess which runs the MS program to calculate the molecular surface. The density of points on the surface may be altered from 3D2 depending upon the quality of display required. This is mainly of use where shading is required since the shaded surfaces require a higher point density to look effective. When the surface calculation has finished, control is passed back to 3D2 whereupon the MEP calculation is started. Calculation of the MEP is performed using the method described in 3.5.

When the calculation is finished the results can either be stored in a file, for later display, or can be displayed directly in one of the formats described above. It is also possible to request a simple analysis of the MEP which can either be printed or displayed in graphical form. This latter option will be described below.

Provision has also been made in 3D2 to allow MEP data, calculated in the same way as for 3D1, to be read in and processed to produce a 3D2 MEP file which can then be displayed as normal. This function does not involve any calculation of the MEP but merely involves processing the external data which may be calculated by any means available.

4.6.3 Calculation of charges

In some cases where a structure file is available but where charges have not yet been calculated, it is possible to use 3D2 to perform a CNDO or INDO single point calculation to obtain a set of charges. Although the charges obtained from such methods are not always ideal they can often be used in place of charges calculated by

some more reliable method. Once the charges have been calculated the MEP can be calculated as described above.

In many cases a MEP map obtained in this way may be similar to the MEP calculated using more reliable AM1 or MNDO charges. At present this option is limited to molecules containing up to 30 atoms, and is not intended to be used routinely.

4.6.4 Analysis of MEP

The MS program calculates the surface points in such a way that it is able to associate each point with an individual atom This information is potentially useful since it allows the MEP to be analysed on an atom by atom basis. This feature has been included in 3D2 to allow an analysis of the MEP data to be performed. The analysis itself takes the form of calculation of the average MEP on each atom using the MEP data for the individual surface points associated with that atom.

Although this may seem a drastic oversimplification, it does provide a method whereby the MEP's of similar molecules can be compared and should certainly be more reliable than our present method of visual comparison. The MEP's of benzamide and 3-methoxybenzamide could be compared, for example, by comparing the average MEP values for the atoms common to both molecules. Thus it is possible to compare the MEP on the ring atoms and the carboxamide group by comparing the average MEP values for both molecules.

A look at the MEP for benzamide in the profile map of Figure 4.5 shows that the procedure described above is not too unrealistic since it is clear that the spread of MEP values for most atoms is concentrated over a small range. It would therefore seem feasible to use the average value of these MEP's to give the average MEP for each atom. Figure 4.6

shows this data for benzamide plotted using 3D2 where the X-axis represents the average MEP and the X-axis the individual atoms. The overall shape of this curve is similar to that of the profile map in Figure 4.5.

Figure 4.7 shows the average MEP's on the atoms of the benzamide and 3-methoxy benzamide molecules calculated using AM1 charges and geometries. As would be expected the major differences in the MEP for the two molecules lies around position 11 in the ring. It is also clear from this figure that the introduction of the methoxy group at the meta-position in the benzamide molecule causes very little change in the MEP around the carboxamide group. This may therefore provide a simple method where the MEP of a group of similar molecules may be studied, and compared, in terms of their average MEP values.

4.6.5 Local Printing

The original 3D1 program was able to produce printed output by simply using the screen dump facility of the 4109 terminal to dump the display to the 4695 colour plotter. Unfortunately the 4695 was not compatible with the GPX workstation which meant that it could not be used to produce coloured plots in this case. This problem was solved by the purchase of an LJ250 colour printer which is specifically designed for use with the workstation and is able to use the full 256 available colours.

The current version of 3D2 allows direct printing of MEP maps on this printer in either extended (256 colour) mode or in screen dump (8 colour) mode. Most of the coloured figures presented in this thesis were obtained in this way.

4.6.6 Surface calculation

The 3D2 program can also be used to display the molecular surface of a molecule in a number of formats. One format allows individual atoms to be coloured according to atom type giving a CPK type display. Another, and possibly more useful, format is where individual parts of a surface can be coloured according to their molecular origin. Figure 4.8 shows part of the active site of the enzyme phospholipase A_2 [101] with a small flavone molecule (quercetin) docked in the active site. By colouring the enzyme and quercetin molecule red and green respectively it allows the 3-dimensional arrangement of the enzyme and substrate to be examined more clearly.

4.6.7 Additional output from 3D2

In addition to the normal MEP map files, 3D2 can write out various data files which contain information such as an MEP analysis, CSSR structure or UIS binary display list. The UIS files, in particular, are especially useful since they store all the information contained in a UIS display in a binary encoded format. These files can then be translated to a format suitable for display on a wide range of termninals and hardcopy devices using the hardcopy UIS (HCUIS) "render" command [102]. In this way it is possible to display 3D2 MEP maps on Sixel, Regis, Postscript or HPGL compatible devices. It is also possible to redisplay a UIS file by typing it directly on the GPX workstation terminal.

The MEP analysis file has been mentioned earlier and contains a simple statistical analysis of the MEP in terms of average potentials, standard deviations, percentage positive and negative points on the surface etc. The average MEP's potentials obtained from this analysis

are used in the next chapter for the purpose of comparing the 3D2 MEP maps with those calculated using ab-initio methods.

4.6.8 Utility routines and miscellaneous features

In addition to the features described above 3D2 has a selection of utilities which allow the user to change default colour maps and keyboard layouts to suit their own needs. Colour maps can be stored in a file and read in to replace the default colour map. A maximum of 20 colours can be defined, although only the first 8 colours will be used for MEP display.

Similarly the default keyboard layout can be re-defined by reading in an alternative keyboard definition file. This file can be used to assign 3D2 commands to any of the function keys, or definable keys, on the keyboard. A maximum of 22 keys may be defined on a VT100/VT52 terminal and 48 keys on VT200 series terminals. Keys may also be re-defined interactively.

As well as the utilities mentioned above a number of extra features have been added to 3D2 to make it more flexible. A few of the extra features are listed below:

(a) An interactive subprocess may be "spawned" at any time to allow access to VMS.

(b) Routines such as POP, PUSH, ZOOM may be used to manipulate individual windows on the workstation

(c) A journal file is kept of all 3D2 commands entered during a run of the program. In the event of a "crash" a complete recovery can be made from the journal file.

(d) Various other files such as PDB and Mopac Z-matrix files can be read in and displayed.

(e) An option is available where the MEP can be calculated on a set of grid points instead of the molecular surface.

Table 4.1 **MEP ranges and colours used to represent values within each range. Range values are given in atomic units and kcal/mol.**

Colour	range (atomic units)	range (kcal/mol)
Red	< -0.045	< -28
Orange	≥ -0.045 - < -0.030	≥ -28 - < -19
Buff	≥ -0.030 - < -0.015	≥ -19 - < -9
Yellow	≥ -0.015 - < 0	≥ -9 - < 0
Light Green	≥ 0.015 - < 0.030	≥ 9 - < 19
Dark Green	≥ 0 - < 0.015	≥ 0 - < 9
Cyan	≥ 0.030 - < 0.045	≥ 19 - < 28
Blue	> 0.045	> 28

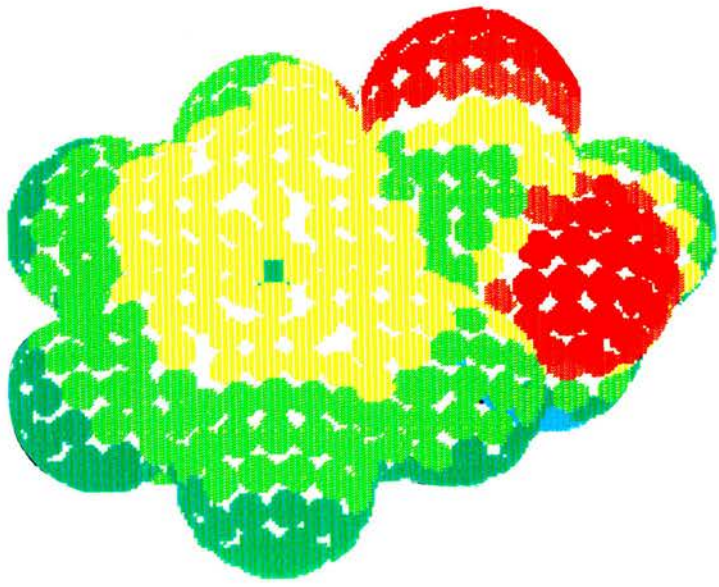


Figure 4.1 *STO-3G MEP of benzamide displayed using the 3D1 program.*

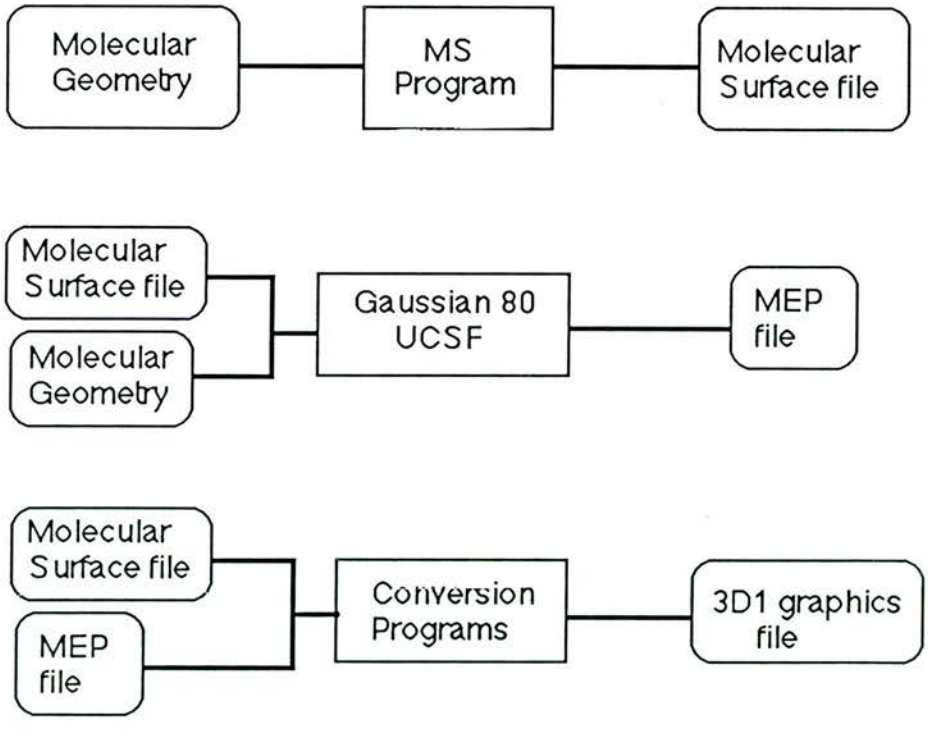


Figure 4.2 *Steps involved in producing an input file for 3D1.*

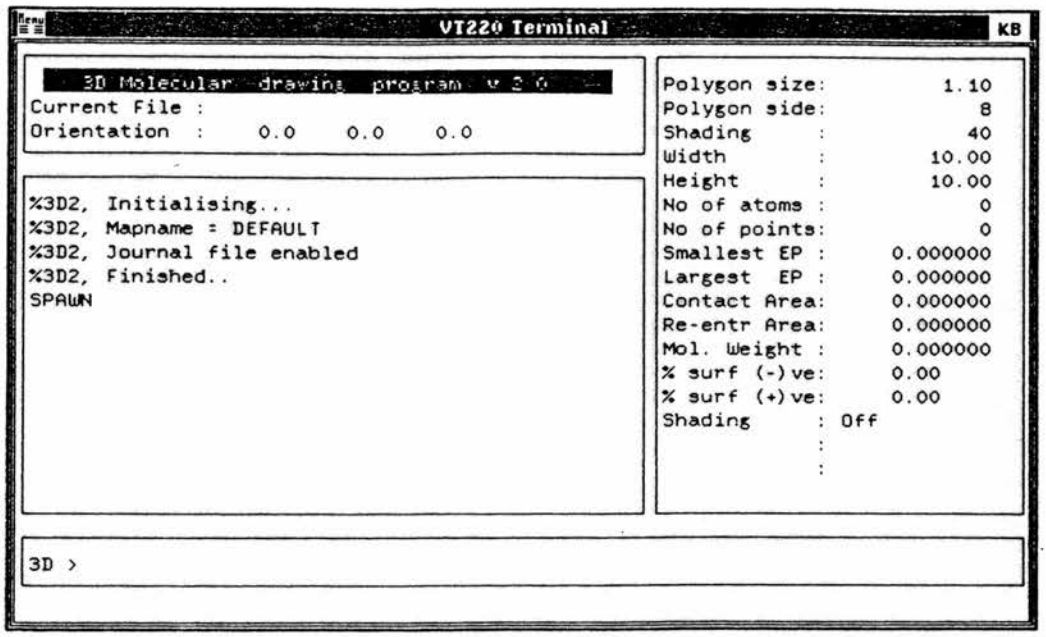


Figure 4.3a

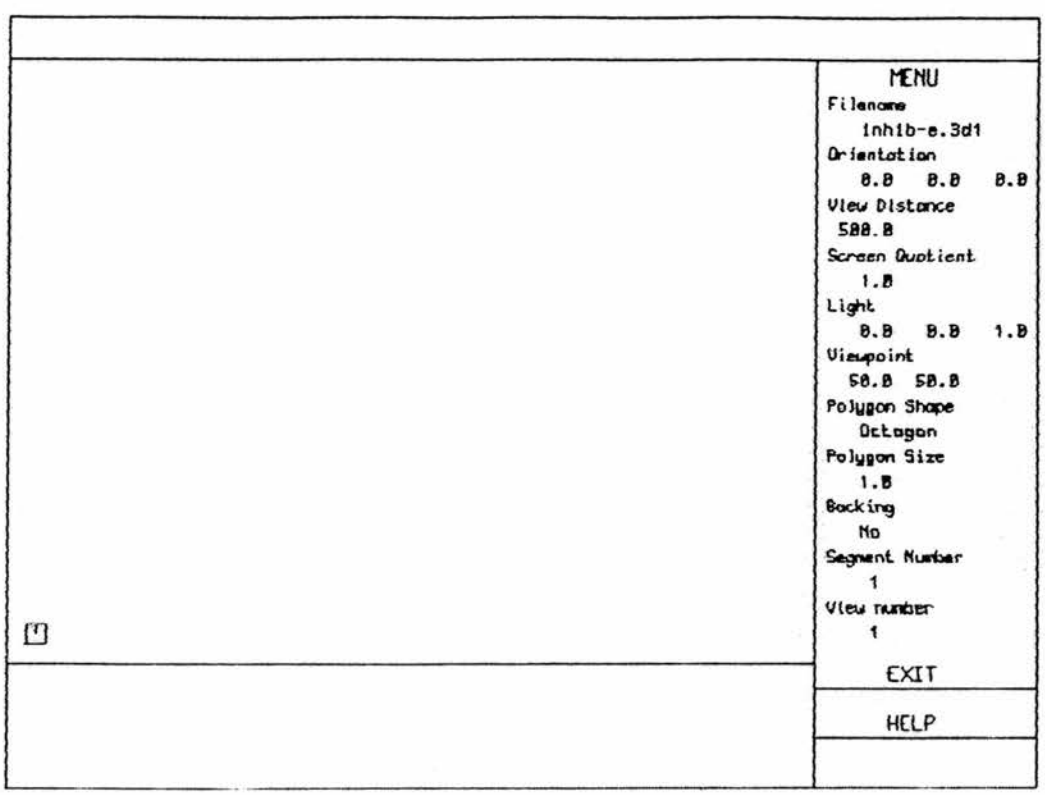


Figure 4.3b

Figure 4.3 Screens layout of (a) the 3D2 program and (b) the 3D1 program.

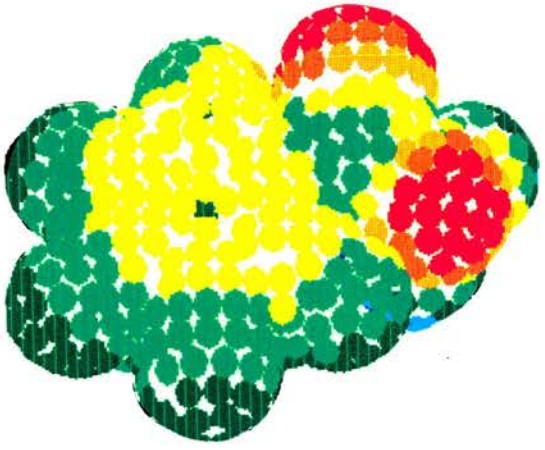


Figure 4.4a

Figure 4.4b

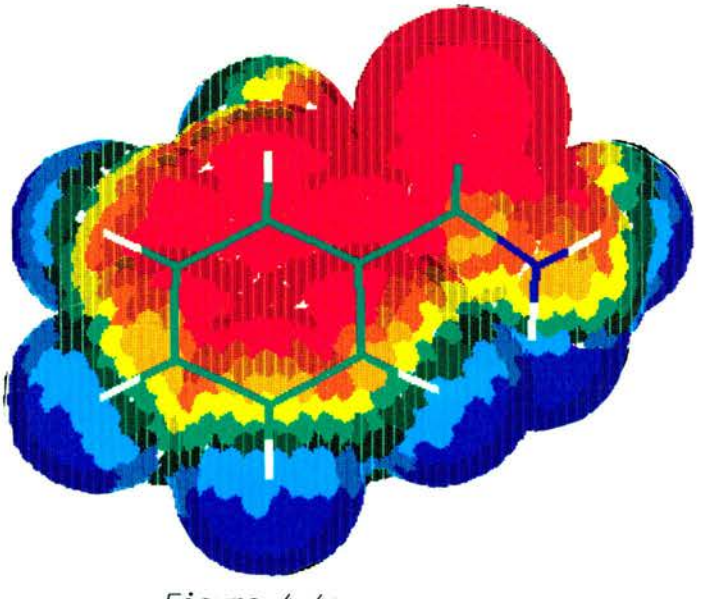


Figure 4.4c

Figure 4.4 *The MEP of benzamide calculated (a) at the STO-3G//AM1 level, (b) at the STO-3G//AM1) level on the double van der Waals surface and (c) by 3D2 using AM1 charges and geometry.*

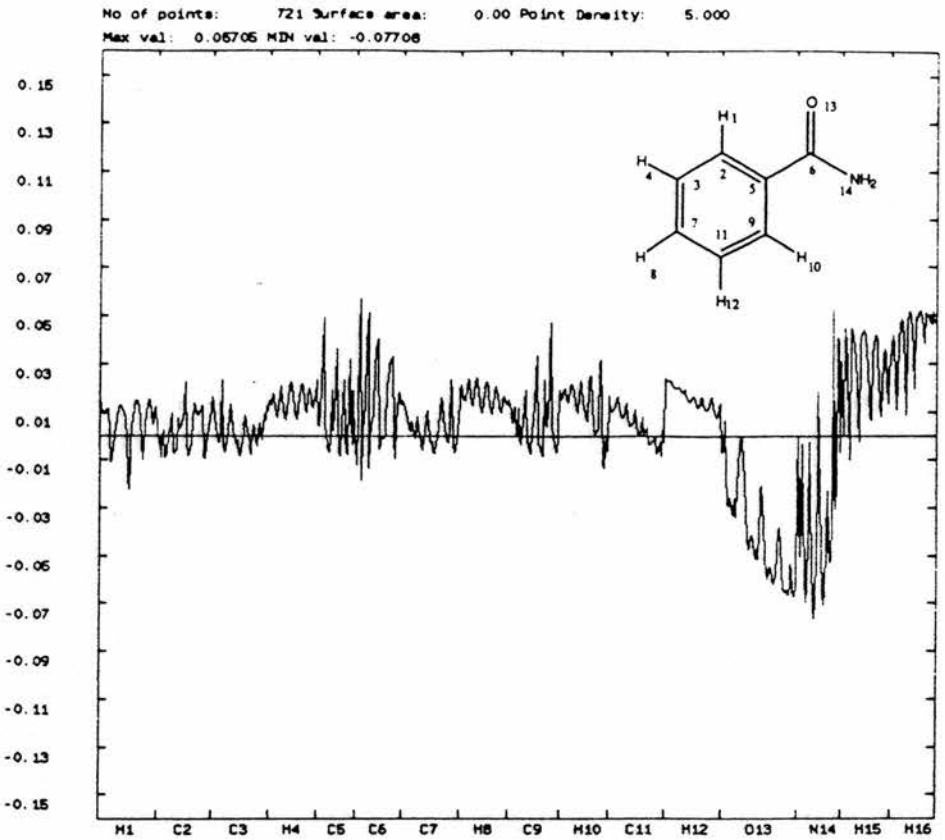


Figure 4.5 *The STO-3G//AM1 MEP of benzamide displayed in the form of a profile map.*

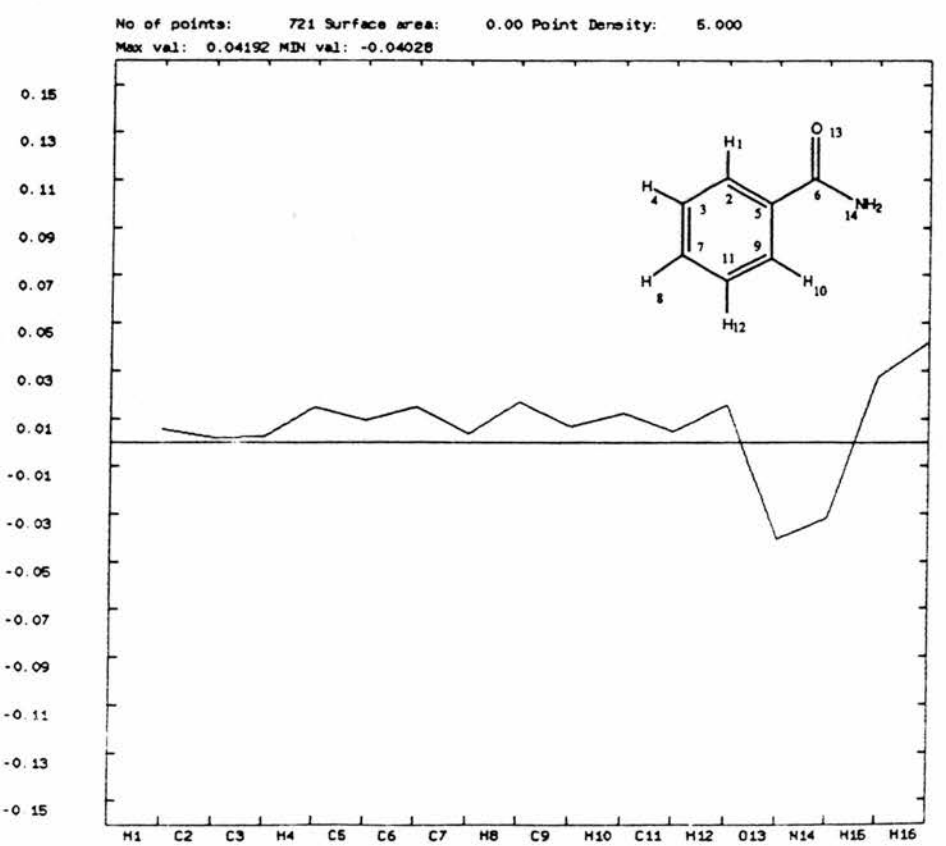
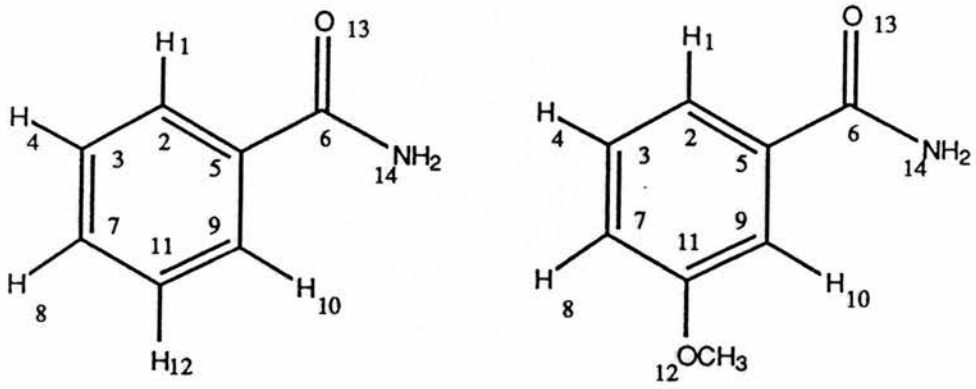


Figure 4.6 *The average MEP values for each of the atoms in the benzamide molecule calculated from the data in figure 4.5.*



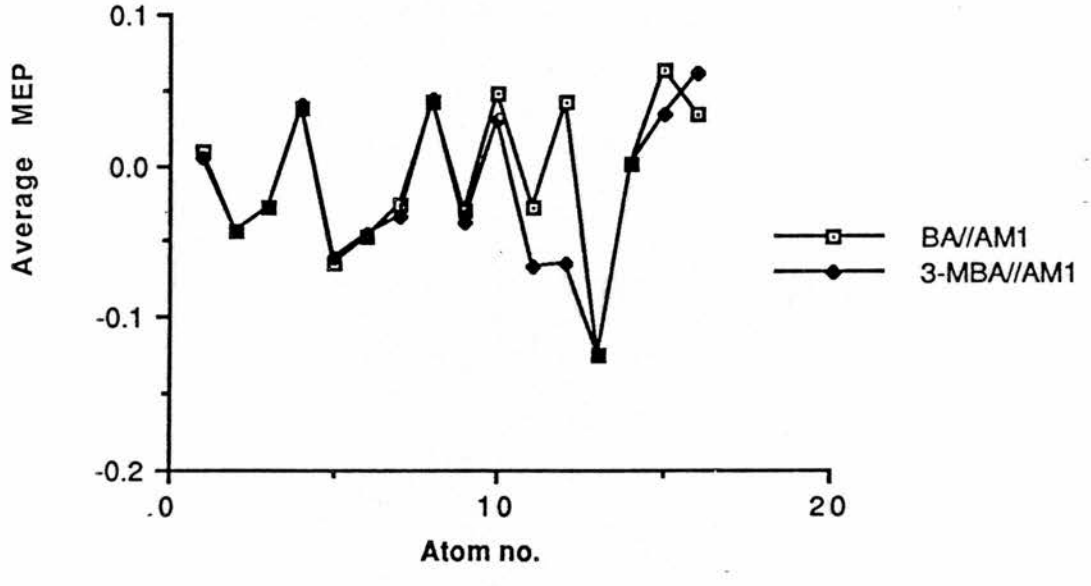


Figure 4.7 *Comparison of the average MEP's on the ring and carboxamide atoms of benzamide and 3-methoxybenzamide. All MEP's calculated using 3D2 from AM1 charges and geometries. Numbering schemes are shown for benzamide (**top left**) and 3-methoxybenzamide (**top right**).*

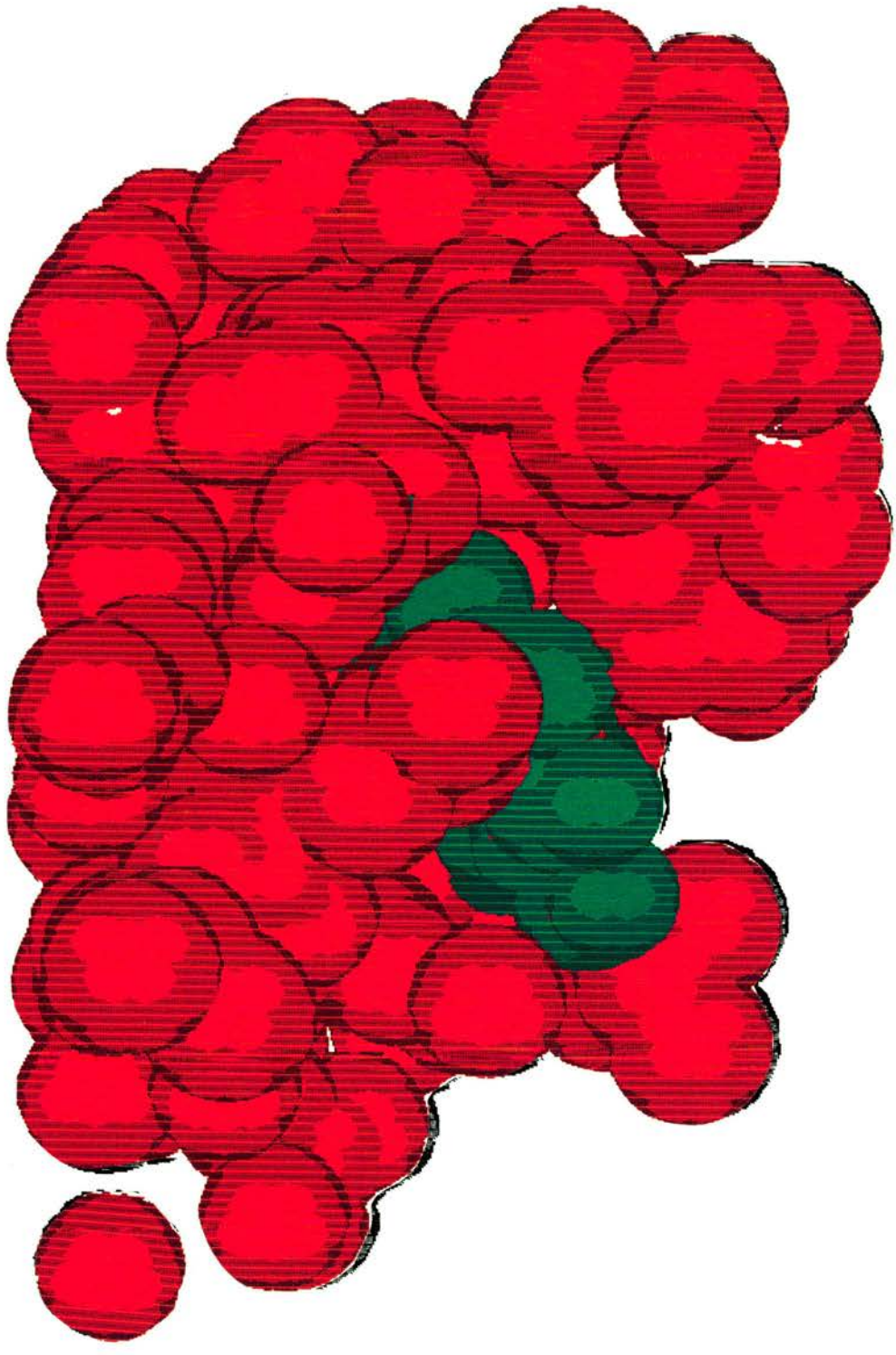


Figure 4.8 *The molecular surface of part of the active site of the enzyme phospholipase A_2 (red), showing a small flavone molecule, quercetin, located in a possible docking site.*

5 MEP calculations: comparison of ab-initio and AM1 point charge methods.

This chapter compares the results of MEP calculations carried out using the simple point charge model described in 3.5 and ab-initio methods using a range of basis sets. The purpose of these calculations was to test the applicability of the point charge model by comparing results for a small set of test molecules. Results are presented in the form of the average MEP maps described in chapter 4 where the average MEP's for both ab-initio and point charge methods are compared in graphical form. A few coloured 3-D MEP maps are also included. All MEP and average MEP values reported in this chapter are in atomic units unless otherwise specified. All coloured MEP maps have been displayed using the contour levels given in Table 4.1.

5.1 Choice of AM1 method

The calculation of the MEP using the simple model described in 3.5 requires only the molecular geometry and a set of point charges for each molecule. These may be obtained from any convenient method, either ab-initio or semi-empirical, although it is desirable to choose a method which results in MEP's which are comparable with those calculated by more accurate methods.

Since we are mainly interested in large molecules then this effectively rules out the use of ab-initio methods since the time required for such calculations is prohibitive. We have chosen, instead, to use optimised geometries and charges obtained from the AM1 method described earlier. In all cases reported in this work the approximate MEP's have been calculated using AM1 geometries and charges unless otherwise stated.

Our reasons for using AM1 rather than MNDO, or other

semi-empirical methods, are based upon our experiences of using this method and those of other groups [103]. The original paper by Dewar *et al* [6] describing the AM1 method reports average errors for most properties which are lower than the corresponding errors for MNDO. The error in heats of formation, dipole moments and molecular geometries in particular have been reduced. Baird and Hadley [7] found the RMS error in heats of formation for a series of 52 compounds to be lower for AM1 (6.17) than MNDO (7.2). Previous work by us, comparing the use of semi-empirical methods for open-shell systems and reported elsewhere [104], show the AM1 method to perform better than MNDO giving similar results to the correlated MNDOC method of Thiel [105, 106].

Chapter 9 reports calculations on the rotational isomers of nitrosomethanol using both AM1 and ab-initio methods. The results of this study show AM1 to give geometries of both minima and transition states in this system which are comparable with 6-31G structures. More recently our calculations on flavone acetic acids [107] and polycyclic aromatic hydrocarbons [108] have increased our confidence in results obtained using the AM1 method. The results presented in the next chapter on the benzamide based inhibitors of ADPRT also show favourable agreement between AM1 optimised gometries and experimental structures where these are available.

5.2 Test calculations

The following sections report the results of calculations of the MEP for a series of small molecules using both ab-initio and simple point charge methods.

5.2.1 Choice of test molecules and geometries

The choice of molecules used in the above comparison is given in Table 5.1 and is limited to small systems with the inclusion of a few larger molecules such as benzamide and 5-hydroxytryptamine. To avoid the comparison being influenced by the use of different geometries, all MEP calculations were performed at the AM1 optimised geometry for each molecule. The only exception to this is for the data presented in figure 5.1 which shows the average MEP's for imidazole and pyrazole calculated using a range of basis sets and corresponding optimised geometries. This figure also shows the average MEP for the same molecules calculated using the 3-21G basis set at the AM1 geometry. It is clear from Figure 5.1 that the difference between the 3-21G//3-21G and 3-21G//AM1 MEP's for both molecules is negligible. This is also evident from the 3-D MEP maps (not shown) for these molecules where the 3-21G//AM1 maps are almost indistinguishable from the 3-21G//3-21G maps.

We have found this to be the case, in general, for most molecules which we have studied in this way. This is comforting, and is in agreement with Kitaura and Morokuma [109] who found that the MEP is not affected much by the structure of the molecule but is, however, more sensitive to the quality of the wavefunction used for its calculation.

The AM1 optimised geometries of the test molecules are in good agreement with available experimental geometries and are not reported here since they are not the purpose of the comparision.

5.2.2 Basis sets for ab-initio MEP calculations

Figure 5.1 compares the average MEP's of STO-3G, 3-21G, 6-31G and 6-31G* basis sets at the respective geometries. It is clear from Figure 5.1 that there is little difference between the 3-21G, 6-31G and 6-31G* MEP's:

all of these basis sets give virtually identical MEP maps. The STO-3G MEP's, whilst showing some difference from the larger split-valence MEP's. follow the same trends and lead to the same qualitative results. This is also true for the MEP's of isoxazole and oxazole which are not shown here. The addition of polarisation functions, in the case of the 6-31G* results, does not appear to affect the MEP appreciably.

Our findings are in general agreement with a number of previous studies [110-112] which have compared MEP's obtained from basis sets of different size. The overall conclusion from such studies is that improvement of the quality of the wavefunction, by increasing the size of the basis set, does not appreciably alter the overall form of the MEP for a molecule. Care must be taken, however, where the calculated potentials are near zero since in such cases small differences may alter the sign of the MEP.

In the comparisons described in the following sections we have chosen the 3-21G basis set for our reference ab-initio MEP calculations. This basis set offers better results than the minimal STO-3G basis while still allowing the calculations to be carried out in a reasonable time.

5.3 Choice of surface

We have also considered the choice of surface on which to calculate the MEP. The normal method which we have used is to calculate the MEP on the van der Waals surface of the molecule. Tait [113] and Hall [114] have shown that the MEP calculated using a point charge model becomes more accurate with increasing distance from the molecule, gradually reducing to an exact point-charge potential. The 3D2 program allows the MEP to be calculated on surfaces generated by multiplying the van der Waals radii of the atoms by a scale factor. The comparisons below make

use of this facility to compare the point charge MEP's on both the van der Waals and double van der Waals surfaces.

5.4 Simple five-membered heterocycles

MEP's have been calculated for the four heterocycles imidazole, pyrazole, oxazole and isoxazole at both AM1 point charge and 3-21G//AM1 levels. Figure 5.2 shows the MEP of imidazole calculated as above and also at the MNDO point charge and 6-31G*//6-31G levels. As stated previously there is little difference between the two ab-initio potentials. The AM1 and MNDO potentials are also similar although the AM1 MEP does give better agreement with the ab-initio MEP's than MNDO. We have found this to be the case in general and have made no further use of MNDO potentials.

Figures 5.3 and 5.4 compare the AM1 point charge and 3-21G//AM1 average MEP's for imidazole and pyrazole respectively, calculated on both their van der Waals and double van der Waals surfaces. Figure 5.5 shows the same information for oxazole and isoxazole on the double van der Waals surface only. It is clear from these figures that the potentials calculated on the van der Waals surface show the largest difference between the simple point charge AM1 and 3-21G//AM1 models. Even so the general shapes of the curves are similar, with both methods giving the same qualitative predictions. The MEP's calculated on the double van der Waals surface are in better agreement, in all cases, with small differences between methods and agree with the findings of Tait and Hall.

Previous work by Berthier *et al* [115] has compared ab-initio MEP's for the above heterocycles with their known experimental behaviour. In all heterocycles the N heteroatoms are found to be more reactive towards electrophiles than the O atoms. The Figures 5.3 to 5.5 support this

evidence especially where the double van der Waals surface MEP's are considered. In the case of pyrrole (fig 5.4) the van der Waals surface AM1 point charge MEP predicts the N-H nitrogen to be marginally more reactive than N7 in contrary to the experimental evidence. The MEP on the double van der Waals surface does, however, slightly favour the N7 atom in accordance with experiment.

This suggests that it is not advisable to make predictions of reactivity based solely upon the point charge AM1 potentials calculated on the van der Waals surface, especially where the atoms of interest have average MEP values which are very similar. In such cases it is necessary to calculate the potential on the double van der Waals surface as well.

5.5 Formamide and other small molecules

Formamide contains two distinct sites of protonation, namely the carbonyl oxygen and amine nitrogen, and so provides a useful test of the MEP. Experimental evidence suggests that protonation on the carbonyl oxygen is most favourable [116]. Ab-initio calculations by Hopkinson and Csizmadia [117] also favour oxygen as the likely site for protonation.

Figure 5.6 shows the MEP for formamide calculated at STO-3G//AM1, 3-21G//AM1 and point charge AM1 levels on the van der Waals surface of the molecule. Again it is clear that the simple point charge AM1 MEP is in good agreement with the ab-initio MEP's. Figure 5.7 shows the average MEP's for formamide calculated on the double van der Waals surface. This clearly shows the oxygen atom to be the most likely site of protonation since the average MEP on this atom is far more negative than that on the N atom which is slightly positive. This is also clear in Figure 5.6.

The remainder of the Figures 5.8-5.14 compare the AM1 and 3-21G//AM1 average MEP's calculated on the double van der Waals surface

of some of the small molecules in Table 5.1 In general the agreement between the AM1 point charge and 3-21G//AM1 methods is satisfactory.

5.6 Benzamide and 5-HT

The average MEP's of benzamide and 5-hydroxytryptamine (5-HT) are compared in Figures 5.15 and 5.16 respectively. The agreement between the point charge AM1 and 3-21G//AM1 MEP's for these larger molecules is again satisfactory especially on their double van der Waals surfaces. The 5-HT molecule is particularly interesting since it has been widely studied by Weinstein *et al* [77, 118-119] in relation to drug-receptor recognition in peripheral tissues and in the brain. In his description of the MEP of this molecule, Weinstein refers to an "orientation vector" which describes the "electrostatically preferred alignment of a molecule towards a field generated by positive charges placed in a plane above it". This "orientation vector" is defined by a line connecting the minima through the areas of greatest change in the potential.

Figure 5.17 shows the AM1 point charge and 3-21G//AM1 MEP maps for 5-HT calculated on the double van der Waals surface of the molecule. The molecular structure and numbering scheme is shown at the top along with the direction of Weinstein's orientation vector. This figure also shows an orientation vector which we have "tentatively" drawn on the AM1 MEP map. This is in good agreement with the direction of Weinstein's vector. The agreement between AM1 and 3-21G//AM1 potential maps is also rather good.

5.7 Conclusions

The results presented in this chapter have hopefully shown that the simple point charge MEP's compare well with the more accurate 3-21G//AM1

potentials. The following points should, however, be noted. In most cases comparison of the MEP in molecules of similar structure can be carried out on the van der Waals surface of the molecules. In cases where there is doubt about the results or where the molecules are very different then the double van der Waals surface MEP's should be used.

