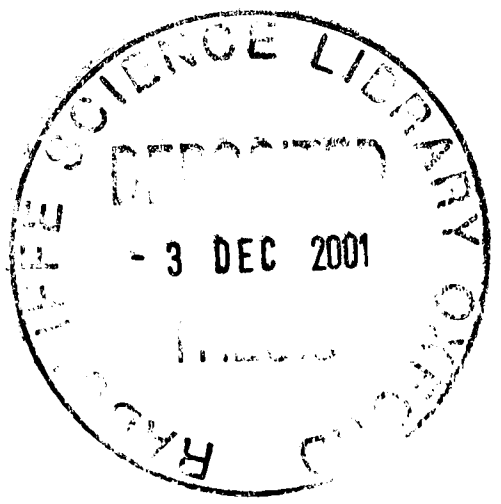
Computer Simulation of Phosphate Ester Hydrolysis Reactions.

Owen Anthony Walsh

A thesis submitted for the degree of
Doctor of Philosophy in the
University of Oxford



Brasenose College

Trinity Term 2001

Computer Simulation of Phosphate Ester Hydrolysis Reactions

Abstract of a thesis submitted for the degree of Doctor of Philosophy

in the University of Oxford

Owen Anthony Walsh Brasenose College

Trinity Term 2001

Phosphate esters and their hydrolysis reactions underpin many of the most important reactions in biology and have therefore been the focus of continued research.

In this thesis, a range of theoretical studies on two phosphate ester hydrolysis reactions is presented. Particular attention is paid to the role of solvent in these reactions. The first reaction studied is the hydrolysis of dimethyl phosphate. A study using density functional theory examines three possible mechanisms for this reaction in the gas phase and the presence of solvent is also considered through the use of reaction field methods. The base-catalysed mechanism for this reaction receives further attention and the reaction energy profile in solution is adiabatically mapped using two different hybrid QM/MM potentials. The use of a QM/MM potential facilitates an atomistic representation of the solvent. Another study determines the potential of mean force for this reaction by simulating the reaction using QM/MM molecular dynamics simulations in conjunction with umbrella sampling methods. Both QM/MM studies demonstrate that, in solution, the rate-determining step for the base-catalysed reaction is the cleavage of the bond between the phosphorous and the leaving group. The QM/MM studies also established that the reaction proceeds via the formation of a short-lived pentacoordinated intermediate species.

The second reaction studied was the base-catalysed hydrolysis of methyl ribose phosphate, a realistic model of the nonenzymatic hydrolysis of RNA. Since the hydrolysis of thio-substituted analogues of RNA have been the study of many experimental studies probing the mechanism of RNA hydrolysis, theoretical studies of the hydrolysis of two thio analogues of methyl ribose phosphate are also presented. The results of these studies provide valuable insight into the mechanism of RNA hydrolysis

and concludes that the cleavage of P–O5′ bond corresponds to the reaction transition state.

Acknowledgements

Given that my D. Phil. research began in the last century, there are many people who helped me along the way and therefore deserve to be acknowledged.

Firstly, I'd like to thank my two supervisors Paul Lyne and Graham Richards. This research topic was Paul's idea and before and after he left for Boston, he shared the vast knowledge he has accumulated in this field with me. I would also like to thank him for teaching me never to bet again, no matter how tempting the odds offered may seem. Graham has also been a very enthusiastic and supportive supervisor whose generosity knows no bounds. I will always fondly remember the Thursday lunchtime trips to the Kings Arms where the group was plied with alcoholic beverages and the odd watermelon Bacardi Breezer.

I could not imagine doing the research that has gone into this thesis without the help and company of the other members of the W.G.R. group. I'd particularly like to mention `benwebb.pcl.ox.ac.uk` for all the help he has given me in my many struggles against obstinate computers and the rarely logical behaviour of CHARMM. He also tirelessly proof-read this thesis (all remaining mistakes are deliberate) and for that alone he deserves great sympathy. Thanks must also go to: Stewart Adcock, who started and finished his D.Phil at the same time as me and is now off to California to join the cast of Baywatch whilst working part-time in UCSD. Maya "P" Topf for being a serene and even-tempered influence in the lab and for also proof-reading much of the thesis; Drs. Aaron Dinner, Xabi Lopez, and Meir "dilithium" Glick for all their help, advice, and hospitality; Péter "would you like to see my profile?" Várnai for never running out of questions; and finally Martin Parretti and Kate Frost for providing many funny moments during my first year despite rupturing my eardrums.

On the rare occasions I left my desk in the lab, I made many friends in Brasenose College whom I'd like to thank for all the great times that were had. I'd also like to acknowledge the financial support given to me by the Wellcome Trust.

Finally, I'd like to thank my mother and father, to whom this thesis is dedicated, for all the support and guidance they have given me thus far.

Contents

Introduction 1

1 Theoretical models 4

1.1 Introduction 4

1.2 Molecular quantum mechanics 5

1.3 Determination of electronic structure 7

1.3.1 The Hartree-Fock approach 8

1.3.2 Basis sets 11

1.4 Density functional theory 13

1.5 Semi-empirical methods 17

1.6 Geometry optimisation 18

1.7 Transition state location 21

1.8 Normal mode analysis 22

1.9 Determination of reaction energy profiles 23

1.9.1 Reaction path following methods 23

1.9.2 Adiabatic mapping methods 24

1.10 Calculation of atomic charge 25

1.11 Molecular mechanics force fields 27

1.11.1 The potential energy function 28

1.11.2 Other features of force fields 30

1.12 Molecular dynamics simulation methods 31

1.12.1 Theory 31

1.12.2 Molecular dynamics algorithms 32

1.12.3 Calculation of free energy 34
1.13 Solvent representations 38
1.13.1 Explicit solvent models 39
1.13.2 Continuum solvent models 39
1.14 Summary 41

2 QM/MM Methods **42**
2.1 Introduction 42
2.2 Simulation of enzymatic reactions 43
2.3 Methodology 44
2.3.1 The quantum-classical boundary 51
2.4 Applications 54
2.4.1 Triosephosphate Isomerase. 55
2.4.2 Bovine Protein Tyrosine Phosphatase. 57
2.4.3 Citrate Synthase 59
2.5 Summary 62

3 Phosphate Ester Hydrolysis **64**
3.1 Introduction 64
3.2 General mechanisms 66
3.3 The hydrolysis of DNA 71
3.3.1 DMP - a model compound for DNA 72
3.4 The hydrolysis of RNA 73
3.4.1 Ribozymes 75
3.4.2 Thio-substituted RNA analogues 76
3.5 Summary 78

4 A quantum chemical study of DMP hydrolysis **79**
4.1 Introduction 79
4.2 Methods 80
4.3 Results 82
4.3.1 Base catalysed hydrolysis of dimethyl phosphate 82

4.3.2 Acid catalysed hydrolysis of dimethyl phosphate 89
4.3.3 Acid catalysed hydrolysis of protonated dimethyl phosphate . . 98
4.4 Summary 106

5 A quantum chemical study of RNA hydrolysis **107**
5.1 Introduction 107
5.2 Methods 109
5.3 Results 111
5.3.1 Reaction A: the hydrolysis of methyl ribose phosphate 111
5.3.2 Reaction B: the hydrolysis of thiomethyl ribose phosphate . . . 116
5.3.3 Reaction C: the hydrolysis of methyl ribose phosphothiolate . . 121
5.4 Summary 127

6 QM/MM minimisations **129**
6.1 Introduction 129
6.2 Methods 130
6.2.1 Construction and Equilibration of the System 130
6.2.2 Determination of the reaction energy profile 132
6.2.3 Choice of QM potential 134
6.2.4 Computational details 135
6.3 Results 136
6.3.1 HF 3-21G*/MM potential energy surface 136
6.3.2 Further examination of phosphorane intermediate 143
6.3.3 AM1d/MM potential energy surface 145
6.4 Summary 153

7 A QM/MD simulation of the hydrolysis of DMP **155**
7.1 Introduction 155
7.2 Methods 156
7.2.1 Preparation of system 156
7.2.2 Umbrella sampling 157
7.2.3 Simulation protocol 157

7.2.4 Generation of the PMF 159
7.3 Results 159
7.3.1 Stationary points 160
7.3.2 Radial distribution functions 163
7.3.3 Energetics 165
7.4 Conclusions 166

8 Conclusions and future research prospects **167**
8.1 Conclusions 167
8.2 Prospects for future research 170

Bibliography **174**

Introduction

What do the maintenance of genetic code, the storage of metabolic energy, and the regulation of cellular signalling have in common? One answer to this question is that each is fundamental to the functioning of organisms. A less obvious answer is that each process is mediated by phosphate esters and their hydrolysis reactions [1]. Although this piece of information may come as a surprise to some, there has been a huge effort by chemists and biochemists to elucidate the features of phosphate ester hydrolysis reactions [2, 3]. However, despite this, many questions remain unanswered.

Experimental studies of phosphate ester hydrolysis reactions have been complicated by two factors. The first is the extraordinary stability of phosphoester bonds which is manifested by the extraordinarily slow rates at which these bonds are hydrolysed in solution. While this feature makes phosphate esters eminently suitable for their role as links between individual nucleotides in DNA and RNA, it complicates the experimental determination of rate constants. Secondly, in solution, phosphate ester hydrolysis reactions can proceed via a combination of different mechanisms, thus making unambiguous determination of reaction mechanism and energetics difficult. This thesis aims to avoid these difficulties by examining the features of phosphate ester hydrolysis reactions using theoretical methods. However, this approach also presents obstacles, albeit in a different form.

Upon inspecting the literature concerning theoretical studies of chemical reactions, one might be forgiven for concluding that most chemical reactions occur in the gas phase. Whilst this is the case in the realms of combustion and atmospheric chemistry,

the majority of chemical reactions actually occur in solution. So why do so many theoretical studies focus on the mechanism and energetics of reactions in the gas phase? One reason is that many reactions, especially those that don't involve the generation of charge, have similar energetics and mechanisms in the gas phase and in solution. Perhaps a more pragmatic answer to the question is that modelling the effects of solution on a reaction is both computationally demanding and academically challenging.

The goal of this thesis is to further our understanding of the mechanism and energetics of phosphate ester hydrolysis reactions. The results presented not only deal with phosphate ester hydrolysis reactions in the gas phase but also examine the important role played by aqueous solvent in determining the mechanistic and energetic features of these reactions. Both continuum and atomistic models are employed to represent the solvent. Hybrid quantum mechanical and molecular mechanical (QM/MM) potentials are used to facilitate the latter description of the solvent.

A number of different reactions are examined. These can be categorised into two groups. The first set of reactions concern the various mechanisms by which the hydrolysis of dimethyl phosphate (DMP) can occur. This reaction is perceived to be a good model for the hydrolysis of DNA and therefore has been the subject of many research studies. However, the exact mechanism of the reaction is still a matter of controversy [4]. The second set of reactions studied concern the hydrolysis of RNA. Since it has been discovered that RNA has catalytic properties, this reaction has been at the forefront of research [5,6]. In this thesis, the base-catalysed hydrolysis of methyl ribose phosphate, a good model for RNA, is examined. Since thio analogues of RNA have been used in experiments to probe the reactivity of RNA, the hydrolysis of equivalent analogues of methyl ribose phosphate is also covered.

Chapter 1 presents a review of the theoretical methods employed in the research presented. Particular attention is paid to the theories that underpin these methods

and the algorithms that are used to implement them. Chapter 2 gives a description of QM/MM potentials. Particular attention is paid to the formation of these potentials as well as some of their more prominent applications in the literature. An introduction to the chemistry of phosphate esters, and in particular their hydrolysis reactions, is presented in Chapter 3. Both experimental and theoretical studies are reviewed. The chapter concludes with a review of research on the hydrolysis of both DNA and RNA.

Chapter 4 details the results of a quantum chemical study of the hydrolysis of DMP. Three possible mechanistic pathways are evaluated both in the gas phase and in solution. Chapter 5 presents a similar study on the base-catalysed hydrolysis of methyl ribose phosphate. The effects of two different thio-substitutions on the reaction mechanism and energetics are also considered.

In Chapter 6, the results of a QM/MM study on the effect of solvent on the reaction energy profile for the base-catalysed hydrolysis of DMP is presented. Chapter 7 details the use of molecular dynamics simulations to examine the free energy profile of the same reaction. The results of these simulations provide a clearer insight into the role of solvent in the reaction.

Chapter 1

Theoretical models

1.1 Introduction

The last 25 years have seen, through extraordinary advances in computer hardware and algorithms, huge increases in the computational power available to scientists. These improvements allowed the fields of theoretical physics and chemistry develop from being relatively esoteric disciplines to being effective and practical tools that are now part of everyday science. This, in turn, has brought about the development of a vast array of computational methods which can be used to determine the properties of molecular systems. This chapter reviews some of these methods and the theories that underpin them. Special attention is paid to the methods employed in this thesis.

Beginning with a brief introduction to the basic tenets of molecular quantum mechanics, the chapter proceeds to outline how molecular electronic structure can be accurately estimated using molecular orbital theory and the Hartree-Fock approach in what are known as *ab initio* methods. As well as employing these methods, the quantum mechanical calculations detailed in this thesis also utilised the relatively modern and increasingly popular density functional theory and the more approximate semi-empirical method to determine electronic structure. Descriptions of these methods are also included.

A key application of electronic structure theory is the determination of the molecular potential energy surface. By exploring the potential energy surface, using geometry optimisation methods, the lowest energy conformations of a system can be determined. A description of these methods constitutes the next part of this chapter.

In the case of large systems, such as condensed phase systems, the use of electronic structure methods is no longer viable for energy determinations due to the huge computational effort required. Instead, classical approaches, which essentially treat a molecule as a collection of interconnected 'balls' and 'springs' can be employed. These methods, known as molecular mechanics methods, can be used in conjunction with standard optimisation methods or better still in conjunction with time-dependent simulations. The theory underpinning these methods and the computational algorithms used to apply them are also described in this chapter.

Finally, the chapter concludes with a description of some of the methods available to model the effect of solvent on chemical reactions. Particular attention is paid to atomistic and continuum descriptions.

1.2 Molecular quantum mechanics

The prime aim of modern quantum chemistry is to obtain an exact solution to the Schrödinger equation for molecular systems. In its simplest form, the time-independent Schrödinger equation is expressed as:

$$\hat{H}(r,R)\Psi(r,R) = E(R)\Psi(r,R) \tag{1.1}$$

Solving this equation yields the molecular wavefunction, $\Psi(r,R)$, which depends explicitly on the coordinates of all the electrons in the system r, and implicitly on the collective nuclear coordinates, denoted as R. The use of the Born-Oppenheimer approximation [7] allows for the implicit rather than explicit dependence on nuclear

coordinates[1]. Therefore, the electronic wavefunction depends only on the position of the nuclei and not their momenta and consequently can be written as

$$\Psi(nuclei, electrons) = \Psi(electons)\Psi(nuclei) \tag{1.2}$$

Solving the Schrödinger equation also yields $E(R)$, which is the energy of the system and is dependent on the positions of the nuclei. $\hat{H}$ is the energy operator and is known as the Hamiltonian operator. The Hamiltonian consists of a kinetic energy term and three potential energy terms, all of which are subject to frozen nuclei,

$$\begin{aligned}\hat{H}(r,R) = &-\frac{\hbar^2}{2m_e}\sum_i^n\left(\frac{\partial^2}{\partial x_i^2}+\frac{\partial^2}{\partial y_i^2}+\frac{\partial^2}{\partial z_i^2}\right)-\sum_i^n\sum_\alpha^N\frac{Z_\alpha}{|r_i-R_\alpha|}\\ &+\frac{1}{2}\sum_i^n\sum_j^n\frac{1}{|r_i-r_j|}+\frac{1}{2}\sum_\alpha^N\sum_\beta^N\frac{Z_\alpha Z_\beta}{|R_\alpha-R_\beta|}\end{aligned} \tag{1.3}$$

where n is the number of electrons in the system, and N the number of nuclei. r and R refer to the positions of the electrons and nuclei respectively, while Z refers to the charge on a given nucleus. The subscripts α and β run over the nuclei and i and j over the electrons.

The first term of the right-hand side of the equation represents the kinetic energy of each electron, of mass m_e, moving through space in the x, y, and z directions. The second, third, and fourth terms are Coulombic terms which represent the electron-nuclear attractions, electron-electron repulsions, and nuclear-nuclear repulsions within the system, respectively. The energy of the system for any given nuclear configuration can be determined by rearranging Equation 1.1 to give

[1]This approximation, which states that the motion of the electrons and nuclei within a system can be considered independently of each other, stands because the mass of nuclei is far greater than that of electrons and thus electrons adjust almost instantaneously to any changes in the positions of nuclei.

$$E = \frac{\int \Psi^* \hat{H} \Psi d\tau}{\int \Psi^* \Psi d\tau} \tag{1.4}$$

where Ψ^* is the complex conjugate of the wavefunction and $d\tau$ indicates that the integration is over all space. How the molecular wavefunction of a system and its energy are determined and the approximations that are necessary to make in order to do so are presented in the next section.

1.3 Determination of electronic structure

The exact resolution of the Schrödinger equation is only possible for one electron systems [8]. As the vast majority of chemical species that we are interested in are polyelectronic, a number of approximations must be made to make solving the Schrödinger equation a tractable exercise. One of these, the Born-Oppenheimer approximation, has already been mentioned. Other approximations ensure that the exact form of Equation 1.3 does not have to be solved in order to determine the molecular wavefunction and energy of a system. One such approximation allows the molecular wavefunction, Ψ, for an N electron system to be written as the product of N individual one-electron wavefunctions ψ:

$$\Psi_{electronic} = |\psi_1 \psi_2 \ldots \psi_N| \tag{1.5}$$

Each of the individual wavefunctions ψ is known as a molecular orbital (MO) and is the product of a spatial term and a spin function. This form of expressing the overall wavefunction is known as a Slater determinant. It has the advantage of satisfying the condition that the overall wavefunction must be antisymmetric under electron exchange. That is, the sign of the wavefunction must change if the space and spin

coordinates of two electrons are interchanged[2]. In a further approximation known as the Linear Combination of Atomic Orbitals (LCAO) approximation, each MO is expressed as a linear combination of atomic orbitals, ϕ, each having an appropriate weighting coefficient $c_{i\mu}$,

$$\psi_i = \sum_{\mu=1}^{L} c_{i\mu}\phi_\mu \tag{1.6}$$

The atomic orbitals are normally known as basis functions. In equation 1.6, the molecular orbital is made up of L basis functions in what is known as a basis set. The composition of basis sets is discussed in detail in Section 1.3.2.

By determining a set of $c_{i\mu}$ values for each MO an approximate solution of the Schrödinger equation can be determined. The most common way of doing this is to use the Hartree-Fock approach.

1.3.1 The Hartree-Fock approach

By replacing the Schrödinger equation with an approximate set of equations describing the electronic structure of a system, it becomes possible to determine the optimal electron distribution for a given system. This is normally done by initially guessing the values of the coefficients, $c_{i\mu}$, and then iteratively changing their values until a satisfactory degree of accuracy is obtained. This process of improving the quality of a wavefunction relies on the use of the variational principle which states that the energy of an approximation to the true wavefunction of a molecule can never be less than that of the true wavefunction. In other words: the better the wavefunction, the lower the energy.

The most commonly used methods of minimising the energy of a wavefunction are based on the Hartree-Fock approach [9, 10] which uses the variational principle to

[2]This condition is derived from the Pauli exclusion principle which states that no two electrons may have the same quantum number.

determine the set of basis set expansion coefficients which minimise the energy of the wavefunction with respect to the coefficients. This is done by using the self-consistent field (SCF) procedure which treats each electron one at a time and assumes that each electron experiences the potential due to the other electrons in the system as a mean field. The calculation proceeds by initially guessing the wavefunction for each electron and from this the mean potential experienced by each electron can be determined. This, in turn, enables the calculation of a new wavefunction. This procedure is continued by iteratively recalculating each one-electron wavefunction until they remain unchanged for one complete cycle.

The Hartree-Fock approximation is applied by deriving a set of equations known as the Roothaan-Hall [11] equations. These are now explained with reference to a closed-shell system containing N electrons and $N/2$ orbitals. In this formulation, the expansion coefficients from Equation 1.6 must satisfy the following set of equations:

$$\sum_{\mu} c_{i\mu}(F_{\mu\nu} - \epsilon_i S_{\mu\nu}) = 0, \qquad i = 1, 2, \ldots n. \tag{1.7}$$

Each equation is specific to a single electron i, and μ and ν represent different basis functions. The validity of this method is subject to the constraint that the molecular orbitals remain orthonormal. This condition is expressed through the use of the overlap integral S_{ij},

$$S_{ij} = \int \phi_i(1)\phi_j(1)d\tau \tag{1.8}$$

where (1) indicates that both wavefunctions apply to the same electron, and where i and j refer to different molecular orbitals.

$F_{\mu\nu}$, the elements of the Fock matrix, are given by

$$F_{\mu\nu} = \int \phi_\mu^*(1)\hat{H}_{\mu\nu}^{core}\phi_\nu(1)d\tau + \sum_{i=1}^{N/2}\sum_{\lambda=1}^{L}\sum_{\sigma=1}^{L} c_{\lambda i}c_{\sigma i}(2J_{\mu\nu} - K_{\mu\nu}) \quad (1.9)$$

where λ and σ are also basis functions and the limit L refers to the number of basis functions present. $\hat{H}_{\mu\nu}^{core}$ is the one-electron Hamiltonian operator for the motion of a single electron moving in a field of bare nuclei. $J_{\mu\nu}$ and $K_{\mu\nu}$ are called the Coulomb (electron-electron repulsion) and exchange operators respectively. In their full forms they are expressed as

$$J_{\mu\nu} = \int \phi_\mu^*(1)\phi_\mu(1)\frac{1}{r_{12}}\phi_\nu^*(2)\phi_\nu(2)d\tau_1 d\tau_2 \quad (1.10)$$

and

$$K_{\mu\nu} = \int \phi_\mu(1)^*\phi_\nu(1)\frac{1}{r_{12}}\phi_\nu(2)^*\phi_\mu(2)d\tau_1 d\tau_2 \quad (1.11)$$

where (1) and (2) refer to different electrons and r_{12} is the distance between electrons 1 and 2.

The Roothann-Hall equations (Equation 1.8) can be conveniently written in matrix form as:

$$\mathbf{FC} = \mathbf{SCE} \quad (1.12)$$

$\mathbf{F}$ and $\mathbf{S}$ are the Fock and overlap matrices and contain the elements of $F_{\mu\nu}$ and $S_{\mu\nu}$ respectively. $\mathbf{C}$ and $\mathbf{E}$ are matrices of the coefficients $c_{\nu i}$ and the orbital energies, respectively. It is not possible to solve the Roothaan-Hall equations using basic matrix methods since the elements of the Fock matrix depend on the molecular orbital coefficients which also appear on the righthand side of the equation. Thus, there

is a need for an iterative method to solve them. The exact methods used are quite intricate and involve many matrix transformations. They are covered in many standard texts [8,12]. Since solving these equations involves the calculation and storage of many integrals, *ab initio* calculations make large demands on computational resources and are therefore not suitable for determining the energies of very large systems.

1.3.2 Basis sets

The choice of the basis set of atomic functions used to represent each MO (Equation 1.6) is an important aspect that needs to be considered in all quantum mechanical calculations. A good basis set allows accurate representation of molecular orbitals with a minimum of terms as well as fast evaluation of the $F_{\mu\nu}$ and $S_{\mu\nu}$ integrals.

There are two types of basis functions that are commonly used. The first of these represents orbitals using Slater-type functions whilst the other does so with Gaussian functions. Whilst the use of Slater-type functions gives very accurate results, they lead to the formation of integrals that are often very difficult and time-consuming to solve. This is especially true when orbitals are based on different atoms.

Therefore the use of Gaussian functions to represent atomic orbitals has become prevalent [13]. A Gaussian function has the following form

$$\phi = x^i y^j z^k e^{-\alpha r_a^2} \tag{1.13}$$

where r_a is the distance of the electron from the nucleus. α determines the radial extent of the basis function. The x, y, and z terms are in Cartesian space and govern the directionality of the function.

The use of Gaussian functions has the advantage that the product of two Gaussians can be expressed as a single Gaussian and also that, by combining different Gaussians, it is possible to produce functions with the same angular symmetry as each of the common atomic orbital types. However, this form of basis function also has its drawbacks. In

contrast to Slater-type functions, Gaussian functions do not have a cusp at the origin and decay towards zero more rapidly. Therefore, in order to be able to approach the level of accuracy achieved by Slater-type functions, each atomic orbital is normally represented by a linear combination of Gaussian functions.

$$\phi_\mu = \sum_{i=1}^{L} d_{i\mu} g_i(\alpha_{i\mu}) \tag{1.14}$$

where $d_{i\mu}$ is the coefficient of the Gaussian function g_i whose exponent is $\alpha_{i\mu}$. These linear combinations are known as contracted Gaussian functions (CGFs) and are said to be made up of a set of primitive Gaussian functions.

The terminology that quantum chemists use to name basis sets is not as confusing as it first may seem. Each name reflects the number of basis functions used to represent the orbitals in a given atom. A 'minimal' basis set consists of one contracted Gaussian function for each occupied orbital. The STO-3G basis set, for example, contains one CGF approximated by 3 Gaussian functions whose coefficients aren't varied for each orbital. This type of basis set does not offer much accuracy since the atomic orbitals cannot contract or expand in response to the surrounding molecular environment.

The use of double zeta basis sets overcomes this problem by representing each atomic orbital as a linear combination of a contracted and a diffuse Gaussian function. Therefore, during a calculation, this allows atomic orbitals to 'expand' and 'contract' in response to their environment. A further, and now more commonly used, alternative is to use split-valence basis sets where the number of functions used to describe valence orbitals is increased and only a single function is used to describe inner shell or core electrons. This approach is appropriate since the core orbitals of an atom do not greatly affect the chemical properties of molecules and tend not to vary from molecule to molecule. An example of a split-valence basis set is the 6-31G basis set [14]. In this basis set a single Gaussian function which is a contraction of six gaussian functions is used to describe the core orbitals. The valence shell orbitals are described by CGFs

with 3 primitive Gaussian functions for the 'contracted' part and a single primitive Gaussian function for the diffuse part.

Further improvements can be made to the quality of basis sets by the inclusion of d orbital-type functions to first row atoms and p orbital type functions to hydrogen atoms. These functions account for the polarisation of atomic orbitals in the presence of the nuclei of other atoms. For example, in the case of hydrogen atoms, the inclusion of p orbital type functions accounts for the distortion of 1s orbitals into orbitals of similar shape to sp hybrid orbitals. In a similar manner, d orbital type functions are added to p orbitals. The presence of these polarisation functions in a basis set is indicated by an asterisk ($*$) e.g. 6-31G*. A single asterisk indicates the presence of a d orbital-type function, and a second asterisk indicates that a p orbital-type function is also present.

A similar modification to basis sets is the addition of highly diffuse functions to a basis set to account for regions of electron density that reside a long way away from nuclei. These functions are particularly important when dealing with anionic species and species with lone pairs. Their presence in a basis set is denoted by the presence of a + sign, e.g. 6-31+G.

1.4 Density functional theory

The advent of Density Functional Theory (DFT) [15–19] provided an alternative method for accurately calculating the electronic properties of molecular systems. Whilst *ab initio* methods use wavefunctions to describe electronic structure, DFT uses electron density as the primary way of describing the system. The foundations of modern DFT were laid down in 1964 by Hohenberg and Kohn [15] who stated that the ground state energy and all the electronic properties of the ground state of a molecule are uniquely determined by its charge density $\rho(r)$.

Hence the ground state energy E_0 is a function of $\rho(r)$, and therefore, with knowledge of the ground state electron density, it is possible to calculate all the ground

state properties of a given nuclear configuration from $\rho(r)$. In a second theorem, Hohenberg and Kohn established a variational principle for the energy functional. That is, for a trial density $\tilde{\rho}(r)$ that conforms to the condition

$$\int \tilde{\rho}(r)\, dr = N \tag{1.15}$$

where N is the number of electrons in the system, the system energy will never be lower than the true energy

$$E_0 \leq E[\rho(r)] \tag{1.16}$$

It is therefore possible to search for the true ground state density in a manner analogous to that used for wavefunctions in the Hartree-Fock approach. However the Hohenberg and Kohn theorems do not give a method for calculating E_0 from $\tilde{\rho}(r)$ since the exact form of the energy functional is not known.

Kohn and Sham [20] invented an indirect approach to solving this problem which has allowed DFT to be used as a practical tool for electronic structure calculations. They showed that the ground state electronic energy $E_0[\rho]$ of an N electron system is given by

$$E_0[\rho] = V_{ex}[\rho] + T[\rho] + U[\rho] \tag{1.17}$$

where $V_{ex}[\rho]$ is the external potential due to the nuclei, and $T[\rho]$ and $U[\rho]$ are the kinetic energy of the electrons and the energy of interaction between the electrons in the system, respectively. Since they are both independent of $V_{ex}[\rho]$, Kohn and Sham grouped both $T[\rho]$ and $U[\rho]$ together as $F_{KS}[\rho]$, which is known as the Kohn-Sham operator for a given system.

$$E_0[\rho] = V_{ex}[\rho] + F_{KS}[\rho] = \int \rho(r)\, V_{ex}\, dr + F_{KS}[\rho] \tag{1.18}$$

The exact form of $F_{KS}[\rho]$ is not known and Kohn and Sham suggested that it could be divided into the following parts.

$$F_{KS}[\rho] = T_0[\rho] + J[\rho] + E_{XC}[\rho] \tag{1.19}$$

where $T_0[\rho]$ is the kinetic energy of the electrons in a system which has the same electron density as the real system but in which there are no electron-electron interactions. $J[\rho]$ is the electron-electron Coulombic energy, and $E_{XC}[\rho]$ is a term that includes all the contributions not included in the other terms which include electron exchange, electron correlation and kinetic energy correction terms.

Kohn and Sham were able to implement this approach by considering the density of the system $\rho(r)$ as the sum of the square of the moduli of a set of one-electron orthonormal orbitals known as Kohn-Sham orbitals.

$$\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \tag{1.20}$$

They also showed that the resolution of this expression in conjunction with appropriate variational conditions results in a set of one-electron equations known as Kohn-Sham equations [17, 20]. As in Hartree-Fock theory, these equations have to be solved iteratively until self-consistency is achieved.

The main drawback of using the the Kohn-Sham equations to determine ρ and E is that the exact form of the functional $E_{XC}[\rho]$ is not known for molecules. However, approximate functionals have been proposed. One way of estimating the value of the exchange functional is to use the local density approximation (LDA) [21]. If ρ varies very slowly with position, one can show, using the LDA, that $E_{XC}[\rho]$ is given by

$$E_{XC}^{LDA}[\rho] = \int \rho(r)\varepsilon_{XC}[\rho]dr \tag{1.21}$$

where $\varepsilon_{XC}[\rho]$ is the exchange plus correlation energy per electron in a homogeneous electron gas of electron density ρ. An accurate expression for $\varepsilon_{XC}[\rho]$ has been found [21]. The local spin density approximation (LSDA) [22] is a similar method in which different densities (ρ^{α} and ρ^{β}) are used for electrons with different spins.

However, in the case of molecules these are poor approximations to the true functionals since the ground state electron density of a molecule is far from being homogeneous. In order to gain a more accurate description of a molecule's electron density, correctional terms based on the gradient of the density are added. These corrections are called non-local gradient corrections, and they are added to both the exchange and correlation terms.

In the density functional studies outlined in this thesis, gradient corrections proposed by Becke (B88) [23] are used for the exchange term. For the correlation term the gradient corrected term of Lee, Yang and Parr [24] was employed. These are combined with a Hartree-Fock exchange term in what is known as the B3LYP functional [25].

Because density-functional calculations do not use the exact value of E_{xc} they are not, strictly speaking, *ab initio* calculations. A drawback is that the correlation effects cannot be explicitly estimated since they are implicitly included within the functional. A further disadvantage is that there is no systematic way to improve the functionals used by using more and more sophisticated terms since the parameters used are determined empirically.

Despite this, DFT have been found to yield results of a quality comparable to MP2 results for the ground state properties of a wide range of systems [19, 26]. Due to its relatively low computational cost, DFT is the method of choice for accurate quantum calculations on large systems, for which the inclusion of electron correlation by Møller-Plesset perturbation theory or configuration interaction methods is prohibitive.

1.5 Semi-empirical methods

Although *ab initio* and DFT methods give very accurate descriptions of the electronic structures of systems, the computational expense involved makes their use for large systems very demanding on computational resources. Semi-empirical methods [27, 28], whilst using the same SCF-MO framework, provide a far more rapid method of calculating electronic structure. In contrast to *ab initio* methods which attempt to rigorously evaluate all the integrals describing inter-electron interactions in a system, semi-empirical methods either replace these integrals with approximations or ignore them altogether. These approximations are usually based on experimental data.

One of the most time-saving approximations made in semi-empirical methods is that only valence electrons are considered. Core electrons are combined with the nucleus of the atom they are centred on to give a positive core. Most modern semi-empirical methods also employ the Neglect of Diatomic Differential Overlap (NDDO) convention [29] which involves ignoring differential overlap between atomic orbitals on different atomic centres. This eliminates the need for three- and four-centre integrals. Electronic structures and energies are calculated using this reduced set of integrals and the Roothaan-Hall equations [11].

There are a large number of semi-empirical methods in use today. The most popular of these are the MNDO [30], AM1 [31], and PM3 [32] methods. The MNDO/d [33,34] method is a newer method which includes functions which cater for the presence of d-orbitals. This improves the quality of results for metals and hypervalent compounds such as phosphate esters. A similar method, based on the AM1 method, but with the inclusion of *d* orbital functions for phosphorus as well as other parameters which are generated especially for phosphate ester hydrolysis reactions, is used in this thesis.

Overall, semi-empirical methods provide good descriptions of molecular energies and geometries with respect to their computational expense. Indeed in some cases, they can be more accurate than lower level *ab initio* methods since the

functions, parameterised against experimental results, implicitly account for exchange, correlation, and vibrational effects. However, it must be borne in mind that these methods are parameterised for molecules in the ground state at 298 K. Therefore due caution must be exercised when dealing with systems in states far from their ground state equilibria and results obtained should be compared to high-level *ab initio* results.

1.6 Geometry optimisation

The potential energy surface (PES) of an N-atom system is a function of many coordinates: $3N$ Cartesian coordinates or $3N - 6$ internal coordinates. Therefore, the number of configurations available to the system is enormous. A molecular system is in a stable conformation when the structure corresponds to a minimum on the PES. At a minimum, the gradient of the energy, E, with respect to the coordinates x is zero:

$$\frac{\partial E(x)}{\partial x} = 0 \tag{1.22}$$

Due to the complicated topology of molecular potential energy surfaces, it is very difficult to minimise potential energy functions analytically. Instead minima are located using numerical methods which gradually change the coordinates of the system in the direction of the energy gradient (Equation 1.22) until an energy minimum is obtained. The progress of the minimisation is guided by both the first and second derivatives of the energy with respect to the coordinates. The direction of the first derivative of the energy indicates the direction in which the minimum lies and the magnitude of the gradient indicates the steepness of the local slope. Second derivatives indicate the curvature of the function which can be used to predict where the function will change direction.

The potential energy function of a system at an arbitrary point x can be written as a Taylor series expansion about the point x_0,

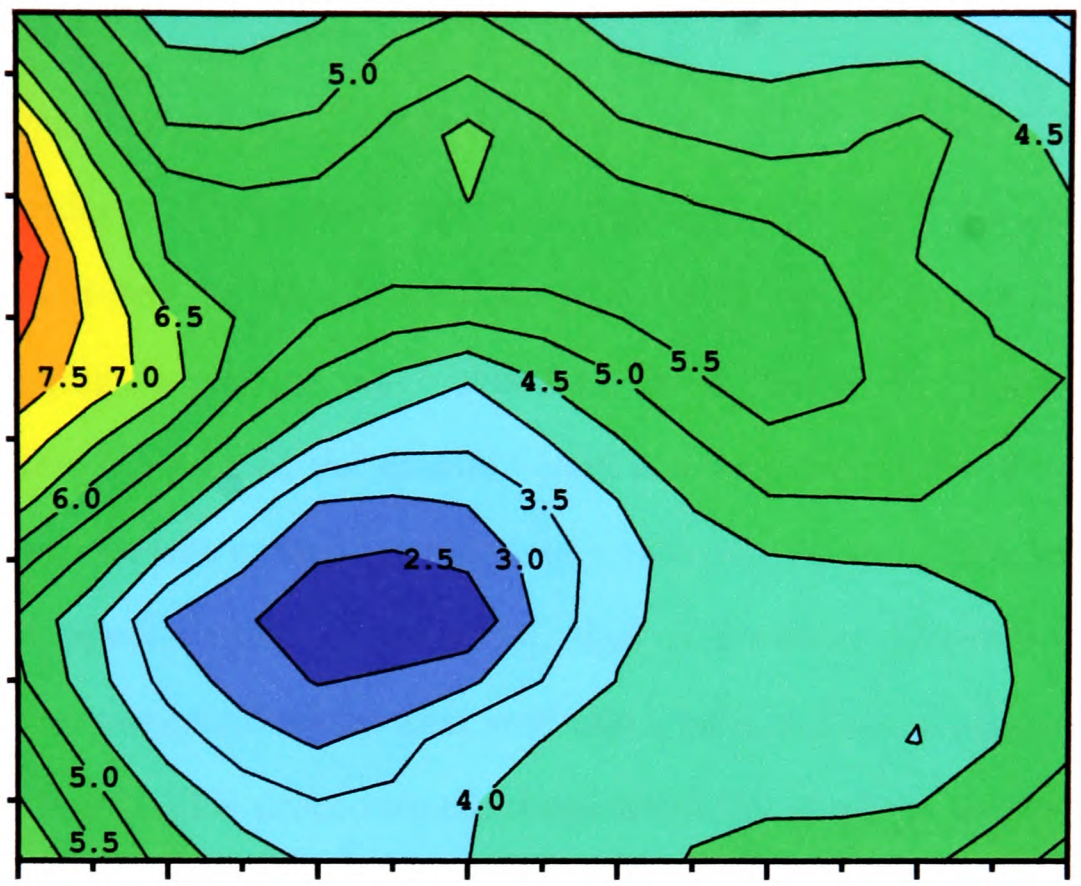


Figure 1.1: A schematic contour plot of a potential energy surface where the energy is dependent on two variables. These could be bond lengths, or bond angles or dihedral angles. The numbers on the diagram indicate the energies associated with particular contour lines.

$$V(x) = V(x_0) + (x - x_0)V'(x_0) + \frac{1}{2}(x - x_0)^2 V''(x_0) + \cdots \tag{1.23}$$

where x is a vector of $3N$ components and x_0 is the current configuration of the system. $V'(x_0)$ is a $3N \times 1$ matrix, each element of which is the partial derivative of V with respect to the appropriate coordinate.

$$V'(i) = \frac{\partial V}{\partial x_i} \tag{1.24}$$

Each element (i,j) of the matrix $V''(x_0)$ is the partial second derivative of the energy function with respect to the two coordinates x_i and x_j,

$$V''(ij) = \frac{\partial^2 V}{\partial x_i \partial x_j} \tag{1.25}$$

$V''(x_0)$ is therefore a $3N \times 3N$ matrix and is known as the Hessian or the force constant matrix.

The majority of minimisation methods in use today can be classified as being either first or second order methods, after the highest derivative used. First order methods, such as the steepest descent [35] or conjugate gradient [36] methods, are fast and simple methods for finding energy minima. These methods iteratively change the nuclear coordinates in the direction of the gradient until a minimum is arrived at.

In the steepest descent method, for example, each step in the minimisation is made in the direction opposite to the gradient, i.e. that of the force. If a step decreases the overall energy, the magnitude of the next step taken is increased. If not, the magnitude of the step is reduced. This procedure continues until the gradient stays below a given threshold. The steepest descent method has the advantage of rapidly decreasing the system energy at points far away from minima but it is found to converge poorly at points close to minima. It is therefore good for relieving bad contacts in systems or for use in the first few steps of long minimisation process. The conjugate gradient method is similar to the steepest descent method except that the history of the gradients from previous minimisation steps is also used to determine the size and direction of a step.

As their name implies, second order methods use not only the first derivative of the energy function but also the second derivative. One such method is the Newton-Raphson method [37]. By taking the first derivative of Equation 1.23 and setting $V'(x) = 0$, the Newton-Raphson step, which is used to determine the next position of the system, is obtained:

$$x - x_0 = -\frac{V'(x_0)}{V''(x_0)} \tag{1.26}$$

where x and x_0 are the next position and current positions of the system respectively. Near a minimum, where the potential is typically quadratic in shape, this minimisation algorithm allows faster convergence. It is, however, considerably more costly than first

order methods since each step necessitates the calculation and inversion of the Hessian matrix. It also encounters problems at points far away from minima when the forces are large but the curvature of the function can be small. This leads the algorithm to make a large step (a large gradient divided by the small curvature) which can lead to a point on the other side of the minimum which may be even further from it.

The adopted basis Newton Raphson (ABNR) algorithm [38], which is implemented in CHARMM [39], is a memory and time saving version of the Newton Raphson method. It avoids the costly calculation of analytical second derivatives by calculating a Hessian matrix for subsets of the coordinates using finite difference methods from the coordinates and gradients calculated in previous steps. The Hessian matrix for the other coordinates is calculated in the normal fashion.

1.7 Transition state location

A transition state (TS) can be defined as the highest energy point on the lowest energy pathway connecting two adjacent minima. It has the property that it is a maximum with respect to one of the degrees of freedom that makes up the potential energy surface a reaction whilst being a minimum with respect to all the other degrees of freedom. This is reflected by its Hessian matrix possessing one negative eigenvalue as opposed to the positive Hessian matrix associated with minima. Transition states are significant features of chemical reactions since the energy difference between the TS and local minima determines the probability or rate of the reaction proceeding from one stable structure to another.

There are various ways in which a TS can be located [40]. These range from a simple search or scan of the reaction coordinate to methods which optimise the potential function whilst keeping one of the Hessian's eigenvalues negative.

More accurate methods usually use a Newton Raphson algorithm or a derivative of it to increase the system energy along the eigenvector with the lowest eigenvalue in the

Hessian matrix if all the eigenvectors are positive. If there is one negative eigenvalue in the Hessian (normally in the quadratic region surrounding the saddle point), then the energy is lowered in the direction of that eigenvalue.

1.8 Normal mode analysis

In order to compare the results of theoretical calculations on a minimised structure with the equivalent experimental results, the thermal motions of the structure must be considered. Knowledge of these motions enables the calculation of thermodynamic quantities such as free energies, entropies, and zero point energy corrections. Whilst the rotational and translational contributions to the energy of a system are usually calculated in a straightforward manner using standard statistical mechanical formulae [41], the vibrational contributions are more difficult to ascertain since calculation of the vibrational frequencies within the system is required. These can be calculated using normal mode analysis.

Normal mode analysis treats the atoms within a system as pairs of harmonic oscillators. In order for this assumption to hold the atoms are treated as point masses and the PES surrounding the structure must be quadratic in form. Therefore, the method is only valid at stationary points. For an N-atom molecule with $3N$-6 vibrational frequencies, the constituent vibrational frequencies are calculated using the Hessian matrix, V'' (Equation 1.25). The Hessian matrix is converted into mass-weighted Cartesian coordinates so that the effect of the forces on atoms of different masses is accounted for. By diagonalising this matrix, its eigenvalues can be determined. Six of the eigenvalues, which correspond to rotational and translational motions, will be zero in value whilst the other $3N$-6 eigenvalues can be used to calculate the vibrational frequencies using the relationship

$$\nu_i = \frac{\lambda_i}{2\pi} \tag{1.27}$$

where λ_i is an eigenvalue of the mass-weighted Hessian matrix and ν_i is the corresponding vibrational frequency.

The values of the frequencies calculated give important insights into the nature of stationary point being studied. If they are all positive then the structure is a true minimum. The presence of a single negative frequency indicates the that the structure is a first order saddle point, i.e. a transition state. Therefore, normal mode analysis calculations are normally used to check the authenticity of optimised geometries determined by minimisations. It is also possible to visualise the vibration corresponding to the imaginary frequency using programs such as MOLDEN [42] in order to make sure that the vibration corresponds to an intuitive transition state for the reaction being studied. Normal mode analysis can also be used to determine the vibrational frequencies of proteins [43] where the lowest frequency vibrations correspond to large scale motions such as hinge-bending.

1.9 Determination of reaction energy profiles

It is often desirable to determine the complete energy profile for a reaction as opposed to just the energies of its stationary points. The enables one to calculate the curvature of the potential energy surface as well as molecular properties at points along the reaction energy profile. Two of the most popular methods of calculating reaction pathways are discussed below.

1.9.1 Reaction path following methods

One of the most accurate ways of determining a reaction energy profile is to use reaction path following methods. The first implementation of this method determined what is known as the reaction's intrinsic reaction coordinate (IRC) by starting at a transition state geometry and minimising in the direction of each of the local minima using a

steepest descent minimisation. The reaction path towards each minimum, described in mass-weighted coordinates, is determined by minimising along the eigenvector that corresponds to the transition state's imaginary frequency [44]. However, this method is not particularly accurate since the steepest descents minimisation gives a path that oscillates about the true reaction path. More rigorous approaches were developed [45–47] that involved the calculation of the reaction path using the gradient at each point calculated. The most popular of these methods, due to its implementation in the Gaussian series of programs [48], was developed by Gonzalez and Schlegel. In this method, steps of preordained size are taken along the reaction coordinate. Each step is taken in the direction of the gradient calculated at the previous point and each new point is minimised subject to the constraint that the distance between it and the last point is equal to the step size value. The reaction path is then approximated by a smooth curve that passes through all points. In general, IRC calculations give very accurate estimates of a reaction energy profile since they do not presume prior knowledge of the reaction coordinate and can implicitly cater for multi-dimensional reaction coordinates.

1.9.2 Adiabatic mapping methods

In the case of large chemical systems such as proteins, the determination of transition states using reaction path following techniques is very difficult. This is because the calculation of the Hessian matrix for systems this size is computationally very expensive. In these cases, the transition states and energy profiles of chemical processes can be determined by calculating the energy of the system at various values of the internal coordinates whose changes correspond to the chemical reaction or process being studied.

These methods are known as adiabatic mapping methods and presume knowledge of the reaction coordinate. The procedure used involves changing the chosen internal

coordinates from their initial values to their final ones in a stepwise fashion. At each step, the other internal coordinates of the system are allowed to relax using conventional minimisation methods. The appropriate internal coordinates are held using a restraint energy term which is usually harmonic. The RESD module in CHARMM [39] facilitates the use of a combination of such terms [49]. The adiabatic mapping method gives good results when the reaction coordinate is simple and corresponds to changes in a small number of internal coordinates.

For more complicated reaction coordinates such as protein folding landscapes pathway methods [50–53] are used instead. These methods attempt to construct a reaction pathway by starting from the end point geometries of the chemical process and minimising intermediate points, subject to constraints that prevent them from going back to the minima, until a minimum energy path is determined.

1.10 Calculation of atomic charge

Since atomic charges are not experimentally observable quantities and therefore cannot be calculated directly from electronic wavefunctions, approximate methods must be used to calculate them. Two of the most common approaches to calculating atomic charges involve the use of population analysis methods or fitting the charges to the molecular electrostatic potential.

Population analysis methods [8] calculate atomic charges by partitioning the system electron density among its nuclei leaving each atom in the system with a portion of the system electron density associated with it. The most commonly used population analysis method was developed by Mulliken [54]. In this method the electron population, q_μ, associated with a given atom is expressed as:

$$q_\mu = P_{\mu\mu} + \sum_{\nu \neq \mu} P_{\mu\nu} S_{\mu\nu} \tag{1.28}$$

where $S_{\mu\nu}$ is the overlap integral (Equation 1.8) and $P_{\mu\mu}$ and $P_{\mu\nu}$ are known as density matrices whose elements are

$$P_{\mu\mu} = 2\sum_{i=1}^{N/2} c_{\mu i} c_{\mu i} \tag{1.29}$$

and

$$P_{\mu\nu} = 2\sum_{i=1}^{N/2} c_{\mu i} c_{\nu i} \tag{1.30}$$

where N is the number of electrons in the system. The $P_{\mu\mu}$ term refers to the electrons directly associated with the basis function ϕ_μ while the $P_{\mu\nu}$ term refers to the portion of the overlap population between ϕ_μ and ϕ_ν that is associated with the electron in question. This is exactly half the sum of the relevant $P_{\mu\nu}S_{\mu\nu}$ and $P_{\nu\mu}S_{\nu\mu}$ terms, i.e. the electron density is exactly halved between the electrons. The total charge on an atom Q_A is then calculated by subtracting q_μ from the nuclear charge, Z_A, of the atom,

$$Q_A = Z_A - q_\mu \tag{1.31}$$

This method has numerous drawbacks. Firstly, the atomic charges are highly dependent on the basis set employed for the calculation. If some atoms are expressed by proportionally more basis functions, then the magnitude of the charges associated with these atoms will be exaggerated. Also the convention by which the overlap populations are halved between atoms fails to account for electronegativity differences between atoms which can cause the overlap population to reside closer to one atom than another. However, whilst not being useful for calculating absolute values, Mulliken population analysis is reasonably reliable for calculating charge differences between atoms and trends within a system.

A more reliable method of estimating atomic charges is by fitting them to the molecular electrostatic potential that is calculated during the SCF procedure. The most

commonly used method that utilises this approach is the CHELPG algorithm [55], a modification of the CHELP algorithm [56]. In the CHELPG method, the electrostatic potential is determined at a selected number of points around the molecule. These points lie on a cubic grid and are not within the van der Waals sphere of any atom nor more than 2.8 Å from its centre. A Lagrange multiplier method is then used to find the partial atomic charges that best reproduce the electrostatic potential. This is subject to the constraint that the sum of the charges is equal to the net charge on the molecule.

1.11 Molecular mechanics force fields

When modelling large molecular systems such as proteins or the atomistically solvated systems encountered in this thesis, the use of the electronic structure methods described earlier in this chapter becomes impossible due to the prohibitive computational expense involved. This is because quantum mechanical methods, when calculating the energy of the system, consider all the electrons in a system, which are normally an order of magnitude more abundant than the number of nuclei present in the system. Therefore, in order to enable the efficient determination of the energy of large systems, approximations must be made. One such approximation, used in semi-empirical methods [27], is to parameterise the behaviour of the electrons in a system. Despite the resultant reduction in computational expense, these methods are still unsuitable for large numbers of energy evaluations on sizeable molecular systems.

Molecular mechanics (MM) methods, known as force field methods, are ideal for modelling such systems. These methods employ a potential that does not explicitly account for the electron distribution within a system and instead merely considers the positions of the nuclei within the system. The resulting potential energy function, also known as a force field, is based on the interactions between the nuclei within the

system and takes into account molecular processes such as bond stretching, bond angle opening and closing, and non-bonded interactions between atoms.

There are a number of different MM potentials that are in common use including the CHARMM [39,57], AMBER [58], GROMOS [59], and OPLS [60] potentials.

1.11.1 The potential energy function

The overall potential energy function, V_{tot}, consists of a linear combination of independent terms each of which describes a particular type of internuclear interaction found within a system. A typical potential energy function has the form:

$$V_{tot} = V_{bond} + V_{angle} + V_{torsion} + V_{vdw} + V_{elec} \tag{1.32}$$

The terms on the righthand side of the equation describe the energetic contributions due to bond stretching, the opening and closing of bond angles, the hindered rotation of groups about a bond, as well as the van der Waals and electrostatic interactions within the system, respectively. Each of these terms is now described in more detail.

The bond stretching energy term, V_{bond}, is a summation of the energies of the bonded pairs of atoms in the system.

$$V_{bond} = \sum_{bonds} \frac{k_l}{2}(l - l_0)^2 \tag{1.33}$$

It imposes an energy penalty when a bond length, denoted by l, deviates from its reference value, l_0. The potential is a harmonic function whose force constant is k_l. V_{angle}, the angle bending term, has a similar form to the bond stretching term. It also imposes a penalty when a bond angle, θ, deviates from its reference value θ_0. In this case, k_θ is the constant for the restoring force,

$$V_{angle} = \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{1.34}$$

The third term in the potential energy function, $V_{torsion}$, describes how the energy of a molecule changes as it undergoes rotation about one of its bonds. This is also known as the torsional energy of the system. Unlike the previous two terms, a harmonic potential is not suitable here since, in many cases, the whole range of possible angles is accessible without substantial energy changes. Thus the energy function has a periodic form, expressed as a cosine series expansion, which is continuous through a range of values accessible to the angle. It has been found that just the first term of the expansion suffices,

$$V_{torsion} = \sum_{torsions} \frac{k_\phi}{2}(1 + cos(n\phi - \delta)) \tag{1.35}$$

where n is the periodicity of the angle, i.e. the number of energy peaks and wells in the potential, ϕ is the phase of the angle, and k_ϕ is the force constant. Another form of torsional potential that is sometimes present in molecular mechanics force fields is known as an improper torsion or an out-of-plane bending term. This term is used to account for the energy cost when atoms that are usually part of a planar arrangement bend out of plane.

The three foregoing terms (Equations 1.33, 1.34, 1.35) are known as the bonded terms of the potential since they refer to atoms that are bonded to each other. Of the three, the torsional term tends to have the biggest bearing on the energy of a system.

The final two terms of the potential energy function are known as the non-bonded terms as they evaluate the interactions between atoms that are not bonded to each other. The first of these, V_{vdw}, refers to the van der Waals interactions between non-bonded atoms in the system. This potential is a combination of attractive and repulsive terms. The repulsive term reflects internuclear electrostatic repulsions and repulsive exchange forces between electrons. This is expressed by a short-range inverse twelfth power term. The second term reflects the attractive interaction that is brought about by changes in the charge distribution about nuclei when they come within close

proximity of each other. This induces attractive dipole interactions between nuclei and is reflected in the potential by a longer range inverse sixth power attractive term. This potential is known as a Lennard-Jones potential [61]:

$$V_{vdw} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{1.36}$$

σ_{ij}, r_{ij}, and ε_{ij} represent the collision diameter (the interatomic distance for which the energy is zero), the interatomic distance, and the well depth, respectively.

The electrostatic interactions between atoms described by a Coulombic term, V_{elec}, which treats the atoms in question as point charges:

$$V_{elec} = \sum_{i=1}^{N} \sum_{j=1+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{1.37}$$

where q_i and q_j are the charges on atoms i and j, and ε_0 is the dielectric constant of the medium. This force, due to its inverse square nature, acts over a considerably longer range than the van der Waals interactions and thus can have an important bearing on the energy of the system.

1.11.2 Other features of force fields

An ideal MM force field should combine both simplicity and accuracy whilst still being transferable to systems other than those it was parameterised for. The various parameters that make up a force field are usually fitted to reproduce structural, thermodynamic, and spectroscopic data gained from experimental studies as well as the results of high level electronic structure calculations on model compounds. As mentioned, transferability is especially important since it is desirable to be able to model a large range of different molecular systems without having to reparameterise the force field in each case. The concept of atom types is also important. In a standard force field, a given atom is represented by more than one atom type. Each atom

type reflects the hybridisation and local environment of the atom and has accordingly adjusted parameters. Typically MM force fields are normally used in conjunction with the minimisation algorithms discussed in Section 1.6 or within molecular dynamics simulations. This latter technique is discussed in the next section.

