Theoretical modelling of transition states for chemical processes: application to enzymic and non-enzymic glycosidic hydrolysis

Barnes, John Ashley

Award date:
1994

Awarding institution:
University of Bath

THEORETICAL MODELLING OF TRANSITION STATES FOR CHEMICAL PROCESSES: APPLICATION TO ENZYMIC AND NON-ENZYMIC GLYCOSIDIC HYDROLYSIS

"It may well wait a century for a reader, as God has waited six thousand years for an observer."

- Johannes Kepler

Submitted by John Ashley Barnes
for the degree of PhD
of the University of Bath
1994

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I wish to acknowledge the financial support of the S.E.R.C and Celltech Research Ltd. Thanks to John Wilkie for software engineering in the early stages of the work, and to Ian Williams and his family for all their help and support.

HELLO Pillan, Jowan, Morwenna and Mum!

Dedicated to Twist

ABSTRACT

Some aspects of glycosidic hydrolysis have been addressed using quantum chemical methods which incorporate solvation effects. This has been done by the exploration of potential energy surfaces (PESs), and comparison of calculated kinetic isotope effects (KIEs) with experimental values. Use has been made of CHARMM, a hybrid QM/MM method which uses the AM1 semiempirical Hamiltonian, combined with the CHARMM force field and the TIP3P model of water, which models the water molecules explicitly.

For non-enzymic hydrolysis of a model for adenosine monophosphate (AMP), the S_N1 and S_N2-like mechanisms are finely balanced in energy, possess similar TS structures, and appear to coexist on the generated PESs. The CHARMM simulation gives calculated KIEs in excellent agreement with experimental values.

A semiempirical QM *in vacuo* study of the S_N1 to S_N2 mechanistic change has been made, using a degenerate series of reactions of H_2O with $CHR'R''OH_2^+$. As the substituents R become more electron-donating, the mechanism switches from S_N2 to S_N1 with no evidence for the two pathways occuring concurrently on the PES, in contrast to the findings for the AMP model.

The mechanisms of the sialidase enzymes from *Salmonella typhimurium* and influenza have been studied. They act on the same substrates by inversion and retention of configuration, respectively, despite a high degree of similarity in their active site structures. The conformational differences on binding leave a water channel into the active site of *S. typhimurium* which is not present on the flu enzyme. This allows water to approach the substrate to give inversion of configuration on the *S. typhimurium* enzyme. The TS structures show that the leaving group for the *S. typhimurium* enzyme may protect the reaction centre from attack by water to give retention of configuration, whereas for the flu enzyme the water can *only* approach to give retention of configuration. These differences are borne out by the shapes of the PESs generated for the enzyme mechanisms. The modelling studies are given support by the good agreement of calculated KIEs to experimental values.

CONTENTS

CHAPTER 1.
PROLEGOMENON

"Year by year we are becoming better equipped to accomplish the things we are striving for. But what are we actually striving for?"

- Bertrand de Jouvenel

1.1 Prolegomenon

This short introduction serves as a guide to why I thought this work should be done, and a brief outline of what you will find in these pages if you read on. Because of the rather broad scope of this work, passing through physical organic chemistry, enzyme kinetics and mechanism, software development and computational chemistry, the structure is possibly a bit unusual. Because the work splits naturally into two sections, I have included a short literature review in each section. Thus, because the first part is concerned with computer simulation of chemical systems, with the underlying progression to the problems of simulating enzyme mechanism, there is a short review of the development of computer models of enzyme mechanism. The second part of the thesis is concerned with the enzymic and non-enzymic mechanism of hydrolysis of sialic acids. Consequently, I have provided a short review of what is known about the mechanism of sialic acid hydrolysis and the mode of action of the sialidase enzymes.

Chapters 5, 6 and 7 each deal with one particular problem. They contain sufficient background information to stand alone as complete sections of work. Despite this, they provide a natural progression of work which leads to the complex balance between competing mechanisms in the organic reactions described. The debate currently surrounding the mechanism of action of sialidase enzymes, which also centres around competing mechanisms not dissimilar to those uncovered in chapters 6 and 7, is therefore seen as an obvious target for subsequent work, especially as little or no theoretical modelling of the mechanisms has apparently been attempted before. Thus the thesis provides a logical progression of work across the boundaries of organic and biological chemistry.

There was an element of software development in this thesis which was not

originally intended, but was necessary for satisfactory completion of the desired calculations. Because I felt that a description of the software development would detract from the flow of the chemistry work which is described, I have used two appendices to describe the important software modifications. Suggestions for further code development are provided in the final chapter of the thesis.

1.2 Why?

Why should we bother to understand chemical processes? This question has many answers, but I think there are two good ones. The first is that the human race has an insatiable appetite for knowledge of all kinds, and it is part of human nature to ask questions and seek answers. The second answer is a more pragmatic one, and is a part of the complex world in which we now live. The ability to understand chemical processes enables us to gain a greater control over our lives and our environment because chemical processes are an essential part of many of our activities. Similarly, biochemical reactions are all around us and inside us and are almost always catalyzed by enzymes. Far less is known about the mechanisms of enzymic reactions, and it is in this area that computer simulations have the potential to contribute a great deal to our understanding of these important processes. The possible applications for techniques able to predict the outcome of enzyme-catalyzed reactions are almost limitless. For example, a detailed knowledge of enzyme reactions could enable new molecules to be designed which will be useful in the pharmaceutical industry as potential drugs.[1] Understanding how enzymes catalyze reactions may enable us to alter their structure to perform specific reactions, making current industrial processes more efficient and less wasteful of resources. The knowledge of how enzymes work may even allow us to design our own catalysts from scratch.

Ever since the first three-dimensional enzyme structures were obtained in

the 1960s[2] it has been a basic challenge for the natural sciences to explain the relationship between enzyme structure and catalytic ability. Pauling's statement that enzymes work by lowering the activation energies of chemical reactions via a stabilization of the transition structure of the reaction relative to the ground state[3] is now widely accepted. It is not easy to see how stabilization is achieved, but in principle many ways can be found to achieve it.[4] Menger[5] has also discussed the importance of enzyme-substrate ground state interactions. Apart from these general statements, the key issue is what the most important contributions are, and how the enzyme structure relates to the catalytic power. Some useful insights into these questions are provided by site-directed mutagenesis of enzyme active sites and other experimental techniques. Unfortunately no current experiment allows the entire energy profile for an enzyme reaction to be established, and it is difficult to experimentally quantify the contributions from different catalytic effects. These effects may include steric strain, electrostatic stabilization and entropy as well as energy contributions associated with each residue of the protein.

This situation presents a challenge to the theoretical chemist to model and explain enzyme catalytic activity. The development of a reliable modelling approach would allow exploration of the details of catalysis that are inaccessible to experiment. It would also be possible to predict the outcome of any protein engineering experiment, thus enabling 'designer catalysts' to be made for specific reactions. The excitement of this particular type of research has resulted in much computational effort in enzyme modelling and simulation. Although computer simulations are nowhere near an adequate substitute for experiments, useful insights are being provided. It is clear that computer modelling is becoming an essential part of the process of interpreting experimental findings about enzyme catalysis.

1.3 What?

The ultimate goal of a computer calculation on reaction mechanisms is the energy profile of the reaction, giving energies and geometries of all stationary points along the reaction coordinate, including the transition state (TS). When ground state and TS structures are available it is possible to calculate rate constants and isotope effects for the reaction. These calculations, when used alongside experimental results, should give useful chemical insights into catalytic mechanisms. It would seem reasonable that modelling of biochemical processes must be begin from an understanding of the basic underlying chemistry. A computational technique which presumes to model enzyme systems should be capable of describing simpler systems. It is therefore useful to start with simple systems and build upon the insights gained from these when interpreting the modelling of more complex systems. This is the way in which the work presented in this thesis will progress.

The TS has been relatively undervalued and the relationship between vibrational force constants and molecular structure not fully exploited in computational studies of reaction mechanism. All too often stationary points are not characterized by force constant calculations, and where experimental kinetic isotope effect (KIE) data are available they are rarely compared to the values obtained *via* calculated TS structures. (The link to KIE data is made because these experimental measurements are powerful probes of TS structure, as will be shown in Chapter 2.) Although work excluding these checks is valuable for qualitative insights, it is clear that no deeper analysis is feasible without appropriate comparison to experimental work. Without comparison, there is no real evidence to support mechanisms postulated from computational work, and they remain firmly in the realm of thought experiments until such a comparison is made. Comparison may of course come from experiments done to establish the

credentials of the theoretical work.

The first aim of this thesis is to explore ways in which equilibrium isotope effects (EIEs) calculated from cutoff chemical structures can be made more reliable. This is important because one of the ways in which theory and experiment complement each other is in the use of calculated EIEs to aid the interpretation of experimental KIEs. This relies quite heavily on the accuracy of cutoff models of molecules in so far as how well calculations using them can predict EIEs. Cutoff models are used because in general interesting chemical reactions involve too many atoms for a computational technique to deal with at the level of theory required. Chapter 5 provides a reevaluation of some work done a few years ago on the acid-catalyzed hydrolysis of methyl glucopyranosides. This involves the calculation of some EIEs, which are generally much simpler to calculate than the KIEs because there is no need to identify a TS structure. An attempt is made to establish the reliability of a semiempirical quantum mechanical (QM) solvation model to provide realistic EIEs using a range of model systems for the methyl glucopyranosides. This serves to establish in principle the usefulness of some of the computational methods used in this thesis and provides an example of the utility of using isotope effect data to probe TS structure.

The second aim of this thesis is to show that the complementary nature of experimental and calculated measures of TS structure, principally KIE values, should be fully exploited wherever possible to gain the deepest insight into the nature of reaction mechanisms for both chemical and biochemical systems. The check that this provides on the computational model is also a valuable means of refining the model for use on systems where no experimental comparison is possible. In such cases, the greater the predictive ability of the theoretical model, the more confidence there is in its application to novel systems. To satisfy this aim, calculations of KIEs from computed structures along alternative

reaction pathways will be compared with experimental values. This will allow a choice to be made regarding the most likely pathway. The type of reaction chosen for this work is the hydrolysis of glycosides, since there is ample experimental KIE data available for both enzymic and non-enzymic hydrolysis. Examples of both types of hydrolysis will be modelled. Chapter 6 deals with the calculation of KIEs for the acid-catalyzed hydrolysis of adenosine monophosphate (AMP), using a more sophisticated chemical system to approximate the AMP than was the case for the methyl glucopyranosides. A range of semiempirical solvation models are used to calculate TS structures for the reaction, and comparison of calculated to experimental KIEs provides a guide to which models are likely to be useful for predictive work in the area of mechanistic chemistry. This chapter also demonstrates the need for good methods to provide realistic geometries for TS structures, from which KIEs can be calculated. The protocol developed for this work is extended in chapters 9 and 10 to the hydrolysis of PNA, a sialic acid derivative, both in solution and by the sialidase enzyme.

A third and equally important aim of this thesis is to establish the limitations in currently available computational chemistry software for the modelling of solution and enzyme reactions. Specifically, the combined quantum and molecular mechanical approach used in this thesis is still under development and there are almost certainly deficiencies in either the model, its implementation or some aspects of the software performance. As this is a fairly new area of computation it is hoped that, by throwing the software in at the deep end, some pointers can be provided to improvements and modifications that will enable future users to more fully exploit the potential of the software.

During the course of the work, some interesting diversions were taken whilst software development held up the main work. These diversions followed from the results of the AMP hydrolysis work described in Chapter 6, and are

incorporated into this thesis because they provide useful contributions to ideas about isotope effects and how reaction mechanisms may change. These diversions appear in Chapter 7 and the Interlude.

1.4 When? Where? Who?

The work in this thesis was carried out between October 1991 and November 1994. I was based at the University of Bath and the computer facilities at the University of Bath and ULCC were used to run the computational chemistry software. You can find my name on the title page.

1.5 References

1. W. G. Richards in Computer-aided Molecular Design, Ed W. G. Richards; IBM Technical Services, London 1993
2. C. F. F. Blake et al, *Nature* , 1965, **206**, 757; *Proc. R. Soc. London* , 1967, **B167**, 378
3. L. Pauling, *Am. Sci.*, 1948, **36**, 51-58
4. M.I. Page in Enzyme Mechanisms, Ed. M. I. Page, A. Williams; RSC London 1987.
5. F. M. Menger, *Biochemistry*, 1992, **31**, 5368-73

CHAPTER 2.
CHEMISTRY, CONCEPTS AND MODELS

"No good model even accounted for all the facts, since some data was bound to be misleading if not plain wrong."

- James Dewey Watson

2.1 Introduction

Chemistry is the study of the properties and reactions of atoms and molecules. The majority of the chemical processes studied are in the condensed phase, where the interaction between a solute and its environment plays a critical role in the outcome of a reaction. The environment may be a solvent, a solid surface or an enzyme. To investigate chemistry we require basic concepts which underpin our understanding of the processes being studied. To use computational simulation methods there must be a model that adequately reproduces the phenomena of interest. This chapter briefly introduces the concepts and models which are used in this thesis.

2.2 The Potential Energy Surface

The potential energy of a system is usually calculated as a function of the positions of the atomic nuclei. With quantum mechanical (QM) calculations as described later in this chapter, this is only possible because the motions of atomic nuclei can be separated from the motion of their electrons. This independence is called the Born-Oppenheimer approximation. Essentially the nuclei of atoms are given fixed positions and the interactions of the electrons with the nuclei and each other are calculated. The total energy is called the potential energy of the system, even though it may also include the kinetic energy of the electrons if the energy is calculated by QM methods. If the nuclei are then moved and the potential energy recalculated then another value is obtained. Repeating this many times gives values for the potential energy of the system with respect to the nuclear coordinates. If a plot of just two nuclear coordinates and the calculated potential energy were drawn, it would be a potential energy surface (PES) like the one in Figure 2.1. In reality a

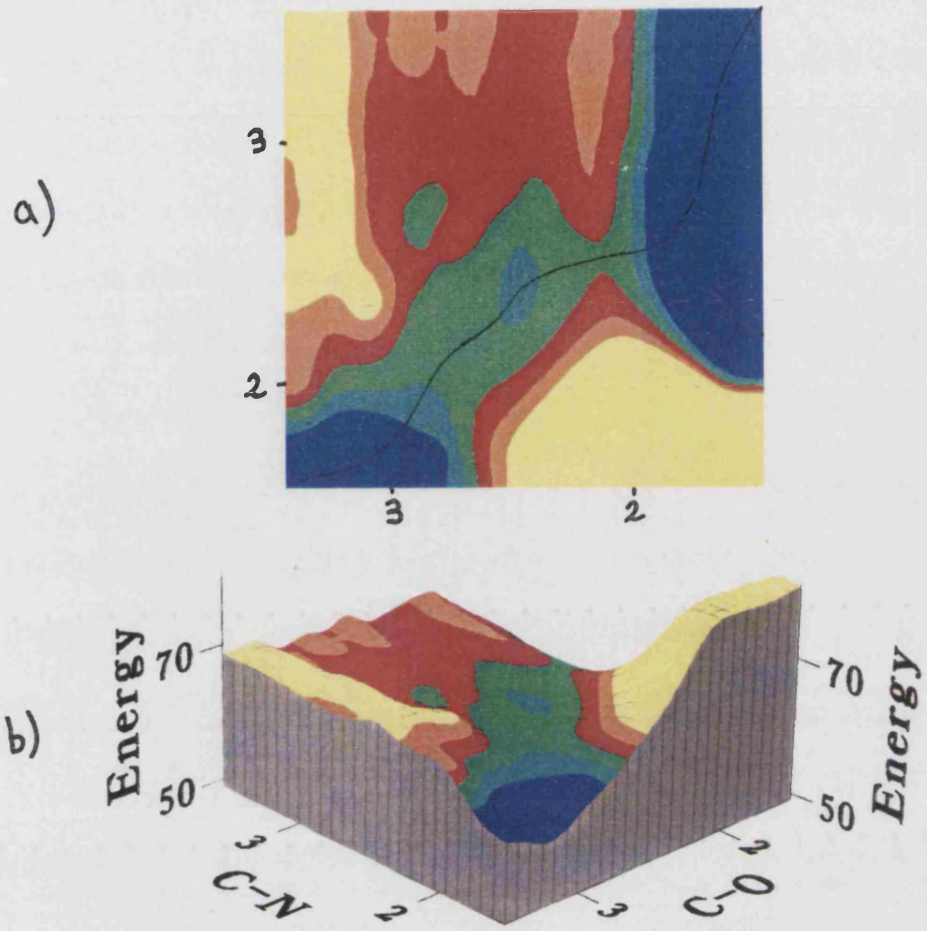


Figure 2.1. Representation of a two-dimensional (a) and three-dimensional (b) potential energy surface for an S_N1-like reaction. The line on (a) is the reaction coordinate whose energy profile is shown in Figure 2.2.

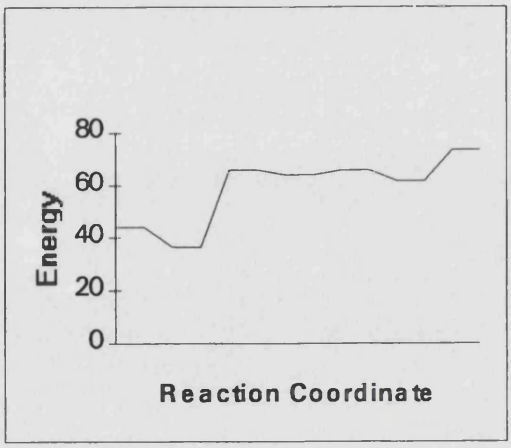


Figure 2.2. The energy profile of the reaction coordinate on Figure 2.1a.

full set of calculations involving movement of all nuclear coordinates in each of the three dimensions available would be required to give a complete multi-dimensional surface. This would be an almost intractable calculation for many-atom systems and would yield a great deal of uninteresting data. With appropriate choice of coordinates diagrams such as Figure 2.1 are useful ways of representing chemical reactions. The idea of using a PES in this way was suggested by Marcelin[1a] in 1915, but it was Eyring and Polanyi[1b] in 1931 who first calculated one.

The PES gives a description of the chemical reaction from beginning to end. By using the contours of the surface to follow the lowest energy path from the reactants to products it is possible to pick out a reaction profile (Figure 2.2). The highest energy point on this profile is associated with the specific structure of the activated complex, and it is this structure which is the goal of calculations to find the transition structure of a reaction. It must be remembered that the PES that is generated for this type of representation is a function of the nuclear coordinates that have been chosen to be explicitly measured, and that the wrong choice of coordinates will give a misleading PES for the reaction.

2.3 The Transition State

The transition state (TS) should really be thought of as a group of species of varying energy around the location of a first order saddle point on the PES which represents the transformation of a chemical system from reactants to products. The actual saddle point is a specific structure which should be called the activated complex or transition structure. Because transition state and transition structure are used interchangeably in normal chemical usage, it is important to clarify the meaning in this thesis. When TS is mentioned in association with structure, this refers to the calculated structure of the activated complex. The use

of TS by itself will refer to the less specific normal usage where the activated complex and diffuse group of species are thought of as a single entity.

The theory of TS stabilisation espoused by Pauling[2] has been a major influence in the study of enzyme mechanism. However, a deeper understanding at the molecular level is desirable, so that the knowledge can be applied to manipulate the reactions that are taking place. This means that an explicit mechanism for the particular reaction is needed, together with structural and energetic information about the intermediates and transition states along the reaction path. It is also desirable to identify the specific interactions responsible for TS stabilisation, since the TS structure is the gateway through which the reaction must pass.

The TS structure is a desirable goal for many reasons. Knowledge of the structure and energy of the TS gives insight into the route taken by a reaction and the ease with which the reaction occurs. Mechanism-led drug design implicitly aims to produce TS analogues which will act as inhibitors of enzyme reactions, and design of synthetic catalysts is aided by a knowledge of the TS of the process. Unfortunately for the experimental chemist, the TS is like the Scarlet Pimpernel: rather elusive! It has become feasible to use femtosecond laser spectroscopy[3] to observe activated complexes for simple gas phase processes, but for most practical purposes TS structure is generally inferred from methods such as experimental kinetic isotope effects (KIEs) and structure-reactivity relationships. The interpretation of these methods is not straightforward. It was Eyring and Polanyi[1b] who first connected the TS to the saddle point of a PES, so the TS is essentially a theoretical concept. The idea of a discrete structure whose properties held the key to the observed reactions of molecules was appealing, and experimental chemists incorporated it into their thinking very quickly. Soon, identifying the mechanism for a reaction became a matter of finding the TSs along a reaction path between reactant and product

states,[4] using information gained from the indirect methods mentioned above. Theoretical determination of TS structures requires the location of a saddle point on a PES. The details of this task will be presented shortly. Theory and experiment can be used in a complementary way to clarify the nature of the TS, particularly with the use of KIEs. These are amenable to calculation as well as experimental evaluation, and when multiple isotopic substitution is done KIEs provide a good test of a proposed TS structure.

2.4 Isotope Effects

The theory of isotope effects is based on the assumption that the PES of a reaction is unchanged by isotopic substitution. This makes isotopic substitution a much subtler tool for investigating reaction mechanisms than the substitution of substituent groups. The magnitude of a KIE is largely determined by the changes in vibrational force constants in going from the reactant state to the transition state. By calculating the structures of reactant and transition states it is therefore possible to calculate KIEs. In a similar manner, an equilibrium isotope effect (EIE) measures changes between two stable states that are at chemical equilibrium. The magnitudes of EIEs can be used in conjunction with KIEs to provide estimates of the position of a TS along the reaction coordinate. An application of this is the work on glucopyranosides described in Chapter 5.

To use KIEs to draw conclusions about reaction mechanisms, it is necessary to know the magnitude of a KIE for a particular type of reaction. This is normally done by building on the knowledge of effects from previously measured values of known mechanisms. Measurable effects are mainly observed when the site of isotopic substitution is near the site of bond rupture. There is then a significant effect on the energy barrier which controls the reaction rate, as will now be explained.

The way in which a KIE arises can be seen from Figure 2.3. The black curve represents a cross-section through a PES following the reaction coordinate of a chemical reaction. The coloured curves represent the energy of a C-H bond which is being monitored through the reaction (a C-D bond when isotopic substitution occurs). The horizontal levels are the lowest vibrational energies of the C-H(D) bond. A bond involving deuterium (D) is lower in energy than the equivalent protium (H) bond, which can be deduced from the period of vibration of the simple harmonic oscillator model,

$$\nu = \left(\frac{1}{2\pi}\right)\sqrt{\frac{F}{M_r}} \qquad (2.1)$$

The force constant F is independent of isotopic substitution, so the expression is mainly determined by the reduced mass M_r. When the atom to which the H(D)

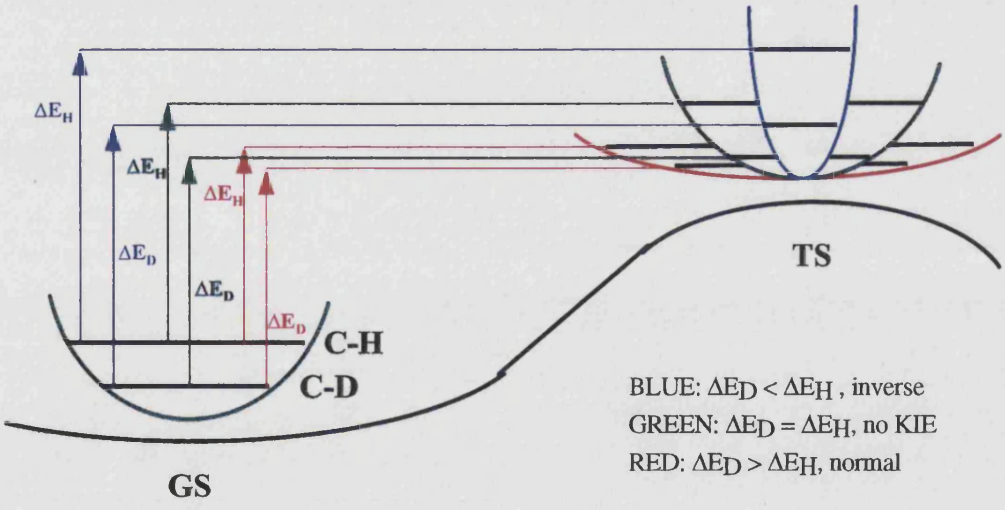


Figure 2.3. The origin of kinetic isotope effects, showing how normal and inverse effects arise. The curve for the C-H(D) bond energy is tighter the stronger the bond is, resulting in larger differences in vibrational energy levels when comparing a C-H and C-D bond.

is attached is much more massive, the reduced mass approximates to the mass of H(D) which means that the vibrational frequency of a bond involving H is about 1.4 times that of the equivalent bond involving D.

As the system moves from the ground state to the transition state, it can be seen that the change in energy for the two isotopic reactions can be different, depending upon the differences in zero point energy. As a rough guide, the weaker the bond in the TS compared to the ground state the larger the difference in activation energy between the two isotopic reactions, $\Delta E_D - \Delta E_H$. This leads to a positive or normal KIE and the reaction of the system with the heavier isotope proceeds more slowly. In the extreme case that the bond is completely dissociated in the TS then the isotope effect is a maximum. If the bond in the TS is stronger than the ground state then $\Delta E_D - \Delta E_H$ is smaller or even negative. When $\Delta E_D - \Delta E_H$ is negative this gives rise to an inverse KIE and the reaction of the system with the heavier isotope proceeds more quickly.

The TS structure differs from the ground state in that one of the vibrational modes is the reaction coordinate and can be thought of as being transformed into a translational motion. This mode has a negative rather than positive restoring force constant and, as can be seen from equation 2.1, leads to an expression containing the square root of -1. Therefore the transition frequency is described as imaginary.

2.5 Enzyme Catalysis and Transition State Theory

The incredible rate accelerations of enzymes have fascinated scientists ever since enzymic activity was first observed.[5] Over twenty hypotheses for enzyme catalysis have been noted,[6] but as mentioned earlier Pauling's view is now generally accepted. The basic idea is simple and results from a combination of

two principles of physical chemistry: absolute reaction rate theory and thermodynamic cycles. The enzyme is a flexible template which has evolved to be complementary to the TS geometry of a substrate rather than its ground state. Therefore an enzyme binds the TS more strongly than it binds the ground state.

2.5.1 A Potted History of Enzyme Catalysis

Modern theories of enzyme catalysis began with Haldane's work entitled 'Enzymes'. He introduced the idea that an enzyme-substrate (ES) complex requires an activation energy before reacting, and suggested that Fischer's famous lock-and key simile[7] be reworded to "the key does not fit the lock quite perfectly but exercises a certain strain on it".[8] Shortly after this came Eyring and Polanyi's work and Eyring began the development of absolute reaction rate theory, or transition state theory as it is commonly called.[9] This approach was based on treating the TS complex as if it were in equilibrium with the reactants and laid the groundwork for Pauling's contribution.

An expression relating catalytic rate acceleration and the relative binding strength between ground and transition states was given by Kurz,[10] but it had no specific mention of enzymes. He had combined Eyring's work with a thermodynamic cycle, leading to a quantitative description of Pauling's statement. In 1966, Jencks first postulated the existence of TS analogue molecules as inhibitors of enzyme catalysis and gave several possible examples from the literature.[11] From this point, transition state theory began to have more impact on enzyme studies. Wolfenden[12] emphasized its broad general applicability and argued that TS analogues should bind more strongly than natural substrates. Many examples were given, and he proposed that the analogues could give important clues to catalytic mechanisms.

With the X-ray structure of hen egg-white lysozyme,[13] the complementary nature of the active site to transition state geometry became virtually visible.

Model building studies[14] led to the conclusion that a sugar residue would only be strongly bound in the half-chair conformation, which is precisely the conformation expected for a transition state resembling a glycosyl oxocarbenium ion. The few years after this saw many X-ray structures reported, but the concepts of strain and distortion rather than transition state complementarity[15] were used to explain the enzyme action. Several serine proteases were amongst this vanguard of X-ray structures, but it was not until 1977 that the common feature of a number of potent inhibitors was noted: they all resemble the expected tetrahedral transition state for substrate hydrolysis and they all bind in the same complementary oxyanion pocket.[16] Even though the serine proteases are amongst the most widely studied class of enzymes, there is still a great spread of opinion about the role of many structural features in the catalysis.[17] This shows the difficulties of interpreting the functions of enzyme structural features and emphasizes the challenge that theoretical models face.

2.5.2 Two Minute Transition State Theory

Transition state theory is not exact but is based on some assumptions and approximations.[18] It has been shown to work very well during its six decades of use and is widely accepted as conceptually accurate. The basic assumptions are that the rate is controlled by the decomposition of an activated complex, and that the system can be treated as if this complex were in equilibrium with the reactants. The resulting fundamental equation is

$$k = qvK' \qquad (2.2)$$

where k is the observed rate constant, q is the transmission coefficient, v is the

frequency of the oscillation of the complex along the reaction coordinate and K' is the equilibrium constant for formation of the activated complex from reactants.[9b] The quantity q can for this description include all correction factors such as tunneling, barrier recrossing and so forth. In general q is thought to fall in the range of 0.1 to 1.0 for reactions in solution at room temperature.

The equilibrium constant K' can be written in terms of partition functions, and a factor corresponding to the unique reaction coordinate mode of vibration, v, can be approximated as

$$\frac{k_b T}{hv} \tag{2.3}$$

where k_b is Boltzmann's constant, T the absolute temperature and h is Planck's constant. This expression is valid so long as it evaluates to a number much greater than 1. The v from equations 2 and 3 now cancel to give

$$k = q\frac{k_b T}{h} K'' \tag{2.4}$$

K'' is a pseudo-equilibrium constant as it now contains all transition state vibrational modes except for the transition vector.

2.5.3 Application of Transition State Theory to Enzyme Catalysis

Comparing the reaction for a single substrate enzyme with the equivalent solution reaction in the absence of enzyme needs careful consideration. To obtain the greatest possible difference requires that the enzyme reaction be measured under saturating conditions to give the optimal rate for the enzyme (see section 2.6). The equivalent non-enzymic reaction rate is then measured using the lowest concentration of substrate required to give rise to saturation of the enzyme. Alternatively, rate can be measured at the substrate concentration equivalent to the Michaelis constant K_m (see section 2.6) if this is known.

TS theory can be applied to enzymes by invoking the use of thermodynamic cycles[10,12] as shown in figure 2.4. Using equation 2.2,

$$\frac{k_e}{k_n} = \frac{q_e v_e K'_e}{q_n v_n K'_n} \qquad (2.5)$$

The e and n subscripts refer to enzymic and non-enzymic reactions. This equation can be written as

$$\frac{k_e}{k_n} = \frac{q_e v_e K_S}{q_n v_n K_T} \qquad (2.6)$$

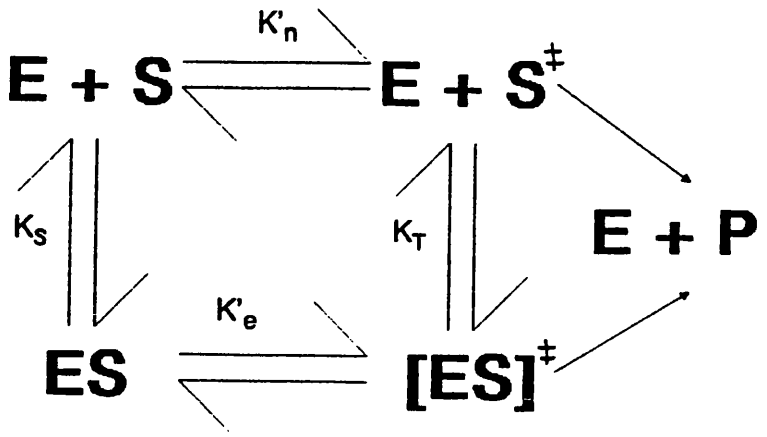


Figure 2.4. A thermodynamic cycle used to obtain a rate constant ratio between an enzymic reaction and its non-enzymic equivalent. E is enzyme; S is substrate; P is product; ‡ denotes transition state. The upper path is the non-enzymic pathway.

by noting from the thermodynamic cycle that the ratio of transition state formation constants K'_e and K'_n is equal to the ratio of the substrate and transition state dissociation constants K_S and K_T.

This ratio can be very large and values of 10^{10} to 10^{14} are common between the enzymic and non-enzymic reaction. From the assumptions in the transition state theory, $q_e\nu_e/q_n\nu_n$ cannot be as large as this[15], although no experimental evidence is available for this statement. The transition vibration is highly unlikely to vary by a significant factor, given that it is essentially one vibration of the molecular system, and as q is in the range 0.1 to 1 this contributes at most one order of magnitude to the ratio value. Therefore a working approximation is

$$\frac{k_e}{k_n} \approx \frac{K_S}{K_T} \qquad (2.7)$$

This equation can be interpreted as saying that the enzyme binds the transition state much more strongly than the substrate in its ground state, by a factor equivalent to the rate acceleration.

2.6 Single Substrate Enzyme Kinetics

For a simple first order reaction S $\rightarrow$ P the rate is given by k[S], so the rate is proportional to the concentration of S. The plot of initial rate against concentration of S would be a straight line of slope k. This relationship is different when enzyme catalysis is involved, as Figure 2.5 shows. The significance of this initial rate pattern for enzymes was first realised by Brown[19] in 1902. He suggested that the substrate forms a complex with the enzyme and it is only the breakdown of the complex to products, also freeing the enzyme, that is rate determining. The reaction can be represented as shown in Figure 2.5 and by Brown's reasoning the initial rate is proportional to the concentration of the enzyme-substrate complex [ES]. At low concentrations of S compared to total enzyme, [ES] is proportional to [S] so the rate is proportional to [S]. When [S] exceeds total enzyme concentration all enzyme is present as ES and the enzyme is said to be saturated with S. As fast as ES breaks down to release E and either S or P, more S combines with the enzyme to re-form the ES complex. Since the

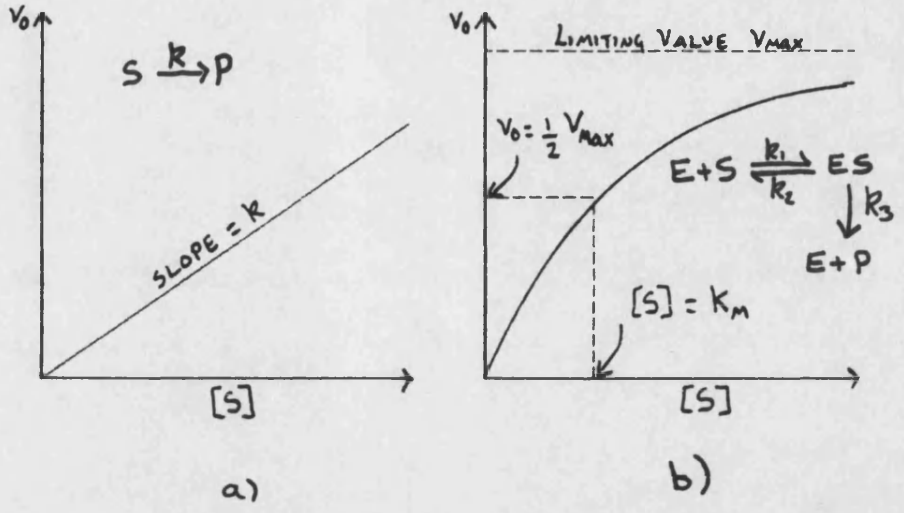


Figure 2.5. Dependence of initial reaction rate on substrate concentration for a) a simple first order reaction and b) a general single substrate enzyme-catalyzed reaction.

total enzyme concentration, $[E]_{tot}$, is fixed, the maximum rate cannot exceed $k_3[E]_{tot}$ and so the rate reaches a plateau.

Michaelis and Menten[20] gave a simple treatment of this one-substrate scheme, based upon the following assumptions:

1. The enzyme concentration is small compared to S so that formation of ES does not significantly reduce [S].
2. The concentration of P is effectively zero. This is the initial rate assumption and implies that no P is present at the start and that insufficient P is present during initial rate measurement to cause the reverse reaction.
3. Product release is fast, but much slower than the release of S from ES, so E and ES can be considered to be at equilibrium.

Two expressions can be written down from these assumptions,

$$k_1[E][S] = k_2[ES] \tag{2.8}$$

and

$$[E] + [ES] = [E]_{tot} \tag{2.9}$$

where e is the total enzyme concentration. Therefore,

$$[E]_{tot} = [ES]\frac{k_2}{k_1[S]} + [ES] = [ES]\left(1 + \frac{k_2}{k_1[S]}\right) \tag{2.10}$$

The rate is $k_3[ES]$, so the rate expression can be written as

$$v = \frac{k_3[E]_{tot}}{1 + \frac{k_2}{k_1[S]}} = \frac{V_{max}[S]}{[S] + K_m} \tag{2.11}$$

In equation 2.11 K_m is the ratio of k_2 to k_1 and V_{max} is k_3 $[E]_{tot}$, the maximum rate.

The constant K_m is the Michaelis constant and is usually interpreted as an

empirical value that gives the substrate concentration which yields $V_{max}/2$ under defined experimental conditions. It gives a useful measure of the range over which an enzyme responds to changes in [S]. For example from equation 2.11 it is easily seen that when [S] is four times K_m, the rate is 80% of V_{max}.

2.7 Chemical Models

Computational chemistry attempts to simulate chemical processes, and therefore there must be a model for the chemistry that can adequately reproduce experimental phenomena. There are two broad categories of computational chemistry model: molecular mechanics (MM) and electronic structure theory[21] (EST). They are both capable of computing energies and structures of molecules as well as the molecular vibrations. They are however fundamentally different in their approach, as the next three sections will show.

2.8 Electronic Structure Theory (Without much Maths!)

The basis of EST is quantum mechanics. Theoretical methods can be divided into two classes, ab initio and semiempirical, depending on whether or not they contain parameters whose values are chosen to match calculated results to experiments. In a sense the "from first principles" claim of ab initio work is misleading because many basis sets used in this work contain parameters chosen to optimize results for a set of molecules and so are themselves partly empirical.

Quantum mechanics seeks solutions to the Schrodinger equation

$$H\Psi = E\Psi \tag{2.12}$$

which was first formulated in 1926.[22] H represents the Hamiltonian operator, Ψ the wave function and E the energy of the electronic system. Ψ and E are often referred to as eigenfunctions and eigenvalues. The wavefunction is a function of the spatial coordinates and of time. For systems which remain in one state, for example stay in the ground state, the time dependency is not of much interest, since in the absence of any other interactions the ground state does not change with time. This means that in the vast majority of cases, the Schrodinger equation is solved for time-independent wave functions,and takes the form

$$\left(-\frac{h^2}{8\pi^2 m}\nabla^2 + V\right)\Psi = E\Psi \tag{2.13}$$

where the Laplacian operator, a differential operator over all space, is

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.14}$$

The usual interpretation of Ψ is to say that the integral of Ψ^2 over a volume of space dV represents the probability of finding an electron in the volume dV. An alternative interpretation is to say that the same integral represents the electron density in dV.

The first serious methods for solving this equation for chemical systems by use of computers were developed by Boys[23] using the Roothan-Hall[24] self consistent field (SCF) linear combination of atomic orbitals (LCAO) molecular orbital[25] method. Boys introduced most of the basic techniques used in ab initio calculations, including the use of Gaussian orbitals[26] to simplify integral calculations and configuration interaction to allow for electron correlation[27].

The most common way to implement QM calculations is by the molecular orbital (MO) theory. According to this model, the electrons in a molecule move independently from each other. Each electron is described by a MO which is a LCAO of the atoms of the molecule. It is assumed that each MO can house two electrons provided that the electrons have opposite spins. This model can be evaluated by using a basis set of functions to represent the atomic orbitals (AOs) of the atoms in a molecule and calculating its energy. The energy is a time-average of the kinetic energy of the whole system and the electrostatic energies of interaction between pairs of particles, i.e electron-electron repulsions (EE), electron-nuclear attractions (EN) and nuclear-nuclear repulsions (NN). The restriction of two electrons per orbital is met by representing each MO by a pair of spin orbitals φs, where s can take the value of α or β, corresponding to the alternative spin states of an electron. The total wave function of the molecule is the determinant of all spin orbitals $\varphi\alpha$ and $\varphi\beta$ for each MO (see Slater Determinant box). The coefficients of the AOs in each MO are chosen to minimize the electronic energy of the molecule.

The Slater Determinant

The wavefunction is the determinant rather than the product of the spin orbitals because of the Pauli Principle, which is discussed in section 2.8.2. One consequence of the Pauli Principle is that the wavefunction must change sign whenever the coordinates of two electrons are interchanged. A simple product does not have this property. When the wavefunction is written as a linear combination of the electron spin orbitals then it can be expressed as a determinant. The change of sign of the wavefunction then follows automatically from the property of determinants which causes the determinant to change sign if any pair of rows or columns are swapped.

2.8.1 Calculation of Molecular Energy

For all but the simplest of molecules, MO theory requires a number of approximations to be made.[28]

1. Time independence. The Hamiltonian operator H corresponds to a stationary state and depends only on spatial coordinates.
2. Neglect of relativity. The electrons are assumed to have the same mass regardless of their velocity.
3. Born-Oppenheimer approximation. The nuclear coordinates are assumed to remain static since the greater mass of the nuclei compared to the electrons means that electronic motion is much faster.
4. The orbital approximation. The real wavefunction is represented as a combination of orbitals each dependent on the position of only one electron.

5. Linear Combination of Atomic Orbitals (LCAO). The molecular orbitals are described by a combination of atomic orbitals centred on each of the atoms in the molecule.

As already stated, if an electron occupies an MO, the probability that it will be at a given point in space is proportional to Ψ^2. Time averaged electrostatic interactions involving the electron can be evaluated by using this probability as an electron density at a point in space. Since each MO is also a LCAO, Ψ^2 can be expressed as a sum of terms which each contain a product ϕ^2 or $\phi_i\phi_j$ of the AOs which comprise the MOs. For example, the MO $\phi_a + \phi_b$ would, when multiplied by itself, be the sum of terms $\phi_a^2 + 2\phi_a\phi_b + \phi_b^2$. Thus an EN attraction is a double sum of terms each involving two AOs and corresponding to the attraction between a given nucleus and one of the parts into which the electron cloud has been split.

EE repulsions between electrons occupying different MOs are calculated as a quadruple sum of terms. These electron repulsion integrals are represented as $(ij|kl)$ for convenience, with $\phi_i\phi_j$ being the terms for one electron and $\phi_k\phi_l$ being the terms for the second electron.

The remaining electronic terms correspond to exchange correlation. These are necessary to correct for the probability that two electrons would occupy similar positions in space if the Pauli Principle was inoperative (see section 2.8.2). The correction is equal to the repulsion between two identical overlap clouds of negative charge. If the MOs do not overlap then there is no cloud overlap and the correction vanishes.

The NN terms are the summation of the potential energy for interactions

between point charges. Each nucleus is assigned an effective charge based on the screening effect of its valence electrons for semiempirical methods. *Ab initio* calculations use the actual nuclear charge. There is no kinetic energy term for the nuclei because of the Born-Oppenheimer approximation. The NN terms are simply added to the expectation value of the electronic Hamiltonian.

The above terms are evaluated to give the energy and wavefunction of a molecular system using the Hartree-Fock[29] SCF equation,

$$FC = SCe \qquad (2.15)$$

where F is the Fock matrix of elements representing the kinetic energy of the electrons, EE repulsion integrals and EN attractions. C is the matrix of LCAO coefficients, S is the matrix of overlap integrals between the atomic orbitals and e is the eigenvalue matrix. The goal of an MO calculation is to find the values of the coefficients in matrix C that produce the lowest electronic energy. An initial guess of C is made and then the coefficients are varied in a series of iterations until the C obtained from diagonalizing the Fock matrix is sufficiently close to the C used to formulate the Fock matrix. At this stage the average electrical field produced by all other electrons on a given electron is self consistent, and the energy is (usually) a minimum. The best possible MO function will yield a value known as the Hartree-Fock limit, but in practice the calculated value is always higher than this. Even the Hartree-Fock limit is higher than the real energy, because the SCF method does not allow for electron correlation effects, only allowing each electron orbital to distort in response to an average electric field, instead of in response to each individual electron. The errors in high level ab initio energies are caused almost entirely by inadequate treatment of coulombic correlation.[30]

2.8.2 Electron Correlation

The simple MO model neglects electron correlation, which is the tendency for electrons to remain as far apart as possible. There are two elements to this: exchange correlation and coulombic correlation.

Exchange correlation arises from the Pauli Exclusion Principle, which states that two electrons with the same spin cannot occupy the same point in space simultaneously. The wave function for a molecule must vanish if this ever occurs, and the wave function must be small when the coordinates and spin are nearly identical. The probability of finding two electrons close together is therefore smaller than it would be if the exclusion principle did not operate. Exchange correlation corrects for this by reducing the EE repulsions and leading to a decrease in energy.

As an example, take a helium atom where the electrons occupy two spin orbitals $\varphi\alpha$ and $\varphi\beta$. The atom can be represented as $[\varphi\alpha(1)][\varphi\beta(2)]$ or $[\varphi\alpha(2)][\varphi\beta(1)]$, the numbers representing each electron. Individually, neither state disappears if the electronic coordinates are the same so they violate the exclusion principle. However the combination of these two states $\{[\varphi\alpha(1)][\varphi\beta(2)] - [\varphi\alpha(2)][\varphi\beta(1)]\}$ does disappear because both terms would be identical if the coordinates were swapped. This approach can be extended easily to molecules in which electrons are paired, and leads to only a small time penalty upon calculation.

Coulombic correlation is a very much greater problem. It arises from the neglect of local distortion of the electronic orbitals in favour of the averaged field that is part of the Hartree-Fock method. The result of this neglect is to cause the calculated potential energy curve to lie above the true curve. The true shape of the curve is approximated, except that it is too narrow in the region of the minimum. The minimum itself occurs at about the same position as the true minimum, which means that equilibrium geometries are predicted well.[31]

Methods used to estimate this correlation energy are very inefficient and require a lot of extra computation. The fraction of energy that can be recovered is also dependent on the basis set used and decreases with increasing molecular size, rarely exceeding 75% of the true correlation energy. Since coulombic correlation energies of organic molecules are usually greater than their heats of atomization, the residual errors are enormous in absolute chemical terms. Even in the case of benzene errors in the correlation energy amount to hundreds of kcal mol^{-1}.[30] Since a 1 kcal mol^{-1} error corresponds to 0.001% of the total energy, the percentage accuracy needs to be extremely large to reproduce the absolute molecular energy of even this modestly sized molecule. Fortunately, chemical behaviour depends on differences between pairs of closely related systems, so chemical behaviour may be predicted accurately due to cancellation of errors!

2.9 Semiempirical Methods

Using the above methods with even the simplest ab initio basis sets requires too much time for chemical systems of any reasonable size. In the semiempirical methods, some simplifying assumptions are introduced which reduce the computational effort. At the same time, parameters are introduced to help offset the loss of accuracy. These parameters are chosen to optimize the prediction of heats of formation of the set of test molecules which are used in the parameterization process. The calculations used in this thesis are done with the AM1 semiempirical model, described in section 2.9.3.

2.9.1 The NDDO Approximation

Perhaps the most widely used approximation is the neglect of diatomic differential overlap (NDDO). If atomic orbitals of different atoms did not overlap with each other, overlap integrals would vanish and so would the EE integral

$(ij|kl)$ if the orbital combinations occurred in either of the pairs ij or kl. The number of EE integrals to be calculated would be far less. This is exactly the approximation introduced by Pople.[32] The neglect applies only to overlap integrals and EE repulsion integrals, not to the EN integrals. The sum of terms involving interaction of an overlap integral with a nucleus is called a resonance integral.

2.9.2 Core Approximation

It is assumed that the valence electrons move in a field generated by a fixed core. This core is made from the atomic nuclei and a cloud of negative charge which represents the averaged motion of the inner electrons. This further reduces the EE integrals that are required.

2.9.3 The AM1 model[33]

The quantities that must be calculated using the above approximations are the one-centre and two-centre EE integrals, one-centre electron-core (EC) attractions, the resonance integrals and the core-core (CC) repulsions. The one-centre integrals are treated as parameters. The sum of terms comprising a resonance integral is estimated using the parametric expression $(B_i + B_j)S_{ij}$ where the B_i and B_j are parameters and S is the overlap integral between two AOs.

The two-centre EE integrals are also estimated from a parametric expression[34] which allows for electron correlation. The remaining EC terms, which correspond to the attraction of an electron occupying the AO of one atom and the core of a different atom, are estimated by assuming the field of the core to be -Z times that of a corresponding valence shell s electron, where Z is the net

charge of the core. A repulsion term is also required for the CC term to correct for the fact that the net EE and EC terms are similar if the two atoms involved are neutral. AM1 uses a gaussian-type internuclear distance function for this purpose which introduces additional parameters.

2.9.4 Parameterization

The values for all the parameters mentioned above are chosen by a least squares fit to experimental values for selected properties of a chosen basis set of molecules.[35] The parameters are then tested by carrying out calculations for various properties of a large number of molecules, including some not used in the parameterization. The errors for individual molecules are assessed on the basis of their chemical significance. The parameters should therefore vary in a regular way across the periodic table, and calculated atomic charges should correspond to current ideas of electronegativity. The best forms for the parametric functions are found by trial and error.

2.10 Molecular Mechanics

Molecular mechanics (MM) is based on a classical potential energy,

$$V = V_b + V_a + V_d + V_\omega + V_{nb} \qquad (2.16)$$

where there are contributions towards the potential energy from bonding, angle, dihedral, non-bonding and hydrogen-bonded energy terms. The precise value of these contributions depends upon the particular force field which has been used. In the CHARMM force field used for the majority of MM calculations in this thesis, the contributions are evaluated as empirical functions of the

following types,

Bond: $$V_b = \sum k_r(r - r_0)^2 \quad (2.17)$$

Bond angle: $$V_a = \sum k_\theta(\theta-\theta_0)^2 \quad (2.18)$$

Dihedral: $$V_d = \sum |k_\phi| - k_\phi \cos(n\phi) \quad (2.19)$$

Improper torsion: $$V_\omega = \sum k_\omega(\omega - \omega_0)^2 \quad (2.20)$$

Van der Waals: $$V_{nb} = \sum \left\{ \frac{A}{r^{12}} - \frac{B}{r^6} \right\} \quad (2.21)$$

The first two terms represent bond and angle deformations which in most cases at ordinary temperatures and in the absence of chemical reactions are small enough for the harmonic approximation to apply. The dihedral potential is a four atom term based on the dihedral about an axis defined by the middle pair of atoms. There may be several contributions to this term with different k_ϕ and periodicities for a given set of four atoms. The improper torsion maintains planarity about planar atoms and provides a better force field near minimum energy geometries, which is important for molecular dynamics simulations. The various force constants k_x are obtained from fitting to vibrational data or using

existing literature values. The geometric constants r_0, θ_0, n and ω_0 are derived from crystallographic data using the average bond length, angle and so forth as appropriate. In addition, the CHARMM force field has a comprehensive set of electrostatic terms which incorporate hydrogen bonding interactions. These may be evaluated using individual atoms or groups of atoms and have a choice of switching functions and distance dependencies. The choice of these options will be dealt with in Chapter 4.

For a given geometry of a molecule, the contributions to the potential energy are evaluated, and in a series of iterations the energy is minimized by one of a number of methods into a local minimum on the PES. Molecular mechanics is suitable for treatment of molecules consisting of up to several thousand atoms. With suitable parameterization, it is also suitable for the study of solute-solvent interactions and can be used to study condensed phase systems. The technique is fast and extremely accurate for molecules within the scope of the force field parameterization used, but should be used with extreme caution in the investigation of novel structures for which there is no parameter set available. It is the method of choice when dealing with large systems such as enzymes or other biomolecules when bond making and breaking events are not being considered.