The next two chapters of this thesis describe the application of this simple point charge model to the study of the inhibitors of the enzyme poly (ADP-ribose) transferase.

Table 5.1

The selection of molecules used in comparison of ab-initio and point charge MEP's.

ethylene	imidazole
formamide	pyrazole
formaldehyde	oxazole
methanol	isoxazole
methylamine	formic acid
glyoxal	water
methane	benzamide
5-hydroxytryptamine	

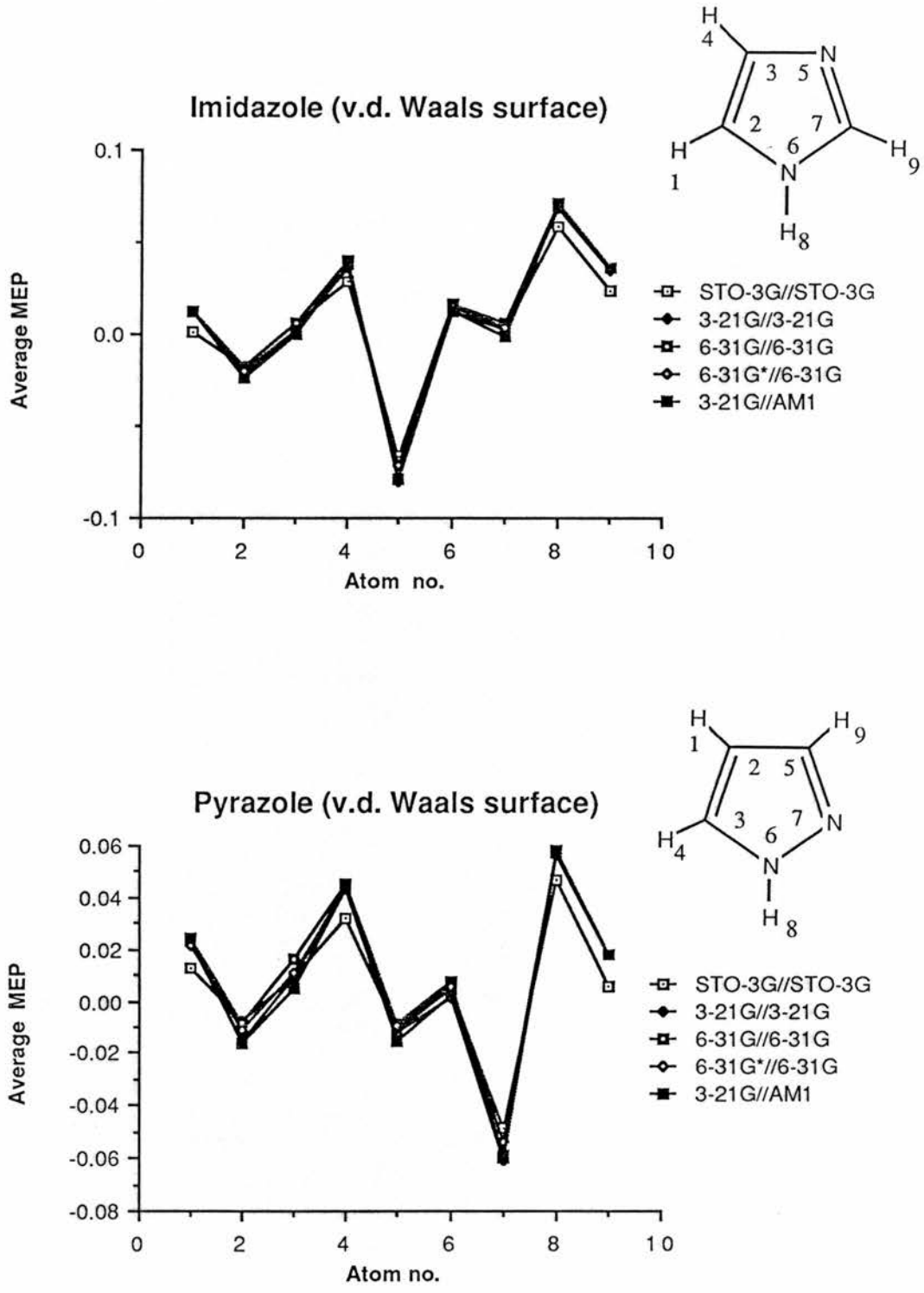


Figure 5.1 Comparison of the average atomic MEP's of imidazole and pyrazole calculated on the van der Waals surface of the molecules using various basis sets.

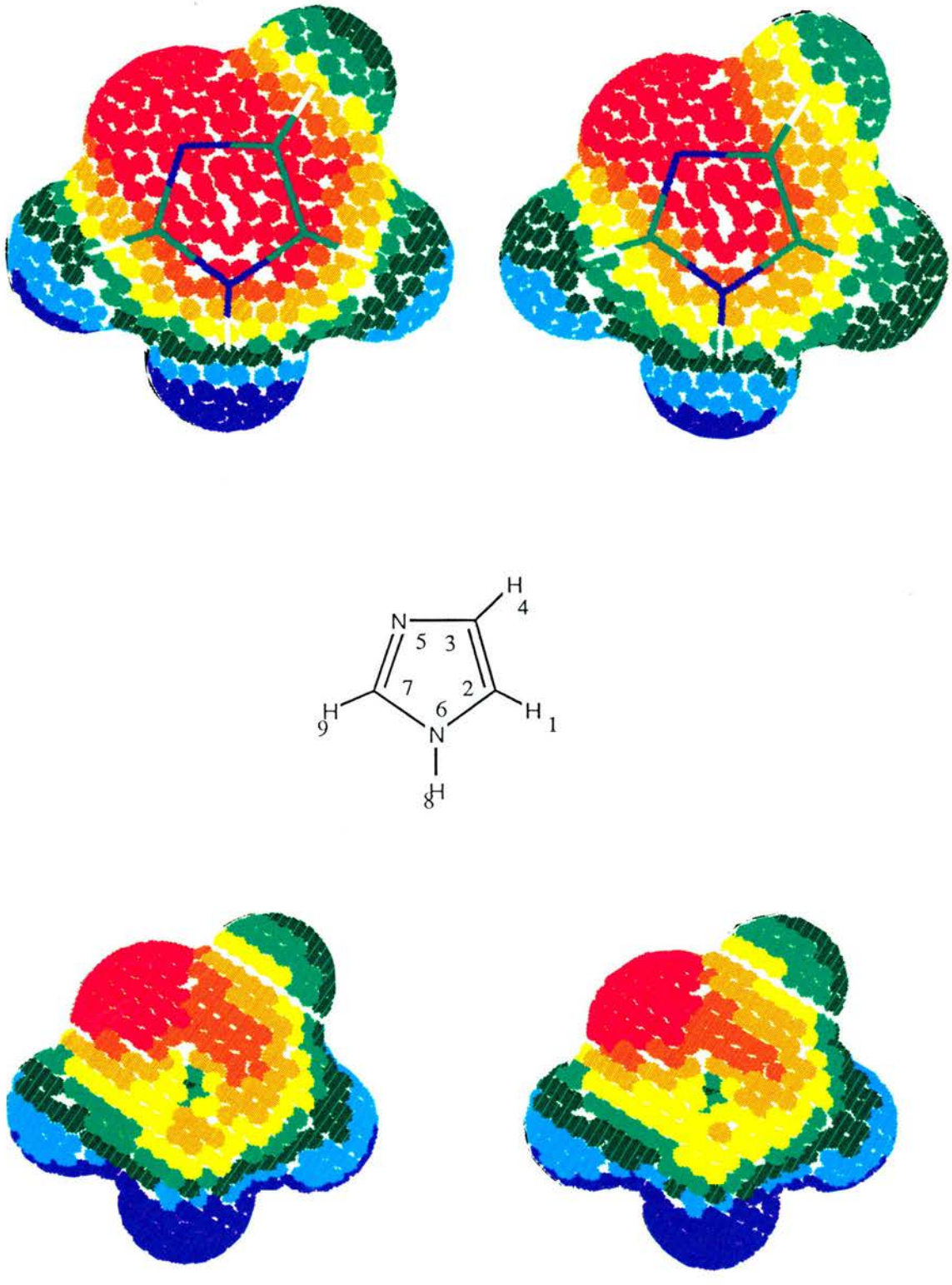
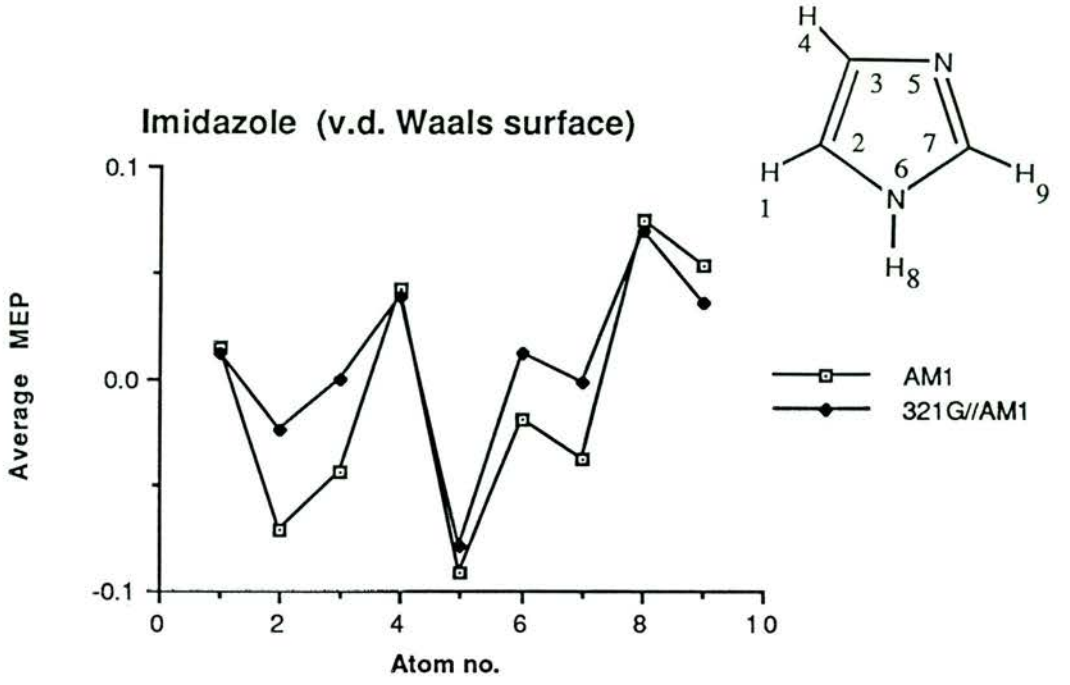


Figure 5.2 MEP's of imidazole, calculated on the van der Waals surface of the molecule, at both point charge and ab-initio levels. Top left shows AM1 point charge MEP, top right MNDO point charge MEP, bottom left 3-21G//AM1 MEP and bottom right 6-31G*//6-31G MEP.



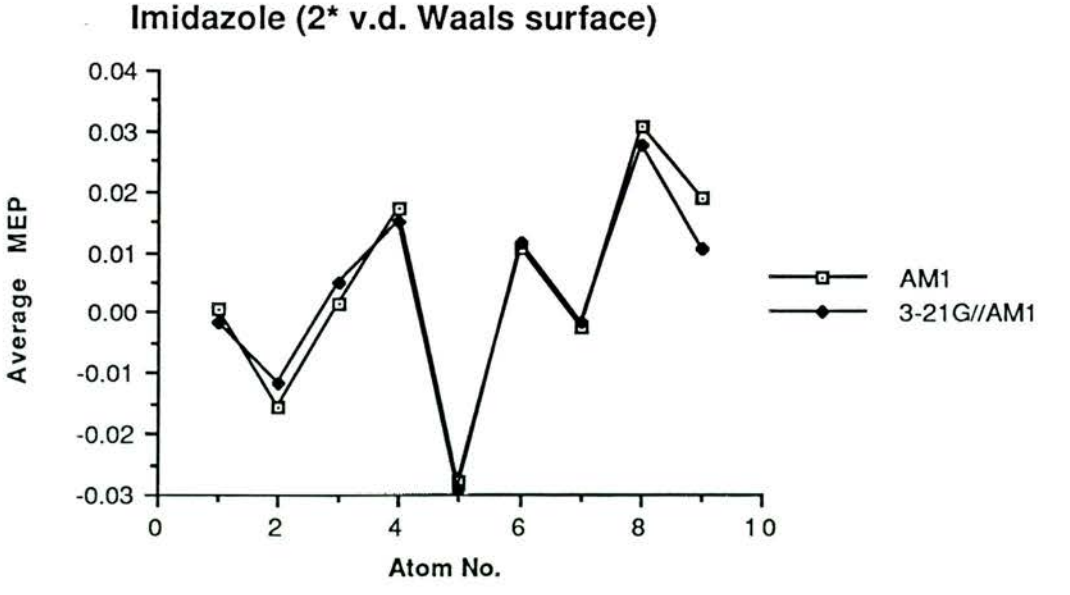
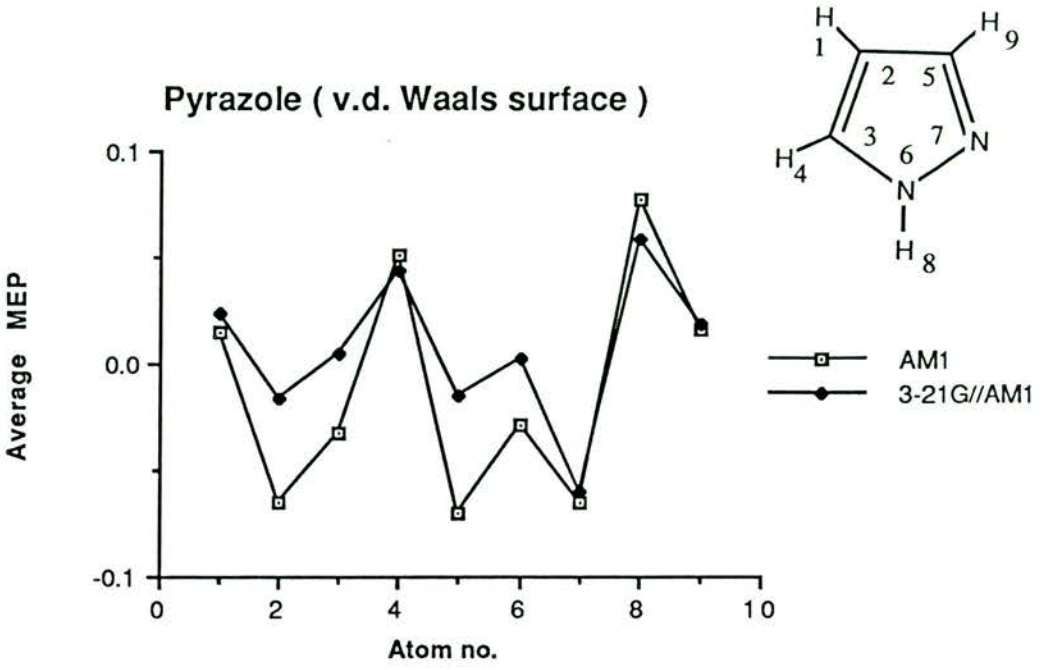


Figure 5.3 Comparison of the average atomic MEP's of imidazole calculated at the AM1 point charge and 3-21G//AM1 levels.



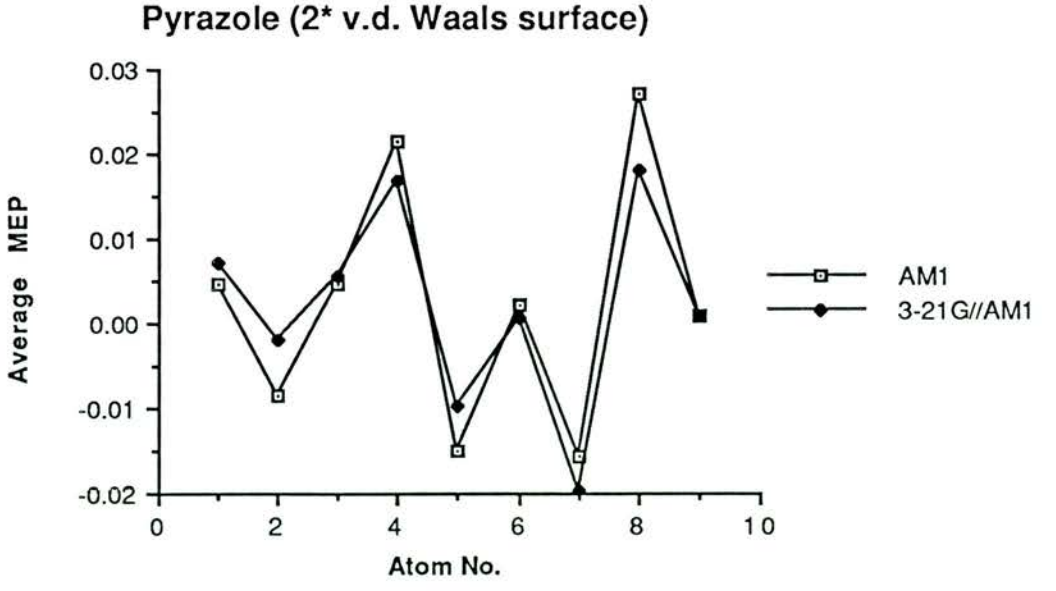


Figure 5.4 Comparison of the average atomic MEP's of pyrazole calculated at the AM1 point charge and 3-21G//AM1 levels.

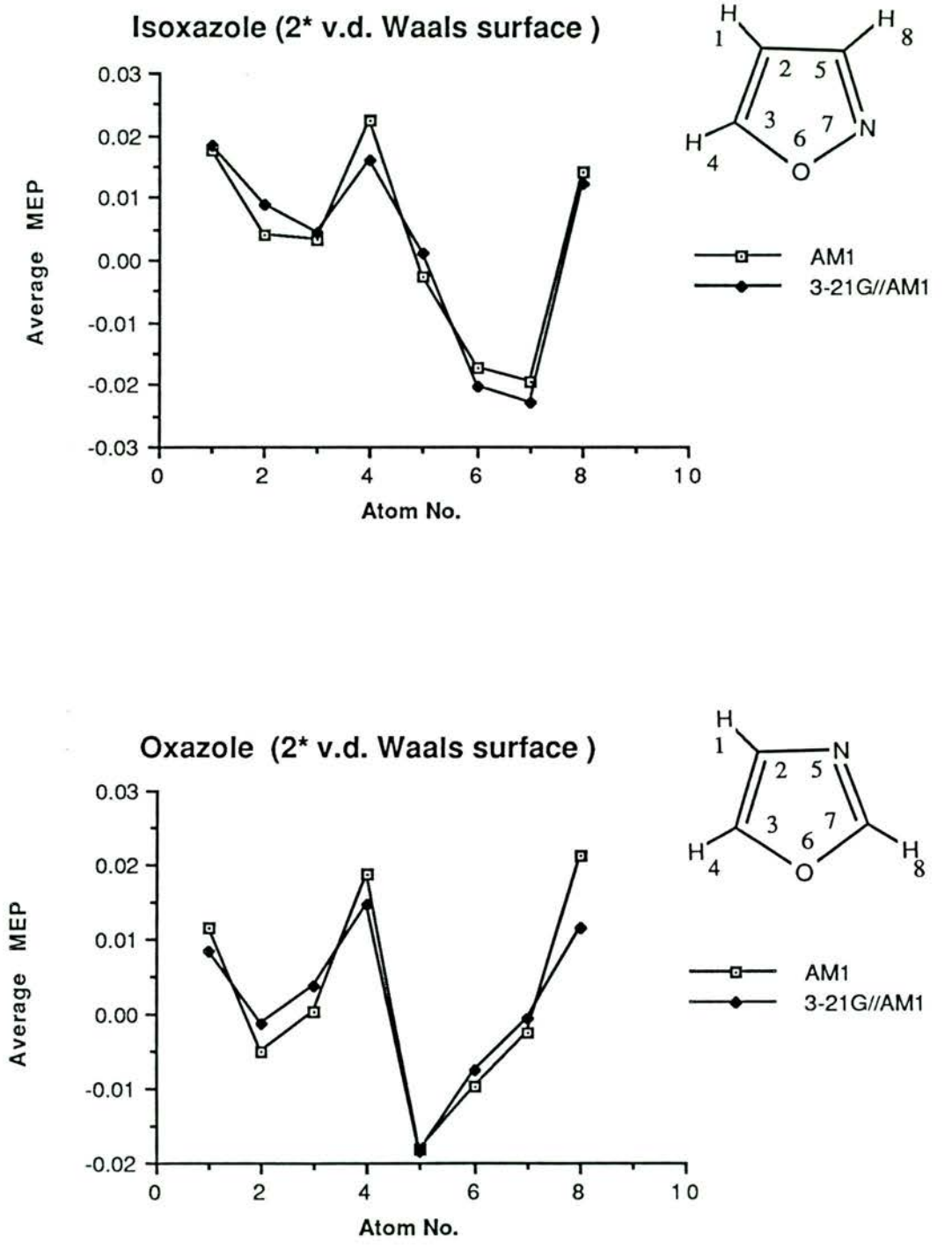
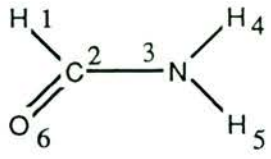
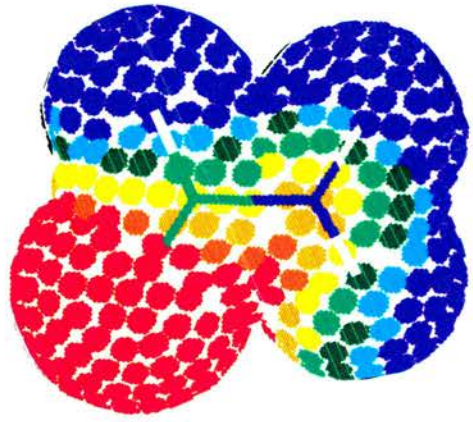
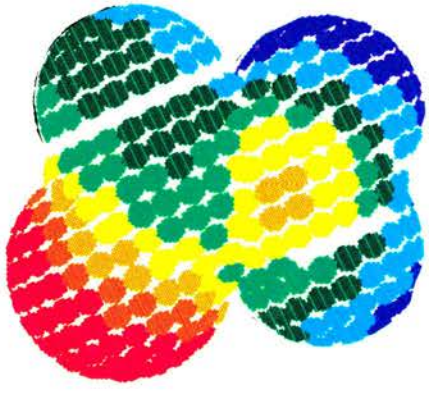


Figure 5.5 Comparison of the average atomic MEP's of isoxazole and oxazole calculated at the AM1 point charge and 3-21G//AM1 levels.







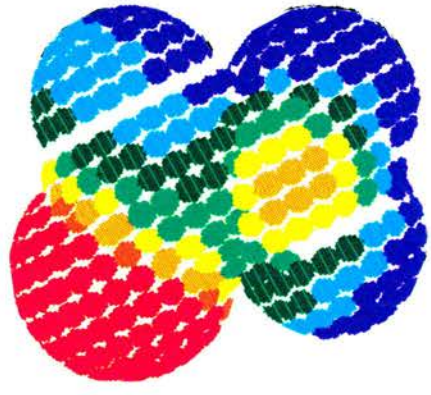


Figure 5.6 MEP's of formamide calculated at AM1 point charge (top right), STO-3G//AM1 (bottom left) and 3-21G//AM1 levels.

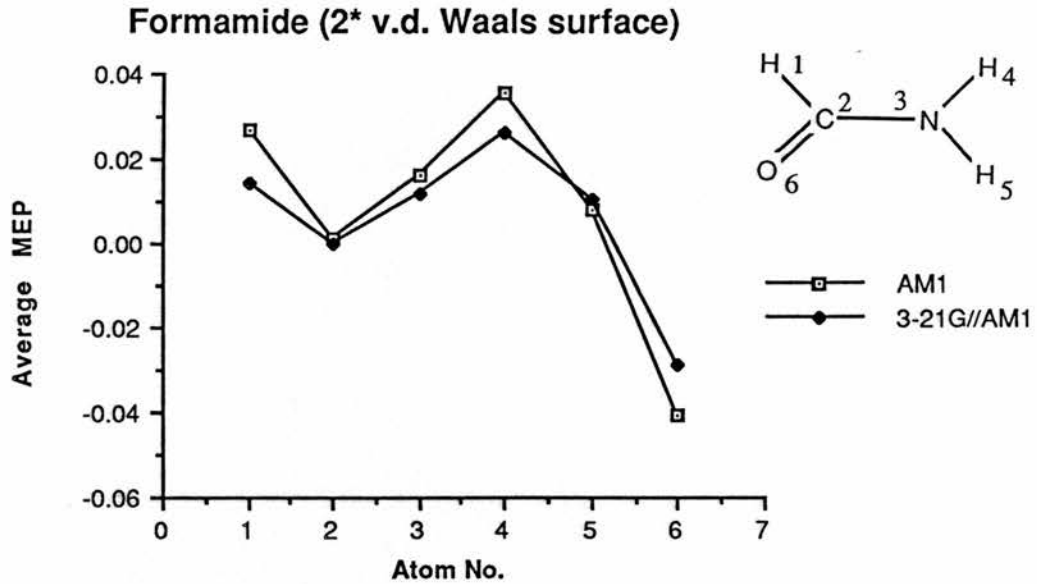


Figure 5.7 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of Formamide.

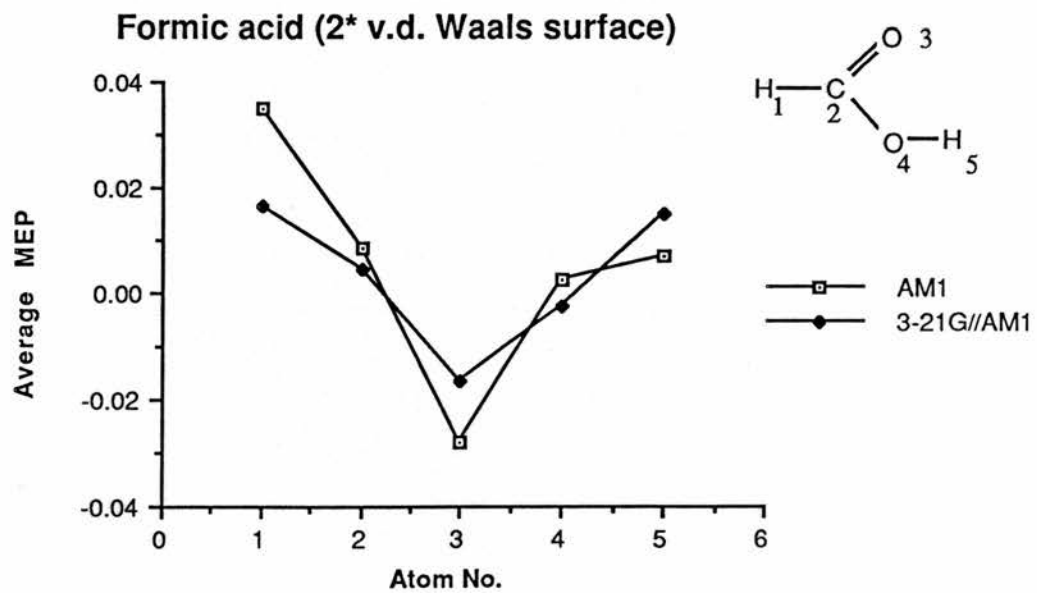


Figure 5.8 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of Formic acid.

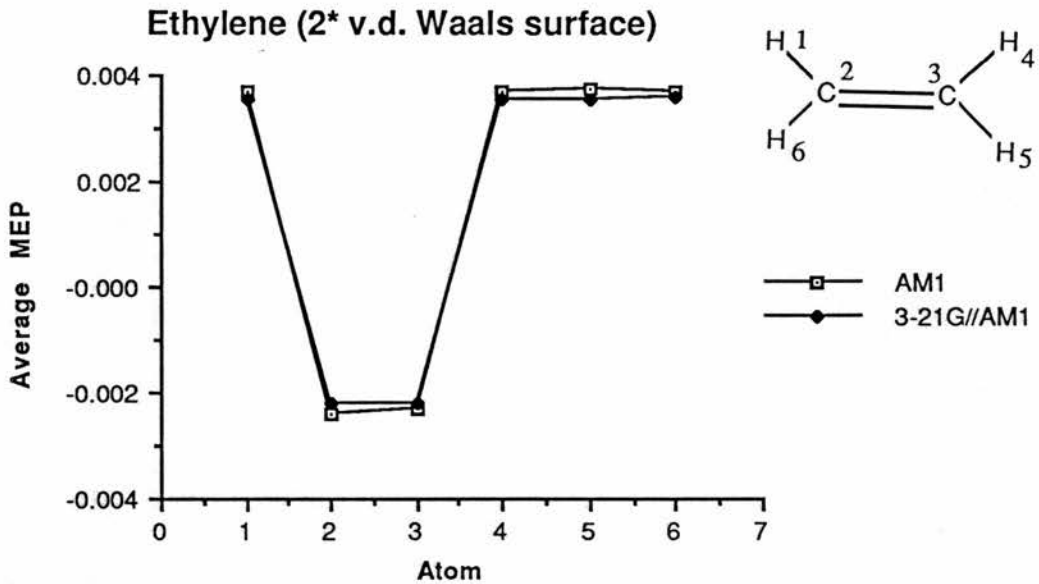


Figure 5.9 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of ethylene.

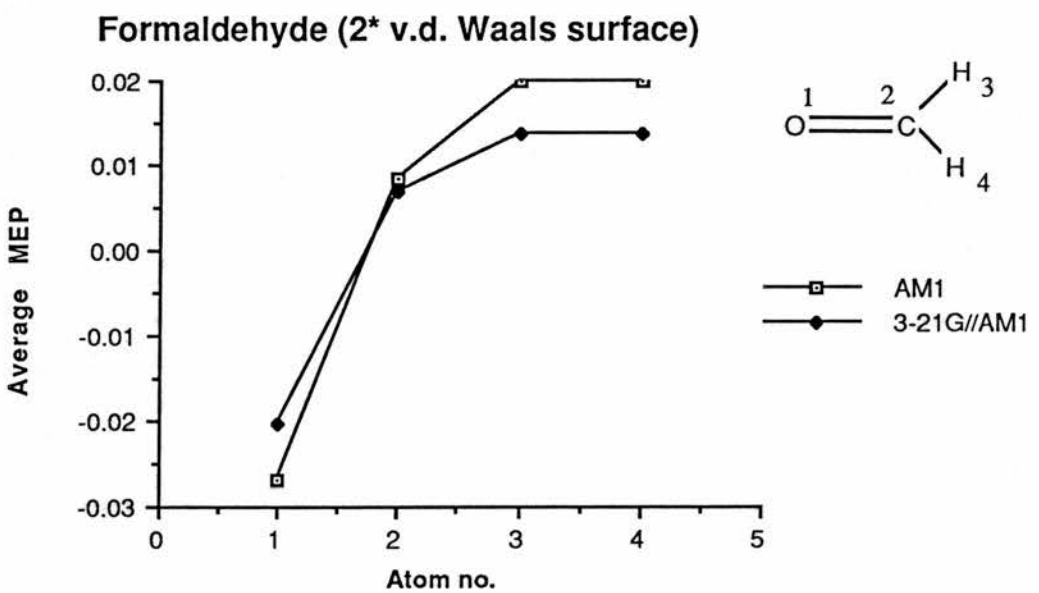


Figure 5.10 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of formaldehyde.

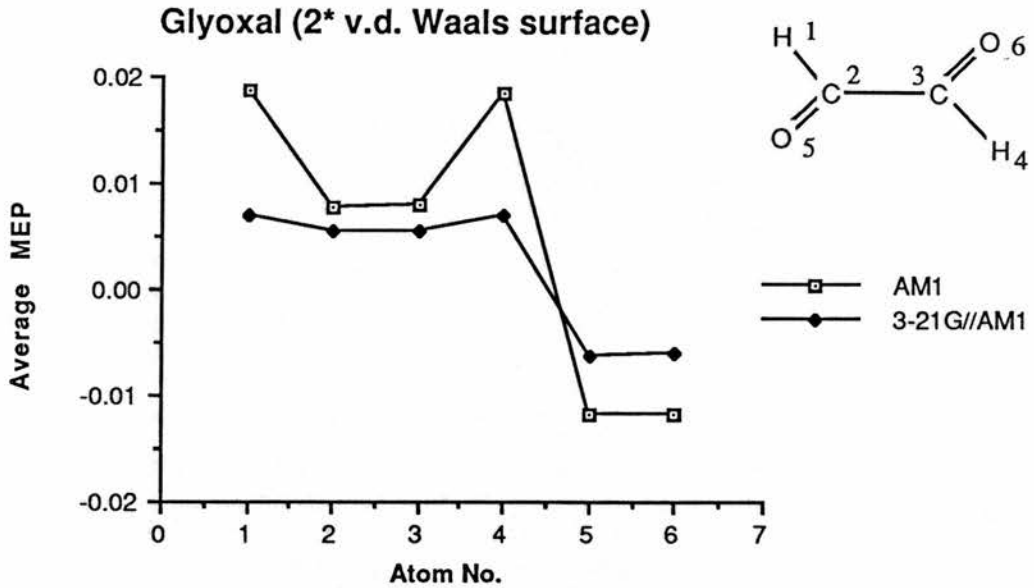


Figure 5.11 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of glyoxal.

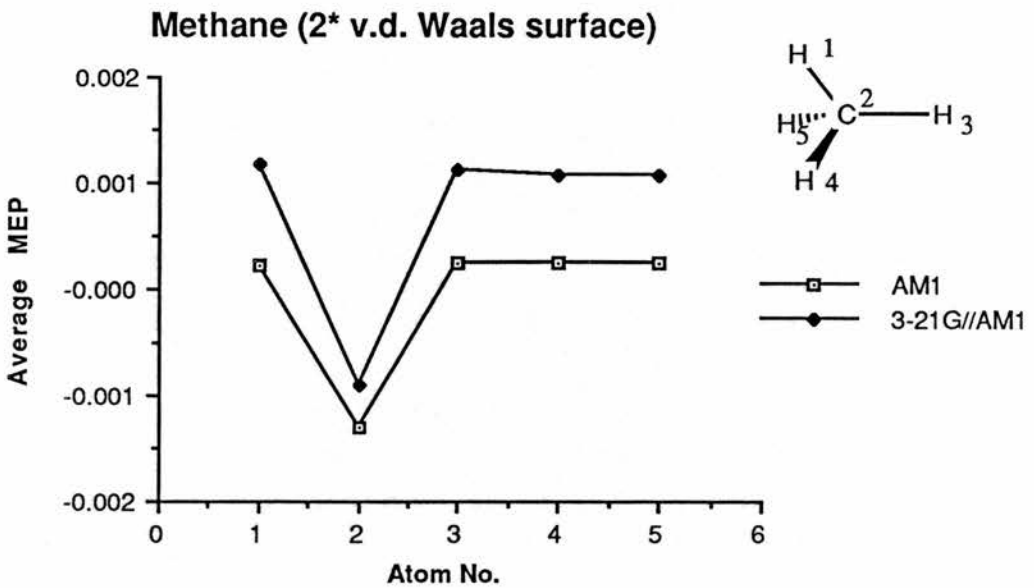


Figure 5.12 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of methane.

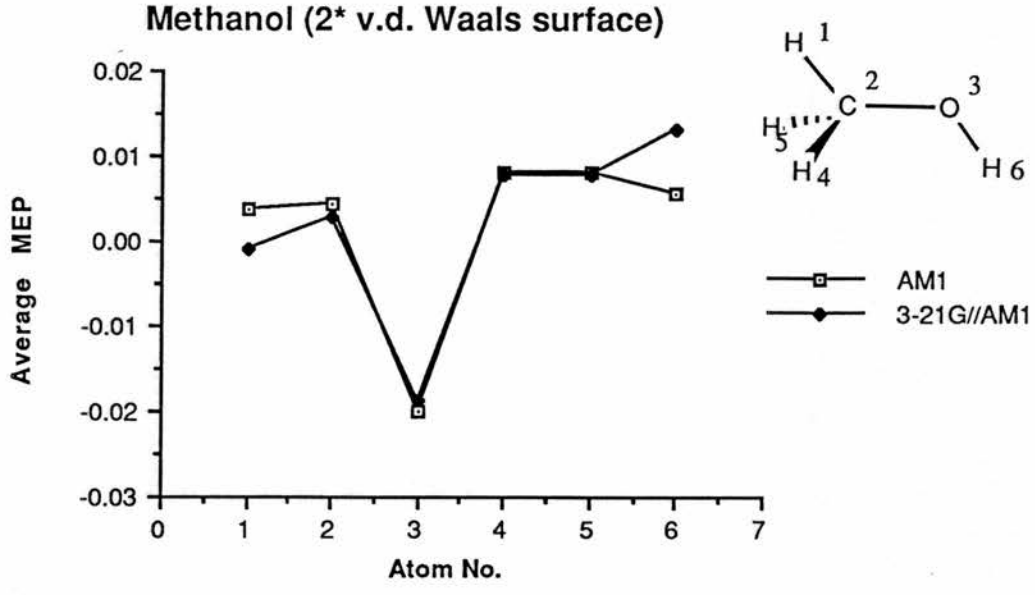


Figure 5.13 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of methanol.

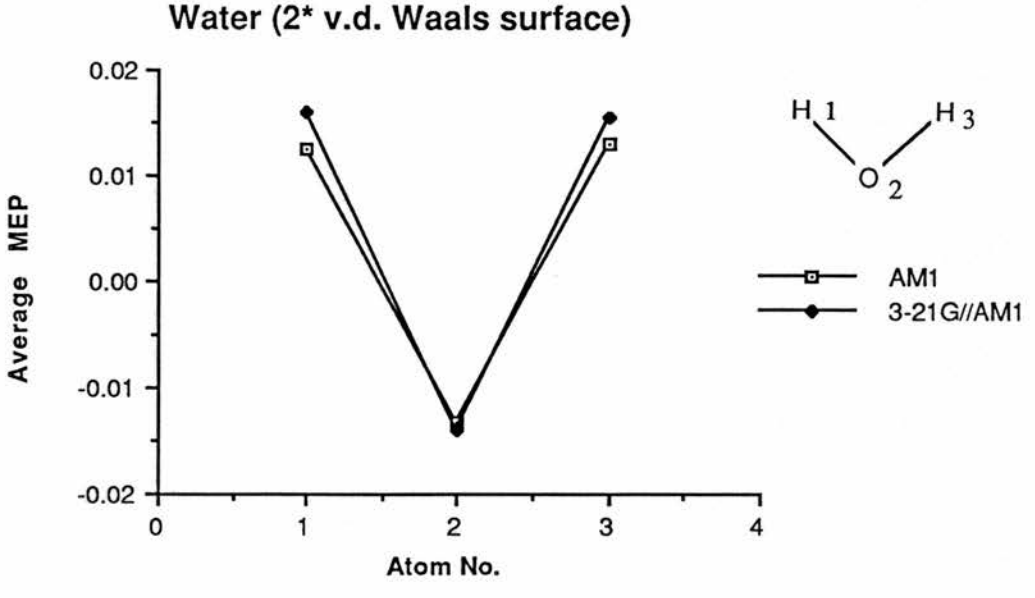


Figure 5.14 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of water.