1.12 Molecular dynamics simulation methods

Molecular dynamics (MD) simulations [62, 63] have become a major tool in computational chemistry and are used to simulate the time-dependent behaviour of many-body systems. In essence, an MD simulation involves the calculation of the system energy, which is then differentiated with respect to the atomic coordinates to obtain the forces acting on each molecule. These forces are then applied to the system over a suitable timestep through the resolution of Newton's second law, $F{=}ma$, to produce successive configurations of the system, a trajectory, which gives the evolution the system geometry in time. The trajectory can be considered as an ensemble of low energy conformations of the system from which average values of many of the system properties may be calculated.

1.12.1 Theory

In an N particle system where the position of each particle i of mass m_i is represented by the vector $\mathbf{r}_i$, the relationship between the velocity of the particle $\dot{\mathbf{r}}_i$, where the dot denotes differentiation with respect to time, and the momentum $\mathbf{p}_i$ is denoted by

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \tag{1.38}$$

Also, if $\mathbf{F}_i$ represents the net force exerted on the particle i by the rest of the particles in the system and is given by the derivative of the system's potential energy function V (Equation 1.32) with respect to the change in the atom's position then

$$\mathbf{F}_i = -\frac{dV}{d\mathbf{r}_i} \tag{1.39}$$

Since the equation of motion of a particle can be expressed as

$$\dot{\mathbf{p}}_i = \mathbf{F}_i \tag{1.40}$$

then by combining Equations 1.38, 1.39, and 1.40 the following expression is arrived at:

$$m_i \ddot{\mathbf{r}}_i = -\frac{dV}{d\mathbf{r}_i} \tag{1.41}$$

Thus, it is possible to calculate the new position of each atom in the system at the end of each time step. In order to propagate a trajectory for the system of particles Equation 1.41 must be integrated for all the particles in the system at each timestep of the simulation. This is not a trivial task since the motions of all the atoms in the system are coupled together and thus analytically solving the equations of motion is impossible. Therefore, these equations must be integrated using numerical methods. The algorithms employed to do this are described next.

1.12.2 Molecular dynamics algorithms

There are many algorithms available for solving the equations of motion for large molecular systems [64]. A good algorithm should be able to conserve energy and momentum, be computationally efficient, and should permit the use of a reasonably large time step. One of the most popular methods of solving the equations of motion for a large system is the Verlet integrator [65], as it conforms well to each of the above guidelines.

The Verlet integrator uses the atomic positions and accelerations at time t and the atomic positions, $\mathbf{r}_i(t - \delta t)$, from the previous step to calculate the atomic positions

at time $t + \delta t$, $\mathbf{r}_i(t + \delta t)$. By adding the expressions for $\mathbf{r}_i(t - \delta t)$ and $\mathbf{r}_i(t + \delta t)$ the following expression can be arrived at:

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \delta t^2 \ddot{\mathbf{r}}_i(t) \tag{1.42}$$

It should be noted that the velocities of the atoms are not required in the calculation of the new atomic positions by this particular algorithm. If knowledge of velocities is required, e.g. for calculation of temperature, they can be explicitly calculated using the formula

$$\dot{\mathbf{r}}_i(t) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)}{2\delta t} \tag{1.43}$$

Most MD software packages, including CHARMM [39], use a variety of methods based on the Verlet algorithm for propagation of atomic coordinates, which are designed to allow velocities to be calculated at each step. These include the 'leap-frog' Verlet [66] and the 'velocity Verlet' [67].

One of the biggest drawbacks of MD simulations is the computational expense involved. Therefore, algorithms that reduce the CPU time required for simulations are much sought after. One of the more obvious ways of doing this is to increase the time step δt. The extent to which this can be done is limited by the fact that the magnitude of the time step must be shorter than the time required by the highest frequency motion in the system. These motions often correspond to the motion of hydrogen atoms. By fixing the value of the internal coordinates corresponding to this motion, it is possible to increase the time step of a simulation. The SHAKE algorithm [68, 69] applies such constraints to the system in an iterative manner until all the constraints are satisfied. When used in conjunction with the Verlet algorithm, the SHAKE algorithm has been found to reduce computational expense significantly [69].

MD simulations can sample the conformations of a system from a variety of different ensembles, for example, the canonical ensemble for which the number of particles, the

volume and the temperature are constant. The ensemble, which is a collection of different configurations for the system, can be used to calculate the thermodynamic properties of the system since the configurations generated in a simulation will be distributed according to their Boltzmann probabilities. Other features such as radial distribution functions and time-dependent properties of the system can also be calculated from the trajectory.

In general, MD simulations are a very powerful tool in the computational chemist's arsenal which allow one to observe the dynamical properties of a system at a given temperature. This is in contrast to minimisation algorithms which produce a static representation of how the system would be at 0 K.

1.12.3 Calculation of free energy

The free energy of a chemical system can be thought of as the probability of finding the system in a given state. By extension, the free energy difference between two states gives the relative probability of the system being in one state as opposed to another. The free energy profile of a chemical reaction, also known as its potential of mean force (PMF) [70] describes the changes in the system free energy as a function of the reaction coordinate.

The PMF, $\mathcal{W}(\xi)$, along a given reaction coordinate, ξ can be defined in terms of the average distribution function $\langle\rho(\xi)\rangle$,

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - k_B T \ln \frac{\langle\rho(\xi)\rangle}{\langle\rho(\xi^*)\rangle} \tag{1.44}$$

where ξ^* and $\mathcal{W}(\xi^*)$ are arbitrary constants. The average distribution function along the reaction coordinate ξ is obtained from a Boltzmann weighted average,

$$\langle\rho(\xi)\rangle = \frac{\int dR\, \delta(\xi'[R] - \xi) \exp[-U(R)/k_B T]}{\int dR \exp[-U(R)/k_B T]} \tag{1.45}$$

where $U(R)$ is the total energy of the system as a function of the coordinates R and $\xi'[R]$ is a function depending on a number of the degrees of freedom of the system. It may depend on, for example, a combination of bond distances and angles.

At normal temperatures, the probability density of low energy states are much more heavily weighted than high-energy states. Therefore, it is difficult for a normal MD or Monte Carlo simulation to sample the entire reaction coordinate within the computer time that is normally available. This is especially true within the context of a chemical reaction, where large energy barriers exist along the reaction coordinate. A number of methods have been devised to overcome this problem by biasing the progress of a simulation so as to ensure that the entire reaction coordinate is sampled effectively. These include free energy perturbation [71], thermodynamic integration [72], and umbrella sampling [73,74] methods. The last of these methods is employed in Chapter 7 of this thesis to determine the PMF for the hydrolysis of dimethyl phosphate and therefore will be focussed on for the remainder of this section.

In order for the entire potential energy surface of a chemical reaction to be adequately sampled during the course of a simulation, the reaction potential function is altered so that unfavourable conformations can be sampled sufficiently. This is achieved by replacing the original potential function, $U(R)$, with a modified potential, $U'(R)$, which includes an additional biasing term, $w(\xi)$, that has a low value for the high energy regions of the reaction coordinate.

$$U'(R) = U(R) + w(\xi) \tag{1.46}$$

The biasing potential may take a number of forms. Ideally, the potential would be the negative of the PMF of the reaction being studied, but this would require knowledge of the free energy profile beforehand and therefore is rarely an option. In practice, it is easier to divide the reaction coordinate into a series of discrete regions or 'windows'. Each window is sampled in the presence of a biasing potential which confines the

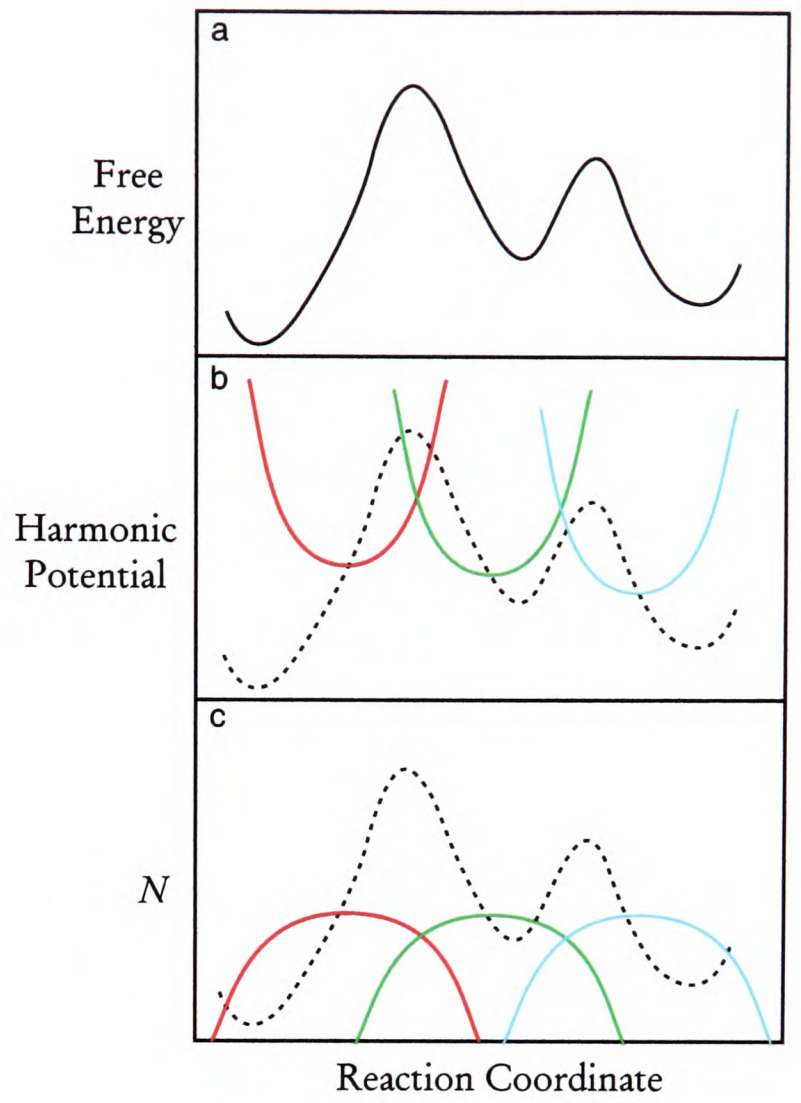


Figure 1.2: The use of umbrella sampling to obtain the PMF of a typical chemical reaction. (a) The potential energy surface of the reaction. (b) The implementation of the overlapping biasing potentials in order to sample the entire reaction coordinate. In practice, the reaction coordinate would be divided into more than three regions. (c) The resulting frequency distribution showing how the entire reaction coordinate has been well sampled.

simulation to a small region of phase space centred about a given value of ξ. This procedure is performed repeatedly along the reaction coordinate so that a series of overlapping windows span the entire potential energy surface of the reaction. This concept is illustrated in parts (a) and (b) of Figure 1.2.

The biasing potential used is normally harmonic and forces the value of the reaction coordinate in a given window, ξ, towards a reference value, ξ_{ref}, with a restoring force of k_{umb},

$$w(\xi) = \frac{1}{2} k_{umb} (\xi - \xi_{ref})^2 \tag{1.47}$$

This procedure of dividing the reaction coordinate into a series of windows yields a series of PMFs, each one of which samples the conformational space associated with a region of the potential energy surface very well. This ensures that energetically less favoured regions of the potential energy surface are now sampled well enough to ensure statistical reliability and that, overall, the entire potential energy surface is adequately sampled (Figure 1.2(c)). In order to obtain a single PMF for the entire reaction coordinate, all the individual PMFs must be joined together. The procedure for doing this is discussed later in this section. However, the inclusion of the biasing potential within the modified potential function means that the configurations sampled in each window no longer have appropriate Boltzmann weights thus making the calculation of free energies problematic. This means that each distribution has to be unbiased in order for its true PMF to be calculated.

Using a method devised by Torrie and Valleau [73, 74] it is possible to extract corresponding Boltzmann averages from a non-Boltzmann distribution. Using the relationship in Equation 1.45, the biased distribution function, $\langle \rho(\xi) \rangle_i$ for the ith window can be expressed as

$$\langle \rho(\xi) \rangle_i = \frac{\exp[-w_i(\xi)/k_B T] \, \langle \rho(\xi) \rangle}{\langle \exp[-w_i(\xi)/k_B T] \rangle} \tag{1.48}$$

The unbiased PMF for the ith window can then be expressed as

$$\mathcal{W}_i(\xi) = \mathcal{W}(\xi^*) - k_B T \, \ln \frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi)^* \rangle} - w_i(\xi) + F_i \tag{1.49}$$

where F_i represents the change in the free energy of the system due to the imposition of the biasing potential. This value depends on the ensemble average and cannot readily be calculated. Because of this, the individual PMFs must be matched empirically by adjusting the values of F_i for adjacent windows in the region in which they overlap until they match (Figure 1.3). The procedure can be done manually or by using a more

sophisticated method such as the weighted histogram analysis method (WHAM) [75]. The merits of the available methods have been compared in the literature [76].

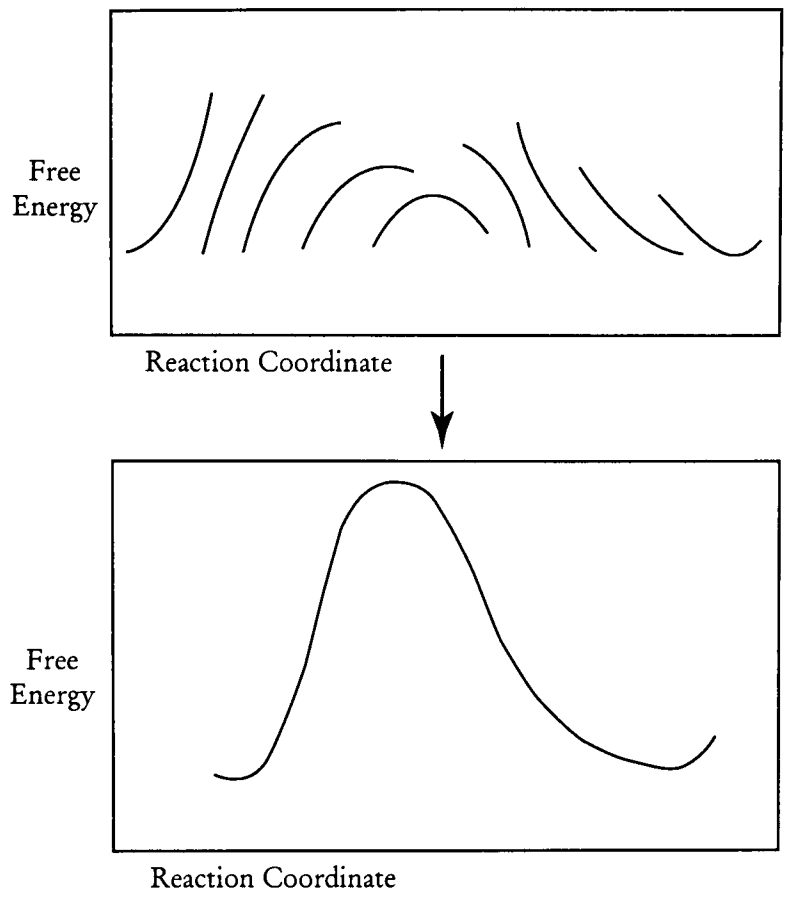


Figure 1.3: The construction of a single PMF from the series of PMFs obtained by a standard set of umbrella sampling simulations. These separate PMFs are joined together empirically to give the PMF for the entire reaction.

1.13 Solvent representations

So far this chapter has detailed a variety of methods that facilitate the calculation of a variety of molecular properties. Up to this point, no mention has been made of approaches used to model solvation effects. These effects are, of course, crucial when modelling a protein active site or a reaction in solution. This section summarises approaches involving the representation of solvation using both explicit and continuum models.

1.13.1 Explicit solvent models

The ideal way to model solvent is to represent each solvent molecule explicitly. To do so using a quantum mechanical potential would rapidly become intractable when one progresses beyond the first solvation shell of a substrate. Therefore, it is more common to model water molecules using a classical potential. To this end, specific models for water have been developed.

The most simple models in common use are the simple point charge (SPC) [77] model of Berendsen *et al.* and the three site interaction potential (TIP3P) model [78] of Jorgensen *et al.* Both these models represent each water molecule as having a rigid geometry, as well as a charge site on each of the three atoms and a single van der Waals interaction site on the oxygen atom. In both cases the parameters have been adjusted so that experimental properties of liquid water such as density, dipole moments, and radial distribution functions are reproduced as accurately as possible. A further modification of the TIP3P model is the TIP4P model which adds an extra interaction site by moving the charge from oxygen towards the hydrogen atoms along a line bisecting the H-O-H angle. Another interesting development to arise is the advent of fluctuating charge models [79] which allow charge transfer within each water molecule thus better modelling polarisation effects.

1.13.2 Continuum solvent models

Another alternative to modelling the effect of solvation is the use of continuum solvent representations. This approach is more commonly used when the solvent does not directly interact with the solute but instead alters its properties through electrostatic interactions. This approach can be used in conjunction with both classical and quantum mechanical methods. For the purposes of this discussion, the latter case will be focussed on.

One of the first such methods to be used is the Born solvation model [80, 81] which treats the solute as a point charge, q,within a spherical cavity of radius a. The surrounding solvent is represented by an electrical field and the effect of solvation is calculated as the work required to move the point charge from a vacuum to within the electrical field. This corresponds to the electrostatic contribution, ΔG_{elec}, to the solvation free energy and can be expressed as:

$$\Delta G_{elec} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon}\right) \tag{1.50}$$

where ϵ is the dielectric constant of the electric field. A further development of this method is the Onsager dipole method [82] which which treats the solute as a dipole. The solute dipole interacts with the electric field which, in turn, induces an electric field within the cavity popularly known as a reaction field. In electronic structure calculations, the reaction field acts as a perturbation to the gas phase Hamiltonian for the solute. Thus the effect of solvent is gauged by calculating the energy of the modified molecular wavefunction using standard methods. The most obvious weakness of the Born and Onsager models is approximation of the solute cavity as being spherical when, in fact, molecules are almost never spherical in shape.

The polarised continuum method (PCM) [83, 84] allows the solute cavity to adopt a more realistic shape that is based on the van der Waals radii of the atoms that make up the solute. The effect of the solvent is represented by the charge density of the surface of the solute cavity. To facilitate this, the cavity surface is divided into a series of discrete segments each with its own point charge. The electrostatic potential at each surface point consists of the potential due to the solvent and the potential due to other surface charges. The potential from the cavity surface is then added to the solute Hamiltonian, as in the Onsager method, and a standard SCF calculation is performed. This procedure continues, with the solute wavefunction modifying the charges on the cavity surface and they in turn modifying the solute Hamiltonian until both the solute

wavefunction and the surface charges achieve self-consistency. A further modification of the PCM method is the self-consistent isodensity (SCI-PCM) [85] method in which the solute cavity is based on an isosurface of the total electron density. This makes the cavity surface smoother in shape and thus decreases the CPU time required to integrate it when calculating the reaction field.

1.14 Summary

Many of foregoing methods described have been used to obtain the results described in Chapters 4, 5, 6, and 7 of this thesis. Although it is clear that these methods have limitations, they represent a powerful arsenal with which the elusive details of phosphate ester hydrolysis reactions can be elucidated. DFT calculations were used to examine model reactions in the gas phase and also in conjunction with a continuum solvent representation. In order to be able to combine a quantum mechanical description of the reactive species with an atomistic solvent representation a combination of these methods was called for. The next chapter describes the theory behind hybrid quantum mechanical and molecular mechanical potentials and reviews some of their most powerful applications.

Chapter 2

Computer Simulation of Chemical Reactions with QM/MM Methods

2.1 Introduction

Chapter 1 introduced the wide range of powerful computational methods available to theoretical chemists today. It was seen that *ab initio*, DFT and semi-empirical methods are very well suited to studying chemical reactions in the gas phase but are unsuitable for larger systems due to the high computational cost involved. Similarly, it was noted that classical force field methods enable one to simulate very large molecular systems at a relatively low cost. This chapter reviews the theory and application of combined quantum mechanical and molecular mechanical (QM/MM) potentials [86,87]. As their name implies, these hybrid potentials combine both quantum mechanical and classical potentials and are particularly powerful when used to model chemical reactions in enzymatic and explicitly solvated systems.

After briefly reviewing the difficulties encountered in determining the mechanism of enzymatic reactions, the chapter explains how a hybrid QM/MM potential is constructed as well as the potential problems that must be avoided. The chapter

then concludes with a review of a varied set of QM/MM studies conducted by other authors.

2.2 Simulation of enzymatic reactions

Enzymes are phenomenally proficient molecular machines that are able to catalyse chemical reactions many orders of magnitude faster than the corresponding non-catalyzed reactions in solution [88, 89]. How enzymes achieve these large rate enhancements is a matter of continued research [90]. Recently our understanding of the function of enzymes has been greatly increased by advances in structural biology. X-ray crystallography [91] studies have produced atomic resolution structures of many enzymes. From the structures of these enzymes it is possible to infer mechanisms from the orientation of amino-acids at the active site of the enzyme. In some cases, using the technique of Laue diffraction [92], it is possible to study an enzymatic reaction crystallographically. This technique is, however, far from being widely applicable. Despite having a detailed three dimensional view of the active sites of many enzymes, it is often not possible to unambiguously predict one reaction mechanism over another on the basis of the structure. Other experimental methods such as kinetic studies [93] or site-directed mutagenesis [90] can augment the information available, but currently it is not possible to study the entire energy profile of a reaction in the active site of an enzyme with experimental techniques.

Computer simulation techniques offer the ability to study the potential energy surfaces of chemical reactions to a high degree of quantitative accuracy [94]. Theoretical studies of chemical reactions in the gas phase is a major field and can provide detailed insights into a variety of processes of fundamental interest. In the past decade theoretical methods have been extended to study reaction processes in mesoscopic systems such as enzymatic reactions in solution, albeit to a more approximate level when compared to the most accurate gas phase studies.

The best computational approach to studying chemical reactions involves the use of quantum mechanics. However, in practice, the size of typical enzymatic systems precludes the use of traditional quantum mechanical methods [95]. A typical simulation of a solvated enzyme could include approximately 10000 atoms, which makes the simulation using quantum mechanics intractable even with current computing facilities. A simplification of the problem can be achieved by only considering the substrate and selected active site residues, while neglecting the remainder of the enzyme and solvent. However, the remainder of the enzyme and solvent, while not having a covalent effect on the reaction, can significantly alter the energetics of the system through nonbonded interactions and therefore need to be included in the simulation.

2.3 Methodology

The emergence of hybrid quantum mechanical-molecular mechanical (QM/MM) methods in recent years addresses this problem. The first implementation of these methods was by Warshel and Levitt [86] in 1976. The method entails the division of the system being studied into a small region that is treated quantum mechanically, with the remainder of the system treated with computationally less expensive classical methods. The quantum region includes all the atoms that are directly involved in the reaction being studied and the remainder of the system, believed to change little during the reaction, is treated with a molecular mechanics force field [63]. The atoms in each region influence those in the other region through a coupled potential that involves electrostatic and van der Waals interactions [87, 96–105]. The partition between the two region is made in such as way as to place the part of the system which undergoes electronic changes during the reaction in the quantum region and to leave the remainder of the system, which only influences the reaction through electrostatic interactions, to be modelled using a classical force field.

Several molecular mechanics programs have been adapted to perform hybrid QM/MM simulations. In the majority of the implementations the quantum region has been treated either by using empirical valence bond (EVB) methods [106] or with a semi-empirical method (usually AM1 [31]). These implementations have been applied to study solvation [99,107]; condensed phase spectroscopy [108]; conformational flexibility [109], and chemical reactivity in solution [110, 111], enzymes [105, 112–122], and DNA [123].

Although semi-empirical methods have the advantage of being computationally inexpensive, they have a number of limitations [124–126], the most noteworthy of which concern the accuracy and reliability of these methods. In general, they are less accurate than high level methods, and since they have been parameterised to reproduce the ground state properties of molecules, they are often not well suited to studying chemical reactions. A further disadvantage of the semi-empirical methods is the limited range of elements for which parameters have been determined.

To overcome these limitations, the hybrid QM/MM potential can employ *ab initio* or density functional methods in the quantum region. Both of these methods can ensure a higher quantitative accuracy and the density functional methods offer a computationally less expensive procedure for including electron correlation [95]. Several research groups have reported the development of QM/MM programs that employ *ab initio* [49, 96, 97, 100, 103] or density functional methods [97, 127–130].

The focus of the remainder of this chapter is on the implementation and application of QM/MM methods for studying chemical reactions. The application of QM/MM methods to study solution phase reactions has been extensively reviewed by Gao [131, 132] and is also the subject of Chapters 6 and 7 of this thesis and will therefore not be reviewed here. Similarly, empirical valence bond methods, which have been successfully applied to studying enzymatic reactions by Warshel, Åqvist and co-workers [106, 133], will not be covered in this chapter.

In the combined QM/MM methodology the system being studied is partitioned into a quantum mechanical region and a molecular mechanical region (Figure 2.1). The quantum region will normally include the substrate, sidechains of residues believed to be involved in the reaction and any cofactors.

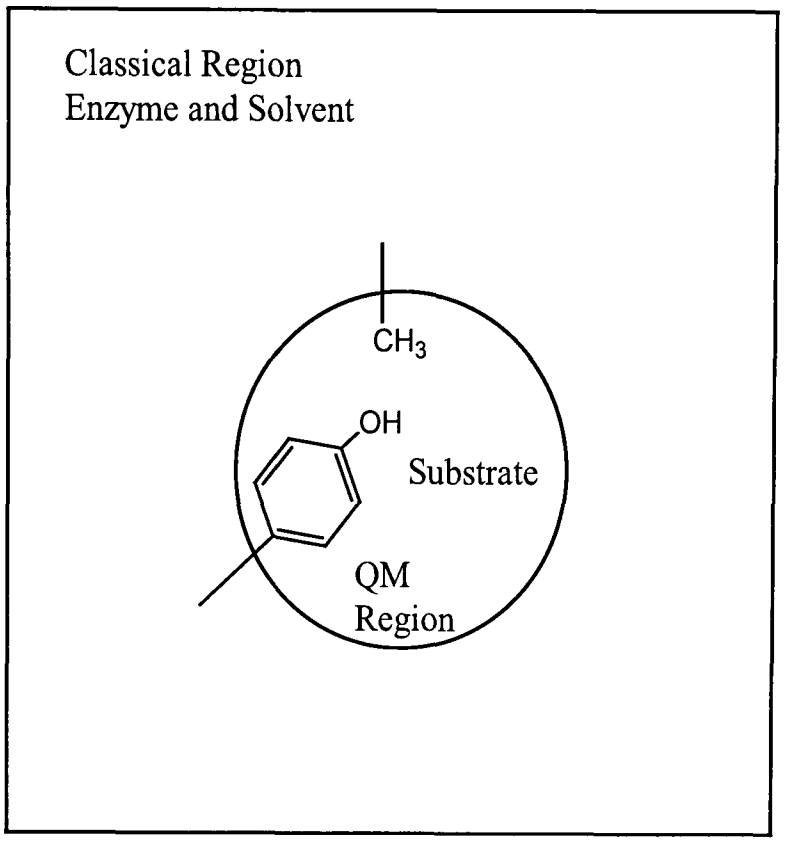


Figure 2.1: Schematic diagram depicting the partitioning of an enzymatic system into quantum and classical regions. The sidechains of a tyrosine and valine are treated quantum mechanically with the remainder of the enzyme and added solvent modelled by a classical force field.

The remainder of the protein and solvent is included in the molecular mechanics region. For the QM region, the wavefunction of the system of N electrons, Ψ, is a Slater determinant of one electron molecular orbitals, ψ_i, (or Kohn-Sham orbitals in the case of density functional theory [16]),

$$\Psi(r, R_q, R_c) = |\psi_1\alpha(1)\,\psi_1\beta(2)\ldots\psi_N\beta(2N)| \tag{2.1}$$

where α and β refer to spin eigenfunctions, r the coordinates of the electrons, and R_q and R_c the coordinates of the QM nuclei and the atoms in the MM region respectively.

The total energy of the system is evaluated by solving the Schrödinger equation with an effective Hamiltonian $\hat{H}_{\text{eff}}$ for the system:

$$\hat{H}_{\text{eff}}\Psi\left(r, R_q, R_c\right) = E\left(R_q, R_c\right)\Psi\left(r, R_q, R_c\right) \tag{2.2}$$

For QM/MM methods it is assumed that the effective Hamiltonian can be partitioned into quantum and classical components by writing,

$$\hat{H}_{\text{eff}} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \tag{2.3}$$

where $\hat{H}_{QM}$ is the pure quantum Hamiltonian, $\hat{H}_{MM}$ is the classical Hamiltonian and $\hat{H}_{QM/MM}$ is the hybrid QM/MM Hamiltonian. From equations 2.2 and 2.3 the energy of the system is given as:

$$E(R_q, R_c) = \frac{\int \Psi^* \hat{H}_{QM}\Psi d\tau + \int \Psi^* \hat{H}_{QM/MM}\Psi d\tau}{\int \Psi^*\Psi d\tau} + E_{MM} \tag{2.4}$$

$$= E_{QM} + E_{QM/MM} + E_{MM} \tag{2.5}$$

To date, QM/MM applications have employed density functional methods, *ab initio* or semi-empirical methods in the quantum region. The energy terms evaluated in these methods are generally similar but there are specific differences. The relevant equations for the density functional based methods are described first, and this is followed by a description of the specific differences associated with the other methods.

For density functional based QM/MM methods the electronic energy terms depend explicitly on the electron density, $\rho\left(r\right)$, of the atoms in the quantum region [16]. The electron density is determined by solving the one-electron Kohn-Sham equations [16, 20]:

$$\hat{H}_{DF}\psi_i(r) = e_i\psi_i(r) \qquad i = 1, \ldots, n \tag{2.6}$$

where e_i are the eigenvalues associated with the Kohn-Sham orbitals. The quantum Hamiltonian $\hat{H}_{DF}$ is given by[1]

$$\hat{H}_{DF} = -\sum_i \frac{\hbar^2}{2m_i}\nabla_i^2 - \sum_{i,q}\frac{Z_q}{r_{iq}} + \sum_{q<q'}\frac{Z_q Z_{q'}}{R_{qq'}} + \int \frac{\rho(r')}{|r-r'|}dr' + \frac{\partial E_{XC}}{\partial \rho(r)} \tag{2.7}$$

In this equation E_{XC} is the exchange correlation functional [16], Z_q is the nuclear charge of an atom in the quantum region, r_{iq} is the distance between an electron and quantum atom q; $R_{qq'}$ is the distance between two quantum nuclei, and r' is the coordinate of a second electron. The QM/MM interaction Hamiltonian, $\hat{H}_{DF/MM}$, is given by:

$$\hat{H}_{DF/MM} = -\sum_{i,c}\frac{q_c}{r_{ic}} + \sum_{q,c}\frac{Z_q q_c}{R_{qc}} + \sum_{q,c} V_{qc} \tag{2.8}$$

where V_{qc} is the van der Waals interaction energy between the quantum and classical regions, q_c is the partial charge of an atom in the classical region, r_{ic} is the distance between an electron and a classical atom c. This is expressed in terms of a Lennard-Jones potential which can be expressed as:

$$V_{qc} = 4\varepsilon_{qc}\left[\left(\frac{\sigma_{qc}}{R_{qc}}\right)^{12} - \left(\frac{\sigma_{qc}}{R_{qc}}\right)^{6}\right] \tag{2.9}$$

where ε_{qc} and σ_{qc} are empirical parameters and how these terms are obtained is discussed later in this section. This term is essential to obtain correct geometries since there is no Pauli repulsion between quantum and classical atoms.

[1] In this section, for ease of legibility, the abbreviation ∇_i^2 is used to denote $\frac{\partial}{\partial x_i^2} + \frac{\partial}{\partial y_i^2} + \frac{\partial}{\partial z_i^2}$

The molecular mechanics Hamiltonian, $\hat{H}_{MM}$, is calculated with the standard potential energy term from, for example, the CHARMM [39], AMBER [134], or GROMOS [135] programs.

The implementation of the method using *ab initio* methods for the quantum region is straightforward. The analogous equations for the electronic Hamiltonian and the corresponding energies in this case are [8]:

$$\hat{H}_{QM} = \frac{1}{2}\sum_{\mu}\sum_{\nu} P_{\mu\nu}\left(H_{\mu\nu}^{core} + F_{\mu\nu}\right) + \sum_{\alpha}\sum_{\beta>\alpha}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \tag{2.10}$$

where $H_{\mu\nu}^{core}$ is the core Hamiltonian. The indices μ and ν refer to the basis set orbitals ϕ, and $P_{\mu\nu}$ and $F_{\mu\nu}$ are elements of the density and Fock matrices. The composition of the QM/MM Hamiltonian is identical to equivalent DF/MM term given in Equation 2.8.

For QM/MM methods employing semi-empirical methods in the quantum region, the implementation is not as straightforward. In semi-empirical methods such as MNDO [30], AM1 [31], or PM3 [124], a minimal basis set representation of the valence levels of each atom is employed. Thus, for example, for first row main group elements, the basis set comprises just four atomic orbitals, namely an s and three p orbitals. The remainder of the electrons in the atoms is grouped into core terms. Therefore, the added complication for QM/MM methods involving semi-empirical Hamiltonians, is how to treat the interactions between the core terms on the quantum atoms and the partial charges on atoms in the classical region. The treatment that is currently used follows that suggested by Field *et al.* [87], where the interaction between the semi-empirical cores and the classical atoms are treated in the same manner as pure quantum interactions with the atoms in the classical regions carrying an s orbital for this purpose. For AM1 or PM3 the interaction between the classical atoms and the cores of the quantum atoms is given as [31]:

$$E^{core}_{QM/MM} = q_c Z_q \left(s_q s_q | s_c s_c\right) \left[1 + exp\left(-\alpha_c R_{qc}\right) + exp\left(-\alpha_q R_{qc}\right)\right]$$
$$+ \frac{q_c Z_q}{R_{qc}} \sum_{i=1}^{4} \left[K^c_i exp(-L^m_i (R_{qc} - M^m_i)^2) + K^q_i exp(-L^s_i (R_{qc} - M^q_i)^2)\right] \qquad (2.11)$$

where $(s_q s_q | s_c s_c)$ represents the one electron integrals for the interaction between the s orbitals on the quantum atoms s_q and the pseudo s-orbitals s_c on the classical atoms.

From this equation, it can be seen that parameters have been introduced to the QM/MM method: K^c_i, L^c_i, M^c_i, and α_c corresponding to the pseudo-s orbital on the classical atom. These parameters are atom-dependant and can be optimized to reproduce experimental or high-level theoretical data. Field *et al.* [87] performed extensive investigations of the values of these extra parameters and suggested that the parameters K^c_i, L^c_i and M^c_i $(i = 1 \ldots 4)$ can be set to zero, and that α_c should take a value of 5.0. These are generally the values used in most current QM/MM implementations that employ semi-empirical methods in the quantum region.

Finally, the parameterisation of the van der Waals part of the QM/MM interaction must be considered. This applies to all QM/MM implementations irrespective of the quantum method being employed. From Equation 2.9 it can be seen that each quantum atom needs to have two Lennard-Jones parameters associated with it in order to have a van der Waals interaction with classical atoms. Generally, there are two approaches to this problem. The first is to derive a set of parameters, ε_q and σ_q for each common atom type and then to use this standard set for any system that is being studied using a QM/MM potential. This is the most common approach and the 'derived' Lennard-Jones parameters for the quantum atoms are simply the parameters found in the MM force-field for the analogous classical atom types. For example, a study that employed a QM/MM method implemented in the program CHARMM [39] would

use the appropriate Lennard-Jones parameters of the CHARMM force-field [57] for the atoms in the quantum region.

The second approach is to derive Lennard-Jones parameters for the quantum atoms that are specific to the system being studied. This is a less common approach, but has been shown to improve the quantitative accuracy of the QM/MM approach in specific cases [136, 137]. The disadvantage of this approach however, is the necessity to derive Lennard-Jones parameters for the quantum region for every different study. Since the derivation of Lennard-Jones parameters is not a trivial exercise, this method of finding van der Waals parameters for the QM/MM interaction term is the less popular of the two.

2.3.1 The quantum-classical boundary

In the previous section, the essentials of the implementation of QM/MM methods were presented. For studies on systems where there is a discrete natural boundary between the quantum and classical regions, a QM/MM potential can be employed as outlined above, without any modifications. An example of such a system might be the study of a chemical reactions in water, such as the S_N2 substitution of CH_3Cl by Cl^- or the phosphate ester hydrolysis reaction described later in this thesis. In the first of the two reactions, the quantum region could comprise chloroform and the chloride ion with the solvent being treated by classical methods [110].

For QM/MM studies on the active sites of enzymes there is no natural boundary between the quantum and the classical regions. An example of this would be a situation where the quantum region might consist of a substrate and the side-chains of a number of active-site residues. The active site residues thus span both the quantum and classical regions, where the side-chain is in the quantum region and the main-chain atoms are in the classical region (Figure 2.2). The boundary between the quantum and classical regions falls across the covalent bond between the alpha and beta carbons

of the residue. This is an extra complication for QM/MM implementations and it is necessary to devise a method that can model the interface between the quantum and classical regions properly.

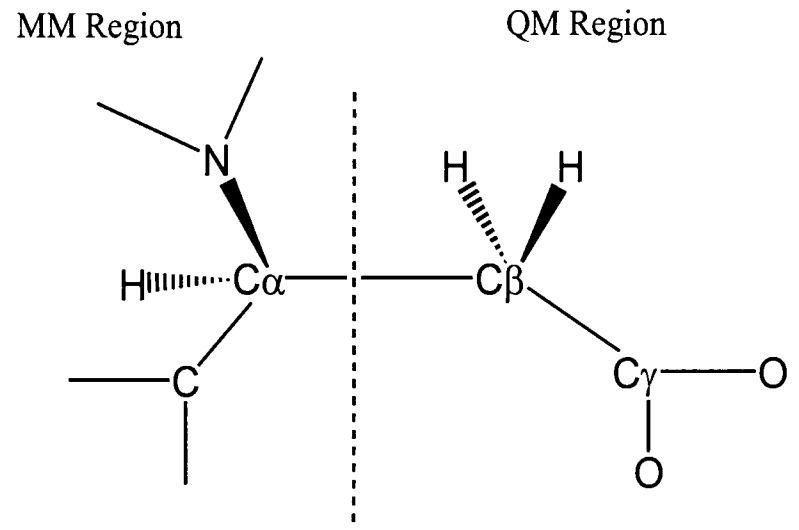


Figure 2.2: A aspartate sidechain partitioned into quantum and classical regions. The terminal $CH_2CO_2^-$ group lies within the quantum region and the backbone atoms are treated using a molecular mechanics force field.

Since the electrons on atoms in the classical region are not treated explicitly a method has to be devised that allows the electron density along the QM/MM bonds to be terminated satisfactorily. This is to ensure that the electronic structure of the quantum region is preserved and the valence requirements of all the atoms within are met. There are two common approaches that have been proposed to deal with this situation.

The first of these is termed the 'link atom' approach. In this approach a 'dummy' atom, called a link atom, is introduced to the quantum system at the boundary between the quantum and classical regions. This link atom, which is bonded to QM region atoms that are bonded to classically treated atoms, serves the purpose of satisfying the truncation of the electron density in the quantum region and is not seen by the atoms in the classical region. There are two common implementations of the link atom approach. The first of these was used by Field *et al.* [87] in the original CHARMM implementation of QM/MM methods. In this case, the QM/MM interaction term, $E_{QM/MM}$, is evaluated with all interactions between link atoms and

MM atoms discounted. As usual, each of the QM atoms interact with all of the MM atoms.

In recent *ab initio* [49] and density functional studies [97], a slightly different implementation of the link atom approach is employed. Here, the link atom is treated as a proper quantum atom and, as such, interacts with the charges of the atoms in the classical region. The charges on the classical functional group or atom that is replaced by the link atom do not contribute to the one-electron integrals in the SCF calculation for the quantum region. For example, if we were to consider a QM/MM study of CH_3CH_2OH with the OH moiety treated by quantum mechanics and CH_3CH_2 treated by molecular mechanics as an example, the charges on the CH_2 group are not included in the QM Hamiltonian; all other partial charges on classical atoms, i.e. from the CH_3 group are included. The link atom is initially positioned 1Å along the original bond but is not constrained during the simulations. The latter of the two approaches has been shown to give more reliable results since the former leads to the QM atom bonded to the link atom experiencing the charges of the MM region atoms whilst the link atoms do not. This often leads to a hyperpolarisation of the bond between the QM atom and the link atom.

The location of the quantum-classical boundary across a covalent bond also has implications for the energy terms evaluated in the E_{MM} term. All classical energy terms that involve only quantum atoms are not evaluated. These are accounted for by the quantum Hamiltonian. Classical energy terms that include at least one classical atom are evaluated. Referring to Figure 2.2, the Cα-Cβ bond term; the N-Cα-Cβ, C-Cα-Cβ, Hα-Cα-Cβ, Cα-Cb-Hβ_1 and Cα-Cb-Hβ_2 angle terms; and the proper dihedral terms involving a classical atom are all included.

An alternative approach to the link atom method is to use the frozen orbital approach developed by Rivail and co-workers [138]. This is also known as the local self-consistent field (LSCF) approach. The continuity of the electron density at

the boundary region is maintained by a frozen orbital along the bond between the quantum and classical atoms. This frozen orbital is derived from calculations on model compounds, based on the assumption that the orbitals from model compounds are transferable to the enzymatic system. A more generalised form of this implementation has been presented by Gao *et al.* [139] where a set of hybrid orbitals are used at the boundary region. This method is termed the generalised hybrid orbital (GHO) method by the authors. The set is divided into auxiliary and active orbitals and act as a basis set for the boundary atoms of the MM fragment. The active orbitals are optimised along with other orbitals on the QM atoms in the SCF cycle. In essence this method is an expansion of the approach of Rivail but has the advantage that the orbitals do not need to be parameterised for each specific problem.

There is concern in the field about current approaches for treating the boundary region in QM/MM calculations. As mentioned, the original implementation of the link atom approach [87] results in arbitrary charge polarisation in the QM region and the development of unrealistic partial charges on the link atom [140]. The implementation of the LSCF and GHO methods appears to improve the treatment of the boundary region compared to similar calculations using the original link atom approach. However, the more recent implementation of the link atom method does not appear to introduce unphysical electrostatic perturbations to the QM region, and in a limited number of test cases has been found to give very reasonable results for geometries, energetics and charge distributions when compared with pure quantum calculations [97, 140].

2.4 Applications

In this section the applications of QM/MM methods to three enzymatic reactions are presented with the intention of illustrating the techniques that are employed to simulate enzyme catalysis. The specific enzymes chosen cover applications involving both semi-empirical and *ab initio* methods in the quantum region.

2.4.1 Triosephosphate Isomerase.

The chemical reaction catalyzed by Triosephosphate Isomerase (TIM) was the first application of the QM/MM method in CHARMM to the study of enzyme catalysis [112]. The study calculated an energy pathway for the reaction in the enzyme and decomposed the energetics into specific contributions from each of the residues of the enzyme. TIM catalyses the interconversion of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (GAP) as part of the glycolytic pathway. Extensive experimental studies have been performed on TIM and it has been proposed that Glu-165 acts as a base for deprotonation of DHAP, and that His-95 acts as an acid to protonate the carbonyl oxygen of DHAP, forming an enediolate (see Figure 2.3) [141].

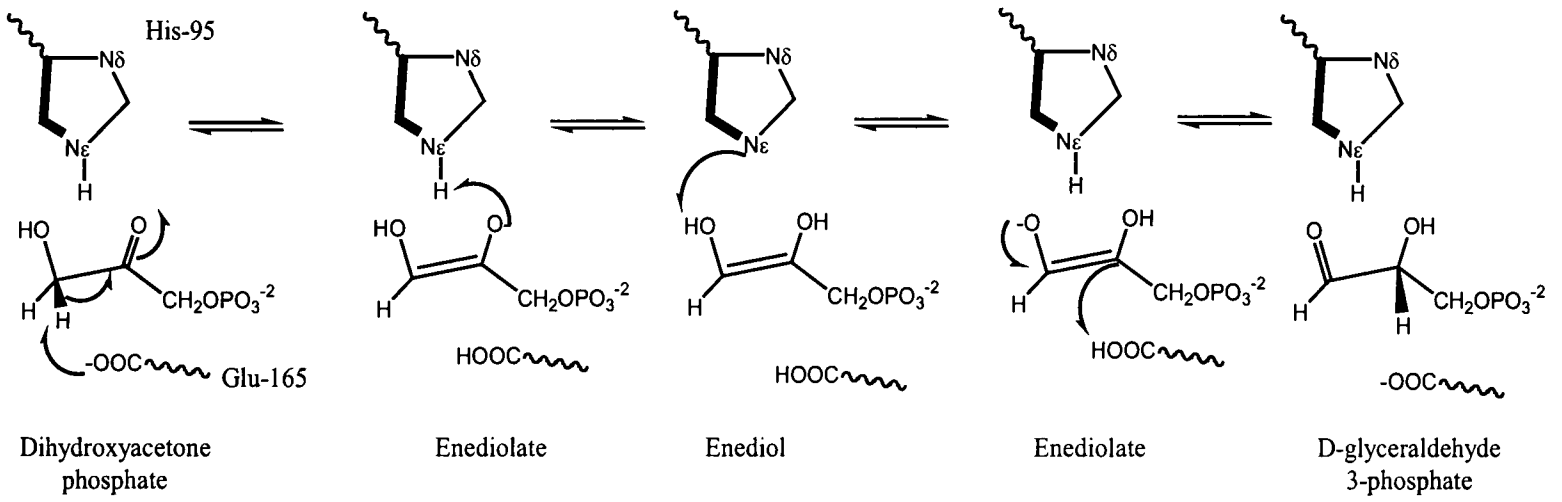


Figure 2.3: A possible mechanism for the isomerisation of dihyroxyacetone phosphate (DHAP) to D-glyceraldehyde 3-phosphate (GAP) by the enzyme triosephosphate isomerase (TIM). The general acid (Glu-165) and general base (His-95) are shown.

Bash *et al.* used QM/MM methods to follow the reaction path in TIM using a simulation protocol that has been employed in many subsequent simulations of enzyme catalysis with QM/MM methods. Firstly, the initial coordinates used for the calculations came from a high resolution X-ray crystal structure of TIM [142]. This is an important point since the validity of the results of a simulation of this nature on an enzyme is questionable in the absence of an accurate structure from crystallographic studies. The authors then chose to include only residues within a 16 Å radius sphere of the active site, with the remainder of the enzyme deleted. This saves on computation

time whilst still maintaining the bulk of the protein around the active site. In any given study of enzyme catalysis it is important to evaluate the effect on the results of different sized model systems (i.e. determine the radius of the sphere to include in the study) in order to prevent any size dependency being introduced to the study. Subsequently, Bash *et al.* surrounded the truncated protein with a sphere of pre-equilibrated water molecules, and applied stochastic boundary conditions [143] to the system to mimic the effects of bulk water. The reaction pathway was mapped adiabatically, by fixing one degree of freedom corresponding to the reaction coordinate, and fully minimising all other degrees of freedom. The quantum region was chosen to include the substrate and the side chains of Glu-165 and His-95, with link atoms used at the quantum-classical boundary. The AM1 Hamiltonian [31] was used for the quantum region, and the CHARMM force-field [57] for the classical region.

The results supported the proposal of Glu-165 as the general base, and suggested the novel possibility of neutral histidine acting as an acid, contrary to the expectation that His-95 was protonated [112, 141]. The conclusion that the catalytic His-95 is neutral has been confirmed by NMR spectroscopy [144]. The selection of neutral imidazole as the general acid catalyst has been discussed in terms of achieving a pK_a balance with the weakly acidic intermediate. This avoids the 'thermodynamic trap' that would result from a too stable enediol intermediate, produced by reaction with the more acidic imidazolium [141].

Using the QM/MM method, it is also possible to quantify the effects of individual residues on the reaction occurring at the active site. By performing a perturbative analysis involving sequential deletion of residues followed by a QM/MM interaction energy calculation, starting with the furthest residue from the active site and working in, an energy decomposition profile is generated [112]. This illustrated the effect that long range electrostatics can have on the active site, with residues up to 14 Å away have an affect on the energy of the system in the active site. The most important

contribution was found for Lys-12 which lies close to the active site, and the role of this residue is the stabilization of the enediolate. This explicitly explains the inactivity of a Lys-12 to Met mutant determined in an experimental study [145].

The QM/MM study of TIM was the first illustration of the potential of these methods for studying enzyme catalysis, and has served as a reference for the protocol required for subsequent studies of enzyme reactions.

2.4.2 Bovine Protein Tyrosine Phosphatase.

Recently, the study of the reaction catalyzed by Bovine Protein Tyrosine Phosphatase (BPTP) using QM/MM methods has been reported by Alhambra *et al* [116]. This study represents a progression from the techniques applied to the study of TIM and other enzymatic systems, since the reaction was simulated using molecular dynamics techniques and a QM/MM potential (QM/MD). QM/MD studies are a more powerful technique for studying chemical reactions in condensed phases since they allow for sampling of configuration space as the reaction pathway is followed, and the generation of statistics that can be used to calculate reaction activation parameters such as the enthalpy and free energy of activation of various steps along the reaction pathway.

The use of QM/MD as opposed to QM/MM minimization techniques is computationally intensive and thus precluded the use of an *ab initio* or density functional method for the quantum region. This study was performed with an AM1 Hamiltonian, and the first step of the dephosphorylation reaction was examined (Figure 2.4). Due to the important role that phosphorus has in biological systems, phosphatase reactions have been studied extensively [1]. From experimental data it is believed that Cys-12 and Asp-129 residues are involved in the first step of the dephosphorylation reaction of BPTP [146, 147]. The authors included the sidechains of the phosphorylated tyrosine, Cys-12 and Asp-129 in the quantum region, with link atoms used at the quantum-classical boundaries. In this study the protein was not

truncated, and was surrounded with a 24 Å radius sphere of water molecules. Stochastic boundary methods were applied.

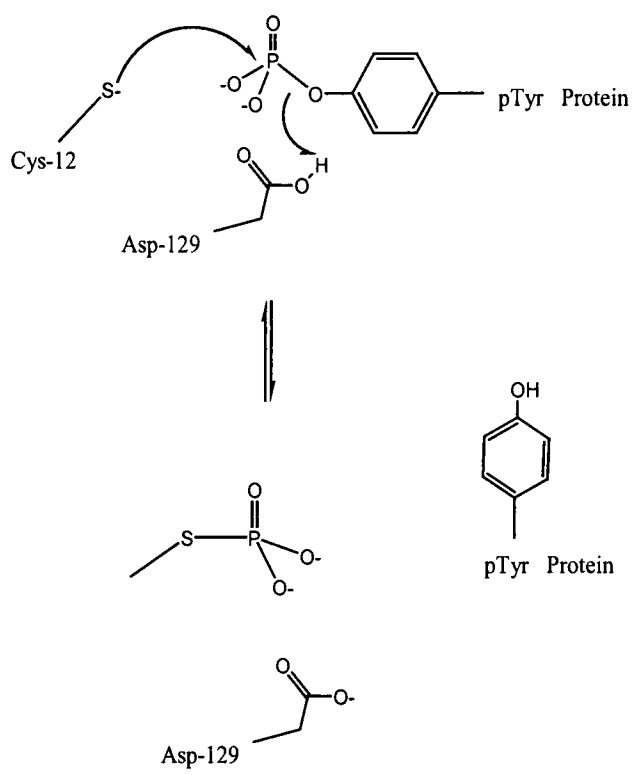


Figure 2.4: Schematic diagram of the first step of the reaction catalysed by bovine protein tyrosine phosphatase (BPTP); formation of the cysteinyl-phosphate intermediate.

The reaction pathway in the enzyme was calculated by using the method of umbrella sampling, which has been widely used in the study of chemical reactions in solution [106,143]. The simulations were performed using between 10 and 12 overlapping regions to cover the entire reaction coordinate, with 25 ps equilibration and 20 ps sampling for each region.

The study found that the transition-state structure for the initial step of the dephosphorylation step is preferentially stabilized over the ground state through a Walden inversion-enforced hydrogen-bonding mechanism at the active site, and suggested that a dianionic mechanism is preferred in the reaction over a monoanionic mechanism since the latter involves the breakage of a hydrogen bond between the nucleophile and the phosphoryl group, which causes the overall energy barrier to be raised.

This study is particularly noteworthy in the evolution of QM/MM studies of enzyme reactions in that a number of technical features have enhanced the accuracy of the technique. Firstly, the authors explicitly optimised the semi-empirical parameters for this specific reaction based on extensive studies of model reactions. This approach had also been used with considerable success in a QM/MM simulation of the proton transfer reaction between methanol and imidazole in solution. Secondly, molecular dynamics techniques were employed that allowed the accurate determination of reaction activation energies and the dynamics of active site residues to be followed along the trajectory of the reaction. This enabled the determination of a free energy of activation of 14 kcal/mol, which was in excellent agreement with the results from stopped-flow studies [116]. In addition the dynamics of the hydrogen-bonding network of the active site could be followed throughout the reaction.

2.4.3 Citrate Synthase

The final application considered in this chapter is chosen to illustrate the application of a QM/MM study of an enzyme reaction that employs an *ab initio* Hamiltonian in the quantum region [148]. Due to the computational intensity of such calculations there are currently very few examples in the literature of QM/MM studies that use a quantum mechanical technique that is more sophisticated than a semi-empirical method. Mulholland *et al.* [148] have recently reported a study of part of the reaction catalyzed by citrate synthase where the quantum region is treated by Hartree-Fock and MP2 methods [8,97] and this serves as a useful illustration of the current state-of-the-art in the field of QM/MM applications to enzyme catalysis.

Citrate Synthase (CS) catalyzes the metabolically important formation of citrate from acetyl-CoA and oxaloacetate [149]. The Asp-375 residue has been shown to be the base for the rate-limiting deprotonation of acetyl-CoA (Figure 2.5) [150]. An intermediate is believed to be formed in this step and is thought to be stabilised by

a hydrogen bond with His-274. It is not clear from experimental data whether this intermediate is the enolate or enol of acetyl-CoA. From the relative pK_as of Asp-375 and acetyl-CoA, it appears that the enolate (or enol) can only be an intermediate in the enzymatic reaction if it is stabilised by the enzyme. It has been proposed that the necessary stabilisation is provided by a low barrier hydrogen bond (LBHB) [151, 152].