2.11 Geometry Optimization and Stationary Points

A major use of both EST and MM calculations is to find a minimum energy conformation of a molecular system. The conformation of a system is related to the wavefunction in EST calculations and to the empirical force field in MM calculations. In both cases, if the forces acting on an isolated molecule are reduced to zero, the molecule will be at a stationary point. The method of Pulay[36]

allows the forces acting on a molecule to be calculated analytically. The forces are given by the negative of the first derivative of the energy with respect to nuclear coordinates. The geometry of the molecule is altered using one of several minimization techniques to reduce these derivatives, leading to a stationary point on the PES of the molecule.

A stationary point on a molecular PES can be amongst other things a minimum or a saddle point. A minimum is the bottom of a valley on the PES, which means that the molecular potential energy will initially increase however the molecule is displaced. It is at a minimum in all dimensions and is in a locally stable conformation. A saddle point is a minimum with respect to some dimensions of the PES and a maximum in one or more others. Of special interest is the TS structure which links two minimum energy structures. The TS structure is a maximum in only one dimension and is a first order saddle point. To identify which type of point has been found, it is necessary to do a frequency calculation. Molecular vibrational frequencies are proportional to the square root of the second derivatives of the energy with respect to the normal coordinates. If the structure has all real frequencies it is at a minimum energy conformation. If it has exactly one imaginary frequency it is a transition state.

Gradient methods for optimization are based on two techniques, the steepest descent and conjugate gradient methods. They both use a first order Taylor expansion of the energy,

$$E(x + \Delta x) = E(x) + g(\Delta x) \qquad (2.22)$$

Steepest descent methods use the fact that the greatest change in E is obtained when Δx is in the negative direction of the gradient, so Δx is found by

$$x_k = x_{k-1} + \lambda_k g_k \qquad (2.23)$$

λ_k is the step size, which is reduced if at any stage the energy increases. This method is good for rapidly reducing a large gradient but has very poor convergence. The conjugate gradient method converges much more rapidly as it can locate the minimum energy along a given displacement vector. This means that for an n-dimensional surface it will converge within n steps.

2.12 Molecular Dynamics

Energy minimization is useful for determining molecular structure but gives no information about the behaviour of a molecular system over a period of time. Molecular dynamics (MD) is a deterministic simulation technique which provides this information. It was originally used for fluids, then bulk solvents and large biomolecules. The atoms move by the laws of Newtonian mechanics, and a MM force field is used to calculate the potential energy at each step. The basic equations are

$$x_i(t + \Delta t) = x_i(t) + v_i(t)\Delta t \qquad (2.24)$$

$$v_i(t + \Delta t) = v_i(t) + \frac{f_i}{m_i}(t)\Delta t \qquad (2.25)$$

where Δt is an infinitely small time step, x_i is the position of the i^{th} atom, v_i is

the velocity of the i^{th} atom, f_i is the force acting on the i^{th} atom and m_i is the mass of the i^{th} atom. These equations assume that the force is constant throughout the time step. This becomes less valid as the time step increases. To overcome this, a simple and accurate leap-frog algorithm was developed by Verlet[37] which updates position and velocity alternately,

$$x_i(t + \Delta t) = x_i(t) + v_i\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{2.26}$$

$$v_i\left(t + \frac{\Delta t}{2}\right) = v_i\left(t - \frac{\Delta t}{2}\right) + \frac{f_i(t)}{m_i}\Delta t \tag{2.27}$$

The time step Δt must be much smaller than the fastest vibration in the molecular system. Since vibrational frequency is related to the force constant, which is highest for bond stretching, it is common to fix bond lengths in MD simulations to allow for a larger time step. The SHAKE algorithm is used by the CHARMM program to fix the bond lengths which involve protons, since these bonds have the highest frequencies.

The initial positions for a MD calculation can be from an approximate arrangement of the atoms, or more usefully from an energy minimization calculation. The initial velocities are assigned randomly to give a Boltzmann distribution of energies, and the desired temperature is attained over a few steps. The random velocity assignment means that the initial energy distribution is

usually incorrect, so a period of equilibration is necessary before any useful information can be obtained. The length of equilibration depends on the system under simulation.

2.13 References

1. a) Marcelin, *Ann. Phys.*,1915, **3**, 158; b) H. Eyring & M. Polanyi, *Z. Physik. Chem. B* , 1931, **12**, 279
2. L. Pauling *Am. Sci.* , 1948, **36**, 51-58
3. *Faraday Disc. Chem. Soc.* , 1991, **91**
4. I. H. Williams, *Chem. Soc. Rev.* , 1993, 277-83
5. Possibly L. Planch in 1810 - see Enzymes by J. B. S. Haldane, MIT Press 1965 reprint, 8-9
6. M. I. Page in Enzyme Mechanisms, Ed M. I. Page & A. Williams, RSC London 1987, 1-13
7. E. Fischer, *Ber.* , 1894, **27**, 3479
8. Enzymes by J. B. S. Haldane, MIT Press 1965 reprint, 182
9. a) H. Eyring, *J. Chem. Phys.* , 1935, **3**, 107; b) The Theory of Rate Processes by S. Glasstone, K. J. Laidler & H. Eyring, McGraw-Hill NY 1941, 184-91
10. J. L. Kurz, J. *Am. Chem. Soc.*, 1963, **85**, 987
11. W. P. Jencks in Current Aspects of Biochemical Energetics, Ed. E. P. Kennedy, Academic Press NY 1966, 273-98
12. R. Wolfenden, *Acc. Chem. Res.*, 1972, 5, 10
13. C. F. F. Blake et al, *Nature*, 1965, **206**, 757; *Proc. R. Soc. London* , 1967, **B167**, 378
14. D. C. Phillips, P*roc. Nat. Acad. Sci. USA*, 1967, **57**, 484
15. J. Kraut, *Science*, 1988, **242**, 533-40
16. J. Kraut, *Ann. Rev. Biochem.*, 1977, **46**, 331

17. R. L. Schowen in Molecular Structure & Energetics, Vol. 9, Ed. J. F. Liebman & A. Greenberg, VCH 1989

18. Physical Chemistry by W. J. Moore, 4th edition, Longman London 1978, 381

19. A. J. Brown, *J. Chem. Soc.*, 1902, **81**, 373

20. L. Michaelis & M. L. Menten, *Biochem. Zeitschr.*, 1913, **49**, 333

21. Exploring Chemistry with Electronic Structure Methods by J. B. Foresman & A. Frisch, Gaussian Inc. 1993, 3

22. Physical Chemistry by W. J. Moore, 4th edition, Longman London 1978, 601

23. a) S.F. Boys, *Proc. R. Soc. London Ser A*, 1950, **200**, 543; b) S.F. Boys, C.M. Reeves & I. Shavitt, *Nature* , 1956, **178**, 1207

24. a) C.C.J. Roothan, *Rev. Mod. Phys.*, 1951, **23**, 69

b) G.G.Hall, *Proc. R. Soc. London Ser A*, 1951, **205**, 541

25. Linear Combination of Atomic Orbitals Molecular Orbital method.

$U_i = \sum C_{ij}\phi_j$, where U_i is the i^{th} molecular orbital, ϕ_j is the j^{th} atomic orbital and C_{ij} is the coefficient of the j^{th} atomic orbital contributing to the i^{th} molecular orbital.

26. S. F. Boys & G. B. Cook, *Rev. Mod. Phys* ., 1960, **30**, 285

27. J. M. Foster & S. F. Boys, *Rev. Mod. Phys.*, 1960, **30**, 300

28. Reviews in Computational Chemistry, Ed. Lapkowitz & Boyd, Pub. VCH 1990, p326.

29. The two men who made the solution of the Schrodinger equation a realistic proposition, based on the variational method in quantum mechanics. See Quantum Chemistry, I.N. Levine, 3rd ed. 1983, p172-192

30. M. J. S. Dewar, Caoxian Jie & Jianguo Yu, *Tetrahedron*, 1993, **49**, 5003

31. Correlation energy entry in Quanta: A handbook of concepts by P. W. Atkins, Clarendon Press, Oxford, 1979

32. Approximate Molecular Orbital Theory by J. A. Pople & D. L. Beveridge, McGraw-Hill NY 1970

33. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy & J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902

34. M. J. S. Dewar & W. J. Thiel, *Theor. Chim. Acta.*, 1979, **46**, 89

35. This paragraph appears in ref 30.

36. P.Pulay, *Mol. Phys.* , 1969, **17**, 197

37. L. Verlet, *Phys. Rev.*, 1967, **159**, 98

CHAPTER 3.
COMPUTER SIMULATION OF ENZYME MECHANISM

"[The great supercomputer, asked what is the answer to] the great problem of life, the universe and everything [replied, after many years of computation] 42."

- Douglas Adams

3.1 Introduction

The difficulty of describing chemical processes on enzymes and in solution is due to the complexity of these systems. They have a large number of degrees of freedom and do not usually have any symmetry. Enzyme reactions always have at least two phases, the enzyme itself and the environment. An understanding of enzyme mechanism at the molecular level requires the evaluation of various interactions which lead to transition state (TS) stabilization.

Electrostatic interactions have often been cited as the most important contributors to the lowering of the activation free energy of reaction.[1] The stabilization can come from the permanent dipoles of the protein or from induced dipoles in the protein and surrounding solvent. The opposite has also been proposed, that desolvation to a non-polar active site occurs to give a gas phase reaction,[2] but this has been argued against on thermodynamic grounds.[3] The solvent may play a part in catalysis. The difference in free energy between bound substrate and transition state must certainly be compared with the energy in solution to calculate where the catalytic interactions are important.[4] Entropic effects have also been proposed.[4,5] Dynamic effects[6] and quantum mechanical tunnelling[7] may also play their part in catalysis in some cases. When steric strain is involved, van der Waals forces[4] and stereoelectronic effects[8] can be important. For all of the foregoing interactions, a computer simulation should be able to give some indication of the relative contributions in an ideal world. This is a big challenge, but it appears that semi-quantitative calculations of some enzyme reactions are possible.[9]

The complexity outlined above poses problems for computer simulation.

First, it is impossible to obtain a potential energy surface (PES) for a complete system by rigorous quantum mechanical (QM) methods. Secondly, the dimensionality of the problem is enormous and so reliable minimum energy paths are not readily accesible using standard optimization methods. Finally, the sheer weight of calculation demands that various approximations must be used in addition to those required to overcome the first two problems.

As it is not possible even with semiempirical MO methods to solve the Schrödinger equation for an entire protein system, some form of partitioning must be done. The division into quantum mechanically and non-quantum mechanically treated parts is common to the newer hybrid QM/MM calculations of solvated systems. The general Hamiltonian for this type of system is

$$H_{tot} = H_r + H_{rs} + H_s \qquad (3.1)$$

where H_r is the vacuum Hamiltonian for the quantum system, H_{rs} the interactions between the quantum and solvent systems and H_s the interactions solely in the solvent.

3.2 Simulation in Vacuum

The easiest way to study enzyme reactions from a theoretical viewpoint is to focus on the chemistry of a small part of the system in the gas phase. It is often possible to use high level ab initio calculations. Where the structure of the enzyme is known it is possible to place small model molecules in the same orientation as the amino acid residues they represent. For example, a histidine may be replaced by an imidazole ring. The same approach can be used to model large substrates. Several studies have been done which include some basic

elements of an enzyme active site. Some examples are carbonic anhydrase,[10] serine protease,[11] lactate dehydrogenase[12] and triose phosphate isomerase.[13] These types of vacuum studies allow the intrinsic gas phase chemistry to be examined.

One of the most useful aspects of this small model work is in the characterization of the TS for a reaction. Efficient computer algorithms like the eigenvector following routine[14] are found in most MO programs. TS features, for instance in hydride transfers[15] can be used to predict the binding in TS complexes of an enzyme, given the crystal structure of the enzyme with bound substrate.[16]

The gas phase QM calculations are limited in that they do not give much useful information about the energetics of enzyme reactions. It is reasonable to think of an enzyme active site as a special case of solvation, and as there is no environmental contribution in gas phase calculations there is no reason to suppose that the energies have any significance. This may be true even if some of the active site residues are in the model if no extra effort is made to include the overall dielectric properties of the whole system.[17]

3.3 Molecular Dynamics and Enzymes

Molecular dynamics (MD) simulations of enzymes have provided useful information about conformational changes and flexibility of the system as substrates are bound. They can be combined with free energy perturbation (FEP) methods to gain free energy differences for the binding of different substrates, or to investigate the effect of site directed mutagenesis.[18] The MD

and FEP methods have been reviewed recently[19] so will only be mentioned briefly here.

Simulation of enzymes using MD methods have revealed motions associated with ligand binding.[20] The likely catalytic effect of active site mutation has been investigated for TIM isomerase[21] and the 'loop closing' motion of this enzyme has also been examined[22] by MD. Other studies have included simulations of non-covalent enzyme-substrate complexes[23] and hydrogen-bonding networks.[24]

A useful method called the stochastic boundary method has been used to simulate only a limited area around the enzyme active site.[25] In this method, an explicit system of the enzyme, solvent and substrate is divided into the active site region, a buffer region and the reservoir. The reservoir is deleted. Atoms in the active site are treated in the normal MD way with the empirical force field. Atoms in the buffer region are influenced by the empirical force field and mean boundary forces which help to keep them near to their equilibrium positions. In addition, stochastic forces are applied to the buffer atoms in the form of frictional forces which allow for energy flow into and out of the modelled region. This energy exchange is done *via* an external temperature bath. The mean boundary forces and stochastic friction forces mimic the effect of the deleted reservoir region atoms. My version of this method is described in Chapter 10.

Relative free energy is now calculable for binding and catalysis from the combination of MD and FEP techniques. The implementation is done by calculating a mixed Hamiltonian from two Hamiltonians which represent the systems for which you want to find the free energy difference. The mixed Hamiltonian takes the form

$$H_m = mH_a + (1\text{-}m)H_b \qquad (3.2)$$

where m is a mixing parameter which is varied from 0 to 1. For each value of m a MD simulation is done to equilibrate the hybrid system and to obtain time-averaged properties such as energy. The summation of these energies gives an estimate of the free energy of the change being modelled. Figure 3.1 shows how a thermodynamic cycle is used to calculate the relative free energy changes. This technique has been used to calculate binding energy changes on trypsin and a mutant,[26] and also for the binding difference between different substrates on the same enzyme.[27] In many cases, binding free energy differences can be calculated to a good degree of accuracy.[18]

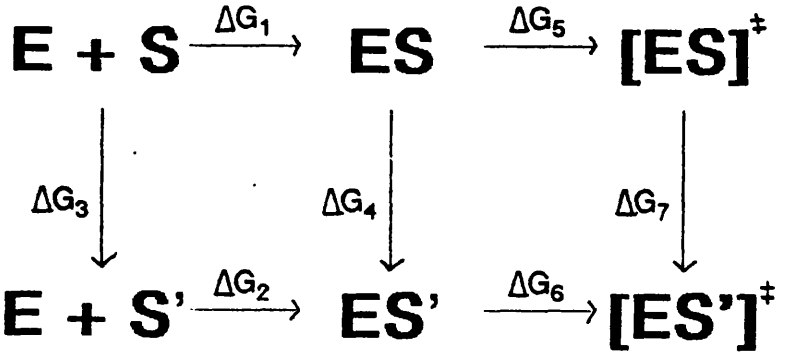


Figure 3.1. A thermodynamic cycle for FEP calculations. ‡ denotes a TS. The free energies of binding ΔG1 and ΔG2 and activation free energies ΔG5 and ΔG6 can be obtained from the K_m and k_{cat} values for the enzyme reaction with the substrates S and S'. The non-physical processes ΔG3, ΔG4 and ΔG7 can be simulated using the FEP method. The calculation of these latter values allows the differences in binding free energy or activation energy between the different substrates to be obtained *via* the thermodynamic cycle. The same technique can be applied to simulation of mutant enzymes without the need to make the mutant enzyme. In this case, E would represent a single substrate and S and S' the native and mutated enzyme.

3.4 Reaction Field Methods

Onsager developed the concept of the reaction field[28] to describe ions in solution. The idea is to place charges inside a spherical cavity within a homogeneous solvent that has a specific dielectric constant. When this model is applied within MO theory it is known as the self-consistent reaction field (SCRF) method.[29] SCRF methods can be applied at the ab initio and semiempirical levels of MO theory. These methods have been fairly successful in predicting solvent effects for small molecules.[30] The cavity radius at the core of the model is unfortunately an empirical parameter, but can be successfully treated.[31] The SCRF model has been extended to enzymes by including the heterogeneous potential of the protein,[32] but a better description requires that the inhomogeneity of the whole reaction field be considered. The direct reaction field (DRF)[33] includes not only point charges for the protein but also allows for polarization. In the DRF method, atoms not in the QM system are represented as point charges and atomic polarizabilities. The method has been used in *ab initio* calculations applied to cysteine protease[34] and actinidin.[35]

Reaction field methods give a good representation of the effects of polarization in the protein environment. By themselves, however, they do not explicitly consider bonding and steric interactions unless parts of the enzyme are incorporated into the model system which is treated quantum mechanically.

3.5 Hybrid Quantum and Molecular Mechanical Simulations

At the expense of increased complexity and computing time, it is possible to treat the enzyme and solvent molecules explicitly, and this is potentially the most reliable way of obtaining accurate geometry optimizations. The combination of a QM description of the active site of an enzyme plus substrate, together with a classical description of the rest of the system would seem to be a good idea.

Over the last twenty years there have been various attempts to merge the strengths of QM and MM to give an improved model of chemical behaviour. The consistent force field methods of Warshel and Karplus[36] and Birge *et al*[37] were the first such attempts. Warshel and Levitt[38] combined a semiempirical QM approach with MM to calculate the strain energies in the steps of carbohydrate hydrolysis by lysozyme. Warshel has also used an empirical valence bond (EVB) method with MM to simulate complex systems.[39] Allinger and co-workers have used QM to provide parameters for MM optimizations.[40] Houk and co-workers have used ab initio methods to determine transition state geometries for reacting parts of larger systems.[41] These are then used as the basis for a MM parameterization which can be combined with standard parameters for the non-reacting part of the system, in order to estimate strain in the transition state and relative chemical reactivity. More recently, Weiner *et al*[42] did a study of amide hydrolysis in the gas phase, in aqueous solution and in the active site of an enzyme. Chandresekhar *et al*[43] combined QM with Monte Carlo methods to describe the symmetric S_N2 reaction between chloride ion and chloromethane in both gas and solution. With the exception of Warshel and Levitt's QM/MM approach, which used semiempirical QM, none of the above addressed the

problem of coupling the optimizations of the QM and MM systems. The first attempt to couple MM with ab initio methods was by Singh and Kollman in their QUEST program[44] which combined the Gaussian 80 quantum chemistry program[45] with the AMBER force field.[46] Similar methodologies have followed, notably Waszkowycz *et al*[47] who have done a series of studies on catalysis by the Phospholipase A_2 enzyme. A new semiempirical QM/MM implementation has been developed and incorporated into the CHARMM program.[48] It uses the MOPAC[49] program for the MO part of the calculation and the CHARMM empirical force field for the molecular mechanics. It has the additional facility of using the calculations in conjunction with molecular dynamical methods. This is almost certainly the first attempt to explore anything other than a static potential surface. The CHARMM approach has been used to look at the triosephosphate isomerase reaction pathway.[1b] Since it is the CHARMM method which is to be used for much of the work in this thesis, it is appropriate to describe this method in greater detail.

3.5.1 The CHARMM Method

The approach in CHARMM is to divide the system into three regions: a QM region, an MM region and a boundary region. The QM and MM atoms interact using the empirical force field when there are both QM and MM atoms in a force field term. If a molecule extends across the QM/MM regions so that it has both QM and MM atoms, then link atoms are defined. These are hydrogen atoms added to the QM system to provide the correct valence. They are invisible to the MM system because no link/MM atom interactions are calculated. The important van der Waals interactions between QM/MM atoms are calculated in

the same way as for MM/MM interactions via a Lennard-Jones function, and only over pairs of atoms which are separated by three or more bonds. Evaluation of electrostatic energy is more difficult because the semiempirical QM method separates the electrons into valence and core electrons, with the core electrons treated as invariant. The implementation has been to treat the MM atoms as partial positive atomic cores just as if they were electron-less atoms in the standard QM treatment and to calculate core-core repulsion and one-electron attractions between the QM and MM sets of atoms. For the AM1 model this requires evaluation of the two centre repulsion integrals, which have five parameters for each element. Each MM atom must be assigned relevant parameters. After much testing it was found that the gaussian functions in AM1 made no difference to the QM/MM potential, which meant that three of the parameters could be given a value of zero. Further testing showed that only the parameter for an extra core-core repulsion term was significant, and that this could take the same value for each MM atom regardless of type. The MM atoms are therefore distinguished only by their partial charges and van der Waals parameters when interacting with the QM system. For the QM/MM electrostatic interactions, all pairs of interactions are considered.

The boundary region is the edge of the molecular system that is being modelled. The two usual methods in MM to treat edges are the periodic boundary[50] and the stochastic boundary[51] approaches. In the periodic boundary method the system is replicated by translating it in three dimensions so that the "virtual system" is a lattice of the central group of atoms and exact images of the central group. This requires that the original system is defined as a box to avoid "holes" in the virtual system. The atoms in the images interact in the usual manner with the original atoms to calculate the energy of the system. The problem with this approach is that the charges on the image QM atoms need

calculating separately from the original ones and so much extra computing is required. The periodic cell could be chosen so that the original and image QM atoms are separated by more than the nonbond cutoff distance used, so they would not interact. However, as explained in section 10.3, for effective calculations there must be no nonbonded cutoffs in a QM/MM system using the current implementation of CHARMM, so the periodic boundary method is not satisfactory.

The stochastic boundary method is the preferred approach for QM/MM systems. This has been briefly mentioned in section 3.3 and my implementation is explained in Chapter 10. It is this method which is used in calculations throughout this thesis which use the CHARMM program to simulate enzymic or solution reactions.

The CHARMM method allows for MD simulations to be done at the same time as consideration of the electronic changes, which allows for the possibility of modelling dynamic reaction pathways. This has been attempted for the solvated CH_3Cl-Cl^- system.[52] The method has also been used in the calculation of absolute free energies of solvation.[53]

3.6 References

1. a) A. Warshel & M. Levitt, *J. Mol. Biol.* , 1976, **103**, 227-49; b) P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe & M. Karplus, *Biochemistry* 1991, **30**, 5826-32
2. M. J. S. Dewar & D. M. Storch, *Proc. Nat. Acad. Sci. USA*, 1985, **82**, 2225; *J. Chem. Soc. Perkin II* , 1989, 877-85

3. A. Warshel, J. Aqvist & S. Creighton, *Proc. Nat. Acad. Sci. USA* , 1989, **86**, 5820

4. Enzyme Structure & Mechanism, A. Fersht, 2nd Edn, pub Freeman NY 1985

5. M. I. Page & W. P. Jencks, *Proc. Nat. Acad. Sci. USA*, 1971, **68**, 1678-83

6. J. Kraut, *Science* , 1988, **242**, 533-40

7. W. J. Bruno, W. Bialek, *Biophys. J..*, 1992, **63**, 689-99

8. S. A. Benner in Mechanistic Principles of enzyme activity, Ed J. F. Liebman, A. Greenberg, pub VCH NY 1988, p27-74

9. A. J. Mulholland, G. H. Grant & W. G. Richards, *Prot. Eng.* , 1993, **6**, 133-47

10. O. Jacob, R. Cardenas & O. Tapia, *J. Am. Chem. Soc.*, 1990, **112**, 8692

11. P. A. Kollman & D. M. Hayes, *J. Am. Chem. Soc.* , 1981, **103**, 2955

12. J. Wilkie & I. H. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 5423

13. G. Alagona, P. Desmeules, C. Ghio & P. A. Kollman, *J. Am. Chem. Soc.* 1984, **106**, 5710

14. J. Baker, *J. Comp. Chem.*, 1986, **7**, 385

15. a) Y. D. Wu & K. N. Houk, *J. Am. Chem. Soc.*, 1987, **109**, 2226; b) I. H. Williams, A. B. Miller & G. M. Maggiora, *J. Am. Chem. Soc.*, 1990, **112**, 530

16. C. Bystroff, S. J. Oatley & J. Kraut, *Biochemistry*, 1990, **29**, 3263

17. J. Aqvist & A. Warshel, *Chem. Rev.*, 1993, **93**, 2523

18. C. Lee, *Curr. Opinion Struct. Biol.*, 1992, **2**, 217

19. a) W. L. Jorgensen, *Acc. Chem. Res.*, 1989, **22**, 184; b) J. A. McCammon, *Curr. Opinion Struct. Biol.*, 1991, **1**, 196

20. M. Karplus & G. A. Petsko, *Nature*, 1990, **347**, 631

21. V. Dagget & P. A. Kollman, *Prot. Eng.*, 1990, **3**, 677

22. F. K. Brown & P. A. Kollman, *J. Mol. Biol.*, 1987, **198**, 533

23. V. Dagget, S. Scroeder & P. A. Kollman, *J. Am. Chem. Soc.*, 1991, **113**, 8926

24. K. M. Merz Jr., *J. Mol. Biol.*, 1990, **214**, 799

25. a) A. Brunger, C. L. Brooks III & M. Karplus, *Proc. Nat. Acad. Sci. USA*, 1985, **82**, 8458; b) C. L. Brooks III & M. Karplus, *J. Mol. Biol.*, 1989, **208**, 159

26. C. F. Wong & J. A. McCammon, *J. Am. Chem. Soc.*, 1986, **108**, 3830

27. B. G. Rao & U. C. Singh, *J. Am. Chem. Soc.*, 1991, **113**, 6735

28. L. Onsager, *J. Am. Chem. Soc.*, 1936, **58**, 1486

29. O. Tapia & O. Goscinski, *Mol. Phys.*, 1975, **29**, 1653

30. H. S. Rzepa & M. Yi, *J. Chem. Soc. Perkin II*, 1991, 531

31. a) C. J. Cramer & D. G. Truhlar, *J. Am. Chem. Soc.*, 1991, **113**, 8305; b) V. Luzhkov & A. Warshel, *J. Comp. Chem.*, 1992, **13**, 199

32. O. Tapia, F. Sussman & E. Poulain, *J. Theor. Biol.*, 1978, **71**,49

33. B. T. Thole & P. Th. van Duijnen, *Theor. Chim. Acta*, 1980, **55**, 307

34. J. A. C. Rullmann, M. N. Bellido & P. Th. van Duijnen, *J. Mol. Biol.*, 1989, **206**, 101

35. B. T. Thole & P. Th. van Duijnen, Biophys. Chem. 1983, 18, 53

36. A. Warshel & M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612

37. R. Birge, M. Sullivan & B. Kohler, *J. Am. Chem. Soc.*, 1976, **98**, 358

38. A. Warshel & M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227

39. A. Warshel & R. Weiss, *J. Am. Chem. Soc.*, 1980, **102**, 6218

40. N. L. Allinger & J. T. Sprague, *J. Am. Chem. Soc.*, 1973, **95**, 3893

41. K. N. Houk, S. R. Moses, Y. D. Wu, N. G. Rondan, V. Jäger, R. Schoke & F. R. Fromczek, *J. Am. Chem. Soc.*, 1984, **106**, 3880

42. a) S. Weiner, U. C. Singh & P. Kollman, *J. Am. Chem. Soc.*, 1985, **107**, 2219; b) S. Weiner, G. Seibel & P. Kollman, *Proc. Nat. Acad. Sci. USA*, 1986, **83**, 649

43. J. Chandresakhar, S. Smith & W. Jorgensen, *J. Am. Chem. Soc.*, 1985, **107**, 154

44. U. C. Singh & P. A. Kollman, *J. Comp. Chem.*, 1986, **7**, 718

45. U. C. Singh & P. Kollman, *QCPE Bull.* , 1982, **2**, 17

46. P. Weiner & P. Kollman, *J. Comp. Chem.*, 1981, **2**, 287

47. a) B. Waszkowycz & I. H. Hillier, *J. Chem. Soc. Perkin 2*, 1990, 1259; b) B. Waszkowycz, I. H. Hillier, N. Gensmantel & D. W. Payling, *J. Chem. Soc. Perkin 2*, 1991, 225 & 2025

48. M. J. Field, P. Bash & M. Karplus, *J. Comp. Chem.*, 1990, **11**, 700

49. J. J. P. Stewart, *J. Comp. Aided Mol. Des.*, 1990, **4**, 1

50. M. Born & Th. Von Karman, *Physik. Z.*, 1912, **13**, 297

51. C. L. Brooks III & M. Karplus, *J. Chem. Phys.*,1983, **79**, 6312

52. P. A. Bash, M. J. Field & M. Karplus, *J. Am. Chem. Soc.*, 1987, **109**, 8092

53. J. Gao, *J. Phys. Chem.*, 1992, **96**, 537

CHAPTER 4.
STANDARD CALCULATION METHODS

"Why", said the Dodo, "the best way to explain it is to do it."

- Lewis Carroll (C. L. Dodgson)

4.1 Introduction

Each computational method used in this thesis requires the use of keywords and initialization of variables for effective operation. This chapter describes the standard conditions under which most calculations were done. Any variation from these will be noted in the descriptions of the calculations performed.

4.2 Geometry Input

For all computational methods used on small molecules, that is AM1, AMSOL and COSMO, geometry input was done *via* a graphical interface. This was either the COSMIC program[1] or the graphical interface AGUI supplied with AMPAC v4.0[2] or higher.

The larger systems explored by CHARMM were specified in several ways. The protein structures were read directly from Brookhaven .pdb files which contain the cartesian coordinate data for heavy atoms. Water molecules for solvation were taken from pre-equilibrated molecular dynamics simulations provided with the CHARMM program. The original files contained coordinates for cubes of water molecules, but for the calculations in this thesis a sphere of water molecules 15Å in radius was extracted using CHARMM commands and saved to a file for regular use. Small molecules could be input to CHARMM via a conversion program I wrote to convert the cartesian coordinates output by the 0SCF command of MOPAC to the CHARMM .pdb format. Before a molecule can be recognised by CHARMM it must have an associated topology description. Since my decision was to read cartesian coordinates at all times, the topology descriptions for new molecules could contain the bare minimum of information. This required the atoms to be grouped into selections of nominally

zero or unit charge; each atom to be assigned an atom type for identification in force field calculations; a list of the bonds; a list of the proton acceptors in the molecule; and a list of the proton donors in the molecule. If any of the molecules were to be partitioned into QM and MM parts, link atoms also had to be defined. Each link atom must be in a group by itself. Figure 4.1 shows a description for an alanine residue with link atoms, together with the structure that is being described.

```
RESI ALAQ    0.00000    ! alanine with 2 QM/MM link atoms
GROUP
ATOM N   NH1   -0.47000 ! ALLOW PEP
ATOM HN  H      0.31000 ! ALLOW PEP
ATOM CA  CT     0.07000 ! ALLOW PEP
ATOM HA  HB     0.09000 ! ALLOW PEP
GROUP
ATOM CB  CT    -0.27000 ! ALLOW ALI
ATOM HB1 HA     0.09000 ! ALLOW ALI
ATOM HB2 HA     0.09000 ! ALLOW ALI
ATOM HB3 HA     0.09000 ! ALLOW ALI
GROUP
ATOM C   C      0.51000 ! ALLOW PEP
ATOM O   O     -0.51000 ! ALLOW PEP
GROUP
ATOM HQN QQ     0.00000 ! link atom on nitrogen
GROUP
ATOM HQC QQ     0.00000 ! link atom on carbon
BOND CB CA N HN N CA O C
BOND C CA C +N CA HA CB HB1 CB HB2 CB HB3
IMPR N -C CA HN C CA +N O
DONOR HN N
ACCEPTOR O C
! These lines are not necessary if you can guarantee that you will
! always supply coordinates for your systems.
IC -C  CA  *N  HN   1.3474 124.1100 180.0000 114.3400  0.9977
IC -C  N   CA  C    1.3474 124.1100 180.0000 107.2500  1.5185
IC N   CA  C   +N   1.4482 107.2500 180.0000 117.7500  1.3474
IC +N  CA  *C  O    1.3474 117.7500 180.0000 120.6600  1.2297
IC CA  C   +N  +CA  1.5185 117.7500 180.0000 124.0300  1.4478
IC N   C   *CA CB   1.4482 107.2500 121.8800 110.7900  1.5432
IC N   C   *CA HA   1.4482 107.2500 -117.4000 107.3400  1.0835
IC C   CA  CB  HB1  1.5185 110.7900 179.7600 109.5400  1.1119
IC HB1 CA  *CB HB2  1.1119 109.5400 118.9500 110.9400  1.1113
IC HB1 CA  *CB HB3  1.1119 109.5400 -119.3900 111.7000  1.1106
```

Figure 4.1. A typical CHARMM topology entry. The example shows an alanine residue, with added link atoms, that can be treated as a part of a quantum mechanical system.

4.3 AM1, AMSOL and COSMO Semiempirical QM Methods

The AM1 Hamiltonian[3] was used in its implementation in the MOSOL program. This is the version of the MOPAC v6 program[4] which I modified to include the free energy of solvation calculations of AMSOL. The modifications are described in Appendix I. The AM1 Hamiltonian does semiempirical *in vacuo* QM calculations.

AMSOL calculates the free energy of solvation in aqueous solution using the method of Cramer and Truhlar.[5] Appendix I gives details of this model and its implementation in MOSOL. The AM1-SM1 model was used at all times because the other available model, AM1-SM1a, was potentially misleading in the important structures near an activated complex.

COSMO[6] is a conductor-like screening model which calculates the heat of formation in solution. It is implemented in MOPAC 93.[7]

Table 4.1. Keywords Used for Semiempirical QM Calculation

Method	Geometry Optimization	Force Calculation	Transition State Search
AM1	AM1 GNORM=0.05 PRECISE	AM1 FORCE ISOTOPE	AM1 NLLSQ[a] XYZ PRECISE GNORM=0.05
AMSOL	AM1 GNORM=5 AQUO PRECISE EF[b]	AM1 AQUO FORCE ISOTOPE AQFAST[c]	AM1 AQUO PRECISE NLLSQ[a] XYZ GNORM=0.05
COSMO	AM1 GNORM=0.05 PRECISE EPS=78.4	AM1 FORCE EPS=78.4 ISOTOPE	AM1 XYZ GNORM=0.05 EPS=78.4 PRECISE NLLSQ[a]

[a] If this failed then SIGMA was used. [b] This keyword omitted in second optimization stage. See text. [c] This keyword not used when force constants were to be used for KIE calculations with CAMVIB

Table 4.1 summarises the keywords used for geometry optimization and force constant calculations. If a species was charged then the keyword CHARGE was used with the appropriate value. AMSOL optimizations required a two stage process. The first stage used the keywords as in Table 4.1. For stage two, the keyword EF was omitted and GNORM=0.05 used. This optimization protocol was developed during testing of the program. It is believed to work because the discontinuity of the free energy calculation causes instability in the minimization algorithms. The EF method nearly always failed when used with a low GNORM value; the other minimization procedures nearly always gave high GNORM values when starting from a structure which had not been optimized at all, and would not optimize further.

4.4 Hybrid QM/MM Calculations with CHARMM

The CHARMM molecular dynamics program allows the mixing of quantum and molecular mechanical descriptions of atoms within a system. It does this by partitioning the system of interest into QM and MM regions. The QM/MM calculations in this thesis fall into two groups: small molecules solvated within a sphere of water; and a substrate in an enzyme active site. They share some common features in the way that calculations are performed. It may be useful to refer to section 3.5.1 which contains a description of how CHARMM is implemented.

4.4.1 Solvated Molecules

To set up a QM/MM calculation on a solvated small molecule the procedure outlined below was used. It may be useful to look at Appendix III, example A which performs the tasks outlined below.

1. Read in the topology and parameter files that CHARMM requires for building structures and assigning MM parameters, then define the set of molecules that make up the system being simulated.
2. Read in a file containing the coordinates of a water sphere of radius 15Å. The coordinates are centred at the origin. The water molecules will be modelled with the MM force field using the TIP3P model[8] for water.
3. Read in the coordinates of the solute system that is to be modelled by the AM1 Hamiltonian. Translate the solute so that it is centred at the origin to maximise the depth of water in all directions.
4. Delete all water molecules that have an oxygen atom within 2.5Å of any heavy atom of the solute. This distance is sufficient to allow H-bonding interactions to be preserved without including too many bad van der Waals contacts.
5. Identify the QM atoms to CHARMM. If part of the solute is to be treated by the MM force field then the positions of any link atoms must be defined. This is usually done by placing them half way along the bond that connects a QM atom to an MM atom.
6. Set up a solvent boundary potential on the MM water molecules. This provides a potential which helps to maintain the spherical structure of the water at the edges of the sphere and prevents the waters from flying away. The potential falls off rapidly inside the sphere but is very large outside the sphere, encouraging the water molecules at the sphere surface to stay close to the bulk of the water. The potential has spherical symmetry and is centered on the origin of the water sphere.

Once the above steps have been carried out, the calculations can be started. The non-bonded interactions are calculated on an atom-atom basis and no cutoff function is used to restrict them. This has been found necessary to obtain stable results with the QM/MM simulation.[9] A distant-dependent dielectric was used in the calculation of electrostatic interactions. Although this has no real physical meaning, protein simulations show best results when such a dielectric is used[10] and it was thought unnecessary to change this default calculation method.

4.4.2 Enzyme systems

For enzyme simulations, the whole system needs to be defined, then the system is truncated to make it more manageable. Some constraints need to be applied. I use a modified version of stochastic boundary molecular dynamics (SBMD) which has been shown to perform satisfactorily (Chapter 10). Enzyme systems are set up in the manner described below. Example B in Appendix III shows the commands that CHARMM requires to perform the following sequence of manipulations and do a simple minimization:

1. Read in the topology and parameter files that CHARMM requires for building structures and assigning MM parameters, then define the set of molecules that make up the system being simulated.
2. Read in the enzyme and substrate coordinates. It may be necessary to add hydrogens if you are starting from scratch using X-ray or NMR data; if so then CHARMM has a function HBUILD which allows protons to be added automatically. It does this by reading information from the topology file.
3. If you want to solvate the active site, read in the water sphere and translate the enzyme-substrate coordinates so that the desired atom

is at the centre of the water sphere. Then delete all water molecules that have an oxygen atom within 2.5Å of any heavy atom of the enzyme-substrate system. This distance is sufficient to allow H-bonding interactions to be preserved without including too many bad van der Waals contacts. Finally, set up a solvent boundary potential on the MM water molecules. This provides a potential which helps to maintain the spherical structure of the water at the edges of the sphere and prevents the waters from flying away. The potential falls off rapidly inside the sphere but is very large outside the sphere, encouraging the water molecules at the sphere surface to stay close to the bulk of the water. The potential has spherical symmetry and is centered on the origin of the water sphere.

4. Identify the QM atoms to CHARMM. If only part of the substrate is to be treated quantum mechanically, or parts of the enzyme are to be included in the QM system, then the positions of any link atoms must be defined. This is usually done by placing them half way along the bond that connects a QM atom to an MM atom.
5. Delete all the residues of the enzyme that have no atoms within 17Å of the substrate. The substrate is defined for this purpose as all the atoms of the substrate which are included in the quantum system. This truncation deliberately leaves residues intact so that fragments of residues are not left "floating" around. The deleted set is the equivalent of the reservoir in the stochastic boundary method (section 3.3). The truncation is performed to reduce the computational overhead (section 10.4). The 17Å truncation radius has been chosen because it is 2Å further than any solvent water that may be included in the system. This extra 2Å allows for some constraints explained below.

6. Constrain the enzyme atoms that are in the outer 2Å of the system. This serves to keep the structure of the enzyme, especially where there may be loops of protein disconnected from the main chain because of the truncation. The atoms are constrained with the harmonic constraint force in CHARMM, which mimics the effect of mean boundary forces in the stochastic boundary method, and is easier to implement (section 10.4).

The overall effect of the above is to give a zone of 15Å radius in which atoms move according to the combined QM/MM force field, with structural integrity maintained by a combination of a solvent boundary potential and harmonic forces. The system is now ready for QM/MM calculations. Non-bonded interactions are calculated as for solvated molecules.

4.5 Molecular Dynamics Calculations with CHARMM

Molecular dynamics calculations were performed on enzyme systems. The preliminary set up was the same as for QM/MM enzyme systems except that no QM atoms were defined. In addition, because a variation of stochastic boundary dynamics was used, friction coefficients had to be assigned to the protein and water atoms to allow thermal coupling of the system to an external temperature bath, within the Langevin dynamics algorithm implemented in CHARMM. The friction coefficients used were taken from reference 11, and a short test simulation showed them to perform well (Chapter 10). The friction coefficient values were 120 for the protein heavy atoms and 33 for the water oxygens. The friction coefficients were used on all atoms beyond 15Å of the origin of the molecular system, that is the buffer region where atoms were also constrained with harmonic forces. The atom positions were checked at every fifty dynamics

steps and friction terms applied appropriately. The frictional terms and the harmonic constraints mimic the effect of the reservoir region of atoms which is deleted in the standard SBMD method.

To enable a larger time step to be used, the bonds involving hydrogen atoms were constrained by the SHAKE algorithm. The leapfrog algorithm of Verlet (section 2.12) was used to propagate the motion of the system. The equilibrium temperature was 300K. Simulations were generally run for 40ps in steps of 0.002ps, the first 10ps of which was taken to be equilibration time. This was checked by using the internal correlation coefficient generated after every 1000 steps of dynamics: a low value was taken to be a good indicator of equilibration.

4.6 Calculation of Isotope Effects

Both KIEs and EIEs were calculated using the CAMVIB[12] and CAMISO[13] programs. CAMVIB was given as input the force constant matrix calculated by MOPAC or GAUSSIAN and a file containing the atomic masses in the system and a set of valence coordinates which specify the stretches, bends, wags and torsions within the system. CAMVIB removes the spurious translational and rotational contributions to the force constants and outputs a file containing these cleaned-up force constants. This file, slightly modified, is used by CAMISO to calculate rate constants for a reaction containing different isotopic species. Sample input for CAMVIB and CAMISO is given in Appendix III.

4.7 References

1. J. G. Vinter, A. Davis & M. R. Saunders, *J. Comput. Aided Mol. Des.*, 1987, **1**, 31

2. AMPAC v4.0 and AGUI, Semichem, 12715 W. 66th Terrace, Shawnee, KS 66216, 1992

3. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy & J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902

4. MOPAC 6.0, J. J. P. Stewart, *QCPE Bull.*, 1990, **10**, 86

5. C. J. Cramer & D. G. Truhlar, *J. Am. Chem. Soc.*, 1991, **113**, 8305

6. A. Klamt & G. Shuurmann, *J. Chem. Soc. Perkin 2*, 1993, 799

7. MOPAC93, J. J. P. Stewart and Fujitsu Ltd, Tokyo, 1993

8. W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey & M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926

9. P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe & M. Karplus, *Biochemistry*, 1991, **30**, 5826

10. R. J. Loncharich & B. R. Brooks, *Proteins*, 1989, **6**, 32

11. C. L. Brooks III & M. Karplus, *J. Mol. Biol.*, 1989, **208**, 159

12. CAMVIB, I. H. Williams, University of Bath, Bath BA2 7AY, 1992; I. H. Williams, *Chem. Phys. Lett.*, 1982, **88**, 462; *J. Mol. Struct.*, 1983, **94**, 275

13. CAMISO, I. H. Williams, University of Bath, Bath BA2 7AY, 1992; I. H. Williams, *J. Am. Chem. Soc.*, 1984, **106**, 7206

CHAPTER 5.
THE ACID-CATALYZED HYDROLYSIS OF METHYL α- AND β-GLUCOPYRANOSIDES

"The best material model of a cat is another, or preferably the same, cat."

- A. Rosenblueth

5.1 Introduction

The work in this thesis relies upon the accurate calculation of isotope effects. One of the concerns with the majority of the work is the use of approximate or cutoff structures for substrates which are too large to be easily modelled using quantum mechanical methods. What are the minimum structural features necessary to provide a good model of the compound of interest insofar as the calculation of isotope effects is concerned? Some of the features can be predicted from chemical knowledge, for example retention of substituents in an aromatic system which are known to affect rates of reaction because of their electronic effects. These are not necessarily related to isotope effect measurements but are required for a complete description of the electronic reorganisation which occurs during the reaction. What is required is knowledge of how much of the molecular framework around the site of isotopic substitution needs to be retained. It has been suggested[1,2] that as many as three bonds around the site of isotopic substitution need to be retained in any cutoff model. This proposal came from Stern and Wolfsberg who carried out a series of calculations on full and partial models. This 'three bonds model' is often not done especially in high level *ab initio* calculations where such large systems are still not often modelled.

This chapter summarises some earlier work that used experimental KIEs and EIE values calculated using minimal cutoff models of glucopyranosides to draw conclusions about TS structure. The work is extended using larger cutoff models and includes some preliminary calculation of EIEs using a semiempirical solvation model. It is hoped that the work done here will allow the question of model size to be at least partially answered.

5.2 Previous work

Bennet and Sinnott[3] measured KIEs for acid catalyzed hydrolyses of methyl α and β-glucopyranosides at 80°C. Figure 5.1 shows the pattern of seven isotopic substitutions they made. These experimental results were interpreted with the aid of some EIEs calculated by Williams[4] from ab initio HF/4-31G MO calculations. The model for these EIEs was a simple acyclic model: methanediol was used for the glucoside and protonated formaldehyde for the oxocarbenium ion. Figure 5.2 indicates the relationship between the actual and the model system and gives the original experimental KIEs and calculated EIEs.

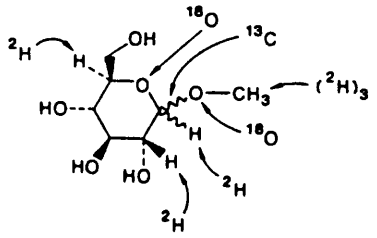
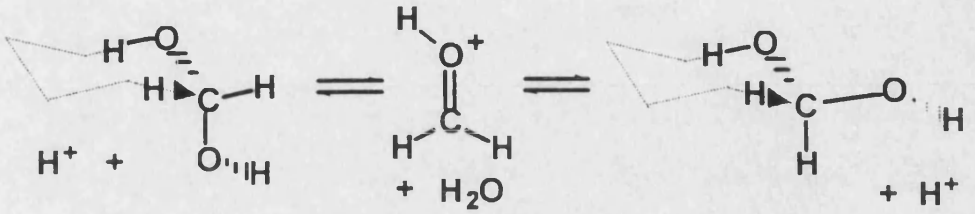


Figure 5.1. The pattern of isotopic substitution used to obtain KIEs for the methyl glucopyranosides

The acid-catalyzed hydrolysis of glucopyranosides is known to proceed by specific acid catalysis.[5] The leaving group oxygen is reversibly protonated before the bond to the leaving group is cleaved. The magnitude of the leaving group ^{18}O KIE is therefore a direct measure of the extent to which the glycone-aglycone bond is broken at the TS. The experimental KIE is similar to the calculated EIE, and this was interpreted as showing substantial glycone-aglycone bond cleavage in the rate determining TS structure.

The ring ^{18}O effect arises because the endocyclic C-O bond acquires some double bond character as the charge on the anomeric carbon is delocalized by



α-anomer Expt KIE	α-anomer Calc EIE		β-anomer Expt KIE	β-anomer Calc EIE
1.137	1.188	alpha-D	1.089	1.137
1.026	1.019	leaving gp ^{18}O	1.024	1.021
0.996	0.978	ring ^{18}O	0.991	0.976
1.007	1.004	anomeric ^{13}C	1.011	1.004
1.073		beta-D	1.045	

Figure 5.2 Experimental KIEs for the acid-catalyzed hydrolysis of methyl α- and β-glucopyranosides and HF/4-31G calculated EIEs for the illustrated model system. Dotted lines help to show the part of the structure which is being modelled.

the oxygen lone pairs. Because the bond strength is increased, the effect will be inverse (see section 2.4) and its magnitude reflects the double bond character and thus the charge on the oxygen. The delocalization is due to the overlap of the vacant p-orbital on the anomeric carbon and a lone pair of the ring oxygen, as illustrated in Figure 5.3. Since the anomeric carbon is carrying an almost full positive charge,[3] the EIE should be close to the maximum effect possible because it is calculated for the oxocarbenium ion species at equilibrium and this has the optimal p-p orbital overlap. The overlap of p orbitals is proportional to the square of the cosine of their dihedral overlap. If the approximation is made that the isotope effect can be approximated just by the zero point energy

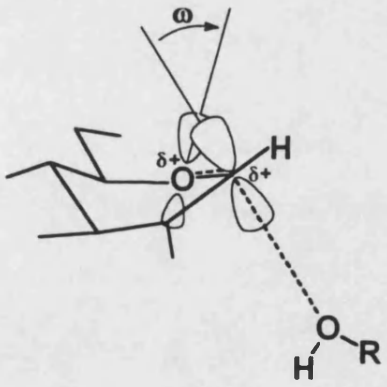


Figure 5.3 The origin of the ring ^{18}O isotope effect.

contribution, which is reasonable at lower temperatures, then the isotope effects can be expressed as pure exponentials.[6] The angle between these orbitals at the TS can therefore be estimated by the expression

$$\ln(KIE) \approx \ln(EIE) \times \cos^2\omega. \qquad (5.1)$$

Values of 66° and 52° for ω were obtained by this means for the α- and β-anomer respectively. Bennet and Sinnott deduced lower bounds for this overlap angle by consideration of an acyclic system and obtained 54° and 24° for the α- and β-anomer respectively.

If there is no nucleophilic participation at the TS, the ^{13}C KIE will reflect the compensating effect that the strengthening endocyclic C-O bond has on the cleavage of the glycone-aglycone bond. The fact that the KIE was larger than the EIE was interpreted as a lagging behind of the electronic and geometric reorganization of the glycone compared to the almost cleaved bond to the aglycone.The comparatively small value was also seen as confirmation that there is no nucleophilic assistance in the TS.

The α-D effect arises from changes in the bending force constants involving the hydrogen. For the reaction considered here it can be regarded as a measure of the planarity of the bonds about the anomeric carbon, or "flatness", with the maximum "flatness" being in the oxocarbenium ion. The maximum KIE would therefore be expected to be close to the EIE value. The "flatness" ratio ln(KIE)/ln(EIE) is approximately 0.75 for the α- and 0.66 for the β-anomer. This represents progress along the reaction coordinate for re-hybridization at the anomeric centre. This measurement supports the asynchronous sequence of events suggested above.

The β-D 2^0 effect originates from changes in vibrational frequency involving the hydrogen and has two components, hyperconjugation and the inductive effect of deuterium. β-D effects can be expressed as follows

$$\ln (K_H/K_D) = \cos^2\emptyset.\ln(K_H/K_D)_{max} + \ln (K_H/K_D)_i \qquad (5.2)$$

The first term represents hyperconjugation of the C-H sigma orbital with an electron deficient p orbital on an adjacent carbon atom, where Ø is the dihedral angle between the orbitals. The maximum ratio is obtained when the orbitals are eclipsed, and also increases as the positive charge on the adjacent atom increases with the associated weakening of the C-H bond. The second term represents a geometry independent inductive effect of the deuterium atom, reflecting the electron-donating effect of deuterium. (This apparent electron-donating ability has a vibrational origin, since the PES is identical for both isotopomers within the Born-Oppenheimer approximation). Using the above equation, Bennet and Sinnott estimated ranges for Ø of 31°-43° for the α-anomer and 41°-50° for the β-anomer based on previously reported isotope effects[7] and an estimate of 0.96 for $\ln (K_H/K_D)_i$.[8]

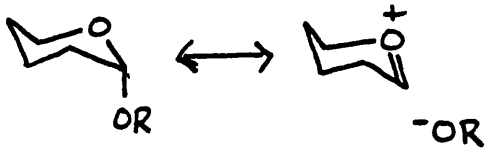
All the information from these interpretations was used to produce constraints on the TS structure which were used with Allinger's MM2 molecular mechanics force field[9] to generate possible conformations for the TS structure. The resulting structures (see section 5.4) showed that the α-anomer adopts a skew-boat conformation and the β-anomer a chair, both of which contradict the theory of stereoelectronic control or antiperiplanar lone pair hypothesis (ALPH, see ALPH box). They concluded that the use of ALPH as a

guide to glycoside hydrolysis is misleading. This assumes that the calculated EIEs are reliable, and it is this assumption that the rest of this chapter sets out to test.

