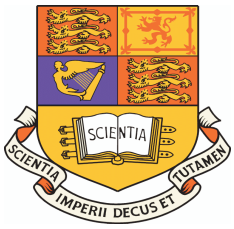
# The Development of a QM/MM Based Linear Response Method and its Application to Proteins

**Ross Walker**




**Thesis submitted in partial fulfilment**
**of the degree of Doctor of Philosophy**
**&**
**the Diploma of Imperial College**

**Imperial College London**
**Department of Chemistry**
**October 2003**

# Abstract

The equilibrium dynamics of a protein are of fundamental importance in understanding the relationship between its function and structure. Electronic structure calculations employing the CIS method for calculating excited states have been coupled with MD simulations to track the fluctuations in the ground to excited state energy of LADH as a function of time. Through the use of a QM/MM framework for predicting optical spectra based on linear response theory the UV/VIS absorption and emission spectra for NADH in LADH have been calculated yielding a Stokes shift prediction that is within 13 % of experiment and widths that are within 1.2 %. The Stokes shift would appear to be produced from intramolecular interactions on a timescale of around 6 to 8 fs. The spectral widths are sensitive to the amplitude and time scale of the protein fluctuations, while the Stokes shift is a direct measure of the extent of relaxation of the molecular system. Accurate reproduction of these properties acts as an indicator of the performance of the methodology for probing equilibrium fluctuations and the large Stokes shift of NADH tests the limit of the linear response assumption. The quality of the results very much depend on the quality of the MD trajectory used and as such the method provides a way to validate MD force field parameters for chromophores via direct comparison with experiment.

The results demonstrate that an extension of the QM/MM approach to the calculation of free energy surfaces is a viable objective even for very large and very fast fluctuations in systems as complex as proteins.

Simulations using semi-empirical and TD-DFT evaluation of the energy gap fluctuations have cast doubt on the ability of such methods to accurately evaluate the energy gap fluctuations of equilibrium structures.

Extension of the method to zinc-myoglobin has uncovered the existence of correlation between two degenerate excitations. Work is currently in progress to use these results to predict anisotropy experiments.

*For my parents*

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

Paul Dirac, 1929

# Acknowledgements

Thanks go to: Ian Gould, for his help, enthusiasm and computers; David Klug, for lively and interactive discussion; Sam Hughes for her help with AMBER and for keeping me sane; Halima Amer, Byung Moon Cho and Salinthip Thipayang for their help, friendship and for providing me endless opportunities to hone my computer support skills; Martin Burke and Daniel Ormsby for friendship and proof reading; Gabriel Asseily for timely escapades, welcome diversions and for being a thoroughly nice chap; Laura Campbell for just being there; and my parents, for everything else.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| 3PEP | 3-Pulse Echo Peak Shift |
| AO | Atomic Orbital |
| B3LYP | Becke 3-parameter functional; Lee, Yang, Parr, exchange |
| BSSE | Basis Set Superposition Error |
| CGF | Contracted Gaussian Function |
| CI | Configuration Interaction |
| CID | Configuration Interaction Doubles |
| CIS | Configuration Interaction Singles |
| CISD | Configuration Interaction Singles and Doubles |
| CNDO | Complete Neglect of Differential Overlap |
| CSF | Configuration State Function |
| CXF | *n*-Cyclohexyl Formamide |
| DFT | Density Functional Theory |
| DZ | Double Zeta Basis Set |
| FFTs | Fast Fourier Transforms |
| GTO | Gaussian-Type Orbital |
| HF | Hartree-Fock |
| HOMO | Highest Occupied Molecular Orbital |
| INDO | Intermediate Neglect of Differential Overlap |
| LADH | Liver Alcohol Dehydrogenase |
| LCAO | Linear Combination of Atomic Orbitals |
| LUMO | Lowest Unoccupied Molecular Orbital |

| | |
|---|---|
| MD | Classical Molecular Dynamics |
| MO | Molecular Orbital |
| NADH | di-Hydro Nicotinamide Adenine Dinucleotide |
| NICH | di-Hydro Nicotinamide |
| PME | Particle Mesh Ewald |
| QM/MM | Quantum Mechanical/Molecular Mechanical |
| QM/MM MD | QM/MM driven Molecular Dynamics |
| RESP | Restrained Electrostatic Potential |
| RMSD | Root Mean Square Deviation (distance) |
| RPA | Random Phase Approximation (eqv. TD-HF) |
| SCF | Self-Consistent Field |
| SIMD | Single Instruction Multiple Data |
| SSE | Streaming SIMD Extensions |
| STO | Slater-Type Orbital |
| SV | Split Valence Basis Set |
| TD-B3LYP | Time Dependent DFT using the B3LYP functional |
| TD-DFT | Time Dependent DFT |
| TD-HF | Time Dependent Hartree Fock |
| TDM | Transition Dipole Moment |
| TZ | Triple Zeta Basis Set |
| ZDO | Zero-Differential Overlap |
| ZINDO | Zerner's Intermediate Neglect of Differential Overlap |
| Zn-Mb | Zinc Substituted Myoglobin |

# Chapter 1

# Introduction

## 1.1 A Theoretical Approach to Relaxation Dynamics

A common misconception in chemistry is to consider systems to be static in nature. We are used to building a static picture of systems: X-ray structures; ball- and stick- models; optimised geometries *etc*. However, while this limited picture is sufficient to predict some properties accurately, there are many situations where this approach is not suitable and time dependent behaviour must be taken into account.

Disorder in molecular systems is induced by thermal fluctuations which are characterised by deviations from a mean structure. These fluctuations can occur over a huge number of different time scales, ranging from femtoseconds (C-H stretches) to hours or even days (protein folding) [Brooks III *et al.* 1988]. Protein dynamics has been a field of interest to spectroscopists and computational chemists for some time now but it has only been with the advent of cheaply available high speed computers in the last couple of decades that routine simulation of protein systems has been possible. One of the first such studies was carried out on the oxygen storage and transport protein: myoglobin [Case and Karplus 1979]. Since then increasingly complex simulations have been carried out to the point where modelling of protein folding, albeit for a very small protein system, can be simulated [Duan and Kollman 1998]. MD simulations are now a common

computational simulation method for studying molecular systems including liquids, solutions, electrolytes, liquid crystals, crystals, zeolites, membranes and polymers such as proteins, DNA, and polysaccharides [Allen and Tildesley 1987; Catlow *et al.* 1990; Meyer and Pontikis 1991; van Gunsteren *et al.* 1993].

MD simulation of proteins are now common place with a large array of systems being routinely studied. Some very recent examples include the study of acylation mechanisms in the active sites of serine penicillin-recognising proteins [Oliva *et al.* 2003], the study of the stability of X-ray and NMR based protein structures [Fan and Mark 2003], and the study of unfolded polypeptide chains as a model for the first steps in protein folding [Krieger *et al.* 2003].

In order to study large systems, classical molecular mechanics (MM) protocols are routinely employed since the evaluation of the equations involved is very rapid. This computational efficiency has lead to classical MD simulations becoming an important tool in the study of biomolecular systems, and proteins in particular. However, despite the ubiquity and importance of MD simulations, it is not always clear how well such simulations actually represent the dynamics of proteins. Indeed it is partly due to the lack of available experimental methodologies that MD has become such a popular and important tool. One of the problems with studying large biomolecular systems is that they can be very susceptible to small errors in the potentials used to represent the energy of the system and hence the force on the atoms that drive the equilibrium fluctuations. Small subtle interaction energies, such as $\pi$ stacking energies or hydrogen bonding can have a very large cumulative effect on the system. This is further compounded by the fact that such systems cannot normally be treated in isolation. It is typically necessary to include the effects of solvent, either implicitly or explicitly, on the system. Validation of MD simulations and methods is thus an important aspect of this field of research. A good review of some of the aspects involved in the validation of molecular dynamics simulations is provided by Van Gunsteren [van Gunsteren and Mark 1998].

In this work a linear response based method has been developed and applied to a protein in order to investigate a range of assumptions. One of which is how well the equilibrium fluctuations of a protein active site are represented by a classical MD simulation. The development of a method that allows the relaxation dynamics and equilibrium fluctuations of protein systems sampled by molecular

dynamics to be tested via a direct comparison with experiment provides a useful tool in the validation of MD simulations.

***The Problems with a Classical Approach***

The use of classical molecular mechanics force fields pose a number of problems. MM techniques require carefully calculated parameters, the assumption that these may be transferred to a new system is not always valid. MM techniques are also not good for studying reactions since the harmonic potentials used for bond stretching and bending, and the neglect of electron rearrangement, prohibit the breaking of bonds. The neglect of quantum effects also make the investigation of subtle reaction mechanisms, such as proton tunnelling, impossible.

### 1.1.1 A Combined QM/MM Approach

One method for addressing the inadequacies of molecular mechanics, while preserving much of the speed advantage over pure quantum mechanical methods, is to use a hybrid quantum mechanical/molecular mechanical (QM/MM) approach. For simple chemical systems it is often possible to "prune" the model to a more manageable size by removing inactive groups. For enzymes, however, the whole system is often important for controlling the active site conformation and hence activity and so neglect of specific regions can lead to very poor or unexpected results. The hybrid QM/MM method is ideal for these types of systems as the active site, or region of interest, can be treated using electronic structure methods, either semi-empirical or full *ab initio*, while the rest of the protein can be modelled using classical molecular mechanics [Field *et al.* 1990]. Both semi-empirical and *ab initio* techniques have been implemented with the QM/MM method, allowing a number of properties to be calculated such as kinetics [Gao and Truhlar 2002], reaction pathways [Monard *et al.* 2003], free energies [Stanton *et al.* 1998] and mechanisms [Burton *et al.* 1998; Hillier 1999] associated with chemical and enzymatic reactions. Recent developments have extended the method to allow a fully integrated quantum mechanical calculation of properties such as vibrational frequencies and chemical shifts within conventional molecular mechanical and quantum mechanical program packages [Cui and Karplus 2000a; Cui and Karplus 2000b].

The main problem with QM/MM schemes is deciding how to couple the two parts of the system together. The approach employed in this work, is to incorporate the charges of the MM atoms into the one electron Hamiltonian of the QM calculation in an analogous fashion to the nuclear charges. This results in the QM atoms experiencing the electric potential due to all the MM atoms. Details of this approach are provided in section 3.2.5. Van der Waals potentials can also be included in order to prevent QM and MM atoms from bumping into each other. In this work, however, the QM/MM method has only been used to perform single point calculations and so van der Waals interactions have been neglected.

The big problem with coupling the two systems in the QM/MM approach comes at the interface of the two where any bonds that cross the interface must be cut and replaced by suitable link atoms, typically hydrogen. Fortunately the area of interest in the protein system chosen for this study, LADH, is the coenzyme NADH which is not covalently bound to the protein matrix and thus the problem of link atoms is not applicable. Some examples of studies on LADH using the QM/MM approach are discussed in section 1.1.3.

### 1.1.2 Overview of this Work

The extent of a chemical reaction is determined by the free energy difference between reactants and products. The rate constant for a catalytic step is greatly affected by the free energy of activation. It would be useful therefore to be able to predict the extent, origin and rate of development of this free energy difference in condensed-phase chemical reactions. One of the problems with studying protein systems is that they possess an almost infinite number of reaction paths, all of very similar energy, and thus knowledge of the active site structure does not necessarily lead to an accurate calculation of the rate of reaction. Thus an approach for calculating free energies of reaction that is independent of the transition state structure is important for studying reactivity in proteins.

For this we turn to the fluctuation-dissipation theorem of Callen and Welton [Callen and Welton 1951; Callen 1960]. This theorem, discussed in detail in Chapter 4, provides a way of linking microscopic equilibrium fluctuations, the type that can be sampled by an MD simulation, to macroscopic observables. The fluctuation-dissipation theorem is valid within the linear response regime (§4.2.1)

which requires the forces acting to restore a system to equilibrium to be linearly related to the system's distance from equilibrium.

Before attempts can be made to use equilibrium fluctuations to probe free energies of reaction in proteins it is necessary to investigate a number of factors. The first is to test the range of reorganisation energies over which the linear response approximation holds. The second is to evaluate how well molecular dynamics simulations probe equilibrium fluctuations in protein systems and the third is to test how sensitive the predicted fluctuations are to the parameters and calculation methods used.

To this end the QM/MM method developed by Mercer *et al.* [Mercer *et al.* 1997; Mercer *et al.* 1999] has been extended in this work to study proteins. The linear response approximation, implicit in this approach, has been tested and found to hold for a protein system with a large reorganisation energy, suggesting that the linear response approximation is suitable for studying a whole range of biological systems. The effects of the parameters and computational methods used have also been investigated.

### 1.1.3 The QM/MM Approach to Calculating Optical Spectra

A liquid or protein will respond to a rapid change in charge distribution by structural rearrangements that seek to minimise the energy of the system. The energy associated with such rearrangements is generally termed the reorganisation energy and has a well established role in, for example, non-adiabatic electron-transfer theory. The reorganisation energy also appears in the response of a condensed-phase chemical system to optical excitation. The creation of an excited electronic state rearranges the electron density in the molecule, leading to a rearrangement of the surrounding medium and a lowering of the excited-state energy and concomitant raising of the ground-state energy. If a system is fluorescent then the reorganisation energy is revealed by the fluorescence Stokes shift (§4.1.2) which is proportional to the reorganisation energy of the system. A prediction of the Stokes shift from theory provides an excellent experimental validation that the system's equilibrium dynamics are being adequately sampled.

In one sense, the creation of an excited state and subsequent Stokes shift of the fluorescence can be regarded as an extremely simple type of chemical reaction. The reactant in this case being the excited state of the molecule initially created by the applied optical field, and the product being the state created by relaxation of the system. The free energy gap between reactants and products is indicated by the shift in the peak of the steady-state absorption band from that of the steady-state emission band. The steady-state emission spectrum is a good representation of the product energy because most of the relaxation is complete well before the excited-state emits a photon. Moreover, as discussed in Chapter 4, the spectral widths of the absorption and emission spectra reflect the time-scale of the fluctuations.

Attempts to calculate the free energy difference between initial and relaxed states in optically excited systems are a natural first step prior to calculating free energy gaps, free energies of activation, and relaxation rates during more complex chemical processes.

In this work, configuration space has been sampled with a classical molecular dynamics simulation (Chapter 2) and the electronic energy gap between ground and excited state as a function of time using an *ab initio* CIS [Foresman *et al.* 1992] based QM/MM approach (Chapter 3) or a semi-empirical/DFT based QM/MM approach (Chapter 5). In so doing the method of exploring phase space has been separated from that of evaluating each contribution of phase space to the observable. If time-ordered behaviour were not required, then phase space could have been sampled, in principle, by a form of Monte Carlo procedure with a quantum mechanical evaluation of each element. In a similar way the method employed in this work separates the two aspects of the problem, but maintains time ordering.

Classical molecular dynamics in conjunction with classical evaluation of energy gaps has previously been applied to the calculation of optical properties of smaller solvated molecules. This yielded predictions of the shifts of the absorption and emission maxima between vacuum and solvent of indole and 3MI [Muino and Callis 1994], the Stokes shift of coumarin 153 [Kumar and Maroncelli 1995], and the Stokes shift and dynamic fluorescence shift of coumarin 343 [Jimenez *et al.* 1994; Fleming and Cho 1996]. More recently semi-empirical methods have been used to investigate the link between the optical spectra of Fe(II) cytochrome *c* and distortions of the heme and surrounding protein [Prabhu *et al.* 2002] and for investigating tryptophan fluorescence in proteins [Vivian and Callis 2001]. Very

recently, after the publication of this work [Walker *et al.* 2002], a quantum classical approach, utilising the GROMOS force field [van Gunsteren *et al.* 1996] and CIS energy gap evaluations has been used to study how the conformational flexibility of a protein affects the electronic properties of a chromophore [Spezia *et al.* 2003]. In the work of Spezia *et al.*, however, the chromophore geometry was kept fixed throughout the simulations.

The QM/MM method used in this work has previously been successfully used to calculate the relaxations of solvated chlorophylls [Mercer *et al.* 1997; Mercer *et al.* 1999] reproducing the reorganisation energies and spectral widths to within 20 % of their experimental values. In this work the method has been automated and optimised for studying proteins and has been successfully used to calculate the much larger relaxations in the enzyme liver alcohol dehydrogenase (LADH) (Chapters 2-5). The QM/MM method of Mercer *et al.* has also been used to look at zinc-myoglobin uncovering the existence of correlation between two degenerate excited states, this ongoing work is discussed in Chapter 6.

A number of groups have successfully employed QM/MM methods for the study of LADH. Gao, Truhlar and co-workers [Gao *et al.* 1998; Alhambra *et al.* 2000] used a QM/MM (Semiempirical AM1/CHARMM-22) potential function in semi-classical quantum dynamics simulations of LADH to investigate Swain-Schaad kinetic isotope effects of hydride transfer in LADH catalysed benzyl alcohol oxidation. Cui, Karplus and co-workers have similarly studied hydride transfer using a DFT based QM/MM method [Cui *et al.* 2002]. Hammes-Schiffer and co-workers [Agarwal *et al.* 2000; Billeter *et al.* 2001] have similarly looked at quantum dynamics of LADH using an empirical valence bond potential to represent electronic quantum effects, with nuclear quantum effects of the transferring hydrogen included by representing the hydrogen nucleus by a three-dimensional vibrational wave function. In this way they have been able to probe the reactivity of LADH without specifically identifying the reaction coordinate. Ryde has performed QM/MM simulations on the catalytic zinc atom and surrounding residues of LADH [Ryde 1996b] that confirm previous molecular dynamics studies [Ryde 1995] that showed four coordinate zinc to be favoured over 5 coordinate. While recently Tresadern, Hillier and co-workers [Tresadern *et al.* 2002; Tresadern *et al.* 2003] have used an AM1 and a Hartree Fock based QM/MM approach to study kinetic isotope effects in the reduction of benzyl alcohol by LADH.

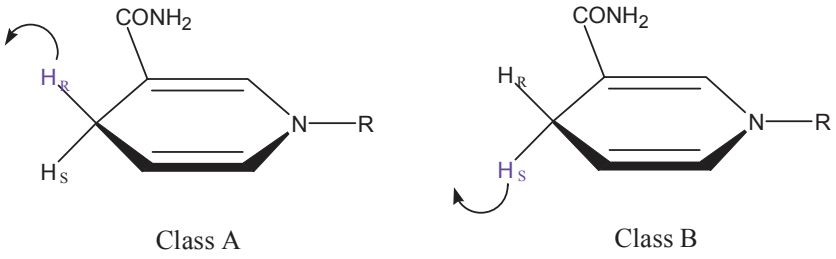



**Figure 1-1** Illustration of the difference in the stereospecificity of class A and class B dehydrogenases. The hydrogen that is transferred from NADH in each case is shown in blue.

The choice of LADH for this work was based on the decision to test both how the QM/MM linear response approach would perform for protein systems and also whether the linear response approximation would hold for a system with a large Stokes shift. LADH possesses both these attributes and being a well studied protein was an ideal candidate for this work. LADH is discussed below.

# 1.2 Liver Alcohol Dehydrogenase – An Overview

## 1.2.1 Structure and Function

### 1.2.1.1 The Dehydrogenases

Liver alcohol dehydrogenase (LADH) (Figure 1-2) belongs to a class of enzymes known as the dehydrogenases [Clake and Dafforn 1998]. These enzymes utilise either $NAD^+$ or $NADP^+$ as coenzymes for the oxidation of alcohols to carbonyl compounds. Of these enzymes most are specific to just one of the coenzymes but a few do use both. The oxidation reaction is readily reversible, with carbonyl compounds being reduced by either NADH or NADPH. The reduced coenzyme has a characteristic UV absorption at 340 nm and a characteristic fluorescence when excited at 340 nm that makes the appearance or disappearance of the reduced coenzyme easy to quantify. This in turn provides a simple measure of reaction rate [Fersht 1999].

**Table 1-1** Coenzyme specificity and stereo class for some common dehydrogenases.

| Dehydrogenase | Coenzyme Used | Class | Example PDB |
|---|---|---|---|
| Alcohol | NAD+ | A | 1LDY |
| Glucose 6-phosphate | NADP+ | B | 1DPG |
| Glutamate | NAD+ or NADP+ | B | 1AUP |
| Glyceraldehyde 3-phosphate | NAD+ | B | 1B7G |
| Lactate | NAD+ | A | 1IOZ |
| Malate | NAD+ | A | 1B8P |

The enzyme + coenzyme complex is termed the *holo*enzyme while the free enzyme is called the *apo*enzyme. Overall there are a large number of different dehydrogenases, divided into classes based on the coenzyme they use and their stereospecificity. There are two classes of stereospecificity. Class A transfers the pro-*R* hydrogen from NADH while class B transfers the pro-*S* hydrogen (Figure 1-1). The rationale for these two classes is based on the observation that the more reactive carbonyls are reduced by class A enzymes while class B enzymes are associated with the less reactive carbonyls [Benner 1982; Oppenheimer 1984; Benner *et al.* 1985]. Several dehydrogenase structures have now been solved and reviewed in depth in the literature [Eklund and Brändén 1987; Ohno and Ushio 1987; Clake and Dafforn 1998]. The crystal structures of the dehydrogenases solved to date are consistent with NMR studies and show that for class A enzymes the nicotinamide ring is in an anti configuration about the glycosidic bond while for class B enzymes the nicotinamide ring is bound in a syn conformation [Gronenborn and Clore 1982]. Table 1-1 provides a summary of the coenzyme specificity and stereospecificity class of some common dehydrogenases [Fersht 1999].

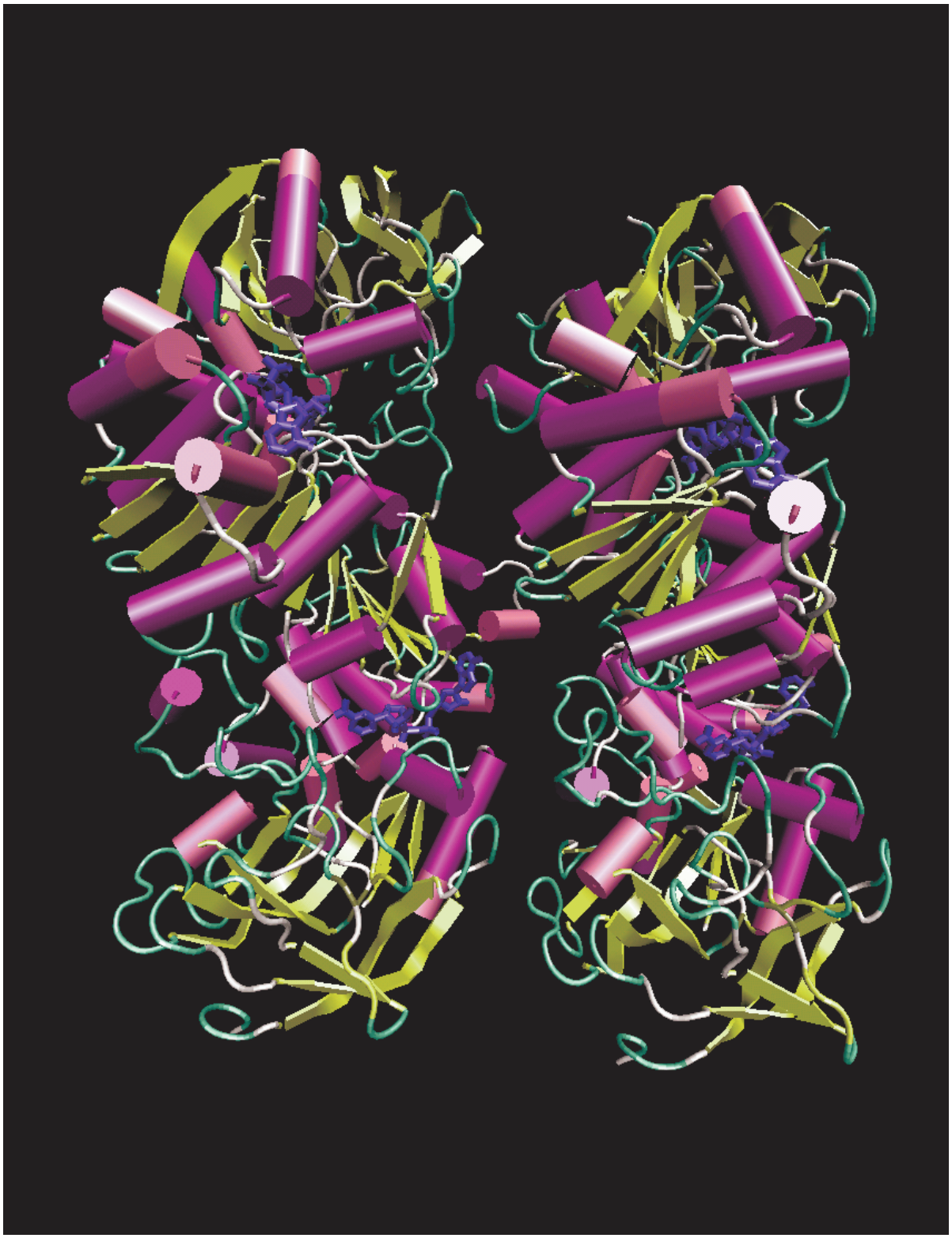


**Figure 1-2** Cartoon representation of the crystal structure of horse-liver alcohol dehydrogenase (Brookhaven Protein Databank 1LDY). Illustration shows NADH coenzyme in blue indicating that the X-ray crystal structure consists of two dimers in close proximity.

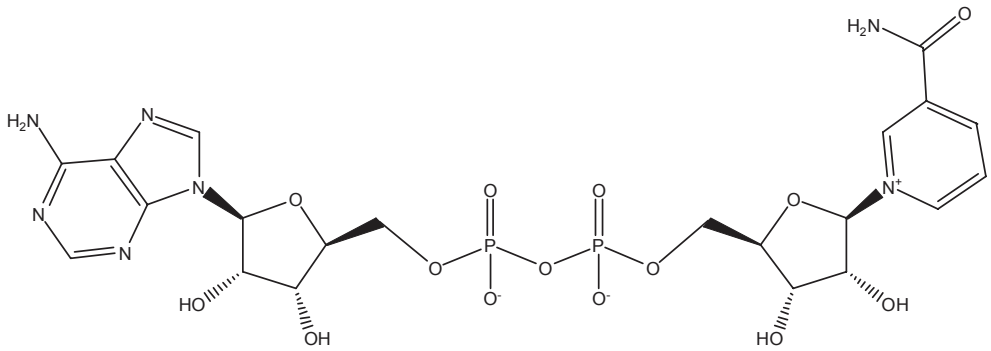


**Figure 1-3** 2-D Schematic representation of the NAD+ coenzyme.

### 1.2.1.2 The Alcohol Dehydrogenases

The protein chosen for study in this work, based on its spectral properties (§1.2.2), is the Equine variant of LADH called horse liver alcohol dehydrogenase. This is part of the alcohol dehydrogenase family of enzymes, a class A sub section of the broader class of enzymes termed the dehydrogenases (§1.2.1.1).

The alcohol dehydrogenases are zinc metalloenzymes of broad specificity which can oxidise both aliphatic and aromatic alcohols to their corresponding aldehydes and ketones. All alcohol dehydrogenase enzymes function via the transfer of hydrogen using the nicotinamide adenine dinucleotide ($NAD^+$) coenzyme (Figure 1-3) [Fersht 1999]. In humans, LADH is especially important in the breakdown of ingested ethanol as well as alcohols produced by bacteria in the gut. The toxic ethanol is oxidised to acetaldehyde, even more toxic, which is subsequently oxidised to acetate by aldehyde dehydrogenase. Acetate is then metabolised in cells. Alcohol dehydrogenases also act on a broader range of substrates in humans such as steroids, fatty acids and retinol. Genetic polymorphisms of alcohol metabolising enzymes, prevalent in people of Japanese descent, have been related to alcohol sensitivity and alcoholic diseases [Yoshida 1994] as well as increased risk of head and neck cancers [Yokoyama and Omori 2003].

The reduction chemistry of $NAD^+$ has been well characterised [Popják 1970]. The oxidation of an alcohol requires the removal of two hydrogen atoms (Figure 1-4). One is believed to be transferred directly to the 4 position of the nicotinamide

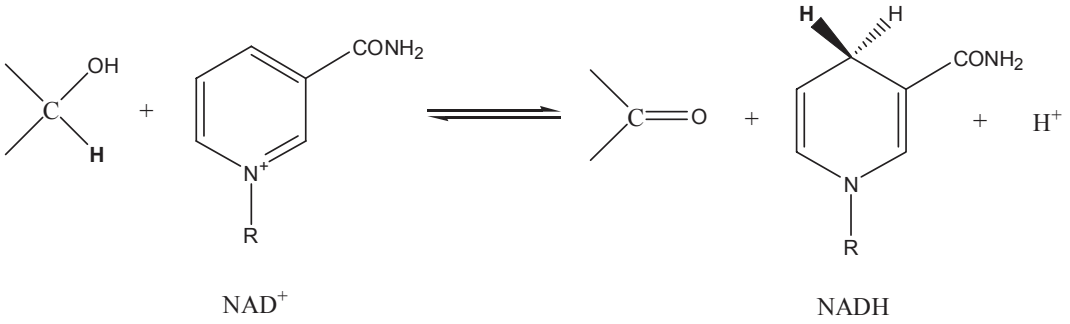



**Figure 1-4** Oxidation protocol of alcohol dehydrogenases, showing the role of the $NAD^+$ coenzyme.

ring of $NAD^+$ while the other is released as a proton [Fisher *et al.* 1953; Pullman *et al.* 1954]. There is still much debate over the exact mechanism by which oxidation occurs; it is generally believed that the hydrogen is transferred as a hydride ion, $H^-$, but it is also possible that a free radical intermediate is formed.

Of the alcohol dehydrogenases, the two most studied are those of yeast and horse liver. In this work the horse liver variant of alcohol dehydrogenase has been used due to the availability of good quality crystal structures and its relatively cheap availability in pure crystal form.

The crystal structures of the horse liver *holo*- and *apo*-enzymes were first solved at 2.9 Å resolution [Eklund *et al.* 1981] and 2.4 Å resolution respectively [Eklund *et al.* 1976b]. Since then a large amount of study into the structure of LADH has been undertaken and a brief search of the Brookhaven Protein Data Bank[1] [Berman *et al.* 2000] for liver alcohol dehydrogenase, today yields 34 structures covering numerous analogues and mutants ranging from 3.20 Å [Plapp *et al.* 1983] to 1.13 Å [Rubach and Plapp 2003] resolution.

The biologically significant LADH unit exists in solution as a symmetrical dimer. The dimer is composed of two identical chains of approximately 5700 atoms (horse liver LADH = 5707 atoms) and *MW* 40,000 per chain. Each chain contains a single binding site for $NAD^+$ but two for $Zn^{2+}$. The two zinc ions are separated by about 20 Å with only the one closest to the coenzyme playing a catalytic role. The other has only a structural role.

The LADH enzyme exists in two forms, E and S, which differ by only six residues [Eklund *et al.* 1976a], none of which are at the junction of the dimer. Only

[1] http://www.rcsb.org/pdb/

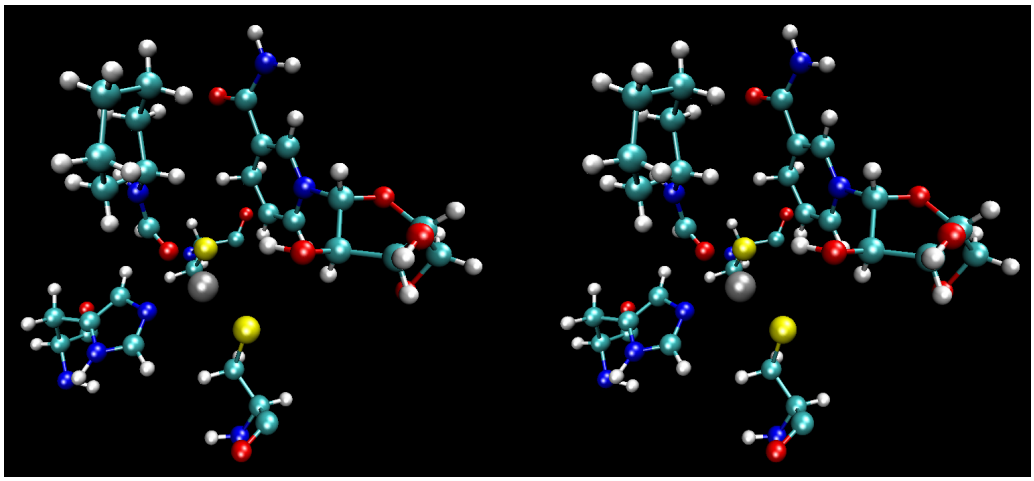


**Figure 1-5** Cross eyes stereo image of the catalytic zinc atom and surrounding residues of horse liver alcohol dehydrogenase. Adapted from PDB ID 1LDY, showing part of the reduced form of the coenzyme (NADH) together with a substrate inhibitor, *n*-cyclohexylformamide and the bound histidine and two bound cysteine residues.

the S form of LADH is active towards 3-$\beta$-hydroxysteroids, but both forms are active toward ethanol. Since none of the residue differences occur at the dimer interface the E and S chains can combine to form SS, EE and ES dimers termed *isozymes*. In liver the enzyme consists of between 40 – 60 % EE form with the remainder being a mixture of ES and SS forms. The crystal structure used in this work (1LDY - §2.3) is the EE form of the enzyme.

#### 1.2.1.3 LADH Active Site Structure

Crystallographic studies of horse liver alcohol dehydrogenase and its complexes with coenzyme and various substrates suggest that the catalytic zinc ion is tetrahedrally four-coordinate, with the coenzyme in close proximity but not explicitly bound to the zinc atom. However, some spectroscopic studies [Dworschack and Plapp 1977; Schmidt *et al.* 1979; Andersson *et al.* 1981; Makinen and Yim 1981; Bauer *et al.* 1991] indicate a five coordinate zinc ion with a water molecule occupying the fifth coordination site, although other spectroscopic investigations contradict these results [Maret *et al.* 1983; Bertini *et al.* 1984; Maret and Zeppezauer 1986; Corwin *et al.* 1987]. Work by Ryde [Ryde 1994] using HF quantum chemical geometry optimisations of active site models suggests that a zinc ion in vacuum with the same ligands as in alcohol dehydrogenase prefers four

coordination by about 20 kJ $mol^{-1}$. Indeed, inspection of the bond lengths, from crystallographic data, between various residues close to the catalytic zinc ion suggests that tetrahedral coordination is the preferred geometry. This is further supported by molecular dynamics simulations of the catalytic zinc site which shows a 36 $KJmol^{-1}$ preference for four coordination [Ryde 1995]. In this configuration the catalytic zinc binds to the enzyme through one histidine and two cysteine residues, with a water molecule or substrate forming the fourth first sphere ligand (Figure 1-5) [Ryde 1995].

### 1.2.2 Spectroscopic and Electronic Properties

A plethora of spectroscopic methods exists for the investigation of molecules, a number of which have been used for the study of LADH and it's coenzyme (NADH). These include Infrared [Nadolny and Zundel 1996; Nadolny and Zundel 1997; Hellwig *et al.* 2000; Trovaslet *et al.* 2003], Raman [Jagodzinski and Peticolas 1981; Jagodzinski *et al.* 1982; Yue *et al.* 1985; Chen *et al.* 1987a; Martin *et al.* 1987], Difference Raman [Chen *et al.* 1987b; Callender and Deng 1994; Deng *et al.* 1998] and Surface Enhanced Raman [Chen *et al.* 2002], Two Photon Excitation [Kierdaszuk *et al.* 1996], NMR [Cui and Karplus 2000a; Li and London 2002; Makowska-Grzyska *et al.* 2002; Rubach and Plapp 2002; Zhang and Pionnier 2002], Circular Dichroism [Reisbig and Woody 1978; Piersma *et al.* 1998], Circularly Polarised Luminescence [Schauerte *et al.* 1995] and X-Ray Crystallography [Ramaswamy *et al.* 1999; Meijers *et al.* 2001; Pocker *et al.* 2001; Makowska-Grzyska *et al.* 2002; Rubach and Plapp 2002]. The most relevant to this work, and that which will be discussed here, is that of solution phase optical absorption and emission (fluorescence) spectroscopy of NADH embedded in the LADH protein matrix [Scott *et al.* 1970; Gafni and Brand 1976; Piersma *et al.* 1998].

When in the reduced (NADH) form, the unbound coenzyme displays an absorption maximum in water, at pH 7, at 340 nm, a near identical emission maximum at 470 nm [Scott *et al.* 1970], and a Stokes shift of 130 nm (1.01 eV). For such a large Stokes shift it is surprising that the absorption and emission spectra are so similar with no obvious evidence of side bands (§4.3). More interestingly, when complexed with the liver alcohol dehydrogenase enzyme the absorption and emission spectra remain more or less identical in shape, merely blue shifted by

approximately 15 nm [Gafni and Brand 1976; Piersma *et al.* 1998]. The similarity between the enzyme bound and unbound emission spectrum of NADH suggests that the spectral shape is due largely to intramolecular effects making the enzyme bound system, with its reduced degrees of freedom over solution phase NADH and the lack of covalent bonds to the protein matrix, an ideal candidate for study using the QM/MM methodology. The reduced form of the enzyme was also chosen as an ideal candidate for this work as it has one of the largest Stokes shifts known for a protein based system. This large Stokes shift provided an opportunity to test the applicability of linear response theory (§4.2.1) to systems with large reorganisation energies.

## 1.3 Overview of this Thesis

This thesis consists of 6 chapters and a supplementary CD-ROM, the contents of which are summarised as follows:

**Chapter 2 - Molecular Mechanical Simulations of LADH**, describes the theory behind classical MD simulations and their applications to modelling protein systems. The theory described is used to generate four procedurally identical, but parametrically different, stable classical molecular dynamics simulations of horse liver alcohol dehydrogenase (LADH) containing the coenzyme di-hydro nicotinamide adenine dinucleotide (NADH) and an inhibitor, *n*-cyclohexyl formamide (CXF). The equilibrium dynamics sampled by the MD simulations are then used to generate structural 'snapshots' at 2 fs intervals for incorporation within the QM/MM protocol. The methods employed to obtain the time-ordered 'snapshots' along with evidence that validates the protocols used is discussed. It is found that the dynamics and structural preferences of the NADH coenzyme are sensitive to the MD parameters used but all simulations would, using 'traditional' MD validation techniques, be considered to be 'correct'. Evidence of structural asymmetry in the two active sites of LADH is found for one of the simulations but is believed to be a factor of the poor parameterisation.

**Chapter 3 - QM/MM Calculations on LADH**, uses the structural snapshots from Chapter 2 to calculate time ordered ground to excited state energy gap

fluctuation data for the two NADH chromophores in the LADH molecule. Aspects of molecular orbital theory are discussed and an overview of the Hartree Fock procedure is provided along with a discussion of the configuration interaction singles method for excited states and the implementation of a QM/MM method for evaluating the energy gaps. An overview of the processes implemented to tackle the logistics of running large numbers of QM/MM calculations is provided along with flow charts illustrating the construction of the QM/MM input files and their distributed computation on heterogeneous computer clusters. The performance of the CIS method for predicting the excited state energy gaps of LADH is discussed with data for a range of different basis sets along with the reasons for selecting the 3-21G* basis set. Finally the calculated energy gaps over 5 ps of the MD production runs are shown for each of the four MD simulations and the stability and dynamics of each energy gap trace is discussed.

**Chapter 4 - The Linear Response Approach** uses the energy gap traces produced in Chapter 3 to calculate theoretical UV/VIS absorption and emission spectra for LADH that can be compared directly with experiment. The experimental measurement of the UV/VIS absorption and emission spectra of CXF inhibited LADH+NADH is discussed and the results are found to compare well with those published in the literature. The application of elements of linear optical response theory to the energy gap fluctuations for the calculation of optical spectra is discussed and the theory behind the generation of optical observables from equilibrium fluctuations is explained in the context of this implementation. It is shown that the linear response theory holds for a system with a reorganisation energy as large as LADH and that the Stokes shift can be recovered to an accuracy better than 13 %. It is shown that the method is highly sensitive to the parameters used for the MD simulations and as such the linear response approach to relaxation dynamics has the potential to validate MD parameters for chromophores via direct comparison with experiment.

**Chapter 5 - Performance of Semi-Empirical Methods** repeats the energy gap evaluations for the 'correctly parameterised' MD simulation using different methods for calculating the ground to excited state energy gaps. Energy gap traces are calculated using ZINDO, TD-HF and TD-DFT methods. A brief description of the theory behind each of the methods is given along with their computational performance compared to the CIS approach. It is found that the performance of the *ab initio* TD-HF method is comparable to the CIS approach while the semi-

empirical ZINDO and TD-DFT approaches perform poorly for calculating equilibrium fluctuations. It is suggested that this is due to the parameterisation of the semi-empirical methods being designed for calculating optimised structures and as such these methods are not suitable for evaluating equilibrium fluctuations.

Finally **Chapter 6 - Work in Progress** provides an overview of the work currently in progress using the QM/MM based linear response method. Some brief examples of the extensions of the linear response approach are given along with a more in-depth discussion of work involving the zinc-substituted form of the oxygen storage protein myoglobin.

***CD-ROM*[2]**

The CD-ROM provided with this thesis contains, along with a pdf of this thesis in searchable form, supplementary data and software as follows:

1) Self installing executables for the Computational Chemistry Tools Software developed for this research.
2) Source code and description of the software used to generate Figure 2-12.
3) Patch Files for adding Bzip2 compression support to Carnal v6 and v7.
4) Examples of the Perl scripts used for creating the QM/MM Gaussian input files from the Amber MD trajectory files.
5) Static images and Mathematica [Wolfram Research 1999] files of the zinc-myoglobin TDM plots shown in Figure 6-6.
6) Details on converting SGI Crystal Eyes stereo glasses for use on a PC with a NVIDIA GeForce 2 or 3 card.[3]

[2] If you do not find a CD-ROM accompanying this thesis an iso file can be downloaded from http://www.rosswalker.co.uk/

[3] The ability to view crystal structures and dynamic data in 3D stereo on a large screen PC based system greatly aided the analysis of the MD data generated in this work.

# Chapter 2

# Molecular Mechanical Simulations of LADH

## 2.1 Introduction

The aim of the work described in this chapter was to generate a number of stable classical molecular dynamics simulations of horse liver alcohol dehydrogenase (LADH) containing the coenzyme di-hydro nicotinamide adenine dinucleotide (NADH) and an inhibitor (*n*-cyclohexyl formamide CXF) that would accurately describe the equilibrium fluctuations experienced by LADH at room temperature. The equilibrium dynamics sampled by these MD simulations were used to generate structural 'snapshots' at regular intervals which could then be incorporated into the QM/MM protocol described later. The methods employed to obtain these time-ordered 'snapshots', in the form of molecular dynamics trajectories, along with evidence that validates the protocols used is discussed below.

In order to test the sensitivity of the linear response method, discussed later, to the equilibrium dynamics represented by the classical MD simulations a series of MD runs were conducted using four different approaches to the parameterisation of the nicotinamide chromophore. The four separate but procedurally identical molecular dynamics simulations were all validated in the same way and while the results in Chapter 4 show that only one accurately

represents the equilibrium fluctuations of LADH correctly they would all, using 'traditional' MD validation techniques, be considered to be 'correct'.

## 2.2 Theory

This section contains an outline of the theory underpinning, and the implementation of classical molecular dynamics simulations [Allen and Tildesley 1987; Grant and Richards 1995; Frenkel and Smit 1996; Leach 1996; Jensen 1999; Leach 2001] and in particular the Amber force field [Cornell *et al.* 1995].

### 2.2.1 The Classical Model

Despite recent advances quantum mechanical calculations, while offering structurally and electronically vigorous descriptions of molecules, suffer as a result of their computational cost. Thus for the study, at an atomic level, of systems the size of proteins force field methods present the only practical technique widely available. In contrast to quantum mechanics calculations, force field (or molecular mechanics) methods do not take into account electronic effects considering only nuclear motions. Such a separation is facilitated by the Born-Oppenheimer approximation [Born and Oppenheimer 1927]. This approximation, also used in quantum mechanical calculations and discussed in §3.2.1.1, simplifies the calculations involved in simulating molecular systems by assuming that the motion of electrons and nuclei is decoupled. It is assumed that because the mass ratio between electrons and nuclei is large the electrons can effectively reorganise instantaneously. This separation allows the nuclear positions to be described by classical Newtonian mechanics. The use of classical mechanics greatly reduces the complexity of the calculations involved in simulating large molecular systems and makes possible simulations involving hundreds of thousands of atoms. The drawback of the force field approach is that electronic effects cannot be examined; calculations are therefore restricted to the ground state, although excited state MD simulations have been attempted using specially parameterised force fields [Debolt and Kollman 1990]. A classical approach also excludes the direct study of chemical reactions or other phenomena where changes in electron distribution are significant.

In order to study large systems in which electronic effects are relevant it is necessary to use a combined quantum mechanical / molecular mechanical (QM/MM) approach. The approach taken in this work is to couple single point QM/MM calculations to classical MD simulations in order to study excited states and other electronic effects while retaining the molecular environment. The QM/MM approach used in this work is discussed in detail in Chapter 3.

#### 2.2.1.1 The Force Field

It is difficult to accept that the behaviour of atomistic systems, which behave according to quantum rather than classical laws, could be accurately described by the application of classical Newtonian mechanics. The justification for this can be made by considering the de Broglie expression for the thermal wavelength $\Lambda$,

$$\Lambda = \sqrt{\frac{2\pi\hbar^2}{Mk_BT}} \tag{2.1}$$

where $T$ is the temperature and $M$ is the atomic mass. The approximation of classical behaviour holds if $\Lambda \ll \alpha$, where $\alpha$ is the mean nearest neighbour separation. This holds for 'heavy' liquid systems at all but the lowest temperatures, at which quantum effects become important.

To describe the molecular dynamics of a system using a classical model a function that represents the potential energy of the system together with the related parameters is required. The energy is typically calculated from the sum of steric and non-bonded interactions, such that the total energy is described as the sum of bond, angle, dihedral and non-bonded energies.

$$E_{total} = E_{bond} + E_{angle} + E_{dihedral} + E_{non-bond} \tag{2.2}$$

Typically each bond, angle and dihedral energy contribution is calculated individually and then summed; while the non-bonded interactions represent the influence of non-covalent effects such as electrostatics and van der Waals forces.

The exact form of the terms in the above potential function and the associated parameters varies across different molecular mechanical force fields. Some force fields also include a cross term representing coupling between the first three terms in equation (2.2). Some examples of commonly used force fields include the Allinger

MM2 and MM3 series [Allinger 1977; Allinger *et al.* 1989], CHARMM [Mackerell *et al.* 1998], AMBER 1984 [Weiner *et al.* 1984] and GROMOS [Hermans *et al.* 1984]. Each force field has a slightly different ethos and is typically suited to the study of one class of molecules. In this work the AMBER 1995 force field and parameter set [Cornell *et al.* 1995] has been used. This force field is designed for the simulation of proteins and nucleic acids and uses the following potential function,

$$V\left(r^n\right) = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_r R_{ij}}\right] \tag{2.3}$$

where the potential energy $V$ is expressed as a function of the positions $r$ of $n$ atoms. $K_r, r_{eq}, K_\theta, \theta_{eq}, V_n, n, \phi, \gamma, A_{ij}, B_{ij}, \varepsilon_r, q_i$ and $q_j$ are all empirically defined parameters. The first three terms of equation (2.3) correspond to the bond, angle and dihedral term respectively, while the last term describes the non-bonded van der Waals and electrostatic interactions.

### 2.2.1.2 Interatomic Potentials

#### 2.2.1.2.1 The Stretch Energy

In the AMBER force field there are three terms describing the interactions between atoms that are covalently bonded either directly or separated by two or three bonds. The first term ($K_r\left(r - r_{eq}\right)^2$) describes a simple harmonic potential obeying Hooke's Law and is used to represent the energy involved in the stretching of a bond between two directly linked atoms. $2K_r$ is the force constant for the bond and $\left(r - r_{eq}\right)$ the distortion from the equilibrium bond length. It should be immediately obvious from this expression that such an approach prohibits the study of bond breaking since the expression tends to infinity as the bond length is increased significantly beyond the equilibrium value.

This expression is derived from a Taylor expansion around an equilibrium bond length $r_{eq}$. Truncating the expansion at second order gives, for two atom types A and B, the expression,

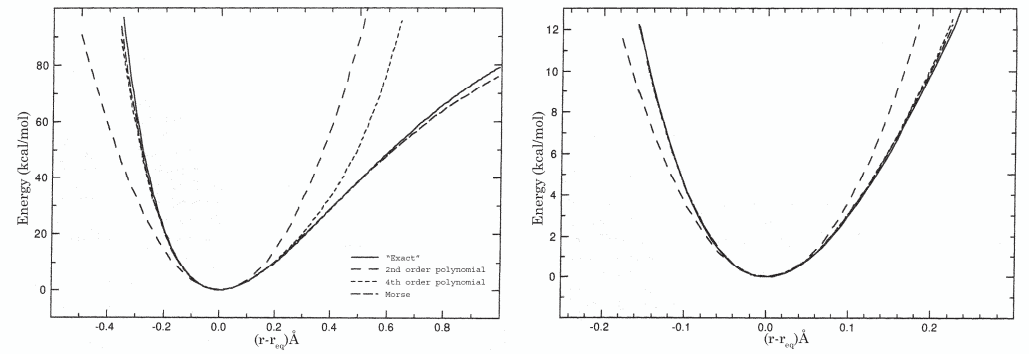



**Figure 2-1** The stretch energy for $CH_4$ showing the various functional forms in comparisson to a CASSCF/6-311++G(2df,2pd) electronic structure calculation ("exact"). Adapted from [Jensen 1999]

$$E_{stretch}\left(r^{AB}-r_{eq}^{AB}\right)=E(0)+\frac{dE}{dr}\left(r^{AB}-r_{eq}^{AB}\right)+\frac{1}{2}\frac{d^2E}{dr^2}\left(r^{AB}-r_{eq}^{AB}\right)^2 \tag{2.4}$$

The $E(0)$ term is set to zero since it represents the zero point on the energy scale and the derivatives are evaluated about $r_{eq}$. Since the expansion is about the equilibrium position the second term in (2.4) is zero and so in the simplest form the stretch energy can be represented by the Hooke's Law expression given above.

This harmonic representation of the potential is the simplest possible but provides a fair description of the energetics of bond stretching and compression when the bond length remains close to the equilibrium value (Figure 2-1). If however, the bond is distorted significantly away from equilibrium the harmonic expression fails to accurately represent the anharmonicity. By including higher order terms in the Taylor expansion,

$$E_{stretch}\left(\Delta r^{AB}\right)=K_{r2}^{AB}\left(\Delta r^{AB}\right)^2+K_{r3}^{AB}\left(\Delta r^{AB}\right)^3+K_{r4}^{AB}\left(\Delta r^{AB}\right)^4+\ldots \tag{2.5}$$

the functional form of $E_{stretch}$ can be improved as illustrated by the order 4 polynomial curve in Figure 2-1. Including extra terms in the bond stretching expression comes at a price: more parameters need to be included and there are more expressions to calculate. The polynomial expressions for bond stretching also possess the wrong limiting behaviour tending either towards $+\infty$ or $-\infty$ at long bond lengths. This behaviour can result in the molecule essentially 'blowing up' if a poor starting geometry is chosen. The correct limiting behaviour is for the bond

energy to converge to the dissociation energy when stretched to infinity. A function that shows this behaviour is the Morse potential [Morse 1929],

$$E_{Morse}\left(\Delta r^{AB}\right) = D\left[1 - e^{-\alpha\Delta r^{AB}}\right]^2 \tag{2.6}$$

where $D$ is the dissociation energy and $\alpha$ the force constant. The Morse potential thus requires three parameters per bond ($r_{eq}, D, \alpha$) rather than the two ($r_{eq}, K_r$) required by the harmonic expression. The exponential expression also makes computation more complex resulting in lower performance.

Since the force constants for bond stretching are generally large it is rare for a bond length in a molecular mechanical calculation to deviate significantly from equilibrium. Hence using a more complex function in (2.3) would "make the model unnecessarily cluttered with little gain in representing the properties of interest" [Pearlman *et al.* 1995]. The simple harmonic expression therefore suffices.

#### 2.2.1.2.2 The Bend Energy

The second term in equation (2.3) describes the angle bend energy between three covalently bound atoms. This also utilises Hooke's law using the same harmonic potential used for bond stretching. Here $\theta_{eq}$ is the equilibrium bond angle and $K_\theta$ the angular force constant. The angular force constants are typically smaller in magnitude than the stretching force constants as the energy penalty for angle bending is generally lower. As with the bond stretching term, a more accurate description is possible by including higher order terms. The simple harmonic approximation used in the AMBER force field is accurate to about $\pm 30°$ from the equilibrium angle while a 3rd order polynomial is accurate up to about $\pm 70°$. Since most systems lie within the $\pm 30°$ range higher order terms are generally only required for highly strained systems.

#### 2.2.1.2.3 The Torsional Energy

The bond-stretching and angle-bending terms, discussed above, are often referred to as 'hard' degrees of freedom since quite large energies are required to cause significant deviations from the equilibrium geometries. Most of the complex variations in structure and relative energies observed in biological systems are due to the 'softer' torsional and non-bonded contributions. These potentials are also

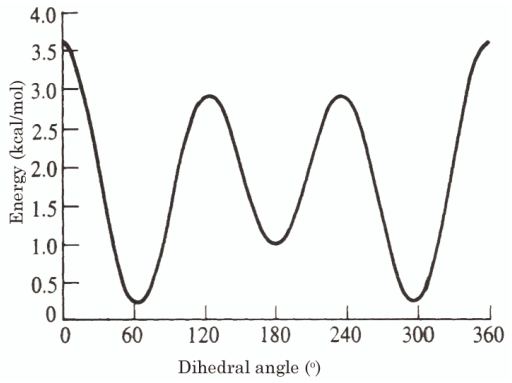



**Figure 2-2** Variation in torsional energy with O-C-C-O torsion angle for an $OCH_2$-$CH_2O$ fragment. Adapted from [Leach 2001].

unfortunately the more difficult to accurately describe. The importance of getting the torsional potentials correct is illustrated by the results discussed in section 4.4.2.

The barriers to rotation about a bond can be modelled in one of two ways. In very early force fields it was believed that rotational barriers could be omitted. The gauche-trans energy differences would be recreated by the non-bonded interactions. For organic molecules it was quickly realised that this was an almost impossible task and so dihedral angle terms were explicitly included. The AMBER force field, in common with a number of other organic molecule force fields, uses a Fourier series expansion (2.7) for the torsional potential.

$$V(\phi) = \sum_{n=0}^{N} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] \tag{2.7}$$

One term of this expansion appears in the AMBER force field equation (2.3) where $V_n$ is the relative barrier height to rotation, $n$ is the multiplicity (number of minima in a 360° rotation), $\phi$ is the dihedral angle and $\gamma$ is the phase factor which determines the location of the minima. $V_n$ is referred to as the relative barrier height since other terms in the force field equation contribute to the barrier height as the bond is rotated, especially the 1-4 non bonded interactions discussed in section 2.2.1.2.5. While most dihedral angles in the AMBER force field can be described adequately by a single two or three-fold term ($n = 2,3$) some dihedrals require a combination of terms to accurately describe the profile of the torsions. A good example of this is the O-C-C-O dihedral angle of an $OCH_2$-$CH_2O$ fragment.

This torsion angle is described by the combination of a 3-fold term and a 2-fold term,

$$\upsilon\left(\omega_{C-O-O-C}\right) = 0.25\left(1+\cos 3\omega\right) + 0.25\left(1+\cos 2\omega\right) \tag{2.8}$$

giving the energy profile shown in Figure 2-2.

Improper torsion angles, also known as out of plane bending, are defined for four atoms that are not bonded in a serial manner. They are used to maintain planarity where necessary, for example if a carbon atom is $sp^2$ hybridised then there is a significant energy penalty associated with deviations from planarity, and also to preserve the correct stereo chemistry at tetrahedral centres. The AMBER force field accounts for improper torsions in the same way as regular torsion angles but using a two-fold multiplicity.

#### 2.2.1.2.4 Cross Terms

Separating atomic motions into isolated bond, angle and dihedral terms provides a simplistic approach to calculating the potential energy. In reality the internal degrees of freedom within a molecule are not isolated from each other but are in fact coupled. For example, as a bond angle increases, the adjacent bonds often contract. Such effects are typically incorporated within molecular mechanics force fields via the application of cross terms (stretch-bend, stretch-torsion *etc*.). The use of cross terms is used to categorise force fields into three separate classes [Hwang *et al.* 1994]. A class I force field is restricted to harmonic terms and does not possess any cross terms. A class II force field contains some anharmonic terms and explicit cross terms while a class III force field would go beyond this taking into account chemical effects such as hyper-conjugation. The AMBER force field (2.3) by adopting a minimalist approach falls into the category of a class I force field. The AMBER force field takes the approach that by concentrating on optimising the non-bonded interactions the use of cross terms is unnecessary [Cornell *et al.* 1995].

#### 2.2.1.2.5 Non-Bonded Interactions

The accurate treatment of non-bonded interactions is paramount in a force field description. The philosophy behind the AMBER force field is to deal with

these interactions in the most accurate way possible within the limitations of the available hardware. Non-bonded interactions do not depend upon a specific bonding relationship between atoms. They are 'through-space' interactions, the number of which scales roughly as the square of the number of atoms. Unsurprisingly it is the non-bonded interactions which form the most time-consuming component of molecular mechanics simulations. Molecular mechanics force fields typically deal with non-bonded interactions in two groups, one comprising electrostatic interactions and the other van der Waals interactions.

***Electrostatic Interactions***

The last term in (2.3) describes the electrostatic interactions within the system. There are a number of ways to represent the charge distribution within a molecule, the simplest being via the use of point charges. This is the method utilised by the AMBER force field. In the point charge model a series of fractional charges are distributed throughout the molecule. If the charges are centred on atoms then they are referred to as partial atomic charges. In this model the interaction energy is calculated using Coulomb's law,

$$V(elec) = \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \tag{2.9}$$

where $\varepsilon_0$ is the permittivity of a vacuum, $\varepsilon_r$ is the relative permittivity of the medium, $q_i$ and $q_j$ are the partial charges on the two atoms and $r_{ij}$ is the distance between them. Since the charge on an atom is not experimentally observable the partial atomic charges have to be assigned in an analogous fashion to the parameters used in the bonding interaction terms. The partial charges are generally obtained by fitting to a charge potential obtained from an electronic structure calculation. The method for calculating the partial charges used in the AMBER force field is discussed in more detail in section 2.2.1.6.3.

An alternative approach is to represent the charge distribution using the central multipole expansion. In this method the electrostatic interaction energy is then a truncated series consisting of charge-charge, charge-dipole, dipole-dipole, dipole-quadrupole, quadrupole-quadrupole and so on. This method can be very useful for systems with a high degree of anisotropy such as aromatic $\pi$-systems where the charge distribution cannot be treated as spherically symmetric.

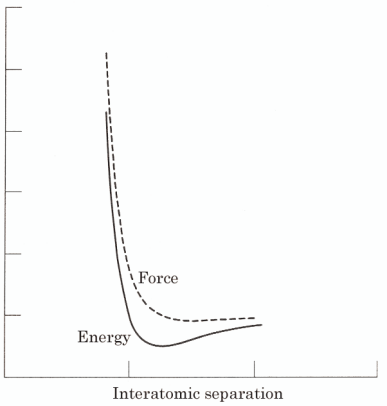



**Figure 2-3** The interaction energy and force between two argon atoms as a function of their separation. Adapted from [Leach 2001].

Unfortunately the evaluation of forces is not trivial and so this method does not lend itself to molecular mechanical simulations. The limitations of the multipole method and the simplicity of the partial charge model has led to widespread use of the latter.

***Van der Waals Forces***

Electrostatics cannot account for all of the non-bonded interactions in a system. Good examples of this are provided by the noble gases whose multipole moments are all zero, and yet have liquid and solid phases. The forces responsible for this are termed van der Waals forces after their discoverer van der Waals who was the first to quantify the deviations from ideal gas behaviour that are due to these forces. Van der Waals interactions are electrostatic in nature, but the magnitude of the interactions are much smaller. Figure 2-3 shows the force (equal to the negative of the energy gradient) between two isolated argon atoms as a function of their separation. The curve in Figure 2-3 is considered to arise from a combination of repulsive and attractive forces. The repulsive term originates from the Pauli exclusion principle. At very short distances the energy varies steeply with $r$ but at larger separations the decay is exponential in nature.

The attractive forces, which dominate at longer distances, arise from dispersion, or London [London 1930], forces and are due to the formation of instantaneous dipoles. Such an instantaneous dipole can in turn induce a dipole in nearby atoms, resulting in an attractive force. A simple model of this attractive

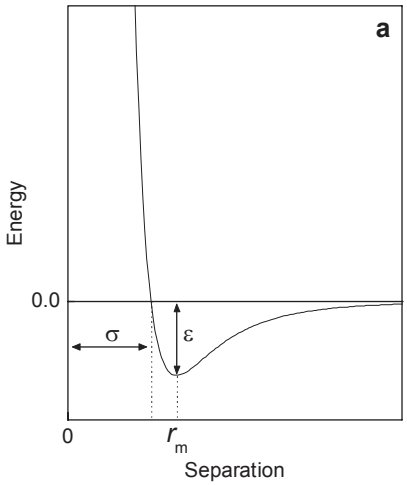



**Figure 2-4** The Lennard-Jones 12-6 potential and associated parameters $\varepsilon$, $r_m$ and $\sigma$

force, proposed by Drude, shows it to be proportional to $r^{-6}$, where $r$ is the interatomic distance.

A number of expressions exist for the modelling of van der Waals forces. Arguably the best known of these, and the one used within the AMBER force field, is the Lennard-Jones 12-6 function,

$$V(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \tag{2.10}$$

where $A = \varepsilon r_m^{12}$ and $B = 2\varepsilon r_m^6$. The Lennard-Jones potential therefore contains two adjustable parameters; the depth of the well $\varepsilon$ and the minimum energy location $r_m$ (Figure 2-4). The function can also be written in terms of a collision diameter $\sigma$ in which case $A = 4\varepsilon\sigma^{12}$ and $B = 4\varepsilon\sigma^6$. Unlike the attractive $r^{-6}$ term which has a clear physical basis the choice of $r^{-12}$ for the repulsive term is purely arbitrary and chosen for ease of computation. In the AMBER force field 1-4 van der Waals interactions are treated slightly differently from the other van der Waals interactions. Coupled with the torsional potential terms, these 1-4 non-bonded interactions contribute to the torsional energy profile. The use of the $r^{-12}$ expression for repulsion is too severe for the 1-4 interactions and these are handled via the use of an empirical scaling factor.

Alternative expressions such as the Buckingham or Hill [Hill 1948] type potential (2.11) replace the $r^{-12}$ term with a more realistic exponential expression.

$$V(r) = Ae^{-Br} - \frac{C}{r^6} \tag{2.11}$$

Since the number of van der Waals interactions that need to be evaluated is high the Lennard-Jones potential is generally the preferred form since evaluating the exponential function in the Buckingham potential is computationally expensive.

***Many-Body Effects***

The electrostatic and van der Waals interactions described by equations (2.9) and (2.10) calculate the non-bonded interaction energies using the implicit assumption that the total interaction can be expressed as the sum of all the atom pairs. This picture, however, is not physically realistic since the energy of interaction between two atoms is influenced by the presence of a nearby third atom. Such three-body interactions are not additive and can have a significant effect on the dispersion energy. It is therefore important to account for these effects but to calculate them explicitly would require the evaluation of an extra $N/3$ non-bonded terms. Fortunately a computationally efficient way of dealing with these effects is to incorporate them within the parameterisation of the pair wise potentials. Thus the AMBER force field in common with many other force fields makes use of "effective" pair potentials when calculating the non-bonding interactions.

***Non-bonded Cut-offs***

It is the non-bonded interaction terms that are the most computationally demanding aspect of a force field calculation. There are $N(N-1)/2$ interactions, where $N$ is the number of atoms in the system. A reduction in the number of non-bonded interactions would thus have a dramatic effect on the speed of the calculation. Since the size of the van der Waals interaction between atoms decreases rapidly with distance it is possible to truncate the Lennard-Jones potential without introducing significant errors in the calculation. Unfortunately the electrostatic interactions are longer ranged and so truncating them can introduce significant errors into the calculation. Section 2.2.1.3 describes a more comprehensive method for treating the long range electrostatic interactions.

The distance used for the cut-off can be determined by evaluating the energy of the system for a range of different cut-off values. Beyond a certain cut-off size the energy begins to converge and so the errors introduced by the use of a cut-off diminish. See section 2.3.3.2.1 for an example of this procedure. Typical cut-off distances are between 8 Å and 15 Å with larger cut-offs resulting in slower performance but less errors in the non-bonded energy evaluation. The use of periodic boundary conditions (§ 2.2.1.4) limit the maximum cut-off size to half the length of the smallest cell dimension since the minimum image convention applies wherein each atom should "see" only one image of every other atom in the system.

Although establishing a non-bonded cut-off reduces the number of interactions that need to be calculated it does not in itself lead to improved computational efficiency. The reason for this is that in order to determine whether or not the potential between two atoms should be calculated it is first necessary to evaluate the distance between them and compare this to the cut-off size. This process is almost as time consuming as evaluating the non-bonded interactions. The solution lies in a 'book-keeping' technique involving the creation of a Verlet or non-bonded pair list [Verlet 1967b]. In this method two radii values are used, the first ($r_c$) being the cut-off distance and the second ($r_v$) being a slightly larger value referred to as the neighbour list cut-off. All atoms within the radius $r_c$ are stored in the list along with atoms at the larger radius $r_v$ allowing neighbours of atoms to be identified without having to calculate the distance between atom pairs. Updating the non-bonded pair list requires of the order of $N^2$ work for $N$ atoms but since the positions of the atoms change by only a small amount at each step it is only necessary to update the pair list every 10-20 steps. Thus while updating the pair list is computationally intensive the overall efficiency is increased.

### 2.2.1.3 Long-Range Forces

The handling of long-range forces, such as those described by electrostatic interactions (eq. (2.9)), pose a significant problem in simulations of molecular systems since they cannot be dealt with effectively using the conventional methods outlined above since their range can often exceed half the box length. Techniques for dealing with these forces in a computationally efficient manner include the cell multipole method [Ding *et al.* 1992a; Ding *et al.* 1992b], the reaction field method [Friedman 1975] and the Ewald summation [Ewald 1921]. In this work only the

Ewald summation technique has been employed, and a brief discussion of the theory follows[4].

First introduced in 1921 [Ewald 1921], the Ewald summation method was originally devised to study ionic crystals. It has since been widely incorporated, in a range of different forms, into molecular simulation techniques [Heyes 1981; Smith and Fincham 1993; Darden *et al.* 1999]. In this method each particle interacts with all of the particles within the simulation box and all of the image boxes in an infinite periodic array. The position of each of these boxes, with reference to the central box, assumed to be a cube for simplicity, of side $L$ containing $N$ charges, can be described by a translation vector $(\pm iL, \pm jL, \pm kL)$, where $(i, j, k)$ are integral numbers. The electrostatic interactions between atoms within the central box are calculated via the usual Coulomb term (2.9) while the interactions between atoms in the central box and those in the image boxes are accounted for in the same way but with the distance modified by the appropriate translation vector $\boldsymbol{n}(= n_x L, n_y L, n_z L)$. The overall electrostatic interaction energy is therefore,

$$V(elec) = \frac{1}{2} \sum_{|\boldsymbol{n}|=0}^{N'} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r \left| R_{ij} + \boldsymbol{n} \right|} \tag{2.12}$$

where the prime on the first summation indicates that the interaction $i = j$ is not included when $\boldsymbol{n} = 0$. Contributions to the total electronic interaction energy therefore come from interactions within the central box together with the interactions between the central box and all image boxes. The problem with the Ewald sum (2.12) is that it converges extremely slowly. (2.12) is actually a conditionally convergent series as it consists of two divergent series whose sum is dependent on the order in which the summation is conducted. The solution to this problem is to make use of the identity (2.13) which allows conversion of (2.12) into

---

[4] For a more detailed description see de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1980). "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants." *Proceedings of the Royal Society of London Series a- Mathematical Physical and Engineering Sciences* **373**: 27-56, de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1983). "Simulation of Electrostatic Systems in Periodic Boundary- Conditions .3. Further Theory and Applications." *Proceedings of the Royal Society of London Series a- Mathematical Physical and Engineering Sciences* **388**(1794): 177-193.

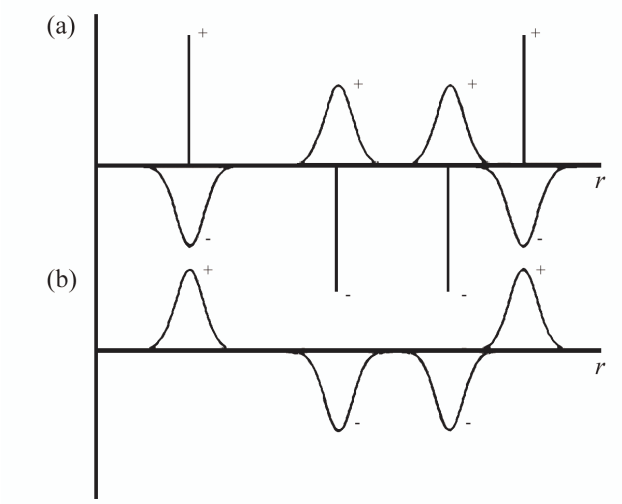



**Figure 2-5** An example of the charge distribution in the Ewald sum. (a) original point charges plus screening distribution (b) cancelling distribution.

the sum of two series, each of which converges much more rapidly [Darden *et al.* 1999].

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r} \tag{2.13}$$

By considering each point charge to be surrounded by a neutralising charge distribution (usually Gaussian) of equal magnitude but opposite sign (Figure 2-5) the sum over point charges can be converted into a sum of interactions between point charges plus the neutralising distributions. The interaction between the charges is therefore dependent on the amount of charge that is not screened. This interaction rapidly converges to zero at long distances, hence the direct sum converges quickly. The width of the shielding Gaussian functions is responsible for the rate of convergence; the wider the function, the quicker the convergence. The width of the shielding functions are typically chosen so that only interactions within the cut-off distance are included.

It is then necessary to include a second charge distribution in order to cancel the neutralising distribution. This second charge distribution is a periodic sum of Gaussians which by conversion to reciprocal space can be computed more easily since the reciprocal sum converges much more rapidly than its real space counterpart.

It is also necessary to subtract a third correction term from the potential to correct for the spurious self-term introduced by the interaction of each Gaussian

with itself. Finally it may be necessary to include the interaction of the periodic array with the surrounding medium to completely describe the potential. Despite some artefacts arising from the application of periodic boundary conditions the Ewald sum is still "the most 'correct' way yet devised to accurately include all the effects of long-range forces in a computer simulation" [Leach 2001].

The increased convergence of both the real space summation and the reciprocal sum results in greatly increased computational efficiency in evaluating the long-range electrostatic interactions. Despite this the Ewald sum is still computationally intensive since the reciprocal space sum scales as $N^2$. By utilising fast Fourier transforms (FFTs) the scaling of the reciprocal sum can be reduced to $N \ln N$. In order for FFTs to be used, however, it is necessary to have discrete data. The particle-mesh Ewald (PME) [Darden *et al.* 1993] approach used in this work replaces the point charges with their continuous coordinates by a grid based charge distribution. The atom centred point charges are then distributed on the grid in such a way as to reproduce the potential of the charges at the original locations. The PME method is widely used in the study of charged or polar systems and is considered routine for many types of solid-state materials. It has been shown to be very effective for molecular dynamics simulations of large molecular systems, including DNA and proteins [Darden *et al.* 1999].

#### 2.2.1.4 Periodic Boundary Conditions

A realistic model of a solution requires a very large number of solvent molecules to be included along with the solute. Simply placing the solute in a box of solvent is not sufficient, however, since while some solvent molecules will be at the boundary between solute and solvent and others will be within the bulk of the solvent a large number will be at the edge of the solvent and the surrounding vacuum. This is obviously not a realistic picture of a bulk fluid. In order to prevent the outer solvent molecules from 'boiling' off into space, and to allow a relatively small number of solvent molecules to reproduce the properties of the bulk, periodic boundary conditions are employed. In this method the particles being simulated are enclosed in a box which is then replicated in all three dimensions to give a periodic array, a two-dimensional representation of which is shown in Figure 2-6. In this periodic array a particle at position $\boldsymbol{r}$ represents an infinite set of particles at position,

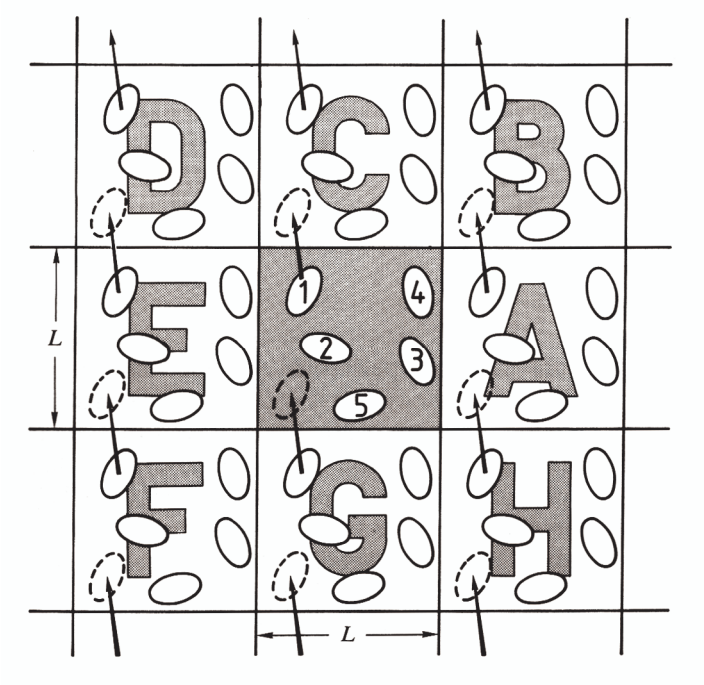



**Figure 2-6** A two-dimensional array of boxes. As molecule 1 moves from the central box into box C it is replaced by it's image which moves from box G into the central box. This movement is replicated across all the boxes. Adapted from [Allen and Tildesley 1987]

$$\boldsymbol{r} + i\boldsymbol{x} + j\boldsymbol{y} + k\boldsymbol{z} \qquad (i, j, k \to -\infty, \infty) \tag{2.14}$$

where $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$ are the vectors corresponding to the box edges. During the simulation only one of the particles is represented, but the effects are reproduced over all the image particles with each particle not only interacting with the other particles but also with their images in neighbouring boxes. Particles that leave one side of the box re-enter from the opposite side as their image. In this way the total number of particles in the central box remains constant.