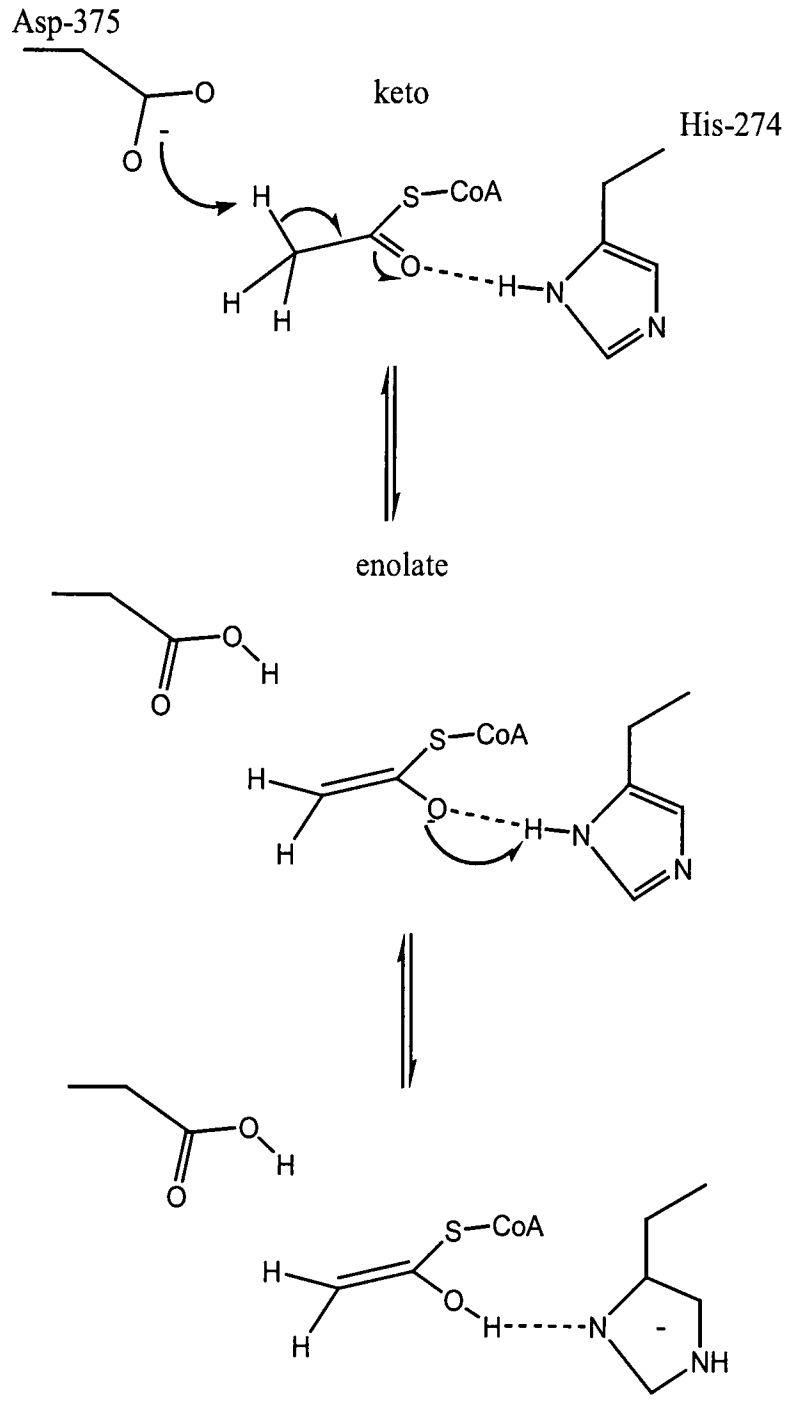


Figure 2.5: A suggested mechanism for the enolisation of acetyl CoA by the enzyme citrate synthase (CS). The keto, enolate, and enol forms of the substrate are shown.

A low barrier hydrogen bond is a covalent interaction between a hydrogen bond donor and the transition state of an enzymatic reaction, and it is believed by experimentalists [152] to be a key source of catalysis. The hydrogen in a LBHB is

almost equidistant from the heavy atoms, and the donor and acceptor are closer to each other than in normal hydrogen bonds. The potential energy barrier for the transfer of the hydrogen between the donor and acceptor atoms is very small leading to the appearance of a broad single energy well or two very shallow energy wells of equal depth [153]. For normal hydrogen bonds, the potential energy for hydrogen transfer is characterised by two energy wells corresponding to the hydrogen being bonded to either of the heavy atoms. The suggested LBHB is between His-274, which is neutral, and an 'enolic' acetyl-CoA intermediate [152,154]. To resolve the question of the nature of the intermediate and its stabilisation by the enzyme, the first reaction step in CS mechanism was investigated by *ab initio* QM/MM calculations with the CHARMM program.

The system contained all residues within 17 Å of the terminal carbon of acetyl-CoA, and 23 crystallographic water molecules. All *ab initio* QM/MM calculations were performed using the CHARMM program in conjunction with the GAMESS-US electronic structure program [155]. The reaction pathway was followed using the adiabatic mapping technique used previously for TIM and other enzymes. The approach used by Mulholland *et al.* [148] was to perform a series of different calculations, beginning with a computationally inexpensive method and progressively increasing the level of calculation used in the quantum region. This is analogous to the approach often used in pure quantum mechanical studies that initially use a low-level basis set to get a rough estimate of the potential energy surface and then refine this with a more accurate higher-level basis set. The quantum region included the sidechains of Asp-375, His-274, the thioester portion of acetyl-CoA and the substrate.

For citrate synthase, the reaction profile was initially extensively studied with by QM/MM with an AM1 Hamiltonian [114]. The points along the reaction pathway were subsequently subjected to QM/MM minimisation using a HF/3-21G* potential for the quantum region and finally the minimisation of the reaction points on the pathway

was completed with the employment of HF/6-31G* potential to treat quantum region. This reaction pathway was further refined by performing single point calculations at the MP2/6-31G* level for the quantum region. This allowed the system to be fully minimised without having a large computational cost that would be incurred had the quantum region been treated at the HF/6-31G* level from the start

The study found that the enolate of acetyl-CoA is the intermediate in the rate-limiting step of the reaction, in agreement with previous experimental studies [150]. The reaction catalyzed by CS has previously been proposed to employ a mechanism that uses low-barrier hydrogen bonds. Such bonds can be exceptionally strong in the gas-phase, and have been proposed to have energies of up to 20 kcal/mol in enzymes. However, the debate about their role in enzyme catalysis is controversial. A characteristic of low-barrier hydrogen bonds is that the hydrogen is shared between atoms of approximately equal pKa. In CS it has been proposed that the hydrogen bond between His-274 and an enolic (the proton is shared equally) intermediate, is responsible for stabilizing the intermediate. The calculations on CS indicated that the enolate of acetyl-CoA is significantly more stable than the enol or a proton sharing enolic form, and thus do not support the proposal that a low-barrier hydrogen bond is involved in catalysis in CS. This study demonstrates the practical application of high level QM/MM studies to elucidating mechanistic details of an enzymatic reaction that are otherwise unclear.

2.5 Summary

The field of QM/MM simulations of chemical reactions has grown considerably from the initial proposals of Warshel *et al.* [86] in the 1970s to a technique that can deliver quantitatively accurate reaction pathways for reactions in solution and in the active sites of enzymes. Currently, the computational chemist has the option of using semi-empirical, density functional or *ab initio* methods for the quantum region.

The method that is employed in any given situation is dependent on a number of factors. The computational expense of the density functional and *ab initio* methods dictate that studies using these methods need to be carried out on parallel computers. Naturally, these methods are more accurate for studying the chemistry of the process under consideration, and in the case of metalloenzymes with transition metals it is almost essential to use density functional methods. Nonetheless, it is possible to get quantitative accuracy with semi-empirical based QM/MM studies, particularly when the quantum atoms and the van der Waals parameters are parameterised for the specific reaction being studied. Additionally, semi-empirical based QM/MM methods allow a dynamical study of the chemical reaction, which is currently beyond the higher-level methods, even with parallel computers. The field continues to grow and it is to be expected that advances in the speed of the quantum calculations, the accuracy of the treatment of the quantum-classical boundary region, and computational speed will come over the next decade and enable further insight of the mechanisms of chemical reactions to be gained.

Chapter 3

Phosphate ester hydrolysis: an introduction.

3.1 Introduction

The material presented in the previous two chapters of this thesis give an account of the development and application of theoretical methods to determine the chemical features and reactivities of molecular systems. While the development of these methods is an interesting and worthwhile pursuit, their use when solving chemical 'puzzles', which defy the capabilities of experimental techniques, is often a far more stimulating and rewarding exercise.

The later chapters of this thesis present a detailed account of the application of these methods to elucidate the characteristics of one of nature's most important reactions. Despite its importance, many of the features of this reaction remain opaque to both experimental and theoretical methods. The reaction in question is the hydrolysis of phosphate esters.

To the majority of chemists, phosphate esters and their reactions are of no apparent consequence. In the living world, however, phosphate esters are ubiquitous and their reactions underpin some of the most important processes in biology [1]. The most

prominent phosphate esters in nature, and arguably the most important molecules in nature, are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) [88, 156, 157]. These molecules are responsible for storing and maintaining the integrity of a living organism's genetic code. Both molecules are huge polymeric chains consisting of subunits known as nucleotides which are linked together by phosphodiester bonds. In order to maintain genetic integrity and prevent the occurrence of dangerous mutations, these linkages must be incredibly strong and impervious to chemical attack. This has been found to be the case. The half-lives of the phosphodiester linkage in DNA and RNA are 200 million years and 1000 years respectively [158]. In recent years, RNA and latterly DNA have also been shown to catalyse their own cleavage [157, 159–162]. Particular attention will be paid to the catalytic properties of RNA in Section 3.4.1 of this chapter.

Another key biological phosphate ester is adenosine triphosphate (ATP), the universal currency of energy in all forms of life [163]. A large amount of energy is released on the hydrolysis of ATP to adenosine diphosphate (ADP) and this energy is used to drive many key biological reactions [88]. Other important biological metabolites such as creatine phosphate and phosphoenolpyruvate are also phosphate esters.

The hydrolysis of the phosphate guanosine triphosphate (GTP), another phosphate ester, to give guanosine diphosphate (GDP) is used to 'switch off' G-proteins such as RAS at the beginning of many very important signal transduction cascades [88, 164]. As a result, there is considerable interest in determining ways to inhibit this type of phosphate ester hydrolysis reaction.

It can be seen from the examples cited above that phosphate ester hydrolysis reactions can be found throughout nature and that a precise knowledge of their features would be of considerable therapeutic benefit. Therefore, it should not be surprising that there is a vast body of research, both experimental and theoretical, that has sought to elucidate the chemistry of phosphate esters [2–4, 165–171]. Despite this huge collective

effort, many questions remain unanswered concerning the mechanism and energetics of phosphate ester hydrolysis reactions. The next section of this chapter examines the principal mechanistic routes available to phosphate ester hydrolysis reactions.

3.2 General mechanisms

The hydrolysis of phosphate esters can proceed by one of two major mechanistic pathways, both giving rise to the same products from given reactants. The two pathways, in their extreme forms, are comparable to the well-known mechanisms for substitution reactions at a tetrahedral carbon: the S_N1 and S_N2 mechanisms [165]. In phosphate ester hydrolysis chemistry, the two mechanistic alternatives are respectively termed 'dissociative' and 'associative'. Both are shown in Figure 3.1.

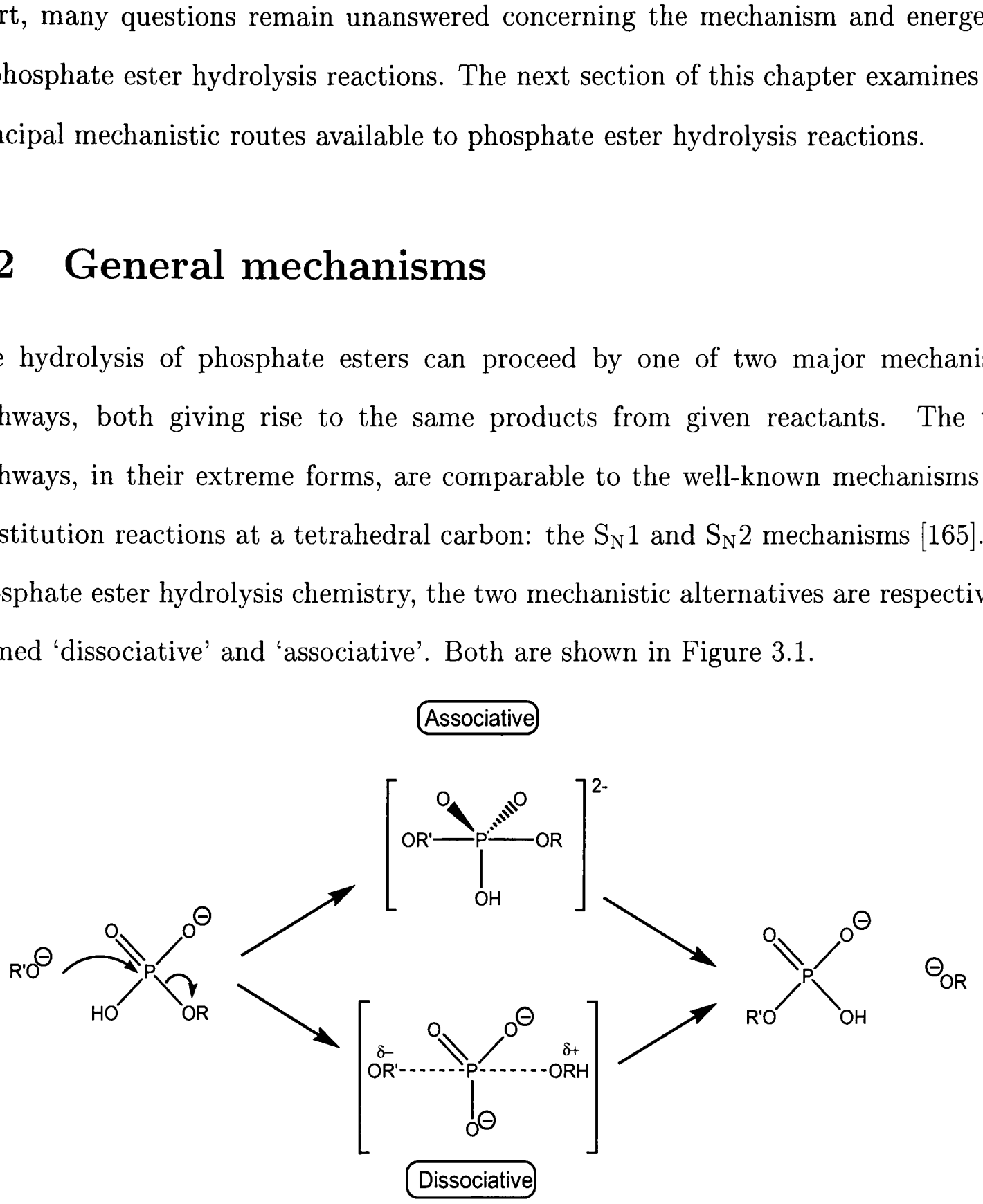


Figure 3.1: A model base-catalysed phosphate ester hydrolysis reaction. Both the associative and dissociative reaction pathways are shown. Partial positive and negative charges are denoted by $\delta+$ and $\delta-$ respectively.

The dissociative mechanism is dominated by bond cleavage and is characterised by the dissociation of a ligand from the phosphorus to leave a tricoordinated metaphosphate species which is then susceptible to attack from a nucleophile to form product [172–174]. The metaphosphate species is highly unstable due its high

charge density and the electron deficiency at phosphorus makes it very susceptible to nucleophilic attack.

The associative mechanism proceeds via the formation of a pentacoordinated species in which there is a large degree of bond formation between the nucleophile and the phosphorus and very little bond cleavage between the phosphorus and the leaving group [166, 169, 170, 172]. The pentacoordinated species, known as a phosphorane, can often be an intermediate and has a trigonal bipyramidal (TBP) geometry. The mechanism is completed by the decomposition of the phosphorane to form the reaction products.

There has been a recent surge of interest, led by Florián and Warshel, in the theoretical investigation of these mechanistic alternatives [167, 175–177]. They have challenged the consensus view that, while the hydrolysis of phosphodiesters and phosphotriesters proceeds via an associative mechanism, the hydrolysis of monoesters follow a dissociative pathway [172, 178, 179]. They determined, through calculations and reevaluation of experimental data that, the hydrolysis of phosphate monoesters can proceed via associative or dissociative pathways. They also concluded that it is extremely difficult to interpret experimental data differentiating these mechanisms and that the use of methods combining quantum mechanical potentials and reliable solvent models is therefore the best way to elucidate mechanistic pathways for the hydrolysis of phosphate esters. In Chapter 6, a detailed study of the relative energy barriers for both pathways in a particular phosphate ester hydrolysis is presented.

A further complication is that the hydrolysis reaction can be either acid- or base-catalysed. The reaction shown in Figure 3.1 is an example of a base-catalysed reaction [6]. The monoanionic nucleophile, which may have been deprotonated by another base, attacks the phosphate to form a dianionic transition state. If the reaction were acid-catalysed, the phosphate would have been protonated either simultaneous to or before nucleophilic attack. This leads to the formation of a more stable phosphorane species.

Whether the reaction undergoes acid- or base-catalysis is largely dependent on the prevalent pH environment in which the reaction takes place.

A great deal of experimental research has been conducted in the field of phosphate ester hydrolysis. The findings are too vast to be reviewed in this chapter and only selected studies will be mentioned. A lengthy and excellent review has been written by Thatcher and Kluger which covers a wide range of experimental studies [2].

The nature of the phosphorane species

As mentioned in the previous section, the hydrolysis of phosphodiesters and -triesters, including DNA and RNA, are thought to follow an associative pathway [2,3,180]. This has led to a great deal of interest in the nature of the pentacoordinated phosphorane species that is encountered on the associative reaction pathway [2,166]. The main point of interest is whether or not the phosphorane species exists as a stable intermediate on the reaction potential energy surface.

The chemistry of trigonal bipyramidal phosphoranes is further complicated because two types of ligand positions exist: apical (also known as axial) and equatorial. A set of rules governing their existence has been proposed by Westheimer and Holmes [166,170]. The key points can be summarised as follows:

- Ligands may occupy either equatorial or apical positions within the TBP structure and they may swap positions via a reorganisation process known as pseudorotation [181]. This process is covered later in this section.

- Nucleophiles attack and leaving groups leave the complex from apical positions. This can be facilitated by pseudorotation processes.

The existence of phosphorane intermediates is governed by many factors which include the nature of the substrate, i.e. its size and pK_a, and the chemical environment in which the reaction takes place [2–4,182]. However, the key factor, from which many

other factors are ultimately derived, is the degree of negative charge residing on the phosphorane species. A monoanionic phosphorane is inherently more stable than a dianionic phosphorane since the latter contains two full negative charges which must be delocalised throughout the molecule in order to gain stability. The stability of phosphoranes increases with molecular size and solvation since both facilitate charge delocalisation. Theoretical calculations by the Taira group have shown that the substitution of hydrogens by simple bridging alkyl groups [183–185] and the presence of an explicitly modelled first solvation shell [186] substantially increase the stability of dianionic phosphoranes to the extent where they may exist as reaction intermediates in the gas phase. Often, even if dianionic phosphorane intermediates exist, the depth of the potential energy well in which they reside may only be of the order of $k_{\mathrm{B}}\mathrm{T}$, and they are thus unlikely to be kinetically significant. Intermediates of this type have been identified in theoretical studies [168, 184, 187]. It has also been found that the use of low accuracy basis sets, such as the STO-3G basis set, can overestimate the stability of phosphorane intermediates [182, 187, 188].

The overall conclusions reached by theoretical calculations on the nature of phosphorane intermediates indicate that, while stable monoanionic phosphoranes exist in the gas phase [189–191], dianionic phosphoranes are not stable in the gas phase and if they exist as intermediates, they will be probably be kinetically insignificant and will not have a lifetime long enough to be experimentally distinguished from transition states [4, 168]. Dianionic phosphoranes have been found to exist in solution but they too may only be metastable [168].

One of the aims of the research presented in this thesis is to determine the stability of monoanionic and dianionic phosphoranes generated in a series of three related reactions in both the gas phase and in solution. As an improvement on previous theoretical studies, a more accurate Hamiltonian, which accounts for electron correlation, is

employed in conjunction with both continuum and atomistic solvent representations. The results of these calculations are outlined in Chapter 4.

Pseudorotation

Pseudorotation is a chemical process that occurs in stable trigonal bipyramidal phosphorane species [166, 181]. Through pseudorotation, the TBP species is converted into another isomer through ligand reorganisation. The conversion is achieved by the simultaneous inward bending of both axial ligands and the outward bending of two of the equatorial ligands (Figure 3.2).

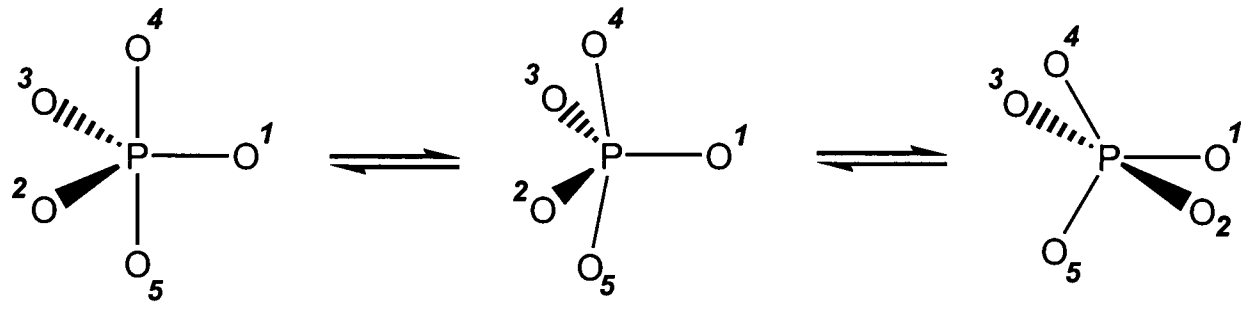


Figure 3.2: The pseudorotation process for an arbitrary phosphorane. (For the sake of clarity only the phosphorus and oxygen atoms are displayed.) Equatorial ligands 2 and 3 rotate to become axial ligands whilst ligands 4 and 5 become equatorial. Ligand 1 remains stationary.

Pseudorotation, which can be experimentally detected using IR spectroscopy, can only occur in stable intermediate species. The lifetime of a phosphorane transition state is thought to be shorter than the timescale of a pseudorotation process and thus it is thought that pseudorotation does not occur within transition states [3].

The hydrolysis of cyclic phosphates

The hydrolysis of cyclic phosphate esters has been the subject of much attention since the reaction is very similar to the reverse of the transesterification step in the hydrolysis of RNA [3, 6]. The hydrolysis of methylene phosphate and ethylene phosphate, both of which are small cyclic phosphates, has been the focus of many research studies [2, 166, 189, 192–194]. Experimental studies have shown the hydrolysis of cyclic phosphate

esters to be 10^6 to 10^8 times faster than the hydrolysis of equivalent acyclic phosphate esters [193, 195]. It was initially thought that this higher reactivity was due to loss of instability due to ring-strain on going from the ground state to a TBP transition state. However, more recent theoretical studies have suggested that the faster rate of hydrolysis encountered for cyclic phosphates is because of greater solvent stabilisation of the transition state relative to the ground state [168].

3.3 The hydrolysis of DNA

Over the years, the hydrolysis of DNA has received far less attention than the analogous RNA reaction. The main reason for this is that, in solution the non-enzymatic hydrolysis of DNA is virtually non-existent due to the great stability of the phosphodiester linkages in the molecule which makes research into the reaction a difficult venture [158].

In the past ten years, the excellent catalytic properties of the lanthanide ions for the hydrolysis of nucleic acids has become apparent [196, 197] and in the presence of Ce^{IV} ions the half-life of the phosphodiester bond in DNA is reduced by a factor of 10^{11} [198]. These advances allowed mechanistic studies of this reaction for the first time and the reaction has become the subject of substantial research.

The mechanism for DNA hydrolysis proceeds in two steps (Figure 3.3) [180]. The first step is characterised by the attack of an external nucleophile, often a hydroxide ion, on the phosphorus atom. This leads to the formation of a pentacoordinate species postulated to be an intermediate. The second step involves the cleavage of the P-O5′ bond[1]. It is difficult to determine which of these steps is rate limiting since the mechanistic studies used either lanthanide ions or DNA analogues with especially good leaving groups to promote the reaction [180]. Since it is impossible to determine the exact nature of the transition states without recourse to methods such as the use of

[1]The atom numbering scheme used for both DNA and RNA is shown in Figure 3.4.

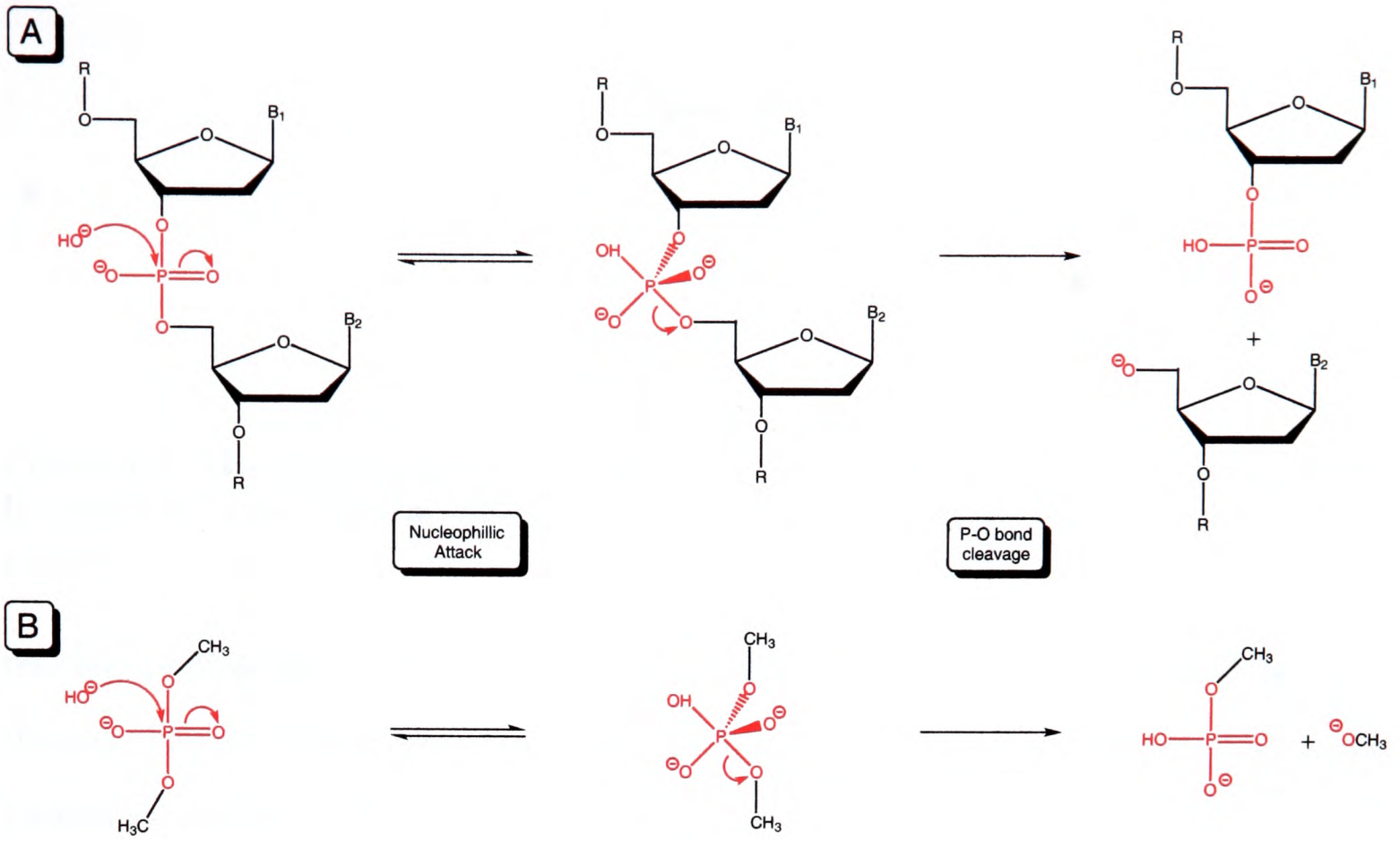


Figure 3.3: The reaction mechanism for the hydrolysis of both DNA (part A) and DMP (part B). Both the nucleophilic attack and P–O bond cleavage steps are shown. By visually comparing the structures involved in both reactions, it becomes apparent why the hydrolysis of DMP is considered an excellent model for the hydrolysis of DNA.

metal ions or modified leaving groups, theoretical studies using DNA analogues offer great promise.

3.3.1 DMP - a model compound for DNA

Since the hydrolysis of DNA is initiated by the attack of an external nucleophile, acyclic phosphate esters are used as model compounds for DNA. Dimethylphosphate (DMP) is the most commonly used of the available models (Figure 3.3). Like DNA, its hydrolysis in solution proceeds at an exceptionally slow rate (2×10^{-14} s^{-1} at 25° C [199])

Early theoretical studies indicated that the base-catalysed hydrolysis of DMP in the gas phase proceeded via a pentacoordinate intermediate species [188, 200, 201]. However, subsequent studies, employing more accurate basis sets, concluded that no such intermediate existed in the gas phase [168, 184, 187, 194]. Dejaegere and Karplus showed, that by estimating the solvation free energy of geometries along the gas phase

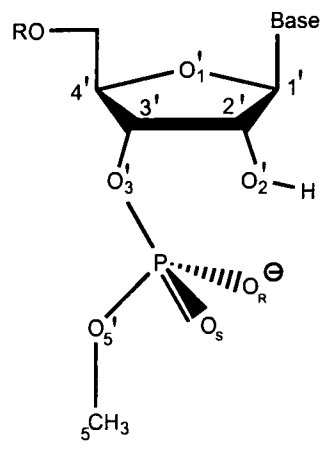


Figure 3.4: The conventional atom numbering scheme used for both DNA and RNA. It should be noted that, in DNA, O'_2 is replaced is not present and is replaced by a proton.

reaction profile, the energy barrier for the reaction was lowered from 97 kcal mol^{-1} in the gas phase to 37 kcal mol^{-1} in solution. No presence of an intermediate species was located in solution [168].

A great deal of the research presented in this thesis concerns the determination of the energy profile for both the acid- and base-catalysed hydrolysis of DMP in the gas phase and in solution using some of the most accurate methods available.

3.4 The hydrolysis of RNA

The hydrolysis of RNA has attracted considerable interest since the discovery of catalytic RNA molecules and the subsequent realisation of their potential as tools for gene therapy [3, 4]. Just as with other phosphate ester hydrolysis reactions, the hydrolysis of RNA can occur via acid- or base-catalysed mechanisms [3, 4, 6]. Both mechanisms are illustrated in Figure 3.5. In the base-catalysed reaction, a general base deprotonates the 2′-OH and the resulting 2′-hydroxide acts as an internal nucleophile and attacks the phosphorus leading to the formation of a dianionic phosphorane. This is followed by the expulsion of the 5′-oxygen of the next nucleotide to form a cyclic phosphodiester. In the second step, the cyclic phosphodiester is hydrolysed to

regenerate the 2′-OH and form a 3′-phosphate. The first and second steps are known as the transesterification and hydrolysis steps, respectively.

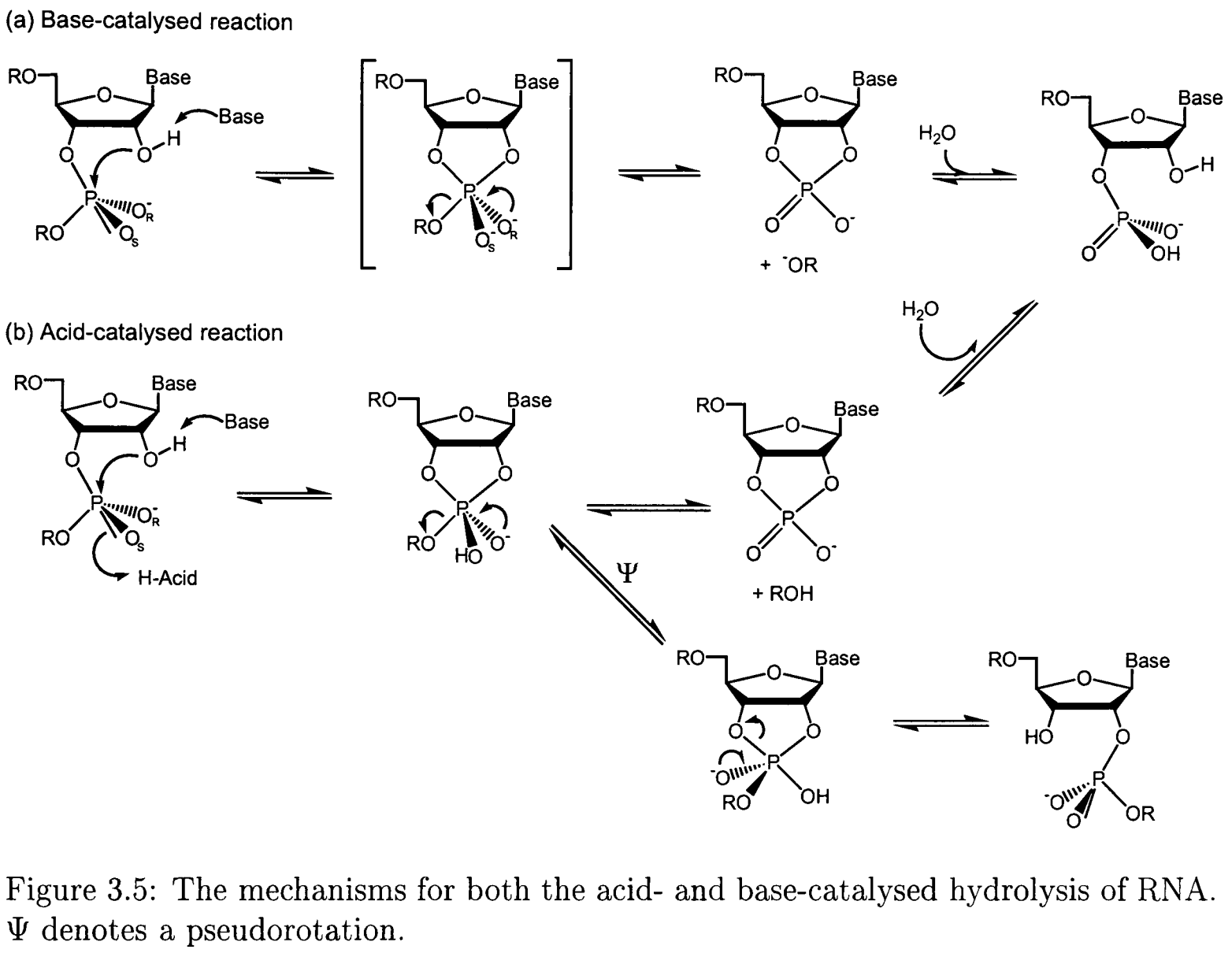


Figure 3.5: The mechanisms for both the acid- and base-catalysed hydrolysis of RNA. Ψ denotes a pseudorotation.

The acid-catalysed mechanism is more complicated and involves activation of the phosphodiester by proton transfer either prior or simultaneous to nucleophilic attack. This results in the formation of a monoanionic phosphorane intermediate. The intermediate can undergo pseudorotation, denoted by Ψ, to produce an intermediate in which the 3′-oxygen is in an apical position and is ultimately expelled. Otherwise, without the occurrence of pseudorotation, the 5′-O leaving group is expelled and the resulting cyclic phosphodiester is hydrolysed just as in the base-catalysed mechanism.

3.4.1 Ribozymes

In the early 1980s the discovery of RNA molecules that can catalyse catalyse both their own cleavage and that of other RNA molecules shattered the belief that only proteins were capable of enzymatic activity [5, 202]. These RNA molecules are known as ribozymes and their existence has provoked thought that RNA enzymes predate the evolution of proteinaceous enzymes and that they once catalysed all biochemical reactions [157]. Since their initial discovery, many different ribozymes have been identified [159, 203, 204].

It had always been thought that all ribozymes are metalloenzymes, i.e. they require the presence of a metal ion to facilitate their catalytic function [205–208]. Recently this conjecture has been challenged by the discovery of ribozymes whose catalytic activity is independent of metal ions [209,210]. However, the vast majority of ribozymes do utilise divalent metal ions in a catalytic role. The exact role of the divalent metal ions has still not been properly ascertained but possible functions include: a metal-coordinated hydroxide ion acting as a general base and abstracting the proton from the 2′-OH; the negative charge on the O5′ leaving moiety being stabilised by either direct coordination with a metal ion or by a water molecule bound to a metal; and direct coordination of a metal ion to one of the non-bridging oxygens stabilising the charge in the phosphate and lower the energy barrier to nucleophilic attack.

Ribozymes have been shown to have great potential as therapeutic agents that can inhibit gene expression [211, 212]. Therefore knowledge of their mode of action would be especially advantageous. Chapter 5 presents a detailed theoretical study of the transesterification of the base-catalysed hydrolysis of RNA.

3.4.2 Thio-substituted RNA analogues

In order to probe the mechanism of the RNA hydrolysis, several experiments have studied the hydrolysis of thio-analogues of RNA and dinucleotides [213–215]. This section introduces the two most common types of thio substitutions used.

5′-thio substitutions

The non-enzymatic hydrolysis of RNA (Figure 3.5) is thought to proceed via two putative transition states. The first corresponds to the approach of the 2′-oxygen to the phosphorus and the second to the cleavage of the P–O5′ phosphodiester bond. The second transition state is thought to have a higher energy than the first [4,6] and therefore P–O5′ bond cleavage is thought to be rate-limiting. In order to confirm this, a number of groups have synthesised RNA analogues with 5′-thio leaving groups [205,214, 216–218] with a view to determining whether the cleavage of the P–O5′ phosphodiester bond is the rate-determining step.

If the approach of the 2′-oxygen were the rate limiting step, the presence of a 5′-phosphorothiolate linkage would have no effect on reaction rate since the energy barrier for this step would not be reduced [219]. On the other hand, if TS2 is the rate-determining transition state, it would be expected that the thiolate would be hydrolysed far more quickly since the pK_a of the thiol is 5 units lower than that of the equivalent alcohol [220, 221]. The results of the experiments have shown that the hydrolysis of the phosphorothiolate RNA analogue is 10^5 times faster than that of the corresponding RNA analogue with a natural phosphodiester bond at the scissile linkage. Thus, the rate-limiting step for the non-enzymatic hydrolysis of RNA corresponds to the cleavage of the P–O5′ bond.

pro-R oxygen substitution

One of the possible roles of divalent metal ions in catalysing the RNA cleavage by ribozymes is through direct interaction with one of the nonbridging scissile phosphate oxygens. This is thought to have the effect of stabilising the negative charge that builds up in the phosphorane transition state. To test this hypothesis, experiments where both stereochemically unique non-bridging oxygens, *pro*-R oxygen and *pro*-S oxygen, in the hammerhead ribozyme were, in turn, replaced by sulphur atoms [222]. It was observed that when *pro*-R oxygen was replaced in the presence of Mg^{2+} ions, catalytic activity was reduced dramatically, while for the *pro*-S oxygen substitution, this was not observed to be the case [222–224]. This effect is known as the 'thio effect' and is thought to indicate that an Mg^{2+} ion binds to the *pro*-R-oxygen.

These experiments also showed that when the Mg^{2+} ions were replaced by Mn^{2+} the catalytic rate of the *pro*-S substrate was increased back to normal levels. This phenomenon has been explained in terms of a 'hard acid' such as the Mg^{2+} ion having a greater binding affinity for a 'hard base' such as oxygen rather than a 'soft base' like sulphur. The 'soft acid' Mn^{2+} binds preferentially to sulphur. Thus, it was concluded that, once metal coordination was restored by the addition of Mn^{2+}, the phosphorane transition state was once more stabilised and the reaction proceeds at a rate comparable to that of the unmodified ribozyme.

This hypothesis has been challenged by experiments which have shown that the substitution of a sulphur for *pro*-S oxygen shows a similar effect on the addition of Mn^{2+} [225,226]. A more recent experiment shows that, when both nonbridging oxygens are substituted with sulphur atoms in the presence of Mg^{2+} ions, there is far less rate reduction than for the single *pro*-R oxygen substitution [227]. These two conclusions cast doubt over the binding of a divalent metal ion to a non-bridging oxygen in the hammerhead ribozyme active site. It has instead been suggested that the single *pro*-R

oxygen substitution merely causes a charge asymmetry due to the inability of sulphur to accommodate more than a single negative charge [224].

However, experimental studies have determined that the substitution of sulphur for either the *pro*-R or *pro*-S oxygen in RNA and RNA analogues has only a small effect on the reaction rate in alkaline solution without the presence of metal ions [215, 228–230].

A study examining the effect of both *pro*-R or 5′ sulphur substitutions on the hydrolysis of an RNA analogue in the gas phase is presented in Chapter 5.

3.5 Summary

There can be no doubt as to the importance of phosphate esters and their hydrolysis reactions. A comprehensive knowledge of the chemistry of these species would greatly further medical and scientific research. While a great deal of effort has gone into the elucidation of the mechanistic features of phosphate ester hydrolysis reactions, there is still uncertainty regarding the exact nature of these reactions. This is because we are dealing with a complicated reaction that can follow a number of different mechanistic pathways simultaneously. For example, acid and base catalysis can occur together in solution at a certain pHs. This makes unambiguous experimental determination of mechanistic pathways and reaction energies very difficult. Therefore, with this many variables present for this class of chemical reaction, there is a challenge to theoretical chemistry to determine precisely the mechanistic features of these reactions and the parameters that govern them. The next four chapters present separate studies of various phosphate ester hydrolysis reactions in the gas phase and in solution.

Chapter 4

A quantum chemical study of the hydrolysis of dimethyl phosphate

4.1 Introduction

This chapter presents the results of a quantum chemical study of the hydrolysis of dimethyl phosphate (DMP). This reaction, discussed in Section 3.3.1, is considered to be a good model for the hydrolysis of DNA (Figure 3.3) since the involvement of an external nucleophile and the formation of an acyclic transition state are involved.

Because DMP strongly resists the hydrolysis of its phosphodiester bond in aqueous solution ($k = 1.6 \times 10^{-13}$ s^{-1} at 25° C) [231], the majority of studies of the uncatalysed hydrolysis of DMP have been theoretical [168, 171, 180, 187]. Another difficulty concerning the viability of experimental studies is that, in solution, phosphate esters such as DMP can be attacked by either H_2O or OH^- acting as a nucleophile in what are known as acid- and base-catalysed mechanisms, respectively [232]. There is also the added complication of whether the phosphate is initially protonated or not. This confusion is compounded by the fact that alternative mechanisms can occur simultaneously in solution making it is very difficult to distinguish between them experimentally [2, 232].

Therefore, in this study, three different mechanisms for the hydrolysis of DMP are examined: both the acid- and base-catalysed hydrolysis of the monoanionic, unprotonated substrate, and the acid-catalysed hydrolysis of the protonated, neutrally charged substrate. Each reaction was modelled in the gas phase and the effect of solvent was considered by estimating the free energy of solvation of each reaction stationary point using a continuum solvent representation. In addition, the solution reaction profile for the base-catalysed reaction was also determined through full geometry optimisation of each stationary point in the presence of the continuum solvent representation.

4.2 Methods

Each gas phase reaction profile was determined using density functional theory [17] Hamiltonians. All calculations used a combination of the Becke three-parameter hybrid method [23, 25] and the Lee, Yang, and Parr correlation functional [24] in what is called the B3LYP functional. The use of DFT instead of conventional Hartree-Fock methods [8, 12] has the advantage of accounting for electron correlation effects while being computationally less demanding than standard perturbation or configuration interaction methods [8]. The inclusion of correlation effects has been shown to be important for studying both hydrogen bonded [233] and dianionic [234] chemical systems and the B3LYP functional has also been shown to particularly effective for the prediction of reaction geometries and activation energies [26, 235, 236].

Previous *ab initio* studies of this reaction [187, 194] have shown that the use of diffuse and polarisation functions is advantageous when studying the chemistry of phosphoranes and phosphates due to the highly charged molecules involved and therefore the 6-31+G* basis set [12] was used for all calculations.

The geometries of all the stationary points were fully optimised using standard gradient optimisation methods. The reactants and products for each reaction were

treated at infinite separation, i.e. separate calculations were performed for each molecule involved. The validity of each stationary point was verified by frequency analysis to confirm a positive Hessian matrix for minima and a Hessian matrix with one negative eigenvalue for transition states. The frequencies were used to calculate thermodynamic quantities including zero-point energy (ZPE) corrections, and Gibbs Free energies (G). A study by Wong [237] has shown that B3LYP calculations employing the 6-31G(d) basis set provides harmonic frequencies that can be effectively scaled for comparison with experimental wavenumbers by using an empirical scaling factor of 0.96. All vibrational frequencies were scaled accordingly in this study.

An IRC profile [47] was calculated from each transition state obtained to confirm that the transition state was on the correct potential energy surface for the reaction. The electronic charge distribution was derived from the molecular electrostatic potential using the CHELPG [55] scheme.

The effect of solvation was estimated by calculations performed at the same level of theory in conjunction with the Polarizable Continuum Model (PCM) [83, 84] representation of solvent[1]. The values of the dielectric constant were set to the standard values of 78.39 for the solvent (water) and 1.00 for inside the solute cavity. The dimensions of the solvent cavity in which the solute resides were calculated using the Pauling van der Waals radii [238] which were scaled by a factor of 1.2 for all atoms except acidic hydrogens. The free energies in solution, $G_{solution}$, for each stationary point were calculated by adding the free energy of solvation , ΔG_{sol}, to the gas phase free energy, G_{gas}. All calculations were carried out using the Gaussian 98 [48] suite of programs on a range of individual workstations and parallel supercomputers.

[1]The use of the PCM model at the B3LYP/6-31+G* level of theory will be referred to in the text as the B3LYP(PCM)/6-31+G* level of theory.

4.3 Results

4.3.1 Base catalysed hydrolysis of dimethyl phosphate

The first reaction studied was base-catalysed hydrolysis of unprotonated DMP (Figure 4.1). The geometries of each reaction stationary point were fully optimised in the gas phase and in solution. In both cases, the reaction was found to proceed via an associative pathway with the formation of a single transition state.

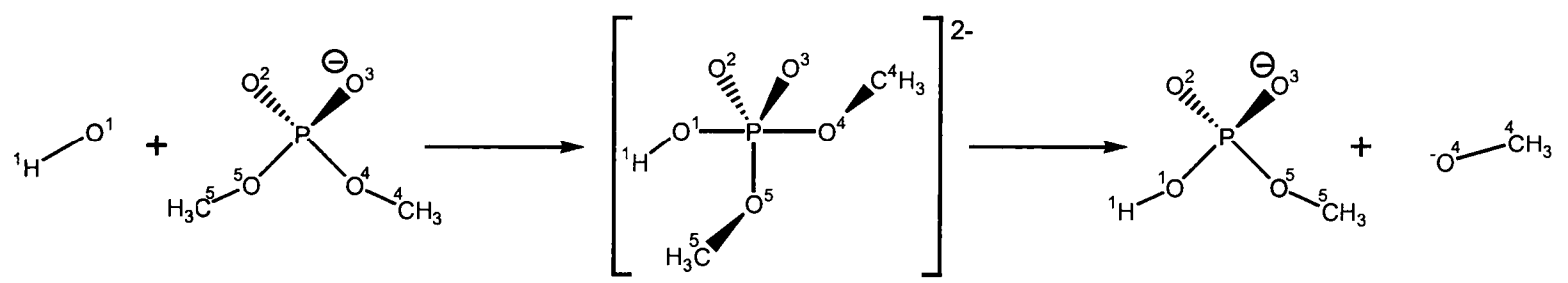


Figure 4.1: The reaction scheme for the base-catalysed hydrolysis of DMP. The atom numbering scheme referred to in the text is also shown.

The structure of the gas phase and solution transition states are shown in Figure 4.3 and key bond lengths, bond angles, and dihedral angles of the stationary points are given in Tables 4.1, 4.2, and 4.3 respectively.

Gas phase reaction

The geometry of dimethyl phosphate in the gas phase (Figure 4.2) is tetrahedral and the two methoxy groups in a *gauche, gauche* conformation. Both P–OCH_3 bonds are fully formed with bond lengths of 1.67 Å.

The geometry of the pentacoordinated gas phase transition state is distorted trigonal bipyramidal with the nucleophile and leaving groups occupying axial positions. This indicates that the reaction proceeds by an associative rather than a dissociative pathway. This transition state, shown in Figure 4.3(a), is reactant-like and corresponds to the formation of the P–OH bond. The bond between the incoming oxygen and the phosphorus is partially formed at 2.22 Å and the axial P–OCH_3 bond has lengthened by only 0.18 Å to 1.85 Å. The angle of approach between the nucleophile and leaving

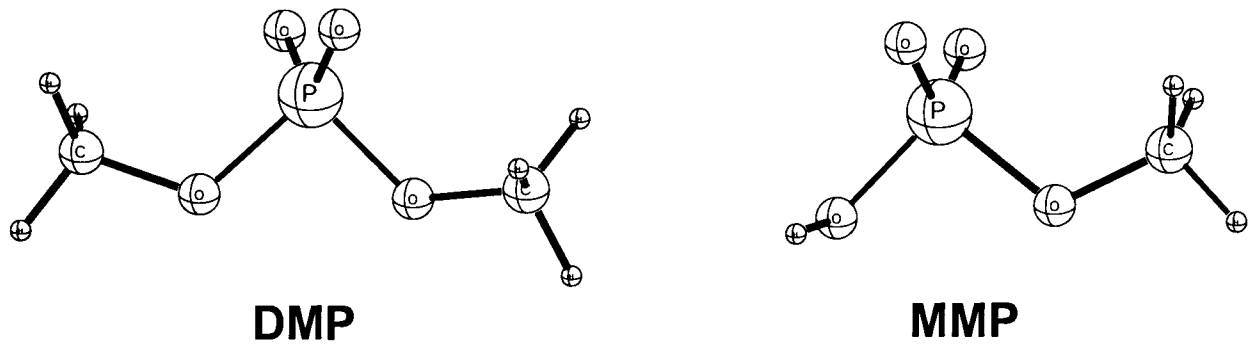


Figure 4.2: The gas phase geometries for DMP and MMP as determined at the B3LYP/6-31+G* level of theory.

group, O1–P–O4, is almost linear (171°) confirming the angle of attack is 'in-line' as predicted by experimental studies [3, 4, 6].

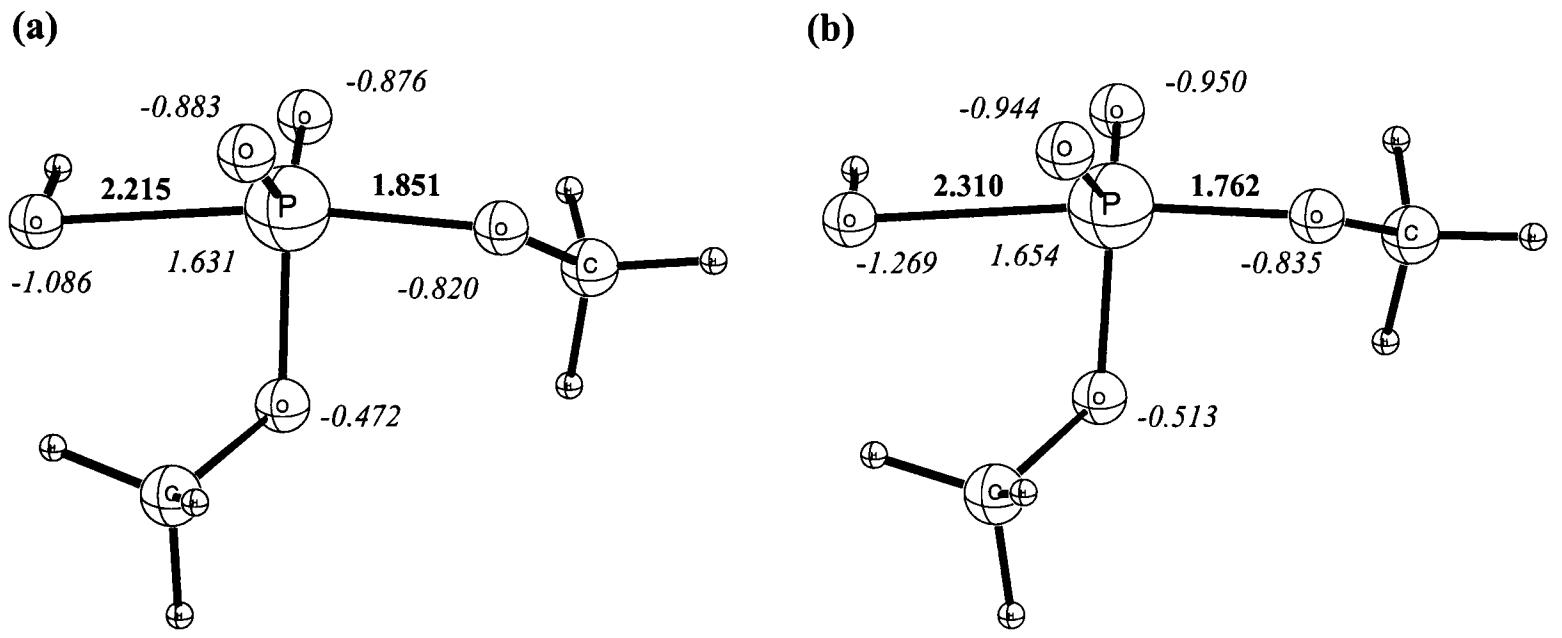


Figure 4.3: The gas phase (a) and solution (b) transition state geometries determined for the DMP + OH^- reaction at the B3LYP/6-31+G* and B3LYP(PCM)/6-31+G* levels of theory. Key bond lengths are denoted in bold type and CHELPG atomic charges are given in italicised type.

The equatorial methoxy group in the transition state was determined to be oriented towards the incoming OH^- nucleophile. This may lead to the formation of an ion-molecular electrostatic interaction between the CH_3 group and the OH^- anion which would alleviate some of the Coulombic repulsion associated with the approach of the hydroxyl anion towards the negatively charged phosphate.