The Antiperiplanar Lone Pair Hypothesis

The antiperiplanar lone pair hypothesis (ALPH) suggests that leaving groups from sp^3-hybridized carbon, which is substituted with one or more heteroatoms, depart from conformations in which the leaving group makes a dihedral angle of 180° with an sp^3 lone pair on the heteroatom.

The application of ALPH to ground state conformations has been generally supported by theoretical studies. In a tetrahydropyranyl derivative, the C-O ring bond will shorten and the axial orientation of electronegative substituents at the anomeric carbon will be stabilized, despite an increased bond length, by the interaction of a p-type lone pair on the oxygen and the C-X σ^* orbital. This could arise from the following type of resonance:



There is still a debate about how useful ALPH is in the consideration of reactive conformations.

5.3 Calculation of EIEs for Larger Models of Glucoside Hydrolysis

Model systems of increasing complexity were used to calculate EIEs at 80°C for the hydrolysis of methyl α-glucopyranosides. *Ab initio* HF/4-31G and the semiempirical AM1 and AM1-SM1 models were used to calculate force constants for the reactant and oxocarbenium ion structures. EIEs at 80°C were then calculated using the resultant force constant matrix as input to the CAMVIB and CAMISO programs. All calculations used the standard conditions as described in Chapter 4. Consideration of the various calculated EIEs allows a "best" estimate for the isotope effect upon the methyl α-glucopyranoside hydrolysis to be obtained from the AM1 *in aquo* calculated EIE. The estimate is obtained by assuming the value of the *ab initio* scaled 4-31G EIE to be the most accurate calculation. This is because the scaling factor used was a value which has been found to reproduce experimental vibrational frequencies of small organic molecules.[10] The ratio of the scaled 4-31G EIE to the scaled AM1 *in vacuo* value gives the factor required to bring the AM1 value to the 4-31G value. Similarly, a ratio is needed to connect the unscaled AM1 *in vacuo* to the scaled *in vacuo* value. Therefore the best estimate is obtained by

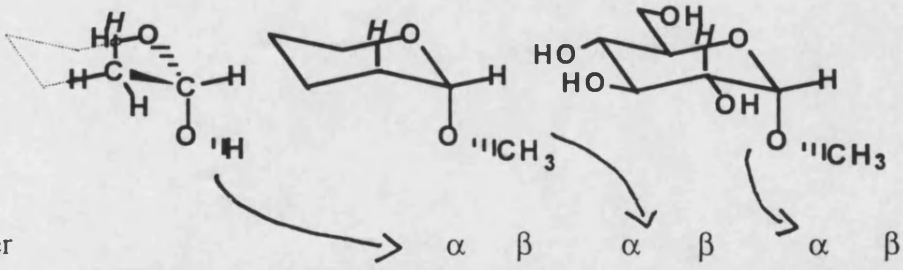
$$EIE_{\text{AM1 in aquo}} \times \frac{EIE_{\text{scaled AM1 in vacuo}}}{EIE_{\text{unscaled AM1 in vacuo}}} \times \frac{EIE_{\text{scaled 4-31G in vacuo}}}{EIE_{\text{scaled AM1 in vacuo}}}$$

For example, the best estimate for the α-anomer in Figure 5.4 is calculated by 1.117 x (1.134/1.139) x (1.091/1.031). Note that the ratios are calculated from results obtained for the same sized model. The "best" estimate represents the EIE

of the relevant isotopic substitution upon the reaction when the reaction occurs in aqueous solution.

5.3.1 Calculated β-D EIEs

Figure 5.4 gives the calculated β-D EIEs for each model system below an illustration of the α-anomer of the model. The original methanediol-protonated formaldehyde model system could not be used to provide a calculated β-D EIE. Comparison of the ethanediol-protonated acetaldehyde and methoxytetrahydropyranoside-tetrahydropyranosyl cation systems at the *in vacuo* AM1 level shows that the "cut-off" approximation employed in the former choice is too severe. Scaling of vibrational frequencies has only a small influence on the calculated EIEs, and the AM1-SM1 in aquo treatment reduces the effects by a small amount also. The AM1 method significantly underestimates the EIE relative to HF/4-31G. Utilization of the "best" estimate of 1.177, together with the experimental KIE of 1.073, allows a value of 44° to be calculated for the dihedral angle between the vacant p-orbital at the anomeric carbon and the β-C-H bond on the adjacent atom; this compares not too



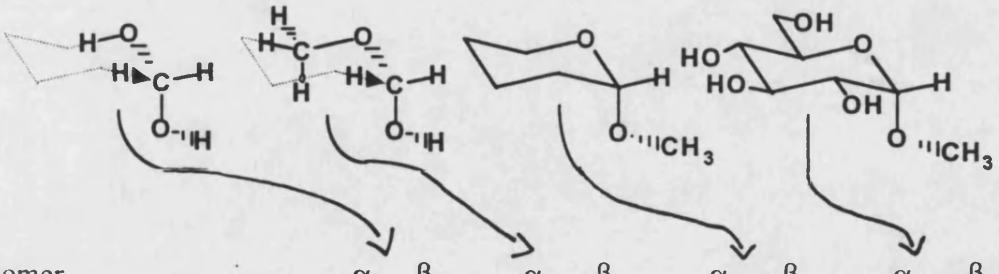
Anomer	α	β	α	β	α	β
Scaled 4-31G *in vacuo*	1.091	1.089				
Scaled AM1 *in vacuo*	1.031	1.030	1.134	1.147		
Unscaled AM1 *in vacuo*			1.139	1.151	1.129	1.117
Unscaled AM1 *in aquo*			1.131	1.145	1.117	1.106
"Best" estimate					1.177	1.165

Figure 5.4. Calculated β-D EIEs at 80°C for hydrolysis of methyl α- and β-glucopyranosides. The illustrated models are the α-anomer. Dotted lines help to show the part of the structure which is being modelled.

unfavourably with the range of 31° to 43° obtained by Bennet and Sinnott. A similar analysis for the β-anomer produces a "best" estimate of 1.165 for the EIE which, with the experimental KIE, yields a dihedral angle of 52° as compared with the range of 41° to 50° obtained previously.

5.3.2 Calculated ring ^{18}O EIEs

The value of the ring ^{18}O EIE was a crucial datum for Sinnott's determination of TS structure. Figure 5.5 shows the ^{18}O EIEs calculated using the illustrated models. For comparison the earlier model and calculated value is also shown. The results in Figure 5.5 raise doubts concerning the reliability of the methanediol-protonated formaldehyde model system. While there is no significant difference between the values calculated for methoxymethanol and methoxytetrahydropyranoside as models for methyl glucopyranoside, the EIE calculated for the smallest model is appreciably too inverse. The "best" estimate



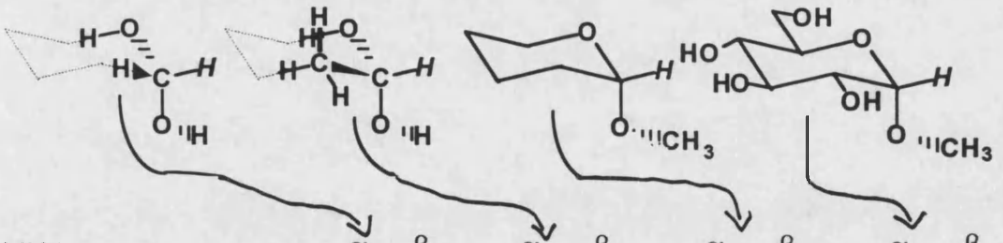
Anomer	α	β	α	β	α	β	α	β
Scaled 4-31G *in vacuo*	0.978	0.976	0.988	0.987				
Scaled AM1 *in vacuo*			0.994	0.994	0.996	0.995		
Unscaled AM1 *in vacuo*					0.996	0.995	0.996	0.998
Unscaled AM1 *in aquo*					0.995	0.994	0.993	0.995
"Best" estimate							0.987	0.988

Figure 5.5. Calculated ring ^{18}O EIEs at 80°C for hydrolysis of methyl α- and β-glucopyranosides. The illustrated models are the α-anomer. Dotted lines help to show the part of the structure which is being modelled.

for the ring ^{18}O EIE for formation of the glucosyl cation from methyl α-glucopyranoside in aqueous solution is 0.987. Combination of this revised value with the experimental KIE, in the manner described above, gives a dihedral angle of 56°, as compared with 66° obtained previously. Similarly a revised dihedral angle of 30° may now be estimated for the β-anomer, rather less than the original value of 52°.

5.3.3 Calculated α-D EIEs

Figure 5.6 presents the calculated α-D EIEs for the pictured models. The AM1 results are considerably larger (more normal) than the HF/4-31G results, but the most surprising finding is that the AM1-SM1 method leads to very much larger values for these isotope effects. This is puzzling, for it is not obvious why this should be so. However, if these results are accepted then a "best" estimate of 1.332 is obtained for the α-anomer, which would suggest a TS much earlier along the reaction coordinate for re-hybridization at the anomeric centre. The "flatness" factor is 0.45, as compared with the earlier estimate of 0.75; for the β-anomer the result would be 0.39 as against 0.66.



Anomer	α	β	α	β	α	β	α	β
Scaled 4-31G *in vacuo*	1.188	1.137	1.181	1.131				
Scaled AM1 *in vacuo*			1.274	1.285	1.267	1.248		
Unscaled AM1 *in vacuo*					1.277	1.258	1.302	1.277
Unscaled AM1 *in aquo*					1.365	1.344	1.448	1.425
"Best" estimate							1.332	1.244

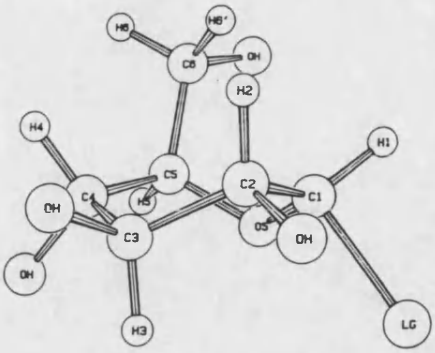
Figure 5.6. Calculated α-D EIEs at 80°C for hydrolysis of methyl α- and β-glucopyranosides. The illustrated models are the α-anomer. Dotted lines help to show the part of the structure which is being modelled.

5.4 Discussion

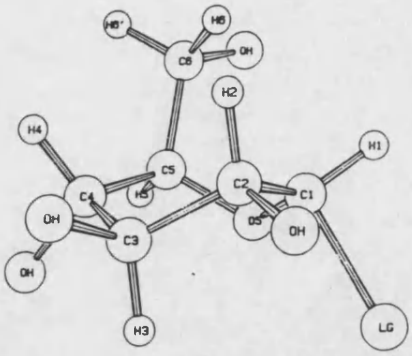
There are two points that require attention in light of the new EIE calculations. The first is a re-evaluation of the original conclusions of Bennet and Sinnott regarding the ALPH hypothesis and the second is the more general one of model size in the context of isotope effect calculations.

The main geometric constraints on Bennet and Sinnott's calculated TS structures were the two dihedral angles estimated from the ^{18}O and β-D isotope effects. They calculated structures using the extremes of the ranges given above and so had four structures for each anomer. In addition, the values for the β-D dihedral for the β-anomer allowed valid structures for the negative angle as well as the positive, giving a total of six possible conformations for the β-anomer. All 10 conformations found are shown in Figure 5.7. The conformations for the α-anomer TS structure are skew boats. Those for the β-anomer fall into three pairs: $^4C_1$ chair, half chair and flattened $^{2,5}B$ boat in order of stability. In each case, the dihedral angle derived from the calculated β-D EIEs is sufficiently close to the range originally considered that one can conclude that the models used are adequate.

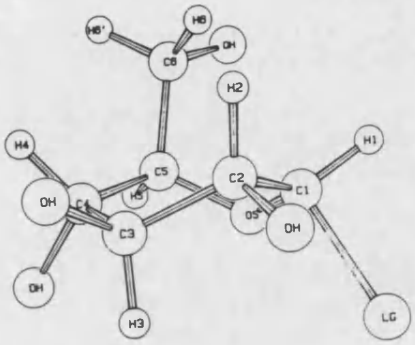
The crucial result is the new ^{18}O EIE value. In both the α- and β-anomer the new calculations lead to a lowering of the associated dihedral angle. For the α-anomer, the revised dihedral angle of 56° compares favourably to the original limits of 54° to 66° although it must be remembered that the upper limit was calculated from the old EIE value. In the light of the revised value, two of the original TS structures can be discarded (5.7a and 5.7b), but the conclusions of Bennet and Sinnott about the ALPH hypothesis when applied to α-anomers is



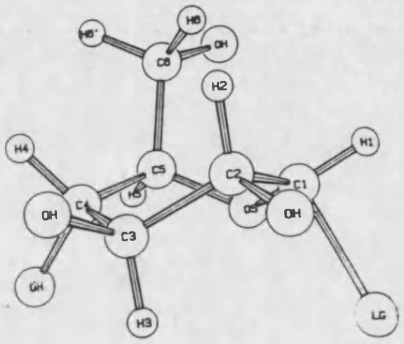
a. $\varnothing = 66^{o}$, $\omega = 43^{o}$



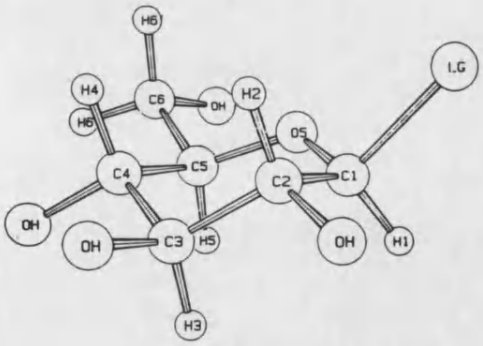
b. $\varnothing = 66^{o}$, $\omega = 31^{o}$



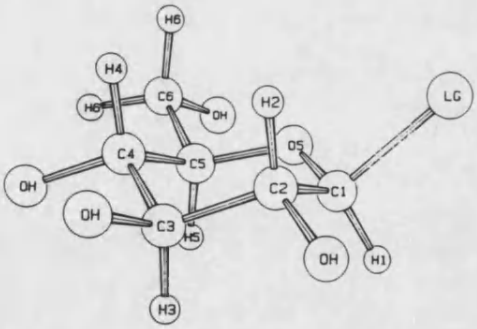
c. $\varnothing = 54^{o}$, $\omega = 43^{o}$



d. $\varnothing = 54^{o}$, $\omega = 31^{o}$

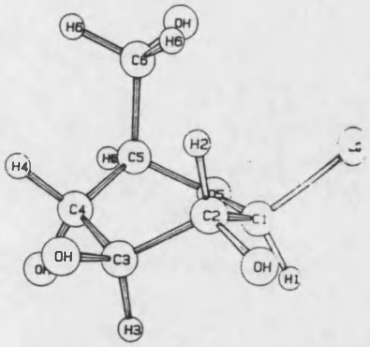


e. $\varnothing = 52^{o}$, $\omega = 50^{o}$

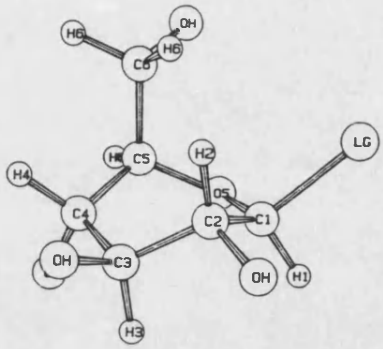


f. $\varnothing = 52^{o}$, $\omega = 41^{o}$

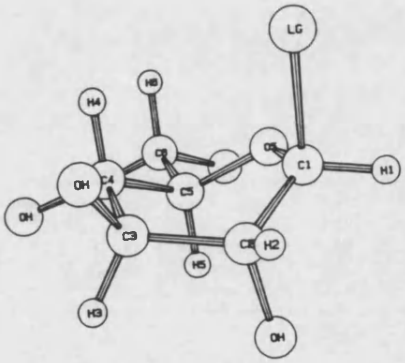
Figure 5.7. Legend on next page.



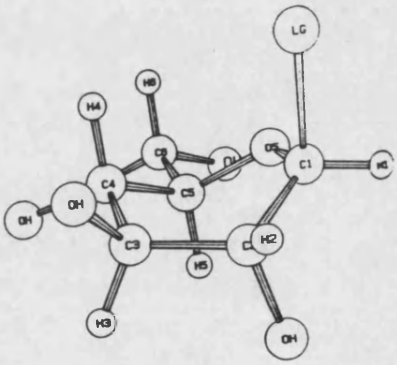
g. $\varnothing = 24^{o}$, $\omega = 50^{o}$



h. $\varnothing = 24^{o}$, $\omega = 41^{o}$



i. $\varnothing = 24^{o}$, $\omega = -50^{o}$



j. $\varnothing = 24^{o}$, $\omega = -41^{o}$

Figure 5.7. Bennet and Sinnott's TS structures for the hydrolysis of methyl α- and β-glucopyranosides. a-d are the α–anomer and e-j the β-anomer TS structures. ∅ is the dihedral angle calculated from the β-D effect and ω the dihedral angle calculated from the ^{18}O effect.

still valid. For the β-anomers, the picture is less clear. The revised EIE gives a new dihedral angle of 30° which means that the pair of structures 5.7e and 5.7f that they considered to be most probable must be discarded. The remaining structures are the half chair and flattened boat. Given that the ALPH hypothesis requires β-anomers to hydrolyse through a boat-like TS[11] this result

unfortunately chooses to sit on the fence: there is certainly movement towards a boat conformation but the most likely structures are the half-chairs 5.7i and 5.7j.

From consideration of the EIEs calculated using the above models there appear to be no hard and fast rules about the size of model. It certainly seems that at least all heavy atoms directly attached to the atom involved in bond making or breaking are required. For the systems here, the larger acyclic models performed better for calculation of H/D isotope effects whereas the cyclic models were better for the heavy atom effects. This could be partly due to the tendency of the AM1 model to make ring systems flatter than they should be. This would affect the bending frequencies of the protons.

It is unfortunate that the present work was not able to provide more useful information regarding the size of cutoff model required for accurate calculation of isotope effects. One of the ways in which this situation could be improved is to fill the EIE tables given for the various isotopes. Use of values for a single model system to calculate the "best" estimates seems a logical way to improve the estimates. In this way the simpler models could be tested individually by calculation of the dihedral angles described above to provide more information about the types of model required for accurate calculation of EIEs.

Given the success of CHARMM in the study of AMP hydrolysis which is described in the next chapter, it would be interesting to use CHARMM for the calculation of the EIEs presented here. As discussed in Chapter 1, EIEs are relatively easy to calculate but difficult to measure experimentally, whereas KIEs are relatively easy to measure but difficult to calculate because of the problems in finding the TS structure for a reaction. The use of CHARMM to provide reliable EIEs would further enhance interpretation of the experimental KIEs. Given sufficient computer resource, a complete PES for the reaction could be generated using CHARMM to provide what could be definitive TS structures and calculated KIEs for the hydrolysis. In this way a valuable

contribution to the knowledge of glucopyranoside hydrolyses can be made, and as a bonus a more definitive test of ALPH versus traditional views of the reactivity of these compounds follows automatically.

5.5 References

1. M. Wolfsberg & M. J. Stern, *Pure Appl. Chem.* , 1964, **8**, 225
2. M. J. Stern & M. Wolfsberg, *J. Chem. Phys.*, 1966, **45**, 4105
3. A. J. Bennet & M. L. Sinnott, *J. Am. Chem. Soc.*, 1986, **108**, 7287
4. I. H. Williams, unpublished work quoted in reference 3.
5. B. Capon & W. G. Overend, *Adv. Carbohydr. Chem.*, 1960, **15**, 33
6. L. Melander & W. H. Saunders in Reaction Rates of Isotopic Molecules, pub Wiley 1980, p26
7. a) H. U. Siehl & H. J. Walter, *J. Am. Chem. Soc.*, 1977, **99**, 5000; b) A. J. Kresge & D. P. J. Weeks, *J. Am. Chem. Soc.*, 1984, **106**, 7140
8. I. H. Williams, *J. Chem. Soc. Chem. Comm.*, 1985, 510
9. N. L. Allinger, *J. Am. Chem. Soc.*,1977, **99**, 812
10. I. H. Williams, *J. Phys. Org. Chem.*, 1990, **3**, 181
11. P. Deslongchamps in Stereoelectronic Effects in Organic Chemistry, pub. Pergamon, Oxford 1983

CHAPTER 6.
COMPUTER SIMULATION OF THE ACID-CATALYZED HYDROLYSIS OF A MODEL FOR ADENOSINE MONOPHOSPHATE (AMP)

"9 cannot do it without comp[u]tens."

- William Shakespeare

6.1 Introduction

Glycoside hydrolyses are important bioorganic processes whose mechanisms are of considerable interest to physical organic chemists. A simple glycoside may undergo reaction with a nucleophile by the three processes in Figure 6.1. The stepwise D_N + A_N (S_N1) mechanism involves rate-determining heterolysis of the bond to the aglycone leaving group, forming a glycosyl cation which is then subject to nucleophilic attack. The concerted A_ND_N (S_N2) mechanism involves a bond to the nucleophile being formed as the bond to the leaving group is broken. The stepwise D_N*A_N pre-association mechanism is similar to the D_N + A_N one, with the nucleophile appearing in the rate determining step as a "spectator".

The basic mechanistic chemistry of the non-enzymic hydrolysis of simple pyranosidic and furanosidic O-glycosides was established by the mid-1970s.[1] The mechanistic picture appears to be that the glucopyranosyl cation just has a real existence in water but not in less polar solvents. Even in water the cation is too unstable to exist in the presence of anionic nucleophiles.[2] For simple

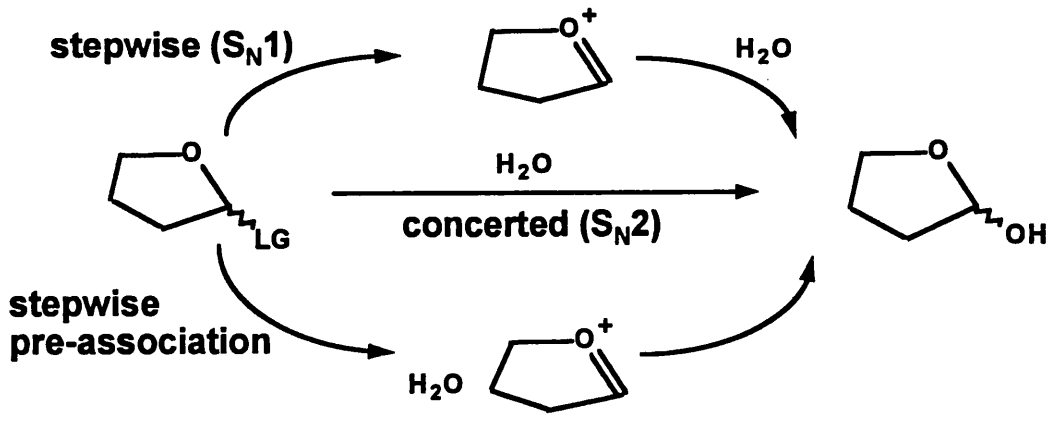


Figure 6.1. Pathways for the hydrolysis of simple glycosides

tetrahydropyranyl derivatives in which the cation is not destabilized by attached hydroxyl groups, the reaction appears to go by simple S_N1 or A_1 pathways in water.[3] The approximate lifetime of typical glycosyl oxocarbenium ions is about 10^{-12}s, and experiment suggests that glycoside hydrolysis occurs at the borderline of the S_N1 and S_N2 mechanisms.[2,4] Characterization of these processes requires knowledge of the transition state (TS) structures. Experimentally, TS structure is most elegantly obtained from kinetic isotope effect (KIE) measurements. Primary (1^o) KIEs indicate atoms which are directly involved in bond making or breaking in the rate-determining TS. Secondary (2^o) KIEs measured for atoms which are not directly involved in bond making or breaking can be used to estimate the position of the TS along the reaction coordinate. The best information from KIEs usually comes from studies where the effects of several isotopic substitutions are examined. A simple way to obtain TS structures is to use the bond-energy-bond-order (BEBO) method to reproduce experimental KIEs using a model in which the TS bond orders are adjustable parameters. Fundamentally, however, study of TS structure requires a quantum mechanical (QM) description of the electronic behaviour of the system. Moreover, for comparison with experiments in aqueous solution it is necessary to employ a theoretical method which can model solvation effects.

Since one of the goals of this thesis is to satisfactorily reproduce experimental KIE data as a measure of the successful modelling of a reaction, it is timely to assess the merits of some modelling approaches using an appropriate system for which good experimental data is available. The chosen reaction is the acid-catalyzed hydrolysis of adenosine monophosphate (AMP) because of the availability of high quality experimental KIEs and the importance of the

equivalent enzyme-catalyzed reaction. In this section three different semiempirical MO methods for modelling TS structure and mechanism in aqueous solution are compared to the equivalent gas phase calculation and to the results of a BEBO method. The criteria for successful modelling will be the accuracy to which experimental KIEs are reproduced and the plausibility of the obtained TS structures within the current state of knowledge about glycosidic hydrolysis.

6.2 Modelling of AMP hydrolysis

Schramm and co-workers measured KIEs for acid-catalyzed hydrolysis of AMP (Figure 6.2) with isotopic substitutions at four positions[5] and also performed BEBOVIB calculations[6] on a cut-off model system to obtain estimates for the TS bond orders.[7] By modelling this reaction using several semiempirical MO methods that simulate aqueous solvation, the effectiveness of each method can be assessed and the TS structures obtained can be compared to the BEBOVIB method. Figure 6.3 shows the cut-off model system used and indicates the positions of isotopic substitution. In the simulations, the AMP model and one water molecule were modelled to simulate hydrolysis with inversion of configuration at the reaction centre. The AM1 Hamiltonian[8] was chosen

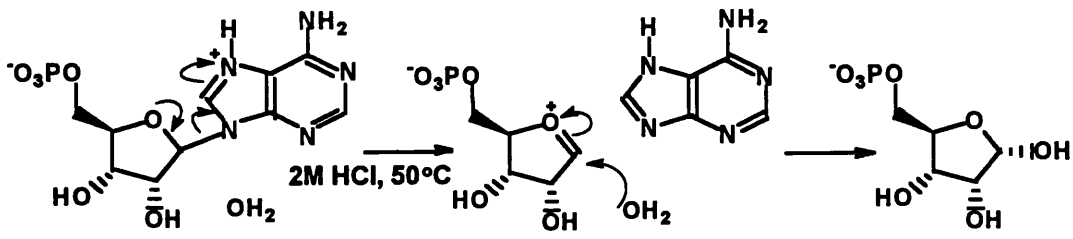


Figure 6.2 The hydrolysis of adenosine monophosphate (AMP)

because it was the only parameterization common to all three solvation treatments employed, viz. AMSOL[9], COSMO[10] and CHARMM.[11] AMSOL (AM1-SM1) adds a free energy of solvation term to the *in vacuo* heat of formation. The version of AMSOL used was a locally modified version which was tested and used as recommended in Appendix 1. COSMO uses a continuum screening method to calculate the heat of formation in solution; in the present simulation the solvent is water. CHARMM performs hybrid QM/MM calculations, combining a QM description of the reacting solute and nucleophile with a molecular-mechanical (TIP3P, see section 4.4.1) description of explicit water molecules within a sphere of defined radius. Each method was used in the standard way as described in Chapter 4. The first step for each method was to generate a grid of points by fixing the forming and breaking bond lengths at appropriate values and then minimizing the energy of the system with respect to all other variables. This grid was used to generate an energy surface on which

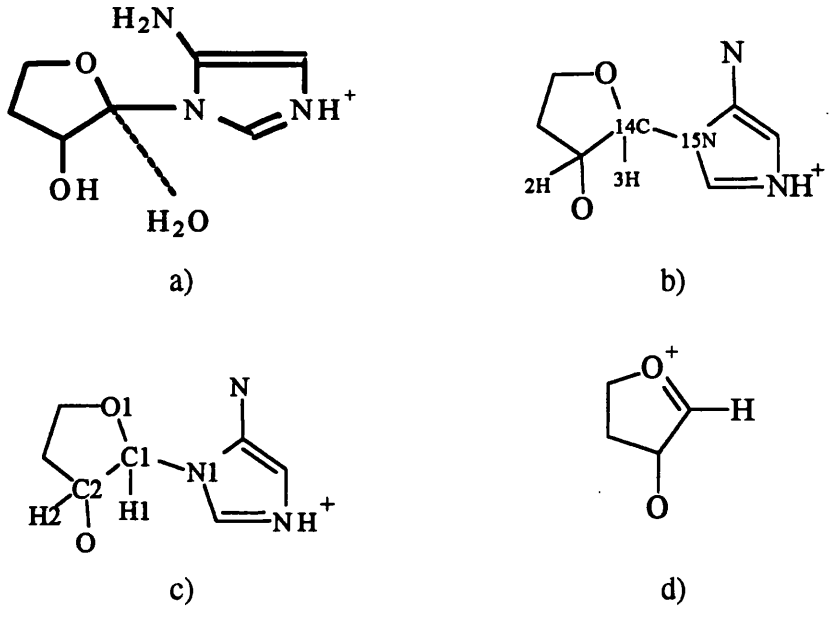


Figure 6.3. Descriptions of the AMP system. a) The model system; b) Positions and type of isotopic substitution; c) Atom labels referred to in Tables; d) The oxocarbenium ion.

stationary points could be approximately located. The closest appropriate grid point was used to initiate a gradient search for an accurate stationary structure. All TS structures had a single imaginary frequency. For the CHARMM surface the energy plotted is the sum of the QM electronic and QM/MM van der Waals energies. The MM-only energy is very large compared to the QM components because of the far greater number of atoms in the MM system, and is effectively constant. The major potential energy changes are reflected in the variation of the QM components of the energy. The energy surfaces are reproduced in Figure 6.4 and energy profiles for the identified reaction paths are in Figure 6.5. The TS structures are shown in Figure 6.6.

6.3 Comparative Evaluation of Solvation Models

The relative calculation times for the entire simulation with each method were: AM1, 1.0; AMSOL, 5.4; COSMO, 2.5; CHARMM, 5.0.

The AM1 *in vacuo* and COSMO simulations gave no problems.

The AMSOL method often gave very high gradients for structures, requiring considerable additional effort to reduce these to acceptably small values. This problem was noted in our preliminary tests (Appendix I) and is also accepted by the authors of the program[9] in their documentation. Stationary points were difficult to locate and characterize using AMSOL and often did not seem to correspond with features observed on the interpolated surface. This did not happen with the other methods used. These difficulties with AMSOL are manifested in the very irregular surface generated. Many of the AMSOL surface features are probably artefacts caused by these problems.

CHARMM has the capability to perform molecular dynamics with the combined QM/MM potential. This option was not used because of the large

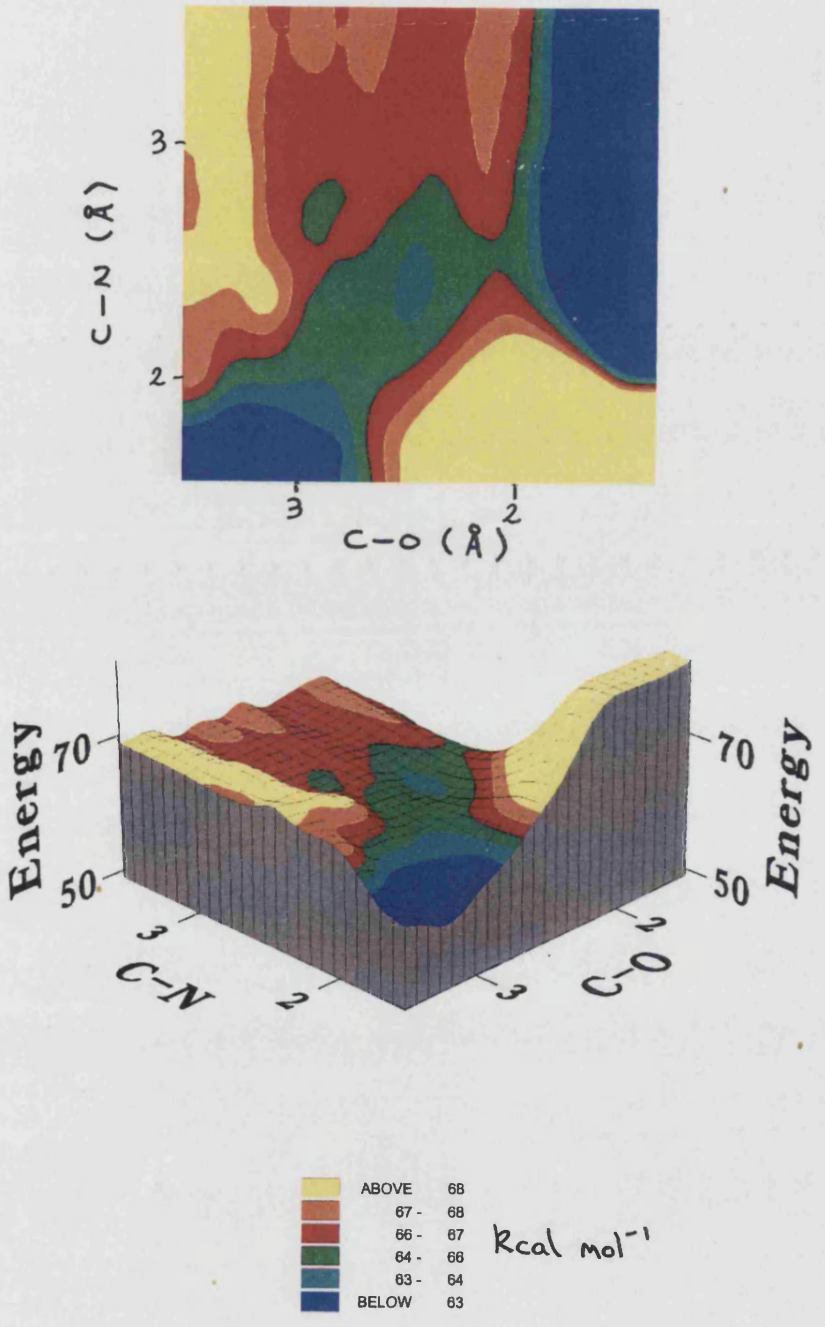


Figure 6.4a. The AM1 PES shown in 2d and 3d projections.

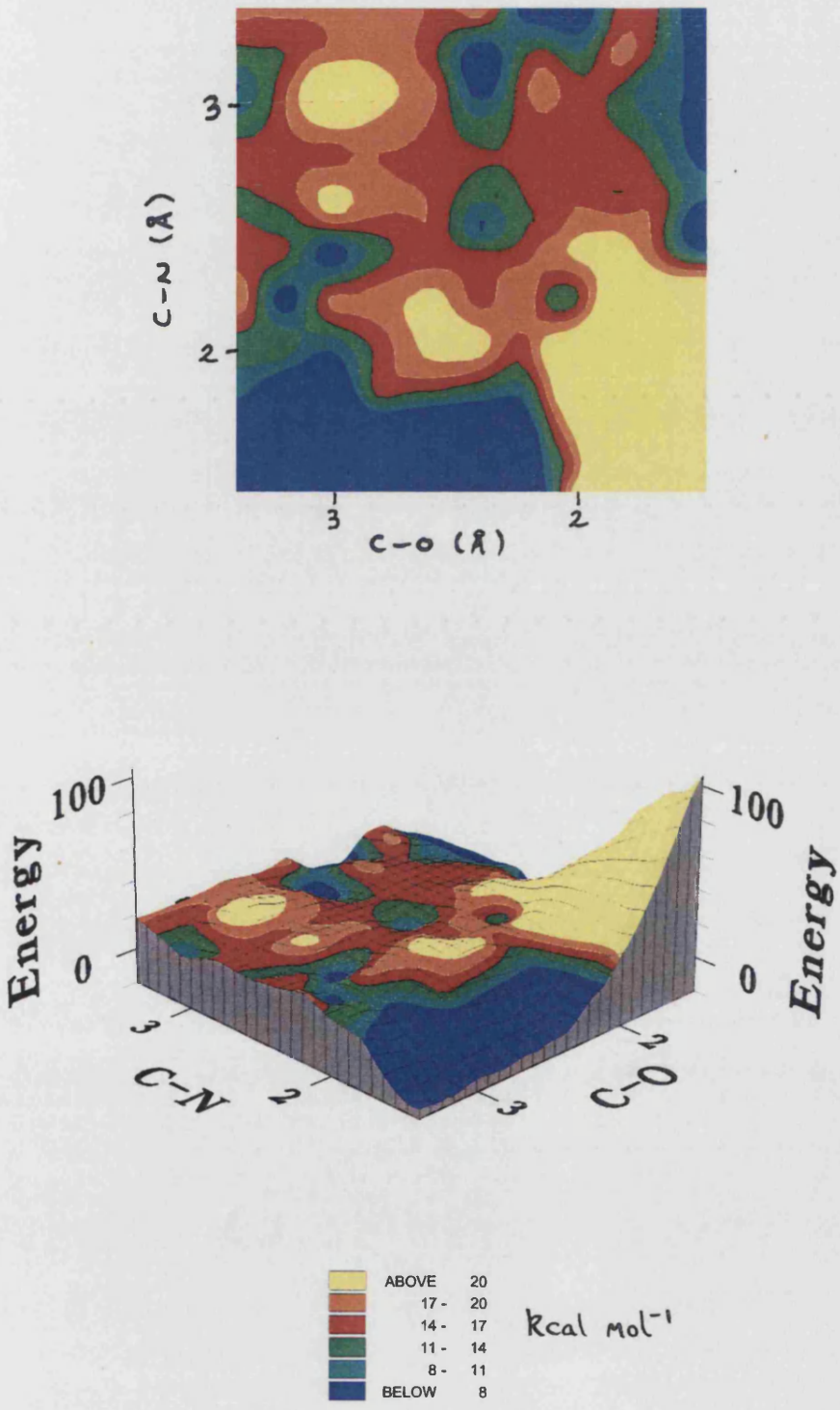


Figure 6.4b. The AMSOL PES shown in 2d and 3d projections

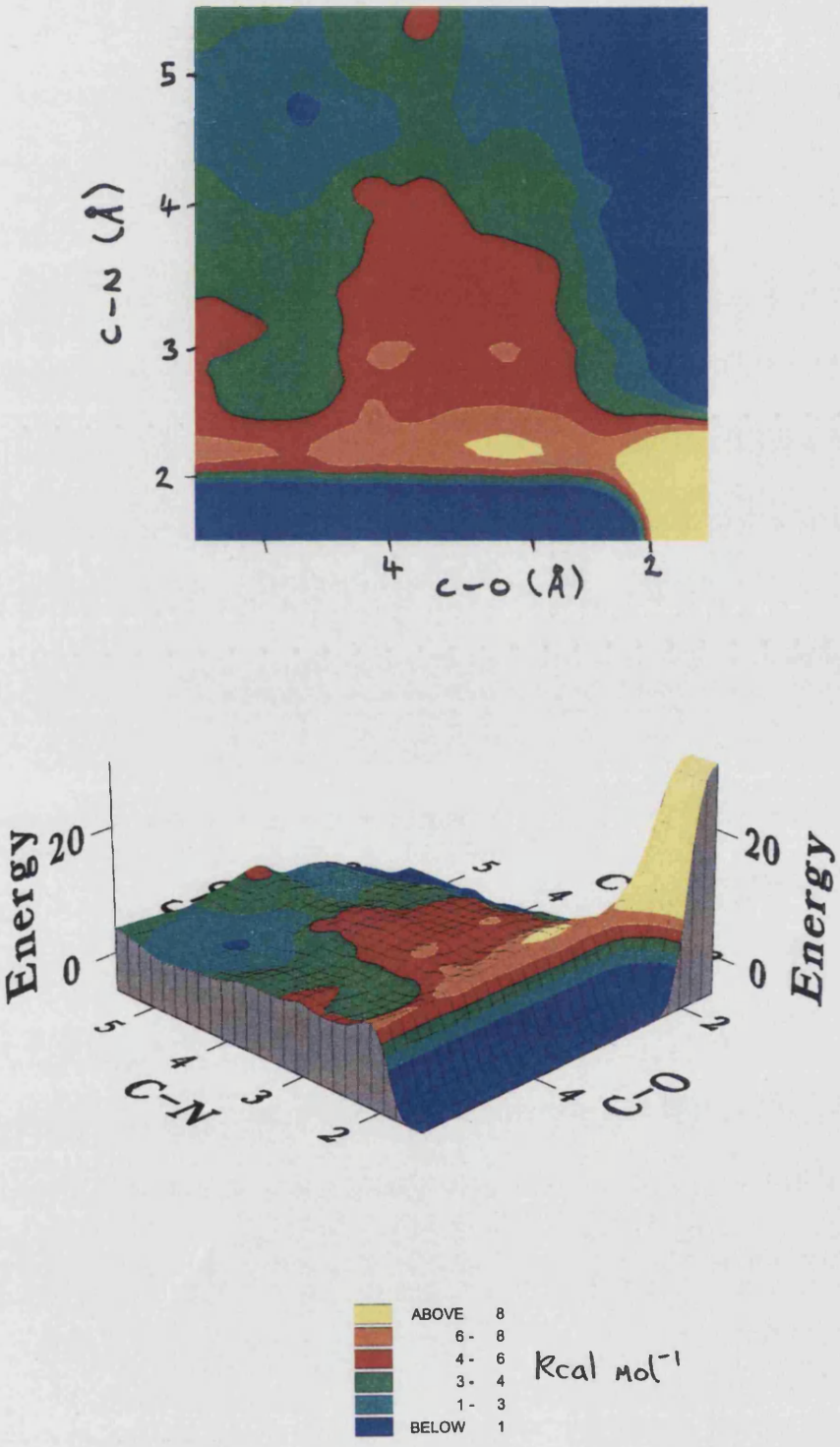


Figure 6.4c. The COSMO PES shown in 2d and 3d projections.

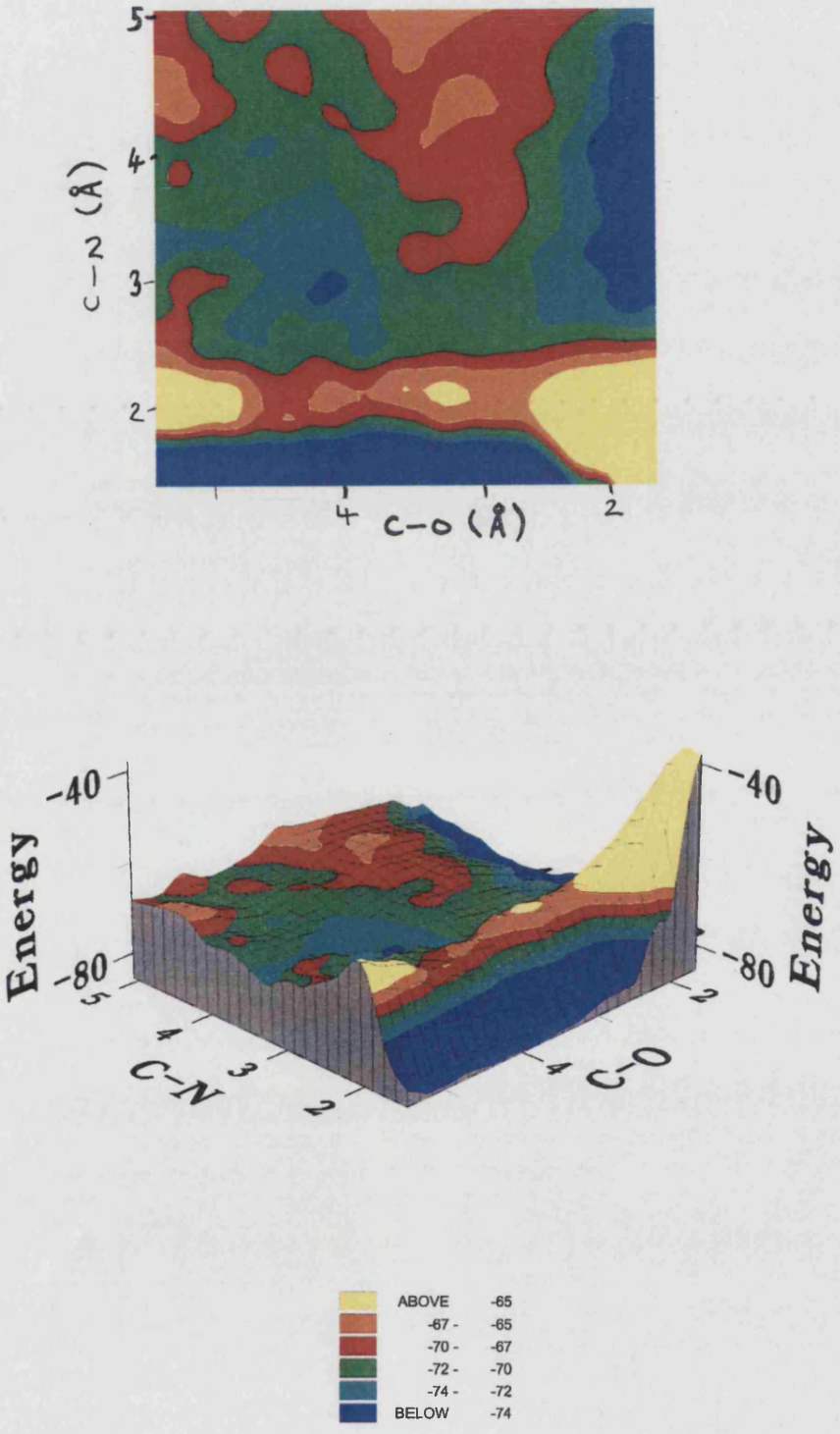


Figure 6.4d. The CHARMM PES shown in 2d and 3d projections.

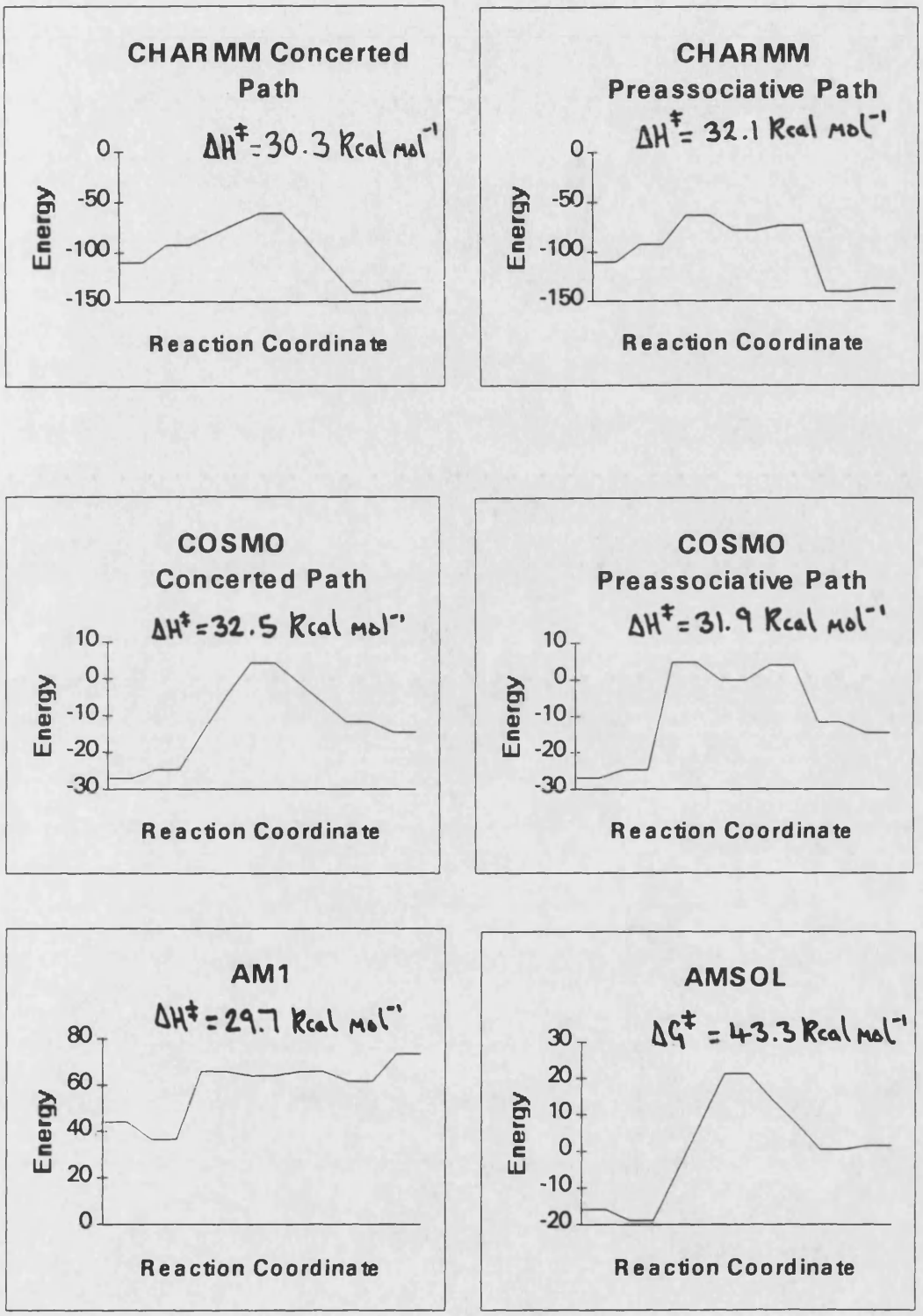


Figure 6.5. Energy profiles for the reaction pathways identified using AM1, AMSOL, COSMO and CHARMM.