Upon initial inspection such a method would appear to be very computationally intensive requiring the evaluation of a very large number of interacting pairs. However, as explained in section 2.2.1.2.5 the non-bonded cut-off distance is chosen such that each atom 'sees' only one image of all of the other atoms. Hence as long as the box size is more than twice the cut-off distance it is impossible for a particle to interact with any two images of the same particle simultaneously.

***Non-periodic Boundary Methods***

Periodic boundary conditions may not be applicable in some cases. Some systems, such as liquid droplets, inherently contain a defined boundary while others are too big to be simulated efficiently. The alternative is to use a non-periodic boundary where the solvent is included as a shell or 'skin'. If the solvent skin is sufficiently thick then the system is akin to a solute molecule inside a 'drop' of solvent. The use of non-periodic boundaries is susceptible to artificial effects. Periodic boundary conditions are considered the 'safest' way to minimise boundary conditions [Leach 2001] and since their use is applicable to simulations the size of small proteins periodic boundary conditions have been used in all of the molecular dynamics simulations in this work.

### 2.2.1.5 The Solvent Model

The periodic boundary method discussed in section 2.2.1.4 makes it possible to simulate a solute in explicit bulk solvent, however, a model for the solvent is still required. An obvious consequence of including a solvent would be the electrostatic interactions between the solute and the solvent as well as the increased frictional drag. The solvent chosen will, in a fashion analogous to someone working in an experimental lab, depend on the nature of the system being studied. For biological systems such as proteins and nucleic acids the solvent of choice is water. This is the only solvent used in this work and so discussion of solvent models will be restricted to the various water models.

Water models can be divided into three types, consisting of a simple rigid model, a model that incorporates some flexibility and finally models that explicitly account for polarisation and many-body effects. The most common of these is, unsurprisingly, the simple water model. Simple water models use between three and five interaction sites per molecule and the water molecules are kept rigid such that no bond or angle deformations can occur. Commonly encountered models are the TIP3P [Jorgensen *et al.* 1983] and SPC [Berendsen *et al.* 1969] models which use three interaction sites per water molecule. In these models charges are placed on each of the two hydrogens and an opposite balancing charge is placed on the oxygen. The van der Waals interaction between two water molecules is calculated using a single site centred on the oxygen atom. By reducing the degrees of freedom available to the water molecules and thus the number of interactions that require

calculating it is possible to include a significant number of solvent molecules without making the simulation too computationally expensive.

The TIP3P and SPC models differ slightly in their geometries, charges and Lennard-Jones parameters but otherwise are very similar. In this work, the TIP3P water model has been used.

#### 2.2.1.6 Parameters and Atom Types

It should be obvious from the proceeding discussion that a classical force field is only as good as its parameter set. Even if it is intended for calculations on only a small subset of molecules a force field can contain a large number of parameters. Parameterisation of a force field is a non-trivial task requiring a significant amount of effort. In order for a force field to be practical and predictive the parameters must be transferable. Parameters for one type of atom in a particular environment should be sufficient to describe all the atoms in the molecule that exist in similar environments. This leads to the concept of an "atom type", a fundamental concept in force field calculations. Without the ability to transfer parameters between molecules every atom would require unique parameters leading to an infinite parameter set. This would also result in the force field having no predictive power for molecules on which it had not been parameterised. For example the AMBER force field uses the same parameters for all tetrahedral carbons (type CT) regardless of the substituents.

The desire to maintain a small, highly transferable, parameter set means that it is necessary to have a balance between the accuracy with which a specific system is described and the number of parameters that are included. It is also necessary to ensure that the errors in the force field are balanced, spending a long time optimising the bond-stretching terms would be a waste if the van der Waals parameters give rise to large errors.

Although 'guessing' parameters can, if done properly, occasionally give quite reasonable results it is better to construct the force field in such a way that the parameters can be derived from experimental data or quantum mechanical fitting. Typically a test set of molecules is used to perform least squares fitting to experimental and quantum mechanical simulations. Sections 2.2.1.6.1, 2.2.1.6.2 & 2.2.1.6.3 detail the various parameters required for each of the components that make up the AMBER force field potential function (2.3).

***Parameter Accuracy Considerations***

One of the big problems with classical force fields is knowing if the parameters are 'correct'. One aspect of this work, illustrated by §2.3.3 concerned the simulation of LADH using several different but very similar parameters for describing the equilibrium dynamics of the NADH coenzyme within LADH. While all the parameter sets used gave, at face value, very good results it was only by direct comparison with experimental data that the accuracy with which the parameters described the equilibrium fluctuations could be checked. Chapter 4 describes the development of a QM/MM based linear response method which offers a way to validate molecular mechanics force fields for chromophores via direct comparison with experimental steady state spectroscopic data.

#### 2.2.1.6.1 Bonded Interaction Parameters

As described in section 2.2.1.2 the bonded interactions are described by the first 3 terms of equation (2.3). For the bonding term all pairs of atom types that are directly bonded require a force constant $K_r$ and an equilibrium bond length $r_{eq}$ while all permutations of atom types involving angle bending require an angle force constant $K_\theta$ and an equilibrium angle $\theta_{eq}$.

The torsional interactions present more of a problem since experimental information on torsional barriers is generally very sparse. Quantum mechanical calculations are thus widely used to determine torsional potentials by systematically rotating about a bond within a test structure while calculating the energy quantum mechanically at set intervals. The classical expression for the torsional potential is then fitted to this quantum mechanical profile. To account for every different torsional interaction in this way would require a large number of parameters and consequently a very large amount of computational 'effort'. The approach adopted in the AMBER force field is to use, in most cases, generalised torsion parameters that reduce the dihedral dependence from 4 atoms down to the two central atoms.

#### 2.2.1.6.2 Van der Waals Parameters

As explained previously the van der Waals interactions are calculated using a Lennard-Jones 12-6 potential function. This expression requires two parameters

per atom type; the well depth $\varepsilon$ and the minimum energy location $r_m$. Simplification of the parameterisation of the Lennard-Jones interactions is essential since for $N$ atom types a total of $N(N-1)$ different parameters would be required. The AMBER force field employs atom based mixing rules in order to reduce the number of parameters required. By invoking the Lorentz-Berthelot mixing rules the combined Lennard-Jones parameters for two atom types A & B ($r_m^{AB}$, $\varepsilon_{AB}$) can be obtained from individual atomic parameters,

$$\begin{aligned} r_m^{AB} &= \frac{r_m^A + r_m^B}{2} \\ \varepsilon^{AB} &= \sqrt{\varepsilon^A \varepsilon^B} \end{aligned} \tag{2.15}$$

reducing the parameter requirement to one per atom. The 1-4 van der Waals interactions are scaled by a factor of 1/2.0 to reduce the otherwise exaggerated short-range repulsions.

#### 2.2.1.6.3 Electrostatic Interaction Parameters

The derivation of charges in molecular mechanics force fields is a controversial issue in computational chemistry. A number of different approaches have been taken to derive electrostatic interaction parameters for simulating condensed-phase properties. The AMBER 1995 force field, uses a partial atomic charge model described by equation (2.9). Since partial atomic charges cannot be determined empirically the Restrained Electrostatic Potential (RESP) method [Bayly *et al.* 1993; Cornell *et al.* 1993] is used to fit the charges to quantum mechanical potentials.

***The RESP Method***

The RESP method involves calculating the electrostatic potential from the charge density $\rho(\boldsymbol{r})$ of the molecule

$$\phi(\boldsymbol{r}) = \phi^{nuc}(\boldsymbol{r}) + \phi^{elec}(\boldsymbol{r}) = \sum_{j=1}^{N} \frac{Z_j}{|\boldsymbol{r} - \boldsymbol{R}_j|} - \int \frac{d\boldsymbol{r}' \rho(\boldsymbol{r})}{|\boldsymbol{r}' - \boldsymbol{r}|} \tag{2.16}$$

The electrostatic potential is a continuous property and is not easily represented by an analytical function and so it is necessary to derive discrete

points for numerical fitting. In the RESP method the charge density is evaluated using the Hartree Fock *ab initio* method with a 6-31G* basis set[5]. The electrostatic potential is then evaluated using a large number of points (with a density of 1 point / Å$^2$) defined by 4 shells at 1.4, 1.6, 1.8 and 2.0 times the van der Waals radii of the atoms [Besler *et al.* 1990]. Least-squares fitting is then used to derive the atom centred point charges using a two stage approach. In the first stage all of the charges are optimised. The methyl hydrogens are then constrained to have identical charges and then refit while the charges on the other atoms are fixed at their values from stage one. The rationale behind this two stage procedure is that forcing methyl hydrogens to have equivalent charges during the first stage of the fit can adversely affect the heteroatom charges since such hydrogens are not equivalent in a fixed conformation. By using a two stage procedure it is possible to accurately represent both the charges and the dipole moment of the molecule. The AMBER 1995 parameters use partial atomic charges derived from the RESP procedure on a residue by residue basis such that identical protein residues share identical charges. This is more accurate than forcing each atom type to have a specific charge associated with it in an analogous fashion to the bonded parameters.

### 2.2.2 The Energy Surface & Energy Minimisation

It should be apparent from the proceeding discussion that the potential energy surface represented by equation (2.1) is multidimensional for all but the simplest systems. A non-linear molecule consisting of $N$ atoms has a total of $3N-6$ degrees of freedom in Cartesian space. For example, the energy conformation of ethane ($CH_3$-$CH_3$) is a function of 18 internal coordinates or 24 Cartesian coordinates.

The potential energy surface (often called a *hypersurface*) which is a function of the $3N$ Cartesian coordinates thus has a huge dimensionality. Visualisation of the entire energy surface is therefore impossible, however visualisation of small parts of the surface are possible by representing the energy function in terms of only a few coordinates, while keeping the remainder fixed. This allows the energetic effect of varying specific degrees of freedom, such as a single bond stretch, to be calculated and visualised.

---

[5] See section 3.2 for a discussion of *ab initio* theory.

Molecular simulations are typically concerned with potential energy surface minima and saddle points since these are the regions of chemical interest. Energy minima correspond to stable structures while saddle points, linking minima, correspond to transition states. Due to the high dimensionality there may be a huge number of minima on the potential energy surface and while one will be of lowest energy, referred to as the 'global energy minimum', biological systems rarely exist at the global minimum but instead occupy local energy minima. Fortunately, the purpose of molecular dynamics simulations are normally to explore the region of the potential energy surface around the particular minimum that the system inhabits in a physiological environment and so location of the global minimum is not required. An exception to this is the study of protein folding which requires that a large amount of conformational space is sampled. The aim of the molecular dynamics simulations on LADH in this work, discussed in §2.3, were to obtain trajectories that would accurately represent the equilibrium fluctuations experienced by a protein at room temperature and so the importance of locating the lowest energy conformations was minimal. The way in which a protein explores the energy surface due to thermal fluctuations was of far greater interest.

Minimisation methods were still employed, however, for obtaining suitable starting structures for the molecular dynamics simulations. While crystal structures are often assumed to correspond to stable states of the system it is usually necessary to manipulate them (addition of solvent, hydrogens etc.) prior to using them as starting structures for MD simulations. The main use of minimisation in this work was thus to allow the starting structures to essentially 'relax' prior to beginning molecular dynamics equilibration.

Molecular dynamics simulations are covered in detail in §2.2.3. The discussion below covers some of the methods for finding minimum energy structures [Schlegel 1987; Jensen 1999; Leach 2001].

### 2.2.2.1 Energy Minimisation

Given a function $f$ which depends on a series of independent variables $x_1, x_2, ..., x_{n,}$, energy minimisation involves finding the values of the variables such that $f$ has a minimum value. At an energy minimum two conditions are fulfilled:

1) The first derivative of the energy function with respect to all of the variables is zero.

2) The second derivatives are all positive.

In functional form this is:

$$\frac{\partial f}{\partial x_n} = 0; \qquad \frac{\partial^2 f}{\partial x_n^2} > 0 \tag{2.17}$$

If these conditions are fulfilled then a local energy minimum has been located. Thus a procedure for finding a local energy minimum involves finding an analytical expression for the first order differential of the energy with respect to each of the system coordinates. Standard calculus methods can then be used to locate the point at which all of the partial differentials are zero. Unfortunately while this is possible for well behaved functions the energy of molecular systems varies with respect to the coordinates in such a complicated fashion that analytical derivatives cannot be found, a total of $3N$ simultaneous equations would need to be solved. Hence a numerical procedure that gradually changes the coordinates to produce a series of configurations with lower and lower energies must be used.

There are two types of numerical minimisation methods typically used, named according to their use of derivatives. Non-derivative methods, such as the Sequential Univariate Method or the Simplex Method, do not require the evaluation of derivatives. These methods can be computationally demanding and slow to converge but are useful when analytical first derivatives are not available. Since analytical gradients are available for the AMBER force field, for example the expression for the Lennard-Jones force (2.18), derivative based methods are preferred.

$$\frac{\partial v}{\partial r_{ij}} = \frac{6\varepsilon}{r_{ij}}\left[-2\left(\frac{r_m}{r_{ij}}\right)^{12} + \left(\frac{r_m}{r_{ij}}\right)^{6}\right] \tag{2.18}$$

Derivative methods are named according to the highest order derivatives used, with the common methods being either first or second order derivative methods.

### 2.2.2.2 Derivative Minimisation Methods

The use of derivatives in energy minimisation can improve the efficiency with which the local minimum is located. The first derivative, or gradient, indicates the

direction in which the minimum lies and its gradient indicates the steepness of the slope. The negative of the gradient is the force and so by moving each atom in response to the force acting on it the energy of the system can be reduced. While efficient, first order minimisation methods do not take into account the curvature of the energy function and so can sometimes optimise to transition states rather than an energy minimum. Second-order methods use both the first and second derivatives to locate a minimum. The second derivatives provide information about the curvature of the function allowing the points where the energy surface changes direction to be predicted and hence both minima and transition states can be found.

Locating energy minima using derivative techniques is made simpler by writing the energy function as a Taylor series expansion about a point $x_k$ on the potential energy surface,

$$V(x) = V(x_k) + (x - x_k)V'(x_k) + \frac{(x - x_k)^2}{2}V''(x_k) + \ldots \tag{2.19}$$

where $x$ is a vector consisting of $3N$ components in Cartesian space and $x_k$ corresponds to the current system configuration. $V'(x_k)$ is a $3N \times 1$ matrix of vectors, each element of which is the partial first derivative of $V$ with respect to the appropriate coordinate $(\partial V/\partial x_i)$ while $V''(x_k)$ is a $3N \times 3N$ matrix of partial second derivatives with respect to two coordinates $x_i$ and $x_j$ $\left(\partial^2 V/\partial x_i x_j\right)$. $V''(x_k)$ is commonly referred to as the *Hessian* or *force constant* matrix.

Truncating the Taylor expansion at the quadratic term, the first derivative of $V(x)$ is:

$$V'(x) = xV'(x_k) + (x - x_k)V''(x_k) \tag{2.20}$$

As stated in §2.2.2.1 at the energy minimum $(x = x_{\min})$ the first derivative of the energy is zero $V'(x_{\min})$ and so (2.20) can be written in terms of $x_{\min}$ the nearest minimum energy configuration.

$$x_{\min} = x_k - \frac{V'(x_k)}{V''(x_k)} \tag{2.21}$$

If the energy surface were truly quadratic a second-derivative method such as the Newton-Raphson method §2.2.2.2.2 would locate the minimum in a single step. Unfortunately energy surfaces in molecular systems are rarely quadratic and so (2.19) is only an approximation. This has the consequence that the minimum energy has to be located via an iterative method. Typically a starting structure is chosen $(x_k)$, the first and second derivatives are calculated and used to produce a guess of the minimum energy location. This is then used as the starting structure for the second step. This process is then repeated until certain predetermined convergence criteria are met.

#### 2.2.2.2.1 First-order Methods

Since the *Hessian* is a $3N \times 3N$ matrix of second derivatives which, for the above method, has to be stored and then inverted the use of second derivatives in the minimisation can prove very computationally demanding for large systems. For example a procedure that used 8 bytes for each double precision floating point number would require, for a system comprising 1,000 atoms, 69MB of secondary storage to store half of the *Hessian* matrix (since it is symmetric) and half of the inverted *Hessian*. For 5,000 atoms this rises to 1,720MB, for 10,000 atoms this is 6.9GB and for a system the size of LADH (*ca.* 76,500 atoms) this is 392GB!

Fortunately the objective of most minimisations on proteins are simply to "clean up" the starting structure prior to performing molecular dynamics. This implies that an accurate minimum need not be found and so less rigorous minimisation procedures which do not utilise the second derivatives can be used. Two first derivative minimisation methods that are frequently used in molecular dynamics, and have been employed in this work, are the *steepest descent* and *conjugate gradient* methods. These methods gradually change the coordinates of the atoms, slowly moving the system closer and closer to the minimum. The second derivatives are approximated as constants and only the gradient needs to be calculated.

***Steepest Descent***

The steepest descent method works by always moving in the direction parallel to the net force. Essentially this is 'straight' downhill moving in the direction that

is the negative of the gradient. For each configuration a line search is performed along the line defined by the negative of the gradient $(v_k)$ and the minimum found.

$$v_k = -V'(x_k) \tag{2.22}$$

This minimum is then used as the starting point for the next step. A line search is then performed perpendicular to the previous search and the minimum point on this line found. Figure 2-7 shows an example of the path a steepest descent method might take in locating a minimum. The advantage of the steepest descent algorithm is that it is very good when the system is a long way from equilibrium as it quickly eliminates the largest strains. Close to the equilibrium though the method is inefficient since it does not converge well. The steepest descent method is therefore typically used for the first few stages of a minimisation before switching to a more robust method such as a *conjugate gradients* or *Newton-Raphson* method.

***Conjugate Gradients***

The conjugate gradients method attempts to improve on the steepest descent method by removing the biggest inefficiency in its search algorithm. In the steepest descent method each successive move is orthogonal to the previous move. This can result in oscillatory behaviour in narrow valleys. The conjugate gradients method works by performing the line search not along the current gradient, but along a line determined by using a mixture of the current negative gradient and the previous search direction,

$$v_k = -V'(x_k) + \gamma_k v_{k-1} \tag{2.23}$$

where $\gamma_k$ is a scale factor that can vary between methods. Two examples are:

$$\gamma_k = \frac{V'(x_k) \bullet V'(x_k)}{V'(x_{k-1}) \bullet V'(x_{k-1})}\ ^{6} \text{ or } \gamma_k = \frac{\left(V'(x_k) - V'(x_{k-1})\right) \bullet V'(x_k)}{V'(x_{k-1}) \bullet V'(x_{k-1})}\ ^{7} \tag{2.24}$$

[6] Fletcher-Reeves method

[7] Polak and Riviere method

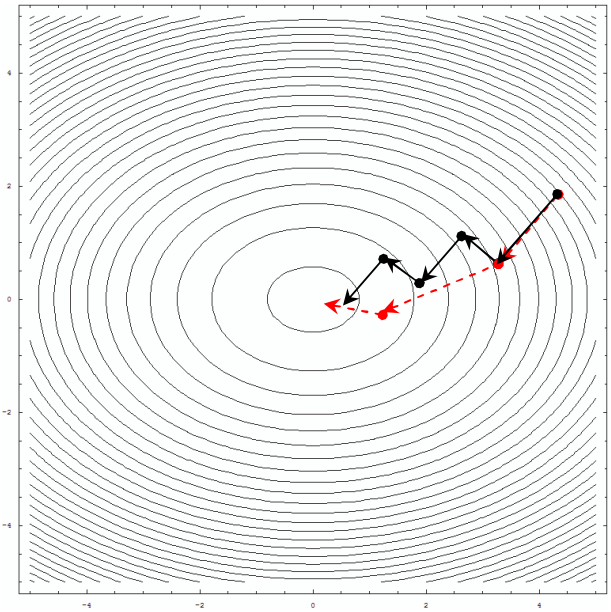


**Figure 2-7** Comparison of potential minimisation paths for steepest descent (solid line) and conjugate gradients (dashed line) minimisation methods.

The use of information from the previous step gives search lines which are "conjugate" to the previous search directions. The distance to be moved at each step is determined in the same way as the steepest descent method. Removing the restriction to always move in an orthogonal direction greatly improves the performance of the minimisation when close to the equilibrium position. Figure 2-7 shows a comparison between the paths that would be taken by the steepest descent and conjugate gradients methods when minimising the function $x^2 + 2y^2$.

#### 2.2.2.2.2 Second-order Methods

Due to computational complexity considerations only first order minimisation methods have been employed in this work, however, a brief description of second order methods is given here for completeness.

Second order minimisation techniques make use of the second derivatives of the energy, the *Hessian* matrix, in order to optimise the minimisation procedure. One example of a second order minimisation routine is the Newton-Raphson Method.

***The Newton-Raphson Method***

The Newton-Raphson Method is the simplest second-order method. It makes use of the *Hessian* matrix, which describes the curvature of the function, in order to solve equation (2.21) iteratively. This makes the method very efficient when close to the minimum but the problems associated with storing and inverting the *Hessian* matrix make this method impractical for large systems.

***Quasi-Newton Methods***

Attempts to reduce the computational effort associated with manipulating the *Hessian* matrix have led to the development of so called Quasi-Newton methods. These methods work by gradually building up the inverse *Hessian* matrix via successive iterations rather than calculating it explicitly. By initiating the minimisation procedure with a reasonable guess of the *Hessian* matrix the performance of Quasi-Newton methods can be improved over the Newton-Raphson method. For the minimisation of systems the size proteins however, first derivative methods are still the only practical method available.

## 2.2.3 The Molecular Dynamics Method

While energy minimisation techniques can generate individual minimum energy conformations of a system that are sometimes sufficient to predict accurately certain properties of that system, they do not provide any time ordered data. In order therefore to investigate the dynamics of a system it is necessary to simulate how the system evolves with time. For this purpose molecular dynamics simulations can be employed.

The requirements for a molecular dynamics simulation are:

1) A suitable starting conformation such as a minimised X-Ray crystal structure.
2) A force field, such as AMBER (§2.2.1.1), describing potential energy with respect to conformation.
3) A suitable algorithm for propagating the system through time.

It is the third requirement that will be discussed here.

#### 2.2.3.1 Newton's Laws of Motion

Classical molecular dynamics simulations generate successive configurations of a system by integrating Newton's laws of motion, specifically Newton's second law,

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{2.25}$$

where $x_i$ is the position of particle $i$ of mass $m_i$, operated on by a force $F_{x_i}$ for a particle moving in one dimension. The force is evaluated from the first differential of the potential energy expression described by the force field.

It is helpful here to consider three examples where Newton's laws of motion can be applied. In the simplest case, a particle experiences no force between collisions. In this example the position of the particle changes by $v_i\delta t$, where $\delta t$ is the time between collisions and $v_i$ is the constant velocity. In this simple situation (2.25) can be solved analytically. A second situation is where the particle experiences a constant force between collisions. An example of this type of motion is that of a charged particle moving through a uniform electric field. The final example is the most complex and corresponds to the motion of molecules. Here the force on the particle depends on its position relative to all the other particles. This type of motion is almost impossible to describe analytically and so (2.25) has to be solved numerically

#### 2.2.3.2 Solving the Differential Equations

Numerous numerical algorithms exist for solving the differential equations that arise from Newton's laws of motion. Examples include the predictor-corrector methods [Gear 1971] and finite difference methods. The AMBER program [Case *et al.* 1999], used in this work, implements finite difference methods for solving the differential equations and so the following discussion will centre on these methods.

##### 2.2.3.2.1 Finite Difference Methods

Finite difference methods work by splitting the integration into many small steps, each separated by a fixed time $\delta t$. At a time $t$ the force on each atom is calculated from the vector sum of the forces arising from its interactions with the

other atoms. From this the acceleration ($d^2x_i/dt^2$) can be calculated using a vector form of equation (2.25). The accelerations are then combined with the position and velocity data for time $t$ in order to calculate the positions and velocities at a time $t+\delta t$. The iteration is repeated until sufficient time steps have been sampled.

Over the period of the time step the force is assumed to be constant. This places severe constraints on the length of a time step. Ideally the time step should be as large as possible since this allows more phase space to be explored for a given computational effort. However, if the time step is too long it is possible for two atoms to approach too closely. The subsequent repulsive force calculated at the next step will thus be very large leading to a large acceleration and hence velocity. This can cause unrealistic oscillations in the system which can rapidly multiply resulting in an unstable molecular dynamics trajectory. In practice the time step is limited to an order of magnitude lower than the highest frequency motion. In flexible molecules these are typically bond stretches involving hydrogen (e.g. C-H *ca.* 10 fs period). Several methods such as SHAKE [Ryckaert *et al.* 1977] and RATTLE [Andersen 1983] exist that, via the use of constraints on high frequency oscillations, allow longer time steps to be used. In this work, however, the aim has been to accurately model the equilibrium dynamics. The use of constraints would have biased the model and so no oscillations were restricted and a time step of 1fs was used for all MD simulations.

***The Leap-frog Algorithm***

There are a number of different schemes for integrating the equations of motion in molecular dynamics. A common one that is used by AMBER 6.0 is the *leap-frog* algorithm [Hockney 1970]. This is a variation on the well known *Verlet algorithm* [Verlet 1967a] and assumes that the positions, velocities and accelerations can be approximated as Taylor series expansions. The *leap-frog* algorithm makes use of the following relationships:

$$r_i\left(t+\delta t\right) = r_i\left(t\right) + \delta t v_i\left(t+\frac{1}{2}\delta t\right) \tag{2.26}$$

$$v_i\left(t+\frac{1}{2}\delta t\right) = v_i'\left(t-\frac{1}{2}\delta t\right) + \delta t a_i\left(t\right) \tag{2.27}$$

where $r_i$ is the position of particle $i$, $v_i$ is the velocity and $a_i$ the acceleration.

The algorithm is implemented by first calculating the velocity $v_i(t+\delta t/2)$ from the velocity at time $t-\delta t/2$ and the acceleration at time $t$ using (2.27). The position $r_i$ at time $t+\delta t$ is then calculated using the position at time $t$ and the velocity just calculated via equation (2.26). The velocities thus 'leap-frog' over the positions and vice versa, hence the name. This method has advantages over the *Verlet* method since it does not require the calculation of the small difference between two large numbers and it explicitly includes the velocity. The disadvantage is that the positions and velocities are not synchronised, meaning that kinetic and potential energies cannot be calculated at the same time. This is easily remedied by using the relationship in equation (2.28) to calculate the velocity at time $t$.

$$v_i(t) = \frac{v\left(t+\frac{1}{2}\delta t\right) - v\left(t-\frac{1}{2}\delta t\right)}{2} \tag{2.28}$$

Despite the disadvantage of having unsynchronised positions and velocities, the simplicity and computational efficiency of the *leap-frog* form of the *Verlet* algorithm have ensured its popularity in molecular dynamics simulations.

#### 2.2.3.3 Molecular Dynamics at Constant Temperature and Pressure

The isothermal-isobaric, constant temperature and pressure (NPT), ensemble is widely used in simulations of biomolecules since it reproduces experimental conditions. In this ensemble the number of molecules, the pressure and the temperature are constant and so it is necessary to monitor both the temperature and pressure over the period of an MD simulation to ensure that they remain constant. The next two sections discuss the methods used in this work for maintaining constant temperatures and pressures.

##### 2.2.3.3.1 Maintaining a Constant Temperature

The temperature is related to the kinetic energy of the system and is kept constant by adjusting the velocities. The Berendsen temperature control scheme [Berendsen *et al.* 1984] works by coupling the system to an external heat bath,

fixed at the desired temperature. The velocities are scaled by equation (2.29) which results in the change in temperature being proportional to the difference between the bath temperature and the system temperature.

$$\lambda_T = \sqrt{1+\frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)}-1\right)} \tag{2.29}$$

Here $\tau$ represents the coupling constant, the magnitude of which determines the degree of coupling between the bath and the system. A large $\tau$ indicates weak coupling, while a small $\tau$ implies that the bath and system are strongly coupled. The velocities are scaled at each step and there is an exponential decay of the system towards the desired temperature. The advantage of this approach is that it allows the system to fluctuate around the desired temperature and eliminates the 'hot solvent, cold solute' phenomenon.

#### 2.2.3.3.2 Maintaining a Constant Pressure

The pressure is maintained in an analogous fashion to the temperature control. The volume is scaled by coupling to a 'pressure' bath [Berendsen *et al.* 1984]. The volume at each step is scaled by a factor $\lambda_p$,

$$\lambda_P = 1-\kappa\frac{\delta t}{\tau_P}\left(P(t)-P_{BATH}\right) \tag{2.30}$$

where the constant $\kappa$, is the isothermal compressibility, the pressure equivalent of the heat capacity, and $\tau_P$ is the pressure coupling constant.

### 2.2.3.4 The Molecular Dynamics Procedure

Molecular dynamics simulations are typically carried out in four stages. The first stage is to construct an initial configuration and minimise it prior to beginning the dynamics. The starting structure is taken preferably from crystallographic (X-ray / Neutron) or NMR data, or if not available from a theoretical model. Hydrogens, where missing, are added at idealised bond lengths and angles, the system is solvated and then minimised to remove any bad contacts created from the hydrogenation and solvation. The second step is to begin the MD simulation and slowly heat the system over a period of say 20 ps from 0 K to the

desired temperature (often 300 K). This allows the system time to move away from any unfavourable conformations in a controlled fashion. The initial velocities are normally assigned to the atoms via random selection from a Maxwell-Boltzmann distribution at the starting temperature. The third step is to run sufficient MD simulation to allow the system to equilibrate at the desired temperature prior to the collection of 'production' data. A very inhomogeneous system, often resulting from the use of a theoretical model, may require a more complex equilibration procedure involving several stages of slow heating, followed by possible heating-cooling cycles and equilibration.

## 2.2.4 Monitoring the Equilibrium

The QM/MM based linear response approach to protein relaxation dynamics developed in this work and discussed in detail in Chapter 4 works by relating the microscopic fluctuations experienced by a system at equilibrium to macroscopic observables. This requires that the protein be properly equilibrated during the production phase of the molecular dynamics simulation, such that the structures recorded accurately reflect the dynamics of the protein at equilibrium. It is necessary therefore to monitor various parameters and system properties during the MD simulations to verify that the system is successfully equilibrated. This section discusses some of the properties that can be calculated and monitored.

### 2.2.4.1 Temperature

Since the Berendsen temperature control scheme (§2.2.3.3.1) is used to maintain the temperature during the simulation it is essential to test that this is maintaining the temperature to within a specified tolerance. Although in the *NPT* ensemble in which the simulations are run the temperature is strictly constant, in practice it fluctuates about a constant mean value. The temperature is directly related to the kinetic energy of the system by,

$$E_K = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m_i} = \frac{k_B T}{2}\left(3N - N_c\right) \tag{2.31}$$

where $\boldsymbol{p}_i$ is the total momentum of particle $i$ of mass $m_i$, $N$ is the number of particles and $N_c$ the number of constrained degrees of freedom. According to the

equipartition theorem each degree of freedom contributes $K_B T/2$ to the energy. The temperature can therefore be calculated from the kinetic energy.

#### 2.2.4.2 Pressure

The pressure is also controlled via coupling to a bath and so it is essential to check that this is similarly constant to within a specified tolerance. The pressure is calculated via the virial theorem and can found using the equation,

$$P = \frac{1}{V}\left[ Nk_B T - \frac{1}{3}\sum_{i=1}^{N}\sum_{j=i+1}^{N} r_{ij} f_{ij} \right] \tag{2.32}$$

where $f_{ij}$ is the force acting between atoms $i$ and $j$. Once the forces on each atom have been evaluated and the temperature calculated it is then possible to calculate the pressure.

#### 2.2.4.3 RMS Deviations

One of the most useful properties for evaluating if the system is at equilibrium is the root mean square deviation, or distance, (RMSD) from the starting structure. Often measured with reference to the protein active site, the alpha carbon backbone or individual residues, the RMS deviation can give a very good indication of whether a simulation is progressing satisfactorily. Wild oscillations in the RMSD or very large deviations from the crystal structure are common signs of a malfunctioning simulation. An RMSD of less than 2 Å for non-side chain atoms would typically be considered satisfactory for a simulation initiated from a crystal structure [van Gunsteren and Mark 1998; Beveridge and McConnell 2000]. In this work the RMSD's have been calculated for the protein backbone, the NADH active site and each individual residue (§2.3.4.2).

The RMS deviation between atoms is given by:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} d_i^2}{N_{atoms}}} \tag{2.33}$$

where $N_{atoms}$ is the number of atoms for which the RMSD is being measured and $d_i$ is the distance between the coordinates of atom $i$ when the two structures are

overlaid. The aim of RMSD fitting is to find relative orientations of the two molecules in which the RMSD is a minimum. A number of different algorithms exist for performing such fits including the numerical method of Ferro and Hermans [Ferro and Hermans 1977] which locates the minimum RMSD in a step wise fashion, and the direct method of Kabsch [Kabsch 1978] which locates the minimum RMSD directly.

# 2.3 Simulations

## 2.3.1 The Model

LADH [Brändén *et al.* 1975] is a very well characterised enzyme which, with the use of $NAD^+$ as a cofactor, catalyses the reversible oxidation of a wide range of alcohols to their corresponding aldehydes. It has a molecular weight of 80,000 and is composed of a dimer of two identical parts as reported in the X-ray structure [Ramaswamy *et al.* 1997]. Each subunit of the dimer binds one molecule of NADH and two Zn(II) ions. One zinc is in the active site and is associated with the aldehyde oxygen of the substrate or substrate analogue, while the other zinc performs a structural role. At pH 7, the equilibrium position of the chemistry favours $NAD^+$ and alcohol [Shearer *et al.* 1993]. Hence crystallisation of LADH containing aldehyde and the NADH coenzyme at physiological pH is inherently difficult. For this reason *n*-cyclohexylformamide was used as an inhibitor. The system of study was therefore composed of the LADH-NADH-*n*-cyclohexyl formamide ternary complex.

## 2.3.2 Software and Platform Details

All MD simulations discussed in this section were performed using the *AMBER 6* [Case *et al.* 1999] suite of programs implementing the Cornell *et al.* [Cornell *et al.* 1995] force field. The simulations were run on commodity Intel and AMD PC clusters designed and built from scratch specifically for this project. All machines ran custom installations of RedHat Linux versions 7.1, 7.2, 7.3 or 8.0[8].

[8] http://www.redhat.com/

The protein models were built using xLEaP [Schafmeister *et al.* 1995], and subsequent minimisations and molecular dynamics simulations performed with SANDER. Partial atomic charges for non-standard units were derived using the RESP procedure [Bayly *et al.* 1993]. *Gaussian 98* [Frisch *et al.* 1998] was used to generate the electrostatic potential required for the RESP fits. Trajectory analysis was performed using CARNAL and PTRAJ. Graphical representations were produced using VMD 1.8.1 [Humphrey *et al.* 1996], MOLSCRIPT [Kraulis 1991], Raster3D [Merritt and Murphy 1994], DSViewer Pro v5.0[9], PovChem v2.1.1[10] and WinPov v3.1g[11].

## 2.3.3 Computational Methodology

### 2.3.3.1 Amber Force Field Parameterisation

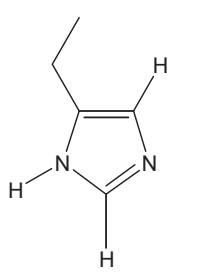
The initial starting structure of LADH included *n*-cyclohexyl formamide as an inhibitor and was obtained from the crystal structure of horse liver LADH by Ramaswamy *et al.* [Ramaswamy *et al.* 1997] (RCSB Protein Data Bank[12] ID: 1LDY). This X-ray structure is reasonably well defined with a resolution of 2.5 Å, reported *R*-value of 0.207, and has all residues and heavy atoms present as well as 1,012 crystallographic waters. A higher resolution structure of LADH containing NADH at 1.66 Å, *R*-value: 0.185, (RCSB Protein Data Bank[12] ID: 3BTO) published by Ramaswamy *et al.* [Cho *et al.* 1997] using 3-butylthiolane-1-oxide as the inhibitor was also available at the time this work was conducted. While a higher resolution structure would have been more desirable there is published evidence that thiol compounds quench nicotinamide fluorescence [Cheng and Lek 1992] and therefore it was decided to use the lower resolution structure.

[9] http://www.accelrys.com/

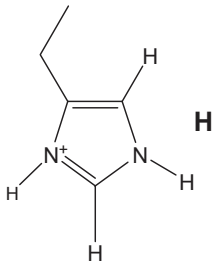
[10] http://www.chemicalgraphics.com/PovChem/

[11] http://www.povray.org/

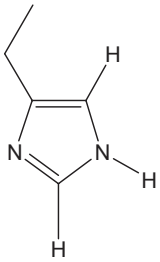
[12] http://www.rcsb.org/




HID$^{\delta}$




HIP




HIE$^{\varepsilon}$

| Residue | LADH Histidine Tautomeric States |
|---|---|
| His34 | HID |
| His51 | HIP |
| His67 | HID |
| His105 | HIE |
| His138 | HID |
| His139 | HID |
| His348 | HIP |

**Figure 2-8** The tautomers of histidine, and the tautomeric states of the histidine residues in liver alcohol dehydrogenase.

The asymmetric unit contains two dimers in close proximity. The first of the two dimers (chains A and D in the pdb) was used as the initial configuration. The extracted coordinates were prepared for the molecular dynamics simulations using the xLEaP module [Schafmeister *et al.* 1995] which forms part of the AMBER 6.0 suite of molecular dynamics programs [Case *et al.* 1999]. On the basis of an examination of the hydrogen bonds by Ryde [Ryde 1995], histidine residues 34, 67, 138 and 139 were defined as being protonated on the $\delta$ nitrogen (HID), histidine residue 105 on the $\varepsilon$ nitrogen (HIE), and histidine residues 51 and 348 on both nitrogens (HIP) and thus positively charged (Figure 2-8). All of the six cysteine zinc ligands, Cys46, Cys97, Cys100, Cys103, Cys111, and Cys174, were assumed to be negatively charged (CYM) [Ryde 1995; Ryde 1996a], while the remaining eight cysteine ligands were assumed to be neutral (CYS). Because the structural zinc is a large distance from the NADH active site (~ 16 Å), it was defined as a hard sphere of charge 2+ and was not explicitly bonded to any of the four surrounding cysteine residues. The catalytic zinc atom, however, is close to the NADH active site and so was parameterised explicitly.

**Table 2-1** Details of NADH parameterisation for the various LADH MD simulations.

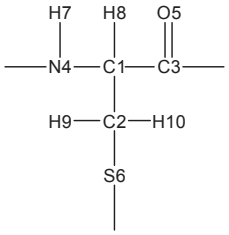
| *Simulation* | *NICH Parameterisation* | *Ave Protein RMSD*[a] | *Ave NADH RMSD*[b] |
|---|---|---|---|
| **Simulation 1** | All nicotinamide parameters based on Pavelites *et al.* [Pavelites *et al.* 1997] published parameters. | 0.9794 Å | 0.4170 Å |
| **Simulation 2** | Nicotinamide parameters derived by analogy with existing residues but Pavelites *et al.* [Pavelites *et al.* 1997] parameters used for the $CONH_2$ moiety torsions. | 0.9497 Å | 0.4414 Å |
| **Simulation 3** | All nicotinamide parameters derived by analogy with existing residues. | 0.9672 Å | 0.5566 Å |
| **Simulation 4** | All nicotinamide parameters derived by analogy with existing residues but $CONH_2$ moiety torsions set to zero. | 0.9261 Å | 0.4883 Å |

[a]Calculated for the alpha carbons (CA) forming the protein backbone over all equilibration steps, Å
[b]Mean RMSD for both NADH residues averaged across all NADH atoms.

Four-coordinate zinc is believed to be more stable than five-coordinate [Ryde 1994; Ryde 1995; Ryde 1996b], so the catalytic zinc atom was bonded to the two nearby cysteine residues (residues 46 and 174), the oxygen of the *n*-cyclohexyl formamide inhibitor (residue 378), and the $\varepsilon$ nitrogen of the nearby histidine (residue 67).

Force field parameters not defined in the standard Amber force field [Cornell *et al.* 1995] were obtained from existing published parameters or by analogy with the standard residues. For the catalytic zinc and attached ligand, these were taken from published computational studies of the catalytic zinc ion by Ryde [Ryde 1995] (Table 2-2 & Table 2-3), while those for *n*-cyclohexyl formamide were derived from comparison with the parameters for standard amino acids (Table 2-4). Other parameters for the zinc ions have been published [Clementi *et al.* 1980; Vedani *et al.* 1986; Vedani and Huhta 1990; Hoops *et al.* 1991; Stote and Karplus 1995] but they either lack data for some of the zinc ligand parameters or contain what is believed to be unreliable data. The parameters developed by Ryde were considered to be more reliable since they were developed specifically for the LADH active site.

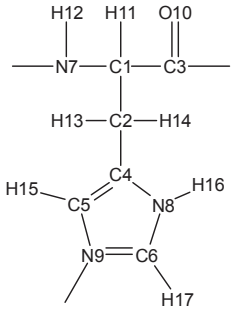
**Table 2-2** Non-standard charges for catalytic zinc bound cysteines and histidine




Catalytic Zinc Bound Cysteines, Atom Types and Charges

| *atom* | *type* | *charge CYA* | *charge CYB* | *atom* | *type* | *charge CYA* | *charge CYB* |
|---|---|---|---|---|---|---|---|
| C1 | CT | +0.01995 | -0.00767 | S6 | SH | -0.54271 | -0.56920 |
| C2 | CT | -0.09617 | -0.20192 | H7 | H | +0.25298 | +0.29837 |
| C3 | C | +0.63948 | +0.60835 | H8 | H1 | +0.13291 | +0.11811 |
| N4 | N | -0.48926 | -0.46442 | H9 | H1 | +0.09345 | +0.12708 |
| O5 | O | -0.59208 | -0.57108 | H10 | H1 | +0.09345 | +0.12708 |

a CYA = residue 46; CYB = residue 174.




Catalytic Zinc Bound Histidine, Atom Types and Charges

| *atom* | *type* | *charge* | *atom* | *type* | *charge* |
|---|---|---|---|---|---|
| C1 | CT | -0.096079 | O10 | O | -0.498915 |
| C2 | CT | -0.088318 | H11 | H1 | +0.135500 |
| C3 | C | +0.529423 | H12 | H | +0.209094 |
| C4 | CC | -0.034573 | H13 | HC | +0.092702 |
| C5 | CV | -0.029486 | H14 | HC | +0.092702 |
| C6 | CR | -0.034573 | H15 | H4 | +0.188232 |
| N7 | N | -0.335847 | H16 | H | +0.272147 |
| N8 | NA | -0.038078 | H17 | H5 | +0.161570 |
| N9 | NB | -0.242772 | | | |

**Table 2-3** Ryde *et al.* [Ryde 1995] parameters used for catalytic zinc and attached residues.

Catalytic Zinc, Missing Bond Parameters

| *bond type* | $K_r$ *(kcal/mol Å²)* | $r_{eq}$ *(Å)* |
|---|---|---|
| Zn-SH | 81.7 | 2.293 |
| Zn-NB | 85.5 | 2.101 |
| Zn-O | 45.1 | 2.359 |

Catalytic Zinc, Missing Angle Parameters

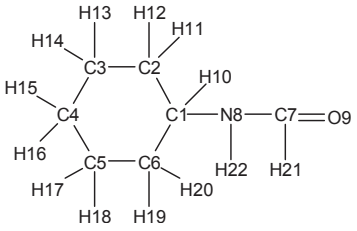
| *angle type* | $K_\theta$*(kcal/mol rad²)* | $\theta_{eq}$ *(°)* |
|---|---|---|
| SH-Zn-O | 31.0 | 104.4 |
| SH-Zn-SH | 49.5 | 146.5 |
| SH-Zn-NB | 30.4 | 103.0 |
| CT-SH-Zn | 18.9 | 111.6 |
| CV-NB-Zn | 44.5 | 123.6 |
| NB-Zn-O | 16.1 | 97.9 |
| CR-NB-Zn | 44.5 | 123.6 |
| Zn-O-C | 19.9 | 131.2 |

Catalytic Zinc, Missing Torsion Parameters

| *torsion type* | $V_n$ [a] | $\gamma$ *(°)* | *n* | *torsion type* | $V_n$ [a] | $\gamma$ *(°)* | *n* |
|---|---|---|---|---|---|---|---|
| SH-Zn-NB-CV | -2.82 | 0.00 | 2.0 | CV-NB-Zn-O | 2.82 | 0.00 | 2.0 |
| SH-Zn-NB-CR | -2.82 | 0.00 | 2.0 | NB-Zn-O-C | -0.77 | 0.00 | 3.0 |
| SH-Zn-SH-CT | -0.03 | 0.00 | 2.0 | CR-NB-Zn-O | 2.82 | 0.00 | 2.0 |
| SH-Zn-O-C | -0.77 | 0.00 | 3.0 | Zn-O-C-H2 | -1.82 | 0.00 | 2.0 |
| CT-SH-Zn-NB | -0.03 | 0.00 | 3.0 | Zn-O-C-N | -1.82 | 0.00 | 2.0 |
| CT-SH-Zn-O | -0.03 | 0.00 | 3.0 | | | | |

[a] *kcal/mol*

**Table 2-4** Non-standard MD parameters and charges for the *n*-cyclohexyl formamide inhibitor




CXF Atom Types and Charges

| *atom* | *type* | *charge* | *atom* | *type* | *charge* |
|---|---|---|---|---|---|
| C1 | CT | -0.047093 | H12 | HC | +0.014835 |
| C2 | CT | -0.013558 | H13 | HC | +0.003181 |
| C3 | CT | +0.005532 | H14 | HC | +0.003181 |
| C4 | CT | +0.018703 | H15 | HC | +0.001954 |
| C5 | CT | -0.021191 | H16 | HC | +0.001954 |
| C6 | CT | -0.038157 | H17 | HC | +0.016289 |
| N8 | N | -0.251314 | H18 | HC | +0.016289 |
| C7 | C | +0.298933 | H19 | HC | +0.029578 |
| O9 | O | -0.387555 | H20 | HC | +0.029578 |
| H10 | H1 | +0.143500 | H21 | H2 | +0.069656 |
| H11 | HC | +0.014835 | H22 | H | +0.240420 |

Missing CXF Bond Parameters

| *bond type* | $K_r$ *(kcal/mol Å²)* | $r_{eq}$ *(Å)* |
|---|---|---|
| C-H2 | 367.0 | 1.080 |

Missing CXF Angle Parameters

| *angle type* | $K_\theta$ *(kcal/mol rad²)* | $\theta_{eq}$ *(°)* |
|---|---|---|
| N-C-H2 | 35.0 | 113.2 |
| O-C-H2 | 35.0 | 122.3 |

Four different parameter sets were used for the NADH residue. The NADH was defined as a non-standard residue, and in all four cases, the adenine dinucleotide, diphosphate and the two ribose sections were parameterised from analogy with the standard amino acid residue library in AMBER 6.0. The adenine moiety of NADH was taken from the adenine residue defined in the Amber force field. The two ribose and attached phosphate groups were copied from the default template for a DNA backbone, while the nicotinamide section was parameterised in a different way for each of the four classical MD simulations conducted.

The four MD simulations conducted are summarised in Table 2-1. In the first simulation, the parameters used for the nicotinamide (NICH) section of NADH were based on those published by Pavelites *et al.* [Pavelites *et al.* 1997] for the CHARMM force field [Mackerell *et al.* 1998]. The AMBER and CHARMM force fields are sufficiently similar that parameters can be easily transferred between them. The Pavelites parameters, reproduced in Table 2-7 & Table 2-8, were derived with "strong emphasis [was placed] on the successful reproduction of experimental geometries and crystal data". For this reason it was expected that incorporation of all of the Pavelites NICH parameters would lead to the best reproduction of the equilibrium fluctuations of LADH+NADH. In the second, third and fourth simulations the NICH parameterisation was based on simple analogy using the atom types and parameters given in Table 2-5, Table 2-6, Table 2-7, and Table 2-8. The difference between the last three simulations concerned the approach towards the parameterisation of the NICH amide group torsions. Initial trial CIS calculations (§3.3.2) showed the molecular orbitals involved in the UV/VIS absorption and emission of LADH to be centred on the nicotinamide moiety within the NADH co-factor. For this reason it was believed that getting the equilibrium fluctuations of the nicotinamide correct would be essential to obtaining good results. The vast majority of the fluctuations in this moiety concern the amide arm of the nicotinamide (Atoms C17, C18, O14, N7, H67 & H73 in Table 2-5). The fluctuations of this amide arm are largely controlled by the torsion parameters and so three different sets of amide torsion parameters were employed in the second, third and fourth MD simulations to investigate how the parameter accuracy would affect the linear response results discussed later. In the second MD simulation the NICH group parameters were found by analogy but the Pavelites torsional parameters were used for the amide arm of the nicotinamide (Table 2-8). This was expected to estimate the amide arm fluctuations reasonably well but the

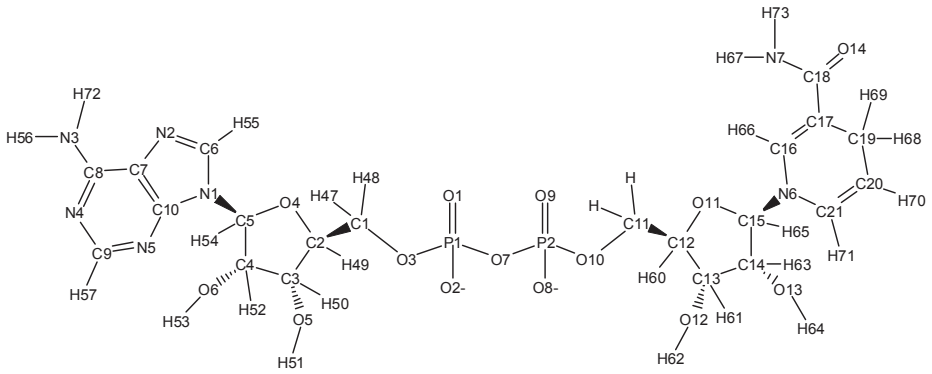
remainder of the nicotinamide ring would be kept in a slightly too rigid geometry. This was expected to result in a small underestimation of the magnitude of the equilibrium fluctuations. In the third simulation all of the NICH group parameters were found by analogy with the amide arm being modelled as a truncated peptide linkage. The torsion parameters here were far 'stiffer' than the Pavelites parameters and it was believed that this would result in the equilibrium dynamics being under estimated. In the final simulation the amide arm torsional barriers were set to zero. This was expected to result in the magnitude of the equilibrium dynamics being overestimated and potentially allow the exploration of parts of phase space that would not normally be accessible to the NICH group.

Equilibrium angles that were missing from the default Amber force field were found by conducting *ab initio* geometry optimisations on dihydronicotinamide truncated with a methyl group on the ring nitrogen (atom N6 in Table 2-5) using Hartree-Fock theory [Roothaan 1951] with the SV 3-21G* [Binkley *et al.* 1980] basis set implemented in the program Gaussian 98 [Frisch *et al.* 1998].

In all four cases the hydrogen atoms, not present in the crystal structure, were added at idealised bond lengths and angles as necessary to fulfil valence requirements. The LADH dimer was then neutralised by the addition of four $Cl^-$ ions distributed in a shell around the LADH dimer using a Coulombic potential on a 1 Å grid. The whole system was then solvated using the TIP3P [Jorgensen *et al.* 1983] water model by placing it in a box of dimensions 83.51 x 128.12 x 86.55 Å containing approximately 21,000 water molecules to yield a system of 76,723 atoms, 11,464 from the LADH solute and 65,259 from the water, with an average density of 0.849 g $cm^{-3}$ (Figure 2-9).

Atom-centred point charges for the non-standard residues were derived for the ground state using density matrices obtained from Hartree-Fock SV 6-31G* [Binkley *et al.* 1980] calculations in accordance with the RESP method [Bayly *et al.* 1993] for the Amber force field.

**Table 2-5** Non-standard charges for NADH, common to all four MD simulations




NADH Charges Common to All Four MD Simulations

| *atom* | *charge* | *atom* | *charge* | *atom* | *charge* | *atom* | *charge* |
|---|---|---|---|---|---|---|---|
| P1 | +0.942262 | N4 | -0.740753 | C16 | -0.109296 | H57 | +0.073409 |
| O1 | -0.751784 | C9 | +0.512274 | C17 | -0.106001 | H58 | +0.101734 |
| O2 | -0.751784 | N5 | -0.683413 | C18 | +0.651558 | H59 | +0.101734 |
| O3 | -0.393707 | C10 | +0.287632 | O14 | -0.617698 | H60 | +0.074556 |
| C1 | +0.018558 | O7 | -0.389268 | N7 | -0.864974 | H61 | +0.062116 |
| C2 | +0.064412 | P2 | +1.085819 | C19 | +0.003584 | H62 | +0.443688 |
| O4 | -0.408559 | O8 | -0.772815 | C20 | -0.257754 | H63 | +0.160887 |
| C3 | +0.291031 | O9 | -0.772815 | C21 | -0.066683 | H64 | +0.393708 |
| O5 | -0.656568 | O10 | -0.430385 | H47 | +0.082207 | H65 | +0.163577 |
| C4 | +0.173011 | C11 | +0.027083 | H48 | +0.082207 | H66 | +0.116496 |
| O6 | -0.605416 | C12 | +0.060268 | H49 | +0.110945 | H67 | +0.374674 |
| C5 | +0.148989 | O11 | -0.391394 | H50 | +0.074971 | H68 | +0.035385 |
| N1 | +0.003740 | C13 | +0.249868 | H51 | +0.399284 | H69 | +0.035385 |
| C6 | +0.074646 | O12 | -0.688179 | H52 | +0.013375 | H70 | +0.130701 |
| N2 | -0.594769 | C14 | +0.085110 | H53 | +0.390589 | H71 | +0.162816 |
| C7 | +0.164769 | O13 | -0.624314 | H54 | +0.120526 | H72 | +0.374708 |
| C8 | +0.558570 | C15 | +0.037068 | H55 | +0.182914 | H73 | +0.374674 |
| N3 | -0.777933 | N6 | +0.011980 | H56 | +0.374708 | | |

**Table 2-6** NADH atom types for MD simulations 1 to 4

NADH Atom Types Common to All Four MD Simulations

| *atom* | *type* | *atom* | *type* | *atom* | *type* | *atom* | *type* | *atom* | *type* |
|---|---|---|---|---|---|---|---|---|---|
| P1 | P | C5 | CT | O7 | OS | O13 | OH | H56 | H |
| O1 | O2 | N1 | N* | P2 | P | C15 | CT | H57 | H5 |
| O2 | O2 | C6 | CK | O8 | O2 | H47 | H1 | H58 | H1 |
| O3 | OS | N2 | NB | O9 | O2 | H48 | H1 | H59 | H1 |
| C1 | CT | C7 | CB | O10 | OS | H49 | H1 | H60 | H1 |
| C2 | CT | C8 | CA | C11 | CT | H50 | H1 | H61 | H1 |
| O4 | OS | N3 | N2 | C12 | CT | H51 | HO | H62 | HO |
| C3 | CT | N4 | NC | O11 | OS | H52 | H1 | H63 | H1 |
| O5 | OH | C9 | CQ | C13 | CT | H53 | HO | H64 | HO |
| C4 | CT | N5 | NC | O12 | OH | H54 | H2 | H65 | H2 |
| O6 | OH | C10 | CB | C14 | CT | H55 | H5 | H72 | H |

NADH Atom Types Specific to MD Simulation 1

| *atom* | *type* | *atom* | *type* | *atom* | *type* | *atom* | *type* | *atom* | *type* |
|---|---|---|---|---|---|---|---|---|---|
| N6 | NF | C17 | CF | C20 | CF | H67 | HE | H70 | HF |
| N7 | NE | C18 | CE | C21 | CG | H68 | HG | H71 | HF |
| O14 | OE | C19 | CH | H66 | HF | H69 | HG | H73 | HE |
| C16 | CG | | | | | | | | |

NADH Atom Types Specific to MD Simulations 2,3 & 4

| *atom* | *type* | *atom* | *type* | *atom* | *type* | *atom* | *type* | *atom* | *type* |
|---|---|---|---|---|---|---|---|---|---|
| N6 | N* | C17 | CM | C20 | CM | H67 | HZ | H70 | HA |
| N7 | NZ | C18 | CZ | C21 | CM | H68 | HC | H71 | H4 |
| O14 | OZ | C19 | CT | H66 | H4 | H69 | HC | H73 | HZ |
| C16 | CM | | | | | | | | |

**Table 2-7** NADH missing bond and angle parameters

Pavelites *et al.* [Pavelites *et al.* 1997] NADH, Missing Bond Parameters

| *bond type* | $K_r$ *(kcal/mol Å²)* | $r_{eq}$ *(Å)* | *bond type* | $K_r$ *(kcal/mol Å²)* | $r_{eq}$ *(Å)* |
|---|---|---|---|---|---|
| NE-CE | 302.0 | 1.480 | CF-CG | 420.0 | 1.320 |
| NE-HE | 434.0 | 1.010 | CF-HF | 374.0 | 1.090 |
| NF-CG | 420.0 | 1.320 | CG-HF | 374.0 | 1.090 |
| OE-CE | 860.0 | 1.230 | CH-HG | 340.0 | 1.090 |
| CE-CF | 302.0 | 1.480 | CT-NF | 337.0 | 1.475 |
| CF-CH | 225.0 | 1.490 | | | |

Analogy Based NADH, Missing Bond Parameters

| *bond type* | $K_r$ *(kcal/mol rad²)* | $r_{eq}$ *(Å)* | *bond type* | $K_r$ *(kcal/mol rad²)* | $r_{eq}$ *(Å)* |
|---|---|---|---|---|---|
| CZ-OZ | 570.0 | 1.229 | CZ-NZ | 490.0 | 1.335 |
| CM-CZ | 410.0 | 1.444 | NZ-HZ | 434.0 | 1.010 |

Pavelites *et al.* [Pavelites *et al.* 1997] NADH, Missing Angle Parameters

| *angle type* | $K_\theta$ *(kcal/mol rad²)* | $\theta_{eq}$(°) | *angle type* | $K_\theta$ *(kcal/mol rad²)* | $\theta_{eq}$(°) |
|---|---|---|---|---|---|
| HE-NE-HE | 35.0 | 120.0 | CH-CF-CG | 43.5 | 128.0 |
| HE-NE-CE | 35.0 | 120.0 | CH-CF-HF | 30.0 | 116.0 |
| NE-CE-OE | 85.0 | 128.5 | CG-NF-CG | 20.0 | 114.0 |
| NE-CE-CF | 85.0 | 113.0 | HF-CG-NF | 42.0 | 119.0 |
| CE-CF-CG | 5.0 | 107.8 | CG-CF-HF | 42.0 | 116.0 |
| CE-CF-CH | 125.0 | 124.2 | HG-CH-HG | 35.0 | 109.5 |
| OE-CE-CF | 85.0 | 118.5 | NF-CT-H2 | 50.0 | 109.5 |
| CF-CH-CF | 125.0 | 108.0 | CT-NF-CG | 70.0 | 121.2 |
| CF-CH-HG | 55.0 | 110.1 | CT-CT-NF | 50.0 | 109.5 |
| CF-CG-HF | 42.0 | 119.0 | OS-CT-NF | 50.0 | 109.5 |
| CF-CG-NF | 60.0 | 122.0 | | | |

Analogy Based NADH, Missing Angle Parameters

| *angle type* | $K_\theta$ *(kcal/mol rad²)* | $\theta_{eq}$(°) | *angle type* | $K_\theta$ *(kcal/mol rad²)* | $\theta_{eq}$(°) |
|---|---|---|---|---|---|
| CT-CM-HA | 35.0 | 118.0 | CM-CM-CZ | 63.0 | 120.7 |
| CM-CT-CM | 63.0 | 110.2 | CM-CZ-NZ | 70.0 | 117.3 |
| CM-N*-CM | 70.0 | 117.8 | OZ-CZ-NZ | 80.0 | 122.9 |
| CM-CZ-OZ | 80.0 | 125.3 | CZ-NZ-HZ | 30.0 | 120.0 |
| CT-CM-CZ | 70.0 | 119.7 | HZ-NZ-HZ | 35.0 | 120.0 |

**Table 2-8** NADH missing torsion parameters

Pavelites *et al.* [Pavelites *et al.* 1997] NADH, Missing Torsion Parameters

| *torsion type* | $V_n$ [c] | $\gamma$ (°) | *n* | *torsion type* | $V_n$ [c] | $\gamma$ (°) | *n* |
|---|---|---|---|---|---|---|---|
| HE-NE-CE-OE | 2.50 | 180.0 | 2.0 | CE-CF-CH-CF | 1.00 | 180.0 | 3.0 |
| NE-CE-CF-CG | 1.10 | 180.0 | 1.0 | CF-CG-NF-CG | 0.10 | 180.0 | 2.0 |
| NE-CE-CF-CG | 1.95 | 180.0 | 2.0 | CF-CH-CF-HF | 1.00 | 180.0 | 3.0 |
| NE-CF-CF-CH | 0.50 | 180.0 | 2.0 | CF-CH-CF-CG | 4.00 | 180.0 | 3.0 |
| NE-CF-CF-CH | 0.35 | 180.0 | 3.0 | CH-CF-CG-HF | 0.10 | 180.0 | 2.0 |
| NE-CE-CF-CH | 0.40 | 0.0 | 6.0 | CH-CF-CG-NF | 0.10 | 180.0 | 2.0 |
| OE-CE-CF-CG | 0.30 | 0.0 | 1.0 | HG-CH-CF-CG | 1.00 | 180.0 | 3.0 |
| OE-CE-CF-CG | 1.95 | 180.0 | 2.0 | HG-CH-CF-HF | 1.00 | 180.0 | 3.0 |
| OE-CE-CF-CH | 1.00 | 190.0 | 2.0 | HF-CF-CG-HF | 0.40 | 180.0 | 2.0 |
| OE-CE-CF-CH | 1.00 | 0.0 | 3.0 | HF-CF-CG-NF | 7.00 | 180.0 | 2.0 |
| OE-CE-CF-CH | 0.40 | 0.0 | 6.0 | HF-CG-NF-CG | 7.00 | 180.0 | 2.0 |
| CE-CF-CG-HF | 0.10 | 180.0 | 2.0 | CF-CE-NE-HE | 2.50 | 180.0 | 2.0 |
| CE-CF-CG-NF | 2.50 | 180.0 | 2.0 | X-CG-NF-X | 7.40 | 180.0 | 2.0 |
| CE-CF-CH-HG | 1.00 | 180.0 | 3.0 | | | | |

Analogy Based NADH, Missing Torsion Parameters (Simulations 2, 3 & 4)[a]

| *torsion type* | $V_n$ [c] | $\gamma$ (°) | *n* | *torsion type* | $V_n$ [c] | $\gamma$ (°) | *N* |
|---|---|---|---|---|---|---|---|
| NZ-CZ-CM-CT | 0.50[b] | 180.0 | 2.0 | CM-CM-CZ-NZ | 1.10[b] | 180.0 | 1.0 |
| NZ-CZ-CM-CT | 0.35[b] | 180.0 | 3.0 | CM-CM-CZ-NZ | 1.95[b] | 180.0 | 1.0 |
| NZ-CZ-CM-CT | 0.40[b] | 0.00 | 6.0 | CM-CM-CZ-OZ | 0.30[b] | 180.0 | 1.0 |
| OZ-CZ-NZ-HZ | 2.50[b] | 180.0 | 2.0 | CM-CM-CZ-OZ | 1.95[b] | 180.0 | 2.0 |
| OZ-CZ-CM-CT | 1.00[b] | 180.0 | 2.0 | CZ-CM-CM-H4 | 0.10 | 180.0 | 2.0 |
| OZ-CZ-CM-CT | 1.00[b] | 0.00 | 3.0 | CZ-CM-CM-N* | 2.50 | 180.0 | 2.0 |
| OZ-CZ-CM-CT | 0.40[b] | 0.00 | 6.0 | CZ-CM-CT-CM | 1.00 | 180.0 | 3.0 |
| CM-CZ-NZ-HZ | 2.50[b] | 180.0 | 2.0 | CZ-CM-CT-HC | 1.00 | 180.0 | 3.0 |

[a] Simulation 2 used the torsion parameters given in the Pavelites *et al*. table above for all torsions involving NADH atoms C17, C18, O14, N7, H67 & H73

[b] These values were set to zero for simulation 4.

[c] *kcal/mol*

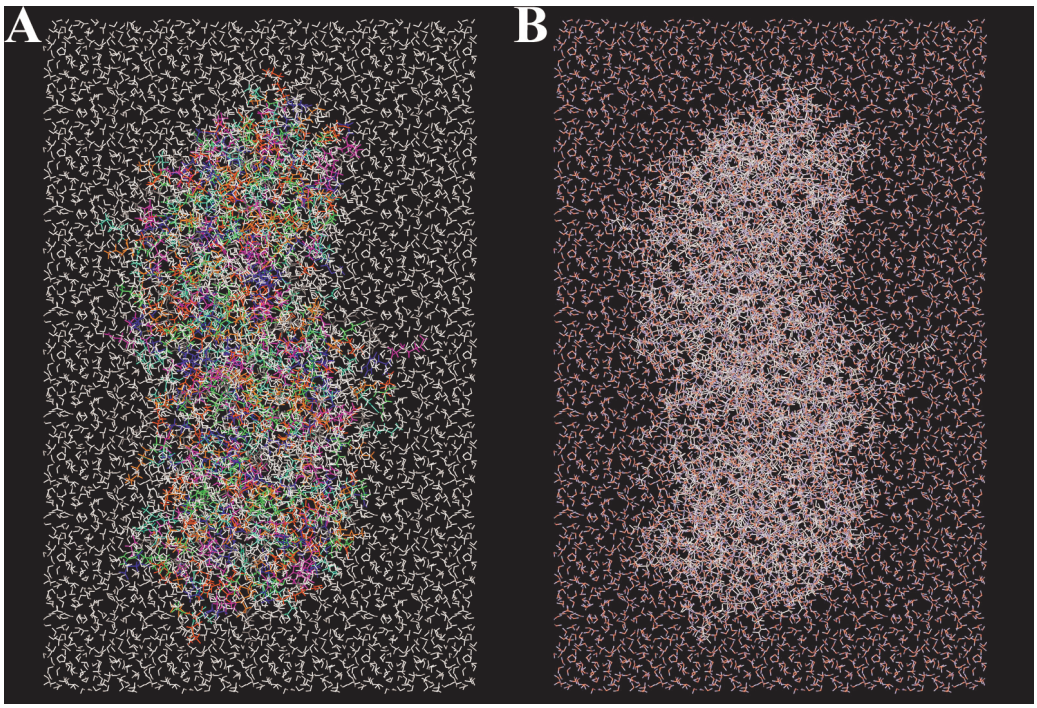



**Figure 2-9** 1LDY starting structure after charge neutralisation and solvation. Coloured by residue (**A**) and by charge (**B**).

### 2.3.3.2 Molecular Dynamics Protocol

Four separate but procedurally identical molecular dynamics (MD) simulations were performed on the dimer system in a bath of water molecules for the four force field configurations detailed above. The protocol used for the MD simulations was as follows. The system was first subjected to 200 steps of steepest descent followed by 300 steps of conjugate-gradient minimisation on just the solvent molecules followed by 200 steps of steepest descent and 800 steps of conjugate-gradient minimisation on the whole system to alleviate incorrect van der Waals contacts created by hydrogenation and solvation of the system. Based on the results of the calculations discussed in §2.3.3.2.1 an atom based non-bonded cut-off of 12 Å was used for this and all subsequent simulations. The system was subjected to 20 ps of slow heating from 0 to 300 K using independent scaling of the velocities of the solute and solvent following the method of Berendsen [Berendsen *et al.* 1984] to control the temperature. In all the MD simulations, constant pressure periodic boundary conditions using the particle mesh Ewald (PME)

[Essmann *et al.* 1995] method were employed. The integration time step was set at 1 fs, and all interactions were calculated at every step. No atoms had their positions fixed or their motions damped.

After slow heating, the system was equilibrated for 100 ps at 300 K. Equilibration was deemed to have been successfully obtained when the RMSD of the protein α-carbons was reasonably small (< 1.0 Å) and fluctuated by less than 10 % over a timescale of approximately 20 ps. An energy gap production run was then performed for the subsequent 10 ps with the complete system coordinates being recorded every 2 fs, resulting in a trajectory of 5,000 coordinate sets (*ca.* 10 GB of data). These structures were later used to quantum mechanically evaluate the ground to 1st excited state energy gap as a function of time (§3.3.3). This entire process was then repeated for each of the 4 parameter sets.

#### 2.3.3.2.1 Cut-off Determination

As mentioned in section 2.2.1.2.5 a non-bonded cut-off can be used to speed up the MD simulation at the expense of accuracy. The choice of cut-off is therefore an important factor in carrying out an MD simulation. Figure 2-10 shows the variation in total system energy after 100 steps of steepest descent minimisation of simulation 1 as a function of cut-off distance. It can be seen that the total energy converges quickly and that by 12 Å the gain in going to large cut-offs is small and no longer justifies the extra computational cost. Hence for this work a cut-off of 12 Å was utilised.

## 2.3.4 Results and Discussion

The results discussed here form the basis of the simulations used in the work published as [Walker *et al.* 2002]. The figures at the end of this chapter summarise the results of the four MD simulations conducted.

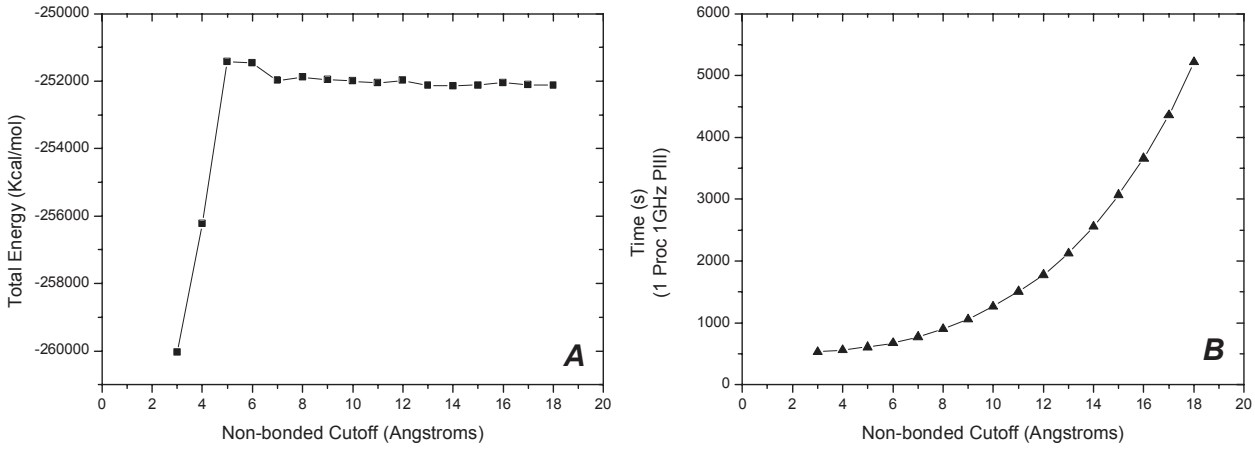



**Figure 2-10** Energy of LADH after 100 steps of steepest descent minimisation (**A**), and processor time (2 x 1GHz PIII) in seconds (**B**), as a function of cut-off distance

### 2.3.4.1 Comparing the Models

Figure 2-14 and Figure 2-15 show the calculated temperatures, pressures and system volumes for all four simulations. It can be seen from Figure 2-14 that all four simulations show steady temperatures over the equilibration and production phases indicating that the Berendsen thermostat was performing correctly. All simulations showed average temperatures of 300.0 ± 0.25 K with standard deviations of 0.714, 0.675, 0.710 and 0.706 K respectively. Figure 2-15 shows that the *NPT* ensemble has been successfully enforced with the volume being varied as necessary to maintain a constant pressure over the equilibration and production phases of all four MD simulations.

The RMS deviations, from the crystal structure, of the α-carbons are given by the black lines in Figure 2-16 and Figure 2-17. It has been said that "[an] overall RMS of *ca.* 1 Å is well within the norm for simulations; beyond 2 Å would be alarming".[13] In all four simulations there are no wild oscillations in the α-carbon RMSDs and all have averages of less than 1 Å over the equilibration period. Over the production phase the simulations (Figure 2-17) show good stability in the α-carbon RMSDs with averages of 1.037, 1.190, 0.896 and 0.837 Å respectively. Interestingly simulations 3 and 4, where the nicotinamide parameters were

[13] Ross. W. S., http://amber.scripps.edu/tutorial/streptavidin/index.html

obtained by analogy, show the lowest RMSDs. There is no obvious explanation for this other than the suggestion that using analogy based parameters for the nicotinamide resulted in a reduction in stress on the surrounding protein and an overall lowering of the α-carbon RMSDs. However, the simulations are too short to make any firm conclusions about this. A lower RMSD would typically be associated with a *better* simulation but as we see by the results in Chapter 4 the RMS deviations do not directly reflect the accuracy with which equilibrium fluctuations are reproduced.

Figure 2-18 and Figure 2-19 show the average RMS deviations from the crystal structure over the 10 ps production runs for each of the residues in the LADH dimer. It can be seen that the vast majority of residues have RMSDs below 2 Å. Figure 2-20 and Figure 2-21 show ribbon representations of the average production phase structures of all four simulations. The NADH coenzyme and surrounding active site residues are shown in cyan while residues with individual average RMSDs greater than 3 Å are shown in yellow. Figure 2-22 shows superimposed snapshots of the structures of LADH for each of the four simulations. Each representation consists of 50 structures produced at 200 fs intervals over the course of the 10 ps production runs. Residues with average RMSDs greater than 3 Å are highlighted in white. It is apparent from Figure 2-20, Figure 2-21 and Figure 2-22 that the disordered residues are all at the solvent interface of the protein and sufficiently far from the active site to be judged not to greatly affect its dynamics.

#### 2.3.4.2 Modelling of the NADH Residue

The RMS deviations from the crystal structure of all atoms of the two NADH residues (Res ID 375 [red line] & 753 [blue line]) in LADH are shown in Figure 2-16 and Figure 2-17. Figure 2-23 shows a ball and stick representation of the structures of the NADH residues (Res ID 375 [a] & 753 [b]) averaged over the 10 ps production runs for each of the four simulations while Figure 2-24 and Figure 2-25 show structural snapshots of the NADH residues taken at 200 fs intervals over the 10 ps production runs.