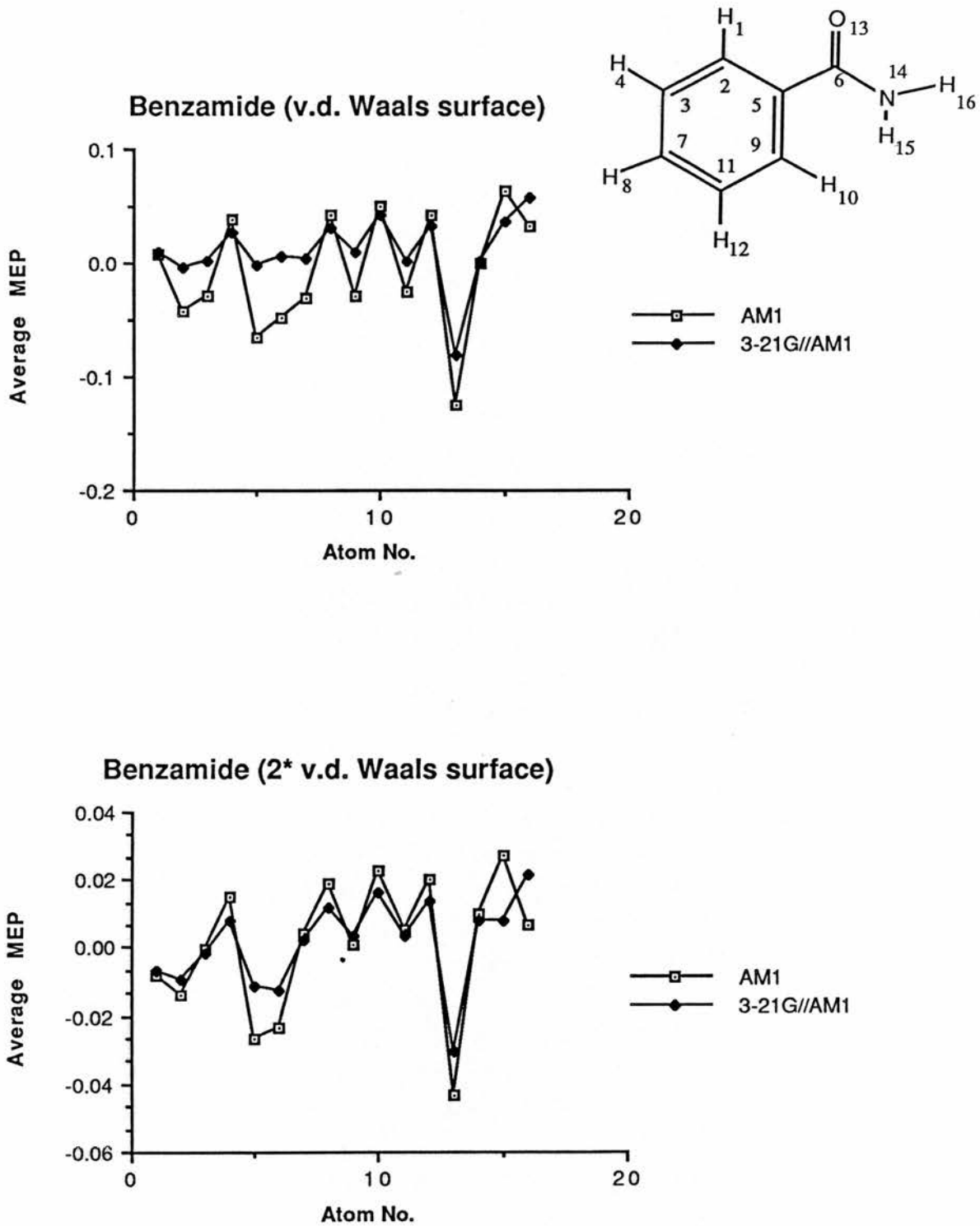
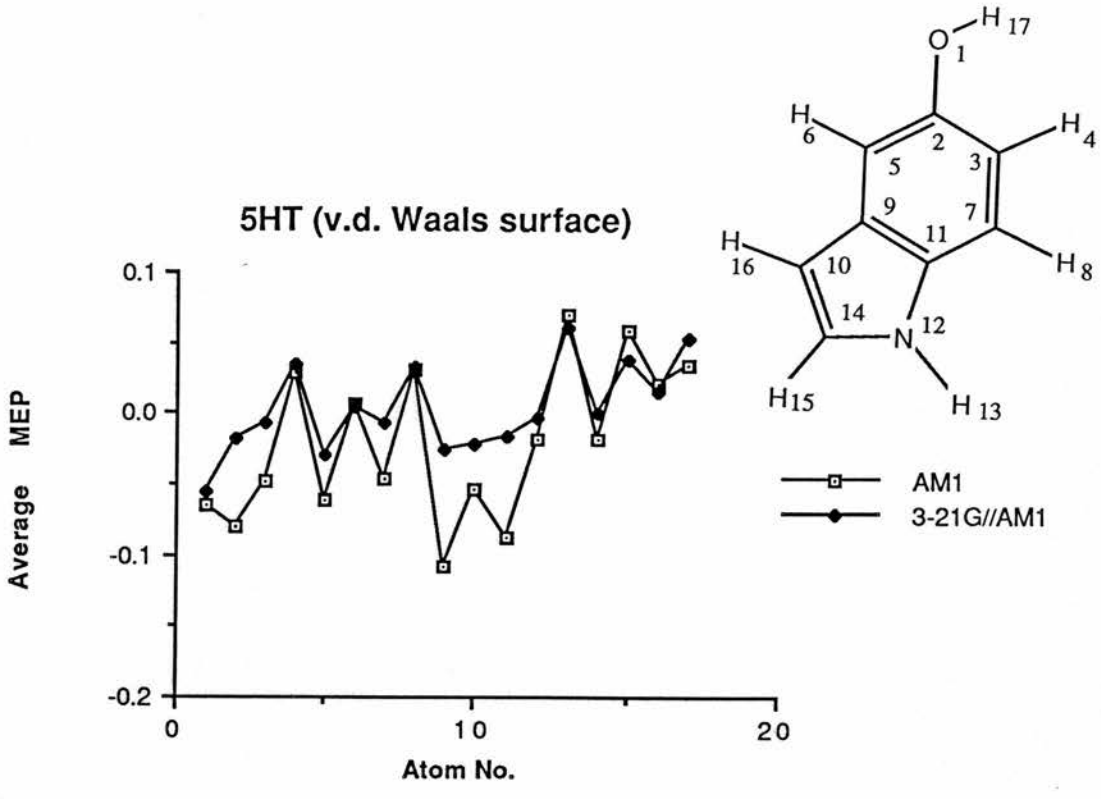


Figure 5.15 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of benzamide.



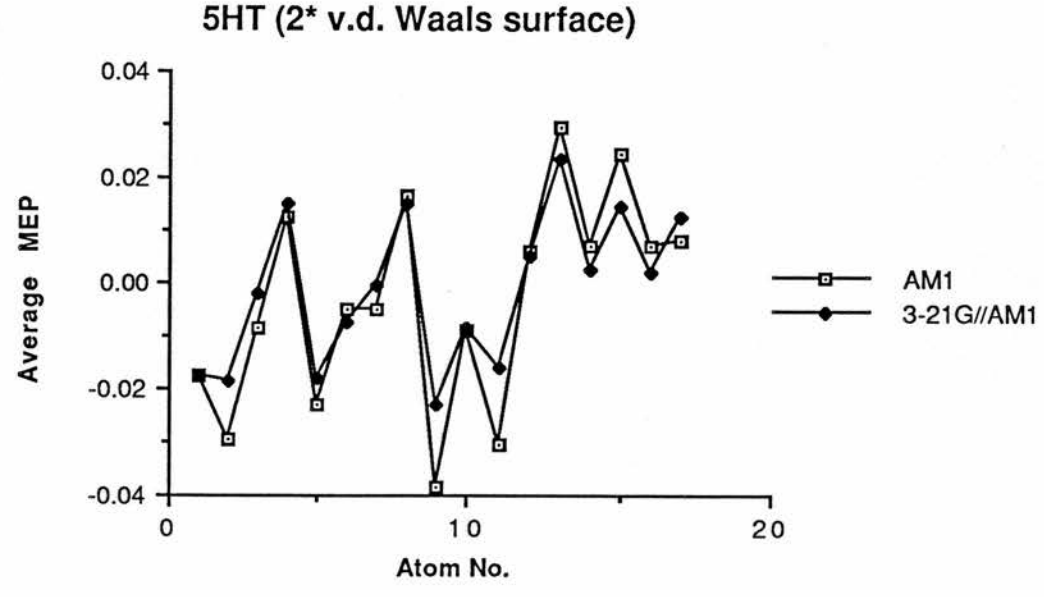


Figure 5.16 Comparison of the AM1 point charge and 3-21G//AM1 average atomic MEP's of 5-hydroxytryptamine.

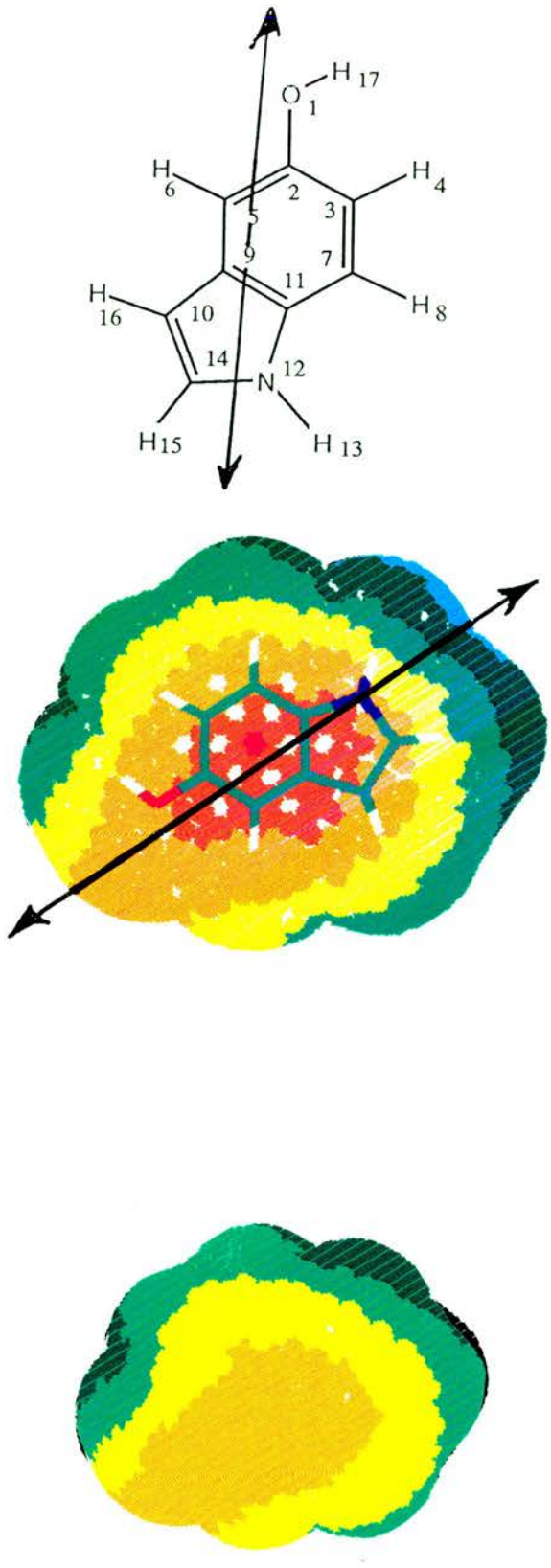


Figure 5.17 The AM1 point charge MEP (middle) and 3-21G//AM1 MEP (bottom) of 5HT calculated and displayed on the double van der Waals surface of the molecule. See text for more information.

6 ADPRT inhibitor calculations

This chapter presents the results of our calculations on the inhibitors of the enzyme ARPRT described in chapter 2. Figure 6.1 shows the inhibitors which we have studied in detail and for which results are reported here. The first ten compounds shown in this figure (A-F, I, K, L and 3-HB) were suggested to us by Cantoni [120], who has studied the effect of these compounds as inhibitors of ADPRT in Chinese hamster ovary (CHO) cells.

Cantoni's experimental system involves treatment of the cells with 1mM methyl methane sulphonate (MMS) for 30 minutes followed by a post incubation period of 5 hours in the presence or abscence of the ADPRT inhibitors. The level of DNA breakage is then measured immediately after exposure to MMS and after the repair interval by the alkaline elution assay. If the compound under test functions as an ADPRT inhibitor, then more DNA breaks will be present after the time allowed for repair when compared with cells which have been allowed to repair in absence of the test compound.

Figure 6.2 shows a typical elution profile of DNA from cells which were treated as indicated in the figure. The DNA from cells treated with MMS elutes very rapidly since the alkylating agent introduces many single strand breaks in the DNA. The profile of cells which have been incubated for the 5 hr recovery period is similar to that of the untreated cells since most of the DNA breaks have been repaired by ADPRT. When ADPRT inhibitors are added, the repair process is affected with more breaks being detectable at the end of the repair time. In this way the amount of breaks present is a function of the ability of the inhibitor to inhibit the enzyme.

Table 6.1 shows the results of Cantoni's experiments with different

concentrations of the benzamide derivatives in Figure 6.1. The results in this table are shown as strand scission factor (SSF) values. The SSF value is the relative number for the amount of DNA breakage present, the higher the SSF the greater the level of DNA breaks. The SSF value for cells treated with MMS but allowed to repair in the absence of ADPRT inhibitors was 0.1, therefore values in this table close to 0.1 indicate that the tested compound is not active as an inhibitor.

The last two inhibitors in Figure 6.1, theobromine and theophylline, have not been studied by Cantoni but have been studied by other groups [25, 59]. They are particularly interesting since they do not have the basic benzamide structure common to the other inhibitors shown but are quite potent inhibitors of ADPRT (see Table 2.1). Both theobromine and theophylline have a purine type structure where the two rings are coplanar. They are thought to inhibit ADPRT by competing for a different portion of the NAD binding site on the enzyme than the benzamide based inhibitors.

We have included the two purine analogues in our calculations in order to compare their MEP maps with those of the benzamide inhibitors and to hopefully reveal which molecular features are necessary for inhibition.

6.1 Notation used for rotamers

The following section describe the notation used in the remainder of this chapter to describe the rotamers of each inhibitor.

6.1.1 Notation for inhibitors A and B

Inhibitors A and B have three main torsion angles, denoted by R_1, R_2 and R_3 in Figure 6.3, around which rotations have been considered.

Each angle can be denoted by a single letter code which describes its value: therefore each of the individual rotamers can be identified by a simple three letter code.

Values of R_1 which result in the carboxamide carbonyl oxygen being *trans* to the the substituent on position 3 are denoted by the letter T, and *cis* conformations by the letter C. Similarly, values of R_2 which have the amine hydrogen cis or trans to the carboxamide group are denoted by the letters C and T. Values of R_3 which result in the carbonyl group being cis or trans to the amine hydrogen are denoted by C and T respectively. Examples of these three letter codes are given in Figure 6.3 for the conformations TTT and TCT. A total of 8 conformations are possible using this scheme although not all have been studied in detail.

6.1.2 Notation for inhibitor C

The notation used for the rotamers of inhibitor C is similar to that for inhibitors A and B described above. In this case only two torsion angles, R_1 and R_2 in Figure 6.3, are used. The notation used for angles R_1 and R_2 is the same as that described above for inhibitors A and B. We have not considered the possibility of inversion at the amine nitrogen which would effectively double the number of possible conformations. In this case four distinct conformations are possible.

6.1.3 Notation for inhibitors D and 3-HB

The notation used for inhibitors D and 3-HB is again similar to that for inhibitor C with the two angles R_1 and R_2 (Figure 6.3) being used. This leads to four distinct conformations.

6.1.4 Notation for inhibitor F

The rotamers of inhibitor F which have been studied can be described by the two torsion angles R_1 and R_2 in Figure 6.3. The notation used for angle R_1 is the same as for the other inhibitors. The notation used for the torsion angle R_2 is slightly different because of the need to distinguish between the two identical amino hydrogens. Instead of using one of the amino hydrogens to indicate C and T conformations as before we have used the symbols α and β to denote whether the hydrogens are above or below the plane of the ring respectively. The Tα and Tβ rotamers of inhibitor F are shown in Figure 6.3.

6.2 Optimised geometries of the inhibitors

As mentioned above, each inhibitor may exist in a number of different rotameric forms due to rotation around those bonds shown in Figure 6.3. This results in eight rotamers for inhibitors A and B and four rotamers for each of the inhibitors C, D, F and 3-HB.

Molecular geometries of each conformation were optimised using the AM1 method available in the MOPAC program package [61]. The reasons for the use of AM1 have already been discussed. In all cases the PRECISE option was used to ensure proper minimisation of gradients. Since the energy difference between individual rotamers was often very small, (~0.5 kcal/mol), we have also performed full geometry optimisations on each rotamer using the AM1 semi-empirical option available in Gaussian 86. This has the advantage of allowing the more powerful Berny and MS optimisation routines, available in Gaussian 86, to be used which ensures that the individual forces on each atom are properly minimised. All heats of formation and energy differences quoted in this

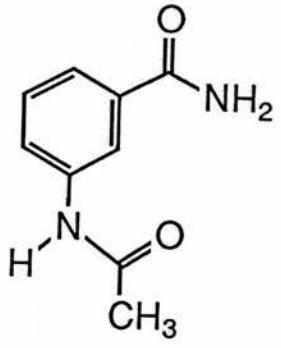
chapter and in Appendix A have been calculated from Gaussian 86 AM1 optimised structures. In all cases each rotamer has been characterised as a true minimum on the potential energy surface by inspection of the force constant matrix.

6.2.1 Rotamers of Inhibitors A and B

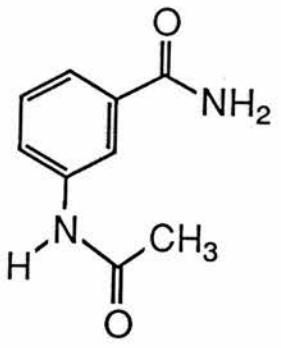
Inhibitors A and B differ only in the nature of the substituent X shown in Figure 6.3 and give rise to similar conformations where rotations are considered around R_1, R_2 and R_3. Of the eight possible rotamers of inhibitors A and B we shall only consider the four lowest energy rotamers here. The code letters used to describe each of the rotamers are given in Table 6.2 along with the energy difference between the lowest energy rotamer and the three rotamers of higher energy. The values of R_1, R_2 and R_3 for each rotamer are given in Table 6.3 along with their respective heats of formation and dipole moments.

In both cases the lowest energy rotamer corresponds to the TTT conformation shown in Figure 6.3. The energy difference between the TTT conformation and the next lowest in energy, the CCT conformation, is only around one tenth of a kcal/mol however. As is well known [121], the amido group is planar and gives rise to *cis* and *trans* isomers. In the case of inhibitors A and B these isomers correspond to rotations around torsion angle R_3.

If we consider the TTT conformation then this is the *trans* isomer with the TTC conformation corresponding to the *cis* isomer. In both inhibitors A and B the *cis* (TTC) isomer is around 2.2 kcal/mol less stable than the *trans* (TTT) isomer. This agrees with the work of Brown and Corbridge [122] on acetanilide where they found the *trans* conformation shown below to be more stable than the *cis* conformation.



(trans)



(cis)

We have also performed calculations on a related molecule where **X** in Figure 6.3 is hydrogen. In this case our calculations predicted the *cis* form to be more stable than the *trans* form. This result is also consistent with the experimental work of Buorn and coworkers [123] and La Planche and Rogers [124] on substituted formamides where they found the *cis* conformation to be more favoured than the *trans* conformation.

Our calculations predict the substituted amido group in inhibitors A and B to deviate from planarity by around 6 degrees with the angle made between the plane of the amido group and the ring to be 11 degrees. Experimental estimates of the angle between the ring and amido group in acetanilide range from 17° to 80°. Brown [125] originally found a dihedral angle of ~38° by X-ray diffraction but subsequently found a value of ~18° using more refined methods [126]. The larger experimental values for this angle were obtained from polarizability and dipole moment

measurements [127, 128].

The AM1 geometries of the most stable (TTT) rotamers of inhibitors A and B agree well with most of the experimental data on similar molecules. The calculations predict a near planar amido group with the angle between the amido plane and the ring of the same order as that for acetanilide in crystal form. The calculations also indicate the possibility of an intramolecular hydrogen bond being formed between the amido oxygen and H10 on the ring since the distance between the two is ~2.1 Å which is within the normal range for hydrogen bonds. Once again this is consistent with the crystal structure of acetanilide [126].

This re-affirms our confidence in the AM1 method since it seems to model this type of molecule satisfactorily.

6.2.2 Rotamers of inhibitor C

With inhibitor C we have only considered rotation about the two torsion angles R_1 and R_2 shown in Figure 6.3. This gives four distinct rotamers for this molecule. The order of stability of the four rotamers is given in Table 6.2 with R_1, R_2 values and heats of formation in Table 6.4. As mentioned previously we have not considered conformations arising from inversion at the nitrogen atom of the substituent $NHCH_3$ group. Figure 6.3 shows the conformation of the most stable of the rotamers (CT) which has the carboxamide group in the *cis* conformation.

The conformation of the $NHCH_3$ group agrees with calculations by Hehre *et al* [129, 130], on the conformation of the $NHCH_3$ group in N-methylaniline using the STO-3G basis set. The experimental value for the inversion barrier of the $NHCH_3$ group in N-methylaniline has also been determined [131, 132] and is found to be ≃0.5 kcal/mol.

6.2.3 Rotamers of inhibitor D and 3-HB

Since inhibitor D and 3-HB have a similar structure they will be considered together. There are four possible rotamers of each inhibitor arising from rotation around torsion angles R_1 and R_2 shown in Figure 6.3. The order of stability of the rotamers of both inhibitors is identical and is given in Table 6.2 along with respective energy differences. Tables 6.5 and 6.6 give the corresponding values of R_1 and R_2 for the rotamers of inhibitors D and 3-HB respectively.

In both inhibitors the AM1 method predicts the $-OCH_3$ and -OH groups to lie in the plane of the ring. This is consistent with results of previous calculations on anisole and phenol using PC1LO [133] and also ab-initio STO-3G calcuations by Hehre [129]. Experimental studies on anisole and phenol also show the planar conformation to be most stable.

6.2.4 Rotamers of inhibitor F

The order of stability of the four rotamers of inhibitor F is given in Table 6.2 along with the energy difference between rotamers. Figure 6.3 shows the torsion angles R_1 and R_2 used to define the rotamers and Table 6.7 give the respective values of R_1 and R_2 for each rotamer. As before inversion of the NH_2 group has not been considered. The structure of the substituent $-NH_2$ group in each rotamer is in good agreement with previous calculations on aniline [129, 130].

6.2.5 Structure of other inhibitors

The structures of all of the other inhibitors predicted by our AM1 calculations are in good agreement with available experimental or ab-initio structures. We should, at this point, mention the optimised

geometry of the benzamide molecule (inhibitor E) since it is the basic structure of all of the other inhibitors.

The experimental crystal structure [134] of benzamide shows that the amide group lies out of the plane of the benzene ring: the dihedral angle between the ring and amide group being 26 degrees. Our calculations predict a dihedral angle of 33 degrees in benzamide. This value shows little variation from this value for all of the above inhibitors.

6.2.6 Optimised geometries

Since there are a large number of rotamers for each inhibitor it is not appropriate to include the optimised geometry of each rotamer here. Instead we have included the optimised geometries of the rotamer of each inhibitor which we believe to be the active conformation in Appendix A. In some cases this is not necessarily the most stable rotamer. In any case the geometries of the other rotamers can be obtained by substituting the appropriate torsion angles from Tables 6.3-6.7.

6.3 Comparison of inhibitor A with purine inhibitors

As mentioned earlier the purine analogues theobromine and theophylline were included in our study of ADPRT inhibitors since they are believed to compete for a different portion of the NAD binding site on the enzyme than the benzamide inhibitors. It is possible therefore, that there are structural features common to both benzamide and purine inhibitors which are necessary for binding to the receptor.

Figure 6.4 shows the MEP maps for inhibitor A, the most potent of Cantoni's inhibitors, and for theobromine and theophylline. In each case the MEP has been calculated and displayed on the double van der Waals

surface of the molecule using the colours and contour levels given in Table 4.1. In this figure the TTT rotamer of inhibitor A is shown since this rotamer bears the closest resemblance to the structures of theobromine and theophylline. The TTT rotamer is also the most stable conformation of inhibitor A.

The most striking feature about all three MEP maps is their similarity. This is surprising since, at first glance, there does not appear to be any obvious similarity between their structures. On closer examination we found that fragments of the purine analogues were similar to the particular part of the structure of inhibitor A from O13 to H17. This is shown more clearly in Figure 6.5. In all cases each fragment has a similar structure consisting of a chain of 7 atoms. At one end of the chain lies a carbonyl oxygen which has a large negative EP and at the other end a hydrogen atom which has a large positive EP. The atoms in each fragment which connect the oxygen and hydrogen atoms are not identical.

Figure 6.5 also shows a graph of the average MEP's on the atoms of each fragment. It is clear from this graph that the MEP on each of the purine fragments is similar to the MEP on the fragment of inhibitor A. Thus although the structures are different we can see that each molecule does exhibit a particular region which has similar electrostatic properties.

Figure 6.6 shows the common fragments of theobromine and theophylline superimposed upon the fragment of inhibitor A. In this case both molecules show a reasonable fit to inhibitor A, although theophylline does seem to have the best fit. It is interesting to note that theophylline is a slighly more powerful inhibitor than theobromine (see Table 2.1)

Based upon the above results we can propose a tentative binding site with suitable topography and electrostatic features to accommodate both purine and benzamide inhibitors. Figure 6.7 shows one possibility where the shape of the site and location of charges have been arranged to complement those of the inhibitors shown in Figures 6.4 and 6.6. This binding site has charged regions at the locations denoted by site A, site B and site C in Figure 6.7. It is clear from this figure that both benzamide and purine inhibitors can easily be accommodated within the binding site. Binding could occur through electrostatic interaction at the three sites mentioned above.

If we consider the structure and MEP of the benzamide inhibitors then the most probably sites for binding are site A and site B. It is also possible that some long range attraction can occur at site C. The structure and MEP of the purine inhibitors is such that they can interact with all three sites with strongest binding occuring at sites A and C.

6.4 MEP's of benzamide inhibitors

This section deals with the remainder of the benzamide based inhibitors shown in Figure 6.1 and attempts to explain the ability of these compounds to inhibit ADPRT based upon their MEP's and the requirements of the hypothetical binding site proposed above.

6.4.1 Inhibitor B

Figure 6.8 shows the MEP of the TTT rotamer of inhibitor B along with the same rotamer of inhibitor A. As with inhibitor A, the TTT rotamer of inhibitor B is the most stable of all the rotamers studied and is the rotamer which we consider to be the active conformation. As can

be seen from this figure there is little difference between the MEP maps of both molecules when compared visually. This is also clear from the average MEP's for each inhibitor, which are shown in Figure 6.8 where the average MEP curves for both inhibitors are identical.

This suggests that inhibitor B should have similar activity to inhibitor A: this is confirmed by the results given in Table 6.1. The fact that inhibitor A is slightly more powerful than inhibitor B (SSF 0.69 and 0.61 respectively) may be attributable to the longer chain length of the propionyl group, in inhibitor A, compared with the acetyl group in inhibitor B. Since the enzyme is located in the nucleus, then any potential inhibitor will first have to penetrate the cell membrane before it can affect the enzyme. A longer substituent chain length may increase the lipophilicity of an inhibitor thus allowing it easier passage across the cell membrane. The interaction of inhibitor B with the proposed binding site in Figure 6.7 should be similar to inhibitor A.

6.4.2 Inhibitor C

Figure 6.9 shows the MEP of the **TT** rotamer of inhibitor C. This rotamer is the highest energy rotamer of inhibitor C, being 0.6 kcal/mol less stable than the lowest energy **CT** rotamer. We have chosen this rotamer since it is the only one which has a similar conformation to the TTT rotamer of inhibitor A discussed earlier and also shown in Figure 6.9.

The MEP of inhibitor C is similar to that of inhibitor A when compared on the carboxamide and benzene ring parts of both molecules. The greatest difference occurs on the $-NHCH_3$ group of inhibitor C. The most obvious difference is that the large positive region (blue) around H17 of inhibitor A is less positive, in inhibitor C. Similarly the area of

negative EP around O19 in inhibitor A has a corresponding area in inhibitor C which is positive. These differences are most likely to be attributable to the different electronic properties of the $NHCH_3$ and $NHCOC_2H_5$ groups.

If we now consider the effect which these differences in MEP have on the binding of inhibitor C to the proposed binding site, in Figure 6.7, we can see that binding *via* the carboxamide group at site A will not be affected appreciably. The effect of binding at sites B and C will however be significantly reduced. When we also consider that this rotamer is the least stable of the rotamers of inhibitor C which we have studied, then the proportion of molecules in the TT conformation will be small. We would not then expect this molecule to be a very potent inhibitor. The results in Table 6.1 show that inhibitor C is one of the least effective inhibitors of ADPRT.

6.4.3 Inhibitor D

Figure 6.10 shows the MEP of the TT rotamer of inhibitor D along with the TTT rotamer of inhibitor A for comparison. We have chosen the TT rotamer of inhibitor D since it bears the closest resemblance to inhibitor A both in terms of conformation and MEP. Once again the MEP on the ring and carboxamide parts of both molecules is similar. The MEP of inhibitor D is also positive on the $-OCH_3$ methyl group which is in a similar position to the positive region around H17 on inhibitor A although of reduced magnitude. The negative MEP around O19 on inhibitor A has no counterpart in inhibitor D since the $-OCH_3$ group has no atoms in this position. The MEP on atom O12 in inhibitor D is very negative which may offset this slightly.

The TT rotamer of inhibitor D is the second most stable of the

four rotamers being only 0.2 kcal/mol less stable than the lowest energy CC rotamer. This should ensure that a reasonable proportion of molecules are present in the TT conformation. We would therefore expect this compound to be a reasonably active inhibitor of ADPRT. Table 6.1 shows that this molecule is a reasonably potent inhibitor. Interaction with our proposed binding site in Figure 6.7 should again occur at sites A and B.

6.4.4 Inhibitor E

The MEP of inhibitor E is shown in Figure 6.11 along with the MEP of inhibitor A. In this case there is only one conformation for inhibitor E since it does not have a substituent in the meta position. The MEP of both inhibitors is similar on the ring and carboxamide parts of the molecules. It is clear from Figure 6.11 that the area of positive MEP around H17 in inhibitor A does not exist in inhibitor E since there is no substituent group. This also applies to the area of negative MEP around O18 in inhibitor A. Thus binding of inhibitor E will only occur at carboxamide site A in Figure 6.7. The efficiency of benzamide as an ADPRT inhibitor will therefore be solely dependent upon the extent of binding at this site.

Table 6.1 shows that benzamide is a more powerful inhibitor than inhibitor C but less powerful than inhibitors A, B or D. This suggests that the ability of inhibitors A, B and D to interact with the other two binding sites shown in Figure 6.7 brings about an increase in potency: the stronger the interaction the more powerful the inhibitor.

6.4.5 Inhibitor F

Figure 6.12 shows the MEP of inhibitor F along with that of

inhibitor A. The rotamer of inhibitor F shown is the Tβ rotamer which is the least stable of the rotamers studied. Once again the MEP on the ring and carboxamide parts of both molecules are similar. Inhibitor F has a region of positive MEP around H17, which is slighly less positive than the same region in inhibitor A. The MEP around H18 is also positive whereas the corresponding MEP in inhibitor A is negative. This would favour binding at sites A and B of the binding site in Figure 6.7. This is similar to inhibitor C which like inhibitor F should bind at sites A and B. The similarity also extends to the stability, where the Tβ rotamer of inhibitor F is also the least stable of all the rotamers studied. This suggests the inhibitor F should have similar potency to inhibitor C. Table 6.1 shows that inhibitor F is slighly more potent than inhibitor C.

6.4.6 Inhibitor I

Figure 6.13 shows the MEP of inhibitor I along with that of inhibitor A. It is immediately obvious that the MEP maps of these two molecules differ considerably from each other. Examination of the average MEP graph shows that the shape of the curve is similar to that for inhibitor A but is shifted towards more positive values. It is also clear that both oxygen atoms of the nitro group have a large negative EP. This is consistent with the nitro group withdrawing electrons from the ring system thus making it more positive. The binding of this molecule to the proposed active site in Figure 6.7 should be reduced at site A and repulsive at site B. Binding at site C should be favourable. Since the molecule is a very poor inhibitor (see Table 6.1) this would suggest that site C is of less importance for binding benzamide type inhibitors than sites A or B.

6.4.7 Inhibitor 3-HB

Figure 6.14 shows the MEP map for the TT rotamer of inhibitor 3-HB. This inhibitor is similar to inhibitor D in both MEP and stability of rotamers. Consideration of the MEP map shows that this molecule should also favour binding at sites A and B of the proposed binding site. Table 6.1 shows that 3-HB is similar in activity to inhibitor D.

6.4.8 Other inhibitors

Figure 6.15 shows MEP's of nicotinamide and 4-aminobenzamide. The MEP of nicotinamide is similar to that of benzamide which is what we would expect since nicotinamide also inhibits ADPRT and is present as part of the substrate NAD. We have included the MEP of nicotinamide for purposes of comparison only.

The MEP of 4-aminobenzamide, inhibitor K, is interesting since it has electrostatic properties which suggest that it would bind to our proposed binding site at site A and to some extent at site B. The activity based upon the above binding should be similar to that of benzamide. Cantoni's results in Table 6.1 show that this molecule does not act as an inhibitor in his particular cell system. Other groups, however, have indicated that 4-aminobenzamide does inhibit ADPRT. Sims and co-workers [59] report that 4-aminobenzamide acts as a weak inhibitor of ADPRT in nucleotide-permeable human lymphocyte cells. This conflict of opinion possibly arises from the different techniques or cell lines used in these studies. Cantoni, for example, measures ADPRT activity after a 5 hour period as described in the introduction to this chapter. Sims, on the other hand, measures activity following stimulation by 2 mg/mL of L-phytohemagglutinin for a period of three days.

In general the results obtained by Cantoni compare well with those

of other groups. Compounds which have similar structure to inhibitor B (3-acetamidobenzamide), in particular, are generally found to be the most potent inhibitors.

6.5 Dipole Moments

Cantoni has recently suggested [135] that the ADPRT binding activity of a molecule is an inverse function of its dipole moment. These conclusions were based upon experiments carried out on structural analogues of benzamide containing a sulphur atom, such as benzene sulphonamide and thiobenzamide. Most of the compounds included in these experiments were poor ADPRT inhibitors.

While the ADPRT activity may be inversely related to the dipole moment in the case of the sulphur analogues, this theory does not agree with our calculations on the benzamide inhibitors of ADPRT. Table 6.8 shows the calculated dipole moment for the rotamer of each inhibitor which we belive to be the active rotamer and also the calculated dipole moment of the most stable rotamer of each inhibitor. This data clearly shows that the activity of the inhibitor is not a function of the inverse of the dipole moment. If any conclusion can be made from our results it is that the opposite is true where the least potent inhibitors are those which have small dipole moments.

This trend is based upon the calculated dipole moment of the most stable rotamer of each inhibitor which should be comparable with the experimental dipole moment. Since it may not always be the most stable conformation of a molecule which is responsible for binding, and hence activity, then we would suggest that experimental dipole moments are not a reliable indicator of ADPRT activity. See for example inhibitors C and F in Table 6.8

Table 6.1 Strand Scission factors of Cantoni's benzamide based inhibitors.

		Strand Scission Factors (SSF)[a]		
Inhibitor	Inhibitor name	1.0 mM	3.0 mM	5.0 mM
A	3-propionamido-BA	0.38±0.034	0.69±0.058	-
B	3-acetamido-BA	0.34±0.030	0.61±0.060	-
C	3-methylamino-BA	-	-	0.25±0.03
D	3-methoxy-BA	0.21	0.38±0.037	0.51±0.051
E	Benzamide	0.189	0.22±0.028	0.39±0.035
F	3-amino-BA	-	0.16±0.012	0.31±0.033
I	3-nitro-BA	0.12	0.12	0.14
K	4-amino-BA	0.1	0.1	0.1
L	2-amino-BA	0.1	0.1	0.1
3-HB	3-hydroxy-BA	0.29±0.032	0.42±0.05	0.52±0.07
Control		0.1	0.1	0.1

(a) See text for definition of relation between SSF and inhibitor activity

Table 6.2 **Conformation and energy differences for the four lowest energy rotamers of benzamide inhibitors. See text for explanation of notation.**

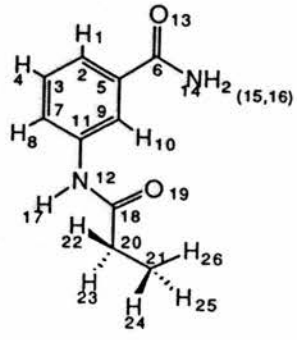
	Conformation			
Inhibitor	1	2	3	4
A	TTT	CCT	TCT	CTC
B	TTT	CCT	TCT	CTC
C	CT	TC	CC	TT
D	CC	TT	TC	CT
F	Cβ	Tα	Cα	Tβ
3-HB	CC	TT	TC	CT

	Energy difference (kcal/mol)			
Inhibitor	1	2	3	4
A	0	0.088	0.785	1.125
B	0	0.087	0.788	1.123
C	0	0.256	0.313	0.582
D	0	0.203	0.429	0.608
F	0	0.235	0.344	0.622
3-HB	0	0.153	0.384	0.583

Table 6.3 Torsion angles R_1, R_2 and R_3, heats of formation and dipole moments for the four lowest energy rotamers of inhibitor A (top) and inhibitor B (bottom)

Rotamer	R_1	R_2	R_3	ΔHf (kcal/mol)	Dipole Moment (Debye)
TTT	33.2	-11.4	2.2	-59.608	5.75
CCT	-146.8	174.1	0.2	-59.520	1.62
TCT	40.6	-171.2	0.0	-58.823	5.18
CTC	-143.3	-32.0	-178.5	-58.482	5.07

R_1 = C2C5C6O13
R_2 = C7C11N12H17
R_3 = C11N12C18O19



Rotamer	R_1	R_2	R_3	ΔHf (kcal/mol)	Dipole Moment (Debye)
TTT	33.0	-11.1	2.5	-53.868	5.76
CCT	-146.4	174.5	0.1	-53.785	1.60
TCT	40.6	-172.6	0.5	-53.080	5.22
CTC	-143.6	-29.4	-176.4	-52.745	5.23

R_1 = C2C5C6O13
R_2 = C7C11N12H17
R_3 = C11N12C18O19

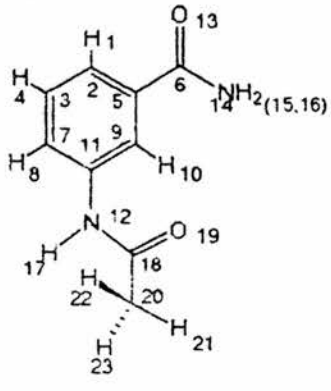


Table 6.4 Torsion angles R_1 and R_2, heats of formation and dipole moments of the four lowest energy rotamers of inhibitor C.

Rotamer	R_1	R_2	R_3	ΔHf (kcal/mol)	Dipole Moment (Debye)
CT	-142.3	25.7	-	-14.199	2.83
TC	40.4	-156.9	-	-13.943	4.09
CC	-145.4	-161.9	-	-13.886	4.19
TT	37.9	25.1	-	-13.618	5.17

R_1 = C2C5C6O13
R_2 = C7C11N12H17

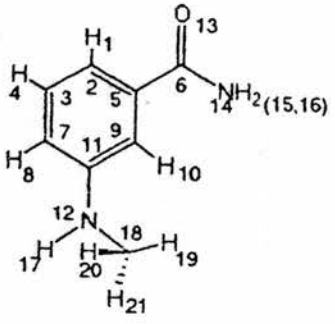


Table 6.5 Torsion angles R_1 and R_2, heats of formation and dipole moments of the four lowest energy rotamers of inhibitor D.

Rotamer	R_1	R_2	R_3	ΔHf (kcal/mol)	Dipole Moment (Debye)
CC	-144.7	176.7	-	-54.015	2.67
TT	34.3	-3.3	-	-53.813	4.07
TC	39.1	-175.3	-	-53.587	3.75
CT	-140.0	2.7	-	-53.407	4.61

R_1 = C2C5C6O13
R_2 = C7C11O12C12H17

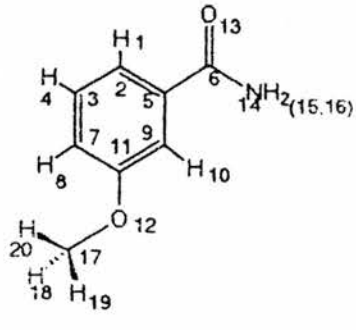


Table 6.6 Torsion angles R_1 and R_2, heats of formation and dipole moments of the four lowest energy rotamers of inhibitor 3-HB.

Rotamer	R_1	R_2	R_3	ΔHf (kcal/mol)	Dipole Moment (Debye)
CC	-145.9	177.4	-	-60.222	3.05
TT	34.6	-1.9	-	-60.070	3.59
TC	37.7	-176.8	-	-59.839	3.34
CT	-140.7	2.2	-	-59.639	4.67

R_1 = C2C5C6O13
R_2 = C7C11O12H17

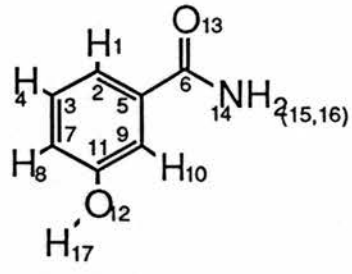


Table 6.7 Torsion angles R_1 and R_2, heats of formation and dipole moments of the four lowest energy rotamers of inhibitor F.