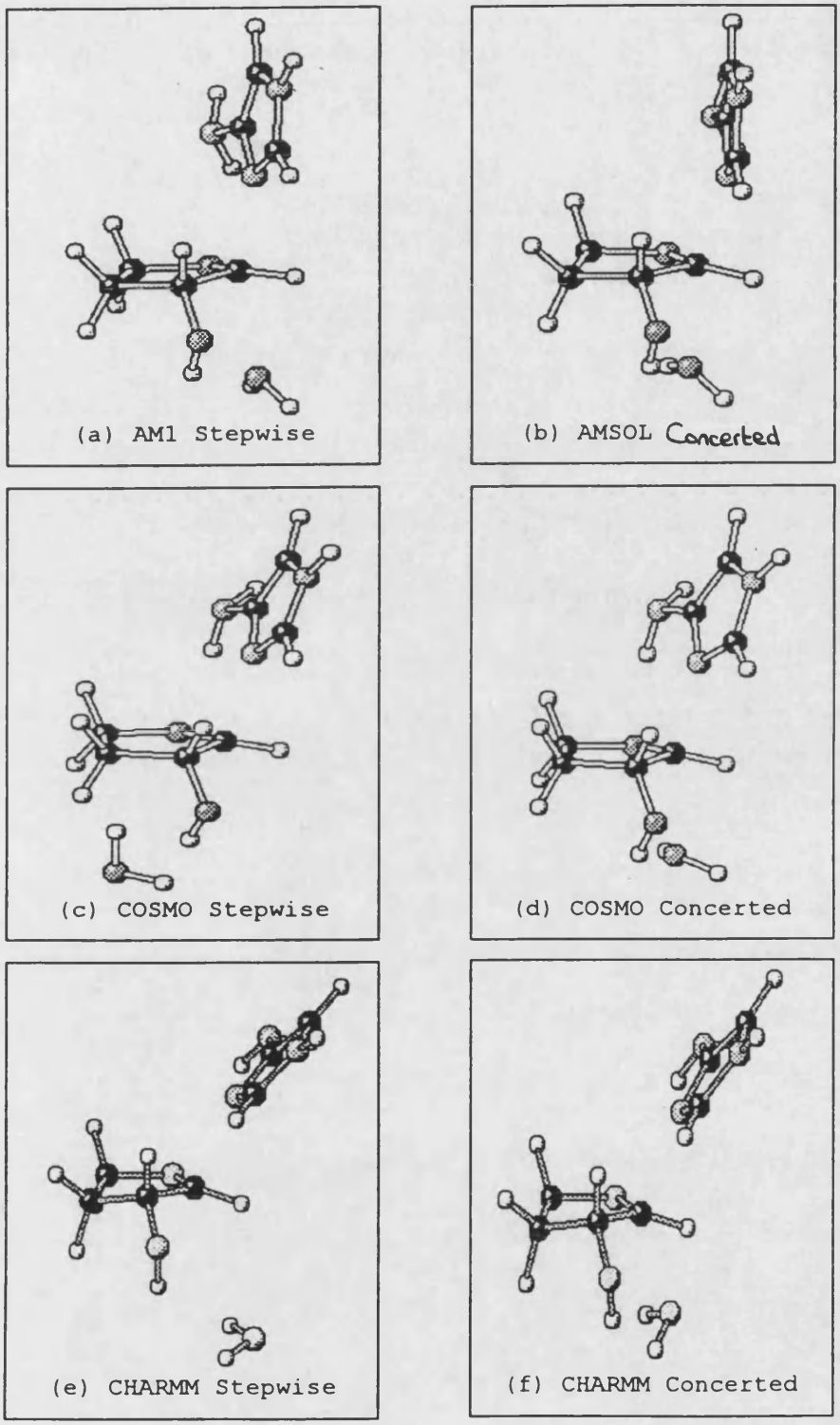


Figure 6.6. TS structures found for the reaction pathways identified using AM1, AMSOL, COSMO and CHARMM.

computational resource required. Instead, geometry optimization to a local minimum was performed at each point on the grid. For this reason, the stationary points may not have the best solvent distribution around the solute. The minimization algorithms in CHARMM do not locate saddle points, so it was usually necessary to do a very fine search around the approximate location of a TS, identified from the PES generated; effectively this was a second grid calculation. The time required for this is reflected in the relative calculation time reported. A locally modified version of CHARMM was used so that the matrix of second derivatives for the QM atoms could be calculated; this is not a standard feature. Details of my modifications can be found in Appendix II.

6.4 Results

The surfaces generated by each method have enabled the major stationary points along the reaction coordinate to be identified. The energies of these points have been profiled in Figure 6.5. COSMO and CHARMM both show an increase in energy going from separated reactants to associated reactants, which is presumably the energy required to release a water from bulk solvent to a position where it can act as a nucleophile.

6.4.1 Transition State Structures and Reaction Pathways

The surfaces generated using AM1, COSMO and CHARMM show the existence of a stepwise reaction path. In the case of AM1 *in vacuo* this mechanism would appear to involve pre-association, since there is a small but distinct Pauling bond order to the nucleophile in the TS (Table 6.1). In contrast, the COSMO and CHARMM methods suggest a classical S_N1 mechanism with essentially zero bond order to the incoming water nucleophile in the rate-

Table 6.1. Pauling Bond Orders[a] for Stationary Points Along The Reaction Coordinate of The AMP Hydrolysis

Point	Bond	AM1[b]	AMSOL	COSMO	CHARMM	BEBOVIB[c]
Concerted TS	C1-N1	-.--	0.13	0.12	0.10	0.02
	C1-H1	-.--	1.11	1.09	1.09	1.00
	C1-O1	-.--	1.43	1.38	1.41	1.71
	C1-C2	-.--	1.16	1.15	1.21	1.05
	Water-C1[d]	-.--	0.03	0.05	0.02	0.02
	ΣC1	-.--	3.86	3.79	3.83	3.80
Stepwise TS	C1-N1	0.13	-.--	0.11	0.07	0.10
	C1-H1	1.11	-.--	1.08	1.09	1.00
	C1-O1	1.37	-.--	1.41	1.43	1.85
	C1-C2	1.16	-.--	1.16	1.21	1.05
	Water-C1[d]	0.03	-.--	0.00	0.00	-.--
	ΣC1	3.81	-.--	3.76	3.80	4.00
Stepwise Intermediate	C1-N1	0.02	-.--	0.00	0.02	
	C1-H1	1.11	-.--	1.09	1.11	
	C1-O1	1.46	-.--	1.51	1.54	
	C1-C2	1.19	-.--	1.21	1.18	
	Water-C1[d]	0.08	-.--	0.00	0.00	
	ΣC1	3.86	-.--	3.81	3.85	
Cation[e]	C1-H1	1.09	1.05	1.09	1.11	
	C1-O1	1.51	1.50	1.51	1.52	
	C1-C2	1.19	1.22	1.21	1.18	
	ΣC1	3.79	3.77	3.81	3.81	

[a] bond order = $\exp[(r_1 - r_n)/0.3]$ where r_1 is bond length for unit bond order and r_n is bond length for bond order n. [b] *in vacuo*. [c] Stepwise model did not have a nucleophile. [d] Ground state bond length taken as the $[H_3C\text{-}OH_2]^+$ bond length calculated by each method. [e] The oxocarbenium ion, see Figure 6.3.

determining TS; the water is sufficiently far from the reaction centre to have no special interaction to distinguish it from bulk solvent. The charge distribution in the stationary points along the reaction coordinate are set out in Table 6.2. COSMO predicts that the intermediate structure is a fully fledged oxocarbenium ion, as comparison of the charges between the intermediate and cation show,

Table 6.2. Charge Distribution in the Stationary Points Along The Reaction Coordinate of The AMP Hydrolysis

Point	Atom	AM1[b]	AMSOL	COSMO	CHARMM
Ground State	N1	-0.16	-0.23	-0.18	-0.17
	C1	0.10	0.11	0.11	0.11
	O1	-0.28	-0.29	-0.30	-0.28
	C2	0.01	0.01	0.00	0.00
	H1	0.16	0.17	0.19	0.17
	H2	0.16	0.15	0.17	0.18
	N-ring	0.67	0.72	0.72	0.73
	O-ring	0.33	0.28	0.28	0.27
	Water	0.00	0.00	0.00	0.00
Concerted TS	N1	-.--	-0.31	-0.31	-0.32
	C1	-.--	0.31	0.30	0.31
	O1	-.--	-0.13	-0.17	-0.14
	C2	-.--	0.01	0.00	-0.02
	H1	-.--	0.25	0.27	0.25
	H2	-.--	0.18	0.19	0.20
	N-ring	-.--	0.14	0.03	0.09
	O-ring	-.--	0.85	0.88	0.91
	Water	-.--	0.01	0.09	0.00
Stepwise TS	N1	-0.29	-.--	-0.31	-0.31
	C1	0.30	-.--	0.30	0.29
	O1	-0.13	-.--	-0.15	-0.14
	C2	0.01	-.--	-0.01	-0.01
	H1	0.23	-.--	0.28	0.25
	H2	0.18	-.--	0.20	0.20
	N-ring	0.17	-.--	0.15	0.09
	O-ring	0.83	-.--	0.85	0.91
	Water	0.00	-.--	0.00	0.00
Stepwise Intermediate	N1	-0.27	-.--	-0.34	-0.26
	C1	0.30	-.--	0.29	0.27
	O1	-0.09	-.--	-0.08	-0.03
	C2	0.01	-.--	-0.02	-0.01
	H1	0.24	-.--	0.30	0.25
	H2	0.19	-.--	0.21	0.19
	N-ring	0.01	-.--	0.01	0.01
	O-ring	0.95	-.--	0.99	0.99
	Water	0.04	-.--	0.00	0.00
Cation[a]	C1	0.24	0.32	0.32	0.24
	O1	-0.05	-0.05	-0.10	-0.04
	C2	-0.01	0.01	-0.02	0.01
	H1	0.26	0.35	0.29	0.26
	H2	0.17	0.17	0.20	0.16

[a] The oxocarbenium ion, see Figure 6.3

with zero bond orders to the nucleophile and the leaving group. However, CHARMM suggests the intermediate is an ion-dipole complex in which the ion is stabilized by the leaving group before the approach of the nucleophile.

The AMSOL PES is unreliable due to the difficulties already mentioned. One TS structure was identified and characterized. Use of the IRC path following option showed that the TS was on a concerted pathway and did connect reactants to products. However, the IRC path does not share the features of the PES when superimposed. The general structure of the AMSOL TS for this concerted path shares the general features of the CHARMM and COSMO TS for the same path, and these are described below.

The AMSOL, COSMO and CHARMM surfaces show the presence of a concerted reaction pathway. However, the TS structures for these S_N2 mechanisms do not quite correspond with the usual expectation: inspection of their bond orders reveals extremely low values to both the leaving group and the nucleophile. The positive charge on the central moiety of the TS is high, indicating an oxocarbenium ion-like structure similar to the intermediate of a stepwise process. Chapter 7 provides an explanation of how the destabilization of an intermediate can reduce its well-depth to zero and thereby cause three stationary points on a surface (two saddle points and a minimum) to coalesce to a single saddle point corresponding to an S_N2 TS with essentially the same structure as an S_N1 intermediate.[13] Here, however, the S_N2 TS on the COSMO and CHARMM *in aquo* surfaces has a structure similar to the S_N1 intermediate on the *in vacuo* surface, and exists together on the same surface with an S_N1 intermediate of distinctly different structure. The S_N2 TS structures for our simulated AMP hydrolysis appear to resemble the "exploded" TS for the

solvolysis of D-glucopyranosyl derivatives suggested by Sinnott and Jencks.[14] The activation enthalpies of the two processes are similar for both the COSMO and CHARMM simulations (Figure 6.5), suggesting that both mechanisms could be followed concurrently. It is important to note that these surfaces are enthalpy only and do not explicitly include entropic effects. Consideration of entropy would lead to the expectation that association complexes would be disfavoured. It could be argued that because the apparent reaction coordinate on the PES for the concerted path involves the approach of the water before any significant movement of the leaving group, there is an entropic cost which is not reflected on the PES: this may bias the surface to favour the S_N1 pathway.

The main difference in the TS structures for the two distinct paths in both the COSMO and CHARMM simulations is the low bond order to the nucleophile for the concerted path. This shows that subtle solvation changes between the ground and transition states may play a role in deciding which pathway dominates. This could possibly be measured experimentally with a solvent isotope effect, although none have so far been reported for this reaction.

6.4.2 Kinetic Isotope Effects

The test of the theoretically determined structures is in how well they reproduce experimental KIEs. Each isotopic substitution gives information about the change in bonding of that atom in the TS. The 1^0 ^{14}C and ^{15}N isotope effects indicate the reaction coordinate motion of the atoms. The larger the ^{14}C effect, the more S_N2 character is reflected in the carbon motion, since the carbon would be moving towards the nucleophile and would have a significant bond order to the leaving group. In an S_N1 mechanism the reaction coordinate is mainly movement of the leaving group away from the oxocarbenium ion. Thus for the

1^0 ^{15}N effect, a large value reflects S_N1-like character with a low bond order to the leaving group. The α-^3H 2^0 effect reflects the out of plane bending of the hydrogen in the TS. This motion would be more restricted in an S_N2-like mechanism because of the relative proximity of nucleophile and leaving group, thus a low value would reflect S_N2 character and a high value S_N1 character.

The β-^2H 2^0 effect originates from changes in vibrational frequency involving the hydrogen and has two components, hyperconjugation and the inductive effect of deuterium. β-deuterium effects can be expressed as follows

$$\ln (K_H/K_D) = \cos^2\emptyset.\ln(K_H/K_D)_{max} + \ln (K_H/K_D)_i$$

The first term represents hyperconjugation of the C-H sigma orbital with an electron deficient p orbital on an adjacent carbon atom, where $\emptyset$ is the dihedral angle between the orbitals. The maximum ratio is obtained when the orbitals are eclipsed, and also increases as the positive charge on the adjacent atom increases with the associated weakening of the C-H bond. The second term represents a geometry independent inductive effect of the deuterium atom, reflecting the electron-donating effect of deuterium. Significant oxocarbenium ion character should therefore be reflected in a high β-H secondary KIE. Magnitudes of KIEs can be found in several texts[15] and a summary of expected values is presented in Table 6.3. Table 6.4 shows the calculated KIEs for each method. The AM1 *in vacuo* model overestimates the contribution of the ^{14}C motion to the reaction coordinate. The low β-2^o ^2H value is mainly due to the dihedral angle between the C-H bond and the anomeric carbon p-orbital: a

Table 6.3. Expected Range of Kinetic Isotope Effects for Glycosidic Hydrolysis

	Character of TS	
Isotope	S_N1	S_N2
1^o ^{14}C	1.00-1.05	1.09-1.14
1^o ^{15}N	1.02-1.04	1.01-1.02
α-2^o ^{3}H	1.10-1.40	1.00-1.10
β-2^o ^{2}H	1.05-1.15	1.00-1.02

Table 6.4. Kinetic Isotope Effects for the AMP Hydrolysis

Isotope/Process	Model				
	AM1	AMSOL	COSMO	CHARMM	EXP'T
Stepwise					
1^0 ^{14}C	1.075		1.091	1.045	1.044
α-^{3}H	1.146		1.282	1.211	1.216
1^0 ^{15}N	1.026		1.027	1.028	1.030
β-^{2}H	1.046		1.053	1.074	1.077
Concerted					
1^0 ^{14}C		1.083	1.091	1.044	
α-^{3}H		1.116	1.189	1.164	
1^0 ^{15}N		1.027	1.028	1.028	
β-^{2}H		1.029	1.047	1.071	

dihedral of 35.2° compared to 16.4° in the TS structure from the more accurate CHARMM simulation. The AM1 TS is too S_N2-like in character. These deficiencies are also observed in the AMSOL predictions. The COSMO

stepwise path also overestimates ^{14}C motion in the reaction coordinate at the TS. This is reflected in the bond order to the leaving group which is still over 0.1. The fact that the hydrogen motion is unhindered by the nucleophile at the TS accounts for the high α-^{3}H value. The concerted path is similar, but the extra hindrance by the incoming nucleophile attenuates the α-^{3}H effect. The CHARMM results are quite impressive, with good agreement to experiment for both pathways. The low bond order to leaving group in both pathways relieves the carbon from motion in the TS. The β-^{2}H effect is reproduced well, due mainly to the low dihedral angle of 16.4° between the C-H bond and the anomeric carbon p-orbital. The ^{15}N value is also in agreement. The stepwise pathway reproduces the α-^{3}H effect better than the concerted pathway, suggesting that the stepwise pathway is the most likely one. This contrasts with the prediction made from consideration of entropy made earlier. The fact that the TS for the concerted path is "exploded" and resembles more a classical S_N1 TS, plus the good general agreement with the calculated and experimental KIEs suggests that the two mechanisms could occur concurrently. It is possible that measurement of a solvent isotope effect may allow one path to be positively identified.

6.4.3 Comparison with BEBO TS Structure

The bond energy bond order (BEBO) method used in the BEBOVIB-IV[6] program was developed from the ideas of Johnston.[16] The method essentially consists of the assignment of a Pauling-type bond order to each bond in the model system. Atomic masses are also input because it is the variation of these

which gives rise to the isotope effect. Chemical knowledge is used to supply constraints, which can come from X-ray or other structural data, so only a few bonds are actually independent parameters. The independent bonds are usually chosen to be those that are made or broken during the reaction being modelled. Reactant and transition state vibrations are approximated by assuming harmonic potentials for each vibrational degree of freedom. These are input as the diagonal elements of a force constant matrix and may be varied. Off-diagonal elements may also be used, but have been shown to have little effect on calculation of KIEs.[17] Once the basic model is set up, the bond orders for the model TS are varied. Contributions to isotope effects from changes in zero point energies (ZPE), mass moment of inertia (MMI) and populations of excited states (EXC) are calculated by the program. The structure which best succeeds in reproducing experimental KIEs is taken to be the TS structure for the reaction. The success of the method lies in the predictable effects of an isotopic substitution, with calculated KIEs derived from transition state theory[18] as expressed by Bigeleisen,[19]

$$KIE = MMI * EXC * ZPE$$

The BEBOVIB modelling of Schramm and co-workers assumed a collinear arrangement of the bonds to the nucleophile and leaving group. As the results in Table 6.1 show, there is a reasonable degree of agreement between the MO and BEBOVIB TS bond orders for these two bonds and for the bond between the carbon at the reaction centre and the oxygen in the ring. This suggests that when used carefully BEBOVIB can provide useful estimates of the TS bond orders. It cannot of course answer questions about charge distribution and reactivity; that is the preserve of QM calculations.

6.5 Conclusions

These simulations are exploring molecular behaviour right at the S_N1/S_N2 mechanistic borderline. The AM1 method predicts the existence of a stepwise pathway and AMSOL a concerted pathway. COSMO and CHARMM predict that a stepwise and a concerted pathway can occur. The energy balance between these two mechanisms is a fine one, and quite possibly both occur concurrently. This is supported by the CHARMM simulation, with its similar TS for both pathways and the calculated KIEs for both paths which are in excellent agreement with experiment. It is possible that measurement of a solvent isotope effect may allow one path to be positively identified.

The calculated KIEs improve in the order AMSOL, AM1, COSMO, CHARMM. The AMSOL results are poor, due to the problems with obtaining stationary points with low gradients. It is not recommended that this method be used to model solvated reaction mechanisms using the grid approach taken here. COSMO seems adequate for general work. CHARMM is recommended for the most accurate work. BEBOVIB can be used in conjunction with experimental KIE data to obtain information about TS bond orders which is comparable in quality to the semiempirical MO methods presented here.

6.6 References

1. M. L. Sinnott in The Chemistry of Enzyme Action, Ed. M. I. Page, Elsevier 1984
2. N. S. Benait & W. P. Jencks, *J. Am. Chem. Soc.*, 1991, **113**, 7051
3. G. A. Craze & A. J. Kirby, *J. Chem. Soc. Perkin 2*, 1978, 354
4. T. L. Amyes & W. P. Jencks, *J. Am. Chem. Soc.*, 1989, **111**, 7888

5. D. W. Parkin & V. L. Schramm, *Biochemistry*, 1987, **26**, 913

6. L. B. Sims, G. W. Burton & D. E. Lewis, BEBOVIB, *QCPE #337* (1977)

7. F. Mentch, D. W. Parkin & V. L. Schramm, *Biochemistry*, 1987, **26**, 921

8. M.J.S.Dewar, E. G. Zoebisch, E. F. Healy & J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902; MOPAC v6.0 implementation

9. C. J. Cramer, D. G. Truhlar, *J. Am. Chem. Soc.*, 1991, **113**, 8305, 9901; AMSOL v1, QCPE #606

10. A. Klamt & G. Schuurmann, *J. Chem. Soc. Perkin 2*, 1993, 799; MOPAC 93 implementation

11. M. J. Field, P. A. Bash & M. Karplus, *J. Comp. Chem.*, 1990, **11**, 700; CHARMM v22, locally modified to produce second derivatives for the QM atoms. See Appendix II for details.

12. C. L. Brooks III & M. Karplus, *J. Chem. Phys.*, 1983, **79**, 6312

13. J. A. Barnes, J. Wilkie & I. H. Williams, *J. Chem. Soc. Faraday Trans.*, 1994, **90**, 1709

14. M. L. Sinnott & W. P. Jencks, *J. Am. Chem. Soc.*, 1980, **102**, 2026

15. a) W. W. Cleland, M. H. O'Leary & D. B. Northrop, Eds. (1977) Isotope Effects on Enzyme-catalyzed Reactions, pub Univ. Park Press, Baltimore; b) L. Melander & W. H. Saunders Jr., (1980) in Reaction Rates of Isotopic Molecules, pub Wiley, NY

16. H. S. Johnston, Gas Phase Reaction Rate Theory, pub Ronald Press, NY, 1966, p179

17. G. W. Burton, L. B. Sims, J. C. Wilson & A. Fry, *J. Am. Chem. Soc.*, 1977, **99**, 3371

18. S. Glasstone, K. J. Laidler & H. Eyring, The Theory of Rate Processes, pub McGraw-Hill, NY, 1941

19. J. Bigeleisen, *J. Chem. Phys.*, 1949, **17**, 675; J. Bigeleisen & M. G. Mayer, *J. Chem. Phys.*, 1947, **15**, 261

CHAPTER 7.
MECHANISTIC CHANGE AT THE S_N1/S_N2 BORDERLINE

"We used to think that if we knew one, we knew two, because one and one are two. We are finding that we must learn a great deal more about and."

- Sir Arthur Stanley Eddington

7.1 Introduction

How does a mechanism change, and how is it related to transition state (TS) structure? The previous chapter has shown that the simple mechanistic categories of S_N1 and S_N2 become blurred when the intermediate of an S_N1 mechanism becomes unstable and requires some stabilizing influence from the nucleophile. It becomes a matter of interpretation as to whether the process is S_N1 or S_N2. This chapter aims to show how the change from S_N2 to S_N1 can occur, and to show how TS structure can vary as the relative energies along the reaction coordinate change. This will be done by reference to More O'Ferrall - Jencks (MOFJ) diagrams and the use of semiempirical AM1 MO calculations for two sets of degenerate displacement reactions on α-substituted alkyl substrates.

7.2 More O'Ferrall - Jencks Diagrams

The MOFJ diagram[1,2] is essentially a two-dimensional plot with an implied third dimension of energy, rather similar to a two-dimensional PES without the explicit contouring of the energy. Figure 7.1 shows one example of a MOFJ diagram. The axes are Pauling bond orders (defined as $n = \exp\{[d_1 - d_n]/c\}$, where d_1 and d_n are bond lengths for bonds of order unity and n respectively, and the constant $c = 0.3$) rather than the interatomic distances used in the PESs in the previous chapter. Progress from reactants (bottom-left corner of Figure 7.1) to products (top-right corner) may occur in principle by two-step mechanisms involving an intermediate formed either by dissociation (top-left

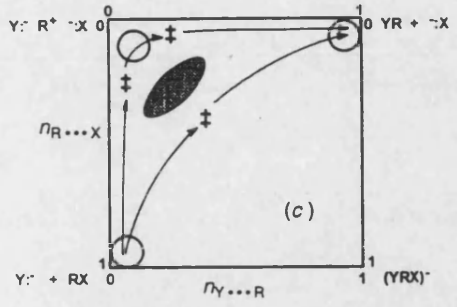


Figure 7.1. A MOFJ diagram for the exchange reaction Y^- + RX -> YR + X^-, showing the S_N1 and S_N2 pathways.

corner); alternatively, reaction may occur by a one-step process in which dissociation of the nucleofuge is concerted with association of the nucleophile - the classical S_N2 mechanism. The S_N1 and S_N2 routes are shown on Figure 7.1. It was probably Ingold's diagram[3] that was the first example of this approach to discussion of nucleophilic substitution mechanisms on saturated carbon substrates, but more recently Young and Jencks[4] and Harris et al[5] have also provided examples of its use.

7.3 Transition State Structural Variation

To use MOFJ diagrams as a qualitative tool for prediction of TS structural changes, it is necessary only to imagine the contours corresponding to a saddle-shaped PES: both reactants and products are relatively low in energy, while both the dissociative and associative intermediates are relatively high in energy. With reference to Figure 7.2, the TS is located at the saddle point, as marked by the symbol ‡; the energy profile along the reaction coordinate (solid line with arrows) shows a maximum at this point whereas the energy profile in the perpendicular direction (dashed line) shows a minimum here.

Substitution of electron-donating groups at α-C should stabilize the carbocation R^+ component of the dissociative intermediate (top-left corner) without significantly affecting the relative energetics of the reactants and

products. Application of the Hammond Postulate[6] in the direction parallel to the reaction coordinate suggests no change in the location of the TS, but consideration of the "anti-Hammond" effect of stabilizing the carbocationic intermediate suggests that the TS structure should shift in this perpendicular direction towards the top-left corner of the diagram.

7.3.1 Theoretical Modelling of Transition State Structural Variation

The following model system was used to test the above MOFJ prediction about the change in TS structure,

$$Cl^- + CLR'R''Cl \rightarrow ClCLR'R'' + Cl^- \qquad (7.1)$$

where L = H or D, R' = H or Me, and R" = H, Me or MeO. This reaction was modelled *in vacuo* using the AM1 method, using the standard conditions described in Chapter 4. The distance between the Cl^- nucleophile and α-C was used to define a reaction coordinate, and points were optimized along this reaction coordinate. Approximate TS structures were identified and refined.

Table 7.1 contains Pauling bond orders for TS structures for the range of α-substituted alkyl chlorides with chloride anion which were modelled. It also contains relative energies and semiclassical kinetic isotope effects for substitution of deuterium at α-C. The KIEs were calculated for an investigation[7] into the origin of secondary α-D KIEs that appears in the Interlude following this chapter. As the stability of the carbocationic intermediate increases (as

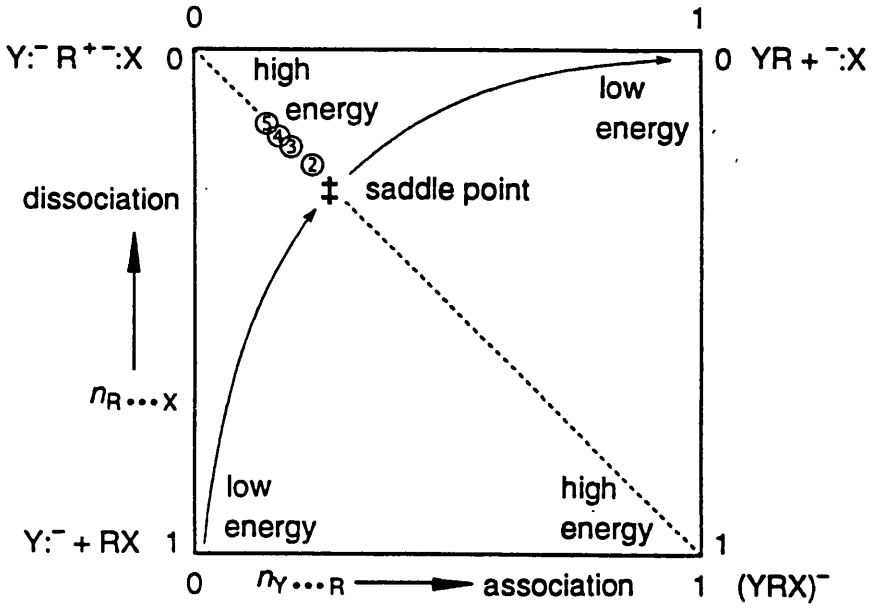


Figure 7.2. MOFJ diagram for nucleophilic substitution showing location of AM1 calculated TSs for degenerate S_N2 reactions of Cl^- with CHR'R"Cl, eqn (7.1), as indicated in Table 7.1.

measured by a decrease in the energy, ΔH_{hetero}, of heterolysis) so the bond orders $n_{C\cdots Cl}$ in the symmetrical TS diminish, indicative of a progressively "looser" S_N2 TS. The symbol ‡ in Figure 7.2 marks the location of the TS for the degenerate S_N2 reaction of chloromethane (R' = R" = H), and the encircled numbers mark the progression of TSs whose properties are listed in Table 7.1. Clearly, the shift in TS structure is just as predicted by the MOFJ diagram. Furthermore, the calculated secondary α-D KIE $(k_H/k_D)_s$ increases in magnitude, becoming more normal, as the stability of the carbocationic intermediate increases, in agreement with a wide range of experimental isotope effects[6,8] which were interpreted as evidence for an increasingly "loose" S_N2 TS. It is gratifying that the quantum-chemical results confirm the predictions of

Table 7.1. AM1 calculated semiclassical 2° α-D KIEs at 298K for degenerate S_N2 reaction of Cl^- with CLR'R"Cl (L = H, D), together with TS partial-bond lengths and Pauling bond orders, and energies (kJ mol^{-1}) of TS relative to reactant ion-molecule complex and of heterolysis CHR'R"Cl → CHR'R" + Cl^-

Mark on Fig. 7.2	R'	R"	C...Cl/Å	nC...Cl	$(k_H/k_D)s$	$\Delta H^{\dagger}(S_N2)$	ΔH_{hetero}
†	H	H	2.154	0.27	0.982	37.9	978
2	H	Me	2.202	0.23	1.015	69.0	859
3	H	MeO	2.252	0.19	1.048	68.2	776
4	Me	Me	2.270	0.18	1.090	91.3	759
5	Me	MeO	2.335	0.15	1.156	81.7	700

the MOFJ diagram and accord with the conclusions of experimental probes for TS structural change.

The MOFJ diagram predicts the change in TS structure successfully, but apparently not the change in energy. The values of $\Delta H^{\dagger}$, the intrinsic barrier height, shown in Table 7.1 do not vary in a simple manner, but are determined by the relative energies of the trivalent dissociative and pentacovalent associative intermediates that make the top left and bottom right extremes of the diagram, and also by specific contributions not modelled by the diagram. The quantitative data of Table 7.1 would very probably be altered if a different theoretical method were employed. However, it is unlikely that the trend in the results would be affected. For the purposes of the current study, it is sufficient to use an economical semiempirical MO treatment, provided that it yields topologically correct information regarding critical points of interest. The object is to examine the response of the topographical features of energy surfaces for organic systems to changes in reaction variables, not to obtain accurate descriptions of each individual reaction. The focus is on changes occurring

within families of reactions, rather than on one specific reaction.

7.4 A Limitation of MOFJ Diagrams

Despite the above success for a prediction made from consideration of the MOFJ diagram, there is a danger in using MOFJ diagrams to represent mechanistic change. This arises from the fact that (at best) the imagined PES is a projection of a multi-dimensional energy hypersurface onto the two-dimensional plane of the two coordinates chosen to construct the diagram. Each point on the diagram notionally represents the minimum-energy structure for the pair of coordinates used, having optimized the values of all other coordinates of the system. If, however, optimization of all the remaining coordinates yields more than one solution, then the PES corresponding to the MOFJ diagram is not uniquely defined, since it is not single-valued at every point. It is possible, for instance, to find an apparent TS structure which does not lead to the intermediate or product which defines the appropriate corner.[9] Williams and Maggiora have previously reported a two-dimensional energy-contour diagram for a reaction involving nucleophilic addition and proton transfer.[10] In this example there was a substantial region of the "surface" which was double-valued; selection of only the lower-energy structure at each point yielded an apparently reasonable "surface" showing a fine saddle. However, this concealed the existence of a profound discontinuity between reactant-like and product-like structures. The manner in which this artefact arose, by projection of reactant and product valleys lying in separate regions of hyperspace onto the chosen two-dimensional plane, was discussed in some detail.

7.5 Exploration of the S_N1/S_N2 Mechanistic Borderline

How does a mechanism change? This was the very question considered by More O'Ferrall in his seminal paper on the application of MOFJ diagrams to elimination reactions.[1] Jencks[11] has discussed possible ways by which an S_N1 mechanism could change to an S_N2 mechanism, which may be illustrated by the MOFJ diagrams shown in Figure 7.3. The S_N1 mechanism (Figure 7.3a) involves an intermediate in the top-left corner corresponding to an energy well on the PES, represented by an open circle; there is one TS for its formation and a second for its breakdown. Destabilization of the intermediate may cause the depth of this energy well to diminish and, as the lifetime of the intermediate becomes comparable with the period of a molecular vibration, the three critical points (two saddle points and one minimum) on the stepwise reaction path coalesce to a single saddle point, corresponding to a TS for a concerted reaction having a structure and energy essentially the same as that of the now-disappeared S_N1 intermediate. This possibility, illustrated by Figure 7.3b, was discussed by Doering and Zeiss[12] and may hold experimentally for nucleophilic substitution by azide anion on ring-substituted cumyl derivatives.[13] Alternatively, the S_N1 intermediate may co-exist with the S_N2 TS, being separated from it by an energy maximum, represented by the shaded region on Figure 7.3c. The S_N1 and S_N2 pathways occur in parallel and traverse different regions of the MOFJ diagram, despite being of equal energy at the point of mechanistic change. There is evidence that nucleophilic substitution by azide anion on 4-methoxybenzyl chloride in 70:30 (v/v) acetone/water proceeds by

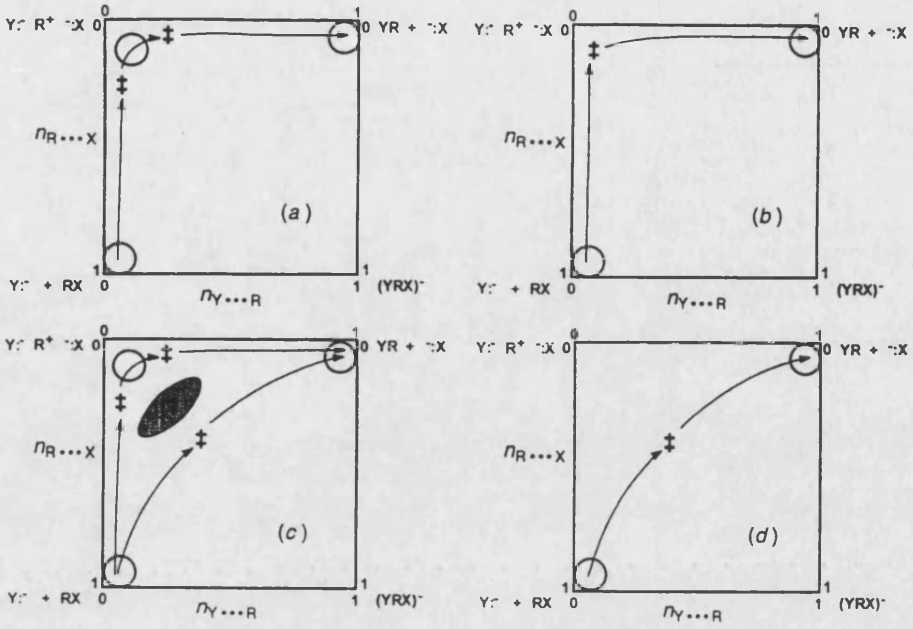


Figure 7.3. MOFJ diagrams depicting mechanistic change from (a) S_N1 to (d) S_N2 by means of either (b) S_N2 with TS having essentially the same structure as the disappeared S_N1 intermediate or (c) concurrent S_N1 and S_N2 pathways.

concurrent stepwise and concerted mechanisms.[14] Finally, the topological catastrophe accompanying the disappearance of the energy well on the S_N1 path may be such that the resulting saddle point for the enforced concerted S_N2 mechanism may appear at a location quite distinct from that of the destabilized S_N1 intermediate. There is, therefore, an abrupt change in mechanism (Figure 7.3a to 7.3d) with a structural discontinuity between the S_N1 intermediate and the S_N2 TS. Reactions of various nucleophiles with 1-phenylethyl derivatives seem to involve a mechanistic changeover of this nature.[25]

7.5.1 Theoretical Modelling of the S_N2 to S_N1 Mechanistic Change

If the above considerations are turned around, the question becomes how does the S_N2 mechanism change to S_N1? The degenerate reactions of α–substituted alkyl chlorides with chloride anion, which were modelled in section 7.3, involve a single, symmetrical TS in a one-step process. A two-step mechanism ought to be favoured by replacing Cl^- by the poorer nucleophile and better nucleofuge H_2O,

$$H_2O + CHR'R''OH_2^+ \rightarrow {}^+H_2OCHR'R'' + OH_2 \qquad (7.2)$$

choosing the substituent groups R' and R" to stabilize a carbocation. If the MOFJ diagrams in Figure 7.3 are accurate, the energy profile along the diagonal from the pentacovalent associative intermediate (bottom right) to the trivalent dissociative intermediate (top left) will show a single minimum for MOFJ diagrams like Figure 7.3a, 7.3b and 7.3d, corresponding to either an intermediate or a TS. The situation in Figure 7.3c, where both S_N1 and S_N2 pathways co-exist, should mean that the same diagonal profile will contain two minima, one corresponding to an intermediate and the other to a TS. Whichever situation is found, the minima may be characterised as either intermediates or saddle points on the full PES by calculating the vibrational frequencies of the structures found.

The calculations were done on system 7.2, using the same set of substituents as for the earlier chloride exchange reaction. The reaction coordinate was taken to be the distance between the nucleophile water oxygen and α-C. The leaving water's oxygen to α-C distance was set equal to the

reaction coordinate distance at each point that was optimized to ensure that the correct diagonal path was followed.

Figure 7.4 shows the AM1 in vacuo calculated diagonal energy profiles for reaction (7.2) with increasingly electron-donating α-substituents. Each profile contains only a single minimum, indicating that for this reaction in vacuo there are not concurrent S_N1 and S_N2 pathways. The position of the minimum shifts to longer partial C...O bond lengths with increasing alkyl substitution, consistent with stabilization of the top-left corner of the MOFJ diagram as found in the chloride exchange case. Introduction of an α-methoxy substituent leads to

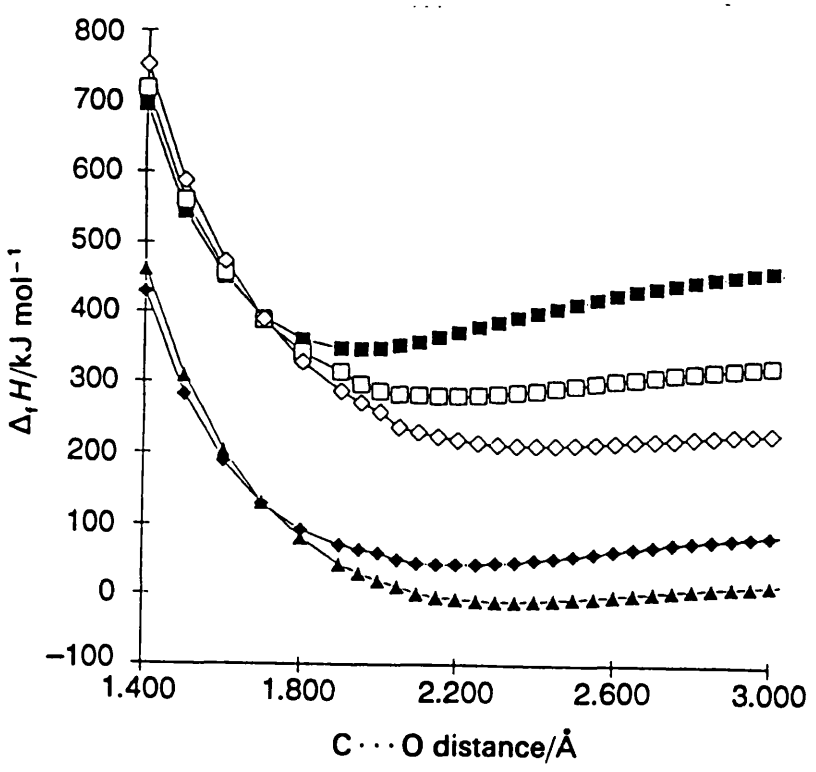


Figure 7.4. AM1 calculated diagonal energy profiles for degenerate S_N2 reaction of H_2O with $CHR'R''OH_2^+$ as a function of the making and breaking C...O bond distance (Å). Solid squares: R' = R" = H; open squares: R' = H, R" = Me; open diamonds: R' = R" = Me; solid diamonds: R' = H, R" = OMe; solid triangles: R' = Me, R" = OMe.

shifts in the C...O bond lengths out of step with those arising from simple alkyl substitution. It may be noted, however, that a methoxy group stabilizes not only the trivalent dissociative intermediate but also the pentacovalent associative intermediate in the bottom-right corner, thereby giving a different curvature along the diagonal profile. It is not immediately apparent from inspection of the profiles in Figure 7.4 whether a change in mechanism has occurred, but the vibrational frequencies reported in Table 7.2 for the asymmetric C...O stretching mode of the species at each of the minima reveal this information clearly. While the methyl and ethyl substrates react by the S_N2 mechanism with a symmetrical TS having an imaginary reaction-coordinate frequency, the remaining substrates react by the S_N1 mechanism involving an intermediate with all real vibrational frequencies. The curvature in this coordinate increases from substantially negative to substantially positive along the series listed in Table 7.2,

Table 7.2. AM1 calculated TS partial-bond lengths and Pauling bond orders for degenerate S_N2 reaction of H_2O with $CHR'R''OH_2^+$, together with vibrational frequency for reaction coordinate mode (asymmetric C...O stretching) and energies (kJ mol^{-1}) of S_N2 TS or S_N1 intermediate relative to reactant ion-molecule complex and well-depth of S_N1 intermediate relative to S_N1 TS.

R'	R"	C...O/Å	nC...O	$\nu^{\ddagger}$/cm^{-1}	$\Delta H^{\ddagger}(S_N2)$	$\Delta H^{\ddagger}(S_N1)$	ΔH_{well}
H	H	1.943	0.17	533i	55.6		
H	Me	2.144	0.09	238i	38.6		
H	MeO	2.194	0.07	165		5.0	0.6
Me	Me	2.407	0.04	289		10.4	4.4
Me	MeO	2.332	0.05	406			52.7

corresponding to longer C...O distances in the symmetrical stationary-point structure and a trivalent dissociative intermediate of increasing stability.

Plate 1 shows AM1 *in vacuo* calculated energy surfaces for degenerate reaction of H_2O with (a) $MeCH_2OH_2^+$ and (b) $MeOCH_2OH_2^+$ as a function of the making and breaking C...O bond distances (Å). The two surfaces are very similar in overall form, but the coloured energy contours reveal a striking topographical dissimilarity in the region of the S_N2 TS and S_N1 intermediate. The surface for the methoxymethyl substrate clearly contains a shallow well, only 0.6 kJ mol^{-1} deep, at essentially the same location as the saddle point on the surface for the ethyl substrate. In this case the change from S_N1 to S_N2 occurs in the manner of Figure 7.3a to 7.3b, without any structural discontinuity or the existence of concurrent S_N2 and S_N1 pathways. The intrinsic reaction coordinates for the two mechanisms are virtually superimposable forwards from reactants (or backwards from products) up to the point of the S_N1 TS; the paths diverge only minimally as the arcs through either the S_N2 TS or the S_N1 intermediate follow slightly different curvatures. The picture that emerges is a mechanistic change occurring according to the classical model of Doering and Zeiss.[12]

Inspection of Figure 7.4 shows that the methyl, ethyl and isopropyl substrates form one family, whereas the methoxymethyl and methoxyethyl substrates belong to another. The well-depth of the S_N1 intermediate for reaction of the isopropyl substrate, that is the difference in energy between this species and the TSs for its formation and destruction, is 4.4 kJ mol^{-1}. In view of

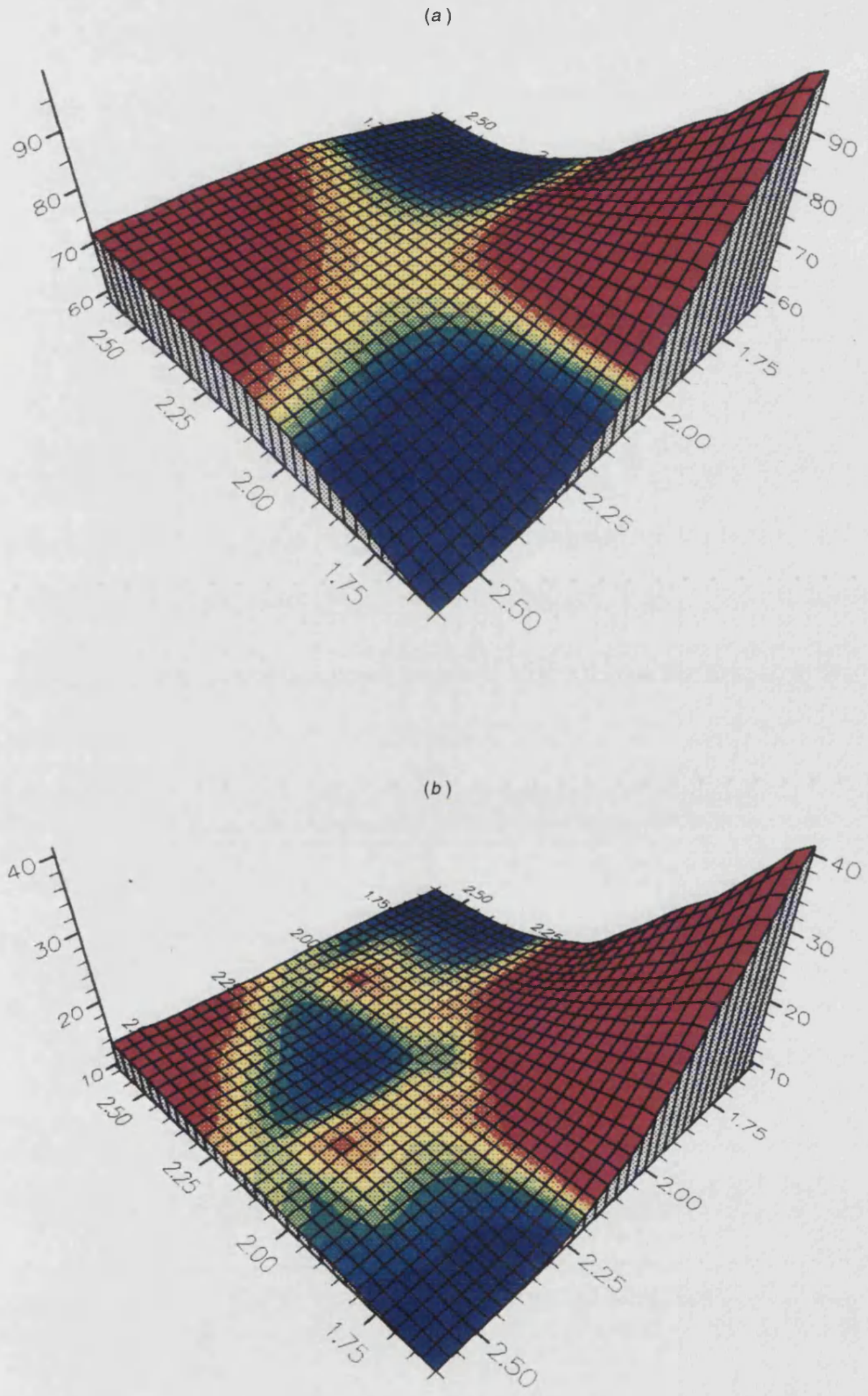


Plate 1 AM1 *in vacuo* calculated energy surfaces for degenerate S_N2 reaction of H_2O with (*a*) $MeCH_2OH_2^+$ and (*b*) $MeOCH_2OH_2^+$ as a function of the making and breaking C···O bond distances (Å). Vertical energy scale is in kcal mol^{-1}; low energies are coloured blue, high energies red.

the nature of the mechanistic change described above, it is reasonable to suppose that the energy of the (non-existent) S_N2 TS for this reaction would be approximately 2 x 4.4 kJ mol^{-1} higher than that of the S_N1 intermediate, corresponding to a negative curvature over the barrier the same as the positive curvature within the energy well of the intermediate. If so, then an intrinsic barrier height of about 19.2 kJ mol^{-1} would be expected for the isopropyl substrate, as compared with 38.6 kJ mol^{-1} for ethyl and 55.6 kJ mol^{-1} for methyl, suggesting a barrier-height reduction of about 18 kJ mol^{-1} for each α–methyl substituent. Note also that the intrinsic barrier decreases as the symmetric S_N2 TS becomes looser. This finding contrasts with the results of Wolfe and co-workers for degenerate S_N2 methyl transfer: variation of the nucleophile and nucleofuge leads to smaller intrinsic barriers for tighter TSs.[16] Comparison of the data in Tables 7.1 and 7.2 indicates a difference in behaviour between the two series of reactions. The S_N2 reactions with chloride anion as nucleophile and nucleofuge involve tighter TSs for corresponding substrates than those with water. The intrinsic barrier for methyl transfer is smaller for the tighter TS with chloride anion, in agreement with Wolfe's results, and successive α-methyl substituents raise the intrinsic barrier by 20 or 30 kJ mol^{-1}, for reactions of the ethyl and isopropyl substrates by means of looser TSs. This difference in behaviour may arise from the different overall charges for the two series of reactions: negative for reaction with chloride anion but positive for reaction with water as nucleophile and nucleofuge.

Finally, it may be noted that the S_N1 intermediate is the only stationary point

on the basin-shaped energy surface for the methoxyethyl substrate, being lower in energy than the isolated reactants H_2O + $CHMe(MeO)OH_2^+$ by 52.7 kJ mol^{-1}; despite the similarity of its diagonal energy profile (Figure 7.4) with those of the other substrates, the topography of its surface away from the diagonal is dramatically different.

7.6 References

1. R. A. More O'Ferrall, *J. Chem. Soc. B*, 1970, 274
2. W. P. Jencks, *Chem. Rev.*, 1972, **72**, 705
3. E. D. Hughes, C. K. Ingold & U. G. Shapiro, *J. Chem. Soc.*, 1936, 225
4. P. R. Young & W. P. Jencks, *J. Am. Chem. Soc.*, 1979, **101**, 3288
5. J. M. Harris, S. G. Shafer, J. R. Moffatt & A. R. Becker, J. *Am. Chem. Soc.*, 1979, **101**, 3295
6. G. S. Hammond, *J. Am. Chem. Soc.*, 1955, **77**, 334
7. J. A. Barnes & I. H. Williams, *J. Chem. Soc., Chem. Commun.*, 1993, 1286
8. T. Ando, H. Tanabe & H. Yamataka, *J. Am. Chem. Soc.*, 1984, **106**, 2084
9. J. A. Barnes, J. Wilkie & I. H. Williams, *J. Chem. Soc., Faraday Trans.*, 1994, **90**, 1709
10. I. H. Williams & G. M. Maggiora, *J. Mol. Struct., THEOCHEM*, 1982, **89**, 365
11. W. P. Jencks, *Acc. Chem. Res.*, 1976, **9**, 425; *ibid.*, 1980, **13**, 161
12. W. E. Doering & H. H. Zeiss, *J. Am. Chem. Soc.*, 1953, **75**, 4733
13. J. P. Richard, T. L. Amyes & T. Vontor, *J. Am. Chem. Soc.*, 1991, **113**, 5871
14. T. L. Amyes & J. P. Richard, *J. Am. Chem. Soc.*, 1990, **112**, 9507
15. J. P. Richard & W. P. Jencks, *J. Am. Chem. Soc.*, 1984, **106**, 1383
16. S. Wolfe & C.-K Kim, *J. Am. Chem. Soc.*, 1991, **113**, 8056

INTERLUDE.
THEORETICAL INVESTIGATION OF THE ORIGIN OF SECONDARY α-DEUTERIUM KINETIC ISOTOPE EFFECTS

"Isotopes...whatever they may be."
- Richard Howard S. Crossman

Introduction

The following piece of work is reproduced from the Journal of the Chemical Society, Chemical Communications. It stands by itself as a contribution to the debate about the origins of secondary α-deuterium kinetic isotope effects. The trends shown in this work have recently been independently confirmed[a] using the *ab initio* HF/6-31G level of theory.