Over the equilibration and production phases of simulation 1 we see that the two NADH residues have very similar RMS deviations averaging around 0.417 Å with very similar standard deviations of 0.0577 Å and 0.0596 Å. The structures are

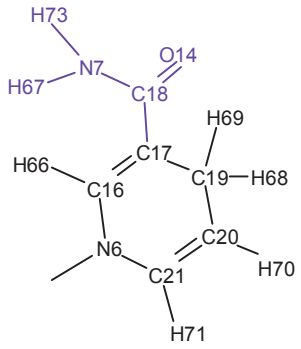



**Figure 2-11** 2D representation of the nicotinamide component of NADH. The amide moiety is highlighted in blue.

also very similar suggesting that there are very few differences in the dynamics of the two residues. In the second simulation, however, when only the $CONH_2$ moiety of the NADH residues were based on the published parameters of Pavelites *et al.* [Pavelites *et al.* 1997] the average RMSD of the two NADH residues increased to 0.441 Å and differences in the RMS deviations can be observed over the equilibration phase of the simulation. NADH residue 753 consistently has a larger standard deviation of 0.0916 Å in its RMSD than NADH residue 375 which has a standard deviation of 0.0536 Å in its RMSD. This would suggest that residue 753 shows a larger degree of structural fluctuations during the simulation. Over the duration of the production run, however, the two structures are very similar and the RMS deviations are also equivalent.

In simulations 3 and 4 it was expected that the equilibrium fluctuations would not be successfully reproduced since the force field parameters used were not ideal for the system. In simulation 3 the whole of the NADH residue was parameterised by analogy. Importantly the $CONH_2$ moiety of the nicotinamide section (Figure 2-11) of NADH was modelled using parameters based on a peptide backbone. Section 2.3.4.3 illustrates that this is a massive over estimation of the torsional barriers of this group and the result is that the amide moiety is forced into a planar geometry and the fluctuations of this group are reduced. The overall average RMSD for the NADH residues increased to 0.557 Å suggesting that artificially restricting the fluctuations increases the strain on the rest of the NADH residue so increasing its deviation from the crystal structure. This is apparent from

the difference in the average NADH structure, from the first two simulations, shown in Figure 2-23.

In the final simulation the NADH amide torsions were switched off entirely. This had the expected effect of increasing the fluctuation of this group. Initially this was not expected to have a marked effect on the results since the amide moiety torsions in the fully parameterised simulation (simulation 1) were already very small (see §2.3.4.3). The net result of switching the amide torsions off, however, was that it lowered the barrier to rotation around the N-C-C to a point where parts of phase space that were not available, for energetic reasons, to the amide moiety became available. While this resulted in a lowering of the average NADH RMSD to 0.4883 Å it also, as discussed in the next two sections, gave rise to what appear to be physically unrealistic structures. This is further highlighted by the steady state spectra results shown in Chapter 4.

### 2.3.4.3 Modelling of the NADH Amide Torsions

In order to illustrate parts of the potential energy surface for rotation about the N7-C18-C17 bond a series of normal mode frequency calculations were run for various CT-CM-CZ-NZ and CT-CM-CZ-OZ torsion parameters (Figure 2-12) utilising a program written for automating the generation of AMBER 6.0 coordinate and topology files[14]. Comparison with a HF [Roothaan 1951] SV 6-31G** [Petersson *et al.* 1988; Petersson and Al-Laham 1991] frequency calculation on gas phase nicotinamide, which gives a prediction of 74.25 $cm^{-1}$ for rotation about the C17-C18-O14-N7 bond, allows the fluctuations predicted by different dihedral parameters to be compared. From Figure 2-12 it can be seen that while not strictly correct, values of zero for the barrier heights ($V_n$ in equation (2.7)) of dihedrals CT-CM-CZ-NZ and CT-CM-CZ-OZ give a rotational frequency prediction that is closer to the *ab initio* frequency prediction than do peptide analogous dihedral barrier heights (10.5 kcal $mol^{-1}$ for each dihedral torsion angle). The Pavelites *et al.* dihedral barrier heights are 0.40 kcal $mol^{-1}$ and 1.00 kcal $mol^{-1}$ respectively and give C17-C18-O14-N7 rotational frequency predictions that are closer to the zero barrier height and *ab initio* predictions than to the peptide barrier height predictions.

[14] The software used for generating Figure 2-12 is available on the supplemental CD-ROM provided with this work.

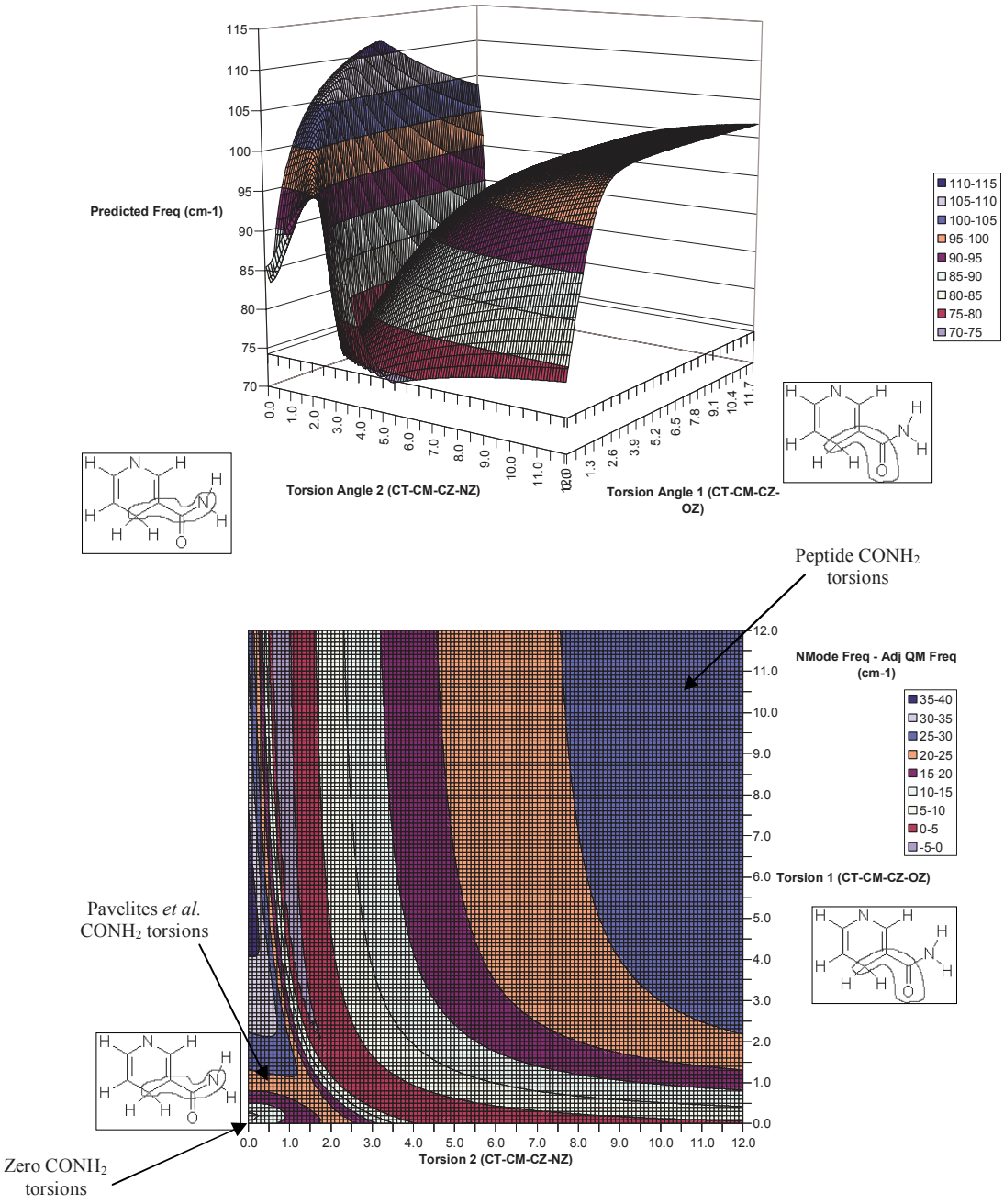



**Figure 2-12** 3D plot (**a**) of Nmode predicted C17-C18-O14-N7 rotation frequency and 2D contour plot (**b**) of the difference between the Nmode predicted rotational frequency and a HF SV 6-31G** prediction as a function of CT-CM-CZ-OZ and CT-CM-CZ-NZ torsion parameters. HF SV 6-31G** scaled prediction is 74.25 $cm^{-1}$.

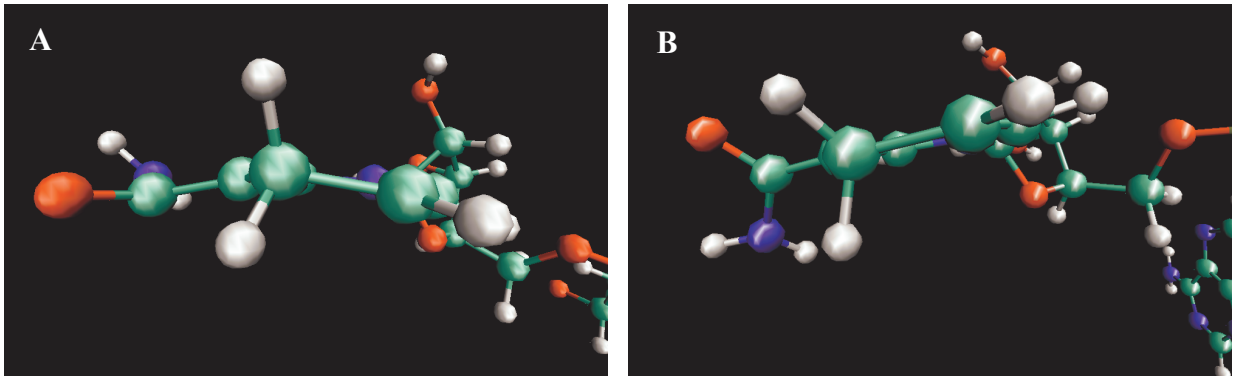



**Figure 2-13** Typical structure from 10 ps production run for NADH residue **A**) 753, **B**) 375, showing the two distinctly different conformations observed for the nicotinamide amide moiety.

Thus using either the values predicted by Pavelites *et al.* or no dihedral barriers for the amide moiety of NADH would be expected to result in a better representation of the equilibrium fluctuations of the NADH amide moiety than would the use of barrier heights based on a protein peptide linkage.

It should be noted, however, that this calculation provides only a qualitative representation of the torsional potential since only a small subset of the torsions affecting the $CONH_2$ group have been investigated and all calculations here have been carried out on a small fragment of the entire system in a gas phase environment.

#### 2.3.4.4 Active Site Structure and Asymmetry

One of the most interesting observations of the MD simulations is that when the NADH amide torsions were switched off the dynamics of the two NADH active sites were found to be distinctly different over the timescale of the simulation. This led to two distinctly different predictions for the Stokes shift discussed in Chapter 4. While simulations 1 to 3 showed only minor differences in the structure of the two NADH residues over the period of the production run, simulation 4 showed a distinct structural difference between the two sites. During the heating phase of the MD simulation it was found that the two NADH residues became in-equivalent with the amide arm of the NADH either residing mainly in the same plane as the nicotinamide ring (residue 753), or out of the plane of the nicotinamide ring (residue 375) (Figure 2-13). It is also apparent from Figure 2-23 that the two NADH structures are markedly different in simulation 4 (Orange). From

illustration **d** in Figure 2-24 and Figure 2-25 it is apparent that while the overall degree of fluctuation of the nicotinamide moiety is increased in simulation 4 there is a marked difference in the size of the fluctuations of the $CONH_2$ moiety between the two NADH residues. Residue 375 shows much greater fluctuation than residue 753. This is as a result of the two residues occupying different parts of phase space and as discussed in Chapter 4 this leads to two different results being obtained from the linear response implementation described in this work.

This difference in the NADH structures would suggest that the two sites of the LADH dimer do not react in an identical fashion. There has been considerable controversy in the literature with regards to the symmetry of dehydrogenases, concerning cooperativity of ligand binding and half-of-the-sites versus full-site reactivity. Although the LADH enzyme binds two moles of NADH with equal affinity, it has been claimed that only one site reacts during the turnover of the enzyme [Bernhard *et al.* 1970]. There is also evidence of interaction between sites in other dehydrogenases, particularly in glyceraldehyde 3-phosphate dehydrogenase where the NADH binding is found to be cooperative [Conway and Koshland Jr 1968; Schlessinger and Levitzki 1974; Gennis 1976], though half-of-the-sites reactivity is only observed for artificial substrates. However, structural [Eklund *et al.* 1982] and NMR [Anderson and Dahlquist 1982] data indicates that the subunits are equivalent when substrates are bound and kinetic data indicates that both sites do react simultaneously [Shore and Gutfreund 1970; Hadorn *et al.* 1975; Kvassman and Pettersson 1976; Weidig *et al.* 1977; Ehrig *et al.* 1991].

Although classical molecular dynamics simulations have recently demonstrated asymmetry in enzymes [Hughes *et al.* 2003] the fact that asymmetry of the NADH coenzyme is only observed in simulation 4, where the nicotinamide amide torsions were deliberately set to inappropriate values, suggests that in this case the asymmetry is an artefact of the simulation protocol rather than a true representation of the dynamics of LADH. While this does not rule out active site asymmetry for the NADH - *n*-cyclohexyl formamide inhibitor ternary complex the lack of evidence for asymmetry in the first three simulations would suggest that it does not exist in this system.

For the purposes of this work the QM/MM calculations were carried out on both NADH sites, independently for each of the simulations and then the results were combined when simulating the absorption and emission spectra in Chapter 4.

### 2.3.5 Conclusions

Based on the RMS deviations, temperatures, pressures and observation of the molecular dynamics trajectories it was concluded that all four simulations were suitable for use in the QM/MM evaluation of ground to excited state energy gaps discussed in Chapter 3. The four different approaches to the NADH parameterisation all resulted in stable MD trajectories that after 120 ps of simulation appear to be sufficiently equilibrated to accurately reproduce the equilibrium fluctuations of LADH. The distinct structural differences observed between the dynamics of the two NADH residues within the LADH dimer appear to be isolated to the fourth simulation and are an artefact of the force field parameterisation. More in-depth molecular dynamics simulations of LADH could serve as a basis for investigation into asymmetry within LADH but this was outside of the scope of this work. The slight differences in NADH dynamics observed for the other three simulations are attributed to the active sites simply exploring different parts of normally accessible phase space over the timescale of the simulations. There is no evidence that the structures explored during the first three MD simulations are physically unrealistic. It was decided, however to calculate the ground to excited state energy gaps for both NADH residues in order to maximise the degree of phase space sampled when producing the spectra shown in Chapter 4

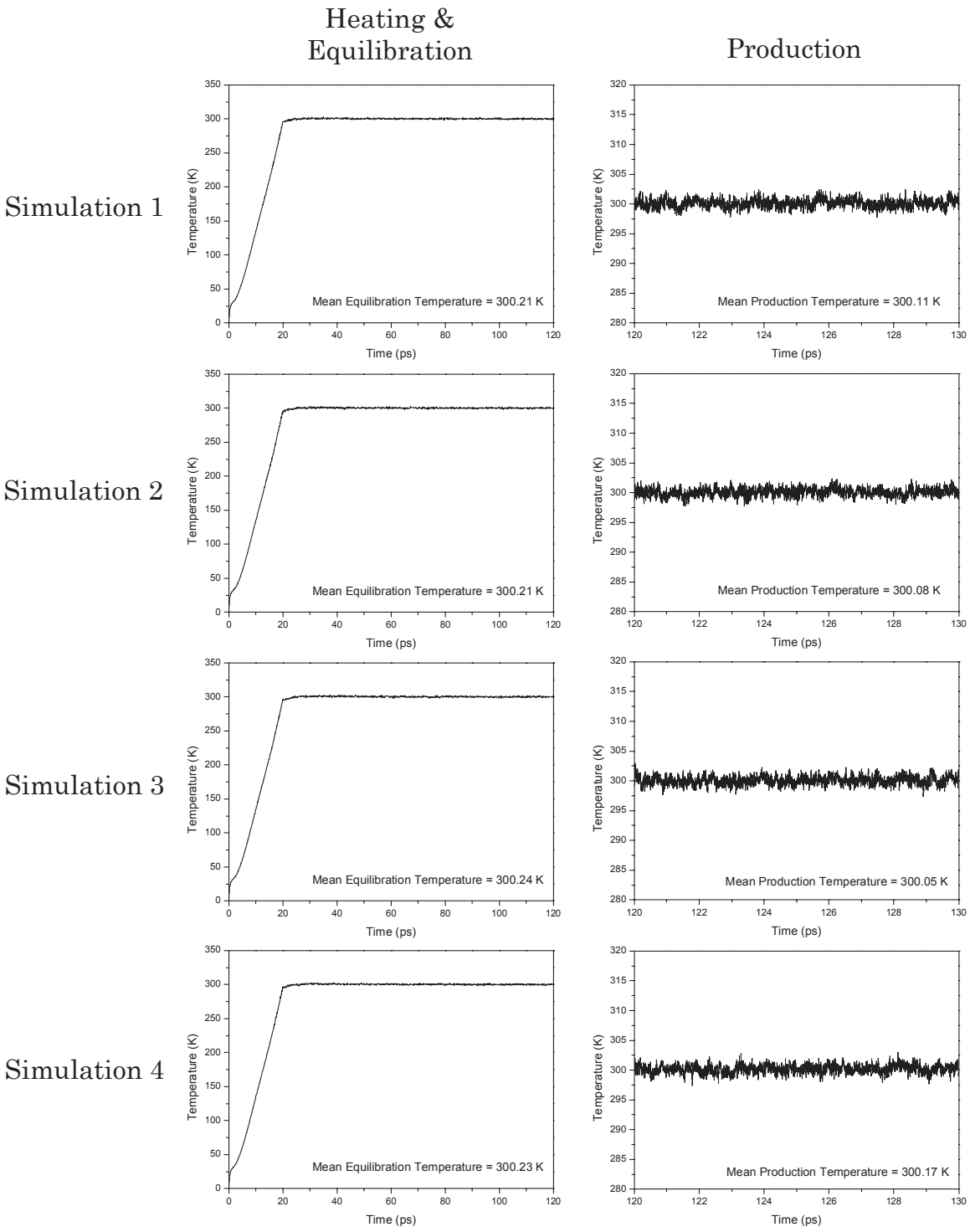



**Figure 2-14** Calculated simulation temperature, as a function of time, for heating, equilibration and production phases of simulations 1 - 4.

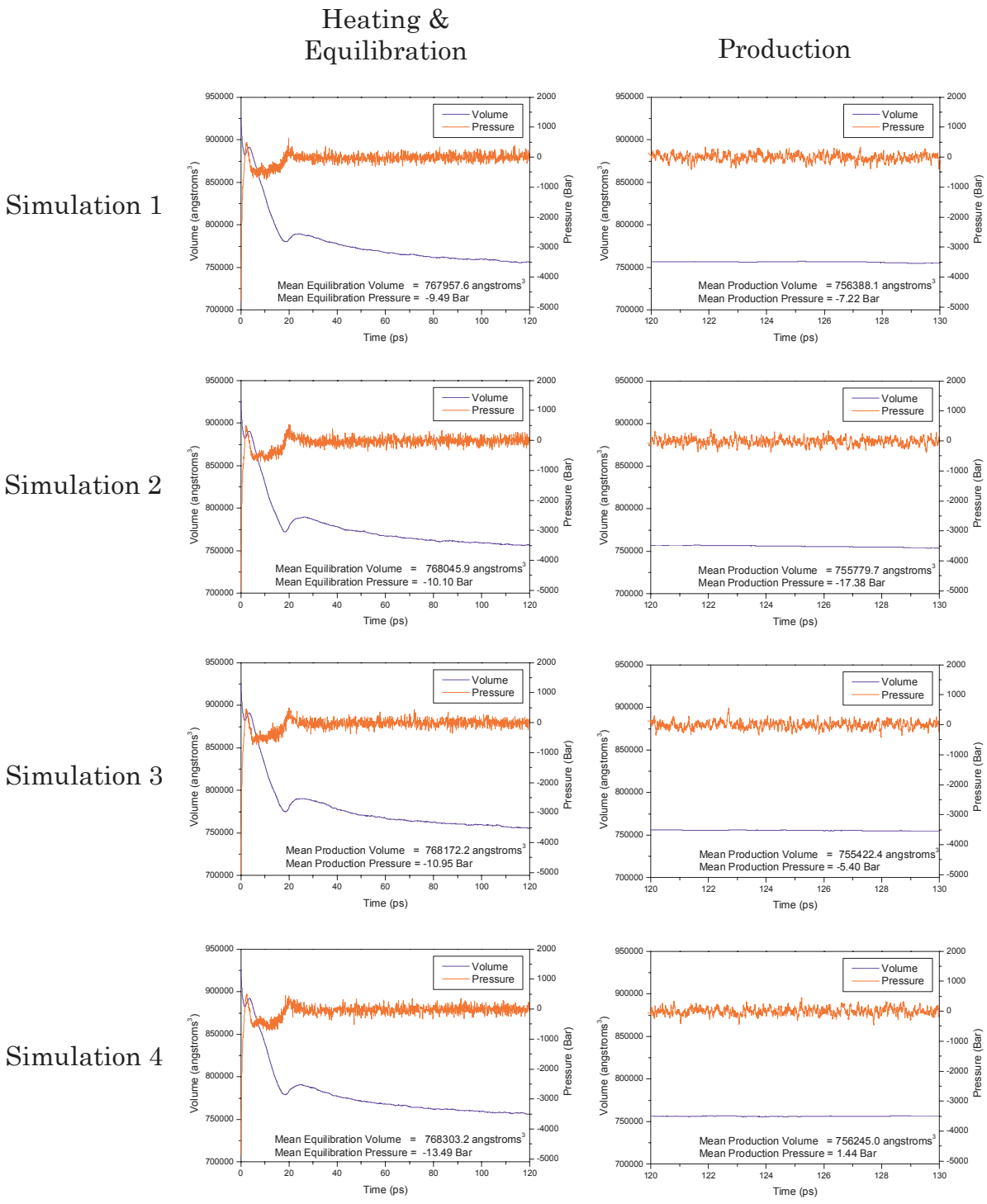



**Figure 2-15** Calculated simulation volume (left hand scale) and pressure (right hand scale), as a function of time, for heating, equilibration and production phases of simulations 1 - 4.

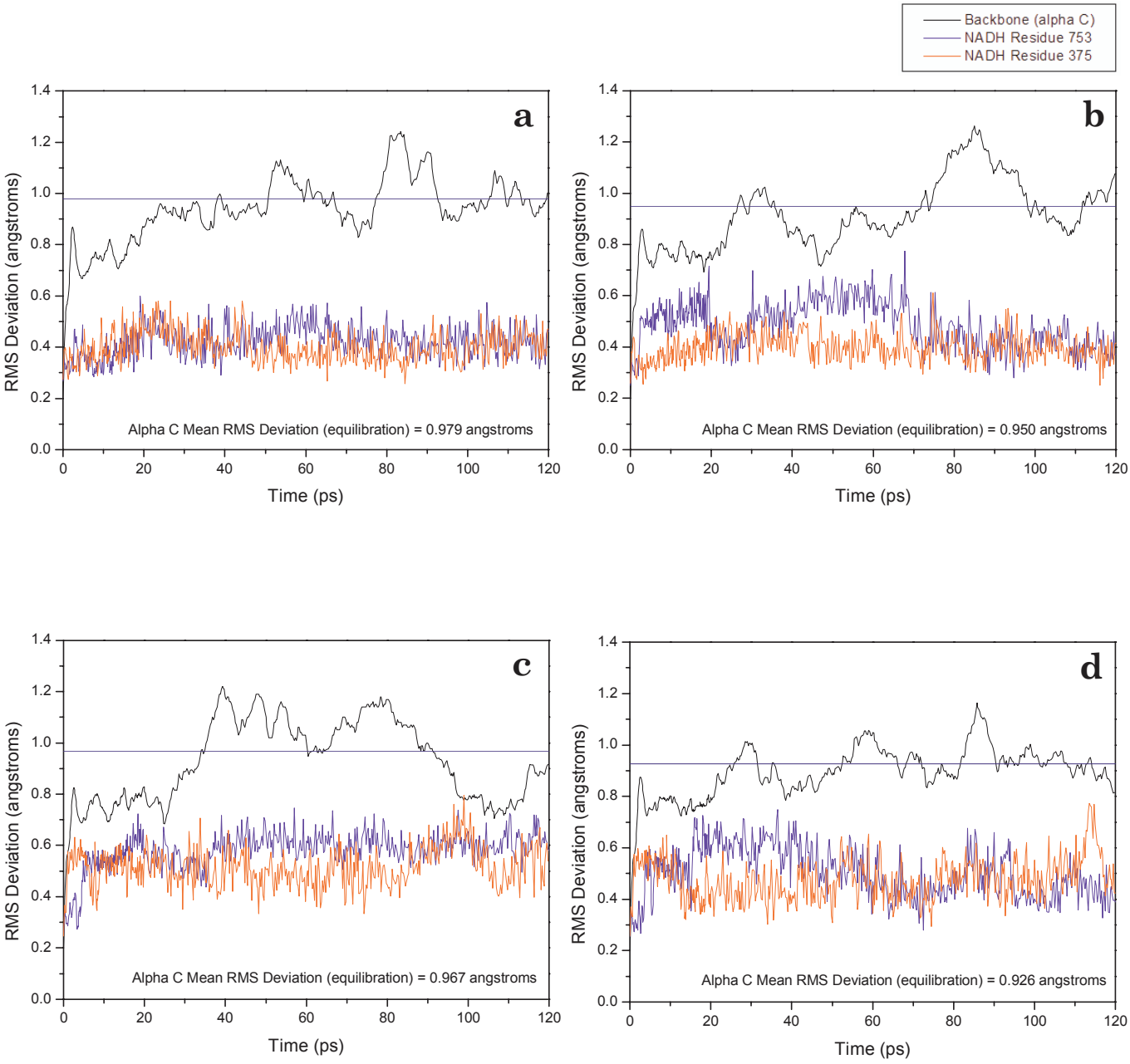



**Figure 2-16** RMS deviations of LADH α-carbons (black line) and all atoms of NADH residues (753 = blue line, 375 = red line) over 120 ps of heating and equilibration runs for **a.** Simulation 1, **b.** Simulation 2, **c.** Simulation 3, **d.** Simulation 4.

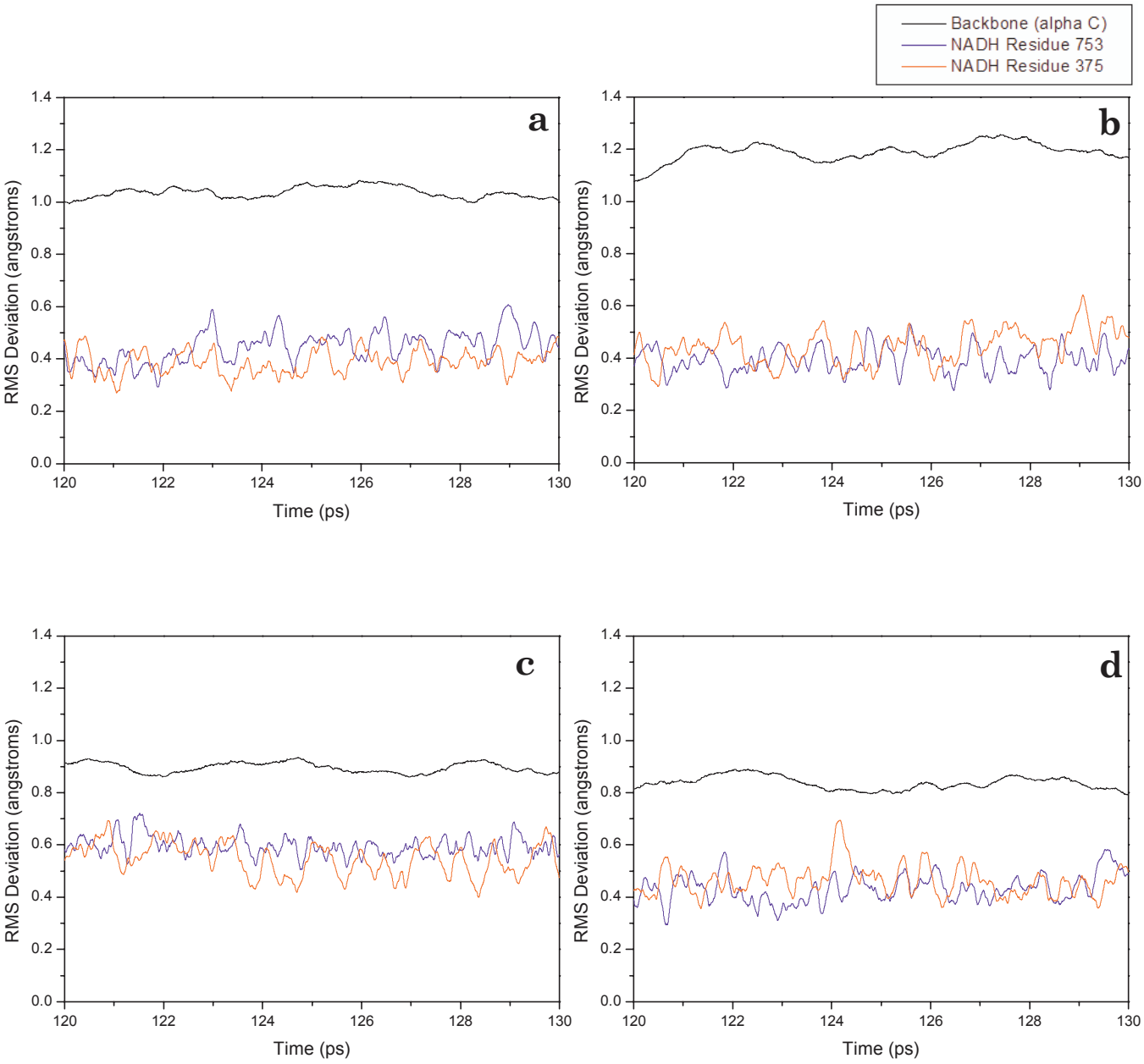



**Figure 2-17** RMS deviations of LADH α-carbons (black line) and all atoms of NADH residues (753 = blue line, 375 = red line) over 10 ps production runs for **a.** Simulation 1, **b.** Simulation 2, **c.** Simulation 3, **d.** Simulation 4.

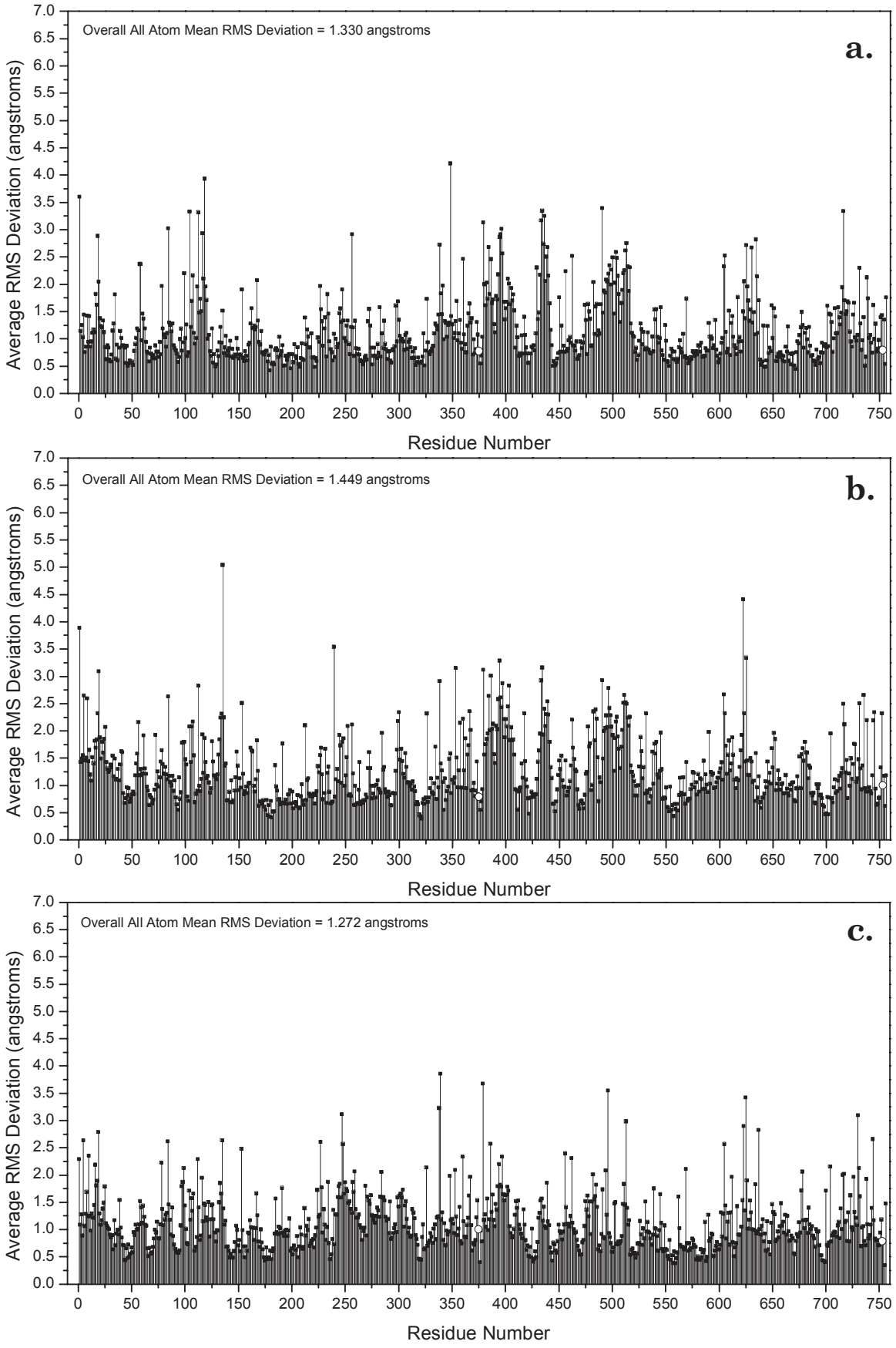



**Figure 2-18** Mean RMS deviations of residues over 10 ps production runs for **a**. Simulation 1, **b**. Simulation 2, and **c**. Simulation 3. NADH residues are labelled (o).

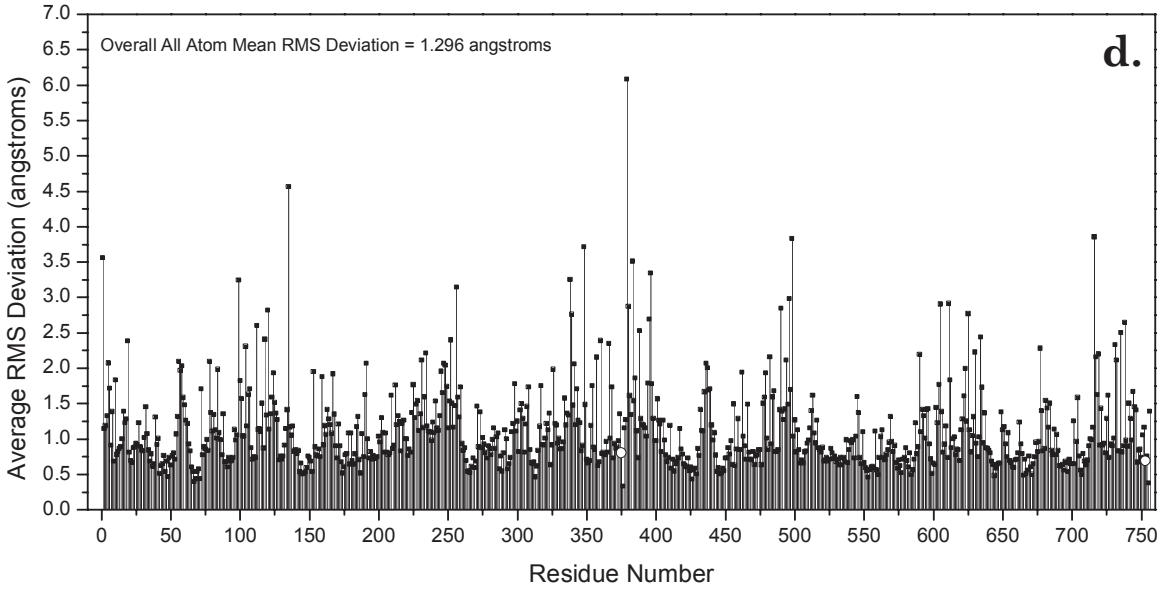



**Figure 2-19** Mean RMS deviations of residues over 10 ps production run for **d**. Simulation 4. NADH residues are labelled (o).

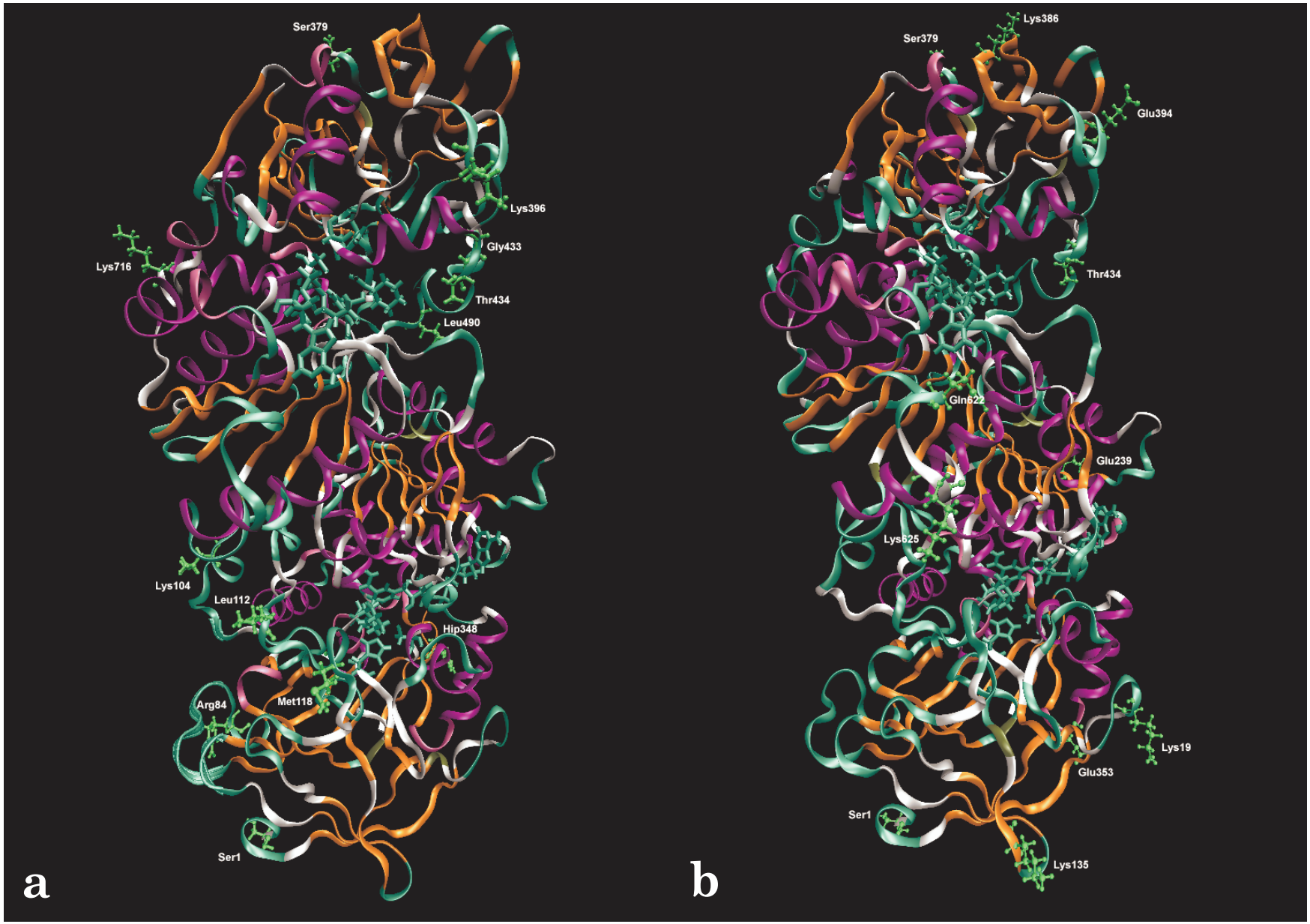



**Figure 2-20** Ribbon representations of the structures of **a**. Simulation 1 and **b**. Simulation 2 averaged over the 10 ps production runs. Residues with average RMS deviations ≥ 3 Å are shown in green. α-helices are in purple and ß-sheets in orange. The NADH active sites and surrounding residues are shown in cyan.

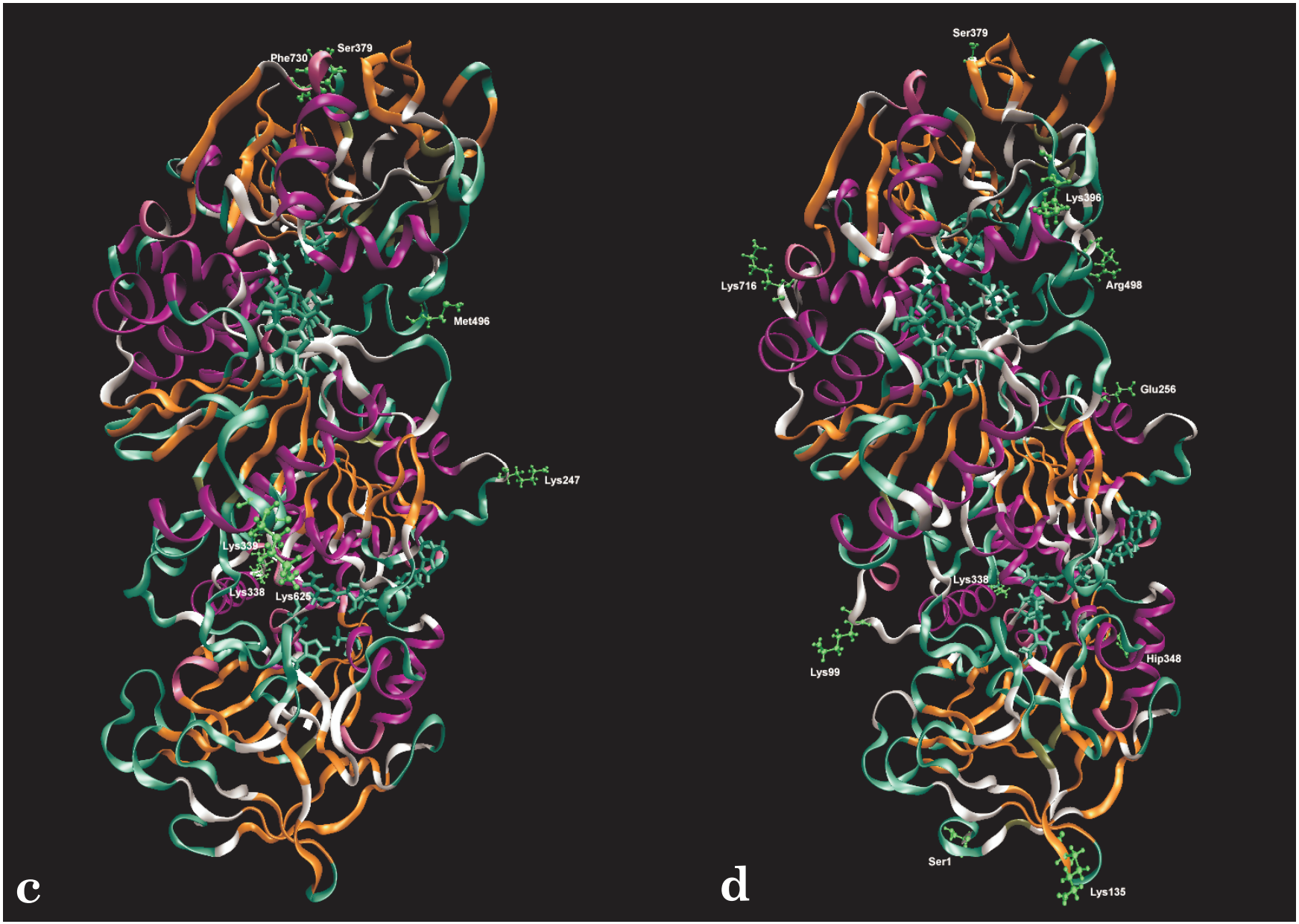



**Figure 2-21** Ribbon representations of the structures of **c**. Simulation 3 and **d**. Simulation 4 averaged over the 10 ps production runs. Residues with average RMS deviations ≥ 3 Å are shown in green. α-helices are in purple and ß-sheets in orange. The NADH active sites and surrounding residues are shown in cyan.

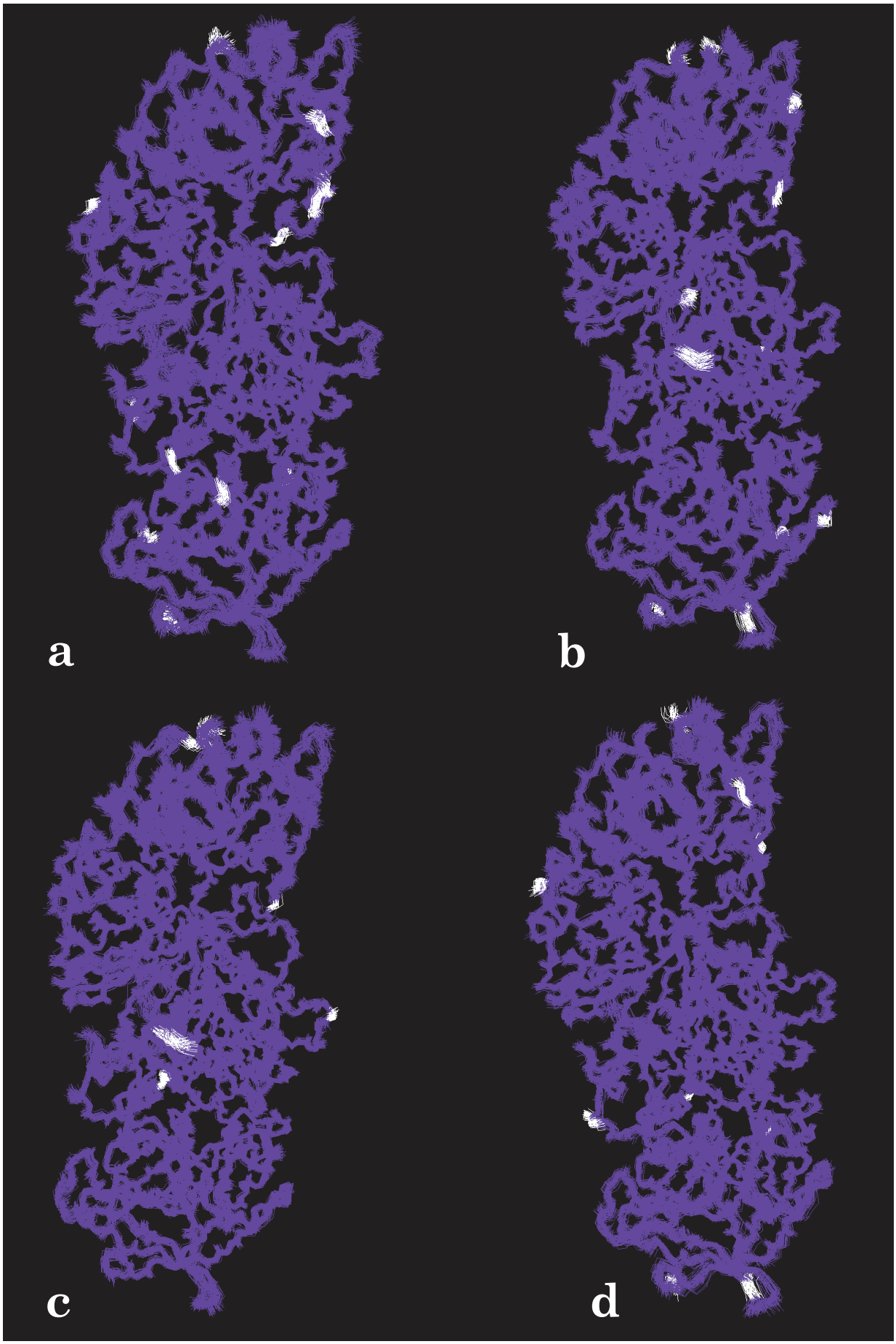



**Figure 2-22** Stick representations of the structures of **a**. Simulation 1, **b**. Simulation 2, **c**. Simulation 3, **d**. Simulation 4, produced at 200 fs intervals over the 10 ps production runs. Residues with average RMS deviations ≥ 3 Å are shown in white.

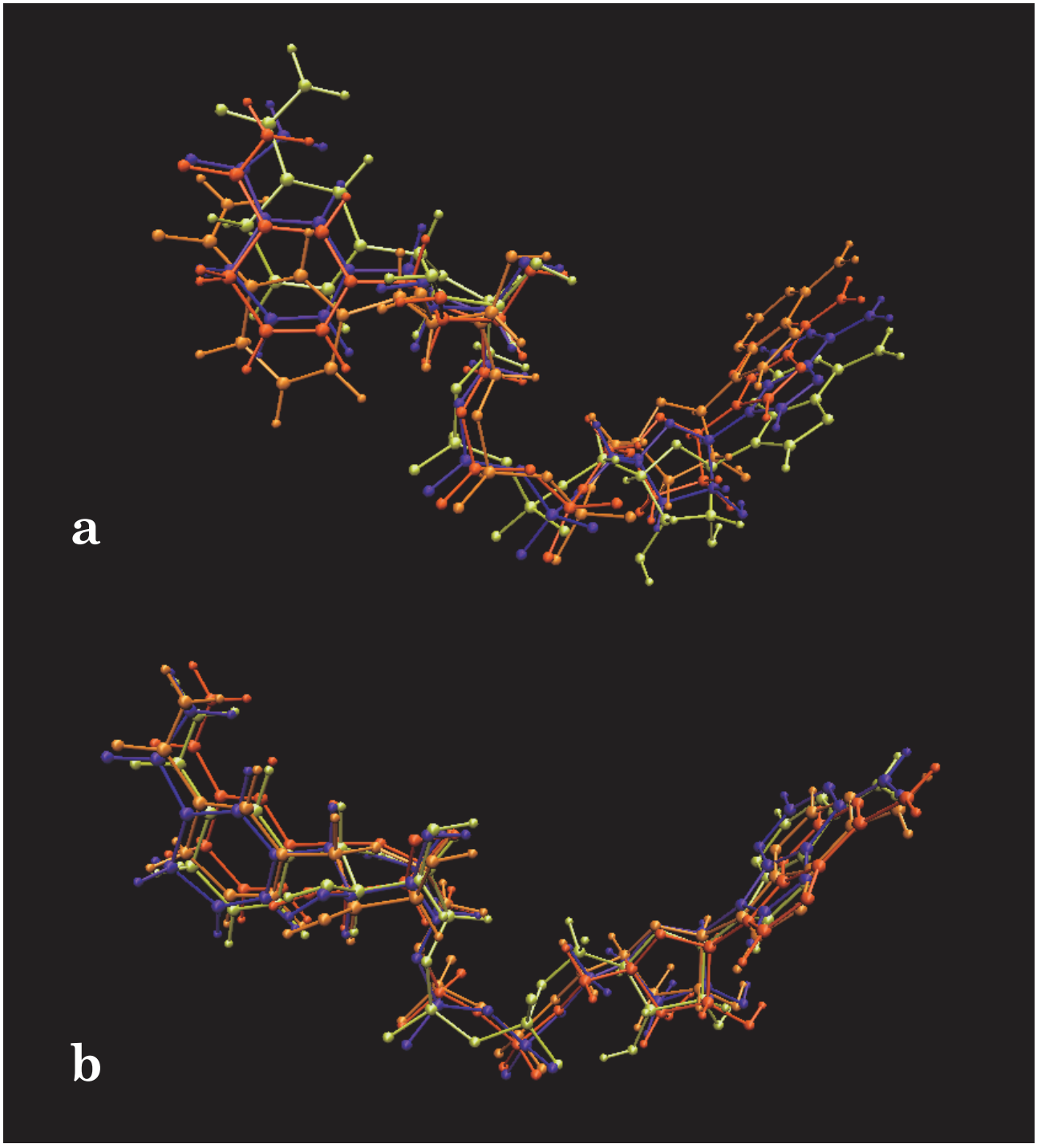



**Figure 2-23** Ball and stick representation of the structures of the NADH residues **a**. NAD375, **b**. NAD753, averaged over the 10 ps production runs. Simulation 1 = Blue, Simulation 2 = Red, Simulation 3 = Yellow and Simulation 4 = Orange.

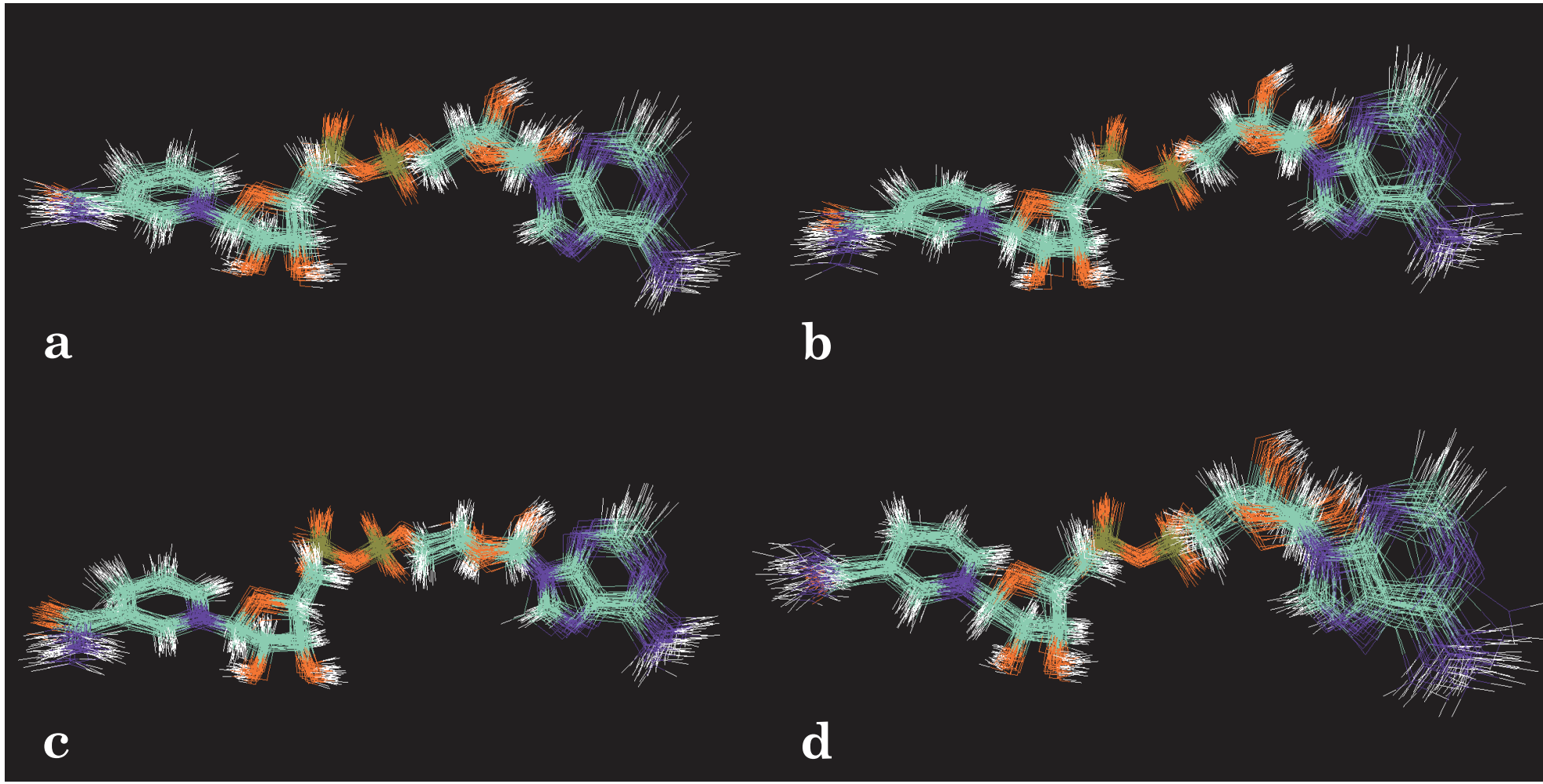



**Figure 2-24** NADH (residue **375**) structure snapshots taken at 200 fs intervals over the 10 ps production runs for **a**. Simulation 1, **b**. Simulation 2, **c**. Simulation 3, **d**
Simulation 4.

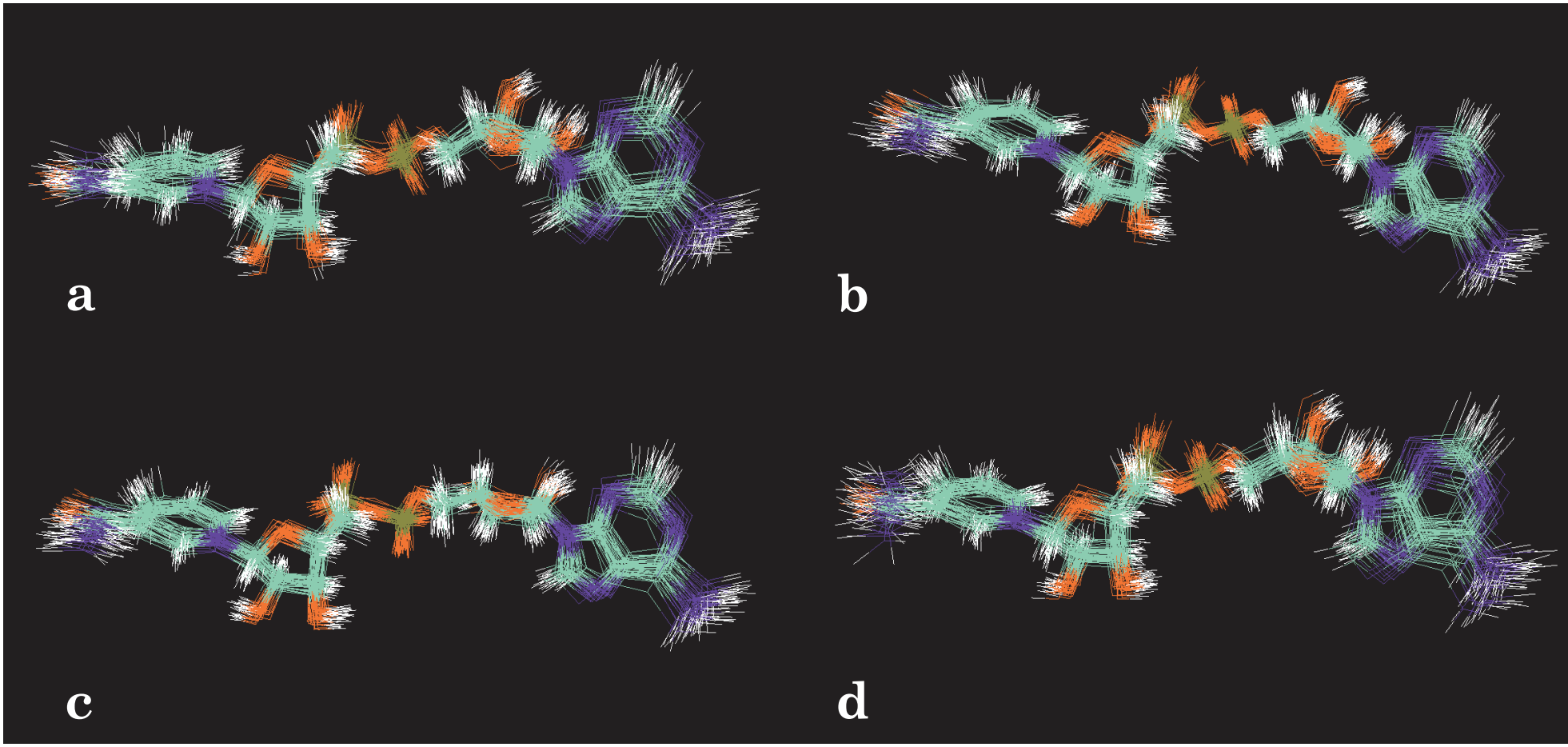



**Figure 2-25** NADH (residue **753**) structure snapshots taken at 200 fs intervals over the 10 ps production runs for **a**. Simulation 1, **b**. Simulation 2, **c**. Simulation 3, **d**. Simulation 4.

# Chapter 3

# QM/MM Calculations on LADH

## 3.1 Introduction

The linear response approach to calculating optical observables for biological systems employed in this work and discussed in Chapter 4 requires the auto-correlation of the chromophore energy gap between the ground and excited state as a function of time. Trial calculations, discussed in section 3.3.2, showed that configuration interaction singles (CIS) calculations, utilising the SV3-21G* basis set, on optimised NADH structures could accurately reproduce the peaks in the absorption spectra when adjusted by a fixed offset. Thus single point CIS SV3-21G* calculations were used in this work. By employing a QM/MM methodology whereby the NADH residue was treated quantum mechanically while the rest of the protein and surrounding waters were treated by including their charges within the one electron Hamiltonian it was possible, within the limitations of the available computing power, to evaluate the energy gaps for each of the structures obtained during the 10 ps production runs discussed in Chapter 2. The aim of this approach was to obtain information about how the ground and excited states of NADH fluctuated within the protein matrix as a function of time while preserving the time ordering. This information was then used for generating the theoretical spectra shown in Figure 4-12 and Figure 4-13.

This chapter discusses the theory and methodology used to generate the time ordered energy gap fluctuation data from the four trajectories produced in Chapter 2 as well as the methods used to tackle the logistics involved with calculating and processing the thousands of single point CIS calculations required to achieve convergence of the energy gap correlation functions. A large proportion of this work involved the design, construction and implementation of a number of custom built computer clusters as well as the design of software for managing the execution and processing the data generated from the large number of simultaneously executing calculations. An overview of this work is given in this chapter with a brief description of the software. The source code and executables of the various pieces of software developed for this work are provided on the accompanying CD-ROM.

# 3.2 Theory

This section contains an overview of the QM/MM method employed in this study along with an outline of aspects of molecular orbital theory [Levine 1991; McWeeny 1992; Grant and Richards 1995; Szabo and Ostlund 1996; Atkins and Friedman 1997; Jensen 1999].

## 3.2.1 Schrödinger's Equation

The Schrödinger equation is a well known entity in modern science. In its barest, non-relativistic, time-independent form it is represented by the eigenvalue relationship in (3.1) [Schrödinger 1926]:

$$\hat{H}\psi = E\psi \tag{3.1}$$

where $E$ is the total molecular energy and $\hat{H}$ is the Hamiltonian which operates on the molecular wavefunction $\Psi$ to give the system energy $E$. For a time-independent, multi-electron system the molecular Hamiltonian can be represented as:

$$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{electrons}\nabla_i^2 - \frac{\hbar^2}{2}\sum_{\alpha=1}^{nuclei}\frac{1}{m_\alpha}\nabla_\alpha^2 - \sum_{\alpha=1}^{nuclei}\sum_{i=1}^{electrons}\frac{Z_\alpha e^2}{4\pi\varepsilon_0 r_{\alpha i}} + \sum_{\alpha=1}^{nuclei}\sum_{\beta>\alpha}^{nuclei}\frac{Z_\alpha Z_\beta e^2}{4\pi\varepsilon_0 r_{\alpha\beta}} + \sum_{i}^{electrons}\sum_{j>i}^{electrons}\frac{e^2}{4\pi\varepsilon_0 r_{ij}} \quad (3.2)$$

where $m_\alpha$ is the mass of a nucleus of charge $Z$ and $m_e$ is the mass of an electron whose charge is $e$. The distances separating pairs of nuclei, pairs of electrons, and electrons from nuclei are $r_{\alpha\beta}$, $r_{ij}$ and $r_{\alpha i}$ respectively. The first two terms in (3.2) are summations of the differential operator $\nabla^2$ over the entire molecule. The first term describes the electron kinetic energy ($T_e$) while the second describes the nuclear kinetic energy ($T_n$). The last three terms represent the potential energies arising from coulombic interactions between nuclei and electrons ($V_{ne}$), nuclei and nuclei ($V_{nn}$), and electrons and electrons ($V_{ee}$) respectively. The total Hamiltonian operator can thus be represented as the summation:

$$\hat{H}_{tot} = T_e + T_n + V_{ne} + V_{nn} + V_{ee} \quad (3.3)$$

Due to the complexity of the calculations involved in solving the Schrödinger equation only the simplest of molecular systems can be treated in the exact form. A number of approximations must therefore be made, and these are outlined below.

#### 3.2.1.1 The Born-Oppenheimer Approximation

Oppenheimer and Born showed, in 1927 [Born and Oppenheimer 1927], that (3.2) could be greatly simplified by exploiting the fact that the mass of the nuclei are much greater than the mass of the electrons. This means that the nuclei can be considered as fixed whilst the electrons carry out their motions, allowing the ($T_n$) term in (3.2) to be neglected and $V_{nn}$ to be added to the total energy at the end of the calculation. It is then possible to define a purely electronic Schrödinger equation,

$$\hat{H}_{el}\psi_{el} = E_{el}\psi_{el} \quad (3.4)$$

where $\hat{H}_{el}$ is the electronic Hamiltonian,

$$\hat{H}_{el} = -\frac{\hbar^2}{2m_e}\sum_{i}^{electrons}\nabla_i^2 - \sum_{\alpha}^{nuclei}\sum_{i}^{electrons}\frac{Z_\alpha e^2}{4\pi\varepsilon_0 r_{\alpha i}} + \sum_{i}^{electrons}\sum_{j>i}^{electrons}\frac{e^2}{4\pi\varepsilon_0 r_{ij}} \quad (3.5)$$

(3.5) is thus solved to find the electronic energy $E_{el}$ and then $V_{nn}$ is added to obtain the total energy,

$$E_{tot} = E_{el} + V_{nn} \tag{3.6}$$

The problem is therefore reduced to finding two terms, the electronic energy and the nuclear repulsion energy. Unfortunately the electron-electron repulsion term ($V_{ee}$) is still present. The value of $V_{ee}$ depends on the distance between each of the electrons which cannot be known prior to calculation. Hence iterative methods have to be employed to find the wavefunction for a many electron system.

### 3.2.2 The Hartree-Fock SCF Method

In 1928 Hartree devised a method that made it possible to find, to a reasonable approximation, the most accurate wavefunction, and hence solution to the Schrödinger equation for a poly-electronic system [Hartree 1928]. By treating the electron-electron repulsions in an average way the many electron wavefunction can be replaced by the product of one-electron wavefunctions. This separates the variables making their calculation much easier. For a system consisting of *n*-electrons the simplest form of this wavefunction is the Hartree product [Hartree 1928],

$$\psi(\boldsymbol{r}) = \psi_1(\boldsymbol{r}_1)\psi_2(\boldsymbol{r}_2)...\psi_n(\boldsymbol{r}_n) \tag{3.7}$$

where $\psi_n(r_n)$ is an orbital whose value is a function of the position vector $\boldsymbol{r}$ such that the probability of finding an electron in the element $d\boldsymbol{r}$ surrounding $\boldsymbol{r}$ is $|\psi_n(\boldsymbol{r}_n)|^2 d\boldsymbol{r}$. This does not, however, include electron spin and must be adjusted to satisfy the Pauli exclusion principle [Feynman 1987]. This states that "the wavefunction must be antisymmetric with respect to interchange of any two electrons." [Atkins and Friedman 1997]. This is done by introducing the concept of spin orbitals $\phi(\boldsymbol{x},\boldsymbol{r})$ where $\boldsymbol{x}$ describes the spin and $\boldsymbol{r}$ the spatial coordinates,

$$\phi(\boldsymbol{x};\boldsymbol{r}) = \begin{cases} \phi(\boldsymbol{r})\alpha(\omega) \\ \text{or} \\ \phi(\boldsymbol{r})\beta(\omega) \end{cases} \tag{3.8}$$

where $\alpha(\omega)$ and $\beta(\omega)$ are orthonormal spin functions describing electron spin as either up ($\alpha$) or down ($\beta$).

### 3.2.2.1 Slater Determinants

By writing the many-electron wavefunction as a single-determinant of spin orbitals known as a Slater determinant [Slater 1929] it is possible to incorporate electron spin into the product of one-electron wavefunctions,

$$\psi(\boldsymbol{x};\boldsymbol{r}) = \frac{1}{\sqrt{n!}} \det \begin{vmatrix} \phi_a(1) & \phi_b(1) & \cdots & \phi_z(1) \\ \phi_a(2) & \phi_b(2) & \cdots & \phi_z(2) \\ \vdots & \vdots & & \vdots \\ \phi_a(n) & \phi_b(n) & \cdots & \phi_z(n) \end{vmatrix} \tag{3.9}$$

where $n$ is the number of electrons, $1/\sqrt{n!}$ is a normalisation constant to ensure that the probability of finding an electron in space is unity and the letter subscripts correspond to each spin orbital.

### 3.2.2.2 Orthonormality of Spin Orbitals

An important property of spin orbitals is that they are orthogonal and normalised. This orthonormality condition is expressed as,

$$\int \phi_i^* \phi_j \, d\tau = \delta_{ij} \tag{3.10}$$

where $d\tau$ indicates that the integration is performed over all space and $\delta_{ij}$ is the Kronecker delta which posses the following property,

$$\delta_{ij} = 1 \quad \text{if} \quad i = j, \qquad \delta_{ij} = 0 \quad \text{if} \quad i \neq j \tag{3.11}$$

### 3.2.2.3 Electron Correlation and the HF Method

Electron-electron repulsions are important and need to be included for an accurate treatment of a molecular system. The Hartree Fock method treats electron correlation in an average way. Each electron is considered to move in the field generated by each of the nuclei and the average field of the other $n-1$ electrons. In practice this average electron repulsion approach manifests itself in bond distances

being underestimated since electrons are allowed to approach too closely. The aim is to find a product wavefunction of the form of equation (3.9). The 'best' *n*-electron wavefunction is found by adjusting the spin orbitals using variation theory.

### 3.2.2.4 The Variational Principle

The variational principle is the cornerstone of the HF-SCF approach to solving molecular wavefunctions. The variational principle states that the energy of the exact wavefunction is always lower than the energies calculated using any trial normalised antisymmetric wavefunction, that is

$$E = \frac{\int \psi_{trial}^{*} \hat{H} \psi_{trial} d\tau}{\int \psi_{trial}^{*} \psi_{trial} d\tau} \geq E_0 \tag{3.12}$$

where $E_0$ is the energy of the true ground state wavefunction of the system. The equality only holds if the trial wavefunction is identical to the true ground state wavefunction.

The significance of the variation theorem is that it provides a way of calculating the upper bound to the true energy of the system. Trial wavefunctions can thus be created and modified so as to minimise (3.12). This approach leads to the formation of the Hartree-Fock equations, solved using the self-consistent field approach (SCF) discussed in §3.2.2.5.

### 3.2.2.5 The Hartree-Fock Equations

The normalised ground state Hartree-Fock wavefunction $\psi_0$ is given by the *n*-electron Slater determinant (3.9). The ground-state HF energy is given by (3.13)[15]

$$E = \langle \psi_0 | \hat{H} | \psi_0 \rangle \tag{3.13}$$

where $\hat{H}$ is given by,

$$\hat{H} = \sum_i h_i + \frac{1}{2} \sum_{i,j}{}' \left( \frac{e^2}{4\pi\varepsilon_0 r_{ij}} \right) \tag{3.14}$$

[15] Where Dirac's bracket notation has been used. In this notation a function $f$ (a ket vector) is represented by $|f\rangle$ and its complex conjugate (a bra vector) is represented by $\langle f|$. Thus $\langle \psi_0 | \hat{H} | \psi_0 \rangle = \int \psi_0^{*} \hat{H} \psi_0 d\tau$.

thus,

$$E = \langle \psi_0 | \sum_i h_i + \frac{1}{2} \sum_{i,j}{}' \left( \frac{e^2}{4\pi\varepsilon_0 r_{ij}} \right) | \psi_0 \rangle \tag{3.15}$$

where $h_i$ represents the one electron term corresponding to the electron nuclei interaction energy and the electron kinetic energy.