An IRC profile was calculated from the transition state and is shown in Figure 4.4. The calculation only generated 5 points in the direction leading towards the formation

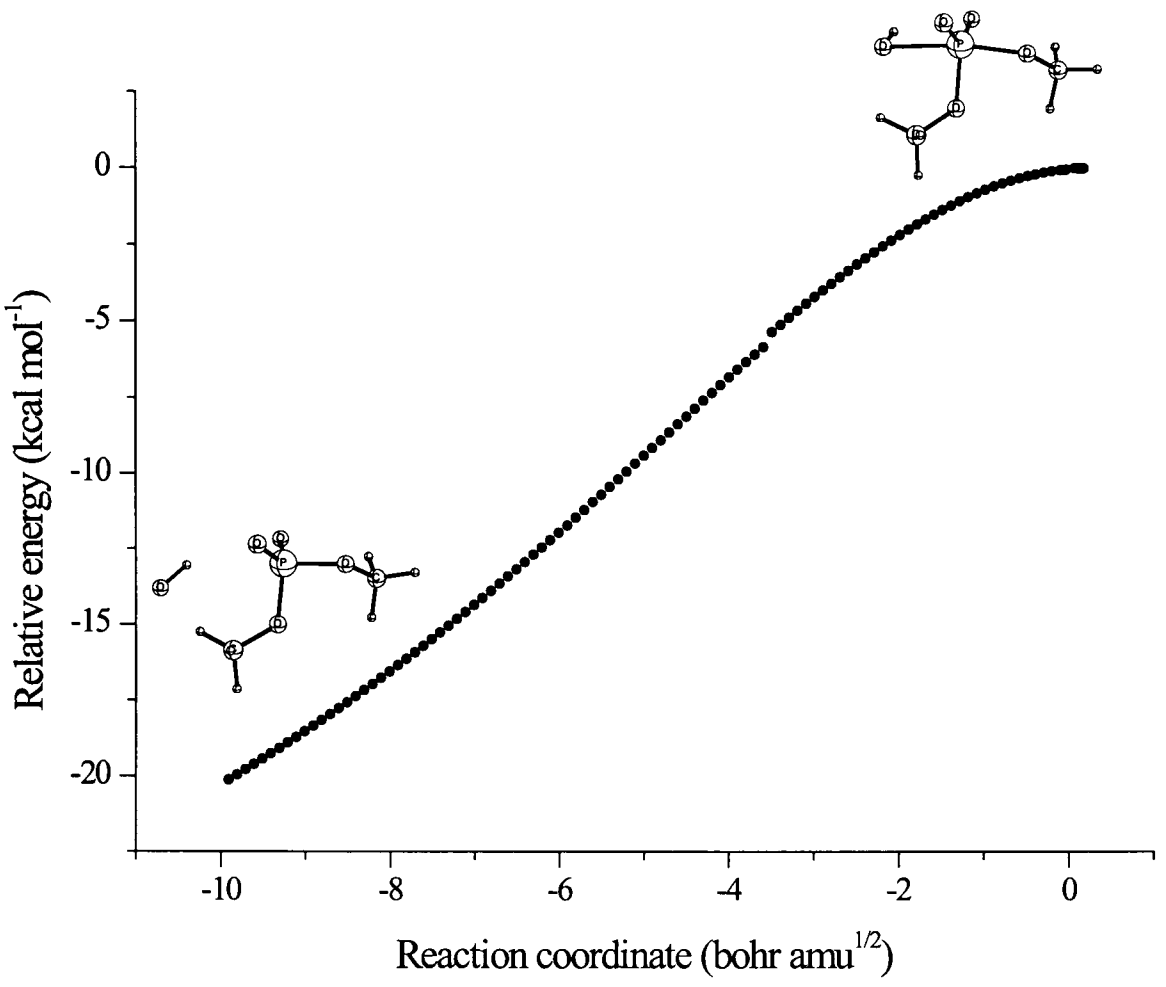


Figure 4.4: The IRC profile obtained for the base-catalysed hydrolysis of DMP. The structures of the transition state and the endpoint in the reactant direction are also shown.

of the reaction products. It was initially thought that this might indicate the presence of an intermediate along the productive trajectory. A series of minimisations scanning the reaction coordinate in this direction were performed but no evidence of the formation of a dianionic intermediate species was found.

The dissociation of the axial P–OCH_3 bond leads to the formation of the reaction products, monomethyl phosphate (MMP) and a methoxy anion. The geometry of MMP (Figure 4.2) is similar to that of DMP except that one of the methoxy ligands is replaced by a hydroxyl ligand.

Solution phase stationary points

It proved possible to fully optimise the structures of each of the reaction stationary point in solution. This proved to be extremely difficult, especially in the case of the transition state. This is because the use of the PCM method can often result in the

	P-O1	P-O2	P-O3	P-O4	P-O5	O1-H1	O4-C4	O5-C5
Gas Phase								
DMP	N/A	1.508	1.511	1.672	1.679	0.974	1.417	1.416
TS	2.215	1.532	1.518	1.852	1.713	0.969	1.387	1.407
MMP	1.668	1.504	1.514	N/A	1.668	0.970	N/A	1.414
MeO^-	N/A	N/A	N/A	N/A	N/A	N/A	1.341	N/A
Solution								
DMP	N/A	1.514	1.514	1.649	1.655	0.996	1.432	1.433
TS	2.310	1.525	1.523	1.762	1.671	0.979	1.414	1.428
MMP	1.649	1.510	1.514	N/A	1.659	0.976	N/A	1.434
MeO^-	N/A	N/A	N/A	N/A	N/A	N/A	1.375	N/A

Table 4.1: Key bond lengths (Å) for the stationary points in the DMP + OH^- reaction, calculated using the the B3LYP/6-31+G* level of theory, in the gas phase and in solution using the PCM representation of solvent.

	O1-P-O4	O2-P-O3	O2-P-O5	O3-P-O5	O1-P-O5	O5-P-O4	O1-P-O2	O3-P-O4
Gas Phase								
DMP	N/A	122.29	108.09	109.86	N/A	94.64	N/A	108.59
TS	170.99	131.42	113.30	114.47	83.57	87.55	86.10	92.27
MMP	N/A	123.48	108.51	108.24	96.61	N/A	109.10	N/A
Solution								
DMP	N/A	119.79	109.83	109.52	N/A	95.47	N/A	109.52
TS	174.74	128.62	114.32	114.30	82.47	91.71	83.67	96.00
MMP	N/A	120.90	109.29	109.40	97.77	N/A	107.55	N/A

Table 4.2: Key bond angles, in degrees, for the stationary points in the DMP + OH^- reaction, calculated using the the B3LYP/6-31+G* level of theory, in the gas phase and in solution, using the PCM representation of solvent.

formation of discontinuities in the reaction field which give rise to non-continuous gradients that make geometry optimisations very time-consuming and challenging. The validity of the transition state was confirmed by vibrational analysis which confirmed that the Hessian matrix had a single negative eigenvalue. Selected geometric parameters are presented in Tables 4.1, 4.2, and 4.3.

It was found that, in solution, the reaction also proceeds via an associative mechanism and the formation of a single transition state. The transition state geometry, shown in Figure 4.3(b), is more reactant-like than the gas phase equivalent with the partial bond between the incoming oxygen and the phosphorus being 2.31 Å long compared to 2.22 Å in the gas phase. Similarly, the distance between the

	O1-P-O5-C5	O1-P-O4-C4	H1-O1-P-O4
Gas Phase			
DMP	N/A	N/A	N/A
TS	305.34	59.33	257.94
MMP	174.94	N/A	N/A
PCM			
DMP	N/A	N/A	N/A
TS	302.14	69.29	257.28
MMP	176.01	N/A	N/A

Table 4.3: Key dihedral angles, in degrees, for the stationary points in the DMP + OH^- reaction, calculated using the the B3LYP/6-31+G* level of theory, in the gas phase and in solution using the PCM representation of solvent.

phosphorus and the expelled oxygen (O4) decreases from 1.85 Å in the gas phase to 1.76 Å in solution. The angle of approach between the nucleophile and the leaving group is almost linear at 175°. The orientation of the equatorial methoxy group is is similar to that encountered in the gas phase transition state.

It can be concluded that the presence of solvent facilitates the formation of an 'earlier' transition through electronic stabilisation of the incoming hydroxyl anion. Therefore, solvent should also have an effect on the relative electronic structures of the gas phase and solution transition states.

Atomic Charges

The extent to which solvation alters the electronic structure of the reaction stationary points can be gauged by calculating the electrostatic potential fitted atomic charges using the CHELPG scheme. These are shown in Table 4.4. It is apparent that, in solution, the magnitude of the partial charges in the transition states, especially for the phosphorus and the oxygens, have increased due to the polarising effect of the solvent. The greatest difference occurs between the two transition states where the incoming oxygen (O1) which has 0.2 e more negative charge in solution. This is due to a combination of the polarising effects of the solvent and the 'earlier' nature of the solution transition state.

	P	O1	O2	O3	O4	O5
Gas Phase						
OH-	N/A	-1.294	N/A	N/A	N/A	N/A
DMP	1.329	N/A	-0.814	-0.814	-0.540	-0.540
TS	1.631	-1.086	-0.876	-0.883	-0.820	-0.611
MeO^-	N/A	N/A	N/A	N/A	-1.135	N/A
MMP	1.442	-0.759	-0.877	-0.818	N/A	-0.553
Solution						
OH-	N/A	-1.426	N/A	N/A	N/A	N/A
DMP	1.508	N/A	-0.917	-0.905	-0.624	-0.594
TS	1.654	-1.269	-0.944	-0.905	-0.834	-0.513
MeO^-	N/A	N/A	N/A	N/A	-1.328	N/A
MMP	1.457	-0.800	-0.882	-0.917	N/A	-0.580

Table 4.4: Key electrostatic potential derived atomic charges (e) calculated using the B3LYP/6-31+G* level of theory for the stationary points in the DMP + OH^- reaction in both the gas phase and in solution (PCM).

Energetics

The relative energies of both the gas phase- and solution-optimised stationary points are shown in Table 4.5. In the gas phase, the free energy barrier, ΔG_{gas}, for the reaction was calculated to be 95.54 kcal mol^{-1}. When the effects of solvation were accounted for by the determination of the solvation free energy at the B3LYP(PCM)/6-31+G* level of theory, the energy barrier was reduced to 38.33 kcal mol^{-1}. The overall free energy change for the reaction was computed to be -12.46 kcal $mol^{-1}$and 0.07 kcal mol^{-1}in the gas phase and with solvation free energies included, respectively.

When ΔG_{gas} and $\Delta G_{solution}$ were calculated in a similar manner for the solution optimised geometries, different values were obtained. The value for ΔG_{gas} was determined to be 96.36 kcal mol^{-1} which is 0.8 kcal mol^{-1} higher than the equivalent value for the gas phase optimised reaction profile. However, the value of $\Delta\Delta G_{solv}$ for the solution optimised transition state was determined to be 3.6 kcal mol^{-1} larger than the gas phase transition state value. This had the effect of reducing the solution free energy barrier for the reaction to 35.53 kcal mol^{-1}. The free energy change for the reaction in solution was computed to be -1.39 kcal mol^{-1}.

	ΔE	ΔZPE	ΔG_{gas}	$\Delta\Delta G_{solv}$	$\Delta G_{solution}$
Gas phase geometries					
Reacts.	0.00	0.00	0.00	0.00	0.00
TS	85.69	0.89	95.54	-57.21	38.33
Prods.	-9.24	-0.96	-12.46	-12.39	0.07
Solution geometries					
Reacts.	0.00	0.00	0.00	0.00	0.00
TS	84.35	1.57	96.36	-60.83	35.53
Prods.	-8.41	0.34	-10.23	8.85	-1.39

Table 4.5: Relative energies (kcal mol^{-1}) of the stationary points of the DMP + OH^- reaction determined at B3LYP/6-31+G* and B3LYP(PCM)/6-31+G* levels. ΔE relative electronic energy of each point relative to the reactants, ΔZPE gives the relative zero-point energy corrections, and ΔG_{gas} denotes the relative gas phase free energies. $\Delta\Delta G_{solv}$ is the relative free energy of solvation and $\Delta G_{solution}$ is the relative free energy in solution which is calculated by adding ΔG_{gas} and $\Delta\Delta G_{solv}$.

Therefore, it can be concluded, using the relevant energy profiles that the presence of solvent reduces the free energy barrier for the reaction by 60.0 kcal mol^{-1} from 95.5 kcal mol^{-1} in the gas phase to 35.5 kcal mol^{-1} in solution. This is in agreement with the Born model of solvation [80,81].

Conclusions

The base-catalysed hydrolysis of monoanionic phosphate esters has been extensively studied using both experimental [89,165,166] and theoretical [168,171,184,186,187,194] methods. Early theoretical studies of the hydrolysis of DMP [171], employing the STO-3G basis set, indicated the presence of a stable dianionic intermediate along the reaction pathway. However, subsequent studies using the 3-21+G* basis set indicated that this intermediate was merely an artifact of the low-level basis set employed. The results presented here confirm that the base-catalysed hydrolysis of DMP in the gas phase does not involve the formation of a stable dianionic intermediate.

The optimisation of the transition state geometry in solution for this reaction represents a considerable improvement on previous studies of this reaction where the effect of solution was estimated by single-point energy evaluations on gas phase

geometries. The fact that the transition state geometry was earlier along the reaction coordinate in solution highlighted the importance of optimising geometries in solution.

The value of 35.53 kcal mol^{-1} obtained for the free energy of activation in solution is in good agreement with the experimental value of approximately 34.8 kcal mol^{-1} determined by Radzicka and Wolfenden[2] [89]. However, it not known to what extent this experimental free energy barrier includes the participation of an acid-catalysed mechanism as well as a base-catalysed mechanism.

Finally, contrary to speculation [4], no stable dianionic intermediate was shown to exist in solution. Whether or not this is due to the use of a continuum solvent representation which does not account for individual solute-solvent interactions will be will be examined in Chapters 6 and 7 where the reaction is studied using an atomistic representation of solvent.

4.3.2 Acid catalysed hydrolysis of dimethyl phosphate

The acid catalysed hydrolysis of DMP, shown in Figure 4.5, involves nucleophilic attack on DMP by a water molecule as opposed to the hydroxyl anion involved in the base-catalysed reaction. It is thought that the reaction can proceed via one of two alternate mechanistic pathways [3]. In the first mechanism, a proton is transferred from the water molecule to the phosphate prior to nucleophilic attack by the resultant OH^- anion. In the second, the proton transfer is concerted with the nucleophilic attack. In other words, the proton is transferred as the new P–O bond is being formed.

The reaction stationary point geometries were optimised in the gas phase and continuum based solvation calculations were carried out to evaluate the effect of solvent on the reaction energy profile. The optimisation of stationary point geometries in solution proved not to be possible.

---

[2]This value was arrived at by combining the two experimentally determined values $\Delta H = +26.5$ kcal mol^{-1} and $-T\Delta S = -8.3$ kcal mol^{-1}.

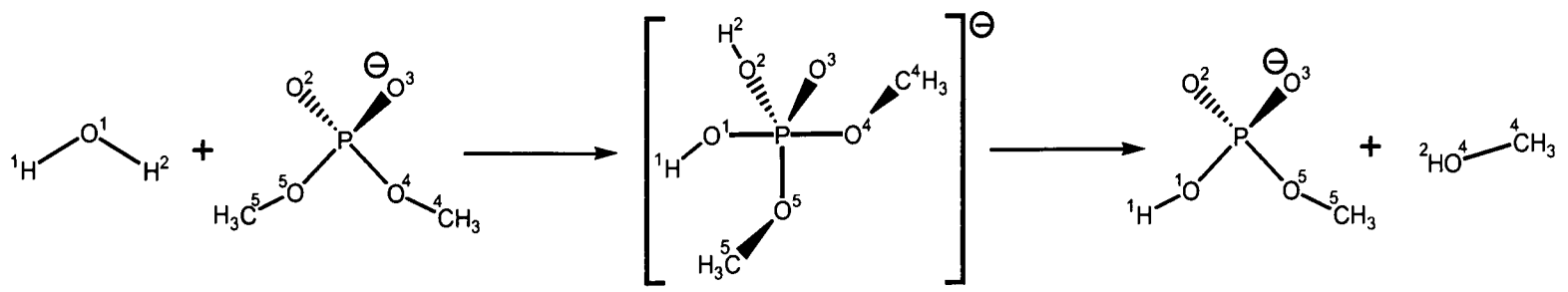


Figure 4.5: The general reaction scheme for the acid catalysed hydrolysis of dimethyl phosphate. The atom numbering scheme referred to in the text is also shown.

Stationary points

It was shown that, in the gas phase, the acid-catalysed hydrolysis of DMP proceeds via the formation of three transition states and two intermediates. The transfer of a proton from the water was observed to occur as the new P–O1 bond is being formed. The stationary point geometries are shown in Figure 4.6 and key bond lengths, bond angles, and dihedral angles are given in Tables 4.6, 4.7, and 4.8 respectively.

The geometry of the first transition state (TS1) was optimised by starting from an initial structure where the water molecule oxygen was placed 3.0 Å away from the phosphorus atom, therefore not biasing the reaction mechanism towards either of the two alternatives mentioned earlier in this section. In TS1, the bond between the incoming oxygen and the phosphorus is partially formed at 2.22 Å. The bond between the phosphorus and the leaving oxygen (P–O4) has only slightly lengthened by 0.02 Å to 1.69 Å. The structure of the transition state is distorted trigonal bipyramidal and the angle of approach is almost linear ($|\angle$O1-P-O4$| = 164.7°$), indicating axial attack and departure. Just as in the transition state determined for the base-catalysed reaction, the equatorial methoxy group is oriented towards the incoming oxygen.

A proton (H2) has begun to dissociate from the incoming water and has almost completely transferred to one of the phosphate non-bridging oxygens, O2. This new bond is almost fully formed with a bond length of 1.01 Å. The distance between the proton and the water oxygen has lengthened from 0.97 Å to 1.64 Å . Thus, it can be concluded that proton transfer occurs before the transition state is reached and

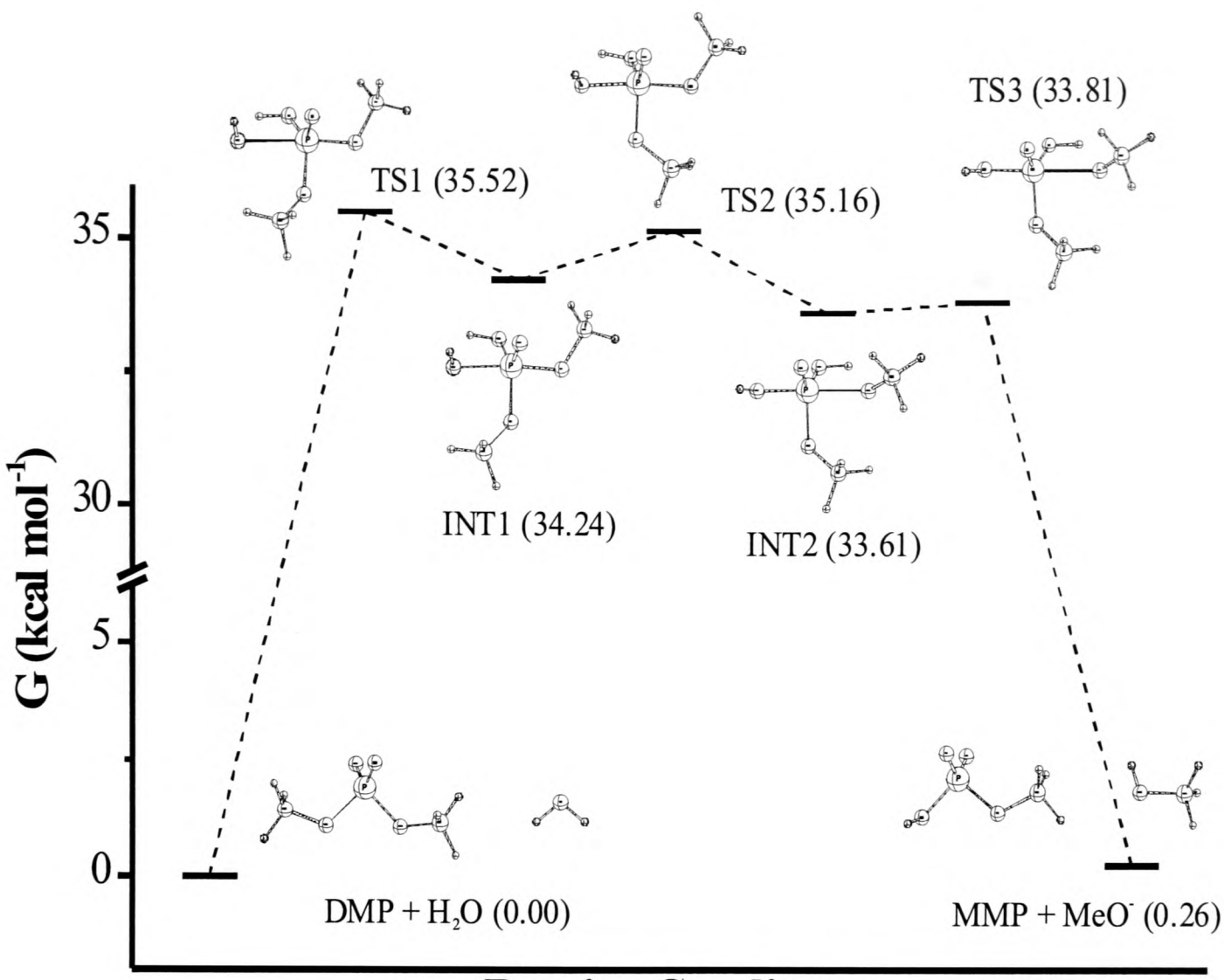


Figure 4.6: The reaction coordinate for the gas phase reaction of DMP + H_2O calculated using the B3LYP/6-31+G* level of theory. The relative gas phase free energy (kcal/mol) of each stationary point is shown in brackets alongside its structure.

that the transition state TS1 corresponds to the formation of P–O1 bond. An IRC calculation was started from the transition state and the profile is shown in Figure 4.7 along with the geometries of selected points in this region of the PES. The reaction profile obtained confirmed that the transition state is on the correct PES and also showed that proton transfer from the water to O2 occurs before the transition state is formed.

By continuing along the reaction coordinate from TS1, an intermediate species, INT1, was located. In the structure of this stationary point, the bond between the attacking oxygen, O1, and the phosphorus is almost formed being, at 1.87 Å, 0.35 Å

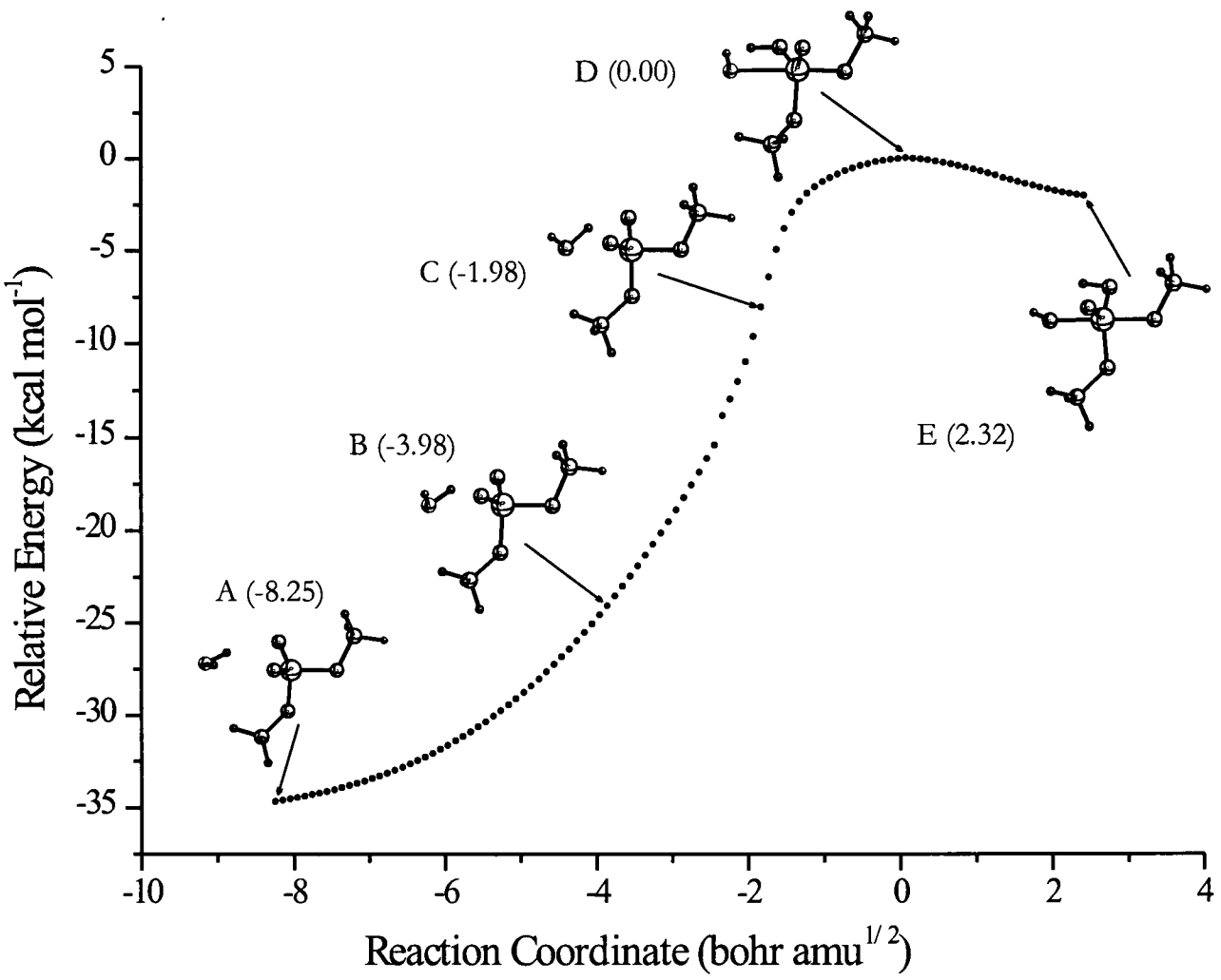


Figure 4.7: The energy profile constructed from the results of B3LYP/6-31+G* IRC calculation starting from TS1. Selected structures are depicted with the value of the reaction coordinate at the particular point they occur shown in brackets alongside.

shorter than it is in TS1. The bond between the phosphorus and the leaving oxygen is only 0.06 Å longer than it is in DMP confirming the reaction has an associative mechanism. The transferred proton, H2, is oriented towards the incoming oxygen and the bond between it and the non-bridging oxygen O2 is fully formed at 0.97 Å.

Further movement along the reaction coordinate resulted in the location of another transition state, TS2. The structure of this transition state is very similar to that of INT1 and the bond between the attacking oxygen and the phosphorus has only shortened by 0.06 Å while the bond between the phosphorus and the leaving group has lengthened by 0.02 Å. The proton transferred from the attacking water is still oriented towards the attacking oxygen. The main difference between the geometry of TS2 and the preceding intermediate and transition state is that the equatorial methoxy group is now oriented towards the leaving methoxy ligand. This is evidenced by the value of

	P-O1	P-O2	P-O3	P-O4	P-O5	O1-H2	O2-H2	O4-H2	O4-C4	O5-C5
DMP	N/A	1.508	1.511	1.672	1.679	0.969	N/A	N/A	1.417	1.416
TS1	2.215	1.633	1.509	1.691	1.637	1.642	1.008	3.268	1.413	1.426
INT1	1.866	1.675	1.526	1.731	1.665	1.921	0.974	3.244	1.405	1.416
TS2	1.802	1.682	1.527	1.757	1.670	1.948	0.974	3.243	1.405	1.418
INT2	1.713	1.654	1.520	1.962	1.670	3.200	0.978	1.804	1.396	1.417
TS3	1.694	1.633	1.512	2.140	1.656	3.216	0.992	1.694	1.388	1.423
MMP	1.668	1.504	1.514	N/A	1.668	N/A	N/A	0.969	1.425	1.414

Table 4.6: Key bond lengths from the stationary points of the reaction of DMP + H_2O using the B3LYP/6-31+G* level of theory. All values are given in Angstroms (Å).

	O1-P-O4	O2-P-O3	O2-P-O5	O3-P-O5	O1-P-O5	O5-P-O4	O1-P-O2	O3-P-O4
DMP	N/A	122.29	108.09	109.86	N/A	94.64	N/A	108.59
TS1	164.74	122.98	113.33	120.03	86.18	89.86	74.76	104.98
INT1	167.25	124.42	113.12	122.31	89.56	85.11	83.57	100.00
TS2	166.48	123.99	114.26	121.58	85.66	88.76	84.75	98.37
INT2	165.14	129.29	112.31	117.83	90.26	80.40	99.55	94.15
TS3	163.65	127.86	112.81	116.81	92.59	86.18	89.60	92.43
MMP	N/A	123.48	108.51	108.24	96.61	N/A	109.10	N/A

Table 4.7: Key bond angles from the stationary points of the reaction of DMP + H_2O in the gas phase using the B3LYP/6-31+G* level of theory. All values are given in degrees.

the dihedral angle O1-P-O5-C5 changing from 323° to 189° between INT1 and TS2. It can be postulated that the reason for this rotation is that the CH_3 moiety can form a positive electrostatic interaction with the leaving oxygen which would stabilise its departure and thus lower the energy barrier corresponding to the cleavage of the axial P–OCH_3 bond.

	O1-P-O5-C5	O1-P-O4-C4	H1-O1-P-O4	O4-P-O2-H2
DMP	N/A	N/A	N/A	N/A
TS1	310.51	107.04	204.23	175.04
INT1	323.31	120.28	189.52	177.73
TS2	189.47	124.60	177.35	178.64
INT2	201.89	163.69	181.96	356.88
TS3	205.50	158.87	178.23	356.04
MMP	174.94	N/A	N/A	N/A

Table 4.8: Key dihedral angles from the stationary points of the reaction of DMP + H_2O in the gas phase using the B3LYP/6-31+G* level of theory. All values are given in degrees.

The next stationary point located was determined to be another intermediate, INT2. Compared to TS2, the P–O1 bond in INT2 is 0.09 Å shorter and is almost fully formed at 1.71 Å. The distance between the phosphorus and the leaving oxygen has lengthened by 0.20 Å to 1.96 Å. Therefore, it can be concluded that the reaction coordinate is now in the region of the PES which corresponds to the cleavage of the P–OCH_3 bond. The P–O2 bond has rotated so as to orient the proton that originally resided on the water molecule towards the leaving methoxy group and poised for transfer at 1.80 Å away from O4.

Finally, the reaction proceeds towards product formation via a third and final transition state. By now the P–O1 bond has almost fully formed at 1.69 Å and the axial methoxy group has begun to dissociate from the phosphorus (|P-O4| = 2.14 Å). Proton transfer to the leaving group has not yet occurred. However, the distance between the proton and the leaving oxygen has decreased by 0.11 Å compared to INT2 to 1.69 Å. This is accompanied by a slight dissociation of the O2-H2 bond.

In order to ensure that proton transfer does indeed occur, an IRC profile, shown in Figure 4.8, was calculated from this transition state. Selected geometries from the profile are also shown to illustrate the chemical processes occurring in this region of the PES. The profile confirms that, as the leaving group dissociates, proton transfer occurs from the remaining MMP moiety to the leaving group.

Charge distribution

Table 4.9 lists the gas phase electrostatic potential derived (CHELPG) charges for a selection of key atoms at reaction stationary point. As the reaction proceeds towards product formation, it is observed that electron density is transferred from the incoming oxygen and the phosphorus to the equatorial oxygens and the leaving group.

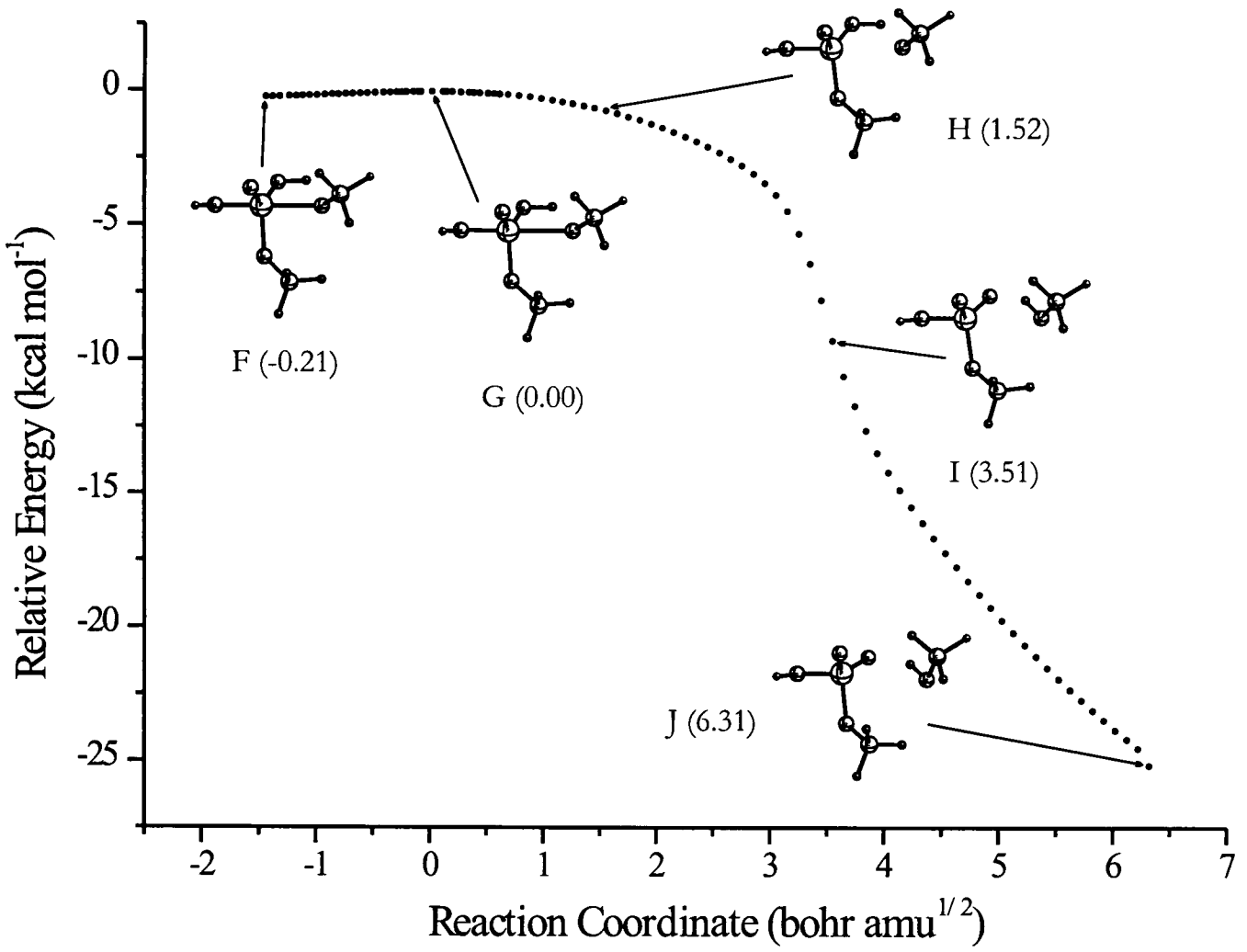


Figure 4.8: Energy profiles constructed from the results of B3LYP/6-31+G* IRC calculations starting from TS3. Selected structures are depicted with the value of the reaction coordinate at the particular point they occur shown in brackets alongside.

The magnitude of the charges are larger in solution than in the gas phase. This is indicative of the polarization introduced by solvation and is particularly noticeable for the oxygen atoms.

Energetics

The energies of each stationary point relative to the reactants are shown in Table 4.10. The relative free energy in solution, $\Delta G_{solution}$, of each stationary point was calculated by adding the relative free energy of solvation, $\Delta\Delta G_{solv}$, to the gas phase free energy, ΔG_{gas}.

	Gas phase						*PCM*					
	P	O1	O2	O3	O4	O5	P	O1	O2	O3	O4	O5
H_2O	N/A	-0.81	N/A	N/A	N/A	N/A	N/A	-0.92	N/A	N/A	N/A	N/A
DMP	1.33	N/A	-0.81	-0.81	-0.54	-0.54	1.46	N/A	-0.88	-0.89	-0.61	-0.59
TS1	1.50	-0.96	-0.67	-0.78	-0.67	-0.50	1.55	-1.03	-0.70	-0.83	-0.72	-0.56
INT1	1.50	-0.84	-0.68	-0.82	-0.68	-0.50	1.56	-0.89	-0.72	-0.89	-0.76	-0.57
TS2	1.54	-0.93	-0.79	-0.88	-0.48	-0.49	1.63	-1.00	-0.84	-0.95	-0.54	-0.59
INT2	1.43	-0.80	-0.64	-0.82	-0.64	-0.50	1.50	-0.88	-0.73	-0.88	-0.70	-0.56
TS3	1.42	-0.78	-0.66	-0.80	-0.71	-0.51	1.48	-0.86	-0.74	-0.85	-0.78	-0.55
MeOH	N/A	N/A	N/A	N/A	-0.70	N/A	N/A	N/A	N/A	N/A	-0.80	N/A
MMP	1.44	-0.76	-0.88	-0.82	N/A	-0.55	1.42	-0.79	-0.86	-0.90	N/A	-0.57

Table 4.9: Key electrostatic potential derived atomic charges (e) for the DMP + H_2O reaction in the gas phase and in solution (PCM) calculated using a B3LYP/6-31+G* potential.

In the gas phase, the first transition state, TS1, has the highest energy relative to the reactants (35.52 kcal mol^{-1}) thus confirming that the formation of the new P–O bond is the rate-limiting step for the reaction.

	ΔE	ΔZPE	ΔG_{gas}	$\Delta\Delta G_{solv}$	$\Delta G_{solution}$
Reacts.	0.00	0.00	0.00	0.00	0.00
TS1	22.96	2.29	35.52	13.63	49.15
INT1	22.70	3.20	34.24	12.21	46.45
TS2	21.65	2.91	35.16	11.54	46.70
INT2	21.55	2.62	33.61	13.11	46.72
TS3	22.96	2.29	33.81	14.07	47.88
Prods.	0.95	1.27	0.26	0.57	0.83

Table 4.10: Relative energies (kcal mol^{-1}) of the stationary points of the DMP + H_2O reaction determined at B3LYP/6-31+G* level of theory. ΔE relative electronic energy of each point relative to the reactants, ΔZPE gives the relative zero-point energy corrections, and ΔG_{gas} denotes the relative gas phase free energies. $\Delta\Delta G_{solv}$ is the relative free energy of solvation and $\Delta G_{solution}$ is the relative free energy in solution which is calculated by adding ΔG_{gas} and $\Delta\Delta G_{solv}$.

The central region of the potential energy surface, which is dominated by the rotation of the equatorial methoxy group and the proton transfer proton to the leaving group, is very flat. The first intermediate, INT1, is probably not kinetically significant since it is only 1.3 kcal mol^{-1} and 1.1 kcal mol^{-1} more stable than TS1 and TS2 respectively. The same can be said for the later intermediate, INT2, which has only 1.6 kcal mol^{-1} and 0.2 kcal mol^{-1} lower free energy than TS2 and TS3 respectively.

The overall free energy barrier for the reaction in the gas phase is 35.5 kcal mol^{-1}. When compared to an equivalent value of 95.5 kcal mol^{-1} for the base-catalysed

hydrolysis, it can be concluded that the monoanionic, rate-limiting phosphorane transition state formed in this reaction is considerably more stable than the dianionic phosphorane transition state encountered in the base-catalysed reaction.

When the relative free energies of the stationary points in solution are compared it is observed that solvation has the effect of increasing the energy barrier for the reaction from 35.5 kcal mol^{-1} to 49.2 kcal mol^{-1}. This increase in the free energy barrier on going from the gas phase to solution can be explained in terms of the larger collective stabilisation experienced by a molecule with no net charge and a monoanionic molecule compared to that experienced by a single larger monoanionic species.

Solvation also has an effect on the relative energies of the reaction intermediates and transition states. The intermediate INT1 becomes more stable relative to TS1 with the energy difference increasing to 2.7 kcal mol^{-1} compared to 1.3 kcal mol^{-1} in the gas phase. However, the energy difference between INT1 and TS2 decreases from 0.9 kcal mol^{-1} in the gas phase to 0.3 kcal mol^{-1} in solution.

The same pattern is observed for the other reaction intermediate, INT2. The energy difference between this intermediate and TS3, the transition state corresponding to the cleavage of the P–OCH_3 bond increases from 0.2 kcal mol^{-1} in the gas phase to 1.2 kcal mol^{-1} in solution. However, the effect of solvation causes INT2 to be 0.02 kcal mol^{-1} less stable than TS2 thus asking the question whether an equivalent intermediate exists in solution. The overall free energy change for the reaction was determined to be 0.26 kcal mol^{-1} in the gas phase and 0.83 kcal mol^{-1} in solution.

Conclusions

The reaction energy profile determined is considerably more intricate than that determined for the base-catalysed hydrolysis reaction. However, both are similar in that the rate-limiting transition state for each reaction is 'early' and corresponds to the formation of the new P–O bond.

The activation free energy for this reaction was computed to be 35.5 kcal mol^{-1} in the gas phase, 64 kcal mol^{-1} lower than the equivalent barrier for the base-catalysed reaction. This difference can be attributed to the greater stability of the monoanionic phosphorane transition state compared to the dianionic transition state encountered in the base-catalysed reaction which has more electron density to distribute.

The 'middle' region of the reaction profile is very flat and the relative energies of the transition states and intermediates do not differ by more than 2 kcal mol^{-1} in the gas phase. This stands to reason since the chemical processes which dominate this region of the PES, i.e. proton transfer and bond rotation, have very shallow energy wells. Therefore, it is doubtful that a stable monoanionic phosphorane is formed. This is in agreement with the results of a similar study by Florià*n et al.* which concluded that the acid-catalysed hydrolysis of monomethyl phosphate was not likely to result in the formation of a stable intermediate [167].

4.3.3 Acid catalysed hydrolysis of protonated dimethyl phosphate

The final mechanism considered for the hydrolysis of DMP is nucleophilic attack by a water molecule on a DMP molecule that has already been protonated by a general acid. The reaction scheme is depicted in Figure 4.9. The protonation of the substrate is assumed to occur before nucleophilic attack occurs and it is also assumed that the added proton resides on one of the phosphate non-bridging oxygen atoms. It should be noted that, in this reaction, neither reactant has a net charge and therefore any phosphorane species formed during the reaction will also be uncharged. The reaction was examined in the gas phase at the B3LYP/6-31+G* level of theory and and the effect of solvent on the reaction was estimated using the PCM solvent representation. Again, it was not possible to optimise the stationary point geometries in solution.

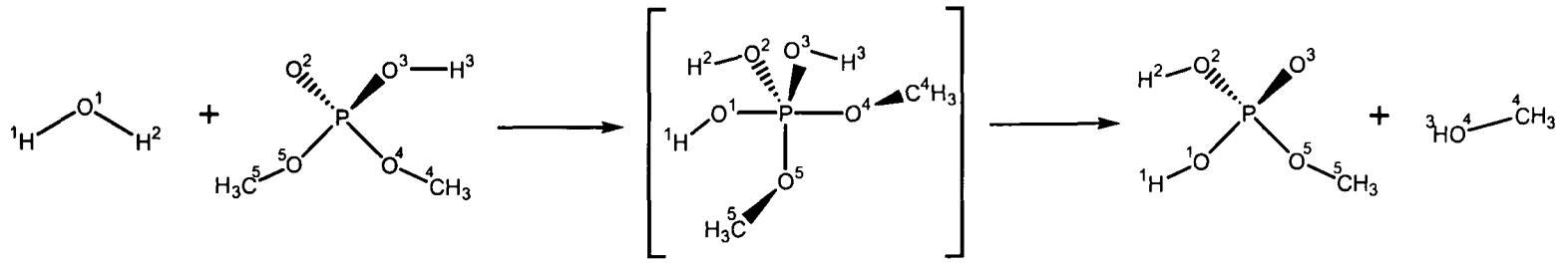


Figure 4.9: The reaction scheme for the base-catalysed hydrolysis of DMP. The atom numbering scheme referred to in the text is also shown.

In the gas phase, the reaction was again found to proceed via an associative pathway which involved the formation of two transition states and a stable intermediate. Each of the pentacoordinated phosphorane species has a trigonal bipyramidal geometry with the attacking and departing oxygens occupying the two axial positions. The stationary point geometries are depicted in Figure 4.10 and selected geometrical parameters are given in Tables 4.11, 4.12, and 4.13.

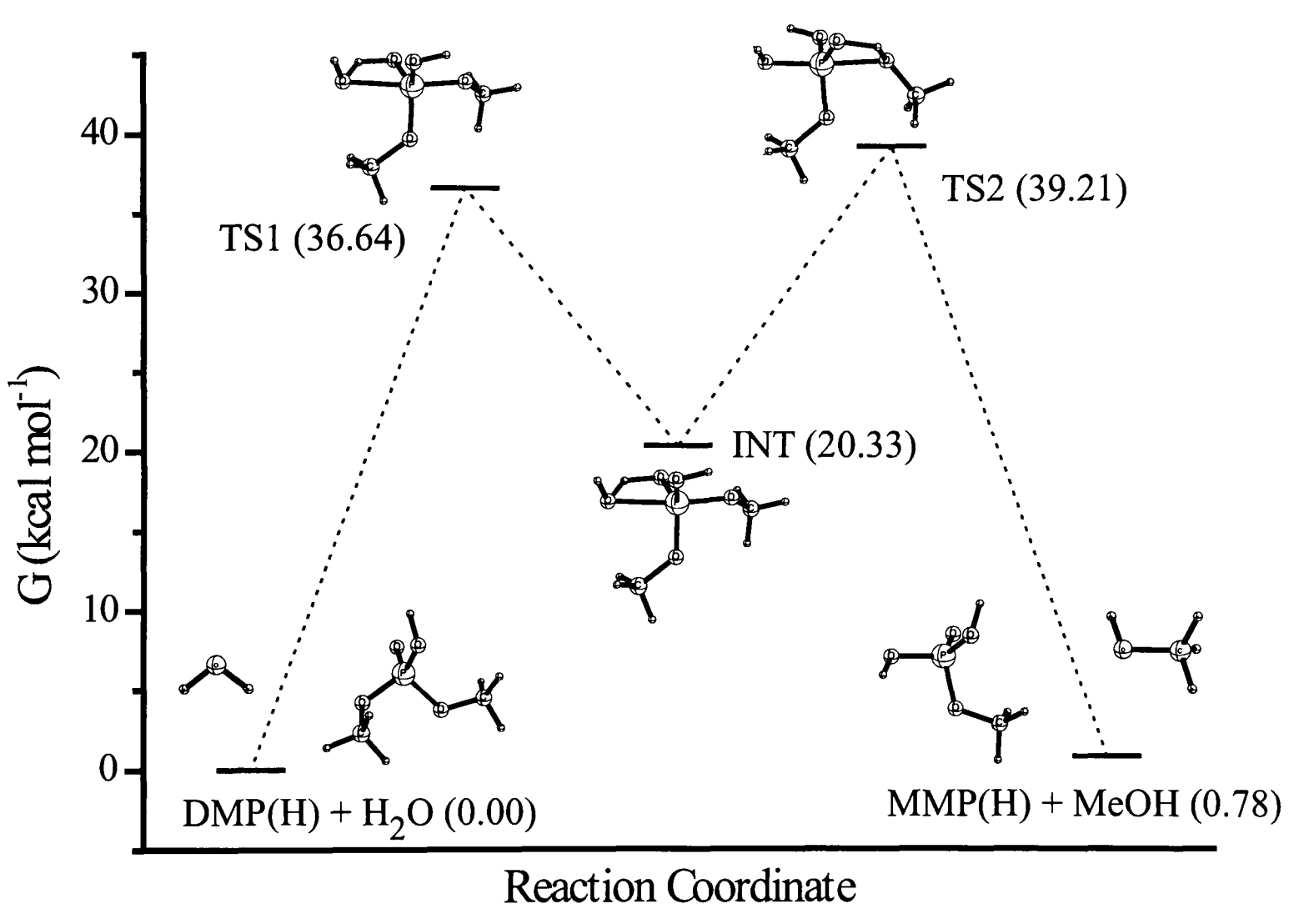


Figure 4.10: The reaction energy profile and stationary point geometries for the acid-catalysed hydrolysis of DMP(H). The free energy of each structure relative to the reactants is shown in brackets.

Stationary points

The optimised geometry of protonated dimethyl phosphate, DMP(H), has the four oxygen atoms arranged tetrahedrally about the phosphorus. The methoxy groups adopt a *trans, gauche* conformation which differs from the *gauche,gauche* conformation determined for unprotonated DMP in the gas phase.

By following the reaction coordinate, a saddle point was located and was confirmed to be a transition state by frequency analysis. This transition state, TS1, which corresponds to the formation of the new P–O bond between the phosphorus and the incoming oxygen, O1, which is 2.07 Å away from the phosphorus. The bond from the phosphorus to the axial methoxy oxygen is, at 1.64 Å, just 0.03 Å longer than it is in DMP(H). The angle of attack is linear ($|\angle$O1–P–O4$| = 173.1°$).

Just as in the acid-catalysed hydrolysis of unprotonated DMP, a proton is transferred from the incoming water to one of the non-bridging oxygens (O2). In this reaction however, the proton transfer occurs further along the reaction coordinate and is more concerted with the approach of the oxygen atom. This is evidenced by the bond between the proton and the non-bridging oxygen not being fully formed (|O2-H2| = 1.20 Å). An IRC calculation was started from this transition state and the profile obtained is shown in Figure 4.11.

	P-O1	P-O2	P-O3	P-O4	P-O5	O2-H2	O3-H3	O4-H3	O4-C4	O5-C5
React.	N/A	1.479	1.634	1.606	1.596	N/A	0.972	3.316	1.443	1.443
TS1	2.074	1.566	1.627	1.639	1.604	1.201	0.975	2.145	1.441	1.435
INT	1.716	1.639	1.647	1.681	1.627	0.973	0.974	2.020	1.427	1.436
TS2	1.678	1.625	1.570	1.941	1.616	0.973	1.249	1.216	1.425	1.440
Prod.	1.613	1.617	1.484	N/A	1.599	0.972	N/A	0.948	1.405	1.444

Table 4.11: Key bond lengths (Å) from the stationary points of the gas phase acid-catalysed hydrolysis of DMP(H) determined at the B3LYP/6-31+G* level of theory.

The next stationary point located on the reaction coordinate was confirmed to be an intermediate. The structure of this intermediate (INT) is again approximately trigonal bipyramidal. The bond between the incoming oxygen and the phosphorus is, at 1.72 Å,

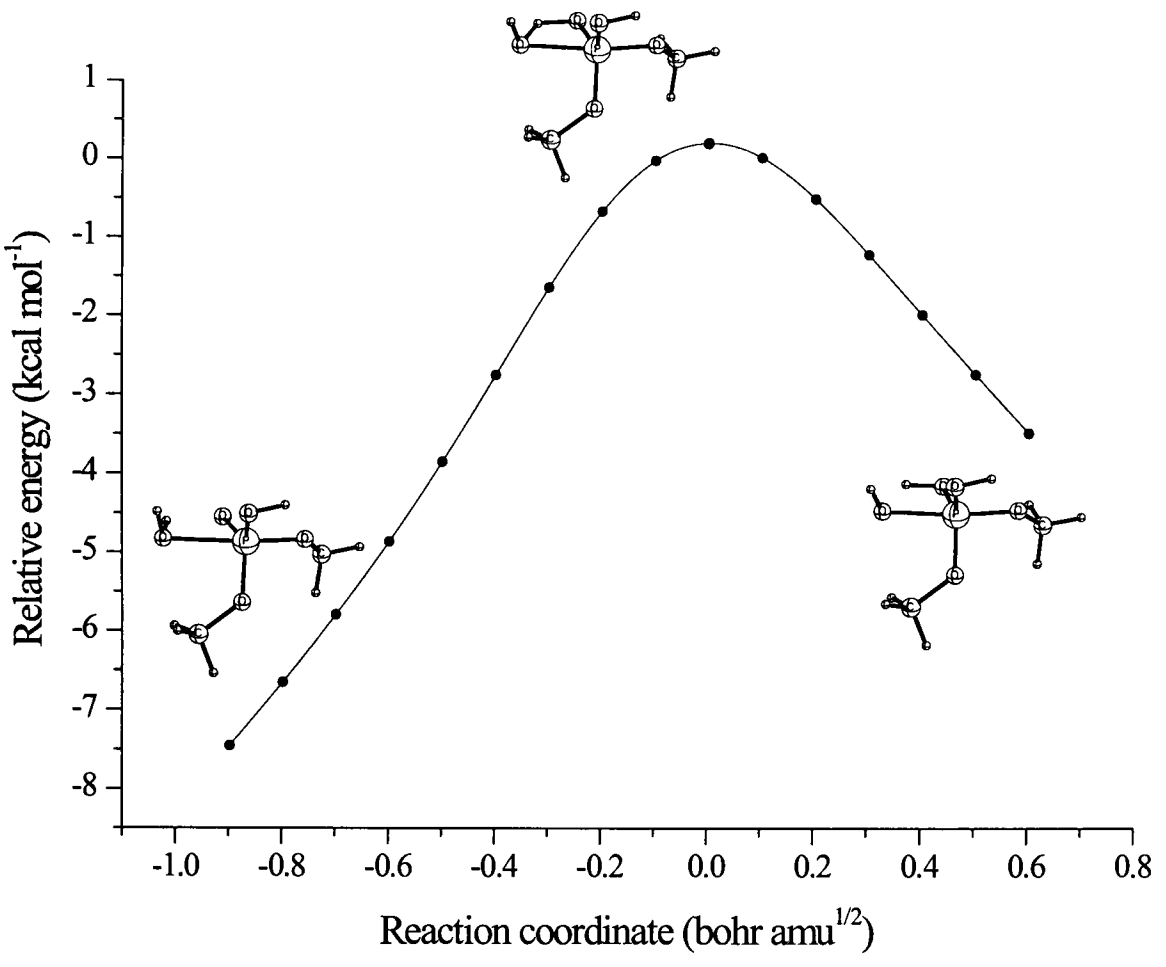


Figure 4.11: The energy profile constructed from the results of B3LYP/6-31+G* IRC calculation starting from the first transition state for the acid catalysed hydrolysis of protonated DMP. The structures of the transition state and the endpoints are also shown.

0.35 Å shorter than it is in TS1 and the bond from phosphorus to the leaving oxygen is only slightly dissociated being just 0.07 Å longer than it is in DMP(H). This indicates that both ligands are tightly bound to the phosphorus at this stage of the reaction and that an extreme form of the associative pathway is being followed. The proton (H2) that originally resided on the water molecule has fully bonded to the non-bridging oxygen (O2) with a typical O–H bond distance of 0.97 Å. The proton H3, bonded to the

	O1-P-O4	O2-P-O3	O2-P-O5	O3-P-O5	O1-P-O5	O5-P-O4	O1-P-O2	O3-P-O4
React.	N/A	113.91	114.40	105.45	N/A	101.78	N/A	101.42
TS1	173.08	121.67	114.42	119.06	90.67	96.18	74.73	93.04
INT	173.62	124.92	115.21	119.86	94.65	91.08	88.96	87.26
TS2	173.27	123.94	111.30	120.51	98.66	87.59	91.12	76.76
Prod.	N/A	113.69	100.74	119.36	101.17	N/A	106.45	N/A

Table 4.12: Key bond angles, in degrees, from the stationary points of the gas phase acid-catalysed hydrolysis of DMP(H) determined at the B3LYP/6-31+G* level of theory.