Just a brief explanation of the calculation of the specific contribution of a group of vibrational modes to the zero point energy (ZPE). As explained in Chapter 2, the kinetic isotope effect is largely determined by the ZPE factor. This factor reflects differences in energy between the ground and transition structures involved, comparing the isotopically substituted structures with the "normal" structures. The ZPE is the sum of vibrational contributions to the energy of the molecule. For a given molecule and its isotopically substituted counterpart, their ZPE difference is just the difference in the energies of those vibrational modes that are affected by the isotopic substitution. By extension, any subset of vibrations that can be isolated in a system can be compared to their equivalent set in a second system. Therefore, the difference in energy between a subset of vibrations in two systems is related to their contribution to the overall ZPE. This situation is shown schematically in Scheme 1.

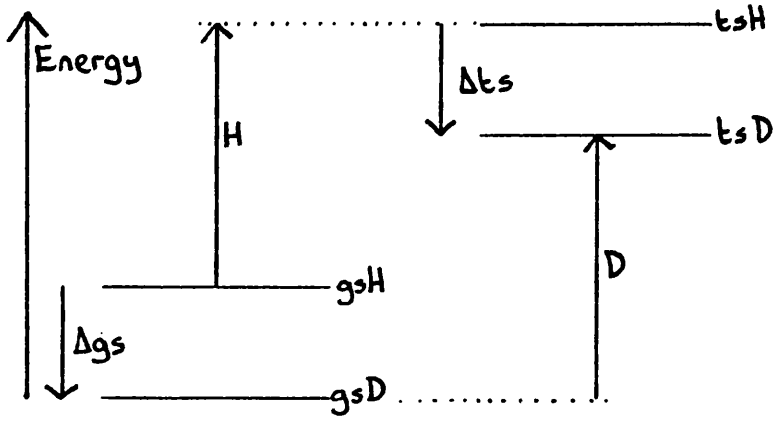
By calculating the vibrational frequencies for the ground and transition states of a system and its isotopically substituted counterpart, and identifying those vibrations which contain a contribution from the desired vibrational modes - C-H(D) stretches for example - it is possible to apply the formula in Scheme 1 to just those selected frequencies. This can then be used to show the

---

[a] R. A. Poirier, Y. Wang & K. C. Westaway, *J. Am. Chem. Soc.*, 1994, **116**, 2526

contribution that a particular set of vibrational frequencies makes to the overall ZPE factor, and hence to the isotope effect.

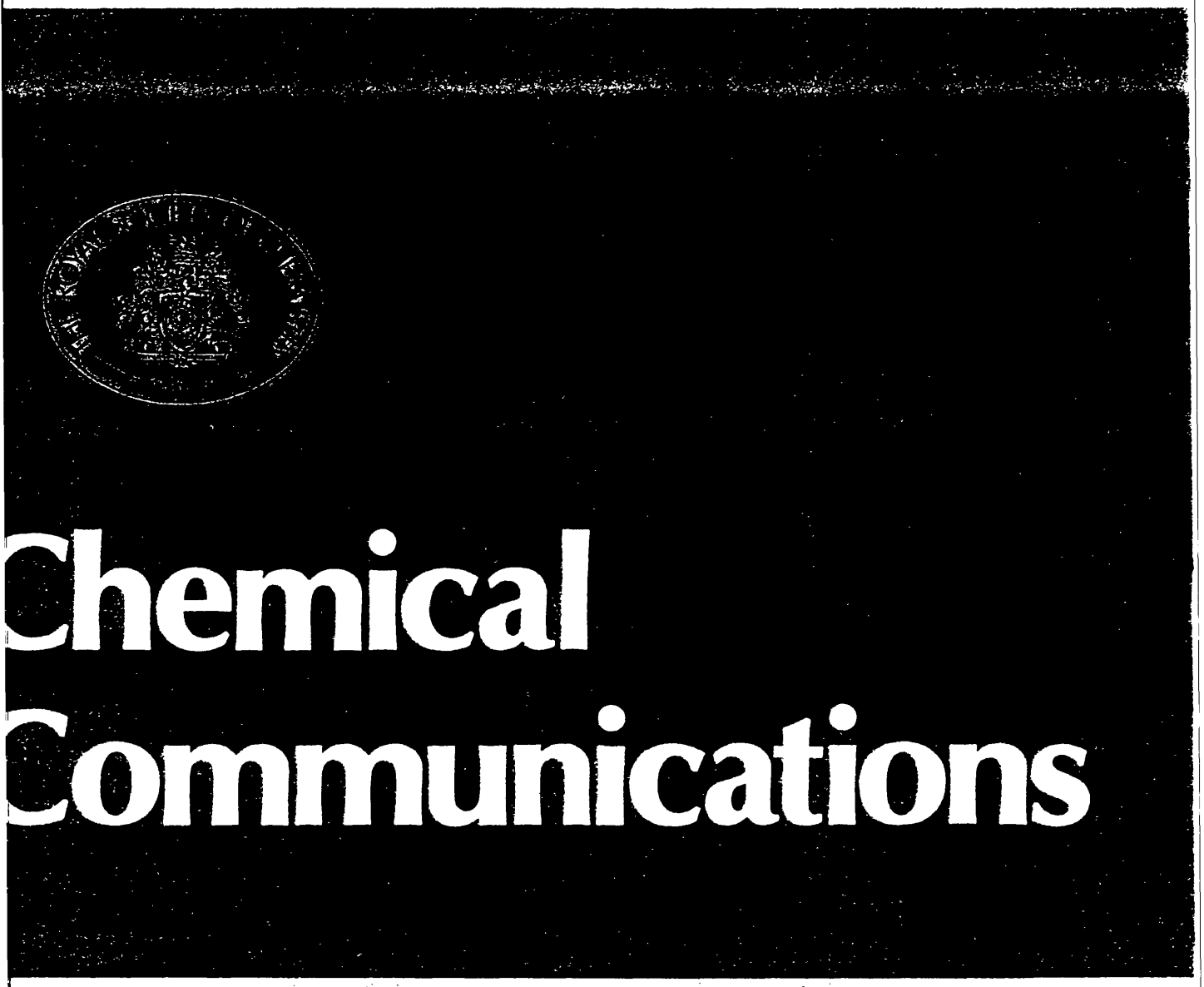
The work in this paper was done using the AM1 Hamiltonian and used CAMVIB and CAMISO for the calculation of vibrational frequencies from the initial force constants obtained from the AM1 method. Calculations were carried out as described in Chapter 4.



$$\text{ZPE contribution} = \exp\left(\frac{hc\Delta}{2kT}\right), \text{ where } \Delta = H - D = \Delta gs - \Delta ts, \text{ cm}^{-1} \text{ units}$$

$$\Delta = -\{\sum \text{modes}(gsH + tsD) - \sum \text{modes}(tsH + gsD)\}$$

Scheme 1. A representation of the contribution of vibrational energy levels to the zero point energy of a molecule, and how the difference between isotopically substituted species is calculated. The energy levels shown can represent some or all of the vibrational levels of the system.

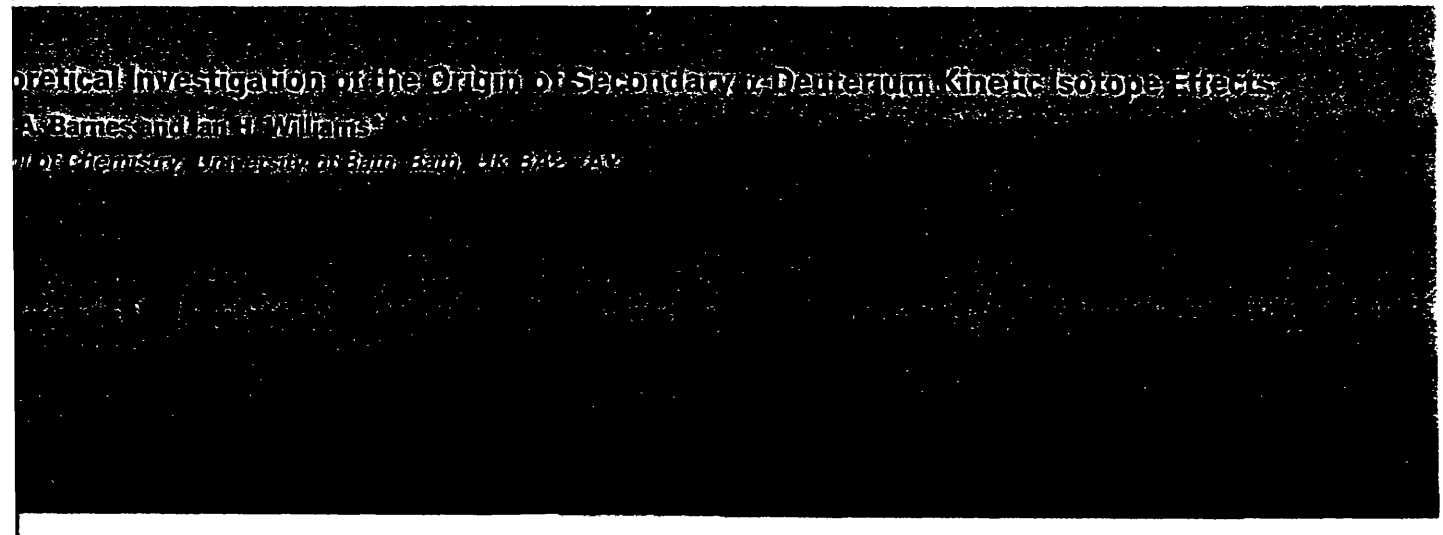


Chemical Communications

Reprinted from

J. Chem. Soc., Chemical Communications

Issue 16 1993



Theoretical Investigation of the Origin of Secondary α-Deuterium Kinetic Isotope Effects

John A. Barnes and Ian H. Williams*

School of Chemistry, University of Bath, Bath, UK BA2 7AY

Although the secondary α-deuterium kinetic isotope effect upon the degenerate S_N2 reaction of chloromethane with chloride anion arises from changes in α-CH stretching force constants, α-substitution by methyl and/or methoxy groups shows this inverse inductive contribution to be approximately constant and reveals that the trend in the isotope effect (calculated by semiempirical molecular orbital theory) is determined by changes in bending force constants, in accord with the conventional view.

Recent theoretical studies[1,2] of S_N2 methyl transfer reactions have questioned the conventional view[3] of the origin of secondary α-deuterium kinetic isotope effects (2° α-D KIEs) as arising from changes in bending-force constants HCX {(where X is the nucleofuge and/or nucleophile) in a degenerate process [eqn. (1)]} between reactant and transition state.

$$X^- + MeX \rightarrow XMe + X^- \quad (1)$$

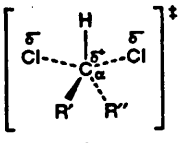
Instead, it has been pointed out that the inverse α-D KIEs for eqn. (1) (X = Cl, F, OF, NC, OMe, CN and CCH) calculated by means of *ab initio* molecular orbital theory arise from an increase in the α-CH bond stretching force constant accompanying the change from tetrahedral to trigonal geometry about C_α along the reaction coordinate. We now report preliminary results of semiempirical molecular orbital theory calculations of 2° α-D KIEs for degenerate S_N2 reactions involving α-substitution by methyl and/or methoxy groups [eqn. (2), R', R" = H, Me, OMe]: these confirm the recent

$$Cl^- + CHR'R"Cl \rightarrow ClCHR'R" + Cl^- \quad (2)$$

findings for methyl transfer (R' = R" = H) but suggest that the overall trend in the α-D KIEs may be understood within the framework of the conventional view.

The trend in the AM1[4] semiclassical[5] 2° α-D KIEs calculated at 298 K (Table 1; results are for replacement of a single H by D) for the range of degenerate S_N2 reactions [eqn. (2)] is clearly dominated by the zero-point energy factor (ZPE), the magnitude of which is offset by the product (MMI × EXC) of the mass/moment-of-inertia and excitational factors; there is an excellent linear correlation between $(k_H/k_D)_s$ and ZPE. Further factorization of the overall ZPE term reveals an almost constant inverse contribution from the CH stretching modes but a contribution from all the other modes ('the rest'), which varies proportionally with $(k_H/k_D)_s$. The inverse CH-stretching contribution agrees with the findings of both Wolfe[1] and Truhlar[2] and indeed with our own previously reported *ab initio* calculated α-D KIEs for methyl transfer.[6] The range of values from 0.952 to 0.960 are consistent with this contribution being regarded as inductive: a typical value[5,7] for an inductive β-D KIE is 0.985 from which, if an attenuation factor of about three per C–C bond is applied,[5,8] a value of *ca.* 0.955 may be estimated for an inductive α-D KIE.

The contribution to the α-D KIE from all the rest of the vibrational modes, but especially including α-CH bending modes, tends to be normal. Increasing the number of electron-donating methyl or methoxy substituents attached to C_α causes more carbocation character to develop in the transition state **1** and the $C_\alpha\cdots Cl$ partial bonds to increase in length (Table 1). This opening up of the transition-state structure does indeed involve a trend towards loosening of the α-CH bending motions, in turn giving rise to a normal ZPE contribution to the α-D KIE. The value of this factor ('the rest') is just slightly greater than unity with a single α-methyl substituent (R' = H, R" = Me), but is just slightly less than unity for the unsubstituted system (R' = R" = H); in these cases the product of the ZPE factors appears to be dominated by the α-CH stretching contribution. Analysis of the molecular vibrational force fields is complicated by the presence of redundancies involving the bending modes about C_α in both the tetrahedral reactant structures and the trigonal bipyramidal transition-state structures, with the consequence that a valence force constant for HCCl angle bending is not



1

Table 1 AM1 calculated semiclassical 2° α-D KIEs at 298 K for degenerate S_N2 reaction of Cl^- with CLR'R"Cl (L = H, D), together with zero-point energy factors and transition-state partial-bond lengths

			ZPE factors			
R'	R"	$(k_H/k_D)_s$	ZPE	CH stretches	The rest	C···Cl/Å
H	H	0.982	0.935	0.959	0.975	2.154
H	Me	1.015	0.977	0.954	1.024	2.202
H	MeO	1.048	1.027	0.954	1.077	2.252
Me	Me	1.090	1.114	0.952	1.170	2.270
Me	MeO	1.156	1.196	0.960	1.246	2.335

Table 2 AM1 calculated force constants [mdyn Å rad^{-2} (1 dyn = 10^{-5} N)] for α-CH bending in reactants and transition states

		Method A			Method B		
		F_{HCCl}		ΔF_{HCCl}	F_{umbr}	F_{wag}	$\Delta F_{wag\text{-}umbr}$
R'	R"	Reactant	Transition state		Reactant	Transition state	
H	H	0.551	0.375	−0.176	0.632	0.769	0.137
H	Me	0.583	0.402	−0.181	0.780	0.627	−0.153
H	MeO	0.589	0.424	−0.165	0.860	0.606	−0.254
Me	Me	0.621	0.410	−0.211	0.996	0.626	−0.370
Me	MeO	0.648	0.415	−0.233	1.061	0.493	−0.568

unambiguously defined. However, Table 2 contains two different representations of the AM1 calculated force constants, each of which suggests a trend towards looser α-CH bending with increasing electron-donating α-substitution. In method A the redundancies are eliminated as previously reported for distorted tetrahedra[9] and trigonal bipyramids[6,10] and HCCl force constants are obtained from sets of six or nine redundant valence coordinates for bending about C_α in reactants and transition states, respectively. In method B the force constant for the 'umbrella' symmetry coordinate for bending about C_α in the reactants is compared with the force constant for wagging of the α-CH bond out of the plane of the trigonal C_α moiety of the transition states. Linear regression of the changes ΔF in these force constants from reactant to transition state against the ZPE factor (Table 1, 'the rest') gives correlation coefficients of 0.952 and 0.961 for methods A and B, respectively.

In a recent experimental study[11] of deuterium isotope effects in gas-phase reactions of alkyl halides, inverse values of k_H/k_D were interpreted as characterizing the S_N2 mechanism and all significantly normal values were taken to indicate reactions proceeding mainly by the E2 mechanism. Our present results suggest, however, that it is quite reasonable for S_N2 reactions of α-substituted alkyl chlorides to manifest significantly normal α-D KIEs, in accord with conventional views. Reactions of *N*-(methoxymethyl)-*N*,*N*-dimethylanilinium[12] or 1-(methoxymethyl)-2,4-dinitrobenzene[13] in solution show 2° α-D KIEs (per D) in the ranges 0.99–1.18 and 1.05–1.16, respectively, depending upon the choice of nucleophile, for bimolecular substitutions proceeding by means of very loose transition states.

We thank the SERC for support.

Received, 28th April 1993; Com. 3/02449K

References

1 S. Wolfe and C.-K. Kim, *J. Am. Chem. Soc.*, 1991, **113**, 8056.
2 X. G. Zhao, S. C. Tucker and D. G. Truhlar, *J. Am. Chem. Soc.*, 1991, **113**, 826.
3 A. Streitwieser, R. H. Jagow, R. C. Fahey and S. Suzuki, *J. Am. Chem. Soc.*, 1958, **80**, 2326; K. T. Leffek, J. A. Llewellyn and R. E. Robertson, *Can. J. Chem.*, 1960, **38**, 1505; M. Wolfsberg and M. J. Stern, *Pure Appl. Chem.*, 1964, **8**, 235.
4 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902.
5 L. Melander and W. H. Saunders, *Reaction Rates of Isotopic Molecules*, Wiley Interscience, New York, 1980.
6 I. H. Williams, *J. Am. Chem. Soc.*, 1984, **106**, 7206.
7 V. J. Shiner, in *Isotope Effects in Chemical Reactions*, ed. C. J. Collins and N. S. Bowman, Van Nostrand Reinhold, New York, 1970.
8 J. E. Leffler and E. Grunwald, *Rates and Equilibria of Organic Reactions*, Wiley, New York, 1963.
9 I. H. Williams, *Chem. Phys. Lett.*, 1982, **88**, 462.
10 I. H. Williams, *Bull. Soc. Chim. Belg.*, 1982, **91**, 356.
11 S. Gronert, C. H. DePuy and V. M. Bierbaum, *J. Am. Chem. Soc.*, 1991, **113**, 4009.
12 B. L. Knier and W. P. Jencks, *J. Am. Chem. Soc.*, 1980, **102**, 6789.
13 G. A. Craze, A. J. Kirby and R. J. Osborne, *J. Chem. Soc., Perkin Trans. 2*, 1978, 357.

CHAPTER 8.
SIALIC ACIDS AND SIALIDASES

"...all the work of crystallographers serves only to demonstrate that there is only variety everywhere where they suppose uniformity... that in nature there is nothing absolute, nothing perfectly regular."

- Georges Leclerc [Comte de] Buffon

8.1 Sialic Acid

The term sialic acid covers a series of compounds containing the natural derivatives of neuraminic acid (NMA), an acid amino sugar. Neuraminic acid exists in pyranose form and contains nine carbon atoms (Figure 8.1). Unsubstituted NMA does not occur naturally; the amino group is substituted by an acetyl or glycolyl residue and the hydroxyl groups can be methylated or esterified. Sialic acids are the most varied of the natural sugars.

Sialic acids have been found in most higher animals and in some viruses and bacteria. The most usual sialic acid to be found is N-acetylneuraminic acid (NANA, Figure 8.1). Sialic acids are most often found in the outer cell membrane where they are components of glycoproteins, gangliosides or polysaccharides. The electronegative charge of a human red blood cell is mainly due to a dense coat of about twenty million sialic acid molecules.[1] The acid sugars are also secreted bound to glycoproteins in serum and mucus, and have been found in larger quantities in some patients[1] compared to healthy people.

When bound to macromolecules, sialic acids are α-glycosidically linked to other sugars, most frequently galactose or N-acetylgalactosamine.[2] They are most often the terminating sugar of a chain, and can be cleaved from the chain by a

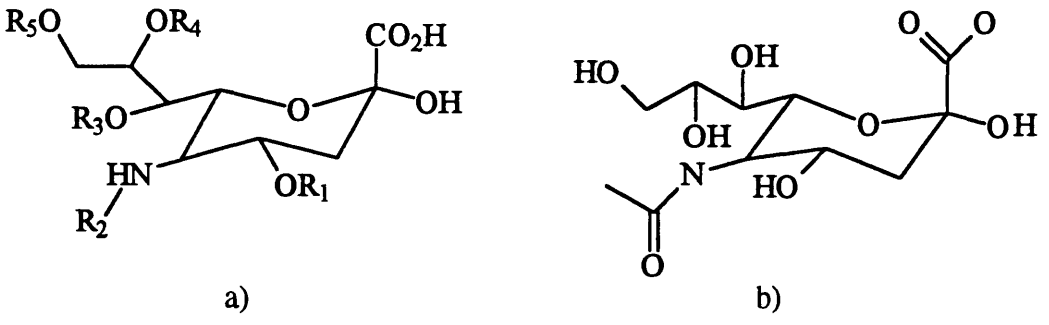


Figure 8.1. The possibilities for natural variation in sialic acid (a) and the most abundant natural variant, N-acetylneuraminic acid (b).

family of enzymes called sialidases. The sialic acid substituent and its type of linkage affects the rate of enzyme reaction. β-anomers of sialic acids are not substrates of sialidases[3] and with the exception of sialidase from *Arthrobacter ureafaciens*, α(2,3)-glycosidic linkages are hydrolyzed much faster than α(2,6) ones. The recent proposal that (2,3) sialosides are more rigid than (2.6) ones[4] may be one factor that explains this observation.

The wide occurrence of sialic acids in cell surfaces implies that they have a significant role in biological functions and it is true that enzymic removal of them leads to changes in the biological behaviour of cells and molecules.[5] Five main functions have been established for sialic acids:[6]

1. Because of their negative charge they are involved in the binding and transport of positively charged compounds.
2. They influence the conformation of glycoproteins which is important for the correct arrangement of glycoproteins in cell membranes and for their resistance to proteases. This resistance is thought to be important for preventing the spread of bacteria in infected tissue.
3. Sialic acids contribute to the specificity of blood group substances and have been found to be antigenic determinants in some mammalian systems.
4. Sialic acids are essential components of receptors for peptide hormones, toxins and viruses. Infection of cells by some microorganisms is exclusively dependent on the presence of sialic acids on cell membranes.
5. Sialic acids have been found to prevent recognition of antigenic sites by the immune defence system.

It is clear that sialic acids are important compounds in biochemical processes. In so far as this thesis is concerned, knowledge of the basic chemistry of glycosides of NANA is desirable because these have a glycosidic structure and are substrates which are hydrolysed by the sialidase enzymes.

8.1.1 Aqueous Hydrolysis of N-Acetylneuraminic Acid Glycosides

The hydrolysis of glycosides of NANA might be expected to be similar to that of glucopyranosides and related structures. However, there are complications. Compared to sugar-like rings, the carbon adjacent to the anomeric centre does not have a hydroxyl group, and the anomeric carbon itself has a carboxylate group rather than a hydrogen. In simple pyranosides, removal of the 2-OH group stabilizes the glycosyl cation that would form in a hydrolytic process, so the acid-catalyzed hydrolyses of these pyranosides are a thousand-fold faster[7] than those of glucopyranosides. It is also debatable whether or not the cation of NANA would form, since the glucopyranosyl cation is too unstable to exist in the presence of an anionic nucleophile[8] and of course NANA has its own in-built anionic group. It follows that knowledge of glycoside hydrolysis, while useful, cannot be used confidently when applied to sialic acid hydrolysis.

Ashwell *et al* have identified three main processes[9] in the hydrolysis of aryl α-glycosides of NANA:

1. The acid catalyzed reaction of the neutral molecule.
2. The spontaneous reaction of the neutral molecule or its kinetic equivalent, the proton-catalyzed reaction of the anion derived by ionization of the carboxylic acid group.
3. The spontaneous reaction of the anion.

There is also a fourth process, which is base catalyzed but insignificant compared to the first three. It is processes 2 and 3 which are of most interest since they are likely to occur at physiological pH and involve the ionised carboxylate group. These conditions are also met by the enzymic process catalyzed by sialidases. The work mentioned above used secondary deuterium KIEs to probe TS structure and concluded that some nucleophilic assistance by the carboxylate group was required.[9] This may involve a full α-lactone ring structure. Clearly this has implications for sialidase catalysis, and indeed for many glycosidases, and merits investigation as part of the current thesis.

Before describing the work carried out on sialic acid derivatives, a brief outline of the sialidase enzymes and their postulated mechanisms of hydrolysis of sialic acid derivatives is appropriate to clarify the reasons for the types of investigation undertaken.

8.2 The Sialidase Enzymes

Early in the 1940s, Hirst[10] noticed that liquid from chicken eggs infected by influenza virus caused red blood cells to agglutinate. When the red cells were then warmed to 37°C they dispersed as the virus eluted and were not subsequently affected by fresh infection. The virus which eluted from the red cells could agglutinate fresh red blood cells. Hirst interpreted this as the result of some enzymic action of the virus where the substrate was a part of the red cell surface. In the next decade, Gottschalk[11] discovered a product that the virus removed from the cell surface. The product was NANA and the enzyme was called sialidase or neuraminidase (NA). Its formal name is acylneuraminyl hydrolase (EC 3.2.1.18).

Sialidases are glycohydrolases which catalyze the cleavage of terminal sialic acids which are linked α-ketosidically to glycoproteins, glycolipids and polysaccharides. Figure 8.2 shows the general reaction. As mentioned above, only α-anomers of sialic acid derivatives are substrates for the enzymes. All known NAs act with retention of configuration of the substrate with the exception of the *Salmonella typhimurium* NA, which yields the β-anomer as the initial product. Many biological roles have been attributed to sialidases, including cell-cell recognition processes and the pathogenicity of some infections by sialidase-bearing microorganisms.[5] In influenza virus, which is the most widely studied sialidase, the enzyme is considered important for transporting the virus through mucin[12] and for the elution of new virus from infected cells.[13] NA is also produced by bacteria in the presence of host sialic

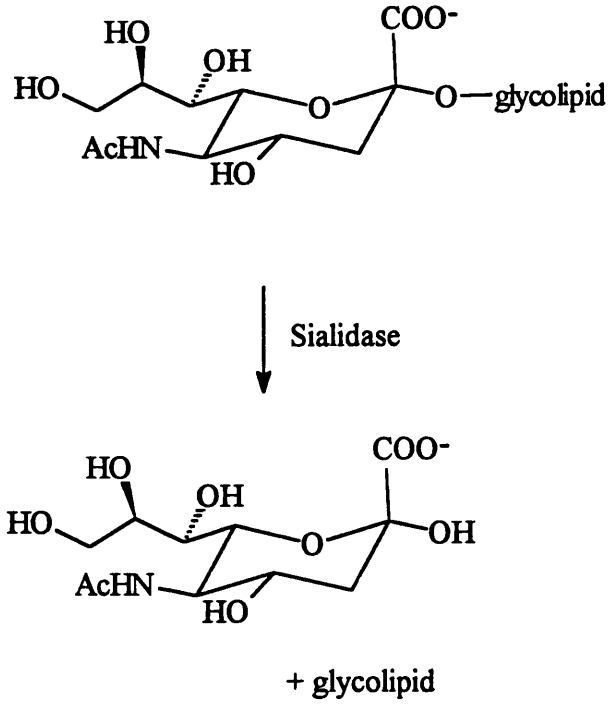


Figure 8.2. The general scheme for enzymic hydrolysis of sialic acid derivatives by sialidases.

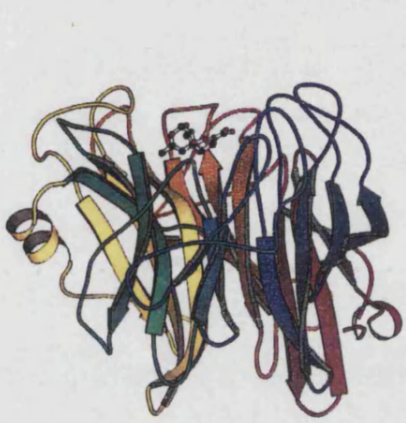
acid where it may aid infection or be part of the nutrition cycle, since some bacteria utilise sialic acid.

8.2.1 Structure of Sialidases

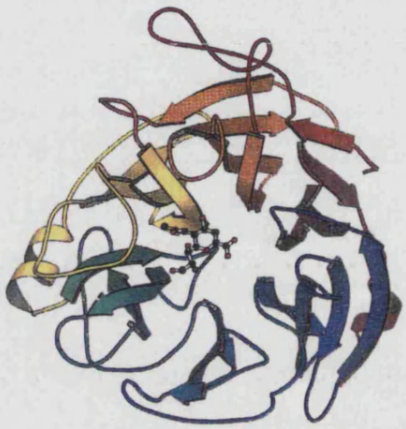
There is a growing body of information about the structure of NA enzymes from both sequence and crystallographic studies. Most work to date has been on the NA from influenza virus. There are nine identified subtypes of influenza A NA, called N1 to N9: Only subtypes N1 and N2 are on viruses infecting humans. The structures of N2,[14] N9[15] and a human influenza B[16] NA are known. On the flu virus surface, NA appears as a mushroom-shaped spike.[17] Each of the characterized NAs shares a similar topology: a tetramer with a box-shaped head, 100 x 100 x 60Å, made of four coplanar spherical subunits and a central stalk by which it is embedded in the viral membrane. The monomers are arranged around a four-fold axis, with each monomer containing six 4-stranded B-sheets arranged like the blades of a propeller[14] (Figure 8.3). Each monomer also contains a calcium ion which appears to have a structural role, and although the calcium is not essential for catalysis, when present the calcium does enhance activity. Residues in the active site are highly conserved across subtypes even though sequence homology can be as low as 25%.

The three-dimensional structure of the bacterial NA from *S. typhimurium* has recently been obtained.[18] The enzyme is a monomer which is similar in size to the flu NA monomer (~380 residues), but the sequence homology is only 15%. Despite this, the two enzymes show a good similarity in the general folding and spatial arrangement of catalytic residues (Figure 8.3). Across the bacterial NAs there is evidence for two families dependent on the presence of a divalent ion for optimal activity. Those not requiring a cation are in the "small" family with

around 400 residues and have some sequence homology. Those that do require a cation are in the "large" family and show no significant homology either to the "small" family or to the flu NA. *S. typhimurium* NA belongs to the "small" family.



S. typhimurium sialidase with DANA



S. typhimurium sialidase with DANA

Influenza virus sialidase Subtype N9

Influenza virus sialidase Subtype N9

Figure 8.3. Orthogonal views of the sialidases from *S. typhimurium* and influenza virus subtype N9. Left hand views are from the side and right hand views from above the active site. The main chain is coloured from red at the N terminus to violet at the C terminus. The influenza structure shows just one of the four monomers that make the whole enzyme. The pictures were produced using the MOLSCRIPT program.[29]

8.2.2 Active Site Comparison of Influenza and *S. Typhimurium* NA

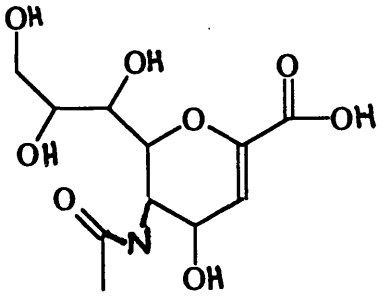


Figure 8.4. 2-deoxy-2,3-dehydro-N-acetylneuraminic acid, DANA

The catalytic site of the influenza A and B NA has been characterized by binding of the inhibitor 2-deoxy-2,3-dehydro-N-acetylneuraminic acid (DANA, Figure 8.4) or of the product NANA. Kinetic studies on *S. typhimurium* enzyme have shown inhibition by DANA similar to the viral enzyme but no inhibition by NANA. Influenza NA is weakly inhibited by NANA. The catalytic sites share several features:

1. Three Arg residues stabilize the carboxylic acid group common to all natural sialic acid derivatives.
2. A Glu stabilizes the position of one of the Arg residues.
3. A Tyr lies close to the sugar ring of the sialic acid, on the opposite side to the aglycone.
4. A hydrophobic pocket accommodates the N-acetyl group, although the pocket is formed by a Met, Trp, Trp, Leu in the *S. typhimurium* and by a Trp, Ile in the influenza enzyme.

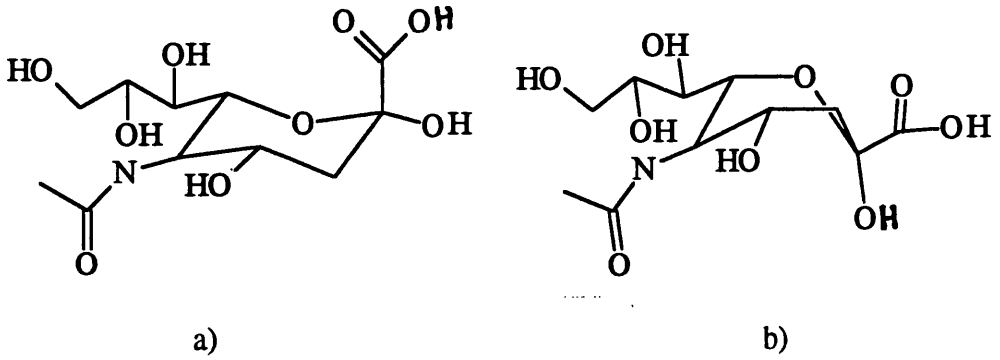
The main differences between the two sites are the interactions with the glycerol side chain and the sialic acid ring hydroxyl oxygen. In the viral enzyme a Glu forms two strong H-bonds with two of the three oxygens of the glycerol chain; in *S. typhimurium* there is one weak H-bond made by a Trp residue. The large difference in turnover number between the two enzymes (*S. typhimurium*

2700 s^{-1}, influenza 10 s^{-1}) could be explained by this interaction. The viral enzyme has a weak H-bond from an Asp to the ring hydroxyl oxygen; *S. typhimurium* has two strong H-bonds from Asp and an Arg. This difference has been exploited recently to produce the most effective viral NA inhibitors yet made.[19]

8.2.3 Catalytic Mechanism

There is not a great deal of mechanistic information about NA enzymes. Isotope effects have been used to show that some glycosidases have mechanisms with a glycosyl-like TS.[20-22] Lentz *et al*[23] used site-directed mutagenesis studies and proposed that flu NA has a lysozyme-like mechanism, with a histidine donating a proton to a glutamic acid which then attacks the glycosidic linkage. However, this is not consistent with the observed mode of binding of sialic acid in the active site of NA as reported by Burmeister *et al.*[16] A study by von Itzstein and co-workers[24] applied KIEs, NMR and molecular dynamics to the influenza A NA. This confirmed that the initial product was the α-anomer of NMA, so the flu NA retains the configuration of the substrate. They postulated a mechanism with a sialosyl cation TS. The sequence of events proposed was:

1. The binding of substrate such that the sialoside distorted from a $^2C_5$ chair to a $B_{2,5}$ boat conformation (Figure 8.5).
2. Generation of a sialosyl cation via general acid catalysis by Asp 151 through a water molecule.
3. General base catalysis by Asp 151 of the attack by water on the cation.



a) b)

Figure 8.5. The $^{2}C_{5}$ chair (a) and $B_{2,5}$ boat (b) conformations of sialic acid postulated for enzymic hydrolysis by sialidases.

This mechanism was subsequently refined to involve Arg 371 for step three when modelling studies showed that a water molecule could fit between the Asp and Arg.[25] A KIE study of flu A by Guo and co-workers[26] concluded that for the flu NA the conformational change occurred after initial binding and that there is significant proton donation to the leaving group. This is broadly in agreement with their recent findings about non-enzymic hydrolyses of glycosides of NMA.[9] More interestingly, the same workers have shown[27] that *S. typhimurium* acts with inversion of configuration, unlike any other of the known NAs. They interpret their KIE studies of this enzyme[26] by postulating that the mechanism is a single displacement reaction with a TS derived from the $^{2}C_{5}$ chair conformation with little proton donation to the leaving group and substantial charge development at the anomeric carbon of the substrate.

The oxocarbenium ion formed by the hydrolysis of N-acetylneuraminyllactose (NANL) by influenza B NA has apparently been observed in an X-ray structure.[28] This observation has been used to explain the low turnover rate of the enzyme, since stable binding of a positively charged intermediate in a planar conformation is expected to be very inefficient

because very strong electrostatic interactions would be required. The proposed mechanism in this case has no nucleophile or proton donor, and reaction is driven purely by the induction and stabilization of the oxocarbenium ion.

The finding that NA enzymes have different mechanisms despite similar secondary structure makes the basis for an interesting modelling study using computational techniques. Small changes in the positions of amino acid residues and charges may be enough to alter the enzyme reaction from one pathway to another. One reason for this may be the instability of the TS or intermediate involved in the reaction. Whether it is a sialosyl cation, an α-lactone or some other species, it is on the borderline of a real existence.

In particular, the parallels in the proposed reaction mechanisms between the *S. typhimurium* and flu B NAs and the two pathways in the AMPH study are fascinating and demand investigation. To recap, KIE studies of *S. typhimurium* NA suggest a highly charged TS, and the enzyme acts with a single displacement and inverts the substrate configuration. The flu B NA has a two-step process with an oxocarbenium ion intermediate and retains the substrate configuration. There is no obvious structural difference that can adequately explain the contrast in these mechanisms,[26] but perhaps the use of computer simulation techniques can provide some suggestions. As shown in Chapter 6, the PES for the AMP hydrolysis contained two finely balanced pathways which were distinguished by the amount of pre-association of a water molecule. That such a fine distinction could be made demonstrates the power of the CHARMM QM/MM method in investigating these types of mechanism. It is the comparison of these two enzyme mechanisms that forms the final study in this thesis.

8.3 References

1. R. Schauer, *Adv. Carbohyd. Chem. Biochem.*, 1982, **40**, 131
2. A. P. Corfield & R. Schauer in Sialic Acids: Chemistry, metabolism & Function, Ed. R. Schauer, pub. Springer 1982
3. J. N. Varghese, J. L. McKimm-Breschkin, J. B. Caldwell, A. Kortt & P. M. Colman, *Proteins*, 1992, **14**, 327
4. S. Sabesan, K. Bock & J. C. Paulson, *Carbohyd. Res.*, 1991, **218**, 27
5. R. Schauer, *Trends Biochem. Sci.*, 1985, **10**, 357
6. W. Reutter, E. Kottgen, C. Bauer & W. Gerok in Sialic Acids: Chemistry, metabolism & Function, Ed. R. Schauer, pub. Springer 1982
7. a) E. H. Cordes & H. G. Bull, *Chem. Rev.*,1974, **74**, 581; b) T. H. Fife, *Acc. Chem. Res.*, 1972, **5**, 264
8. N. S. Banait & W. P. Jencks, *J. Am. Chem. Soc.*,1991, **113**,7051
9. M. Ashwell, X. Guo & M. L. Sinnott, *J. Am. Chem. Soc.*, 1992, **114**, 10158
10. G. K. Hirst, *Science*, 1941, **94**, 22
11. A. Gottschalk, *Biochim. Biophys. Acta* ., 1957, **23**, 645
12. H. D. Klenk & R. Rott, *Adv. Virus Res.*, 1988, **34**, 247
13. P. Palese, K. Jobita, M. Ueda & R. W. Compans, *Virology*, 1974, **61**, 397
14. P. M. Colman, J. N. Varghese & W. G. Laver, *Nature*, 1983, **303**, 35
15. W. R. Tulip, J. N. Varghese, A. T. Baker, A. van Donkelaar, W. G. Laver, R. G. Webster & P. M. Colman, *J. Mol. Biol.*, 1991, **221**, 473
16. W. P. Burmeister, R. W. H. Ruigrok & S. Cusack, *EMBO J.*, 1992, **11**, 49
17. G. M. Air & W. G. Laver, *Proteins*, 1989, **6**, 341
18. S. J. Crennel, E. F. Garman, W. G. Laver, E. R. Vimr & G. L. Taylor, *Proc. Nat. Acad. Sci. USA*, 1993, **90**, 9852

19. M. L. von Itzstein, W. Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. V. Phan, M. L. Smythe, H. F. White, S. W. Oliver, P. M. Colman, J. N. Varghese, D. M. Ryan, J. M. Woods, R. C. Bethell, V. J. Hotham, J. M. Camerson & C. R. Penn, *Nature*, 1993, **363**, 418

20. F. W. Dahlquist, R. Rand-Mier & M. A. Raftery, *Biochemistry*, 1969, **8**, 4214

21. M. L. Sinnott, *FEBS Lett.*, 1978, **94**, 1

22. D. W. Parkin & V. L. Schramm, *Biochemistry*, 1987, **26**, 913

23. M. R. Lentz, R. G. Webster & G. M. Air, *Biochemistry*, 1987, **26**, 5351

24. A. K. J. Chong, M. S. Pegg, N. R. Taylor & M. von Itzstein, *Eur. J. Biochem*, 1992, **207**, 335

25. N. R. Taylor & M. von Itzstein, *J. Med. Chem.*, 1994, **37**, 616

26. X. Guo, W. G. Laver, E. Vimr & M. L. Sinnott, *J. Am. Chem. Soc.*, 1994, **116**, 5572

27. X. Guo & M. L. Sinnott, *Biochem. J.*, 1993, **296**, 291

28. M. N. Janakiraman, C. L. White, W. G. Laver, G. M. Air & M. Luo, *Biochemistry*, 1994, **33**, 8172

29. P. J. Kraulis, *J. Appl. Crystallog.*, 1991, **24**, 946

CHAPTER 9.
MODELLING STUDIES OF SIALIC ACID DERIVATIVES

"Never say I tried it once and it did not work."

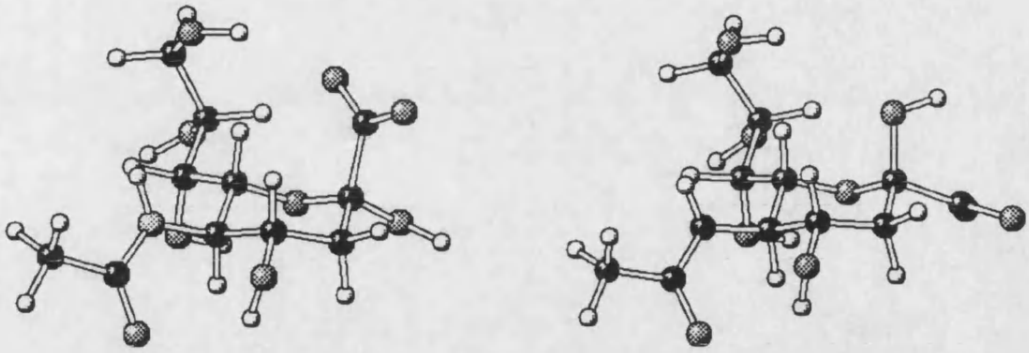
- Lord Ernest Rutherford

9.1 Introduction

There are several aspects of the chemistry of sialic acid derivatives which can be addressed by computer simulation and which it is appropriate to investigate given the background in the previous chapter and the remaining aims of this thesis. A conformational search will be used to provide a starting structure for QM studies. The QM studies will assess the stability of possible α-lactonic intermediates in the hydrolysis of NMA derivatives compared to the zwitterionic structure. In the light of the findings that the flu enzyme mechanism involves a conformational change, the path of this conversion will also be explored. There are experimental KIE data available for the hydrolysis of p-nitrophenyl N-acetylneuraminide (PNA) in water. This reaction is amenable to modelling using a similar methodology to the AMPH study in Chapter 6. The study of PNA hydrolysis will provide a useful basis from which to progress to simulation of the enzymic reaction and will provide a baseline from which to assess the novel components of the enzymic process.

9.2 Conformational Analysis of N-Acetyl Neuraminic Acid

A preliminary conformational analysis of N-acetyl neuraminic acid (NANA) was carried out using the MacroModel[1] molecular mechanics program. Both α and β anomeric forms were modelled. The ring system was treated by the default MacroModel settings which automatically reject any conformations which disrupt the ring system. The dihedral angles of all side chains were stepped through 360° in 15° increments. The favoured conformations without exception had the ring in the $^2C_5$ conformation as shown in Figure 9.1. The β-anomers



α-anomer, 11.00 kcal mol^{-1} β-anomer, 19.82 kcal mol^{-1}

Figure 9.1. Lowest energy MacroModel conformations found for NANA.

were less stable than the equivalent α-anomers, according to the calculated MacroModel energies shown in Figure 9.1. MacroModel appears to predict the conformation successfully, but not the relative energy of the α and β forms. As the experimental KIE data which is going to be used to assess the effectiveness of modelling studies is based on the hydrolysis of the α-anomeric form, the most stable α-anomeric conformation from this search was taken as the basis for subsequent QM calculations. This conformation is shown in Figure 9.1.

9.3 Preliminary Quantum Mechanical Modelling of NANA

Several issues were addressed by QM modelling of NANA and PNA: the relative stability of the α and β anomers of NANA; the barrier to interconversion of chair and boat forms of PNA; and the possibility of cyclic intermediates in hydrolysis. The reasons for these points have already been discussed. Except for the stability question, the studies have only been done on the α-anomer since it is this anomer for which all the experimental KIE data is available. Each of these investigations was done using the AM1 *in vacuo* Hamiltonian. The AMSOL program was also used to assess the potential intermediate structures.

The starting α and β anomers were the lowest energy conformations found in the MacroModel search described previously. These two structures were geometry optimized using the standard operating conditions (Chapter 4) appropriate to the AM1 and AMSOL methods.

Figure 9.2 shows the chair and boat forms of the α-anomer of PNA which are of interest in the sialidase hydrolysis. This molecule was modelled because it is the substrate which has been used in experimental studies of enzymic and non-enzymic hydrolysis of the neuraminic acids. The dihedral angle made by C5O6-C1C2 can be seen to be an approximate reaction coordinate for this conversion and has been used in previous studies of glycoside hydrolysis.[2] This dihedral was used to define a path calculation using the AM1 and AMSOL methods, with the dihedral angle varied from -60° to +60° in 10° steps. Local minimum energy structures along the path were geometry optimized to give the most stable conformations for this particular interconversion.

To investigate the possibility of α-lactone intermediates, the structures shown in Figure 9.3 were optimized using the AM1 and AMSOL models. The

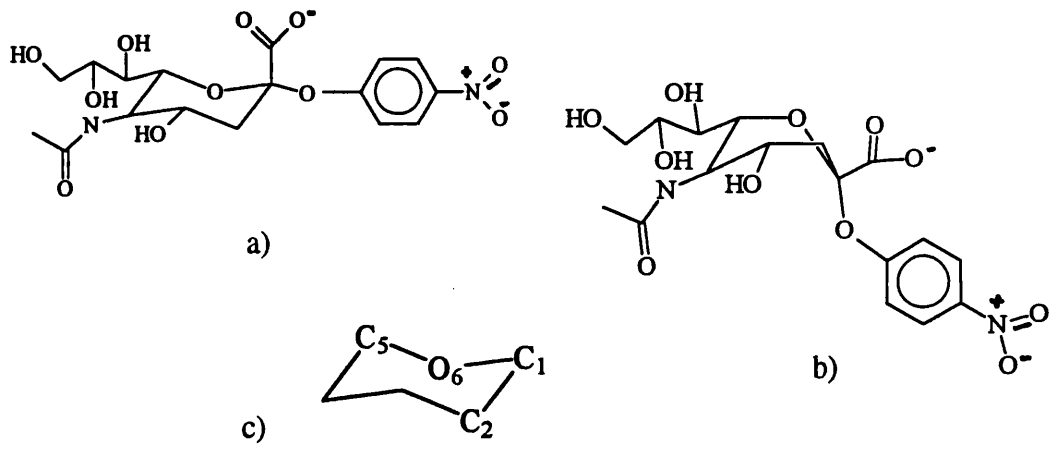


Figure 9.2. The chair (a) and boat (b) conformations of the α-anomer of PNA, and the dihedral angle (c) used to define a reaction coordinate for their interconversion.

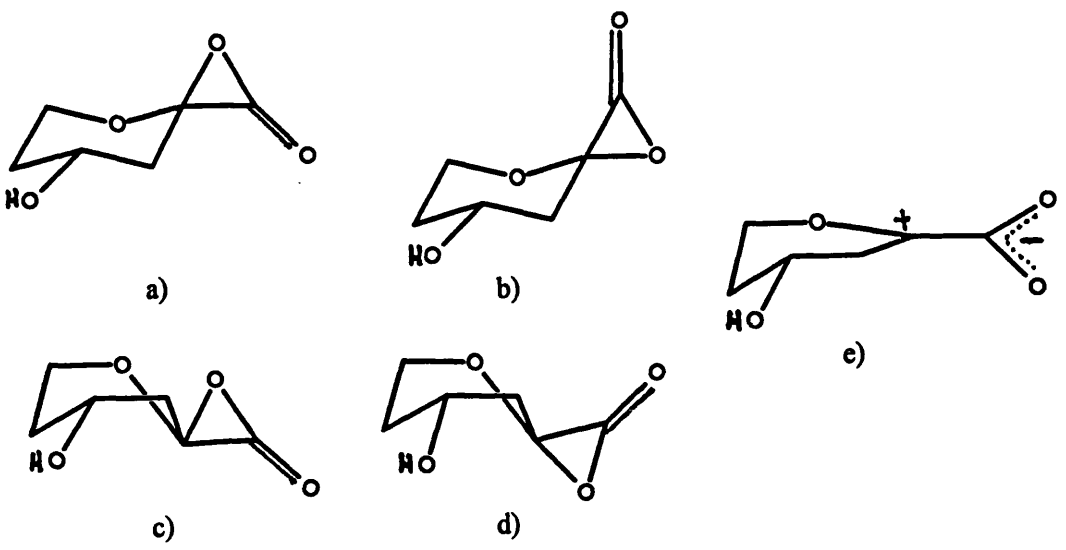
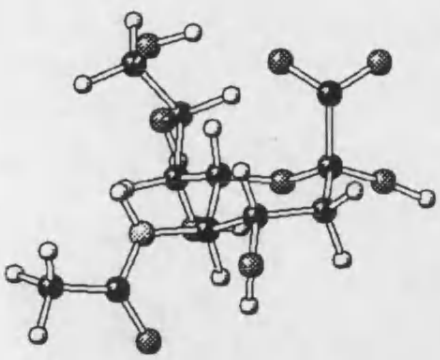


Figure 9.3. Possible α-lactonic (a-d) and zwitterionic (e) intermediate structures for PNA hydrolysis.

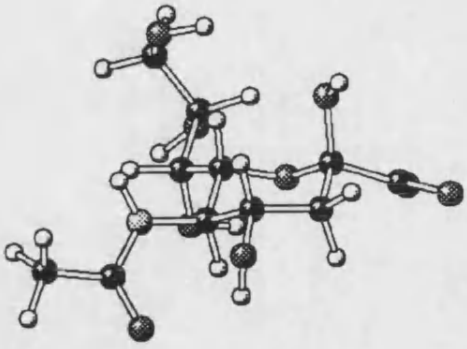
starting structures were derived from the most stable chair and boat conformers identified from the AM1 path calculation. As indicated, the N-acetyl and glyceryl sidechains were not included for these calculations. This is an acceptable compromise to save time because it is unlikely that either sidechain could be implicated in the formation or stabilization of the structures of interest. The α-lactone structures were compared with the alternative oxocarbenium ion intermediate, also shown in Figure 9.3.