The fact that all of the electrons are indistinguishable means that all of the matrix elements of $h_i$ are the same and so the first term of (3.15) can be written in terms of the sum of spin orbitals by expanding $\psi_0$ in terms of the spin orbitals, utilising equation (3.9), to give,

$$\langle \psi_0 | \sum_i h_i | \psi_0 \rangle = \sum_{i=1}^{n} \langle \phi_i(1) | h_1 | \phi_i(1) \rangle \tag{3.16}$$

which can be rewritten using one-electron notation[16] as,

$$\langle \psi_0 | \sum_i h_i | \psi_0 \rangle = \sum_{i=1}^{n} \left[ \phi_i | h | \phi_i \right] \tag{3.17}$$

The second term in (3.15) is over all $n(n-1)/2$ unique pairs of electrons. The electrons, however, are indistinguishable in HF theory and so each term of the second sum in (3.15) yields the same result. Thus

$$\langle \psi_0 | \frac{1}{2} \sum_{i,j}{}' \left( \frac{e^2}{4\pi\varepsilon_0 r_{ij}} \right) | \psi_0 \rangle = \frac{n(n-1) \langle \psi_0 | \frac{e^2}{4\pi\varepsilon_0 r_{12}} | \psi_0 \rangle}{2} \tag{3.18}$$

Expanding $\psi_0$ in terms of spin orbitals using equation (3.9) turns (3.18) into,

$$\frac{1}{2} \sum_{i,j}{}' \int \phi_i^*(1) \phi_j^*(2) \left( \frac{e^2}{4\pi\varepsilon_0 r_{12}} \right) \{ \phi_i(1) \phi_j(2) - \phi_j(1) \phi_i(2) \} d\mathbf{x}_1 d\mathbf{x}_2 \tag{3.19}$$

Using the relationship

[16] $\left[ \phi_i | h | \phi_i \right] = \langle \phi_i(1) | h_1 | \phi_i(1) \rangle$

$$\left[\phi_a\phi_b\middle|\phi_c\phi_d\right] = \int \phi_a^*(1)\phi_b(1)\left(\frac{e^2}{4\pi\varepsilon_0 r_{12}}\right)\phi_c^*(2)\phi_d(2)\,d\boldsymbol{x}_1 d\boldsymbol{x}_2 \tag{3.20}$$

and equations (3.17) and (3.19), the energy (eq. (3.15)) can be written as a function of spin orbitals such that for every set of functions $\phi$, there will be an associated value $E$,

$$E = \sum_{i=1}^{n}\left[\phi_i\middle|h\middle|\phi_i\right] + \frac{1}{2}\sum_{i,j}^{n}\left\{\left[\phi_i\phi_i\middle|\phi_j\phi_j\right] - \left[\phi_i\phi_j\middle|\phi_j\phi_i\right]\right\} \tag{3.21}$$

The technique of functional variation, discussed in section 3.2.2.4, can then be used to derive the Hartree Fock equations. The determinant $\psi$ is sought for which a small change $\delta\psi$ yields no change in energy,

$$\delta\langle\psi|H|\psi\rangle = \langle\delta\psi|H|\psi\rangle + \langle\psi|H|\delta\psi\rangle = \delta E = 0 \tag{3.22}$$

The constraint that the spin orbitals be orthonormal (§3.2.2.2), however, means that the spin orbitals cannot be varied in an independent fashion. This means that the technique of undetermined multipliers[17] must be used giving:

$$\sum_{i=1}^{n}\left[\delta\phi_i\middle|h\middle|\phi_i\right] + \sum_{i,j}^{n}\left\{\left[(\delta\phi_i)\phi_i\middle|\phi_j\phi_j\right] - \left[(\delta\phi_i)\phi_j\middle|\phi_j\phi_i\right] - \lambda_{ij}\langle\delta\phi_i|\phi_j\rangle\right\} = 0 \tag{3.23}$$

where $\lambda_{ij}$ represents a set of undetermined multipliers. By introducing the concept of a Coulomb operator, $J_u$ (3.24), and an exchange operator, $K_u$ (3.25), where the Coulomb operator represents repulsion between electrons and the exchange operator represents the effects of spin correlation,

$$J_u(1)\phi_a(1) = \left\{\int \phi_u^*(2)\left(\frac{e^2}{4\pi\varepsilon_0 r_{12}}\right)\phi_u(2)\,d\boldsymbol{x}_2\right\}\phi_a(1) \tag{3.24}$$

$$K_u(1)\phi_a(1) = \left\{\int \phi_u^*(2)\left(\frac{e^2}{4\pi\varepsilon_0 r_{12}}\right)\phi_a(2)\right\}\phi_u(1) \tag{3.25}$$

[17] For more details see *Further Information 14* in Atkins, P. W. (1994). "Physical Chemistry". Oxford, Oxford University Press.

equations (3.17) and (3.20) can be used to obtain, after factoring out the common $\delta\phi_i^*$ term,

$$\sum_{i=1}^{n}\int \delta\phi_i^*(1)\left( h_1\phi_i(1)+\sum_{j=1}^{n}\left\{J_i(1)\phi_i(1)-K_j(1)\phi_i(1)-\lambda_{ji}\phi_j(1)\right\}\right)d\boldsymbol{x}_1 = 0 \tag{3.26}$$

Since $\delta\phi_i^*$ is arbitrary each term within the parentheses is zero and thus for each spinorbital,

$$h_1\phi_i(1)+\sum_{j=1}^{nn}\left\{J_i(1)\phi_i(1)-K_j(1)\phi_i(1)\right\}=\sum_{i=1}\lambda_{ij}\phi_j(1) \tag{3.27}$$

Defining a Fock operator $f_1$ as,

$$f_1 = h_1+\sum_u\left\{J_u(1)-K_u(1)\right\} \tag{3.28}$$

equation (3.27) can be rewritten as,

$$f_1\phi_i(1)=\sum_{i=1}^{n}\lambda_{ij}\phi_j(1) \tag{3.29}$$

The set of spin orbitals $\phi$ is not unique since it is possible to form a new set of spin orbitals, each a linear combination of $\phi$, without changing the minimum energy. One possible transformation is to convert the original set of spin orbitals into a new set of orthonormal canonical spin orbitals, $\phi'$, where the transformed fock operator $f_1'$ is the same as $f_1$ and the matrix of multipliers $\lambda_{ij}$ is a diagonal matrix of elements $\varepsilon_i'$. (3.29) then becomes what is referred to, when the prime labels are discarded, as the Hartree-Fock equation for spinorbital $\phi_a(1)$, where electron 1 has been arbitrarily assigned to spinorbital $\phi_a$,

$$f_1\phi_a(1)=\varepsilon_a\phi_a(1) \tag{3.30}$$

### 3.2.2.6 The Self-Consistent Field Approach

Each spin orbital can be obtained from (3.30) with the corresponding Fock operator $f_i$. A fundamental problem remains however. Each Fock operator $f_i$

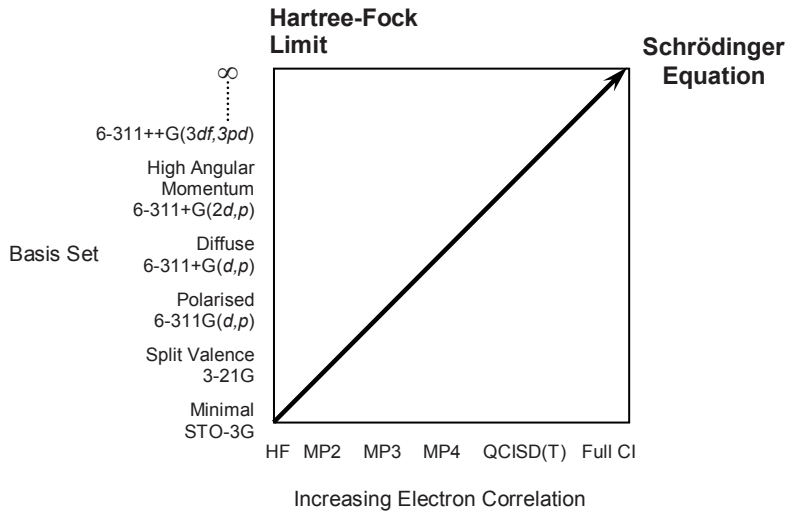



**Figure 3-1** Basis set size versus electron correlation. As the basis set increases the Hartree-Fock limit is approached. Similarly with improved treatment of electron correlation an exact solution to the Schrödinger equation is approached.

depends on the spin orbitals of the other $n-1$ electrons. Hence to set up the Hartree-Fock equations one would need to know the spin orbitals prior to their calculation. The method for tackling this dilemma is to solve the equations in an iterative fashion that stops when the solutions are self-consistent. This approach is referred to as the self-consistent field (SCF) approach and forms the cornerstone of quantum chemistry calculations. The SCF procedure involves forming a trial set of spin orbitals which are used to construct the Fock operator. The HF equations are then solved to yield a new set of spin orbitals that are used to construct a revised Fock operator. This cycle is repeated until predefined converge criteria are satisfied.

### 3.2.2.7 The Roothaan-Hall Equations

The HF procedure is fairly straight forward to implement for atoms since their spherical symmetry means that (3.30) can be solved numerically. Numerical solution of (3.30) is not computationally feasible for molecules, however, and so modifications are required to the HF procedure for the treatment of molecules. In 1951 C.C.J. Roothaan [Roothaan 1951] and G.G. Hall [Hall 1951] independently proposed a solution to the treatment of molecules via the expansion of the molecular orbitals (MOs) $\psi_i$ as linear combinations of known basis functions $\theta_j$,

$$\psi_i = \sum_{j=1}^{M} c_{ji}\theta_j \tag{3.31}$$

where $c_{ij}$ are as yet unknown coefficients. If the basis set were infinite then the $\theta_j$'s would form a complete set and the representation would be exact. Using a complete set of basis functions with the Hartree-Fock approach results in an energy equal to that given by the variational expression in equation (3.12). This energy, known as the Hartree-Fock limit, is not the exact ground state energy of the molecule because the effects of electron correlation are not included. However, in practice the basis set is limited to a finite set of $M$ basis functions, and the error due to the incompleteness of the basis set is called the basis-set truncation error (Figure 3-1). Substitution of the expansion into (3.30) yields,

$$f_1 \sum_{j=1}^{M} c_{ja}\theta_j(1) = \varepsilon_a \sum_{j=1}^{M} c_{ja}\theta_j(1) \tag{3.32}$$

The problem is thus reduced to computing the coefficients $c_{ja}$. Multiplication of (3.32) by the basis function $\theta_i^*(1)$ followed by integration over $d\boldsymbol{r}_1$ yields,

$$\sum_{j=1}^{M} c_{ja} \int \theta_i^*(1) f_1 \theta_j(1) d\boldsymbol{r}_1 = \varepsilon_a \sum_{j=1}^{M} c_{ja} \int \theta_i^*(1)\theta_j(1) d\boldsymbol{r}_1 \tag{3.33}$$

This expression is one in a set of $M$ simultaneous equations that are known as the *Roothaan-Hall* equations. (3.33) can be expressed in a more compact notation via the use of matrices,

$$\boldsymbol{Fc} = \boldsymbol{Sc}\varepsilon \tag{3.34}$$

where $\varepsilon$ is an $M \times M$ diagonal matrix of the orbital energies $\varepsilon_a$, $\boldsymbol{c}$ is an $M \times M$ matrix of the coefficients $c_{ja}$ and $\boldsymbol{S}$ and $\boldsymbol{F}$ are the overlap and Fock matrices respectively,

$$\begin{aligned} S_{ij} &= \langle \theta_i | \theta_j \rangle \\ F_{ij} &= \langle \theta_i | f | \theta_j \rangle \end{aligned} \tag{3.35}$$

(3.34) has a non-trivial solution only when the determinant of the coefficients is zero,

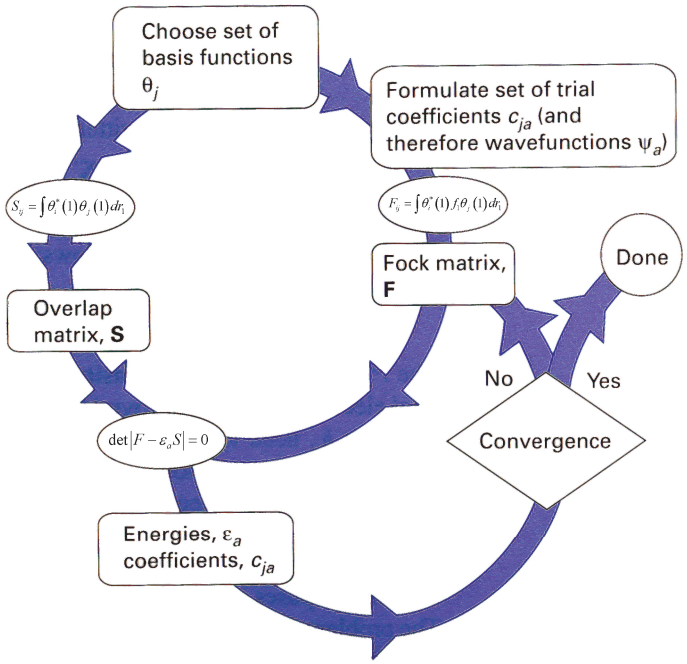



**Figure 3-2** Graphical illustration of the HF-SCF procedure. Adapted from [Atkins and Friedman 1997]. A set of basis functions are chosen and, with a set of trial coefficients, used to create a set of trial wavefunctions. These are then used to form the Fock matrix, yielding a new set of energies and coefficients. The process is repeated until the convergence criteria are met.

$$\det\left|\boldsymbol{F} - \varepsilon_a \boldsymbol{S}\right| = 0 \tag{3.36}$$

Unfortunately (3.36) cannot be solved directly since the Fock matrix $\boldsymbol{F}$ is composed of Coulomb and exchange integrals whose values depend on the spatial wavefunctions that are trying to be found. Thus an iterative procedure (§3.2.2.6) has to be used. The HF-SCF procedure, shown graphically in Figure 3-2, involves choosing a set of basis functions, taking a trial set of coefficients and thus wavefunctions and then using these to construct a trial Fock matrix, giving a set of orbital energies and new coefficients. This procedure is then repeated until the pre-defined convergence criteria are fulfilled.

## 3.2.3 Basis Sets

By using the Roothaan-Hall equations to solve the HF equations it is necessary to express the molecular orbitals $\psi_i$ as linear combinations of known basis functions $\theta_j$

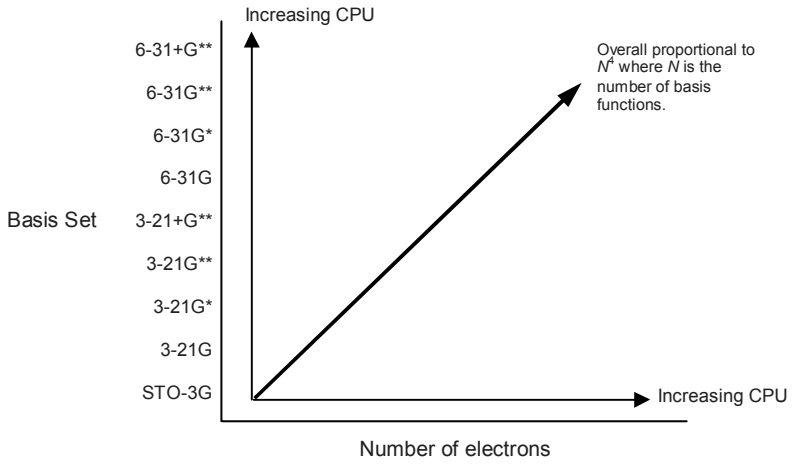



**Figure 3-3** Increase in computational time with increasing basis set complexity.

(3.31). In principle any well behaved function could be used as a basis function $\theta$, for example if the atomic orbitals from each of the constituent atoms are used then this becomes known as a Linear Combination of Atomic Orbitals (LCAO). The choice of basis set is an important factor in any electronic structure calculation. Choosing a more complex basis set containing more functions will give a better representation of the MOs but at the expense of computational resources (Figure 3-3).

#### 3.2.3.1 Basis Functions

In practice atomic orbitals are not generally used in molecular orbital theory and instead basis functions are chosen according to two overriding factors. The first is the degree to which they qualitatively describe the orbitals and the second is the ease with which they can be computed. Two functions that have found common usage in MO theory are the Slater Type Orbital (STO) and the Gaussian Type Orbital (GTO). The functional form as well as the advantages and disadvantages of each will be discussed below.

***Slater Functions***

Slater functions are the ideal function to use in a basis set for describing molecular orbitals. Slater-type orbitals (STOs) have the following functional form, in spherical harmonics:

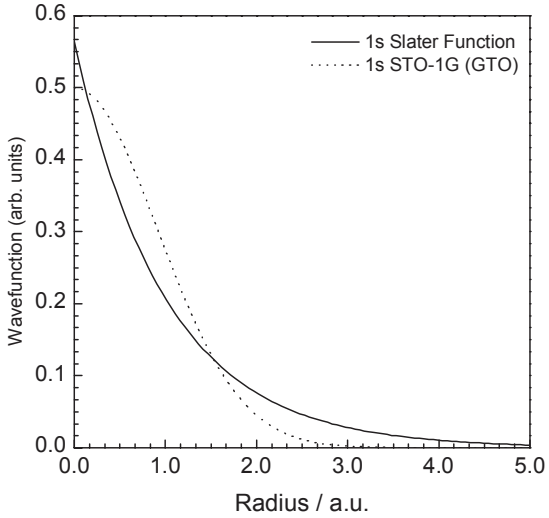



**Figure 3-4** Comparison of a 1s Slater function (solid line) and a 1s STO-1G (GTO) function (dotted line)

$$\chi_a^{STO} = N_{STO} r_a^{n-1} Y_{l,m}(\theta_a, \upsilon_a) e^{-\varsigma r_a} \tag{3.37}$$

where $N_{STO}$ is a normalisation constant, $Y_{l,m}$ is the spherical harmonic function, $n$ is a quantum number and $\varsigma$ is a parameter known as the Slater orbital exponent. STOs have a finite value at zero which correctly describes the cusp of an orbital at the nucleus. They also decay slowly describing the wave function accurately at long distances from the nucleus (Figure 3-4). Unfortunately the use of STOs leads to difficulties in the evaluation of the two-electron integrals for polyatomic systems.

***Gaussian Functions***

An alternative to STOs proposed by Boys in 1950 [Boys 1950] uses Gaussian type orbitals (GTOs) to make evaluation of one- and two- electron integrals more efficient. These have the functional form:

$$\chi_a^{GTO} = N_{GTO}\left(x_a^i y_a^j z_a^k\right) e^{-\alpha r_a^2} \tag{3.38}$$

where $N_{GTO}$ is a normalisation constant, $x_a, y_a, z_a$ are the Cartesian coordinates for atom $a$, $i+j+k$ is the total angular momentum of the basis function and $\alpha$ is the Gaussian orbital exponent. GTOs differ from STOs in two important respects (Figure 3-4). Firstly at a distance of zero from the nucleus a GTO has zero gradient whilst an

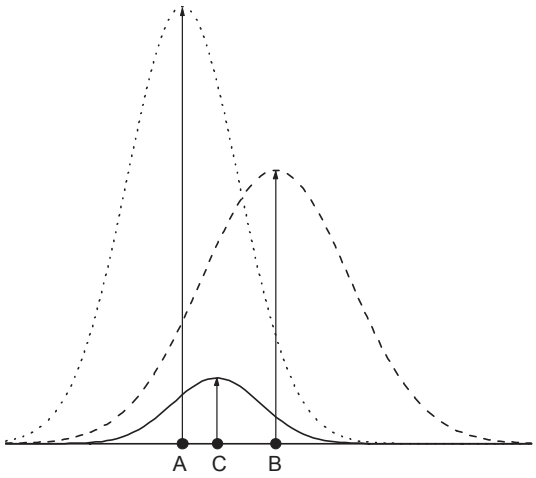



**Figure 3-5** The product of two 1s Gaussian functions centred on points **A** and **B** is a third Gaussian centred between them on point **C**.

STO has a finite gradient. Thus GTOs fail to describe the cusp at the nucleus correctly. Secondly at large distances GTOs decay with an exponential quadratic dependence while STOs show a linear dependence. It is known that the wavefunction decays in the form of a Slater function rather than a Gaussian function. The exact solution for a hydrogen atom is of Slater form, hence using a STO is akin to using a hydrogen atomic orbital.

At first sight STOs appear to be far superior to GTOs, however, GTOs have one distinct advantage; they allow the three- and four-centre two-electron repulsion integrals to be calculated much faster by exploiting the Gaussian product theory. This states that the product of two Gaussians on different centres is a single Gaussian situated on a third centre located at a point between them as shown in Figure 3-5. This allows the three- and four-centre two-electron repulsion integrals to be reduced to two-centre integrals thus giving GTOs a massive computational advantage over STOs. The disadvantages caused by the poor representation of the orbitals by GTOs is overcome by the use of basis functions that are constructed from linear combinations of primitive Gaussian type functions. These combinations produce what are termed contracted Gaussian functions (CGFs),

$$\theta_j^{CGF} = \sum_{p=1}^{L} d_{pj} \chi_j^{GTO} \tag{3.39}$$

where $d_{pj}$ is a contraction coefficient and $L$ is the length of the contraction.

### 3.2.3.2 Standard Basis Sets

Standard basis sets use CGFs to approximate STOs. The accuracy of the basis set, and conversely the amount of cpu time required to evaluate it, is determined by the number of GTOs used to form each CGF and the number of CGFs employed for describing each orbital. Figure 3-3 illustrates the increase in computation time with increasing basis set sophistication. The scaling is formally $N^4$ since there are $N$ ways of replacing each basis function in the two electron integrals. However, by symmetry relationships this is formally reduced to $N^4/8$. The evaluation of the two electron repulsions still remains the bottleneck in HF calculations, however, and so the choice of basis set is important since it involves a choice of trade-off between speed and accuracy.

Generally combinations of between one and six GTOs are used to form the basis functions with the orbital exponents being chosen to give the best least-squares fit to the STO. Standard basis sets are generally named in a systematic fashion that describes the number of Gaussian primitives used to construct each basis function making up the set. The various different types of basis set available, together with some common examples are discussed below.

***Minimal Basis Sets***

Minimal basis sets are the simplest type of basis set available and consist of a single STO that is approximated by a series of primitive Gaussians. An example is the STO-3G basis set [Hehre *et al.* 1969; Collins *et al.* 1976] which uses three Gaussian primitives to construct a single STO for each core and valence shell orbital of each atom. For first row elements, such as carbon, the basis set consists of 15 Gaussian primitives contracted to 5 basis functions, namely $1s2s2p_x2p_y2p_z$. For $H_2O$ a minimal basis set would therefore contain a total of 7 basis functions, one each for the hydrogens and five for the oxygen. Such minimal basis sets are referred to generally as STO-*L*G where *L* represents the number of primitives used to construct each basis function. STO-*L*G basis sets also have the restriction that the exponents used in the least squares fitting to the STO are constrained to be identical. This means that the radial behaviour is the same and so they can be treated as a single function in the integral evaluations, making them very efficient. Minimal basis sets are relatively computationally inexpensive and thus can be used for calculations on quite large molecules. Their small size, however, results in wavefunctions and energies that are

not very close to the Hartree-Fock limit and so more accurate calculations require more extensive basis sets.

***Double and Triple Zeta Basis Sets***

More flexibility can be incorporated in the basis set description of the atomic orbitals via the use of double- (DZ) or triple-zeta (TZ) basis sets. In a DZ basis set, each AO is replaced by two STOs each made up from a CGF that differs in its orbital exponent $\varsigma$. A triple-zeta basis set adds even more flexibility, at the expense of computational cost, by replacing each AO with three CGF constructed STOs. An example of a double-zeta basis set is the Dunning DZ basis [Dunning Jr. 1970; Dunning Jr. and Hay 1976] set which consists, for first row atoms, of 24 primitives contracted to 10 basis functions [4s,6p]. For $H_2O$ a DZ basis set would involve 14 basis functions, two for each hydrogen and ten for the oxygen. The doubling in the number of basis functions over minimal basis sets greatly improves the accuracy with which the wavefunction is represented but at the expense of a 16 fold increase in the amount of processing power required.

***Split Valence Basis Sets***

Split valence (SV) basis sets such as the double split valence basis sets 3-21G [Binkley *et al.* 1980; Pietro *et al.* 1982], 6-21G [Binkley *et al.* 1980; Gordon *et al.* 1982] and 6-31G [Hehre *et al.* 1972; Francl *et al.* 1982] offer a compromise between the inadequacy of STO-*L*G minimal basis sets and the high computational demands of double or triple zeta basis sets. In double SV basis sets the valence shell atomic orbitals are described using two basis functions, as in double zeta basis sets, while each core atomic orbital is represented by a single basis function. This is a reasonable approximation to make since it is generally only the valence orbitals that are responsible for the chemical properties of an atom or molecule.

In the 3-21G basis set, two Gaussian primitives, contracted to a single basis function, and an additional diffuse function describe the valance orbital while the inner electrons are described by three primitive Gaussians contracted to a single function. For first row atoms this basis set consists of 15 primitives contracted to 9 basis functions [3s6p]. For $H_2O$ the 3-21G basis set would involve 13 basis functions, two for each hydrogen and nine for the oxygen.

Triple split valence basis sets such as 6-311G [Krishnan *et al.* 1980] describe the valence electrons by three functions and the inner electrons by one function. For first row atoms this basis set consists of 26 primitives contracted to 13 basis functions.

***Diffuse and Polarisation Functions***

Split valence basis sets can be improved further by including additional polarisation or diffuse functions to describe the valence electrons. These are known as extended basis sets. Polarised basis sets work by adding orbitals with angular momentum greater than those needed to describe the ground state orbitals. These orbitals are referred to as polarisation functions because they mimic the experience of an atom in the non-uniform electric field that arises from an aspherical molecular environment. The 3-21G* [Pietro *et al.* 1982] basis set adds *d*-type functions to second row atoms while the 6-31G* [Hariharan and Pople 1973] basis sets adds these functions to all heavy atoms. The 6-31G** [Hariharan and Pople 1973] basis set extends the polarisation treatment by also adding *p*-functions to hydrogen. These polarisation functions are required for the proper description of electron correlation for *p*-type electronic orbitals [Sadlej 1988; Sadlej 1991b; Sadlej 1991a; Sadlej and Urban 1991; Sadlej 1992].

Diffuse functions are large *s*- or *p*-type functions that increase the flexibility of the basis set by allowing the valence orbitals to occupy larger regions of space. An example of this type of basis set is the 6-311++G basis set [Hehre *et al.* 1972; Clark *et al.* 1983]. Diffuse basis functions have a large effect in reducing the phenomenon known as basis set superposition error (BSSE). This effect results in the stabilisation of a monomer in a donor-acceptor complex being over estimated due to the presence of the basis sets of the other [Stone 1996]. Thus without BSSE correction bond strengths will often be over estimated.

## 3.2.4 Electron Correlation and Excited States

Using a complete or infinite basis set would yield the lowest energy using the Hartree-Fock ($E_{HF}$) method since it is variational. This would, however, still be considerably above the non-relativistic, spin-independent ground state energy within the Born-Oppenheimer approximation ($E_0$). The reason for this is the neglect of electron correlation. This difference in energy is known as the correlation energy,

$$E_{corr} = E_0 - E_{HF} \tag{3.40}$$

and typically accounts for an error of about 0.5 – 1 % of the total Hartree-Fock energy. This is roughly the energy of a single chemical bond.

There exist two types of electron correlation. Dynamical correlation is the instantaneous repulsion between electrons of opposite spin while Non-dynamical correlation occurs because of the degeneracy that occurs when orbitals of similar energy are partly filled [Shavitt 1977].

HF theory accounts for some non-dynamical correlation of electrons with the same spin, however no correlation of electrons with different spins is included. The result of this electron correlation neglect is that bond lengths are often underestimated, electrons can come too close together, and dissociation products are often predicted incorrectly. For example the dissociation of $H_2$ is predicted to be $H^+$ and $H^-$ rather than the correct 2x$H^{\cdot}$. This problem can be partially resolved by using an unrestricted description whereby the spatial orbitals for each MO are allowed to vary independently. This method is still insufficient to accurately deal with the problem however.

There exist a number of post-SCF methods that can improve the Hartree-Fock wavefunction significantly. Møller-Plesset (MP) perturbation theory [Møller and Plesset 1934] accounts for dynamical correlation by adding higher excitations to the ground state wavefunction as linear corrections but at the expense of not being a variational method.

Configuration interaction (CI) theory, used in this work for calculating excited states, corrects for both dynamical and non-dynamical correlation effects and is described below. The basic concept behind the CI method is that a single Slater determinant is inadequate to accurately describe the wavefunction and so alternative electronic configurations need to be considered and included in the wavefunction description.

#### 3.2.4.1 Configuration State Functions

The use of a finite basis set with the HF method yields a finite set of spin orbitals ($\phi$). If a basis set consists of $N$ functions then it will yield $2N$ different spin orbitals. The Hartree-Fock wavefunction ($\Phi_0$) is formed by ordering the spin orbitals in terms of energy and taking, for a system of $n$ electrons, the $n$ lowest. There remain, however, $2M - n$ virtual orbitals. By populating the $2M$ spin orbitals in different

ways, other than energy ordering, it is possible to form a large number of different Slater determinants. By reference to the Hartree-Fock wavefunction ($\Phi_0$) the various Slater determinants that can be formed can be classified according to how many electrons have been promoted from occupied to virtual orbitals. $\Phi_0$ is therefore,

$$\Phi_0 = \left|\phi_1\phi_2 \ldots \phi_a\phi_b \ldots \phi_n\right| \tag{3.41}$$

where $\phi_a$ and $\phi_b$ are among the $n$ occupied spin orbitals for the HF ground state.

A singly excited determinant is one in which a single electron in an occupied spinorbital ($\phi_a$) has been promoted to a virtual orbital $\phi_p$,

$$\Phi_a^p = \left|\phi_1\phi_2 \ldots \phi_p\phi_b \ldots \phi_n\right| \tag{3.42}$$

Similarly a doubly excited determinant corresponds to one in which two electrons have been promoted to virtual orbitals. In a similar manner any number, up to $n$, of electron promotions can be used to form multiply excited determinants. Each determinant, or a linear of combination of a number of them, is called a configuration state function (CSF). These excited CSFs can be taken to approximate excited-state wavefunctions, as used in the CIS method for calculating excited states (§3.2.4.3), or can be used as linear combinations to improve the ground-state (or any excited state) wavefunction.

### 3.2.4.2 Full Configuration Interaction

The exact ground-state and excited-state wavefunctions can be expressed as a linear combination of all possible *n*-electron Slater determinants arising from a complete set of spin orbitals [Löwdin 1959]. The exact electronic wavefunction $\psi$ for any state of a system can thus be expressed in the form,

$$\psi = C_0\Phi_0 + \sum_{a,p} C_a^p\Phi_a^p + \sum_{\substack{a<b\\p<q}} C_{ab}^{pq}\Phi_{ab}^{pq} + \sum_{\substack{a<b<c\\p<q<r}} C_{abc}^{pqr}\Phi_{abc}^{pqr} + \ldots \tag{3.43}$$

where the $Cs$ are expansion coefficients and where the limits in the summation ensure the spin orbitals are summed over all unique pairs in doubly excited determinants and all unique triplets in triply excited determinants and so on. The *ab initio* implementation of (3.43) is called full CI. In practice the use of an infinite basis

set with the full CI method is not computationally feasible and so limited basis sets must be used. Even so the full CI method is very computationally demanding. For an $n$-electron system described by a basis set of $N$ functions the summation in (3.43) is over $(2N)!/\left[n!(2N-n)!\right]$ substituted determinants. It should be apparent from this that full CI calculations are only feasible for systems with fewer than about 10 electrons. For example, a full CI calculation involving 10 electrons and 20 basis functions requires a total of $8.4766 \times 10^8$ determinants to be considered. In practice full CI is rarely used and (3.43) is often truncated to include only a limited set of spin orbitals and excited determinants. Some examples of the various truncation schemes are discussed in section 3.2.4.3. Nonetheless, CI is a popular method for the calculation of accurate molecular wavefunctions.

In CI methods the energies corresponding to the wavefunction are found by linear variation of equation (3.43) which gives the following relation,

$$\sum_{J=1}^{L} H_{IJ} C_{Js} = E_s \sum_{J=1}^{L} S_{IJ} C_{Js} \tag{3.44}$$

where the sum is over $L$ determinants, $S_{IJ}$ is the overlap integral, $E_s$ is the energy of the state $s$ and $C_{Js}$ represent the expansion coefficients. The ground state energy is given by the lowest root of (3.44). $H_{IJ}$ is a configurational matrix element, and the full CI matrix can be represented symbolically as,

$$\begin{vmatrix} H_{00} & 0 & H_{0\mathrm{D}} & 0 & 0 & \cdots \\ & H_{\mathrm{SS}} & H_{\mathrm{SD}} & H_{\mathrm{ST}} & 0 & \cdots \\ & & H_{\mathrm{DD}} & H_{\mathrm{DT}} & H_{\mathrm{DQ}} & \cdots \\ & & & H_{\mathrm{TT}} & H_{\mathrm{TQ}} & \cdots \\ & & & & H_{\mathrm{QQ}} & \cdots \\ & & & & & \vdots \end{vmatrix} \tag{3.45}$$

where the subscripts $S$, $D$, $T$ and $Q$ denote single, double, triple and quadruple substitutions respectively. Selection rules result in a number of blocks in (3.45) being zero. Brillouin's theorem [Brillouin 1932] results in the coupling between ground and singly-substituted wavefunctions being zero. Similarly, there are no couplings between the ground state and triple or quadruple excitations, and between singles

and quadruples. The reason for this is that all matrix elements of the Hamiltonian between Slater determinants which differ by three or more spin orbitals are zero.

By calculating all possible permutations of the wavefunction in this way the correlation energy, within the basis set, is included and the full CI method approaches an exact solution to the Schrödinger equation. The full CI method is size-consistent, meaning the relative errors increase in proportion to molecular size, and it is variational, meaning that the calculated energy is always larger than the true energy. Unfortunately, as discussed above the large number of determinants mean that full CI is only practical with systems of less than 10 electrons. Hence in calculations of molecular systems a limited form of CI is generally applied.

#### 3.2.4.3 Limited Configuration Interaction

The complexity of the full CI method can be reduced in a number of ways. The simplest involves truncating the series in (3.43). This approach is effective in reducing the computational complexity of CI calculations but is plagued by the lack of size-consistency. However, using the Davidson correction [Langhoff and Davidson 1974], which allows the contribution of quadruply excited determinants to be estimated and included without explicitly calculating them, it is possible to reduce the size consistency error significantly.

Restricting the determinant series to only single substitutions results in the CI-singles method (CIS), to only double substitutions results in the CI-doubles method (CID), and to single and double substitutions results in the CI-singles and doubles method (CISD). The CIS wavefunction therefore has the following form,

$$\psi = C_0\Phi_0 + \sum_{a,p} C_a^p \Phi_a^p \tag{3.46}$$

while the CID wavefunction is represented by,

$$\psi = C_0\Phi_0 + \sum_{\substack{a<b \\ p<q}} C_{ab}^{pq} \Phi_{ab}^{pq} \tag{3.47}$$

The CISD wavefunction is simply the sum of the substituted terms in these two expressions. It is only the double substitutions that improve the Hartree-Fock wavefunction by incorporating electron correlation since in the CIS method the matrix

in (3.45) is reduced to a single term and there is no lowering of the energy. Although the CIS method does include some non-dynamical electron correlation, it is typically used for the study of excited states [Foresman *et al.* 1992; Hadad *et al.* 1993; Wiberg *et al.* 1993]. This is the approach used in this work and it is discussed briefly below.

***The CIS Method for Calculating Excited States***

The CIS method is a state based, post SCF, method that attempts to calculate the wavefunction and energy of a given excited state without restrictions on the physics connecting the states. CIS gives wavefunctions of roughly HF quality for excited states, since no orbital optimisation is involved. This is reasonable for valence excited states such as those arising from excitations between $\pi$-orbitals in unsaturated systems. For excitation from low-lying orbitals, however, there is considerable contribution from double excitations and so these require at least inclusion of single and double excitations (CISD). The molecules studied in this work all involved excitation from delocalised $\pi$-systems and so the CIS method was deemed appropriate.

The basis of the CIS method involves applying the CI procedure to the orbitals of a ground state HF wavefunction to solve for each higher state. The CIS wavefunction is formed from a linear combination of Slater determinants formed by replacing each spinorbital in turn by a virtual orbital (eq. (3.46)). For single excitations this gives $n(N-n)$ singly excited determinants, the coefficients of which are eigenvectors of the Hamiltonian,

$$\left\langle \phi_{ia} \middle| \hat{H} \middle| \phi_{jb} \right\rangle = \left[ E_{HF} + \varepsilon_a - \varepsilon_i \right] \delta_{ij} \delta_{ab} - \left\langle ja \middle| ib \right\rangle \tag{3.48}$$

where equation (3.48) is expressed in the notation of Foresman *et al.* [Foresman *et al.* 1992] in which $i$ & $j$ denote molecular orbitals that are occupied in the ground state, and $a$ & $b$ denote virtual molecular orbitals, unoccupied in the ground state. $\varepsilon_i$ and $\varepsilon_a$ are the one-electron energies of the occupied orbitals $\phi_i$ and unoccupied virtual orbitals $\phi_a$ respectively, and $\langle ja|ib\rangle$ are the two-electron repulsion integrals associated with them. $\delta_{ij}$ and $\delta_{ab}$ specify the orthonormality of the orbitals $\phi_i, \phi_j$ and $\phi_a, \phi_b$ respectively. Brillouin's theorem (§3.2.4.2) means that,

$$\left\langle \phi_{ia} \middle| \hat{H} \middle| \phi_{HF} \right\rangle = 0 \tag{3.49}$$

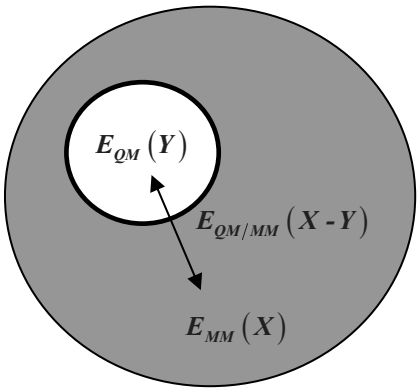



**Figure 3-6** Schematic of the QM/MM approach showing the division of a molecular system into a QM region (Y) and a MM region (X).

and so the CIS wavefunction is orthogonal to the ground state wavefunction. The total energy of the CIS wavefunction is given by,

$$E_{CIS} = E_{HF} + \sum_{ia} a_{ia}^2 \left(\varepsilon_a - \varepsilon_i\right) - \sum_{ij,ab} a_{ia} a_{jb} \left\langle ja \middle| ib \right\rangle \tag{3.50}$$

from which the excited state properties, and hence energy gap between the ground state and each excited state, can be determined [Foresman *et al.* 1992].

## 3.2.5 The QM/MM Method

The aim of the work described in this section of the thesis was to calculate the ground to excited state energy gap of the chromophore of the protein LADH as a function of time. From the proceeding discussion it is obvious that application of *ab initio* methods to a system the size of LADH (*ca.* 76,000 atoms when solvated) is not feasible using present computing resources. Hence truncation of the system of interest is required. This involves selecting the region of the protein that is mostly responsible for the property that is to be studied. The chromophore for absorption and emission spectra and non-linear spectroscopies, or the active site for reactivity. In LADH the chromophore (nicotinamide) is part of the NADH coenzyme which forms a sub-section of the active site. However, simply truncating the entire solvated protein to the system of interest would not give a reasonable approximation of the bulk system since the QM calculation would now be in gas phase and, while accounted for in the

molecular dynamics and hence structures used in the QM calculations, the effects of the electrostatic potential generated by the surrounding protein and solvent would not be included. In order to incorporate these effects within the CIS evaluation of the ground to excited state energy a combined QM/MM approach has been employed. The formalism of the method used in this work is discussed below.

#### 3.2.5.1 The Hybrid QM/MM Formalism

The QM/MM approach allows large systems to be treated using a method that combines the speed of classical molecular mechanics with the accuracy of quantum mechanical approaches. The use of a hybrid QM/MM approach involves partitioning a large chemical system into a quantum mechanically described electronically important region (usually the part of chemical interest such as an active site), and a remaining area which has only perturbative effects on the quantum region and can be dealt with classically (Figure 3-6). Models that utilise this description include those based on a semi-empirical description of the electronically important region [Warshel and Levitt 1976], a Hartree-Fock *ab initio* method [Singh and Kollman 1986; Field *et al.* 1990; Cui and Karplus 2000a; Cui and Karplus 2000b], and a density functional theory approach [Harrison 1999; Lyne *et al.* 1999; Murphy *et al.* 2000b; Murphy *et al.* 2000a].

Various methods exist for treating the two regions in a QM/MM calculation and more importantly the interface between the quantum and the classical region [Bakowies and Thiel 1996a; Bakowies and Thiel 1996b; Monard and Merz 1999]. The simplest technique for combining the QM and MM regions is called 'mechanical' embedding. In this method the total energy of two separate molecules X and Y are combined (3.51) to yield a total combined energy.

$$E_{total}\left(X-Y\right)=E_{MM}\left(X\right)+E_{QM}\left(Y\right)+E_{QM/MM}\left(X-Y\right) \tag{3.51}$$

where $E_{QM/MM}\left(X-Y\right)$ is the intermolecular interaction energy. All of the interactions between the QM and MM regions are included within the $E_{QM/MM}\left(X-Y\right)$ term and so these interactions are calculated classically. In practice it is rare to consider two separate molecules and instead a molecular system is generally split into a QM and a MM fragment. In this situation it becomes necessary to account for the intersection between the two fragments. Clearly where the intersection between the QM and MM

regions cuts covalent bonds it is necessary to ensure that the correct valencies are maintained. The most common method for achieving this is through the use of link atoms [Singh and Kollman 1986; Field *et al.* 1990; Cunningham *et al.* 1997; Harrison *et al.* 1997; Ranganathan and Gready 1997; Gao *et al.* 1998; Stanton *et al.* 1998; Antes and Thiel 1999; Kollman *et al.* 2002]. These are incorporated in a chemically intuitive way into the QM fragment, a $CH_2$-O-C link might be replaced by a $CH_2$-O-H group for example. The energy of the fragments including the link atoms must be considered and accounted for when determining the total energy of the system. Fortunately the section of interest in the LADH model used here, the NADH coenzyme, is not covalently bound to the enzyme. Similarly for the myoglobin models discussed in Chapter 6 the system of interest is the porphyrin based zinc substituted heme unit that can conveniently be described as an unbound residue within the protein matrix. Thus by choosing the QM section carefully it has been possible to avoid the need for link atoms. All further discussion will therefore treat the QM and MM sections as being essentially separate units.

Covalent interactions across the QM/MM boundary are not the only interactions between the two units that need to be considered. It is also necessary to account for the long range van der Waals and electrostatic interactions (§2.2.1.3). In the simple 'mechanical' embedding model discussed above (eq. (3.51)) these long range interactions are calculated classically. In this work a hybrid QM/MM model, implemented within Gaussian 98 [Hall and Smith 1984; Smith and Hall 1986], has been used that improves upon the simple 'mechanical' model by calculating the interactions quantum mechanically. During the single point excited state energy evaluations, carried out on each structure from the MD production runs, the partial atomic charges from the molecular mechanical region (§2.2.1.6.3) are explicitly included in the one-electron Hamiltonian,

$$\hat{H}_{el}(Y;X) = \hat{H}_{el}(Y) - \sum_{i}^{X}\sum_{j}^{Y}\frac{q_j}{r_{ij}} \tag{3.52}$$

where $\hat{H}_{el}(Y;X)$ and $\hat{H}_{el}(Y)$ are the electronic Hamiltonians for $Y$ with and without the external field respectively. The external field is due to the partial charges $q_j$ of the MM region. These charges are the ones derived from the RESP calculations (§2.3.3.1) and used for the classical MD simulations. Although the magnitude of these charges are fixed over the period of the MD simulations, due to computational

complexity considerations, the positions of these point charges vary as a function of time and so the electrostatic field, due to the MM region, experienced by the QM region changes as a function of time. The inclusion of the MM charges modifies the core Hamiltonian as follows,

$$\bar{H}_{\mu\upsilon}^{core} = H_{\mu\upsilon}^{core} - \sum_{j}^{Y} q_j V_{\mu\upsilon}^{j} \tag{3.53}$$

and the resulting Coulombic (or electrostatic) energy $E^{Coul}\left(X-Y\right)$ is described by,

$$E^{Coul}\left(X-Y\right) = \sum_{\mu}^{X}\sum_{\upsilon}^{X}\sum_{j}^{Y} P_{\mu\upsilon} q_j V_{\mu\upsilon}^{j} + \sum_{A}^{X}\sum_{j}^{Y} Z_A q_j V^{Aj} \tag{3.54}$$

where $P_{\mu\upsilon}$ are the density matrix elements and $Z_A$ the nuclear charges. $V_{\mu\upsilon}^{j}$ are the nuclear attraction integrals and $V^{Aj}$ the Coulomb repulsion terms describing the interaction between a unit charge on an MM region atom $j$ and a QM electron or nucleus $A$ respectively.

This approach results in the SCF energy of the QM region $\bar{E}_{QM}\left(X\right)$ being modified as a result of the altered core Hamiltonian, resulting in a net gain in energy $E^{pol}\left(X\right)$ referred to as the polarisation or induction energy due to the inclusion of the MM partial charges within the QM region,

$$E^{pol}\left(Y\right) = \bar{E}_{QM}\left(Y\right) - E_{QM}\left(Y\right) - E^{Coul}\left(X-Y\right) \tag{3.55}$$

Additionally the van der Waals interactions could also be included in the QM/MM treatment but since these do not change in going from the ground to excited state it was concluded that they could be neglected, in order to keep the approach as simple as possible, without adversely affecting the results. Conceptually similar implementations have also been developed [Thompson *et al.* 1994; Thompson 1995; Thompson and Schenter 1995]. More complete treatments extend the method to include the polarisation of the MM term as well as that of the QM term. In this work only single point energy calculations have been conducted although analytic gradients on similar methods have been developed that are suitable for use in molecular dynamics simulations [Thompson 1996]. Such an approach would allow the energy gap to be calculated at the same time as the MD using a hybrid QM/MM MD simulation. Unfortunately the linear response approach, discussed in detail in

Chapter 4 requires the fluctuation in energy gap for a system at equilibrium and while trial simulations using a QM/MM MD approach, implemented in the program ROAR 2.1 [Cheng *et al.* 2002], showed it to be feasible for the 10 ps production phases of the MD simulations it's computational demand was too high to enable effective equilibration. Besides, the approach used in this work, classical MD coupled to QM/MM evaluation of energy gaps via single point energy calculations, describes the equilibrium fluctuations between ground and excited state sufficiently well to recover the reorganisation energy to within 13 % of experiment.

## 3.3 Calculations

The aim of the calculations described here were to evaluate the LADH chromophore energy gap between ground and excited state as a function of time. A number of factors influenced the choice of methodology used in these calculations. One aspect, the choice of the CIS method for evaluating the excited state energy gaps is discussed in detail in Chapter 5. One of the biggest influences on the procedure was the computational requirement. The results discussed in Chapter 4 show that convergence of the energy gap auto-correlation function requires a minimum sampling of approximately 5 ps of equilibrated phase space. Section 4.2.2 outlines the method used for calculating time correlation functions. The important point is that a number of discrete Fourier transforms are required in the generation of the auto-correlation function and spectral density. In order to avoid aliasing problems in the Fourier transforms it is necessary to sample at sufficiently small intervals so as to accurately describe the waveform of the highest frequency oscillation. In the energy gap calculations it was found that the highest frequency oscillation had a period of just under 10 fs. It was therefore necessary to sample the ground to excited state energy gap at a frequency no less than once every 2 fs. This put a lower bound of 2,500 on the number of CIS calculations required to produce a converged auto-correlation function. The large number of CIS calculations required posed several problems. These are discussed below.

### 3.3.1 Logistics and Computational Considerations

The logistics of running upwards of 2,500 single point QM/MM CIS calculations on structures obtained from a MD simulation of a solvated protein are immense. In order to perform the CIS calculations it was necessary to extract each structural snapshot from the MD production phase and convert this to a Gaussian 98 [Frisch *et al.* 1998; Frisch *et al.* 2001] input file. Each input file then needed to be calculated, checked for correct completion, and then the various properties extracted.

The size of system involved (*ca.* 76,000 atoms) made any attempt at manually performing this process impossible. Thus one of the biggest tasks, in the application of the QM/MM based linear response method to proteins, has been the development of a framework for automating the process between running an MD simulation and calculating the spectra. In the next few sections will be discussed the various procedures used and the programs written to overcome the logistical problems.

#### 3.3.1.1 Extraction of MD Structures and Creation of Input Files

The AMBER driven MD simulations produce a single coordinate file containing each structure produced during the MD production phase. Each 'frame' of this coordinate file contains the X, Y and Z coordinates, in 8 block wide single byte ASCII format. Thus each atom in a frame requires 24 bytes of storage. Each 76,000 atom frame of LADH thus requires 1.824 MB of disk storage. A 5,000 frame coordinate file is therefore 9.1 GB in size. This poses a problem if 32 bit *i*x86 based Linux machines are used since the maximum file size when using unsigned 32 bit integer pointers is 4 GB. Although 64 bit integer offsets are now available with most compilers for Linux kernel versions later than 2.4.0, this was not available at the time this work was initially implemented.

Implementing a binary read-write routine, using single precision floats, for the coordinate files allows the file size to be reduced to 912 KB per frame. This is still over 4.5 GB for a 10 ps production run, however. The use of a binary file format for the coordinate file can also prove troublesome since it is not supported by most visualisation programs and post MD manipulation of the file is difficult. Writing a single file for each frame is a possibility, although 9.1 GB of storage per simulation would still be required, but this would require significant alterations to the read-write routines within the AMBER software. Fortunately it is fairly simple to implement an

on the fly compression system based on the bzip2 compression algorithm [Seward 2003]. This is a "freely available, patent free, high-quality data compressor" [Seward 2003]. The library implements streaming compression which means data streams can be compressed and decompressed on the fly without the need to go via a fully decompressed file. By producing a filter at the input and output stage of all programs used in this work it was possible to keep all data files in a compressed state on disk. In this way the 9.1 GB coordinate files never exceeded 2.4 GB in size, a compression ratio of 3.8 : 1. Although the compression routines added computational overhead to any file IO the impact was minimal since the computational bottleneck was the time taken to carry out the QM/MM calculations.

In order to create each Gaussian input file from the MD coordinate data a series of perl scripts were created. These used a 'Bzip2' modified version of Carnal[18] to extract each individual structure from the compressed coordinate file. The script divides the structure into the QM and MM sections and then writes out, along with the relevant Gaussian header information, a series of sequentially numbered compressed Gaussian input files, one for each structure in the coordinate file. A flow chart of the process for generating the Gaussian input files is shown in Figure 3-7.

Once the input files were obtained an automated procedure, discussed below, was used to process them via custom built computer clusters.

#### 3.3.1.2 Automation of Calculations and Speed Improvements

The most time consuming aspect of this work involved the post processing of the MD runs to produce time ordered ground to excited state energy gaps via a QM/MM method. This area of the work had the greatest scope for optimisation and so a number of aspects of the calculations were examined. Optimisation of the calculations in ways that had no effect on the result are discussed here. Changes involving the result generated by each calculation, such as changing the basis set, are discussed in section 3.3.2.

[18] Patch scripts for adding Bzip2 support to Carnal v6.0 and v7.0 are available on the accompanying CD.

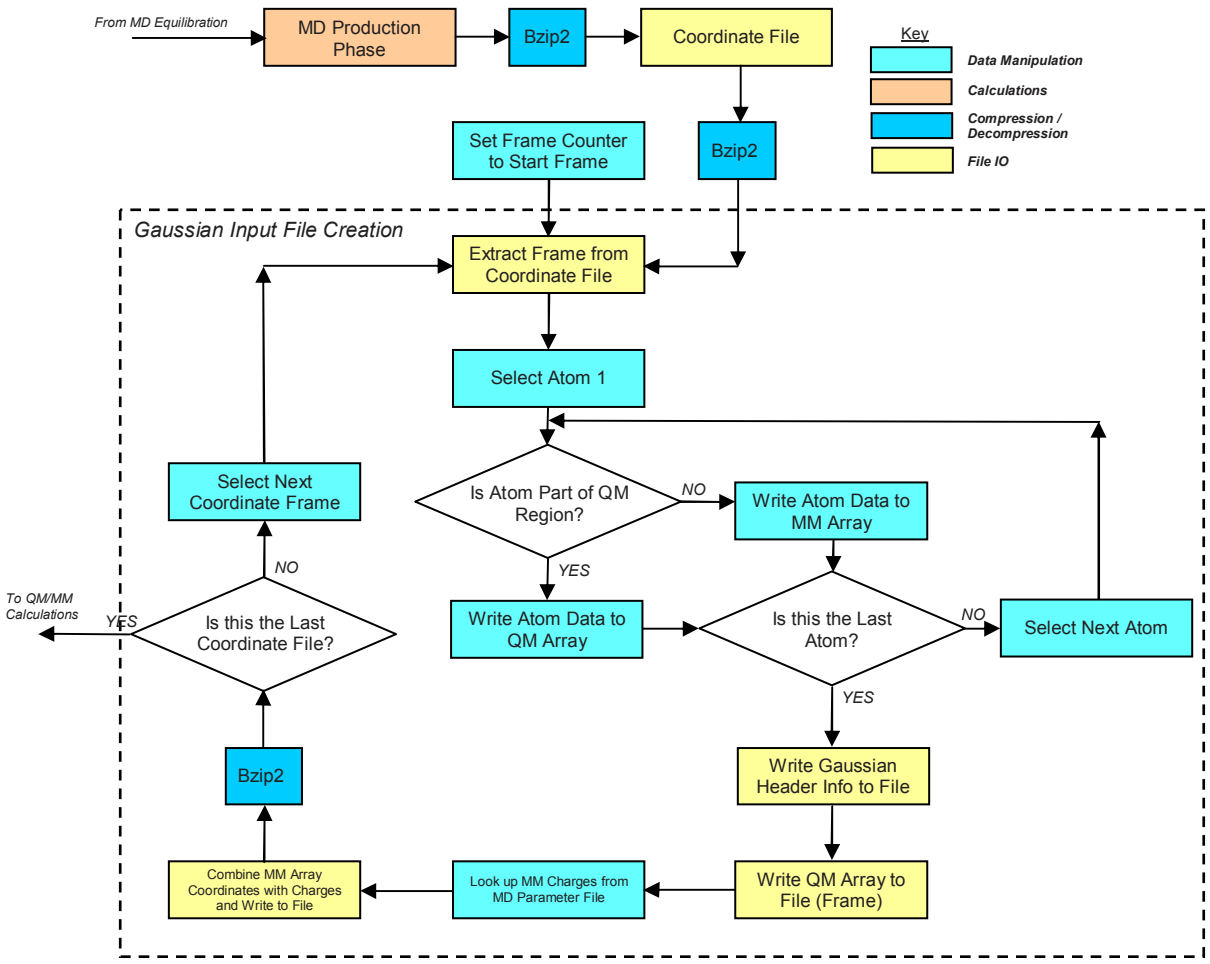



**Figure 3-7** Flow chart illustrating the procedure for converting the MD production run coordinates to QM/MM Gaussian input files.

### *Cluster Load Balancing*

The first problem to address involved a way of utilising all of the available computing power. While the jobs to be run could have simply been divided up between the available machines, in the heterogeneous environment available load balancing issues were a problem. This was made more acute by the fact that a very large number of idle desktop machines had sufficient power to run the single point QM/MM calculations. Initial attempts to use widely available "*fork and forget*" load balancing programs such as Mosix [Barak and La'adan 1997; Barak and La'adan 1998; Amir *et al.* 2000; Keren and Barak 2003] and ClumpOS [Marrow 2002] proved unsuccessful as their demands on the networking infrastructure were too great.

By creating a purpose written automated system for distributing the jobs utilising a server-client approach, but using local disk for temporary files, it was possible to efficiently use any available machine that could be booted into Linux. In this way users desktops could be utilised overnight, while idle, for running calculations. When the user returned in the morning they simply reset their machine, the currently running job would stop and be farmed out by the server to an alternative machine. Figure 3-8 shows a flow chart of the load balancing system used for the QM/MM calculations.

***Optimising the QM/MM Performance***

One way of reducing the time required to perform each energy gap evaluation was to optimise the speed of convergence of the SCF and CIS components of the calculations. By exploiting the fact that the structures changed by only a small amount between each frame it was possible to reduce the number of steps required in the SCF, by an average of about 30 %, by reusing the Hessian matrix from previous calculations. An initial guess is normally made for the Hartree-Fock wavefunction via the semi-empirical INDO method [Pople *et al.* 1967]. While this provides a reasonable guess, the SCF optimised wavefunction from a calculation on an electronically similar system is much closer and so quicker to converge. This was achieved in practice by having the server distribute a checkpoint file from the closest completed calculation with each new job. In a similar fashion the time taken for the CIS section of the calculation was reduced by reading an initial guess for the CI-Singles states from the checkpoint file of a previous calculation.

The time taken for the CIS portion of the QM/MM calculation was also reduced by relaxing the excited state energy convergence criterion in the iterative matrix diagonalisation routine used by Gaussian 98 from the default of $10^{-6}$ eV to $10^{-3}$ eV. Since the calculations were single points this was acceptable. Had the calculations been optimisations a convergence criterion of only $10^{-3}$ eV would have led to a magnification of the errors at each step. Trial calculations showed that relaxing the CIS convergence criterion had no measurable effect on the spectral densities of the auto-correlation functions calculated from the energy gap fluctuations. Attempting to reduce the SCF density matrix convergence criterion from the default of 8 decimal places to 6 decimal places was found to affect the results, however, and so a convergence criterion of 8 decimal places in the density matrix was used for the SCF portion of the QM/MM calculations.

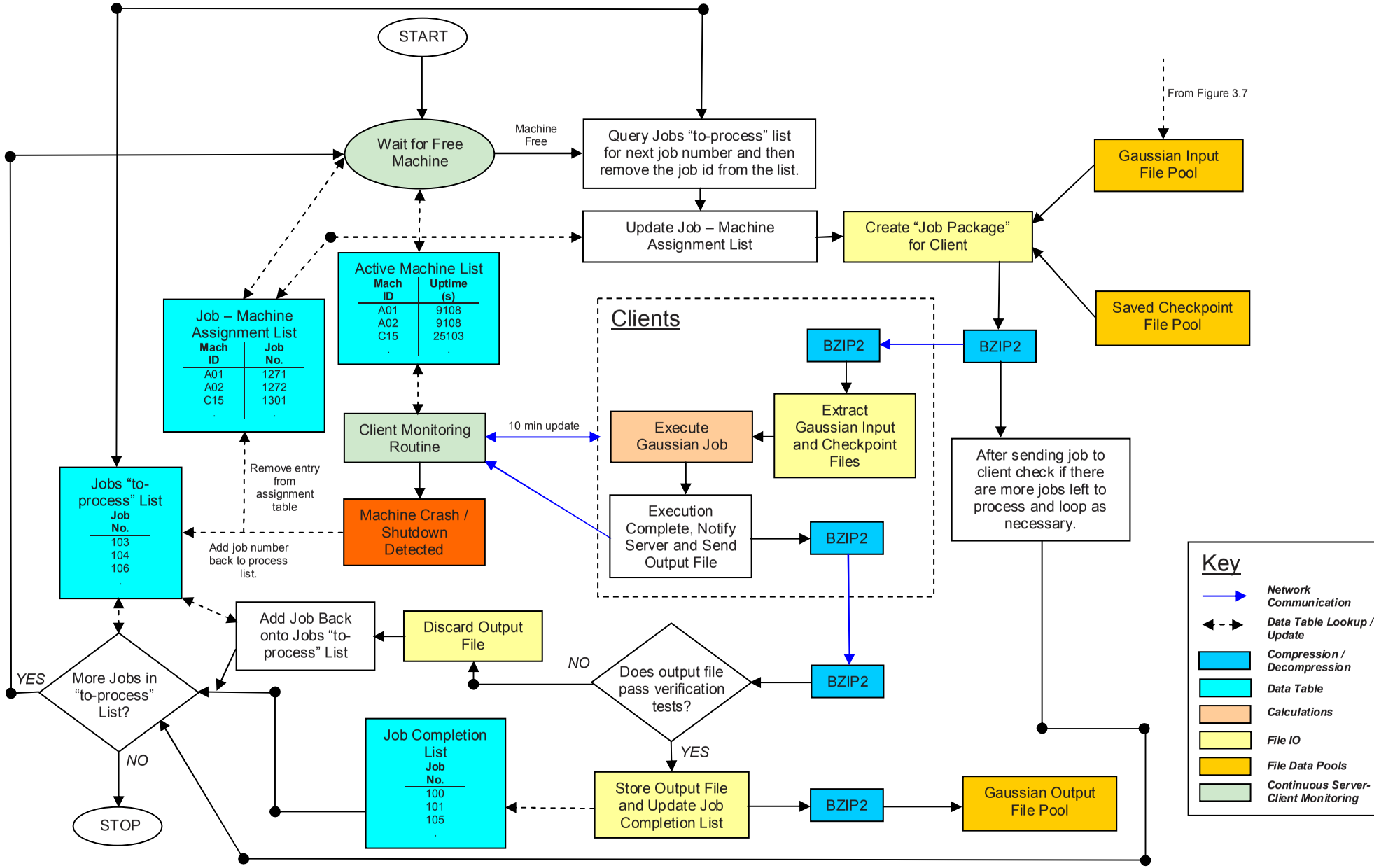



**Figure 3-8** Schematic showing the monitoring processes, data organisation and load balancing involved in the client-server based heterogeneous cluster environment, created maximising the utilisation of the available computing resources.

### 3.3.2 Optimisation of CIS Method

Computational complexity considerations meant that the choice of basis set and QM fragment size was paramount. The aim of accurately calculating the energy gaps drove the methodology towards a large QM fragment size and a large basis set while considerations of the available computing power drove the methodology towards a small QM fragment and a small basis set. To optimise the protocol used, in order to give a system that would give good agreement with experiment while utilising the minimum in computing power, a number of trial CIS calculations were carried out on the NADH co-enzyme. In order to avoid the complications of covalent bonds crossing the QM/MM boundary it was decided to use the full NADH structure for the CIS calculations. While some recent trial work has shown that a truncated NADH unit can produce the same results as using the full NADH, in less computing time, at the time these calculations were run it was decided to minimise the scope for error by avoiding unnecessary complications. The trial CIS calculations are discussed below.

#### 3.3.2.1 Computational Procedure

##### 3.3.2.1.1 Software and Platform Details

All calculations discussed in this section were performed using an SSE optimised version of Gaussian 98 Revision A11 [Frisch *et al.* 2001] compiled using the Linux optimised Intel Fortran compiler v7.0[19]. This gives a performance improvement of between 10 % and 20 % over the default Portland Group[20] compiled version. In order to allow the calculation times to be directly compared the calculations were all run on the same hardware platform, a dual Intel 1 GHz PIIIEB with 1GB SDRAM running RedHat Linux v7.3.

##### 3.3.2.1.2 Computational Methodology

Single point CIS calculations using various basis sets were carried out on gas phase NADH structures to evaluate the performance of each basis set, both in terms of the results obtained and the memory and wall clock time requirements.

---

[19] http://www.intel.com/software/products/compilers/flin/

[20] http://www.pgroup.com

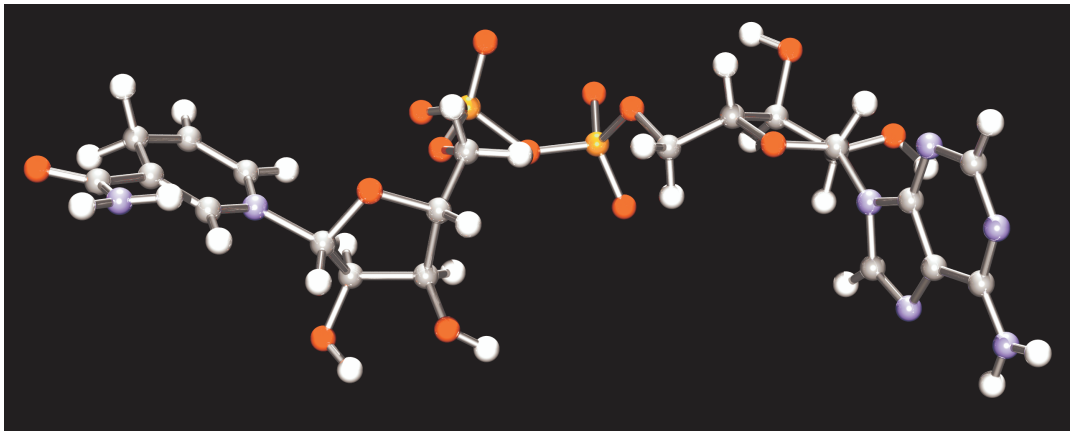


**Figure 3-9** The structure of the NADH co-enzyme, with the hydrogen atoms optimised at the HF level of theory using the STO-3G basis set.

The initial structure for the NADH unit was taken from crystallographic data [Ramaswamy *et al.* 1997]. Hydrogen atoms were added manually at standard values for the bond lengths and angles [Lide 1996]. In line with the methodology used for the MD simulations (Chapter 2) the two phosphate groups were assumed to be deprotonated giving the system an overall charge of -2. In all calculations the multiplicity of the system was specified as a singlet. The positions of the hydrogen atoms in the resultant structure were optimised at the HF level using the STO-3G basis set (Figure 3-9). The positions of the heavy atoms were fixed throughout the computation.

A series of single point CIS calculations were then performed on the resulting structure. The CIS calculations were restricted to the first 4 excited states. In all calculations the orbital space was not restricted and no symmetry constraints were placed on the NADH unit. The influence of the basis set size on the results obtained and the wall clock time required was investigated by comparing the excited state results obtained by running the single point CIS calculation using the STO-3G, STO-3G* [Hehre *et al.* 1969; Collins *et al.* 1976], 3-21G [Binkley *et al.* 1980], 3-21G*[Pietro *et al.* 1982], 6-31G* and 6-31G** [Hariharan and Pople 1973] basis sets. It should be noted that while the 6-31G* and 6-31G** basis sets include polarisation functions for all first and second row atoms, the STO-3G* and 3-21G* basis sets provide polarisation functions only for second row elements, in this case the two phosphorus atoms.

### 3.3.2.2 Results and Discussion

Table 3-1 summarises the results obtained from the single point CIS calculations. The energies for all excited states with oscillator strengths greater than zero are presented for each basis set used. States with oscillator strengths higher than 0.1000 are highlighted. The time required for each 4 state CIS single point calculation is also given. Figure 3-10 shows stick representations of the main peaks from the 3-21G*, 6-31G* and 6-31G** calculations. The molecular orbitals that form the chief components of the two transitions are shown in Figure 3-11.

It is immediately obvious from Table 3-1 that the use of the STO-3G and STO-3G* minimal basis sets are insufficient for excited state calculations on NADH. In both cases all states have oscillator strengths of zero or very close to zero. Expanding the state restriction to 16 states does not fix this problem with STO-3G yielding no states with oscillator strengths greater than 0.001. The split valence basis sets all yield reasonable results giving two strong excitations with a relative separation of approximately 1.2 eV. In going from 3-21G to 6-31G** the relative separations and oscillator strengths of these two excitations is largely unaffected. The larger basis set just results in a shift to lower energy. The two major excitations found originate from distinctly different parts of the NADH molecule. This is made apparent by considering the molecular orbitals involved in the two excitations. Figure 3-11 shows that the lower energy excitation involves orbital 169 (HOMO-4) exciting to orbital 176 (LUMO+2), both of which are centred on the nicotinamide section of the NADH. This figure also shows that the higher energy excitation involving orbital 164 (HOMO-9) exciting to orbital 174 (LUMO) is centred on the adenine end of the NADH molecule. Since it is the nicotinamide portion of the NADH that is responsible for the fluorescence it is the excitation corresponding to this part of the NADH that is required when calculating the absorption and emission spectra using the linear response approach.

The experimental absorption maximum for NADH in water is 3.64 eV (340 nm) for NICH absorption and 4.76 - 4.84 eV (260nm – 255nm) for the adenine. This gap is very similar to that calculated in the single point CIS calculations although the absolute locations of the peaks are overestimated by approximately 1.1 - 1.3 eV. This is to be expected since the calculations were single point energy evaluations on a crystal structure and so represent only a single structural snapshot. This calculation

was also carried out in gas phase while the experimental data is for NADH in aqueous solution.

#### 3.3.2.3 Conclusions

While the calculations show a large offset from the experimental data the basis of this work has been to investigate the magnitude of equilibrium fluctuations rather than absolute energies. Hence obtaining a energy fluctuation mean that equated directly with experiment was not an overriding factor in the choice of methodology. Since the size of the basis set had little effect on the magnitude or relative separation of the NADH excitations past 3-21G the use of large basis sets was deemed unnecessary for calculating the equilibrium fluctuations. The calculation times shown in Table 3-1 clearly show that in going from the 3-21G* basis set to the 6-31G* basis the computational cost is increased by a factor of over 4.1. Since computational efficiency considerations were paramount in this section of the work it was decided that the use of a basis set larger than 3-21G* would require an excessive amount of cpu time with little gain in the accuracy of the results. The 3-21G* basis set was therefore used for all of the LADH+NADH equilibrium energy gap evaluations discussed in section 3.3.3.

### 3.3.3 QM/MM Calculation of Energy Gaps

#### 3.3.3.1 Software and Platform Details

All QM/MM calculations discussed in this section were performed using Gaussian 98 Revisions A7 [Frisch *et al.* 1998] and A11 [Frisch *et al.* 2001]. The calculations were run, using custom load balancing software, on commodity Intel (PIII & P4) and AMD (XP & Opteron) PC clusters designed and built from scratch specifically for this project as well as a cluster of 24 Compaq DS10 Alpha EV6 machines and a 44 processor R10K SGI Infinite Reality Onyx 2. The Linux machines ran RedHat Linux versions 7.1, 7.2, 7.3 and 8.0 while the Compaq Alpha based machines ran Compaq-Tru64 Unix Version 4.0D[21]. The SGI Onyx 2 ran SGI's Irix operating system v6.5[22]. The resulting Gaussian output files were processed, and the relevant data extracted,

[21] http://www.tru64unix.compaq.com/

[22] http://www.sgi.com

using software developed specifically for this task. This software forms part of a computational chemistry tools suite developed by the author of this work. Version 1.8B3 (Windows XP) of this software, along with a description of each component, is available on the CD-ROM included with hard copies of this thesis. The latest version of this software can be downloaded free of charge from http://www.rosswalker.co.uk/.

#### 3.3.3.2 Computational Methodology

The trajectories obtained from the 10 ps production runs of the MD simulations, discussed in Chapter 2, were used to create time-ordered structures for the QM calculations. Single point CIS [Foresman *et al.* 1992] calculations utilising the 3-21G* [Binkley *et al.* 1980] basis set were performed to calculate the singlet energy gap between the ground and first excited state for the system of interest at 2 fs intervals. The use of the 3-21G* basis set, as discussed above, was to ensure that the calculations were tractable to the available computing resources. For reasons already discussed the NADH residue was treated quantum mechanically, while the rest of the protein and surrounding waters were treated classically. This gave a QM system consisting of 71 atoms corresponding to 470 basis functions (777 primitive Gaussians). The point charges of the classical system were included in the one-electron Hamiltonian of the quantum element of the calculations (§3.2.5.1).

For MD simulations 1, 2 and 4 a total of 2,500 points representing the first 5 ps of the 10 ps production runs were calculated for both NADH residues (ID numbers 375 and 753) yielding a total of 5,000 CIS calculations per MD simulation. For simulation 3 it was obvious after 1,800 points had been calculated for NADH residue 753 that the equilibrium dynamics had not been accurately reproduced by the MD simulation and so in the interests of not wasting cpu time the calculations were halted. Thus only limited QM/MM data for simulation 3 is available. 1,800 points is sufficient, however, to make broad conclusions concerning the parameterisation of classical MD simulation 3.

### 3.3.4 Results and Discussion

For simulations 1, 2 and 4 the properties from the equilibrium energy gap calculations were calculated over the first 2,500 points (5 ps) of each of the production runs for both NADH residues independently. For simulation 3 data for NADH residue

753 for only the first 1,800 points (3.6 ps) of the production run was available and so the properties for simulation 3 were calculated over the available data.

Figure 3-12 shows the energy gap fluctuations with time for simulations 1 to 4. The energy gaps due to each NADH residue in the LADH dimer were calculated separately and are shown by separate plots in Figure 3-12. The mean energy gap as a function of time is shown in Figure 3-13 along with the standard deviations. These were calculated by considering a series of boxes consisting of 25 points (50 fs) either side of the point of interest. Figure 3-14 and Figure 3-15 show discrete Fourier transforms of the excited state energy gap fluctuations for all four MD simulations.

The most striking thing about the energy gap fluctuations is the differences not just between the different simulations but between the two NADH residues in the same simulation. This difference is most striking in simulation 4 where residue 375 shows much more low frequency dynamics than residue 753. The differences between the two sites is made more apparent by considering the FFTs of the energy gap fluctuations. All of the energy gaps appear to show a very high frequency oscillation of around 3,000 $cm^{-1}$ to 3,500 $cm^{-1}$ which can be attributed to various different hydrogen oscillations. These high frequency oscillations are distributed over a large range of different frequencies, however, and there is no distinctive mode. Consideration of the auto-correlation functions, discussed in Chapter 4, show that these high frequency oscillations do not show a correlation over time and so contribute very little to the resulting spectra.

It is the lower frequency oscillations that are more interesting. Most of these show a correlation over time. The most obvious peak that is common to all of the simulations is around 1,600 $cm^{-1}$ to 1,900 $cm^{-1}$. This is believed to be due to the amide section of the NADH. Simulations 1 and 2 both used carefully calculated parameters for this section of the NADH and they both show sharp, intense peaks. Simulation 3 shows a similarly sharp peak but there are also several intense peaks around 500 $cm^{-1}$. The differences between the FFTs of the two NADH energy gap traces is most marked in simulation 4. Here it can be seen that there are a lot more very low frequency (< 100 $cm^{-1}$) oscillations in the residue 375 trace than there are in the residue 753 trace. This can be attributed to the distinctly different parts of phase space that the two NADH residues occupy over the production phase of the MD simulation (§2.3.4.4).

***Time Average of the Energy Gap***

An important aspect of the linear response approach to relaxation dynamics is that the energy gap fluctuations should represent the fluctuations about equilibrium. The system should be sufficiently equilibrated and the sample window should be large enough that an adequate portion of equilibrated phase space is sampled. What this means in practice is that the mean of the energy gap and its auto-correlation should not change over time. This property allows the auto-correlation functions to be calculated from the inverse Fourier transform of the spectral power density (Wiener-Khinchin Theorem [Haykin 1995] §4.2.2). This property is verified by the data in Figure 3-13. All of the plots, with the exception of simulation 4, residue 375, show a steady energy gap mean for the whole of the 5 ps of data. The mean energy gap for simulation 4, residue 375, however, is much more unstable reflecting the fact that the poor parameterisation of the original MD simulation has allowed the dynamics of the nicotinamide moiety of NADH to be over estimated. This data suggests that the spectral properties calculated from simulation 4 cannot be expected to yield valid results. All of the other simulations, however, would appear to show convergence of the energy gap averages.

## 3.3.5 Conclusions

In conclusion it can be seen that the energy gap fluctuation profile obtained originates almost entirely from the orbitals centred on the nicotinamide of the NADH coenzyme. It can be seen that varying the parameters on the nicotinamide moiety has an effect on the low frequency dynamics within the energy gap fluctuations and consequently is expected to affect the spectral predictions.

All of the calculations converged without trouble taking an average of around 16 scf cycles for each single point calculation. This, along with the stable averages in the energy gap traces suggests that the CIS method is capable of evaluating excited state properties from equilibrium fluctuations. The energy gap traces were thus deemed to be suitable for use in calculating optical observables using the linear response approach discussed next.

The approach used here could be improved by increasing the amount of phase space sampled, although the convergence of the Stokes Shift shown in section 4.4 suggests that the effect on the results would not be large. It could also be improved by

the use of periodic boundaries within the QM/MM evaluation of the energy gaps. The effect the accurate treatment of long range effects would have on the energy gap fluctuations is unclear, however. The computational overhead would also be large and it is unlikely that the extra effort would yield much improvement in the overall results. Concentrating on the accuracy of the classical MD simulations is likely to be more beneficial.

