Calculation of the Vibrational States of Small Biomolecules

A thesis presented to the University of London in partial fulfilment

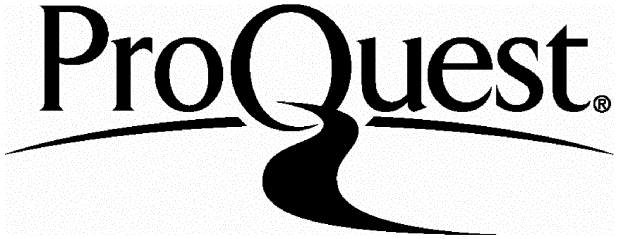
of the requirements for the degree of Doctor of Philosphy

Shervin Moghaddam

The Department of Chemistry

University College London

July 2004



"If we knew what we were doing, it wouldn't be called research, would it?"

— Albert Einstein

Acknowledgements

I would like to sincerely thank my supervisor, David Clary, for all his patient help and inspiration.

I would also like to express my gratitude to all members of the Clary group for making the experience so enjoyable. In particular, Anthony, Adam, Andy, Carmen, Chris, Davids Benoit and Charlo, Dominic, Erasmo, Sergei, Tanja and Tommy provided extremely helpful comments and suggestions during the course of my research.

I am forever indebted to my friends for their extraordinary support. My sincere thanks to my contemporaries at Epsom, Merton and UCL and of course to all at Le Grand Bornand and the Ten Bells. Extra special mentions must be given to Sophie for her patience, to Anthony for being such a huge inspiration and friend and to Joe for his inexplicable yet undying enthusiasm for this project.

Finally, thanks go to my family for their continued support throughout my education.

Abstract

The present thesis documents the development of a general method for calculating the vibrational energies of small biomolecules. Starting with the general time-dependent molecular Hamiltonian, assumptions allowing the separation of coordinates are outlined. The variational Configuration Interaction technique for calculating ab initio electronic eigenstates is adapted for vibrational systems, and a general method for solving the Hamiltonian in normal coordinate space is described. A computer program is developed, implementing this formulation of the quantum problem. The code is tested against benchmark systems including the Morse Oscillator, the Hénon-Heiles 3D potential and water monomer. It is seen that the results agree well with those of DVR and variational calculations. The vibrational CI method is applied to tryptamine and its water clusters, systems of great biological importance. Results are used to test the assignment of experimental IR spectra for the monomer and are compared with ab initio results to probe the accuracy of anharmonic terms in the molecular modelling potential. An attempt is made to assign the conformational isomers found in the solvated spectra. The results of CI calculations are found to be an improvement on harmonic calculations, agreeing well with experimental spectra in some cases. High order terms in the potential are found to model the anharmonicity of the system accurately. The results demonstrate that anharmonic effects in widely-used potential functions for biomolecules are large and need to be treated accurately in interpreting vibrational spectra.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Molecular Vibrations

1.1 Biomolecules

Amino acids are the building blocks of life. These bio-molecules are crucial precursors in the synthesis of peptides and proteins. They are involved at the basic level of the biochemical pathways of life. There are 20 naturally occurring amino acids, eight of which are essential in nutrition. The other 12 can be synthesised from these. All stereoisomeric amino acids in nature are found in the laevorotatory form. Two amino acids can react to form a peptide bond. The acid group from one molecule reacts with the amine group of the other to form an amide linking group with the loss of water. Many amino acids can link in a sequence via a polymerisation process

to form proteins. These can be thought of as biological machines. For example, enzymes are proteins that catalyse reactions and hormones, such as insulin, act as messengers in biological systems.