Rotamer	R_1	R_2	R_3	ΔHf (kcal/mol)	Dipole Moment (Debye)
Cβ	-141.9	26.3	-	-17.850	2.89
Tα	39.8	-157.1	-	-17.615	3.80
Cα	-145.6	-161.4	-	-17.506	4.23
Tβ	37.1	25.1	-	-17.228	5.15

R_1 = C2C5C6O13
R_2 = C7C11N12H17

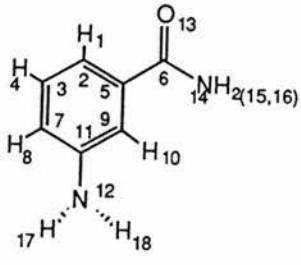


Table 6.8 **Calculated dipole moments of the benzamide inhibitors where (A is the rotamer refered to in text, and B is the most stable rotamer)**

Inhibitor	Dipole Moment (Debye)	
	A	B
A	5.8	5.8
B	5.8	5.8
3HB	3.6	3.1
D	4.1	2.7
E	3.7	3.7
F	5.2	2.9
C	5.2	2.8
I	2.8	2.8
THEOB[a]	4.1	4.1
THEOP[a]	3.3	3.3

[a] These purine inhibitors have been included for comparison.

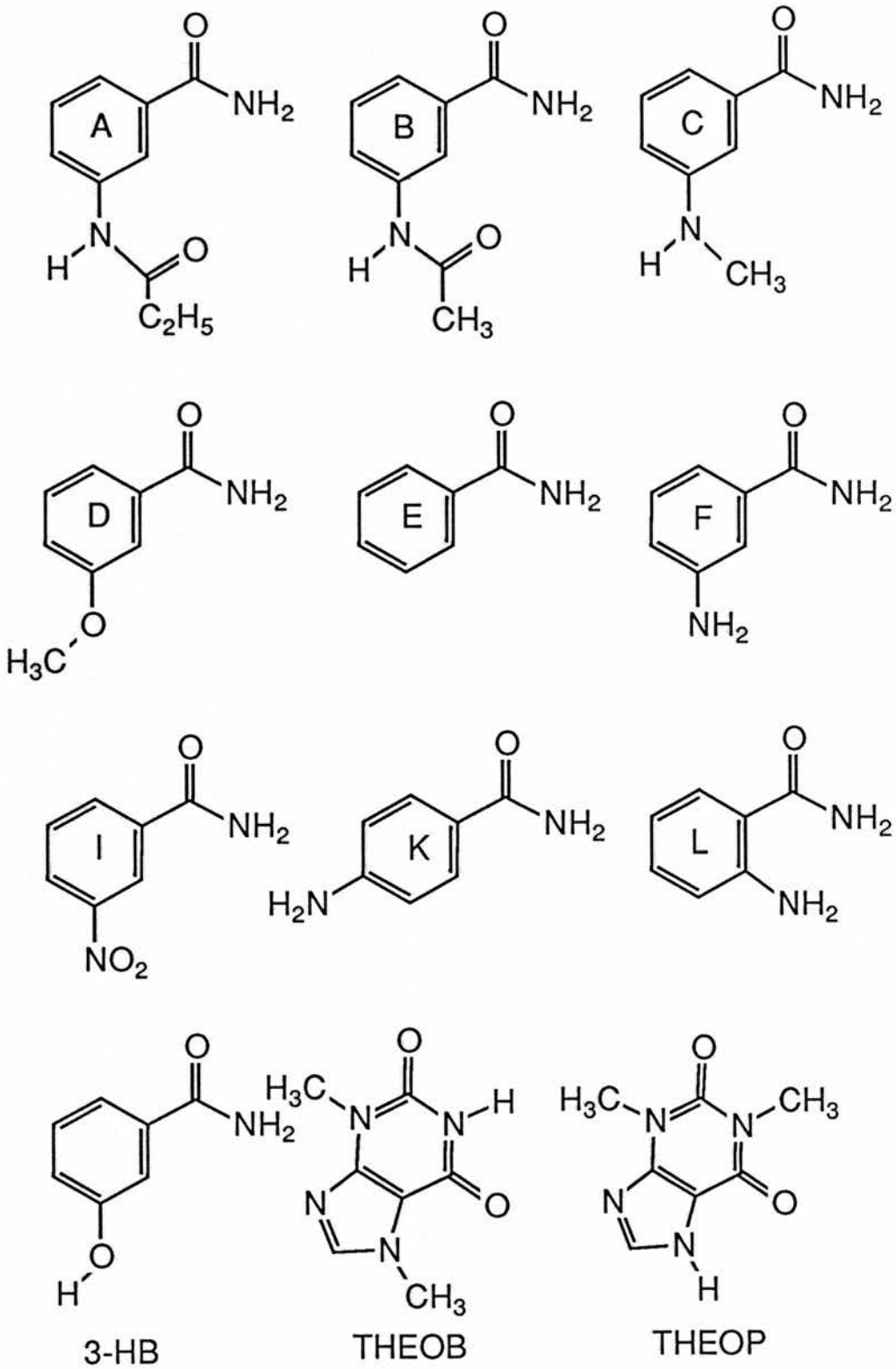


Figure 6.1 Enzyme inhibitors of ADPRT which have been studied in this work.

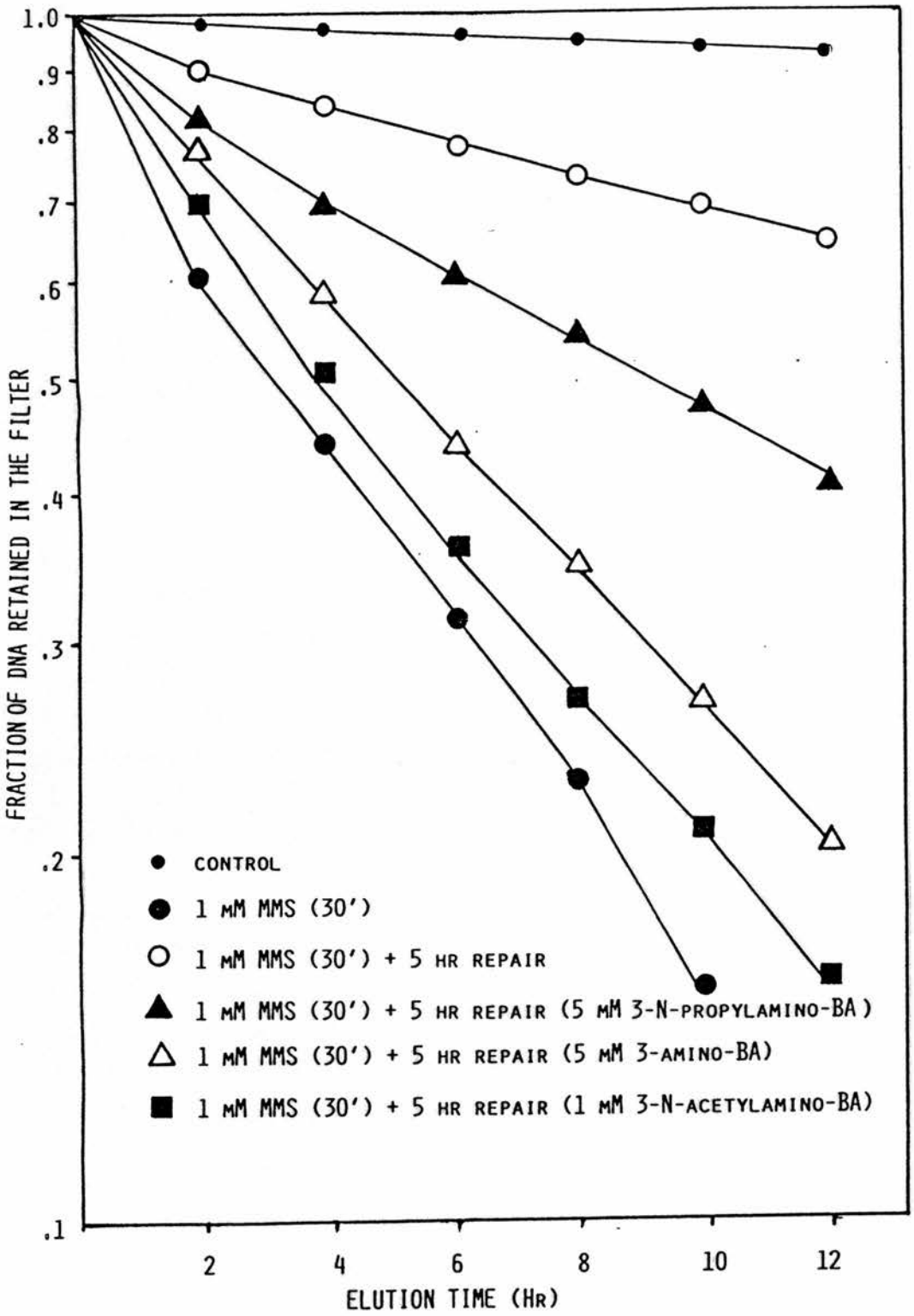


Figure 6.2 Typical elution profile of DNA from CHO cells treated with various combinations of MMS and ADPRT inhibitors.

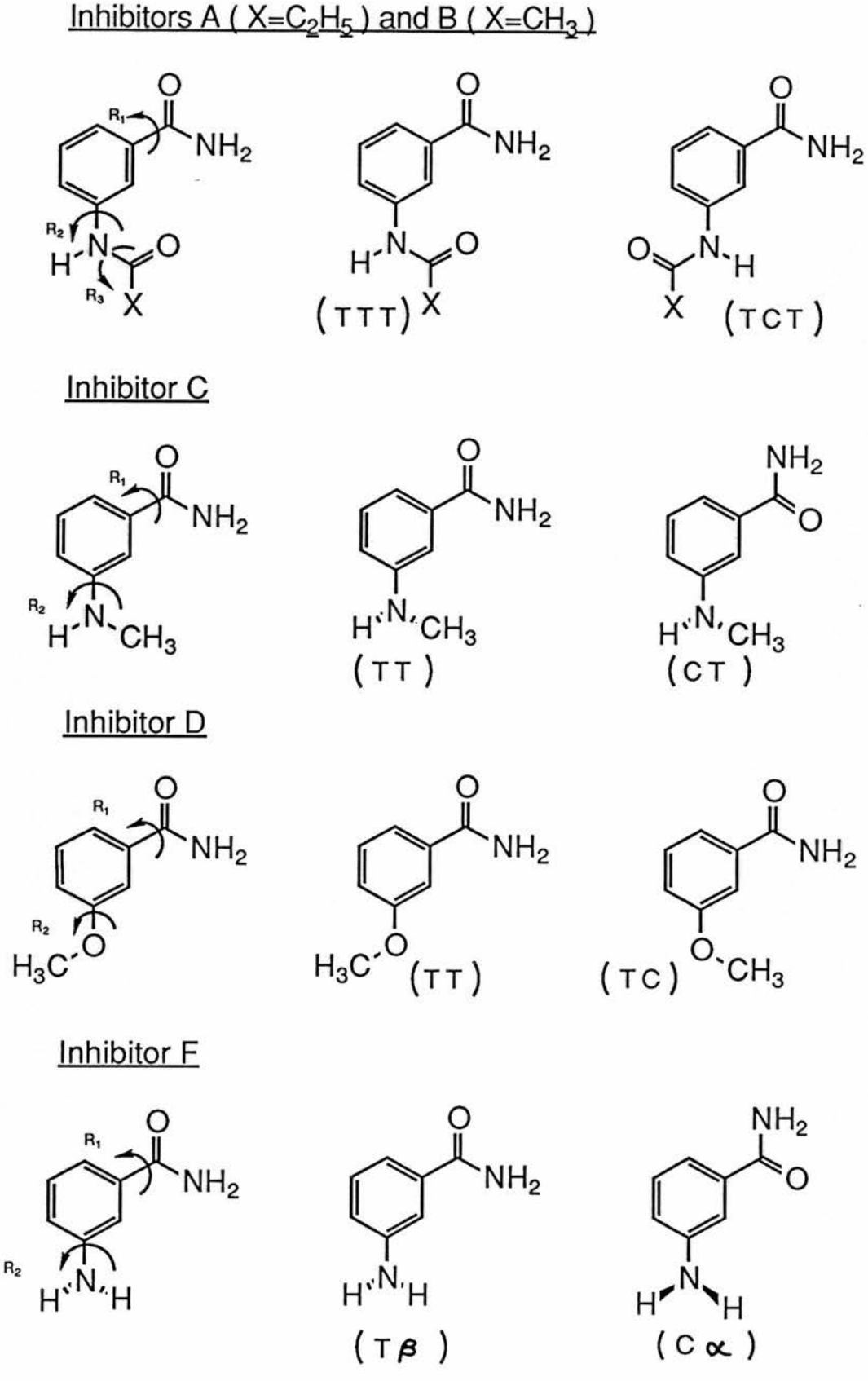


Figure 6.3 Dihedral angles around which rotations have been considered for inhibitors A, B, C, D and F.

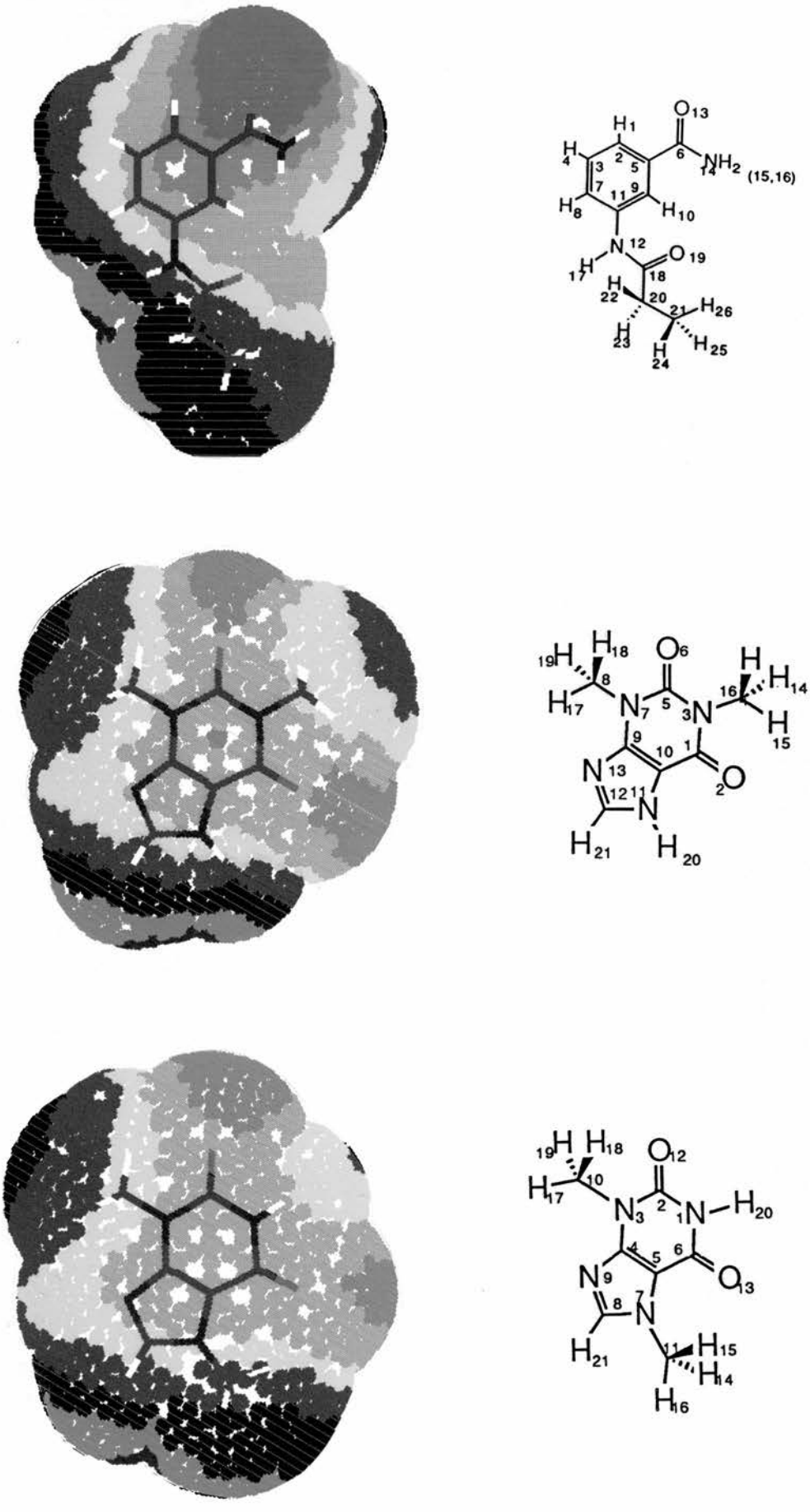
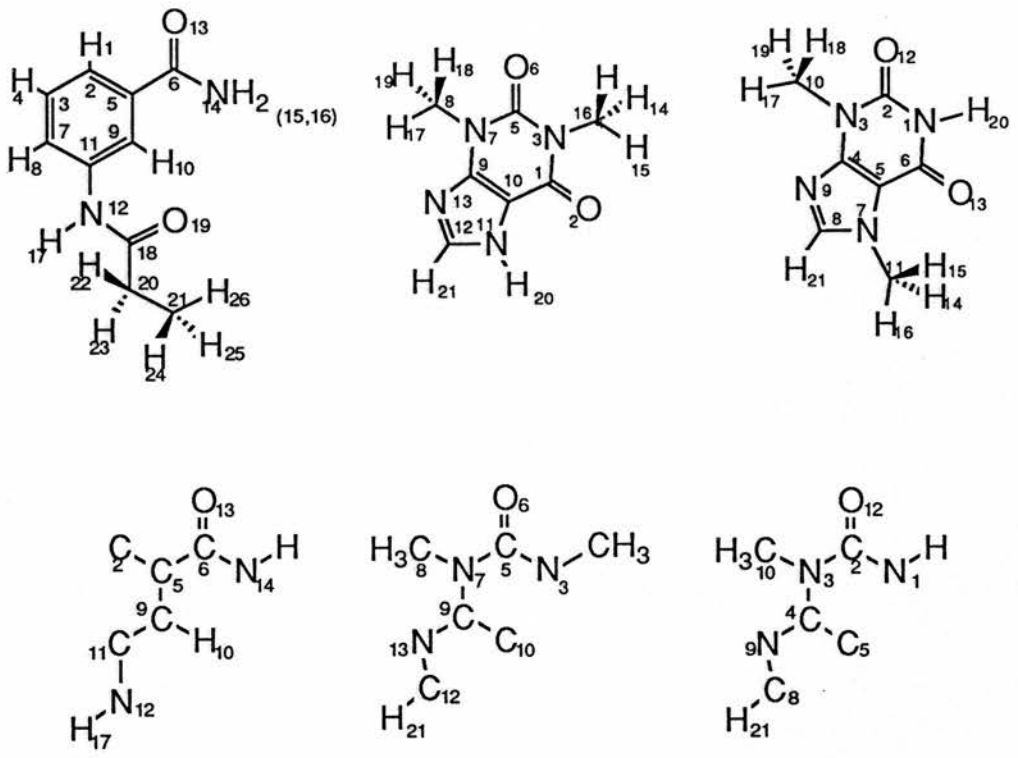


Figure 6.4 The AM1 point charge MEP's for inhibitor A (top), theophylline (middle) and theobromine (bottom), calculated on the double van der Waals surface of the molecules.



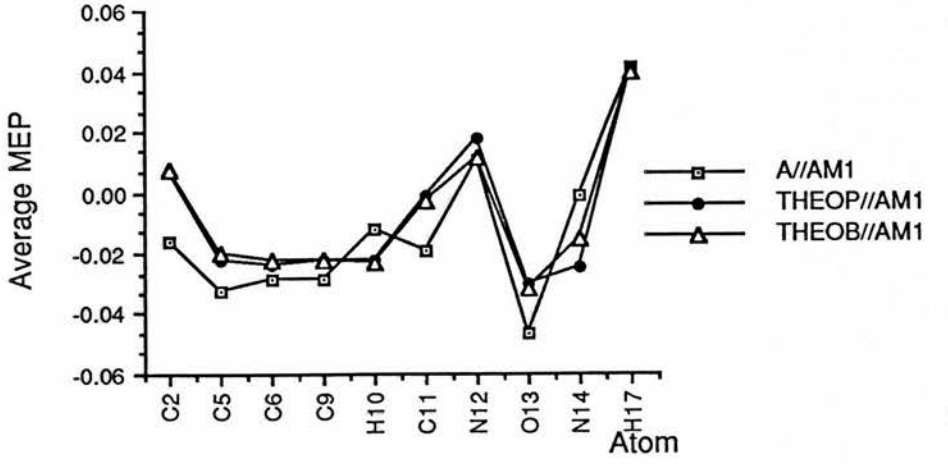


Figure 6.5 Comparison of AM1 point charge average atomic MEP's calculated on the double van der Waals surface of the molecules indicated. MEP's are compared on the fragments shown in the middle of the figure only.

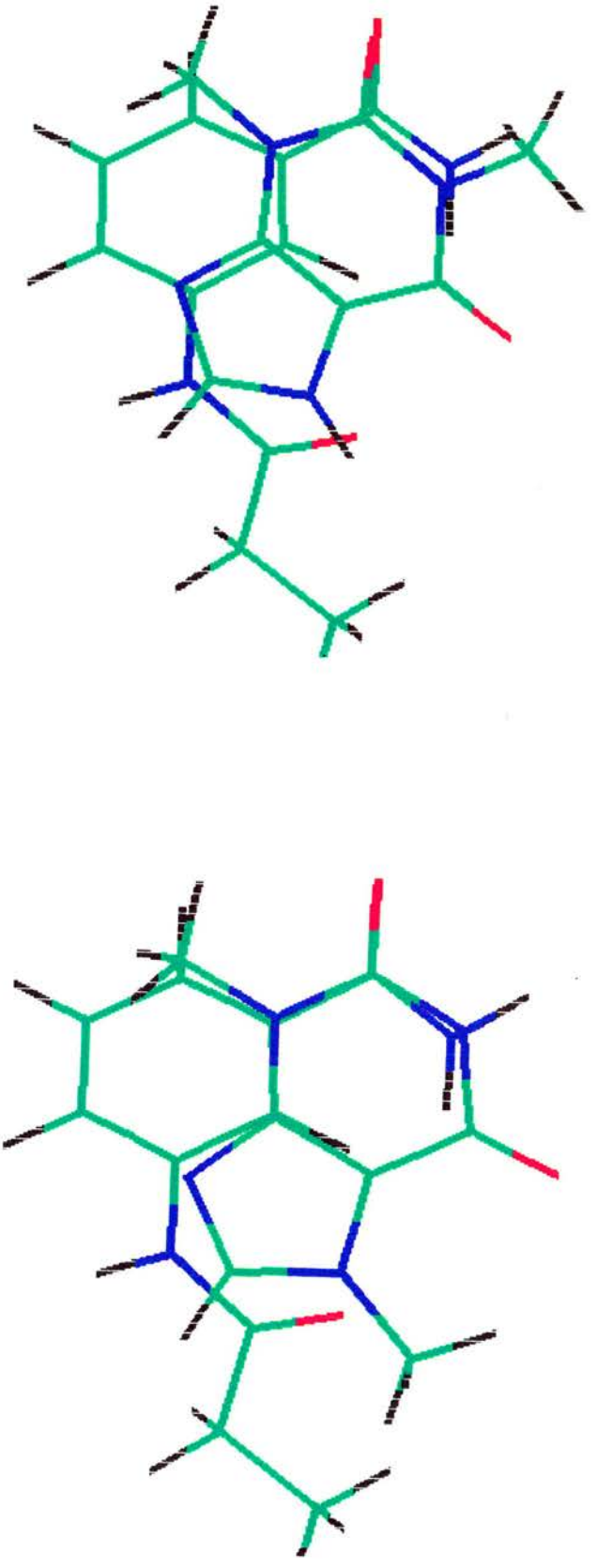


Figure 6.6 Common fragments of theophylline (top) and theobromine (bottom) superimposed on inhibitor A. See Figure 6.5 for identification of fragments.

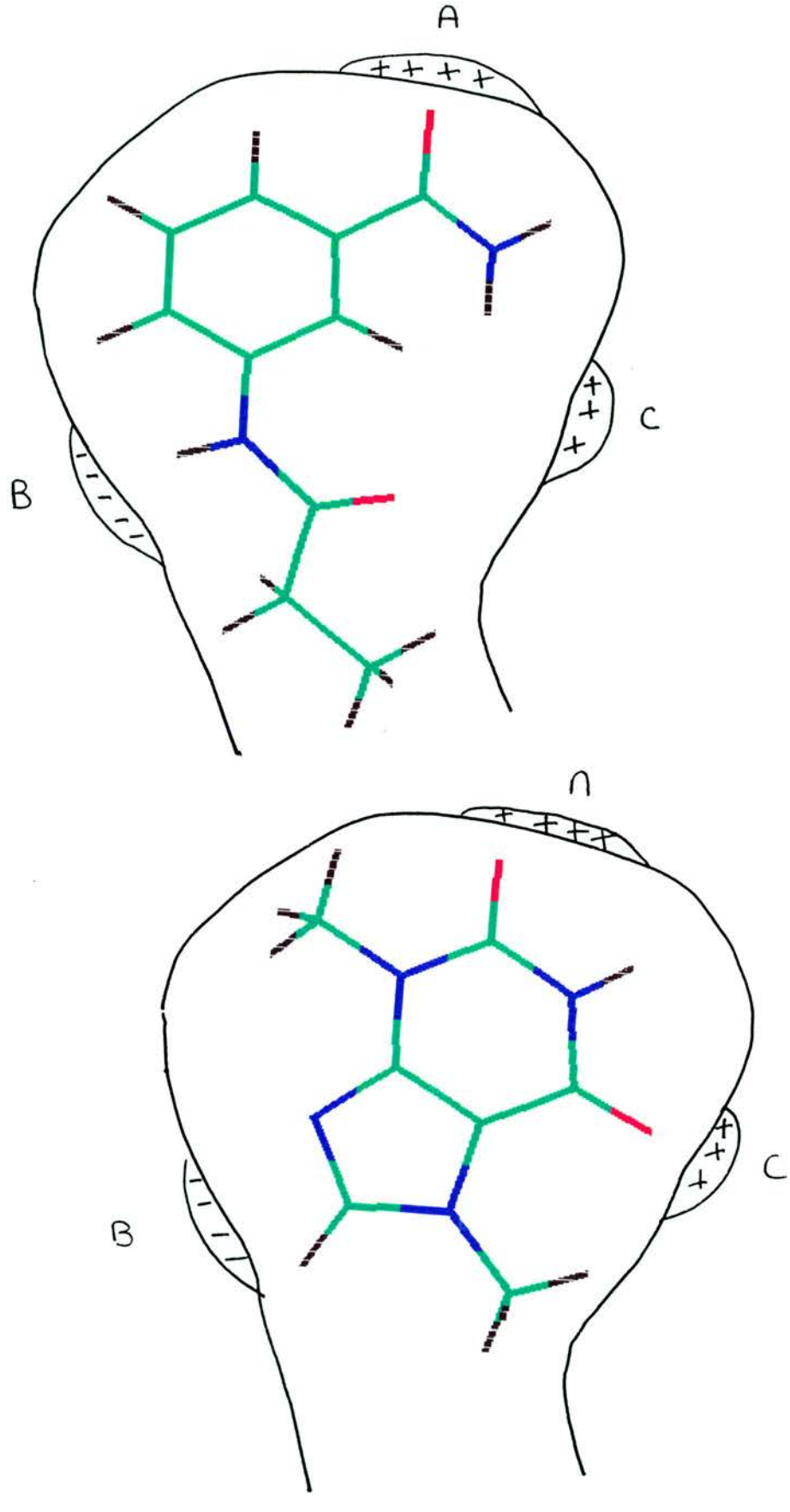


Figure 6.7 Possible topography of the NAD binding site of ADPRT, showing inhibitor A (top) and theobromine (bottom) fitted in the site.

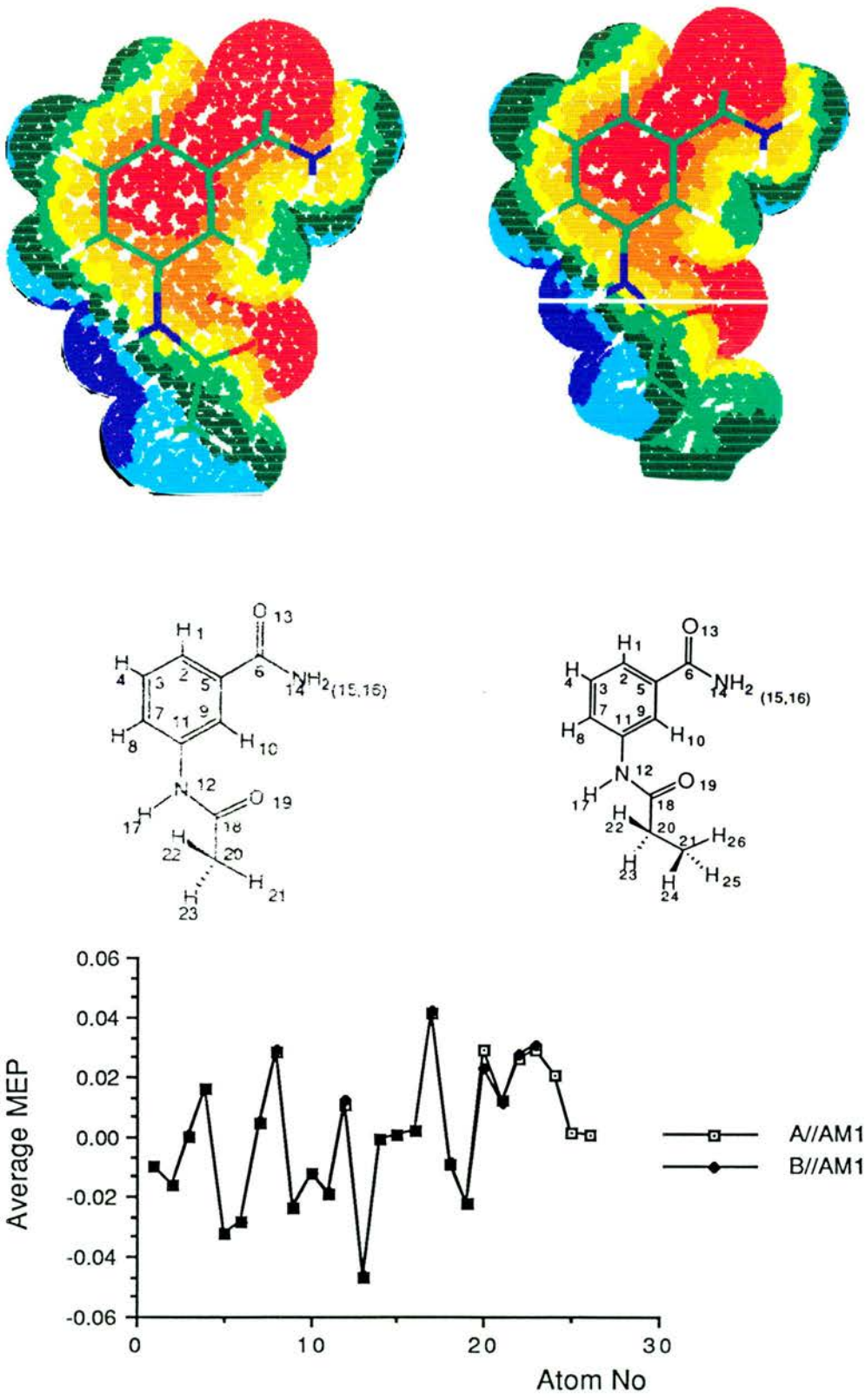


Figure 6.8 Comparison of AM1 point charge MEP's of inhibitor B (top left) and inhibitor A (top right). Average atomic MEP's were calculated on the double van der Waals surface.

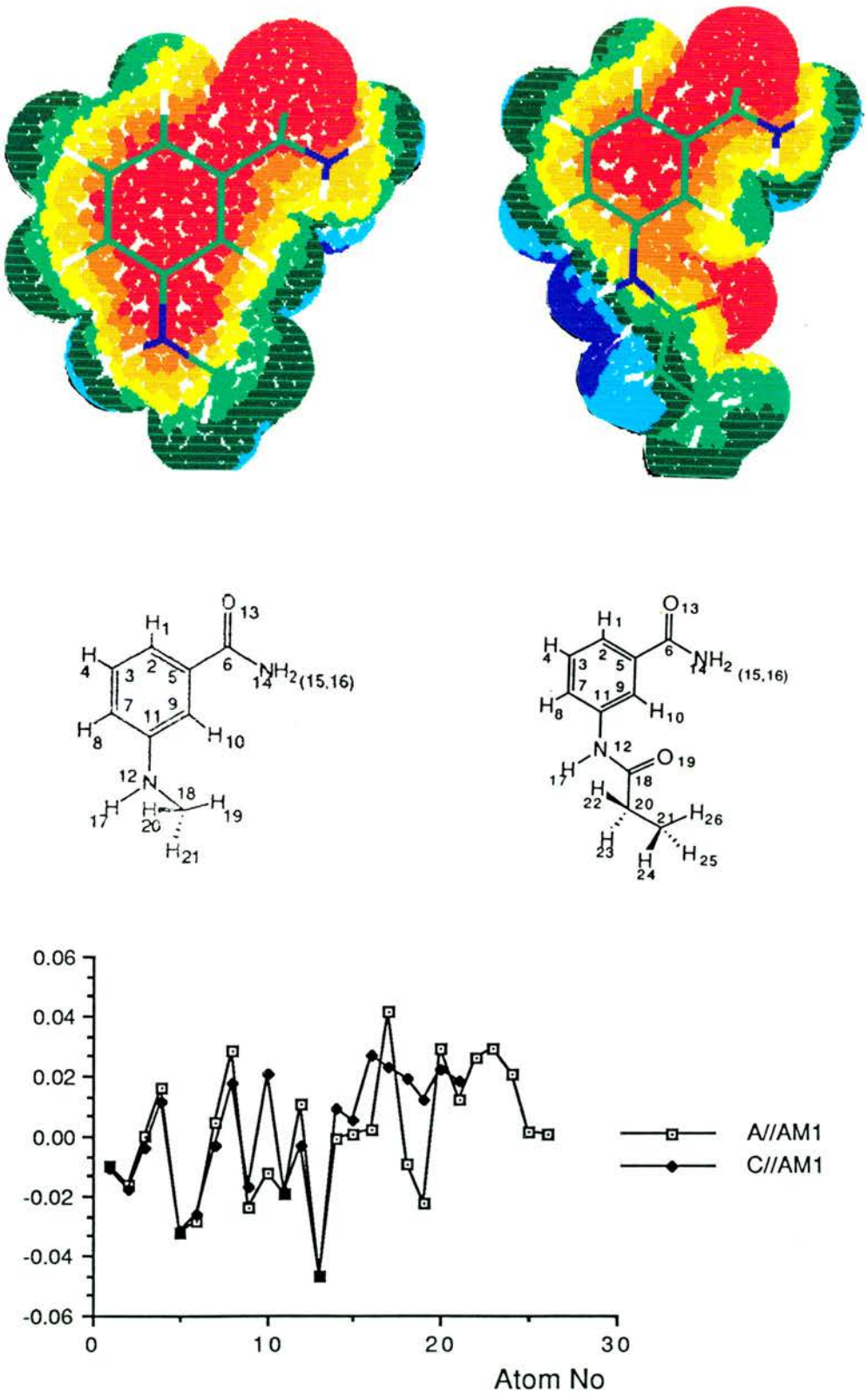


Figure 6.9 Comparison of AM1 point charge MEP's of inhibitor C (top left) and inhibitor A (top right). Average atomic MEP's were calculated on the double van der Waals surface.

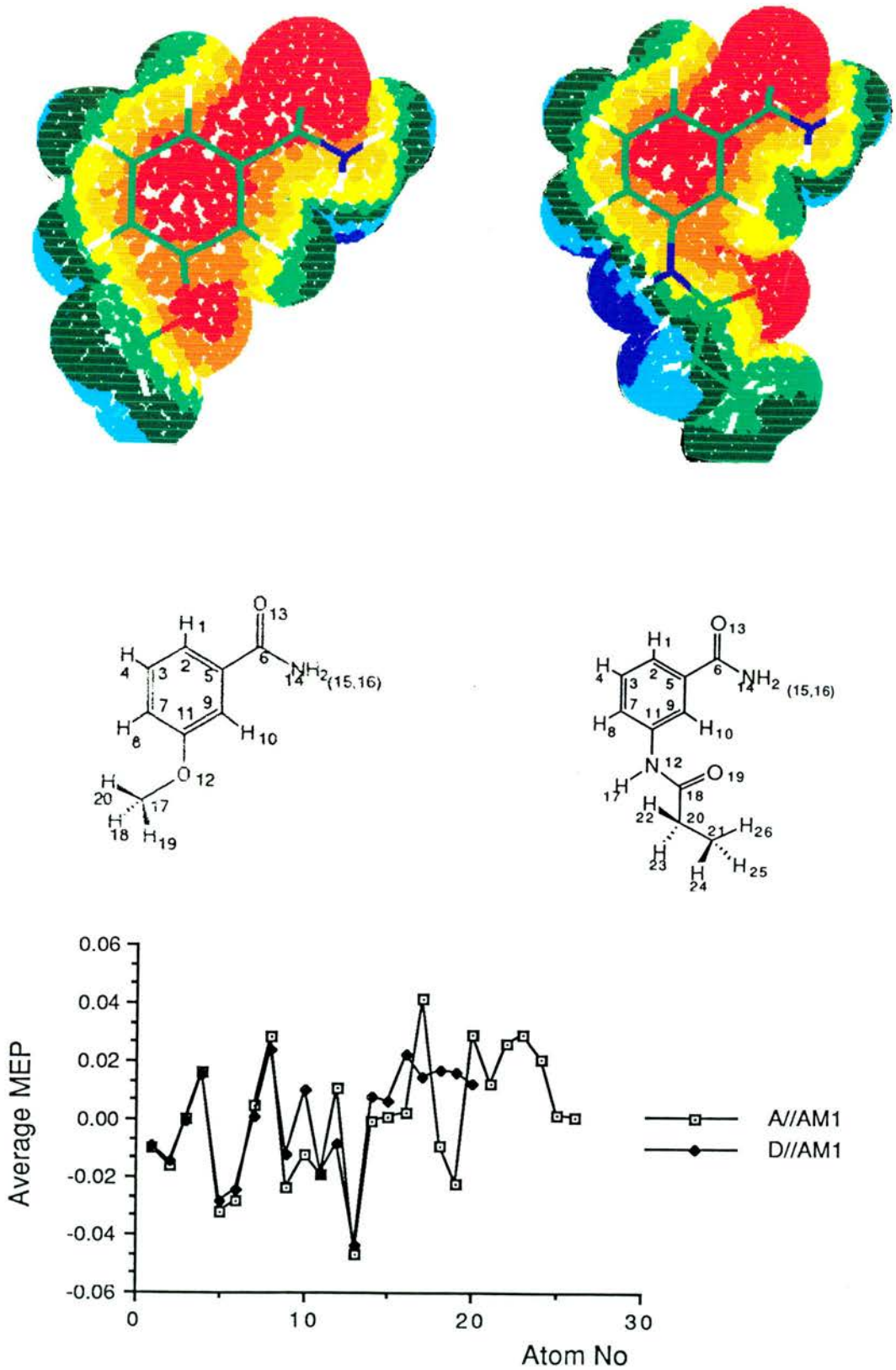
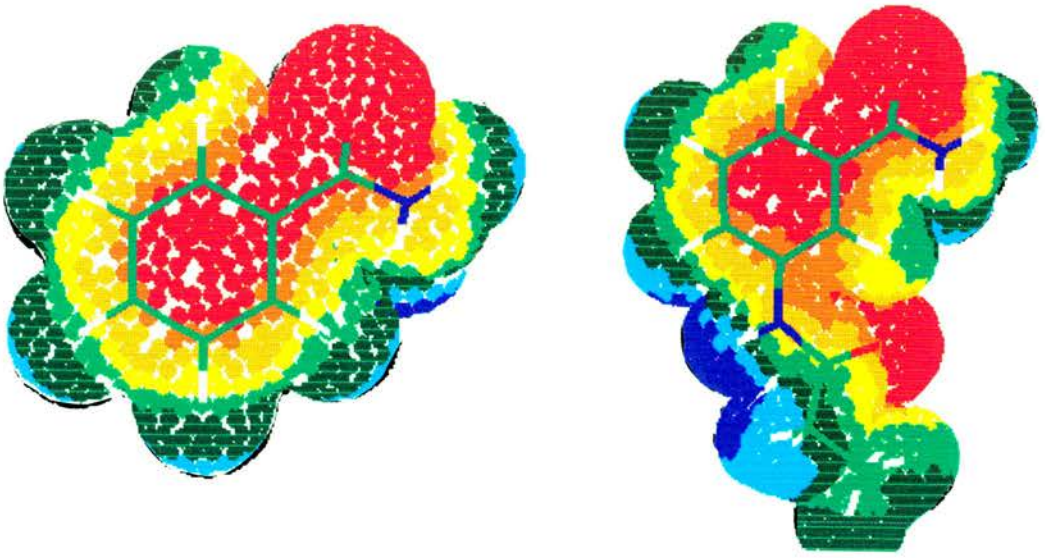
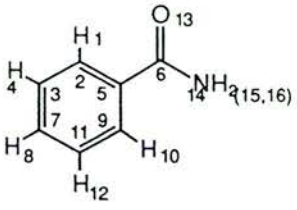
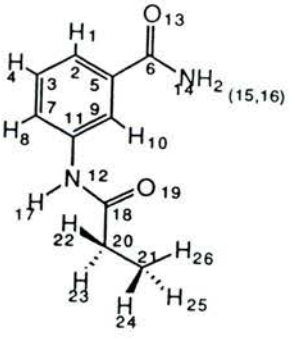


Figure 6.10 Comparison of AM1 point charge MEP's of inhibitor D (top left) and inhibitor A (top right). Average atomic MAP's were calculated on the double van der Waals surface.