	O1-P-O5-C5	O1-P-O4-C4	H1-O1-P-O4	O4-P-O2-H2	O4-P-O3-H3
React.	N/A	N/A	N/A	N/A	148.03
TS1	8.09	-118.90	130.52	-171.17	-0.68
INT1	5.90	-146.22	-39.91	-173.38	-2.07
TS2	5.77	158.95	-38.41	-168.16	-2.83
Prod.	-72.00	N/A	N/A	N/A	N/A

Table 4.13: Key dihedral angles, in degrees, from the stationary points of the gas phase acid-catalysed hydrolysis of DMP(H) determined at the B3LYP/6-31+G* level of theory.

other non-bridging oxygen (O3), is now oriented towards the leaving, axial methoxy group but it is still strongly bonded to O3 with bond length of 0.97 Å.

The reaction then proceeds via the formation of another transition state. The geometry of this transition state, TS2, is product-like and graphical examination of the vibration corresponding to the imaginary frequency confirms that it corresponds to the cleavage of the axial P–OCH_3 bond. The bond between the attacking oxygen and the phosphorus is almost fully formed with the interatomic distance being 1.68 Å. The leaving group has began to dissociate from the phosphorus and the P–O4 bond length, 1.94 Å, is 0.26 Å longer than it was in the preceding intermediate. The proton transfer from O3 to the leaving group has been initiated and the proton H3 lies halfway between both oxygens (|O3-H3|= 1.25 Å and |O4-H3|= 1.22 Å). An IRC profile, shown in Figure 4.12, was calculated from this transition state and it demonstrated that the transition state is located on a potential energy surface appropriate to the reaction being studied. Finally, the reaction then proceeds to the formation of the products: methanol and protonated monomethyl phosphate, MMP(H).

Energetics

The relative free energies of the reaction stationary points in both the gas phase and in solution are shown in Table 4.14. The reaction was shown to proceed via the formation of two transition states and an intermediate.

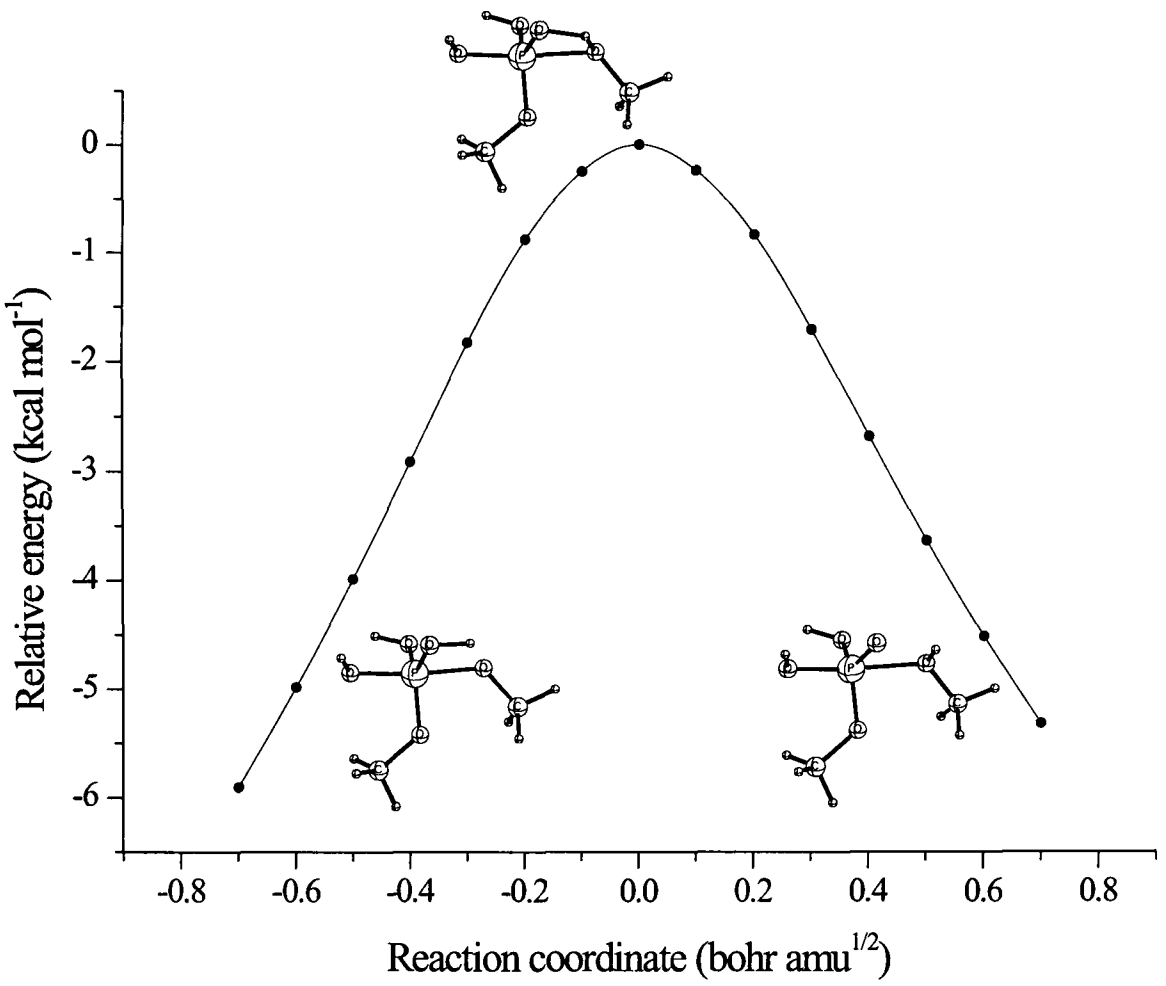


Figure 4.12: The energy profile constructed from the results of B3LYP/6-31+G* IRC calculation starting from the second transition state for the acid catalysed hydrolysis of protonated DMP. The structures of the transition state and the endpoints are also shown.

The free energy difference between the first transition state, which corresponds to the formation of the new P–O bond, and the reactants was computed to be 36.6 kcal mol^{-1} in the gas phase. This value is 1.1 kcal mol^{-1} higher than the barrier to the formation of the equivalent transition state in the acid-catalysed hydrolysis of the unprotonated substrate. Considering that, in this reaction, an uncharged phosphorane is being formed as opposed to a less stable monoanionic phosphorane, this value is, on first inspection, somewhat surprising. However, the relative differences can be reconciled to the relative ground state stability of protonated dimethyl phosphate compared to the the unprotonated substrate and the higher energetic cost of transferring a proton to a neutral species compared to a monoanionic species.

However, once proton transfer occurs the stability of the system increases and the intermediate, INT, is only 20.3 kcal mol^{-1} less stable than the reactants and is 16.3 kcal mol^{-1} more stable than the first transition state.

	ΔE	ΔZPE	ΔG_{gas}	$\Delta\Delta G_{solv}$	$\Delta G_{solution}$
R_B	0.00	0.00	0.00	0.00	0.00
TS1	25.23	0.72	36.64	12.09	48.73
INT	5.81	3.74	20.33	11.84	32.17
TS2	28.31	0.42	39.21	10.31	49.52
Prods.	1.46	0.70	0.77	-1.01	-0.24

Table 4.14: Relative enthalpies, E_{gas}, and free energies in the gas phase (G_{gas}) and in solution (G_{sol}) of stationary points of the DMP + H_2O reaction in both the gas phase. All energies are in kcal mol^{-1}.

The final transition state, TS2, which corresponds to the cleavage of the axial P–OCH_3 bond is 18.9 and 39.2 kcal mol^{-1} less stable than the intermediate and reactants respectively. This makes the formation of this transition state the rate-determining step for the reaction. Again, a large energy barrier has to be overcome to form this transition state. This barrier may correspond to the energetic cost of destroying the highly stable neutral phosphorane intermediate.

The second transition state, TS2, still has, in solution, the higher energy being 49.52 kcal mol^{-1} less stable than the reactants. TS1 is 48.73 kcal mol^{-1} less stable in solution. The free energy of the intermediate relative to the ground state increases by 11.84 kcal mol^{-1} in solution to a value of 32.17 kcal mol^{-1}. These increases in magnitude of the relative free energies can be explained in terms of the greater stabilisation experienced by two small neutrally-charged molecules compared to a single larger neutrally-charged molecule.

Finally, the overall free energy change for the reaction was computed to be 0.77 kcal mol^{-1} and -0.24 kcal mol^{-1} in the gas phase and in solution respectively.

Charges

The partial charges of each stationary point were calculated in both the gas phase and in solution. The values obtained are shown in Figure 4.15. Overall, the charge distributions show very similar trends to ones observed in the acid-catalysed hydrolysis of the unprotonated phosphate. The partial charges on the oxygen atoms are less negative due to the presence of neutrally charged species throughout the reaction. As usual, the presence of solvent causes more pronounced charge polarisation.

	Gas phase						*PCM*					
	P	O1	O2	O3	O4	O5	P	O1	O2	O3	O4	O5
H_2O	N/A	-0.81	N/A	N/A	N/A	N/A	N/A	-0.92	N/A	N/A	N/A	N/A
DMP(H)	1.16	N/A	-0.65	-0.62	-0.45	-0.35	1.22	N/A	-0.72	-0.64	-0.51	-0.38
TS1	1.25	-0.82	-0.62	-0.65	-0.55	-0.34	1.32	-0.87	-0.67	-0.68	-0.59	-0.38
INT	1.32	-0.81	-0.62	-0.72	-0.59	-0.36	1.38	-0.86	-0.67	-0.76	-0.64	-0.39
TS2	1.28	-0.78	-0.65	-0.74	-0.55	-0.29	1.35	-0.81	-0.72	-0.80	-0.62	-0.31
MeOH	N/A	N/A	N/A	N/A	-0.70	N/A	N/A	N/A	N/A	N/A	-0.80	N/A
MMP(H)	1.29	-0.66	-0.65	-0.65	N/A	-0.45	1.36	-0.69	-0.72	-0.72	N/A	-0.48

Table 4.15: Key electrostatic potential derived atomic charges (e) for the DMP(H) + H_2O reaction in the gas phase and in solution (PCM) calculated using a B3LYP/6-31+G* potential.

Conclusions

Of the three reactions examined in this study, this is the only one in which the formation of a stable phosphorane intermediate is observed. The reaction mechanism is very similar to that of the acid-catalysed hydrolysis of unprotonated DMP. In both reactions a proton is transferred from the incoming water molecule to the phosphate. In the hydrolysis of unprotonated DMP, the same proton is transferred from the water to the phosphate and from the phosphate to the leaving group. However, for the hydrolysis of the neutral phosphate, different protons are involved in each of the proton transfer steps.

This reaction mechanism is, however, biologically less important than the preceding two mechanistic alternatives presented since it has been found that neutral forms of phosphate diester exist in solution at very low concentrations under physiological conditions.

4.4 Summary

Each of the three possible mechanisms for the hydrolysis of dimethyl phosphate were examined. The base-catalysed hydrolysis reaction was shown to proceed via a single transition state in the gas phase and in solution. The geometries of both transition states were found to be reactant-like, although the solution transition state is found to be more reactant-like than the gas phase transition state. This result highlights the importance, where possible, of optimising geometries in solution.

The acid-catalysed hydrolysis of the unprotonated, monoanionic substrate is facilitated by a single proton being transferred from the nucleophile to the phosphate and then to the leaving group. These two transfers stabilise the approach of the nucleophile and the departure of the leaving group respectively and help bring about a significant reduction in the energy of activation for the reaction.

The final reaction considered was the nucleophilic attack by water on the neutral substrate. The reaction mechanism was determined to be similar to that of the acid-catalysed hydrolysis of the unprotonated substrate except that the proton transfers involved are effected by two different protons.

The presence of solvent, represented by a dielectric continuum, was found to greatly reduce the free energy of activation for the base-catalysed hydrolysis reaction and to increase the free-energy barriers of the two acid-catalysed reactions.

Chapter 5

A quantum chemical study of RNA hydrolysis

5.1 Introduction

This chapter presents the results of a quantum chemical study of the transesterification step of the base-catalysed hydrolysis of RNA. The overall reaction mechanism for the non-enzymatic hydrolysis of RNA is discussed in Chapter 3 (Figure 3.5). Up to now, theoretical studies of the hydrolysis of RNA have used small model compounds such as ethylene phosphate (EP) and methyl ethylene phosphate (MEP) as models for RNA [168,182,194,239]. In this study, methyl ribose phosphate, shown in Figure 5.1, is used as a model for RNA. This constitutes a considerably more realistic representation than both MEP and EP.

The hydrolysis of three different analogues of methyl ribose phosphate are considered in this study. These are referred to as A, B, and C and are shown in Figure 5.2. The three corresponding hydrolysis reactions are referred to as Reactions A, B, and C, respectively.

Experimental evidence suggests that for the non-enzymatic hydrolysis of RNA, the cleavage of the P-O5′ bond is the rate determining step [4,6] for the transesterification

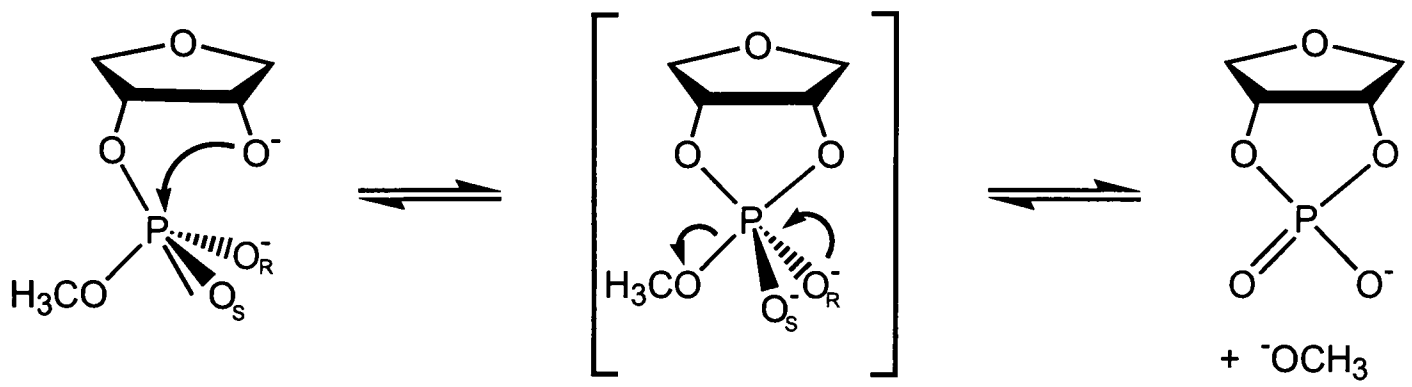


Figure 5.1: The mechanism of the first, transesterification step in the hydrolysis of the unmodified methyl ribose phosphate; the RNA analogue used in this study.

reaction. To examine this hypothesis, a theoretical study of the hydrolysis of unmodified methyl ribose phosphate, Reaction A, is presented.

In the second reaction studied, Reaction B, the 5′-oxygen is substituted by a sulphur atom to give 5′-thiomethyl ribose phosphate. The reasoning behind this substitution is that if the cleavage of the P-O5′ bond is indeed the rate-determining step in the transesterification reaction, the replacement of the leaving methoxy group with a thiomethyl moiety, a better leaving group [220, 221], will reduce the overall energy barrier for the reaction [135, 240].

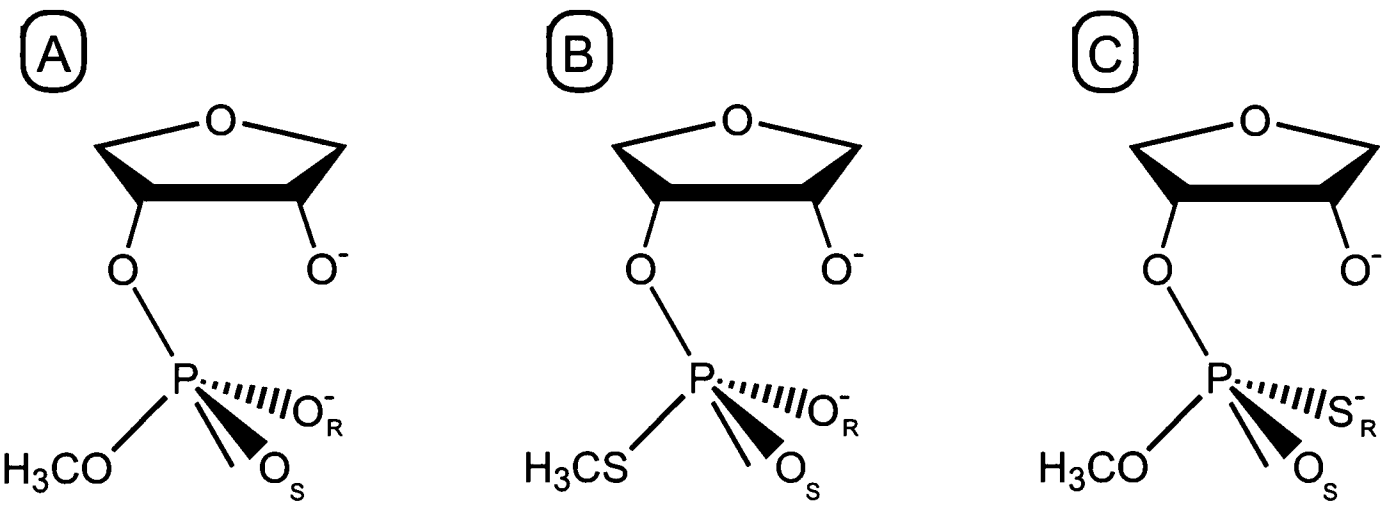


Figure 5.2: The three RNA analogues examined in this study. Molecules A, B and C correspond to the unmodified methyl ribose phosphate, the 5′ thiomethyl ribose phosphate, and the *pro*-R phosphothiolate respectively.

The analogue in the final reaction studied, Reaction C, has the *pro*-R non-bridging oxygen replaced by a sulphur atom to methyl ribose thiophosphate. This substitution, discussed in Chapter 3, Section 3.4.2, is also a commonly used mechanistic probe for

the hydrolysis of RNA. This study aims to quantitatively determine the effect of both of these thio substitutions on the reaction mechanism and energetics by comparing the results obtained for Reactions B and C to those obtained for Reaction A, the hydrolysis of the unmodified RNA analogue.

5.2 Methods

Initial structures

The starting point for phosphate ester hydrolysis in ribozymes is deprotonation of the adjacent 2′-OH group by a general base [4, 6]. There is considerable debate about the nature of this proton transfer step, in particular over whether the step is concerted with nucleophilic attack on the adjacent phosphorus. Kinetic studies of the hammerhead ribozyme-catalysed reaction indicate that the proton transfer is nonconcerted with nucleophilic attack [241]. Contrary to this, a mechanism for the RNAse-catalysed of RNA has been proposed in which the initial proton transfer is concerted with nucleophilic attack [242].

For the purposes of this study, only the nonconcerted mechanism is considered, and it is assumed that proton transfer from O2′ has already occurred. Therefore, the first point on the reaction profile corresponds to the deprotonated form of the nucleoside.

For each of the three reactions examined, the initial structure for optimisation of the reactant was based on the back bone configuration of an RNA A-form duplex with single adenosine bulges from a crystal structure determined by Egli *et al* [243]. In that structure, a modelled 2'-OH of the bulged nucleotide is ideally poised for nucleophilic attack on the adjacent phosphorus, with the 2′-oxygen 3.8 Å away from the phosphorus. The atom numbering scheme used throughout the rest of the chapter is shown in Figure 5.3.

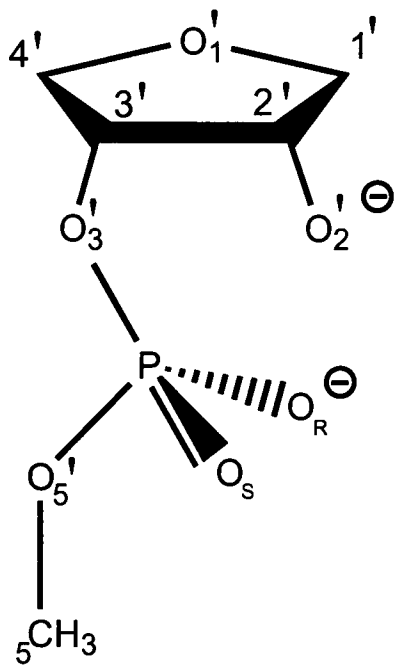


Figure 5.3: The atom number scheme used for description of the three reactions under examination. For the two reactions involving the thio substitutions, the respective sulphur atoms will be denoted according to the atom they are replacing. That is, the sulphur replacing O'_5 is referred to as S'_5, and that replacing O_R as S_R

Computational details

Since the methods used in this study are essentially the same as those described in Chapter 4, only a brief recap is given here with emphasis on any differences in protocol.

All calculations employed the B3LYP density functional [23–25] in conjunction with the 6-31+G** basis set. Each stationary point geometry was fully optimised and vibrational analysis was employed to verify the validity of each optimised structure. For each transition state obtained, an IRC [47] calculation confirmed that optimising in the direction of the eigenvector corresponding to the structure's imaginary frequency led to both reactant- and product-like structures. Frequencies were scaled by an empirical factor of 0.96 when used for the calculation of thermodynamic parameters. Partial atomic charges were derived from the molecular electrostatic potential using the CHELPG [55] scheme.

The calculations estimating the effect of solution of the reaction utilised the PCM continuum solvent representation [83, 84]. The Pauling atomic radii, which were scaled by a factor of 1.2 for all atoms except acidic hydrogens [238], were used to

determine the dimensions of the solute cavity. All calculations were performed using the Gaussian 98 electronic structure program [48] on a variety of workstations and parallel supercomputers.

5.3 Results

5.3.1 Reaction A: the hydrolysis of methyl ribose phosphate

Stationary points

For the first reaction studied, the hydrolysis of the unmodified RNA analogue methyl ribose phosphate, the geometry of each stationary point was fully optimised in the gas phase. The structure of each stationary point is depicted in Figure 5.4 and selected geometrical parameters are given in Tables 5.1 and 5.2.

The stationary point corresponding to the reactant geometry, R_A, has a distance of 4.6 Å between the attacking oxygen, O2′, and the phosphorus. Because of Coulombic repulsion between O2′ and the scissile phosphate group, the geometry of the system is significantly distorted from the original crystal structure geometry. This is reflected in the value of 108° obtained for the angle of attack, O2′–P–O5′.

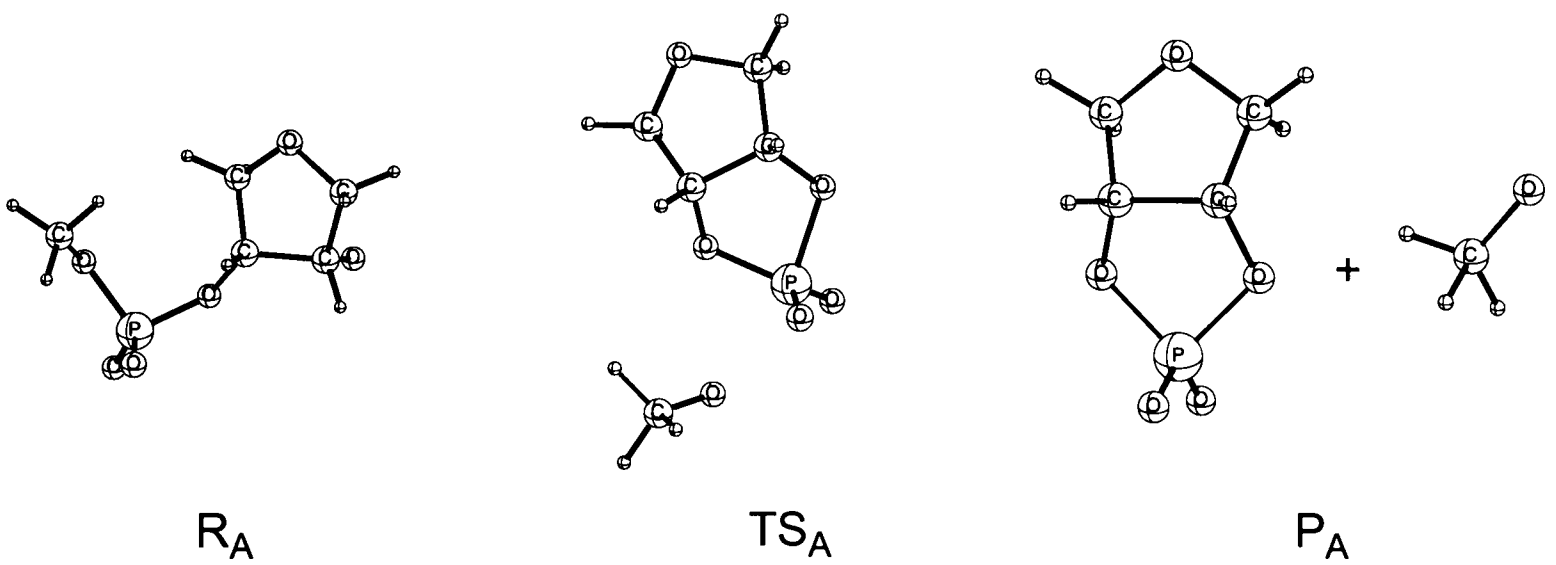


Figure 5.4: Stationary point geometries determined at the B3LYP/6-31+G** level of theory for Reaction A: the base-catalysed hydrolysis of methyl ribose phosphate.

To facilitate the location of the neighbouring transition state on the reaction PES, the reaction coordinate was scanned by decreasing the distance between O2′ and the phosphorus atom in 0.2 Å increments. At each point, the system geometry was minimised with the P–O2′ bond length held fixed. The highest energy structure obtained was optimised with a view to locating a transition state.

This resulted in the location of a saddle point, TS_A, which was confirmed to be a transition state by a frequency calculation. Graphical examination of the normal mode vibrations of the transition state, using the MOLDEN program [42], confirmed that the imaginary frequency corresponds to a vibration synonymous with the formation and cleavage of the bond between the phosphorus and the leaving group.

In TS_A, the distance between the nucleophile and the phosphorus has closed to 1.84 Å, almost a normal P–O bond length. The angle between the approaching nucleophile and the bond angle between the phosphorus and the leaving group, O2′–P–O5′ has opened to 162°, corresponding to an axial attack of the nucleophile on the phosphorus, in agreement with experimental findings for both enzymatic and nonenzymatic reactions [3, 4, 6]. The geometry of the transition state is trigonal bipyramidal with the attacking 2′-oxygen and the leaving 5′-oxygen occupying the apical points. The bond between the phosphorus and the leaving methoxide group has weakened from 1.65 Å to 2.54 Å.

	P-O2′	P-O5′	P-O_S	P-O_R	P-O3′	C-O5′	C2′-O2′	C3‘-O3‘
R_A	4.602	1.707	1.517	1.505	1.650	1.411	1.336	1.428
TS_A	1.843	2.544	1.511	1.499	1.716	1.344	1.381	1.399
P_A	1.706	N/A	1.505	1.496	1.707	1.341	1.410	1.410

Table 5.1: Key interatomic distances for the B3LYP/6-31+G** optimised stationary points of Reaction A. All values are given in Angstroms (Å).

This ‘late’ transition state corresponds to the cleavage of the bond between the phosphorus and the leaving group. An IRC calculation from the transition state was carried out and the reaction energy profile obtained (Figure 5.5) confirmed that structure TS_A is the only transition state located on the reaction PES. Therefore,

	O2′-P-O5′	O2′-P-O_R	O2′-P-O_S	O3′-P-O2′	C-O5′-P	C-O5′ -P-O_R	C-O5′ -P-O2′	C-O5′ -P-O3′
R_A	108.78	108.94	103.68	8.87	119.57	144.08	-99.03	-104.48
TS_A	162.1	100.33	99.54	85.84	114.28	47.34	-74.46	-65.36
P_A	N/A	109.49	108.58	90.95	N/A	N/A	N/A	N/A

Table 5.2: Key bond and dihedral angles for the B3LYP/6-31+G** optimised stationary points of Reaction A. All values are given in degrees.

neither a transition state corresponding to the approach of the nucleophile nor a dianionic intermediate were found to be present in the gas phase. As a precaution, the structures corresponding to the slight shoulder on the IRC were subjected to further energy evaluations using the PCM solvent representation in order to ascertain if an intermediate existed in solution but no evidence was found.

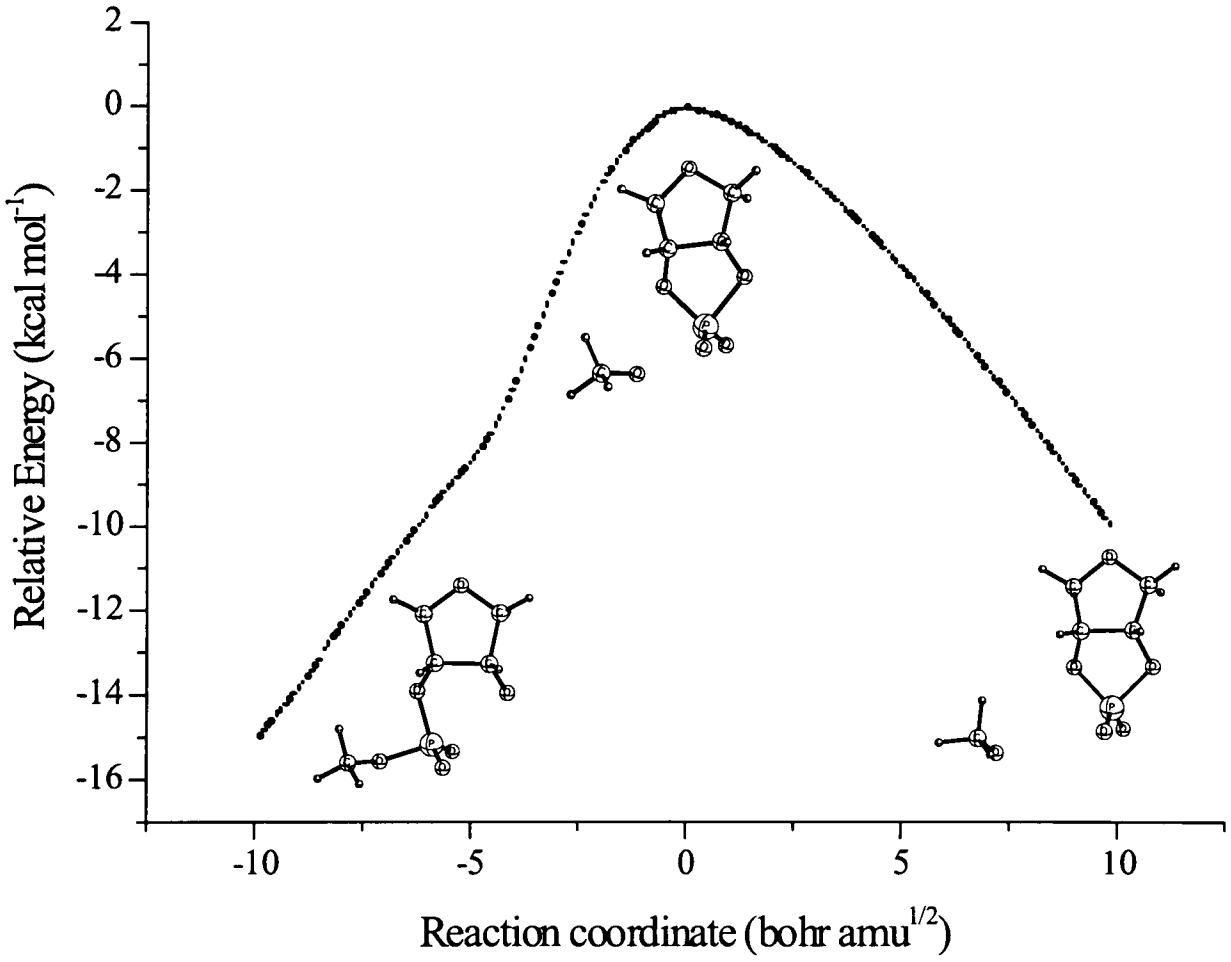


Figure 5.5: The energy profile of Reaction A determined by IRC calculation at the B3LYP/6-31+G** level of theory. The geometries of the two extreme points on the profile and of TS_A are also shown.

The reaction product, P_A, was considered as two structures at infinite separation. The structure of the phosphorylated ribose is very symmetrical with each corresponding pairs of bond lengths having practically the same value.

Energetics

The relative energies of the stationary points in the gas phase and in solution are shown in Table 5.3. The transition state was calculated to be 30.0 kcal mol^{-1} less stable than the reactant. Since this is the only transition state for the reaction, this value corresponds to the free energy barrier for the reaction. The inclusion of solvent effects, through a PCM representation, lowered the barrier by 11.9 kcal mol^{-1} to 18.2kcal mol^{-1} showing that the transition state is more favourably polarised by solvent than the reactants.

	ΔE	ΔZPE	ΔG_{gas}	$\Delta\Delta G_{solv}$	$\Delta G_{solution}$
R_A	0.00	0.00	0.00	0.00	0.00
TS_A	31.40	-1.74	30.03	-11.85	18.18
P_A	-48.22	-1.79	-59.82	49.41	-10.58

Table 5.3: The relative energies of the stationary points calculated on the Reaction A energy profile calculated at the B3LYP/6-31+G** level of theory. ΔZPE gives the relative zero-point energy corrections, and ΔG_{gas} denotes the relative gas phase free energies. $\Delta\Delta G_{solv}$ is the relative free energy of solvation and $\Delta G_{solution}$ is the relative free energy in solution which is calculated by adding ΔG_{gas} and $\Delta\Delta G_{solv}$. All values are given in kcal mol^{-1}.

The overall free energy change for the reaction was computed to be -59.82 kcal mol^{-1} making the products considerably more stable than the reactants. This highly exergonic value reflects the entropic gain on going from a unimolecular to a bimolecular system. Consideration of solvent effects reduces the overall free energy difference to -10.6 kcal mol^{-1}. However, the preferential solvation of a single dianionic molecule over two monoanionic molecules must be borne in mind when considering this value.

Charge distribution

The partial atomic charges, shown in Figure 5.4, exhibit the usual trends encountered for this type of reaction. The partial charge of O2′ decreases in magnitude while that of the leaving oxygen, O5′, increases as the reaction progresses. The presence of solvent is reflected in the polarisation of the stationary point atomic charges. In general, the partial charges on many of the oxygen atoms increases in solution. In particular, the partial charge of O2′ in the reactant species increases from -1.02 e to -1.15 e. Similarly, the partial charge on O5′ increases from -1.14 e to -1.31 e on going from the gas phase to solution.

	Gas Phase			Solution		
	R_A	TS_A	P_A	R_A	TS_A	P_A
C1′	0.36	0.23	0.18	0.27	0.21	0.19
C2′	0.64	0.48	0.37	0.58	0.41	0.34
C3′	0.53	0.34	0.37	0.42	0.26	0.34
C4′	0.28	0.25	0.18	0.29	0.24	0.19
O1′	-0.63	-0.62	-0.55	-0.63	-0.61	-0.58
O2′	-1.02	-0.82	-0.65	-1.15	-0.83	-0.68
O3′	-0.65	-0.54	-0.65	-0.73	-0.56	-0.68
P	1.46	1.60	1.48	1.47	1.55	1.49
O_R	-0.87	-0.84	-0.82	-0.88	-0.85	-0.88
O_S	-0.88	-0.85	-0.84	-0.88	-0.86	-0.86
O5′	-0.62	-0.98	-1.14	-0.66	-1.15	-1.31
C5′	0.40	0.97	1.06	0.29	0.85	0.92

Table 5.4: Partial atomic charges (e) for the Reaction A stationary points. The charges were calculated from the molecular electrostatic potential using the CHELPG scheme at the B3-LYP/6-31+G** level of theory.

Conclusions

The finding that the transition state for this reaction corresponds to cleavage of the bond to the leaving group is in agreement with solution phase experimental studies of the hydrolysis of the nucleic acid phosphodiester linkage. Experimental studies have shown that by varying the pK_a of the leaving group, it has been found that cleavage of the bond to the leaving group is the rate determining step in the alkaline hydrolysis of

an RNA dimer, adenyl(3′-5′)adenosine [244]. There is also evidence for cleavage of the bond to the leaving group being the rate limiting step in ribonuclease and ribozyme catalysis [4, 240].

There has been considerable speculation about the existence of dianionic phosphorane intermediates along the pathway of base catalyzed phosphate ester hydrolysis [3, 4] Previous theoretical studies of the hydrolysis of simpler phosphate esters in the gas phase have not found evidence to support the existence of phosphorane intermediates [168, 187, 194]. The IRC calculation from the transition state did not find any evidence for an intermediate on the transesterification pathway for the phosphorylated ribose in the gas phase.

5.3.2 Reaction B: the hydrolysis of thiomethyl ribose phosphate

The effect of changing the reaction leaving group from a methoxide to a methylthiolate was considered. Like the hydrolysis of methyl ribose phosphate (Reaction A), the reaction was found to proceed via a single pentacoordinated transition state. The stationary point geometries are shown in Figure 5.6 and selected geometric parameters are given in Tables 5.5 and 5.6.

Stationary points

The reactant geometry, R_B, was found to be similar to that of R_A with the distance between the nucleophile and the phosphorus atom at 4.56 Å. The bond distance between phosphorus and the leaving group is 2.24 Å which is a normal P–S bond length [245]. The structure is again distorted from the original crystal structure with the angle of attack, O2′–P–S5′, having a value of 115.13°

Scanning the reaction coordinate from R_B towards the reaction product resulted in the location of a stationary point that was confirmed by frequency calculation to be

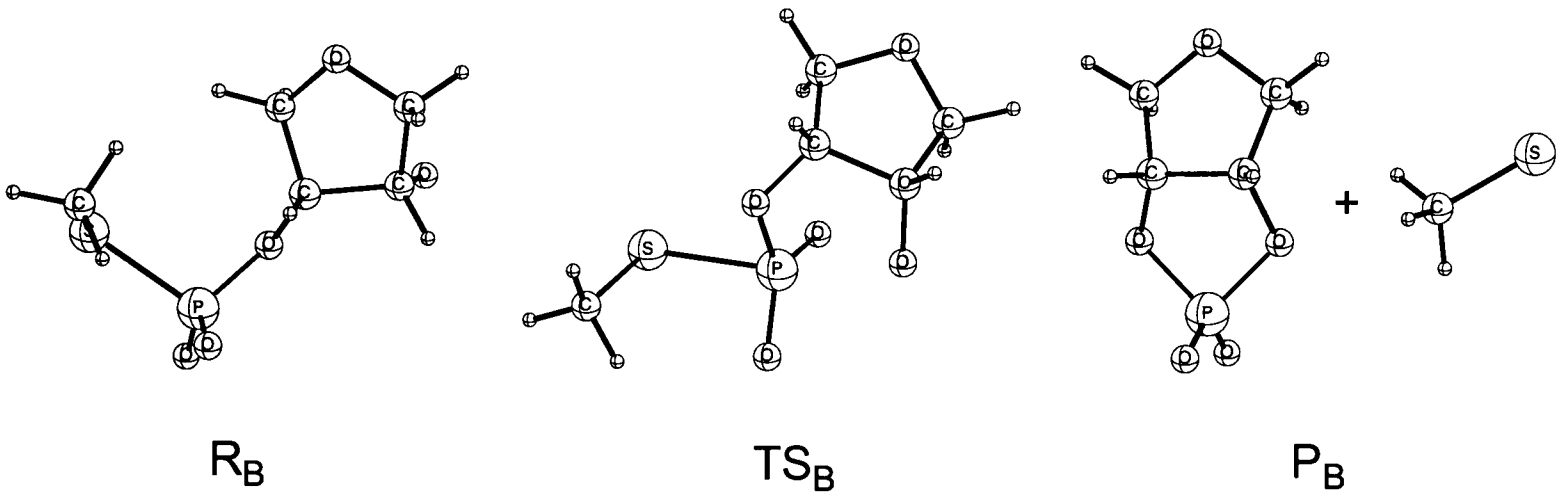


Figure 5.6: Stationary point geometries determined at the B3LYP/6-31+G** level of theory for Reaction B: the base-catalysed hydrolysis of thiomethyl ribose phosphate.

a transition state. The geometry of the transition state, TS_B, is symmetrical about the phosphorus atom with the incoming and leaving oxygens at the apices and almost equidistant from the phosphorus: |P–O2′| = 2.60 Å and |P–O5′| = 2.45 Å. The angle between the axial groups is 165.62° confirming that the nucleophilic attack is 'in-line'.

The geometry of TS_B is considerably different to that of TS_A the transition state determined for Reaction A. TS_A was found to be a 'late' transition state which relates to the cleavage of the bond between phosphorus and the leaving group. In this reaction, the transition state occurs much earlier on the reaction coordinate and corresponds to the formation of the bond between the nucleophile and the phosphorus atom. The bond between phosphorus and the thiloate leaving group has only weakened slightly with a 0.2 Å increase in the bond length P–O′ compared to the reactant.

	P-O2′	P-S5′	P-OS	P-OR	P-O3′	C-S5′	C2′-O2′	C3‘-O3‘
R_B	4.563	2.240	1.515	1.506	1.644	1.839	1.334	1.432
TS_B	2.605	2.446	1.506	1.510	1.680	1.838	1.338	1.416
P_B	1.706	N/A	1.505	1.496	1.707	1.852	1.410	1.410

Table 5.5: Key interatomic distances from the B3LYP/6-31+G** optimised Reaction B stationary points. All values are given in Angstroms (Å).

The difference in reaction pathway for the transesterification of methyl ribose phosphate, Reaction A, and thiomethyl ribose phosphate, Reaction B, is a result of the

	O2′-P-S5′	O2′-P-OR	O2′-P-OS	O3′-P-O2′	C-S5′-P	C-S5′ -P-OR	C-S5′ -P-O2′	C-S5′ -P-O3′
R_B	115.13	102.88	107.22	11.67	101.38	24.55	-87.58	-92.18
TS_B	165.62	90.13	89.27	76.42	99.62	41.15	-77.29	-73.16
P_B	N/A	109.49	108.58	90.95	N/A	N/A	N/A	N/A

Table 5.6: Key bond and dihedral angles for the B3LYP/6-31+G** optimised Reaction B stationary points. All values are given in degrees.

relative strengths of the bonds between the phosphorus atom and the leaving groups, and the relative pK_a s of the leaving groups. An average P–O single bond has a bond energy of approximately 91 kcal mol^{-1} compared to the average P–S single bond strength of approximately 50 kcal mol^{-1} [246]. Also, the pK_a of a 5′-thiol (pK_a = 11) is more than 5 units lower than that of the corresponding 5′-hydroxyl (pK_a = 16) making the 5'-thiol a considerably better leaving group than the 5'-hydroxyl [220,221].

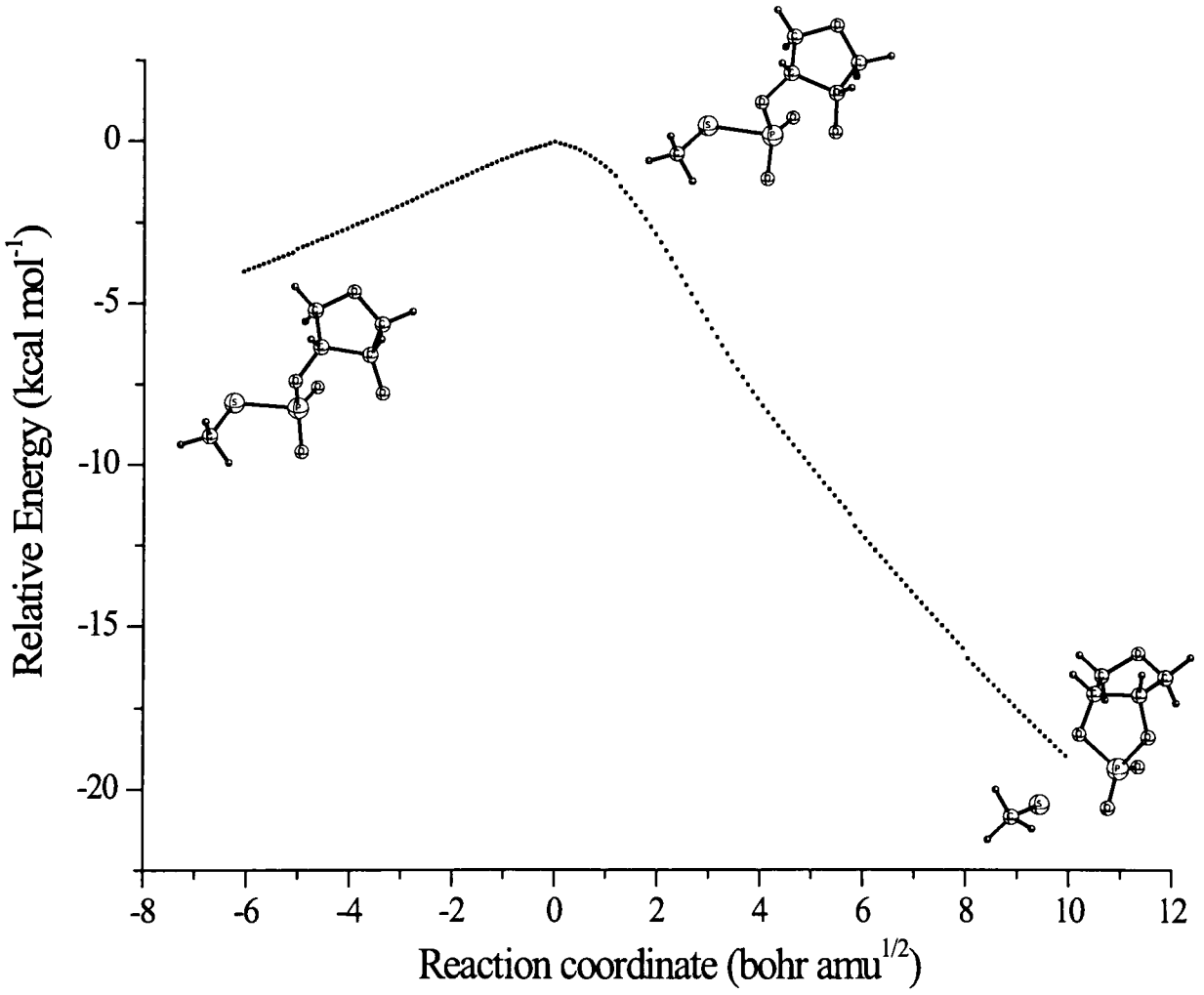


Figure 5.7: The energy profile of Reaction B determined by IRC calculation at the B3LYP/6-31+G** level of theory. The geometries of the two extreme points on the profile and of TS_B are also shown.

Both these factors explain why cleavage of the P-S5′ bond was not found to be rate-determining for this reaction. This is confirmed by an IRC calculation from TS_B. The profile obtained (Figure 5.7) shows that there is no energy barrier along the reaction pathway which corresponds to P-S5′ bond cleavage.

Energetics

The relative energies of the reaction stationary points relative to the reactant are shown in Table 5.7. The gas phase free energy of activation was calculated to be 17.2 kcal mol^{-1}. This value is 12.8 kcal mol^{-1} lower than the corresponding value for Reaction A. The overall free energy change for the reaction is -86.4 kcal mol^{-1} making the reaction considerably more exergonic than Reaction A.

	ΔE	ΔZPE	ΔG_{gas}	$\Delta\Delta G_{solv}$	$\Delta G_{solution}$
R_B	0.00	0.00	0.00	0.00	0.00
TS_B	16.99	-0.39	17.15	-2.65	14.50
P_B	-76.66	0.81	-86.43	56.16	-30.27

Table 5.7: The relative energies of the stationary points calculated on the Reaction B energy profile calculated at the B3LYP/6-31+G** level of theory. ΔZPE gives the relative zero-point energy corrections, and ΔG_{gas} denotes the relative gas phase free energies. $\Delta\Delta G_{solv}$ is the relative free energy of solvation and $\Delta G_{solution}$ is the relative free energy in solution which is calculated by adding ΔG_{gas} and $\Delta\Delta G_{solv}$. All values are given in kcal mol^{-1}.

The inclusion of solvent effects on the gas phase geometries had the effect of lowering the reaction energy barrier by 2.7 kcal mol^{-1} to 14.5 kcal mol^{-1}. This contrasts with the effect of solvation on TS_A, the transition state for the reaction of the unmodified ribose phosphate, Reaction A. In that reaction, the incorporation of solvent effects lowers the energy barrier by 9.5 kcal mol^{-1} which is a much more pronounced stabilisation reflecting the differing electronic structures of the leaving groups and the inability of sulphur to accommodate more than a single full negative charge.

Charge distribution

In general, the charge distributions for the stationary points in both the gas phase and in solution are similar to those found in Reaction A. In the transition state TS_B, the charge on the sulphur atom is less negative than of O5′, the equivalent atom in Reaction A. This is especially noticeable in solution where the S5′ partial charge is -0.74 *e* compared to -1.15 *e* for O5′ in Reaction A. Solvation has very little effect on the magnitude of the S5′ partial charge compared to O5′ in R_A.

The partial charge of O2′ in TS_B is more negative than the equivalent charge in Reaction A reflecting the earlier transition state encountered for this reaction.

	Gas Phase			Solution		
	R_B	TS_B	P_B	R_B	TS_B	P_B
C1′	0.36	0.14	0.18	0.27	0.13	0.19
C2′	0.69	0.69	0.37	0.59	0.63	0.34
C3′	0.51	0.33	0.37	0.40	0.31	0.34
C4′	0.25	0.04	0.18	0.24	0.00	0.19
O1′	-0.61	-0.55	-0.55	-0.61	-0.55	-0.58
O2′	-1.02	-0.96	-0.65	-1.14	-1.09	-0.68
O3′	-0.61	-0.52	-0.65	-0.67	-0.55	-0.68
P	1.36	1.47	1.48	1.28	1.37	1.49
O^R	-0.82	-0.79	-0.82	-0.83	-0.81	-0.88
O^S	-0.83	-0.79	-0.84	-0.83	-0.82	-0.86
S5′	-0.56	-0.79	-1.12	-0.47	-0.74	-1.23
C5′	0.14	0.31	0.70	0.09	0.23	0.56

Table 5.8: Partial atomic charges (*e*) for the Reaction B stationary points. The charges were calculated from the molecular electrostatic potential using the CHELPG scheme at the B3-LYP/6-31+G** level of theory.

Conclusions

The results of this quantum chemical study are in good agreement with similar experimental studies [214,216,217,240,247]. These studies show that the hydrolysis of thio analogues of RNA is 10^5 times faster than the hydrolysis of unmodified substrates.

The character of the transition state for this reaction is different to that of the Reaction A transition state. This is further confirmation that the P-O5′ bond cleavage

is the rate-limiting step for the hydrolysis of RNA. If this was not the case, it would be expected that the introduction of a thiomethyl moiety, a better leaving group [220,221], would have had no effect on the overall reaction free energy barrier [240].

5.3.3 Reaction C: the hydrolysis of methyl ribose phosphothiolate

The final reaction examined in this study is Reaction C, the base-catalysed hydrolysis of methyl ribose *pro*-R phosphothiolate. This analogue of methyl ribose phosphate has the *pro*-R oxygen substituted by a sulphur. The reaction profile was determined to have two transition states and a high energy intermediate in the gas phase. The geometries of the stationary points are shown in Figure 5.8 and selected structural parameters given in Tables 5.9 and 5.10.

Stationary points

The minimum corresponding to the reactant, R_C, is similar to the unmodified reactant encountered in Reaction A. The attacking oxygen, O2′, is 4.56 Å away from the phosphorus. The presence of the *pro*-R sulphur has little effect on the geometry of the atoms bonded to the phosphorus. The P–O3′ bond length is 1.42 Å and the P–S_R bond length is 2.01 Å; shorter than a typical P–S single bond [245]. The bond between the phosphorus and the leaving oxygen is longer than it is in R_A, the equivalent structure in Reaction A. Thus it can be concluded that the presence of the *pro*-R sulphur destabilises this bond.

A scan of the reaction coordinate led to the location of a stationary point which was confirmed to be a transition state by vibrational analysis. The geometry of this transition state, $TS1_C$, is reactant-like with the bond between the nucleophile and the phosphorus partially formed, |P–O2′| = 2.22 Å, and distance between the leaving group and the phosphorus only 0.07 Å longer than in the reactant. An IRC calculation from

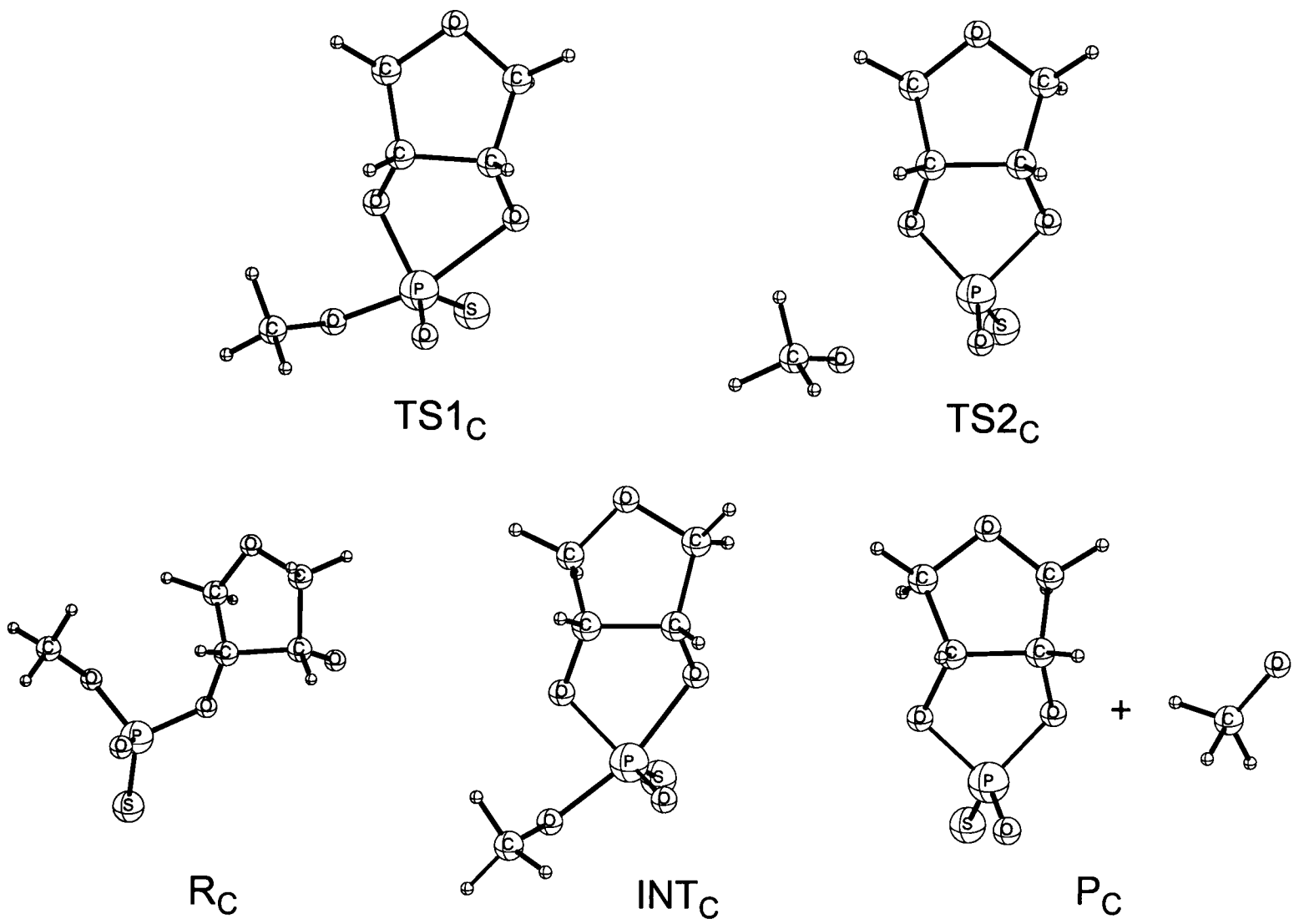


Figure 5.8: Stationary point geometries determined at the B3LYP/6-31+G** level of theory for Reaction C: the base-catalysed hydrolysis of methyl ribose R_P-phosphothiolate.