Many biochemical processes are governed by non-bonded interactions between, for example, proteins and sugars. The 20 amino acid differ only in their side chains. It is these side chains that control the non-bonded interactions in proteins. Thus, the sequence of amino acids defines the structure and, therefore, bioactivity of the protein. Such interactions are found to play a central role in stabilising solvated structures and in the activation or inhibition of physiological processes by proteins. In many cases, man-made drugs are designed to maximise non-bonded interactions so as to block biological pathways such as the insertion mechanism of the HIV virus into human cells.[1–7] Clearly, study of intermolecular interactions in such systems is of great biological and medical interest.[8–16]

Systems of biological or physiological importance are mainly found in aqueous media. Therefore, studies of the effects of biomolecule-water complexation are of great importance.[17–19] Applications are wide-ranging and include the function of biomolecules[18] and simulations of water effects on protein folding structures and mechanisms.[19] Hydrogen bonds are weak compared to covalent interactions and, therefore, exhibit highly anharmonic behaviour. On forming a hydrogen bond, the donor molecule experiences a weakening of the donor-hydrogen bond. This results in a red-shift in corresponding infra-red absorption. Thus, accurate models of water-

biomolecule interactions are needed to interpret and assign experimental vibrational spectra.

Interest in computer simulation of biomolecules has grown dramatically in the last decade as computers have grown more powerful, enabling calculations on more complex systems. Molecular modelling force fields have been developed with atom types optimised for amino acid residues in proteins.[20–42] For these surfaces, the time needed to calculate each potential point is greatly reduced with respect to ab initio methods. Thus, calculations on larger systems are now feasible. These force fields are used in the calculation of structure and bonding. However, studies of the internal and inter-molecular vibrations of biomolecules are also of great interest. It is important to test the accuracy of these force fields by comparison with precise experimental data such as infra-red spectra.

1.2 Vibrational States

In the classical view of molecular systems, molecules can have no vibrational energy; the atoms are stationary in their equilibrium positions. The properties of the bond, such as strength and length, are measured from the bottom of the potential well. In the quantum mechanical interpretation of bonding, however, even bound atoms in the ground vibrational state oscillate about the potential minimum. The zero-point energy must be considered when discussing the bond properties, reducing

the energy needed to reach the vibrational continuum and break the bond and, in many cases, increasing the average bond length. Furthermore, in the classical picture of vibrators, the atoms are most likely to be found at the extremes of the motion. However, in the real quantum world, the probability density is maximised at the potential minimum. Consequently, a rigorous approach to the calculation of vibrational states in the quantum frame is essential.

The vibrational zero-point energy is implicit in all bound systems and is found to be important in fields as diverse as proton transfer in photosynthesis[43] and to the formation of hydrogen in the interstellar medium.[44–46]

Modelling vibrations can be vital in predicting barrier heights and rate constants in transition state theory. Consideration of the zero point energies of reactants and transition states is important as these can affect calculated reaction pathways and rate constants.[47–50] Furthermore, such research can aid understanding of chemical stability.[50–52]

In molecular spectroscopy, accurate calculation of ground and excited vibrational states can specifically aid in assigning infra-red, electronic and rotational spectra.[53–58] Assignment, in turn, enables a better understanding of energetically preferred structures and binding sites for solvation. Recent spectroscopic studies provide a wide range of biomolecules of interest to study.[59–78] High quality, quantitative information is available from new spectroscopic techniques and thus a theoretical

method is required to analyze the data. Of particular interest are experimental spectra of amino acids[61] such as tryptophan,[72–75] and their analogues.[63–71] For example, spectroscopic studies have yielded information on the role of water in directing conformational preferences in biomolecules.[69] Analysis of these spectra can aid in understanding the preference of one solvent binding site over another and how this affects the structure and reactivity of the system. The work of Kleinermanns *et al.* is particularly fascinating, as, in measuring IR spectra of monomer and bound DNA bases,[76–78] they are probing the inner workings of the genetic code.