9.3.1 Relative Stability of the α- and β-anomers of NANA

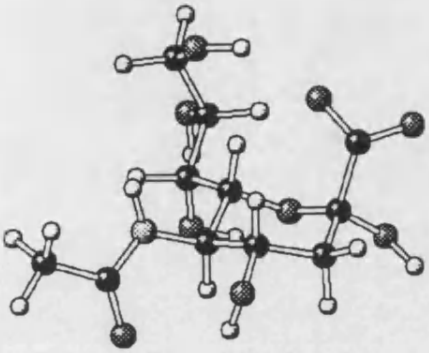
Figure 9.4 shows the structures of the α and β anomers of NANA calculated by the AM1 and AMSOL methods. As expected the β anomer is the more stable, with a heat of formation of -468.87 kcal mol^{-1} as compared to -467.95 kcal mol^{-1} for the α-anomer using AM1 and -533.94 kcal mol^{-1} as compared to -532.77 kcal mol^{-1} using AMSOL. The geometric differences between the AM1 and AMSOL structures are very small: in general the AMSOL solvation model does



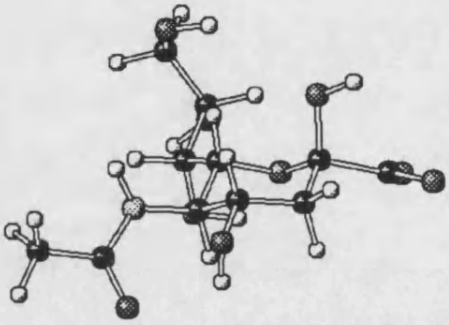
AM1, α-anomer



AM1, β-anomer



AMSOL, α-anomer



AMSOL, β-anomer

Figure 9.4. Geometry optimized structures of the α and β anomers of NANA calculated by AM1 and AMSOL.

not appear to induce large structural changes from the *in vacuo* geometry. This calculation was done to demonstrate that the semiempirical AM1 quantum-chemical calculations can successfully predict the relative stability of the α and β anomers: this may be important in later studies of the enzyme reaction where different types of sialidase produce either the α- or β-anomer of sialic acid as the initial product of hydrolysis.

9.3.2 The Chair to Boat Conformational Change of PNA

The energy profile of the AM1 *in vacuo* conformational change of the α-anomer of PNA from chair to boat is shown in Figure 9.5. Only one stable conformer can be identified along the path and this was fully optimized to give

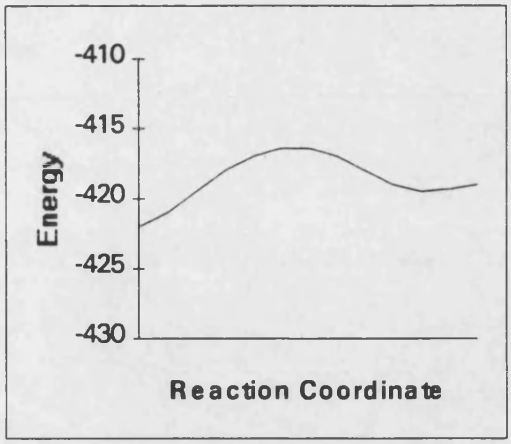


Figure 9.5. AM1 energy profile for the chair to boat conversion for the α-anomer of PNA.

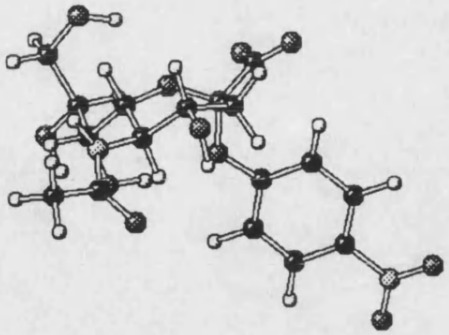


Figure 9.6. The boat conformation found for the α-anomer of PNA.

the boat-like structure in Figure 9.6. As expected, the chair conformation is more stable than the boat, at -422 kcal mol^{-1} compared to -419.5 kcal mol^{-1}. The attempt to generate a similar profile using AMSOL was unsuccessful. The first path found with AMSOL was one in which the sugar ring opened *via* cleavage of the ring C-O bond. When this bond was constrained at each point to equal the bond length found for the AM1 path, the energy profile became jagged, as shown in Figure 9.7. There were no obvious reasons for this beyond the applied constraint; however, low gradients could not be achieved for the majority of the points along the path in either of the attempts to generate the energy profile, and it was decided that these results would be unreliable.

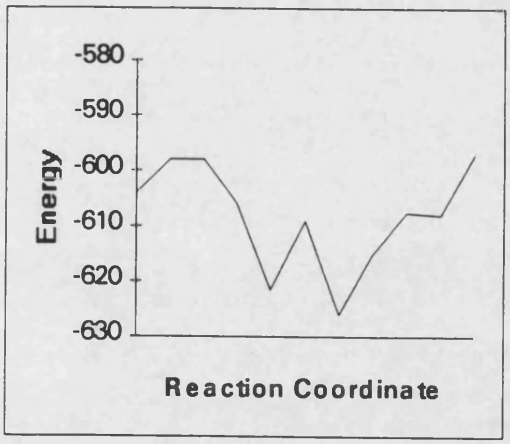
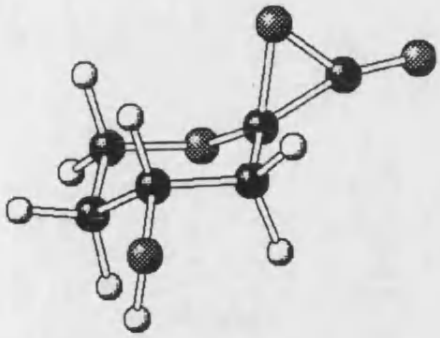


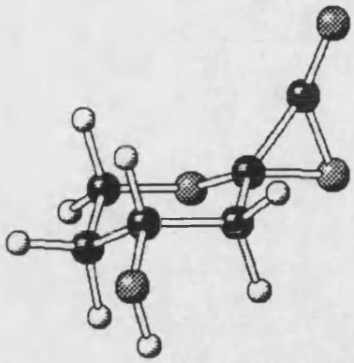
Figure 9.7. AMSOL energy profile for the chair to boat conversion for the α-anomer of PNA.

9.3.3 Relative Stability of the Possible Hydrolysis Intermediates

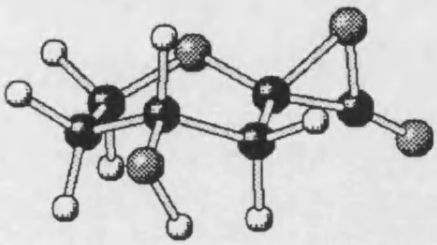
Figure 9.8 shows the optimized structures and energies found using the AM1 method for the possible intermediates depicted in Figure 9.3. The AMSOL structures are not shown, since once again the geometric differences between the two methods are very small. The relative energies of the different possibilities are the same within each method. However, for the *in vacuo* simulation there is only a 2.5 kcal mol^{-1} difference between the zwitterionic and covalent intermediates, compared to a difference of 11.4 kcal mol^{-1} for the *in aquo* simulation. This is not unexpected given the ability of water to stabilize charge-separated structures and the high strain that may be attributed to the alternative α-lactonic structures. The surprise is that the two structures are so similar in energy, given that the strain energy of oxirane, the three-membered heterocycle containing one oxygen atom, has been estimated[3] at 27.3 kcal mol^{-1} and it might be expected that the structures presented here would possess similar strain. Assuming that they do, it is by no means certain that any intermediate would be the zwitterionic species in an enzymic reaction, since the energy difference between the two types of structure is not an insurmountable barrier for an enzyme: it would depend on how the enzyme has evolved to catalyze the reaction. It is likely that the aqueous hydrolysis would proceed *via* the zwitterionic species in a two-step process, if such a process occurs, because the estimated 11.4 kcal mol^{-1} is a significant energy difference in solution.



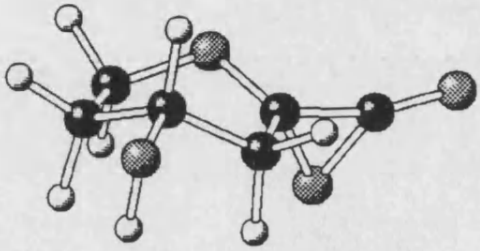
a) AM1 : -237.3
AMSOL : -255.7



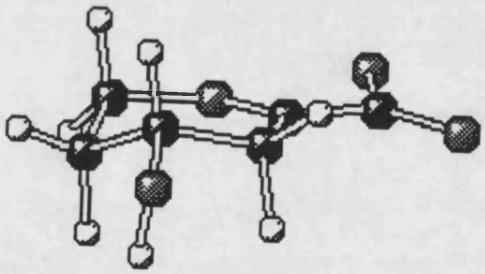
b) AM1 : -235.2
AMSOL : -252.6



c) AM1 : -235.9
AMSOL : -253.4



d) AM1 : -235.4
AMSOL : -252.9



e) AM1 : -239.8
AMSOL : -267.1

Figure 9.8. Optimized structures of the possible α-lactonic and zwitterionic intermediate structures for the hydrolysis of PNA. The numbers shown are the heat of formation for AM1 and the free energy of solvation for AMSOL, both in kcal mol^{-1}.

9.4 Hydrolysis of p-nitrophenyl N-acetylneuraminide

As stated in section 8.1.1, the hydrolysis of aryl a-glycosides of N-acetyl neuraminic acid occurs *via* the three main processes[4] below:

1. The acid catalyzed reaction of the neutral molecule.
2. The spontaneous reaction of the neutral molecule or its kinetic equivalent, the proton-catalyzed reaction of the anion derived by ionization of the carboxylic acid group.
3. The spontaneous reaction of the anion.

I have chosen to model the hydrolysis of PNA by process 3, using CHARMM in a similar manner to that used in Chapter 6 for the adenosine monophosphate model. Process 3 is the easiest to model and occurs at pH 6.79, close to physiological pH. The optimum activity of neuraminidase (NA) occurs at about pH 6.5 for the influenza enzyme[5] and the pH range 5.5-7.0 for the S. typhimurium enzyme.[6] These factors make process 3 the most appropriate comparison for any modelling of the enzyme mechanism. The choice of PNA as a model arises from the availability of experimental KIEs[4] for the hydrolysis of this compound.

9.4.1 Theoretical Modelling of PNA Hydrolysis

Since both the chair and boat forms of PNA are hydrolyzed by the NA enzyme, both conformations have been modelled. The starting structures were taken directly from the optimized coordinates of modelled enzyme-substrate (ES) systems (Chapter 10). The model consists of the PNA molecule plus one water

molecule which was identified from the ES system as being in a good position either to act as a nucleophile or to become the hydroxyl group of the product. Three simulations were considered. From the influenza ES system it was only possible to identify the substrate and water molecule which could undergo reaction leading to retention of the configuration of the substrate: no water molecules were present on the side of the substrate necessary for an inversion of configuration. This simulation begins with the sugar ring in a boat conformation, and is referenced PFR (PNA, flu, retention). The *S. typhimurium* ES system had water molecules available on both sides of the substrate, so it was possible to extract starting coordinates for both an inverting and retaining reaction. These two simulations begin with the sugar ring in a chair conformation and are labelled PSI and PSR for the inverting and retaining models.

The model was partitioned into QM and MM parts as shown in Figure 9.9, and was then set up for a solvated calculation as described in section 4.3.1. The bonds shown as dotted lines in Figure 9.9 were used to define two coordinates that were varied to generate a set of points suitable for interpolation into a PES. The spacing between each point generated was 0.2Å along each reaction coordinate. For each simulation the distance between each carboxylate oxygen and the anomeric carbon was monitored for signs of nucleophilic participation

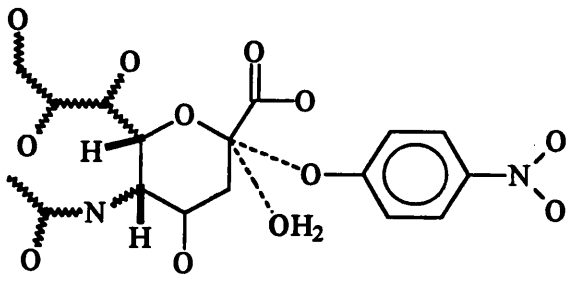


Figure 9.9. The model system for PNA hydrolysis. Dotted lines indicate the distances defined as reaction coordinates; thick lines indicate the two hydrogen link atoms; and the zigzag lines indicate those atoms which were defined as part of the MM system.

in the reaction, as has been suggested by Ashwell *et al.*[4] The simulation assumes that any intermediate will require solvent stabilization, hence the inclusion of the solvent nucleophile to anomeric carbon distance as a reaction coordinate. If the carboxylate is involved as an integral part of the reaction mechanism then it is hoped that the geometry optimization will deform the structure appropriately, because in this situation movement of the carboxylate group must lead to the lowest energy structure within the constraints applied to the two reaction coordinates. (Since the TS structure should be the highest energy point on the lowest energy path, a TS structure should be an end point for optimization from a structure geometrically close to but above it in energy.)

The PESs for the simulations described above are displayed in Figure 9.10. Potential stationary points corresponding to intermediates or transition states were identified from these surfaces. These points are identified on the two dimensional surfaces in the figures. Intermediate points were allowed to minimize freely from the nearest appropriate calculated grid point, removing the constraints on the reaction coordinate distances. Because the CHARMM program lacks an effective TS searching algorithm, approximate transition structures were sought as follows:

1. The calculated gridpoint closest to the apparent TS on the interpolated surface is taken as a starting structure.
2. A mini grid of points is calculated around this point, for a distance of 0.12A in each reaction coordinate direction, in steps of 0.02A.
3. The lowest energy structure identified is taken to be a TS structure, and is characterized by a force constant calculation.
4. Any imaginary vibrational frequencies are analyzed using CAMVIB to assess whether or not the structure represents a TS structure.

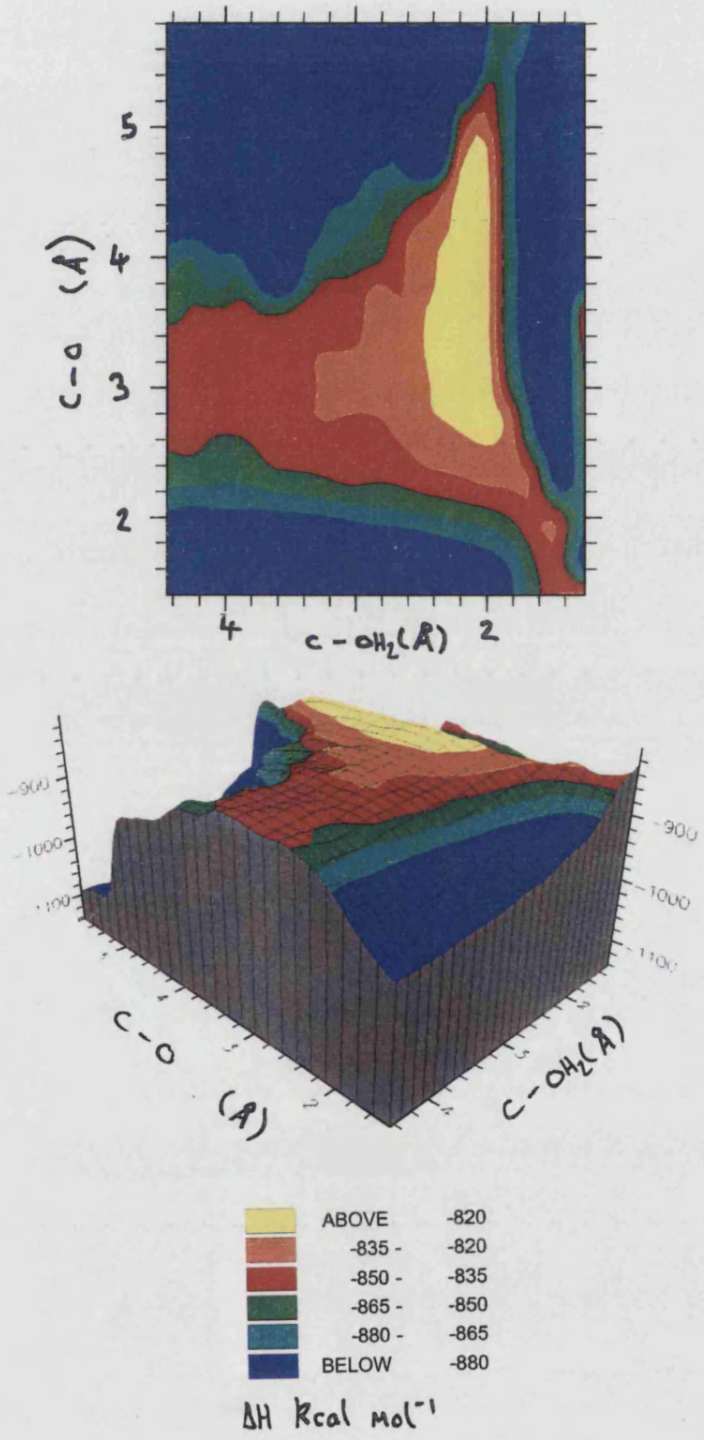


Figure 9.10a. PES for the PFR simulation (Retention of configuration, starting from a boat conformation).

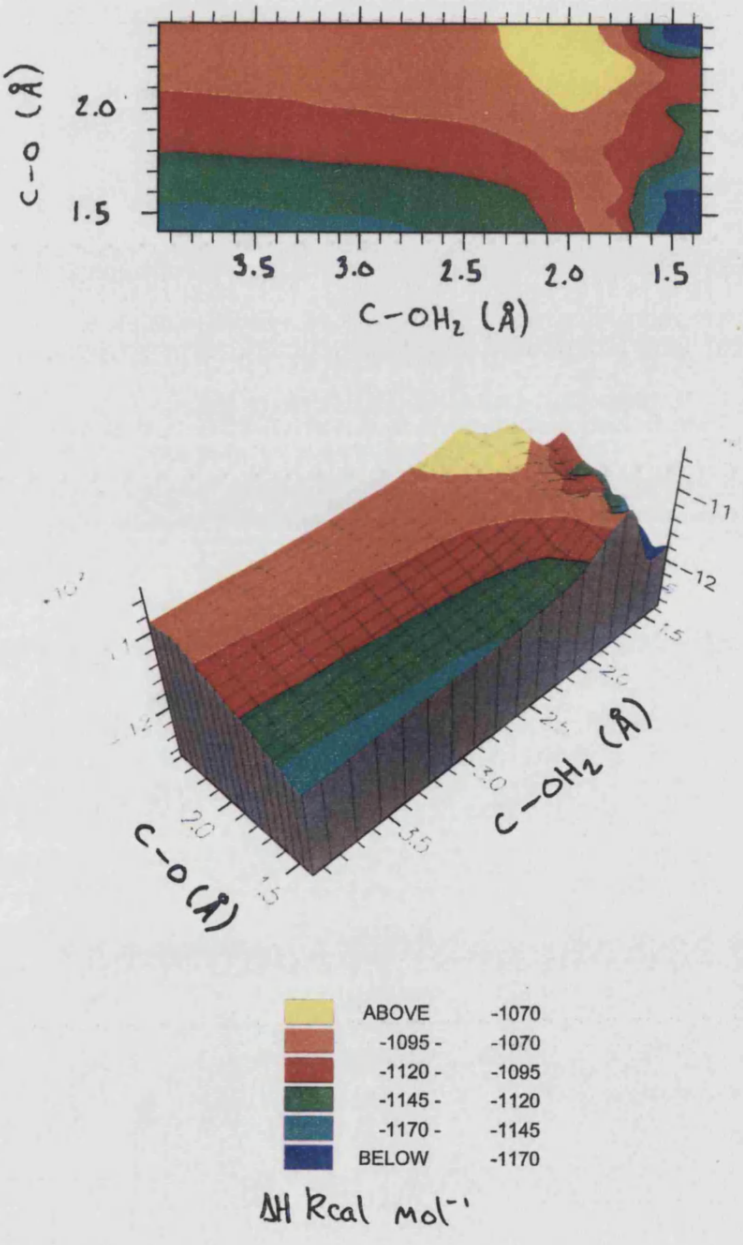


Figure 9.10b. PES for the PSR simulation (Retention of configuration, starting from a chair conformation).

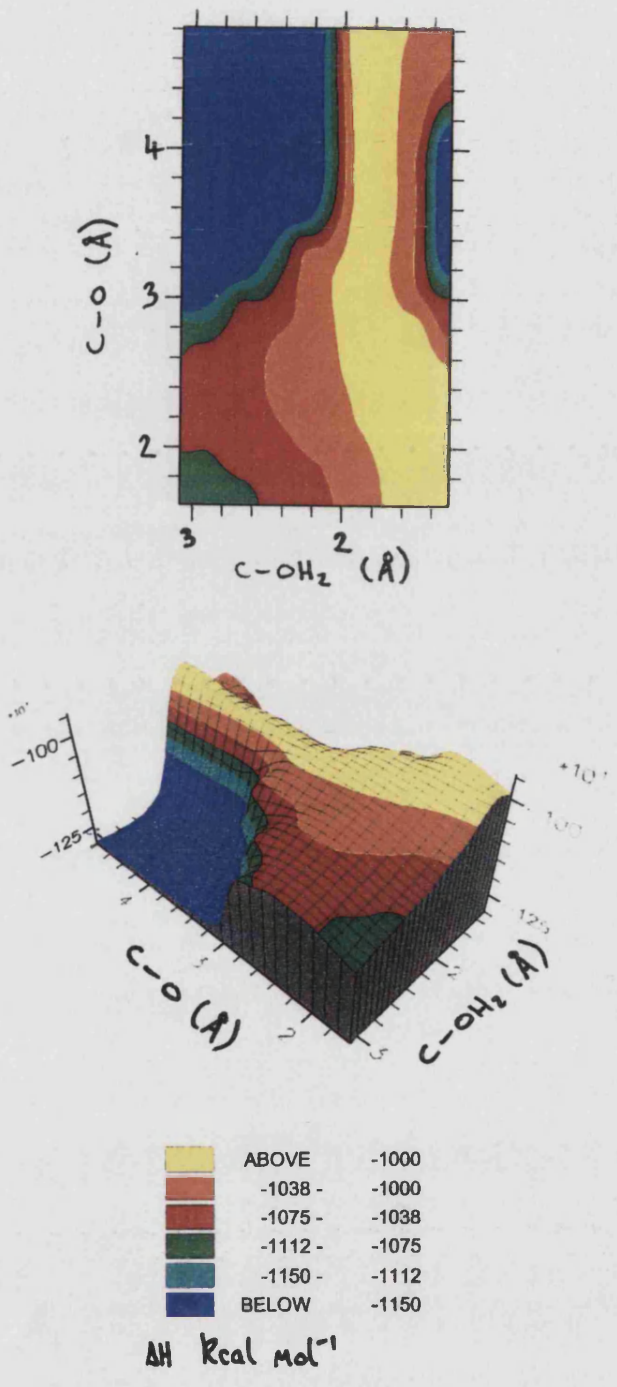


Figure 9.10c. PES for the PSI simulation (Inversion of configuration, starting from a chair conformation).

5. Steps 3-4 are repeated if necessary for each point in the mini grid in the order of ascending energy value.

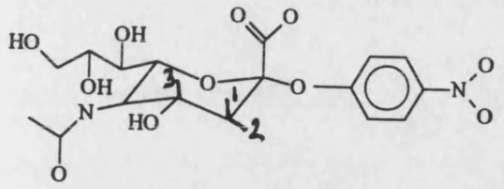


Figure 9.11. Positions of isotopic substitution. 1 is the pro R, 2 the pro S and 3 the γ- position for H or D.

Kinetic isotope effects were calculated at 333K for the atoms indicated in Figure 9.11, using TS structures identified in this way.

9.4.2 Results from the Theoretical Modelling of PNA Hydrolysis

During the generation of the PESs, the anomeric carbon to carboxylate oxygens distances were monitored for signs that the carboxylate may be involved in the reaction. No significant shortening of these distances occurred, indicating that for this simulation no nucleophilic effect is involved.

The TS searching procedure, which worked well for the AMPH system, identified only one feasible TS structure, TS2 from the PFR model. This TS structure is shown in Figure 9.12. The lack of success in identifying further TS structures is partly due to the lack of a suitable algorithm within CHARMM for searching for such points. This is one of the areas that has been identified for improvements to the software and is discussed in Chapter 11. The calculated KIE values using this TS structure, along with the experimental values, are given in

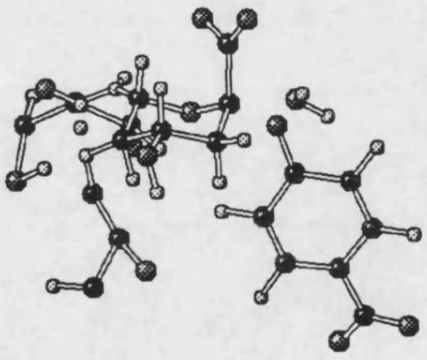


Figure 9.12. TS structure found for the PFR simulation.

Table 9.1. Secondary Kinetic Isotope Effects k_H/k_D for PNA Hydrolysis at 60°C[a]

Substituent	Calculated KIE	Experimental KIE
Pro R D	1.045	1.073
Pro S D	0.870	1.000
Pro R+S D2	0.909	1.075
γ D	0.992	1.000

[a] Experimental work done at pH 6.67, reference 4

Table 9.1. The calculated KIEs are disappointing after the previous successful work. The fact that the pro R and pro S positions give KIEs of opposite effect indicates that the TS structure is fundamentally flawed and cannot represent a point along the reaction coordinate for this particular reaction. Since the modelling has failed to reveal any respectable intermediate structures it is not safe to attempt to produce structural information based on EIE calculations.

There are a few comments regarding the TS structure found. The reaction was started at a boat conformation, which in solution is unfavourable. It is therefore not unexpected that the conformation has flipped to a chair-like structure at the TS. The experimental KIEs are not reproduced, since the important dihedral angles, which can be estimated from the isotope effects, must of necessity be wrong, assuming that the reaction does proceed *via* a flattened $^2C_5$ chair conformation, as suggested by Ashwell and co-workers[4]. The calculated TS structure in fact has a nearly perfect $^2C_5$ chair conformation, and has dihedral angles, made by the C-H pro R and pro S bonds with the anomeric carbon to leaving group bond, of 90.2° and -28.4°. From consideration of the

experimental KIEs they should be close to 0° and 90° respectively. This makes it extremely doubtful that this is the TS for the intended target of the modelling, especially since no explicit account was taken of the carboxylate interaction. This simulation was included to demonstrate the unfavourable nature of reaction through the boat conformation, and apparently it does do that in that the conformation changes to a chair.

Hindsight always seems very wise in situations like this. Given the importance placed upon the carboxylate participation by the experimental workers, it may have been wise to simulate this motion in some manner. However, the problems inherent in attempting to cater for three potentially interdependent nuclear motions would have greatly increased the computational cost of the project. It was also thought reasonable that if the carboxylate motion was important then the optimization process might cause the correct deformation of the structure as this would presumably be on the lowest energy path. The fact that it did not suggests either that the carboxylate does not participate nucleophilically, or that there is at least one "dead end" pathway whose energy is lower than that of the actual reaction. The current approach cannot eliminate such pathways. One potentially useful scheme which could be tried would be to simulate the formation of an α-lactone structure using one carboxylate oxygen to anomeric carbon distance as one reaction coordinate and the anomeric carbon to leaving group as the second. If the "nucleophilic carboxylate" has a real existence then the partial or complete formation of the α-lactone would be a distinct event either requiring or being the TS structure, and might occur prior to addition of water. The alternative that I modelled may over-emphasise the role of the water molecule which subsequently completes the reaction. This was perhaps a dangerous extrapolation from glucoside chemistry.

A preliminary calculation on the PNA molecule without a water molecule as

nucleophile, using AM1 *in vacuo* with the carboxylate oxygen to anomeric carbon distance as a reaction coordinate, did find a TS structure. This corresponded to initiation of the α-lactone formation. Subsequent shortening of the reaction coordinate to complete the α-lactone formation caused opening of the glycosidic ring rather than the loss of p-nitrophenolate anion, although there was some evidence for a lengthening of the bond to the p-nitrophenyl group prior to ring opening. There is the possibility that the inclusion of a water molecule at that stage may have led to the desired reaction. There was insufficient time to investigate this possibility further. Chapter 11 contains some suggestions for additional work which could usefully be done to model this reaction more successfully.

Given that the CHARMM approach has been successful for the AMP hydrolysis, it is probable that one of the causes for failure in this case was a poor choice of reaction coordinate. As discussed above, alternative choices can be selected which may provide a better model of the true reaction. Even with a reaction coordinate that seems to be appropriate, the many degrees of freedom involved in the system can cause an unintended deviation from the expected path. The lesson appears to be that it is unwise to suppose that a simulation will follow the intended path: the experimenter must navigate more warily!

9.5 References

1. MacroModel reference

2. C. Andrews, B. Fraser-Reid & J. Bowen, *J. Am. Chem. Soc.*, 1991, **113**, 8293

3. J. F. Liebman & A. J. Greenberg, *J. Org. Chem.*, 1974, **39**, 123

4. M. Ashwell, X. Guo & M. L. Sinnott, *J. Am. Chem. Soc.*, 1992, **114**, 10158

5. M. R. Lentz, R. G. Webster & G. M. Air, *Biochemistry*, 1987, **26**, 5351

6. L. Hoyer, P. Roggentin, R. Schauer & E. R. Vimr, *J. Biochem.*, 1991, **110**, 462

CHAPTER 10.
MODELLING THE MECHANISM OF SIALIDASE ENZYMES

"The protein molecule model resulting from the X-ray crystallographic observations is a Platonic protein, well removed in its perfection from the kicking and screaming stochastic molecule that we must infer must exist in solution."

- G. Weber

10.1 Introduction

The overview of sialidases in Chapter 8 makes it clear that their catalytic mechanism is an important and interesting area of research in biological chemistry. As far as I am aware, no QM modelling studies of the mechanistic aspects of the sialidase enzymes have yet been published. The evidence from the work reported so far in this thesis, showing the subtle differences between S_N1-like and S_N2-like mechanisms at their interface for glycosidic hydrolysis, has some bearing on sialic acid chemistry. In all probability, the different mechanisms of the *S. typhimurium* and influenza sialidases are likely to be caused by small effects that are unlikely to be seen using MM modelling alone, even using the relatively high resolution X-ray structures currently available. Although enzyme mechanisms are in general more complex than their solvated counterparts, in principle the PES approach used for solvated systems in this thesis should be applicable to enzyme-substrate systems, provided that care is taken to isolate the individual steps of any reaction scheme which is to be tested. The failure of the PES approach for the hydrolysis of PNA in water was due in part to the neglect of a potentially important molecular motion, as was discussed in Chapter 9. The reaction steps which are to be modelled in this chapter will be shown to be different in character and not subject to that type of limitation.

10.2 Models of the Enzyme Mechanism of NA

The proposed NA mechanisms described in section 8.2.3 indicate the uncertainty that currently exists about these enzymes. There are two schemes which appear amenable to the PES approach, one for the flu B NA and the other

for the *S. typhimurium* NA. A short justification for the choice of each reaction scheme to be modelled follows.

The evidence for a stabilized sialosyl cation as an intermediate in the influenza B NA reaction[1] makes for an obvious study, because there is no nucleophile or proton donor in the proposed mechanism. The reaction is therefore driven solely by the cation formation, the driving force presumably coming *via* the active site structure. Starting from the PNA substrate bound in the boat conformation to the enzyme, the use of the anomeric carbon to leaving group oxygen as one reaction coordinate and the approach of a water molecule as the second coordinate should provide a reasonable model for this postulated mechanism. There are two caveats to this simple sounding step. The first is the evidence for a chair to boat conformational change after binding, which has been given by Guo[2] and co-workers in a study on the influenza A NA. This step may be kinetically important; if so, and if influenza B NA exhibits a similar step, then it may not be possible to calculate accurate KIEs for the reaction using the proposed model. There is little evidence to suggest that influenza A and B NA are significantly different, so this point must be borne in mind. The second caveat relates to the possible nucleophilic participation of the carboxylate group. There is the danger of missing any movement of the carboxylate group, as may have been the case for the study in Chapter 9. However, there are good reasons for supposing this to be less critical on the enzyme. The most obvious is that in the active site one assumes that the environment is near optimal for stabilization of the TS structure. This implies that the reaction is more or less driven by its environment. A good chemical model for the atomic interactions would therefore be expected to reproduce this because there is little chance of an alternative outcome to the reaction: the reaction is more structured than in solution. Also, in the boat conformation on the enzyme, the carboxylate group is

held in position by hydrogen bonding interactions with three arginine (Arg) residues, making significant movement highly unlikely. Finally, if the conformational change mentioned above involves the carboxylate group acting as a nucleophile, then the model need not consider it anyway.

The *S. typhimurium* NA mechanism is thought to proceed by a single chemical TS,[2] with little proton donation to the leaving group. The leaving group oxygen is virtually cleaved from the anomeric carbon at the TS. Once again, this description indicates that there is little interaction from either a nucleophile or proton donor, and that a relatively simple selection of the reaction coordinates, identical to that of the influenza B model, should suffice to test this reaction scheme. The reservation about the nucleophilic participation of the carboxylate group remains for this simulation. Once again, however, the group is hydrogen bonded to three Arg residues.

Within the above reaction schemes, some active site residues have been identified as having a role to play, and these must be included in the QM system for modelling. A tyrosine (Tyr) residue has been identified in both NAs as being involved in stabilizing a sialosyl cation or similar highly charged structure. For the flu B NA, it has been noted that Tyr409, Asp149 (aspartic acid) and Glu275 (glutamic acid) are all important for enzyme activity, and by implication all stabilize the proposed intermediate cation. The equivalent residues in the S. typhimurium NA have also been implicated in various reaction schemes, the Glu being a proton donor and the Asp making a hydrogen bond to the hydroxyl group of the substrate ring. There is therefore a core of three active site residues which are important for the activity of both enzymes and which play a role in both reaction schemes which have been chosen as the basis of the modelling studies to be undertaken. These three residues will be incorporated into the QM system. In my flu model, these residues are Tyr333,

Asp73 and Glu200; for the *S. typhimurium* they are Tyr341, Glu230 and Asp61. The three Arg residues mentioned earlier are kept in the MM system because they appear to function as anchors for the structure rather than to take part in the electronic reorganisation of the system.

10.3 Construction of Initial Enzyme-Substrate Structures

The X-ray coordinates[3] for the influenza B NA enzyme with bound DANA were obtained from the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory. The resolution of this structure was 1.8Å. The X-ray coordinates for the *S. typhimurium* NA enzyme with bound DANA were obtained from the originators of the data.[4]

To make these X-ray systems suitable for the proposed modelling, a number of manipulations were necessary. These manipulations were done on both systems and are described below.

The CHARMM program was used for the modelling, so the X-ray coordinates and enzyme residue sequence were read into CHARMM to provide a starting point. These X-ray structures will be identified as F0 and S0 for the flu and salmonella NAs respectively. The CHARMM function ADDH was used to build protons onto the heavy atoms, using the standard ionisation state of each residue as defined in the CHARMM topology files. Using an MM-only force field, these structures were then energy minimized for 100 steps of steepest descent to remove any bad contacts due to the added protons, followed by geometry optimization using the Powell minimization algorithm until the rms gradient was less than 0.01. The resulting structures were labelled F1 and S1 following the naming system above. The rms deviation from the X-ray heavy atom protein coordinates was 0.05Å for F1 and 0.07Å for S1.

The X-ray structures contained DANA, whereas the modelling studies require PNA in the active site. The DANA coordinates were extracted and used as a template to orient the AM1 optimized structures for the chair (*S. typhimurium* NA) or boat (flu B NA) PNA into the enzyme active site. The PNA was overlayed on DANA as shown in Figure 10.1. Atoms C1, C2 and O3 in the ring of DANA were used as successive anchor points in the overlay facility of HyperChem[5], to orient the PNA model. The ring and carboxylate atoms of DANA were than replaced by the ring, carboxylate and p-nitrophenyl atoms from the PNA, maintaining the X-ray coordinate system. The new enzyme-PNA structures were then geometry optimized as above until an rms gradient of less than 0.05 was achieved. These structures were labelled F2 and S2.

The enzyme active site of each model was solvated by adding a sphere of water molecules, as described in Chapter 4. The anomeric carbon of PNA was taken to be the centre of the water sphere. Any solvating water molecules that had an oxygen atom within 2.5Å of any heavy atom of the enzyme-substrate (ES) complex were deleted. The systems were geometry optimized as before until an rms gradient of less than 0.05Å was obtained. These structures were labelled F3 and S3. Table 10.1 provides a quick description of each of these preliminary structures.

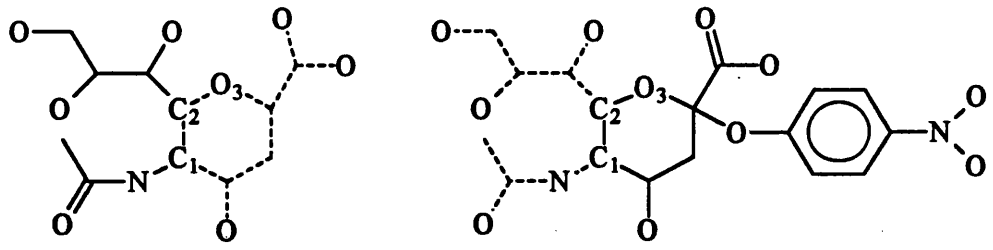


Figure 10.1. Orienting the PNA structure onto the enzyme active site, using the DANA substrate as a template. Dotted lines indicate the atoms that were deleted after the overlaying process. The absolute values of the cartesian coordinates of the DANA atoms were preserved.

Table 10.1. Description of the Preliminary Structures Generated for the Enzyme-Substrate Complexes of NA

Structure[a]	Description
F0, S0	Original X-ray structure
F1, S1	Optimized structure after adding hydrogens
F2, S2	Optimized enzyme-PNA structure
F3, S3	Optimized enzyme-PNA structure with solvated active site

[a] F labels are for flu , S labels for *S. typhimurium* NA structures

10.3.1 Preliminary Observations of the Enzyme-Substrate Models

As an essential part of the planning for the modelling work, the constructed models were viewed to observe any obvious structural features which could lead to some ideas regarding the different mechanisms of the two enzymes. In particular, a model of the *S. typhimurium* with a substrate in the active site has not previously been available. Figure 10.2 shows a representation of the two active sites with the PNA substrate.

The *S. typhimurium* NA picture suggests that, due to the chair conformation of the PNA ring, the leaving group obstructs the solvent access to the anomeric carbon but at the same time allows water to migrate into the active site. This would allow water access to the correct face of the ring to allow inversion of configuration. At the same time, restricting access of the solvent to the anomeric carbon would inhibit the retention of configuration. Because of the shape of the active site, it can be imagined that as the leaving group migrates into the solvent it obstructs the approach of water to the moiety remaining on the enzyme; this time could be sufficiently long for any water in the active site to complete the next step in any reaction sequence. Alternatively, the water could already have acted from the inverting side of the glycosidic ring to release the leaving group.

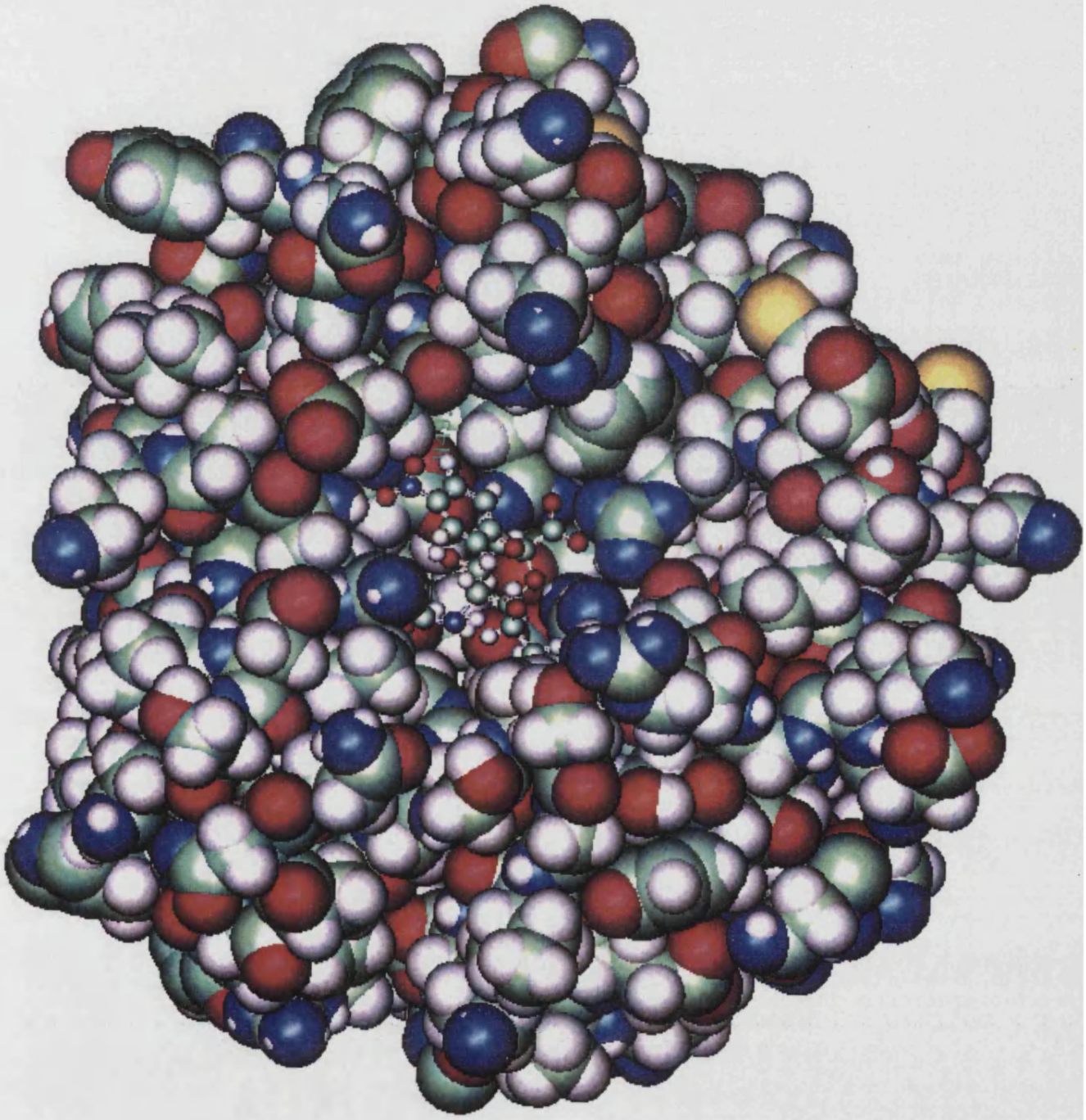


Figure 10.2a. View of the active site of the initial influenza NA enzyme with bound PNA substrate.

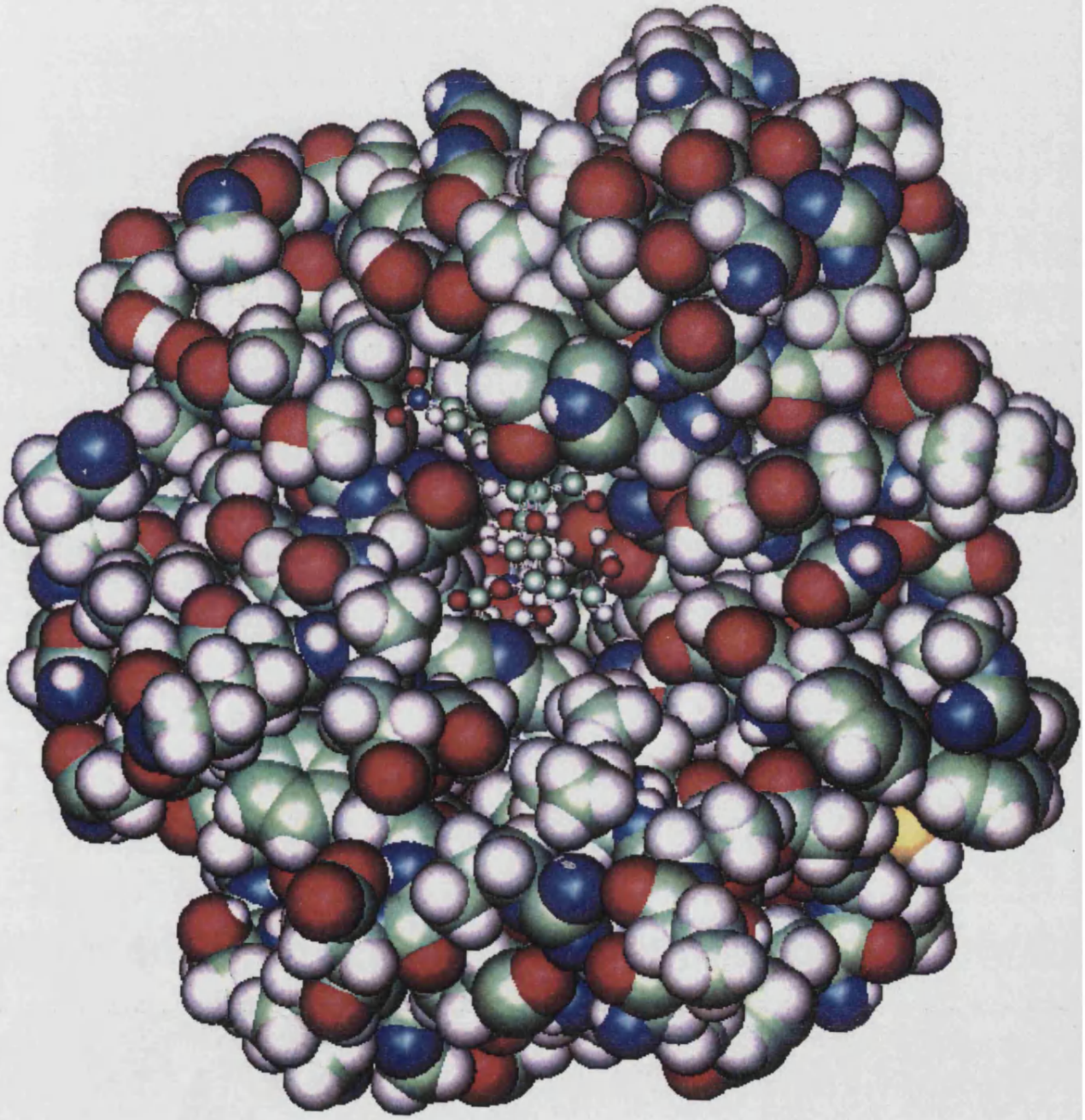


Figure 10.2b. View of the active site of the initial *S. typhimurium* NA enzyme with bound PNA substrate.

The boat conformation of the PNA on the flu NA can be seen to allow solvent access to the anomeric carbon on the side necessary for retention of configuration. At the same time, access to the active site by solvent is restricted. This is the opposite situation to that just described for the *S. typhimurium* enzyme.

There is thus a structural difference which might provide at least a partial explanation of the different mechanisms. Could this also explain the large difference in turnover number for the two enzymes? If the *S. typhimurium* active site contains a site for water suitably positioned for preassociation with the substrate, then the answer must be a qualified yes, if only because the β-anomer that results is the preferred anomer in solution. This may be an additional driving force behind product release, since product release is rate-determining for good substrates of this enzyme. Although it is not possible to identify such a site, it is possible to identify suitable water molecules in the *S. typhimurium* model that are placed in positions where an inverting reaction is possible: this is not the case for the flu model. There is a water molecule at the opposite end of the ring to the anomeric carbon in the flu model, but this is surely an unfavourable position from which to approach. In contrast, the *S. typhimurium* model contains a water molecule almost directly above the anomeric carbon, on the side required for inversion of configuration. Water cannot approach in this manner on the flu model for the reasons discussed earlier.

For both models it is possible to identify water molecules that are suitably placed for a retaining reaction. In the case of the *S. typhimurium* model, the water cannot act as a nucleophile because of hindrance by the leaving group, but it could add to the anomeric carbon after the leaving group has migrated into the solvent. The flu model shows the water molecule to have direct access to the anomeric carbon.

10.4 The Modified Stochastic Boundary Molecular Dynamics Method for Enzyme Systems

As indicated in Chapter 4, the calculations performed on enzyme systems use a modified version of the stochastic boundary molecular dynamics (SBMD) method. Although there are strictly speaking no dynamics calculations involved in my simulations, the SBMD method does involve the use of constraining forces and it is appropriate to test my modified method to ensure that it performs satisfactorily. There is also the possibility that MD calculations will be done in future work so this initial test serves to demonstrate the utility of the method.

The standard SBMD method[5] partitions the system into three regions, centred on the active site of the system. The reaction region, which incorporates the active site, is spherical. Around this is a shell of atoms called the buffer region, and the rest of the system is called the reservoir region. Atoms in the reaction region are described by the full force field methods being used. The buffer region atoms provide a method of coupling the reaction region and reservoir region. They move as interacting Langevin particles via a Langevin dynamics algorithm.[6] This provides a method by which energy fluctuations can be included in the model via coupling to a temperature bath, using friction coefficients within the Langevin algorithm. The buffer region atoms also have associated boundary forces which simulate the effect of the reservoir region atoms. In the standard SBMD method these forces would be assigned by calculating the mean forces acting on atoms by analysis of a full MD simulation. The reservoir region atoms themselves are deleted. Since these atoms would account for a large part of the system, their deletion allows for a faster MD simulation than would otherwise be possible.

Results from previous use of the SBMD method have suggested that the

choice of mean forces is not too critical for localised protein motions: indeed forces have been averaged over side chain heavy atoms.[7] This suggested that it should be adequate to replace these mean forces with the CHARMM harmonic constraint forces, which are mass-weighted and give an equivalent fluctuation about the mean atomic position of about 8Hz. The friction coefficients used are derived from velocity correlation functions and have been taken from reference 5. The comparatively minor modification to the assignment of mean boundary forces means that a preliminary MD simulation is not necessary to obtain the mean forces.

10.5 Validation of the Modified SBMD Method

The modified SBMD method was tested by doing a set of MD simulations on the flu B sialidase enzyme with DANA substrate. The X-ray structure F0 was used as a reference structure. Three short simulations were performed. Simulation M1 was a standard MD simulation in that the whole system was included. The system was not partitioned and no friction or mean forces were used. This was the standard simulation to which the SBMD method should compare favourably. It was started from the optimized F1 structure. M2 was a SBMD simulation which was carried out as described in section 4.5, except that no solvent was added around the active site. This serves to demonstrate the effect of the friction coefficients and harmonic constraints imposed on the buffer atoms. This simulation was started from the truncated F1 structure. Before the MD simulation was started, this truncated structure had the harmonic constraints applied and was geometry optimized until an rms gradient of less than 0.05 was achieved. M3 was also a SBMD simulation, this time with the active site solvated. This shows the additional effect of the solvent around the reaction

region. This also used F1 as a start point and after addition of solvent and truncation was geometry optimized prior to the MD simulation. Simulation M1 contained 3247 heavy atoms, M2 1407 and M3 1536. Each simulation was run for 40ps, the first 10ps of which was equilibration time.

Table 10.2 summarises the rms differences and fluctuations of the protein coordinates across the simulations. As a comparison, the rms differences of the F1 structure are given. The values are compared across the heavy atoms of the protein that remain in the cutoff models; there are 1407 of these. The rms values were calculated from the averaged coordinates of the last 25ps of each dynamics simulation compared to the appropriate reference structure.

The coordinate rms differences between the SBMD simulations and the full dynamics MD are small compared to their difference from the X-ray structure. This suggests that the integrity of the dynamics simulation is maintained by the SBMD simulations. For each simulation, the rms differences of the buffer zone atoms are lower than those of the reaction zone atoms, and similarly the backbone atom differences are lower than the residue atoms.

The rms atomic fluctuations provide an indication of the stiffness of the protein structure. The buffer zone atoms in the SBMD simulations are less mobile than the equivalent full MD simulation; however the reaction zone atoms appear to have similar mobility across all three simulations. Overall, the solvent SBMD simulation is nearer the full simulation than the vacuum SBMD, using the rms comparisons as the guide.

From the point of view that precise reproduction of MD behaviour is not of the utmost importance for the current research, the above results give sufficient confidence that the use of the harmonic constraints in CHARMM will not harm the modelling of the reaction and do allow the reaction zone atoms as much movement as in a full system that is not truncated.