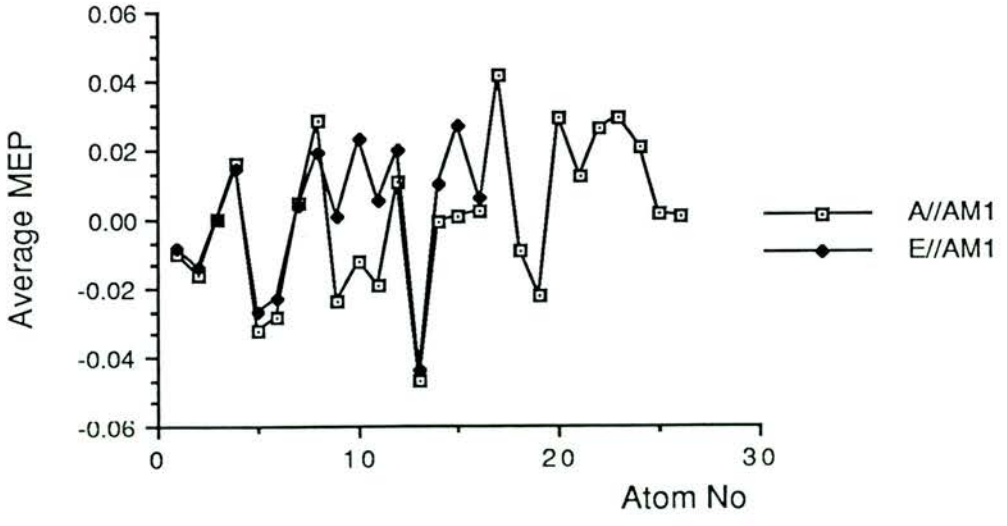


Figure 6.11 Comparison of AM1 point charge MEP's of inhibitor E (top left) and inhibitor A (top right). Average atomic MEP's were calculated on the double van der Waals surface.

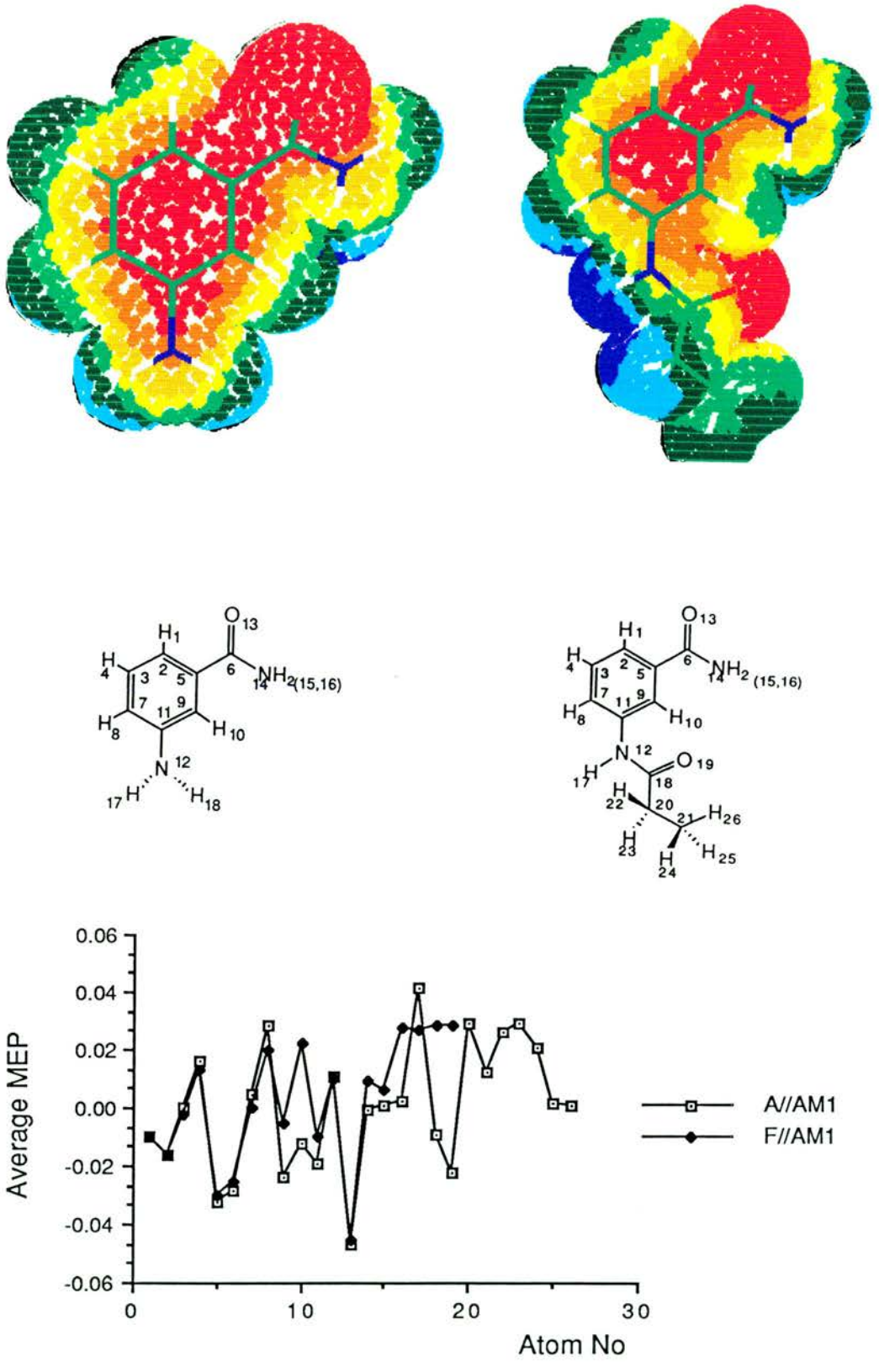


Figure 6.12 Comparison of AM1 point charge MEP's of inhibitor F (top left) and inhibitor A (top right). Average atomic MEP's were calculated on the double van der Waals surface.

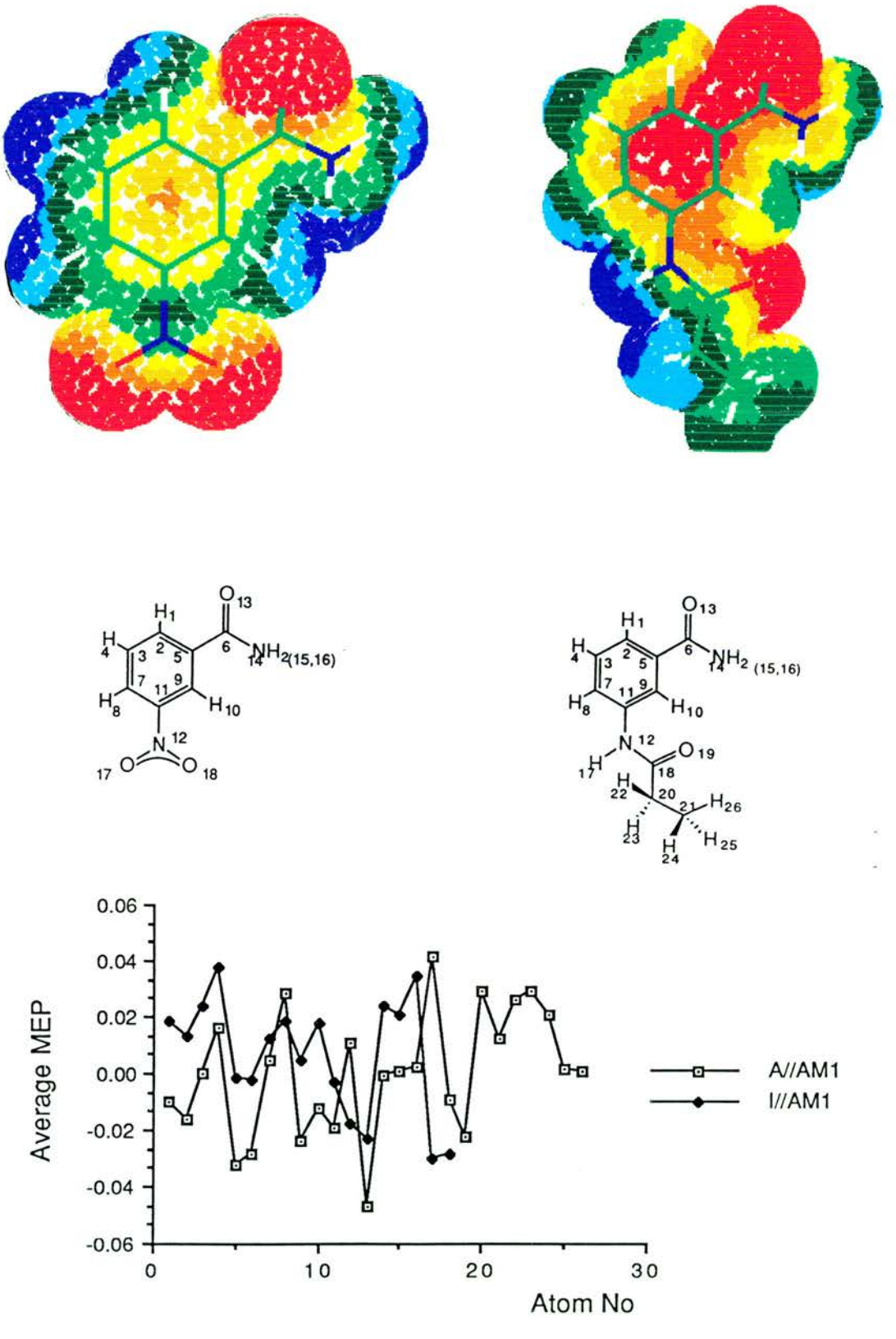
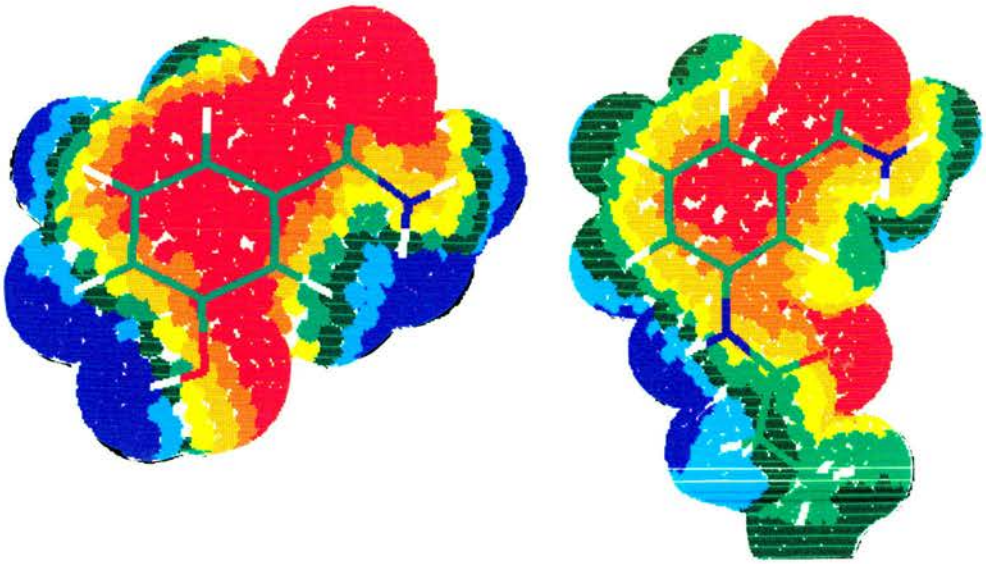
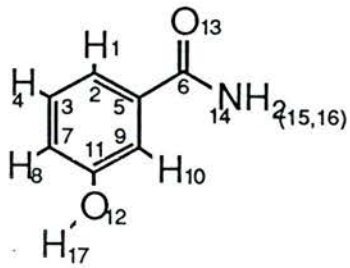
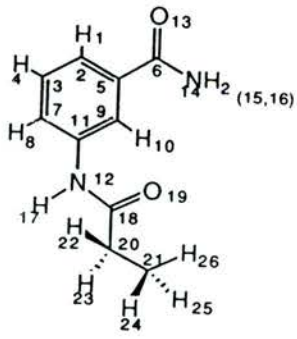


Figure 6.13 Comparison of AM1 point charge MEP's of inhibitor I (top left) and inhibitor A (top right). Average atomic MEP's were calculated on the double van der Waals surface.







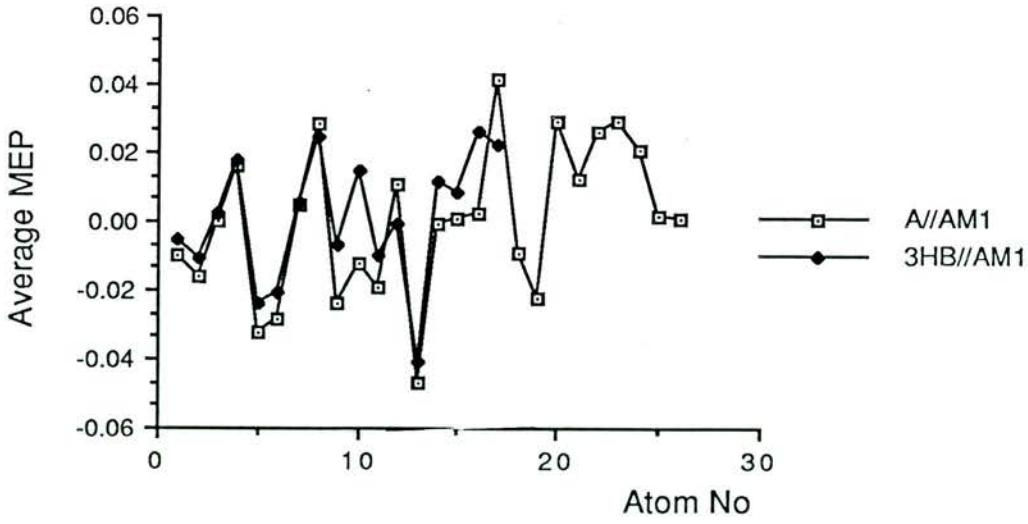


Figure 6.14 Comparison of AM1 point charge MEP's of inhibitor 3HB (top left) and inhibitor A (top right). Average atomic MEP's were calculated on the double van der Waals surface.

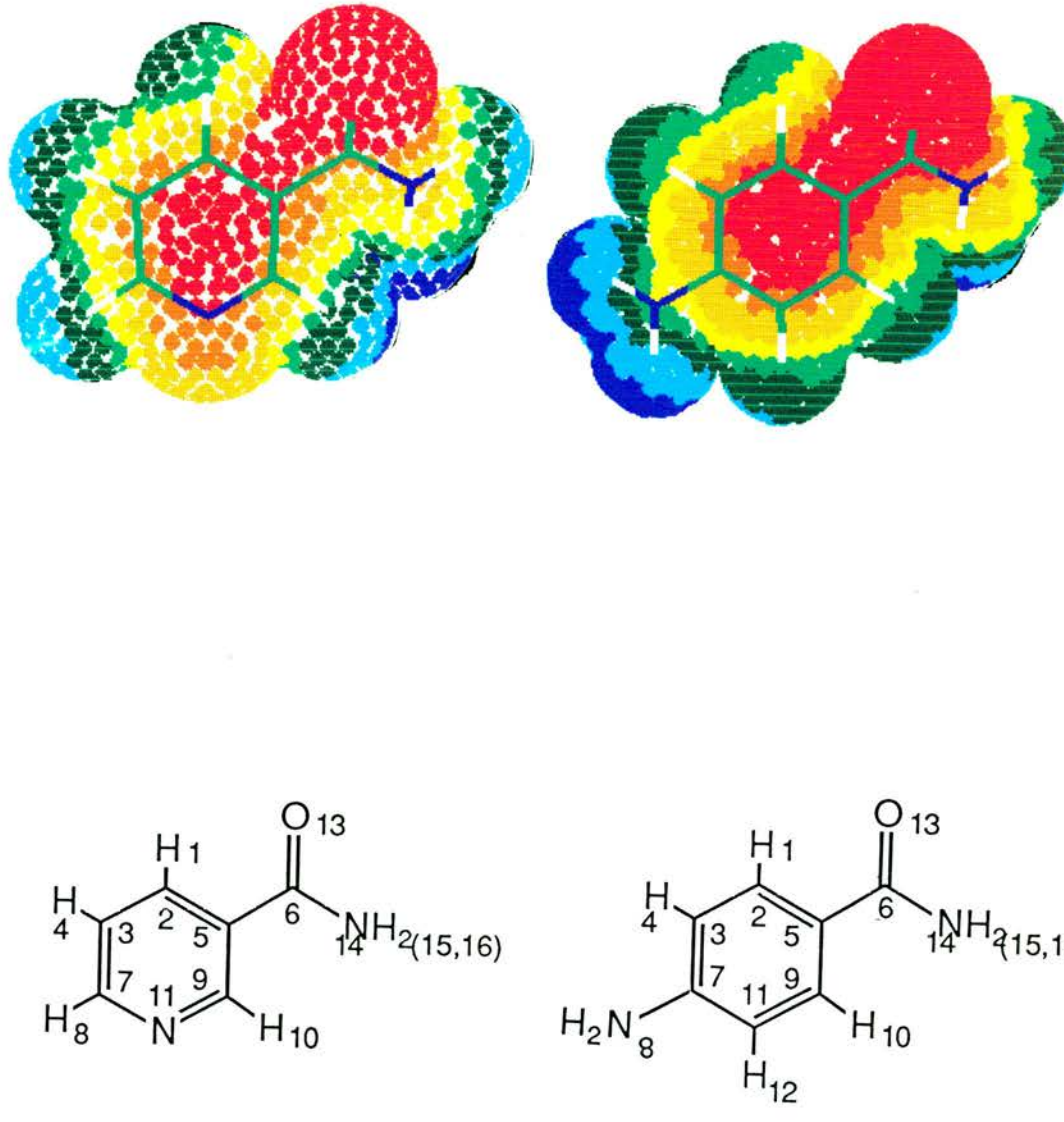


Figure 6.15 AM1 point charge MEP's of nicotinamide (top left) and 4-aminobenzamide (top right).

7 Prediction of new ADPRT inhibitors

This chapter describes the results of our calculations aimed at predicting molecules which may be more powerful inhibitors of ADPRT than those currently available. The strategy behind our calculations was simply to study the MEP's of a range of molecules and compare them with the MEP of the most potent inhibitor of ADPRT, inhibitor A. Although this does not enable us to assign an absolute activity to each of the molecules studied purely from the results of our calculations, we can at least give an indication of which molecules are likely to be active. This is based upon the assumption that molecules which are active have similar electrostatic properties, therefore by comparing the MEP of a molecule with that of inhibitor A, we may be able to decide whether a molecule is likely to be active or not.

As described in chapter 6, the conformation of a molecule is important since not all conformations which a molecule can adopt will be able to interact with the active site of an enzyme. It is desirable, then, to have inhibitors where the active conformation is also the most stable. This will ensure that the greatest proportion of molecules will be present in the active conformation. Some amount of flexibility should be allowed, however, since this will enable molecules to alter their conformation slightly to adapt to the requirements of the binding site.

7.1 Molecules included in this study

The types of molecules which we have studied are shown in Figure 7.1. All of the molecules are benzamide based with various structural modifications made either to introduce rigidity into the molecule or to allow the effects of various functional groups on the MEP to be studied. As with the inhibitors described in chapter 6, all of the molecules

studied here have been fully optimised using the AM1 method and characterised as true minima on their respective PE surfaces by inspection of their force constant matrices. The MEP's were calculated and displayed using the 3D2 program described earlier. The following sections describe the results of our calculations.

7.1.1 Molecules in group (i)

This group of molecules were all derivatives of benzamide with different functional group attached at the meta position in the ring (see Figure 7.1). The following substituents were included in our study: -COH, $-CH{=}CH_2$, $-CO_2H$, $-COCH_3$, $-CH_2OH$, $-CH_2CN$, $-CH_2NH_2$, $-CH_3$, $-C_2H_5$, -CN, -NC, -F,which includes most of the common functional groups.

The MEP maps for the above molecules did not show much promise when compared with the MEP of inhibitor A. In most cases the MEP on the ring and carboxamide parts of each molecule were similar to that on inhibitor A: the MEP on and around the substituent groups, however, did not show much similarity. In most cases the MEP on the substituent group was either of opposite sign to that in the corresponding region on inhibitor A, or it was neutral.

We would therefore expect the molecules in this group to bind at the carboxamide site only with little or no binding at sites B and C shown in Figure 6.7. This means that we would not expect the molecules in this group to be potent ADPRT inhibitors.

7.1.2 Molecules in group (ii)

Figure 7.1 shows the basic structure of molecules in this group which also includes inhibitors A and B described in the previous chapter. The molecules which we have studied from this group all have a

similar structure to inhibitor A ($R=C_2H_5$) with the substituent R including the alkyl groups from $-C_3H_7$ to $-C_7H_{15}$. Our reasoning behind this choice is based upon the small increase in activity which is gained on moving from inhibitor B ($R=CH_3$) to inhibitor A ($R=C_2H_5$). This increase in ADPRT activity does not appear to be a result of changes in the MEP since both inhibitors have almost identical MEP maps (see chapter 6). The change may result from the increase in chain length of the R substituent. The effect of a longer chain might be to increase the lipophilicity of the molecule thereby allowing it easier transport across cell membranes and possibly increased diffusion to the active site on the enzyme.

We have studied the four lowest energy rotamers for each of the molecules mentioned above. We did not consider rotations around the individual $-CH_2$ groups making up the chain, but only considered rotation around the bonds described for inhibitor A in Figure 6.3. The order of stability of the rotamers was found to be the same as that for inhibitors A and B with the TTT rotamer being the most stable.

Figure 7.2 (top) shows the MEP maps for the molecules where $R=C_3H_7$ and $R=C_7H_{15}$ respectively. It is clear that both maps are similar to that of inhibitor A shown earlier. The increase in chain length does not appear to affect the MEP on the ring at all. We should therefore expect this group of molecules to bind to the active site in a similar manner to inhibitor A, and if our suggestion concerning increased lipophilicity is correct they may be more potent inhibitors.

7.1.3 Molecules in group (iii)

Figure 7.1 shows the basic structure of the molecules in this group. The structure is derived from inhibitor A with the formation of a

ring between atoms N12 and C7 (see Figure 6.4 for numbering scheme of inhibitor A). Substituent groups R which we have considered are the simple alkyl groups from CH_3 to C_7H_{15}. These molecules are similar to those described in group **(ii)** except that rotation around the C–N bond (R_3 in Figure 6.2) is eliminated by the presence of the constraining ring. Some degree of flexibility will be present, however, since the ring is not aromatic.

Figure 7.2 (bottom) shows the MEP maps of two molecules in this group with R=C_2H_5 and R=C_4H_9 respectively. It is clear that the MEP's of both molecules compare well with that of inhibitor A. The area of positive EP around N12 in these molecules is larger than that in inhibitor A and extends onto the atoms of the unsaturated ring. The conformation of this ring is twisted upwards. As with molecules in group (ii) an increase in the length of the substituent R has very little effect on the MEP.

Since the electrostatic properties of this group of molecules are similar to those of inhibitor A we would expect them to function as potent ADPRT inhibitors. The addition of the constraining ring might, however, interfere with binding to the active site because of its bulk. Since the ring is flexible it may be able to compensate for this by changing its conformation slightly to improve the fit.

7.1.4 Molecule (iv)

Figure 7.1 shows the structure of molecule **(iv)**. This molecule is similar to inhibitor A with a ring formed between between oxygen **O19** and carbon **C9** (see Figure 6.4 for numbering scheme). This fixes both nitrogen **N12** and oxygen **O19** in similar positions to those which they occupy in the **TTT** rotamer of inhibitor A. Figure 7.3 (top) shows the MEP of molecule (iv).

The MEP has a similar appearance to the MEP of inhibitor A with a positive region around the ring N–H atoms, and a negative region around the ring oxygen. These are in appropriate positions for binding to the active site proposed in Figure 6.7. Th MEP on the benzene ring and carboxamide parts of molecule (iv) is also similar to that of inhibitor A. The orientation of the carboxamide group is different, however, with the angle between the ring and the plane of the carboxamide group increasing from 33° in inhibitor A to 71°. This change is due to the presence of the $-CH_2$ hydrogen atoms, next to the ring oxygen, which lie above and below the plane of the ring and interfere with the carboxamide NH_2 hydrogens. This interference is minimised by rotation of the carboxamide group.

From consideration of its MEP we would expect this molecule to have an activity similar to inhibitors A and B. Since the carboxamide group is essential for activity, then the slight change in its conformation in this molecule might decrease the binding of the molecule to the active site and therefore reduce its activity.

Another factor which might also lead to reduced activity is the lack of an alkyl side chain such as the CH_3 and C_2H_5 groups in inhibitors A and B. It should be possible to introduce a similar group in this molecule on the methylene carbon between the ring nitrogen and oxygen atoms. This may mimic the effect of this group in inhibitors A and B but it should be pointed out that the conformation will be different because of the hydrogen atom which is also present at this position.

7.1.5 Molecule (v)

Molecule (v) is essentially the indole molecule with a carboxamide group attached to position 4 of the benzene ring. This molecule was chosen since it has a similar structure to the purine based inhibitors with both rings co-planar. The MEP of this molecule is shown in Figure 7.3 (bottom). Once again this is similar to the MEP of the more active benzamide based inhibitors with a positive region around the ring -NH atoms. The MEP on the benzene ring and carboxamide parts of the molecule are also consistent with the other benzamide inhibitors.

If we consider the binding of this molecule with the proposed binding site in Figure 6.7, it is likely that binding would be strongest at sites A and B. The MEP on the part of the molecule which would bind to site C is slightly positive which suggests that binding would not occur at this site to any appreciable extent. It is possible, then, that this molecule would be active as an inhibitor of ADPRT but we would not expect it to be as potent an inhibitor as the molecules described above.

7.2 Conclusion

In this chapter we have suggested molecules which may be active as inhibitors of ADPRT and which may even be more potent than those in current use. The molecules which we have included in this study are based upon the results of our calculations described in chapter 6 and were chosen for the various reasons mentioned above. They represent, therefore, only a small fraction of the possible molecules which could have been chosen.

It would be interesting to obtain experimental data regarding the actual activity of the molecules since this would either strengthen our

arguments or dispute them altogether. Either way it will lead to more information about the requirements of the active site.

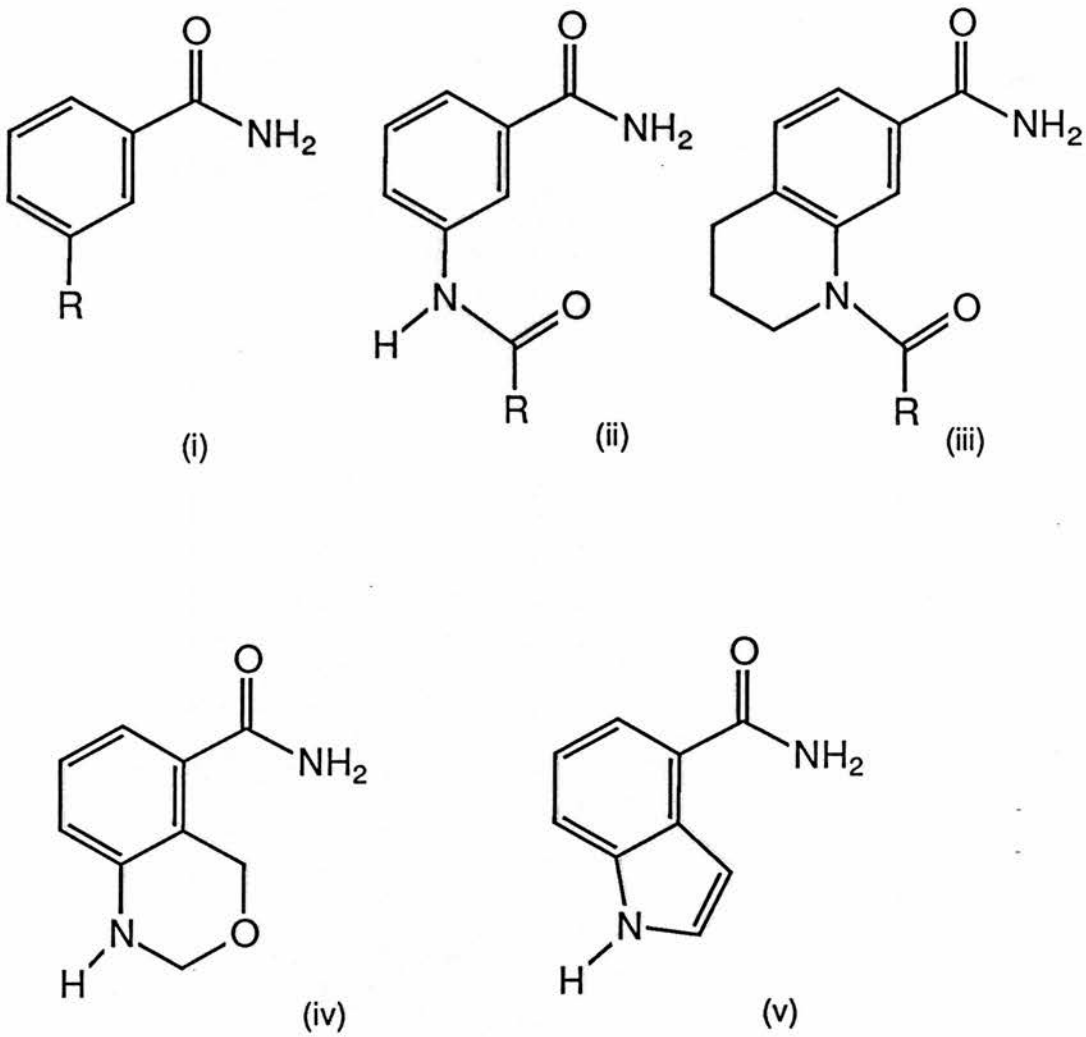


Figure 7.1 Structure of molecules which have been studied in this chapter.

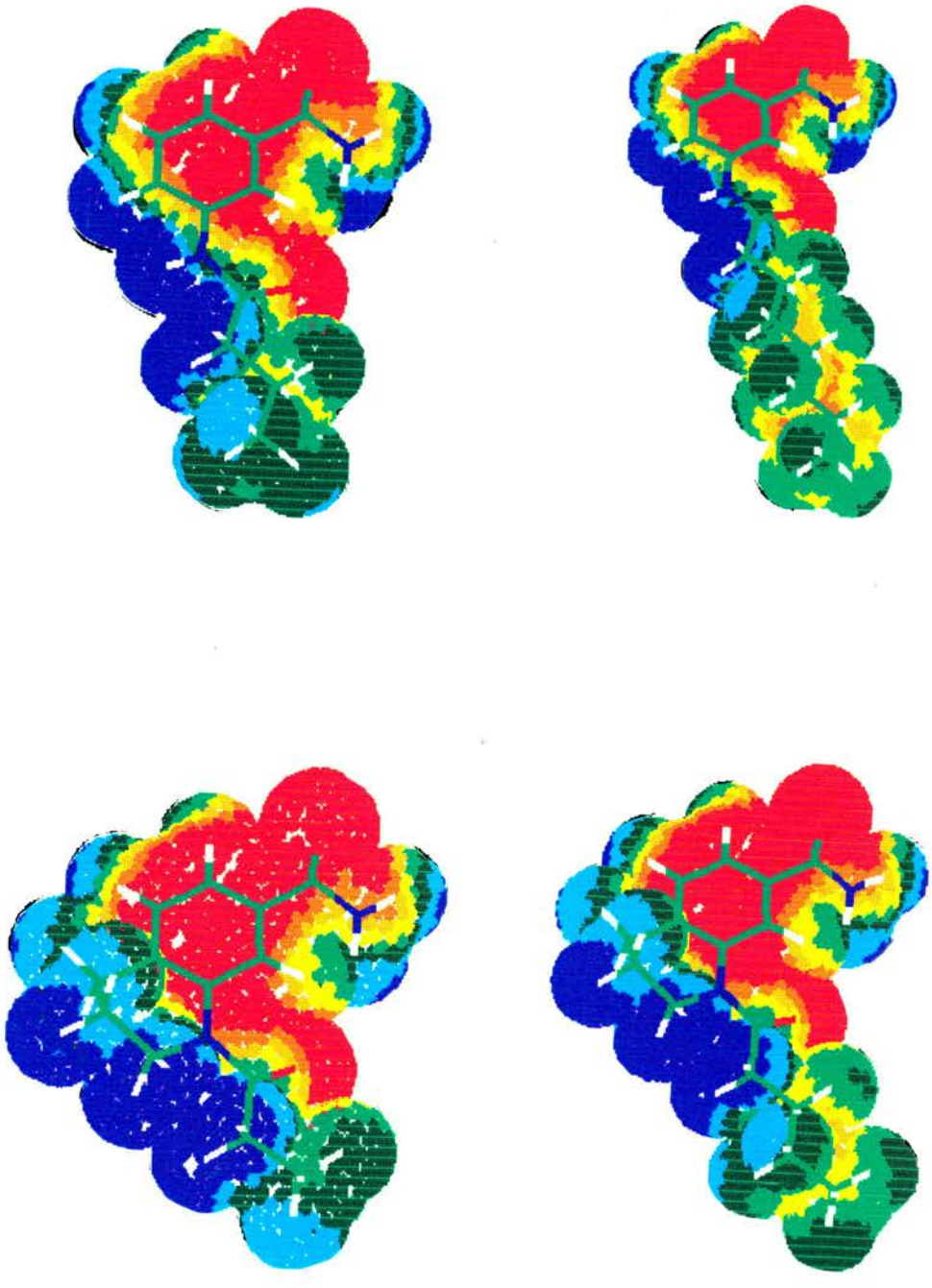


Figure 7.2 AM1 point charge MEP's for molecules in group (i) (top) and group (ii) (bottom).

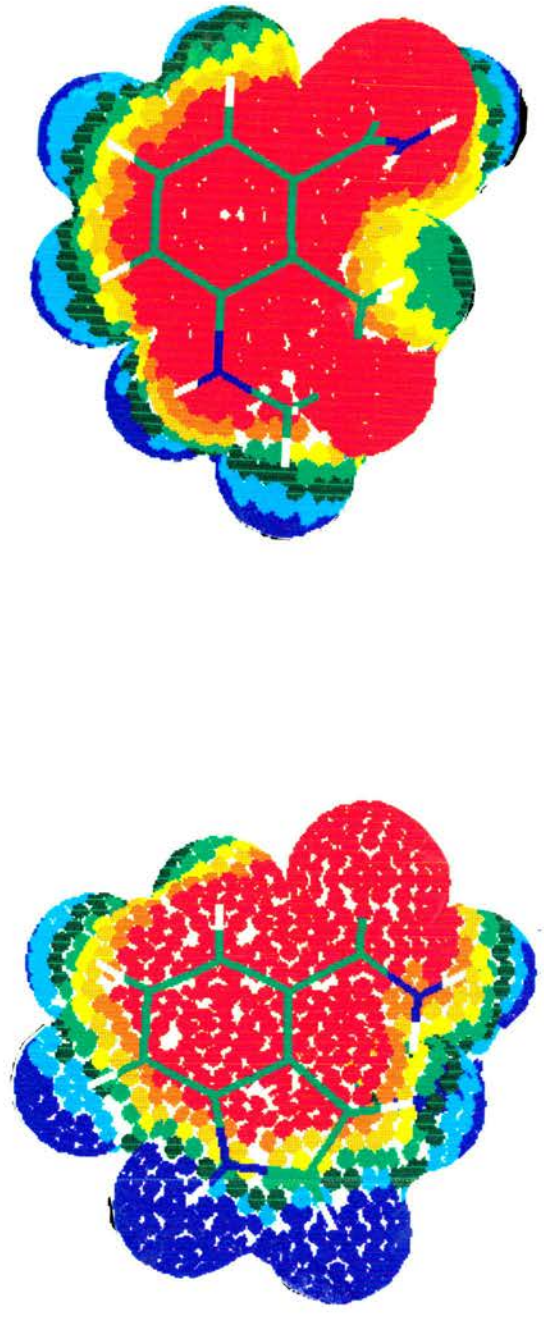


Figure 7.3 AM1 point charge MEP's of molecule (iv) (top) and molecule (v) (bottom).

8 Conclusions

This chapter summarises the results presented in the preceding chapters of Part I and discusses the limitations of the methods used.

8.1 3D2 Program

The 3D2 graphics program described in chapter 4 was designed to allow the Molecular Electrostatic Potential to be calculated for both small (~30 atoms) and large systems (>30 atoms) using a simple point charge model to approximate the MEP. The results obtained using this method are dependent upon the method used to obtain the point charges, and also upon the location of the charges. The method which we have used is to obtain optimised geometries and charges using the AM1 method which are then used as input for the MEP calculation. It is, of course, possible to use any suitable method for obtaining charges and molecular geometries. We have chosen to use the AM1 method since we believe that this method represents the best semi-empirical method currently available for obtaining reliable molecular geometries and associated properties.

Since the program can effectively deal with systems of up to 500 atoms then the limitations imposed upon this method of calculating MEP's will be those related to the method used to obtain the charges and geometries. In our case this limit was 100 atoms since this was the size of the largest system which the MOPAC program, used to perform the AM1 calculations, could handle. The molecules studied in this thesis have all been much smaller than this limit being typically of around 30 atoms: other members of our group, however, use the 3D2 program regularly for systems of ~80-90 atoms using both AM1 and MNDO geometries and charges.

Larger molecules could, of course, be handled by increasing the

dimensions of the MOPAC program although this may lead to problems with performance and resources imposed upon the computer. The major benefit gained by using this approach to calculating MEP's is that it requires much less time than that for a comparible ab-initio calculation and produces similar results.

8.2 Tests of 3D2 and AM1 methods

Chapter 5 describes the results of calculations comparing point charge MEP's, using AM1 geometries and charges, with more accurate ab-initio MEP's for a range of small molecules. The point charge MEP's compare favourably with the ab-initio MEP's, showing similar trends when compared using the average MEP values calculated by the 3D2 program.

The choice of surface used for calculation of the MEP is also important since the simple point charge MEP's are less accurate when calculated close to the molecule. Although the MEP calculated on the van der Waals surface of a molecule is generally adequate for most purposes, more accurate comparisons between MEP's can be made when the MEP is calculated further away from the molecule. The double van der Waals surface is generally adequate for this purpose.

This chapter also shows that the geometry used for location of the charges is less important than the source of the charges. We have found that charges obtained from the AM1 method give the best results.

8.3 ADPRT Inhibitors.

Chapter 6 illustrates the use of MEP's in the study of the enzyme inhibitors of ADPRT. Although there has been much experimental work done on this enzyme and its inhibitors (see chapter 2), we are not aware of any previous calculations on the inhibitors.

In our calculations we have used MEP's, calculated as described above, to study a range of benzamide based inhibitors of ADPRT in an attempt to understand the molecular features necessary for binding to the NAD binding site. We have found similarities in the MEP's of these molecules which suggest that the conformation of an inhibitor is important in determining its activity along with the presence of three areas of negative and positive MEP on the surface of the inhibitor.