1.3 Modelling Vibrations

1.3.1 The Harmonic Approximation

In many studies of vibrational states, cubic and higher order terms in the Taylor series expansion of the potential energy function are assumed to be negligible.[65–67,69–71,76–78] This is the harmonic approximation, the assumption that, near the minimum, the form of the potential is effectively quadratic. This results in a simplified system where no overtones are seen in the spectrum and vibrational modes do not mix. Calculating harmonic vibrational frequencies has proved effective; enabling, for example, assignation of experimental spectra.[66,67] However, high resolution spectroscopy invariably shows the inaccuracies of this model. Even

low resolution spectra can display features attributed to the effects of anharmonicity, particularly in weakly bound systems and for transitions between high energy vibrational states.

1.3.2 Anharmonicity

Strongly bound molecular complexes can be modelled relatively well using a simple harmonic approach. However, this method can be inaccurate for some vibrations. For example, the calculated harmonic frequencies for OH stretching modes in water dimer can deviate from experiment by 100-200cm^{-1}.[79] Two kinds of anharmonicity can be observed. Mechanical anharmonicity is the deviation of the potential from the harmonic model. In the majority of cases, this causes a decrease in the energy of the vibrational eigenstate. Electronic anharmonicity is the result of the dipole moment changing as the bond length is altered. Electronic anharmonicity contributes to the presence of overtones in the spectrum albeit at reduced intensities.

Early attempts to quantify anharmonicity concentrated on small systems.[80–98] As a key molecule in many fields, such as astro-, geo- and bio-chemistry, water was a focal point for many studies.[94–97] Small clusters of water molecules[88–93] have also come under scrutiny as a stepping stone to the ultimate goal of an accurate model of the bulk liquid.[99–102] Variational,[92–97] local mode[90,91] and Discrete Variable Representation, DVR,[86] approaches have been used in modelling anharmonic vibrational states.

Bowman suggested a self consistent field, SCF, approach for a problem of two coupled oscillators.[87] The SCF procedure was shown to accurately predict eigenenergies and described the wavefunction well; better, in fact, than harmonic wavefunctions. Consequently, SCF wavefunctions have been employed as the basis for variational calculations to reduce the number of basis functions needed for the calculation to converge.[96–98]

Extremely detailed calculations have been performed on the vibrations of water[103–124] and other small molecules.[125–128] Even relativistic effects[112,116] and the breakdown of the Born-Oppenheimer approximation[106] have been included. These are very accurate calculations but the complexity of the techniques involved mean that extension to larger systems can prove difficult.

Gerber and co-workers have developed a perturbation technique for calculating the vibrational eigenstates.[58,129–131] The procedure involves a change to normal coordinates followed by a vibrational self-consistent field, VSCF, calculation. A Møller-Plesset type perturbation theory approach, typically of second order, is used to improve on the self consistent field wavefunctions. This technique is fast and general and can be applied with no rigid body approximation i.e. to all atoms. All modes are included in the basis, therefore the coupling between inter- and intra- molecular modes can be studied.[132] The VSCF-perturbation procedure has been applied to many systems, such as water[79,133–136] and argon[137,138] clusters, biomolecules such as glycine[58,130,139,140] and simple organic systems.[58,130,132,134,141] Accurate eigenstates

of rare gas compounds have been calculated to allow identification of species from IR spectra.[53–55,57,142] In the study of the transition from covalent to ionic bonding in hydrated hydrogen-halide clusters, harmonic calculations were shown to be inaccurate as the proton stretching motions are highly anharmonic.[49,56,143] In early studies, Gerber and co-workers used empirical and semi-empirical potentials.[133] This approach renders the direction of research dependent on the availability of high quality potentials. To remove this reliance, the VSCF-perturbation method was adapted to work with ab initio[134] and density functional, DFT,[135] potentials. This facilitated the testing of anharmonic terms in empirical and semi-empirical potentials by direct comparison of results with those of ab initio calculations.[58,139,140]