Table 10.2. Protein Coordinate[a] rms Differences and Fluctuations for the SBMD Validation Simulations (Å)

Description	Atoms				
	Buffer	Reaction	Backbone	Residue	Overall
rms difference from F0					
F1[b]	0.01	0.02	0.01	0.03	0.03 (0.05)[c]
M1	1.41	2.31	1.21	2.53	1.98
M2	1.63	2.57	1.22	2.77	2.19
M3	1.44	2.28	1.19	2.48	2.00
rms difference from M1					
M2	0.33	0.68	0.35	0.74	0.54
M3	0.34	0.87	0.40	0.94	0.74
rms fluctuation about F0					
M1	0.41	0.46	0.37	0.51	0.43
M2	0.27	0.41	0.28	0.45	0.34
M3	0.28	0.43	0.29	0.51	0.37

[a] Only the 1407 heavy protein atoms that are in the truncated system.
[b] Shown for comparison only. [c] Over all the protein heavy atoms.

10.6 Computer Simulation of the Enzyme Mechanism of Sialidase

Now that the effectiveness of the constrained and truncated enzyme model has been established, the combined QM/MM force field can be used as it has been before to generate a grid of points from which to interpolate a PES. The only remaining preparation is to identify the QM system and enter any link atoms into the system that are necessary.

The PNA molecule will be partitioned in the same way as for the previous simulation in water, as shown in Figure 9.9. The three residues that have previously been identified for incorporation into the QM system will be partitioned into QM and MM parts as shown in Figure 10.3. Each model has 65 QM atoms in total, and about 1500 MM atoms.

It has been possible to identify water molecules in the *S. typhimurium* model that could be used to simulate both retaining and inverting reactions, and this will be done. For the flu model, only water suitable for simulating a retaining

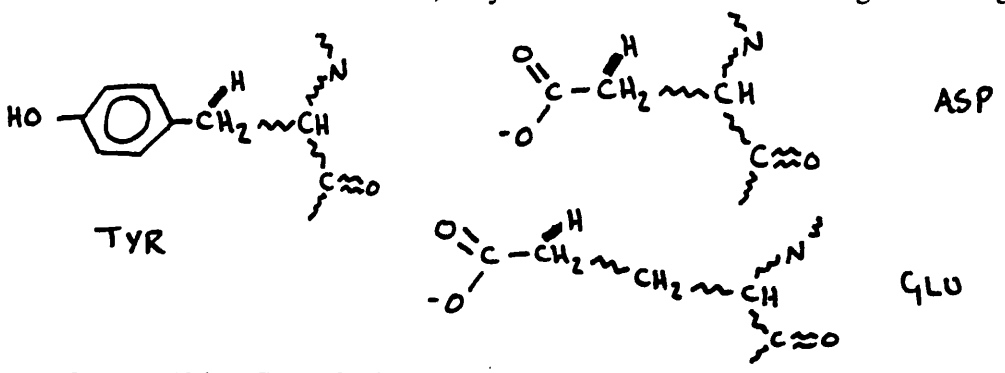


Figure 10.3. The QM/MM partitioning for the three protein residues incorporated into the QM system. Zigzag lines represent MM atoms and the thick lines indicate the added link atoms.

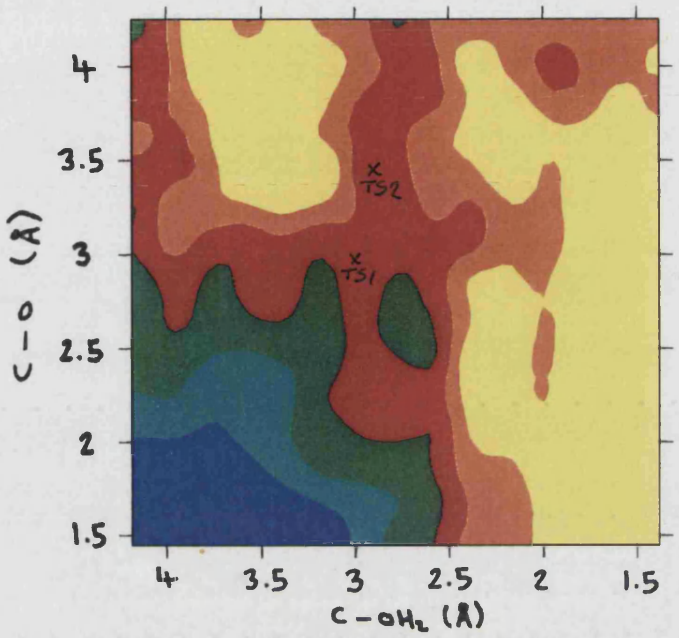
reaction has been identified. There are therefore three simulations which will be done. One reaction coordinate is the distance between the anomeric carbon and the leaving group oxygen, the other is the distance between the anomeric carbon and the incoming water molecule. As already discussed, the flu

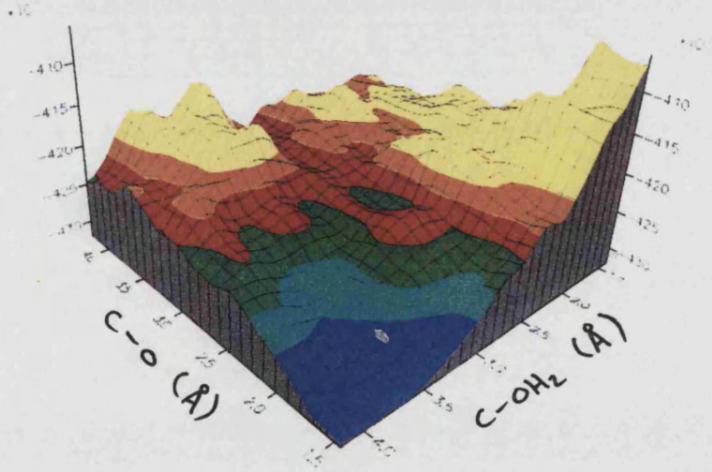
retaining and *S. typhimurium* inverting mechanisms under simulation have no proton donating or non-water nucleophiles postulated, so the reactions are apparently driven by the stabilizing forces within the active site. The simulation used here is therefore appropriate. For the retaining *S. typhimurium* simulation, which does not actually occur, the purpose is to provide information as to why it doesn't occur. The assumption inherent in this simulation is that the reaction would occur in a similar way to the flu NA reaction. The flu enzyme simulation will be identified as EFR, and the *S. typhimurium* ones as ESR for the retaining simulation and ESI for the inverting simulation.

The simulations were carried out as described in section 4.3.2. The reaction coordinates were varied in steps of 0.2Å. As for the solution reaction of PNA, the distance between each carboxylate oxygen and the anomeric carbon was monitored for signs of nucleophilic participation in the reaction. Figure 10.4 shows the PESs generated in this way. Approximate stationary points were identified and refined as described in section 9.4.1.

10.6.1 Results of the Simulation of Enzyme Mechanism

The preliminary stationary points identified from the PESs in Figure 10.4 are marked on the two-dimensional surfaces. It was possible to characterize one TS structure on each of the EFR (label TS2, Figure 10.4a) and ESI (label TS2, Figure 10.4c) surfaces: none could be found on the ESR surface. These TS structures are shown in Figure 10.5. It should be noted that these TS structures are only approximate. They each contain some spurious negative modes; however, analysis within CAMVIB shows that these are due to bending modes in the aromatic ring of the p-nitrophenyl group. Analysis of some random points on the original grid revealed that this appears to be a general problem throughout the structures found by CHARMM in the enzyme simulations, and will be discussed in the next chapter. The decision was taken to continue with





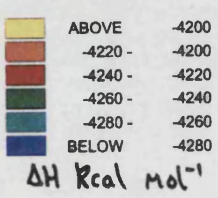


Figure 10.4a. PES for the EFR simulation (flu NA, retaining mechanism). The two dimensional plot indicates preliminary identification of stationary points.

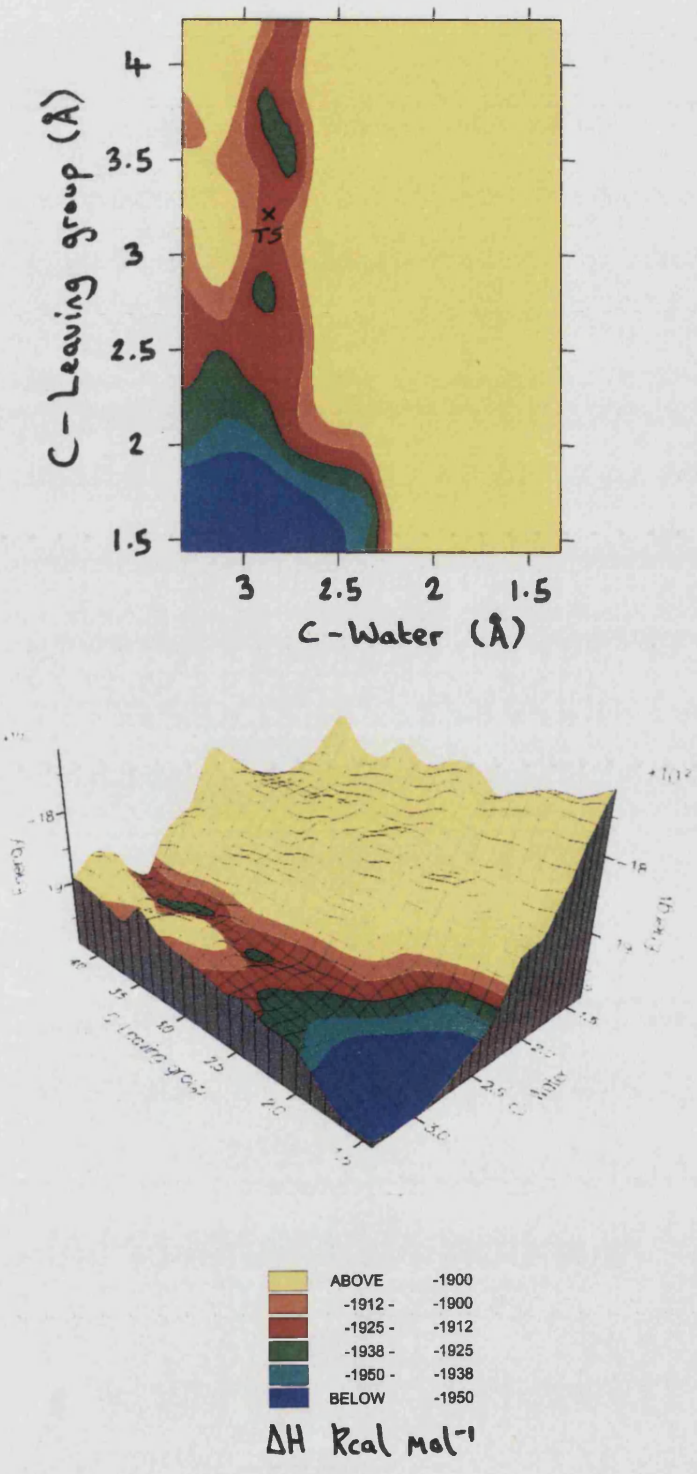


Figure 10.4b. PES for the ESR simulation (S. typhimurium NA, retaining mechanism). The two dimensional plot indicates preliminary identification of stationary points.

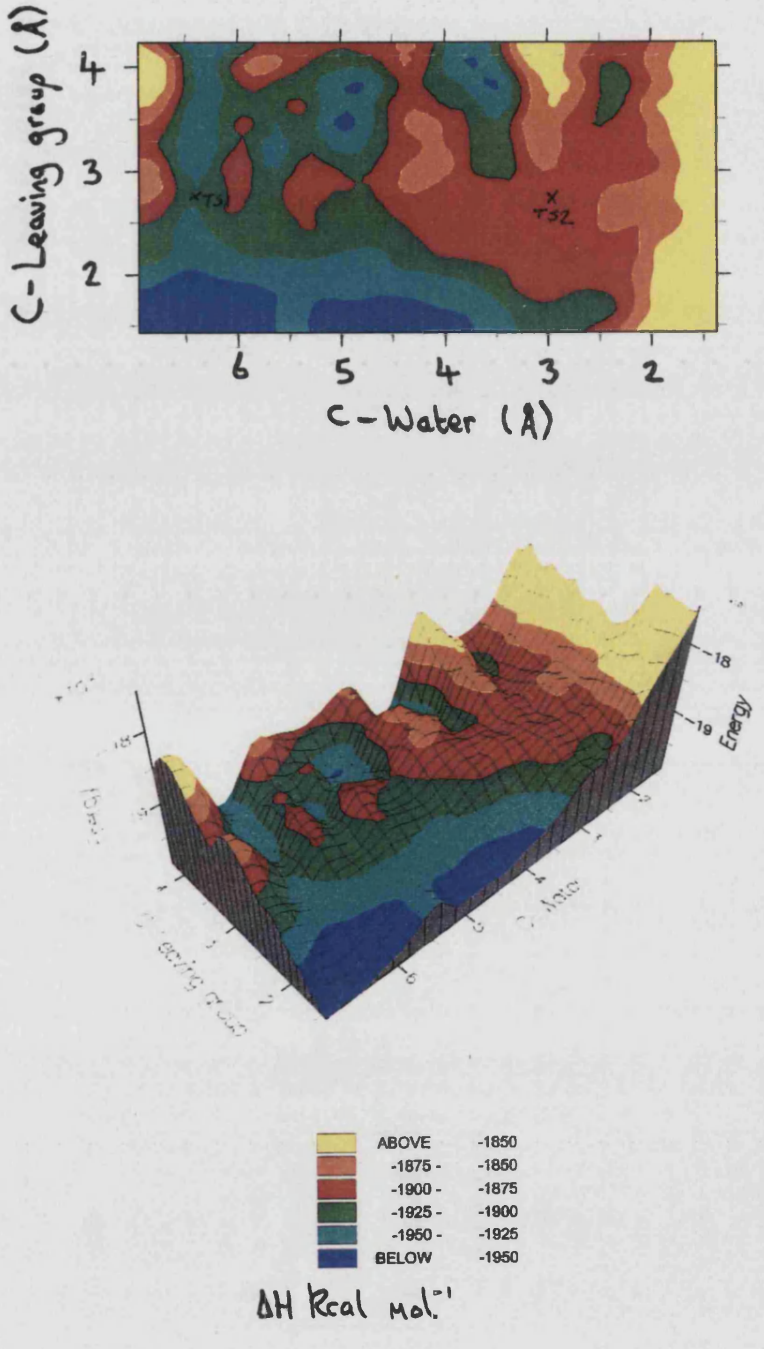


Figure 10.4c. PES for the ESI simulation (S. typhimurium NA, inverting mechanism). The two dimensional plot indicates preliminary identification of stationary points.

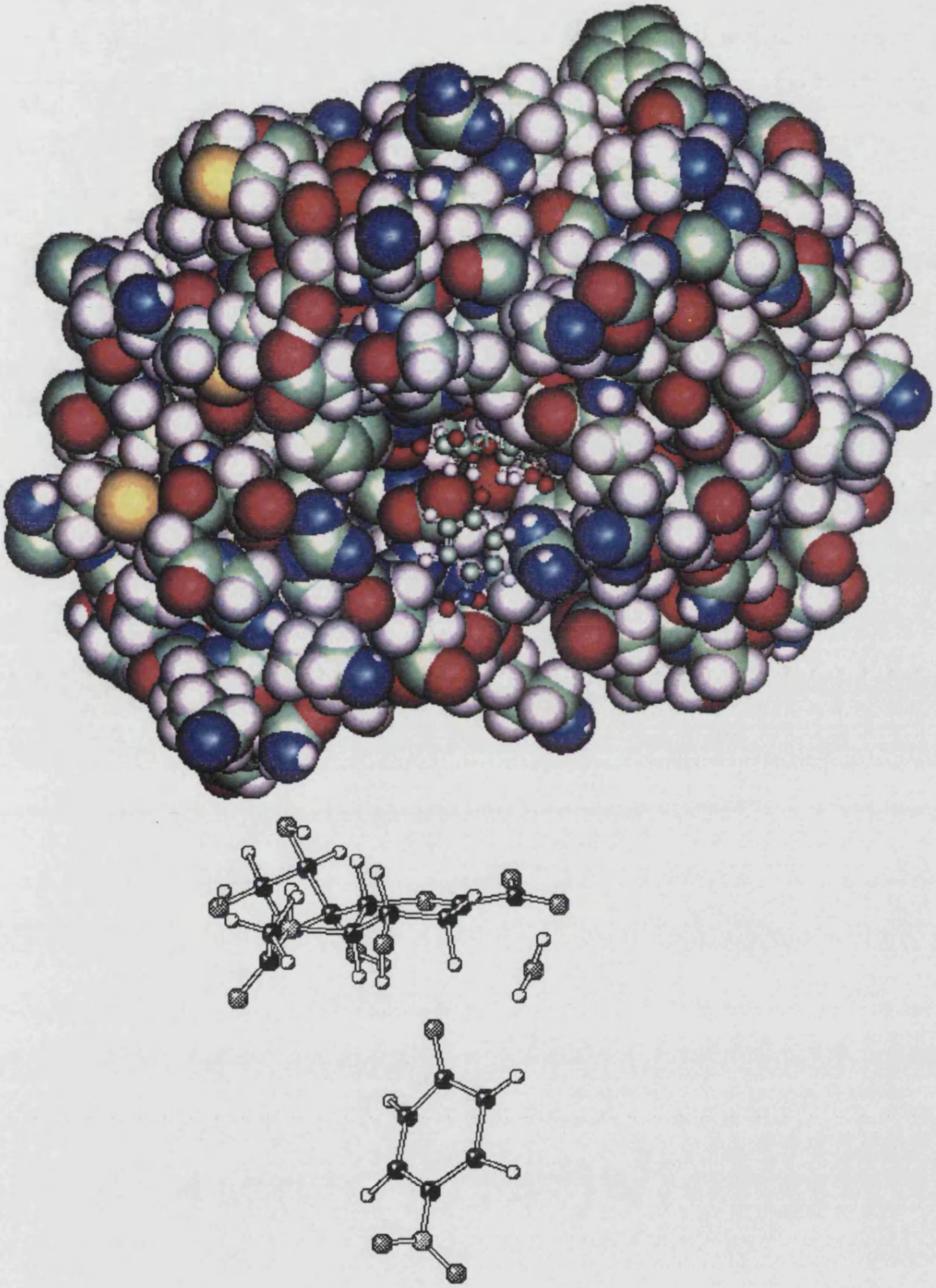


Figure 10.5a. TS structure for the enzymic hydrolysis of PNA by flu NA. The upper picture shows some active site detail.

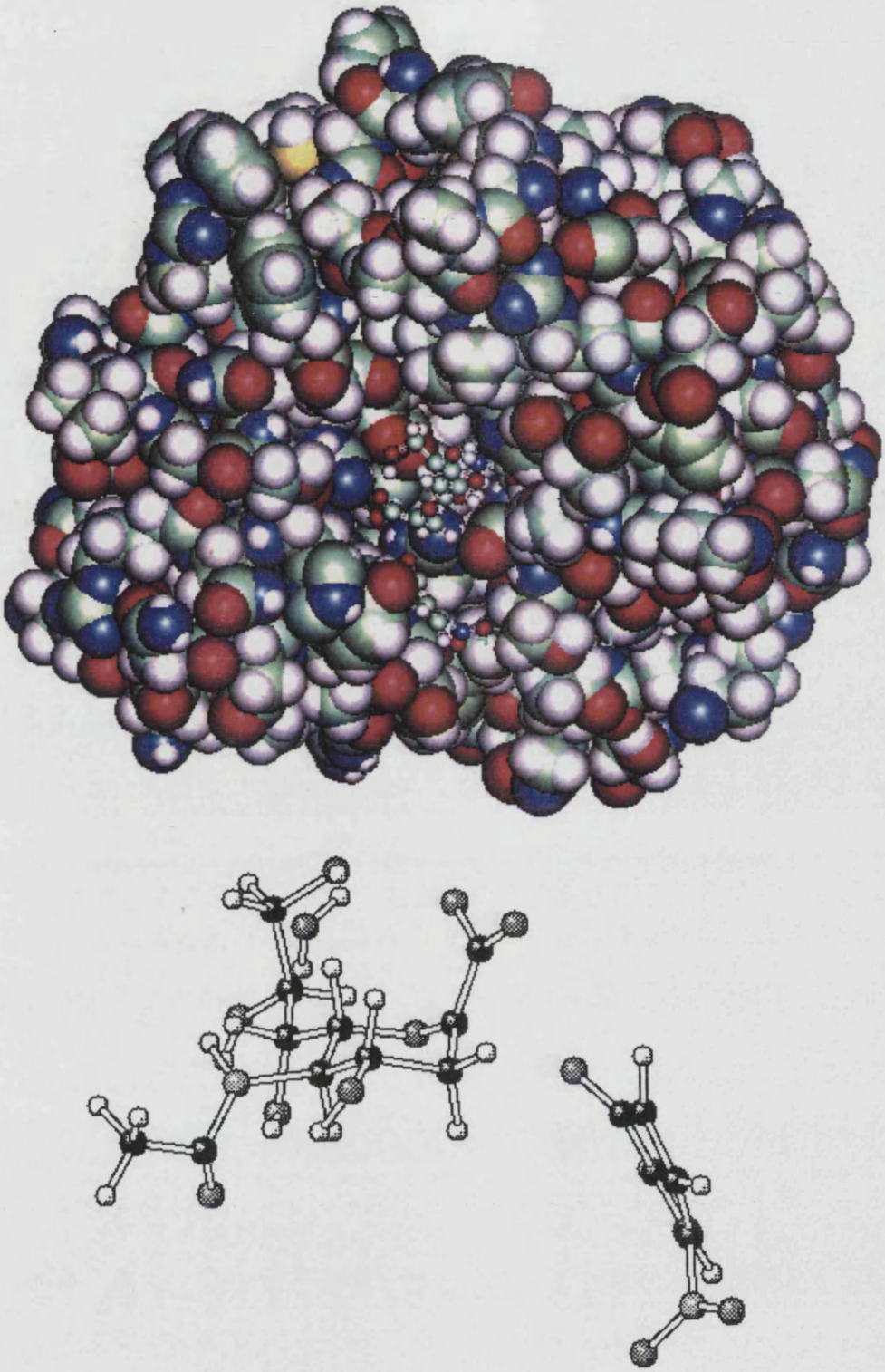


Figure 10.5b. TS structure for the enzymic hydrolysis of PNA by *S. typhimurium* NA. The upper picture shows some active site detail.

calculation of KIEs because for each TS structure, the only other significant contributions to the negative modes came from the stretches and bends associated with movements of the anomeric carbon and its bonds to the leaving group and incoming water.

Table 10.3 shows the calculated and experimental KIEs for the enzymic hydrolysis of PNA. It should be noted that the experimental flu values are for the flu A NA and not the flu B NA that I have modelled, but this is the only available data. Since both types bind NANA in the boat conformation and have a similar turnover number, it seems reasonable to suppose that they share a common mechanism. The calculated values do bear some resemblance to the experimental ones, but there is not good agreement. The calculated KIE for the

Table 10.3. Calculated and Experimental KIEs for the Enzymic Hydrolysis of PNA by influenza and *S. Typhimurium* Sialidases

Isotopic Substitution	Influenza NA		*S. Typhimurium* NA	
	Calculated	Experimental	Calculated	Experimental
Pro-R[a] ^{2}H	1.024	1.035±0.010	1.037	1.051±0.011
Pro-S[a] ^{2}H	1.142	1.058±0.006	1.083	1.050±0.014
Both ^{2}H	1.177	1.087±0.015	1.128	1.097±0.017
Lg ^{18}O	1.047	1.045±0.021	1.041	1.050±0.010

[a] See Figure 9.11 for definition

pro S hydron in the flu simulation is much higher than the experimental value, and exceeds the maximum estimate of 1.12 given by Guo and coworkers.[2] Their estimates for the dihedral angle ∅ (section 5.2) are 23° for the pro S hydron and 143° for the pro R hydron: these compare with 30.0° and 145.5° for my

calculated TS structure. Similarly, for the *S. typhimurium* simulation, Guo and coworkers estimate -60° for the pro S hydron and 60° for the pro R hydron: these compare with -42.1° and 84.3° for my calculated TS structure.

The calculated and experimental ^{18}O leaving group KIE values agree reasonably well. I have calculated the EIE of the ionization of p-nitrophenol to be 1.043, and Rosenburg and Kirsch[8] have calculated 1.0425 as the ionization of p-nitrophenyl β-galactopyranoside to a galactosyl cation and a p-nitrophenolate anion. This suggests that the KIE values obtained both experimentally and *via* calculation for both enzyme mechanisms point to virtually complete cleavage of the bond to the leaving group. This is borne out in my calculated TS structures.

My TS structure for the flu mechanism is incompatible with the experimental results. It should be noted that the experimental results were interpreted as suggesting that there is some degree of acid catalysis being applied to the leaving group oxygen. My model takes no explicit account of any interaction of this nature. Also, as previously mentioned, the experimental work was done on NA from a different subtype of flu. In view of this, I would suggest that comparison of calculated and experimental KIE values for this simulation is inappropriate. However, the postulated mechanism[1] which was tested cannot be ruled out. The calculated TS structure corresponds well to the postulated mechanism: there is a central moiety, which is zwitterionic in nature, that has an electron deficient anomeric carbon and has effectively zero bond order to the leaving group and a bond order of around 0.02 to the incoming water. Given that any number of TS structures could have been found using the PES method, it is encouraging that the one that did occur has some resemblance to one postulated from experimental evidence. There are of course other potential TS structures on the surface which could not be characterized, and these need to be

investigated; nevertheless, this is an encouraging beginning.

The calculated TS structure and KIEs for the *S. typhimurium* model are quite respectable. The imbalance in the pro R and pro S KIE values compared to experiment can almost be eradicated by the experimental error in KIE measurements, as can the ^{18}O leaving group effect. This, together with the observations made in section 10.3.1 about the steric effects of the substrate conformation in the active site, gives some degree of confidence that the model is performing satisfactorily. The TS structure is similar to that found for the solution simulation, which raises some interesting possibilities which are discussed in the next chapter. The bond order to the leaving group is about 0.02, and to the incoming water about 0.01. This is reminiscent of the structures found for the AMP hydrolysis. This TS structure is in line with the interpretation of the experimental results.

The PES for the ESR simulation suggests that there is a path which leads to loss of p-nitrophenolate. However, there appears to be no easy way for the water to approach sufficiently close to the anomeric carbon to continue the reaction. Some exploratory points beyond the pictured surface show that the high energy ridge which prevents the approach of water continues for some distance along the leaving group axis. This is in line with my earlier comments about the structure, in which the leaving group would act to prevent water access to the reaction centre. The potential TS structure identified on the surface could not be characterized with sufficient confidence to be used in KIE calculations. It appears from the surface that reaction could not proceed as for the flu enzyme, and is unlikely to occur because of the large energy barrier. My hypothesis about the protecting effect of the leaving group remains to be tested experimentally, but is not ruled out from these observations.

10.7 References

1. M. N. Janakiraman, C. L. White, W. G. Laver, G. M. Air & M. Luo, *Biochemistry*,, 1994, **33**, 8172

2. X. Guo, W. G. Laver, E. Vimr & M. L. Sinnott, *J. Am. Chem. Soc.*, 1994, **116**, 5572

3. W. P. Burmeister, B. Henrissat, C. Bosso, S. Cusack & R. W. H. Ruigrok, *Structure*, 1993, **1**, 19

4. S. J. Crennel, E. F. Garman, W. G. Laver, E. R. Vimr & G. L. Taylor, *Proc. Nat. Acad. Sci. USA*, 1993, **90**, 9852

5. C. L. Brooks & M. Karplus, *J. Mol. Biol.*, 1989, **208**, 159

6. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaninathan & M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187

7. B. R. Brooks, A. Brunger & M. Karplus, *Biopolymers*, 1985, **24**, 843

8. S. Rosenberg & J. F. Kirsch, *Biochemistry*, 1981, **20**, 3196

CHAPTER 11.
PROPOSALS

"Would you tell me, please, which way I ought to go from here?" said Alice. "That depends a good deal on where you want to get to", said the Cat. "I don't much care where," said Alice. "Then it doesn't matter which way you go", said the Cat. "As long as I get somewhere", Alice added as an explanation. "Oh, you're sure to do that", said the Cat, "if you only walk long enough."

- Lewis Carroll (C. L. Dodgson)

11.1 Introduction

This final chapter consists of a short discussion about the major findings of the previous two chapters, and provides some ideas for future work which could usefully be done to continue the investigation of these important and interesting reactions of neuraminic acid derivatives. The current limitations and scope for improvements to CHARMM are also mentioned. The other aim of this thesis, the provision of good models for use in calculation of isotope effects, has been covered in Chapter 5.

11.2 Comparison of Enzymic and non-Enzymic Hydrolysis of PNA

The results from the simulations done to date are compatible with the idea that the mechanisms of the enzymic and non-enzymic hydrolysis of PNA are fundamentally different. This can be seen from comparison of the PESs for simulations done from the same starting conformation for PNA for each environment. For example, the potential TS structures for the PSI and ESI reactions (Figures 9.10c and 10.4c respectively) are in different positions along the defining reaction coordinates. Since the ESI PES has been used to refine and characterize a feasible, if approximate, TS structure, it can be said that the interpolated PES must reflect at least to an approximate extent the true PES. The PSI surface has been constructed using the same assumptions about the mechanism, yet the potential TS structures are in significantly different areas of the surface. There is no obvious reason why even approximate TS structures could not be located for the solution simulation, except to say again that the model used was inappropriate for the reaction. The interpretation of the experimental results, that the carboxylate group must participate nucleophilically, still remains to be tested, but from the simulation results the

solution reaction does not appear to proceed *via* a TS structure similar to the equivalent enzyme reaction, as seen from the interpolated PESs.

The two enzyme reactions both appear to have a TS structure which does possess a highly charged anomeric carbon. Despite this, the mechanisms, and the TSs themselves, are different. As noted in Chapter 10, the flu NA TS is a flattened half-chair zwitterionic structure, whereas the *S. typhimurium* NA TS is a near perfect $^2C_5$ chair with a highly charged anomeric carbon. Observations from the enzyme-substrate models suggested that the difference could be explained by the conformation of the bound substrate. Comparison of the EFR and ESR surfaces (Figures 10.4a and 10.4b) provides graphic evidence to support this: although the ESI surface does appear to have a path similar to the one through which the EFR reaction proceeds, it is not possible for the water molecule to approach close enough to the reaction centre for completion of the reaction until the leaving group has moved a long distance. It would seem that the time this takes is sufficient for reaction with water from the opposite side of the ring to occur, giving inversion of configuration for the *S. typhimurium* NA. It has not been possible to provide any reasons as to why the flu NA requires reaction from the solution-disfavoured boat conformation, especially as this is probably the cause of the much lower turnover rate for this NA when compared with *S. typhimurium* NA.

On the assumption that the solution reaction does involve carboxylate participation at physiological pH, it is interesting to see how the two NAs adopt separate approaches to increase the reaction rate. The flu NA alters the starting conformation but the *S. typhimurium* NA does not. In solution, the acid-catalyzed hydrolysis of the neutral PNA cannot involve carboxylate participation, yet the reaction still proceeds *via* a flattened $^2C_5$ conformation and is an order of magnitude faster than the spontaneous reaction of the anion. It

could be that the *S. typhimurium* NA is at least partially adopting this strategy, eliminating carboxylate participation by binding rather than protonation of the carboxylate group. The active site would then serve both to stabilize a TS structure similar to the most favoured solution process, and provide access for water to attack on the β-face of the sugar ring to produce the favoured β-anomer product. This provides a great incentive for rapid turnover of the substrate.

The flu NA strategy is to induce conformational change so that the formation of an α-lactonic structure is even less favourable than it already was. It also binds the carboxylate group for good measure. The boat conformation actually makes the β-face of the sugar ring more "open", but the shape of the active site means that the surrounding water molecules cannot take advantage of this. The flu virus may not actually require a rapid turnover of the substrate, since the NA function is apparently to ensure that fresh virus does not reinfect the same cell. If the NA was too efficient it might be detrimental to the infection process, possibly by cleaving substrate before initial infection could occur, since the sialic acid residues are also required to anchor the virus to the cell prior to infection.

11.3 Proposals for Further Work on PNA Hydrolysis

There are a number of unanswered questions that could be addressed by further computer simulation of the hydrolysis of PNA. There is also some experimental work which could be done to provide some missing KIE values for comparison with calculated ones.

11.3.1 PNA Hydrolysis in Water

The outstanding question for the solution hydrolysis is the role of the carboxylate group. The preliminary AM1 *in vacuo* modelling, mentioned in

Chapter 9, has shown that there is some indication that the carboxylate motion does have an effect on the bond to the leaving group. There is an obvious need to incorporate the carboxylate motion explicitly into modelling of the reaction. The experimental work suggests that the intramolecular nucleophilic attack of the carboxylate group would probably be the rate-determining step of the reaction, given the high strain energy that even a partially-formed α-lactonic structure would have.

The *in vacuo* simulation of this intramolecular nucleophilic attack led to cleavage of the glycosidic ring C-O bond. The solution reaction may be quite different and could be modelled effectively using CHARMM. Defining a reaction coordinate for the carboxylate group, and one for the leaving group, a PES could be obtained. Several water molecules close to the reaction centre could be included in the QM system, instead of one explicit water molecule as in previous simulations, to provide an idea of the motion of water as the leaving group departs.

An alternative modelling technique could be used which makes use of CHARMM's unique QM/MM force field in MD simulations. It is possible to perform an activated dynamics simulation. This would involve the definition of a reaction coordinate, in this case the distance of one of the carboxylate oxygens from the anomeric carbon. An umbrella potential would then be applied to this coordinate, which enables the MD simulation to surmount small energy barriers which it would normally do only infrequently. This allows the MD simulation to progress along a reaction coordinate, and gives some information about the structural changes as the carboxylate group approaches the anomeric carbon. The advantage that this would give over the ordinary constrained reaction coordinate approach is that one could obtain a sample of conformations near the top of any energy barrier found which would not be artificially constrained. The disadvantage is the large amount of computer resource which would be required.

11.3.2 PNA Hydrolysis by NA

There are a number of issues that arise from the simulations done so far of the enzymic hydrolysis of PNA. One of my suggestions was that water was prevented from reaching the active site of the flu enzyme, but not the *S. typhimurium* enzyme. This suggestion can be tested by using a MD simulation to explore the movement of water around the active site. There may be some flexibility in the static structure from which my observation came, and this may allow a sufficiently large space to appear between the substrate and enzyme to allow water through. [a]

My simulations showed that for the *S. typhimurium* enzyme, water cannot approach the anomeric carbon until the leaving group has moved some distance from the reaction centre. Examination of the movement of the leaving group away from the active site could establish the distance required before water can approach to give the hypothetical reaction leading to retention of configuration. This can initially be done using molecular graphics alone, without the need for large amounts of computer time. Once an approximate distance has been established, a simulation of the imaginary reaction could be done. This may show whether or not retention of configuration is possible by the enzyme from the chair conformation. The outcome of this simulation may help to explain why the flu enzyme requires reaction through the boat conformation.

One useful calculation that could be done for both enzymes would be to remove the leaving group from the system and see if the remaining part of the substrate is sufficiently stabilized by the active site to allow it a real existence. Experimental evidence from both X-ray and KIE data suggests that for the flu B NA this should certainly be the case. The evidence for the same outcome on the

---

[a] It has since been established that the S. typhimurium enzyme does indeed have a solvent channel which allows water access to the b-face of the sugar ring
S. Crennell, E. Garman, G. Laver. E. Vimr & G. Taylor, *Structure*, 1994, 2, 535

S. typhimurium enzyme is less certain, although the possibility of a highly charged TS makes it likely that the zwitterionic structure would be stable. This work would show how similar the enzyme "machinery" is for TS stabilization within the two mechanisms, since at this stage the two reactions would be identical, the conformational differences having disappeared with the leaving group. Does one enzyme require support from the water surrounding the active site? This is one question that may be answered.

The chair to boat conformational change on the enzyme is another important aspect of the flu reaction which needs to be addressed if the mechanism is to be fully established. Does the change occur after initial binding, as has been suggested by Guo and coworkers; [b] does the enzyme selectively bind the low proportion of molecules in the boat conformation in solution; or does the conformation change because of a requirement of the binding process?

The conformational change on the enzyme could be explored using the same reaction coordinate that was used in section 9.3 for the *in vacuo* simulation. This would provide information on the energetics and protein motions required. This is another instance where the activated dynamics approach of CHARMM could also be used to good effect.

One possibility that can be seen from graphical manipulation of the structure of the two NAs under consideration is the difference in the volumes of the active site. The three Arg residues that stabilize the carboxylate group must have a different spatial arrangement and relationship to the active site if they are to interact with the carboxylate group in two different conformations. This must mean that the active sites are a different shape. Whilst I do not have access to docking software, nor a means of visualizing the volume of the active sites, there does appear to be a difference in the way that the substrate would have to

---

[b] X. Guo, W. G. Laver, E. Vimr & M. L. Sinnott, *J. Am. Chem. Soc.*, 1994, **116**, 5572

slot into the active site. This requires that the substrate on the flu enzyme has to approach in such a way that the carboxylate group has to successively pass within hydrogen bonding distance of three Arg residues, whereas the docking of the substrate on the *S. typhimurium* enzyme does not. If the H-bonding interaction were sufficiently strong at each stage, could this not distort the ring structure sufficiently to flip it into a boat conformation as part of the binding process?

11.4 Limitations and Potential for the CHARMM Software

The CHARMM software has proved very useful for the modelling of small molecules in solution. In particular, the success of the AMP hydrolysis simulation showed that the combined QM/MM force field has great utility in the modelling of reaction mechanisms. However, it was less successful for the modelling of enzyme reactions. Of course, enzyme systems are more complex than their solution counterparts, but in principle I like to think that it is only the number and variety of atoms that make the difference!

There was evidence in the enzyme simulations of some problems that appear to be related to the calculation of the QM/MM components of the energy and their derivatives. This was manifested in the inability of the software to properly optimize the p-nitrophenyl ring. The gradients on the atoms within the ring were consistently high, and the ring was substantially non-planar. In addition, the nitro group was generally well-removed from the approximate plane of the aromatic ring. These observations were only found in QM/MM simulations. MM-only simulations showed a normal structure for the p-nitrophenyl group, and the ring distortions were less severe in the solution simulations. The good behaviour of the MM-only simulations suggests that the problem is not due to poor choice of parameters for the atoms. This deficiency

requires investigation if there is to be full confidence in future enzyme simulations using the QM/MM method.

A major deficiency of CHARMM is the lack of a TS searching algorithm. There is a reaction path following algorithm, but it is unable to cope with the many degrees of freedom in the atoms not immediately involved at the reaction centre. This is a general problem for large systems. It may be possible to perform the same trick that I used for the calculation of force constants, namely take only the energy components of a selected set of atoms and use these to drive a searching algorithm. The problem with this approach is that the effects on the rest of the system are not easily predictable, and particularly with flexible enzymes could lead to misleading structures.

One of the other major problems with CHARMM is the optimization process for large QM/MM systems. Although the QM and MM gradients are coupled, the minimization is done on the whole system. For large systems, the rms gradient as a criterion for the end of the optimization process is not always a good choice, especially since in CHARMM the QM gradients are scaled. This sometimes leads to instances where the QM system is left with quite large gradients despite the target rms gradient for optimization being satisfied. This generally occurs as the result of a large but uninteresting gradient at the extreme of a system being reduced at the expense of an interesting gradient near the reaction centre. What is required is an optimization algorithm that can selectively drive a set of atoms to an optimal geometry, allowing gradients at the extremes of a large system to be ignored, or perhaps to be weighted as less important.

With the additional algorithms mentioned above, CHARMM would be a formidable piece of software for the investigation of enzyme mechanism.

11.5 That's All Folks!

The above recommendations for further work on the hydrolysis of PNA, and the suggestions for additions to the CHARMM program serve to show that the work presented here is merely the first part of what could be a substantial body of work. There are obviously two strands to the work, and they do not easily sit together.

The QM aspects of the work are clearly hindered by the lack of optimization and TS searching procedures appropriate to a combined QM/MM force field which, as noted above, have requirements not usually available in these types of algorithm. The challenge is there for a software engineer to help create a uniquely tailored product for the simulation of enzyme mechanisms.

There are many elements of the proposed simulation work that do not rely on the all-singing, all-dancing software that would be the ideal for this type of work, and these provide a clear path forward to the deeper understanding of the enzyme mechanisms which have been explored in this thesis. The undertaking of some computationally intensive QM/MM dynamics simulations is one aspect that has not been feasible with the available resources. It might make a realistic project if the code was suitable for running on parallel processing machines, and this is one possibility which could be investigated. Most of the other ideas in the previous section could be explored using suitable software on workstations, and could be started tomorrow.

Anyone interested?

APPENDIX I.
A LOCALLY MODIFIED VERSION OF AMSOL

APPENDIX II.
ADDITION OF THE FORCE CONSTANT CALCULATION FOR QUANTUM MECHANICAL ATOMS TO THE CHARMM PROGRAM

"With my two algorithms, one can solve all problems - without error, if God will!"

- Al-Khorezmi

I.1 Introduction

A recent modification to the AMPAC program has been produced which introduces further parameters to account for free energies of solvation in water.[1] The program is called AMSOL. The incorporation of solvation free energy has been done by including terms for solute-solvent polarization (P),and a joint term for cavity creation in the solvent (C) and dispersion energy (D),

$$G_{SOLV} = E_{EN} + G_P + G_{CD}$$

where E_{EN} is the ground state electronic energy and nuclear repulsion for the solute, as normally calculated by AMPAC. The cavity and dispersion terms are derived from surface tension and solvent accessible solute surface area, and the polarization term is from the generalized Born model. In the grand tradition of the semiempirical methods, parameters have been derived in order to speed up calculation, and a test set of 141 neutral molecules and 27 ions was used for this purpose.

I.2 MOPAC Modification to incorporate the Solvation Calculation

The calculations for this study were carried out using a modified version of the program MOPAC version 6.0.[2] The modifications are due to the work of Cramer and Truhlar, who have introduced a new parameterization for aqueous solvation free energies of molecules and ions, as described in the previous section. Their modification to AMPAC is called AMSOL.[3] The supplied version of AMSOL would not run on our computers so I added the relevant new

parts of the program to our current version of MOPAC to produce a program called MOSOL.[4] This program has all the functionality of MOPAC, and in addition by using the extra keywords available it is possible to calculate free energies of solvation, and hence simulate the effect of aqueous solvation on chemical processes. The solvation model is only implemented for the AM1 Hamiltonian. There are currently two solvation models available, AM1-SM1 and AM1-SM1a. The former is independent of chemical environment, taking account of the atom type only. The second model distinguishes between environments, for example an aldehydic oxygen is distinct from an alcoholic oxygen. The two models are invoked by use of the two new keywords AQUO and ENVAQ. It should be borne in mind that the solvation model cannot take into account specific interactions of the solute with water molecules. If specific interactions need to be considered as part of a mechanism, water molecules must be included explicitly in the solute model.

I.3 The $CH_3Cl + Cl^-$ Exchange Reaction: A Test Case

To assess the good working order of the MOSOL program, a test system was used for which there are ample data available in the literature: the symmetric CH_3Cl + Cl- exchange reaction. The reaction was simulated in the gas phase and also done using the new aqueous free energy calculation. The approach taken was to investigate the reaction by using the carbon to chloride ion distance as the reaction coordinate, with a constraint on the chloride ion-carbon-chlorine angle of 180 degrees. The same parameters were kept for both investigations, the only difference being in the use of a keyword to invoke the aqueous phase calculation. Since it is not possible in general to have knowledge of the

environment of exotic species such as transition states the solvation calculations were done using the AM1-SM1 model, using the AQUO keyword. Also, all gas phase calculations have been done using AM1 to enable a comparison to be made between the gas phase and aqueous phase calculations.

I.3.1 Results and Discussion

The results of the calculations are shown in Figure A1. The gas phase curve represents internal energy change at 298K, whereas the solvation energy profile is a "free energy" curve at 298K, with the entropy of the solute species neglected. The major difference between the two is the lack of a double well profile in the solution calculation. There is a step in the calculated free energy in going from a carbon to chlorine distance of 3.6Å to 3.8Å, which is about 1.5Å along the reaction coordinate. This is due in the MOSOL calculation to a significant increase in the accessible surface area of the leaving chloride ion. In the formalism of the energy calculation this causes an increase in the solvation free energy by its contribution to the solute-solvent dispersion interactions. It should be noted that the effective Born radius at this geometry goes from 4.1Å to 4.24Å, as noted from the output provided by the program. This is similar to the distance at which it would be expected that a water molecule could fit between the two parts of the reacting system by consideration of the ionic and covalent radii of Cl^-, O and C, so it seems that in this case at least the model is providing a realistic picture of the mechanics of the reaction in solution. The calculated results for the solution reaction are consistent with the view that S_N2 energy profiles are approximately unimodal in solution. The large increase in activation energy over the gas profile correctly reflects the large difference in rate for the process in solution. Table A1 gives some calculations for the same system from a selection of theoretical calculations. It should be noted that in reference 5 there

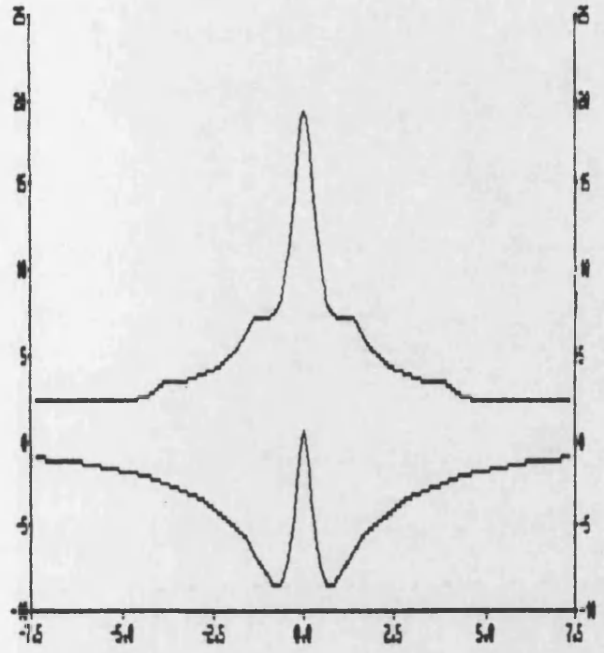


Figure A1 : Relative energy along the reaction coordinate for the CH3Cl + Cl- exchange reaction in the gas phase (lower curve) and in aqueous solution (upper curve).

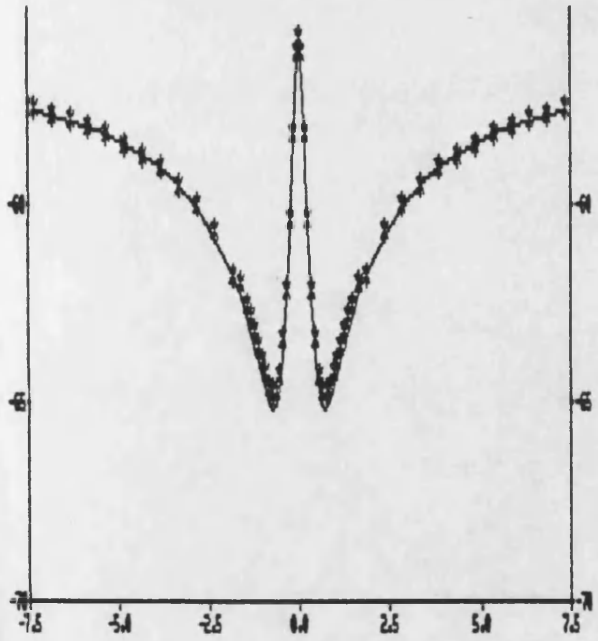


Figure A2 : Heat of Formation for the solute along the Reaction Coordinate for the CH3Cl + Cl- exchange reaction in the gas phase (-) and in aqueous solution(*).

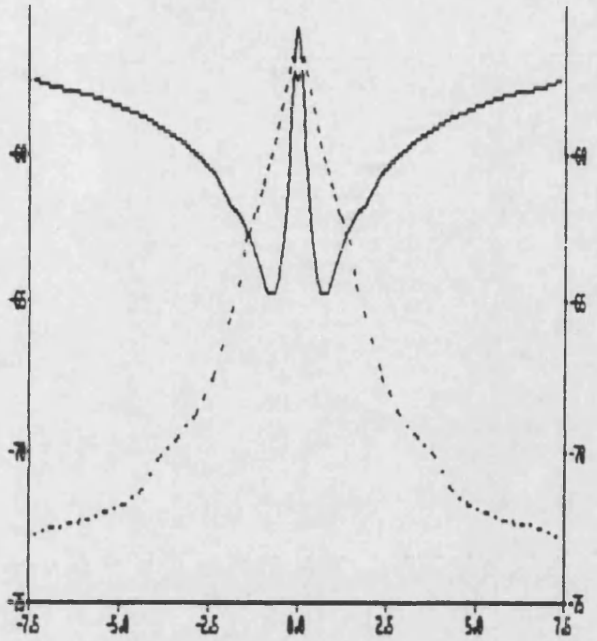


Figure A3 : Heat of Formation (—) and Solvation Free Energy (- -) along the Reaction Coordinate for the CH3Cl + Cl- exchange reaction in aqueous solution.

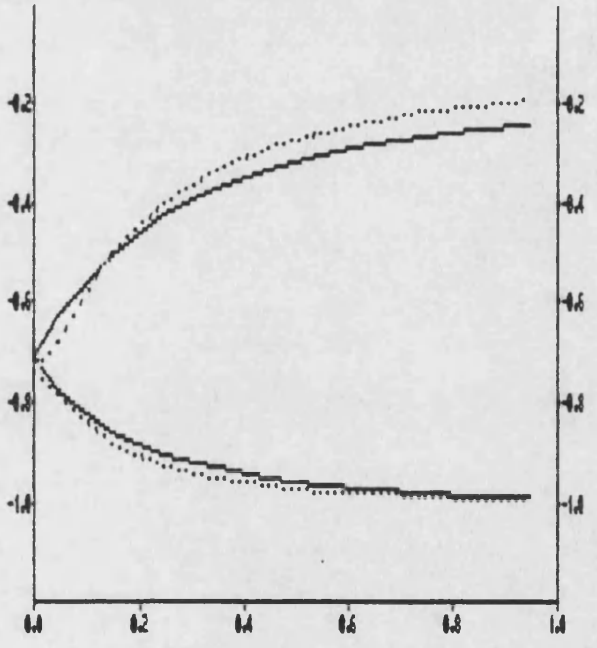


Figure A4 : Charge transfer along the reaction coordinate for the CH3Cl + Cl- exchange reaction in the gas phase (-) and in aqueous solution (...). Lower curves are the attacking Cl-, upper curves the Cl of CH3Cl.

Table A1. Energies of $CH_3Cl + Cl^-$ Reaction (kcal mol^{-1})

Property	Exp'tal	Calc1[a]	Calc2[b]	Calc3[c]	This work
$\Delta H^{\ddagger}(g)$	10-14[c]	13.9[d]	9.1[e]	---	9.1
$\Delta G^{\ddagger}(aq)$	26.6[a]	28 ± 6	27.0	23.0[f]	19.2
$\Delta Gsolv^{g}$	-73.6[h]	---	---	---	-75.9

a: Ref 6 b: Ref 7 c: Ref 5 d: 6-31G* e: AM1 f: Using only the first coordination sphere of the solute to eliminate an artificial minimum of the initial calculation g: Solvation free energy of separated reactants Cl^- and CH_3Cl h: From data in Ref 8

appears to be an artefact in the calculation, due to an unexpected minimum of the ion-dipole complex on the potential surface. Without a correction to the calculation, $\Delta G^{\ddagger}$ for the reaction comes out as 18.3 kcal mol^{-1} which is close to the value found in the present study. It can be seen that the MOSOL calculation is in reasonable agreement with calculations performed by hybrid ab initio quantum mechanics and molecular mechanical methods,[5] FEP techniques[7] and a Monte Carlo simulation.[6] Although the predicted $\Delta G^{\ddagger}$ of solution falls short of the experimental value of 26.6 kcal mol^{-1}, the MOSOL method clearly predicts the large increase in barrier height in the solution reaction.