In all cases, the presence of a carboxamide group is essential: this is presumably necessary for initial binding or recognition. The potency of the inhibitor can be affected by the nature of a substituent attached in the meta position with respect to the carboxamide group. In the most potent of the inhibitors which we have studied, this group is a substituted amine where the amine substituent determines the potency. The most favourable of the amine substituents seem to be those which are capable of withdrawing electrons from the amine group such as the $COCH_3$ and COC_2H_5 groups. This results in areas of negative and positive MEP on either side of the amine nitrogen which we think are responsible for the strength of binding of the inhibitor.

These conclusions are also consistent with the experimental data on the chemically unrelated purine type molecules, theobromine and theophylline, which also function as potent ADPRT inhibitors. Our study of the MEP on the surface of these molecules shows that they have regions where their MEP is similar to that on the most potent of the benzamide inhibitors. The similarity between the MEP's of these molecules and of the benzamide inhibitors has lead us to propose a tentative binding site which can accommodate both types of inhibitors and which is shown in Figure 6.7. The arrangement of charges around this binding site offers some degree of explanation for the relative potency of the

inhibitors when considered along with their MEP's.

Since the crystal structure of the enzyme has not yet been determined the actual structure of the NAD binding site can only be guessed.

8.4 Predicted new inhibitors.

In chapter 7 we have suggested some molecules which might be potent inhibitors of ADPRT, based upon our study of the known inhibitors in chapter 6. All of the molecules suggested are based upon, and have similar MEP's to, the most potent of the benzamide inhibitors. The main areas of difference are in the introduction of rings to constrain conformations and the increase of chain lengths to increase lipophilicity. Since most of the molecules suggested could be easily synthesised it would be interesting to test their efficiency as ADPRT inhibitors. This would at least furnish us with useful information regarding the validity of our proposals and more optimistically may lead to more potent inhibitors being discovered.

PART II

9 Rotational isomers and vibrational frequencies of nitrosomethanol

9.1 Introduction

Recent studies on the photolysis of matrix isolated methylnitrite (CH_3ONO) by Jacox and Rook [136] and Muller and Huber [137], have shown that irradiation at 365 nm produces a complex between H2CO and HNO. Ab-initio calculations using the 6-31G** basis set [138] have predicted the most favourable structures of this complex to be those in which a hydrogen bond is formed. This agrees with experimental data obtained from comparing vibrational frequencies of the complexed H_2CO and HNO with those of matrix isolated H_2CO and HNO [139].

Nitrosomethanol ($CH_2(OH)(NO)$) may be formed from this complex by selective excitation of either the H_2CO moiety at 345 nm or the HNO moiety at 650 nm. Excitation at 650 nm produces the *cis* rotamer while excitation at 345 nm produces the *trans* rotamer. Interconversion between the *cis* and *trans* rotamers may also be induced by selective excitation. Irradiation of the *trans* rotamer at >645 nm has been found to induce *trans* → *cis* rotamerisation, while irradiation of the *cis* rotamer at 510 nm induces *cis* → *trans* rotamerisaton. The *cis* rotamer is found to have an intramolecular hydrogen bond between the hydroxyl hydrogen and nitroso oxygen atoms. If a plane of symmetry is assumed to exist in the molecule, passing through the H-O-C-N-O group of atoms, then 4 distinct rotamers are possible, each possessing C_S symmetry. The four rotamers are shown in Figure 9.1 with the two letter code used to identify them. The notation is that used by Muller and Huber [140]. The first letter refers to the arrangement of the H-O-C-N atoms within the plane of symmetry while the second letter refers to the arrangement of the O-C-N-O atoms. A *cis* arrangement is denoted by C with *trans* denoted by T. Reference [139] contains an excellent review of all of the

light induced transformations mentioned above.

As far as we are aware, the only previous calculations on nitrosomethanol are those reported by Muller and Huber [140] where they give results of geometry optimisations, using the minimal STO-3G basis set, on the four nitrosomethanol rotamers shown in Figure 9.1. The purpose of their calculations was to obtain the equilibrium geometries of the four C_S rotamers for subsequent use in construction of harmonic force fields to enable them to assign experimental vibrational frequencies to the appropriate structure. As a result they assigned the CC structure to the *cis* rotamer and the CT structure to the *trans* rotamer.

This chapter presents a detailed investigation of the rotational isomers of nitrosomethanol using the comparatively new AM1 method and also by more sophisticated ab-initio calculations using the 6-31G, 6-31G*, and 6-31G** basis sets. Some preliminary calculations were also carried out using the STO-3G, 3-21G and 4-31G basis sets, the results of which are also included. The aims of our calculations were threefold.

Firstly we wished to explore the potential energy surface as a function of the two internal torsion angles (NIC2O3H4 and O3C2NIO5 in Figure 9.1) to allow us to locate the minima and transition states and thus predict the probable structures of the *cis* and *trans* rotamers which are found experimentally.

Calculations of this nature are too time consuming to be undertaken purely by ab-initio methods, being feasible only for small systems, especially where the use of large basis sets is desired. For this reason, we decided to calculate points on the potential surface using one of the more economical semi-empirical methods. Of the various semi-empirical approaches available we chose to use the comparatively

new AM1 [6] method which is based on the NDDO approximation and has been designed to overcome the main weaknesses inherent in MNDO. The reasons for choosing AM1 have been discussed in chapter 5.

The optimised structures of the minima and saddle points on the AM1 surface were used as starting points for the ab-initio calculations using the 6-31G basis set thus eliminating the need for extensive surface searching at this level of theory.

Secondly we were interested in calculating the minimum energy paths for conversion between the *cis* and *trans* rotamers. The prediction of barrier heights, with a reasonable degree of accuracy, requires the use of large polarised basis sets with complete optimisation of the geometry at each conformation. For this reason we chose to use the 6-31G* and 6-31G** basis sets. The optimised 6-31G structures were used as initial guesses at the geometry in the 6-31* and 6-31G** calculations.

Thirdly we wished to calculate the harmonic vibrational frequencies for the *cis* and *trans* rotamers in order to compare them with experimental frequencies [140]. This chapter reports the results of our calculations.

9.2 Method

All the ab-initio calculations reported here were carried out using the Gaussian 82 system of programs [62]. The structures of the minima were fully optimised using the analytical gradient techniques contained in the program. Transition states were located initially using the AM1 method and the geometries refined either by using the Berny algorithm, which is available as an option in Gaussian 82, or by using the algorithm developed by Baker [83]. This algorithm is designed for use with the Gaussian 82 package and is more reliable and faster than the Berny

algorithm for location of transition states. It can also be used to locate minima.

The nature of all stationary points on the AM1 and 6-31G surfaces were characterised by inspection of the force constant matrix. Vibrational frequencies were calculated analytically except in the case of the 6-31G* frequencies which were obtained numerically. The basis sets used were those available internally within Gaussian 82.

The semi-empirical AM1 calculations were carried out using the MOPAC program package [61]. All optimised structures were obtained using the PRECISE option. Transition states were located using the SADDLE option available within the program. In each case a force calculation was performed on the optimised structures to determine the nature of the stationary point.

Throughout this chapter, bond lengths are given in Angstroms and bond angles and torsion angles in degrees.

9.3 Results and Discussion

The results of our calculations are presented in the following sections.

9.3.1 Conformational Energy Surface

The conformational potential energy surface is presented in the form of the contour maps in Figures 9.2 and 9.3. As mentioned previously the points on this surface were calculated using the AM1 method, by systematic variation of the two torsion angles. The molecular geometry was allowed to optimise for each value of the two angles and the heat of formation calculated at each point. The PRECISE option was not used for this part of the calculation to allow the optimisations to proceed faster. A

total of 1369 points on the surface were obtained in this way. The data from the surface calculation was processed by the SURFACE II program [141] to produce the contour maps shown in the figures.

The individual maps in Figure 9.2 were all produced from the same data by appropriate selection of the scales used for the axes. They show the various regions of interest on the surface. Figure 9.3 was produced by combining one of the individual maps (Figure 9.2 top right) to form a larger map which allows the separate features on the surface to be seen in relation to each other. Minima on the surface are indicated by the letters A - D and transition states by the numbers 1 - 10. Optimised geometries have been obtained for all minima and transition states on the AM1 surface and for the corresponding points on the HF/6-31G,6-31G*, and 6-31G** surfaces. Due to lack of resources we were not able to determine the nature of the stationary points on the 6-31G* and 6-31G** surfaces except in the case of the 6-31G* minima which we were able to characterise as such. The transition states numbered 9 and 10 have been characterised by AM1 only and correspond to second order saddle points (see later). We have not attempted to locate corresponding points on the other surfaces.

As mentioned above, we have also carried out some preliminary calculations on the four rotamers in Figure 9.1 using the STO-3G, 3-21G and 4-31G basis sets. These calculations involved complete geometry optimisation and frequency calculation for each rotamer. The energies and nature of the stationary points are given in Table 9.1 with the 6-31G results included for comparison. From this table it is clear that the nature of the four rotamers changes as the size of the basis set is increased, with the larger 4-31G and 6-31G basis sets predicting all four rotamers to be transition states. As we were unable to determine the

nature of the stationary points on the 6-31G* and 6-31G** sufaces we have assumed that these points will be of the same nature as the corresponding points on the 6-31G surface. The results obtained using the 4-31G and 6-31G basis sets suggest that this is not too unreasonable an assumption to make.

By inspection of the contour maps it is clear that there are 4 minima on the surface which may be further divided into two types. Those labelled A and B we have called *cis* minima while those labelled C and D we have called *trans* minima in keeping with the *cis/trans* notation of Muller. The *cis* and *trans* rotamers which are found experimentally are thus predicted to correspond to the *cis* and *trans* minima on the potential surface. Each pair of minima are degenerate. The *cis* minima are connected by the transition states labelled 1 and 7 (Figure 9.2) and the *trans* minima by transition states labelled 2 and 8.

The structure of the *cis* B and *trans* D minima are given in Tables 9.2 and 9.3 respectively. Transition states 1 and 2 correspond to the CC and CT rotamers depicted in Figure 9.1. The members of each pair of minima are related by reflection in the C_s plane of symmetry possesed by the transition state structure. The structure of the *cis* A and *trans* C minima may therefore be obtained from Tables 9.2 and 9.3 by swapping the values given for the C-H1 and C-H2, N-C-H1 and N-C-H2, and O-N-C-H1 and O-N-C-H2 pairs. The two internal torsion angles remain unchanged except for their signs which are reversed. It is also clear that paths exist connecting the *cis* and *trans* minima *via* transition states marked 3, 4, 5 and 6 on the contour maps. As with the minima, these transition states also occur as degenerate pairs. The members of pairs 3, 5 and 4, 6 are related to each other by reflection in the C_s plane of symmetry.

If we now look at the process of interconversion between the *cis* and *trans* rotamers then we can consider the following cases. If the *cis/trans* interconversion does not involve crossing transition states 1 or 2, as in the case of *cis* A ⟷ *trans* C, then it is likely to proceed *via* TS 3 or TS 4. If, on the other hand, the *cis/trans* interconversion does involve crossing TS 1 or TS 2, as in the case of *cis* A ⟶ *trans* D, then it is apparent that a number of paths are possible connected by relevant transition states. A few paths are given below

cis A ⟶ TS4 ⟶ *trans* C ⟶ TS2 ⟶ *trans* D

cis A ⟶ TS1 ⟶ *cis* B ⟶ TS5 ⟶ *trans* D

trans D ⟶ TS2 ⟶ *trans* C ⟶ TS3 ⟶ *cis* A

some of which will be more energetically favourable than others. The most favourable paths for interconversion between *cis* and *trans* rotamers will be discussed later.

The remaining transition states on the AM1 surface are labelled 7 - 10 (Figure 9.2). Corresponding stationary points have been found for 7 & 8 on the 6-31G, 6-31G* and 6-31G** surfaces although, as mentioned earlier, only the 6-31G structures have been characterised as transition states. The points labelled 9 & 10 have been located on the AM1 surface only.

Transition state 7 corresponds to the TC rotamer in Figure 9.1. We have found additional saddle points lying on either side of TS7 on both AM1 and 6-31G surfaces. These saddle points had virtually the same energy as TS7 which suggests that this particular transition state is very broad.

Transition state 8 corresponds to the TT rotamer in Figure 9.1.

This has been characterised as a second order saddle point, on the AM1 surface, by virtue of having two negative eigenvalues of the force constant matrix. Transition states 9 and 10 were also found to be second order saddle points on the AM1 surface. This type of saddle point represents a maxima or hilltop on the potential surface and does not appear on a minimum energy pathway. Paths of lower energy always exist which lead around such maxima (Figure 9.3). On the 6-31G surface TS8 was found to be a simple first order saddle point lying on one of the minimum energy pathways between the *trans* minima. The other pathway crosses saddle point 2. For a more detailed discussion of second order saddle points see references [142] and [143].

9.3.2 Optimised structures of minima

Tables 9.2 and 9.3 give the optimised structures of the *cis* B and *trans* D rotamers respectively. The optimised structures are in good agreement with each other with AM1 bond lengths being slightly longer than the ab-initio bond lengths. The parameters which we expected to vary most were the N-C-O-H and O-N-C-O torsion angles. These angles effectively fix the position of the minima on each surface.

In the case of the *trans* rotamer the N-C-O-H torsion angle calculated by the AM1 method agrees with the 6-31G value, both differing by about 3 degrees from the 6-31G* and 6-31G** values. The AM1 value for the O-N-C-O angle is around 9 degrees less than the ab-initio values which are in good agreement with each other. Inclusion of polarisation functions on the heavy atoms gives rise to a slight increase in the N-C-O-H torsion angle with additional p-functions on the hydrogen atoms having little affect on this angle. In the case of the *cis* rotamer the biggest change was found in the N-C-O-H torsion angle. The value

calculated by AM1(56) is around 12 degrees greater than the 6-31G value (44). The effect of adding polarisation functions is more pronounced than in the case of the *trans* rotamer. This is reflected in the change in value of this angle from 44 degrees (6-31G) to 27 degrees (6-31G*) on addition of d-functions to the heavy atoms. The addition of p-functions to the hydrogen atoms reduces this angle to 24 degrees. No similar effect is observed for the O-C-N-O torsion angle which remains at around 350 degrees.

The effect of decreasing the N-C-O-H torsion angle is to bring the hydroxyl hydrogen closer to the nitrosyl oxygen atom and thus increase the likelyhood of hydrogen bonding. This O----H distance is calculated to be 2.574 Angstroms by the AM1 method and decreases to 2.366 Angstroms using the 6-31G basis set. The addition of d-functions to the heavy atoms reduces this by around 0.3 Angstroms to 2.09 Angstroms. The additional p-functions on hydrogen reduce this distance to 2.08 Angstroms and thus have little effect.

9.3.3 Optimised transition state structures

The optimised structures of transition states 1, 2, 7 and 8 correspond to the CC, CT, TC and TT rotamers (Figure 9.1) respectively.Both AM1 and ab-initio calculations predict the structures to have C_s symmetry.

It is interesting to note that some preliminary calculations which we carried out using the STO-3G and 3-21G basis sets (table 9.1) did not agree as to the nature of the rotamers shown in Figure 9.1. It was only by using the larger split valence basis sets (4-31G or better) that the 4 rotamers were found to be transition states on their respective surfaces, this being in agreement with the AM1 results. The optimised

geometries of the transition states are not included here.

9.3.4 Relative energies and conversion barriers

The energies of all stationary points on their respective surfaces are given in Table 9.4 with energy differences and conversion barriers in Table 9.5.

As mentioned previously interconversion between *cis* and *trans* forms of nitrosomethanol can occur by more than one route. The most probable, and energetically favourable routes, are those which involve conversion *via* one, or at most two, transition states. If we first consider *cis/trans* interconversion involving a single transition state then ab-initio calculations predict both *cis* → *trans* and *trans* → *cis* processes to proceed *via* TS 4 on the relevant surface. The *cis* → *trans* energy barrier lies between 5.8(6-31G) and 6.5 (6-31G**) kcal/mol. The *trans* → *cis* barrier lies between 2.3 (6-31G) and 3.0 (6-31g*) kcal/mol. In contrast the AM1 calculations predict the *cis/trans* interconversions to proceed via TS 3 on the AM1 surface. The AM1 energy barriers are slightly lower than the ab-initio barriers: the *cis* → *trans* barrier is 4.3 kcal/mol and the *trans* → *cis* barrier 1 kcal/mol. Both AM1 and ab-initio calculations predict the tran → *cis* conversion to require about half the energy needed for the *cis* → *trans* conversion.

If we now consider *cis/trans* interconversion involving two transition states then the most favourable paths involve interconversion between *cis* and *trans* rotamer pairs *via* TS 1 and TS 2 respectively. It is interesting to note that the ab-initio calculations predict the conversion between *cis* A and B rotamers to have almost no energy barrier while the AM1 calculations predict the barrier to be 0.8 kcal/mol (table 9.5). Both AM1 and ab-initio calculations predict the barrier to

conversion between the *trans* C and *trans* D rotamers to require between 1.2 (AM1) and 1.4 (6-31G**) kcal/mol.

The energy required for *cis/cis* or trans/*trans* interconversion is much lower than that required for *cis/trans* interconversion. It is likely, therefore, that *cis/trans* interconversion can occur either directly by way of a single transition state or the conversion can also involve *cis/cis* or trans/*trans* interconversion as well as the *cis* ⟷ *trans* step.

9.3.5 Vibrational Frequencies

The 6-31G and 6-31G* harmonic vibrational frequencies for *cis* and *trans* nitrosomethanol are given in Tables 9.6 and 9.7 respectively. The AM1 vibrational frequencies are not reported here. The assignments given in the tables correspond to the modes which contributed most to the particular vibration and give an approximate description of the components involved. The calculated frequencies were matched with the experimental frequencies by comparing the vibrational modes with those given in ref. [140]. In some cases a satisfactory match was not possible in which case the calculated frequency was assigned to the experimental mode which matched best. Luckily this was only necessary at low wavenumbers corresponding to torsional and bending modes. One exception to this was for the CH_2 wagging vibration which the ab-initio calculations predict to occur higher in frequency than the COH bending mode. The experimental assignments give the COH bending vibration higher in frequency than the CH_2 wagging vibration.

Tables 9.6 and 9.7 also include two columns containing scaled 6-31G and 6-31G* frequencies. This practice of scaling frequencies has been suggested by Hehre *et al* [144, 145] who have found that frequencies calculated at the HF/6-31G* level are on average 13% higher than

experimental values. Scaling such frequencies by a factor of 0.87 gives frequencies which are in better accord with experimental results. We have also found this to be true for the *cis* and *trans* nitrosomethanol HF/6-31G* and HF/6-31G frequencies where the average value of $\nu_{expt.}/\nu_{calc.}$ was around 0.87 in both cases. This is clear from the tables where it can be seen that the scaled frequencies are in good agreement with the experimental values.

9.4 Conclusions

We have been encouraged to find such good agreement between the structures of minima and transition states on the AM1 and ab-initio surfaces. The AM1 calculations tend to agree well with the ab-initio calculations using the 6-31G basis set. Agreement between the AM1 calculations and the ab-initio calculations using larger polarised basis sets is not as good although still satisfactory. All methods predict the existence of four minima, and four transition states of C_s symmetry, on the potential energy surface for variation of the two internal torsion angles. The relative stabilities of the minima are found to be consistent with known experimental data. The calculated energy barriers for conversion between minima are again similar with ab-initio values being slightly higher than the AM1 values.

Calculated vibrational frequencies, when appropriately scaled, are in reasonable agreement with experimental values with little difference being found between the HF/6-31G and HF/6-31G* values. It would be interesting to study the effect of electron correlation on this system, particularly with regards to the structure of the *cis* minima and the computed vibrational frequencies. We are not able to do this at present but may be able to carry out such work in the near future.

Table 9.1. Energies of stationary points on the HF/STO-3G, HF/3-21G and HF/4-31G surfaces corresponding to the CC, CT, TC, and TT rotamers in Figure 9.1. The HF/6-31G results are included for comparison. All energies are in atomic units. Structures corresponding to transition states are marked with an asterisk.

Structure	STO-3G	3-21G	4-31G	6-31G
CC rotamer	-240.48543	-242.29968	-243.30581*	-242.55635*
CT rotamer	-240.48084*	-242.29340*	-243.30167*	-243.55243*
TC rotamer	-240.48088	-242.88900	-243.29837*	-243.54914*
TT rotamer	-240.47833*	-242.28870	-243.29735*	-243.54821*

Table 9.2. Optimised structures of the *cis* B minimum on the AM1, HF/6-31G, HF/6-31G* and HF/6-31G** surfaces. Atoms are labelled as in Figure 9.1.

parameter	AM1	6-31G	6-31G*	6-31G**
N1C2	1.483	1.475	1.471	1.470
C2O3	1.414	1.400	1.370	1.369
O3H4	0.966	0.952	0.951	0.947
N1O5	1.156	1.200	1.176	1.177
C2H6	1.127	1.084	1.089	1.090
C2H7	1.128	1.076	1.082	1.084
N1C2O3	117.4	115.0	114.9	114.9
C2O3H4	108.3	114.7	109.1	109.1
C2N1O5	119.3	115.3	113.7	113.8
N1C2H6	106.7	104.8	103.9	103.9
N1C2H7	107.2	106.8	106.1	105.9
O3N1C2H6	124.9	124.7	124.3	124.2
O3N1C2H7	242.7	240.3	238.3	238.0
N1C2O3H4	56.3	44.0	26.5	23.7
O3C2N1O5	349.8	348.7	350.5	351.5

Table 9.3. Optimised structures of the *trans* D minimum on the AM1, HF/6-31G, HF/6-31G* and HF/6-31G** surfaces. Atoms are labelled as in Figure 9.1.

parameter	AM1	6-31G	6-31G*	6-31G**
N1C2	1.498	1.478	1.475	1.475
C2O3	1.412	1.401	1.377	1.375
O3H4	0.966	0.952	0.948	0.944
N1O5	1.155	1.200	1.176	1.176
C2H6	1.128	1.084	1.088	1.089
C2H7	1.128	1.076	1.079	1.081
N1C2O3	112.0	109.6	110.3	110.4
C2O3H4	108.6	114.2	110.2	110.4
C2N1O5	117.5	114.7	113.0	113.0
N1C2H6	106.9	106.4	105.7	105.7
N1C2H7	113.1	110.4	109.5	109.3
O3N1C2H6	121.3	122.0	122.4	122.5
O3N1C2H7	242.4	241.3	240.4	240.3
N1C2O3H4	57.2	57.6	60.9	61.0
O3C2N1O5	148.2	156.7	156.8	157.2

Table 9.4. Heats of formation (kcal/mol) and total energies (atomic units) of stationary points on the AM1,HF/6-31G,HF/6-31G* and HF/6-31G** surfaces.

structure	AM1	6-31G	6-31G*	6-31G**
cis A,B minima	-48.351	-243.556778	-243.681932	-243.691873
trans C,D minima	-45.155	-243.553943	-243.678393	-243.688151
TS 1 (CC)	-47.513	-243.556346	-243.681851	-243.691822
TS 2 (CT)	-43.992	-243.552433	-243.676214	-243.686003
TS 3 & TS 5	-44.073	-243.547528	-243.671891	-243.681562
TS 4 & TS 6	-43.408	-243.549190	-243.674659	-243.684272
TS 7 (TC)	-45.210	-243.549143	-243.673869	-243.683815
TS 8 (TT)	-40.633	-243.548205	-243.673870	-243.683720
TS 9 & TS 10	-41.376	———	———	———

Table 9.5. Energy differences between stationary points on the AM1, HF/6-31G, HF/6-31G* and HF/6-31G** surfaces. All energies are in kcal/mol.

	AM1	6-31G	6-31G*	6-31G**
cis A - TS1 - *cis* B	0.838	0.271	0.051	0.032
cis A - TS7 - *cis* B	3.141	4.791	5.060	5.056
trans C - TS2 - *trans* D	1.163	0.947	1.367	1.347
trans C - TS8 - *trans* D	4.522	3.600	2.838	2.780
cis A - TS3 - *trans* C	4.278	5.805	6.301	6.471
cis A - TS4 - *trans* C	4.943	4.762	4.564	4.770
trans C - TS3 - *cis* A	1.082	4.025	4.080	4.135
trans C - TS4 - *cis* A	1.747	2.982	2.343	2.434

Table 9.6. Harmonic vibrational frequencies of *cis*-nitrosomethanol calculated using the 6-31G and 6-31G* basis sets. All values are in cm-1. Scaled frequencies were obtained by multiplying the calculated values by 0.87 (see text). Assignments were made from analysis of normal coordinates and give an approximate description of the components of each vibration. Experimental frequencies were taken from reference [140].

6-31G	6-31G*	6-31G scaled	6-31G* scaled	Expt.	Assignment
4015	4066	3493	3537	3478	O-H str
3338	3294	2904	2866	2916	CH_2 assym str
3219	3201	2801	2785	2906	CH_2 symm str
1824	1966	1587	1710	1559	N=O str, O=N-C bend
1643	1647	1429	1433	1439	CH_2 sciss
1528	1577	1329	1372	1250	CH_2 wag, C-O-H bend, CH_2 sciss
1439	1443	1252	1255	1355	C-O-H bend, CH_2 rock, CH_2 wag
1316	1349	1145	1174	1203	CH_2 twist, CH_2 rock, C-O-H bend
1192	1272	1037	1107	1131	C-O str,C-O-H bend
1036	1041	901	906	888	CH_2 rock, CH_2 twist, C-O-H bend
915	945	796	822	792	C-N str, N-C-O bend, CH_2 wag, CH_2 sciss
828	847	720	737	755	O=N-C bend, N-C-O bend, C-O str
370	398	322	346	334	O-H tors, C-O-H bend, N-C-O bend, CH_2 wag
243	291	211	253	191	N=O tors, O-H tors, N-C-O bend, O=N-C bend
230	150	200	131	347	O-H tors, C-O-H bend

Table 9.7. Harmonic vibrational frequencies of *trans*-nitrosomethanol calculated using the 6-31G and 6-31G* basis sets. All values are in cm-1. Scaled frequencies were obtained by multiplying the calculated values by 0.87 (see text). Assignments were made from analysis of normal coordinates and give an approximate description of the components of each vibration. Experimental frequencies were taken from reference [140].

6-31G	6-31G*	6-31G scaled	6-31G* scaled	Expt.	Assignment
4018	4097	3496	3564	3638	O-H str
3338	3332	2904	2899	2916	CH_2 assym str
3211	3206	2794	2789	2906	CH_2 symm str
1833	1980	1595	1723	1555	N=O str,O=N-C bend
1641	1634	1428	1422	1432	CH_2 sciss, CH_2 wag
1533	1547	1334	1346	1181	CH_2 wag, CH_2 sciss, C-O-H bend
1468	1503	1277	1308	1352	C-O-H bend, CH_2 rock, CH_2 twist
1320	1327	1148	1155	?	CH_2 twist, C-O-H bend CH_2 rock
1196	1258	1041	1095	1107	C-O str, C-O-H bend, CH_2 twist
1061	1101	923	958	865	CH_2 rock, CH_2 twist, C-O-H bend, N-C-O bend
982	991	854	862	847	C-N str, N-C-O bend, CH_2 twist, CH_2 rock
604	617	526	537	542	O=N-C bend, N-C-O bend
450	461	392	401	387	O-H tors, C-O-H bend, N-C-O bend, CH_2 wag
282	294	245	256	220	O-H tors
95	89	83	77	?	N=O tors

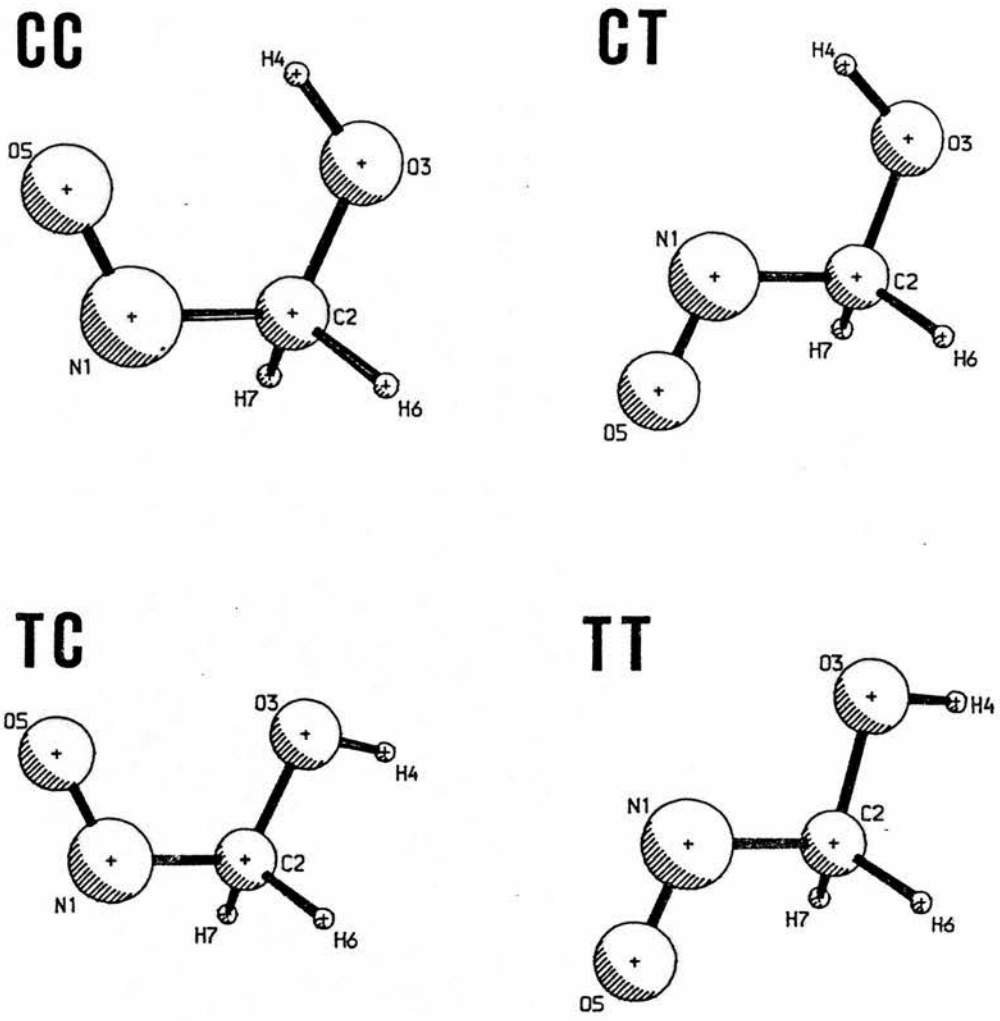


Figure 9.1 The four rotamers of nitrosomethanol which have Cs symmetry. The CC and TT rotamers correspond to transition states 1 and 2 while TC and TT rotamers correspond to transition states 7 and 8 respectively on the AM1 and HF/6-31G surfaces.

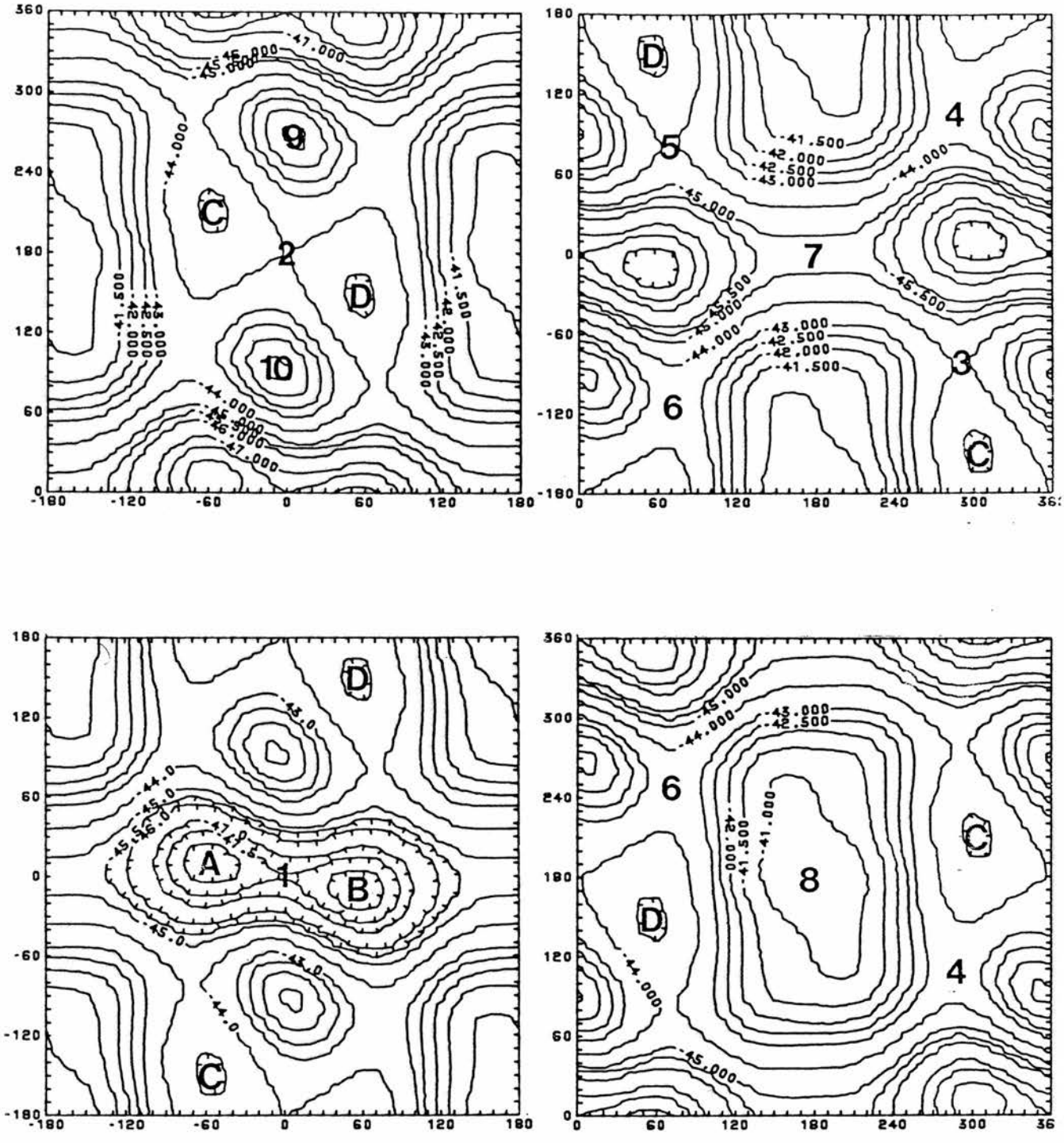


Figure 9.2 The AM1 potential energy surface of nitrosomethanol as a function of the two internal torsion angles. The horizontal axes represent variation of the N-C-O-H torsion angle while the vertical axes represent variation of the O=N-C-O torsion angle. Contour levels are in kcal/mol. All maps contain the same information with scales suitable chosen to show features of interest. Points which are labelled are refered to in the text.

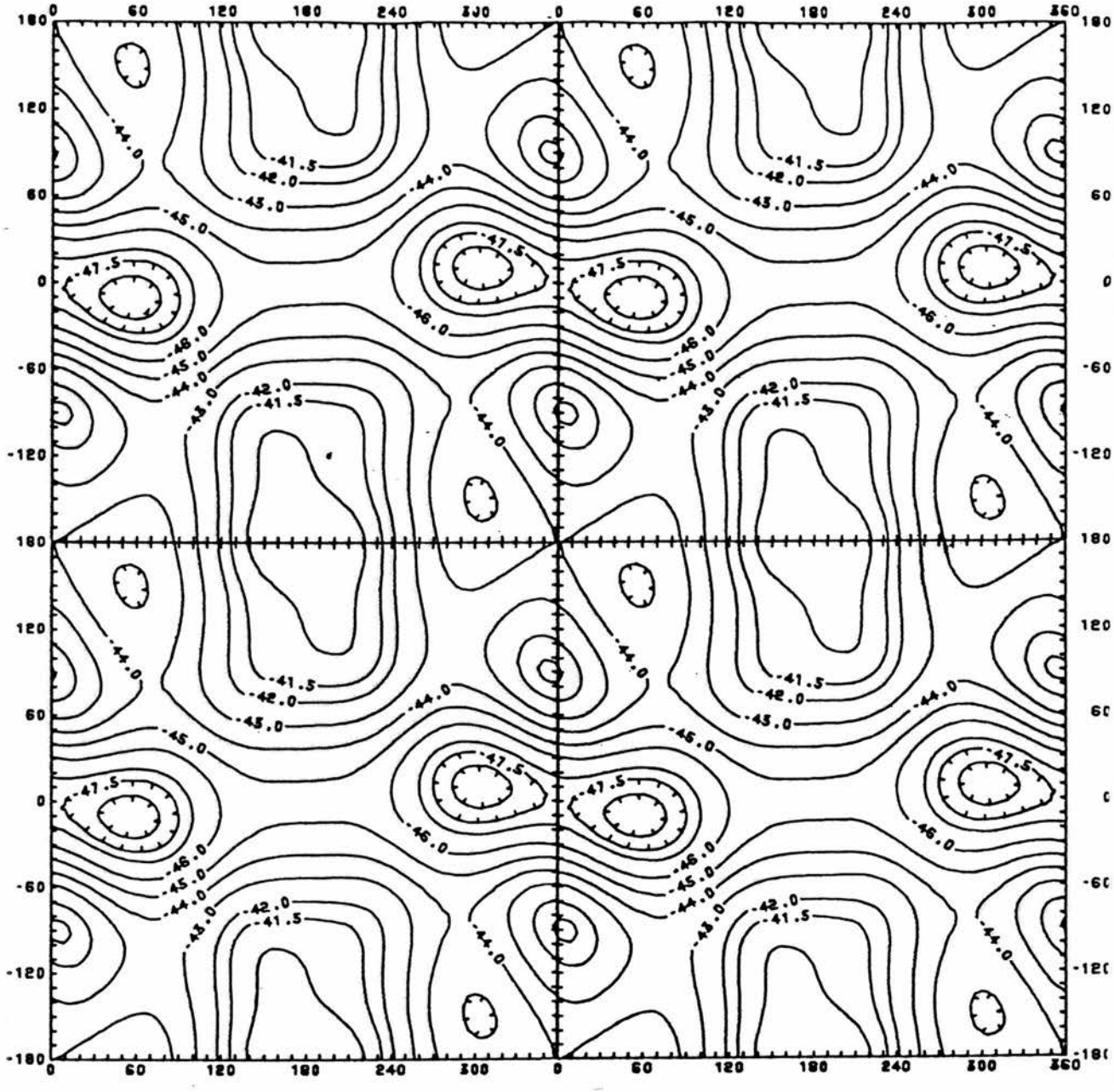


Figure 9.3 Expanded contour map of the AM1 potential energy surface. The horizontal axis represents variation of the N-C-O-H torsion angle and the vertical axis variation of the O=N-C-O torsion angle. Contour levels are in kcal/mol.

10 Development of geminal program

This chapter describes the reasons for our development of a semi-empirical method for calculations on large molecules using a geminal type wavefunction and reports the progress which we have made to date.

10.1 Introduction

Since most chemists think of molecules in terms of two-electron bonds, atomic cores and lone pairs etc, it would be desirable to have a suitable method whereby the resulting molecular wavefunction could be easily interpreted in these terms. One of the shortcomings of the Hartree Fock method is that it produces a molecular wavefunction which is delocalized over the whole molecule and is therefore difficult to interpret, especially where a large number of basis functions are, used.

One solution to this problem is to subject the canonical SCF orbitals to a localization procedure such as those developed by Boys *et al* [146-148]. Such procedures effectively relocalize the canonical SCF orbitals in regions of molecular space which correspond to the classical concepts of atomic core, bonds and lone pairs which is closer to the normal way of thinking about molecular structures. It should be noted, however, that the localized and delocalized wavefunctions are entirely equivalent and lead to the same molecular properties etc.

A more satisfactory solution to the problem however is to recognise the existence of atomic cores, lone pairs and bond pairs from the start, and to construct a molecular wavefunction in terms of an antisymmetrized product of localized two-electron functions, instead of the one-electron functions used in the Hartree Fock type wavefunctions. Such two-electron functions are normally termed "geminals" and their resulting wavefunctions "geminal type wavefunctions". It is worth

mentioning, at this point, that the geminal approach described above is part of a more general theory, where not only distinct pairs of electrons can be considered, but also many-electron "groups". Molecules can then be described in terms of antisymmetrized products of "group functions".

The basic theory behind such methods was developed in the 1950's and can be found in the original papers by Hurley, Lennard-Jones and Pople [149], and Parks and Parr [150], and in the book by Hurley [151].