$TS1_C$ confirmed that $TS1_C$ is on the correct reaction pathway. The energy profile by this calculation is shown in Figure 5.9.

By continuing the scan of the reaction coordinate, an intermediate was located nearby $TS1_C$ on the PES. The structure of this intermediate, INT_C, is very similar to that of $TS1_C$, except that the bond between phosphorus and O2′ has shortened to 1.98 Å. The detection of this dianionic intermediate was somewhat surprising since previous *ab initio* theoretical studies of base-catalysed phosphate ester hydrolysis reactions in the gas phase, including the one presented in Chapter 4, have discounted the possibility of dianionic intermediates in the gas phase [4, 168]. Since a similar

	P-O2′	P-O5′	P-O^S	P-S^R	P-O3′	C-O5′	C2′-O2′	C3‘-O3‘
R_C	4.598	1.698	1.514	2.009	1.641	1.415	1.333	1.438
$TS1_C$	2.215	1.769	1.524	2.056	1.734	1.403	1.344	1.401
INT_C	1.980	1.792	1.534	2.079	1.777	1.399	1.368	1.391
$TS2_C$	1.799	2.725	1.507	2.004	1.706	1.341	1.392	1.402
P_C	1.703	N/A	1.507	1.979	1.703	1.341	1.413	1.413

Table 5.9: Key interatomic distances for the B3LYP/6-31+G** optimised Reaction C stationary points. All values are given in Angstroms (Å).

	O2′-P-O5′	O2′-P-S_R	O2′-P-O_S	O3′-P-O2′	C-O5′-P	C-O5′ -P-S_R	C-O5′ -P-O2′	C-O5′ -P-O3′
R_C	105.66	110.32	105.99	7.00	119.06	131.30	-111.26	-116.57
$TS1_C$	163.89	87.58	90.59	78.13	116.69	171.20	-93.76	-69.38
INT_C	163.72	88.97	93.39	81.55	117.36	168.60	-98.76	-70.67
$TS2_C$	160.51	100.61	102.31	87.30	111.19	175.39	-69.73	-66.94
R_C	N/A	109.83	109.33	91.23	N/A	N/A	N/A	N/A

Table 5.10: Key bond and dihedral angles for the B3LYP/6-31+G** optimised Reaction C stationary points. All values are given in degrees.

intermediate was not located in Reaction A, the presence of INT_C can be attributed to the thio substitution.

A second transition state, $TS2_C$, was found further along the reaction pathway. This transition state corresponds to cleavage of the bond between the phosphorus and the leaving group. The structure of this transition state is very similar to that of TS_A, the rate-limiting transition state for Reaction A. The P–O2′ bond is almost formed at 1.80 Å and the leaving group has almost fully dissociated, |P–O5′| = 2.73 Å. The ‘angle of attack’ is almost linear with the O2′-P-O5′ bond angle having a value of 160.51°. An IRC calculation, shown in Figure 5.10, confirmed that, by following the eigenvector corresponding to the transition state imaginary frequency, structures very similar to INT_C and the reaction products are obtained.

The reaction concludes with the decomposition of $TS2_C$ to give the reaction products, P_C. The geometry of the thio-substituted cyclic phosphate is very similar to the unsubstituted equivalent encountered in Reactions A and B.

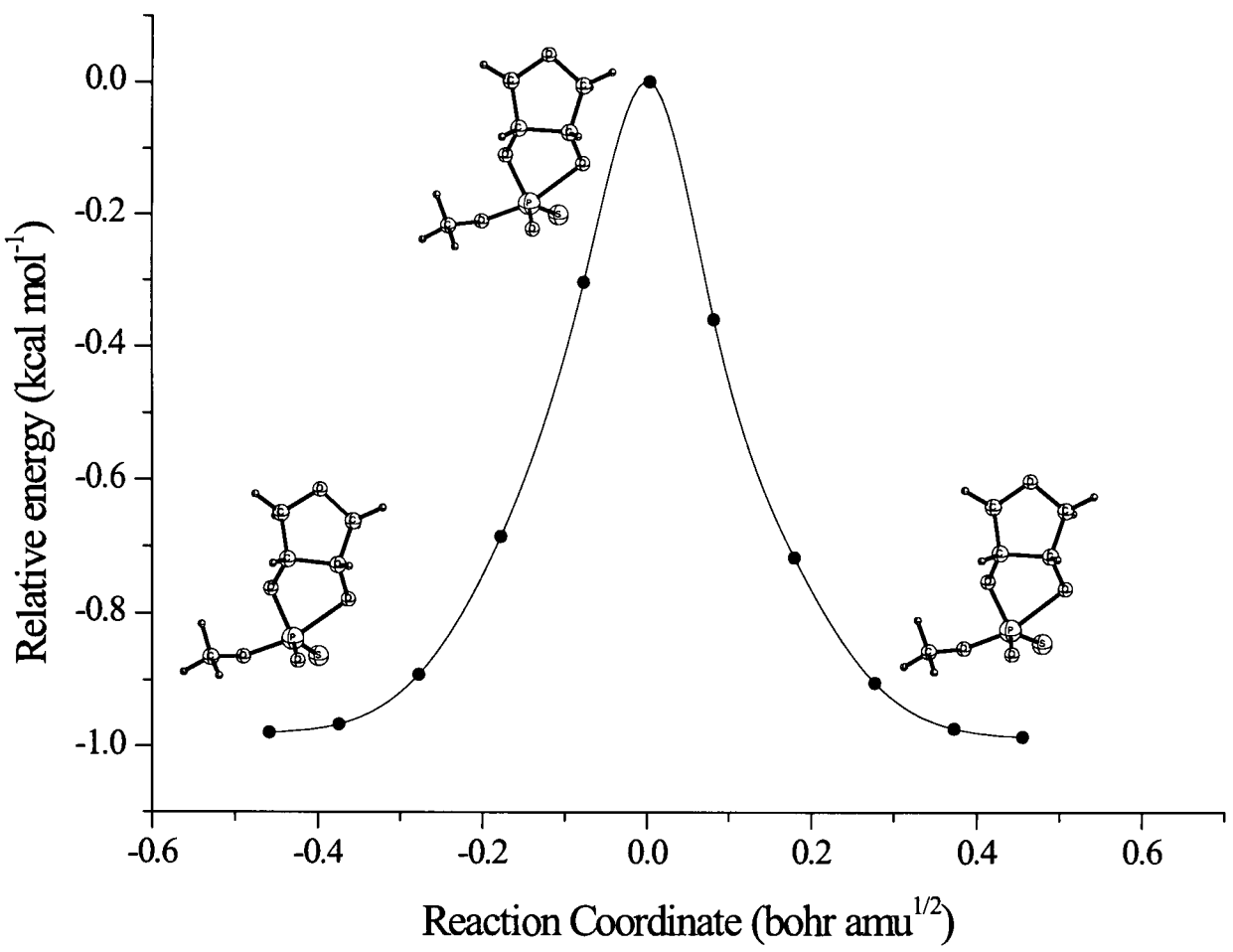


Figure 5.9: The energy profile from the transition state $TS1_C$ determined by IRC calculation at the B3LYP/6-31+G** level of theory. The geometries of the two extreme points on the profile and of $TS1_C$ are also shown.

Energetics

The relative energies of each stationary point relative to the reactant are shown in Table 5.11. The reaction has two transition states, the second of which is rate-limiting. In the gas phase, the first transition state, $TS1_C$ has 20.8 kcal mol^{-1} higher energy than the reactant. The intermediate, INT_C, is only 1.6 kcal mol^{-1} more stable than $TS2_C$. Since the intermediate lies in such a shallow energy well, it is doubtful that it is kinetically significant.

The second transition state, $TS2_C$, is 29.2 kcal mol^{-1} less stable than the reactant. Thus, this transition state constitutes the overall energy barrier for Reaction C. This is just 0.8 kcal mol^{-1} lower than the overall gas phase barrier for the hydrolysis of the unsubstituted analogue thus demonstrating that this thio substitution has very little effect on the reaction energetics in the gas phase. The overall free energy barrier for

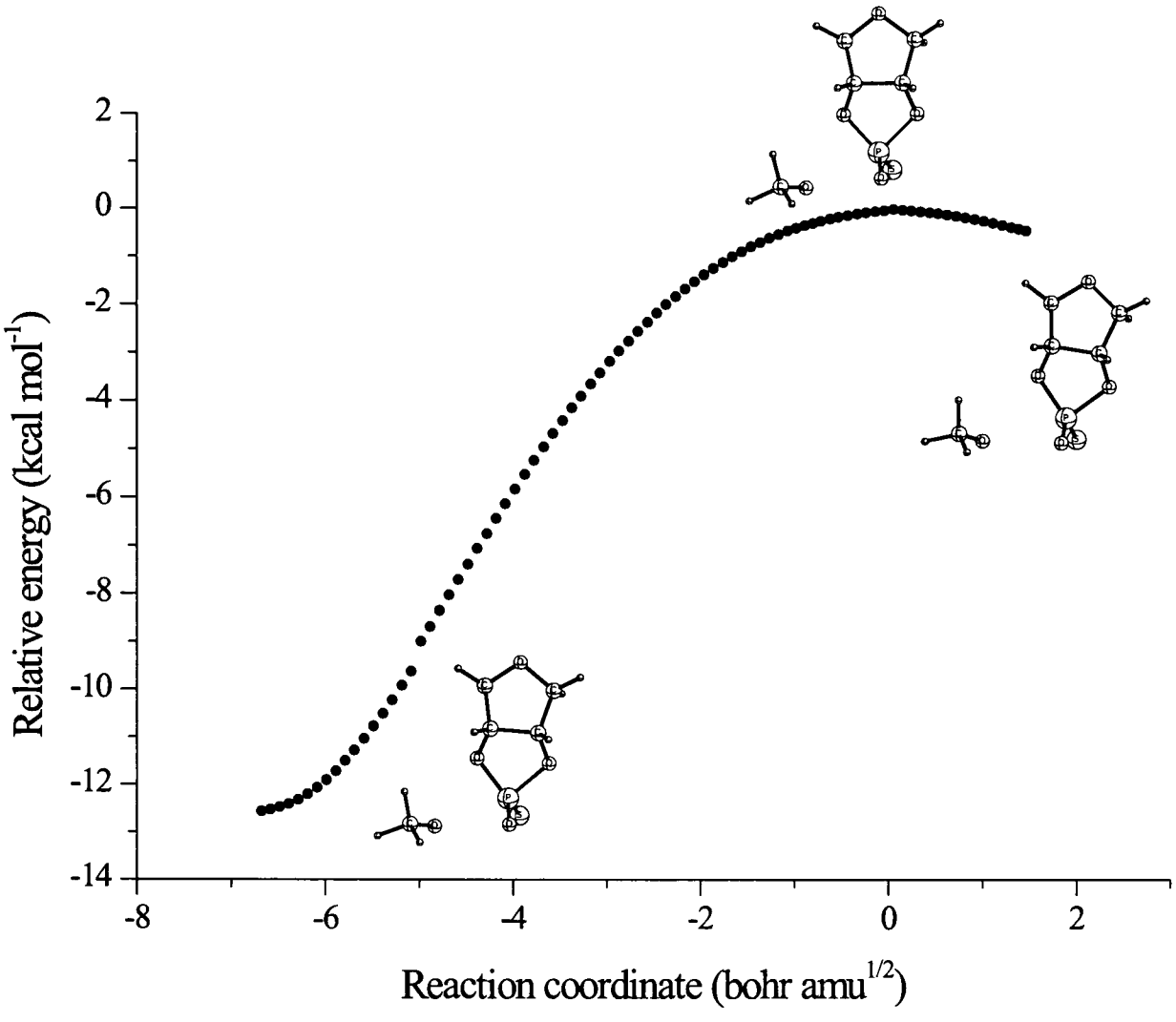


Figure 5.10: The energy profile from the transition state $TS2_C$ determined by IRC calculation at the B3LYP/6-31+G** level of theory. The geometries of the two extreme points on the profile and of $TS2_C$ are also depicted.

the reaction was calculated to be -57.7 kcal mol^{-1} thereby making the reaction highly exergonic.

The inclusion of a PCM solvent representation made significant changes in the reaction energy profile. The free energy difference between the reactant and $TS1_C$ is lowered by 9.6 kcal mol^{-1} to 11.2 kcal mol^{-1}. In solution, INT_C was found to have 9.5 kcal mol^{-1} higher energy than the reactant. The inclusion of solvent effects has very little effect on the stability of the intermediate. The energy difference between $TS1_C$ and INT_C only increases by 0.1 kcal mol^{-1} in solution to 1.7 kcal mol^{-1}.

The incorporation of solvent effects also affects the relative energies of the reactant and $TS2_C$ and decreases the overall free energy barrier for the reaction by 8.6kcal mol^{-1} to 20.6 kcal mol^{-1}. This is 1.8 kcal mol^{-1} higher than the overall barrier for

	ΔE	ΔZPE	ΔG_{gas}	$\Delta\Delta G_{solv}$	$\Delta G_{solution}$
R_C	0.00	0.00	0.00	0.00	0.00
$TS1_C$	19.81	-0.16	20.76	-9.57	11.19
INT_C	18.32	0.53	19.18	-9.65	9.53
$TS2_C$	31.26	-1.51	29.21	-8.64	20.57
P_C	-45.67	-1.59	-57.65	47.09	-10.56

Table 5.11: The relative energies of the stationary points calculated on the Reaction A energy profile calculated at the B3LYP/6-31+G** level of theory. ΔZPE gives the relative zero-point energy corrections, and ΔG_{gas} denotes the relative gas phase free energies. $\Delta\Delta G_{solv}$ is the relative free energy of solvation and $\Delta G_{solution}$ is the relative free energy in solution which is calculated by adding ΔG_{gas} and $\Delta\Delta G_{solv}$. All values are given in kcal mol^{-1}.

Reaction A in solution. This is in agreement with experimental findings which have shown that this substitution has only small effects on the transesterification step in alkaline solution [228–230]

Charge distribution

The charge distributions for the stationary points of Reaction C in the gas phase and in solution are shown in Table 5.12. Overall, the charge distribution is less polarised than that found in the reaction of the unmodified ribose phosphate (Reaction A). In the gas phase, there is a slight charge imbalance between the two non-bridging atoms of the the phosphorothiolate, O_S and S_R. This, however, is balanced out when the polarising effects of solution are taken into account.

Conclusions

Experimentally it has been found that substituting sulphur for either the *pro*-R or *pro*-S non-bridging oxygen has only a slight effect on the transesterification step in alkaline solution, with reported thio effects (k_{P-O}/k_{P-S}) ranging from 0.78 to 2.5 [215, 228–230]. The overall reaction barrier including solvation effects calculated here for the R_pS phosphate ester is 1.8 kcal mol^{-1} higher than the activation free energy for the transesterification of the unsubstituted system.

	Gas Phase					Solution				
	R_C	$TS1_C$	INT_C	$TS2_C$	P_C	R_C	$TS1_C$	INT_C	$TS2_C$	P_C
C1′	0.41	0.13	0.27	0.22	0.10	0.29	0.09	0.16	0.18	0.08
C2′	0.61	0.60	0.44	0.56	0.41	0.54	0.58	0.50	0.49	0.40
C3′	0.57	0.42	0.49	0.27	0.41	0.64	0.33	0.37	0.24	0.40
C4′	0.28	0.14	0.16	0.24	0.14	0.21	0.12	0.10	0.20	0.12
O1′	-0.65	-0.58	-0.58	-0.61	-0.52	-0.64	-0.57	-0.55	-0.59	-0.54
O2′	-1.02	-0.85	-0.76	-0.81	-0.60	-1.14	-1.00	-0.92	-0.80	-0.62
O3′	-0.58	-0.56	-0.60	-0.46	-0.60	-0.71	-0.58	-0.61	-0.48	-0.62
P	1.13	1.37	1.41	1.27	1.06	1.03	1.24	1.29	1.11	1.00
O_S	-0.79	-0.81	-0.84	-0.74	-0.73	-0.79	-0.83	-0.86	-0.76	-0.76
S_R	-0.79	-0.89	-0.94	-0.76	-0.66	-0.69	-0.82	-0.88	-0.68	-0.65
O5′	-0.58	-0.74	-0.72	-0.97	-1.14	-0.59	-0.79	-0.81	-1.19	-1.31
C5′	0.43	0.66	0.68	0.98	1.06	0.36	0.57	0.59	0.95	0.92

Table 5.12: Partial atomic charges (e) for the Reaction C stationary points. The charges were calculated from the molecular electrostatic potential using the CHELPG scheme at the B3-LYP/6-31+G** level of theory.

It has been shown experimentally that the reaction proceeds through a pentacoordinate transition state, with inversion at phosphorus and without rearrangement of the sulfur from an equatorial to an apical position [248]. Under acidic conditions it is believed that a stable phosphorane intermediate is formed since pseudorotation is observed, however no pseudorotation products have been observed under alkaline conditions [248]. The calculated reaction pathway presented here is in agreement with current experimental data. Although an intermediate has been shown to be stable along the gas-phase reaction pathway, its stability with respect to the first transition state along the pathway is negligible (1.6 kcal mol^{-1}).

Solvation would be expected to have an effect on the stability of this intermediate, and the kinetic relevance of a phosphorane intermediate for this reaction can only be addressed properly by performing a geometry optimization of the molecule in solvent.

5.4 Summary

The results presented in this chapter suggest that unmodified RNA is hydrolysed via single transition state which corresponds to the cleavage of the P–O5′ bond. Also,

there is no evidence of the presence of a phosphorane intermediate on the reaction potential energy surface. The presence of solvent is thought to substantially reduce the activation free energy for the reaction.

The hydrolysis of 5′-thiomethyl phosphate has a considerably different reaction pathway which proceeds via the formation of an early transition state corresponding to the formation of the bond between the 2′-oxygen and the phosphorus atom.

The final reaction considered was the hydrolysis of an RNA analogue containing a sulphur atom in place of the *pro*-R non-bridging oxygen. This reaction was observed to have a more complicated reaction pathway which has two transition states and a metastable dianionic intermediate.

Chapter 6

A QM/MM study of phosphate ester hydrolysis in solution

6.1 Introduction

Of all the results presented in the preceding two chapters, the optimisation of the stationary points for the base-catalysed hydrolysis of DMP in solution was among the most noteworthy. Whereas the other condensed phase calculations in these chapters considered the effect of solvation as an energetic perturbation of the gas phase reaction profile, this study also demonstrated how the pathway of a chemical reaction can be altered in solution.

The result was attained by considering the solvent as a dielectric continuum which modifies the solute Hamiltonian [238]. Although this approach gives accurate results, a more ideal approach would consider the specific interatomic interactions between the solute and the solvent. This would involve an atomistic representation of the solvent molecules. While the most obvious way to implement such an approach would be to model the entire system using an empirical force field [63, 143], this would result in unsatisfactory representation of the solute. However, the accuracy of quantum mechanical Hamiltonians and the computational efficiency of force field methods can

be united through the use of hybrid QM/MM potentials which are extensively reviewed in Chapter 2.

This chapter details the results of the use of a QM/MM potential to determine the energy profile for the base-catalysed hydrolysis of DMP in solution. The reaction energy profile was determined by performing a series of minimisations along the reaction coordinate. A variety of quantum mechanical Hamiltonians were employed to describe the quantum region and the results achieved with each potential are compared.

6.2 Methods

6.2.1 Construction and Equilibration of the System

The system was partitioned into quantum (QM) and classical (MM) regions; the substrate occupying the QM region and the MM region comprising the solvent. The solvent was represented atomistically using Jorgensen's three site interaction potential (TIP3P) for liquid water [78]. The need to use link atoms or analogous methods was obviated since no covalent bonds spanned the quantum-classical boundary.

The transition state determined at the B3LYP(PCM)/6-31+G* level of theory for DMP + OH^- reaction was chosen as an initial geometry for the QM region. Considerable care was exercised when 'solvating' the substrate so as to ensure that the solvent would be completely equilibrated and that there were no large 'holes' in the system or bad contacts which led to energy 'hot spots'. The protocol followed is described below.

The QM region was solvated by overlaying it with a pre-equilibrated sphere of radius 16 Å containing TIP3P water molecules. The substrate phosphorus atom occupied the centre of the sphere. Any solvent molecules overlapping the substrate were eliminated by removing those with an oxygen atom within 2.8 Å of a substrate heavy atom and those possessing an oxygen more than 16 Å from the centre of the system were also

removed. The geometries of remaining solvent molecules were then briefly minimised using 200 steps of steepest descent minimisation. Throughout this minimisation and the ensuing equilibration steps, the atoms in the quantum region was held fixed in space.

The solvent molecules were equilibrated using a series of molecular dynamics (MD) simulations. For the purposes of the equilibration procedure the QM region was treated classically using the CHARMM all-hydrogen parameter set for nucleic acids (version c27) [249]. Atomic charges were assigned to the QM region atoms using the results of a CHELPG [55] calculation at the B3LYP(PCM)/6-31+G* level of theory.

In order to prevent the outer molecules of the water sphere from 'boiling off' the edge of the system and to simulate the presence of solvent molecules beyond the system boundary, stochastic boundary conditions [38] were employed. This entailed dividing the system into two regions: a reaction region, which comprised all molecules within 13 Å of the centre of the system, and was simulated using standard Newtonian dynamics; and a buffer region which consisted of the outer 3 Å shell of the system.

The motion of the buffer region atoms was propagated using Langevin dynamics [250]. The atoms in this region maintain the temperature of the system through interaction with an external heat bath which is simulated by both random and frictional forces. The atoms in the buffer region are also subjected to a boundary potential which increases in magnitude as the edge of the system is approached. This potential maintains the shape of the system.

With the stochastic boundary conditions implemented and the phosphate substrate still fixed in space, the water molecules were equilibrated for 30 ps at 300 K followed by a further 50 ps of MD during which the temperature of the system was gradually quenched to 50 K.

The solvent overlay, deletion, minimisation, and MD steps were repeated a further three times, each time with the water sphere rotated 90° about the x, y, and z axis,

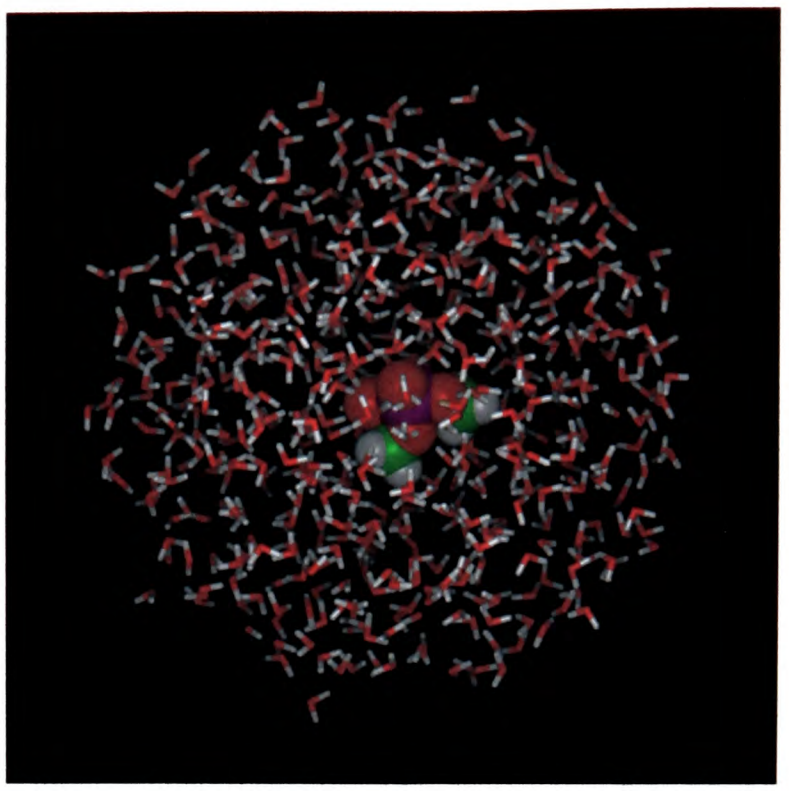


Figure 6.1: A graphical representation of the equilibrated system. The QM region, depicted in CPK form and consisting of DMP and the hydroxide anion, is at the centre of the system. The surrounding 543 water molecules are also represented.

respectively. This was done to eliminate any 'water-sized' holes that may have existed within the system. These further iterations resulted in the addition of just one solvent molecule to the system. Finally, the water molecules were subjected to 200 steps of ABNR minimisation with the solute still fixed. At the end of the construction and equilibration steps, the system, depicted in Figure 6.1, consisted of the substrate and 543 solvent molecules.

6.2.2 Determination of the reaction energy profile

The potential energy surface for the reaction between DMP and OH^- was determined by conducting geometry optimisations at a series of points along the reaction coordinate. This method is known as adiabatic mapping and is described in Chapter 1 (Section 1.9.2). The use of this technique assumes knowledge of the mechanism of the reaction being studied. In the case of the reaction studied here, it is assumed ,on the basis of the results presented in Chapter 4, that the progress of the reaction is dominated by the change in distance between the phosphorus atom of DMP and the two

oxygens involved in the reaction's bond-making and bond-breaking processes, O1 and O4. Attempts were made to define the reaction coordinate in terms of the difference in magnitude between the two distances but it proved very difficult to obtain a smooth reaction energy profile using this implementation.

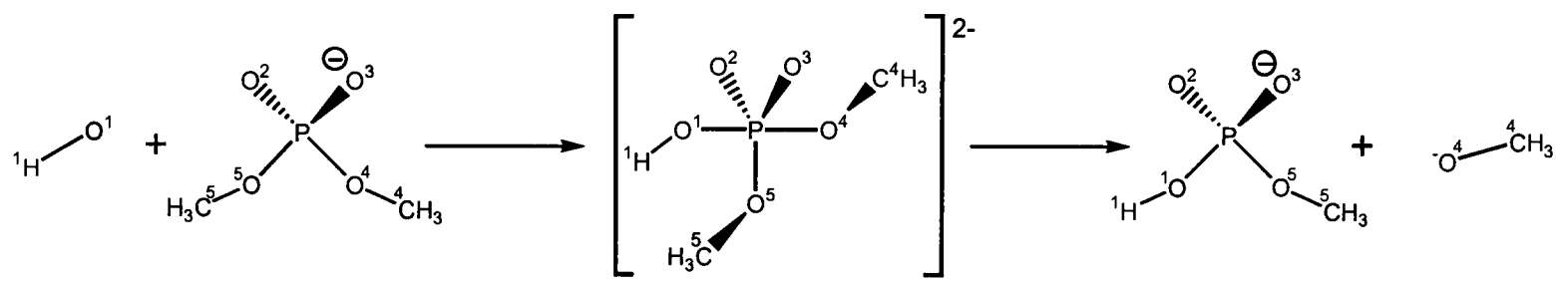


Figure 6.2: The reaction scheme for the base-catalysed hydrolysis of DMP. The atom numbering scheme referred to in the text is shown.

For the first part of the reaction coordinate, the distance between the incoming oxygen, O1, and the phosphorus was used as the reaction coordinate. This was changed to the distance between the phosphorus and the departing oxygen (O4) when the bond between P and O1 was almost formed.

The PES was mapped by placing a harmonic restraint on the bond corresponding to the reaction coordinate. The force constant used was 5000 kcal mol^{-1} $Å^{-2}$. The value of the reaction coordinate was changed in small increments which decreased in the regions of the PES containing stationary points. At each point the system was minimised using the ABNR method until an energy gradient of less than 0.02 kcal mol^{-1} $Å^{-1}$ was attained. Initially, the value of the internal coordinate corresponding to the reaction coordinate was changed in 0.1 Å increments although, in subsequent calculations, further points on the energy profile were calculated in order to gain a more detailed description of the regions containing the stationary points. Throughout the adiabatic mapping procedure, the outer 5 Å shell of the system was fixed to reduce the degrees of freedom in the system and to allow convergence to occur more rapidly. Because solvent molecules often lagged in adjusting to changes in solute conformation, it was necessary to scan parts of the reaction coordinate backwards and forwards to obtain a smooth reaction energy profile.

6.2.3 Choice of QM potential

The vast majority of QM/MM studies reported to date have employed a semi-empirical potential to model the QM region [87]. The main reason for doing so is to avoid the prohibitive computational expense associated with the use of an *ab initio* or DFT potentials to describe the QM region since the use of these types of potentials instead of semi-empirical potentials typically increases the CPU time required for energy evaluations by at least two orders of magnitude. For QM/MM systems in which the QM region contains atoms from the first and second rows of the periodic table, the results achieved using AM1 or PM3 potentials for the QM region are normally comparable with those achieved using more accurate Hamiltonians. However, in the case of phosphate ester hydrolysis reactions, conventional semi-empirical methods fail to give accurate results. The principal reason for this shortcoming is the neglect of *d*-type orbitals which are heavily involved in the formation of pentavalent phosphorane species [245]. Therefore, the use of either *ab initio*, DFT potentials, or specifically modified semi-empirical potentials to model these systems is normally required.

The number of SCF calculations required to map the entire reaction coordinate precluded the use of more accurate basis sets such as 6-31+G* to model the quantum region. It was therefore decided to use a HF/3-21G* potential to model the quantum region[1]. This potential has been shown to give accurate geometries and energies for the gas phase hydrolysis of DMP [168, 187] and, unlike the STO-3G basis set, has been shown not to overestimate the stability of phosphorane intermediate species. A gas phase study conducted by Dejaegere and Karplus demonstrated that the region of the reaction profile containing the phosphorane species had an almost identical shape when either 3-21+G* and 6-31+G* basis sets were used to study the base-catalysed hydrolysis of DMP [168]. A series of single point energy evaluations which utilised the

[1]The hybrid QM/MM potential employing the HF/3-21G* potential for the quantum region and the TIP3P model for MM solvent molecules is referred to as the HF 3-21G*/MM potential from this point on.

HF/3-21+G* and B3LYP/3-21+G* potentials to treat the QM region were also carried out.

Another option for modelling the quantum region has recently became available. A set of AM1 parameters modified to reproduce high level *ab initio* and DFT geometries, energies, and reaction activation energies for reactions involving nucleophilic attack on phosphate esters are in the process of being developed by Lopez and York [251]. The potential, known as the AM1*d* potential, is based on the AM1 Hamiltonian [31] and includes extra terms to account for *d* orbitals on phosphorus, hydrogen bonding interactions, as well as hydrogen bonding interactions involving oxygen atoms, both of which were poorly represented in the original AM1 implementation. Although still under development, early results from the authors are in good agreement with high level *ab initio* and DFT results. A developmental version of the AM1*d* potential was also used to represent the QM region in a hybrid QM/MM potential[2].

6.2.4 Computational details

All the QM/MM calculations described in this chapter were accomplished using locally modified versions of the molecular mechanics software package CHARMM [39]. Those calculations that required an *ab initio* or DFT potentials to treat the quantum region used the GAMESS-US [155] and GAMESS-UK [252] programs, respectively. Both programs were interfaced with CHARMM. Calculations that employed the modified AM1*d* semi-empirical parameters of Lopez and York [251] used a version of CHARMM that was modified to facilitate their inclusion.

The calculations were performed on a variety of computers. Those involving the use of an *ab initio* potential for the QM region were often performed in parallel using various parallel supercomputers and Beowulf [253] clusters. Other calculations were performed on single workstations.

[2]This hybrid potential which unifies the AM1*d* QM potential and the TIP3P parameters for solvent is referred to as the AM1*d*/MM potential from this point on.

6.3 Results

In this section, the results of a series of studies which aim to accurately determine the effect of solution on the energy profile of the base-catalysed hydrolysis of DMP are presented. Both the HF 3-21G*/MM and AM1d/MM hybrid potentials were employed to determine energy profiles. The comparative computational efficiency afforded by the use of the latter potentials facilitated a more thorough examination of the reaction PES which culminated in the formulation of a three-dimensional representation of the middle region of the reaction coordinate.

6.3.1 HF 3-21G*/MM potential energy surface

The potential energy surface for the base-catalysed hydrolysis of DMP in solution was successfully obtained by performing a series of geometry optimisations along the reaction coordinates using the protocol described in Section 6.2.2 while using the HF 3-21G*/MM hybrid potential. The two-dimensional PES obtained is shown in Figure 6.3.

As mentioned in Section 6.2.2, two different internal degrees of freedom were used to describe the reaction. For this scan, the changeover between using the |P-O1| and |P-O4| bond distances was made at the structure corresponding to the phosphorane intermediate where both distances were approximately equidistant[3].

Stationary point geometries

The reaction PES (Figure 6.3) shows that the reaction proceeds via a stable pentacoordinate phosphorane intermediate species. This is flanked, on either side, by a transition state. The reaction mechanism can be described as being associative rather than dissociative. This is evidenced by the presence of a pentavalent rather than a trivalent intermediate species which would be characteristic of a dissociative

[3]The changeover between reaction coordinates occurred when the interatomic distances were at the following values: |P–O1| = 1.87 Å and |P–O1| = 1.88 Å.

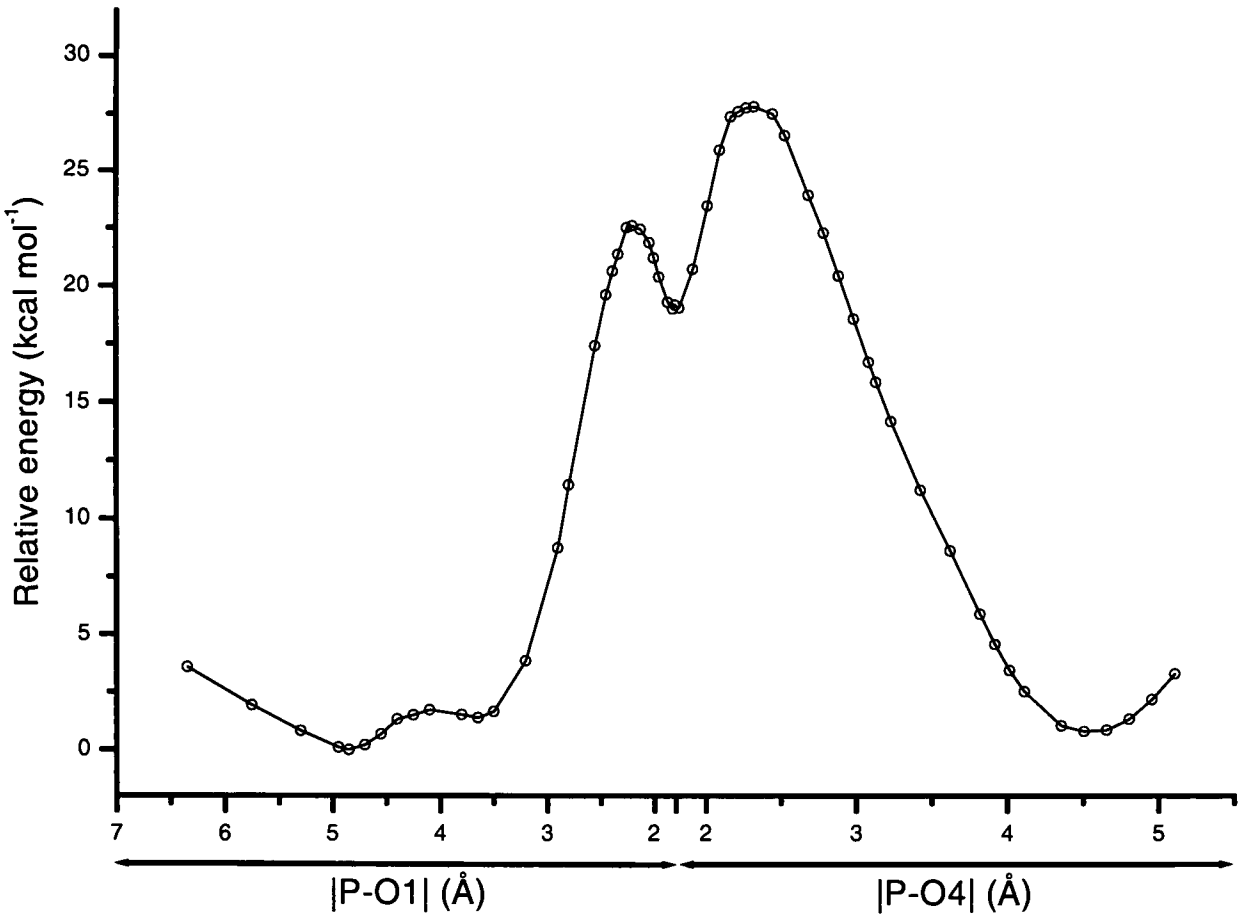


Figure 6.3: The potential energy surface calculated for the reaction DMP + OH^- using a hybrid QM/MM potential which treated the QM region using a HF/3-21G* potential. The x-axis, which describes the reaction coordinate, is divided into two parts reflecting the two different descriptions used in the calculation of the PES. $|P–O_{nuc}|$ is the distance, |P–O1|, between the phosphorus and the incoming hydroxyl oxygen and $|P–O_{lg}|$ is the distance |P–O4| between the phosphorus and the leaving methoxy oxygen.

mechanism. Graphical representations of the QM region stationary point geometries are shown in Figure 6.4 and key geometrical parameters for each are given in Tables 6.1, 6.2, and 6.3.

Instead of considering the reactants and products of this bimolecular reaction as infinitely separated species, it proved possible to locate molecular encounter complexes on the reaction PES which were found to be minima and to correspond to the reactants and product species for the reaction being studied.

The structure of the reactant complex showed that the distance between the incoming hydroxy oxygen (O1) and the phosphorus is 4.85 Å and that the the bond

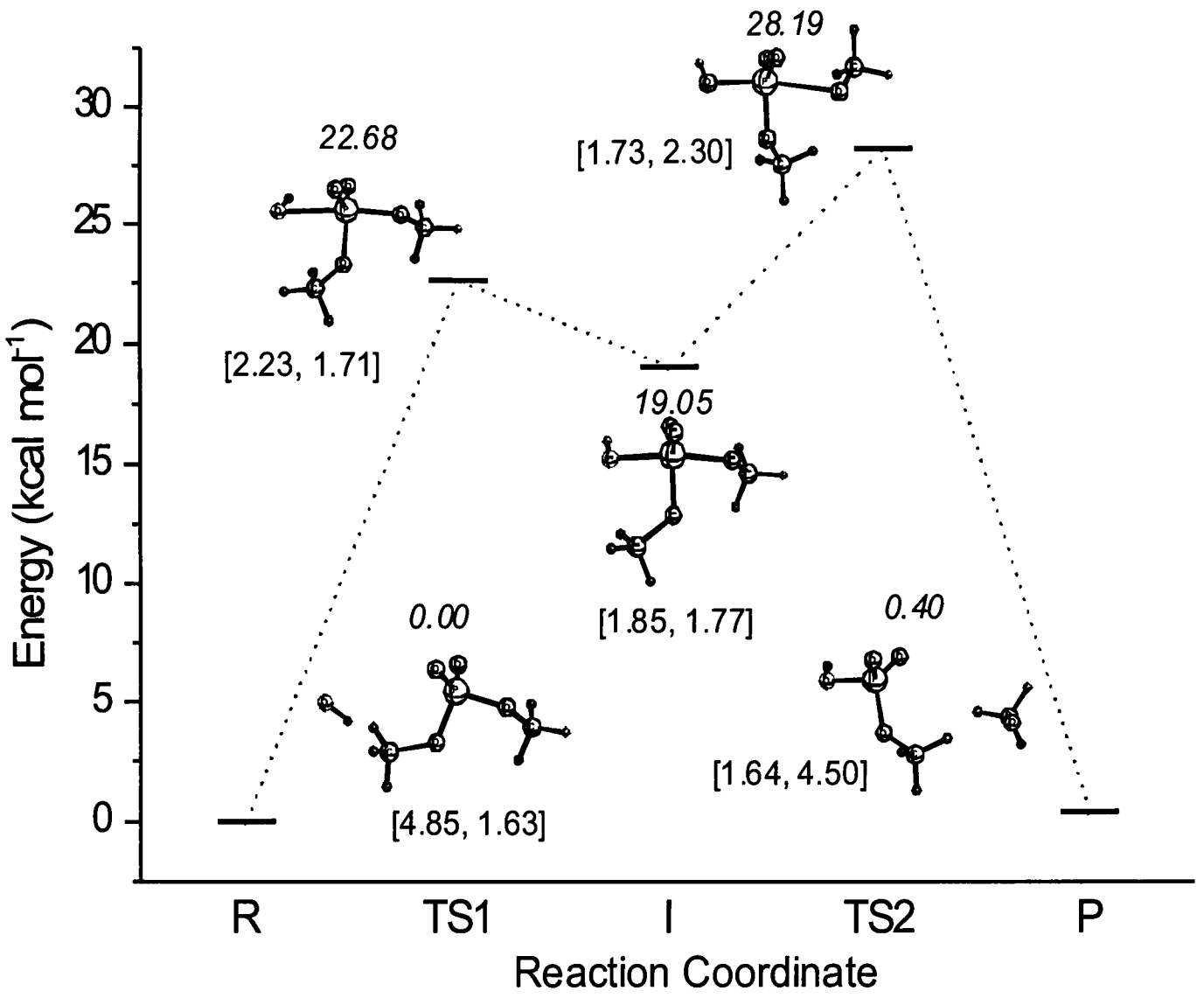


Figure 6.4: Stationary points along the reaction coordinate for DMP + OH^- calculated by adiabatic mapping using a HF/3-21G*/MM potential. The numbers in italics beside each stationary point give the energy relative to the reactant energy. The numbers in square brackets indicate the |P–O1| and |P–O4| distances respectively.

between the phosphorus and the axial methoxy oxygen (O4) was fully formed with a bond length of 1.63 Å. The angle between the incoming oxygen, the phosphorus and the departing oxygen (O1–P–O4) is 155.1°, demonstrating that the nucleophile is poised for axial attack.

As the reaction progresses and the hydroxide anion approaches the phosphorus, the energy of the system rises rapidly as the Coulombic repulsion between the hydroxyl anion and the monoanionic phosphate grows. This culminates in an energy barrier which corresponds to the first reaction transition state, TS1.

	P-O1	P-O2	P-O3	P-O4	P-O5	O1-H1	O4-C4	O5-C5
Reacts.	4.850	1.479	1.482	1.625	1.601	0.975	1.446	1.466
TS1	2.225	1.499	1.502	1.709	1.621	0.969	1.428	1.436
INT	1.849	1.517	1.513	1.769	1.654	0.968	1.419	1.429
TS2	1.729	1.493	1.491	2.300	1.613	0.966	1.404	1.443
Prods.	1.642	1.483	1.483	4.500	1.594	0.969	1.393	1.478

Table 6.1: Key bond lengths (Å) for the approximate stationary points in the DMP + OH^- reaction from the HF 3-21G*/TIP3P reaction energy profile.

	O1-P-O4	O2-P-O3	O2-P-O5	O3-P-O5	O1-P-O5	O5-P-O4	O1-P-O2	O3-P-O4
Reacts.	155.11	120.52	111.31	110.09	61.89	97.94	70.52	106.85
TS1	172.46	127.00	115.41	115.48	83.33	89.39	81.15	96.49
INT	171.03	125.78	118.30	115.90	87.52	84.02	87.65	93.16
TST2	170.11	126.41	118.32	112.57	90.80	81.02	94.73	87.99
Prods.	160.84	120.32	112.66	109.31	98.91	63.50	108.61	85.62

Table 6.2: Key bond angles for the approximate stationary points of the DMP + OH^- reaction from the HF 3-21G*/TIP3P reaction energy profile. All values are given in degrees.

TS1 has a reactant-like, trigonal bipyramidal geometry and represents the barrier to the formation of the P–O1 bond. The bond between the attacking hydroxyl oxygen and the phosphorus is partially formed, |P-O1| = 2.23 Å, and the bond between phosphorus and the leaving group has lengthened by 0.08 Å to 1.71 Å. The equatorial methoxy group is oriented towards the incoming anion. It is thought that the favourable electrostatic interaction between the positively charged CH_3 moiety and the anion helps stabilise the transition state and reduces the height of the energy barrier.

The next stationary point located on the PES was an energy minimum and corresponds to a phosphorane intermediate. The overall geometry of the intermediate

	O1-P-O5-C5	O1-P-O4-C4	H1-O1-P-O4
Reacts.	-31.11	34.17	-14.21
TS1	28.39	38.05	-77.74
INT	9.84	52.18	-98.40
TST2	113.64	96.40	-95.17
Prods.	-153.40	67.68	-95.88

Table 6.3: Key dihedral angles for the approximate stationary points in the DMP + OH^- reaction from the HF 3-21G*/TIP3P reaction energy profile. All values are given in degrees.

is trigonal bipyramidal with the incoming and departing ligands occupying the two apical positions, $|\angle$O1–P–O4$| = 171°$. Both of these ligands are partially bonded to the phosphorus with the P–O1 distance being 1.85 Å and the P–O4 distance at 1.77 Å. The equatorial methoxy group is still oriented towards the nucleophile.

The reaction then progresses via the formation of another transition state which constitutes the barrier for the second step of the reaction, the cleavage of the axial P–OCH_3 bond. The structure of this transition state, TS2, is also trigonal bipyramidal. The bond between the incoming oxygen and the phosphorus is almost fully formed, $|$P–$O_1| = 1.73$ Å, and the bond between the phosphorus and the leaving group has almost dissociated, $|$P–$O_1| = 2.30$ Å. Rotation of the P–O2 bond to orient the equatorial methoxy group towards the axial methoxy group has occurred. This results in a favourable electrostatic interaction which lowers the energy of TS2.

The reaction then embarks on a steep energetically downhill trajectory as it proceeds towards product formation. This culminates in the formation of another minimum which corresponds to the reaction product. The geometry of the product complex somewhat mirrors the reactant complex geometry. The P–O4 bond has fully dissociated and the interatomic distance between the phosphorus and the departing oxygen has extended to 4.50 Å and the bond between the nucleophile and the phosphorus is now fully formed with a bond length of 1.64 Å.

Energetics

The reaction was shown to proceed via the formation of two transition states and a single intermediate. The relative energies of each stationary point are shown in Figure 6.4. The energy difference between the reactants and TS1 is 22.68 kcal mol^{-1}, while TS2 is 5.5 kcal mol^{-1} higher in energy being 28.19 kcal mol^{-1} less stable than the reactants. Therefore, it can be concluded that the cleavage of the axial P–OCH_3 is the rate limiting step for the reaction.

The dianionic phosphorane intermediate species located on the PES is 3.6 kcal mol^{-1} and 9.1 kcal mol^{-1} more stable than TS1 and TS2 respectively. Thus, the depth of the energy well in which intermediate lies suggests that the intermediate has a more than a transient existence and is kinetically significant. The overall energy change for the reaction was negligible with the product being 0.40 kcal mol^{-1} higher in energy than the reactants.

It should be noted that all of these relative energies are not free energies but enthalpies. Bearing this in mind, the energy barrier for the reaction (28.19 kcal mol^{-1}) compares very favourably with the experimental values of 26.5 kcal mol^{-1} and 28 kcal mol^{-1} for $\Delta H^{\ddagger}$ determined by Radzicka *et al.* [89] and Haake *et al.* [195] respectively.

Charge distribution

The atomic charge distribution for each of the stationary points was calculated from the molecular electrostatic potential using the CHELPG scheme. The key atomic charges are presented in Table 6.4.

	P	O1	O2	O3	O4	O5
R	1.64	-1.32	-0.92	-0.98	-0.72	-0.56
TS1	1.77	-1.17	-0.95	-1.05	-0.81	-0.61
INT	1.87	-1.08	-1.02	-1.09	-0.88	-0.62
TS2	1.74	-1.04	-0.96	-1.01	-0.92	-0.61
P	1.92	-1.01	-1.04	-1.02	-1.07	-0.72

Table 6.4: Key electrostatic potential derived charges (e) for the reaction between DMP and OH^- as studied using a 3-21G*/MM hybrid potential.

The general trends in the magnitude of the atomic charges are similar to those described in Chapter 4. The atomic charge of the incoming oxygen O1 decreases by 0.3 e from the reactant complex to when it's fully bonded to the phosphorus as part of the MMP moiety. The same trend is observed for the leaving oxygen O4 whose charge increases by 0.35 e from when it is fully bonded in the reactant to when it is fully dissociated.

The charge distribution of the phosphorane intermediate is also of interest. The charges on the phosphorus and each of the oxygen atoms are more polarised than in either of the neighbouring transition states. The reason behind this large charge distribution could lie in the partially formed nature of both the axial P–O bonds which could facilitate greater distribution of charge throughout the pentacoordinate phosphorane.

Conclusions

By comparing these results with those obtained for the same reaction using a continuum representation of the solvent (Chapter 4), it can be seen that the use of an atomistic solvent representation leads to significant changes in the reaction energy profile. The results using a PCM solvent representation predicted that the reaction profile contained just one transition state which was reactant-like and corresponded to the formation of the new P–O bond. The use of a QM/MM potential located a geometrically similar transition state but also found another transition state with higher energy and which demonstrated that the cleavage of the bond between the phosphorus and the leaving group govern the rate of the reaction. Therefore, it can be concluded that, when compared to a continuum solvent representation, an atomistic representation of the solvent stabilises the approach of the nucleophile more than it does the cleavage of the scissile P–O4 bond. This may be considered to be a result of a superior representation of the solvation of the incoming hydroxide by an explicit solvent representation.

Also, the results of these calculations also suggested the presence of a stable dianionic phosphorane intermediate. The stability of this intermediate will be further probed in the next section of this chapter.

If the value of 28.2 kcal mol^{-1} obtained for the overall enthalpy barrier for the reaction is combined with the experimentally determined value of 8.3 kcal $mol^{-1}$for $-T\Delta S$ at 298 K [89] to give a free energy barrier of 36.5 kcal mol^{-1}, this estimation

of the reaction free energy barrier in reasonable agreement with the value of 36.5 kcal mol^{-1} obtained at the B3LYP(PCM)/6-31+G* level of theory despite the markedly different structures of the system at the energy barrier.

6.3.2 Further examination of phosphorane intermediate

As mentioned in Chapter 3, the existence of stable dianionic phosphorane species is a much debated topic in the literature [2, 4, 168, 188]. Theoretical calculations have shown the existence of such species to be sensitive to both solvation [182, 186] and the basis set employed [182, 184, 187]. With this in mind, it was decided to examine further the intermediate obtained using the HF 3-21G*/MM hybrid potential using different levels of theory to treat the quantum region.

Focusing on the middle region of the PES (Figure 6.3) which contains the intermediate and transition state species, energy evaluations using different hybrid potentials were performed on each geometry in the region. The different quantum mechanical potentials used to treat the QM region were: the HF/6-31+G* potential; the B3LYP/6-31+G* functional; and the AM1*d* semi-empirical potential. The energy profiles obtained are shown in Figure 6.5.

When the HF/6-31+G* level of theory was used to treat the quantum region, the phosphorane intermediate was observed to be no longer present and the region it occupied on the HF 3-21G*/MM surface was replaced by a 'shoulder' where the PES was almost flat. The use of this potential overestimated the magnitude of the energy barrier for this region of the PES. Although the potential employed included both *d*-type and diffuse functions for heavy atoms, electron correlation effects are not accounted for.

In order to remedy this, a hybrid potential which employed B3LYP/6-31+G* level of theory for the QM region was also used to examine this region of the PES. The B3LYP functional, which accounts for both electron correlation and exchange, has

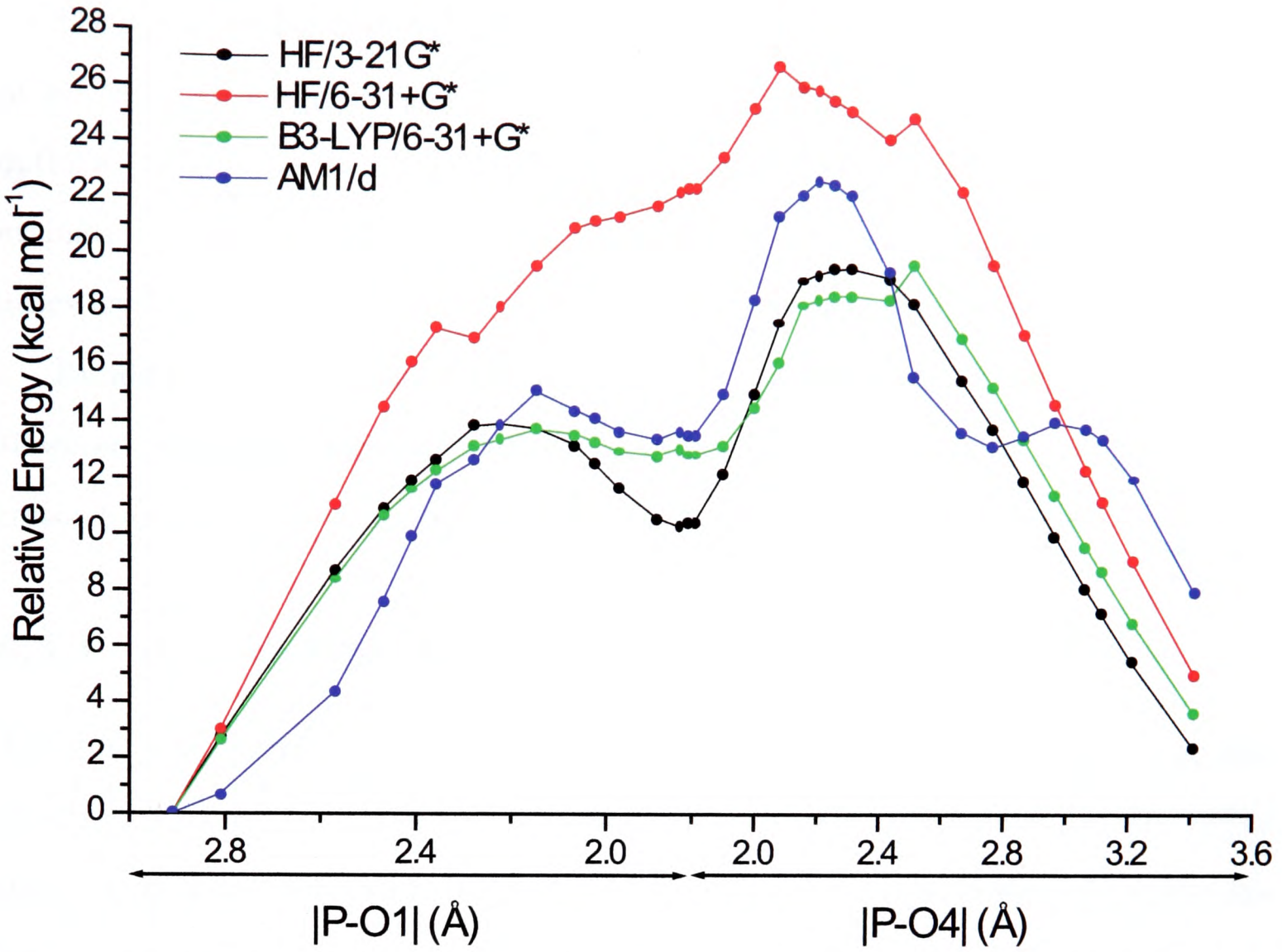


Figure 6.5: Single point energy calculations on the middle part of the 3-21G*/TIP3P PES calculated using different potentials for the quantum region.

been shown to provide accurate results for activation energies [26, 236] and transition state properties [254]. The use of this potential also discounted the existence of a stable phosphorane intermediate. The PES obtained has a very shallow potential energy well in the region corresponding to the intermediate species which would almost certainly be kinetically insignificant. The height of the energy barrier was in very close agreement with the value calculated at the HF 3-21G*/MM level for the corresponding region of the PES.