The Handy group have performed accurate calculations on the vibrational states of small molecules, producing potentials,[144–150] anharmonic constants[151] and spectra.[144–149,152,153] Recently, Carter, Handy and Bowman developed a general method for calculating anharmonic vibrational states. The procedure involved a variational improvement to VSCF basis functions.[154,155] This was extended to a general code, Multimode,[156] and was applied to systems such as benzene using an all-atom potential.[157] The program was also used to investigate the parameterisation of force fields in Morse, Gauss-like and Taylor expansion coordinate spaces.[158,159] Multimode takes into consideration coupling between 2, 3 and 4 modes and is shown to produce calculated spectra of high accuracy. However, the complexity of the calculations mean that the code is only applicable to systems of relatively small molecules

with up to 12 atoms.[157] There is a need for a general technique that can be used on larger systems so that data from spectral studies of important biomolecules can be analysed. Furthermore, it is necessary to test the parameterisation of molecular modelling potentials of such systems. Of particular interest are the importance and accuracy of anharmonic terms in these force fields.

The Diffusion Monte Carlo, DMC, algorithm for electronic structure calculations[160] was adapted for the vibrational problem.[47,48,161–178] Much work has been done on the structure[162–167] and properties[172] of water clusters and aromatic-water clusters.[48,174,175] Accurate vibrational eigenstates have been used to improve calculated rates of reaction and tunnelling.[47,48] However, many of these calculations made the rigid body approximation to reduce the dimensionality of the problem. This method normally gives only the ground vibrational states of molecules which limits its use for interpreting spectra.

1.4 This Work

The research presented in this thesis is an attempt to quantify the anharmonic contributions to vibrational states in small biomolecules.

In Part II, the theoretical background to the problem is presented. A quantum mechanical method is developed with a full description of the techniques used and the approximations made in order to solve the problem.

The construction and testing of a computer program developed specifically for this research is outlined in Part III. Details of the functionality of the code are provided here along with the results of the methods used to ensure the accuracy of the program. The Morse oscillator is used as a 1D benchmark for the code. The extension to three dimensions and the pairwise potential approximation are tested by comparison of results with literature values of vibrational eigenstates of the Hénon-Heiles 3D oscillator and water monomer.

In Part IV of the thesis, the CI method for calculating anharmonic vibrational eigenstates is applied to biomolecular systems of interest. Experimental spectra of tryptamine monomer and water clusters are compared with spectra calculated using the CI technique allowing a test of the method and an investigation of the accuracy of anharmonic terms in the potential.

Part II

Theory

Chapter 2

The Theory of Molecular Vibrations

2.1 The Time Independent Schrödinger Equation

Defined in Equation 2.1, the time independent Schrödinger equation describes a stationary state of energy, E, with wavefunction, Ψ, that contains all necessary information about the system.

$$\hat{\mathscr{H}}\Psi(\mathbf{x}) = E\Psi(\mathbf{x}) \tag{2.1}$$

These are eigenvalue and eigenfunction, respectively of the complete molecular Hamiltonian, $\hat{\mathscr{H}}$. The Hamiltonian can be expressed as a sum of kinetic energy, T, and potential energy, V, operators.

$$\hat{\mathscr{H}} = \hat{\mathbf{T}} + \hat{\mathbf{V}} \tag{2.2}$$

This equation is still very general and very difficult to solve fully for systems of three bodies or more. Approximations must be made to further simplify the problem at hand. The Hamiltonian contains terms depending on, for example, electronic and nuclear coordinates. In order to treat the vibrations of the nuclear frame, it is necessary to separate these contributions to the Hamiltonian to the maximum extent. This will allow a separation of the Schrödinger equation into equations in various coordinate spaces.

2.2 The Born-Oppenheimer Approximation

The first separation arises from the observation that nuclei are at least three orders of magnitude more massive than electrons. Electrons can therefore move much faster and adjust very quickly to changes in nuclear configuration. It is assumed that this electron adjustment is instantaneous, the electrons therefore experiencing a completely frozen nuclear conformation. This is the Born-Oppenheimer, B-O, approximation, a separation of the motion of electrons from the motion of nuclei. It is therefore possible to split the Hamiltonian into electronic and nuclear coordinates.