10.2 Geminal wavefunctions

This section provides a brief description of the form of the geminal type wavefunctions. In order to simplify calculations, and to make the formulas involved more manageable, it is usual to impose the condition of strong orthogonality between the different geminals. This approach is termed the Antisymmetrized Product of Strongly orthogonal Geminals (APSG) [149], and is the method which we have used in this work, and which is described in the remainder of this chapter.

For a closed shell system of 2N electrons the APSG wavefunction can be represented as follows:

$$\Phi^{APSG} = M^{[N]} \, A^{[2N]} \left[\sum_{K=1}^{N} \Lambda (x_{2K-1}, x_{2K}) \right] \qquad (1)$$

where the wavefunction is normalized by the constant $M^{[N]}$ and antisymmetrized by the operator $A^{[2N]}$. The normalization constant $M^{[N]}$ is defined by the equation

$$\langle \Phi^{APSG} | \Phi^{APSG} \rangle = 1 \qquad (2)$$

In equation (1), the x_i represents the combined space and spin coordinates of electron i. The geminals Λ_K are products of an orbital part and a singlet coupled spin function

$$\Lambda_K(x_1,x_2) = \Lambda(r_1,r_2)\theta(s_1,s_2) \qquad (3)$$

The geminals used in the APSG approximation can be defined as follows, in terms of orbitals:

$$\Lambda_K(r_1,r_2) = \sum_{K=1}^{n_K} a_k^K \psi_k^K(r_1)\psi_k^K(r_2) \qquad (4)$$

where the ψ_k represent the orbitals,the a_k represent the expansion coefficients and the n_k represents the number of natural orbitals associated with geminal K. Normalization of the geminals is also performed according to the equation:

$$\langle\Lambda_K|\Lambda_K\rangle = \sum_{k=1}^{n_k} |a_k^K|^2 = 1 \qquad (5)$$

The spin-independent Hamiltonian used in the APSG method has the form:

$$H = \sum_{i=1}^{2N} h(i) + \tfrac{1}{2} \sum_{i\neq j}^{2N} g(i,j) \qquad (6)$$

with the APSG energy then given by the expression:

$$E^{APSG} = \langle\Phi^{APSG}|H|\Phi^{APSG}\rangle$$

$$= \sum_{K=1}^{N} \varepsilon_K + \sum_{K<L} (J_{K,L} - K_{K,L}) \qquad (7)$$

In equation (7) the ε_K can be considered as the energy of an electron pair K in the field of all nuclei, and the terms $J_{K,L}$ and $K_{K,L}$ as the "coulomb" and "exchange" interactions between the pairs of electrons K and L respectively. The total APSG energy is then given by:

$$E_{total}^{APSG} = E^{APSG} + E_{nuc} \qquad (8)$$

where E_{nuc} is the nuclear electrostatic energy. The APSG wavefunction (1) can then be optimized with respect to the orbitals Ψ_k and expansion coefficients a_k in equation (4).

10.3 Advantages of a Geminal type wavefunction

The advantages of a geminal type wavefunction over the standard RHF wavefunction are more than just easier interpretation of the wavefunction itself. The more important benefits of the geminal wavefunctions are listed below:

(a) The molecular wavefunction can be described in terms of chemically significant entities, and the interactions between them.

(b) The wavefunction for a molecule is built up on a geminal-by-geminal basis and therefore can be determined for any size of molecule in theory. The wavefunction of each individual geminal can also be refined indefinitely, by addition of further orbitals to include correlation effects within

the geminal.

(c) A good starting approximation for each geminal can always be obtained from a suitable RHF SCF calculation or from previous calculations.

(d) Finally, because of the nature in which the geminal type wavefunctions are determined it should be possible to develop a semi-empirical geminal method, from a non-empirical method, by the use of approximations and simplications similar to those used in the AM1 and MNDO methods.

The last point mentioned above **(d)** is what we hope to achieve in the near future. Our aim is to develop a semi-empirical geminal program, based upon the **APSG** method, which will be applicable to both large and small systems alike. There are many examples of the applications of non-empirical geminal calculations in the literature [152-160], all of which have been on small molecules, and which are mainly concerned with electron correlation: since it is possible to recover a large fraction of the available correlation energy by using geminal type wavefunctions.

Klessinger and McWeeny [161] have outlined ways in which a semi-empirical geminal method could be developed as early as 1965, but as yet we are not aware of any available program(s) which can carry out such calculations. This, and the fact that such a method should be useful in the study of large molecules, prompted us to attempt to develop a suitable program for performimg semi-empirical geminal calculations. The remaining sections in this chapter describe the results of our efforts.

10.4 Program development

With the above considerations in mind, we decided that a suitable starting point for such work would be to take a suitable, non-empirical geminal based program, based upon the APSG method, to which we could make systematic approximations and simplifications. This approach has the advantage in that it is always possible to check the effects of any changes made by comparing the results of the semi-empirical version with the corresponding calculation performed using the non-empirical program.

We were fortunate, in this respect, in being able to obtain a copy of a suitable program from Dr Inge Röeggen. Röeggens' program (**EXGEM**) was developed for describing the electron correlation in molecules using an extended geminal method. Details of this method can be found in his original papers [155,160,162,163]. Although this program is primarily designed for extensive correlated geminal calculations , it is based upon an APSG wavefunction as the zero-order approximation to the Schrodinger equation. We were able, then, to use the APSG part of this program for the basis of our own system.

10.5 The EXGEM program

This section gives a brief description of the EXGEM program system mentioned above, which consists of five separate program units. A description of each unit and the function which it performs is given below.

MOLPROG - this program is essentially the integral program MOLECULE [164] which generates the necessary one- and two-electron integrals.

INTTAB - this program generates the Beebe-Linderberg integral tables [165,166] for use with the RHF and EXGEM program units.

RHF - this program performs a restricted Hartree-Fock calculation based upon the Beebe-Linderberg integral tables. The purpose of this program is to generate an initial RHF wavefunction which can then be used to construct the initial geminal based wavefunction.

LOC - performs Boys' localization of selected orbitals.

EXGEM - this is the program which performs the geminal calculations and the various correlation calculations which can be requested.

Communication between the individual program units is through a series of disk-files which contain the necessary integrals, wavefunctions, density matrices etc. A data file is also required for each individual program unit so that a single run of the EXGEM program requires the preparation of 5 separate data files.

The next section describes the progress we have made in the development of our semi-empirical geminal program **GEMINI** from the EXGEM system described above.

10.6 GEMINI program

As described above we have used the APSG part of the EXGEM program as the basis of our geminal program. One problem with the EXGEM system was the large number of disk-files (up to 40) which it required for storage of intermediate results and inter-program communication. This severely limited the execution speed of the program since most of its time was spent writing to or reading from these file, thus creating a large I/O bottleneck.

We have chosen to use a program structure similar to that of the GAUSSIAN programs [62, 63], instead of the EXGEM structure, with the GEMINI program being made up from a series of separate links. In this way it was possible to design a more efficient and flexible program structure which is easy to maintain and modify. The problem with disk-files in the EXGEM system has been overcome by the use of a few large files in the GEMINI system which hold most of the control and intermediate data which the program requires.

The basic structure of the GEMINI program, shown in Figure 10.1, is a series of links which execute according to a route specified in the input file, in a similar manner to GAUSSIAN 82. Each link runs as a separate process, using the VAX run time library chaining procedures to load and unload each link. All input to the program is contained in a single data file, using simple keywords to tell the program which functions to perform, in a similar way to the MOPAC program [61]. The order in which the links are called can be specified by including a

route card in the input file thus allowing calculations to be started at any point providing that the necessary files are available.

At present, the program is still non-empirical in nature, since we have not been able to start work on introduction of the approximations necessary to produce a semi-empirical version of the program. The majority of the work already carried out has been aimed at producing a program system which is efficient and easy to use. We have at least succeeded in this respect.

10.7 Test calculations.

An example of a test calculation using the GEMINI program can be found in appendix B. This shows the output from a simple geminal calculation on the CH_2 molecule using the standard 3-21G basis set which illustrates the basic steps involved. The initial wavefunction for this calculation was obtained from a closed shell RHF calculation. The resulting RHF M.O.'s were then localized and re-ordered as follows:

(HOMO + LUMO), (HOMO-1, LUMO +1), ...

which gave the arrangement shown below.

				H		L								
original RHF vectors	1	2	3	4	|	5	6	7	8	9	10	11	12	13
re-ordered vectors	1	8	2	7		3	6	4	5	9	10	11	12	13

The four geminals for CH_2 were obtained by pairing RHF vectors (1,8),

(2,7), (3,6) and (4,5), the remaining vectors 9-13 being included in the correlation space.

The results of the geminal, calculation are shown near the end of Appendix B, where the wavefunctions of the distint electron pairs are given. In this case electron pair 1 corresponds to the carbon 1s core electron, electron pairs 2 and 3 correspond to C-H bond geminals and electron pair 4 corresponds to a carbon lone pair.

The APSG energy for this system is slightly lower than the RHF energy, as expected, since the effect of using two-electron functions increases the flexibility of the wavefunction and hence gives a lower energy. The geminal corrections shown in this appendix result from the intrapair correlation energy associated with each pair of electrons: in this case only single pair corrections have been calculated.

We have used this program to perform geminal calculations with promising results mainly on small systems, the largest being C_2H_6. The CPU time taken for CH_2 test shown in appendix B was around 15 minutes, the time for the C_2H_6 calculation around 5 hours. Therefore, the use of approximations is necessary if calculations or molecules of significant size are to be feasible.

10.8 Recommendations for future work

We intend to continue the work started on this program so that we have a semi-empirical version ready during the next year. The most important area, which we have not looked at so far, is in the nature of the approximations which have to be made to the integral evaluation and geminal parts of the program.

The question of the choice of wave function used to start the geminal calculation is also of importance. Although it is possible to use the RHF wavefunction generated by the program as the initial guess, we hope to study the effects of using wavefunctions from other sources for this purpose. It should also be possible to use previously calculated geminal wavefunctions for this purpose. We also intend to provide suitable routines for geometry optimisation since the current program will only carry out single point calculations.

Figure 10.1 Structure of the GEMINI program

Link No.	Function
0	Sets up all files and logical names required by the program and performs all parsing and checking of the input data.
1	This link reads in the geometry and sets up the appropriate data structures necessary for the calculation.
2	Checks geometry and computes the cartesian coordinates of the molecule.
201	Checks coordinates for any symmetry elements and re-orientates molecules accordingly.
3	This link reads the basis set information and sets up the files required by the integral links.
301, 302	These links calculate the necessary one and two-electron integrals. Link 302 is a faster in-core version of Link 301.
304	This routine calculates the integral tables.
4	Not used at present but reserved for initial guess routines.
5	This link performs a closed shell RHF calculation to obtain suitable starting vectors for the geminal link.
501	Performs Boys' localization.
6	This is the main geminal link which calculates the APSG wavefunction.

11. Appendix A

This appendix contains the AM1 optimised geometries of the ADPRT inhibitors refered to in chapter 6. All bond lengths are in Angstroms, bond angles and dihedral angles in degrees. The average MEP values (Avepot) quoted are in atomic units.

Inhibitor A (3-propionamidobenzamide)

AM1 optimised geometry, atomic charges and average MEP values for the TTT conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.159	-0.010202
C2	1.102	0.0	0.0	1	0	0	-0.103	-0.016344
C3	1.393	121.0	0.0	2	1	0	-0.113	0.000070
H4	1.101	119.8	0.7	3	2	1	0.139	0.016185
C5	1.398	119.5	179.3	2	3	1	-0.094	-0.032583
C6	1.493	117.9	-0.3	5	2	1	0.343	-0.028538
C7	1.390	120.6	180.0	3	2	4	-0.164	0.004530
H8	1.101	119.2	0.2	7	3	4	0.131	0.028585
C9	1.396	120.8	178.8	5	2	6	-0.129	-0.023877
H10	1.103	120.4	0.7	9	5	6	0.170	-0.012306
C11	1.417	120.4	-179.7	7	3	8	0.068	-0.019361
N12	1.405	122.6	-0.4	11	9	10	-0.332	0.010635
O13	1.250	121.7	33.2	6	5	2	-0.375	-0.046562
N14	1.372	118.4	-178.4	6	5	13	-0.430	-0.000937
H15	0.989	119.4	-179.0	14	6	5	0.230	0.000720
H16	0.987	121.3	176.5	14	6	15	0.229	0.002297
H17	0.995	115.5	-11.4	12	11	7	0.229	0.041283
C18	1.391	127.0	173.6	12	11	17	0.309	-0.009267
O19	1.244	122.5	2.2	18	12	11	-0.353	-0.022290
C20	1.521	115.5	179.6	18	12	19	-0.180	0.029432
C21	1.507	111.1	-179.6	20	18	12	-0.205	0.012346
H22	1.123	109.0	121.2	20	18	21	0.105	0.025866
H23	1.122	109.1	-121.2	20	18	21	0.102	0.029205
H24	1.116	109.9	180.0	21	20	18	0.079	0.021044
H25	1.117	110.5	120.4	21	20	24	0.093	0.001331
H26	1.117	110.5	-120.3	21	20	24	0.093	0.000479

Inhibitor B (3-acetamidobenzamide)

AM1 optimised geometry, atomic charges and average MEP values for the TTT conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.159	-0.010076
C2	1.101	0.0	0.0	1	0	0	-0.103	-0.016035
C3	1.393	121.0	0.0	2	1	0	-0.113	0.000468
H4	1.101	119.8	0.7	3	2	1	0.139	0.016430
C5	1.398	119.4	179.3	2	3	1	-0.094	-0.032036
C6	1.493	117.9	-0.4	5	2	1	0.343	-0.028647
C7	1.390	120.7	180.1	3	2	4	-0.164	0.005300
H8	1.101	119.2	0.3	7	3	4	0.131	0.029206
C9	1.396	120.8	178.9	5	2	6	-0.129	-0.023156
H10	1.103	120.3	0.8	9	5	6	0.171	-0.011976
C11	1.416	120.3	-179.7	7	3	8	0.067	-0.018650
N12	1.405	122.5	-0.4	11	9	10	-0.330	0.012336
O13	1.250	121.6	33.0	6	5	2	-0.375	-0.046398
N14	1.372	118.4	-178.4	6	5	13	-0.430	-0.000860
H15	0.989	119.4	-178.9	14	6	5	0.230	0.000668
H16	0.987	121.3	176.1	14	6	15	0.229	0.002315
H17	0.995	115.6	-11.1	12	11	7	0.230	0.042499
C18	1.390	127.0	173.4	12	11	17	0.311	-0.008190
O19	1.244	122.9	2.5	18	12	11	-0.355	-0.021961
C20	1.512	116.0	179.6	18	12	19	-0.246	0.023451
H21	1.118	108.5	-179.6	20	18	12	0.123	0.011657
H22	1.117	110.6	119.4	20	18	21	0.104	0.027789
H23	1.116	110.6	-119.4	20	18	21	0.102	0.030960

Inhibitor C (3-N-methylaminobenzamide)

AM1 optimised geometry, atomic charges and average MEP values for the TT conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.155	-0.011130
C2	1.101	0.0	0.0	1	0	0	-0.121	-0.017682
C3	1.394	121.0	0.0	2	1	0	-0.102	-0.004203
H4	1.101	119.6	0.8	3	2	1	0.137	0.011843
C5	1.397	119.3	179.4	2	3	1	-0.084	-0.031681
C6	1.492	118.0	-0.4	5	2	1	0.343	-0.025950
C7	1.388	120.8	180.3	3	2	4	-0.170	-0.002750
H8	1.101	119.4	0.1	7	3	4	0.134	0.017441
C9	1.395	120.8	-181.0	5	2	6	-0.166	-0.016558
H10	1.100	119.8	0.7	9	5	6	0.136	0.020854
C11	1.418	120.5	-179.2	7	3	8	0.058	-0.018160
N12	1.404	122.0	3.2	11	9	10	-0.289	-0.003209
O13	1.248	122.1	38.2	6	5	2	-0.368	-0.046679
N14	1.375	118.2	-179.2	6	5	13	-0.436	0.009002
H15	0.989	119.5	178.1	14	6	5	0.230	0.005577
H16	0.986	121.1	-175.1	14	6	15	0.220	0.026850
H17	0.999	113.0	25.4	12	11	7	0.194	0.022921
C18	1.433	119.2	137.8	12	11	17	-0.101	0.018872
H19	1.123	109.3	68.0	18	12	11	0.089	0.012178
H20	1.122	108.7	118.2	18	12	19	0.088	0.022545
H21	1.125	113.2	-120.9	18	12	19	0.053	0.018585

Inhibitor D (3-methoxybenzamide)

AM1 optimised geometry, atomic charges and average MEP values for the TT conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.155	-0.008893
C2	1.083	0.0	0.0	1	0	0	-0.111	-0.014899
C3	1.393	120.0	0.0	2	1	0	-0.105	-0.000421
H4	1.101	119.8	0.8	3	2	1	0.139	0.015644
C5	1.401	120.0	179.3	2	3	1	-0.087	-0.028657
C6	1.492	118.1	-0.4	5	2	1	0.344	-0.024415
C7	1.395	120.7	-179.7	3	2	4	-0.183	0.001146
H8	1.099	119.9	0.3	7	3	4	0.140	0.024221
C9	1.394	120.3	178.9	5	2	6	-0.134	-0.012019
H10	1.100	122.1	0.6	9	5	6	0.150	0.009650
C11	1.400	119.0	-179.8	7	3	8	0.071	-0.018493
O12	1.382	114.5	-1.5	11	9	10	-0.212	-0.008509
O13	1.249	121.9	34.3	6	5	2	-0.369	-0.044220
N14	1.374	118.2	-179.2	6	5	13	-0.434	0.007521
H15	0.989	119.4	179.0	14	6	5	0.231	0.005807
H16	0.986	121.2	-177.3	14	6	15	0.223	0.022032
C17	1.423	116.1	-3.3	12	11	7	-0.076	0.014380
H18	1.117	110.6	63.5	17	12	11	0.073	0.016845
H19	1.120	103.4	118.4	17	12	18	0.108	0.016208
H20	1.117	110.8	-123.0	17	12	18	0.075	0.012451

Inhibitor E (benzamide)

AM1 optimised geometry, atomic charges and average MEP values for the conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.133	-0.008113
C2	1.102	0.0	0.0	1	0	0	-0.104	-0.013624
C3	1.393	120.8	0.0	2	1	0	-0.140	-0.000172
H4	1.100	119.8	0.8	3	2	1	0.135	0.014760
C5	1.401	119.9	179.3	2	3	1	-0.121	-0.026641
C6	1.490	118.0	-0.3	5	2	1	0.344	-0.023232
C7	1.395	120.2	-179.7	3	2	4	-0.108	0.003979
H8	1.100	120.0	0.2	7	3	4	0.135	0.019134
C9	1.398	120.0	178.7	5	2	6	-0.078	0.000702
H10	1.100	120.8	0.5	9	5	6	0.156	0.022839
C11	1.395	120.0	-179.7	7	3	8	-0.138	0.005347
H12	1.100	119.8	-0.9	11	9	10	0.138	0.020236
O13	1.249	121.9	33.6	6	5	2	-0.371	-0.043578
N14	1.374	118.3	-179.5	6	5	13	-0.435	0.009777
H15	0.989	119.3	177.8	14	6	5	0.231	0.026926
H16	0.986	121.2	-173.4	14	6	15	0.221	0.006448

Inhibitor F (3-aminobenzamide).

AM1 optimised geometry, atomic charges and average MEP values for the Tβ conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.155	-0.010281
C2	1.101	0.0	0.0	1	0	0	-0.122	-0.016305
C3	1.394	120.9	0.0	2	1	0	-0.100	-0.002496
H4	1.101	119.7	1.0	3	2	1	0.137	0.013208
C5	1.398	119.4	179.2	2	3	1	-0.081	-0.029889
C6	1.492	117.9	-0.6	5	2	1	0.343	-0.025547
C7	1.389	120.8	-179.8	3	2	4	-0.172	0.000011
H8	1.100	119.6	-0.1	7	3	4	0.135	0.019661
C9	1.394	120.7	179.0	5	2	6	-0.171	-0.005750
H10	1.100	120.2	0.9	9	5	6	0.134	0.022581
C11	1.415	120.4	-179.0	7	3	8	0.052	-0.009849
N12	1.399	120.6	2.9	11	9	10	-0.329	0.009874
O13	1.248	122.1	37.1	6	5	2	-0.367	-0.045690
N14	1.375	118.2	-179.2	6	5	13	-0.436	0.009047
H15	0.989	119.5	178.4	14	6	5	0.230	0.005879
H16	0.986	121.2	-175.5	14	6	15	0.220	0.027397
H17	0.996	114.5	25.1	12	11	7	0.187	0.026652
H18	0.995	114.6	133.5	12	11	17	0.185	0.028512

Inhibitor 3-HB (3-hydroxybenzamide)

AM1 optimised geometry, atomic charges and average MEP values for the conformation refered to in the text. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.159	-0.005612
C2	1.102	0.0	0.0	1	0	0	-0.114	-0.010925
C3	1.393	120.7	0.0	2	1	0	-0.099	0.002479
H4	1.100	119.9	1.0	3	2	1	0.140	0.017629
C5	1.401	120.1	179.1	2	3	1	-0.086	-0.024062
C6	1.492	118.0	-0.5	5	2	1	0.343	-0.020585
C7	1.393	120.4	-179.8	3	2	4	-0.193	0.005031
H8	1.099	120.5	0.2	7	3	4	0.137	0.024899
C9	1.395	120.4	178.9	5	2	6	-0.135	-0.006558
H10	1.099	122.2	0.7	9	5	6	0.152	0.014563
C11	1.402	119.2	-179.8	7	3	8	0.071	-0.010342
O12	1.377	116.1	-1.3	11	9	10	-0.251	-0.000515
O13	1.249	121.8	34.6	6	5	2	-0.368	-0.040928
N14	1.373	118.2	-179.2	6	5	13	-0.433	0.011263
H15	0.989	119.4	179.1	14	6	5	0.232	0.008464
H16	0.986	121.3	-177.3	14	6	15	0.224	0.026032
H17	0.968	107.9	-1.9	12	11	7	0.221	0.022342

Inhibitor I (3-nitrobenzamide)

AM1 optimised geometry, atomic charges and average MEP values for inhibitor I. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
H1	0.000	0.0	0.0	0	0	0	0.171	0.018253
C2	1.104	0.0	0.0	1	0	0	-0.038	0.012797
C3	1.394	120.7	0.0	2	1	0	-0.147	0.023668
H4	1.101	120.0	0.6	3	2	1	0.156	0.037327
C5	1.402	120.3	179.3	2	3	1	-0.128	-0.001627
C6	1.493	117.7	-0.1	5	2	1	0.344	-0.002494
C7	1.393	120.4	-179.6	3	2	4	-0.049	0.012629
H8	1.104	120.7	0.3	7	3	4	0.175	0.018813
C9	1.395	120.2	178.6	5	2	6	-0.047	0.004875
H10	1.104	121.5	0.4	9	5	6	0.171	0.017831
C11	1.404	119.0	-179.8	7	3	8	-0.141	-0.002763
N12	1.487	119.3	-1.3	11	9	10	0.569	-0.017831
O13	1.249	121.1	30.3	6	5	2	-0.362	-0.023288
N14	1.371	118.7	-179.1	6	5	13	-0.428	0.023699
H15	0.990	119.2	179.4	14	6	5	0.238	0.020661
H16	0.986	121.6	-177.2	14	6	15	0.226	0.034979
O17	1.201	118.9	-1.6	12	11	7	-0.352	-0.030373
O18	1.202	118.8	179.8	12	11	17	-0.359	-0.028763

Theophylline

AM1 optimised geometry, atomic charges and average MEP values for theophylline. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
C1	0.000	0.0	0.0	0	0	0	0.399	-0.026101
O2	1.244	0.0	0.0	1	0	0	-0.349	-0.028878
N3	1.411	121.7	0.0	2	1	0	-0.313	-0.024925
C4	1.445	120.0	-0.4	3	1	2	-0.075	-0.002835
C5	1.420	123.6	-179.7	3	1	2	0.405	-0.024221
O6	1.250	119.6	179.7	5	3	1	-0.360	-0.030339
N7	1.417	120.8	-0.3	5	3	1	-0.233	-0.022301
C8	1.439	121.0	-179.3	7	5	3	-0.065	0.007032
C9	1.386	118.3	-0.1	7	5	3	0.079	-0.022464
C10	1.432	120.6	0.3	9	7	5	-0.243	-0.022590
N11	1.395	106.3	179.9	10	9	7	-0.153	0.008805
C12	1.390	106.2	0.0	11	10	9	-0.040	0.017933
N13	1.361	113.3	0.0	12	11	10	-0.158	-0.001561
H14	1.122	109.5	5.0	4	3	1	0.120	-0.009369
H15	1.123	109.6	-115.5	4	3	1	0.092	-0.007653
H16	1.123	109.5	125.5	4	3	1	0.097	-0.008124
H17	1.123	109.4	-122.7	8	7	5	0.094	0.006518
H18	1.123	109.5	117.9	8	7	5	0.091	0.006084
H19	1.123	109.7	-2.4	8	7	5	0.124	-0.001021
H20	0.987	125.1	180.0	11	10	9	0.278	0.027168
H21	1.098	122.7	180.0	12	11	10	0.209	0.040730

Theobromine

AM1 optimised geometry, atomic charges and average MEP values for theobromine. Average MEP values were calculated on the double van der Waals surface of the molecule.

Atom	Bond	Angle	Dihedral	NA	NB	NC	Charge	Avepot
N1	0.000	0.0	0.0	0	0	0	-0.360	-0.015082
C2	1.410	0.0	0.0	1	0	0	0.407	-0.022471
N3	1.416	120.2	0.0	2	1	0	-0.233	-0.019643
C4	1.388	118.2	-0.4	3	2	1	0.078	-0.022379
C5	1.435	121.0	0.4	4	3	2	-0.241	-0.022841
C6	1.403	124.9	0.1	1	2	3	0.402	-0.022836
N7	1.398	106.3	179.8	5	4	3	-0.103	-0.011917
C8	1.394	106.0	0.1	7	5	4	-0.044	0.011501
N9	1.361	113.4	-0.1	8	7	5	-0.158	-0.002675
C10	1.439	121.0	-179.3	3	2	1	-0.065	0.007755
C11	1.427	126.6	179.7	7	5	4	-0.089	0.019624
O12	1.250	119.2	179.9	2	1	3	-0.363	-0.031631
O13	1.244	120.3	180.0	6	1	2	-0.354	-0.026595
H14	1.123	109.8	-118.4	11	7	5	0.095	0.026200
H15	1.125	109.4	1.6	11	7	5	0.144	0.001349
H16	1.123	109.7	121.7	11	7	5	0.095	0.028082
H17	1.123	109.4	-123.5	10	3	2	0.095	0.007033
H18	1.123	109.6	-3.3	10	3	2	0.123	-0.001726
H19	1.123	109.6	117.0	10	3	2	0.092	0.006175
H20	0.998	116.2	-179.8	1	2	3	0.272	-0.004631
H21	1.097	122.5	180.0	8	7	5	0.208	0.039293

12. Appendix B

This appendix contains a sample output from the GEMINI program, described in chapter 10, for the CH_2 molecule

```
                    +-----------------------------------+
                    |  GEMINI Ab initio geminal program  |
                    +-----------------------------------+
*****************************************************************************
User -- CHRDH on node -- CRAB     DIGITAL -- MICROVAX   VMS-V4.7
*****************************************************************************

Starting program on  Thursday 28-JUL-1988 at  22:08:03.90
Event flag --  0 -- will be used for this run

GEM_EXEDIR assigned   to -- $DISK3:[CHRDH.GEMINI.EXE]
GEM_SCRDIR assigned   to -- $DISK3:[SCRATCH.GEMINI]
GEM_DEFDIR assigned   to -- $DISK2:[CHRDH]
*****************************************************************************
            ----- Command cards from input file -----
*****************************************************************************
>ROUTE
Card  1 ----  >ROUTE
1=1,2=2,3=201,4=3,5=302,6=304,7=5,8=501,9=6,10=99
( 1)  1=1,2=2,3=201,4=3,5=302,6=304,7=5,8=501,9=6,10=99
END
>INCYC=2 OUTCYC=4 ITMG=3 INTCUT=8 TABCUT=6
Card  2 ----  >INCYC=2 OUTCYC=4 ITMG=3 INTCUT=8 TABCUT=6
>SAVEJNL SAVEINP STOP@501  LOC=ALL
Card  3 ----  >SAVEJNL SAVEINP STOP@501  LOC=ALL
>END
Card  4 ----  >END
*****************************************************************************
            ----- Filenames used for this run -----
*****************************************************************************
INT file --  $DISK3:[SCRATCH.GEMINI]CHRDHINT0.DAT
DMP file --  $DISK3:[SCRATCH.GEMINI]CHRDHDMP0.DAT
JNL file --  $DISK2:[CHRDH]CHRDHJNL0.DAT
INP file --  $DISK2:[CHRDH]CHRDHINP0.DAT
*****************************************************************************
Route taken from input file:
============================

( 1)    Link =    1  Jump =  1
( 2)    Link =    2  Jump =  1
( 3)    Link =  201  Jump =  1
( 4)    Link =    3  Jump =  1
( 5)    Link =  302  Jump =  1
( 6)    Link =  304  Jump =  1
( 7)    Link =    5  Jump =  1

( 8)    Link =  501  Jump =  1
( 9)    Link =    6  Jump =  1
(10)    Link =   99  Jump =  1


METHYLENE 3-21G BASIS SET GEMINAL CALCULATION USING
VECTORS PAIRED AS FOLLOWS
1,8  2,7  3,6  4,5  9,10,11,12,13

H
C 1 1.102103
H 2 1.102103  1 104.617


#BASIS
H  ------------ HYDROGEN -------------------
1.0 1 2
2 1.0 1 1.0
5.447178 0.156285
0.824547 0.904691
0.183192 1.00000
C -------------- CARBON --------------
6.0 2 3 2
3 1.0 2 1.0 1 1.0
172.256 0.061767
25.9109 0.358794
5.53335 0.700713
3.66498 -0.395897
0.770545 1.21584
0.195857 1.00000
2 1.0 1 1.0
3.66498 0.236460
0.770545 0.860619
0.195857 1.00000
#END BASIS

METHYLENE 3-21G BASIS SET GEMINAL CALCULATION USING
VECTORS PAIRED AS FOLLOWS
1,8  2,7  3,6  4,5  9,10,11,12,13
```

```
---------------------------------------------------------------------------
Input Z-MATRIX taken from file -- UNABLE TO DETERMINE FILENAME
---------------------------------------------------------------------------
(  1)        H
(  2)        C       1    1.10210
(  3)        H       2    1.10210     1 104.61700

---------------------------------------------------------------------------

---------------------------------------------------------------------------
                Z-matrix ( Angstroms & Degrees )
 Card  Center  Atom   NA      Bond      NB    Angle     NC    Dihedral
---------------------------------------------------------------------------
   1       1    H
   2       2    C       1    1.10210
   3       3    H       2    1.10210     1   104.617
---------------------------------------------------------------------------
```

Z-matrix Orientation --- Angstroms / Atomic Units

Center	Atom	X	Y	Z	X	Y	Z
1	H	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	C	0.000000	0.000000	1.102103	0.000000	0.000000	2.082673
3	H	1.066433	0.000000	1.380226	2.015266	0.000000	2.608250

Distance Matrix (Angstroms)

		1 H	2 C	3 H
1	H	0.000000		
2	C	1.102103	0.000000	
3	H	1.744220	1.102103	0.000000

```
PTGRP-- Translation Vector:    -0.251908     0.000000    -1.888036
Point Group                  =       CNV Principal Axis  N  =  2
Stoichiometry                =     CH2
Molecular charge             =     1
Multiplicity                 =     0
No. of atoms                 =     3
No. of Alpha orbitals        =     4
No. of Beta  orbitals        =     4
```

Standard Orientation --- Angstroms / Atomic Units

Center	Atom	X	Y	Z	X	Y	Z
1	H	0.000000	0.872110	0.505377	0.000000	1.648049	0.955025
2	C	0.000000	0.000000	-0.168459	0.000000	0.000000	-0.318342
3	H	0.000000	-0.872110	0.505377	0.000000	-1.648049	0.955025

```
Reading BASIS information from input file

No. of entries in basis table =  2
No. of primitives in table    =  12
Entries made in the basis table for the  following atoms:
  H   C
Number of different atoms in molecule =   2
Virtual memory requested                =     40000 real*8 words
Integrals will be computed in Symmetry format for CI and SCF use
Root name for integral files : $DISK3:[SCRATCH.GEMINI]CHRDH0
Threshold to neglect integrals          =  1.0000D-08
Time requested for integral calculation  = 99999.00
Time requested for restart interrupts    =    50.00

Group Multiplication Table
        1
Number of contracted basis functions   =     13


SYMMETRY TRANSFORMATION INFO
----------------------------

Number of Symmetry Adapted basis functions =     13
Irreducible Representation Number          =      1

   1     H1      S     1   1    1.00000
   2     H1      S     1   2    1.00000
   3     H3      S     1   3    1.00000
   4     H3      S     1   4    1.00000
   5     C2      S     1   5    1.00000
   6     C2      S     1   6    1.00000
   7     C2      S     1   7    1.00000
   8     C2      X     1   8    1.00000
   9     C2      Y     1   9    1.00000
  10     C2      Z     1  10    1.00000
  11     C2      X     1  11    1.00000
  12     C2      Y     1  12    1.00000
  13     C2      Z     1  13    1.00000
```

```
Nuclear Repulsion Energy =
Nuclear Repulsion Energy = 0.6065214E+01
Nuclear Repulsion Energy = 0.6065214E+01
Starting 1-electron integrals
Elapsed time =  0 hours  0 minutes  2 seconds.  Total time =  0 hours  0 minutes  2 seconds.

Starting 2-electron integrals
Elapsed time =  0 hours  0 minutes 20 seconds.  Total time =  0 hours  0 minutes 22 seconds.

Using routine : SYMINT

Number of non-zero  Super-Matrix elements =     2260
Number of blocks of Super-Matrix elements =        4
Finished 2-electron integrals
Elapsed time =  0 hours  0 minutes 11 seconds.  Total time =  0 hours  0 minutes 34 seconds.


Integral Tables
---------------


Threshold for integral tables              =    0.100E-05

  10        0.0677048114        0.0000010000
  20        0.0247689766        0.0000010000
  30        0.0021266671        0.0000010000
  40        0.0007759232        0.0000010000
  50        0.0002241094        0.0000010000
  60        0.0000436903        0.0000010000
  70        0.0000055399        0.0000010000

Closed Shell SCF
----------------


Threshold for linear dep.                  =   0.1000E-07


Initial guess taken from Core Hamiltonian

SCF Data
--------
Maximum number of cycles                   =    64
Convergence threshold on density matrix    =   0.1000E-05
```

Cycle	Total Energy	Electronic energy	E. Difference	1-E. energy	2-E. energy
1	-34.160641107	-40.225854985	40.225854985	-69.029772495	28.803917510
2	-37.191015171	-43.256229050	3.030374065	-58.178347860	14.922118809
3	-38.403486866	-44.468700745	1.212471694	-65.185731123	20.717030378
4	-38.636828275	-44.702042153	0.233341409	-62.804936578	18.102894425
5	-38.650873779	-44.716087658	0.014045505	-63.417668872	18.701581214
6	-38.651761330	-44.716975209	0.000887550	-63.284493428	18.567518220
7	-38.651839547	-44.717053426	0.000078217	-63.316072820	18.599019395
8	-38.651850668	-44.717064547	0.000011122	-63.309532119	18.592467572
9	-38.651852780	-44.717066659	0.000002112	-63.311248022	18.594181363
10	-38.651853226	-44.717067104	0.000000446	-63.310972490	18.593905385
11	-38.651853321	-44.717067199	0.000000095	-63.311084986	18.594017787
12	-38.651853341	-44.717067220	0.000000021	-63.311083004	18.594015784
13	-38.651853346	-44.717067224	0.000000004	-63.311094229	18.594027004
14	-38.651853347	-44.717067225	0.000000001	-63.311096700	18.594029474
15	-38.651853347	-44.717067226	0.000000000	-63.311098461	18.594031235
16	-38.651853347	-44.717067226	0.000000000	-63.311099141	18.594031915
17	-38.651853347	-44.717067226	0.000000000	-63.311099488	18.594032262
18	-38.651853347	-44.717067226	0.000000000	-63.311099642	18.594032417
19	-38.651853347	-44.717067226	0.000000000	-63.311099716	18.594032490

```
SCF    Converged in  19 cycles:  convergence  =  0.48788E-06

Electronic Energy  =      -44.7170672257
Nuclear energy     =        6.0652138788

Total Energy       =      -38.6518533468   A.U.
Total Energy       =      -38.6518533468   A.U.
Total Energy       =      -38.6518533468   A.U.


Eigenvalues
-----------

M.o.s (  1 --   4)    -11.224031141821     -0.896191110519     -0.565728313583     -0.375566801957
M.o.s (  5 --   8)      0.075772034440     0.288105329802      0.341041878484      0.922336544103
M.o.s (  9 --  12)      0.991795232045     1.000380175114      1.199793941728      1.295904172019
M.o.s ( 13 --  13)      1.621957335665
```

SCF wavefunction

			1	2	3	4	5	6
			-11.224031	-0.896191	-0.565728	-0.375567	0.075772	0.288105
1	H1	S	-0.001382	-0.173905	-0.241708	-0.101846	0.000000	-0.021711
2	H1	S	0.015202	-0.055669	-0.259406	-0.132102	0.000000	-1.371427
3	H3	S	-0.001382	-0.173905	0.241709	-0.101845	0.000000	-0.021712
4	H3	S	0.015202	-0.055669	0.259406	-0.132102	0.000000	-1.371426
5	C2	S	0.986815	0.200117	0.000000	-0.110467	0.000000	-0.133213
6	C2	S	0.091775	-0.219700	0.000000	0.106626	0.000000	0.055357
7	C2	S	-0.049904	-0.569427	0.000000	0.532075	0.000000	1.782384
8	C2	X	0.000000	0.000000	0.000000	0.000000	-0.381688	0.000000
9	C2	Y	0.000000	0.000000	-0.381266	0.000000	0.000000	0.000000
10	C2	Z	0.004892	-0.132342	0.000000	-0.419335	0.000000	0.226470
11	C2	X	0.000000	0.000000	0.000000	0.000000	-0.744188	0.000000
12	C2	Y	0.000000	0.000000	-0.242372	0.000000	0.000000	0.000000
13	C2	Z	-0.011956	-0.109558	0.000000	-0.464668	0.000000	0.907482

			7	8	9	10	11	12
			0.341042	0.922337	0.991795	1.000380	1.199794	1.295904
1	H1	S	0.051729	0.546324	-0.307015	0.000000	0.879796	-0.743933
2	H1	S	1.385559	-0.042772	-0.247959	0.000000	-0.486277	1.173612
3	H3	S	-0.051729	-0.546324	-0.307015	0.000000	0.879797	0.743932
4	H3	S	-1.385559	0.042771	-0.247958	0.000000	-0.486278	-1.173612
5	C2	S	0.000000	0.000000	-0.058485	0.000000	0.047759	0.000000
6	C2	S	0.000000	0.000000	0.042093	0.000000	0.281333	0.000000
7	C2	S	0.000000	0.000000	0.521333	0.000000	-0.312338	0.000000
8	C2	X	0.000000	0.000000	0.000000	1.114815	0.000000	0.000000
9	C2	Y	-0.332946	0.715585	0.000000	0.000000	0.000000	0.856019
10	C2	Z	0.000000	0.000000	-0.965127	0.000000	-0.476440	0.000000
11	C2	X	0.000000	0.000000	0.000000	-0.913610	0.000000	0.000000
12	C2	Y	-1.288030	-1.228900	0.000001	0.000000	0.000000	-1.018361
13	C2	Z	0.000000	0.000001	1.301906	0.000000	0.050347	0.000000

			13
			1.621957
1	H1	S	-0.105573
2	H1	S	0.988325
3	H3	S	-0.105573
4	H3	S	0.988325
5	C2	S	-0.049418
6	C2	S	1.741272
7	C2	S	-2.677031
8	C2	X	0.000000
9	C2	Y	0.000000
10	C2	Z	0.214452
11	C2	X	0.000000
12	C2	Y	0.000000
13	C2	Z	-0.760621