The use of the AM1*d* potential to treat the QM region indicated the presence of a metastable phosphorane intermediate. The depth of the potential well in this case is 2 kcal mol^{-1}. It is thus debatable whether the intermediate is kinetically significant.

The results outlined above cast doubt on whether an intermediate species does exist in solution. However, it should be borne in mind that the 'high level' calculations outlined in this sections are not geometry optimisations but single-point energy evaluations. It cannot be assumed that similar results would be obtained if full geometry optimisations were performed using these hybrid potentials.

To address this concern, attempts were made to obtain a HF/6-31G*/MM PES. However, it proved extremely difficult to obtain converged results, and this line of research was deemed to be beyond the computational resources available.

6.3.3 AM1*d*/MM potential energy surface

The potential energy surface for the base-catalysed hydrolysis of DMP was also determined using the AM1*d*MM potential. As discussed in Section 6.2.3, the AM1*d* potential is a derivative of the popular AM1 semi-empirical potential [31] which has been parameterised for phosphate ester hydrolysis reactions. The main advantage of using this potential to treat the QM region compared to standard *ab initio* or DFT potentials is the substantial decrease in CPU time required for energy evaluations[4]. Since the parameters used in this potential were derived from the results of high level *ab initio* and DFT calculations, it is hoped its use as part of a hybrid potential will provide further insight into the mechanism of the reaction.

The reaction system was generated and equilibrated using the methods outlined in Section 6.2.1 and the reaction PES was determined as per Section 6.2.2 except that it was possible to use a gradient tolerance of 0.01 kcal mol^{-1} $Å^{-1}$ for the minimisations due to the reduction in computational time required when using this potential.

The reaction was observed to follow an associative mechanism and proceed via the formation of two bipyramidal transition states with the attacking and leaving groups at the apical points (Figure 6.6). A phosphorane intermediate was found to occupy

[4]The CPU time required for one SCF cycle using the AM1*d* potential is approximately two orders of magnitude faster than the time taken using a HF/3-21G* potential.

the potential energy well lying between the two transition states. The geometries of the QM region at each stationary point are depicted in Figure 6.7 and key geometrical parameters are given in Tables 6.5, 6.6, and 6.7.

Stationary Points

The shape of the PES is very similar to that of the HF 3-21G*/MM PES described in Section 6.3.1. The structure of the reactant complex shows that the hydroxyl anion is poised to attack DMP with the the hydroxyl oxygen 5.85 Å away from the phosphorus atom. Even at this stage of the reaction, the angle of approach, 156.2° suggests that the angle of attack will be 'inline'. The structure of the first transition state, TS1, is slightly later on the reaction coordinate than the equivalent transition state on the HF 3-21G*/MM PES. The distance between the hydroxyl anion and the phosphorus has closed to 2.10 Å, and the bond between the phosphorus atom and the departing methoxy group has lengthened by 0.1 Å to 1.81 Å. The geometry of the transition state of trigonal bipyramidal and the angle O1–P–O4 has opened.

By the time the reaction progresses to the formation of the dianionic phosphorane intermediate, INT, both the bonds at the apices of the trigonal bipyramid are partially formed. The distance between the incoming oxygen and the phosphorus is 1.86 Å and the bond between the phosphorus and the departing oxygen has lengthened to 1.85°.

	P-O1	P-O2	P-O3	P-O4	P-O5	O1-H1	O4-C4	O5-C5
Reacts.	5.850	1.531	1.581	1.706	1.701	0.948	1.441	1.452
TS1	2.099	1.553	1.554	1.808	1.721	0.964	1.414	1.434
INT	1.862	1.562	1.569	1.851	1.737	0.972	1.407	1.430
TS2	1.811	1.548	1.555	2.204	1.715	0.976	1.381	1.442
Prods.	1.718	1.530	1.546	4.851	1.692	0.984	1.372	1.451

Table 6.5: Key bond lengths (Å) for the approximate stationary points in the DMP + OH reaction from AM1d/MM minimisations.

The second transition state, TS2, is almost symmetric to TS1. The bond between the oxygen and the phosphorus is almost fully formed, $|\text{P–O}_1| = 1.81$ Å, and the leaving

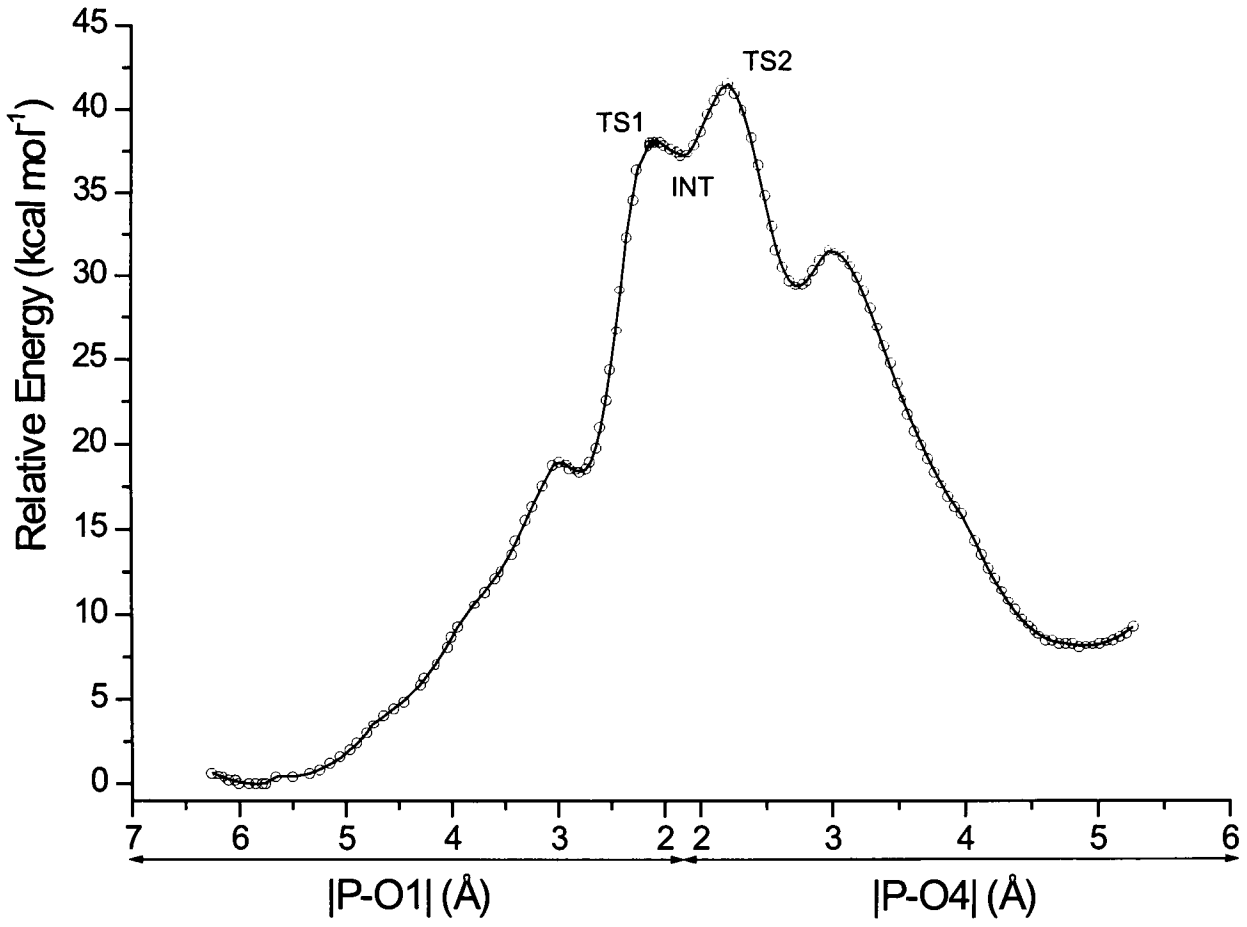


Figure 6.6: The potential energy surface calculated for the reaction DMP + OH^- using the AM1d/MM potential. The x-axis, which describes the reaction coordinate, is divided into two parts reflecting the two different descriptions used in the calculation of the PES. |P–O1| is the distance between the phosphorus and the incoming hydroxyl oxygen and |P–O4| is the distance between the phosphorus and the leaving methoxy oxygen.

group has partially dissociated, $|P–O_4| = 2.20$ Å. It is worth noting that, unlike the equivalent transition state on the HF 3-21G*/MM PES, the equatorial methoxy group has not rotated so as to stabilise the departure of the leaving group. This region of the reaction coordinate was mapped with the equatorial methoxy group *cis* to the leaving methoxy but it was found that the energy of the system increased. Finally, the reaction proceeds to form products. The P-OH bond is fully formed with a bond length of 1.72 Å and the leaving group has fully dissociated, $|P–O_4| = 4.85$ Å.

There are also two further 'shoulders' on the PES. Both of these occur at points on the PES where the P–O1 or P–O4 bond distances are in the region of 3 Å, respectively. The second barrier is considerably more pronounced. The exact nature of these energy barriers are unclear. Since, in each case, one of the two axial bonds has almost fully dissociated while the other is almost completely formed, these barriers may correspond

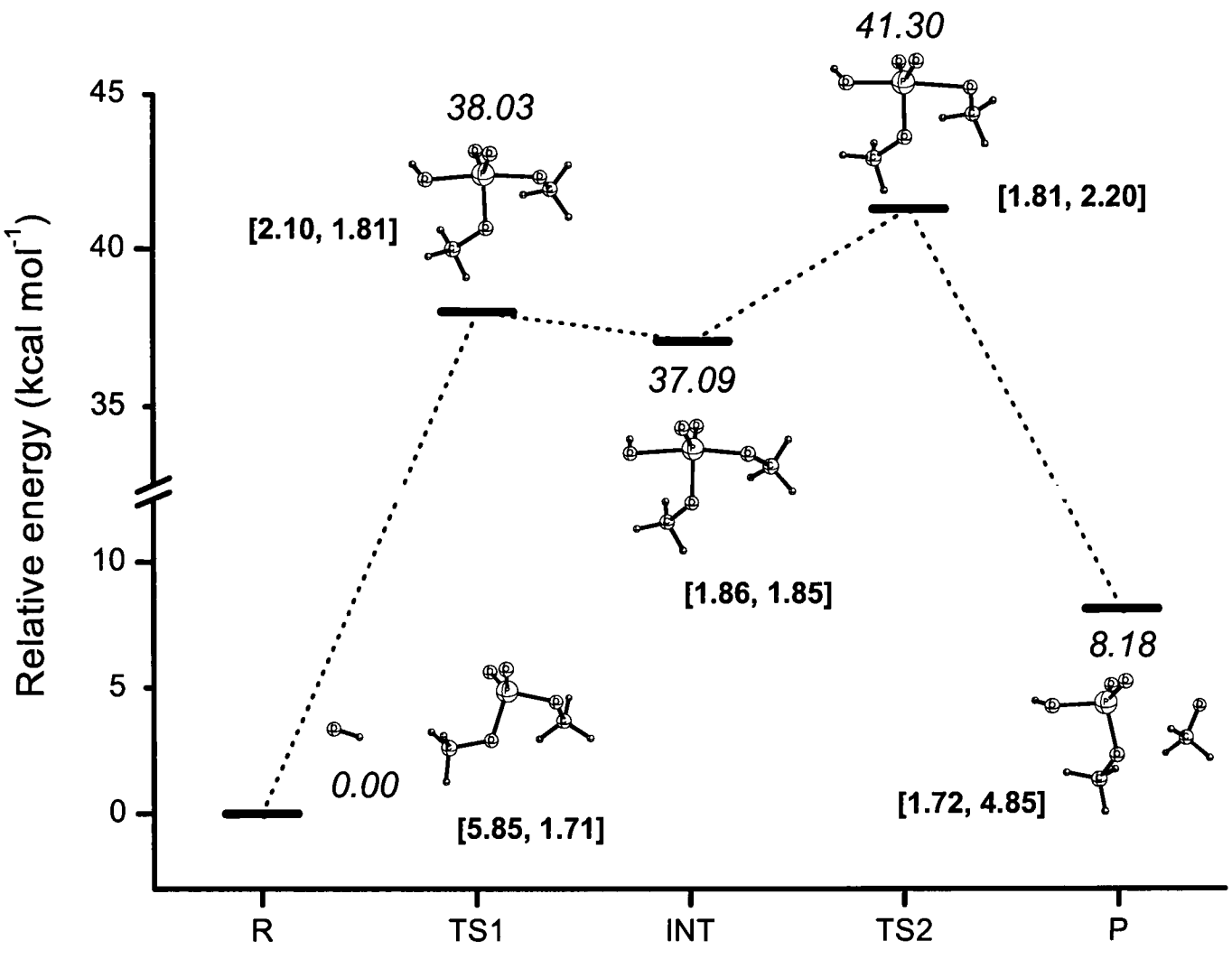


Figure 6.7: The geometries of the quantum region for each stationary point of the base-catalysed hydrolysis of DMP in solution as calculated using the AM1d/MM potential. The energy of each stationary point relative to the reactants is shown in italics and the P–O1 and P–O4 bond lengths are respectively shown in brackets.

to the formation of encounter complexes as the OH^- anion or the axial methoxy anion approaches. Graphical inspections of the system geometry in these regions were made in the expectation that a solvent hydrogen bond may have to break in order for the solute to move but no evidence of this could be found. Therefore, the exact cause of these shallow energy wells cannot be elucidated although personal correspondence with the developers of the potential leads us to believe that the shoulders are either an artifact of the NDDO approximation or the fact that the potential was not well-parameterised for a P–O separation of this magnitude.

	O1-P-O4	O2-P-O3	O2-P-O5	O3-P-O5	O1-P-O5	O5-P-O4	O1-P-O2	O3-P-O4
Reacts.	156.18	122.07	110.94	110.17	59.55	96.66	84.08	106.17
TS1	170.91	126.28	116.94	115.95	85.09	86.01	86.06	95.91
INT	171.67	125.85	117.39	116.76	87.84	83.83	89.96	91.92
TST2	174.44	128.65	117.23	113.02	91.51	84.84	91.96	88.17
Prods.	141.50	122.48	111.61	108.88	99.41	80.48	105.34	110.02

Table 6.6: Key bond angles for the approximate stationary points in the DMP + OH reaction as calculated using the AM1d/MM hybrid potential.

	O1-P-O5-C5	O1-P-O4-C4	H1-O1-P-O4
Reacts.	-3.89	68.88	-38.06
TS1	37.40	60.44	-106.94
INT	34.17	69.89	-122.48
TST2	45.11	2.23	-70.86
Prods.	63.58	-85.93	-18.68

Table 6.7: Key dihedral angles for the approximate stationary points in the DMP + OH reaction as calculated using the AM1d/MM hybrid potential.

Energetics

The relative energies of the reaction stationary points, shown in Figure 6.7, display the same trends as those calculated using the HF 3-21G*/MM hybrid potential. The first transition state, TS1, is 38.0 kcal mol^{-1} less stable than the reactants. This barrier is substantially larger than the equivalent one calculated at the HF 3-21G*/MM level. However, this overestimation of the energy barrier was expected for two reasons. Firstly, semi-empirical methods employ a minimal basis set representation of atomic orbitals. While this normally gives reasonably accurate results, it has been shown that, for highly negative isolated anions such as OH^- and MeO^-, the use of a split valence basis set would be more apt. Secondly, it must also be taken into account that the AM1d potential was not parameterised to deal with systems which contain a phosphorus and oxygen atom at such a large separation and therefore the energies calculated for the quantum region in this area of the PES may not be as accurate. However, neither of these deficiencies are likely to be apparent in the region of the PES containing the reaction transition states and the intermediate and the topology of this region can be considered to be accurately portrayed.

The intermediate species was determined to be 0.9 kcal mol^{-1} more stable than TS1 and 4.2 kcal mol^{-1} more stable than TS2. Therefore, it can be concluded that, at this level of theory, the intermediate is only metastable and probably would not be kinetically significant. The final transition state, TS2 is 41.3 kcal mol^{-1} higher energy than the reactants. Therefore, it can be concluded that the cleavage of the bond to the departing group is the rate-determining step for the reaction.

The overall energy change for the reaction was computed to be +8.2 kcal mol^{-1}. The difference between this value and the equivalent value of 0.4 kcal mol^{-1} calculated at the HF 3-21G*/MM level of theory may be attributable to the AM1*d* potential being more capable of accurately estimating the energy of the free MeO^- anion accurately compared to that of the OH^- ion which has greater charge density.

Is the reaction mechanism associative or dissociative?

The results outlined so far in this chapter and in Chapter 4 indicate that the alkaline hydrolysis of DMP proceeds via an associative mechanism. This conclusion is arrived at because, in both the gas phase and in solution, pentacoordinated phosphorane transition states and intermediates are encountered on the PES as opposed to trivalent phosphate ions which are indicative of a dissociative mechanism.

In a series of articles [167, 175, 176], Florián and Warshel examined both of the mechanistic alternatives for the hydrolysis of methylphosphate in solution. Their results, which contradicted experimental evidence [2,255,256], concluded that, for both the mono- and dianionic protonation states of methylphosphate, the energy barriers for the associative and dissociate mechanisms have a very similar magnitude. This indicated that both mechanistic pathways are equally likely to be followed in solution.

With this in mind, it was decided to determine the relative probability of each of the mechanistic alternatives for the alkaline hydrolysis of DMP in more detail by determining the energy barrier associated with each pathway. This was done by using

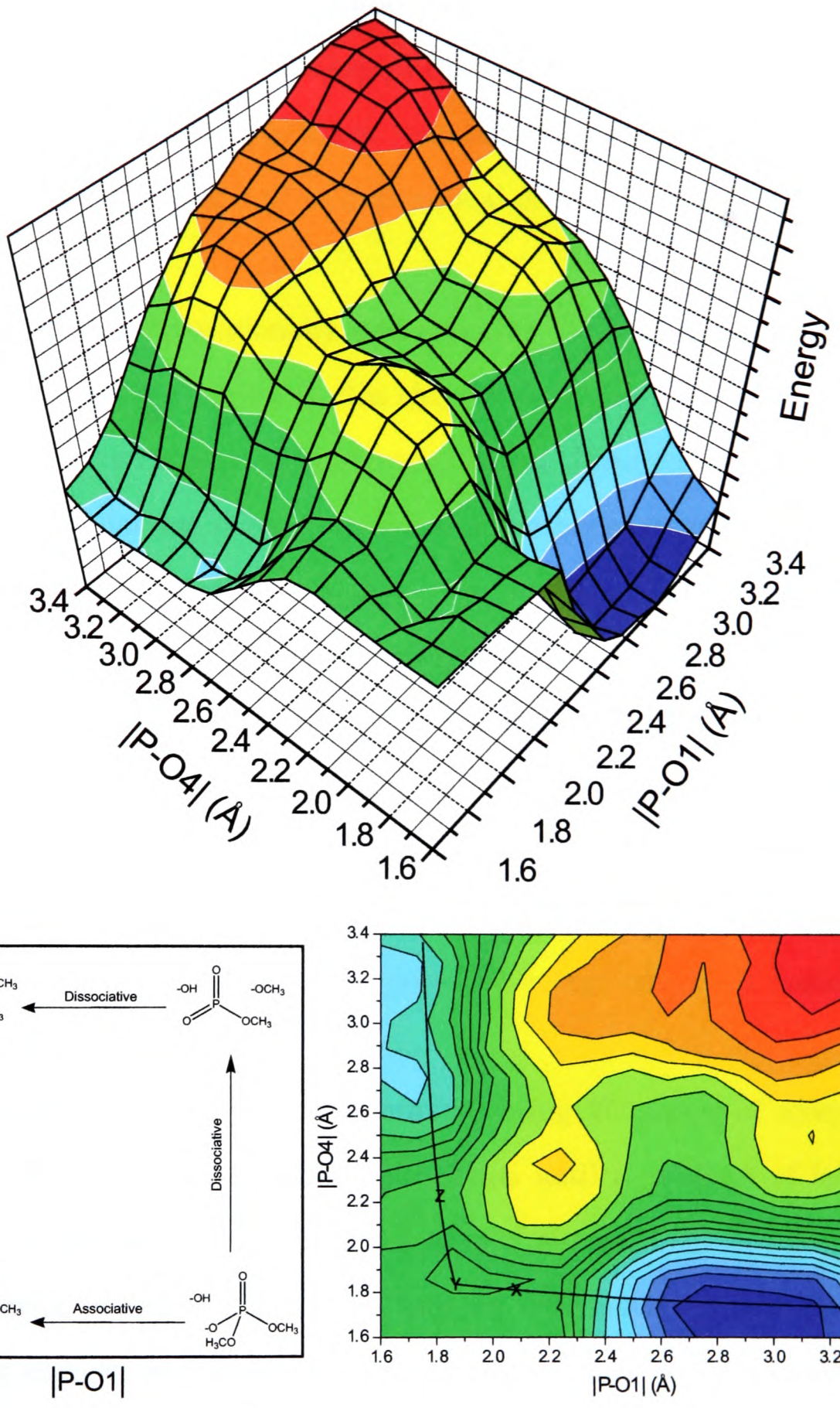


Figure 6.8: The topmost part of the figure shows the 3D PES for the DMP + OH^- reaction in solution as calculated using a QM/MM potential which employed the AM1*d* potential for the quantum region. The bottom left plot is a contour plot of the same data on top of which is superimposed the reaction coordinate of the 2D reaction coordinate determined in Section 6.3.3. The points X, Y, and Z refer to the stationary points TS1, INT, and TS2 respectively. The bottom right plot outlines the two major mechanisms that the reaction can proceed via: the associative and dissociative pathways.

the QM/MM minimisation and adiabatic mapping protocols described earlier in this chapter. A similar protocol was used except that the internal coordinates corresponding to the axial P–OH and P–OCH_3 bonds were simultaneously harmonically restrained to facilitate the generation of a three-dimensional plot of the reaction potential energy surface.

This a far more computationally intensive task than the generation of the two-dimension plots described earlier in this chapter since the calculations of an $N \times N$ matrix of grid points is required to construct a 3D plot. However, since the AM1d allows for rapid SCF calculations relative to *ab initio* methods, it proved computationally feasible to generate such a plot.

The two-dimensional AM1d/MM PES (Figure 6.6) was used to provide the starting geometries required. Each line along the grid of points was generated by fixing the value of one of the two internal coordinates while the value of the other was changed in 0.05 Å increments within the 1.7 Å to 3.4 Å range of values. The endpoints of the reaction coordinate were not included since this would greatly increase the computational expense and because this region of the potential energy surface is common to both mechanistic pathways.

At each point, the system[5] was minimised using 200 steps of steepest descent minimisation followed by an ABNR minimisation until a gradient of less than 0.01 kcal mol^{-1} $Å^{-1}$ was attained. This series of minimisations led to the generation of a 34×34 matrix of values for the two internal coordinates being examined. The corresponding energies were plotted as a 3D energy surface and a contour plot. Both of these are shown in Figure 6.8.

The results obtained show that starting from the point on the PES where the P–OH distance is 3.4 Å and the axial P–OCH_3 distance is 1.73 Å, the overall barrier for the associative pathway is 16 kcal mol^{-1} whilst the barrier for the dissociative pathway is

[5]Just in in the previously described QM/MM studies, the outer 5 Å of the system was fixed in order to reduce the degrees of freedom within the system and facilitate more rapid convergence.

14 kcal mol^{-1} higher at 30 kcal mol^{-1}. The value of the dissociative barrier was arrived at by assuming that the reaction pathway for this mechanism passes through the point on the PES where the length of both the axial bonds is 3.0 Å. This is in line with the approach adopted by Florián and Warshel in their comparison of the associative and dissociative pathways for the hydrolysis of methylphosphate [167]. From these results it can be concluded that the base-catalysed hydrolysis of DMP proceeds via an associative mechanism in solution. The region of the reaction PES corresponding to the formation of a trivalent metaphosphate ion shows that such a species is very unstable even in the presence of the stabilising effect of solvent.

6.4 Summary

The results outlined in this chapter predict that the base-catalysed hydrolysis of dimethyl phosphate in solution proceeds via an associative mechanism and that the reaction potential energy surface contains two transition states and an intermediate. The PES calculated using a HF 3-21G*/MM hybrid potential predicts this intermediate to be stable and thus kinetically significant. However, subsequent energy evaluation employing more accurate representations of the QM region cast doubt on this. The use of the AM1*d* semi-empirical potential gave slightly disappointing but not unexpected results for the endpoints of the reaction. However, it was found to provide accurate geometries and energies for the middle section of the reaction coordinate which included the phosphorane species generated during the course of the reaction and thus the existence of dianionic phosphoranes in solution could be explored in an accurate manner. This hybrid potential was found also to predict the presence of a phosphorane intermediate species. However, the stability and kinetic significance of the intermediate is doubtful. Finally, the generation of a three-dimensional representation of the middle region of the PES predicts that the reaction proceeds via an associative mechanism and

that the relatively large barrier for the dissociative pathway precludes the possibility of it occurring in solution.

Chapter 7

A QM/MD simulation of the base-catalysed hydrolysis of DMP

7.1 Introduction

The results of the QM/MM geometry optimisations presented in Chapter 6 provide valuable insight into the role of solvent in the base-catalysed hydrolysis of DMP through an atomistic representation of the solvent. In this chapter, the results of a study combining molecular dynamics (MD) simulations with the use of a hybrid QM/MM potential are presented. This method, known as QM/MD, is used to determine the potential of mean force (PMF) for the base-catalysed hydrolysis of DMP.

In Chapter 1 (Section 1.12.3), the use of MD simulations to determine the potential of mean force (PMF) of a chemical reaction was discussed. Of the various methods that can be used, particular attention was paid to the umbrella sampling method [73]. This method, used in conjunction with MD simulations, facilitates the exploration of all conformational space along a reaction coordinate and thus accounts for the contribution of high energy states encountered on the reaction coordinate.

As well as presenting the PMF for the base-catalysed hydrolysis of DMP, this results presented here also give an important insight into the average solute and solvent structure at various key points on the reaction PMF.

7.2 Methods

7.2.1 Preparation of system

The rate-limiting transition state from the adiabatically-mapped AM1d/MM reaction profile described in Section 6.3.3 was used as the initial system geometry. Since MD simulations are inherently computationally more expensive than geometry optimisations because of the far greater number of energy evaluations required for the determination of a complete reaction energy profile, the AM1d potential [251] was used to treat the quantum region in the QM/MD simulations. The chemical structure of the QM region is depicted, along with the atom-numbering scheme referred to in the rest of this chapter, in Figure 7.1. The solvent was treated classically using the TIP3P parameters for liquid water [78]. All calculations were performed using a locally-modified version of CHARMM [39].

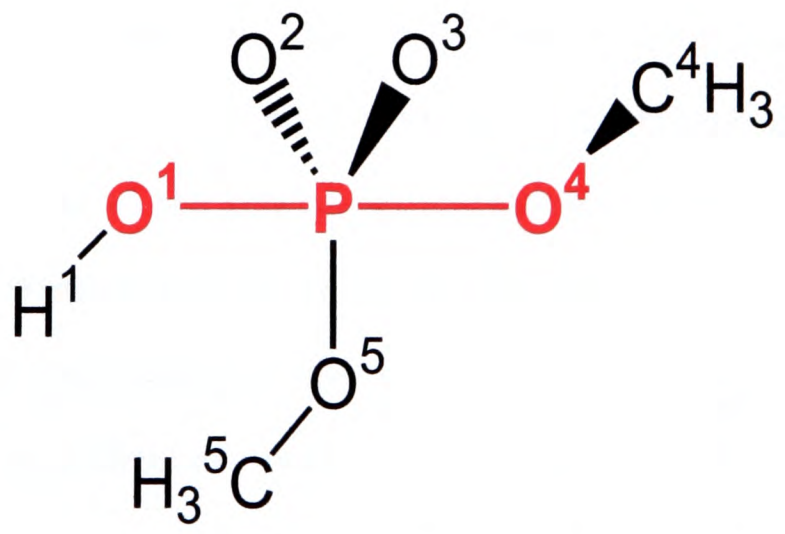


Figure 7.1: The quantum region for the free energy QM/MD calculations. The two bond lengths whose difference comprises the reaction coordinate are highlighted in red.

7.2.2 Umbrella sampling

The reaction PMF was determined using umbrella sampling methods in conjunction with a series of MD simulations. The reaction coordinate used, ξ, was defined as the difference between the P–O4 and the P–O1 interatomic distances ($\xi = |P\text{–}O_4| - |P\text{–}O_1|$). The use of a single reaction coordinate, unlike in Chapter 6, was necessitated since joining PMFs generated using different reaction coordinates such as single bond lengths is very difficult in the context of umbrella sampling calculations. Adequate sampling of the entire reaction coordinate was ensured by dividing it into a series of overlapping 'windows'. Each successive window was centered at points 0.1 Å apart on the reaction coordinate. The conformational space within each window was sampled using a stochastic-boundary molecular dynamics simulation in the presence of a harmonic restraining function which ensured that the value of ξ was kept within 0.2 Å of the centre of the window for the majority of the simulation. The restraining potential was imposed using the RESD module [49] implemented within CHARMM and a force constant of 200 kcal mol^{-1} $Å^{-2}$ was employed.

7.2.3 Simulation protocol

Before sampling of the reaction coordinate was commenced, the system was equilibrated. The benefits of a well-equilibrated system include the absence of energy 'hot-spots' which may lead to the generation of unusually large forces for atoms within the system. This is especially problematic for atoms in the QM region since large gradients can result in convergence problems in the SCF calculations.

The system was equilibrated using an extended QM/MD simulation. Since the initial system geometry was obtained from a geometry optimisation, therefore corresponding to the geometry at 0 K, the system was gradually heated to 298 K in evenly spaced increments during the course of a 100 ps QM/MD simulation. At each temperature increment, new velocities were assigned to the system atoms according

to a random Maxwellian distribution. An umbrella potential which maintained the difference between the P–O4 and P–O1 bond lengths was applied throughout. This was followed by a further 50 ps of equilibration at 298 K.

In order to avoid the computational expense of a full periodic boundary simulation, stochastic boundary methods [38, 43] were used to represent the influence of solvent molecules beyond the boundaries of the system. This entailed the division of the system into a reaction zone and a reservoir region, and the further division of the reaction zone into a reaction region and a buffer region [43]. The reaction region consisted of a sphere of radius r=13 Å about the centre of the system (the phosphate phosphorus); the buffer region comprised the outer shell of the system (13 Å $< r <$ 16 Å). The reservoir region simulated the rest of the system, and was not represented atomistically.

The reservoir region was represented by a combination of a mean force and a dissipative force. The purpose of the buffer region was to act as a heat bath for the thermal fluctuations occurring in the reaction region. The atoms in the buffer region interacted with a stochastic heat bath, via random fluctuating forces and dissipative forces, which were present to account for the dynamics of the neglected reservoir region atoms. A deformable boundary potential is also applied to atoms in the buffer region to simulate the forces exerted on them from the 'atoms' in the reservoir region. The dynamics of the reaction region were propagated using standard Newtonian dynamics. The buffer region was simulated using Langevin dynamics where the friction constant for the solvent molecules was 62 ps^{-1}. A timestep of 1 fs was used throughout the simulations and the bonds within the solvent molecules were fixed using the SHAKE algorithm [68]. Long-range electrostatic forces were smoothly truncated within a range of 8–12 Å using the CHARMM atom-based force shift method [257]. The van der Waals interactions were considered in a similar manner using the shift method [257]. The CHARMM non-bonded interaction list was calculated every ten steps during the simulation using a 13 Å cutoff.

Each window was simulated for 25 ps. The first 10 ps were for equilibration and only the last 15 ps of each trajectory, the production phase, were used for gathering statistics. Each simulation began with the coordinates and velocities from the end of the preceding window.

7.2.4 Generation of the PMF

During the production phase of each simulation, the values of |P–O1| and |P–O4| were recorded at each step. Although CHARMM possesses a module that is capable of calculating PMFs from these data, individual PMFs for each window were calculated using locally written programs. This facilitated experimentation with different parameters for the calculations and also avoided many of the bugs encountered in the equivalent routines within CHARMM. The values of |P–O1| and |P–O4| determined in each window were converted into values of ξ, the reaction coordinate, and distributed into 0.05 Å wide 'bins' and the resulting statistics were used to calculate the PMF for each window according to the formulae detailed in Section 1.12.3 (page 37). The PMF for the entire reaction, W, was obtained by connecting the individual PMFs through the calculation of the weighted average of the overlapping regions of successive PMFs.

7.3 Results

The reaction PMF, shown in Figure 7.2, indicates the presence of two transition states and an intermediate on the reaction pathway. The reaction was also found to proceed via an associative rather than a dissociative mechanism. Average geometric parameters from the windows corresponding to each stationary point are given in Tables 7.1, 7.2, and 7.3.

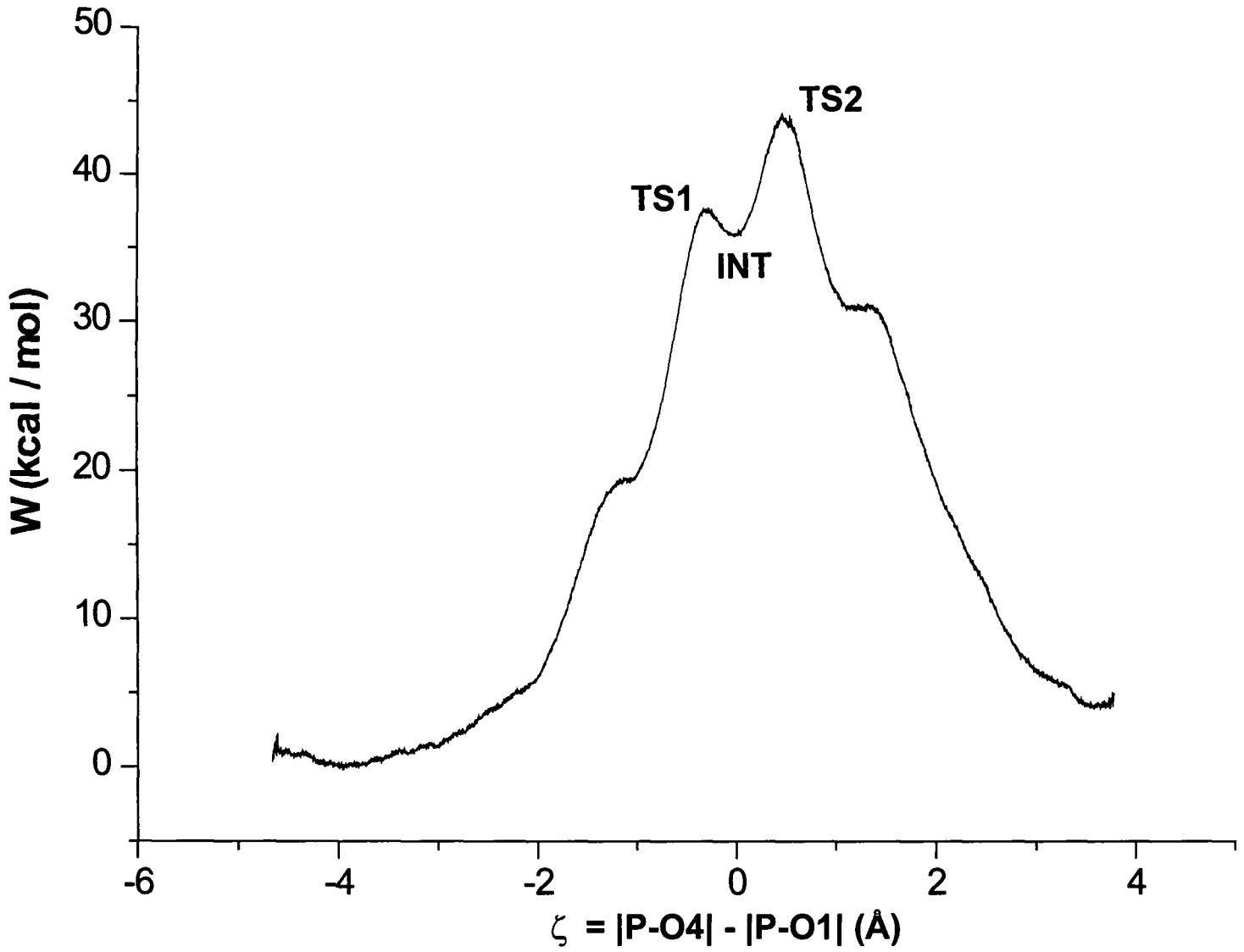


Figure 7.2: The PMF for the base-catalysed hydrolysis of dimethyl phosphate as determined using AM1d/MD simulations in conjunction with umbrella sampling methods.

7.3.1 Stationary points

The minimum located at ξ=-3.95 Å was found to relate to the reactant complex. The average distance between the hydroxyl oxygen (O1) and the phosphorus within the relevant window in was determined to be 5.65 Å. The bond between the phosphorus and the axial methoxy group is completely formed at 1.66 Å. The angle between the incoming oxygen, the phosphorus, and the leaving oxygen (O1–P–O4) has an average value of 128°. However, it is clear from the rms fluctuation of ±11.32° for this angle that the hydroxyl anion moves quite freely in solution at this stage of the reaction.

The first transition state, TS1, occurs at ξ= -0.33 Å indicating that the newly formed P–O1 bond is 0.33 Å longer than the bond between the phosphorus and the departing methoxy group thus indicating an early transition state. The average values

	P-O1	P-O2	P-O3	P-O4	P-O5
Reacts.	5.645 ± 0.060	1.554 ± 0.021	1.572 ± 0.022	1.658 ± 0.029	1.670 ± 0.029
TS1	2.078 ± 0.102	1.568 ± 0.022	1.572 ± 0.023	1.767 ± 0.039	1.710 ± 0.032
INT	1.854 ± 0.049	1.576 ± 0.023	1.581 ± 0.021	1.801 ± 0.033	1.734 ± 0.035
TS2	1.787 ± 0.053	1.565 ± 0.023	1.563 ± 0.025	2.144 ± 0.149	1.706 ± 0.037
Prods.	1.678 ± 0.031	1.553 ± 0.020	1.559 ± 0.021	5.286 ± 0.060	1.667 ± 0.029

Table 7.1: Average values and rms fluctuations of key bond lengths (Å) from the windows corresponding to the stationary points determined by AM1d/MD simulations for the base-catalysed hydrolysis of DMP.

of these two bond lengths are 2.08 Å and 1.77 Å, respectively. This indicates that the bond between the phosphorus and the axial methoxy group has lengthened by 0.11 Å compared to the reactant. The average value of the angle of attack (O1–P–O4) has opened to 169.53° which is indicative of an 'in-line' attack mechanism. The various bond angles detailed in Table 7.2 confirm that the structure of the transition state is trigonal bipyramidal. This was also confirmed through visualisation of the trajectory obtained for the window corresponding to this region of the reaction coordinate. The equatorial methoxy group is oriented towards the leaving group, $|\angle$O1–P–O5–C5$|$ = -121.31°.

	O1-P-O4	O2-P-O3	O2-P-O5	O3-P-O5
Reacts.	128.18 ± 11.32	112.13 ± 3.55	110.77 ± 4.09	106.14 ± 4.07
TS1	169.53 ± 4.19	121.79 ± 4.18	119.54 ± 4.90	115.52 ± 4.70
INT	167.46 ± 5.00	121.78 ± 4.01	122.45 ± 5.22	115.38 ± 5.01
TS2	166.45 ± 4.82	123.36 ± 4.63	119.91 ± 4.46	115.82 ± 4.62
Prods.	137.38 ± 15.41	113.24 ± 3.85	110.90 ± 4.93	107.88 ± 4.19
	O1-P-O5	O5-P-O4	O1-P-O2	O3-P-O4
Reacts.	117.05 ± 8.41	106.09 ± 4.14	96.70 ± 7.56	108.19 ± 3.99
TS1	78.63 ± 3.66	92.52 ± 3.52	88.25 ± 3.84	99.00 ± 3.60
INT	82.95 ± 3.46	86.35 ± 3.73	91.25 ± 3.59	92.11 ± 3.81
TS2	87.48 ± 3.93	80.40 ± 3.97	95.28 ± 3.78	91.68 ± 4.45
Prods.	104.66 ± 4.62	108.09 ± 15.07	109.83 ± 4.49	49.59 ± 19.66

Table 7.2: Average values and rms fluctuations of key bond angles, in degrees, from the windows corresponding to the stationary points determined by AM1d/MD simulations for the base-catalysed hydrolysis of DMP.

The next stationary point located on the PMF is an intermediate. This intermediate, INT, is located at ξ= -0.06 Å on the reaction coordinate and has a structure which is typical for a phosphorane intermediate with the bonds between the

phosphorus and the two apical ligands almost completely formed: |P–O1| = 1.85 Å and |P–O1| = 1.80 Å. The average values of the bond angles indicate that the structure of the intermediate is trigonal bipyramidal.

The rate-limiting transition state, TS2, is product-like and is assumed to correspond to the cleavage of the apical P–OCH_3 bond. The P–OH bond is almost fully formed at this stage of the reaction, |P–O1| = 1.79 Å, and the bond between the phosphorus and the leaving group has lengthened by 0.34 Å relative to the intermediate, |P–O4| = 2.14 Å. Again the structure of the transition state is trigonal bipyramidal. The equatorial methoxy group adopts a position that orients it towards neither of the axial ligands, |∠O1–P–O5–C5| = 92.9°. However the rms fluctuation of ±20.9 Å suggests that this is a very loose degree of freedom.

	O1-P-O5-C5	O1-P-O4-C4	H1-O1-P-O4
Reacts.	168.86 ± 28.00	138.37 ± 17.34	18.01 ± 50.10
TS1	-121.30 ± 17.51	140.67 ± 14.89	-138.59 ± 47.63
INT	-77.93 ± 15.44	-133.73 ± 32.43	99.05 ± 39.53
TS2	-92.88 ± 20.92	-158.63 ± 26.67	37.80 ± 44.47
Prods.	-32.29 ± 46.17	99.15 ± 37.10	-152.83 ± 36.83

Table 7.3: Average values and rms fluctuations of key dihedral angles, in degrees, from the windows corresponding to the stationary points determined by AM1d/MD simulations for the base-catalysed hydrolysis of DMP.

The last stationary point on the PMF is the product complex. As expected, the P–OH bond is fully formed at 1.68 Å and the leaving group has completely dissociated from the remaining MMP moiety as evidenced by a 5.29 Å separation between the phosphorus and the leaving oxygen. The relevant bond angles indicate that the structure of MMP is that of a distorted tetrahedron.

Just as in the AM1d/MM energy profile presented in Chapter 6, two 'shoulders' were observed on the reaction coordinate at the values ξ = -1.1 and ξ = 1.2. Inspection of the trajectories in the region corresponding to these shoulders gave no indication of any discernible chemical processes that would lead to their formation. Therefore, it can only be again assumed that they are artefacts of the QM potential employed.

7.3.2 Radial distribution functions

The average structure of the solvent around a specific solute molecule can be represented by radial distribution functions (RDFs) [62]. Each function, which is denoted by $g(r)$, gives the probability of finding a pair of atoms at a distance r apart, relative to a random distribution with the same density. In this study, the radial distribution functions between O1, the incoming oxygen, and the water oxygens and O4, the departing oxygen, and the water oxygens are calculated from the trajectories corresponding to each stationary point. In order to gain more accurate RDFs each of the relevant simulations were run for a further 10 ps and the resulting trajectory was combined with the existing trajectory to generate the RDFs.

The RDFs obtained for each stationary point are shown in Figure 7.3. Some trends are immediately apparent. In the reactant complex, a very well defined solvation shell is observed around the hydroxide oxygen, O1. This is located at $r \approx 2.7$ Å and at this O1–O_{wat} distance, the probability of finding a water oxygen is 7.35 times more likely than in a random distribution. This indicates the presence of strong hydrogen-bonding interactions which lead to a well-defined first solvation shell. There is also evidence of a secondary solvation shell at $r \approx 4.9$ Å. Since, at this stage of the reaction, the axial methoxy ligand is fully bonded to the phosphorus, very little evidence of an equivalent coordination sphere centred about O4 is found.

Similar solvent behaviour is observed for the product complex. In this case, the strong hydrogen-bonding interactions are between the water molecules and the methoxide oxygen, O4, at $r \approx 2.7$ Å. The value of $g(r) = 5.6$ is slightly lower for this peak compared to the O1-O_{wat} reactant distribution. This reflects the steric bulk of the methoxy anion which reduces the number of hydrogen-bonding interactions formed by O4 compared to the hydroxyl oxygen, O1.

The RDFs calculated for the stationary points TS1, INT, and TS2 give a fascinating account of the change in solvent structure surrounding O1 and O4 as the chemical

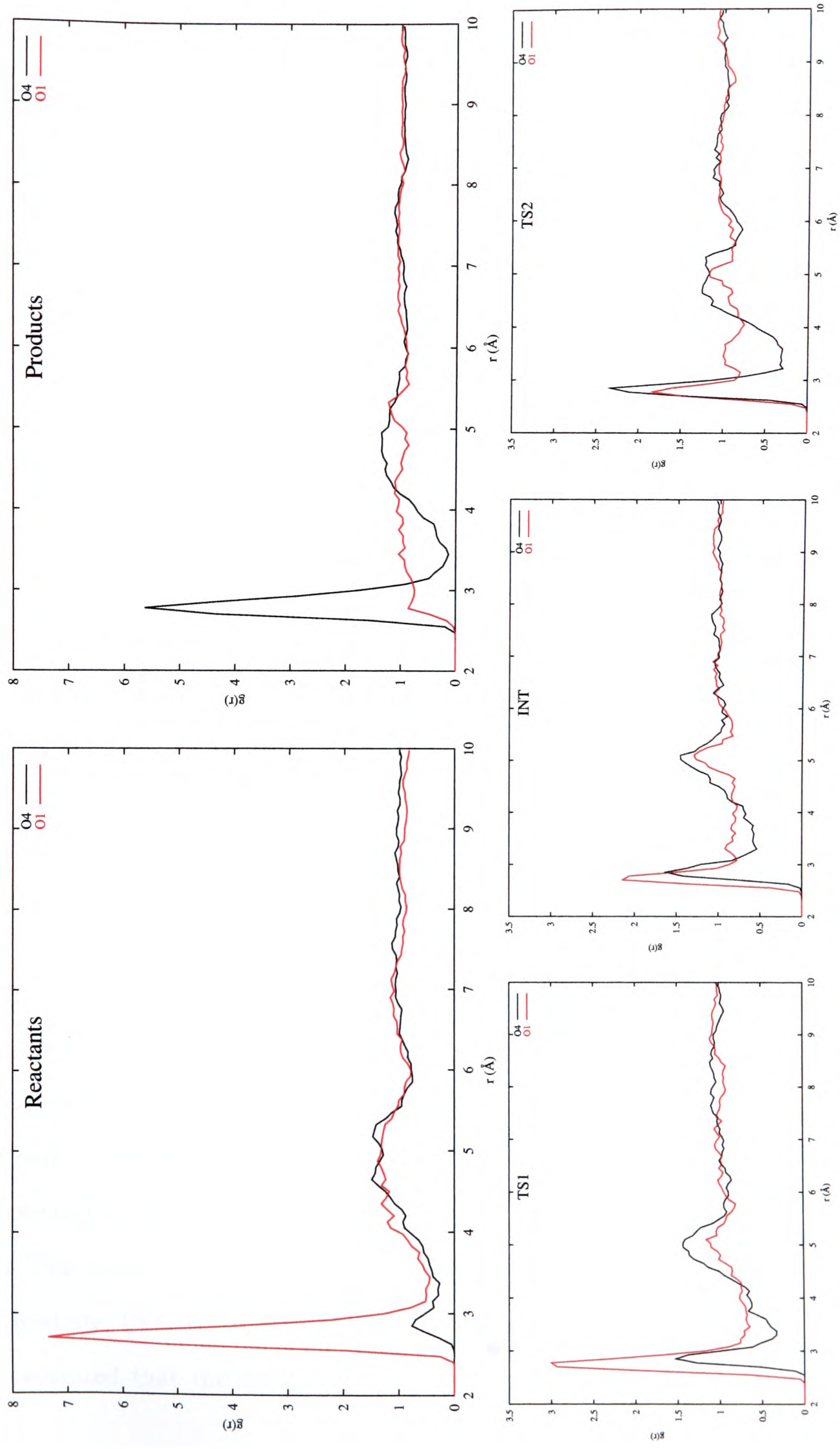


Figure 7.3: The radial distribution functions for each stationary point on AM1*d*/MD PMF. The red lines show the O1-O_{water} RDF and the black lines show the O4-O_{water} RDF.

reaction progress through this region of the PES. The disruption of the once clearly defined solvation shell surrounding the attacking O1 as it forms a bond with the phosphorus is evidenced by a decrease in the value of O1–O_{wat} $g(r)$ value from 3.00 to 2.14 to 1.83 on going from TS1 to INT to TS2.

The opposite trend is observed for O4. As the leaving group departs the solvent aids its departure by forming hydrogen bonding interactions with the departing oxygen. This is illustrated by the increase in $g(r)$ from 1.52 to 1.62 to 2.85 on going from TS1 to INT to TS2.

7.3.3 Energetics

The relative energies of each stationary point relative to the reactants are shown in Table 7.4.

	ΔG
Reacts.	0.00
TS1	37.62
INT	35.83
TS2	43.99
Prods.	4.65

Table 7.4: Relative free energies of each stationary point on the AM1d/MD PMF. All energies are given in kcal mol^{-1}.

Just like the AM1d/MM reaction energy profile presented in Chapter 6 (Section 6.3.3), the rate-limiting transition state for the reaction was determined to correspond to the departure of the leaving group. This transition state, TS2, is 44.0 kcal mol^{-1} less stable than the reactants. The preceding transition state, TS1, which represents the barrier to the formation of the P–O1, has 6.4 kcal mol^{-1} lower energy than TS2 being 37.6 kcal mol^{-1} less stable than the reactants. The phosphorane intermediate, INT, is just 1.8 kcal mol^{-1} more stable than TS1 and, therefore, it may be speculated that the existence of this intermediate is likely to be short-lived and

therefore likely to have little kinetic significance in terms of the overall reaction energy profile.

The overall energy barrier of 44.0 kcal mol^{-1} is 2.7 kcal mol^{-1} larger than the 41.3 kcal mol^{-1} AM1d/MM barrier determined for the reaction in Chapter 6. These results are difficult to compare since one value is related to a static system at 0 K, whilst the value from the MD simulations relate to a dynamic system at 298 K. However, by adding the value of -TΔS = 8.3 kcal mol^{-1} [89] to the AM1d/MM value for the enthalpy of the system gives an estimated free energy barrier of 49.6 kcal mol^{-1}, which is in reasonably close agreement to the values obtained from the AM1d/MD simulations.

7.4 Conclusions

Whilst a sampling time of 15 ps for each window is rather short to fully sample conformational space at each stage on the reaction coordinate, the simulations still represent a considerable improvement compared to simply minimising the system at each point on the reaction pathway. The use of MD simulations also provide an insight into the dynamic behaviour of the solvent during the reaction. This is shown by the RDFs generated at various stages of the reaction coordinate.

Finally, the reaction potential energy surface presented in this chapter has a very similar topology to the HF 3-21G*/MM and AM1d/MM potential energy surfaces presented in the preceding chapter. These results, obtained through the use of different methods and the employment of different potential for the QM region, support each other in the conclusion that, in solution, the base-catalysed hydrolysis of DMP proceeds via an associative mechanism which involves the formation of two transition states. In each case, the rate-limiting step for the reaction was concluded to relate to the cleavage of the bond between the phosphorus and the leaving methoxy group. The results also indicate that the stability of the phosphorane intermediate formed is questionable and that therefore it is unlikely to be of kinetic significance.

Chapter 8

Conclusions and future research prospects

8.1 Conclusions

The research presented in this thesis represents a comprehensive examination of a series of phosphate ester hydrolysis reactions. The results attained elucidate both the mechanism and energetics of these reactions. The two phosphate esters whose hydrolysis reaction are examined, DMP and methyl ribose phosphate, are respectively considered to be realistic models for DNA and RNA. Therefore, the results gained in this study are expected to provide significant insight into the hydrolysis of both DNA and RNA.

The base-catalysed hydrolysis of DMP

One reaction in particular , the base-catalysed hydrolysis of dimethyl phosphate, was the subject of four separate studies which thoroughly examined the manner in which the presence of aqueous solvent alters the reaction energy profile.

In the gas phase, the reaction was found to have an associative mechanism which involved the formation of a single dianionic transition state which corresponded to

the formation of the bond between the nucleophilic oxygen and the phosphorus. No evidence of a dianionic phosphorane intermediate was found. The early nature of the transition state indicates that, in the gas phase, the energy barrier to the approach of the nucleophile is greater than the energetic cost of cleaving the phosphodiester bond between the phosphorus and the leaving group. The use of an implicit solvent representation to modify to the gas phase reaction profile demonstrated that solvent dramatically reduced the free energy of activation for the reaction.

However, further examination of the role of solvent in the base-catalysed hydrolysis of DMP revealed that the presence of solvent not only affects the magnitude of the energy profile for the reaction but also its shape. The piece of evidence suggesting this was found when the geometry of the transition state was determined to be more reactant-like when it was allowed to relax under the influence of a continuum solvent representation.

The employment of an explicit solvent model, through the use of combined QM/MM potentials, to determine the reaction energy profile gave dramatically different results. The reaction mechanism was found to involve the formation of two transition states. In each of the three QM/MM studies (HF 3-21G*/MM, AM1d/MM, and AM1d/MD), the geometries of the two transition states were very similar and the first transition state was found to correspond to the formation of the bond between the attacking oxygen and the phosphorus and the second to the cleavage of the bond between the phosphorus and the leaving group. In each case, this latter step was found to be rate-determining thus indicating that, compared to the gas phase, the presence of solvent molecules provided relatively more stabilisation to the approach of the highly negatively charged nucleophile than to the cleavage of the phosphodiester bond to the leaving group. Each of the three QM/MM reaction profiles also provided evidence of the formation of a pentacoordinated intermediate during the reaction. In each case the intermediate was found to occupy a shallow potential energy well thus leading

to the conclusion that, although the intermediate probably does exist, it is probably neither kinetically significant nor experimentally detectable. This reaction has been he subjuect of many theoretical and experimental studies [89,168,171,180,187,195], each of which have elucidated features of the reaction. It is hoped that, when considered together, the results presented here, represent a definitive analysis of the reaction's chemistry.