**Table 3-1** NADH excited state data calculated from a single point CIS calculation on the crystal structure using a range of different basis sets. States with oscillator strengths > 0.1000 are shown in bold.

| Basis Set | Excited State | Energy / eV | $\lambda$ / nm | Oscillator Strength, *f* | Calculation Time (s) |
|---|---|---|---|---|---|
| STO-3G | 3 | 4.2309 | 293.04 | 0.0002 | 26 mins 14.8 secs (1574.8) |
| STO-3G* | 1 | 4.8027 | 258.15 | 0.0001 | 29 mins 20.9 secs (1760.9) |
| | 3 | 5.2362 | 236.78 | 0.0005 | |
| 3-21G | **1** | **4.8879** | **256.27** | **0.1132** | 3 hrs 59 mins 53.6 secs (14393.6) |
| | 2 | 5.9001 | 210.14 | 0.0034 | |
| | **3** | **6.0931** | **203.48** | **0.4584** | |
| | 4 | 6.5212 | 190.12 | 0.0418 | |
| 3-21G* | **1**[23] | **4.8876** | **256.29** | **0.1130** | 4 hrs 43 mins 53.1 secs (17033.1) |
| | 2 | 5.8991 | 210.17 | 0.0034 | |
| | **3**[24] | **6.0933** | **203.47** | **0.4576** | |
| | 4 | 6.5203 | 190.15 | 0.0409 | |
| 6-31G* | **1** | **4.7143** | **262.99** | **0.1027** | 19 hrs 25 mins 23.7 secs (69923.7) |
| | **2** | **5.8996** | **210.15** | **0.4463** | |
| | 3 | 6.0931 | 203.48 | 0.0027 | |
| | 4 | 6.2011 | 199.94 | 0.0381 | |
| 6-31G** | **1** | **4.6898** | **264.37** | **0.1005** | 1 day 1 hour 16 mins 6.1 secs (90966.1) |
| | **2** | **5.8903** | **210.49** | **0.4494** | |
| | 3 | 6.0903 | 203.91 | 0.0027 | |
| | 4 | 6.1970 | 200.07 | 0.0361 | |

[23] Orbitals involved in this excitation (169 & 176) are shown in Figure 3-11

[24] Orbitals involved in this excitation (164 & 174) are shown in Figure 3-11

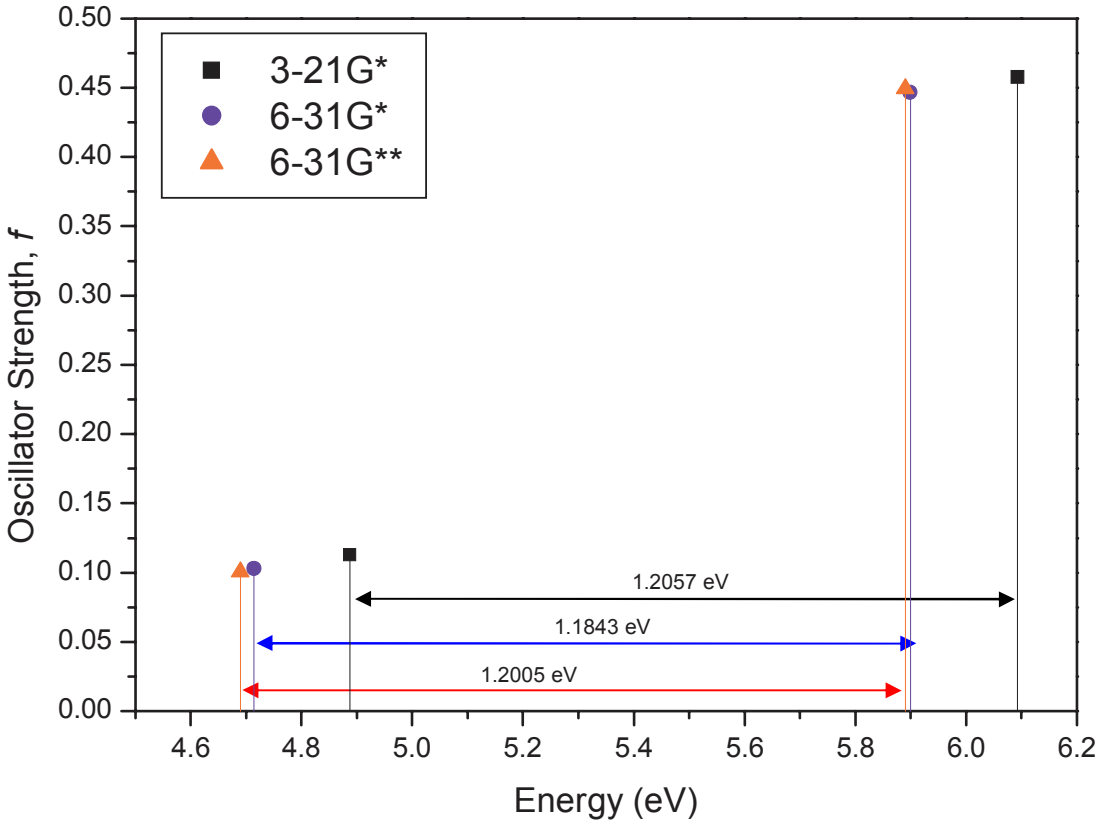



**Figure 3-10** Stick plot of the main peaks in the single point CIS calculations, described in Table 3-1, for the 3-21G*, 6-31G* and 6-31G** basis sets. The experimental absorption maximum for NADH in water is 3.64 eV (340 nm) for NICH absorption and 4.76 - 4.84 eV (260nm – 255nm) for the adenine. The location of the NADH peaks and the gap between them differs when bound to the protein matrix [Piersma *et al.* 1998]. Gas phase experimental data was not available for comparison.

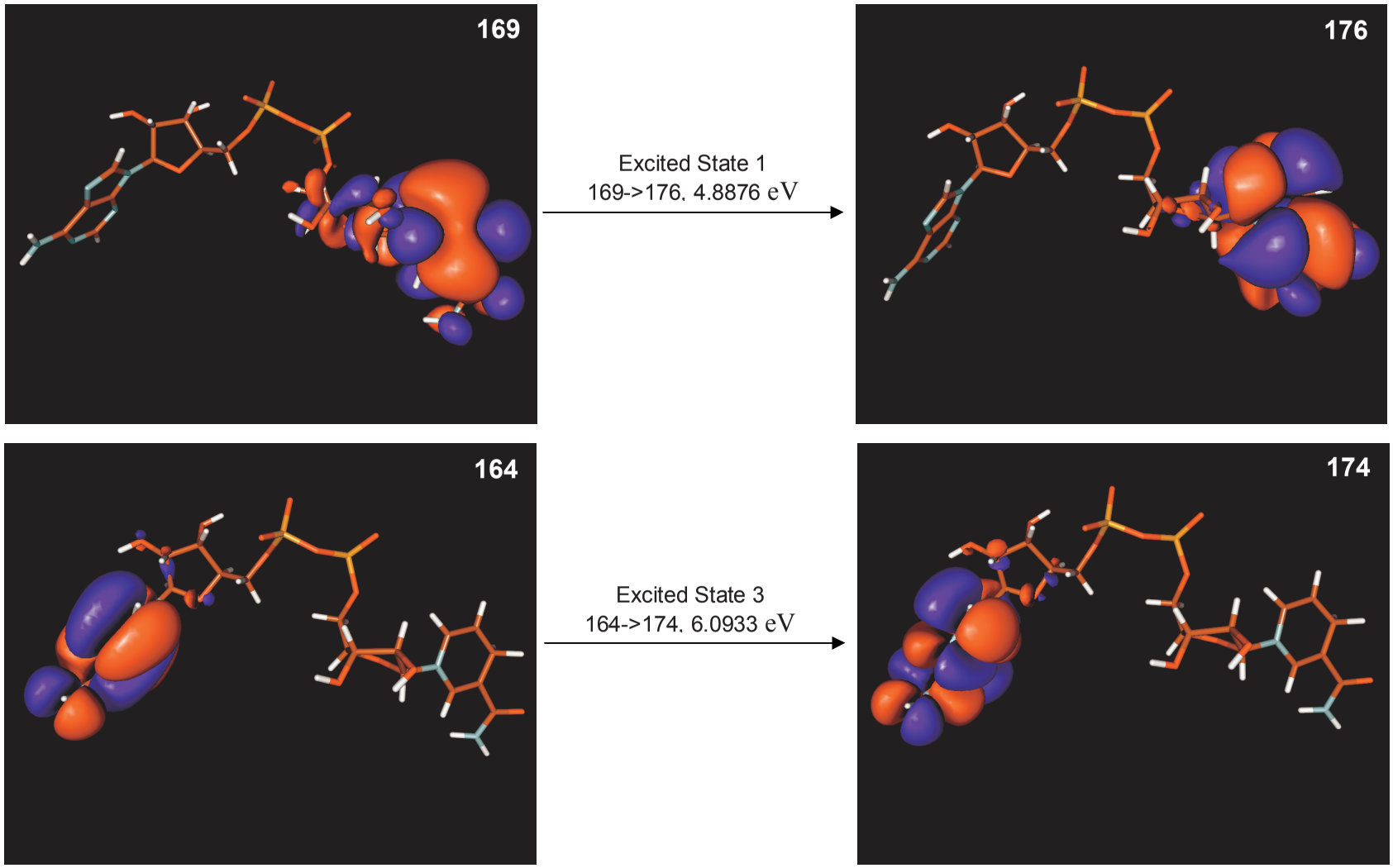



**Figure 3-11** Representations of the molecular orbitals that form the chief components of the two main ground to excited state transitions predicted from a single point CIS/3-21G* calculation.

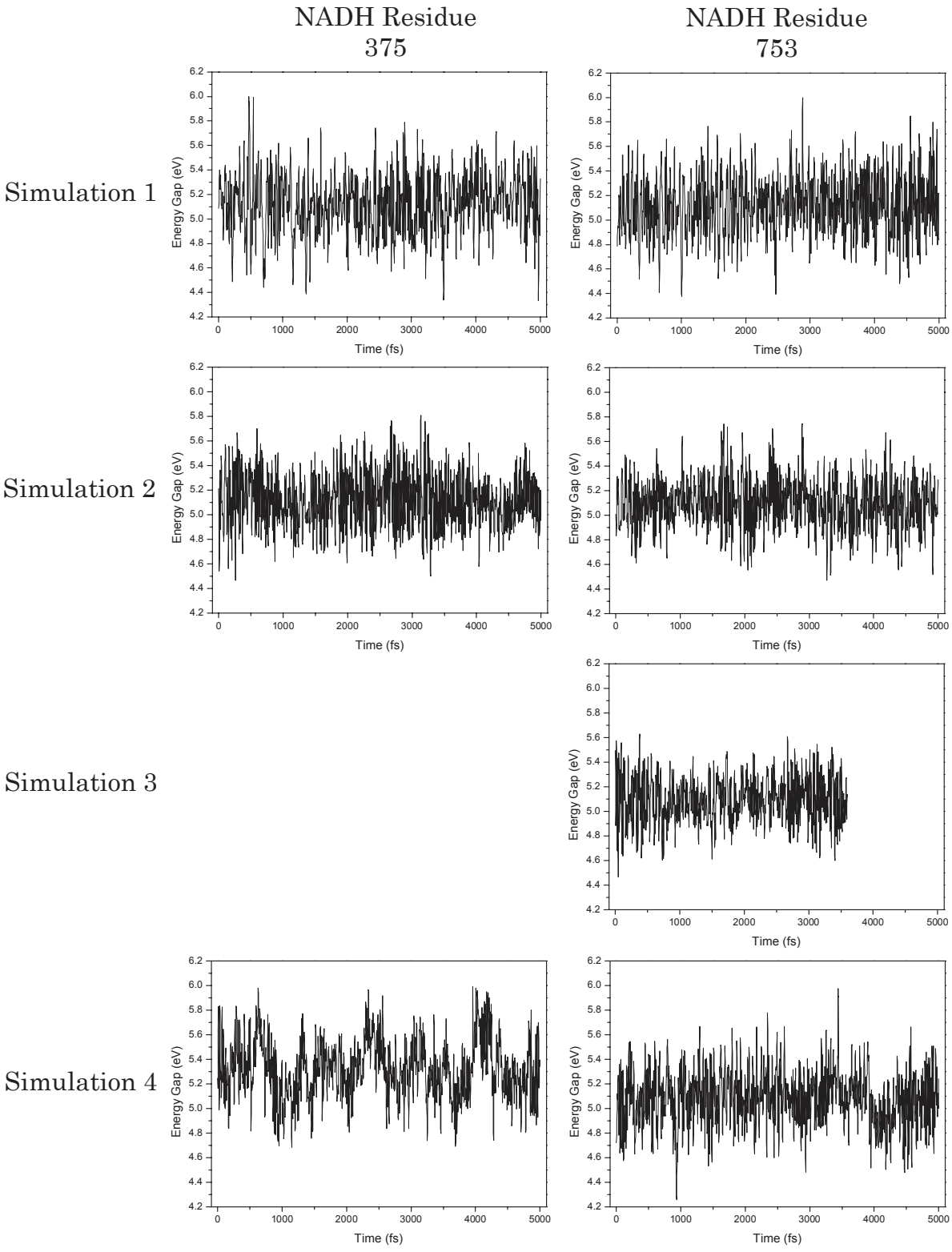



**Figure 3-12** Excited state energy gaps for NADH residue 375 in QM region (left hand column) and NADH residue 753 in QM region (right hand column) for the first 5 ps of the MD production runs. Computational resource constraints meant that only the first 1,800 (3.6 ps) points of residue 753 for simulation 3 were run. 2 fs sampling steps were used for all energy gap traces.

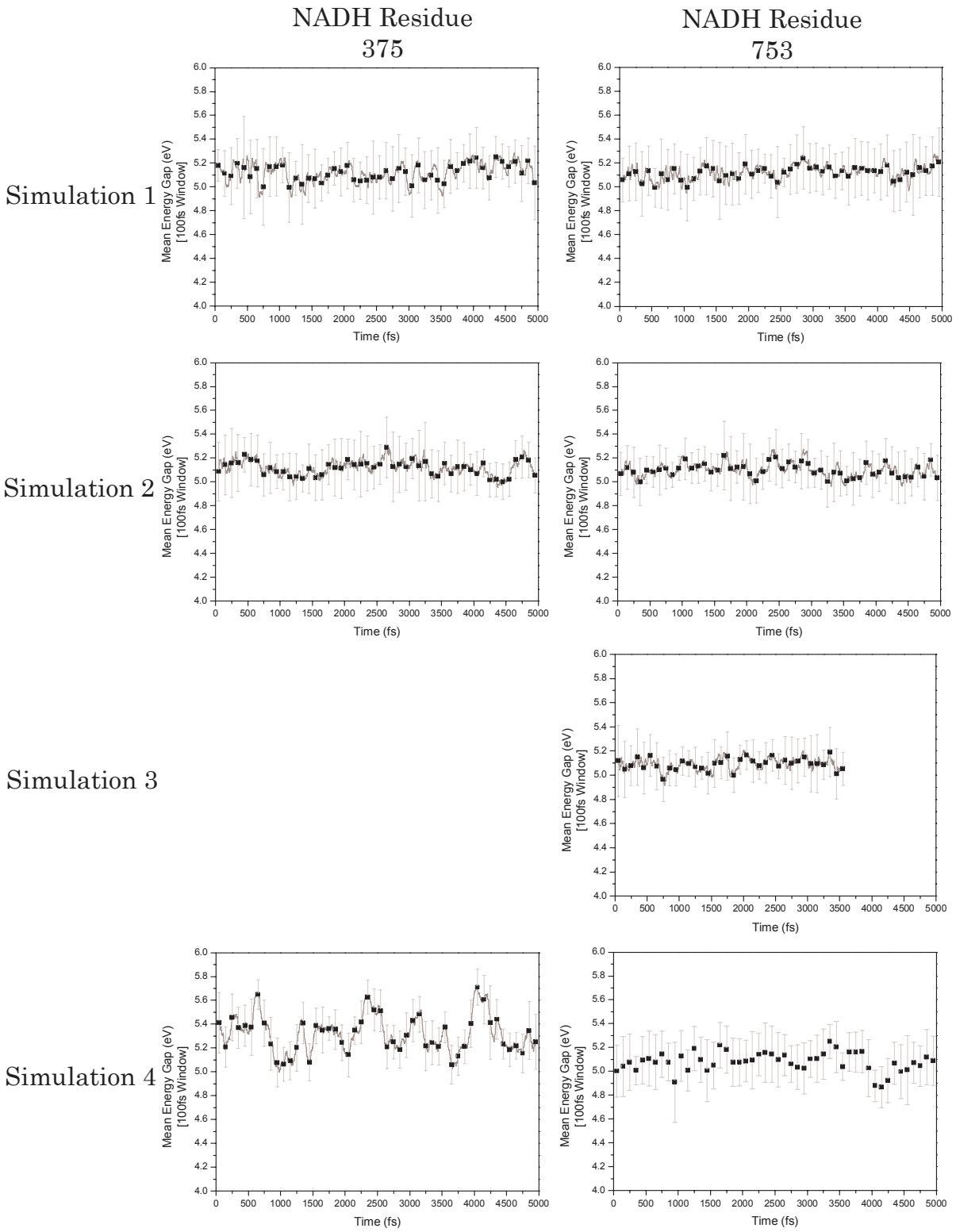



**Figure 3-13** Mean values for the energy gaps (black line) and standard deviations (grey lines) for MD simulations 1 to 4. The energy gap averages and standard deviations were calculated using a +/- 50 fs moving box car average.

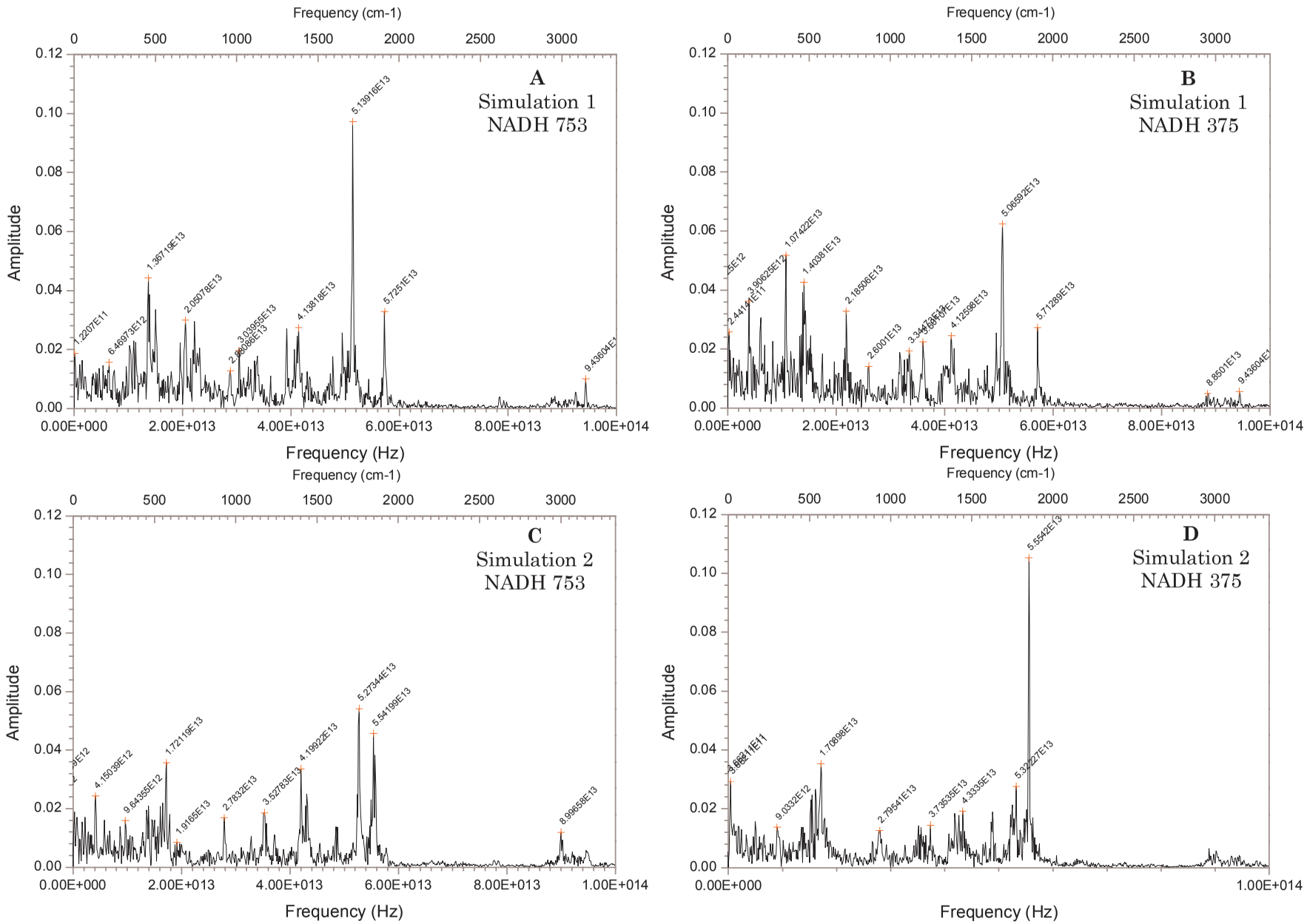



**Figure 3-14** Discrete Fourier transforms of excited state energy gaps for **A**) QM Res 753 Simulation 1, **B**) QM Res 375 Simulation 1, **C**) QM Res 753 Simulation 2, **D**) QM Res 375 Simulation 2, showing the component frequencies, and amplitudes, making up the energy gap traces. All FFTs were performed on the full 2,500 points (5 ps) of each energy gap sample.

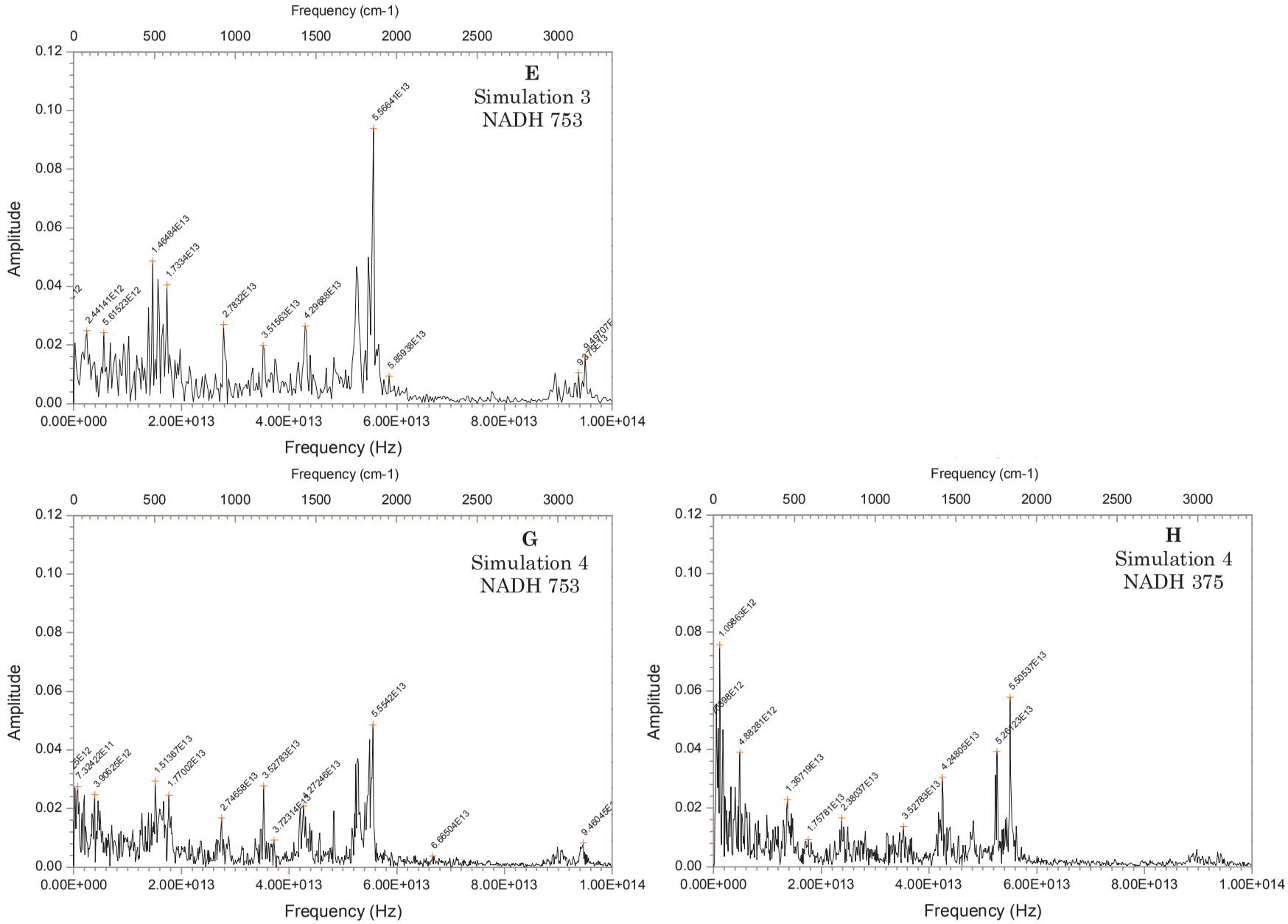



**Figure 3-15** Discrete Fourier transforms of excited state energy gaps for **E**) QM Res 753 Simulation 3, **G**) QM Res 753 Simulation 4, **H**) QM Res 375 Simulation 4, showing the component frequencies, and amplitudes, making up the energy gap traces. Simulation 4 FFTs were performed on the full 2,500 points (5 ps) of each energy gap sample while the FFT for simulation 3 was performed on the available 1,800 data points (3.6 ps).

# Chapter 4

# The Linear Response Approach

## 4.1 Introduction

The post-processing of the equilibrium energy gap fluctuations generated from the time ordered QM/MM calculations of Chapter 3 is discussed here and has been published in [Walker *et al.* 2002].

Central to the study of chemical dynamics in condensed phases is an understanding of how the environment influences the equilibrium energy fluctuations of a solvated macromolecule. It is these fluctuations that are sampled by the coupling of the QM/MM excited state calculations and the ground state classical MD simulations discussed in Chapter 2 & Chapter 3. The aim of the work in this chapter has been to use the excited state energy gaps to generate theoretical spectra that could be compared directly with experimental data in order to validate the protocol used.

The broad and featureless absorption and emission profile, typically observed in the spectra of condensed phase systems contains, in principle, all of the information required to characterise the dynamics of the chromophore. These dynamics represent the collective influence of the many inter- and intra-molecular processes of varying magnitude and time scale that affect the chromophore. The spectral broadening and its equivalent in the time-domain (the dephasing time scale) are an effect that arises from averaging over the ensemble of these processes [Fleming and Cho 1996]. It is these processes that are sampled in the molecular

dynamics simulations and ultimately expressed in the frequencies that make up the energy gap fluctuations. The energy gap fluctuations therefore contain, in theory, sufficient information to not only predict steady state spectra but also to allow the prediction of ultra-fast non-linear spectroscopic results. The identification of the time scales and magnitudes of the individual interactions that are coupled to the chromophore electronic energy gap, that taken together give rise to the overall width of the absorption spectrum, can be realised with ultra-fast non-linear spectroscopy [Cho 2003]. An example of this type of non-linear spectroscopy is 3-pulse echo peak shift (3PEP). While some preliminary theoretical 3PEPs spectra for both LADH and myoglobin [Amer 2001] have been generated from the data produced in this work the lack of experimental data for LADH and the problems associated with degeneracy in myoglobin (Chapter 6) mean that this is the subject of on-going work in this field. The aim of this work has been to show that the QM/MM driven linear response approach [Mercer *et al.* 1999] to relaxation dynamics could be successfully used for simulations of proteins and that the linear response approximation would hold for systems with large reorganisation energies. Thus the work shown in this thesis is restricted to producing accurate steady-state absorption and emission spectra. Comparison of the theoretical Stokes-shift and absorption widths with experimental data, however, can provide valuable information that validates the protocol used and hence the 'quality' of the energy gap fluctuations. With the knowledge that the classical MD simulation coupled to the QM/MM calculations can accurately reproduce the reorganisation energy it is then possible to use this data both to investigate the oscillations that contribute to the spectra (the Fourier transform of the energy gap auto-correlation function) and for explaining the features observed in ultra-fast non-linear spectroscopy experiments.

### 4.1.1 Linking Theory and Experiment

The relationship between theory and experiment allows the creation of a 'feedback loop' whereby the information obtained from one can be used to refine and explain the results of the other. An example of this is discussed in section 4.4.2 where the spectral results obtained from the QM/MM calculations are shown to be very sensitive to the 'quality' of the molecular dynamics parameters.

The link between theory and experiment in this work is provided by the application of elements of linear optical response theory to the energy gap fluctuations. The experimental steady-state absorption and emission spectra were generated by linear fluorescence Stokes-shift experiments while their theoretical counterparts were generated by making a series of assumptions and approximations (section 4.2) that facilitated the calculation of the degree of electronic-vibrational coupling. The theory behind the generation of optical observables from equilibrium fluctuations is discussed in section 4.2. This method has previously been used to investigate optical relaxations within chlorophyll *a* and bacteriochlorophyll *a* [Mercer *et al.* 1997; Mercer *et al.* 1999] which have Stokes-shifts of less than 100 meV. In this work the procedure has been automated, refined and successfully adapted for studying relaxation in proteins and has also been shown to hold for large deviations from equilibrium (Stokes-shift ≈ 0.9 eV).

### 4.1.2 The Origin of the Stokes Shift

The rapid change in electronic distribution which occurs after optical excitation can be regarded as one of the simplest types of chemical reaction. After excitation a liquid or protein system will attempt to re-establish equilibrium by undergoing a series of structural rearrangements that seek to minimise its energy. This leads to a lowering of the excited state energy and a raising of the ground state energy (Figure 4-1). This shift in energy levels results in the emitted light having a lower frequency than the absorbed light giving rise to a shift between the maxima of the absorption and emission spectra which is referred to as the Stokes shift. Strictly speaking the Stokes shift and the peak shift only coincide when the transition is a pure $v''=0$ to $v'=0$ transition. In this situation the absorption and emission spectra will be symmetric and Gaussian. The inclusion of other vibrational transitions within the ground and excited states leads to a series of side-bands within the absorption and emission spectra, known as Franck-Condon progressions. In such a case the Stokes shift is correctly obtained from the difference in the first moments of the absorption and emission profiles. The reorganisation energy upon excitation is thus half the Stokes shift. In the case of LADH the absorption and emission spectra (Figure 4-4) are, rather surprisingly considering the large Stokes shift, highly Gaussian in shape and very symmetric.

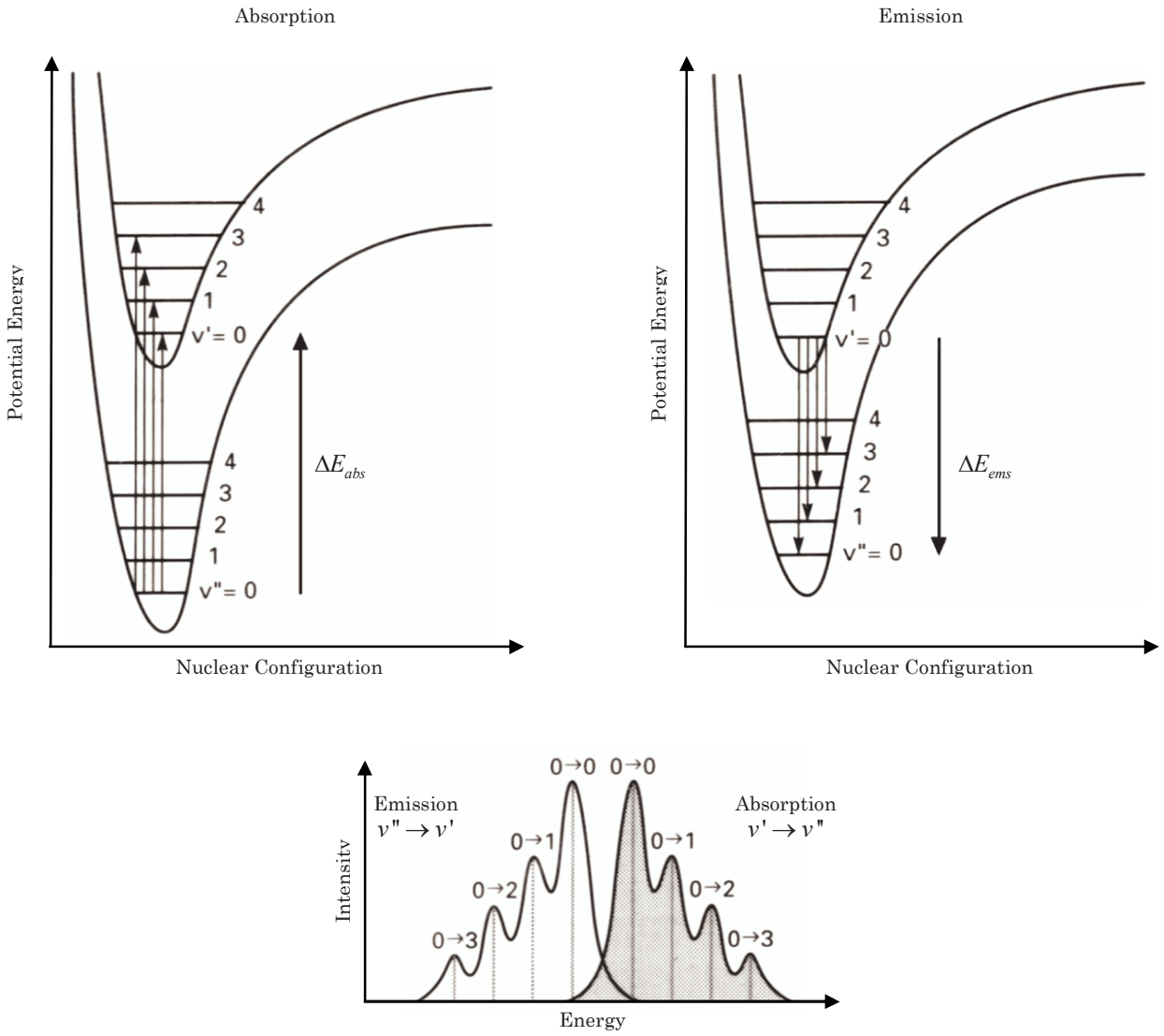



**Figure 4-1** Examples of the absorption and emission processes for a molecular system in a bath of solvent showing the origin of Franck-Condon progressions. Adapted from [Gilbert and Baggott 1991]

Since the peak shift and Stokes shift coincide in this situation all further references, with respect to comparison between theory and experiment, will consider the peak shift since this is quick and simple to measure.

Such a simple treatment of the vibrational levels within the electronic energy profiles would suggest that the absorption and emission spectra should consist of a series of vertical lines of zero width. In conventional excited state calculations on optimised structures this is exactly what is observed. In reality, however, the system is dynamic, both the chromophore structure and the structure of the surrounding medium is changing over time. This results in a broadening of the

absorption and emission profiles giving continuous spectra. The spectral widths are thus also an important parameter for comparison with experiment since these are a good indicator that the degree of coupling between electronic and vibrational levels is being accurately sampled by the methodology employed.

# 4.2 Theory

This section contains an overview of a method for the theoretical prediction of steady state UV/VIS absorption and emission spectroscopy and Stokes shift from energy gap fluctuations and a brief description of aspects of linear optical response theory.

## 4.2.1 Linear Response

The method used for generating steady-state absorption and emission spectra from energy gap fluctuations, discussed in section 4.2.3, employs the fluctuation-dissipation theorem of Callen and Welton [Callen and Welton 1951; Callen 1960]. This theorem relates the relaxation of macroscopic disturbances of a system 'close to equilibrium' to the spontaneous microscopic fluctuations of the system at equilibrium. This theorem is remarkable in its universality; the only assumption is that the response of the system is linear. It is described as the "*cornerstone of linear response theory*" and is a proof of Onsager's regression hypothesis [Onsager 1931a; Onsager 1931b]: "*The relaxation of macroscopic non-equilibrium disturbances is governed by the same laws as the regression of spontaneous microscopic fluctuations in an equilibrium system.*" [Chandler 1987]. This hypothesis forms the basis of almost all modern work in time-dependent statistical and thermal physics and earned Onsager the Nobel Prize for Chemistry in 1968.

In this work, the theory developed by Mukamel [Mukamel 1995] is used to relate the non-equilibrium disturbances due to optical excitation to the calculated energy gap between ground and excited states over time.

Again, the assumption of linear response is the biggest approximation used in this work. It has been shown to hold for a wide variety of cases [Fonseca and Ladanyi 1991; Jimenez *et al.* 1994; Brown 1995; Kumar and Maroncelli 1995; Ando 1997; Mercer *et al.* 1997; Mercer *et al.* 1999]. This is the first time, however, that

the QM/MM methodology developed by Mercer *et al.*[Mercer *et al.* 1999] has been applied to a system with a Stokes shift as large as LADH (*ca.* 0.9 eV) for which the linear response assumption would not be expected to hold.

### 4.2.2 Time Correlation Functions

In order to assign quantitative meaning to Onsager's regression hypothesis it is necessary to consider the time correlation functions of the spontaneous fluctuations. If we take the energy gap fluctuations calculated in this work, where $\langle E \rangle$ is the time independent equilibrium average of the energy gap fluctuations and $\delta E(t)$ is the instantaneous deviation in $E(t)$ from the average such that:

$$\delta E(t) = E(t) - \langle E \rangle \tag{4.1}$$

then the time evolution is governed by microscopic laws. Even for a system at equilibrium the energy gap as a function of time $E(t)$ (Figure 3-12) looks chaotic and could easily be mistaken for simple random noise. In reality, however, the energy gap trace contains all of the information required to investigate the response of the system to optical excitation. By considering the correlation $M(t)$ between $\delta E(t)$ and an instantaneous fluctuation at time zero $\delta E(0)$:

$$M(t) = \langle \delta E(0) \delta E(t) \rangle = \langle E(0) E(t) \rangle - \langle E \rangle^2 \tag{4.2}$$

it is possible to obtain non-chaotic information. In a system at equilibrium the correlation between dynamic variables at various times should depend only on the separation of the times and not on the absolute value of the time since no reaction is occurring. Hence equation (4.2) can be rewritten in terms of two independent time variables:

$$M(t) = \langle \delta E(t_1) \delta E(t_2) \rangle \tag{4.3}$$

where $t = t_2 - t_1$. At small times this becomes:

$$M(0) = \langle \delta E(0) \delta E(0) \rangle = \langle (\delta E)^2 \rangle \tag{4.4}$$

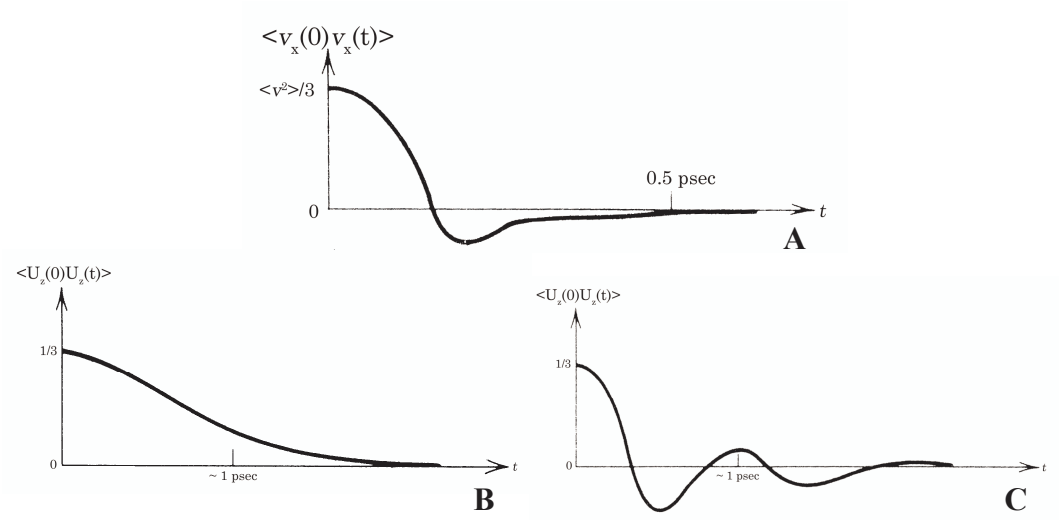



**Figure 4-2** Examples of time correlation functions for **A**) Velocity correlation function for a liquid, **B**) Orientational correlation function for a liquid, **C**) Orientational correlation function for a gas. Adapted from [Chandler 1987].

whereas at large times the correlation between $\delta E(t)$ and $\delta E(0)$ will be zero. Hence as $t \to \infty$ so:

$$M(t) \to \langle \delta E(0) \rangle \langle \delta E(t) \rangle \tag{4.5}$$

and since $\langle \delta E \rangle = 0$, $M(t)$ will tend towards zero. This decay in the correlation function is what Onsager referred to in his hypothesis as the "*regression of spontaneous fluctuations*" [Onsager 1931a; Onsager 1931b].

By employment of the ergodic principle, that given a long enough period of time the ensemble average will be the same as the time average, an alternative method, and the method employed in this work, is available for finding the averages expressed in equations (4.1) to (4.5). If we consider the correlation between $\delta E(t_1)$ and $\delta E(t_2)$, for which there are an infinite number of combinations, we can average over them. Assuming the system is ergodic and sufficient phase space has been sampled, then this average will be the same as averaging over an ensemble of initial conditions for short trajectories of length $t_2 - t_1$. This allows the formation of a correlation function based on a trajectory of energy gaps:

$$M(t) = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} \delta E(\tau + t_1) \delta E(\tau + t_2) d\tau \tag{4.6}$$

where $\tau$ is the length of the trajectory. Figure 4-2 [Chandler 1987] provides some common examples of correlation functions. The limit of $\tau \to \infty$ implies that the system must be sampled for long enough to ensure that all of the available phase space is sampled within a single trajectory. It is this requirement that dictated the run length of the simulations carried out in this work. A short run length is desired due to computational cost, however, the calculations must be run for long enough that sufficient phase space is sampled to reduce the errors which arise from insufficient sampling, to acceptable levels.

Fortunately due to the nature with which discrete correlation functions are calculated they converge with time as each successive point contributes less to the overall profile of the correlation function. This means that the errors are variational with respect to sample length and so each incremental increase in the length of the trajectory will reduce the sampling errors by a diminishing amount each time. For this work the convergence of the peak shift was used as a guide to calculate the amount of phase space to sample and the calculations were stopped when the predicted peak shift was observed to have converged to within pre-determined parameters.

By utilising the notation of time correlation functions it is possible to express Onsager's regression hypothesis mathematically. If a system is excited to a non-equilibrium state at time $t = 0$ and then allowed to relax back to equilibrium then Onsager's principle states that in the linear regime the relaxation obeys[25]:

$$\frac{\Delta \bar{E}(t)}{\Delta \bar{E}(0)} = \frac{M(t)}{M(0)} \tag{4.7}$$

where

$$\Delta \bar{E}(t) = \bar{E}(t) - \langle E \rangle = \bar{\delta} \bar{E}(t) \tag{4.8}$$

and

[25] For a full derivation of these results the reader is referred to Chandler, D. (1987). "Introduction to Modern Statistical Mechanics". New York, Oxford University Press. section 8.5

$$M(t) = \langle \delta E(0) \delta E(t) \rangle \tag{4.9}$$

This implies that the correlation of the property $E(t)$ with the property $E(0)$ at time zero in a system at equilibrium is the same as an average of $E(t)$ given that a specific fluctuation occurred at $t = 0$. This implies that the distribution of the initial phase space is not at equilibrium meaning that for a system close to equilibrium it is not possible to distinguish between the spontaneous fluctuations and externally triggered deviations from equilibrium. This means that knowledge of the equilibrium dynamics is sufficient to predict non-equilibrium properties.

### 4.2.3 Obtaining Optical Spectra from Energy Gaps

Discussion of quantum phenomenon has up to this point been confined to Hilbert (or ordinary) space. In order to evaluate the time evolution of ensemble properties it is necessary to work in Liouville space[26]. The density operator in Liouville space $|\rho\rangle\rangle$, in Dirac notation, is analogous to the Hilbert space operator $|\psi\rangle$, The density operator in Liouville space is a superoperator as indicated by the double bracket notation.

In this work the excitation due to an incident photon is taken to occur between the ground and lowest excited singlet state of the molecule. Due to interactions with the surrounding bath[27] these levels fluctuate leading to the evolution of a time varying Bohr frequency. The time evolution of this system is described by the Liouville-von Neumann equation [Mukamel 1995],

$$\frac{d\rho(t)}{dt} = -\frac{i}{\hbar}\left[\left(H(t), \rho(t)\right)\right] \tag{4.10}$$

where $\rho(t)$ is the density operator for a two-level system with levels g and e, corresponding to the ground and excited states respectively, and $H(t)$ is the Hamiltonian superoperator. For an optical excitation at time zero, where there is no population loss, equation (4.10) yields the solution

---

[26] A detailed discussion of Liouville space, the Liouville-von Neumann equation and its use in deriving the optical response function is beyond the scope of this work, and is not required for an understanding of the theory reviewed here. Reference is made purely for the sake of completeness.

[27] It should be noted that for a chromophore bound to a protein surrounded by water the term "bath" or "solvent" includes not only the water in which the protein is solvated but also the protein matrix itself.

$$\rho_{eg}(t) = \rho_{eg}(0) e^{-i\omega_{eg}t} \left\langle e^{-i\int_0^t \delta\omega_{eg}(\tau)d\tau} \right\rangle \tag{4.11}$$

where $\rho_{eg}(t)$ is an off-diagonal element of the density matrix, $\omega_{eg}$ is the mean angular frequency associated with the electronic energy gap, $\delta\omega_{eg}$ is the energy gap fluctuation from the mean value, $\tau$ is the time of the first interaction with the light field and $t$ is the time of the second interaction with the light field and the angled brackets represent averaging over an ensemble of molecules.

In order to find the polarisation of this ensemble it is necessary to take the trace of the product of the transition dipole operator with the density matrix, yielding the linear optical response function, $R(t)$,

$$R(t) = \left\langle \mu(t)\mu(0) e^{i\int_0^t \delta\omega(\tau)d\tau} \right\rangle \tag{4.12}$$

where $\mu(t)$ and $\mu(0)$ are the transition dipole operators at times $t$ and $0$, respectively.

Predicting optical observables from the energy gap fluctuations obtained in Chapter 3 requires the response function to be related to the auto-correlation of the energy gap fluctuations. This requires the ensemble average in equation (4.12) to be described by the auto-correlation of a single variable, in this case the frequency associated with the mean energy gaps as provided by the QM/MM simulations. Fortunately by assuming that the distribution of the energy gap fluctuations is Gaussian and that sufficient phase space has been sampled, such that the mean does not vary with time, the cumulant expansion [Kubo 1962] can be applied to equation (4.12) which, by assuming that the transition dipole moment is constant in time (Condon approximation[28]) and setting it to unity, allows the response function to be expressed as:

$$R(t) = e^{-g(t)} \tag{4.13}$$

---

[28] Condon approximation: the electronic transition occurs on a very short time-scale with respect to nuclear motion and so it is fair to assume that the transition probability can be calculated at a fixed nuclear position. In other words an electronic transition is always vertical on the potential energy surface.

where $g(t)$, the line broadening function is given by

$$g(t) = \Delta^2 \int_0^t d\tau_1 \int_0^{\tau_1} M(\tau_2) d\tau_2 \tag{4.14}$$

and $\Delta$ is the root mean square deviation of the energy gap fluctuations. Hence we have an expression for the response function in terms of the energy gap correlation $M(t)$ *cf.* (4.6):

$$M(t) = \frac{1}{\Delta^2} \int_{-\infty}^{\infty} \delta\omega_{eg}(t+\tau)\,\delta\omega_{eg}(\tau) d\tau \tag{4.15}$$

where $\delta\omega_{eg}$ is the energy gap deviation delivered by the QM/MM simulations.

The steady-state absorption and emission spectra are Fourier-related to the linear response function in time (eqs (4.12) and (4.13)) and are given by:

$$\begin{aligned} \sigma_{abs}(\omega) &\propto \mathrm{Re}\left[\int_0^t dt R(t) e^{i(\omega-\omega_{eg})t}\right] \\ \sigma_{ems}(\omega) &\propto \mathrm{Re}\left[\int_0^t dt R^*(t) e^{i(\omega-\omega_{eg})t}\right] \end{aligned} \tag{4.16}$$

where Re refers to the real part of the term in brackets.

To find the emission spectrum it is assumed that the excited state potential energy surface has the same profile as that of the ground state but with the minimum shifted along the reaction coordinate. This is probably the biggest approximation made in this work but the agreement between theory and experiment, demonstrated later, shows that this is valid even for a system with a reorganisation energy as large as that of LADH. This assumption is made for convenience and to save computational cost. The fact that the profiles for the experimental absorption and emission spectra are not quite identical demonstrates that this assumption is approximate. In principle an excited state trajectory could be simulated in order to evaluate the gap between excited state and ground state as a function of time. There are inherent problems in this, however, particularly since classical MD force fields were never parameterised for simulating excited states and thus their applicability is questionable. Although attempts have been made at running molecular dynamics simulations of excited states [Debolt and

Kollman 1990] the increased effort and computational cost required is not justified in this work since the results obtained from a simple ground state trajectory already compare well with experimental data.

#### 4.2.3.1 Detailed Balance

Up to this point the correlation function, $M(t)$, contains only a real component since the energy gap data is real. Hence the response function, $R(t)$, is also real and as such $\sigma_{abs}$ and $\sigma_{ems}$ in equation (4.16) are identical; there is no spectral shift between the emission and absorption spectra. To introduce a Stokes shift, $M(t)$ is modified to satisfy detailed balance [Van Kampen 1981], imparting a complex component. By asserting detailed balance the response of the system is made to conform to the fluctuation-dissipation theorem [Chandler 1987].

Detailed balance is included by operating on the spectrum of oscillators, $J(\omega)$, which is given by the Fourier transform of $M(t)$. Since $M(t)$ is real and symmetric so $J(\omega)$ is also real and symmetric. However, to satisfy detailed balance, it is required that positive frequencies should be related to their negative counterparts by a Boltzmann coefficient [Kubo 1966] such that

$$J(-\omega) = e^{\frac{-\hbar\omega}{kT}} J(\omega) \tag{4.17}$$

To satisfy this relationship a modified semi-classical form of the spectral density is required [Mukamel 1985]:

$$J_{SC}(\omega) = \frac{2J(\omega)}{\left[1 + e^{\frac{-\hbar\omega}{kT}}\right]} \equiv \left[1 + \tanh\left(\frac{-\hbar\omega}{kT}\right)\right] J(\omega) \tag{4.18}$$

Equation (4.18) is then back Fourier transformed to yield a modified version of $M(t)$ where the even components of the function are determined by the real components in the Fourier domain, while the odd components are related to the complex components. With the transformation given by eq. (4.18), the even part of the spectrum of oscillators, the real part of $M(t)$, is left unchanged, and only an odd component that delivers a complex part to $M(t)$ is added. This in turn imparts a complex component for the linear response function, resulting in a Stokes shift between the calculated absorption and emission spectra (eq. (4.16)).

Note that in the high-temperature limit[29], the above method yields the same result as the multimode Brownian oscillator (MBO) picture [Mukamel 1995], in which

$$g(t) = i\lambda \int_0^t M(\tau) d\tau + \Delta^2 \int_0^t d\tau_1 \int_0^{t_1} M(\tau_2) d\tau_2 \tag{4.19}$$

and in the high-temperature limit $\lambda = \hbar\Delta^2 / (2kT)$, where $\lambda$ is the reorganisation energy, $\Delta$ is the root-mean-square energy gap fluctuation, $k$ is the Boltzmann constant, and $T$ is the temperature.

It is also possible to include the time-varying transition dipole moment (TDM) (see eq. (4.12)) within the cumulant expansion approach for the calculation of the optical response. When constants in time are ignored, the linear optical response function becomes the product of three response functions derived from the energy gap and TDM auto-correlation functions and the cross correlation between the energy gap and TDM [Khidekel *et al.* 1996], as shown in eq (4.20),

$$R(t) = e^{\left(-g_{dd}(t) + 2i\dot{g}_{\delta d}(t) + \ddot{g}_{\delta\delta}(t)\right)} \tag{4.20}$$

where $g_{dd}$ is the line-broadening function already discussed and is derived from the energy gap auto-correlation function. The other line-broadening functions, $g_{\delta\delta}$ and $g_{\delta d}$, are derived in the same manner but substituting the energy gap auto-correlation function for the TDM auto-correlation function and the cross correlation function between the energy gap and TDM, respectively. Detailed balance is asserted by adjusting the spectrum of oscillators for each individual correlation function using eq. (4.18). It was found that for LADH the inclusion of the time-varying TDM had an insignificant effect on the results. For calculations involving spectral sidebands, however, the effect was found to be more pronounced.

A flow chart illustrating the process of calculating optical spectra from energy gaps is shown in Figure 4-3.

---

[29] In the low temperature limit $k_B T \ll \hbar\omega$ while in the high temperature limit $k_B T \gg \hbar\omega$. Typically the high temperature limit is only reached around 1000 K.

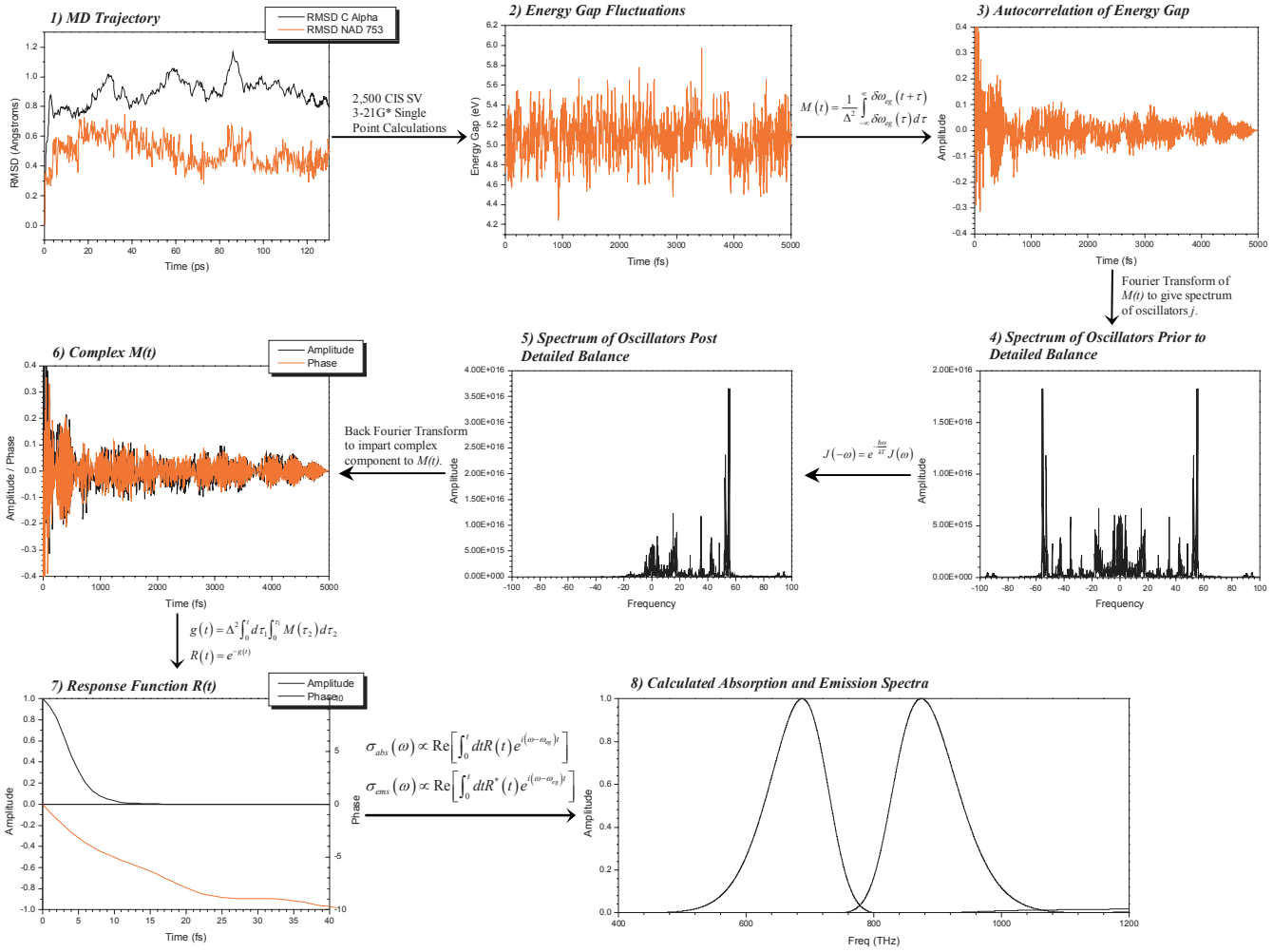



**Figure 4-3** Illustration of the procedure by which steady-state absorption and emission spectra are calculated from ground to excited state energy gap fluctuations.

# 4.3 Experimental Measurement of Spectra

The experimental absorption and emission spectra for NADH bound to LADH, used for comparison in this work, were obtained, with the help of Melanie de Souza, as follows.

LADH (0.52units/mg solid, 0.87 units/mg protein) and NADH (98 % pure) were obtained from Sigma-Aldrich and used without further purification. *N*-Cyclohexyl

formamide (CXF) (Sigma-Aldrich) was purified using a silica gel column developed with ethyl acetate and hexane (60/40). All solutions were made up in phosphate buffer (pH 7.4) and were checked for extraneous fluorescence. To ensure that all of the NADH was in the bound state, an excess of LADH was used in the ternary complex. The absorption and emission spectra of the ternary complex (ratio 3:1:100 of LADH/NADH/CXF in phosphate buffer solution, pH 7.4) are shown in Figure 4-4. To validate that sufficient CXF was present, a number of emission spectra were measured for varying ratios of CXF to protein from 35 % to 93.4 %. It was found that over this range the difference in Stokes shift was 0.034 eV implying that the emission peak position was largely unaffected by the CXF to protein ratio.

The electronic absorption spectrum was recorded on a UV/VIS spectrophotometer (Shimadzu) and the luminescence spectra on a photon-counting spectrometer (SPEX Fluoromax).

The complete absorption spectrum of the nicotinamide moiety of NADH bound to LADH cannot be obtained because of the extremely intense absorption of the NADH adenine group in the range 3.9 to 5.1 eV. To allow accurate comparison between theory and experiment it was necessary to reconstruct the absorption spectrum due to the nicotinamide moiety beyond the 3.95 eV range by taking the profile of the absorption spectrum for unbound NADH in water, which is better resolved from the adenine absorption, and overlaying this on the NADH + LADH ternary complex absorption spectrum. This method was justified because it was found that both spectra are close to Gaussian in shape and have widths within 5 % of each other. The fit used is illustrated in Figure 4-5.

## 4.4 Results and Discussion

### 4.4.1 Experimental Spectra

The measured experimental UV/VIS absorption and emission spectra for the LADH/NADH/CXF ternary complex are shown in Figure 4-4. This data compares well with previously published work [Scott *et al.* 1970; Gafni and Brand 1976; Piersma *et al.* 1998] and shows the expected blue shift in the absorption maximum from 340 nm (3.65 eV) for unbound NADH to 325 nm (3.81 eV) for protein bound

NADH. Figure 4-6 shows both Gaussian and Lorentzian fits to the experimental spectra. It is obvious from these fits that both the absorption and emission spectra are Gaussian in shape and have very similar profiles. This similarity between absorption and emission profiles suggests that NADH has a very simple excited state chemistry; one that does not involving making or breaking bonds or charge separation. The peak shift and Stokes shift coincide in this situation since the spectra are symmetrical. The peak shift is measured to be 0.872 ± 0.05 eV and the emission width at half height is estimated to be 0.59 ± 0.05 eV (Table 4-1). The absorption width cannot be accurately measured experimentally due to the intensity of the adenine absorption but is empirically observed to be similar to that of the emission spectrum.

### 4.4.2 Theoretical Predictions

By taking the energy gaps produced from the QM/MM simulations discussed in Chapter 3 absorption and emission spectra were produced for all 4 MD simulations. This allowed the sensitivity of the linear response approach to the classical MD parameters to be investigated. Chapter 5 considers the sensitivity to the QM/MM method employed to evaluate the energy gaps.

For simulations 1, 2 and 4 the spectral properties from the equilibrium energy gap calculations were evaluated over the first 2,500 points (5 ps) of each of the production runs for both NADH residues independently. For simulation 3, data for NADH residue 753 for only the first 1,800 points (3.6 ps) of the production run was available and so the properties for simulation 3 were calculated over the available data.

Since it was found from the classical MD simulations, discussed in Chapter 2, that over the period of the MD production runs the two NADH residues occupied different regions of phase space it was decided to calculate the spectra independently for each NADH residue. The *raw* predicted spectra for each MD simulation are shown in Figure 4-9. Calculating the energy gaps independently for each NADH residue led to two different absorption and emission spectra being produced, one for each NADH residue. These were subsequently averaged as discussed below (§4.4.2.4) to produce theoretical spectra that could be compared directly with experiment.

Figure 4-7 shows the energy gap auto-correlation functions ($M(t)$) generated from each of the energy gap traces shown in Figure 3-12. The linear optical response functions ($R(t)$) are shown in Figure 4-8 while the convergence of the peak shifts and widths are shown in Figure 4-10 and Figure 4-11 respectively. The final predicted experimental spectra for each of the four MD parameter sets are overlaid with experiment in Figure 4-12 (simulations 1 & 2) and Figure 4-13 (simulations 3 & 4). Table 4-2 provides a summary of the peak shifts and spectral widths for the four MD simulations and experiment while Table 4-3 provides a summary of Gaussian fits to the predicted spectra.

#### 4.4.2.1 Convergence Criteria

An important factor to discuss is how to decide on a convergence criterion for the QM/MM spectral calculations. A corollary of this is the treatment of calculational errors. The degree of convergence was monitored by considering the mean of the energy gap fluctuations (Figure 3-13), the predicted peak shifts (Figure 4-10) for the individual absorption and emission spectra and the predicted spectral widths (Figure 4-11). The peak shift and spectral width data converge as a result of the cumulative nature of the method employed for predicting the optical spectra. Since all of the points from the production run are used to generate the auto-correlation functions so $M(t)$ converges with time. Each successive point has less of an effect on the overall correlation function and energy gap mean square deviation. This cumulative effect means that the *accuracy* of the run improves with time until convergence is achieved. Judging the point at which this convergence occurs is no simple matter and attempts to quantify it in any statistical sense have been ineffective. Since an infinitely long run would be required to achieve true convergence it was necessary to decide on a set of convergence criteria for this work that would balance accuracy against computational cost. The convergence limit chosen was that fluctuations in parameter values of no more than 10 % for a doubling of the run length would be allowed. As illustrated by the peak shift and absorption width plots, convergence within this limit was realised for all four simulations. A longer run would no doubt improve the convergence further and may reduce the observed in-equivalence of the two NADH sites, if indeed they are really equivalent, but with a comparatively small gain in accuracy against considerable computational cost.

The final parameters quoted in Table 4-2 and Table 4-3 are not averaged values, but are derived from optical spectra generated from all data points over the run period.

#### 4.4.2.2 NADH Inequivalency and Relaxation Timescales

The correlation functions for simulations 1 to 3 (Figure 4-7) share similar profiles. All exhibit an ultra fast component, showing initial loss of memory in approximately 6 – 8 fs and a slower decay over the 5 ps of the simulation. The rapidity of the initial drop might lead one to worry that the use of a 2 fs time step between energy gap evaluations might not sample the early time correlation function well. Previous work [Mercer *et al.* 1999] had shown that this was not an issue but to be sure this was checked by performing spline interpolation of the energy gap fluctuations to produce an energy gap step size of 0.5 fs. Sample QM/MM calculations on several hundred of the intervening points showed that the interpolation produced an accurate representation of the intervening data. The optical spectra produced from the interpolated data were indistinguishable from those produced using the 2 fs time step.

In simulation 1 the correlation functions for each NADH residue are very similar, mirroring the similarity in the energy gap fluctuations. The resultant spectra and hence peak shifts predicted from the data in simulation 1 are thus also very similar. The difference in peak shift of 0.23 eV is likely to arise from insufficient sampling since no obvious structural difference is observed in the MD simulations and the second derivative of the experimental emission spectrum shows no evidence of two different emitting species.

In simulation 2 the predicted spectra for the two NADH sites are almost identical even though the long term dynamics of the correlation functions are markedly different. This leads to the prediction that it is the initial decay of the correlation function, coupled with the value of the mean square deviation of the energy gap fluctuations that determines the Stokes shift. At these early times, the auto-correlation function is nominally Gaussian as opposed to being purely exponential. Investigations using analytical functions for the correlation function and varying the mean square deviation support the conclusion that the ultrafast component is necessary to successfully reproduce the experimental spectra, suggesting that it is genuine.

There have been suggestions that the line-shape functions for solvated molecules are dominated by solvent motion. Previous classical simulations for solvated molecules suggest after a change in electronic configuration, a significant degree of relaxation can occur in less than 200 fs [Maroncelli and Fleming 1988; Levy *et al.* 1990; Carter and Hynes 1991; Fonseca and Ladanyi 1991; Bruehl and Hynes 1992; Jimenez *et al.* 1994; Muino and Callis 1994; Rossky and Simon 1994; Stratt and Cho 1994; Brown 1995; Bursulaya *et al.* 1995; Kumar and Maroncelli 1995; Fleming and Cho 1996; Ando 1997]. This fast component of the response has been assigned to motions within the first solvation shell [Fonseca and Ladanyi 1991; Muino and Callis 1994]. Non-linear optical experiments have suggested, however, an even shorter dephasing component of the order of 10 fs duration for solvated DTTCI [deBoeij *et al.* 1996] and 6 fs for IR144 [Passino *et al.* 1997]. Such components are too fast to be associated with dynamics in the first solvation shell and instead are likely to be due to intramolecular dynamics within the solute. This explains the huge differences in peak shifts observed for each of the four MD simulations. In all four simulations only the parameters of the nicotinamide were adjusted, all other parameters were kept fixed and yet the predicted spectra for each simulation are very different (Figure 4-12 & Figure 4-13). The rapidity and intramolecular nature of the fast component of the correlation functions may explain why it has been possible to reproduce the experimental data accurately without taking into account solvent polarisability.

Simulation 4 showed distinct differences in the dynamics of the two NADH sites (§2.3.4) and consequently marked differences in the frequencies contributing to the energy gap fluctuations (Figure 3-15). These differences, largely attributed to the increased phase space available due to the removal of the amide torsions in the nicotinamide chromophore, manifest themselves in markedly different auto-correlation functions. NADH residue 753 was observed to occupy a similar region of phase space to that explored by simulations 1 to 3 (§2.3.4.3) leading to a correlation function and subsequent spectra that are comparable with those obtained for simulations 1 to 3. Residue 375, however, was found to occupy a region of phase space believed to be energetically inaccessible to LADH at room temperature. This artificially increased the dynamics of the NADH residue leading to a correlation function that is markedly different in shape and timescale to that of the other simulations. The initial loss of memory is only partially complete after 8 fs and does not reach zero until almost 200 fs have elapsed. There are also a

number of low frequency components visible within the auto-correlation function. This difference leads to an over estimation of the Stokes shift and spectral widths. The difference between the spectra of the two NADH sites is so different that when they are averaged to produce experimentally comparable spectra the emission profile is significantly different from the experimental data.

#### 4.4.2.3 Homogeneous and Inhomogeneous Broadening

An important concept to introduce here is that of homogeneity. A molecular distribution in which the individual local environments are indistinguishable is said to be homogeneous. Inhomogeneity refers to the different local environments or initial states in a molecular distribution. With reference to absorption, optical profiles can be said to be either homogeneously or inhomogeneously broadened. For example the absorption spectral width could be the result of a number of fast processes that randomise the energy (homogeneous broadening), from a number of slowly inter-converting sites of different energy (inhomogeneous broadening) or from a combination of the two. The optical response function, $R(t)$, is a complex entity and as such has a phase component and an amplitude component. The amplitude of the optical response function gives the dephasing time, or the time taken for the electronic transition to occur [Mercer *et al.* 1999]. If the auto-correlation function, $M(t)$, decays more rapidly than the amplitude of $R(t)$, the optical profiles are homogeneous, implying that there is no differentiation between local molecular environments. If, however, the decay of $M(t)$ is much slower than the amplitude of the optical response function, then the optical profiles are inhomogeneously broadened. This observation can be explained by considering the nature of the auto-correlation and response functions. The correlation function describes the *memory loss* of the system while the response function indicates the time taken to absorb or emit a photon. A comparison of these timescales therefore acts as a measure of the degree of homogeneity. It is thus instructive to compare the response and correlation functions (Figure 4-7 & Figure 4-8).

In all four simulations the decay of the amplitude in the response functions is largely identical, decaying on a timescale comparable to that of the correlation functions. This suggests a combination of homogeneous and inhomogeneous broadening. The *slow* inhomogeneous component could be due to an interconversion between two NADH structures, possibly involving the amide arm

of the nicotinamide group. Time resolved spectroscopy could be used to investigate this further.

#### 4.4.2.4 Spectral Averaging and Comparison with Experiment

When there are two conformational forms of a given chemical species the optical spectra must be reconstructed by correct ensemble averaging. In order to predict the spectra that would be observed in experiment the *raw* theoretical spectra from the two NADH residues have been scaled by the difference in their transition dipole moments, since the optical response of the system is proportional to the square of the TDM, and then averaged to produce absorption and emission spectra that can be compared directly with experiment (Figure 4-12 & Figure 4-13). The absorption peak of the averaged spectra has then been shifted to coincide with the experimental absorption peak. By these means, each method's ability to predict the extent of relaxation becomes apparent by comparing the experimental and calculated Stokes shifts. This emphasises the aim of this method, which is to calculate relaxation energies rather than equilibrium energies.

The major results of this work are summarised by the optical absorption and emission spectra for each of the four simulations shown, overlaid with experiment, in Figure 4-12 & Figure 4-13. Figure 4-12**A** shows the results obtained for the first MD simulation in which the published NADH parameters of Pavelites *et al.* [Pavelites *et al.* 1997] were employed. These show reasonably good agreement with the experimental spectra considering that the assumption of linear response is being used. The peak shift is reported as 0.984 eV, which compared to the experimental value of 0.872 eV agrees to 13 %. The widths are reproduced to within approximately 1.2 % of experiment.

Figure 4-12**B** shows the results obtained from the second MD simulation in which only the NADH amide torsions were based on those of Pavelites *et al.* The remaining NADH parameters were defined by a simple analogy-based approach. It can be seen from Figure 4-12**B** that the calculations do not show such favourable agreement with experiment, yielding a peak shift prediction of only 0.481 eV, an error of 55 %. The absorption and emission spectral widths are also considerably worse, showing a 33 % deviation from experiment. The agreement is similarly poor for simulation 3 (Figure 4-13**C**) in which the whole of the NADH residue was parameterised by analogy. This shows that the method is very sensitive to the

force field parameters used. The reduced peak shift in the case of simulations 2 and 3 suggests that the dynamics of the system have not been adequately explored by the molecular dynamics trajectory.

In Figure 4-13**D** we see that when a force field with zero torsions for the NADH amide moiety is used the results obtained are very poor with two distinct peaks due to the two NADH residues having distinctly different behaviours in the binding sites.

The key result of these calculations is that using a properly parameterised force field, for the molecular dynamics, the Stokes shift/reorganisation energy is recovered to an accuracy better than 13 %. A similar level of accuracy has previously been shown for reorganisation energies in the case of solvated chlorophyll-*a* and bacteriochlorophyll-*a* [Mercer *et al.* 1997; Mercer *et al.* 1999] but this is the first time the methodology has successfully been used for both a large Stokes shift and a protein. It is encouraging to find that the methodology works so well for the large relaxations in alcohol dehydrogenase. It is, at first sight, somewhat surprising that the assumption of linear response holds for a system that shows a Stokes shift of almost 0.9 eV out of an energy gap of approximately 4 eV. The success of the approach used in this work requires the spectrum of coupled oscillators, which provide the electron vibration coupling to be essentially the same in the ground and excited states of NADH. Only one other case is known of in which linear response theory has been tested on a system with a very large reorganisation energy, the work of Chandler *et al.* who successfully showed that the assumption of linear response did not prevent the accurate recovery of free energies of activation for an $Fe^{2+}/Fe^{3+}$ redox couple [Chandler *et al.* 1988]. It is initially somewhat surprising that there is no need to keep recalculating the partial charges as the molecular dynamics calculation progresses; however, by using only one set of ground-state partial charges to drive the MD the linear response restraint is being rigorously applied. The reasonable accuracy that is obtained within this approximation strongly suggests that the electron-vibration coupling (spectrum of oscillators) remains essentially unchanged from the initial absorption of a photon until the relaxation of the system is complete. It may be that the residual inaccuracy of 10-20 % reflects the limitations of the linear response approach. Overall, however, these results provide some hope that this methodology may be successfully applied to determine the free energy of activation in reactive systems (§6.3).

The sensitivity of the reorganisation energy to the parameter set chosen for the NADH chromophore, coupled with the discussion in section 4.4.2.2, suggests that the large Stokes shift observed in LADH comes largely from internal vibrations of the nicotinamide moiety and not from the protein matrix or from solute-solvent interactions. This is further supported by the observation that the absorption and emission profiles of NADH bound to LADH and NADH in water are very similar. The only spectral differences are in the form of a blue shift in the spectra when bound to the LADH protein matrix. The Stokes shift/reorganisation energy is 16.3 % smaller in the LADH + NADH case compared with NADH in water. This again suggests that it is the degrees of freedom internal to the NADH that dominate the reorganisation energy.

#### 4.4.2.5 Force Field Validation

The sensitivity of the results to the MD parameters is also worthy of comment. It is not immediately obvious that the overall electron-vibration coupling calculated by the QM/MM approach would be very sensitive to the details of the molecular dynamics force field. This work shows, however, that the method is actually rather sensitive to the MD parameters. A comparison of the four parameter sets broadly demonstrates this point and is not intended to be an exhaustive assessment of the dominant contributions to the electron-vibration coupling. To achieve such an analysis would require a normal-mode analysis of the spectrum of oscillators and is outside of the scope of this work. The parameter set of Pavelites *et al*. has reasonably low values for the NADH-amide dihedral parameters (0.5 – 2.5 kcal/mol). It may appear fair to assume that a free rotating amide might simulate this rather well; however, Figure 4-13**D** demonstrates that this is not the case. Moreover, when the NADH amide torsion parameters were switched off the production run RMSD for the alpha carbon backbone actually decreased. A lower RMSD is usually associated with a better representation of the system. However, it was found that for the LADH + NADH system this was not the case. The counter example of this was to use very stiff torsional terms. When this variation was employed, Figure 4-13**C**, (using dihedral parameters of ~ 12 kcal/mol) it was found that the predicted Stokes shift becomes far too small (< 0.4 eV). These results demonstrate that the QM/MM methodology is a sensitive, if

expensive, validator for molecular dynamics trajectories of fluorescent chromophores via direct comparison with experiment.