Figure A2 shows the gas phase energy profile together with the solute energy without solvation free energy for the solution reaction. As expected, the two profiles are virtually identical, showing that there is little geometric difference in the course of the two reactions along the reaction coordinate chosen. Figure A3 shows the separate contributions of the heat of formation and

large increase in solvation energy at 1.5Å along the reaction coordinate as discussed above can be clearly seen. Figure A4 shows the extent of charge transfer along the reaction coordinate. As expected for a solvated reaction, a less symmetric transfer is stabilised, the incoming chloride ion losing its charge more slowly than in the gas phase.

I.4 Modification of the Force Constant Calculation

One aspect of MOSOL that required attention was the method of calculating force constants. The standard method displaces each atom in turn along the three cartesian axes and calculates the forces from the resultant energy gradients between the energies of the perturbed geometries. Use of the AQUO option introduces an extra calculation into the energy evaluation, which involves calculation of a so-called Born radius for each atom via the exposed surface area. This is a computationally intensive step which is repeated for each displacement of the atoms in the force calculation. To shorten the time required for a force calculation, two modifications were assessed. The first was to recalculate the exposed surface area of a given atom A only if the atom being displaced was within a distance equal to the sum of the radii for it and atom A. If the displaced atom was not within this distance, the exposed area of atom A would not be affected. Figure A5 illustrates these two situations. The second modification was to calculate the exposed surface area only once, the justification being that the atom displacements used in the force calculation are very small in comparison to the atomic radii, and so should not affect the surface area calculation significantly. Both modifications increased the speed of the calculation, without significant loss of accuracy except for a difference in the negative force constant of transition states, as illustrated in Table A2. The first modification gave a four-fold increase in speed, the second gave a forty-fold

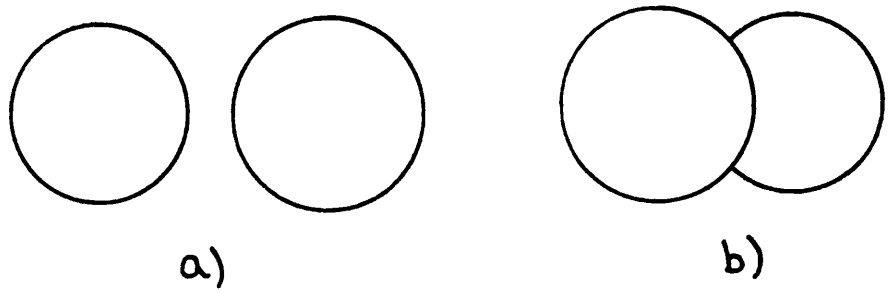


Figure A5. The two atoms in a) do not overlap, so the exposed surface area of atom A is not affected by atom D. In b) the atoms do overlap, so the exposed surface area of A needs to be recalculated.

Table 2. Transition State Vibrational Frequencies for CH3Cl + Cl- using the Three Versions of Force Constant Calculation

Version 1[a]	Version 2[b]	Version 3[c]
-425	-433	-604
217	217	217
217	217	217
258	260	258
985	985	985
985	985	985
1150	1151	1134
1323	1323	1322
1323	1323	1322
3174	3174	3172
3180	3180	3180
3180	3180	3180

a: Original program b: Calculate exposed surface area only when necessary c: Calculate exposed surface area once only.

Table A3. Kinetic Isotope Effects (k_{H3}/k_{D3}) at 298K for the CH_3Cl + Cl^- System Using Frequencies Calculated by the Three Force Constant Calculation Methods[a]

	G.S. orig	G.S. v1	G.S. v2
T.S. orig	0.949	0.949	0.949
T.S. v1	0.943	0.943	0.943
T.S v2	0.966	0.966	0.966

a: orig is the original calculation, v1 the first modification and v2 the second modification to the force constant calculation. The methods were cross-checked with each other to see where errors may occur.

increase, using both the CH_3Cl + Cl^- system and some test geometries supplied with the original code. Table A3 shows the effect on kinetic isotope effect (KIE) calculations of the error in calculation of the transition frequency for the test system. It can be seen that there is a 1.8% error in the calculated KIE when the second modification is used to calculate force constants for the transition state, but there appears to be no significant loss of accuracy when it is used for the ground state calculation. In each case, it seems that there is only an effect on the calculated KIE when the faster methods are used to calculate frequencies for the TS: use of the fast methods on the ground state only does not appear to compromise the calculation. Given this limited evidence, it has been decided that the second modification be implemented in the version of MOSOL used for the calculations presented here, since the large saving in time is considered to outweigh the imprecision in the calculation. The fast calculation is invoked by the use of a new keyword, AQFAST. It is not possible from these results to place likely limits on the accuracy of KIE calculations for other systems, but this will be borne in mind when results are analyzed.

I.5 Conclusions and Recommendations for Use

The methodology of Cramer and Truhlar, implemented here in the MOSOL program, can be a useful method for the investigation of reactions in aqueous solution. As discussed above, it appears to describe the major differences between the gas and aqueous phase chemistry of the test reaction.

It should be noted that care was required to ensure that the gradient norms of many of the structures were kept within reasonable limits and this often meant extra computational effort. Best results were obtained when structures were optimized using the keywords EF and GNORM=5.0, then optimized again using GNORM=0.05 without the EF option. This is the recommended procedure. The problem of gradients is acknowledged by the original authors in the literature accompanying the program from QCPE. Given that the test system is relatively small, the gradient problem may become significantly worse for larger systems.

It is also recommended that the AQFAST keyword be used for routine calculation of force constants for characterization purposes to reduce computing time. The full calculation need only be done when the resulting force constants are to be used for isotope effect calculations.

For calculations that are going to involve transition states, the AQUO keyword should be used throughout since the precise environment of atoms will be unknown.

I.6 References

1. C. J. Cramer & D. G. Truhlar, *J. Am. Chem. Soc.*, 1991, **113**, 8305

2. J. J. P. Stewart, *J. Comp. Aided Mol. Design*, 1990, **4**, 1 (QCPE program 455)

3. AMSOL v1, QCPE program 606

4. Not to be confused with the solid state version of MOPAC!

5. U. C. Singh & P. Kollman, *J. Comp. Chem.*, 1986, **7**, 718

6. J. Chandrasekhar, S. F. Smith & W. Jorgensen, *J. Am. Chem. Soc.*, 1985, **107**, 154

7. P. A. Bash, M. J. Field & M. Karplus, *J. Am. Chem. Soc.*, 1987, **109**, 8092

8. D. J. McLennan, *Aust. J. Chem.*, 1978, **31**, 1897

II.1 Introduction

To allow the calculation of vibrational frequencies for the hybrid QM/MM systems, it was necessary to introduce appropriate code into CHARMM, since the existing code did not support their calculation in hybrid systems. The following Fortran code indicates the route that I took to achieve this. It is essentially the relevant bits of code from MOPAC v6.0, modified to take the first derivatives of the energy that CHARMM calculates. I kept the .RES output file the same so that it can be used in the same way as a standard .RES MOPAC file for a force constant calculation. It was also necessary to adjust the CHARMM quantum select command so that it would accept two atom specifications: this was done using a CHARMM function and is sufficiently trivial that it really does not require documentation.

II.2 The modified MOPAC code

```
      SUBROUTINE QMSECD
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
C
C***** This code has been brutally rippped and butchered from
C***** the original MOPAC routines that calculate the second
C***** derivatives of the energy with respect to cartesian
C***** coordinates. The routine identifies the QM and link atoms
C***** that are in the system and uses the CHARMM forces and
C***** first derivatives on them as input to the MOPAC code to produce
C***** the second derivatives. The force matrix is printed to
C***** channel 6 as in the MOPAC .out file and the second
C***** derivatives are written to channel 27, so you must open a
C***** file for unformatted write to channel 27 before invoking this
C***** code. A vibrational analysis is not done because the system will
C***** have 6 rotational and translational modes mixed in. The
C***** output is designed to be compatible with the CAMVIB program
C***** or similar for later analysis.
C***** information in channel 27 is the same as for a MOPAC .res
C***** file when the keywords FORCE and ISOTOPE are used, except
C***** that only the coordinates and second derivatives appear
C***** after the first three variables TIME,IPT,REFH. These variables
C***** are retained for compatiblity. TIME is the time taken to
C***** calculate the second derivatives, IPT is the famous (3*n*((3*n)+1)/2)
C***** and REFH is the potential energy of the whole system.
```

```
c*****
c***** KNOWN PECULIARLITIES
c***** If you write the standard output to file, it does not always
c***** record the end of the output. I am not sure why! It does not
c***** affect the data written to channel 27, or it hasn't so far
c***** in any of the tests that I have done.
c***** jab Feb 1994
c
c***** The ## are a preprocessor directive in the CHARMM cmpilation
c***** and not some weird Fortran command!
c
##INCLUDE '~/charmm_22/dimens.fcm'
##INCLUDE '~/charmm_22/quantm.fcm'
##INCLUDE '~/charmm_22/psf.fcm'
##INCLUDE '~/charmm_22/coord.fcm'
##INCLUDE '~/charmm_22/deriv.fcm'
      CHARACTER*2 elemnt
      integer LABELS
      common /mostuf/ ELEMNT(107),LABELS(maxqm1)
      REAL*8 qmxyz
      common /qmforc/ natqml,natot,qmxyz(3,maxqm1),nqmpt(maxqm1)
c
c*****DEBUG ONLY
c     character*4 at,rid,ren,sid
c*****DEBUG ONLY
c
c
c     Now the fun starts
c
c*****This is code hacked out from MOPAC to allow the
c*****force calculation. For this to work the qm and link atom
c*****coordinates are identified and displaced in turn within the
c*****normal mopac force routine with some modifications.
c***** 1. The system is not oriented before force. The rotations
c*****    and translations are removed later by CAMVIB.
c***** 2. When COMPFG is called, the CHARMM routine GETE is called
c*****    for energy calculations and the resultant potential
c*****    energy is used for energy.
c***** 3. The CHARMM first derivatives are used instead of calling DCART
c*****    in the DERIV routine. This means that we calculate 2nd
c*****    derivatives directly from CHARMM cartesian 1st derivatives.
c*****The CHARMM energy and first derivatives are thus used directly in
c*****the construction of the force matrix in the FMAT routine, and
c*****the vibrations are done as normal using these. I don't know how
c*****valid this approach is but it's all we have at the moment
c*****JAB Dec 1993
c
c***** MOPAC CODE NOTICE
C
C       Notice of Public Domain nature of MOPAC
C
C     'This computer program is a work of the United States
C      Government and as such is not subject to protection by
C      copyright (17 U.S.C. # 105.)  Any person who fraudulently
C      places a copyright notice or does any other act contrary
C      to the provisions of 17 U.S. Code 506(c) shall be subject
```

```
C      to the penalties provided therein.  This notice shall not
C      be altered or removed from this software and is to be on
C      all reproductions.'
C
C
c
c    element 99 is to be a link atom,symbol QQ, mass 1.000
c
     DATA (ELEMNT(I),I=1,107)/'H','HE',
    1 'LI','BE','B','C','N','O','F','NE',
    2 'NA','MG','AL','SI','P','S','CL','AR',
    3 'K','CA','SC','TI','V','CR','MN','FE','CO','NI','CU',
    4 'ZN','GA','GE','AS','SE','BR','KR',
    5 'RB','SR','Y','ZR','NB','MO','TC','RU','RH','PD','AG',
    6 'CD','IN','SN','SB','TE','I','XE',
    7 'CS','BA','LA','CE','PR','ND','PM','SM','EU','GD','TB','DY',
    8 'HO','ER','TM','YB','LU','HF','TA','W','RE','OS','IR','PT',
    9 'AU','HG','TL','PB','BI','PO','AT','RN',
    1 'FR','RA','AC','TH','PA','U','NP','PU','AM','CM','BK','CF','QQ',
    2 'FM','MD','CB','++','+','--','-','TV'/
c
c
     WRITE(6,'(20X,"                                        ")')
     WRITE(6,'(20X,"                                        ")')
     WRITE(6,'(20X,"****************************************")')
     WRITE(6,'(20X,"****************************************")')
     WRITE(6,'(20X,"          QMFORC CALLED.                ")')
     WRITE(6,'(20X,"This will invoke a force calculation on ")')
     WRITE(6,'(20X,"the QM and link atoms of the system. The")')
     WRITE(6,'(20X,"calculation is only experimental so far.")')
     WRITE(6,'(20X,"****************************************")')
     WRITE(6,'(20X,"****************************************")')
     WRITE(6,'(20X,"                                        ")')
     WRITE(6,'(20X,"                                        ")')
     degree=57.29578
     kone=1
     natot=natom
c
c The coords of the qm system which is to have force constants calculated
c are obtained here. It is not necessary for all the qm atoms to be in
c the calculation. If you give two atom selections to the quantum select
c command, the first set is the qm system, the second the set of qm atoms
c for which force constants are calculated.
c NQMPT points to the atoms in their order in the whole system, and the
c coordinates are put into qmxyz. labels array is filled with the atomic
c number of each qm atom, remembering that a link atom is labelled as
zero
c atomic number in charmm, but we want it to be a hydrogen.
c
     k=1
     do 22222 j=1,natom
                                                  if (iqmfc(j) .eq. 1) then
                                  labels(k)=max(1,qatlab(j))
                                        nqmpt(k)=j
                                      qmxyz(1,k)=x(j)
                                      qmxyz(2,k)=y(j)
```

```
                                qmxyz(3,k)=z(j)
c
c****** DEBUG ONLY
c     WRITE(6,'(/9X,"QM ATOM NUMBER",i5)') k
c     WRITE(6,'(/,4X,"NO.",5X,"ATOM IDENTITY",14X,"X",
c    19X,"Y",9X,"Z",/)')
c        call atomid(nqmpt(k),sid,rid,ren,at)
c        WRITE(6,'("XYZ",I6,3X,A4,2X,2A4,2X,A4,3X,3F10.4)')
c    1   j,at,rid,ren,sid,X(j),Y(j),Z(j)
c        WRITE(6,'("QMXYZ",I6,3X,A4,2X,2A4,2X,A4,3X,3F10.4)')
c    1   nqmpt(k),at,rid,ren,sid,qmxyz(1,k),qmxyz(2,k),qmxyz(3,k)
c*****DEBUG ONLY
c
          k=k+1
        endif
22222 continue
      natqml=k-1
      if ( natqml .eq. 0 ) then
      WRITE(6,'(20X,"                                        ")')
      WRITE(6,'(20X,"                                        ")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"    NO QUANTUM ATOMS ARE DEFINED.       ")')
      WRITE(6,'(20X,"Either you are not doing a calculation  ")')
      WRITE(6,'(20X,"which has QM atoms or the quantum code  ")')
      WRITE(6,'(20X,"is not compiled. Check your job and see ")')
      WRITE(6,'(20X,"your CHARMM manager for further help.   ")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"                                        ")')
      WRITE(6,'(20X,"                                        ")')
      return
      endif
c*****Now the preliminaries are over, call FORCE
c
      call force
      WRITE(6,'(20X,"                                        ")')
      WRITE(6,'(20X,"                                        ")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"   END OF QMFORC FORCE CALCULATION.     ")')
      WRITE(6,'(20X,"If you did not open a file on channel 27")')
      WRITE(6,'(20X,"for the force output it should appear on")')
      WRITE(6,'(20X,"the system default channel 27 File.     ")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"****************************************")')
      WRITE(6,'(20X,"                                        ")')
      WRITE(6,'(20X,"                                        ")')
      return
      END
      SUBROUTINE  FORSAV(TIME,IPT,
     aFMATRX,GRADS,COORD,NVAR,REFH)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      DIMENSION FMATRX(*), COORD(*), GRADS(*)
*************************************************************************
*
```

```
*  FORSAV SAVES THE SECOND DERIVATIVES OF THE QM ATOMS.
*
* ON INPUT TIME = TOTAL TIME ELAPSED SINCE THE START OF THE
CALCULATION.
*       IPT  = LINE OF FORCE MATRIX REACHED, THIS SHOULD ALWAYS
BE
*            3*number of qm atoms, so use it as a checkpoint
*      FMATRX = SECOND DERIVATIVES OF THE QM ATOMS
*      GRADS  = FIRST DERIVATIVES OF THE QM ATOMS
*      COORD  = CARTESIAN COORDINATES OF THE QM ATOMS
*************************************************************************
      IW=27
C
C     WRITE FORCE DATA
C
        WRITE(IW)TIME,IPT,REFH
        LINEAR=(NVAR*(NVAR+1))/2
        WRITE(IW)(COORD(I),I=1,NVAR)
        WRITE(IW)(FMATRX(I),I=1,LINEAR)
        WRITE(IW)(GRADS(I),I=1,NVAR)
                                        close(IW)
      if (ipt .ne. nvar) goto 10
      RETURN
   10 WRITE(6,'(10X,"Internal check indicates that the force ",
     1"calculation",/10X,"did not finish. There is no restart ",
     2"capability for this")')
      WRITE(6,'(10X,"option so you will have to redo the whole ",
     1"calculation. ",/10X,"Sorry. Use a longer cpu time for",
     2" your job.")')
      STOP
      RETURN
      END
      SUBROUTINE COMPFG(XPARAM,ESCF,DXYZ,IDSP,NOCALC)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
C*************************************************************************
C   Adapted to work with CHARMM from the MOPAC original jab Jan 1994
C
C   COMPFG CALCULATES (A) THE POTENTIAL ENERGY  OF THE
SYSTEM, AND
C                 (B) THE GRADIENTS
C
C   ON INPUT  XPARAM = ARRAY OF CARTESIAN COORDINATES
C           IDSP   = THE Ith COORDINATE WHICH IS BEING ADJUSTED IN
FMAT
C           NOCALC = false if energy and gradients are required
C
C   ON OUTPUT ESCF  = POTENTIAL ENERGY OF THE QM/MM SYSTEM
C           DXYZ = FIRST DERIVATIVES OF THE QM ATOMS
C
C
C*************************************************************************
c
c     Stuff that compfg needs to know for using charmm energy routine
c
##INCLUDE '~/charmm_22/dimens.fcm'
##INCLUDE '~/charmm_22/bases.fcm'
```

```
##INCLUDE '~/charmm_22/coord.fcm'
##INCLUDE '~/charmm_22/deriv.fcm'
##INCLUDE '~/charmm_22/quantm.fcm'
##INCLUDE '~/charmm_22/energy.fcm'
c
c*****DEBUG ONLY
c##INCLUDE '~/charmm_22/psf.fcm'
c     character*4 sid,rid,ren,at
c*****DEBUG ONLY
c
      REAL*8 qmxyz
      common /qmforc/ natqml,natot,qmxyz(3,maxqm1),nqmpt(maxqm1)
      dimension dxyz(maxqm1*3),xparam(maxqm1*3)
      logical nocalc
      if (idsp .ge. 1) then
c
c***** adjust the coordinate set of the whole system to reflect
c***** the displacement being made in FMAT for calculating
c***** second derivs. nqmat is the ith qm atom being adjusted
c***** nxyz = 1 for x-coord, 2 for y-coord, 0 for z-coord
c
        nqmat=int((idsp-1)/3)+1
        nxyz=mod(idsp,3)
        if (nxyz .eq. 1) x(nqmpt(nqmat))=xparam(idsp)
        if (nxyz .eq. 2) y(nqmpt(nqmat))=xparam(idsp)
        if (nxyz .eq. 0) z(nqmpt(nqmat))=xparam(idsp)
      endif
c
c****** DEBUG ONLY
c     WRITE(6,'(/9X,"ORIENTATION OF MOLECULE IN FORCE -DEBUG")
c    1')
c     write(6,'(/4x,"idsp",I4,"  nqmat",I4,"  nxyz",I4)')
c    1idsp,nqmat,nxyz
c     WRITE(6,'(/,4X,"NO.",5X,"ATOM IDENTITY",14X,"X",
c    19X,"Y",9X,"Z",/)')
c     DO 100 I=1,NATQML
c        call atomid(nqmpt(I),sid,rid,ren,at)
c        WRITE(6,'(I6,3X,A4,2X,2A4,2X,A4,3X,3F10.4)')
c    1   nqmpt(I),at,rid,ren,sid,X(nqmpt(I)),Y(nqmpt(I)),Z(nqmpt(I))
c 100 CONTINUE
c*****DEBUG ONLY
c
      if (nocalc) return
C
C DO A CALL TO THE ENERGY ROUTINE AND GET POTENTIAL ENERGY

C
      call energy(x,y,z,dx,dy,dz,bnbnd,bimag,0,0,.false.,1)
      escf=eprop(epot)
      call printe(6,eprop,eterm,'ENER','ENR',.true.,0,0.d0,0.d0)
c
c*****Put grads of qm atoms into dxyz
c
      ict=1
      do 3000 i=1,natqml
        dxyz(ict)=dx(nqmpt(i))
```

```
      dxyz(ict+1)=dy(nqmpt(i))
      dxyz(ict+2)=dz(nqmpt(i))
                                    ict=ict+3
 3000 continue
c
c*****DEBUG ONLY
c     WRITE(6,'(//10X,"CARTESIAN COORDs DERIVS - DEBUG",//4X,
c    1"NO.",7x,"ATOM",8X,"DX",8X,"DY",8X,"DZ",/)')
c     ll=-2
c     do 40 l=1,natqml
c       ll=ll+3
c       WRITE(6,'(I6,6X,a4,4X,3F10.4)') nqmpt(l),type(nqmpt(l)),
c    1   dxyz(ll),dxyz(ll+1),dxyz(ll+2)
c  40 continue
c*****DEBUG ONLY
c
      RETURN
      END
      SUBROUTINE VECPRN (A,NUMM)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      DIMENSION  A(*)
C*************************************************************************
C
C  VECPRN PRINTS A LOWER-HALF TRIANGLE OF A SQUARE MATRIX,
THE
C        LOWER-HALF TRIANGLE BEING STORED IN PACKED FORM IN
THE
C        ARRAY "A"
C
C ON INPUT:
C      A      = ARRAY TO BE PRINTED
C      NUMM   = SIZE OF ARRAY TO BE PRINTED
C(REF) NATQML = NUMBER OF ATOMS IN THE MOLECULE
C(REF) LABEL  = LIST OF ATOMIC NUMBERS
C
C  NONE OF THE ARGUMENTS ARE ALTERED BY THE CALL OF
VECPRN
C
C *************************************************************************
##INCLUDE '~/charmm_22/quantm.fcm'
      REAL*8 qmxyz
      common /qmforc/ natqml,natot,qmxyz(3,maxqm1),nqmpt(maxqm1)
      CHARACTER*2 ELEMNT, JTEXT(MAXQM1)
      integer LABELS
      common /mostuf/ ELEMNT(107),LABELS(maxqm1)
      CHARACTER * 6 LINE(21)
      DO 10 I=1,natqml
        JTEXT(I)=ELEMNT(LABELS(I))
   10 CONTINUE
      NUMB=ABS(NUMM)
      DO 50 I=1,21
   50 LINE(I)='------'
      LIMIT=(NUMB*(NUMB+1))/2
      KK=8
      NA=1
   60 LL=0
```

```
      M=MIN0((NUMB+1-NA),6)
      MA=2*M+1
      M=NA+M-1
      WRITE (6,100) (JTEXT(I),nqmpt(I),I=NA,M)
      WRITE (6,110) (LINE(K),K=1,MA)
      DO 80 I=NA,NUMB
         LL=LL+1
         K=(I*(I-1))/2
         L=MIN0((K+M),(K+I))
         K=K+NA
         IF ((KK+LL).LE.50) GO TO 70
         WRITE (6,120)
         WRITE (6,100) (JTEXT(N),nqmpt(N),N=NA,M)
         WRITE (6,110) (LINE(N),N=1,MA)
         KK=4
         LL=0
   70    WRITE (6,130) JTEXT(I),nqmpt(I),(A(N),N=K,L)
   80 CONTINUE
      IF (L.GE.LIMIT) GO TO 90
      KK=KK+LL+4
      NA=M+1
      IF ((KK+NUMB+1-NA).LE.50) GO TO 60
      KK=4
      WRITE (6,120)
      GO TO 60
   90 RETURN
C
  100 FORMAT (1H0/13X,10(2X,A2,I5,2X))
  110 FORMAT (1H ,21A6)
  120 FORMAT (1H1)
  130 FORMAT (1H ,3X,A2,I5,10F11.6)
C
      END
      SUBROUTINE FORCE
*************************************************************************
*
*  FORCE CALCULATES THE FORCE CONSTANTS FOR THE MOLECULE,
AND THE
*       VIBRATIONAL FREQUENCIES. ISOTOPIC SUBSTITUTION IS
ALLOWED.
*
*************************************************************************
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
##INCLUDE '~/charmm_22/dimens.fcm'
##INCLUDE '~/charmm_22/coord.fcm'
##INCLUDE '~/charmm_22/deriv.fcm'
##INCLUDE '~/charmm_22/quantm.fcm'
##INCLUDE '~/charmm_22/psf.fcm'
      REAL*8 qmxyz
      common /qmforc/ natqml,natot,qmxyz(3,maxqm1),nqmpt(maxqm1)
      character*2 ELEMNT
      character*4 sid,rid,ren,at
      integer LABELS
      common /mostuf/ ELEMNT(107),LABELS(maxqm1)
      DIMENSION xparam(3*MAXQM1),GRAD(3*maxqm1),
     1          fmatrx((3*maxqm1*((3*maxqm1)+1))/2),
```

```
     2        store((3*maxqm1*((3*maxqm1)+1))/2)
      LOGICAL nocalc
      equivalence (xparam(1),qmxyz(1,1))
      nocalc=.false.
      nvar=3*natqml
      idsp=0
      CALL COMPFG( XPARAM,ESCF,GRAD,idsp,nocalc)
      WRITE(6,'(//10X,"CARTESIAN COORDINATE DERIVATIVES",//4X,
     1"NO.",5x,"ATOM IDENTITY",14X,"DX",8X,"DY",8X,
     2"DZ",/)')
      II=-2
      DO 40 I=1,NATQML
                                          II=II+3
        call atomid(nqmpt(I),sid,rid,ren,at)
        WRITE(6,'(I6,3X,A4,2X,2A4,2X,A4,3X,3F10.4)')
     1  nqmpt(I),at,rid,ren,sid,
     2  GRAD(II),GRAD(II+1),GRAD(II+2)
   40 CONTINUE
C
C  TEST SUM OF GRADIENTS
C
      GNORM=SQRT(DOT1(GRAD,GRAD,NVAR))
      WRITE(6,'(//10X,"GRADIENT NORM =",F10.5)') GNORM
      IF(GNORM.LT.10.D0) GOTO 70
        WRITE(6,'(///1X,"** GRADIENT IS VERY LARGE, BUT ",
     1" THE CALCULATION WILL CONTINUE")')
        GOTO 80
   70 CONTINUE
C
C NOW TO CALCULATE THE FORCE MATRIX
C
   80 CONTINUE
      WRITE(6,'(/9X,"ORIENTATION OF MOLECULE IN FORCE
CALCULATION")
     1')
      WRITE(6,'(/,4X,"NO.",5X,"ATOM IDENTITY",14X,"X",
     19X,"Y",9X,"Z",/)')
      DO 100 I=1,NATQML
        call atomid(nqmpt(I),sid,rid,ren,at)
        WRITE(6,'(I6,3X,A4,2X,2A4,2X,A4,3X,3F10.4)')
     1  nqmpt(I),at,rid,ren,sid,X(nqmpt(I)),Y(nqmpt(I)),Z(nqmpt(I))
  100 CONTINUE
      CALL FMAT(FMATRX,ESCF)
C
C  THE FORCE MATRIX IS PRINTED AS AN ATOM-ATOM MATRIX
RATHER THAN
C  AS A 3N*3N MATRIX, AS THE 3N MATRIX IS VERY CONFUSING!
C
      IJ=0
      IU=0
      DO 130 I=1,NATQML
        IL=IU+1
        IU=IL+2
        IM1=I-1
        JU=0
        DO 120 J=1,IM1
```

```
      JL=JU+1
      JU=JL+2
      SUM=0.D0
C$DOIT ASIS
      DO 110 II=IL,IU
C$DOIT ASIS
        DO 110 JJ=JL,JU
 110    SUM=SUM+FMATRX((II*(II-1))/2+JJ)**2
      IJ=IJ+1
 120  STORE(IJ)=SQRT(SUM)
     IJ=IJ+1
 130 STORE(IJ)=SQRT(
   1FMATRX(((IL+0)*(IL+1))/2)**2+
   2FMATRX(((IL+1)*(IL+2))/2)**2+
   3FMATRX(((IL+2)*(IL+3))/2)**2+2.D0*(
   4FMATRX(((IL+1)*(IL+2))/2-1)**2+
   5FMATRX(((IL+2)*(IL+3))/2-2)**2+
   6FMATRX(((IL+2)*(IL+3))/2-1)**2))
    WRITE(6,'(//10X," FORCE MATRIX IN MILLIDYNES/ANGSTROM")')
    CALL VECPRN(STORE,NATQML)
    RETURN
    END
    SUBROUTINE FMAT(FMATRX, HEAT)
    IMPLICIT DOUBLE PRECISION (A-H,O-Z)
    DIMENSION FMATRX(*)
*************************************************************************
*
* VALUE CALCULATES THE SECOND-ORDER OF THE ENERGY WITH
*      RESPECT TO THE CARTESIAN COORDINATES I AND J AND
PLACES IT
*      IN FMATRX
*
* ON INPUT natqml  = NUMBER OF QM ATOMS IN THE SYSTEM.
*        qmxyz   = CARTESIAN COORDINATES OF QM ATOMS
*
* VARIABLES USED
*        COORDL  = ARRAY OF CARTESIAN COORDINATES, STORED
LINEARLY.
*        I       = INDEX OF CARTESIAN COORDINATE.
*        J       = INDEX OF CARTESIAN COORDINATE.
*
* ON OUTPUT FMATRX = SECOND DERIVATIVE OF THE ENERGY WITH
RESPECT TO
*                QM CARTESIAN COORDINATES I AND J.
*************************************************************************
##INCLUDE '~/charmm_22/dimens.fcm'
##INCLUDE '~/charmm_22/coord.fcm'
##INCLUDE '~/charmm_22/deriv.fcm'
##INCLUDE '~/charmm_22/quantm.fcm'
    REAL*8 qmxyz
    common /qmforc/ natqml,natot,qmxyz(3,maxqm1),nqmpt(maxqm1)
    DIMENSION GRAD(maxqm1*3), GROLD(maxqm1*3),
   1        COORDL(maxqm1*3),dumy(maxqm1*3)
    EQUIVALENCE (qmxyz(1,1),COORDL(1))
    DATA FACT/6.95125D-3/
C
```

```
      NVAR=natqml*3
      LIN=(NVAR*(NVAR+1))/2
      DO 30 I=1,LIN
   30 FMATRX(I)=0.D0
      WRITE(6,'(/1X,"CHARMM FIRST DERIVATIVES WILL BE USED IN THE"
     1," CALCULATION OF SECOND DERIVATIVES"/)')
      KOUNTF=0
      ISTART=1
C CALCULATE FMATRX
      LU=KOUNTF
      TOTIME=0
      DO 120 I=ISTART,NVAR
        LL=LU+1
        LU=LL+I-1
        L=0
                                  TIME1=secnds(0.0)
                                 DELTA=1.D0/120.D0
        COORDL(I)=COORDL(I)+0.5D0*DELTA
        GROLD(1)=100.D0
        CALL COMPFG(COORDL,ESCF,GROLD,I,.FALSE.)
        WRITE(6,'(/A,I5,A,F12.5/)')' CHECKING POINT',I,
     1  ': GNORM + 0.5*DELTA',SQRT(DOT1(GROLD,GROLD,NVAR))
        COORDL(I)=COORDL(I)-DELTA
        GRAD(1)=100.D0
        CALL COMPFG(COORDL,ESCF,GRAD,I,.FALSE.)
        WRITE(6,'(/A,I5,A,F12.5/)')' CHECKING POINT',I,
     1  ': GNORM - 0.5*DELTA',SQRT(DOT1(GRAD,GRAD,NVAR))
        COORDL(I)=COORDL(I)+0.5D0*DELTA
        CALL COMPFG(COORDL,ESCF,GROLD,I,.TRUE.)
        DO 100 KOUNTF=LL,LU
          L=L+1
          DUMY(L)=((GROLD(L)-GRAD(L)))*0.25D0/DELTA*FACT
          FMATRX(KOUNTF)=FMATRX(KOUNTF)+DUMY(L)
  100   CONTINUE
        L=L-1
        DO 110 K=I,NVAR
          L=L+1
          KK=(K*(K-1))/2+I
          DUMY(L)=((GROLD(L)-GRAD(L)))*0.25D0/DELTA*FACT
          FMATRX(KK)=FMATRX(KK)+DUMY(L)
  110   CONTINUE
                                  TIME2=secnds(0.0)
                                 TSTEP=TIME2-TIME1
                               TOTIME=TOTIME+TSTEP
        WRITE(6,'(/" STEP:",I4," TIME =",F9.2," SECS, INTEGRAL =
     1",F10.2/)')I,TSTEP,TOTIME
  120 CONTINUE
      CALL COMPFG(COORDL,ESCF,GRAD,I,.FALSE.)
      CALL FORSAV(TOTIME,I-1,FMATRX,GRAD,COORDL,NVAR,HEAT)
c
c     This is the end of the code hacked out from MOPAC for this
c     FORCE calculation attempt. This code is NOT guaranteed to work!
c     JAB Dec 1993
c
      RETURN
      END
```

APPENDIX III.
SAMPLE INPUT FILES FOR CHARMM, CAMVIB AND CAMISO

"Let me not understand [it] then;
speak it in Welsh."

- William Shakespeare

III.1 Introduction

This appendix provides two examples of CHARMM input files. They have been commented fairly extensively, and have been designed to illustrate the two major types of system setup that have been used for simulations in this thesis. A written description of what these files do is found in Chapter 4, sections 4.3.1 and 4.3.2.

In the files, a line beginning with an exclamation mark (!) represents a comment which has no effect on the way in which the file controls the execution of instructions. A line beginning with an asterisk (*) denotes a title which is expected by certain CHARMM commands. In particular, the first non-comment lines must constitute a title, and the WRITE command requires a title. A title may consist of any number of lines, ending with a line containing just an asterisk.

After the CHARMM examples there are examples of CAMVIB and CAMISO input. Comments are delineated by exclamation marks.

III.2 Example A : Solvation of a Small Molecule

```
* Setup PNA system for calculations
* Solvate PNA in a sphere of water molecules
*

! Define a parameter
set 1 pna_min

! Open topology and parameter files and the files containing
```

```
! coordinates for the water sphere and the starting structure
! of PNA
OPEN NAME /UC/BARNES/NMA_TOP.INP UNIT 11 CARD READ
OPEN NAME /UC/BARNES/NMA_PAR.INP UNIT 12 CARD READ
OPEN NAME /UC/BARNES/WAT15_SPHERE.PDB UNIT 13 FORM READ
OPEN NAME /WORK/ASL9801/PNA/PDB/PNA_CHAIR.PDB UNIT 14 FORM
READ

! Read topology and parameter files
READ    RTF CARD  UNIT  11
READ    PARAMETER CARD  UNIT  12

! Prepare for coordinate input. In this case, the PNA is being
! defined as a segment called PNAN, and the water sphere as
! a segment called SOLV. The first thing to do is to tell CHARMM
! what sequence of residues make up the segment. This requires
! a title followed by one line containing the number of residues
! followed by a list of all the residues. The generate command is
! then used to build the segment. The data file can then
! be read. Note the shorthand form that is available for water.
READ SEQUENCE CARD
*PNA substrate
*
  1
  PNA
GENERATE PNAN SETUP
READ COOR PDB UNIT 14
```

```
! Now add waters for solvation
READ SEQUENCE TIP3 502
GENERATE SOLV SETUP NOANG NODIHE

! The append option is used because some coordinates of the
! system have already been read in.
READ COOR PDB APPEND UNIT 13

! The anomeric carbon should be placed at the centre of the
! water sphere, which is at the origin (0,0,0). The PNA atoms
! should therefore be translated appropriately.
!Anomeric C is C4 of PNA, atom 17 in the coord file
! quick returns coordinates in xval, yval and zval.
! set gives the variable number a value.
! The @ symbol means substitute the variable's value here.
QUICK 17
SET 5 -1
SET 2 ?XVAL
SET 3 ?YVAL
SET 4 ?ZVAL
MULT 2 BY @5
MULT 3 BY @5
MULT 4 BY @5

! Translate a set of atoms, in this case the atoms which are not
! in the SOLV segment
COOR TRANS SELE .NOT. SEGID SOLV END XDIR @2 YDIR @3 ZDIR @4
```

```
!delete overlapping waters
DELE ATOM SELE ( .BYRES. ( SEGID SOLV .AND. TYPE OH2 .AND. -
   ( ( .NOT. SEGID SOLV .AND. .NOT. HYDROGEN ) -
   .AROUND. 2.50 ) ) ) END

! Set up the QM system. Make sure that the cutoff for non-bonded
! interactions begins outside the water sphere by setting the
! cutoff to be at least twice the water sphere radius.
QUAN SELE SEGID PNAN END AM1 CHARGE -1
NBONDS ATOM SWITCH CUTNB 40.00 CTOFNB 39.00 CTONNB 35.00

!Set up the deformable solvent boundary to maintain the
! structure of water sphere. The .pot file contains the
! preset potential for the size of sphere being used.
OPEN  NAME /UC/BARNES/WAT15.POT UNIT 4 FORM READ
SBOUND READ UNIT 4

! This sets the centre of the potential and tells it which
! atoms to apply the potential to.
SBOUNDARY SET XREF 0.0 YREF 0.0 ZREF 0.0 -
      ASSIGN 1 SELECTION SEGID SOLV .AND. TYPE O* END

! do minimization
! I always do minimizations in short blocks and save data in case of accidents!
! The loop min_loop will do up to 10,000 cycles of minimization in sets of 100.
! For each set, the termination criterion is that each individual gradient is less
! than 0.001, and the whole minimization stops if the gradient norm is below 0.05
! or 10,000 cycles have been done.
```

```
SET 2 0
SET 3 10000
SET 4 0.05

LABEL MIN_LOOP
MINI ABNR NSTEP 100 NPRINT 20 STEP 0.02 TOLGRD 0.001
INCR 2 BY 100

! write structure
OPEN NAME /WORK/ASL9801/PNA/PDB/@1.PDB UNIT 17 FORM WRITE
WRITE COOR PDB UNIT 17
* PNA IN WATER MINIMIZATION
* @2 CYCLES ABNR
* CHECK @1.LOG FOR CONVERGENCE
*
CLOSE UNIT 17
IF 4 GT ?GRMS GOTO ENDLOOP
IF @2 LT @3 GOTO MIN_LOOP

LABEL ENDLOOP
STOP
```

III.3 Example B : Setting up a Truncated Enzyme System

```
* Minimize enzyme + PNA
*
! Read in a saved cutoff structure. This example continues from
! a saved file, so all the coordinates are in one file. If the
```

```
! original system has been truncated, you must still set the
! whole system up and do the truncation to ensure that CHARMM
! recognises all the atoms, so the first part of this example
! does just that.

SET 1 STQM1
OPEN NAME /UC/BARNES/NMA_TOP.INP UNIT 11 CARD READ
OPEN NAME /UC/BARNES/NMA_PAR.INP UNIT 12 CARD READ
!
! The next file should be the original file that was read in and which
! was used as a start point for truncating the system. If you did
! any manipulations of the whole system before truncating it you will
! have to repeat those to guarantee that you can read back the
! truncated system properly. If necessary, save an intermediate
! structure to start from and retain for this purpose.
! You should delete the appropriate atoms in the system
! in exactly the same way as before, and then read the cutdown
! system into the comparison coordinate set before swapping coordinates.
! This will ensure that the system is exactly the same as when it was
! saved. It is tedious but the only way I can think of to
! regain the coordinates of a truncated structure.
!
OPEN NAME /UC/BARNES/STMOD1_START.PDB UNIT 13 FORM READ
!
! Now set up the old,full system in exactly the same way as originally
!
READ    RTF CARD  UNIT  11
READ    PARAMETER CARD  UNIT  12
```

READ SEQUENCE CARD

* SEQUENCE from sue crennel

* The S.typhimurium NA enzyme

* NMA1

*

381

THR VAL GLU LYS SER VAL VAL PHE LYS ALA GLU GLY GLU HSP PHE
THR ASP GLN LYS GLY ASN THR ILE VAL GLY SER GLY SER GLY GLY
THR THR LYS TYR PHE ARG ILE PRO ALA MET CYS THR THR SER LYS
GLY THR ILE VAL VAL PHE ALA ASP ALA ARG HSP ASN THR ALA SER
ASP GLN SER PHE ILE ASP THR ALA ALA ALA ARG SER THR ASP GLY
GLY LYS THR TRP ASN LYS LYS ILE ALA ILE TYR ASN ASP ARG VAL
ASN SER LYS LEU SER ARG VAL MET ASP PRO THR CYS ILE VAL ALA
ASN ILE GLN GLY ARG GLU THR ILE LEU VAL MET VAL GLY LYS TRP
ASN ASN ASN ASP LYS THR TRP GLY ALA TYR ARG ASP LYS ALA PRO
ASP THR ASP TRP ASP LEU VAL LEU TYR LYS SER THR ASP ASP GLY
VAL THR PHE SER LYS VAL GLU THR ASN ILE HSP ASP ILE VAL THR
LYS ASN GLY THR ILE SER ALA MET LEU GLY GLY VAL GLY SER GLY
LEU GLN LEU ASN ASP GLY LYS LEU VAL PHE PRO VAL GLN MET VAL
ARG THR LYS ASN ILE THR THR VAL LEU ASN THR SER PHE ILE TYR
SER THR ASP GLY ILE THR TRP SER LEU PRO SER GLY TYR CYS GLU
GLY PHE GLY SER GLU ASN ASN ILE ILE GLU PHE ASN ALA SER LEU
VAL ASN ASN ILE ARG ASN SER GLY LEU ARG ARG SER PHE GLU THR
LYS ASP PHE GLY LYS THR TRP THR GLU PHE PRO PRO MET ASP LYS
LYS VAL ASP ASN ARG ASN HSP GLY VAL GLN GLY SER THR ILE THR
ILE PRO SER GLY ASN LYS LEU VAL ALA ALA HSP SER SER ALA GLN
ASN LYS ASN ASN ASP TYR THR ARG SER ASP ILE SER LEU TYR ALA
HSP ASN LEU TYR SER GLY GLU VAL LYS LEU ILE ASP ASP PHE TYR

```
PRO LYS VAL GLY ASN ALA SER GLY ALA GLY TYR SER CYS LEU SER
TYR ARG LYS ASN VAL ASP LYS GLU THR LEU TYR VAL VAL TYR GLU
ALA ASN GLY SER ILE GLU PHE GLN ASP LEU SER ARG HSP LEU PRO
VAL ILE LYS SER TYR ASN
GENERATE NMA1 SETUP

! waters of crystal structure.
READ SEQUENCE TIP3 218
GENERATE WAT1 SETUP NOANG NODIHE

! PNA with link atoms separating the molecule

READ SEQUENCE CARD
*PNA SUBSTRATE WITH THE GLYCERYL AND N-ACETYL SIDE CHAINS
*
  1
  PNAQ
GENERATE PNAN SETUP

! now add waters for solvation
READ SEQUENCE TIP3 170
GENERATE SOLV SETUP NOANG NODIHE
READ COOR PDB UNIT 13

! Add the link atoms midway along the bonds which are to be
! the boundaries of the QM/MM system for pna.
! The axis command defines an axis between the two selected atoms
! and puts the coordinates of the middle of the axis in
```

```
! xcen, ycen and zcen. The coor set then assigns coordinates.
COOR AXIS SELE SEGID PNAN .AND. TYPE N11 END -
      SELE SEGID PNAN .AND. TYPE C7 END
COOR SET XDIR ?XCEN YDIR ?YCEN ZDIR ?ZCEN -
   SELE SEGID PNAN .AND. TYPE H52 END
COOR AXIS SELE SEGID PNAN .AND. TYPE C14 END -
      SELE SEGID PNAN .AND. TYPE C6 END
COOR SET XDIR ?XCEN YDIR ?YCEN ZDIR ?ZCEN -
   SELE SEGID PNAN .AND. TYPE H53 END

!delete all residues with no atoms within 17A of pna
DELE ATOM SELE .NOT. ( .BYRES. ( SEGID PNAN .AROUND. 17.00 )) END

!
! This is the previously saved truncated system to be read in at last
!
OPEN NAME /WORK/PDB/STNMA1_MIN.PDB UNIT 22 FORM READ
READ COOR COMP PDB UNIT 22
!
! Note that the coordinates swapped are the set that are NOT deleted -
! this is an easy thing to get wrong!
!
COOR SWAP SELE ( .BYRES. ( SEGID PNAN .AROUND. 17.00 )) END
!
! The truncated system is now restored.
!
! Because the swapped coords don't have the link atoms, repeat this
! Add the link atoms midway along the bonds which are to be
```

```
! the boundaries of the QM/MM system for pna
COOR AXIS SELE SEGID PNAN .AND. TYPE N11 END -
      SELE SEGID PNAN .AND. TYPE C7 END
COOR SET XDIR ?XCEN YDIR ?YCEN ZDIR ?ZCEN -
   SELE SEGID PNAN .AND. TYPE H52 END
COOR AXIS SELE SEGID PNAN .AND. TYPE C14 END -
      SELE SEGID PNAN .AND. TYPE C6 END
COOR SET XDIR ?XCEN YDIR ?YCEN ZDIR ?ZCEN -
   SELE SEGID PNAN .AND. TYPE H53 END

!use deformable solvent boundary to maintain structure of water sphere
!
OPEN  NAME /UC/BARNES/WAT15.POT UNIT 16 FORM READ
SBOUND READ UNIT 16
SBOUNDARY SET XREF 0.0 YREF 0.0 ZREF 0.0 -
ASSIGN 1 SELECTION SEGID SOLV .AND. TYPE O* END

! constrain buffer atoms with harmonic force.
CONS HARM FORCE 1.0 SELE ( (SEGID NMA1 .OR. SEGID WAT1 ) .AND. -
   ( .NOT. TYPE H* ) .AND. -
   ( .NOT. POINT 0.0 0.0 0.0 CUT 15.0 ) ) END MASS EXPO 2

! define the quantum atoms, in this case the PNA moelcule without the glyceryl

! and N-acetyl side chains
QUAN SELE SEGID PNAN .AND. .NOT. (TYPE N11 .OR. TYPE H27 -
 .OR. TYPE C12 .OR. TYPE O17 .OR. TYPE C13 .OR. TYPE H28 -
 .OR. TYPE H29 .OR. TYPE H30 .OR. TYPE C14 .OR. TYPE C15 -
```

```
.OR. TYPE C16 .OR. TYPE O18 .OR. TYPE O19 .OR. TYPE O20 -
.OR. TYPE H31 .OR. TYPE H32 .OR. TYPE H33 .OR. TYPE H34 -
.OR. TYPE H35 .OR. TYPE H36 .OR. TYPE H37 ) END AM1 CHARGE -1

! increase non-bonded cutoff radius inside reaction sphere
NBONDS ATOM SWITCH CUTNB 40.00 CTOFNB 39.00 CTONNB 34.00

! minimize, using a loop as explained in Example A.
SET 2 100
SET 3 5000
SET 4 0.05
LABEL LOOP2

MINI ABNR STEP 0.02 NSTEP 100 NPRINT 20 TOLGRD 0.0005
! Reminder in log file
SET 7 ?QMEL
INCR 7 BY ?QMVD
WRITE TITLE UNIT 6
* STMOD1_START MINIMIZATION
* @2 ABNR
* CHECK @1_MIN.LOG FILE FOR CONVERGENCE
*ENERGIES PE  QMELEC  QMVDW  QMTOT
*ENERGIES ?ENER , ?QMEL , ?QMVD , @7
*
OPEN  NAME  /WORK/ASL9801/STMODELS/PDB/@1_MIN.PDB  UNIT  25
FORM WRITE
WRITE COOR PDB UNIT 25
* STMOD1_START MINIMIZATION
```

```
* @2 ABNR
* CHECK @1_MIN.LOG FILE FOR CONVERGENCE
*ENERGIES PE  QMELEC  QMVDW  QMTOT
*ENERGIES ?ENER , ?QMEL , ?QMVD , @7
*
CLOSE UNIT 25
IF 4 GT ?GRMS GOTO ENDLOOP
INCR 2 BY 100
IF 2 LT @3 GOTO LOOP2
LABEL ENDLOOP
STOP
```

III.4 Example C : CAMVIB Input for Methanol

```
VIB
VECT AMPA REDU BREF VALE IFCS RELX DISO   ! Control words
meoh   ! Title
 1 1 0
! Start of atom descriptors
 12C
 16O
 1H
 1H
 1H
 1H
*
!  Start of normal coordinate definition, stretches, bends etc
S  1  2     0  0  1.0    C1C2   ! 1.0 is a scaling factor, the C1C2 is a label.
```

```
S 1 3
S 1 4
 S 1 5
S 2 6
B 2 1 3              TET3     ! TET3 is a set of tetrahedral bends
B 2 1 4              TET3
B 2 1 5              TET3
B 3 1 5              TET3
B 3 1 4              TET3
B 4 1 5              TET3
B 1 2 6
T 2 1          ! torsional motion about atoms 1 and 2
*
STOP
```

III.5 Example D : CAMISO input

```
ISO
EQUI 2  ! Do EIE calculation, 2 sets of species
papn    ! A label for the run
 1  333.18   ! Temperature
 1  0
 1 papn       ! Labels for the reactant and product states
 1 pcat
 1 1 0 papn     !
 ! Coordinates of the structure, only first and last shown
 16O  -1.2893087  2.1376540  -1.1260989
 |      |        |        |
```

```
 1H  -2.9963994  2.9300920  -.4187149
*
 ! Force constant components, only first and last lines shown
   .628091184794E+00 -.626442501332E-01  .292038561165E+00
   |            |            |
   .532667402908E-01  .962556611644E-02  .617237017668E-01
   ! The zero means a minimum energy stationary point, then the number
   ! of vibrations to ignore followed by their numbers
  0  6  1  2  3  4  5  6
  ! Repeat the above for the product side of the reaction
 1 1 0 pcat
 16O   -.6689165 -1.2213760   .0473244
 |     |       |       |
  1H  -2.3154556  -.2147890  -.5853703
*
   .909839821168E+00 -.603704948772E-01  .392517502881E+00
   |            |            |
   .131821419525E+00  .320627191334E-01  .138178250437E+00
  0  6  1  2  3  4  5  6
  ! Now repeat calculation using the isotopic substitutions shown
  ! Labels for reactants and products
 1 7H
 1 7H
 ! Start reactant side description
 1 1 0 R  D at 7H !
 ! Atomic masses to use for the second calculation, and which vibrations
 ! to omit
 16O  12C  12C  12C  12C  12C   2H   1H   1H   1H   1H  16O  12C  12C   1H  12C
```

1H 12C 14N 16O 16O 12C 1H 12C 1H 1H 1H 1H 1H

0 6 1 2 3 4 5 6

! Repeat for the product side

1 1 0 P D at 7H

16O 12C 12C 12C 12C 12C 2H 1H 1H 1H 1H 1H 1H 1H 1H

0 6 1 2 3 4 5 6

STOP