Other mechanisms for the hydrolysis of DMP

Two other mechanisms for the hydrolysis of DMP were considered: the nucleophilic attack of DMP and protonated DMP by water. Both of these reactions can be termed as being acid-catalysed since the nucleophile transfers a proton to the phosphate as the bond between the water oxygen and the phosphorus is being formed. The gas phase reactions were found to proceed via the formation of two and one phosphorane intermediates respectively. Not surprisingly, the neutral phosphorane intermediate was considerably more stable than either of the two monoanionic intermediates. By comparing these results to those obtained for the gas phase base-catalysed hydrolysis of DMP, it can be concluded that the stability of phosphorane intermediates is directly proportional to the total charge of the species.

The hydrolysis of methyl ribose phosphate

Chapter 5 presents the results of a study of a gas phase study of the base-catalysed hydrolysis of methyl ribose phosphate. The hydrolysis of two thio-analogues of this phosphate were also studied. The combined results, which were in good agreement with existing experimental results, led to the conclusion that the cleavage of the P–O5′ bond is the rate-limiting step for the reaction.

General remarks and considerations

The use of density functional theory, continuum solvent representations, and QM/MM methods have combined well to further elucidate the features of phosphate ester hydrolysis reactions. Since many different levels of theory and approximations were employed, the results achieved are not quantitatively comparable but in many cases, good agreement with experimental results was achieved. With the exception of the AM1d potential, all of the quantum mechanical Hamiltonians and solvent representations used are well established and have already been shown to provide accurate results. Although some aspects of the AM1d-based results exposed the some of the problems usually associated with the use of semi-empirical potentials, this newly developed method was found to give a good representation the geometries and energies of the phosphorane species encountered on the reaction pathway and constitutes a powerful method which has considerable potential for probing the features of many phosphate ester hydrolysis reactions.

8.2 Prospects for future research

The role of metal ions

It has been shown in this thesis that the hydrolysis of phosphate esters such as DNA and RNA is accelerated by the presence of aqueous solvent which stabilises the approach of the nucleophile, the pentacoordinate species that results, and the departure of the leaving group.

Crystal structures of enzymes and ribozymes which catalyse the cleavage phosphodiester bonds, including HIV reverse transcriptase [258], DNA polymerase [259], and the hammerhead ribozyme [260], have revealed the presence of divalent metals ions in the active site. It is thought that the metal ions, which are found to be Mg^{2+} and Zn^{2+} in the majority of cases, reduce the energy barrier to the cleavage of

phosphodiester bonds by forming electrostatic interactions with the nucleophilic and leaving oxygens as well with the non-bridging oxygens [208]. It has also been shown that lanthanide ions such as Ce^{4+} and Zr^{4+} dramatically increase the hydrolysis of DNA in solution [180, 244].

Therefore, a natural progression to the research presented here is to employ the same theoretical methods to quantitatively determine the catalytic contribution of metal ions to the hydrolysis of phosphate esters. There have been *ab initio* studies which have probed the role of metal ions on phosphate ester hydrolysis reactions in the gas phase [239, 261]. However, to the best of my knowledge, there have been no studies which examine the combined roles of solvent and metal ions in reducing the energy barrier to phosphate ester hydrolysis. The use of a combined QM/MM potential would be ideally suited to such a study. The metal ion or ions could either treated classically or as part of the quantum region. The former approach would be based on the assumption that the role of the metals is purely electrostatic and would not account for electronic polarisation of the metal ions. The latter approach would be computationally more expensive but would be expected to yield more accurate results.

Solvent representations

The results presented in Chapters 6 and 7 demonstrate that an atomistic representation of the solvent gives an excellent insight into the role of solvent in chemical reactions. During the discussion of these results, a lot of emphasis was placed on the degree of accuracy afforded by the quantum mechanical potential used to represent the solute. However, the quality of the solvent model employed should also be the focus of attention.

In each of the QM/MM studies presented, the TIP3P solvent representation [78] was employed. While this model is known to provide an accurate description of the complex behaviour of liquid water [262], it does have some drawbacks. The rigid

geometry of each solvent molecule and the lack of charge transfer within and between solvent molecules are among those apparent. The issue of charge transfer within water molecules is especially important and various models have been formulated to address this [263, 264] within the context of QM/MM studies. The most promising of these to date is the Fluctuating Charge (FLUCQ) model [79] of Rick and Berne which is a modification of the TIP4P model that allows charge transfer within molecules according the principle of electronegativity equalisation. This model has been implemented into each of the CHARMM QM/MM modules [265] and the used of this solvent potential in conjunction with the QM/MM studies described in this thesis would allow a polarisable representation of the entire system. This would not only lead to a more realistic simulation of a reaction which involves a highly charged solute but would also make for an intriguing test case for the CHARMM FLUCQ module.

Simulation methods

The greatest single difficulty encountered in this thesis involved the simulation of explicit solvent molecules. Unlike in an enzyme simulation where many internal coordinates are restrained to narrow ranges of values, the conformational space accessible to aqueous solvent is massive. Most of these degrees of freedom have very shallow potential wells and therefore many poorly defined local minima can exist. This presents great difficulties to large scale geometry optimisations like those presented in Chapter 6 and also to a lesser extent molecular dynamics simulations. While MD simulations are inherently capable of sampling all conformation space, very long simulations would be required to adequately sample all possible solvent conformational space. Therefore, a simulation protocol that biases the sampling of conformational space would greatly reduce the CPU time required for simulations such as the one presented in Chapter 7 and also improve the accuracy of the thermodynamic parameters calculated.

The Gao research group has pioneered the use of Monte Carlo simulation methods in conjunction with QM/MM methods to simulate chemical reactions [99,131,136]. Monte Carlo simulations [62, 266] sample the conformational space of a system by making a series of biased 'moves' within the system. Every move that lowers the system energy is accepted and those which do not are accepted if the Boltzmann factor, $\exp(-\Delta V/k_B T)$, is greater than a random value between 0 and 1.

The main advantages of using Monte Carlo instead of MD simulations in conjunction with QM/MM potentials to simulate chemical reactions is that the sampling of some of the more important degrees of freedom in the system can be biased and also that dramatic changes can be made to the system rapidly thus enabling a more efficient sampling of conformational space. Therefore, I feel that studying the reactions studied in this thesis using QM/MC simulations would at least provide an interesting comparison between the advantages and disadvantages of the QM/MD and QM/MC protocols and would probably allow more of the vast conformational space available to the solvent to be sampled.

Lastly, perhaps the most obvious progression that can be made from this research is to simulate phosphate ester hydrolysis reactions within enzyme environments. The successful implementation of QM/MM methods to simulate enzymatic reactions is documented in Chapter 2. It is hoped that the results presented in this thesis would provide a very good point of reference to the results gained from such a study. Although there are many candidates for such a study, HIV reverse transcriptase, DNA polymerase β, and the hammerhead ribozyme are some of the more obvious choices due to their therapeutic importance and the availability of high resolution crystal structures.

Bibliography

[1] F.H. Westheimer, Why nature choses phosphates. Science **235**, 1173 (1987).

[2] G.R.J. Thatcher and R. Kluger, Mechanism and catalysis of nucleophilic substitution in phosphate esters. Adv. Phys. Org. Chem. **25**, 99 (1989), and references therein.

[3] D.M. Perreault and E.V. Anslyn, Unifying the current data on the mechanism of cleavage-transesterification of RNA. Angew. Chem. Int. Ed. Engl. **36**, 432 (1997).

[4] D.M. Zhou and K. Taira, The Hydrolysis of RNA: From theoretical calulations to the Hammerhead Ribozyme-mediated cleavage of RNA. Chem. Rev. **98**, 991 (1998).

[5] T.R. Cech, A.J. Zuag, and P.J. Gradowski, *In vitro* splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell **27**, 487 (1981).

[6] M. Oivanen, S. Kuusela, and H. Lönnberg, Kinetics and mechanisms for the cleavage and isomerization of the phophodiester bonds of RNA by Brønsted acids and bases. Chem. Rev. **98**, 961 (1998).

[7] M. Born and J.R. Oppenheimer, Zur Quantentheorie der Molekeln. Ann. Phys. **79**, 361 (1920).

[8] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).

[9] D.R. Hartree, Proc. Cambridge Philos. Soc. **23**, 328 (1928).

[10] V. A. Fock, Naherungsmethode zur Losung des quantenmechanischen Mehrkorperproblems. Z. Phys. **15**, 126 (1930).

[11] C.C.J. Roothann, New developments in molecular orbital theory. Rev. Mod. Phys. **23**, 69 (1951).

[12] W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).

[13] S.F. Boys, Proc. Roy. Soc. London A **200**, 542 (1950).

[14] W.J. Hehre, R. Ditchfield, and J.A. Pople, Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in MO Studies of Organic Molecules. J. Chem. Phys. **56**, 2257 (1972).

[15] P Hohenberg and W. Kohn, Inhomogeneous electron gas. Phys. Rev. B **136**, 864 (1964).

[16] R.G. Parr, Density Functional Theory. Annu. Rev. Phys. Chem. **34**, 631 (1983).

[17] R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules* (Oxford University Press, New York, 1989).

[18] E. Wimmer, in *Density functional methods in chemistry*, edited by J.R. Labanowski and J.W. Andzelm (Springer Verlag, Berlin, 1991), pp. 7–31.

[19] T. Ziegler, Approximate density functional theory as a practical tool in molecular energetics and dynamics. Chem. Rev. **91**, 651 (1991).

[20] W Kohn and L.J. Sham, Self-Consistent Equations including Exchange and Correlation Effects.. Phys. Rev. A **140**, 1133 (1965).

[21] Vosko S.J., Wilk L., and Nusair M., Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis.. Can. J. Phys. **58**, 1200 (1980).

[22] J.C. Slater, *International Series in Pure and Applied Physics, Vol. 4: The Self-Consistent Field for Molecules and Solids. Quantum Theory* (McGraw-Hill, New York, 1974).

[23] A.D. Becke, Density Functional Exchange-Energy Approximation with Correct Asymptotic Behaviour. Phys. Rev. A. **38**, 3098 (1988).

[24] C. Lee, W. Yang, and R. G. Parr, Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. Phys. Rev. B **37**, 785 (1988).

[25] A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. **98**, 5648 (1993).

[26] B.J. Lynch and D.G. Truhlar, How Well Can Hybrid Density Functional Methods Predict Transition State Geometries and Barrier Heights?. J. Phys. Chem. A **105**, 2936 (2001).

[27] J. J. P. Stewart, in *Reviews in Computational Chemistry*, edited by Kenny B. Lipkowitz and Donald B. Boyd (VCH, New York, 1990), Vol. 1, pp. 45–81.

[28] Michael C. Zerner, in *Reviews in Computational Chemistry*, edited by Kenny B. Lipkowitz and Donald B. Boyd (VCH, New York, 1994), Vol. 4, pp. 313–365.

[29] J.A. Pople and G.A. Segal, Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. J. Chem. Phys. **43**, 136 (1965).

[30] M.J.S. Dewar and W. Thiel, Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. J. Am. Chem. Soc. **99**, 4899 (1977).

[31] M. J. S. Dewar, E. G. Zoebisch, E. A. Healy, and J. J. P. Stewart, AM1: a new general purpose quantum mechanical molecular model. J. Am. Chem. Soc. **107**, 3902 (1985).

[32] J.J.P. Stewart, PM3 parameterization: test applications. J. Comp. Chem. **10**, 221 (1989).

[33] W. Thiel and A. A. Voityuk, Extension of the MNDO formalism to d Orbitals: Integral approx imations and preliminary numerical results.. Theor. Chim. Acta **81**, 391 (1992).

[34] W. Thiel and A. A. Voityuk, Extension of MNDO to d Orbitals: Parameters and Results for th e Halogens. Int. J. Quant. Chem. **44**, 807 (1992).

[35] K.B. Wiberg, A scheme for strain energy minimization. Application to the cycloalkanes. J. Am. Chem. Soc. **87**, 1070 (1965).

[36] J.E. Williams, P.J. Stang, and P von. R. Schleyer, Physical organic chemistry; quantitative conformational analysis; calculation methods. Annu. Rev. Phys. Chem. **19**, 531 (1968).

[37] R.H. Boyd, Method for calculation of the conformation of minimum potential-energy and thermodynamic functions of molecules from empirical valence-force potentials - application to the cyclophanes. J. Chem. Phys. **19**, 2574 (1968).

[38] C.L. Brooks, III and M. Karplus, Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site of lysozyme.. J. Mol. Biol. **208**, 159 (1989).

[39] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comp. Chem. **4**, 187 (1983).

[40] M.L. McKee and M. Page, in *Reviews in Computational Chemistry*, edited by K.B. Lipkowitz and D.B. Boyd (VCH Publishers, Inc., New York, 1993), Vol. 4, pp. 35–65.

[41] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[42] G. Schaftenaar and J.H. Noordik, Molden: a pre- and post-processing program for molecular and electronic structures. J. Comput.-Aided Mol. Design **14**, 123 (2000).

[43] B. Brooks and M. Karplus, Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. Proc. Natl. Acad. Sci. U.S.A. **82**, 4995 (1985), normal mode analysis.

[44] K. Fukui, The Path of Chemical Reactions - The IRC Approach. Acc. Chem. Res. **14**, 363 (1981).

[45] K. Ishida, K. Morokuma, and A. Komornicki, The intrinsic reaction coordinate. An *ab initio* calculation for HNC $\rightarrow$ HCN and $H^- + CH_4 \rightarrow CH_4 + H^-$. J. Chem. Phys. **66**, 2153 (1977).

[46] K. Müller and L.D. Brown, Location of saddle points and minimum energy paths by a constrained simplex optimisation procedure. Theo. Chem. Acta. **53**, 75 (1979).

[47] H.B. Gonzalez, C.and Schlegel, Reaction Path Following in Mass-Weighted Internal Coordinates. J. Phys. Chem. **94**, 5523 (1990).

[48] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *Gaussian 98 (Revision A.7)*, Gaussian, Inc., Pittsburgh PA, 1998.

[49] K.P. Eurenius, D.C. Chatfield, B.R. Brooks, and M. Hodoscek, Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. Int. J. Quantum Chem. **60**, 1189 (1996).

[50] R. Elber and M. Karplus, A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. Chem. Phys. Lett. **139**, 375 (1987).

[51] R. Czerminski and R. Elber, Self-Avoiding Walk Between Two Fixed Points as a Tool to Calculate Reaction Paths in Large Molecular Systems. Int. J. Quantum Chem.: Quantum Chem. Symp. **24**, 167 (1990).

[52] S. Fischer and M. Karplus, Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom. Chem. Phys. Lett. **194**, 252 (1992).

[53] O.S. Smart, A new method to calculate reaction paths for conformational transitions of large molecules. Chem. Phys. Lett. **222**, 503 (1994).

[54] R.S. Mulliken, Electronic Population Analysis on LCAO-MO Molecular Wave Functions. J. Chem. Phys. **23**, 1833 (1955), mulliken population analysis.

[55] C. M. Breneman and K. B. Wiberg, Determining atom-centred monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. J. Comp. Chem. **11**, 361 (1990).

[56] L.E. Chirlian and M.M. Francl, Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. J. Comp. Chem. **8**, 894 (1987).

[57] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, All-atom empirical potential for molecular modeling and dynamics Studies of proteins. J. Phys. Chem. B. **102**, 3586 (1998).

[58] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, An all atom force field for simulations of proteins and nucleic acids. Journal of Computational Chemistry **7**, 230 (1986).

[59] Walter R.P. Scott, P.H. Huenenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Krueger, and W.F. van Gunsteren, The GROMOS Biomolecular Simulation Program Package. J. Phys. Chem. A **103**, 3596 (1999).

[60] W. L. Jorgensen and J. Tirado-Rives, The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin.. J. Am. Chem. Soc. **110**, 1657 (1988).

[61] J. E. Jones, On the determination of molecular fields - II. From the equation of state of a gas. Proc. Roy. Soc. London **A106**, 463 (1924).

[62] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

[63] J.A. McCammon and S.C. Harvey, *Dynamics of proteins and nucleic acids* (Cambridge University Press, Cambridge, 1987).

[64] A. R. Leach, *Molecular Modelling. Principles and application* (Addison Wesley Longman Ltd., Harlow, 1996).

[65] L. Verlet, Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. **159**, 98 (1967).

[66] R. W. Hockney, The potential calculation and some applications. Methods in Computational Physics **9**, 136 (1970).

[67] W.C. Swope, H.C. Anderson, P.H. Berens, and K.R. Wilson, A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.. J. Chem. Phys. **76**, 637 (1982).

[68] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comp. Phys. **23**, 327 (1977).

[69] W.F. van Gunsteren and H.J.C Berendsen, Alogorithms for molecular dynamics and constraint dynamics. J. Comp. Phys. **34**, 1311 (1977).

[70] J.G. Kirkwood, Statistical mechanics of fluid mixtures. J. Chem. Phys. **3**, 300 (1935).

[71] P. Kollman, Free energy calculations: Applications to chemical and biochemical phenomena.. Chem. Rev. **93**, 2395 (1993).

[72] R.W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar Gases.. J. Chem. Phys. **22**, 1426 (1954).

[73] G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling. J. Comput. Phys. **23**, 187 (1977).

[74] J.P. Valleau and G.M. Torrie, in *Statistical Mechanics, Part A*, edited by B.J. Berne (Plenum Press, New York, 1977), pp. 169–94.

[75] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. J. Comp. Chem. **13**, 1011 (1992).

[76] B. Roux, The calculation of the potential of mean force using computer simulations. Comp. Phys. Comm. **91**, 275 (1995).

[77] H.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331–342.

[78] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. **79**, 926 (1983).

[79] S. W. Rick, S. J. Stuart, and B. J. Berne, Dynamical fluctuating charge force fields: Application to liquid water. J. Chem. Phys. **101**, 6141 (1994).

[80] M. Born, Volumen and hydratationswärme der ionen. Zeitschrift für Physik **1**, 45 (1920).

[81] B. Roux, H.-A. Yu, and M. Karplus, Molecular basis for the born model of ion solvation. J. Phys. Chem. **94**, 4683 (1990).

[82] L. Onsager, Electric moments of molecules liquids. J. Am. Chem. Soc. **58**, 1486 (1936).

[83] S. Miertuš, E. Scrocco, and J. Tomasi, Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem. Phys. **55**, 117 (1981).

[84] Roberto Cammi, Maurizio Cossi, Benedetta Mennucci, Christian Silvio Pomelli, and Jacopo Tomasi, Energy and energy derivatives for molecular solutes: perspectives of application to hybrid quantum and molecular methods. Int. J. Quant. Chem. **60**, 1165 (1996).

[85] K.B. Wiberg, T.A. Keith, M.J. Frisch, and M. Murcko, Solvent Effects on 1,2-Dihaloethane Gauche/Trans Ratios. J. Phys. Chem. **99**, 9072 (1995).

[86] A. Warshel and M. Levitt, Theoretical Studies of Enzymic Reactions: Stabilization of the Carbonium Ion in the Reaction of Lysozyme. J. Mol. Biol. **103**, 227 (1976).

[87] M. J. Field, P. A. Bash, and M. Karplus, A combined quantum mechanical and molecular potential for molecular dynamics simulations. J. Comp. Chem. **11**, 700 (1990).

[88] L. Stryer, *Biochemistry*, 4rd ed. (Freeman, New York, 1995).

[89] A. Radzicka and R. Wolfenden, A Proficient Enzyme. Science **267**, 90 (1995).

[90] A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman & Co., New York, 1999).

[91] C.R. Cantor and P.R. Schimmel, *Biophysical Chemistry. Part II: Techniques for the study of biological structure and function* (W. H. Freeman and Co., New York, 1980).

[92] K. Moffat and R. Henderson, Freeze Trapping of Reaction Intermediates. Curr. Opin. Struct. Biol. **5**, 656 (1995).

[93] M.J. Pilling and P.W. Seakins, *Reaction Kinetics* (Oxford University Press, Oxford, 1995).

[94] G.C. Schatz, Scattering Theory and Dynamics: Time-Dependent and Time-Independent Methods. J. Phys. Chem. **100**, 12839 (1996).

[95] M. Head-Gordon, Quantum Chemistry and Molecular Processes. J. Phys. Chem. **100**, 13213 (1996).

[96] U.C. Singh and P.A. Kollman, A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH_3Cl + Cl^- Exchange Reaction and Gas Phase Protonation of Polyethers. J. Comp. Chem. **7**, 718 (1986).

[97] P. D. Lyne, M. Hodoscek, and M. Karplus, A hybrid QM-MM potential employing Hartree-Fock or density funcitonal methods in the quantum region. J. Phys. Chem. A **103**, 3462 (1999).

[98] B.T. Thole and P.Th. van Duijnen, Reaction Field Effects on Proton Transfer in the Active Site of Actinidin. Biophys. Chem. **18**, 53 (1983).

[99] J. Gao and X. Xia, A Priori Evaluation of Aqueous Polarization Effects Through Monte Carlo QM-MM Simulations. Science **258**, 631 (1991).

[100] R. V. Stanton, L. R. Little, and K. M. Merz, An examination of a Hartree-Fock/Molecular Mechanical coupled potential. J. Phys. Chem. **99**, 17344 (1995).

[101] V. Thery, D. Rinaldi, J.-L. Rivail, B. Maigret, and G. G. Frenczy, Quantum mechanical computations on very large molecular systems: The local self-consistent field method.. J. Comp. Chem. **15**, 269 (1994).

[102] I. Tunon, M.T.C. Martins-Costa, C. Millot, M.F. Ruiz-Lopez, and J.L. Rivail, A coupled density functional-molecular mechanics monte carlo simulation method: the water molecule in liquid water.. J. Comp. Chem. **17**, 19 (1996).

[103] F. Maseras and K. Morokuma, A new "integrated ab initio + molecular mechanics" geometry optimization scheme of equilibrium structures and transition states.. J. Comp. Chem. **16**, 1170 (1995).

[104] M. A. Thompson, A QM/MM molecular dynamics study of the potential of mean force for the association of K^+ with dimethyl ether in aqueous solution. J. Am. Chem. Soc. **117**, 11341 (1995).

[105] M.J. Harrison, Burton N.A., and Hillier I.H., Catalytic Mechanism of the Enzyme Papain; Predictions with a Hybrid Quantum Mechanical/Molecular Mechanical Potential.. J. Am. Chem. Soc. **119**, 12285 (1997).

[106] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions.* (Wiley, New York, 1991).

[107] J. Gao, Absolute free energy of solvation from monte carlo simulations using combined quantum and molecular mechanical potentials.. J. Phys. Chem. **96**, 537 (1992).

[108] J. Gao, Monte Carlo quantum mechanical-configuration interaction and molecular mechanics simulation of solvent effects on the n-p* blue shift of acetone.. J. Am. Chem. Soc. **116**, 9324 (1994).

[109] H. Liu, F. Muller-Plathe, and W. F. van Gunsteren, A Molecular Dynamics Simulation Study with a Combined Quantum Mechanical and Molecular Mechanical Energy Function: Solvation Effects on the Conformational Equilibrium of Dimethoxyethane.. J. Chem. Phys. **102**, 1722 (1995).

[110] P.A. Bash, M.J. Field, and M. Karplus, Free energy perturbation method for chemical reactions in the condensed phase: a dynamical approach based on a combined quantum and molecular mechanical potential. J. Am. Chem. Soc. **109**, 8092 (1987).

[111] Jiali Gao, Combined QM/MM simulation study of the Claisen rearrangement of allyl vinyl ether in aqueous solution. Journal of the American Chemical Society **116**, 1563 (1994).

[112] P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe, and M. Karplus, Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. Biochemistry **30**, 5826 (1991).

[113] P.D. Lyne, A.J. Mulholland, and W.G. Richards, Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction. J. Am. Chem. Soc. **117**, 11345 (1995).

[114] A.J. Mulholland and W.G. Richards, Acetyl-CoA Enolization in Citrate Synthase: A Quantum Mechanical/Molecular Mechanical (QM/MM) Study. Proteins: Structure, Function and Genetics **27**, 9 (1997).

[115] M.A.; Cunningham, L.L. Ho, D.T. Nguyen, R.E. Gillilan, and P.A. Bash, Simulation of the Enzyme Reaction Mechanism of Malate Dehydrogenase.. Biochemistry **36**, 4800 (1997).

[116] C. Alhambra, L. Wu, Z.Y. Zhang, and J.L. Gao, Walden-inversion-enforced transition state stabilization in a protein tyrosine phosphatase. J. Am. Chem. Soc. **120**, 3858 (1998), recent QM/MM study in harmony with experiment.

[117] S. Antonczak, G. Monard, M.F. Ruiz-Lopez, and J.L Rivail, Modeling of Peptide Hydrolysis by Thermolysin. A Semiempirical and QM/MM Study. J. Am. Chem. Soc. **120**, 8825 (1998).

[118] P. Varnai, W.G. Richards, and P.D. Lyne, Modelling the catalytic reaction in human aldose reductase.. Proteins: Struct., Funct., Genet. **37**, 218 (1999).

[119] H. Liu, F. Muller-Plathe, and van Gunsteren W.F., A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease.. J. Mol. Biol. **261**, 454 (1996).

[120] S. Ranganathan and J.E. Gready, Hybrid Quantum and Molecular Mechanical (QM/MM) Studies on the Pyruvate to L-Lactate Interconversion in L-Lactate Dehydrogenase.. J. Phys. Chem. B **101**, 5614 (1997).

[121] D.C. Chatfield, K.P. Eurenius, and B.R. Brooks, HIV-1 protease cleavage mechanism: A theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential. Theochem. **423**, 79 (1998).

[122] J. Bentzien, R.P. Muller, J. Florian, and A. Warshel, Hybrid ab Initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymic Reactions: The Nucleophilic Attack in Subtilisin.. J. Phys. Chem. B **102**, 2293 (1998).

[123] A. H. Elcock, P. D. Lyne, A. J. Mulholland, A. Nandra, and W. G. Richards, Combined quantum and molecular mechanical study of DNA cross-linking by nitrous acid. J. Am. Chem. Soc. **117**, 4706 (1995).

[124] J.J.P. Stewart, MOPAC: A semi-empirical molecular orbital program. J. Computer-Aided Mol. Design **4**, 1 (1990).

[125] H.S. Rzepa and M. Yi, A quantitative molecular-orbital study of the structures and vibrational spectra of the hydrogen-bonded complexes $H_2O.NH_3$, $H_2CO.NH_3$ and $(H_2O)n$, n=2-4.. J. Chem. Soc., Perkin Trans. 2 943 (1990).

[126] M.W. Jurema and G.C. Shields, Ability of the PM3 quantum mechanical method to model intermolecular hydrogen bonding between neutral molecules. J. Comp. Chem. **14**, 89 (1993).

[127] I. Tunon, M. F. Ruiz-Lopez, D. Rinaldi, and J. Bertran, Computation of hydration free energies using a parameterized continuum model: study of equilibrium geometries and reactive processes in water solution. J. Comp. Chem. **17**, 148 (1995).

[128] Robert V. Stanton, David S. Hartsough, and Jr. Merz, Kenneth M., An examination of a density functional/molecular mechanical coupled potential. Journal of Computational Chemistry **16**, 113 (1995).

[129] D. Wei and D. R. Salahub, A combined density functional and molecular dynamics simulation of a quantum water molecule in aqueous solution.. Chem. Phys. Lett. **224**, 291 (1994).

[130] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, A QM/MM impementation of the self-consistent charge density functional tight binding (SCC-DFTB) method.. J. Phys. Chem. B **105**, 569 (2001).

[131] J. Gao, in *Reviews in Computational Chemistry*, edited by K.B. Lipkowitz and D. R. Boyd (VCH, New York, 1996), Vol. 7.

[132] J. Gao, Hybrid quantum and molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry. Acc. Chem. Res. **29**, 298 (1996).

[133] A. Warshel, Computer simulations of enzymatic reactions.. Curr. Opinion Struct. Biol. **2**, 230 (1992).

[134] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T Fox, J.W. Caldwell, and P.A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc. **118**, 2309 (1996).

[135] W.G. Scott, Biophysical and biochemical investigations of RNA catalysis in the hammerhead ribozmye. Q. Rev. Biophysics **32**, 241 (1999).

[136] M. Freindorf and J. Gao, Optimization of the Lennard-Jones parameters for a combined *ab initio* quantum mechanical and molecular mechanical potential using the 3-21G basis set.. J. Comput. Chem. **17**, 386 (1996).

[137] P.A. Bash, L.L. Ho, A.D. Mackerell, D. Levine, and P. Hallston, Progress toward chemical accuracy in the computer simulation of condensed phase reactions.. Proc. Nat. Acad. Sci. **93**, 3698 (1996).

[138] G. Monard, M. Loos, V. Thery, K. Baka, and J.-L. Rivail, Hybrid classical quantum force field for modelling very large molecules.. Int. J. Quant. Chem. **58**, 153 (1995).

[139] J.L. Gao, P. Amara, C. Alhambra, and M.J. Field, A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. J. Phys. Chem. A **102**, 4714 (1998).

[140] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, Frontier bonds in QM/MM methods: A comparison of different approaches. J. Phys. Chem. A **104**, 1720 (2000).

[141] J.R. Knowles, To build an enzyme.... Phil. Trans. Roy. Soc. Lond. B **332**, 115 (1991).

[142] R.C. Davenport, P.A. Bash, B.A. Seaton, M. Karplus, G.A. Petsko, and D. Ringe, Structure of the Triosephosphate Isomerase-Phosphoglycolohydroxamate Complex: an Analogue of the Intermediate on the Reaction Pathway. Biochemistry **30**, 5821 (1991).

[143] C. L. Brooks, III, M. Karplus, and B. M. Pettitt, *Proteins* (John Wiley & Sons, New York, 1988).

[144] P. Lodi and J.R. Knowles, Neutral Imidazole is the Electrophile in the Reaction Catalyzed by Triosephosphate Isomerase: Structural Origins and Catalytic Implications. Biochemistry **30**, 6948 (1991).

[145] P.J. Lodi, L.C. Chang, J.R. Knowles, and E.A. Komives, Triosephosphate Isomerase Requires a Positively Charged Active Site: The Role of Lysine-12. Biochemistry **33**, 2809 (1994).

[146] J.A. Stuckey, HL Schubert, E.B. Fauman, Z.Y. Zhang, and Saper M.A. Dixon J.E., Crystal structure of yersinia protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. Nature **370**, 571 (1994).

[147] D. Barford, A.J. Flint, and N.K. Tonks, Crystal structure of human protein tyrosine phosphatase 1B. Science **263**, 1397 (1994).

[148] A.J. Mulholland, P.D. Lyne, and M. Karplus, Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved. J. Am. Chem. Soc. **122**, 534 (2000).

[149] G. Pettersson, U. Lill, and H. Eggerer, Mechanism of interaction of citrate synthase with citryl-CoA. Eur. J. Biochem. **182**, 119 (1989), citryl-CoA hydrolysis by pig enzyme, unlike physiological reaction, has complicated kinetics. Initial rate is consistent with citryl CoA being an intermediate. Account for strange kinetics by formation of a non-producitve complex slowly.

[150] M. Karpusas, B. Branchaud, and S.J. Remington, Proposed Mechanism for the Condensation Reaction of Citrate Synthase: 1.9 Structure of the Ternary Complex with Oxaloacetate and Carboxymethyl Coenzyme A. Biochemistry **29**, 2213 (1990).

[151] J.A. Gerlt and P.G. Gassman, Understanding the Rates of Certain Enzyme-Catalyzed Reactions: Proton Abstraction from Carbon Acids, Acyl-Transfer Reactions, and Displacement Reactions of Phosphodiesters. Biochemistry **32**, 11943 (1993).

[152] W. W. Cleland and M. M. Kreevoy, Low-barrier hydrogen bonds and enzymic catalysis.. Science **264**, 1887 (1994).

[153] A. Warshel, A Papazyan, and P.A. Kollman, On low-barrier hydrogen bonds and enzyme catalysis. Science **269**, 102 (1995).

[154] J.A. Gerlt and P.G. Gassman, An Explanation of Rapid Enzyme-Catalyzed Proton Abstraction from Carbon Acids: Importance of Late Transition States in Concerted Mechanisms. J. Am. Chem. Soc. **115**, 11552 (1993).

[155] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, and Jr. Montgomery, J.A., General Atomic and Molecular Electronic Structure System. J. Comp. Chem. **14**, 1347 (1993).

[156] R.L.P. Adams, J.T. Knowler, and D.P. Leader, *The Biochemistry of the Nucleic Acids*, 11^{th} ed. (Chapman and Hall, London, 1992).

[157] *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World.*, edited by R.F. Gesteland and J.F. Atkins (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993).

[158] M. Komiyama, N. Takeda, and H. Shigekawa, Hydolysis of DNA and RNA by lanthanide ions: mechanistic studies leading to new applications.. Chem. Commun. 1443 (1999).

[159] R. H. Symons, Small Catalytic RNAs. Annu. Rev. Biochem. **61**, 641 (1992).

[160] R.R. Breaker and G.F. Joyce, A DNA enzyme that cleaves RNA. Chem. Biol. **1**, 223 (1994).

[161] R.R. Breaker and G.F. Joyce, A DNA enzyme with Mg^{2+}-dependent RNA phosphoesterase activity. Chem., Biol. **2**, 655 (1995).

[162] B. Cuenoud and J.W. Szostak, A DNA metalloenzyme with DNA ligase activity. Nature **375**, 611 (1995).

[163] D.G. Nicholls and S.J. Ferguson, *Bioenergetics 2* (Academic Press, London, 1992).

[164] M. Barbacid, RAS genes. Annu. Rev. Biochem. **56**, 779 (1987).

[165] J. R. Cox Jr and B. Ramsay, Mechanisms of nucleophilic substitution in phosphate esters. Chem. Rev. **64**, 317 (1964).

[166] F. H. Westheimer, Pseudo-rotation in the hydolysis of phosphate esters. Acc. Chem. Res. **1**, 70 (1968).

[167] J. Florián and A. Warshel, Phosphate Ester Hydrolysis in Aqueous Solution: Associative versus Dissociative Mechanisms. J. Phys. Chem. B **102**, 719 (1998).

[168] A. Dejaegere, X. Liang, and M. Karplus, Phosphate ester hydrolysis: calculation of gas-phase reaction paths and solvation effects. J. Chem. Soc. Faraday Trans. **90**, 1763 (1994).

[169] R.R. Holmes, *Pentacoordinated Phosphorous*, Vol. 1 of *ACS Monograph 176* (American Chemical Society, Washington D.C., 1980).

[170] R.R. Holmes, *Pentacoordinated Phosphorous*, Vol. 2 of *ACS Monograph 176* (American Chemical Society, Washington D.C., 1980).

[171] D.G. Gorenstein, J.B. Findlay, B.A. Luxon, and D. Kar, J. Am. Chem. Soc. **99**, 3473 (1977), stereoelectronic control in C-O and P-O bond-breaking reactions. MO calculations.

[172] F.H. Westheimer, Monomeric metaphosphates. Chem. Rev. **81**, 313 (1981).

[173] W.P. Jencks, When is an intermediate not an intermediate? Enforced mechanisms of general acid-base catalysed, carbocation, carbanion, and ligand exchange reactions. Acc. Chem. Res. **13**, 161 (1980).

[174] A. Todd, Phosphate Chemistry. Proc. Nat. Acad. Sci. **45**, 1389 (1959).

[175] J. Florián and A. Warshel, On the Reactivity of Phosphate Monoester Dianions in Aqueous Solution: Bronsted Linear Free-Energy Relationships Do Not Have an Unique Mechanistic Interpretation. J. Am. Chem. Soc. **120**, 11524 (1998).

[176] J. Åqvist, K. Kolmodin, J. Florián, and A. Warshel, Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experimental data?. Chem. Biol. **6**, R71 (1999).

[177] K.A. Maegley, S.J. Admiraal, and D. Herschlag, Ras-catalysed hydrolysis of GTP: A new perspective from model studies. Proc. Natl. Acad. Sci. **93**, 8160 (1996).

[178] S.J. Benkovic and K.J. Schray, in *The Enzymes*, edited by P.D. Boyer (Academic Press, New York, 1973), p. 201.

[179] S.J. Admiraal and D. Herschlag, Mapping the transition state for ATP hydrolysis: implications for enzymatic catalysis. Chem. Biol. **2**, 729 (1995).

[180] N. Takeda, M. Shibata, N. Tajima, K. Hirao, and Komiyama M., Kinetic and theoretical studies on the mechanism of alkaline hydrolysis of DNA. J. Org. Chem. **65**, 4391 (2000).

[181] R.S. Berry, Correlation of rates of intramolecular tunneling processes, with application to some Group V compounds.. J. Chem. Phys. **32**, 933 (1960).

[182] K. Taira, T. Uchimaru, J.W. Storer, A. Yliniemela, M. Uebayasi, and K. Tanabe, Properties of Dianionic Oxyphosphorane Intermediates: Implication to the Reaction Profile for Base-Catalyzed RNA Hydrolysis. J. Org. Chem. **58**, 3009 (1993).

[183] K. Taira, T. Uchimaru, K. Tanabe, M. Uebayasi, and S. Nishikawa, Rate limiting P-O(5') bond cleavage of RNA fragment: *ab initio* molecular orbital calculations on the base-catalyzed hydrolysis of phosphate. Nucleic Acids Res. **19**, 2747 (1991).

[184] T. Uchimaru, K. Tanabe, S. Nishikawa, and K. Taira, Ab Initio Studies of a Marginally Stable Intermediate in the Base-Catalysed Methanolysis of Dimethyl Phosphate and Nonexistence of the Stereoelectronically Unfavourable Transition State. J. Am. Chem. Soc. **113**, 4351 (1991).

[185] J.W. Storer, T. Uchimaru, K. Tanabe, M. Uebayasi, S. Nishikawa, and K. Taira, Existence of a Marginally Stable Intermediate during the Base-Catalysed methanolysis of Methylene Phosphate and *ab initio* Studies of the Monohydration of the Pentacoordinated Oxyphosphorane Intermediate. J. Am. Chem. Soc. **113**, 5216 (1991).

[186] A. Yelimiena, T. Uchimaru, K. Tanabe, and K. Taira, Do Pentacoordinate Oxyphosphorane Intermediates Always Exist?. J. Am. Chem. Soc. **115**, 3032 (1993).

[187] A. Dejaegere, C. Lim, and M. Karplus, Dianionic Pentacoordinate Species in the Base-Catalysed Hydrolysis of Ethylene and Dimethyl Phosphate. J. Am. Chem. Soc. **113**, 4353 (1991).

[188] D.G. Gorenstein, B.A. Luxon, and J.B. Findlay, Stereoelectronic effects in the reactions of phosphate diesters. *Ab initio* molecular orbital calculations of reaction surfaces. 2.. J. Am. Chem. Soc. **101**, 5869 (1979).

[189] C Lim and P Tole, Concerted Hyroxyl Ion Attack and Pseudorotation in the Base-Catalysed Hydrolysis of Methyl Ethylene Phosphate. J. Phys. Chem. **96**, 5217 (1992).

[190] C. Lim and P Tole, Endocyclic and exocyclic cleavage of phosphorane monoanion: a detailed mechanism of the RNase A transphosphorylation step. J. Am. Chem. Soc. **114**, 7245 (1992).

[191] P Tole and C Lim, New Insights into the Base-Catalysed Hydrolysis of Methyl Ethylene Phosphate. J. Phys. Chem. **97**, 6212 (1993).

[192] R. Kluger, F. Covitz, E. Dennis, L. D. Williams, and F. H. Westheimer, pH-Rate profile for the hydrolysis of some esters of a bicyclic phosphinic acid. Evidence for rate-limiting pseudorotation.. J. Am. Chem. Soc. **91**, 6066 (1969).

[193] R. Kluger and S. D. Taylor, On the origins of enhanced reactivity of five-membered cyclic phosphate esters. The relative contributions of enthalpic and entropic factors.. J. Am. Chem. Soc. **112**, 6669 (1990).

[194] C. Lim and M. Karplus, Nonexistance of Dianionic Pentacovalent intermediates in an *ab initio* study of the base-catalysed hydrolysis of ethylene phosphate. J. Am. Chem. Soc. **112**, 5872 (1990).

[195] P.C. Haake and F.H Westheimer, Hydrolysis and exchange in esters of phosphoric acid. J. Am. Chem. Soc. **83**, 1102 (1961).

[196] K. Matsumoto and M. Komiyama, DNA hydrolysis by rare-earth metal ions. Nucleic Acids Symp. Ser. **27**, 33 (1992).

[197] M. Komiyama, K. Matsumura, and Y. Matsumoto, Unprecedentedly fast hydrolysis of the RNA dinucleoside monophosphates ApA and UpU by rare earth metal ions. J. Chem. Soc., Chem. Commun. **8**, 640 (1992).

[198] M. Komiyama, N. Takeda, Y. Takahashi, H. Uchida, T. Shiiba, T. Kodama, and M. Yashiro, J. Chem. Soc., Perkin Trans, 2 269 (1995).

[199] J.P. Guthrie, Hydration and dehydration of phosphoric acid derivatives: free energies of formation of the pentacoordinate intermediates for phosphate ester hydrolysis and of monomeric metaphosphate. J. Am. Chem. Soc. **99**, 3991 (1977).

[200] D.G. Gorenstein, B.A. Luxon, and J.B. Findlay, Stereoelectronic effects in the reactions of phosphate diesters. *Ab initio* molecular orbital calculations of reaction profiles. J. Am. Chem. Soc. **99**, 8048 (1977).

[201] D.G. Gorenstein, Stereoelectronic Effects in Biomolecules. Chem. Rev. **87**, 1047 (1987).

[202] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, and S. Altman, The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell **35**, 849 (1983).

[203] T.R. Cech, Angew. Chem., Int. Ed. Engl. **29**, 759 (1989).

[204] C. Carola and F. Eckstein, Nucleic acid enzymes. Curr. Opin. Chem. Biol. **3**, 274 (1999).

[205] R.G. Kuimelis and L.W. McLaughlin, Mechamisms of Ribozyme-mediated cleavage. Chem. Rev. **98**, 1027 (1998).

[206] S.C. Dahm, W.B. Derrick, and O.C. Uhlenbeck, Evidence for the Role of Solvated Metal Hydroxide in the Hammerhead Cleavage Mechanism. Biochemistry **32**, 13040 (1993).

[207] A.M. Pyle, Ribozymes - A distinct class of metalloenzymes. Science **261**, 709 (1993).

[208] T.A. Steitz and J.A. Steitz, A General Two-Metal-Ion Mechanism for Catalytic RNA. Proc. Natl. Acad. Sci. **90**, 6498 (1993).

[209] S. Nesbitt, L.A. Hegg, and M.J. Fedor, An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme cleavage. Chem. Biol. **4**, 619 (1997).

[210] A.A. Seyhan and J.M. Burke, Mg^{2+}-independent hairpin ribozyme catalysis on hydrated RNA films. RNA **6**, 189 (2000).

[211] N. K. Vaish, A.R. Kore, and F. Eckstein, Recent developments in the hammerhead ribozyme field. Nucl. Acids Res. **18**, 1971 (1998).

[212] K.J. Scanlon, *Therapeutic applications of ribozymes, methods in molecular medicine* (Humana Press, New Jersey, 1997), Vol. 11.

[213] R. Cosstick and J.S. Vyle, Synthesis and phosphorus-sulfur bond cleavage of 3'-thiothymidylyl(3'-5')thymidine. J. Chem Soc., Chem. Comm. 992 (1988).

[214] R.G. Kuimelis and L.W. Mclaughlin, Ribozyme-mediated cleavage of a substrate analogue containing an internucleotide-bridging 5'-phosphorothioate: evidence for the single-metal model. Biochemistry **35**, 5308 (1996).

[215] H. Almer and R. Stromberg, Base Catalysis and Leaving Group Dependence in Intramolecular Alcoholysis of Uridine 3'-(Aryl phosphorothioate)s. J. Am. Chem. Soc. **118**, 7921 (1996).

[216] D.-M. Zhou, N. Usman, F.E. Wincott, J. Matulic-Adamic, M. Orita, L.-H. Zhang, M. Komiyama, P.K.R. Kumar, and K. Taira, Evidence for the rate-limiting departure of the 5'-oxygen in nonenzymatic and hammerhead ribozyme catalysed reactions. J. Am. Chem. Soc. **118**, 5862 (1996).

[217] X. Liu and C.B. Reese, Uridylyl-(3′-5′)-(5′-thiouridine) – an exceptionally base-labile di-ribodinucleoside phosphate analog. Tetrahedron Lett. **36**, 3413 (1995).

[218] J.B. Thompson, K.B. Patel, V. Jiménez, K. Eckart, and F. Eckstein, Synthesis and properties of diuridine phosphate analogues containing thio and amino modifications. J. Org. Chem. **36**, 6273 (1996).

[219] M. Komiyama and M.L. Bender, The cyclodextrin-accelerated cleavage of thiocarboxylic *S*-esters. Bull. Chem. Soc. Jpn **53**, 1073 (1980).

[220] T.C. Bruice, T.H. Fife, J.J. Bruno, and N.E. Brandon, Hydroxyl group catalysis. II. The reactivity of the hydroxyl group of serine. The nucleophilicity of alcohols and the ease of hydrolysis of their acetyl esters as related to their pKa. Biochemistry **1**, 7 (1962).

[221] Z. Shaked, R.P. Szajewski, and G.M. Whitesides, Biochemistry **19**, 4156 (1980).

[222] S.C. Dahm and O.C. Uhlenbeck, Role of Divalent Metal Ions in the Hammerhead RNA Cleavage Reaction. Biochemistry **30**, 9464 (1991).

[223] H. van Tol, J.M. Buzayan, P.A. Feldstein, F. Eckstein, and G. Bruening, Two Autocatalytic Processing Reactions of Satellite RNA Proceed with Inversion of Configuration. Nucleic Acids Res. **18**, 1971 (1990).

[224] E.C. Scott and O.C. Uhlenbeck, A re-investigation of the thio effect at the hammerhead cleavage site. Nucleic Acids Res. **27**, 479 (1999).

[225] G. Slim and M.J. Gait, Configurationally defined phosphorthiolate-containing oligoribonucleotides in the study of the mechanism of cleavage of hammerhead ribozymes. Nucleic Acids Res. **19**, 1183 (1991).

[226] D.-M. Zhou, K. Penmetch, P.K.R. Kumar, L.-H. Zhang, and K. Taira, Ribozyme mechanism revisited: evidence against direct coordination of a Mg^{2+} ion with the *pro*-R oxygen of the scissile phosphate in the transition state of a hammerhead ribozyme-catalysed reaction. J. Am. Chem. Soc. **118**, 8969 (1996).

[227] W.B. Derrick, C.H. Greef, M.H. Caruthers, and O.C. Uhlenbeck, Hammerhead cleavage of the phosphorodithiolate linkage. Biochemistry **39**, 4947 (2000).

[228] P.M.J. Burgers and F. Eckstein, Diastereomers of 5′-O-adenosyl 3′-O-uridyl phosphorothioate: chemical synthesis and enzymic properties. Biochemistry **18**, 592 (1979).

[229] D. Herschlag, J.A. Piccirilli, and T.R. Cech, Ribozyme-catalyzed and nonenzymatic reactions of phosphate diesters: rate effects upon substitution of sulfur for a nonbridging phosphoryl oxygen atom. Biochemistry **30**, 4844 (1991).

[230] H. Almer and R. Stromberg, Intramolecular transesterification in thiophosphate-analogues of an RNA-dimer. Tetrahedron Lett. **32**, 3723 (1996).

[231] R Wolfenden, C Ridgway, and G Young, Spontaneous Hydrolysis of Ionized Phosphate Monoesters and Diesters and the Proficiencies of Phosphates and Phosphodiesterases as Catalysts. J. Am. Chem. Soc. **120**, 833 (1998).

[232] J. Florián and A. Warshel, A fundamental assumption about OH^- attack in phosphate hydrolysis not fully justified. J. Am. Chem. Soc. **119**, 5473 (1997).

[233] M.J. Frisch, J.E. Delbene, J.S. Binkley, and H.F. Schaefer, Extensive Theoretical studies of the hydrogen-bonded complexes $(H_2O)_2$, $(H_2O)_2H^+$, $(HF)_2$, $(HF)_2H^+$, F_2H^-, and $(NH_3)_2$. J. Chem. Phys. **84**, 2279 (1986).

[234] J. Simons and K.D. Jordan, *Ab initio* electonic structure of anions. Chem. Rev. **87**, 557 (1987).

[235] J.M. Martell, J.D. Goddard, and L.A. Eriksson, Assessment of basis set and functional dependancies in Density Functional theory: Studies of Atomization and Reaction Energies. J. Phys. Chem. A **101**, 1927 (1997).

[236] E. Goldstein, B. Beno, and K.N. Houk, Density Functional Theory Prediction of the relative energies and isotope effects for the concerted and stepwise mechanisms of the Diels-Alder reaction of Butadiene and Ethylene.. J. Am. Chem. Soc. **118**, 6036 (1996).

[237] M.W. Wong, Vibrational frequency prediction using density functional theory. Chem. Phys. Lett. **256**, 391 (1996).

[238] J. Tomasi and M. Persico, Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. Chem. Rev. **94**, 2027 (1994).

[239] M. Uebayasi, T. Uchimaru, T. Koguma, S. Sawata, T. Shimayama, and K. Taira, Theoretical and Experimental Considerations on the Hammerhead Ribozyme Reactions: Divalent Magnesium Ion Mediated Cleavage of Phosphorus-Oxygen Bonds. J. Org. Chem. **59**, 7414 (1994).

[240] Y. Takagi, M. Warashina, W.J. Stec, K. Yoshinari, and K. Taira, Recent advances in the elucidation of the mechanisms of action of ribozymes. Nucleic Acids Res. **29**, 1815 (2001).

[241] S. Sawata, M. Komiyama, and K. Taira, Kinetic Evidence Based on Solvent Isotope Effects for the Nonexistence of a Proton-Transfer Process in Reactions Catalysed by a Hammerhead Ribozyme: Implication to the Double-Metal-Ion mechanism of Catalysis. J. Am. Chem. Soc. **117**, 2357 (1995).

[242] R. Breslow and W.H. Chapman Jr., On the mechanism of action of ribonuclease A: relevance of enzymatic studies with a p-nitrophenylphosphate ester and a thiophosphate ester. Proc. Natl. Acad. Sci. USA **93**, 10018 (1996).

[243] S. Portmann, S. Grimm, C. Workman, N. Usman, and M. Egli, Crystal structures of an A-form duplex with single-adenosine bulges and a conformational basis for site-specific RNA self-cleavage. Chemistry and Biology **3**, 173 (1996).

[244] T. Shiiba and M. Komiyama, Phenyl ester of adenosine 3′-phosphate as a novel probe for the rate-limiting step in RNA hydrolysis. Tetrahedron Lett. **33**, 5571 (1992).

[245] F.A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry* (Wiley, New York, 1990).

[246] R.F. Hudson, *Structure and mechanisms in organo-phosphorous chemistry* (Academic Press, London, 1965).

[247] R.G. Kuimelis and L.W. McLaughlin, Hammerhead Ribozyme-Mediated Cleavage of a Substrate Analog Containing an Internucleotidic Bridging 5′-Phosphorothioate: Implications for the Cleavage Mechanism and the Catalytic Role of the Metal Cofactor. J. Am. Chem. Soc **117**, 11019 (1995).

[248] M. Oivanen, M. Ora, H. Almer, R. Stromberg, and H. Lönnberg, Hydrolytic Reactions of the Diastereomeric Phosphoromonothioate Analogs of Uridylyl(3′,5′)uridine: Kinetics and Mechanisms for Desulfurization, Phosphoester Hydrolysis, and Transesterification to the 2′,5′-Isomers. J. Org. Chem. **60**, 5620 (1995).

[249] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B **102**, 3586 (1998).

[250] C. L. Brooks III and M. Karplus, Deformable stochastic boundaries in molecular dynamics.. J. Chem. Phys. **79**, 6312 (1983).

[251] X. Lopez and D.M. York (unpublished).

[252] M.F. Guest, J.H. van Lenthe, J. Kendrick, P. Sherwood, R.D. Amos, R.J. Buenker, H. van Dam, M. Dupuis, N.C. Handy, I.H. Hillier, P.J. Knowles, V. Bonacic-Koutecky, W. von Niessen, R.J. Harrison, A.P. Rendell, V.R. Saunders, K. Schoffel, A.J. Stone, and D. Tozer, GAMESS-UK, a package of *ab inito programs.*

[253] T. Sterling, D. Becker, and D. Savarese, Beowulf: A parallel workstation for scientic computation, Proceedings of the International Conference on Parallel Processing, 1995.

[254] J.L. Durant, Evaluation of transition state properties by density functional theory. Chem. Phys. Lett. **256**, 595 (1996).

[255] J.M. Buchwald, S.L. Friedman and J.R. Knowles, Stereochemistry of the nucleophilic displacement on 2 phosphoric monoesters and a phosphoguanidine - the role of meta-phosphate.. J. Am. Chem. Soc. **106**, 4911 (1984).

[256] A.J. Kirby and A.G. Varvoglis, J. Am. Chem. Soc. **89**, 415 (1967).

[257] P.J. Steinbach and B.R. Brooks, New spherical-cutoff methods for long-range forces in macromolecular simulations. J. Comp. Chem. **15**, 667 (1994).

[258] H. Huang, R. Chopra, G.L. Verdine, and S.C. Harrison, Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science **282**, 1669 (1998).

[259] M.R. Sawaya, R. Prasad, S.H. Wilson, J. Kraut, and H. Pelletier, Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: Evidence for an induced fit mechanism. Biochemistry **36**, 11205 (1997).

[260] W.G. Scott, J.B. Murray, J.R.P. Arnold, B.L. Stoddard, and A. Klug, Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme.. Science **274**, 2065 (1996).

[261] T. Uchimaru, M. Uebayasi, T. Uchimaru, K. Tanabe, and K. Taira, Theoretical analyses on the role of magnesium ions in ribozyme reactions. FASEB J. **7**, 137 (1993).

[262] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Comparison of simple potential functions for simulating liquid water. J. Phys. Chem. **79**, 926 (1990).

[263] R. A. Bryce, R. Buesnel, I. H. Hillier, and N. A. Burton, A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. Chem. Phys. Lett. **279**, 367 (1997).

[264] M.A Thompson, QM/MM_{pol}: A Consistent Model for Solute/Solvent Polarization. Application to the Aqueous Solvation and Spectroscopy of Formaldehyde, Acetaldehyde, and Acetone. J. Phys. Chem. **100**, 14492 (1996).

[265] B.M. Webb, A fluctuating charge (FLUCQ) module for CHARMM, Manuscript in preparation, 2001.

[266] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087 (1953).