## 4.5 Conclusions

The results produced in this chapter have shown that the previously developed QM/MM method of Mercer *et al.* [Mercer *et al.* 1997; Mercer *et al.* 1999] has been successfully adapted for treating protein systems and is capable of accurately predicting large reorganisation energies. In the case of LADH, the calculated reorganisation energy is found to be within 13 % of the experimental value and that the Stokes shift would appear to be produced as a result of intramolecular interactions on a timescale of around 6 - 8 fs and not due to interactions with the first solvation shell.

This work has also shown that the method is sufficiently sensitive to the parameters used for the molecular dynamics simulation that it provides a method for validating molecular dynamics parameters for fluorescent chromophores.

This work has also concentrated on automating the processes involved in the coupling of MD simulations to QM/MM energy gap calculations and ultimately to the linear response approach for calculating optical observables. In so doing an automated system for carrying out large numbers of electronic structure calculations across heterogeneous computer clusters has been created and successfully deployed. A system for processing the results from thousands of QM/MM simulations has also been developed offering a huge range of different result analysis and extraction options. This software is discussed in greater detail on the accompanying CD-ROM.

The linear response approach to calculating optical observables has now been implemented in a such a way that it can be routinely employed for the study of solvated molecules and proteins. It is currently being used to study reorganisation energies and anisotropy in zinc-myoglobin (Chapter 6) [Walker *et al.* 2003a].

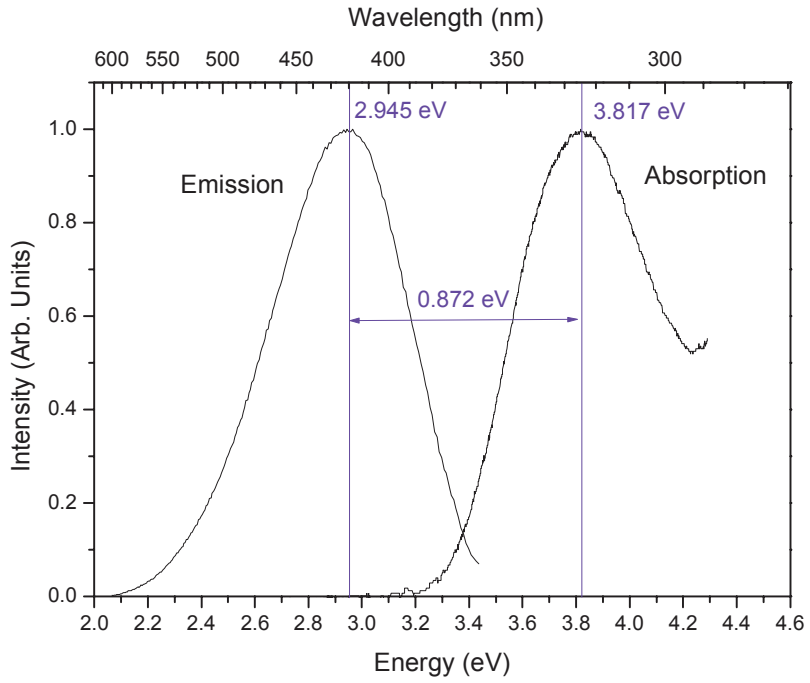



**Figure 4-4** Normalised UV/VIS emission spectrum and reconstructed absorption spectrum for CXF inhibited, NADH containing, horse-liver alcohol dehydrogenase in aqueous solution. The absorption peak is found to be at 3.817 eV (324.8 nm) and the Emission peak at 2.945 eV (421.0 nm) giving a Stokes shift of 0.872 eV.

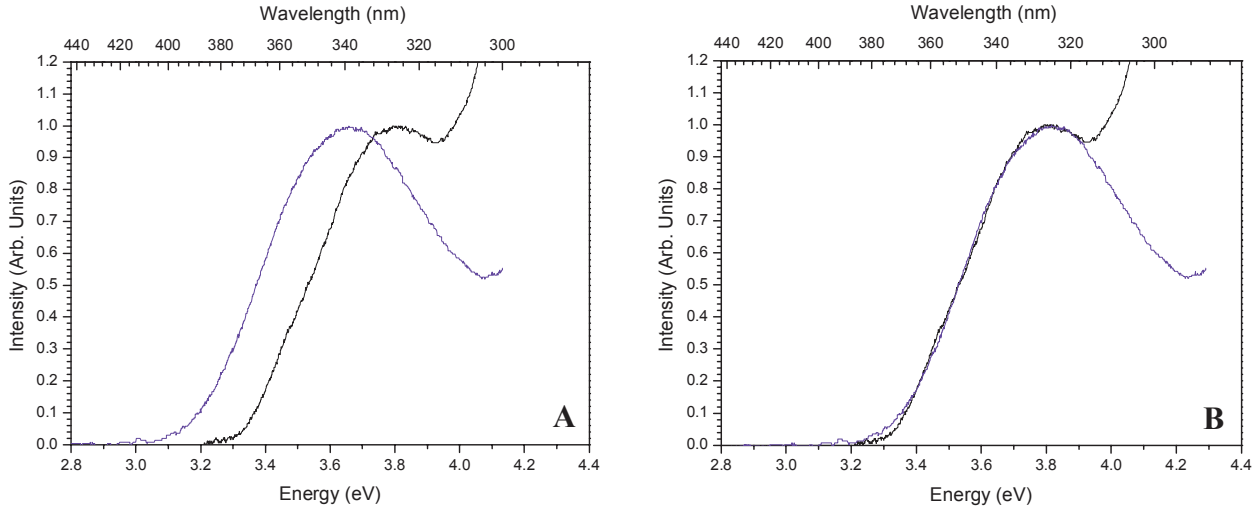



**Figure 4-5 A**) Raw absorption spectra for NADH (blue) and LADH+NADH (black) in water. The similarities in the absorption profile are obvious from the overlay shown in **B**. The shifted NADH spectrum was thus used to reconstruct the LADH+NADH absorption spectrum inside of the adenine absorption envelope.

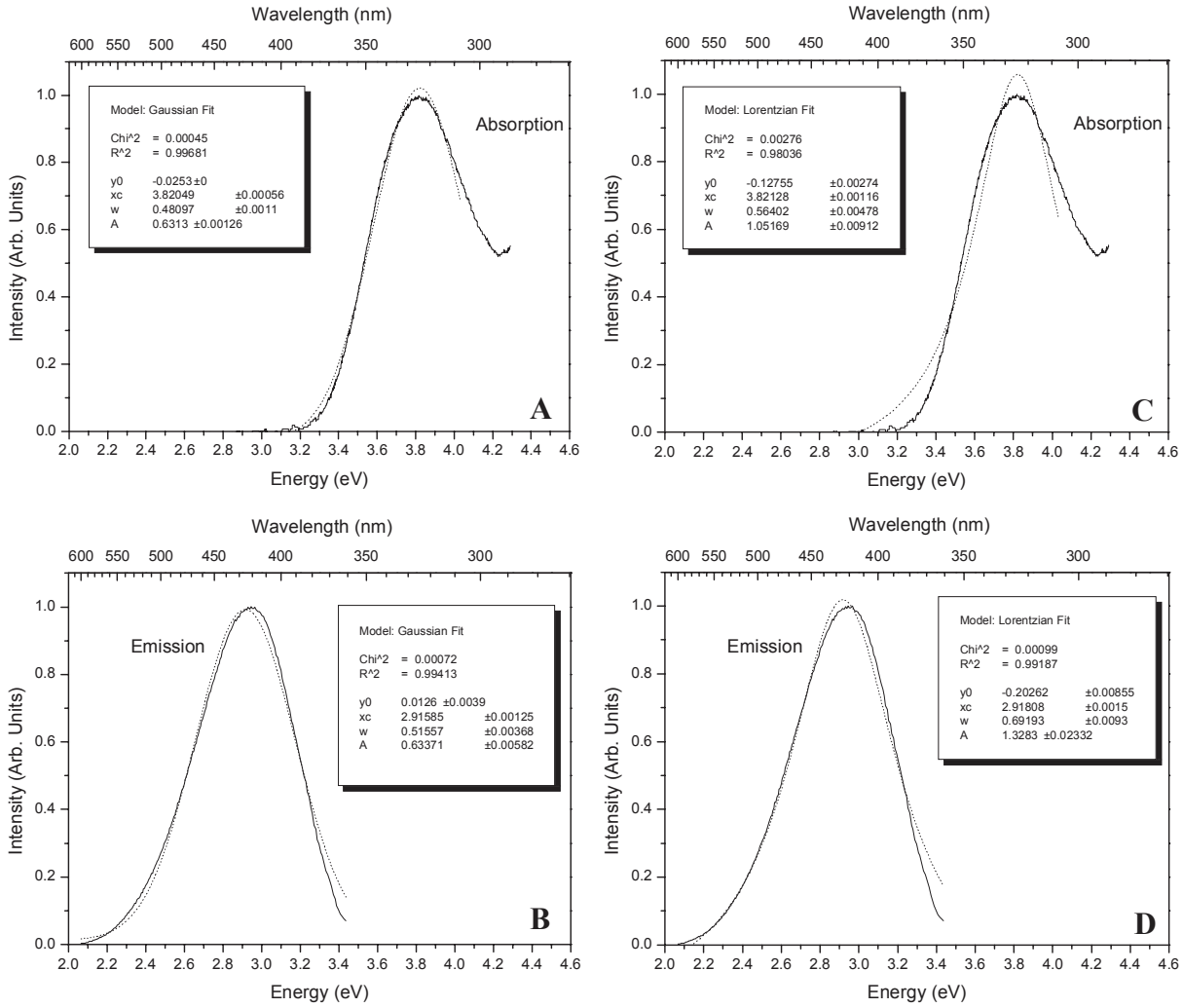



**Figure 4-6** Normalised experimental LADH+NADH absorption and emission spectra (solid lines) with fits (dotted lines) assuming Gaussian (**A** & **B**) and Lorentzian (**C** & **D**) line shapes.

**Table 4-1** Comparison of experimental data (Figure 4-4) and fitted spectral widths of LADH+NADH absorption and emission spectra. All values are in eV.

| System | Peak | | Width | |
|---|---|---|---|---|
| | Gaussian Fit | Lorentzian Fit | Gaussian Fit | Lorentzian Fit |
| **Absorption** | | | | |
| Fit | 3.820 ± 0.001 | 3.821 ± 0.001 | 0.481 ± 0.01 | 0.564 ± 0.02 |
| Experiment | 3.82 ± 0.05 | | N/A | |
| **Emission** | | | | |
| Fit | 2.916 ± 0.001 | 2.918 ± 0.002 | 0.516 ± 0.01 | 0.692 ± 0.02 |
| Experiment | 2.95 ± 0.05 | | 0.59 ± 0.05 | |

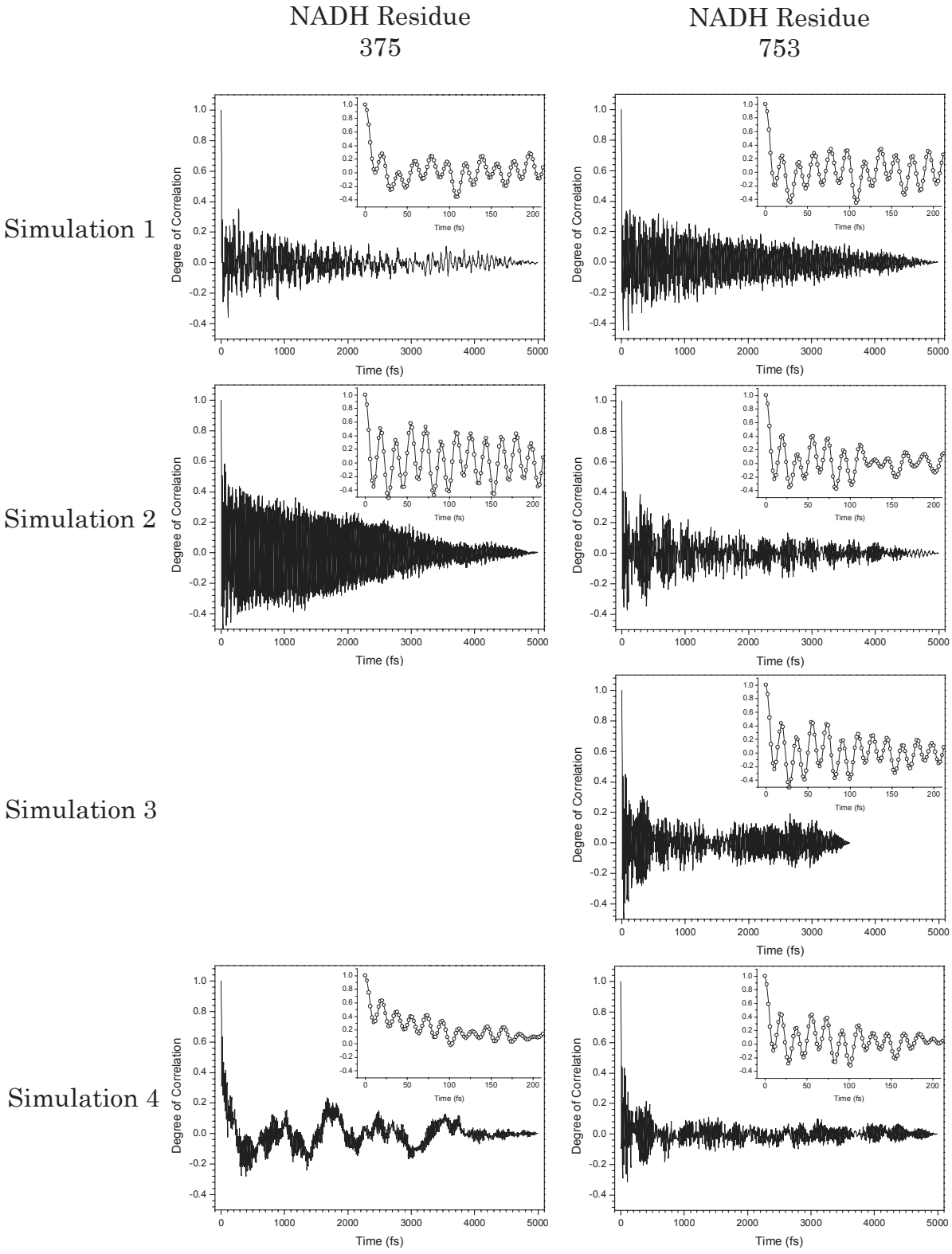



**Figure 4-7** Excited state energy gap correlation functions for Simulations 1 to 4. The auto-correlation functions were calculated separately for each NADH residue using the data shown in Figure 3-12. Insets show magnified views.

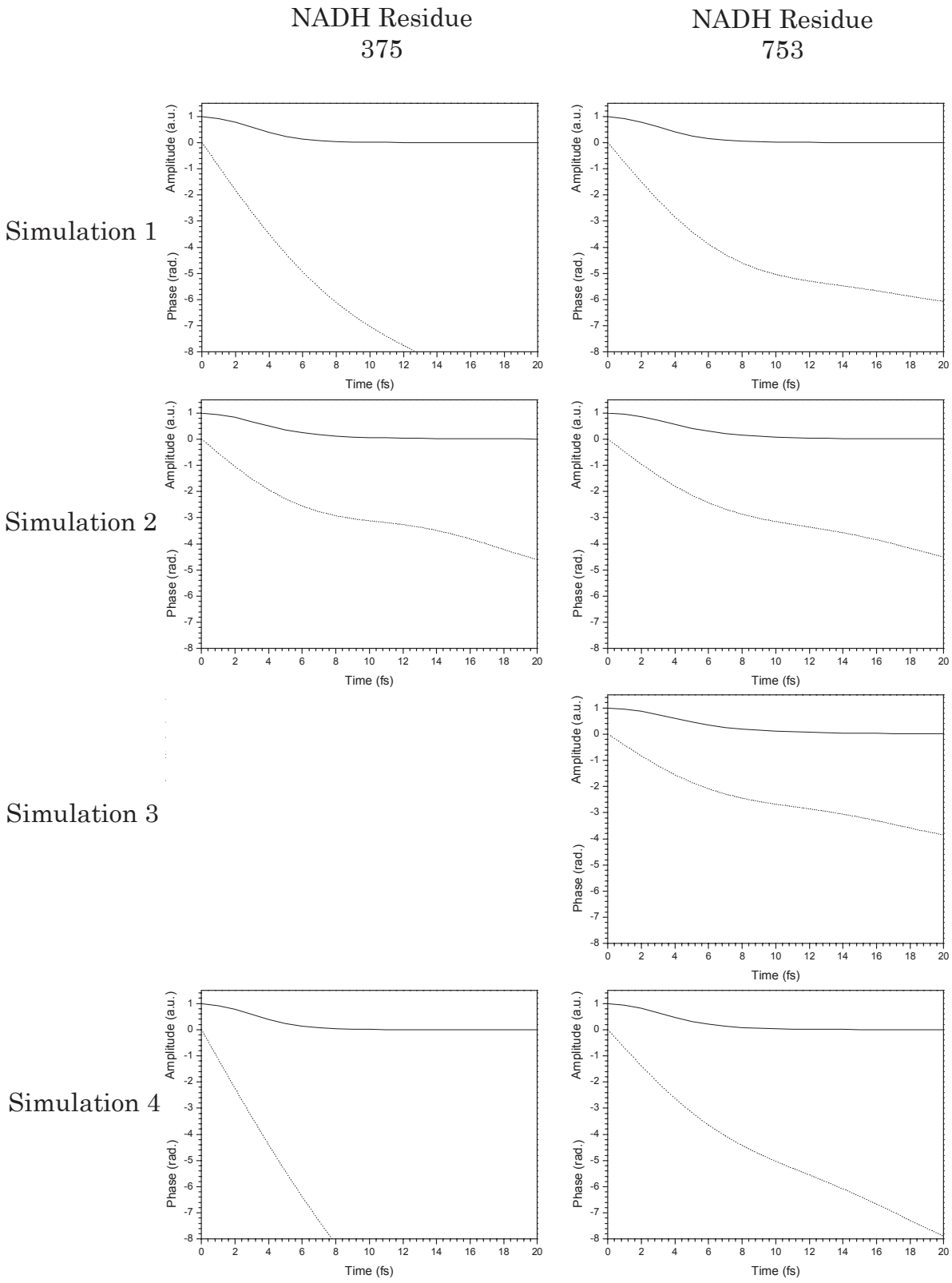



**Figure 4-8** Amplitudes (solid) and phases (dotted) of the linear optical response functions [ $R(t)$ ] for Simulations 1 to 4. The response functions were calculated separately for each NADH residue using the data shown in Figure 3-12.

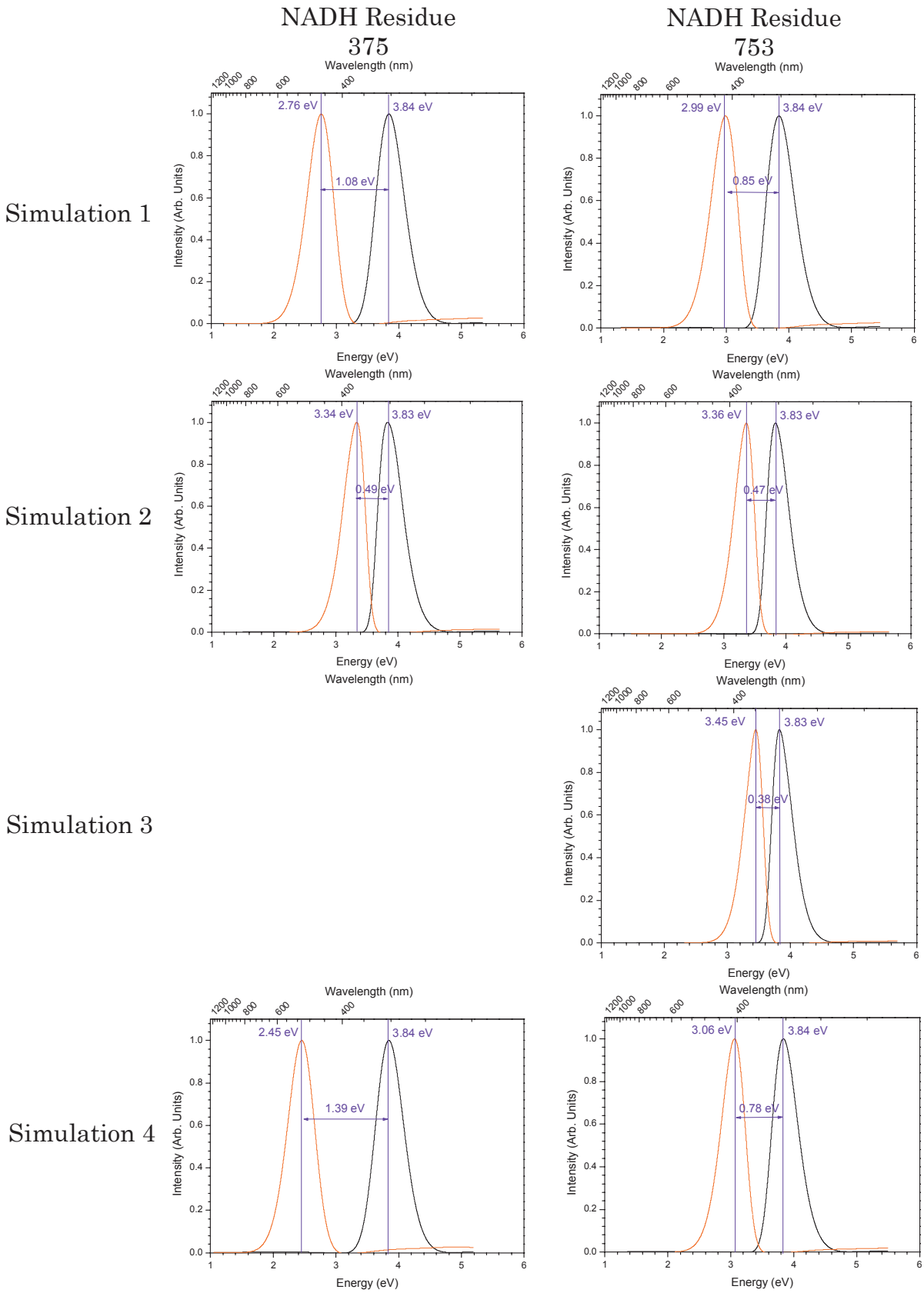



**Figure 4-9** Raw absorption (black line) and emission (red line) spectra, prior to correcting for ensemble averaging, generated from the energy gap fluctuations calculated in Chapter 3. The spectra have been shifted so that the centre of mass of the theoretical and experimental absorption peaks coincide. Peak locations and Stokes shifts are shown in blue.

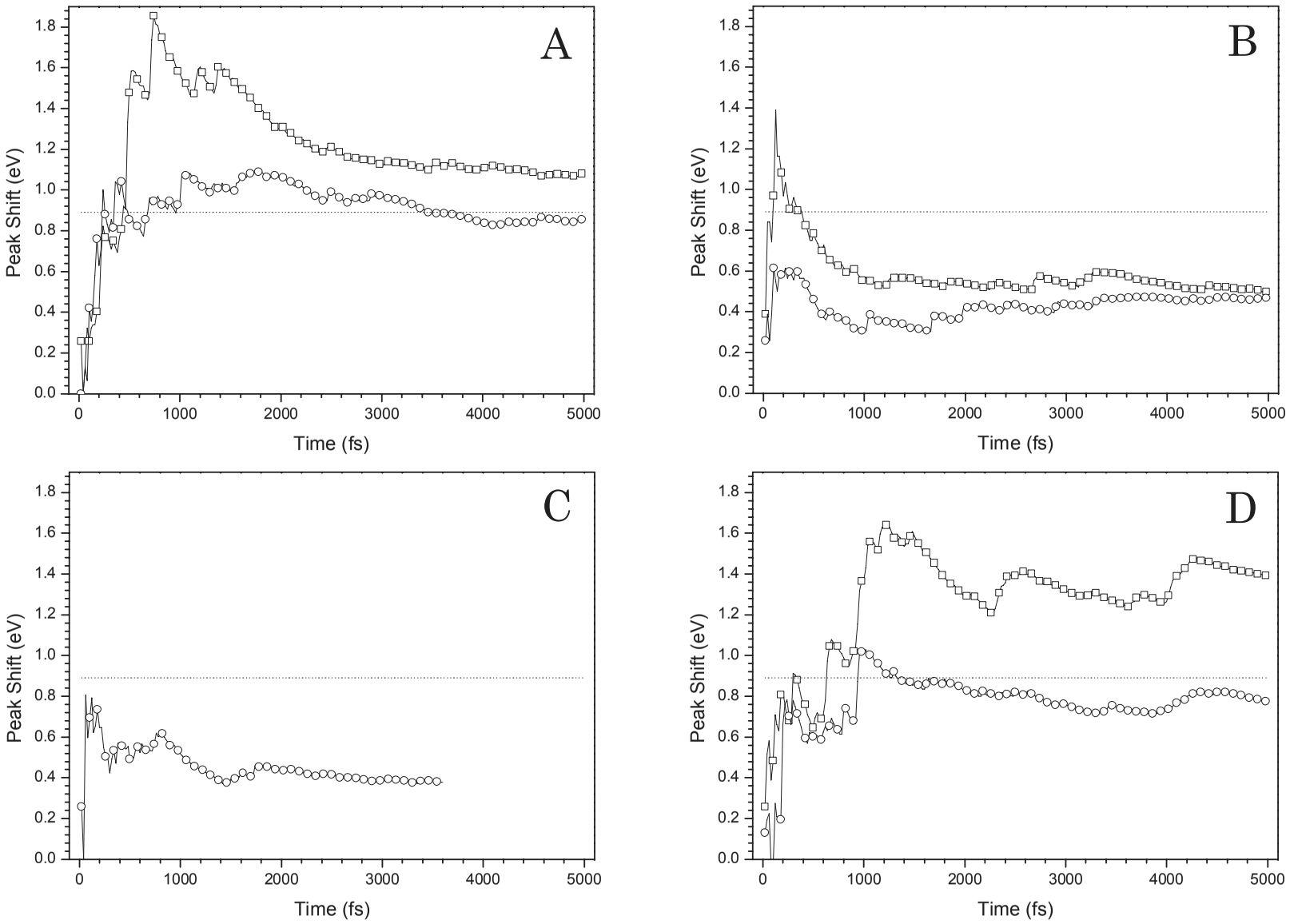



**Figure 4-10** Peak shift convergence for simulations 1 (**A**), 2 (**B**), 3 (**C**) and 4 (**D**). The peak shifts for the two NADH residues, 375 (□) & 753 (○), are shown separately prior to correcting for ensemble averaging. The experimental peak shift is shown by a dotted line.

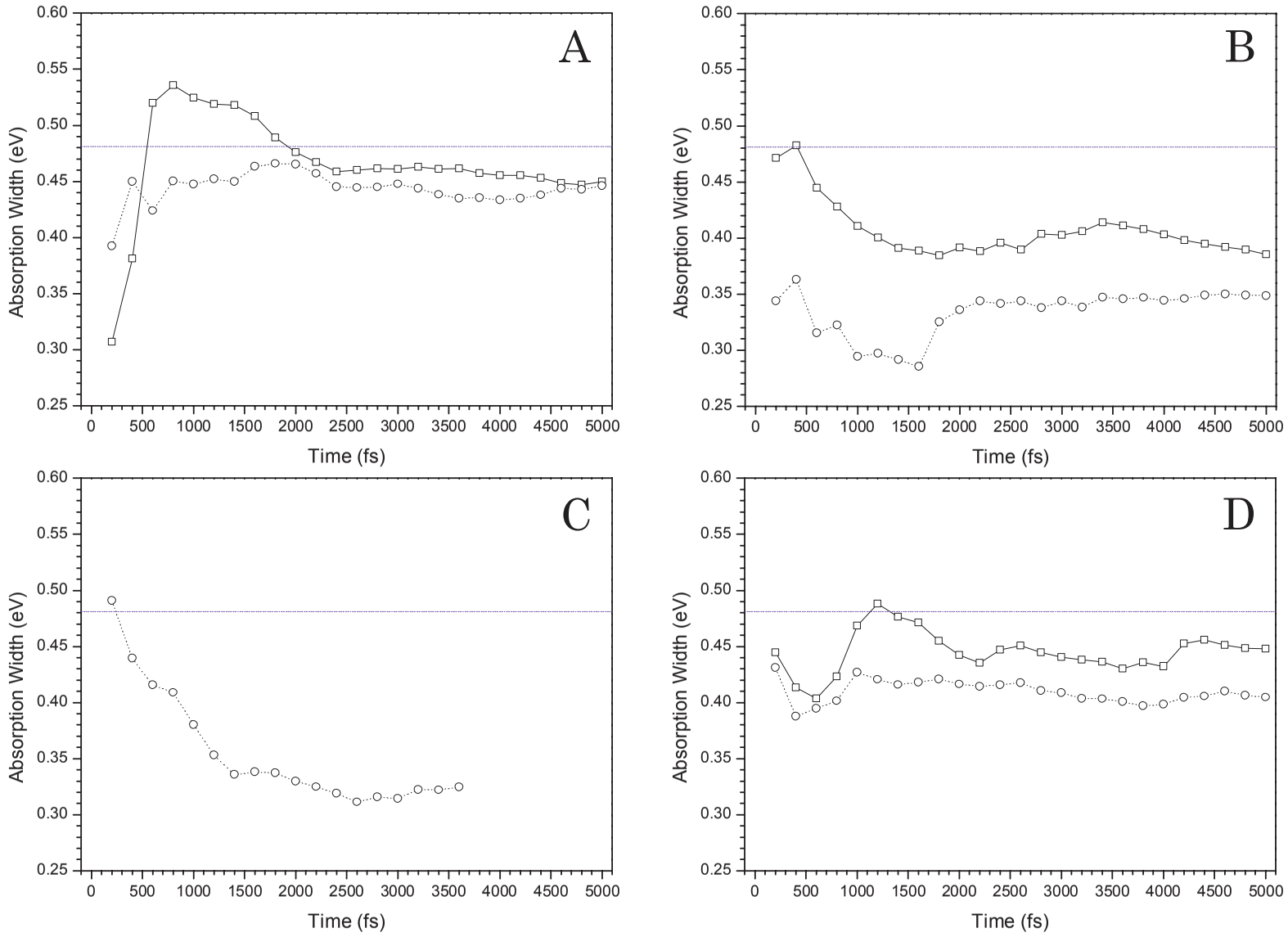



**Figure 4-11** Convergence of absorption spectral widths for simulations 1 (**A**), 2 (**B**), 3 (**C**) & 4 (**D**). Widths are full width, half max widths for Gaussian fits to theoretical spectra prior to correcting for ensemble averaging. The two NADH residues, 375(□) and 753(○), are shown separately. The experimental absorption width is shown as a blue dashed line.

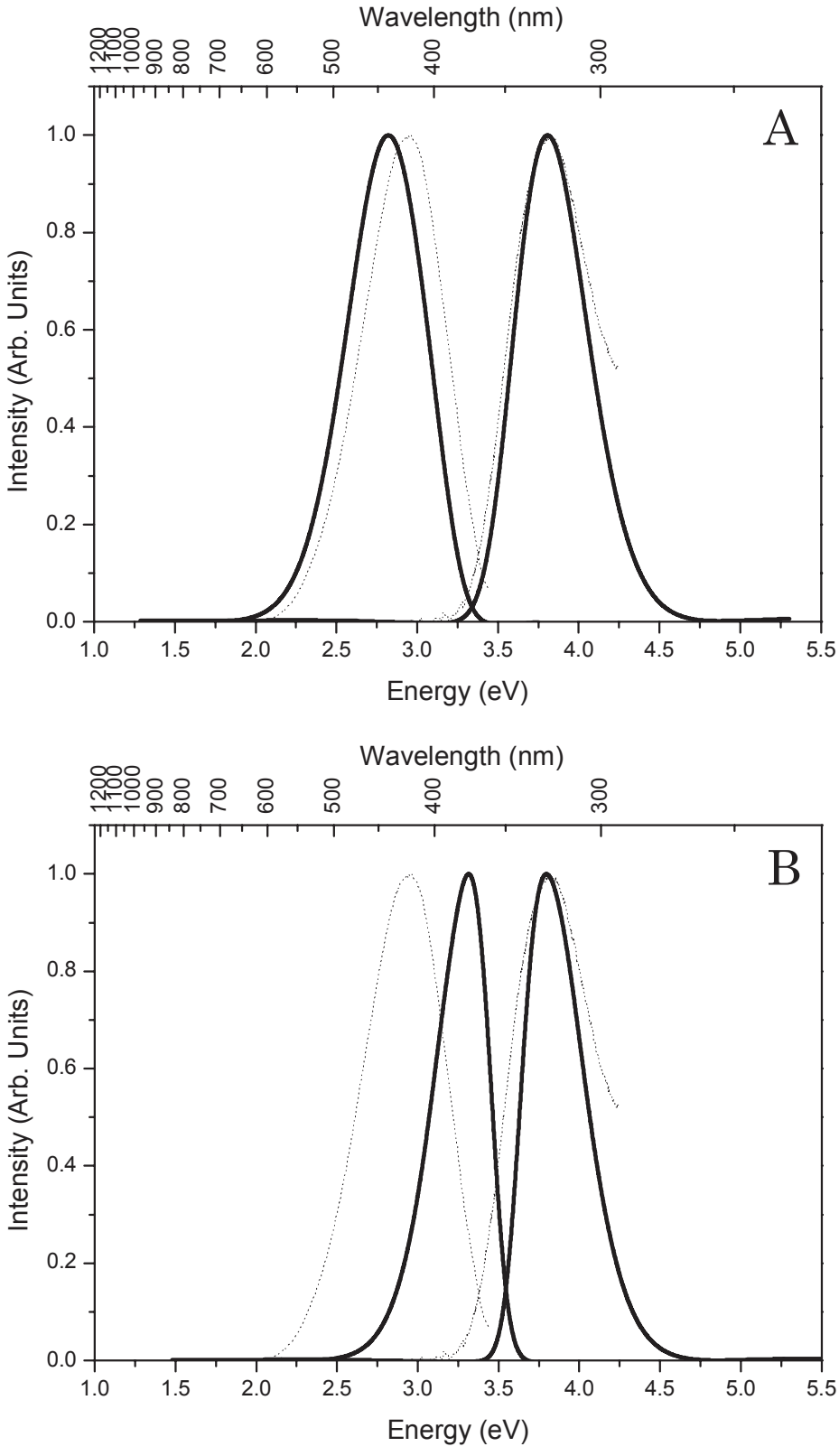



**Figure 4-12** Optical absorption and emission spectra for LADH+NADH in water showing comparison between theory (solid) and experiment (dotted) for simulation 1 (**A**) and simulation 2 (**B**). Theoretical spectra were created via the averaging of the raw spectra shown in Figure 4-9.

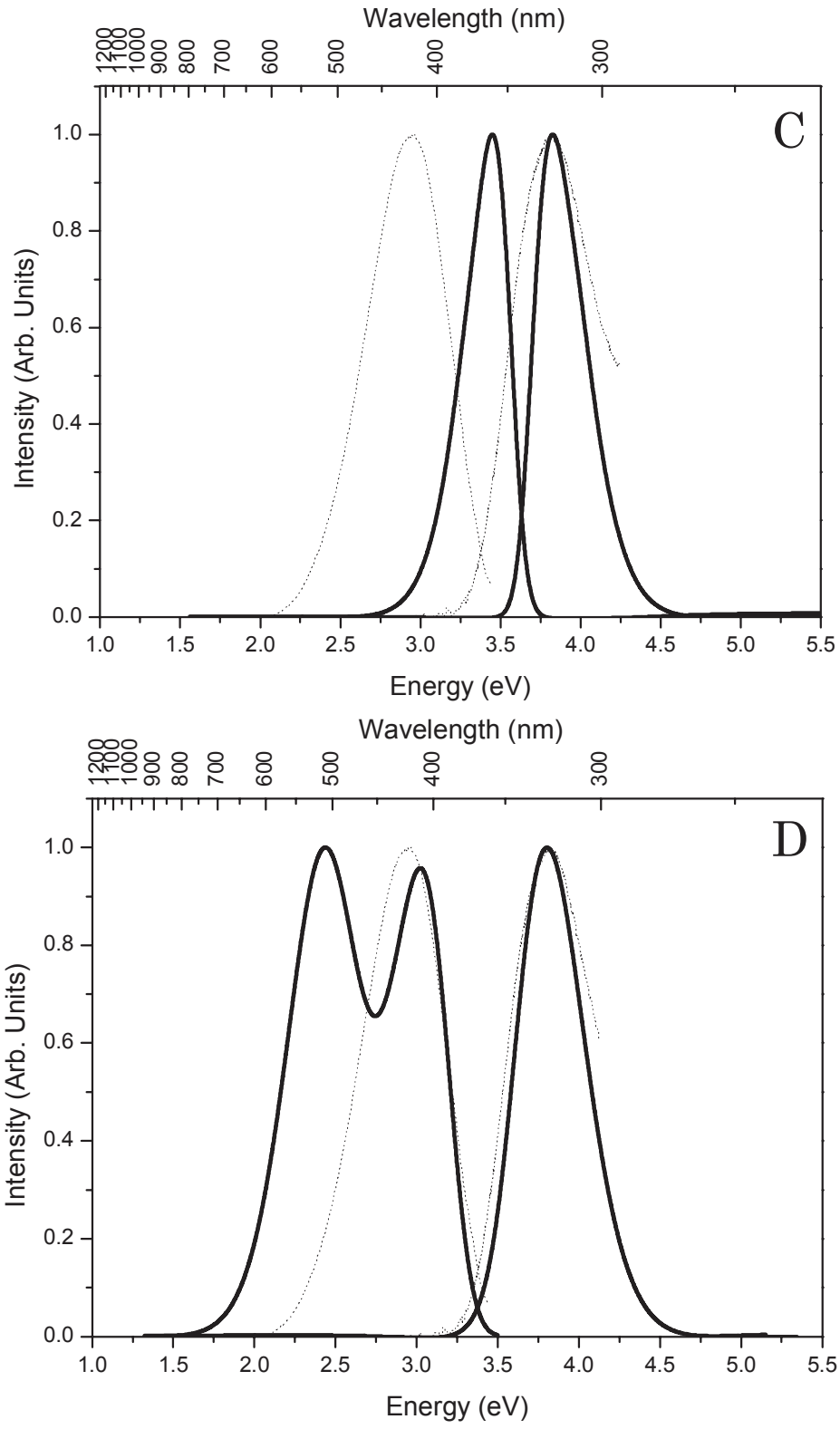



**Figure 4-13** Optical absorption and emission spectra for LADH+NADH in water showing comparison between theory (solid) and experiment (dotted) for simulation 3 (**C) a**nd simulation 4 (**D**). Theoretical spectra were created via the averaging of the raw spectra shown in Figure 4-9. Note: for (**C**) only data for the first 3.6 fs of residue 753 is included.

**Table 4-2** Summary of LADH+NADH theoretical and experimental spectra. Unfitted peak shifts and widths are shown for the experimental data and for simulations 1 to 4. All values are in eV.

| | Unfitted Data | | |
|---|---|---|---|
| **Simulation** | **Peak Shift[a]** | **Absorption Width[b]** | **Emission Width[b]** |
| 1 | 0.984 | 0.532 | 0.583 |
| 2 | 0.481 | 0.437 | 0.395 |
| 3 | 0.376 | 0.371 | 0.371 |
| 4 | 1.366[a] | 0.499 | 1.048 |
| **Experiment** | **0.872 ± 0.05** | **N/A** | **0.59 ± 0.05** |

[a]Measured from highest intensity absorption peak to highest intensity emission peak.
[b]Measured width at intensity = 0.5

**Table 4-3** Summary of LADH+NADH theoretical and experimental spectra. Gaussian fits have been performed on both the experimental and theoretical data and the resultant Stokes shifts and widths are given. All values are in eV.

| | Gaussian Fits | | | | |
|---|---|---|---|---|---|
| **Simulation** | **Stokes Shift[a]** | **Absorption Width** | **Emission Width** | **Abs Fit $R^2$** | **Ems Fit $R^2$** |
| 1 | 1.03 ± 0.02 | 0.448 ± 0.02 | 0.486 ± 0.05 | 0.99311 | 0.99558 |
| 2 | 0.58 ± 0.04 | 0.367 ± 0.04 | 0.336 ± 0.04 | 0.98025 | 0.97758 |
| 3 | 0.47 ± 0.05 | 0.323 ± 0.05 | 0.295 ± 0.05 | 0.97605 | 0.97727 |
| 4[b] | 1.17 ± 0.20 | 0.426 ± 0.02 | 0.965 ± 0.20 | 0.99458 | 0.85500 |
| **Experiment** | **0.904 ± 0.05** | **0.481 ± 0.01** | **0.516 ± 0.01** | **0.99681** | **0.99413** |

[a]Measured from peak of fitted absorption Gaussian profile to peak of emission Gaussian profile.
[b]The Gaussian fit to the emission spectrum in simulation 4 is very poor due to the presence of two distinct peaks in the emission.

# Chapter 5

# Performance of Semi-Empirical Methods

## 5.1 Introduction

The successful prediction of the UV/VIS absorption and emission spectra for LADH discussed in Chapter 4 showed that the linear response approach to relaxation dynamics could be used to predict the Stokes shift of LADH to within 13 % accuracy. The work discussed up to this point has concentrated on the effect played by the classical MD parameters in accurately describing the equilibrium dynamics of the NADH coenzyme and LADH protein matrix. In all cases the *ab initio* HF-CIS method has been used to post process the MD trajectory in order to obtain the time dependence of the energy gap between ground and excited state. One of the questions often raised with this work, however, is how sensitive the approach is to the method used for the energy gap evaluation. In particular how well semi-empirical methods perform for evaluating the energy gaps of equilibrium structures. The essential question here is, since semi-empirical methods are typically parameterised for predicting and investigating optimised structures, can they be expected to work for investigations that require accurate reproduction of equilibrium properties?

The PM3 semi-empirical method has previously been tested for bacteriochlorophyll-*a* and found not to work [Mercer *et al.* 1999]. DFT based methods have not been previously tried with the QM/MM based linear response method however.

The aim of the work described in this chapter was to investigate how the choice of energy gap evaluation method would affect the predicted spectra for LADH. In order to test the sensitivity of the linear response method to the energy gap evaluation technique the energy gaps for Simulation 1, residue 753 (see Figure 3-12) were re-calculated using the semi-empirical ZINDO [Ridley and Zerner 1973] method and the time dependent Hartree Fock (TD-HF) and density functional (TD-DFT) approaches. By using both an *ab initio* HF evaluated SCF component for the TD approach and a semi-empirical DFT (B3LYP) evaluated SCF component it has been possible to relate the differences in the results obtained to the method used for the SCF evaluation rather than the evaluation of the energy gaps post SCF. The results discussed here are published in [Walker *et al.* 2003b].

Section 5.2 provides a brief description and theoretical overview of each of these methods, while section 5.3 discusses their implementation for calculating the energy gap traces and the problems encountered with the semi-empirical methods. Section 5.4 compares the results obtained to those presented in Chapter 4 and speculates on why the performance of the semi-empirical methods is so poor.

## 5.2 Theory

This section provides a brief overview of density functional theory (DFT), the time dependent scheme for evaluating electronic excitations and Zerner's intermediate neglect of differential overlap (ZINDO) model for electronic spectroscopy. The theory discussed here is intended to provide only a brief description of these techniques and is not intended to be an exhaustive discussion[30]. For a more in-depth discussion of these techniques the reader is referred to the referenced papers.

[30] A thorough understanding of the theories involved in the TD-HF, TD-DFT and ZINDO methods is not required for an understanding of the work covered in this chapter. The aim of the calculations in this chapter are purely to illustrate the problems involved with using parameterised methods for studying phenomena that arise from equilibrium fluctuations and are not meant to form an exhaustive investigation of the various semi-empirical methods available.

## 5.2.1 ZINDO

The ZINDO [Ridley and Zerner 1973] method is Zerner's implementation of the Semi-empirical intermediate neglect of differential overlap (INDO) model, developed for electronic spectroscopy. Originally developed by Pople, Beveridge, and Debosh [Pople *et al.* 1967], the INDO model has been extended to include the transition series [Bacon and Zerner 1979; Anderson *et al.* 1986; Anderson *et al.* 1991] and electronic spectroscopy at the CI level of theory [Ridley and Zerner 1973; Ridley and Zerner 1976; Zerner *et al.* 1980] as well as the lanthanide [Culberson *et al.* 1987; Kotzian *et al.* 1991; Kotzian *et al.* 1992] and actinide [Cory *et al.* 1994] elements and their spectra.

The ZINDO model has been shown to be well-suited for the treatment of large biological systems including heme sites [Loew 2000], photosynthetic reaction centres in bacteria [Scherer and Fischer 1989; Thompson *et al.* 1991], and light harvesting antenna aggregates [Cory *et al.* 1998] and has recently been used to study electron tunnelling in proteins [Zheng and Stuchebrukhov 2003]. However, the method does not appear to have been previously employed for evaluating time ordered energy gaps of systems undergoing equilibrium fluctuations.

### 5.2.1.1 Background Theory

Semi-empirical methods, such as ZINDO, differ from *ab initio* methods in that they employ parameters derived from experimental and theoretical data. In *ab initio* methods all elements of the Fock matrix (§3.2.2.5) are calculated using (5.1) irrespective of whether the basis functions $\phi_\mu, \phi_\upsilon, \phi_\lambda$ and $\phi_\sigma$ are on the same atom, on bonded atoms, or on atoms that are not formerly bonded.

$$F_{\mu\upsilon} = H_{\mu\nu}^{core} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K} P_{\lambda\sigma}\left[\left(\mu\upsilon\middle|\lambda\sigma\right) - \frac{1}{2}\left(\mu\lambda\middle|\upsilon\sigma\right)\right] \tag{5.1}$$

As discussed in Chapter 3 the most time consuming aspect of a SCF calculation is the evaluation of the integrals. Hence neglecting or approximating some of these integrals provides a way to reduce the computational effort. One method for doing this is to only consider the valence electrons. The rational behind this being that the electrons involved in bonding and chemical reactivity are those in the valence shell. A feature common to semi-empirical methods is that the

overlap matrix, $S$ (in eq. (3.34)), is set equal to the identity matrix. This means that all diagonal elements of the overlap matrix are set equal to 1 and all off diagonal elements to zero. This means that the overlap between two atomic orbitals on different atoms is zero, with the result that the Roothaan-Hall equations ((3.34)) are simplified to:

$$\boldsymbol{Fc} = \boldsymbol{c}\varepsilon \tag{5.2}$$

It is important to note that setting, $S$, equal to the identity matrix does not mean that all of the overlap integrals are set to zero. In even the simplest semi-empirical models it is necessary to specifically include some of the overlaps. This treatment of the overlaps leads to the naming convention used for common semi-empirical methods.

***Zero-differential Overlap***

Many semi-empirical methods are based upon the so called zero-differential overlap (ZDO) approximation. In this approximation the overlap between pairs of different orbitals is set to zero with the result that all three- and four-centre integrals are set to zero. Applying the ZDO approximation to all orbital pairs results in the Roothaan-Hall equations for a closed-shell molecule being simplified considerably to give, for $\mu \equiv \upsilon$:

$$F_{\mu\mu} = H_{\mu\mu}^{core} + \sum_{\lambda=1}^{K} P_{\lambda\lambda}\left(\mu\mu\middle|\lambda\lambda\right) - \frac{1}{2} P_{\mu\mu}\left(\mu\mu\middle|\mu\mu\right) \tag{5.3}$$

It is not possible to obtain sensible results by simply applying the ZDO approximation to all pairs of orbitals since the total wavefunction should remain the same when a transformed basis set is used and also, major contributors to bond formation are the electron-core interactions between pairs of orbitals and the nuclear cores.

***Complete Neglect of Differential Overlap***

The first method to be developed that implemented the zero-differential overlap approximation in a practical fashion was the complete neglect of differential overlap (CNDO) method of Pople and Segal [Pople and Segal 1965]. In order to overcome the problems of rotational invariance that arise from setting $S$

equal to the identity matrix the two electron integrals, $(\mu\mu|\lambda\lambda)$, were set equal to a parameter $\lambda_{AB}$ which depends only on the nature of the atoms A and B and the distance between them. $\lambda_{AB}$ can be considered to be the average electrostatic repulsion between an electron on atom A and an electron on atom B and as such does not depend on the type of orbital. $\lambda_{AB}$ can be substituted for the two-electron integrals. The value of $\lambda_{AB}$ is a parameter of the calculation, found either from empirical data or from a QM calculation on a simple example. In the CNDO method matrix elements of the core Hamiltonian ($H_{core}$) are also parameterised and in particular, integrals which approximately represent the energy with which an electron is held by an atom are replaced by ionisation potentials.

***Intermediate Neglect of Differential Overlap***

A slightly less dramatic approach to the neglect of differential overlap is offered by the intermediate neglect of differential overlap (INDO) methods [Pople *et al.* 1967]. The CNDO method makes no allowance for the fact that the interaction between two electrons depends on their relative spin states. CNDO thus makes no distinction between singlet and triplet electronic states. The difference in energy between a singlet configuration and a triplet configuration of a molecule is related to the exchange integral, $K$. The effect of neglecting electron spin is most severe when the electrons are on the same atom. The INDO method improves on the CNDO method by not neglecting integrals where the electrons are on the same atom, hence allowing overlap of one-centre integrals. This enables the interaction between electrons on the same atom with parallel spins to have a lower energy than those with opposite spins.

In the ZINDO method the core integrals are obtained by fitting to atomic spectroscopic data while some of the one-centre two-electron integrals are semi-empirical parameters and some are calculated explicitly.

## 5.2.2 Density Functional Theory

In 1964 Hohenberg and Kohn [Hohenberg and Kohn 1964], in work that would ultimately lead to Kohn receiving a 50 % share of the 1998 Nobel prize for chemistry, successfully proved that the ground-state molecular energy wave function is uniquely determined by the ground state electron probability density $\rho(x,y,z)$. Kohn and Sham [Kohn and Sham 1965] then went on to use this

theorem to deduce that it must be possible to use $\rho$ to calculate all the ground state molecular properties. They showed that the purely electronic, exact ground state energy of a molecule containing $n$ electrons with density $\rho$ is given by

$$E = -\frac{1}{2}\sum_{i=1}^{n}\left\langle \psi_i(1)\middle|\nabla_i^2\middle|\psi_i(1)\right\rangle - \sum_{\alpha=1}^{n}\int \frac{Z_\alpha \rho(1)\, d\upsilon_1}{r_{1\alpha}} + \frac{1}{2}\iint \frac{\rho(1)\rho(2)}{r_{12}} d\upsilon_1 d\upsilon_2 + E_{xc}[\rho] \tag{5.4}$$

where $\psi_i(1)$ $(i = 1 \to n)$ are the Kohn-Sham orbitals and $E_{xc}[\rho]$ is the exchange-correlation energy which is a functional of the electron density. Equation (5.5) can be solved to find the Kohn-Sham orbitals

$$\hat{F}_{KS}(1)\psi_i(1) = \varepsilon_{i,KS}\psi_i(1) \tag{5.5}$$

where $\hat{F}_{KS}$ is the Kohn-Sham operator defined in terms of the Coulomb operator, $J_i(1)$ (eq. (3.24)), and the exchange-correlation potential, $V_{xc}$,

$$\hat{F}_{KS} = -\frac{1}{2}\nabla_1^2 - \sum_{\alpha=1}^{N}\frac{Z_\alpha}{r_{1\alpha}} + \sum_{j=1}^{n} J_i(1) + V_{xc}(1) \tag{5.6}$$

$V_{xc}$ is obtained from the functional derivative of $E_{xc}$.

$$V_{xc} = \frac{\delta E_{xc}[\rho]}{\delta \rho} \tag{5.7}$$

The Kohn-Sham operator is thus very similar to the HF operator (§3.2.2) except that the exchange operators ($K_\mu$) in equation (3.28) are replaced by the exchange-correlation potential ($V_{xc}$) which accounts for both exchange and electron correlation. The problem with DFT, and the reasons why it is sometimes referred to as a semi-empirical method, is that the correct form of the functional $E_{xc}[\rho]$ is unknown. Many approximations to $E_{xc}[\rho]$ are possible and the results obtained will be dependent, not only on the choice of basis set, but also on the approximate form of the exchange-correlation term. Typically the exchange correlation term is divided into an exchange functional, allowing for electron exchange, and a correlation functional that deals with the effects of electron correlation. A number

of different functionals exist for representing the exchange and the correlation[31]. There exist true DFT functionals such as BLYP, where the B stands for the Becke exchange functional [Becke 1988] and LYP the Lee, Yang, Parr correlation functional [Lee *et al.* 1988].

Since the exchange-correlation is simply a function it is possible, in theory, to use any type of function as long as it accurately describes the system. Thus it is possible to add empirical amounts of Hartree-Fock exchange that can lead to more accurate results. An example exchange functional that employs this approach, and the functional used in this work, is Becke's 3 parameter exchange functional B3 [Becke 1993] which when coupled with the Lee, Yang, Parr correlation functional gives the DFT method referred to as B3LYP. An overview of the functionals forming B3LYP are given below.

#### 5.2.2.1 Correlation and Exchange Functionals

The correlation functional used in this work is a second order gradient expansion formulated by Lee, Yang and Parr (LYP) [Lee *et al.* 1988] which has the form:

$$E_c^{LYP} = -a\int \frac{4}{1+d\rho^{-1/3}}\frac{\rho_\alpha\rho_\beta}{\rho}d\tau$$
$$-ab\int \omega\left\{\rho_\alpha\rho_\beta\left[\begin{array}{l} 2^{11/3}C_F\left(\rho_\alpha^{8/3}+\rho_\beta^{8/3}\right)+\left(\frac{47}{18}-\frac{7}{18}\delta\right)\left|\nabla\rho\right|^2 \\ -\left(\frac{5}{2}-\frac{1}{18}\delta\right)\left(\left|\nabla\rho_\alpha\right|^2+\left|\nabla\rho_\beta\right|^2\right) \\ -\frac{\delta-11}{9}\left(\frac{\rho_\alpha}{\rho}\left|\nabla\rho_\alpha\right|^2+\frac{\rho_\beta}{\rho}\left|\nabla\rho_\beta\right|^2\right)\end{array}\right]\right.$$
$$\left.-\frac{2}{3}\rho^2\left|\nabla\rho\right|^2+\left(\frac{2}{3}\rho^2-\rho_\alpha^2\right)\left|\nabla\rho_\beta\right|^2+\left(\frac{2}{3}\rho^2-\rho_\beta^2\right)\left|\nabla\rho_\alpha\right|^2\right\}d\tau \qquad (5.8)$$

where

[31] The nature by which the approximate functionals are devised will not be covered here. For background information on each of the functionals used the reader is referred to the original published papers.

$$\omega = \frac{\exp\left(-c\rho^{-1/3}\right)}{1+g\rho^{-1/3}}\rho^{-11/3}$$
$$\delta = c\rho^{-1/3} + \frac{g\rho^{-1/3}}{1+g\rho^{-1/3}} \tag{5.9}$$

and

$$C_F = \frac{3}{10}\left(3\pi^2\right)^{2/3} \tag{5.10}$$

and the values of the parameters are

$$a = 0.04918,\ b = 0.132,\ c = 0.2533,\ g = 0.349 \tag{5.11}$$

which were obtained by Colle and Salvetti [Colle and Salvetti 1975] by fitting to a Helium atom. In this functional the $\rho_\alpha$ and $\rho_\beta$ correspond to the spin-up and spin-down densities respectively.

The exchange functional used in this work is actually a hybrid functional. It is termed hybrid since it comprises a mixture of Hartree-Fock exchange and DFT exchange-correlation. The 3 parameter functional that has been employed is due to Becke [Becke 1993] and has the form:

$$AE_x^S + \left(1-A\right)E_X^{HF} + B\Delta E_x^B + E_c^{VWN} + C\Delta E_c^{non-local} \tag{5.12}$$

where the parameters, which were obtained via a least-squares fit to 565 atomisation energies, 42 ionisation potentials, 8 proton affinities and the 10 first-row atomic energies (G1 molecule set [Pople *et al.* 1989; Curtiss *et al.* 1990]), are:

$$A = 0.20,\ B = 0.72,\ C = 0.81 \tag{5.13}$$

### 5.2.3 Time Dependent DFT and HF

The energy gap fluctuation data discussed in Chapter 3 was calculated by post processing the Hartree-Fock SCF calculations using the CIS method (§3.2.4.3). While this is a popular method for calculating excitation energies it is not the only method available that is computationally tractable to the QM/MM approach to relaxation dynamics employed in this work. Alternative post SCF approaches to

calculating UV/VIS excitation energies are the time dependent DFT (TD-DFT) and HF (TD-HF) methods [Casida 1995]. In this work the TD-HF, also known as the Random Phase Approximation (RPA), and TD-DFT implementations within the computational chemistry code, Gaussian 98 [Bauernschmitt and Ahlrichs 1996; Casida *et al.* 1998; Stratmann and Scuseria 1998], have been used in an identical fashion to the CIS method discussed in Chapter 3. Since the time-dependent Hartree-Fock scheme can be derived along the same lines as the time-dependent DFT method only the later will be discussed here.

The TD-DFT method has been reformulated within the last ten years to compute discrete transition energies and oscillator strengths[32] and has been applied to a number of different atoms and molecules [Bauernschmitt and Ahlrichs 1996; Jamorski *et al.* 1996; Petersilka *et al.* 1996; Tozer *et al.* 1999; Tozer and Handy 2000; Burcl *et al.* 2002; Hirata *et al.* 2003]. TD-DFT is now a standard algorithm in many QM computational packages including CADPAC [Amos *et al.* 1998], TURBOMOL [Ahlrichs *et al.* 1997] and GAUSSIAN [Frisch *et al.* 2001]. The method has recently been extended to include second derivates allowing the optimisation of excited state geometries [Van Caillie and Amos 1999; Van Caillie and Amos 2000]. There are doubts, however, about the validity of the results obtained from TD-DFT calculations. Particularly with reference to excited state energy surfaces. An issue discussed by Tozer *et al.*[Tozer *et al.* 1999] who states that "*in order for DFT to become a useful tool for studies of excited states it should be able to …. give a correct description of excited state energy surfaces …. necessary in order to be able to describe not only vertical excitation spectra, but also adiabatic transitions and emission spectra*". It is this application of TD-DFT and other Semiempirical methods that has proved problematical in this work.

### 5.2.3.1 Background Theory

The foundation of density functional theory is tightly coupled to the calculation of ground states and static external potentials. A time dependent treatment is

[32] For discussion and reviews of TD-DFT the reader is referred to the following references: Casida, M. E. (1995). "Recent Advances in Density Functional Methods". Singapore, World Scientific, Casida, M. E. (1996). Recent Developments and Applications of Modern Density Functional Theory. "Theoretical and Computational Chemistry Vol. 4." Seminario, J. M. Amsterdam, Elsevier Science. **4,** Gross, E. K. U., Dobson, J. F. and Petersilka, M. (1996). Density Functional Theory II. "Topics in Current Chemistry". Nalewajski, R. F. Berlin, Springer. **181**.

required in order to calculate frequency dependent response functions, such as dynamic polarisability and excitation energies. A review of the theory behind the extension of DFT to the treatment of time dependent scalar and/or vector potentials is provided by Gross and Kohn [Gross and Kohn 1990]. A brief overview of this approach, adapted from that provided by Stratmann and Scuseria [Stratmann and Scuseria 1998], is discussed here.

By assuming the existence of a potential, $\upsilon_{eff}(\boldsymbol{r},t)$, for an independent particle system which has the form

$$\upsilon_{eff}(\boldsymbol{r},t) = \upsilon(t) + \upsilon_{SCF}(\boldsymbol{r},t) \tag{5.14}$$

whose orbitals, $\psi(\boldsymbol{r},t)$, yield the same charge density, $\rho(\boldsymbol{r},t)$, as for the interacting system, the time-dependent Kohn-Sham equation (5.15) (*cf.* eq. (5.5)) can be derived.

$$\left[-\frac{1}{2}\nabla^2 + \upsilon_{eff}(\boldsymbol{r},t)\right]\psi(\boldsymbol{r},t) = i\frac{\partial}{\partial t}\psi(\boldsymbol{r},t) \tag{5.15}$$

$\upsilon(t)$ in equation (5.14) is an applied field, while $\upsilon_{SCF}$, the self-consistent field, is defined as

$$\upsilon_{SCF}(\boldsymbol{r},t) = \int \frac{\rho(\boldsymbol{r},t)}{|\boldsymbol{r}-\boldsymbol{r}'|} d\boldsymbol{r}' + \upsilon_{\boldsymbol{xc}}(\boldsymbol{r},t) \tag{5.16}$$

where the exchange-correlation potential is the functional derivative of the exchange-correlation action, $A_{xc}$, represented by

$$\upsilon_{xc}[\rho](\boldsymbol{r},t) = \frac{\delta A_{xc}[\rho]}{\delta\rho(\boldsymbol{r},t)} \approx \frac{\delta E_{xc}[\rho_t]}{\delta\rho_t(\boldsymbol{r})} = \upsilon_{xc}[\rho_t](\boldsymbol{r}) \tag{5.17}$$

In (5.17) the unknown functional $A_{xc}$ of the density over space and time is approximated by the time-independent exchange-correlation function, $E_{xc}$, (*c.f.* eq. (5.7)) which is a function of the density, $\rho$, at a fixed time, $t$. This approximation in time is often referred to as the adiabatic approximation and works best for excited states that have clear valence types and are low lying [Bauernschmitt and Ahlrichs 1996; Casida *et al.* 1998].

***Linear Response of the Density Matrix***

The perturbative effect an applied field, $\delta\upsilon(t)$, has on the Kohn-Sham (or HF) Hamiltonian of a system initially in the ground state is, to a first order approximation

$$\delta\upsilon_{eff}(\boldsymbol{r},t) = \delta\upsilon(t) + \delta\upsilon_{SCF}(\boldsymbol{r},t) \tag{5.18}$$

where $\delta\upsilon_{SCF}(\boldsymbol{r},t)$ is the linear response of the self-consistent field arising from the change in charge density, $\delta\rho$. This can be expressed, in a frequency representation, in terms of the linear response of the Kohn-Sham (or HF) density matrix, $\delta P_{st}(\omega)$, where the subscripts *s* & *t* represent general orbitals, occupied or virtual. By splitting $\delta P$ into its component terms, the only nonzero of which are the particle-hole ($\delta P_{ia}$) and hole-particle ($\delta P_{ai}$) components, $\delta\rho$, can be expressed as:

$$\delta\rho(\boldsymbol{r},\omega) = \sum_{ia}\delta P_{ia}(\omega)\psi_i(\boldsymbol{r})\psi_a^*(\boldsymbol{r}) + \sum_{ai}\delta P_{ai}(\omega)\psi_a(\boldsymbol{r})\psi_i^*(\boldsymbol{r}) \tag{5.19}$$

where the subscripts *a* & *b* represent virtual orbitals and *i* & *j* represent occupied orbitals. From time-dependent perturbation theory[33] the linear response to the applied field of the Kohn-Sham density matrix can be written as:

$$\delta P_{st}(\omega) = \frac{\Delta n_{st}}{(\in_s - \in_t) - \omega}\delta\upsilon_{st}^{eff}(\omega) \tag{5.20}$$

where $\Delta n_{st}$ is the difference in occupation numbers[34]. Unfortunately the potential, $\delta\upsilon_{SCF}$, depends on the response of the density matrix, $\delta P$:

$$\begin{aligned}\delta\upsilon_{st}^{SCF}(\omega) &= \sum_{uv}K_{st,uv}(\omega) \\ &= \sum_{bj}K_{st,bj}(\omega)\delta P_{bj}(\omega) + \sum_{jb}K_{st,jb}(\omega)\delta P_{jb}(\omega)\end{aligned} \tag{5.21}$$

---

[33] For an overview of time-dependent perturbation theory see Atkins, P. W. and Friedman, R. S. (1997). "Molecular Quantum Mechanics". New York, Oxford University Press. p184.

[34] For *st* = *ai* this is +1 and for *st* = *ia* it is -1.

where $u$ & $v$ represent general orbitals and $K$ is a coupling matrix defined in the next section. This makes equation (5.20) more complex than it first appears. Substituting (5.18) and (5.21) into (5.20) yields, after some rearrangement:

$$\left[\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^* & \boldsymbol{A}^* \end{pmatrix} - w\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right]\begin{pmatrix} \delta P \\ \delta P^* \end{pmatrix} = \begin{pmatrix} -\delta\upsilon \\ -\delta\upsilon^* \end{pmatrix} \tag{5.22}$$

where the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are:

$$\begin{aligned} A_{ai,bj} &= \delta_{ab}\delta_{ij}\left(\in_a - \in_i\right) + K_{ai,bj} \\ B_{ai,bj} &= K_{ai,jb} \end{aligned} \tag{5.23}$$

***Excitation Energies from TD-DFT and TD-HF***

The coupling matrix, $K$, can be determined from the chain rule and equations (5.16), (5.17) and (5.19), and is given by:

$$\begin{aligned} K_{st\sigma,uv\tau} &= \frac{\partial \upsilon_{st}^{SCF}}{\partial P_{uv}} \\ &= \frac{\partial \upsilon_{st}^{Coul}}{\partial P_{uv}} + \frac{\partial \upsilon_{st}^{xc}}{\partial P_{uv}} \\ &= \left(\psi_{s\sigma}^*(\boldsymbol{r})\psi_{t\sigma}(\boldsymbol{r})\middle|\psi_{v\tau}^*(\boldsymbol{r}')\psi_{u\tau}(\boldsymbol{r}')\right) \\ &\quad + \int d\boldsymbol{r}d\boldsymbol{r}'\psi_{s\sigma}^*(\boldsymbol{r})\psi_{t\sigma}(\boldsymbol{r})\frac{\delta^2 E_{xc}}{\delta\rho_\sigma(\boldsymbol{r})\delta\rho_\tau(\boldsymbol{r}')} \times \psi_{v\tau}^*(\boldsymbol{r}')\psi_{u\tau}(\boldsymbol{r}') \end{aligned} \tag{5.24}$$

where $\sigma$ and $\tau$ are spin indices. In the adiabatic approximation the coupling matrix is thus independent of time and frequency. It is also real when the molecular orbitals are real. It is possible to derive the time-dependent Hartree-Fock equations in a similar way with the only difference being in (5.24) which for TD-HF is:

$$K_{st\sigma,uv\tau} = \left(\psi_{s\sigma}^*(\boldsymbol{r})\psi_{t\sigma}(\boldsymbol{r})\middle|\psi_{v\tau}^*(\boldsymbol{r}')\psi_{u\tau}(\boldsymbol{r}')\right) - \left(\psi_{s\sigma}^*(\boldsymbol{r})\psi_{u\tau}(\boldsymbol{r})\middle|\psi_{v\tau}^*(\boldsymbol{r}')\psi_{t\sigma}(\boldsymbol{r}')\right) \tag{5.25}$$

In TD-HF the $\boldsymbol{A}$ matrix is thus simply the CIS Hamiltonian (§3.2.4).

The interpretation of the $\boldsymbol{B}$ matrix is that it involves both excitation and de-excitation elements. For a singly-excited state containing the orbitals, $\psi_i\psi_j\psi_a\psi_b$, the swapping of indices for $K$ in (5.23) implies that $\psi_i$ and $\psi_b$ are occupied and

$\psi_j$ and $\psi_a$ unoccupied. By considering the matrix elements of $\left\langle \psi(a \leftarrow i) \middle| \hat{H} \middle| \psi(j \leftarrow b) \right\rangle$ the required matrix elements can be found. Mathematically this is the same as computing the matrix elements between doubly excited states and the ground state. Thus the TD-HF approach includes higher order correlation effects over CIS via the inclusion of double excitations.

The TD-DFT approach includes additional correlation via the exchange-correlation potential.[35]

# 5.3 Calculations

In order to investigate the effects of the method chosen to evaluate the energy gaps, the structures from the MD production run of simulation 1 (Chapter 2) were re-calculated in an identical fashion to that discussed in section 3.3.3 but using ZINDO, TD-HF and TD-DFT methods. In order to allow a direct comparison all 2,500 points were evaluated, but in the interests of lowering computational expense only the energy gaps for NADH residue 753 were evaluated. The basis set used for the TD-HF and TD-DFT cases was the same as that used in the CIS approach, 3-21G*. For the TD-DFT calculations the hybrid B3LYP [Lee *et al.* 1988; Becke 1993] functional was used. While other functionals, including pure exchange functionals, could have been tried B3LYP was chosen because it is a commonly used functional and was also reported by [Bauernschmitt and Ahlrichs 1996] to give the best results when used with the time dependent approach.

For the TD-HF case the calculations were restricted to the first 4 excited states while for the TD-DFT case it was found to be necessary to solve for the first 8 states in order ensure that the results included a state with non-zero oscillator strength. This significantly increased the time required for the TD-DFT calculations making the method less efficient than the CIS method.

---

[35] For a discussion of how the calculation of excited states using the TD-DFT and TD-HF methods is actually implemented the reader is referred to sections c and d of Stratmann, R. E. and Scuseria, G. E. (1998). "An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules." *Journal of Chemical Physics* **109**(19): 8218-8224..

## 5.4 Results and Discussion

The biggest problem encountered when applying the semi-empirical TD-DFT and ZINDO methods to the evaluation of the energy gaps over time was that of convergence. The SCF convergence performance of each method is illustrated by Figure 5-1 and Table 5-1. Overall the ZINDO method performed very poorly with over 20 % of the single point energy gap evaluations failing to converge within 256 SCF cycles. The average number of cycles required for convergence was also high at 107.31. Even when convergence of the SCF was achieved the predicted energy gap was poor in all cases with predictions of between 0.6 and 2.0 eV. Table 5-2 summarises the predicted energy gaps for the first structure from the MD production run. Here it can be seen that the TD-HF and CIS methods are very similar. The ZINDO method is very poor with a large number of very low energy states that have appreciable oscillator strengths. This large number of states made it impossible to select the correct state for use in the linear response approach. This poor energy gap prediction, coupled with the large number of states that failed to converge meant that it was not feasible to produce predicted spectra from the ZINDO calculations. The TD-B3LYP method appears to perform quite well giving a single major excitation at 3.7081 eV which is in much closer agreement with the experimental absorption maximum of 3.64 eV than the TD-HF or CIS methods.

However, while the TD-B3LYP method predicts energy gaps that are closer to the experimental value it still suffers from convergence problems. From Figure 5-1 it can be seen that while the CIS and TD-HF methods always take between 14 and 17 SCF cycles to converge, the TD-B3LYP method ranges from a minimum of 14 cycles to a maximum of 211 cycles, with 1 point failing to converge within 256 cycles. The distribution of the SCF cycles required for convergence with TD-B3LYP, Figure 5-1, are clumped in such a way as to suggest that the convergence issues may be related to the structure. It is possible that certain structural conformations, while accurate representations of equilibrium structures, are at the limits of what would be considered an optimised structure. For example a C-H bond length may appear too long (≈ 1.1 Å) and this may be preventing the TD-B3LYP method from converging successfully. The large number of SCF cycles

required for some of the structures results in the performance of the TD-B3LYP method being less than the TD-HF and CIS methods.

The TD-B3LYP method also suffers from another problem. A large number of the predicted states from the TD-B3LYP calculations have zero oscillator strength. In the vast majority of cases this is not a problem as the NICH excitation is the first state reported. However, some calculations gave a large number of states with zero oscillator strength below the NICH excitation. Increasing the number of states that were evaluated to 8 ensured that in the majority of cases the NICH excitation was present in the reported states. However, for 7 cases out of the 2,500 points all of the reported states had zero oscillator strengths. 2 cases also reported the only non-zero oscillator strength states to be at a low 2.65 eV. The large number of zero oscillator strength states also meant that the adenine absorption was not predicted. The origin of these zero oscillator strength states and low energy excitations is unknown but has been reflected on by Tozer and Handy [Tozer and Handy 2000] although they do not offer an explanation.

### 5.4.1 Theoretical Spectra

Figure 5-2 shows the predicted energy gaps for the CIS, TD-HF and TD-B3LYP methods. The missing points in the TD-B3LYP data, indicated by the vertical lines in the energy gap trace, were replaced with data points found by performing spline interpolation of the surrounding points. In this way the small number of missing points could be accounted for in a way that was not expected to have a major effect on the predicted spectra. The energy gap averages, also shown in Figure 5-2, and all further TD-B3LYP theoretical predictions were made using the interpolated energy gap data. It can be seen from the energy gap averages that the three methods gave stable energy gap traces with an overall mean of 5.12 eV for CIS, 4.92 eV for TD-HF and 3.85 eV for TD-B3LYP. As expected the TD-B3LYP method gives an energy gap mean that is much closer to the experimental value 3.64 eV. However, although the TD-B3LYP method predicts an energy gap that is closer to experiment it does not calculate the energy gap fluctuations correctly and as a result the predicted spectra for the TD-B3LYP method are poor.

The correlation functions, shown in Figure 5-3, for CIS and TD-HF are very similar showing the same initial ultrafast decay of 8 fs followed by a slow component that decays gradually over the remaining 5 ps. The TD-B3LYP

correlation function is different however. The initial ultrafast decay is slightly slower taking approximately 10 fs to reach zero and the long term correlation is much less pronounced. These results manifest themselves in different theoretical spectra.

The peak shift convergence for the CIS, TD-HF and TD-B3LYP methods is shown in Figure 5-4 and the results are summarised in Table 5-3. Figure 5-5 shows the predicted spectra calculated from the 5 ps of data for NADH residue 753 of simulation 1 using the CIS, TD-HF and TD-B3LYP methods.