Density Matrix

			1	2	3	4	5	6
1	H1	S	0.1980804970					
2	H1	S	0.1716292879	0.1761449716				
3	H3	S	-0.0356114737	-0.0791731234	0.1980804935			
4	H3	S	-0.0791733043	-0.0930208558	0.1716294179	0.1761452736		
5	C2	S	-0.0498300492	0.0369091700	-0.0498300241	0.0369092030	2.0521061088	
6	C2	S	0.0544414299	-0.0009195967	0.0544414379	-0.0009195987	0.0696419300	0.13612
7	C2	S	0.0898112015	-0.0786950675	0.0898109491	-0.0786953441	-0.4439502236	0.35451
8	C2	X	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.00000
9	C2	Y	0.1843105916	0.1978054389	-0.1843106636	-0.1978057469	-0.0000000204	-0.00000
10	C2	Z	0.1314310719	0.1256734287	0.1314309488	0.1256732504	0.0493333077	-0.03037
11	C2	X	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.00000
12	C2	Y	0.1171667548	0.1257454888	-0.1171669538	-0.1257458048	-0.0000000042	-0.00000
13	C2	Z	0.1327872423	0.1346017418	0.1327869113	0.1346013386	0.0352151272	-0.05314

			8	9	10	11	12	13
8	C2	X	0.0000000000					
9	C2	Y	0.0000000000	0.2907280691				
10	C2	Z	0.0000000000	0.0000000018	0.3867600758			
11	C2	X	0.0000000000	0.0000000000	0.0000000000	0.0000000000		
12	C2	Y	0.0000000000	0.1848167615	-0.0000001796	0.0000000000	0.1174886052	
13	C2	Z	0.0000000000	0.0000001591	0.4185841402	0.0000000000	-0.0000000906	0.45612

Localisation of orbitals

Number of electron groups = 4
Number of orbitals to be localised = 4

Occupation coefficients - LR GR Oc.Coeff

LR	GR	Oc.Coeff
1	1	1.0000000
1	2	1.0000000
1	3	1.0000000
1	4	1.0000000

Boys Localisation of orbitals

			1	2	3	4
1	H1	S	-0.008426	-0.311086	-0.030741	-0.035349
2	H1	S	0.015766	-0.253259	-0.113597	-0.103826
3	H3	S	-0.008426	0.030741	0.311086	-0.035349
4	H3	S	0.015766	0.113597	0.253259	-0.103826
5	C2	S	0.999419	0.074795	-0.074795	-0.126593
6	C2	S	0.075943	-0.122102	0.122102	0.180207
7	C2	S	-0.097674	-0.244655	0.244655	0.693245
8	C2	X	0.000000	0.000000	0.000000	0.000000
9	C2	Y	0.000000	-0.269596	-0.269596	0.000000
10	C2	Z	0.009313	-0.190942	0.190942	-0.346952
11	C2	X	0.000000	0.000000	0.000000	0.000000
12	C2	Y	0.000000	-0.171383	-0.171383	0.000000
13	C2	Z	-0.004896	-0.186484	0.186484	-0.398103

Charge centeroids of the electron pairs (in Angstroms):

Group	X - component	Y - component	Z - component	R = SQRT(X**2 + Y**2 + Z**2)
1	0.00000000	0.00000000	-0.16776306	0.16776306
2	0.00000000	0.59431195	0.28999849	0.66129103
3	0.00000000	-0.59431212	0.28999837	0.66129112
4	0.00000000	0.00000010	-0.64079188	0.64079188

Distance between charge centeroids: (Angstroms)

	1	2	3	4
1	0.0000000000			
2	0.7501682046	0.0000000000		
3	0.7501682589	1.1886240713	0.0000000000	
4	0.4730288226	1.1043447374	1.1043448278	0.0000000000

Dipole mom.	X - component	Y - component	Z - component	Length
Nuclear components	0.00000000	0.00000000	-0.00000020	
Electronic components	0.00000000	0.00000024	0.86382463	
Total	0.00000000	0.00000024	0.86382443	0.86382443
Total(in Debye)	0.00000000	0.00000062	2.21919133	2.21919133

Redefined localised orbitals

		1	2	3	4
1 H1	S	-0.008426	-0.311086	-0.030741	-0.035349
2 H1	S	0.015766	-0.253259	-0.113597	-0.103826
3 H3	S	-0.008426	0.030741	0.311086	-0.035349
4 H3	S	0.015766	0.113597	0.253259	-0.103826
5 C2	S	0.999419	0.074795	-0.074795	-0.126593
6 C2	S	0.075943	-0.122102	0.122102	0.180207
7 C2	S	-0.097674	-0.244655	0.244655	0.693245
8 C2	X	0.000000	0.000000	0.000000	0.000000
9 C2	Y	0.000000	-0.269596	-0.269596	0.000000
10 C2	Z	0.009313	-0.190942	0.190942	-0.346952
11 C2	X	0.000000	0.000000	0.000000	0.000000
12 C2	Y	0.000000	-0.171383	-0.171383	0.000000
13 C2	Z	-0.004896	-0.186484	0.186484	-0.398103

Charge centeroids of the electron pairs (in Angstroms):

Group	X - component	Y - component	Z - component	R = SQRT(X**2 + Y**2 + Z**2)
1	0.00000000	0.00000000	-0.16776306	0.16776306
2	0.00000000	0.59431195	0.28999849	0.66129103
3	0.00000000	-0.59431212	0.28999837	0.66129112
4	0.00000000	0.00000010	-0.64079188	0.64079188

Distances between charge centeroids: (Angstroms)

	1	2	3	4
1	0.0000000000			
2	0.7501682046	0.0000000000		
3	0.7501682589	1.1886240713	0.0000000000	
4	0.4730288226	1.1043447374	1.1043448278	0.0000000000

Dipole mom.	X - component	Y - component	Z - component	Length
Nuclear components	0.00000000	0.00000000	-0.00000020	
Electronic components	0.00000000	0.00000024	0.86382463	
Total	0.00000000	0.00000024	0.86382443	0.86382443
Total(in Debye)	0.00000000	0.00000062	2.21919133	2.21919133

Geminal Calculation

Input parameters:

Number of basis functions: 13
Number of electron groups: 4
Number of natural orbitals to be used on the left hand side (I,NNO(I),I=1,NELGR)

1 2 2 2 3 2 4 2
Maximum number of iterations - Outer Cycle: 4
Maximum number of iterations - Inner Cycle: 2
Maximum number of iterations - Orbital Mixing: 0 3
The number of orbitals used in the calculation of EPS(K): 20
Number of selected C.Wfns. used in the iterativecalculation of EPS(K): 20

Convergence threshold - Outer Cycle: 0.1E-07

Convergence thresholds - Inner Cycle: 0.1E-04 0.1E-08 0.1E-01
Scale factor for orbital mixing: 1.000

Nuclear electrostatic energy: 0.60652139E+01

Occupation coefficients - LR GR Oc.Coef.

LR	GR	Oc.Coef.
1	1	0.9999984
2	1	-0.0018148
1	2	0.9952995
2	2	-0.0968450
1	3	0.9953030
2	3	-0.0968092
1	4	0.9872375
2	4	-0.1592552

Geminal Optimization

CYCLE	GROUP	ITERATION		GROUP ENERGY	DNORM
1	1				
		1	1	-25.9272036114	0.0000076509
		1	2	-25.9272036121	0.0001427814
1	2				
		1	1	-1.9261761034	0.0000875568
		1	2	-1.9261761127	0.0002059298
1	3				
		1	1	-1.9260459421	0.0000901116
		1	2	-1.9260459520	0.0002405408
1	4				
		1	1	-1.7876207841	0.0002467571
		1	2	-1.7876208618	0.0001624312
1	1				
		1	1	-25.9275631107	0.0000044266
		1	2	-25.9275631110	0.0000747978
1	2				
		1	1	-1.9262987268	0.0000702301
		1	2	-1.9262987336	0.0000949707
1	3				
		1	1	-1.9261143678	0.0000760722
		1	2	-1.9261143753	0.0001394870
1	4				
		1	1	-1.7876258125	0.0000312184
		1	2	-1.7876258138	0.0001456220

Geminal mixing

Iteration	E(APSG)	Norm of gradient
1	-38.7006844823	0.0004006202
2	-38.7006851262	0.0002425283
3	-38.7006852207	0.0000853065

Geminal Optimization

CYCLE	GROUP	ITERATION		GROUP ENERGY	DNORM
2	1				
		1	1	-25.9274150637	0.0000020867
		1	2	-25.9274150637	0.0000393712
2	2				
		1	1	-1.9235167615	0.0000293163
		1	2	-1.9235167625	0.0000739199
2	3				
		1	1	-1.9234936183	0.0000326202
		1	2	-1.9234936195	0.0000847487
2	4				
		1	1	-1.7911172973	0.0001053508
		1	2	-1.7911173114	0.0000631715
2	1				
		1	1	-25.9275772585	0.0000016025
		1	2	-25.9275772585	0.0000204747
2	2				
		1	1	-1.9235668348	0.0000295683
		1	2	-1.9235668360	0.0000361300
2	3				
		1	1	-1.9235183180	0.0000300492
		1	2	-1.9235183191	0.0000461584
2	4				
		1	1	-1.7911200922	0.0000122781
		1	2	-1.7911200923	0.0000544939

Geminal mixing

Iteration	E(APSG)	Norm of gradient
1	-38.7006855712	0.0001648419
2	-38.7006856582	0.0000885124
3	-38.7006856700	0.0000293457

Geminal Optimization

CYCLE	GROUP	ITERATION		GROUP ENERGY	DNORM
3	1				
		1	1	-25.9275112609	0.0000005724
		1	2	-25.9275112609	0.0000108063
3	2				
		1	1	-1.9225501650	0.0000110734
		1	2	-1.9225501651	0.0000268691

3	3				
		1	1	-1.9225491539	0.0000125756
		1	2	-1.9225491541	0.0000296853
3	4				
		1	1	-1.7923997832	0.0000429263
		1	2	-1.7923997855	0.0000236959
3	1				
		1	1	-25.9275778337	0.0000005858
		1	2	-25.9275778337	0.0000056183
3	2				
		1	1	-1.9225693542	0.0000118910
		1	2	-1.9225693544	0.0000134165
3	3				
		1	1	-1.9225579795	0.0000116615
		1	2	-1.9225579797	0.0000157531
3	4				
		1	1	-1.7924010675	0.0000046629
		1	2	-1.7924010676	0.0000198266

Geminal mixing

Iteration	E(APSG)	Norm of gradient
1	-38.7006857120	0.0000654713
2	-38.7006857234	0.0000317667
3	-38.7006857249	0.0000102208

Geminal Optimization

CYCLE	GROUP	ITERATION		GROUP ENERGY	DNORM
4	1				
		1	1	-25.9275506221	0.0000001626
		1	2	-25.9275506221	0.0000029839
4	2				
		1	1	-1.9222046983	0.0000041566
		1	2	-1.9222046984	0.0000097311
4	3				
		1	1	-1.9222062339	0.0000048203
		1	2	-1.9222062339	0.0000104233
4	4				
		1	1	-1.7928620612	0.0000169198
		1	2	-1.7928620615	0.0000087158

```
 4       1
                 1    1          -25.9275766472        0.0000002156
                 1    2          -25.9275766472        0.0000015498


 4       2
                 1    1           -1.9222118499        0.0000046345
                 1    2           -1.9222118499        0.0000049089


 4       3
                 1    1           -1.9222093633        0.0000044562
                 1    2           -1.9222093633        0.0000054754


 4       4
                 1    1           -1.7928626249        0.0000017402
                 1    2           -1.7928626249        0.0000070988
```

Geminal mixing

Iteration	E(APSG)	Norm of gradient
1	-38.7006857300	0.0000253952
2	-38.7006857315	0.0000113281
3	-38.7006857317	0.0000035863

Geminal Corrections: EPS/SUB/K

ELECTRON PAIR: 1

ITERATION	GEMINAL ENERGY	NORM OF RESIDUAL VECTOR
1	-25.9276580743	0.0000000000
2	-25.9276580934	0.0006738611

EPS/SUB/K = -0.0000921760

ELECTRON PAIR: 2

ITERATION	GEMINAL ENERGY	NORM OF RESIDUAL VECTOR
1	-1.9254145105	0.0000000169
2	-1.9255206520	0.0137669644

EPS/SUB/K = -0.0034384858

ELECTRON PAIR: 3

ITERATION	GEMINAL ENERGY	NORM OF RESIDUAL VECTOR
1	-1.9254144135	0.0000000000
2	-1.9255204691	0.0137607465

EPS/SUB/K = -0.0034388386

ELECTRON PAIR: 4

ITERATION	GEMINAL ENERGY	NORM OF RESIDUAL VECTOR
1	-1.7933514373	0.0000000000
2	-1.7933549388	0.0026911932

EPS/SUB/K = -0.0003314841

OPTIMIZED GEMINALS

****

ELECTRON PAIR: 1
****

GEMINAL ENERGY:

-25.9275659174

EXPANSION COEFFICIENTS:

0.99999833 -0.00182935

NATURAL ORBITAL EXPANSION COEFFICIENTS BASED ON FIRST-ORDER W. FN.

0.99999654 -0.00182935 -0.00142730 -0.00093443 -0.00080986 -0.00000060 -

NATURAL ORBITALS:

LK = 1

-0.00174276 0.01597962 -0.00174278 0.01597952 0.98896817 0.08964842 -
0.00000006 0.00692955 0.00000000 -0.00000011 -0.00919894

LK = 2

0.01255852 0.10142987 0.01235559 0.10026861 -0.07246016 1.21142709 -
0.00057265 -0.66860493 0.00000000 -0.00122354 0.57039733

ELECTRON PAIR: 2
****

GEMINAL ENERGY:

-1.9220821663

EXPANSION COEFFICIENTS:

0.99526068 -0.09724285

NATURAL ORBITAL EXPANSION COEFFICIENTS BASED ON FIRST-ORDER W. FN.

0.99467553 -0.09718568 -0.03271972 -0.00872306 -0.00421693 -0.00332036 -

NATURAL ORBITALS:

LK = 1

-0.02720723 -0.08588320 0.31807810 0.27609346 -0.06296028 0.07977245
-0.27341970 0.25521010 0.00000000 -0.16869262 0.25278308

LK = 2

0.01503444 0.03129176 0.71840307 0.60750398 0.09442700 -0.25485346 -
0.54066829 -0.43996533 0.00000000 0.31441233 -0.21608057

ELECTRON PAIR: 3
****

GEMINAL ENERGY:

-1.9220816305

EXPANSION COEFFICIENTS:

0.99526069 -0.09724281

NATURAL ORBITAL EXPANSION COEFFICIENTS BASED ON FIRST-ORDER W. FN.

0.99467523 -0.09718561 -0.03272890 -0.00872382 -0.00421694 -0.00332036 -0.

NATURAL ORBITALS:

LK = 1

0.31807853 0.27609191 -0.02720748 -0.08588349 -0.06296014 0.07977204 0.
0.27342005 0.25521019 0.00000000 0.16869307 0.25278367

LK = 2

0.71840862 0.60749792 0.01503568 0.03128418 0.09442660 -0.25485058 -0.
-0.54066662 -0.43996604 0.00000000 -0.31441698 -0.21607703

ELECTRON PAIR: 4
****

GEMINAL ENERGY:

-1.7930234547

EXPANSION COEFFICIENTS:

0.98757437 -0.15715237

NATURAL ORBITAL EXPANSION COEFFICIENTS BASED ON FIRST-ORDER W. FN.

0.98751584 -0.15714306 -0.01026784 -0.00325976 -0.00126609 -0.00071389 -0.

NATURAL ORBITALS:

LK = 1

0.01964859 -0.08733997 0.01964847 -0.08734072 -0.20196622 0.22071822 0.
0.00000012 -0.24780389 0.00000000 0.00000006 -0.28623351

LK = 2

0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.
0.00000000 0.00000000 -0.52222464 0.00000000 0.00000000

CHARGE CENTEROIDS OF THE ELECTRON PAIRS (IN A.U.):

****

GROUP	X - COMPONENT	Y - COMPONENT	Z - COMPONENT	R = SQRT(X**2 + Y**2 + Z**2)
1	0.00000000	0.00000002	-0.31690082	0.31690082
2	0.00000000	-0.99148349	0.51621112	1.11781637
3	0.00000000	0.99148189	0.51620982	1.11781435
4	0.00000000	0.00000061	-1.11434410	1.11434410

DIPOLMOMENT	X - COMPONENT	Y - COMPONENT	Z - COMPONENT	LENGTH
NUCLAR COMPONENTS	0.00000000	0.00000000	-0.00000020	
ELECTRONIC COMPONENTS	0.00000000	0.00000194	0.79764796	
TOTAL	0.00000000	0.00000194	0.79764776	0.79764776
TOTAL(IN DEBYE)	0.00000000	0.00000499	2.04918144	2.04918144

CHARGE CENTEROIDS OF THE ELECTRON PAIRS (IN A.U.):

****

GROUP	X - COMPONENT	Y - COMPONENT	Z - COMPONENT	R = SQRT(X**2 + Y**2 + Z**2)
1	0.00000000	0.00000240	-0.31985112	0.31985112
	0.00000000	-0.00000237	-0.31395053	0.31395053
2	0.00000000	-1.26116143	0.72125349	1.45283679
	0.00000000	-0.72180554	0.31116875	0.78602114
3	0.00000000	1.26115932	0.72125205	1.45283425
	0.00000000	0.72180447	0.31116759	0.78601969
4	-0.35185201	0.00000061	-1.11434410	1.16857289
	0.35185201	0.00000061	-1.11434410	1.16857289

DIPOLMOMENT	X - COMPONENT	Y - COMPONENT	Z - COMPONENT	LENGTH
NUCLAR COMPONENTS,	0.00000000	0.00000000	-0.00000020	
ELECTRONIC COMPONENTS	0.00000000	0.00000194	0.79764796	
TOTAL	0.00000000	0.00000194	0.79764776	0.79764776
TOTAL(IN DEBYE)	0.00000000	0.00000499	2.04918144	2.04918144

BOND PROFILES

****

ELECTRON PAIR (NUCLEUS, LOCALIZATION)

1
H1 0.96065, H3 0.96065, C2 11.26993,

2
H1 0.71490, H3 1.85380, C2 1.31034,

3
H1 1.85380, H3 0.71490, C2 1.31034,

4
H1 0.73722, H3 0.73722, C2 1.69070,

ENERGY PARTITIONING

****

GROUP	KINETIC ENERGY	ONE-EL. POT. ENERGY	ONE-ELECTRON ENERGY	RED. GEM. EN./APSG
1	31.9181602012	-69.5408780623	-37.6227178611	-34.1105258166
2	2.0058774861	-10.4307334575	-8.4248559714	-7.8083870341
3	2.0058789851	-10.4307359216	-8.4248569364	-7.8083879957
4	2.7668455635	-11.6186188777	-8.8517733142	-8.2397452057
	38.6967622359	-102.0209663191	-63.3242040832	-57.9670460522

GROUPS		COULOMB TERMS	EXCHANGE TERMS
1	1	14.0494270790	7.0247032486
1	2	2.5845728419	0.0271988783
1	3	2.5845736065	0.0271988722
1	4	3.1429125660	0.0747013646
2	2	2.5977744871	1.2810274690
2	3	1.6753037113	0.0356277185
2	4	1.7718537788	0.0825988672
3	3	2.5977746731	1.2810275631
3	4	1.7718548404	0.0825992024
4	4	2.5556830405	1.2330599052
		13.5310713449	0.3299249033

INTRAPAIR CORRELATION ENERGIES (WITH RESPECT TO APSG):

GROUP	CORR. ENERGY
1	-.0000921760
2	-.0034384858
3	-.0034388386
4	-.0003314841

TOTAL INTRAPAIR CORRELATION ENERGY: -0.0073009845

NUCLEAR POTENTIAL ENERGY: 6.0652138788

THE TOTAL ENERGY WITHIN THE APSG-MODEL:

-38.7006857317

THE TOTAL ENERGY INCLUDING SINGLE PAIR CORRECTION TERMS:

-38.7079867162

```
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
TEST STOP in LINK    6
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

g  0 cleared.
%% Logical name -- GEM_INPFIL -- deleted
%% Logical name -- GEM_JNLFIL -- deleted
HRDH        job terminated at 28-JUL-1988 16:16:45.45
```

REFERENCES

1) E. Schrödinger, *Ann. der Physik,* **79**, 361 (1926).

2) C.C.J. Roothaan, *Rev. Mod. Phys.,* **23**, 69 (1951).

3) A good account of the Hartree Fock theory and its applications can be found in: W.J. Hehre, L. Radom, P. von R. Schleyer and J.A. Pople, "Ab initio Molecular Orbital Theory", John Wiley, New York (1986).

4) See for example, the review by E.R. Davidson and D. Feller, *Chem. Rev.,* **86**, 681 (1986), and also reference 3.

5) M.J.S. Dewar and W. Thiel, *J. Am. Chem. Soc.,* **99**, 4899, 4907 (1977).

6) M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).

7) N.C. Baird and G.C. Hadley, *Chem. Phys. Lett.*, **128** (1), 31 (1986).

8) I.I. Merkelbach, H.G.M. Becht and H.M. Bock. *J. Am. Chem. Soc.,* **107**, 4037 (1985).

9) G.P. Ford and J.P. Scribner, *J. Am. Chem. Soc.* **103**, 4281 (1981).

10) K. Kirste and P. Rademacher, *J. Mol. Struct.,* **73**, 171 (1981).

11) T. Oie, G.H. Loew, S.K. Burt, J.S. Binkley and R.D. MacElroy, *J. Am. Chem., Soc.* **104**, 6169 (1982).

12) M.J.S. Dewar and H.S. Rzepa, *J. Am. Chem. Soc.,* **100**, 58 (1978).

13) S. Schneider, *Theor. Chim. Acta,* **57**, 71 (1980).

14) H. Perrin and G. Berges, *THEOCHEM,* **1** 299 (1981).

15) M.J.S. Dewar and E.G. Zoebisch *J. Am. Chem. Soc.,* (to be published).

16) W.D.S. Motherwell, "PLUTO: A program for plotting molecular and

crystal structures", University Chemical Laboratory, Lensfield Road, Cambridge, England.

17) C. Humblet and G.R. Marshall, *Drug Development Research,* 1 409 (1981).

18) D.E. Stewart, P.K. Weiner and J.E. Wampler, *J. Mol. Graph.,* 5 (3), 133 (1987).

19) J.M. Burridge, P. Quarendon, C.A. Reynolds and P.J. Goodford, *J. Mol. Graph.,* 5 (3), 165 (1987).

20) F.H. Allen *et al,* Acta Cryst., Vol, B35, 2331 (1979).

21) Protein Data Bank, Chemistry Department Brookhaven National Laboratory, New York, USA.

22) CHEM-X, developed and distributed by Chemical Design Ltd., Oxford England.

23) The Microvax II GPX workstation was donated to our group by DEC.

24) K. Ueda, M. Kawaichi, J. Oka and O. Hayaishi in "Novel ADP-ribosylations of regulatory enzymes and proteins", (M.E. Smulson and T. Sugimura eds.), Elsevier/North-Holland, New York (1980).

25) K. Ueda, M. Kawaichi and O. Hayaishi in "ADP-ribosylation reactions:
Biology and Medicine", (O. Hayaishi and K. Ueda eds.), New York Academic (1982).

26) E.E. Okolie and N.I. Onyezili, *Biochem. J.,* **209**, 687 (1983).

27) T. Kristensen and J. Holtlund, *Eur. J. Biochem.,* **88**, 495 (1978).

28) D.B. Jump and M. Smulson, *Biochemistry* **19**, 1031 (1980).

29) K. Ikai, K. Ueda, M. Fukushima, T. Nakamura and O. Hayaishi, *Proc. Natl. Acad. Sci.* USA., **77**, 3682 (1980).

30) Y. Nishizuka, K. Ueda, K. Nakazawa and O. Hayaishi, *J. Biol. Chem.*, 242, 3164 (1967).

31) A. Oikawa, Y. Itai, H. Okuyama, S. Hasegawa and T. Sugimura, *Exp. Cell. Res.*, 57, 154 (1969).

32) K. Itai, K. Ueda and O. Hayaishi, *J. Histochem. Cytochem.*, 28, 670 (1980).

33) Y. Kanai, S. Tanuma and T. Sugimura, *Proc. Natl. Acad. Sci.* USA., 78, 2801 (1981).

34) S. Ito, Y. Shizuta and O. Hayaishi, *J. Biol. Chem.*, 254, 3647 (1979).

35) C. Niedergang, H. Okazaki and P. Mandel, *Eur. J. Biochem.*, 102, 43 (1979).

36) H. Ohgushi, K. Yoshihara and T. Kamiya, *J. Biol. Chem.*, 255, 6205 (1980).

37) I. Kameshita, Z. Matsuda, T. Taneguchi and Y. Shizuta, *J. Biol. Chem.*, 259 (8), 4770 (1984).

38) T. Kurosaki, H. Ushiro, Y. Mitsuuchi, S. Suzuki, M. Matsuda, Y. Matsuda, N. Katunuma, K. Kangawa, H. Matsuo, T. Hirose, S. Inayama and Y. Shizuta, *J. Biol. Chem.*, 262, 15990 (1987).

39) H. Ushiro, Y. Yokoyama and Y. Shizuta, *J. Biol. Chem.*, 262, 2352 (1987).

40) R.C. Benjamin and D.M. Gill, *J. Biol. Chem.* 255, 10493 (1980).

41) H. Halldorsson, D.A. Gray and S. Shall, *FEBS Lett.*, 85, 349 (1978).

42) H. Okayama, K. Ueda and O. Hayaishi, *Proc. Natl. Acad. Sci.*, USA, 75, 1111 (1978).

43) K. Ueda, A. Omachi, M. Kawaichi and O. Hayaishi, *Proc. Natl. Acad. Sci.*, USA, 72, 205 (1975).

44) D.B. Jump and M. Smulson, *Biochemistry*, 19, 1024 (1980).

45) D. Creissen and S. Shall, *Nature*, 296, 271 (1972).

46) N. Ogata, M. Kawaichi, K. Ueda and O. Hayaishi, *Biochem. Int.*, 1, 229 (1980).

47) B.W. Durkacz, O. Omidijii, D.A. Gray and S. Shall, *Nature,* 283, 593 (1980).

48) S. Shall in "DNA repair and its inhibition", pp 143-191, IRL Press, Oxford (1984).

49) J.E. Cleaver and W.F. Morgan, *Mutation Res.,* 150, 69 (1985).

50) O. Cantoni, P. Sestili, F. Cattabeni and V. Stocchi, *FEBS Lett.,* 204 (2), 266 (1986).

51) H.M. Wallace, A.M. Gordon, H.M. Keir and C.K. Pearson, *Biochem. J.,* 219, 211 (1984).

52) G. Jackowski and E. Kun, *J. Biol. Chem.,* 256, 3667 (1981).

53) U. Wintersberger and E. Wintersberger, *FEBS Lett.,* 188 (2), 189 (1985).

54) T. Sugimura and M. Miwa, *Carcinogenesis,* 4 (12), 1503 (1983).

55) O. Cantoni, Personal communication.

56) S. Shall in "ADP-ribosylation, DNA repair and cancer", (M. Miwa *et al* eds.), *Japan. Sci. Soc. Press,* Tokyo/VNU Science Press, Utrecht, pp 3-25 (1983).

57) J.C. Gaal and C.K. Pearson, *Biochem. J.* 230, 1 (1985).

58) K. Ueda and O. Hayaishi, *Ann. Rev. Biochem.,* 54, 73 (1985).

59) J.L. Sims, G.W. Sikorski, D.M. Catino, S.J. Berger and N.A. Berger, *Biochemistry* 21, 1813 (1982).

60) M.R. Purnell and W.J.D. Whish, *Biochem. J.,* 185, 775-777 (1980).

61) J.J.P. Stewart, QCPE Bull., 5, 133 (1985).

62) J.S. Binkley, M.J. Frisch, D.J. De Frees, K. Raghavachari, R.A. Whiteside, H.B. Schlegel and J.A. Pople, "Gaussian 82", Department of Chemistry, Carnegie-Mellon University, Pittsburgh, PA, USA

(1982).

63) M.J. Frisch, J.S. Binkley, H.B. Schlegel, K. Raghavachari, C.F. Melius, R.L. Martin, J.J.P. Stewart, F.W. Bobrowicz, C.M. Rohlfing, L.R. Kahn, D.J. De Frees, R. Seeger, R.A. Whiteside, D.J. Fox, E.M. Fleuder and J.A. Pople, "Gaussian 86", Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh PA, (1986).

64) M. Guest, J. Kendrick and S. Pope, "GAMESS", S.E.R.C. Daresbury Laboratory (1982).

65) T.H. Dunning, *J. Chem. Phys.*, 53, 2823 (1970), 55, 716 (1971).

66) U. Chandra Singh and P.A. Kollman, *QCPE Bull.* 2 No 446, p 117 (1982).

67) R.S. Mulliken, *J. Chem. Phys.*, 23, 1833 (1955).

68) D.G. Truhlar in "Chemical Applications of Atomic and Molecular Electrostatic Potentials", (P. Politzer and D.G. Truhlar, Eds.), Plenum Press, New York (1981).

69) R. Bonaccorsi, E. Scrocco and J. Tomasi, in "Theoretical Section Progress Report 1969/1970", Laboratorio di Chimica Quantistica ed Energetica Molecolare del C.N.R. ed Instituto di Chimica-Fisica dell'Università di Pisa, Pisa, Italy, p. 35 (1970).

70) E. Scrocco and J. Tomasi, *J. Chem. Phys.*, 52, 5270 (1970).

71) R. Bonaccorsi, E. Scrocco, J. Tomasi and A. Pullman, *Theoret. Chim. Acta*, 24, 51 (1972).

72) D. Perahia and A Pullman, *Theoret. Chim. Acta*, 48, 263 (1978).

73) L.C. Allen in "Quantum Chemistry in Biomedical Sciences", (H. Weinstein and J.P. Green eds.), New York Academy of Sciences, New York, 383 (1981).

74) W.G. Richards "Quantum Pharmacology (2nd Edition), Butterworths,

London (1983).

75) C. Petrongolo, H.J.T. Preston and J.J. Kaufman, *Int. J. Quant. Chem.,* Vol. XIII, 457 (1978).

76) R. Lavery and A. Pullman, in "New Horizons of Quantum Chemistry", (P.-O. Löwdin and B. Pullman, eds.), D. Reidel Publishing Company, p. 439 (1983).

77) H. Weinstein, R. Osman, J.P. Green and S. Topiol, in "Chemical Applications of Atomic and Molecular Electrostatic Potentials", (P. Politzer and D.G. Truhlar eds.), Plenum Press, New York (1981).

78) H. Weinstein, S. Maayani, S. Srebenik, S. Cohen and M. Sokolovsky, *Mol. Pharmacol.,* 11, 671 (1975).

79) R. Osman and H. Weinstein, *Chem. Phys. Lett.,* 49, 69 (1977).

80) R. Bonaccorsi, C. Petrongolo, E. Scrocco and J. Tomasi, *Theoret. Chim. Acta,* 20, 331 (1972).

81) H.B. Schlegel, *J. Comp. Chem.,* 3, 214 (1982).

82) B.A. Murtagh and R.W.H. Sargent, *Comput. J.,* 13, 185 (1980).

83) J. Baker, *J. Comp Chem.,* 7, 385 (1986).

84) R. Fletcher and M.J.D. Powell, *Comput. J.,* 6, 163 (1963).

85) S. Saebo, L. Farnell, N.V. Riggs and L. Radom, *J. Am. Chem. Soc.,* 106 (18), 5047 (1984).

86) D.J. DeFrees and A.D. McLean, *J. Chem. Phys.,* 82 (1), 333 (1985).

87) E.B. Wilson, J.C. Decius and P.C. Cross, "Molecular Vibrations", McGraw-Hill, New York (1955).

88) D. Higgins, unpublished work (1987).

89) C.M. Edge, Ph.D. thesis, University of St Andrews (1987).

90) C.M. Edge, D. Higgins, and C. Thomson, *J. Mol. Graph.,* accepted.

91) "4107/4109 Computer Display Terminals: Programmer's Reference Manual", Tektronix, Beaverton, Oregon (1983).

92) M.L. Connolly, *Science,* **221**. 709 (1983).

93) M.L. Connolly, *QCPE Bull.,* **1**, 75 (1981).

94) D. Peters and M. Sana, *QCPE* **360**, (1978).

95) C. Giessner-Prettre and A. Pullman, *Theoret. Chim. Acta,* (Berl.), **25**, 83 (1972).

96) D. Higgins, Unpublished work (1986).

97) "VAX/VMS Run-time Library Routines Reference Manual", Digital Equipment Corporation, Maynard, Massachusetts (1985).

98) "MicroVMS Workstation Graphics Programming guide", version 3.0, Digital Equipment Corporation, Maynard, Massachusetts (1986).

99) J. Brickmann, *J. Mol Graph.,* **1** (3), 62 (1983).

100) The Values of atomic and van der Waals radii were obtained from *CRC Handbook of Chemistry and Physics*, 62nd edition, CRC Press, Inc. Boca Raton, Florida, (1982).

101) A. Phin, unpublished work (1988).

102) ""MicroVMS Workstation Guide to Printing Graphics", version 3.0, Digital Equipment Corporation, Maynard, Massachusetts (1986).

103) C.A. Reynolds, personal communication (1987).

104) D. Higgins, C. Thomson and W. Thiel, *J. Comp. Chem.,* accepted (1988).

105) W. Thiel, *J. Am. Chem. Soc.,* **103**, 1413 (1981).

106) W. Thiel. *QCPE Bull.,* **2** (2), 63 (1982).

107) C. Thomson and D. Higgins, *Int. J. Quant. Chem.,* (in press).

108) C. Thomson, *Carcinogenesis.*

109) K. Kitaura and K. Morokuma, *Int. J. Quant. Chem.,* **10**, 325 (1976).

110) C. Ghio and J. Tomasi, *Theoret. Chim. Acta* (Berlin), **27**, 171 (1972).

111) D. Perahia, A. Pullman and H. Berthod, *Theoret. Chim. Acta* (Berlin), **40**, 47 (1976).

112) J. Almlöf and A. Stogard, *Chem. Phys. Lett.*, 29 (3), 418 (1974).

113) A.D. Tait and G.G. Hall. *Theoret. Chim. Acta* (Berlin), 31, 311 (1973).

114) G.G. Hall, *Chem. Phys. Lett.*, 20, 501 (1973).

115) G. Berthier, R. Bonaccorsi, E. Scrocco and J. Tomasi, *Theoret. Chim. Acta* (Berlin), 26, 101 (1972).

116) R.B. Homer and C.D. Johnson, in "Chemistry of amides", (J. Zabicky ed.), John Wiley, New York (1970).

117) A.C. Hopkinson and I.G. Csizmadia, *Can. J. Chem.*, 51, 1432 (1973).

118) J.P. Green, C.L. Johnson, H. Weistein, S. Kang and D. Chou, in "The Psychopharmacology of Hallucinogens", Pergamon Press, New York.

119) H. Weinstein, *Int. J. Quant. Chem.*, Quantum Biol. Symp., 2, 59 (1975).

120) O. Cantoni, personal communication.

121) See for example M. Gomel and H. Lambruso, *Bull. Chim. Soc. Fr.*, 1196 (1962).

122) C.J. Brown and D.E.C. Corbridge, *Nature*, 162, 72 (1948).

123) A.J.R. Buorn, D.G. Gillies and E.W. Randall, *Tetrahedron*, 20, 1811 (1964).

124) L.A. LaPlanche and M.T. Rogers, *J. Am. Chem. Soc.*, 85, 3728 (1963).

125) C.J. Brown and D.E.C. Corbridge, *Acta Crystallogr.*, 7, 711 (1954).

126) C.J. Brown and D.E.C. Corbridge, *Acta Crystallogr.*, 21, 442 (1962).

127) M.J. Aroney, R.J.W. Le Febre and A.N. Singh, *J. Chem. Soc.*, (London), 511 (1963).

128) M. Gomel, H. Lombroso and D. Peltier, *Compt. Rend.*, 254, 3857 (1962).

129) W.J. Hehre, L. Radom and J.A. Pople, *J. Am. Chem. Soc.*, 94 (5), 1496 (1972).

130) R. Cervellati, A. Degli Esposti, D.G. Lister and P. Palmieri, *J. Mol.*

Struct. (THEOCHEM), 122, 173 (1985).

131) R.A. Kydd and A.R.C. Dunham, *J. Mol. Struct.*, 98, 39 (1983).

132) R. Cervellati, G. Corbelli, A.Dal Borgo and D.G. Lister, *J. Mol. Struct.*, 23, 253 (1974).

133) J. Gerhards, T.-K. Ha and X. Perlia, *Helv. Chim. Acta,* 65, (1), 105 (1983).

134) B.R. Penfold and J.C.B White, *Acta Crystallogr.*, 8, 157 (1955).

135) O. Cantoni, P. Sestili, G. Spadoni, C. Balsamini, L. Cucchiarini and F. Cattabeni, *Biochem. Int.*, 15(2), 329 (1987).

136) M.E. Jacox and F.L. Rook, *J. Phys. Chem.*, 86, 2899 (1982).

137) R.P. Muller and J.R. Huber, *J. Phys. Chem.*, 87, 2460 (1983).

138) R.P. Muller, P. Russeger and J.R. Huber, *Chem. Phys.*, 70, 281 (1982).

139) R.P. Muller and J.R. Huber, in "Reviews of chemical intermediates", Elsevier Science Publishers B.V., Amsterdam, 5, 423 (1984).

140) R.P. Muller, J.R. Huber and H. Hollenstein, *J. Mol. Spectrosc.*, 104, 209 (1984).

141) R.J. Sampson, "The **SURFACE** II graphics system", in "Display and Analysis of Spacial Data", (J.C. Davies and M.J. McCullagh eds.), John Wiley and Sons, London, 244 (1975). The implementation of the SURFACE II system used in this work is by the University of Manchester Regional Computer Centre, (UMRCC).

142) P.G. Mezey, "Analysis of conformational energy hypersurfaces", in "Progress in Theoretical Organic Chemistry", (I.G. Csizmadia ed.), Elsevier, 2, 127 (1977).

143) D. Heidrich and W. Quapp, *Theor. Chim. Acta,* 70, 89 (1986).

144) R.F. Hout Jr., B.A. Levi and W.J. Hehre, *J. Comp. Chem.*, 3 (2), 234 (1982).

145) A review on the choice of basis sets for use in ab-initio frequency calculations can be found in the book by Hehre *et al* (see reference 3).

146) S.F. Boys, *Rev. Mod. Phys.*, **32**, 296, 300 (1960).

147) C.E. Edminston and K. Ruedenberg, *Rev. Mod. Phys.*, **35**, 457 (1963).

148) W. von Niessen, *J. Chem. Phys.*, **56**, 4249 (1972).

149) A.C. Hurley, J.E. Lennard-Jones and J.A. Pople, *Proc. Roy. Soc.* (London) A220, 446 (1953).

150) J.M. Parks and R.G. Parr, *J. Chem. Phys.*, **28**, 335 (1958).

151) A.C. Hurley, "Electron Correlation in small molecules" Academic Press, 1976.

152) D.B. Cook, P.C. Hollis and R. McWeeny, *Mol. Phys.*, **13**(6), 553 (1967).

153) R.E. Howard, M. Levy, H. Shull and S. Hagstrom, *J. Chem. Phys., Phys.*, **66** (11), 5181 (1977).

154) D.M. Silver, *J. Chem. Phys.*, **50**(12), 5109 (1969).

155) I. Röeggen, *Int. J. Quant. Chem.*, **22**, 149 (1982).

156) D.D. Ebbing and R.C. Henderson, *J. Chem. Phys.*, **42**(6), 2225 (1965).

157) P.R. Surján, I. Mayer and I. Lukovitz, *Phys. Rev. A*, **32**(2), 748 (1985).

158) R.A. Poirier and P.R. Surján, *J. Comp. Chem.*, **8**(4), 436 (1987).

159) C.F. Bender, E.R. Davidson and F.D. Peat, *Phys. Rev.*, **174**(1), 75 (1968).

160) I. Röeggen, *J. Chem. Phys.*, **79**(11), 5520 (1983).

161) M. Klessinger and R. McWeeny, *J. Chem. Phys.*, **42** (10), 3353 (1965).

162) I. Röeggen, *Int. J. Quant. Chem.*, **20**, 817 (1981).

163) I. Röeggen, *Int. J. Quant. Chem.*, **19**, 319 (1981).

164) "Molecule" J. Almlöf, USIP Tech. Rept., No. 7429, Stockholm,

Sweden (1974).

165) N.H.F. Beebe and J. Linderberg, *Int. J. Quant. Chem.* **12**, 683 (1977).

166) I. Röeggen and E. Wisløff-Nilssen, *Chem. Phys. Lett.*, **132**, 154 (1987).