It can be seen from the peak shift convergence that the CIS and TD-HF results are indeed very similar with the predicted peak shifts tracking each other. The TD-HF method does a slightly better job of predicting the peak shift, possibly due to the inclusion of double excitations, giving a final peak shift of 0.897 eV which is within 2.9 % of the experimental value of 0.872 eV. It should be remembered, however, that this data is only for one of the NADH residues and for strict comparison with experiment both residues should be calculated and the results averaged as was done in Chapter 4. It is possible that the TD-HF method may give a slightly lower peak shift, than the CIS method, for the other NADH residue which would result in a better agreement overall. However, without running the calculations this cannot be known for sure but it is likely that the results will be similar to the CIS results. The peak shift for the TD-B3LYP data shows that it underestimates the peak shift by more than 45 % yielding a final predicted peak shift of only 0.473 eV. The convergence is also not as good for the TD-B3LYP data with the peak shift still showing a downward trend at the end of the 5 ps data window.

### 5.4.2 Why are the Semi-Empirical Results so Poor?

The results discussed above and illustrated by the figures at the end of this chapter show that the time dependent Hartree Fock method is as good as the CIS method for evaluating the energy gap fluctuations of equilibrium structures. The similarity between the results suggests that the linear response method is not overly sensitive to the method used to evaluate the energy gap from the converged SCF. However, the linear response method is sensitive to the SCF method employed. The semi-empirical ZINDO method, while shown to work well for many biological systems including heme sites [Loew 2000], photosynthetic reaction

centres in bacteria [Scherer and Fischer 1989; Thompson *et al.* 1991], and light harvesting antenna aggregates [Cory *et al.* 1998], performs very poorly in this work. The reason for this stems from the fact that the studies highlighted above were concerned with static optimised structures while this work has concentrated on dynamic equilibrium structures. The ZINDO method is a highly parameterised method which while very fast to compute relies on empirical parameters in order to yield accurate results. These parameters were designed for the study of optimised structures and so the whole method breaks down when equilibrium snapshots, which occasionally contain highly strained or elongated bonds, are used. This limitation of the ZINDO method renders it essentially useless for studying equilibrium fluctuations.

The TD-B3LYP method also performed badly, although its convergence problems were a lot less than the ZINDO method. DFT methods are not normally referred to as semi-empirical methods and there is still fierce debate over whether they are semi-empirical or *ab initio* methods. The important point, however, is that DFT methods require both a basis set approximation and a functional for the density, the exact form of which is not known. Thus functionals are essentially fitted in order to obtain good agreement for optimised structures. In this way the DFT method is essentially a semi-empirical one as it is effectively parameterised for studying optimised structures. This explains why the TD-B3LYP method does not yield good results for the linear response approach. The TD-B3LYP method is more tolerant than the ZINDO method as the parameterisation is less severe but the results here still suggest that the need to choose an essentially empirical density functional means that the method cannot deal with the extremes of structural variation seen during an equilibrium simulation. As such TD-DFT methods cannot be trusted for studying equilibrium fluctuations and therefore excited state potential energy surfaces generated with such methods must also be viewed with a degree of scepticism.

## 5.5 Conclusions

The results discussed in this chapter have shown that the *ab initio* CIS and TD-HF methods of evaluating the energy gap fluctuations between ground and

excited state yield very similar results that agree well with the experimental data. However, the semi-empirical ZINDO and TD-DFT methods did not perform so well. The ZINDO method was found to be essentially useless for evaluating energy gaps for equilibrium structures with over 20 % of the calculations failing to converge within 256 SCF cycles and those that did converge yielding very ambiguous energy gaps. This poor performance is considered to be due to the large number of empirical parameters used in the ZINDO method. This result reinforces the failure of the PM3 method to represent the equilibrium fluctuations in bacteriochlorophyl-*a* correctly [Mercer *et al.* 1999]. The TD-B3LYP method was found to perform well for some structures but poorly for others. It is believed that the structures for which poor convergence was encountered are a long way from the average structure and as such the form of the functional used was not appropriate. Overall the TD-B3LYP method is found to be unsuitable for studying equilibrium fluctuations with the linear response approach used in this work.

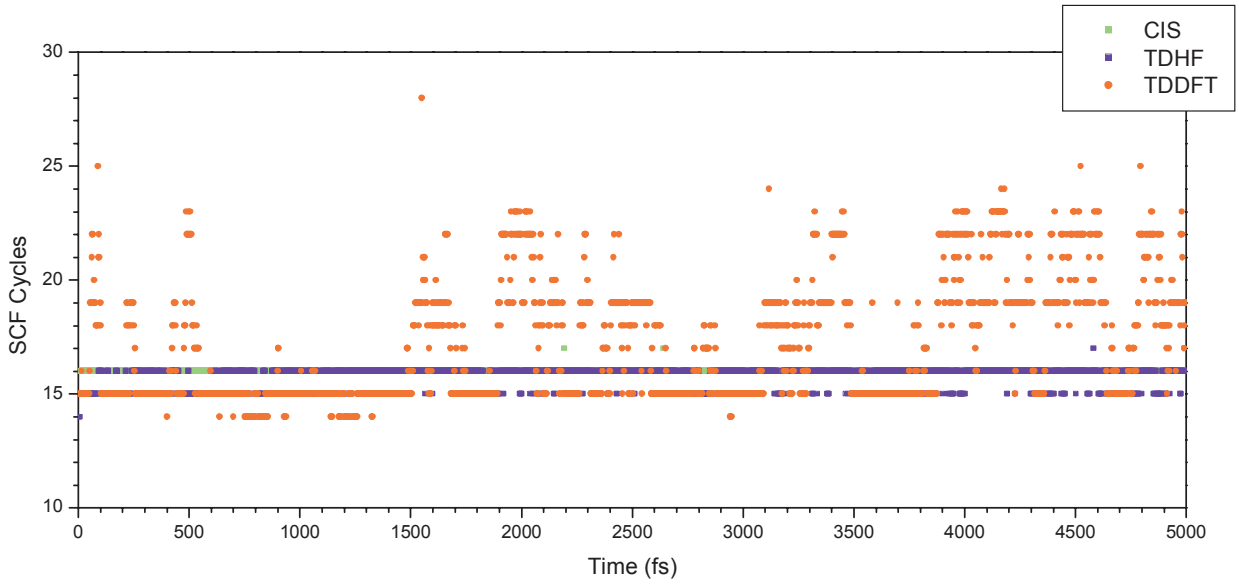



**Figure 5-1** Scatter plot showing the number of SCF convergence cycles required for each point of the 2,500, simulation 1, NADH 753, energy gap evaluations for CIS, TD-HF and TD-B3LYP. Note, ZINDO is not shown due to the very poor convergence and the huge number of points that failed to converge. TD-B3LYP point 1804 fs took 211 cycles and point 4022 fs failed to converge within 256 cycles.

**Table 5-1** Summary of the SCF convergence cycles and failed calculations for simulation 1, NADH 753, using CIS, TD-HF, TD-B3LYP and ZINDO to evaluate the excited state energy gaps.

**SCF Cycles Required for Convergence**

| Method | Minimum | Maximum[a] | Mean[a] | Standard Deviation[a] | No. Failed[b] | No. Zero Osc.[c] |
|---|---|---|---|---|---|---|
| CIS | 16 | 17 | 16.002 | 0.049 | 0 | 0 |
| TD-HF | 14 | 17 | 15.723 | 0.452 | 0 | 0 |
| TD-B3LYP | 14 | 211 | 17.156 | 4.692 | 1 | 7 |
| ZINDO | 38 | 254 | 107.31 | 34.812 | 527 | 0 |

[a.] Excludes calculations which failed to converge.
[b.] Number of calculations for which SCF convergence to $10^{-8}$ was not achieved after 256 cycles.
[c.] Number of converged SCF calculations for which all excited states are reported to have zero oscillator strength.

**Table 5-2** Excited state data for QM/MM single point calculations (NADH 753) on the first structure of the MD production run of simulation 1. All states with oscillator strengths > 0.009 are shown and states with oscillator strengths > 0.1000 are highlighted.

| Method | Excited State | Energy / eV | $\lambda$ / nm | Oscillator Strength, $f$ |
|---|---|---|---|---|
| CIS | **1** | **4.7848** | **259.12** | **0.1756** |
| | 2 | 5.9388 | 108.77 | 0.0123 |
| | **3** | **6.4270** | **192.91** | **0.4459** |
| | 4 | 6.6436 | 186.62 | 0.0135 |
| TD-HF | **1** | **4.5626** | **271.74** | **0.1412** |
| | 2 | 5.7836 | 214.37 | 0.0115 |
| | **3** | **6.1250** | **202.42** | **0.4060** |
| | 4 | 6.4773 | 191.41 | 0.0098 |
| TD-B3LYP | **1** | **3.7081** | **334.36** | **0.1413** |
| | 8 | 4.4583 | 278.10 | 0.0118 |
| ZINDO | 4 | 0.4437 | 2794.50 | 0.0370 |
| | 8 | 0.8072 | 1535.91 | 0.0545 |
| | 10 | 0.9155 | 1354.20 | 0.0095 |
| | 12 | 0.9192 | 1348.81 | 0.0162 |
| | 13 | 0.9859 | 1257.57 | 0.0222 |
| | 14 | 1.0420 | 1189.81 | 0.0105 |
| | 16 | 1.2160 | 1019.61 | 0.0282 |
| | 18 | 1.3339 | 929.47 | 0.0365 |
| | **19** | **1.3853** | **894.99** | **0.1542** |
| | 21 | 1.4015 | 884.64 | 0.0148 |
| | 24 | 1.6134 | 768.46 | 0.0361 |
| | 25 | 1.6345 | 758.55 | 0.0284 |
| | 26 | 1.6858 | 735.46 | 0.0217 |
| | **27** | **1.7509** | **708.11** | **0.1835** |
| | 28 | 1.7943 | 690.97 | 0.0109 |
| | 29 | 1.8302 | 677.42 | 0.0354 |
| | 30 | 1.8439 | 672.39 | 0.0824 |
| | 32 | 1.8697 | 663.11 | 0.0421 |

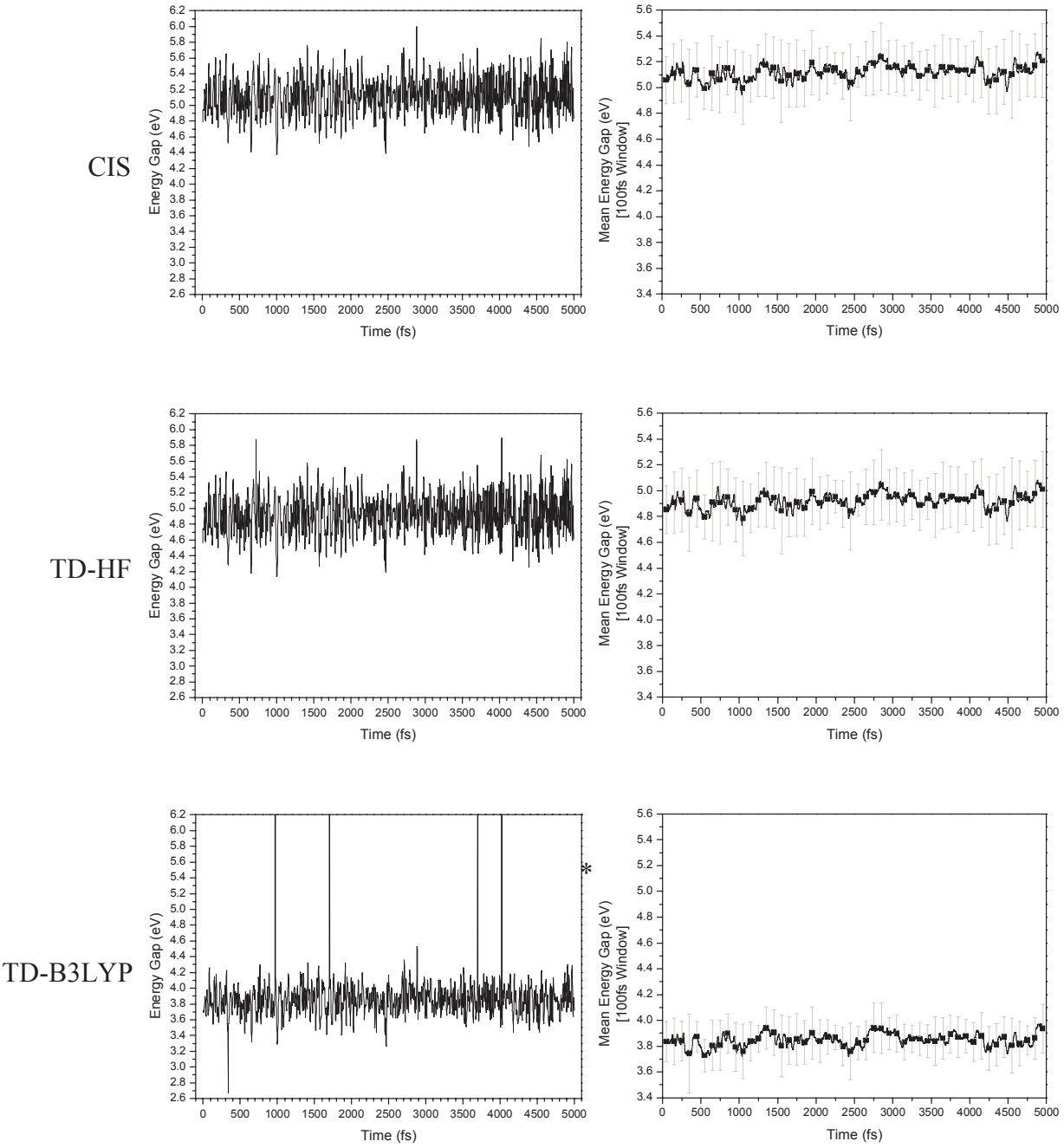



**Figure 5-2** Excited state energy gaps (left hand column) for NADH residue 753 in QM region for the first 5 ps of the MD production run and the mean values of the energy gaps (right hand column black line) and standard deviations (right hand column grey lines), computed using CIS, TD-HF and TD-B3LYP. The energy gap averages and standard deviations were calculated using a +/- 50 fs moving box car average. *The 7 TD-B3LYP points in which all states had zero oscillator strength and the 1 point that failed to converge are shown by the vertical lines in the energy gap trace. These points were interpolated prior to calculating the energy gap averages.

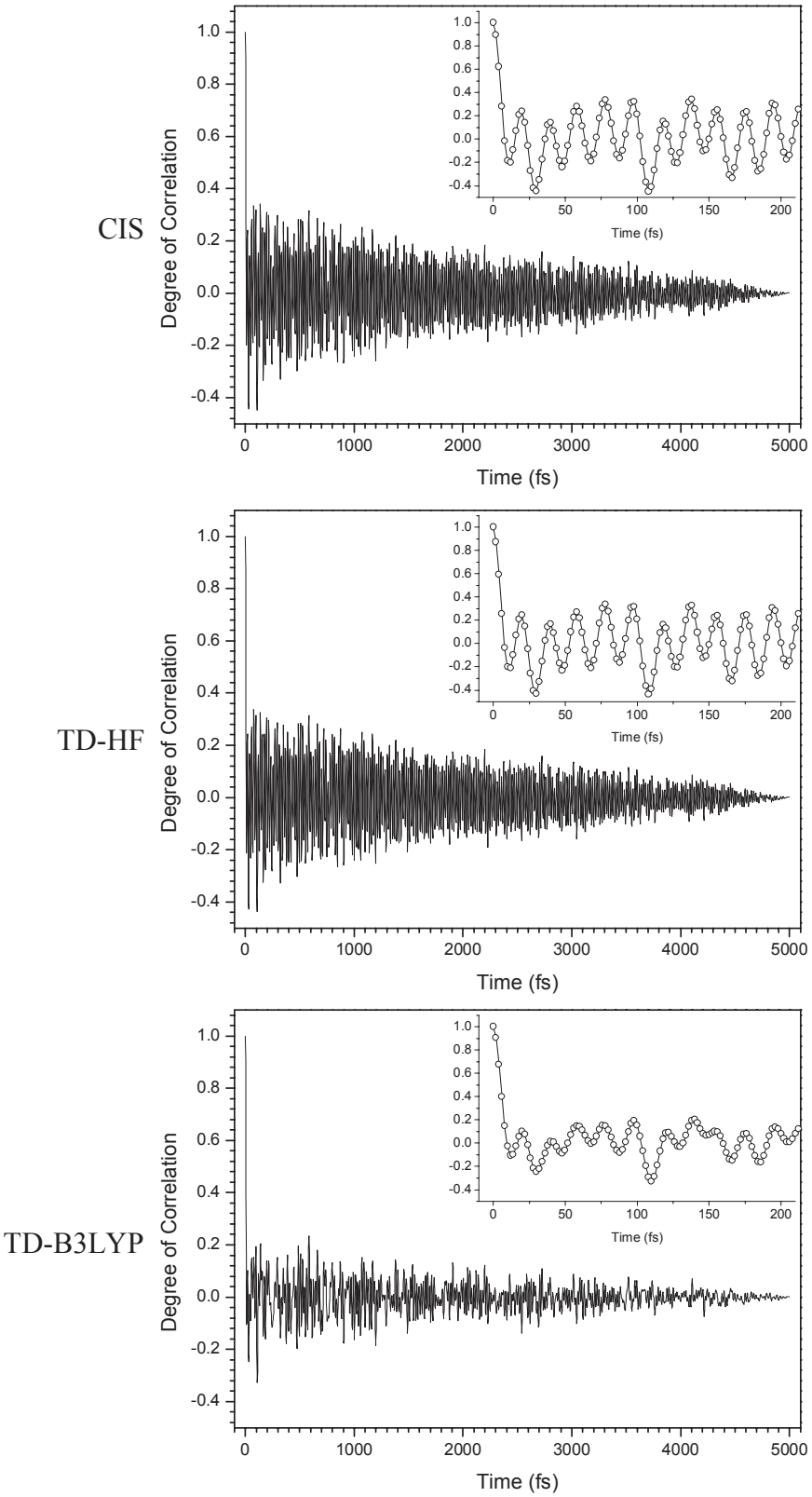



**Figure 5-3** Excited state energy gap correlation functions for simulation 1, NADH residue 753, calculated from CIS, TD-HF and TD-B3LYP data. Insets show magnified views. Missing energy gap points for TD-B3LYP data were interpolated prior to calculating the auto-correlation function.

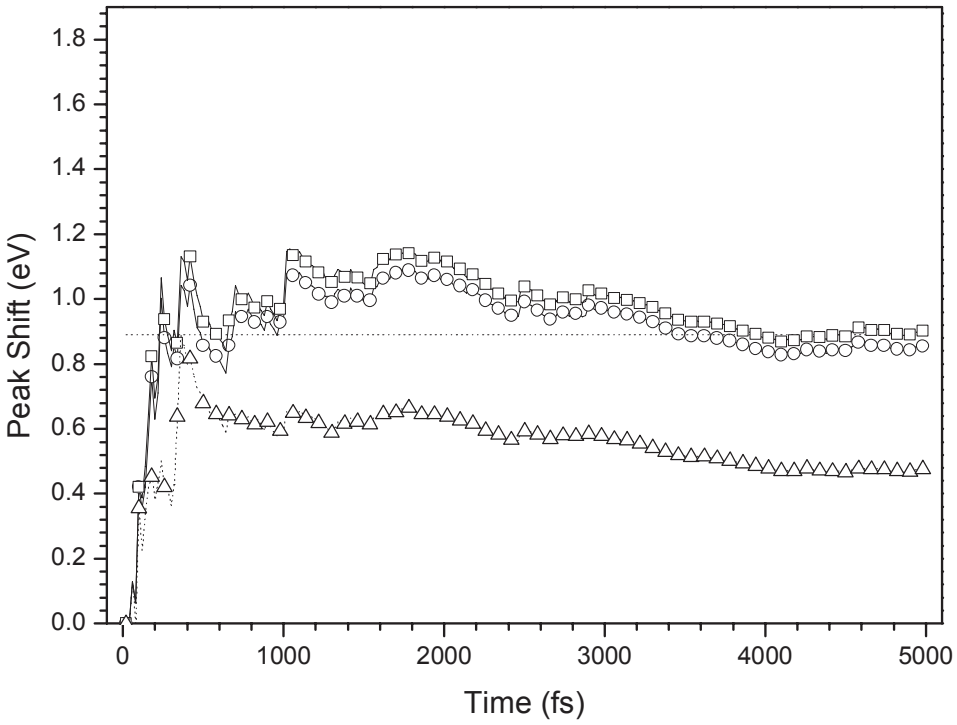



**Figure 5-4** Peak shift convergence for simulation 1, NADH residue 753, calculated using CIS (circles), TD-HF (squares) and TD-B3LYP (triangles). The experimental peak shift is shown by a dotted line.

**Table 5-3** Summary of theoretical spectral data obtained for simulation 1, NADH residue 753, using CIS, TD-HF, TD-B3LYP and ZINDO for the energy gap evaluations.

| | Unfitted Data NADH Residue 753 | |
|---|---|---|
| **Method** | **Peak Shift** | **Absorption Width[b]** |
| CIS | 0.850 | 0.531 |
| TD-HF | 0.897 | 0.551 |
| TD-B3LYP[a] | 0.473 | 0.381 |
| ZINDO[c] | N/A | N/A |
| **Experiment** | **0.872 ± 0.05** | **0.481 ± 0.01[d]** |

[a]A total of 8 missing energy gap data points were spline interpolated in order to predict this data.
[b]Measured width at intensity = 0.5
[c]Convergence problems on over 20 % of the energy gap evaluations prevented theoretical spectra being calculated for the ZINDO method.
[d]Measured from Gaussian fit to absorption data.

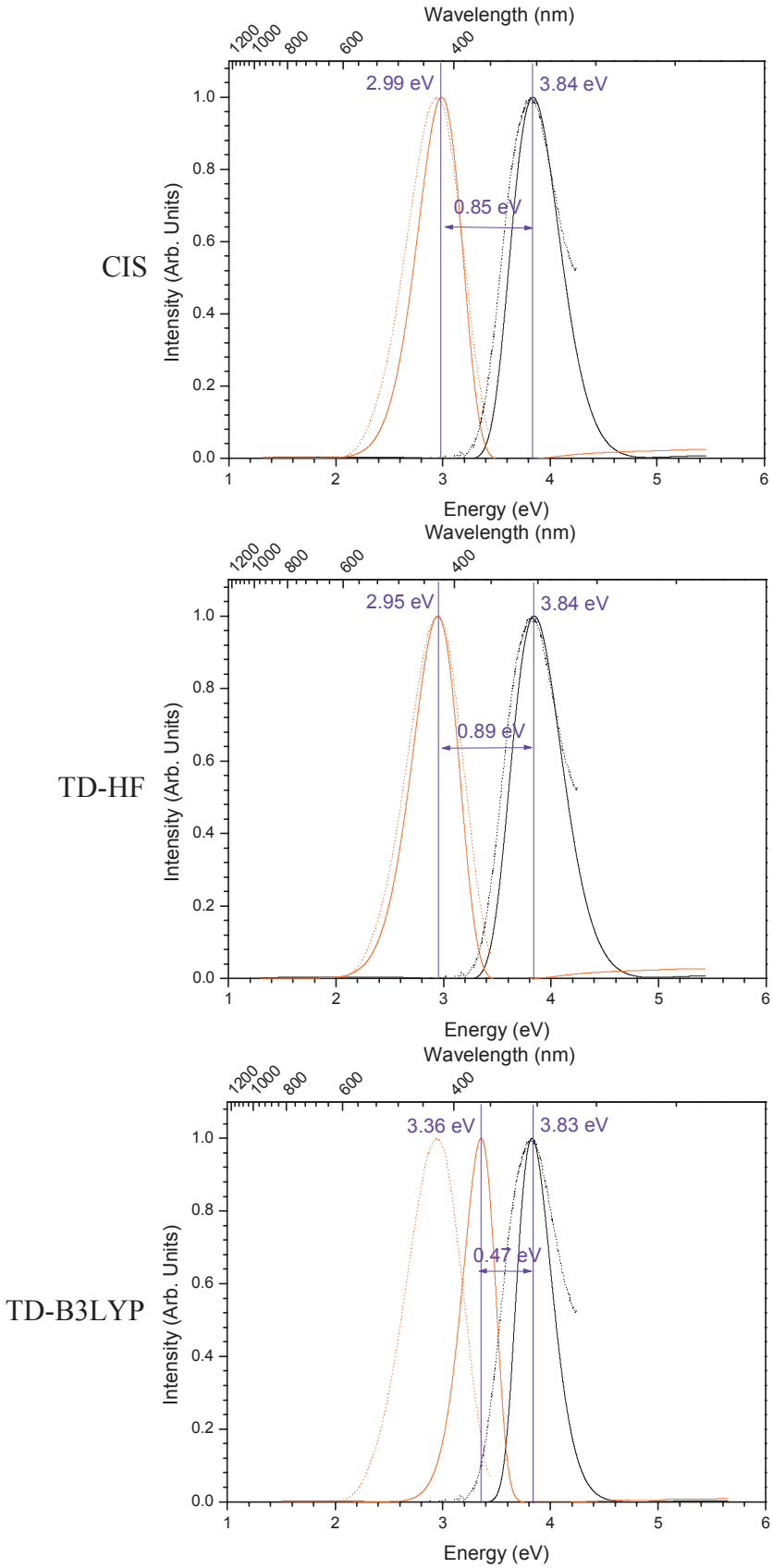



**Figure 5-5** Theoretical absorption (black line) and emission (red line) spectra generated from energy gap fluctuations calculated using CIS, TD-HF or TD-B3LYP for simulation 1, NADH residue 753. Peak locations and shifts are shown in blue. Dotted lines show experimental data.

# Chapter 6

# Work in Progress

## 6.1 Introduction

The work presented in Chapters 1 to 5, based on the methodology developed by Mercer *et al.* [Mercer *et al.* 1997; Mercer *et al.* 1999], represents a significant advance on the previous studies in terms of both computation and complexity. The LADH protein system investigated is considerably larger and more structurally diverse than either the chlorophyll-*a* or bacteriochlorophyll-*a* chromophores, necessitating an increase in computational time and resources as well as the development of automated systems for running the calculations and analysing the results. Nevertheless, as with the previous studies of Mercer *et al.* there is excellent agreement between theory and experiment.

The agreement between theory and experiment obtained for the absorption and emission spectra of LADH (Chapter 4) has shown that the QM/MM based linear response method can be used to study systems with large reorganisation energies. The method has now been used to study both small molecules in solution (chlorophyll-*a* and bacteriochlorophyll-*a*) [Mercer *et al.* 1997; Mercer *et al.* 1999] and a protein (LADH) [Walker *et al.* 2002]. These systems represent Stokes shifts as small as 80 meV and as large as 0.87 eV. As such this work has shown that it should be possible to use the linear response approach to study a broad range of biologically interesting systems. It is hoped that this approach can be extended to study and explain ultra-fast spectroscopy results and also provide an alternative

method for predicting the reactivity of enzymes without the need for knowledge of the reaction pathway or transition state. Work is currently in progress to achieve these aims and this chapter is designed to give a brief overview of the preliminary results that have been obtained at the time of writing.

The work covered here is a direct extension of the work discussed in this thesis. Some brief examples of the extensions of the linear response approach are given along with a more in depth discussion of work involving the zinc substituted form of the oxygen storage protein myoglobin which has lead to some interesting preliminary observations.

## 6.2 Spectral Properties of Zinc-Myoglobin

One of the long term aims of this work is to use the QM/MM approach to predict and explain ultra-fast spectroscopy experiments such as 3-pulse-echo-peak shifts (3PEPs) and transient gratings [Cho 2003]. If steady-state spectra can be successfully reproduced then an extension to ultrafast spectroscopy can be made with confidence. Unfortunately the lack of experimental ultrafast spectroscopic data for LADH meant that theoretically derived spectra could not be compared directly with experiment resulting in it being of limited use. Instead attempts have been made to reproduce the experimental 3PEPs for zinc-myoglobin produced by Cho [Cho 2003]. This was initially tried by Amer [Amer 2001]. Attempts have been made to extend the work of Amer [Amer 2001] by carefully re-calculating the MD simulations on zinc-myoglobin and then evaluating the energy gap fluctuations for the myoglobin zinc-porphyrin unit. The preliminary results, and the problems encountered, are detailed below[36].

[36] For an overview of Zinc-Myoglobin, its chemical and biological properties, its experimental steady state and ultra-fast spectra, and previous attempts to use the QM/MM approach for studying it the reader is referred to the following work: Amer, H. (2001). QM/MM Studies of Myoglobin. *Chemistry*. London, Imperial College of Science, Technology and Medicine, Cho, B. (2003). Protein Dynamics Measured with Non-Linear Spectroscopy. *Department of Chemistry*. London, Imperial College London.

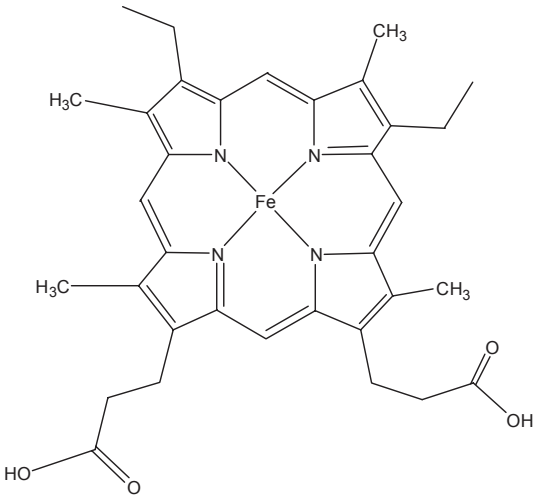



**Figure 6-1** 2D schematic of the myoglobin heme unit (iron-protoporphyrin IX).

The work discussed in this section forms the basis of the paper [Walker *et al.* 2003a] which when complete will be submitted to the Journal of Physical Chemistry B.

## 6.2.1 Myoglobin – Overview[37]

Myoglobin (Figure 6-2) is an iron-based oxygen transport and storage protein found in the muscles of vertebrates. It was one of the first protein structures to be solved using X-ray crystallography by John Kendrew in 1960 [Kendrew 1960]. Since then, technological and methodological advances have enabled an increasingly sophisticated understanding of the structure of myoglobin and its relation to function. A search of the Protein Data Bank [Berman *et al.* 2000] today returns over 200 structures, including numerous mutant variants and ligated structures of the protein, some approaching atomic resolution [Vojtechovsky *et al.* 1999]. Neutron diffraction studies have also allowed the location of the hydrogen atoms to be identified [Shu *et al.* 2000].

[37] The discussion here is designed to simply give a brief overview of Myoglobin. For an in-depth discussion of the structure of Myoglobin and a review of the literature, including previous theoretical simulations of Myoglobin, and discussion of the suitability of the CIS method for studying Zinc-Myoglobin the reader is referred to the work of Amer, H. (2001). QM/MM Studies of Myoglobin. *Chemistry*. London, Imperial College of Science, Technology and Medicine..

Myoglobin is a compact, predominantly (75 %) alpha helical [Kendrew 1961], globular protein consisting of only 153 amino acid residues with a molecular weight of *ca.* 17,000. The 8 alpha helices, labelled A to H in Figure 6-2, form an amphipathic pocket that stabilises the essential prosthetic group, iron-protoporphyrin IX (Figure 6-1). The latter is referred to as heme when in the $Fe^{2+}$ oxidation state and hemin when in the $Fe^{3+}$ state. The $Fe^{2+}$ iron, at the centre of the heme cofactor, binds small molecules such as $O_2$, NO, CO and $H_2$ in a reversible fashion for all ligands except CO.

The small size and important function of myoglobin have contributed to it becoming one of the most widely studied proteins. Indeed it has been described as "*a paradigm for the development of structure-function relationships in many biophysical sciences*" [Phillips and Pettitt 1995] or, more simply, "*the hydrogen atom of biology*" [Frauenfelder *et al.* 1991].

### 6.2.2 Spectroscopic and Electronic Properties

A wide variety of spectroscopic techniques have been used to study myoglobin including resonance Raman [Champion 1976; Stavrov 1993; Spiro and Czernuszewicz 1995; Schelvis and Varotsis 2000], Mössbauer [Ober *et al.* 1997; Trautwein *et al.* 1998], NMR [Lecomte and Lamar 1985; Lukin *et al.* 2000], and ESR [Dickinson and Symons 1983]. The aspects most relevant to this work concern the absorption and emission behaviour of the heme unit in the protein environment. It is generally accepted that "*[porphyrin] fluorescence is not observed in native myoglobin*" [Ahn *et al.* 1993], although the heme unit can accept radiationless energy transfer from the excited states of surrounding tryptophan residues. This phenomenon of 'intrinsic fluorescence' is well documented and acts a useful method for investigating conformational changes in the protein [Postnikova *et al.* 1991; Gryczynski *et al.* 1995; Gryczynski *et al.* 1997; Postnikova 1999; Mataga *et al.* 2000].

Recent work by Stefan Abend [Abend 2000] at Imperial College London has indicated that myoglobin does, in fact, undergo radiative de-excitation from the heme unit, although the non-reproducibility of the data and the likelihood of sample contamination have cast doubt on this conclusion. For this reason the zinc-substituted form of myoglobin, which has well known fluorescent properties [Albani and Alpert 1987; Ahn *et al.* 1993; Murakami and Kushida 1994; Ahn *et al.*

1995], has been used in this work. Fluorescent behaviour is also observed in magnesium-substituted porphyrins [Balog *et al.* 1997].

Recent theoretical attempts to calculate the absorption spectra of porphyrins and identify the excited states have included *ab initio* coupled-cluster [Gwaltney and Bartlett 1998] and configuration interaction singles (CIS) [Yamamoto *et al.* 2000] approaches, both of which reproduce the experimental absorption spectra reasonably well. The computational complexity of such techniques is such that the size of system that can be investigated is limited and there are no reports in the literature of similarly calculated protein absorption spectra.

## 6.2.3 Calculations

The simulation protocol used for studying zinc-myoglobin is consistent with the methodology developed in this work for studying LADH.

### 6.2.3.1 Molecular Dynamics Protocol

The starting structure for zinc-myoglobin was obtained from the horse-heart iron-myoglobin crystal structure determined by Maurus *et al* [Maurus *et al.* 1997] (RCSB Protein Data Bank 1WLA). The Fe atom was replaced with Zn, to give zinc-myoglobin. No bonded interactions between the Zn atom and the heme unit were included, and so the zinc moiety was effectively modelled as a $Zn^{2+}$ ion.

Due to the lack of experimental data for the histidine tautomeric states in horse-heart myoglobin, the states for the sperm-whale variant of myoglobin were used. These have been well characterised by a combination of NMR, pK and modelling studies [Bashford *et al.* 1993; Bhattacharya *et al.* 1997]. Histidine residues 12, 48, 64, 81, 82, 97, 113, 116 and 119 were therefore defined as being protonated on the $\varepsilon$ nitrogen (HIE) and residues 24, 36 and 93 as protonated on the $\delta$ nitrogen (HID). Parameters for heme are not available in the standard Amber force field [Cornell *et al.* 1995] and so the parameters of Giammona [Giammona 1984] were used.

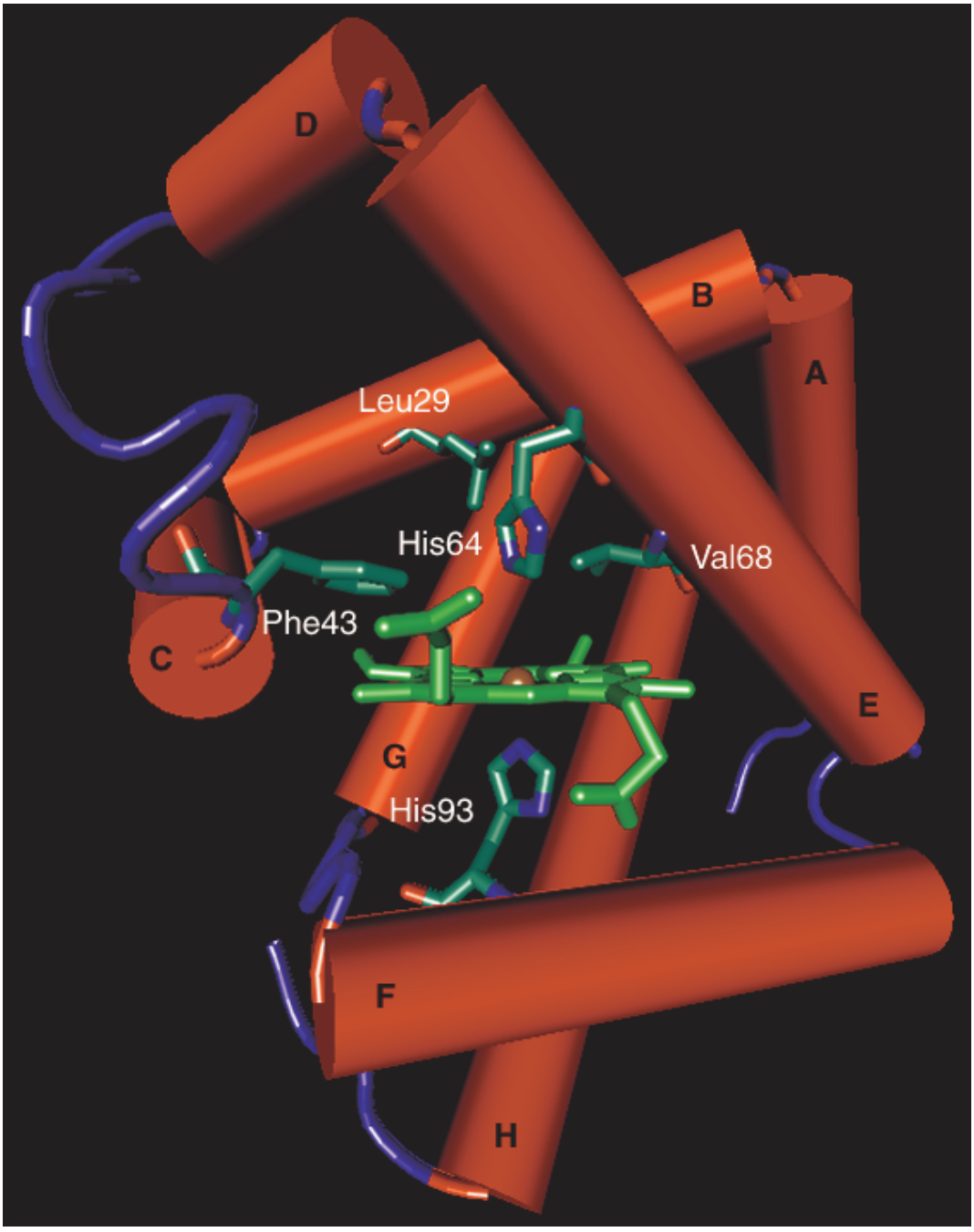



**Figure 6-2** Cartoon representation of the crystal structure of horse-heart myoglobin [Maurus *et al.* 1997], indicating the heme unit (green), the proximal histidine (His93), and the residues of the distal pocket. Alpha helices are shown as cylinders. Adapted from [Amer 2001]

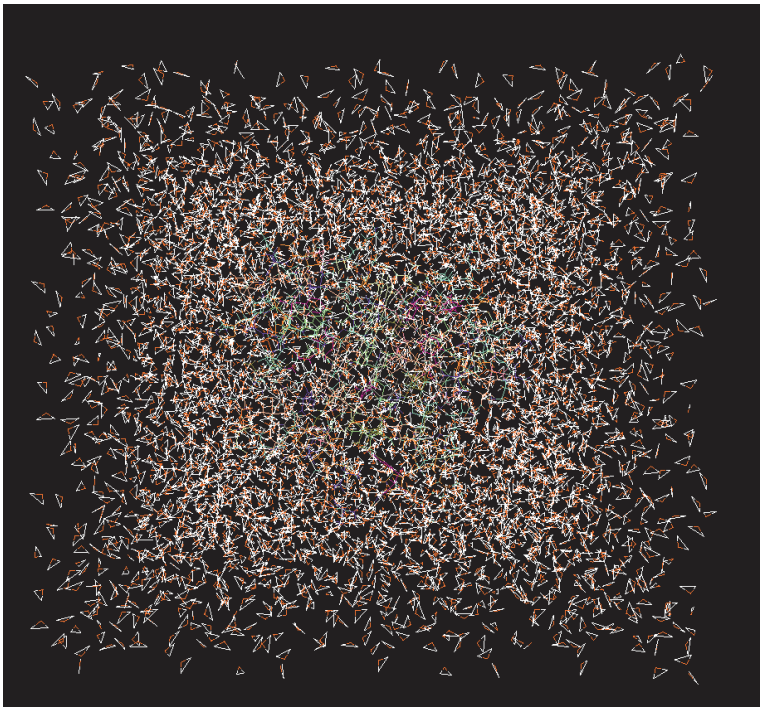


**Figure 6-3** 1WLA starting structure after charge neutralisation and solvation. Coloured by residue.

The hydrogen atoms, not present in the crystal structure, were added at idealised bond lengths and angles as necessary to fulfil valence requirements. The system was then neutralised by the addition of two $Na^+$ ions distributed in a shell around the zinc-myoglobin using a coulombic potential on a 1 Å grid. The whole system was then solvated using the TIP3P [Jorgensen *et al.* 1983] water model by placing it in a box of dimensions 66.2 Å x 60.9 Å x 62.3 Å containing 5,638 water molecules, to yield a system of 19,404 atoms with an average density of 0.788 g $cm^{-3}$ (Figure 6-3).

Atom centred point charges for the non-standard Zn-heme unit were derived for the ground state in accordance with the RESP [Bayly *et al.* 1993] method for the Amber force field. The heme unit hydrogen atoms were optimised using Hartree-Fock [Roothaan 1951] STO-3G [Hehre *et al.* 1969] calculations implemented within the Gaussian 98 package [Frisch *et al.* 2001]. All heavy atoms were kept fixed during these calculations. The electrostatic potential, required for the RESP fitting procedure, was obtained using a single point HF calculation on the hydrogen optimised Zn-heme structure. The SV 6-31G*[Binkley *et al.* 1980]

basis set was used for all atoms. The derived atom centred point charges are given in Table 6-1.

A molecular dynamics simulation of the zinc-myoglobin in water system was performed using similar methodology to the LADH work (Chapter 2). The program Sander, incorporated within the AMBER 6.0 [Case *et al.* 1999] suite of molecular dynamics programs was used to carry out a classical MD simulation of the solvated zinc-myoglobin system. The system was first subjected to 200 steps of steepest descent, followed by 800 steps of conjugate gradient minimisation in order to alleviate incorrect van der Waals contacts created by hydrogenation and solvation of the system. An atom based non-bonded cut-off of 9 Å was used for this and all subsequent simulations. The system was then subjected to 30 ps of slow heating from 0 to 300 K following the method of Berendsen [Berendsen *et al.* 1984] to control the temperature. After slow heating the system was equilibrated at 300 K for 100 ps. Equilibration was deemed to have been successfully obtained when the Root Mean Square Deviation (RMSD) of the protein $\alpha$ carbons was reasonably small ( < 1.4 Å) and both the RMSD and classical energies of the system fluctuated by less than 10 % over a timescale of approximately 20 ps. An energy gap production run was then performed for the subsequent 22 ps with the complete system coordinates being recorded every 4 fs, resulting in a trajectory of 5,500 coordinate sets. The decision to use an increased time step of 4 fs over the 2 fs used for LADH was made in order to reduce the computational cost. The reduced peak shift of myoglobin meant that a large time step could be used without introducing aliasing errors within the Fourier transform routines.

In all stages of the MD simulation constant pressure periodic boundary conditions using the Particle Mesh Ewald method [Essmann *et al.* 1995] were employed, the integration time step was set at 1 fs and all interactions were calculated at every step. With the exception of the rigid water model, no atoms had their positions fixed or their motions damped.

#### 6.2.3.2 Quantum Mechanical Calculation of Energy Gaps

The theory used for calculating optical spectra from a fluctuating energy gap is was discussed in Chapter 4.

The trajectory obtained from the 22 ps production run of the MD simulation was used to create time ordered structures for the QM calculations. The method

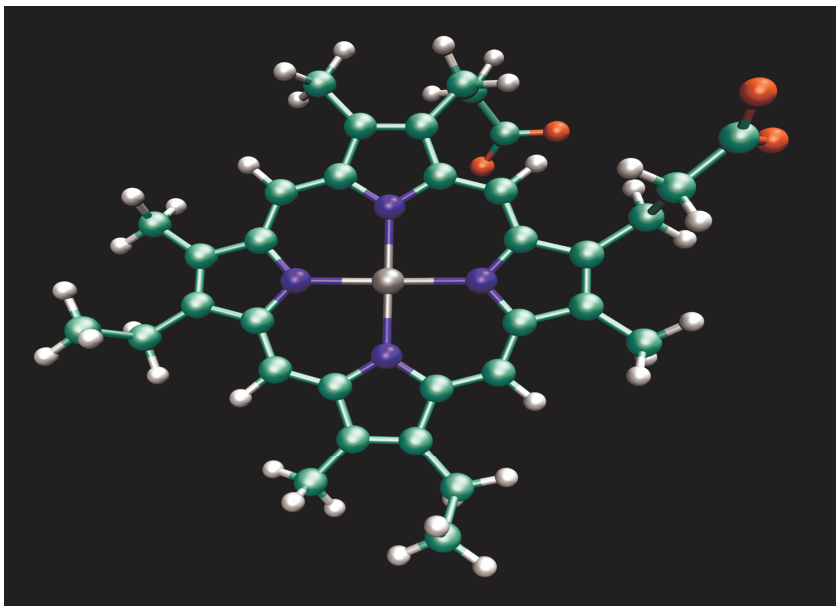


**Figure 6-4** Zinc substituted heme unit used in the QM section of the zinc-myoglobin QM/MM energy gap evaluations. Coloured by atom type. Blue = nitrogen, Red = oxygen, White = hydrogen, Grey = carbon and Silver = zinc.

used for the QM evaluation of the energy gaps was the same as that used for LADH (Chapter 3) but with a 4 fs time step instead of a 2 fs time step. This meant phase space could be covered in half as many calculations. Single point CIS [Foresman *et al.* 1992] calculations utilising the 3-21G* basis set were performed to calculate the singlet energy gap between the ground and first two excited states for the Zn-heme unit at 4 fs intervals. Trial CIS calculations showed that the orbitals involved in the excitation were centred purely of the Zn-heme moiety and thus it was decided to treat the Zn-heme residue quantum mechanically while treating the rest of the protein and surrounding water classically, giving a system of 77 atoms corresponding to 475 basis functions (Figure 6-4). In this way the need for link atoms was negated since the Zn-heme residue is not formally bound to the protein matrix. The point charges of the classical system were included in the one-electron Hamiltonian of the quantum element of the calculations.

A total of 2,500 points representing the first 10 ps of the 22 ps production run were calculated for the Zn-heme residue. The evaluation of the energy gaps yielded two degenerate excitations of very similar energy and oscillator strength (Figure

6-5). This degeneracy in the excited states complicates matters considerably and is the subject of ongoing work. This is discussed below.

### 6.2.4 Degeneracy in the Excited States of Zinc-Myoglobin

The current implementation of the QM/MM based linear response approach requires a single energy gap oscillation as a function of time. This requires the most prominent excitation to be extracted from the QM/MM calculations. For LADH this was a fairly simple task as the excited state corresponding to the nicotinamide excitation is easily identified. For zinc-myoglobin, however, this is not so easy. The problem lies in the existence of two degenerate excited states. Figure 6-5 shows the predicted energy gaps for zinc-myoglobin, calculated in an identical fashion to the LADH work discussed in Chapter 3. It can be seen from Figure 6-5 that there are two excitations of very similar energy. The problem is in choosing how to label these two excitations. The original work of Amer [Amer 2001] attempted to label these two states in terms of their transition dipole moments which were believed to be orthogonal. The assumption that the transition dipole moments were orthogonal warranted further investigation. The extracted TDMs for the two excited states are, after the removal of structural motion, shown in Figure 6-6[38]. From this it can be seen that the TDM directions from the two states occur in the same plane. There is a small avoided region, the origin of which is not currently known. This result shows, however, that TDM direction is not a valid method for identifying the two states. Indeed they do appear to be truly degenerate. Simply using the lowest energy excited state yields the UV/VIS spectral prediction shown in Figure 6-7. Although the widths are reasonably good the peak shift is predicted to be 19 meV which compared to an experimental peak shift of 7 ± 0.5 meV represents and error of over 170 %.

As such in order to accurately reproduce the spectral properties of zinc-myoglobin using the QM/MM linear response method it will be necessary to extend the theory to incorporate degeneracy.

---

[38] The TDM extraction and removal of structural motion via RMS fitting is implemented in the Computational Chemistry Tools package, by Ross Walker, provided on the accompanying CD-ROM. Also provided are Mathematica files that allow the TDMs shown in Figure 6-6 to be rotated in 3 dimensions.

### 6.2.5 Extension of the QM/MM Linear Response Approach

An initial attempt was made to extend the linear response approach to degenerate systems based on work by Khidekel, Chernyak and Mukamel [Khidekel *et al.* 1996]. In this paper they deal with incorporating both multiple electronic levels and transition dipole moments into the construction of the response functions. In the paper they deal only with nonlinear response functions and so it was necessary to convert these to a linear response function. This is summarised as follows:

Equation 2.14 in the Khidekel paper [Khidekel *et al.* 1996]

$$F_{DD}^{(2)}\left(\tau_1,\tau_2\right)=\sum_j D_{j0}^2 \exp\left\{\begin{matrix}-i\Omega_j\tau_{21}+\ddot{g}_{\delta\delta}(0)+\ddot{g}_{\delta\delta}\left(\tau_{21}\right)\\ -g_{d_j d_j}\left(\tau_{21}\right)+2i\ddot{g}_{\delta d_j}\left(\tau_{21}\right)\end{matrix}\right\} \tag{6.1}$$

is believed to collapse, in the case of a single time variable to the response function given in (4.20):

$$R\left(t\right)=e^{\left(-g_{dd}(t)+2i\ddot{g}_{\delta d}(t)+\ddot{g}_{\delta\delta}(t)\right)} \tag{6.2}$$

For the case where the TDMs are neglected the linear response function for a multilevel system is simply the sum of the individual response functions. The overall response function for the multilevel system would thus be

$$R_{12}\left(t\right)=\sum_{j=1}^{2}R_j\left(t\right) \tag{6.3}$$

The problem with this approach, however, is that there is no cross correlation between the energy gaps. This is something that is essential to include if the energy gaps are not independent. Thus while (6.3) is sufficient for treating systems where the excited state energy gaps do not interact it is not suitable for studying zinc-myoglobin where, as discussed in the next section, the energy gaps are not independent. Future work therefore needs to concentrate on adapting the linear response approach for tackling degenerate systems.

### 6.2.6 Anisotropy of Zinc-Myoglobin Probed Using the Linear Response Approach

One aspect that has arisen while considering how to solve the problem of multiply degenerate excitations is the extent to which the two excitations in zinc-myoglobin are independent. One way to investigate this is to look at the angle between the TDMs of the two excited states as a function of time. The angle between two 3 dimensional vectors is:

$$\theta = \cos^{-1} \frac{\boldsymbol{v}.\boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|} \tag{6.4}$$

where

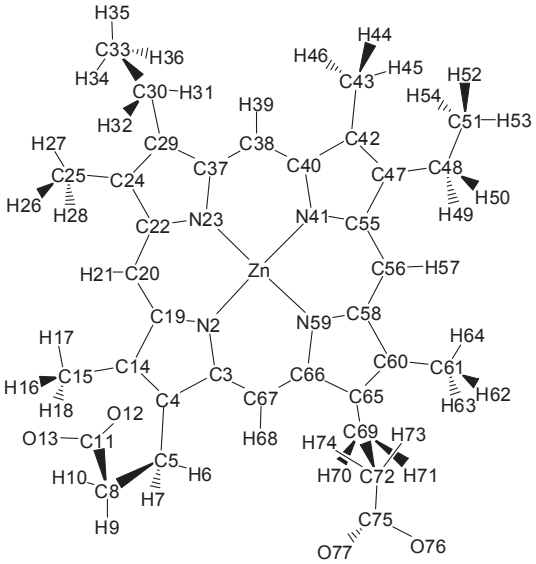
$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}.\boldsymbol{v}} \tag{6.5}$$

The angle between the TDMs as a function of time for the two calculated excitations in Zn-Mb is shown in Figure 6-8. It is obvious from this plot that the TDM angles are not random as would be expected if the excitations were independent. This is made clearer by Figure 6-9 which shows the mean TDM angle as a function of time. From this it can be seen that the two TDMs are on average orthogonal to each other. The interaction between the two excited states is confirmed by taking the auto-correlation of the TDM angles using the equation

$$Corr(g,h)_j = \sum_{k=0}^{N-1} g_{j+k} h_k \tag{6.6}$$

The zero padded auto-correlation function over the first 500 fs is shown in Figure 6-10**A**. For comparison the auto-correlation function for a set of uniform random numbers is shown in Figure 6-10**B**. From these two plots it is immediately obvious that there is memory in the TDM angles that lasts for approximately 1 ps. It is hoped that this data can be used to generate a theoretical prediction of the anisotropy in zinc-myoglobin. Current work is concerned with generating an anisotropy prediction and comparing this to experimental data.

**Table 6-1** Atomic partial charges used for the porphyrin unit of zinc-substituted myoglobin, fitted to the calculated molecular electrostatic potentials according to the RESP method of Bayly *et al.* [Bayly *et al.* 1993]




Zn-Mb porphyrin group atom types and charges

| *atom* | *type* | *charge* | *atom* | *type* | *charge* | *atom* | *type* | *charge* |
|---|---|---|---|---|---|---|---|---|
| Zn | ZN | 1.01669 | H27 | HC | 0.06177 | H53 | HC | 0.00019 |
| N2 | NP | -0.49031 | H28 | HC | 0.04681 | H54 | HC | 0.02801 |
| C3 | CC | 0.14780 | C29 | CB | -0.04446 | C55 | CC | -0.02386 |
| C4 | CB | -0.05709 | C30 | CT | 0.01364 | C56 | CD | -0.04352 |
| C5 | CT | -0.00517 | H31 | HC | 0.01297 | H57 | HC | 0.07141 |
| H6 | HC | 0.00306 | H32 | HC | 0.00814 | C58 | CC | 0.01204 |
| H7 | HC | 0.00306 | C33 | CT | -0.01514 | N59 | NO | -0.43114 |
| C8 | CT | 0.06110 | H34 | HC | -0.00262 | C60 | CB | 0.04520 |
| H9 | HC | -0.02673 | H35 | HC | -0.00262 | C61 | CT | -0.13838 |
| H10 | HC | -0.07037 | H36 | HC | -0.00262 | H62 | HC | 0.03987 |
| C11 | C | 0.70741 | C37 | CC | -0.02130 | H63 | HC | 0.02703 |
| O12 | O2 | -0.78807 | C38 | CD | -0.14138 | H64 | HC | 0.07299 |
| O13 | O2 | -0.78849 | H39 | HC | 0.14707 | C65 | CB | -0.02777 |
| C14 | CB | 0.04791 | C40 | CC | 0.00187 | C66 | CC | 0.08495 |
| C15 | CT | -0.11610 | N41 | NP | -0.38759 | C67 | CD | 0.03505 |
| H16 | HC | 0.04176 | C42 | CB | 0.03162 | H68 | HC | 0.17036 |
| H17 | HC | 0.04176 | C43 | CT | -0.15148 | C69 | CT | -0.08669 |
| H18 | HC | 0.04176 | H44 | HC | 0.04820 | H70 | HC | 0.08332 |
| C19 | CC | 0.02379 | H45 | HC | 0.04840 | H71 | HC | 0.03810 |
| C20 | CD | -0.11630 | H46 | HC | 0.03770 | C72 | CT | 0.03682 |
| H21 | HC | 0.14289 | C47 | CB | -0.03452 | H73 | HC | -0.01265 |
| C22 | CC | 0.02348 | C48 | CT | 0.02005 | H74 | HC | -0.06307 |
| N23 | NO | -0.37815 | H49 | HC | 0.04630 | C75 | C | 0.74898 |
| C24 | CB | 0.00421 | H50 | HC | 0.02166 | O76 | O2 | -0.82350 |
| C25 | CT | -0.15121 | C51 | CT | -0.08751 | O77 | O2 | -0.81543 |
| H26 | HC | 0.04541 | H52 | HC | 0.02275 | | | |

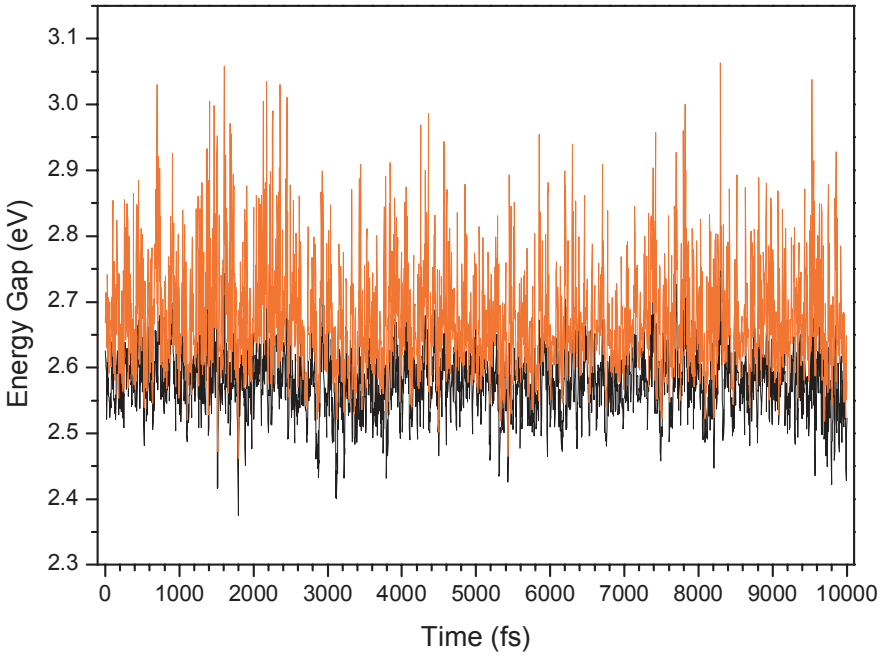



**Figure 6-5** Ground to excited state energy gap, as a function of time, for zinc-myoglobin. The presence of two degenerate states is clear. For clarity the two states have been coloured in terms of their energy.

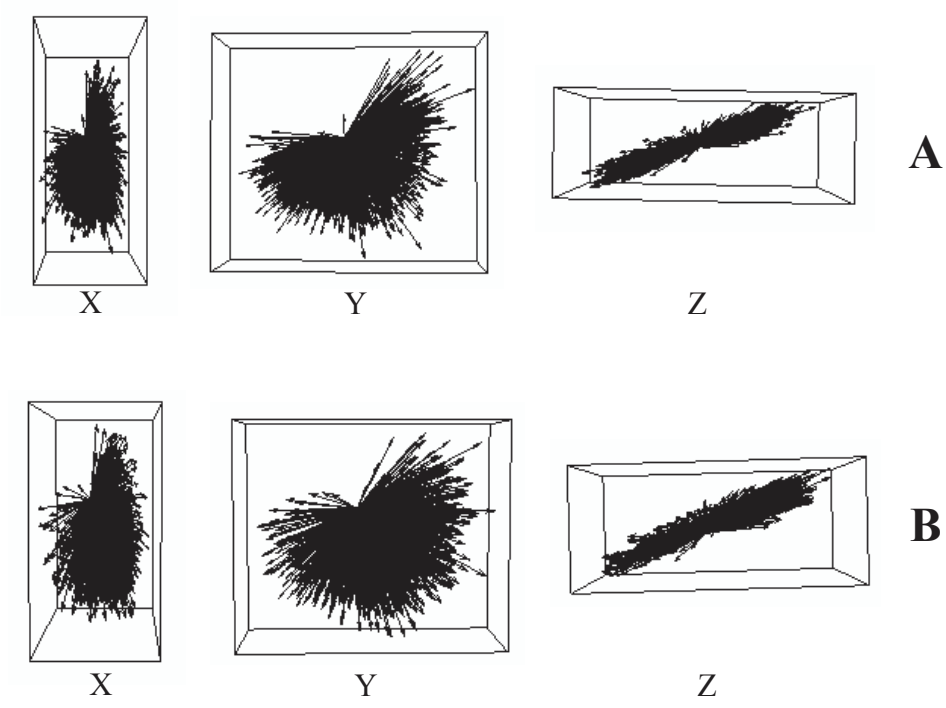



**Figure 6-6** Graphical representations of the Transition Dipole Moments of the two excited states of zinc-myoglobin shown in Figure 6-5. Representations along the three principle axes (X,Y,Z) are shown for **A**) Lowest energy excited state, **B**) Second lowest energy excited state.

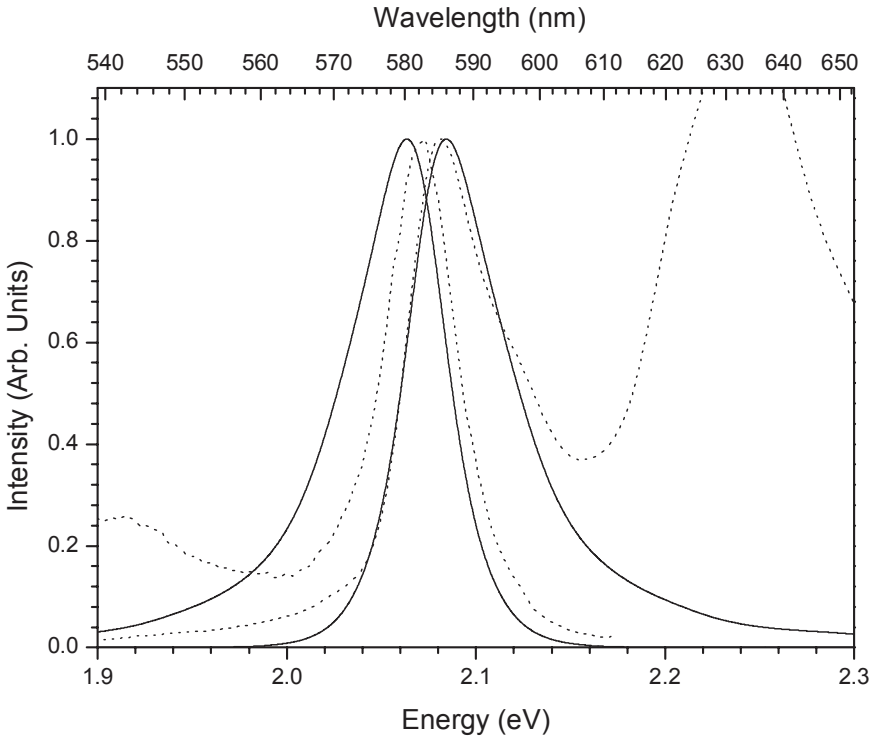



**Figure 6-7** Calculated (solid line) and experimental (dotted line) [Cho 2003] absorption and emission spectra for zinc-myoglobin using the lowest energy excitation. The calculated absorption spectrum has been shifted so that its centre of mass coincides with that of the experimental absorption spectrum.

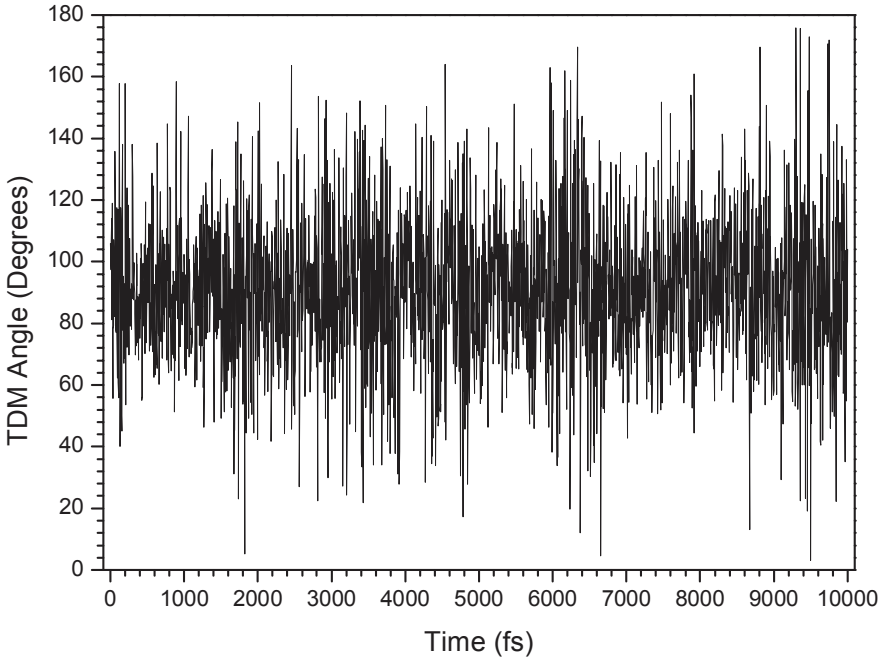



**Figure 6-8** Angle between the transition dipole moments of the two degenerate excitations in zinc-myoglobin as a function of time.

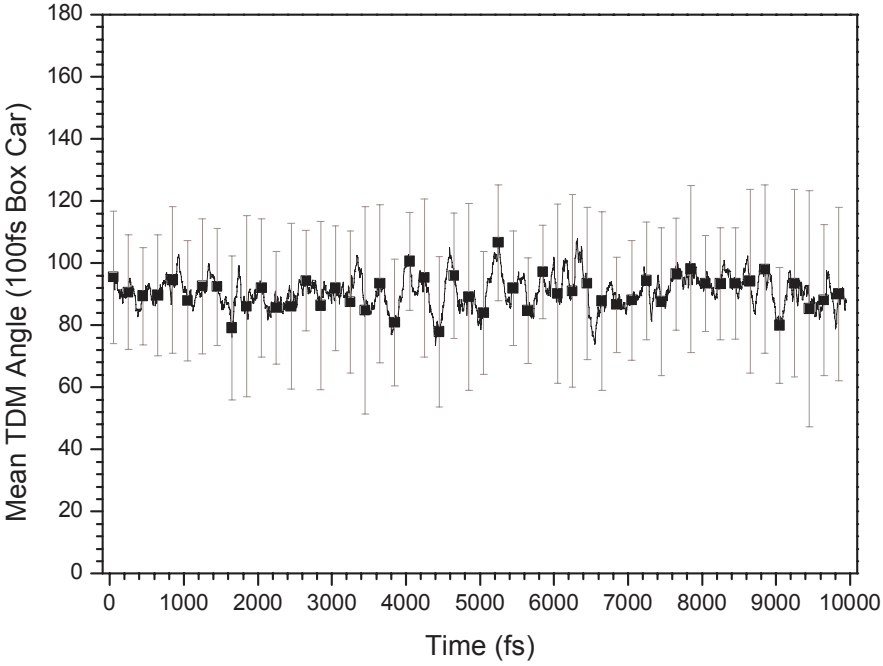



**Figure 6-9** Mean TDM angle between the two Zn-Mb degenerate states (black line) generated with a 100 fs wide moving box car average and standard deviation of TDM angles with 100 fs box (grey line).

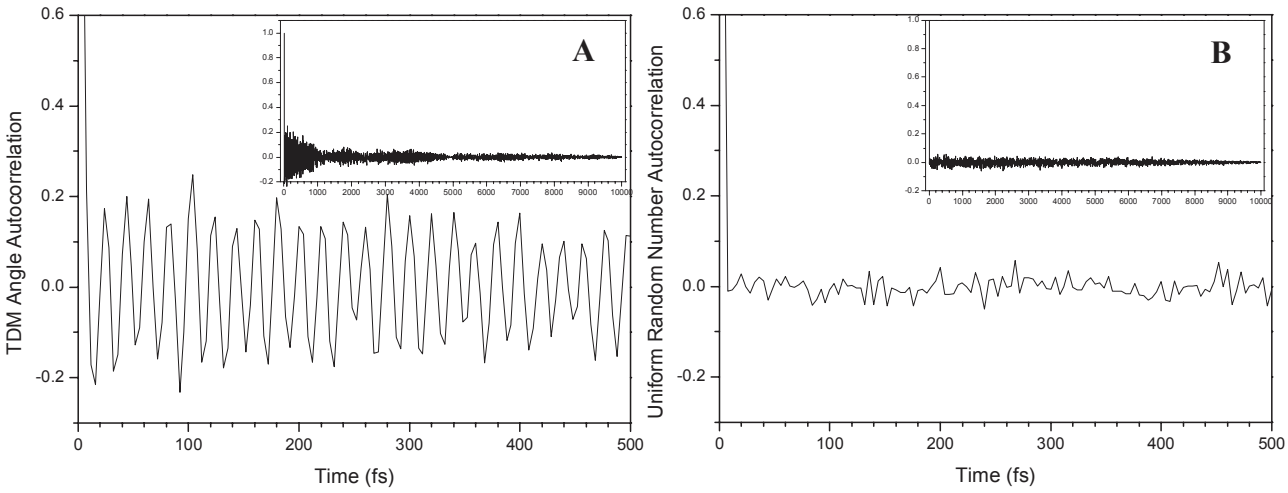



**Figure 6-10** Auto-correlation of the TDM angle between the two degenerate Zn-Mb excited states (A). A correlation between the TDM angles is clearly visible. Inset shows the full 10 ps time window. The auto-correlation of a set of 2,500 uniform random numbers (B) is provided for comparison.

## 6.3 Reactivity in LADH

An obvious extension of this work is to try to use the linear response approach to study the reactivity of LADH. It hoped that by simulating LADH containing benzaldehyde and NADH as well as LADH containing benzyl alcohol and $NAD^+$ it should be possible to calculate the free energy of reaction of LADH without the need for knowledge of the transition state structure. An approach for calculating free energies of reaction that is independent of the transition state structure is important for studying reactivity in proteins since the concept of a potential energy surface linking reactants and products via a transition state is an over simplification of protein activity. The number of different transition states, all of very similar energy, in a protein are essentially infinite and so any method that requires knowledge of the transition state structure is always likely to underestimate the true reaction rate.

The use of the QM/MM based linear response method for calculating reactivity differs from other attempts at studying LADH reactivity [Gao *et al.* 1998; Agarwal *et al.* 2000; Alhambra *et al.* 2000; Billeter *et al.* 2001; Cui *et al.* 2002; Tresadern *et al.* 2002; Tresadern *et al.* 2003], which have concentrated on calculating enzymatic reaction pathways and mechanisms, since the linear response approach employs the fluctuation dissipation theorem (§4.2). This approach allows macroscopic data to be obtained from equilibrium fluctuations. It is hoped therefore that equilibrium fluctuations sampled from simulations of LADH and coenzyme in both oxidised and reduced form can be used to calculate free energy curves from which the crossing point can be obtained. In this way mutations can be made to the LADH protein and their effect on the free energy of reaction calculated without the need to simulate the reaction path.

Preliminary calculations on a simulation of LADH+NADH+Benzaldehyde created from a X-ray structure of LADH+$NAD^+$+2,3,4,5,6-penta-fluorobenzyl alcohol (PDB ID[39]: 1HLD [Ramaswamy *et al.* 1994]) have proved successful, showing both that an MD simulation of an artificially created PDB can successfully reproduce the equilibrium dynamics of LADH in a reactive configuration and that

[39] http://www.rcsb.org/pdb/

the use of 3-21G* on the nicotinamide moiety and STO-3G* on the rest of NADH can still accurately reproduce the electronic energy gap. Figure 6-11 shows that the peak shift prediction is comparable with experiment and the use of a mixed basis set suggests that truncation of the NADH at the nicotinamide phosphate should not adversely affect the results. Future work will concentrate on simulating both the NADH and $NAD^+$ configurations and on running the QM/MM calculations on a region encompassing the main components of the active site, shown in Figure 6-12. From this data it will hopefully be possible to produce free energy surfaces from which the free energy of reaction can be calculated.

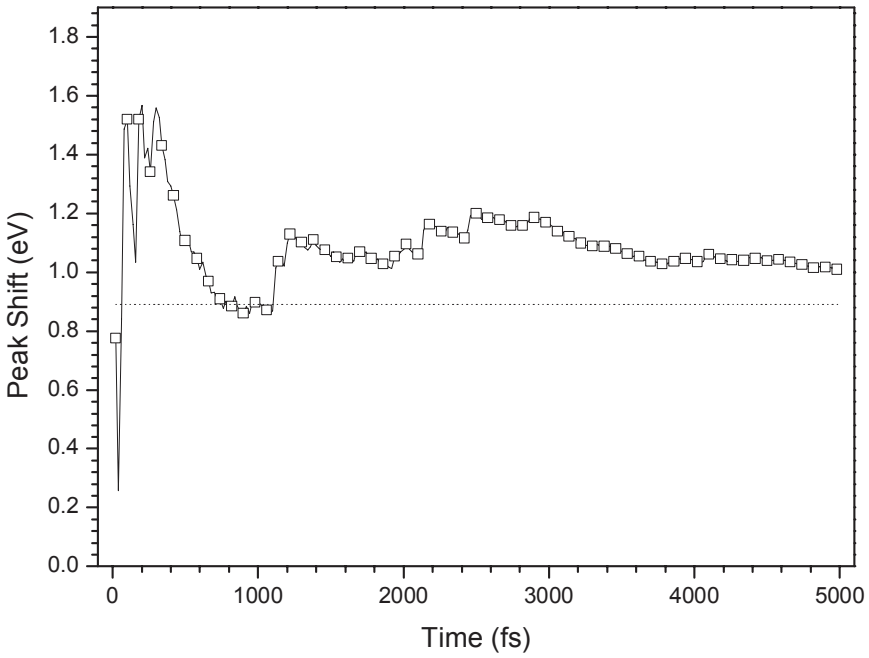



**Figure 6-11** Peak shift convergence of 1HLD+NADH simulation using CIS with 3-21G* on the nicotinamide and STO-3G* on the remainder of the NADH.

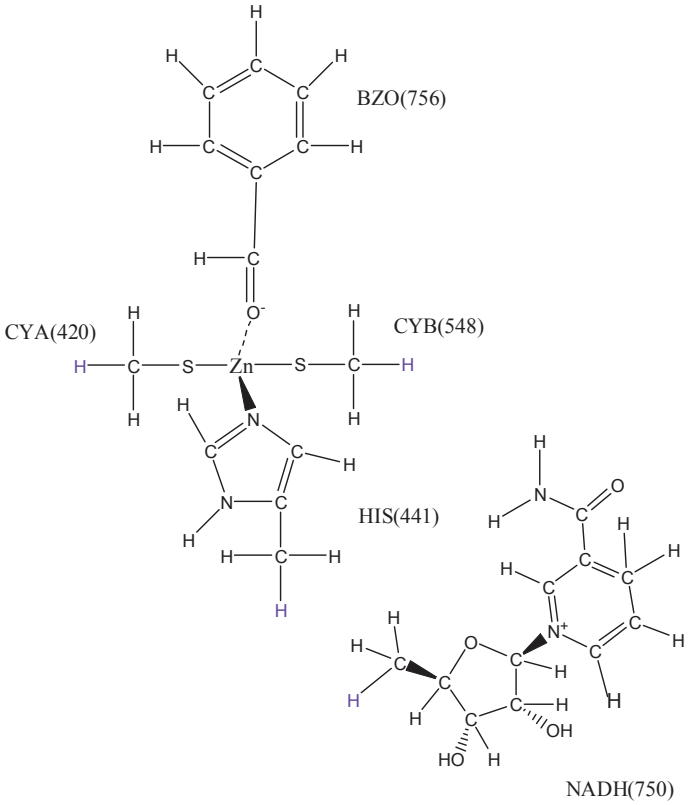



**Figure 6-12** Illustration of the QM region to be used for studying LADH reactivity. Residues are labelled (1HLD numbering) and the truncated atoms are shown in blue.

## 6.4 Summary and Conclusions

As discussed previously the linear response approach has been shown to work for both solvated molecules and proteins and the approximation of linear response has been shown to hold for Stokes shifts as large as 0.87 eV. Initial attempts to extend the work to study reactivity in LADH have shown that the use of a mixed basis set (STO-3G* and 3-21G*) for the NADH is still sufficient to reproduce the Stokes shift to within 20 % of experiment suggesting that the NADH residue can be truncated at the nicotinamide phosphate for calculations involving the substrate and catalytic zinc without adversely affecting the results. The preliminary calculations have also shown that the method still works for LADH in a reactive configuration. It is hoped that two equilibrium energy gap calculations on LADH containing NADH and benzaldehyde and LADH containing $NAD^+$ and benzylalcohol can be used to study LADH reactivity. Generation of theoretical spectra from the LADH+NADH+benzaldehyde simulation will act as a check to ensure the equilibrium dynamics are being correctly sampled. Simulations of LADH with NADH and benzaldehyde, and LADH with $NAD^+$ and benzylalcohol are currently in progress.

Extension of the QM/MM linear response approach to the study of zinc-myoglobin has uncovered some interesting results. The presence of two excitations have meant that the QM/MM linear response approach in its current form cannot be used to predict spectra accurately for myoglobin. It is hoped that work by Mukamel [Mukamel 2003] can be incorporated within the QM/MM methodology of this work in order to enable the study of systems with multiple energy gaps.

Investigation of the two excitations predicted in the zinc-myoglobin simulations has uncovered evidence of correlation between the transition dipole moments of the two excitations implying that they are not independent. It is hoped that data relating to the angle between the TDM vectors of the two excitations can be used to generate a theoretical prediction of the anisotropy in zinc-myoglobin so extending the application of the QM/MM methodology still further.

# Appendix A - Common Unit Conversions

**Appendix A Table 1** Common unit conversion factors 1.

| | |
|---|---|
| 1 **kT*** | 0.0256 **eV** |
| 1 **cm$^{-1}$** | 0.124 **meV** |
| 1 **meV** | 8.1 **cm$^{-1}$** |
| 1 **THz** | 4.14 **meV** |

*Evaluated at room temperature, 297 K.

**Appendix A Table 2** Common unit conversion factors 2.

| **Units** | **kcal mol$^{-1}$** | **kJ mol$^{-1}$** | **eV** | **cm$^{-1}$** |
|---|---|---|---|---|
| **1 kcal mol$^{-1}$** | | 4.18 | $4.34 \times 10^{-2}$ | 350 |
| **1 kJ mol$^{-1}$** | 0.239 | | $1.04 \times 10^{-2}$ | 83.6 |
| **1 eV** | 23.1 | 96.5 | | $8.07 \times 10^{3}$ |
| **1 cm$^{-1}$** | $2.86 \times 10^{-3}$ | $1.20 \times 10^{-2}$ | $1.24 \times 10^{-4}$ | |

# References


Abend, S. (2000). Ultrafast Dynamics of Chlorins and Porphyrins in Proteins and Solution, Investigated by Time-Resolved Three Pulse Photon Echo Spectroscopy. *Department of Chemistry*, Imperial College London.

Agarwal, P. K., Webb, S. P. and Hammes-Schiffer, S. (2000). "Computational studies of the mechanism for proton and hydride transfer in liver alcohol dehydrogenase." *Abstracts of Papers of the American Chemical Society* **220**: 493-PHYS.

Ahlrichs, R., Bar, M., Baron, H. P., Bauernschmitt, R., Bocker, S., Ehrig, M., Eichkorn, K., Elliot, S., Haase, F., Haser, M., Horn, H., Huber, C., Huniar, U., Kattenek, M., Kolmel, C., Kollwitz, M., Ochsenfeld, C., Schafer, O. A., Schneider, U., Treutler, O., von Armin, M., Weigend, F., Weis, P. and Weiss, H. (1997). TURBOMOL, Institut Physikalische Chemie und Elektrochemie, Universitat Karlsruhe.

Ahn, J. S., Kanematsu, Y., Enomoto, M. and Kushida, T. (1993). "Determination of Weighted Density-of-States of Vibrational- Modes in Zn-Substituted Myoglobin." *Chemical Physics Letters* **215**(4): 336-340.

Ahn, J. S., Kitagawa, T., Kanematsu, Y., Nishikawa, Y. and Kushida, T. (1995). "Glass-Transition of Zn-Substituted Myoglobin Probed by Absorption and Site-Selective Fluorescence Spectroscopies." *Journal of Luminescence* **64**(1-6): 81-86.

Albani, J. and Alpert, B. (1987). "Fluctuation Domains in Myoglobin - Fluorescence Quenching Studies." *European Journal of Biochemistry* **162**(1): 175-178.

Alhambra, C., Corchado, J. C., Sanchez, M. L., Gao, J. L. and Truhlar, D. G. (2000). "Quantum dynamics of hydride transfer in enzyme catalysis." *Journal of the American Chemical Society* **122**(34): 8197-8203.

Allen, M. P. and Tildesley, D. J. (1987). "Computer Simulation of Liquids". Oxford, Clarendon Press.

Allinger, N. L. (1977). "Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms." *Journal of the American Chemical Society* **99**: 8127-8134.

Allinger, N. L., Yuh, Y. H. and Lii, J. H. (1989). "Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .1." *Journal of the American Chemical Society* **111**(23): 8551-8566.

Amer, H. (2001). QM/MM Studies of Myoglobin. *Chemistry*. London, Imperial College of Science, Technology and Medicine.

Amir, Y., Awerbuch, B., Barak, A., Borgstrom, R. S. and Keren, A. (2000). "An opportunity cost approach for job assignment in a scalable computing cluster." *Ieee Transactions on Parallel and Distributed Systems* **11**(7): 760-768.

Amos, R. D., Alberts, I. L., Andrews, J. S., Colwell, S. M., Handy, N. C., Jayatilaka, D., Knowles, P. J., Kobayashi, R., Laming, G. J., Lee, A. M., Maslen, P. E., Murray, C. W., Palmieri, P., Rice, J. E., Simandiras, E. D., Stone, A. J., Su, M. D. and Tozer, D. J. (1998). CADPAC: The Cambridge Analytic Derivatives Package. Cambridge, UK.

Andersen, H. C. (1983). "Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Calculations." *Journal of Computational Physics* **52**(1): 24-34.

Anderson, D. C. and Dahlquist, F. W. (1982). "Properties of Bound Trifluoroethanol Complexes with Horse Liver Alcohol-Dehydrogenase." *Biochemistry* **21**(15): 3569-3578.

Anderson, W. P., Cundari, T. R. and Zerner, M. C. (1991). "An Intermediate Neglect of Differential-Overlap Model for 2nd- Row Transition-Metal Species." *International Journal of Quantum Chemistry* **39**(1): 31-45.

Anderson, W. P., Edwards, W. D. and Zerner, M. C. (1986). "Calculated Spectra of Hydrated Ions of the 1st Transition-Metal Series." *Inorganic Chemistry* **25**(16): 2728-2732.

Andersson, I., Maret, W., Zeppezauer, M., Brown, R. D. and Koenig, S. H. (1981). "Metal ion substitution at the catalytic site of horse-liver alcohol dehydrogenase; Results from solvent magnetic relaxation studies. 1. Copper(II) and cobalt(II) ions." *Biochemistry* **20**: 3424-3432.

Ando, K. (1997). "Solvation dynamics and electronic structure development of coumarin 120 in methanol: A theoretical modeling study." *Journal of Chemical Physics* **107**(12): 4585-4596.

Antes, I. and Thiel, W. (1999). "Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods." *Journal of Physical Chemistry A* **103**(46): 9290-9295.

Atkins, P. W. (1994). "Physical Chemistry". Oxford, Oxford University Press.

Atkins, P. W. and Friedman, R. S. (1997). "Molecular Quantum Mechanics". New York, Oxford University Press.

Bacon, A. D. and Zerner, M. C. (1979). "An intermediate neglect of differential overlap theory for transition metal complexes: iron, cobalt and copper chlorides." *Theoretica Chimica Acta* **53**(1): 21-54.

Bakowies, D. and Thiel, W. (1996a). "Hybrid models for combined quantum mechanical and molecular mechanical approaches." *Journal of Physical Chemistry* **100**(25): 10580-10594.

Bakowies, D. and Thiel, W. (1996b). "Semiempirical treatment of electrostatic potentials and partial charges in combined quantum mechanical and

molecular mechanical approaches." *Journal of Computational Chemistry* **17**(1): 87-108.

Balog, E., KisPetik, K., Fidy, J., Kohler, M. and Friedrich, J. (1997). "Interpretation of multiple Q(0,0) bands in the absorption spectrum of Mg-mesoporphyrin embedded in horseradish peroxidase." *Biophysical Journal* **73**(1): 397-405.

Barak, A. and La'adan, O. (1997). Performance of the MOSIX parallel system for a cluster of PC's. "High-Performance Computing and Networking". **1225:** 624-635.

Barak, A. and La'adan, O. (1998). "The MOSIX multicomputer operating system for high performance cluster computing." *Future Generation Computer Systems* **13**(4-5): 361-372.

Bashford, D., Case, D. A., Dalvit, C., Tennant, L. and Wright, P. E. (1993). "Electrostatic Calculations of Side-Chain Pk(a) Values in Myoglobin and Comparison with Nmr Data for Histidines." *Biochemistry* **32**(31): 8045-8056.

Bauer, R., Adolph, H. W., Andersson, I., Danielsen, E., Formicka, G. and Zeppezauer, M. (1991). "Coordination geometry for cadmium in the catalytic zinc site of horse liver alcohol dehydrogenase: studies by PAC spectroscopy." *European Biophysics Journal* **20**: 215-221.

Bauernschmitt, R. and Ahlrichs, R. (1996). "Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory." *Chemical Physics Letters* **256**: 454-464.

Bayly, C. I., Cieplak, P., Cornell, W. D. and Kollman, P. A. (1993). "A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model." *Journal of Physical Chemistry* **97**: 10269-10280.

Becke, A. D. (1988). "Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior." *Physical Review A* **38**(6): 3098-3100.

Becke, A. D. (1993). "A New Mixing of Hartree-Fock and Local Density-Functional Theories." *Journal of Chemical Physics* **98**(2): 1372-1377.

Benner, S. A. (1982). "The Stereoselectivity of Alcohol Dehydrogenases - a Stereochemical Imperative." *Experientia* **38**(5): 633-637.

Benner, S. A., Nambiar, K. P. and Chambers, G. K. (1985). "A Stereochemical Imperative in Dehydrogenases - New Data and Criteria for Evaluating Function-Based Theories in Bioorganic Chemistry." *Journal of the American Chemical Society* **107**(19): 5513-5517.

Berendsen, H. J., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. and Haak, J. R. (1984). "Molecular Dynamics with Coupling to an External Bath." *Journal of Chemical Physics* **81**: 3684-3690.

Berendsen, H. J., Postma, J. P. M., Van Gunsteren, W. F. and Hermans, J. (1969). "Interaction Models for Water in Relation to Protein Hydration." *Nature* **224**: 175-177.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). "The Protein Data Bank." *Nucleic Acids Research* **28**: 235-242.

Bernhard, S. A., Dunn, M. F., Luisi, P. L. and Schack, P. (1970). "Mechanistic studies on equine liver alcohol dehydrogenase. I. The stoichiometry relationship of the coenzyme binding sites to the catalytic sites active in the reduction of aromatic aldehydes in the transient state." *Biochemistry* **9**(1): 185-192.

Bertini, I., Gerber, M., Lanini, G., Maret, W., Rawer, S. and Zeppezauer, M. (1984). "$^{1}$H NMR investigation of the active site of cobalt(II)-substituted liver alcohol dehydrogenase." *Journal of the American Chemical Society* **106**: 1826-1830.

Besler, B. H., Merz, K. M. and Kollman, P. A. (1990). "Atomic Charges Derived from Semiempirical Methods." *Journal of Computational Chemistry* **11**(4): 431-439.

Beveridge, D. L. and McConnell, K. J. (2000). "Nucleic acids: theory and computer simulation, Y2K." *Current Opinion in Structural Biology* **10**(2): 182-196.

Bhattacharya, S., Sukits, S. F., MacLaughlin, K. L. and Lecomte, J. T. J. (1997). "The tautomeric state of histidines in myoglobin." *Biophysical Journal* **73**(6): 3230-3240.

Billeter, S. R., Webb, S. P., Agarwal, P. K., Iordanov, T. and Hammes-Schiffer, S. (2001). "Hydride Transfer in Liver Alcohol Dehydrogenase: Quantum Dynamics, Kinetic Isotope Effects, and Role of Enzyme Motion." *Journal of the American Chemical Society*: Papers ASAP.

Binkley, J. S., Pople, J. A. and Herhe, W. J. (1980). "Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements." *Journal of the American Chemical Society* **102**: 939.

Born, M. and Oppenheimer, R. (1927). "Zur quantentheorie der molekeln." *Annals of Physics* **84**: 457-484.

Boys, S. F. (1950). "A General Method of Calculation for the Stationary States of Any Molecular System." *Proceedings of the Royal Society (London)* **A200**: 542-554.

Brändén, C.-I., Jörnvall, H., Eklund, H. and Furugren, B. (1975). "The Enzymes". New York, Academic Press.

Brillouin, L. (1932). "Free electrons in a crystal lattice: Wave equation and magnetic properties." *Journal de physique et le radium* **3**: 565-581.

Brooks III, C. L., Karplus, M. and Pettitt, B. M. (1988). "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics". New York, John Wiley & Sons.

Brown, R. (1995). "Molecular-Dynamics Modeling of Time-Resolved Fluorescence Shifts in Liquid Solution." *Journal of Chemical Physics* **102**(22): 9059-9068.

Bruehl, M. and Hynes, J. T. (1992). "Dielectric Friction and Solvation Dynamics - a Molecular- Dynamics Study." *Journal of Physical Chemistry* **96**(10): 4068-4074.

Burcl, R., Amos, R. D. and Handy, N. C. (2002). "Study of excited states of furan and pyrrole by time-dependent density functional theory." *Chemical Physics Letters* **355**(1-2): 8-18.

Bursulaya, B. D., Zichi, D. A. and Kim, H. J. (1995). "Role of Solute Electronic Polarizability in Solvation Dynamics." *Journal of Physical Chemistry* **99**(25): 10069-10074.

Burton, N. A., Harrison, M. J., Hart, J. C., Hillier, I. H. and Sheppard, D. W. (1998). "Prediction of the mechanisms of enzyme-catalysed reactions using hybrid quantum mechanical molecular mechanical methods." *Faraday Discussions*(110): 463-475.

Callen, H. B. (1960). "Thermodynamics". New York, Wiley.

Callen, H. B. and Welton, T. A. (1951). "Irreversibility and Generalized Noise." *Physical Review* **83**(1): 34-41.

Callender, R. and Deng, H. (1994). "Nonresonance Raman Difference Spectroscopy - a General Probe of Protein-Structure, Ligand-Binding, Enzymatic Catalysis, and the Structures of Other Biomacromolecules." *Annual Review of Biophysics and Biomolecular Structure* **23**: 215-245.

Carter, E. A. and Hynes, J. T. (1991). "Solvation Dynamics for an Ion-Pair in a Polar-Solvent - Time- Dependent Fluorescence and Photochemical Charge-Transfer." *Journal of Chemical Physics* **94**(9): 5961-5979.

Case, D. A. and Karplus, M. (1979). "Dynamics of ligand binding to heme proteins." *Molecular Biology* **132**: 343-368.

Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J.,

Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. and Kollman, P. A. (1999). AMBER 6. University of California, San Francisco.

Casida, M. E. (1995). "Recent Advances in Density Functional Methods". Singapore, World Scientific.

Casida, M. E. (1996). Recent Developments and Applications of Modern Density Functional Theory. "Theoretical and Computational Chemistry Vol. 4." Seminario, J. M. Amsterdam, Elsevier Science. **4**.

Casida, M. E., Jamorski, C., Casida, K. and Salahub, D. R. (1998). "Molecular excitation energies to high-lying bound states from time-dependent density-functional response theory: Characterization and correction of the time-dependent local density approximation ionization threshold." *Journal of Chemical Physics* **108**(11): 4439-4449.

Catlow, C. R. A., Parker, S. C. and Allen, M. P. (1990). "Computer Modeling of Fluids, Polymers and Solids." Dordrecht, Kluwer.

Champion, P. M. (1976). "Resonance Raman spectra of heme proteins at low temperature." *Journal of the American Chemical Society* **98**: 7114-7115.

Chandler, D. (1987). "Introduction to Modern Statistical Mechanics". New York, Oxford University Press.

Chandler, D., Bader, J. S., Kuharski, R. A., Sprik, M., Klein, M. L. and Impey, R. W. (1988). "Molecular model for aqueous ferrous-ferric electron transfer." *Journal of Chemical Physics* **89**(5): 3248-3257.

Chen, D., Kwok, T. Y., Martin, C., Kee, W. R., Sloan, D. and Callender, R. (1987a). "Classical Raman-Spectroscopic Studies of Nadh and Nad+ Bound to Liver Alcohol-Dehydrogenase by Difference Techniques." *Biochemistry* **26**(15): 4776-4784.

Chen, D. H., Rhee, K. W., Martin, C., Sloan, D., Callender, R. and Yue, K. T. (1987b). "Raman Difference Spectroscopy of Adpr Bound to the Coenzyme

Site of Liver Alcohol-Dehydrogenase." *Biophysical Journal* **51**(2): A311-A311.

Chen, S. P., Hosten, C. M., Vivoni, A., Birke, R. L. and Lombardi, J. R. (2002). "SERS investigation of NAD(+) adsorbed on a silver electrode." *Langmuir* **18**(25): 9888-9900.

Cheng, A., Stanton, R. S., Vincent, J. J., van der Vaart, A., Damodaran, K. V., Dixon, S. L., Hartsough, D. S., Mori, M., Best, S. A., Monard, G., Garcia, M., Van Zant, L. and Merz Jr., K. M. (2002). Roar, The Pennsylvania State University.

Cheng, L.-Y. and Lek, L.-H. (1992). "Inhibition of Alcohol Dehydrogenases by Thiol Compounds." *Federation of European Biochemical Sciences* **300**(3): 251-253.

Cho, B. (2003). Protein Dynamics Measured with Non-Linear Spectroscopy. *Department of Chemistry*. London, Imperial College London.

Cho, H., Ramaswamy, S. and Plapp, B. V. (1997). "Flexibility of liver alcohol dehydrogenase in stereoselective binding of 3-butylthiolane 1-oxides." *Biochemistry* **36**(2): 382-389.

Clake, A. R. and Dafforn, T. R. (1998). in. "Comprehensive biological catalysis". Sinnott, M., Academic Press. **3**.

Clark, T., Chandrasekhar, J., Spitznagel, G. W. and Schleyer, P. V. (1983). "Efficient Diffuse Function-Augmented Basis-Sets for Anion Calculations .3. The 3-21+G Basis Set for 1st-Row Elements, Li- F." *Journal of Computational Chemistry* **4**(3): 294-301.

Clementi, E., Corongiu, G., Jönsson, B. and Romano, S. (1980). "Monte arlo simulations of water clusters around $Zn^{++}$ and a linear $Zn^{++}.CO_2$ complex." *Journal of Chemical Physics* **72**: 260-264.

Colle, R. and Salvetti, O. (1975). "Approximate calculation of the correlation energy for the closed shells." *Theoretica Chimica Acta* **37**(4): 329-334.

Collins, J. B., Schleyer, P. V. R., Binkley, J. S. and Pople, J. A. (1976). "Self-Consistent Molecular Orbital Methods. 17. Geometries and binding energies of second-row molecules. A comparison of three basis sets." *Journal of Chemical Physics* **64**: 5142.

Conway, A. and Koshland Jr, D. E. (1968). "Negative cooperativity in enzyme action. The binding of diphosphopyridine nucleotide to glyceraldehyde 3-phosphate dehydrogenase." *Biochemistry* **7**(11): 4011-4023.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995). "A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules." *Journal of the American Chemical Society* **117**(19): 5179-5197.

Cornell, W. D., Piotr, C., Bayly, C. I. and Kollman, P. A. (1993). "Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of solvation." *Journal of the American Chemical Society* **115**: 9620-9631.

Corwin, D. T., Fikar, R. and Koch, S. A. (1987). "4-Coordinate and 5-Coordinate Cobalt(Ii) Thiolate Complexes - Models for the Catalytic Site of Alcohol-Dehydrogenase." *Inorganic Chemistry* **26**(19): 3079-3080.

Cory, M. G., Kostlmeier, S., Kotzian, M., Rosch, N. and Zerner, M. C. (1994). "An Intermediate Neglect of Differential-Overlap Technique for Actinide Compounds." *Journal of Chemical Physics* **100**(2): 1353-1365.

Cory, M. G., Zerner, M. C., Xu, X. C. and Shulten, K. (1998). "Electronic excitations in aggregates of bacteriochlorophylls." *Journal of Physical Chemistry B* **102**(39): 7640-7650.

Cui, Q., Elstner, M. and Karplus, M. (2002). "A Theoretical Analysis of the Proton and Hydride Transfer in Liver Alcohol Dehydrogenase (LAD)." *Journal of Physical Chemistry B* **106**(10): 2721-2740.

Cui, Q. and Karplus, M. (2000a). "Molecular properties from combined QM/MM methods. 2. Chemical shifts in large molecules." *Journal of Physical Chemistry B* **104**(15): 3721-3743.

Cui, Q. and Karplus, M. (2000b). "Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations." *Journal of Chemical Physics* **112**(3): 1133-1149.

Culberson, J. C., Knapp, P., Roesch, N. and Zerner, M. C. (1987). "An intermediate neglect of differential overlap (INDO) technique for lanthanide complexes: studies on lanthanide halides." *Theoretica Chimica Acta* **71**(1): 29-39.

Cunningham, M. A., Ho, L. L., Nguyen, D. T., Gillilan, R. E. and Bash, P. A. (1997). "Simulation of the enzyme reaction mechanism of malate dehydrogenase." *Biochemistry* **36**(16): 4800-4816.

Curtiss, L. A., Jones, C., Trucks, G. W., Raghavachari, K. and Pople, J. A. (1990). "Gaussian-1 Theory of Molecular-Energies for 2nd-Row Compounds." *Journal of Chemical Physics* **93**(4): 2537-2545.

Darden, T., Perera, L., Li, L. P. and Pedersen, L. (1999). "New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations." *Structure with Folding & Design* **7**(3): R55-R60.

Darden, T., York, D. and Pedersen, L. (1993). "Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems." *Journal of Chemical Physics* **98**(12): 10089-10092.

de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1980). "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants." *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **373**: 27-56.

de Leeuw, S. W., Perram, J. W. and Smith, E. R. (1983). "Simulation of Electrostatic Systems in Periodic Boundary- Conditions .3. Further Theory

and Applications." *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **388**(1794): 177-193.

deBoeij, W. P., Pshenichnikov, M. S. and Wiersma, D. A. (1996). "System-bath correlation function probed by conventional and time-gated stimulated photon echo." *Journal of Physical Chemistry* **100**(29): 11806-11823.

Debolt, S. E. and Kollman, P. A. (1990). "A Theoretical-Examination of Solvatochromism and Solute Solvent Structuring in Simple Alkyl Carbonyl-Compounds - Simulations Using Statistical Mechanical Free-Energy Perturbation-Methods." *Journal of the American Chemical Society* **112**(21): 7515-7524.

Deng, H., Schindler, J. F., Berst, K. B., Plapp, B. V. and Callender, R. (1998). "A Raman spectroscopic characterization of bonding in the complex of horse liver alcohol dehydrogenase with NADH and N- cyclohexylformamide." *Biochemistry* **37**(40): 14267-14278.

Dickinson, L. C. and Symons, M. C. R. (1983). "Electron-Spin Resonance of Hemoglobin and Myoglobin." *Chemical Society Reviews* **12**(4): 387-414.

Ding, H. Q., Karasawa, N. and Goddard, W. A. (1992a). "Atomic Level Simulations on a Million Particles - the Cell Multipole Method for Coulomb and London Nonbond Interactions." *Journal of Chemical Physics* **97**(6): 4309-4315.

Ding, H. Q., Karasawa, N. and Goddard, W. A. (1992b). "The Reduced Cell Multipole Method for Coulomb Interactions in Periodic-Systems with Million-Atom Unit Cells." *Chemical Physics Letters* **196**(1-2): 6-10.

Duan, Y. and Kollman, P. A. (1998). "Pathways to a protein folding intermediate observed in a 1- microsecond simulation in aqueous solution." *Science* **282**(5389): 740-744.

Dunning Jr., T. H. (1970). "Gaussian basis functions for use in molecular calculations. I. Contraction of (9s5p) atomic basis sets for the first-row atoms." *Journal of Chemical Physics* **53**: 2823-2833.

Dunning Jr., T. H. and Hay, P. J. (1976). in. "Modern Theoretical Chemistry". Schaefer III, H. F. New York, Plenum. **3:** 1.

Dworschack, R. T. and Plapp, B. V. (1977). "pH, isotope, and substituent effects on the interconversion of aromatic substrates catalyzed by hydroxybutyrimidylated liver alcohol dehydrogenase." *Biochemistry* **16**: 2716-2725.

Ehrig, T., Hurley, T. D., Edenberg, H. J. and Bosron, W. F. (1991). "General Base Catalysis in a Glutamine for Histidine Mutant at Position-51 of Human Liver Alcohol-Dehydrogenase." *Biochemistry* **30**(4): 1062-1068.

Eklund, H. and Brändén, C. I. (1987). in. "Pyridine nucleotide coenzymes, Part A." Dolphin, D. and Poulson, R., John Wiley**:** 253.

Eklund, H., Brändén, C. I. and Jornvall, H. (1976a). "Structural comparisons of mammalian, yeast and bacillar alcohol dehydrogenases." *Journal of Molecular Biology* **102**: 61-73.

Eklund, H., Nordstrom, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I. and Akeson, A. (1976b). "Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 A resolution." *Journal of Molecular Biology* **102**: 27-59.

Eklund, H., Plapp, B. V., Samama, J. P. and Branden, C. I. (1982). "Binding of Substrate in a Ternary Complex of Horse Liver Alcohol-Dehydrogenase." *Journal of Biological Chemistry* **257**(23): 4349-4358.

Eklund, H., Samama, J. P., Wallen, L., Branden, C. I., Akeson, A. and Jones, T. A. (1981). "Structure of a Triclinic Ternary Complex of Horse Liver Alcohol-Dehydrogenase at 2.9 a Resolution." *Journal of Molecular Biology* **146**(4): 561-587.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. (1995). "A Smooth Particle Mesh Ewald Method." *Journal of Chemical Physics* **103**(19): 8577-8593.

Ewald, P. P. (1921). "Die Berechnung optischer und elektrostatischer Gitterpotentiale." *Annalen Der Physik* **64**: 253-286.

Fan, H. and Mark, A. E. (2003). "Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: A molecular dynamics simulation study." *Proteins-Structure Function and Genetics* **53**(1): 111-120.

Ferro, D. R. and Hermans, J. (1977). "A Different Best Rigid-body Molecular Fit Routine,." *Acta. Crystallographica* **A33**: 345-347.

Fersht, A. (1999). "Structure and Mechanism in Protein Science", W. H. Freemand and Company.

Feynman, R. P. (1987). "Elementary Particles and the Laws of Physics: The 1986 Dirac memorial lectures". Cambridge, Cambridge University Press.

Field, M. J., Bash, P. A. and Karplus, M. (1990). "A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations." *Journal of Computational Chemistry* **11**(6): 700-733.

Fisher, H. F., Conn, E. E., Vennesland, B. and Westheimer, F. H. (1953). "The enzymic transfer of hydrogen. I. The reaction catalysed by alcohol dehydrogenase." *Journal of Biological Chemistry* **202**: 687-697.

Fleming, G. R. and Cho, M. H. (1996). "Chromophore-solvent dynamics." *Annual Review of Physical Chemistry* **47**: 109-134.

Fonseca, T. and Ladanyi, B. M. (1991). "Breakdown of Linear Response for Solvation Dynamics in Methanol." *Journal of Physical Chemistry* **95**(6): 2116-2119.

Foresman, J. B., Headgordon, M., Pople, J. A. and Frisch, M. J. (1992). "Toward a Systematic Molecular-Orbital Theory for Excited-States." *Journal of Physical Chemistry* **96**(1): 135-149.

Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., Defrees, D. J. and Pople, J. A. (1982). "Self-Consistent Molecular-Orbital Methods .23. A Polarization- Type Basis Set for 2nd-Row Elements." *Journal of Chemical Physics* **77**(7): 3654-3665.

Frauenfelder, H., Sligar, S. G. and Wolynes, P. G. (1991). "The Energy Landscapes and Motions of Proteins." *Science* **254**(5038): 1598-1603.

Frenkel, D. and Smit, B. (1996). "Understanding Molecular Simulation: From Algorithms to Applications". London, Academic Press Limited.

Friedman, H. (1975). "Image approximation to the reaction field." *Molecular Physics* **29**(5).

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery Jr., J. A., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Head-Gordon, M., Replogle, E. S. and Pople, J. A. (1998). Gaussian 98 (Revision A. 7), Gaussian, Inc., Pittsburgh P A.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery Jr., J. A., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Salvador, P., Dannenberg, J. J., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox,

D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S. and Pople, J. A. (2001). Gaussian 98 (Revision A. 11), Gaussian, Inc., Pittsburgh P A.

Gafni, A. and Brand, L. (1976). "Fluorescence Decay Studies of Reduced Nicotinamide Adenine Dinucleotide in Solution and Bound to Liver Alcohol Dehydrogenase." *Biochemistry* **15**(15): 3165-3171.

Gao, J. L., Amara, P., Alhambra, C. and Field, M. J. (1998). "A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations." *Journal of Physical Chemistry A* **102**(24): 4714-4721.

Gao, J. L. and Truhlar, D. G. (2002). "Quantum mechanical methods for enzyme kinetics." *Annual Review of Physical Chemistry* **53**: 467-505.

Gear, C. W. (1971). "Numerical Initial Value Problems in Ordinary Differential Equations." Englewood Cliffs, NJ, Prentice Hall.

Gennis, L. S. (1976). "Negative homotropic cooperativity and affinity heterogeneiry: preparation of yeast glyceraldehyde-3-phosphate dehydrogenase with maximal affinity homogeneity." *Proceedings of the National Academy of Science USA* **73**(11): 3928-3932.

Giammona, D. A. (1984). PhD. Thesis. *University of California, Davis.*

Gilbert, A. and Baggott, J. (1991). "Essentials of Molecular Photochemistry". Oxford, Blackwell Scientific Publications.

Gordon, M. S., Binkley, J. S., Pople, J. A., Pietro, W. J. and Hehre, W. J. (1982). "Self-Consistent Molecular-Orbital Methods .22. Small Split- Valence Basis-Sets for 2nd-Row Elements." *Journal of the American Chemical Society* **104**(10): 2797-2803.

Grant, G. H. and Richards, W. G. (1995). "Computational Chemistry". Oxford, Oxford University Press.

Gronenborn, A. M. and Clore, G. M. (1982). "Conformation of Nad+ Bound to Yeast and Horse Liver Alcohol- Dehydrogenase in Solution - the Use of the Proton-Proton Transferred Nuclear Overhauser Enhancement." *Journal of Molecular Biology* **157**(1): 155-160.

Gross, E. K. U., Dobson, J. F. and Petersilka, M. (1996). Density Functional Theory II. "Topics in Current Chemistry". Nalewajski, R. F. Berlin, Springer. **181**.

Gross, E. K. U. and Kohn, W. (1990). "Density Functional Theory of Many-Fermion Systems." *Advances in Quantum Chemistry* **21**: 255.

Gryczynski, Z., Lubkowski, J. and Bucci, E. (1995). "Heme-Protein Interactions in Horse Heart Myoglobin at Neutral Ph and Exposed to Acid Investigated by Time-Resolved Fluorescence in the Picosecond to Nanosecond Time Range." *Journal of Biological Chemistry* **270**(33): 19232-19237.

Gryczynski, Z., Lubkowski, J. and Bucci, E. (1997). Intrinsic fluorescence of hemoglobins and myoglobins. "Fluorescence Spectroscopy". **278:** 538-569.

Gwaltney, S. R. and Bartlett, R. J. (1998). "Coupled-cluster calculations of the electronic excitation spectrum of free base porphin in a polarized basis." *Journal of Chemical Physics* **108**(16): 6790-6798.

Hadad, C. M., Foresman, J. B. and Wiberg, K. B. (1993). "Excited-States of Carbonyl-Compounds .1. Formaldehyde and Acetaldehyde." *Journal of Physical Chemistry* **97**(17): 4293-4312.

Hadorn, M., John, V. A., Meier, F. K. and Dutler, H. (1975). "Kinetic equivalence of the active sites of alcohol dehydrogenase from horse liver." *European Journal of Biochemistry* **54**(1): 65-73.

Hall, G. G. (1951). "The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials." *Proceedings of the Royal Society (London)* **A205**: 541-552.

Hall, G. G. and Smith, C. M. (1984). "Fitting Electron-Densities of Molecules." *International Journal of Quantum Chemistry* **25**(5): 881-890.

Hariharan, P. C. and Pople, J. A. (1973). "The influence of polarisation functions on molecular orbital hydrogenation energies." *Theoretica Chimica Acta.* **28**: 213-222.

Harrison, M. J., Burton, N. A. and Hillier, I. H. (1997). "Catalytic mechanism of the enzyme papain: Predictions with a hybrid quantum mechanical/molecular mechanical potential." *Journal of the American Chemical Society* **119**: 12285-12291.

Harrison, R. W. (1999). "Integrating quantum and molecular mechanics." *Journal of Computational Chemistry* **20**(15): 1618-1633.

Hartree, D. R. (1928). "The Wave Mechanics of an Atom with a Non-Coulomb Central Field." *Proceedings of the Cambridge Philosophical Society* **24**: 89-132.

Haykin, S. (1995). "Adaptive Filter Theory", Prentice Hall.

Hehre, W. J., Ditchfield, R. and Pople, J. A. (1972). "Self-consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for use in Molecular Orbital Studies of Organic Molecules." *Journal of Chemical Physics* **56**: 2257-2261.

Hehre, W. J., Stewart, R. F. and Pople, J. A. (1969). "Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-type Atomic Orbitals." *Journal of Chemical Physics* **51**: 2657-2664.

Hellwig, P., Scheide, D., Bungert, S., Mantele, W. and Friedrich, T. (2000). "FT-IR spectroscopic characterization of NADH : ubiquinone oxidoreductase (complex I) from Escherichia coli: Oxidation of FeS cluster N2 is coupled with the protonation of an aspartate or glutamate side chains." *Biochemistry* **39**(35): 10884-10891.

Hermans, J., Berendsen, H. J., Van Gunsteren, W. F. and Postma, J. P. M. (1984). "A Consistent Empirical Potential for Water-Protein Interactions." *Biopolymers* **23**: 1.

Heyes, D. M. (1981). "Electrostatic Potentials and Fields in Infinite Point-Charge Lattices." *Journal of Chemical Physics* **74**(3): 1924-1929.

Hill, T. L. (1948). "Steric Effects. I. Van der Waals Potential Energy Curves." *Journal of Chemical Physics* **16**: 399-404.

Hillier, I. H. (1999). "Chemical reactivity studied by hybrid QM/MM methods." *Journal of Molecular Structure-Theochem* **463**(1-2): 45-52.

Hirata, S., Head-Gordon, M., Szczepanski, J. and Vala, M. (2003). "Time-dependent density functional study of the electronic excited states of polycyclic aromatic hydrocarbon radical ions." *Journal of Physical Chemistry A* **107**(24): 4940-4951.

Hockney, R. (1970). "The Potential Calculation and Some Applications." *Methods in Computational Physics* **9**(136-211).

Hohenberg, P. and Kohn, W. (1964). "Inhomogeneous Electron Gas." *Physical Review B* **136**: 864-871.

Hoops, S. C., Anderson, K. W. and Merz, K. M. (1991). "Force-Field Design for Metalloproteins." *Journal of the American Chemical Society* **113**(22): 8262-8270.

Hughes, S. J., Tanner, J. A., Hindley, A. D., Miller, A. D. and Gould, I. R. (2003). "Functional asymmetry in the lysyl-tRNA synthetase explored by molecular dynamics, free energy calculations and experiment." *BMC Structural Biology* **3**(5).

Humphrey, W., Dalke, A. and Schulten, K. (1996). "VMD - Visual Molecular Dynamics." *Journal of Molecular Graphics* **14**(1): 33-38.

Hwang, M. J., Stockfisch, T. P. and Hagler, A. T. (1994). "Derivation of Class-II Force-Fields .2. Derivation and Characterization of a Class-II Force-Field, Cff93, for the Alkyl Functional-Group and Alkane Molecules." *Journal of the American Chemical Society* **116**(6): 2515-2525.

Jagodzinski, P. W., Funk, G. F. and Peticolas, W. L. (1982). "Resonance-Enhanced Raman Identification of a Ternary Chemical Intermediate During the Equine Liver Alcohol-Dehydrogenase Reduction of Para-(Dimethylamino)Benzaldehyde." *Biochemistry* **21**(9): 2193-2202.

Jagodzinski, P. W. and Peticolas, W. L. (1981). "Resonance Enhanced Raman Identification of the Zinc-Oxygen Bond in a Horse Liver Alcohol Dehydrogenase-Nicotinamide Adenine Dinucleotide-Aldehyde Transient-Chemical Intermediate." *Journal of the American Chemical Society* **103**(1): 234-236.

Jamorski, C., Casida, M. E. and Salahub, D. R. (1996). "Dynamic polarizabilities and excitation spectra from a molecular implementation of time-dependent density-functional response theory: N-2 as a case study." *Journal of Chemical Physics* **104**(13): 5134-5147.

Jensen, F. (1999). "Introduction to Computational Chemistry". New York, John Wiley and Sons.

Jimenez, R., Fleming, G. R., Kumar, P. V. and Maroncelli, M. (1994). "Femtosecond Solvation Dynamics of Water." *Nature* **369**(6480): 471-473.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. (1983). "Comparison of Simple Potential Functions for Simulating Liquid Water." *Journal of Chemical Physics* **79**: 926-935.

Kabsch, W. (1978). "A Discussion of the Solution for the Best Rotation to Relate Two Sets of Vectors." *Acta. Crystallographica* **A34**: 827-828.

Kendrew, J. C. (1960). "Structure of myoglobin: A three-dimensional synthesis at 2 A resolution." *Nature* **185**: 422-427.

Kendrew, J. C. (1961). "The three-dimensional structure of a protein molecule." *Scientific American* **205**: 96-110.

Keren, A. and Barak, A. (2003). "Opportunity cost algorithms for reduction of I/O and interprocess communication overhead in a computing cluster." *Ieee Transactions on Parallel and Distributed Systems* **14**(1): 39-50.

Khidekel, V., Chernyak, V. and Mukamel, S. (1996). "Interplay of multiple vibrational spectral densities in femtosecond nonlinear spectroscopy of liquids." *Journal of Chemical Physics* **105**(19): 8543-8555.

Kierdaszuk, B., Malak, H., Gryczynski, I., Callis, P. and Lakowicz, J. R. (1996). "Fluorescence of reduced nicotinamides using one- and two-photon excitation." *Biophysical Chemistry* **62**(1-3): 1-13.

Kohn, W. and Sham, L. J. (1965). "Self-Consistent Equations Including Exchange and Correlation Effects." *Physical Review A* **140**: 1133-1138.

Kollman, P. A., Kuhn, B. and Perakyla, M. (2002). "Computational studies of enzyme-catalyzed reactions: Where are we in predicting mechanisms and in understanding the nature of enzyme catalysis?" *Journal of Physical Chemistry B* **106**(7): 1537-1542.

Kotzian, M., Roesch, N. and Zerner, M. C. (1991). "An (INDO/S)-Cl treatment including spin-orbit interaction based on Rumer spin functions. Application to the hydrated cerium ion." *International Journal of Quantum Chemistry* **25**: 545-555.

Kotzian, M., Rosch, N. and Zerner, M. C. (1992). "Intermediate Neglect of Differential-Overlap Spectroscopic Studies on Lanthanide Complexes .1. Spectroscopic Parametrization and Application to Diatomic Lanthanide Oxides Lno (Ln = La, Ce, Gd, and Lu)." *Theoretica Chimica Acta* **81**(4-5): 201-222.

Kraulis, P. J. (1991). "Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures." *Journal of Applied Crystallography* **24**: 946-950.

Krieger, F., Fierz, B., Bieri, O., Drewello, M. and Kiefhaber, T. (2003). "Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding." *Journal of Molecular Biology* **332**(1): 265-274.

Krishnan, R., Binkley, J. S., Seeger, R. and Pople, J. A. (1980). "Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions." *Journal of Chemical Physics* **72**: 650-654.

Kubo, R. (1962). "Generalised cumulant expansion method." *Journal of the Physical Society, Japan* **17**: 1100-1120.

Kubo, R. (1966). "Fluctuation-dissipation theorem." *Reports on Progress in Physics* **29**(1): 255-284.

Kumar, P. V. and Maroncelli, M. (1995). "Polar Solvation Dynamics of Polyatomic Solutes - Simulation Studies in Acetonitrile and Methanol." *Journal of Chemical Physics* **103**(8): 3038-3060.

Kvassman, J. and Pettersson, G. (1976). "Kinteic transients in the reduction of aldehydes catalysed by liver alcohol dehydrogenase." *European Journal of Biochemistry* **69**(1): 279-287.

Langhoff, S. R. and Davidson, E. R. (1974). "Configuration interaction calculations on the Nitrogen molecule." *International Journal of Quantum Chemistry* **8**(61-72).

Leach, A. R. (1996). "Molecular Modelling: Principles and Applications". Singapore, Longman Singapore Publishers Ltd.

Leach, A. R. (2001). "Molecular Modelling: Principles and Applications". Harlow, Pearson Education Limited.

Lecomte, J. T. J. and Lamar, G. N. (1985). "H-1-Nmr Study of Labile Proton-Exchange in the Heme Cavity as a Probe for the Potential Ligand Entry Channel in Myoglobin." *Biochemistry* **24**(25): 7388-7395.

Lee, C. T., Yang, W. T. and Parr, R. G. (1988). "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density." *Physical Review B* **37**(2): 785-789.

Levine, I. N. (1991). "Quantum Chemistry". Englewood Cliffs, Prentice Hall.

Levy, R. M., Kitchen, D. B., Blair, J. T. and Kroghjespersen, K. (1990). "Molecular-Dynamics Simulation of Time-Resolved Fluorescence and Nonequilibrium Solvation of Formaldehyde in Water." *Journal of Physical Chemistry* **94**(11): 4470-4476.

Li, D. W. and London, R. E. (2002). "Ligand discovery using the inter-ligand Overhauser effect: horse liver alcohol dehydrogenase." *Biotechnology Letters* **24**(8): 623-629.

Lide, D. R. (1996). "CRC Handbook of Chemistry and Physics". London, CRC Press.

Loew, G. (2000). "Structure, spectra, and function of heme sites." *International Journal of Quantum Chemistry* **77**(1): 54-70.

London, F. (1930). "Zur Theori und Systematik der Molekularkräfte." *Zeitschrift für Physik* **63**: 245-279.

Löwdin, P. O. (1959). "A Classic Review on Electron Correlation." *Advances in Chemical Physics* **2**: 207.

Lukin, J. A., Simplaceanu, V., Zou, M., Ho, N. T. and Ho, C. (2000). "NMR reveals hydrogen bonds between oxygen and distal histidines in oxyhemoglobin." *Proceedings of the National Academy of Sciences of the United States of America* **97**(19): 10354-10358.

Lyne, P. D., Hodoscek, M. and Karplus, M. (1999). "A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region." *Journal of Physical Chemistry A* **103**(18): 3462-3471.

Mackerell, A. D., Brooks III, C. L., Nilsson, L., B., R., Won, Y. and Karplus, M. (1998). "CHARMM: The Energy Function and Its Parameterization with an

Overview of the Program, in The Encyclopedia of Computational Chemistry, 1, 271-277". Chichester, John Wiley & Sons.

Makinen, M. W. and Yim, M. B. (1981). "Coordination environment of the active-site metal ion of liver alcohol dehydrogenase." *Proceedings of the National Academy of Science USA* **78**: 6221-6225.

Makowska-Grzyska, M. M., Jeppson, P. C., Allred, R. A., Arif, A. M. and Berreau, L. M. (2002). "Modeling substrate- and inhibitor-bound forms of liver alcohol dehydrogenase: Chemistry of mononuclear nitrogen/sulfur-ligated zinc alcohol, formamide, and sulfoxide complexes." *Inorganic Chemistry* **41**(19): 4872-4887.

Maret, W. and Zeppezauer, M. (1986). "Influence of anions and pH on the conformational change of horse liver alcohol dehydrogenase induced by binding of oxidized nicotinamide adenine dinucleotide: Binding of chloride to the catalytic metal ion." *Biochemistry* **25**: 1584-1588.

Maret, W., Zeppezauer, M., Sanders-Loehr, J. and Loehr, T. M. (1983). "Resonance Raman spectra of copper(II)-substituted liver alcohol dehydrogenase: A type 1 copper analogue." *Biochemistry* **22**: 3202-3206.

Maroncelli, M. and Fleming, G. R. (1988). "Comparison of Time-Resolved Fluorescence Stokes Shift Measurements to a Molecular Theory of Solvation Dynamics." *Journal of Chemical Physics* **89**(2): 875-881.

Marrow, J. D. (2002). ClumpOS R6.0 http://clumpos.psoftware.net.

Martin, C., Chen, D. H., Rhee, K. W., Sloan, D., Callender, R., Yue, K. T., Vandersteen, R. and Lugtenburg, J. (1987). "Resonance Raman-Study of the Substrate Daba at the Active-Site of Liver Alcohol-Dehydrogenase." *Biophysical Journal* **51**(2): A311-A311.

Mataga, N., Shibata, Y., Chosrowjan, H., Yoshida, N. and Osuka, A. (2000). "Internal conversion and vibronic relaxation from higher excited electronic state of porphyrins: Femtosecond fluorescence dynamics studies." *Journal of Physical Chemistry B* **104**(17): 4001-4004.

Maurus, R., Overall, C. M., Bogumil, R., Luo, Y., Mauk, A. G., Smith, M. and Brayer, G. D. (1997). "A myoglobin variant with a polar substitution in a conserved hydrophobic cluster in the heme binding pocket." *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1341**(1): 1-13.

McWeeny, R. (1992). "Methods of Molecular Quantum Mechanics". London, Academic Press Limited.

Meijers, R., Morris, R. J., Adolph, H. W., Merli, A., Lamzin, V. S. and Cedergren-Zeppezauer, E. S. (2001). "On the enzymatic activation of NADH." *Journal of Biological Chemistry* **276**(12): 9316-9321.

Mercer, I. P., Gould, I. R. and Klug, D. R. (1997). "Optical properties of solvated molecules calculated by a QMMM method - Chlorophyll a and bacteriochlorophyll a." *Faraday Discussions*(108): 51-62.

Mercer, I. P., Gould, I. R. and Klug, D. R. (1999). "A quantum mechanical/molecular mechanical approach to relaxation dynamics: Calculation of the optical properties of solvated bacteriochlorophyll-a." *Journal of Physical Chemistry B* **103**(36): 7720-7727.

Merritt, E. A. and Murphy, M. E. P. (1994). "Raster3d Version-2.0 - a Program for Photorealistic Molecular Graphics." *Acta Crystallographica Section D-Biological Crystallography* **50**: 869-873.

Meyer, M. and Pontikis, V. (1991). "Computer Simulation in Material Science". Dordrecht, Kluwer.

Møller, C. and Plesset, M. S. (1934). "Note on the approximation treatment for many-electron systems." *Physical Review* **46**: 618-622.

Monard, G. and Merz, K. M. (1999). "Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems." *Accounts of Chemical Research* **32**(10): 904-911.

Monard, G., Prat-Resina, X., Gonzalez-Lafont, A. and Lluch, J. M. (2003). "Determination of enzymatic reaction pathways using QM/MM methods." *International Journal of Quantum Chemistry* **93**(3): 229-244.

Morse, P. M. (1929). "Diatomic molecules according to wave dynamics. II. Vibrational levels." *Physical Review* **34**: 57-64.

Muino, P. L. and Callis, P. R. (1994). "Hybrid Simulations of Solvation Effects on Electronic-Spectra - Indoles in Water." *Journal of Chemical Physics* **100**(6): 4093-4109.

Mukamel, S. (1985). "Fluorescence and Absorption of Large Anharmonic Molecules - Spectroscopy without Eigenstates." *Journal of Physical Chemistry* **89**(7): 1077-1087.

Mukamel, S. (1995). "Principles of Nonlinear Optical Spectroscopy". New York, Oxford University Press.

Mukamel, S. (2003). "Personal Correspondence."

Murakami, H. and Kushida, T. (1994). "Fluorescence Properties of Zn-Substituted Myoglobin." *Journal of Luminescence* **58**(1-6): 172-175.

Murphy, R. B., Philipp, D. M. and Friesner, R. A. (2000a). "Frozen orbital QM/MM methods for density functional theory." *Chemical Physics Letters* **321**(1-2): 113-120.

Murphy, R. B., Philipp, D. M. and Friesner, R. A. (2000b). "A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments." *Journal of Computational Chemistry* **21**(16): 1442-1457.

Nadolny, C. and Zundel, G. (1996). "Protonation, conformation and hydrogen bonding of nicotinamide adenine dinucleotide - An FT-IR study." *Journal of Molecular Structure* **385**(2): 81-87.

Nadolny, C. and Zundel, G. (1997). "Fourier transform infrared spectroscopic studies of proton transfer processes and the dissociation of Zn2+-bound water in alcohol dehydrogenases." *European Journal of Biochemistry* **247**(3): 914-919.

Ober, C., Burkardt, M., Winkler, H., Trautwein, A. X., Zharikov, A. A., Fischer, S. F. and Parak, F. (1997). "Low temperature study of myoglobin ligand rebinding kinetics with Mossbauer spectroscopy." *European Biophysics Journal with Biophysics Letters* **26**(3): 227-237.

Ohno, A. and Ushio, K. (1987). in. "Pyridine nucleotide coenzymes, Part B." Dolphin, D. and Poulson, R., John Wiley**:** 275.

Oliva, M., Dideberg, O. and Field, M. J. (2003). "Understanding the acylation mechanisms of active-site serine penicillin-recognizing proteins: A molecular dynamics simulation study." *Proteins-Structure Function and Genetics* **53**(1): 88-100.

Onsager, L. (1931a). "Reciprocal relations in irreversible processes. I." *Physical Review* **37**: 405-426.

Onsager, L. (1931b). "Reciprocal relations in irreversible processes. II." *Physical Review* **38**: 2265-2279.

Oppenheimer, N. J. (1984). "Stereoselectivity of Enzymatic Transfer of Hydrogen from Nicotinamide Coenzymes - a Stereochemical Imperative." *Journal of the American Chemical Society* **106**(10): 3032-3033.

Passino, S. A., Nagasawa, Y., Joo, T. and Fleming, G. R. (1997). "Three-pulse echo peak shift studies of polar solvation dynamics." *Journal of Physical Chemistry A* **101**(4): 725-731.

Pavelites, J. J., Gao, J. L., Bash, P. A. and Mackerell, A. D. (1997). "A molecular mechanics force field for NAD(+), NADH, and the pyrophosphate groups of nucleotides." *Journal of Computational Chemistry* **18**(2): 221-239.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G. and Kollman, P. (1995). "Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free- Energy Calculations to Simulate the Structural and Energetic Properties of Molecules." *Computer Physics Communications* **91**(1-3): 1-41.

Petersilka, M., Gossmann, U. J. and Gross, E. K. U. (1996). "Excitation energies from time-dependent density-functional theory." *Physical Review Letters* **76**(8): 1212-1215.

Petersson, G. A. and Al-Laham, M. A. (1991). "A complete basis set model chemistry. II. Open-shell systems and the total energies of the first row atoms." *Journal of Chemical Physics* **94**: 6081.

Petersson, G. A., Bennett, A., Tensfeldt, T. G., Al-Laham, M. A., Shirley, W. A. and Mantzaris, J. (1988). "A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row atoms." *Journal of Chemical Physics* **89**: 2193.

Phillips, G. N. and Pettitt, B. M. (1995). "Structure and Dynamics of the Water around Myoglobin." *Protein Science* **4**(2): 149-158.

Piersma, S. R., Visser, A., de Vries, S. and Duine, J. A. (1998). "Optical spectroscopy of nicotinoprotein alcohol dehydrogenase from Amycolatopsis methanolica: A comparison with horse liver alcohol dehydrogenase and UDP-galactose epimerase." *Biochemistry* **37**(9): 3068-3077.

Pietro, W. J., Francl, M. M., Hehre, W. J., DeFrees, D. J., Pople, J. A. and Binkley, J. S. (1982). "Self-Consistent Molecular Orbital Methods. 24. Supplemented Small Split Valence Basis Sets for Second Row Elements." *Journal of the American Chemical Society* **104**: 5039-5048.

Plapp, B. V., Eklund, H., Jones, T. A. and Branden, C. I. (1983). "3-Dimensional Structure of Isonicotinimidylated Liver Alcohol- Dehydrogenase." *Journal of Biological Chemistry* **258**(9): 5537-5547.

Pocker, Y., Page, J. D., Li, H. and Bhat, C. C. (2001). "Ternary complexes of liver alcohol dehydrogenase." *Chemico-Biological Interactions* **130**(1-3): 371-381.

Popják, G. (1970). "The Enzymes".

Pople, J. A., Beveridge, D. L. and Dobosh, P. (1967). "Approximate self-consistent molecular orbital theory. V. Intermediate neglect of differential overlap." *Journal of Chemical Physics* **47**: 2026-2033.

Pople, J. A., Headgordon, M., Fox, D. J., Raghavachari, K. and Curtiss, L. A. (1989). "Gaussian-1 Theory - a General Procedure for Prediction of Molecular-Energies." *Journal of Chemical Physics* **90**(10): 5622-5629.

Pople, J. A. and Segal, G. A. (1965). "Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap." *Journal of Chemical Physics* **43**(S136-S149).

Postnikova, G. B. (1999). "Fluorescence study of conformational transitions in the structure of myoglobin." *Biochemistry-Moscow* **64**(3): 267-286.

Postnikova, G. B., Komarov, Y. E. and Yumakova, E. M. (1991). "Fluorescence Study of the Conformational Properties of Myoglobin Structure .1. Ph-Dependent Changes of Tryptophanyl Fluorescence in Intact and Chemically Modified Sperm Whale Apomyoglobins." *European Journal of Biochemistry* **198**(1): 223-232.

Prabhu, N. V., Dalosto, S. D., Sharp, K. A., Wright, W. W. and Vanderkooi, J. M. (2002). "Optical spectra of Fe(II) cytochrome c interpreted using molecular dynamics simulations and quantum mechanical calculations." *Journal of Physical Chemistry B* **106**(21): 5561-5571.

Pullman, M. E., San Pietro, A. and Colowick, S. P. (1954). "The structure of diphosphopyridine nucleotide." *Journal of Biological Chemistry* **206**: 129-141.

Ramaswamy, S., Eklund, H. and Plapp, B. V. (1994). Structures of Horse Liver Alcohol-Dehydrogenase Complexed with Nad(+) and Substituted Benzyl Alcohols. *Biochemistry*. **33:** 5230-5237.

Ramaswamy, S., Park, D. H. and Plapp, B. V. (1999). "Substitutions in a flexible loop of horse liver alcohol dehydrogenase hinder the conformational change and unmask hydrogen transfer." *Biochemistry* **38**(42): 13951-13959.

Ramaswamy, S., Scholze, M. and Plapp, B. V. (1997). "Binding of formamides to liver alcohol dehydrogenase." *Biochemistry* **36**(12): 3522-3527.

Ranganathan, S. and Gready, J. E. (1997). "Hybrid quantum and molecular mechanical (QM/MM) studies on the pyruvate to L-lactate interconversion in L-lactate dehydrogenase." *Journal of Physical Chemistry B* **101**(28): 5614-5618.

Reisbig, R. R. and Woody, R. W. (1978). "Characterization of a Long-Wavelength Feature in the Absorption and Circular Dichroism Spectra of B-Nicotinamide Adenine Dinucleotide. Evidence for a Charge Transfer Transition." *Biochemistry* **17**(10): 1974-1984.

Ridley, J. E. and Zerner, M. C. (1973). "Intermediate neglect of differential overlap (INDO) technique for spectroscopy. Pyrrole and the azines." *Theoretica Chimica Acta* **32**(2): 111-134.

Ridley, J. E. and Zerner, M. C. (1976). "Triplet states via intermediate neglect of differential overlap: benzene, pyridine and the diazines." *Theoretica Chimica Acta* **42**(3): 223-236.

Roothaan, C. C. J. (1951). "New Developments in Molecular Orbital Theory." *Reviews of Modern Physics* **23**: 69-89.

Rossky, P. J. and Simon, J. D. (1994). "Dynamics of Chemical Processes in Polar-Solvents." *Nature* **370**(6487): 263-269.

Rubach, J. K. and Plapp, B. V. (2002). "Mobility of fluorobenzyl alcohols bound to liver alcohol dehydrogenases as determined by NMR and X-ray crystallographic studies." *Biochemistry* **41**(52): 15770-15779.

Rubach, J. K. and Plapp, B. V. (2003). "Amino acid residues in the nicotinamide binding site contribute to catalysis by horse liver alcohol dehydrogenase." *Biochemistry* **42**(10): 2907-2915.

Ryckaert, J. P., Cicotti, G. and Berendsen, H. J. (1977). "Numerical Integration of the Cartesian Equations of Motion in a System with Constraints: Molecular Dynamics of n-Alkanes." *Journal of Computational Physics* **23**(327-341).

Ryde, U. (1994). "The Coordination Chemistry of the Catalytic Zinc Ion in Alcohol-Dehydrogenase Studies by Ab-Initio Quantum-Chemical Calculations." *International Journal of Quantum Chemistry* **52**(5): 1229-1243.

Ryde, U. (1995). "Molecular-Dynamics Simulations of Alcohol-Dehydrogenase with a 4-Coordinate or 5-Coordinate Catalytic Zinc Ion." *Proteins-Structure Function and Genetics* **21**(1): 40-56.

Ryde, U. (1996a). "The coordination chemistry of the structural zinc ion in alcohol dehydrogenase studied by ab initio quantum chemical calculations." *European Biophysics Journal with Biophysics Letters* **24**(4): 213-221.

Ryde, U. (1996b). "The coordination of the catalytic zinc ion in alcohol dehydrogenase studied by combined quantum-chemical and molecular mechanics calculations." *Journal of Computer-Aided Molecular Design* **10**(2): 153-164.

Sadlej, A. J. (1988). "Medium-Size Polarized Basis-Sets for High-Level Correlated Calculations of Molecular Electric Properties." *Collection of Czechoslovak Chemical Communications* **53**(9): 1995-2016.

Sadlej, A. J. (1991a). "Medium-Size Polarized Basis-Sets for High-Level-Correlated Calculations of Molecular Electric Properties .2. 2nd-Row Atoms - Si through Cl." *Theoretica Chimica Acta* **79**(2): 123-140.

Sadlej, A. J. (1991b). "Medium-Size Polarized Basis-Sets for High-Level-Correlated Calculations of Molecular Electric Properties .4. 3rd-Row Atoms - Ge through Br." *Theoretica Chimica Acta* **81**(1-2): 45-63.

Sadlej, A. J. (1992). "Medium-Size Polarized Basis-Sets for High-Level-Correlated Calculations of Molecular Electric Properties .5. 4th-Row Atoms - Sn through I." *Theoretica Chimica Acta* **81**(4-5): 339-354.

Sadlej, A. J. and Urban, M. (1991). "Medium-Size Polarized Basis-Sets for High-Level-Correlated Calculations of Molecular Electric Properties .3. Alkali (Li, Na, K, Rb) and Alkaline-Earth (Be, Mg, Ca, Sr) Atoms." *Theochem-Journal of Molecular Structure* **80**: 147-171.

Schafmeister, C. E. A. F., Ross, W. S. and Romanovski, V. (1995). Leap, University of San Francisco.

Schauerte, J. A., Schlyer, B. D., Steel, D. G. and Gafni, A. (1995). "Nanosecond Time-Resolved Circular-Polarization of Fluorescence - Study of Nadh Bound to Horse Liver Alcohol-Dehydrogenase." *Proceedings of the National Academy of Sciences of the United States of America* **92**(2): 569-573.

Schelvis, J. P. M. and Varotsis, C. A. (2000). "Picosecond resonance Raman evidence of the structure of a long- lived electronic excited state of low-spin Fe(III) heme o." *Chemical Physics Letters* **321**(1-2): 37-42.

Scherer, P. O. J. and Fischer, S. F. (1989). "Quantum Treatment of the Optical-Spectra and the Initial Electron-Transfer Process with the Reaction Center of Rhodopseudomonas-Viridis." *Chemical Physics* **131**(1): 115-127.

Schlegel, H. B. (1987). "Optimization of Equilibrium Geometries and Transition Structures." *Advances in Chemical Physics* **67**: 249-286.

Schlessinger, J. and Levitzki, A. (1974). "Molecular basis of negative co-operativity in rabbit muscle glyceraldehyde-3-phosphate dehydrogenase." *Journal of Molecular Biology* **82**(4): 547-561.

Schmidt, J., Chen, J., DeTraglia, M., Mikel, D. and McFarland, J. T. (1979). "Solvent deuterium isotope effect on the liver alcohol dehydrogenase reaction." *Journal of the American Chemical Society* **101**: 3634-3640.

Schrödinger, E. (1926). "Quantisierung als eigenwertproblem." *Annals of Physics*(79): 361-376.

Scott, T. G., Spencer, R. D., Leonard, N. J. and Weber, G. (1970). "Emission Properties of NADH. Studies of Fluorescence Lifetimes and Quantum Efficiencies of NADH, AcPyADH, and Simplified Synthetic Models." *Journal of the American Chemical Society* **92**(3): 687-695.

Seward, J. (2003). The Bzip2 compression library, http://sources.redhat.com/bzip2/.

Shavitt, I. (1977). "Methods of Electronic Structure Theory". New York, Plenum Press.

Shearer, G., Kim, K., Lee, K., Wang, C. and Plapp, B. (1993). "Alternative Pathways and Reactions of Benzyl Alcohol and Benzaldehyde with Horse Liver Alcohol Dehydrogenase." *Biochemistry* **32**: 11186-11194.

Shore, J. D. and Gutfreund, H. (1970). "Transients in the reactions of liver alcohol dehydrogenase." *Biochemistry* **9**(24): 4655-4659.

Shu, F., Ramakrishnan, V. and Schoenborn, B. P. (2000). "Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin." *Proceedings of the National Academy of Sciences of the United States of America* **97**(8): 3872-3877.

Singh, U. C. and Kollman, P. A. (1986). "A Combined Abinitio Quantum-Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular-Systems - Applications to the Ch3cl + Cl- Exchange-Reaction and Gas-Phase Protonation of Polyethers." *Journal of Computational Chemistry* **7**(6): 718-730.

Slater, J. C. (1929). "The Theory of Complex Spectra." *Physical Review* **34**: 1293-1322.

Smith, C. M. and Hall, G. G. (1986). "The Approximation of Electron-Densities." *Theoretica Chimica Acta* **69**(1): 63-69.

Smith, W. and Fincham, D. (1993). "The Ewald Sum in Truncated Octahedral and Rhombic Dodecahedral Boundary-Conditions." *Molecular Simulation* **10**(1): 67-71.

Spezia, R., Aschi, M., Di Nola, A., Di Valentin, M., Carbonera, D. and Amadei, A. (2003). "The effect of protein conformational flexibility on the electronic properties of a chromophore." *Biophysical Journal* **84**(5): 2805-2813.

Spiro, T. G. and Czernuszewicz, R. S. (1995). Resonance Raman-Spectroscopy of Metalloproteins. "Biochemical Spectroscopy". **246:** 416-460.

Stanton, R. V., Perakyla, M., Bakowies, D. and Kollman, P. A. (1998). "Combined ab initio and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution." *Journal of the American Chemical Society* **120**(14): 3448-3457.

Stavrov, S. S. (1993). "The Effect of Iron Displacement out of the Porphyrin Plane on the Resonance Raman-Spectra of Heme-Proteins and Iron Porphyrins." *Biophysical Journal* **65**(5): 1942-1950.

Stone, A. J. (1996). "The Theory of Intermolecular Forces". Oxford, Clarendon Press.

Stote, R. H. and Karplus, M. (1995). "Zinc-Binding in Proteins and Solution - a Simple but Accurate Nonbonded Representation." *Proteins-Structure Function and Genetics* **23**(1): 12-31.

Stratmann, R. E. and Scuseria, G. E. (1998). "An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules." *Journal of Chemical Physics* **109**(19): 8218-8224.

Stratt, R. M. and Cho, M. H. (1994). "The Short-Time Dynamics of Solvation." *Journal of Chemical Physics* **100**(9): 6700-6708.

Szabo, A. and Ostlund, N. S. (1996). "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory". New York, Dover Publications Inc.

Thompson, M. A. (1995). "Hybrid Quantum-Mechanical Molecular Mechanical Force-Field Development for Large Flexible Molecules - a Molecular-Dynamics Study of 18-Crown-6." *Journal of Physical Chemistry* **99**(13): 4794-4804.

Thompson, M. A. (1996). "QM/MMpol: A consistent model for solute/solvent polarization. Application to the aqueous solvation and spectroscopy of formaldehyde, acetaldehyde, and acetone." *Journal of Physical Chemistry* **100**(34): 14492-14507.

Thompson, M. A., Glendening, E. D. and Feller, D. (1994). "The Nature of K+ Crown-Ether Interactions - a Hybrid Quantum Mechanical-Molecular Mechanical Study." *Journal of Physical Chemistry* **98**(41): 10465-10476.

Thompson, M. A. and Schenter, G. K. (1995). "Excited-States of the Bacteriochlorophyll-B Dimer of Rhodopseudomonas-Viridis - a Qm/Mm Study of the Photosynthetic Reaction-Center That Includes Mm Polarization." *Journal of Physical Chemistry* **99**(17): 6374-6386.

Thompson, M. A., Zerner, M. C. and Fajer, J. (1991). "A Theoretical-Examination of the Electronic-Structure and Excited-States of the Bacteriochlorophyll B Dimer from Rhodopseudomonas-Viridis." *Journal of Physical Chemistry* **95**(14): 5693-5700.

Tozer, D. J., Amos, R. D., Handy, N. C., Roos, B. O. and Serrano-Andres, L. (1999). "Does density functional theory contribute to the understanding of excited states of unsaturated organic compounds?" *Molecular Physics* **97**(7): 859-868.

Tozer, D. J. and Handy, N. C. (2000). "On the determination of excitation energies using density functional theory." *Physical Chemistry Chemical Physics* **2**(10): 2117-2121.

Trautwein, A. X., Winkler, H., Schwendy, S., Grunsteudel, H., Meyer-Klaucke, W., Leupold, O., Ruter, H. D., Gerdau, E., Haas, M., Realo, E., Mandon, D. and Weiss, R. (1998). "Iron porphyrins reinvestigated by a new method: Mossbauer spectroscopy using synchrotron radiation." *Pure and Applied Chemistry* **70**(4): 917-924.

Tresadern, G., Faulder, P. F., Gleeson, M. P., Tai, Z., MacKenzie, G., Burton, N. A. and Hillier, I. H. (2003). "Recent advances in quantum mechanical/molecular mechanical calculations of enzyme catalysis: hydrogen tunnelling in liver alcohol dehydrogenase and inhibition of elastase by alpha- ketoheterocycles." *Theoretical Chemistry Accounts* **109**(3): 108-117.

Tresadern, G., McNamara, J. P., Mohr, M., Wang, H., Burton, N. A. and Hillier, I. H. (2002). "Calculations of hydrogen tunnelling and enzyme catalysis: a comparison of liver alcohol dehydrogenase, methylamine dehydrogenase and soybean lipoxygenase." *Chemical Physics Letters* **358**(5-6): 489-494.

Trovaslet, M., Dallet-Choisy, S., Meersman, F., Heremans, K., Balny, C. and Legoy, M. D. (2003). "Fluorescence and FTIR study of pressure-induced structural modifications of horse liver alcohol dehydrogenase (HLADH)." *European Journal of Biochemistry* **270**(1): 119-128.

Van Caillie, C. and Amos, R. D. (1999). "Geometric derivatives of excitation energies using SCF and DFT." *Chemical Physics Letters* **308**(3-4): 249-255.

Van Caillie, C. and Amos, R. D. (2000). "Geometric derivatives of density functional theory excitation energies using gradient-corrected functionals." *Chemical Physics Letters* **317**(1-2): 159-164.

van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hunenberger, P. H., Kuger, P., Mark, A. E., Scott, W. R. P. and Tironi, I. G. (1996). "The GROMOS96 Manual and User Guide". Zurich, Switzerland, Biomos.

van Gunsteren, W. F. and Mark, A. E. (1998). "Validation of molecular dynamics simulation." *Journal of Chemical Physics* **108**(15): 6109-6116.

van Gunsteren, W. F., Weiner, P. and Wilkinson, A. J. (1993). "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications." Leiden, ESCOM.

Van Kampen, N. G. (1981). "Stochastic Processes in Physics and Chemistry". Amsterdam, North-Holland.

Vedani, A., Dobler, M. and Dunitz, J. D. (1986). "An empirical potential function for metal centers: application to molecular mechanics calculations on metalloproteins." *Journal of Computational Chemistry* **7**: 701-710.

Vedani, A. and Huhta, D. W. (1990). "A New Force-Field for Modeling Metalloproteins." *Journal of the American Chemical Society* **112**(12): 4759-4767.

Verlet, L. (1967a). "Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lenard-Jones Molecules." *Physical Review* **159**: 98-103.

Verlet, L. (1967b). "Computer 'Experiments' on Classical Fluids. II. Equilibrium Correlation Functions." *Physical Review* **165**: 201-204.

Vivian, J. T. and Callis, P. R. (2001). "Mechanisms of tryptophan fluorescence shifts in proteins." *Biophysical Journal* **80**(5): 2093-2109.

Vojtechovsky, J., Chu, K., Berendzen, J., Sweet, R. M. and Schlichting, I. (1999). "Crystal structures of myoglobin-ligand complexes at near-atomic resolution." *Biophysical Journal* **77**(4): 2153-2174.

Walker, R. C., Byungmoon, C., Amer, H., Mercer, I. P., Klug, D. R. and Gould, I. R. (2003a). "Adiabatic Effects in the Electronic States of Zinc Myoglobin." *Journal of Physical Chemistry B* **To be submitted**.

Walker, R. C., de Souza, M. M., Mercer, I. P., Gould, I. R. and Klug, D. R. (2002). "Large and fast relaxations inside a protein: Calculation and measurement of reorganization energies in alcohol dehydrogenase." *Journal of Physical Chemistry B* **106**(44): 11658-11665.

Walker, R. C., Klug, D. R. and Gould, I. R. (2003b). "A Comparison of the Performance of Ab Initio, DFT and Semi-Empirical Methods for Probing Equilibrium Fluctuations." *Journal of the American Chemical Society* **to be submitted**.

Warshel, A. and Levitt, M. (1976). "Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilisation of the carbonium ion in the reaction of lysozyme." *Journal of Molecular Biology* **103**: 227-249.

Weidig, C. F., Halvorson, H. R. and Shore, J. D. (1977). "Evidence for site equivalence in the reaction mechanism of horse liver alcohol dehydrogenase with aromatic substrates at alkaline pH." *Biochemistry* **16**(13): 2961-2922.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984). "A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins." *Journal of the American Chemical Society* **106**(3): 765-784.

Wiberg, K. B., Hadad, C. M., Ellison, G. B. and Foresman, J. B. (1993). "Butadiene .3. Charge-Distribution in Electronically Excited-States." *Journal of Physical Chemistry* **97**(51): 13586-13597.

Wolfram Research, I. (1999). Mathematica. Champaign, Illinois, Wolfram Research, Inc.

Yamamoto, S., Diercksen, G. H. F. and Karelson, M. (2000). "An ab initio CI study of electronic spectra of substituted free-base porphyrins." *Chemical Physics Letters* **318**(6): 590-596.

Yokoyama, A. and Omori, T. (2003). "Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers." *Japanese Journal of Clinical Oncology* **33**(3): 111-121.

Yoshida, A. (1994). "Genetic polymorphisms of alcohol metabolizing enzymes related to alcohol sensitivity and alcoholic diseases." *Alcohol and Alcoholism* **29**: 693-696.

Yue, K. T., Martin, C. L., Yang, J. P., Lee, S. K., Sloan, D. L. and Callender, R. H. (1985). "A Raman-Study of Reduced Nicotinamide Adenine-Dinucleotide Bound to Liver Alcohol-Dehydrogenase." *Biophysical Journal* **47**(2): A412-A412.

Zerner, M. C., Loew, G. H., Kirchner, R. F. and Mueller-Westerhoff, U. T. (1980). "An intermediate neglect of differential overlap technique for spectroscopy of transition-metal complexes. Ferrocence." *Journal of the American Chemical Society* **102**(2): 589-599.

Zhang, B. L. and Pionnier, S. (2002). "Natural stereospecific hydrogen isotope transfer in alcohol dehydrogenase-catalysed reduction." *Nukleonika* **47**: S29-S31.

Zheng, X. H. and Stuchebrukhov, A. A. (2003). "Electron tunneling in proteins: Implementation of ZINDO model for tunneling currents calculations." *Journal of Physical Chemistry B* **107**(27): 6621-6628.