The electronic Hamiltonian, $\hat{\mathscr{H}}_e$, may be used to solve the Schrödinger equation for a given configuration of nuclei, x_n

$$\hat{\mathscr{H}}_e(x_e; x_n) = \hat{\mathbf{T}}_e(x_e) + \hat{\mathbf{V}}_e(x_e; x_n) \tag{2.3}$$

$$\hat{\mathscr{H}}_e(x_e; x_n)\Psi_e(x_e; x_n) = E_e(x_n)\Psi_e(x_e; x_n) \tag{2.4}$$

The electronic energy, E_e, can be calculated over a range of nuclear conformations. These points can be fitted to form a potential energy surface, a functional form for the electronic energy of the system at different nuclear configurations, $E_e(x_n)$.

Following the Born-Oppenheimer separation, the nuclear Hamiltonian can be expressed solely in nuclear coordinates

$$\hat{\mathscr{H}}_n(x_n) = \hat{\mathbf{T}}_n(x_n) + \hat{\mathbf{V}}_n(x_n) \tag{2.5}$$

and the full Hamiltonian of the system is simply the sum of electronic and nuclear Hamiltonians

$$\hat{\mathscr{H}}(x_e; x_n) = \hat{\mathscr{H}}_e(x_e; x_n) + \hat{\mathscr{H}}_n(x_n) \tag{2.6}$$

Using the potential energy surface obtained by solving Equation 2.4, the Schrödinger equation for nuclear motion is, in principle, soluble

$$\left[E_e(x_n) + \hat{\mathscr{H}}_n(x_n)\right] \Psi_n(x_n) = E_n\Psi_n(x_n) \tag{2.7}$$

where $\psi_n(x_n)$ is the nuclear part of the total wavefunction, Ψ

$$\Psi = \Psi_e(x_e; x_n).\Psi_n(x_n) \tag{2.8}$$

The Born-Oppenheimer approximation is applied extensively for molecular systems that are not in highly excited states.

2.2.1 The Potential Energy Surface

Potential energy surfaces arise from the Born-Oppenheimer separation of electronic and nuclear motion. The B-O approximation resolves the motion of electrons and nuclei allowing electronic and nuclear calculations to be performed independently. Electronic structure calculations of eigenstates at various nuclear conformations can be performed and the energies fitted to functions approximating $E_e(x_n)$. Such potential energy surfaces allow quick and easy calculation of the electronic energy at a given nuclear geometry. A variety of surfaces are used in this study, and the validity and accuracy of our calculations are dependent on them. To aid discussion, the three main types of potential energy surface are introduced briefly.

1. Empirical. The potential is a functional form, usually a polynomial expansion of coordinates or set of Morse functions. All parameters are chosen to fit to experimental data.

2. Semi-Empirical. Some parts of the potential, for which the form is known accurately from quantum calculation, are used alongside empirical corrections.

3. Ab initio. The energy at every point is obtained from quantum mechanical calculations.

In many studies of vibrational eigenstates, the potential is assumed to have a harmonic form when close to the minimum to simplify calculations. This approximation is discussed in Section 2.4. High order anharmonic terms in a potential are important, however, as they allow the molecule to dissociate. Furthermore, in weakly bound and vibrationally excited systems, anharmonic contributions to the eigenenergy can be significant. Clearly, testing anharmonicity in force fields is essential in the development of accurate potentials. Later in this study, experimental spectra and calculated anharmonic vibrational energies are compared, enabling testing and parameterisation of empirical and semi-empirical potential energy surfaces for small biomolecules.

2.3 Rovibratonal Coupling

The complete molecular Hamiltonian has thus far been reduced to a general $3N$ dimensional nuclear Hamiltonian, where N is the number of atoms. The translation of the system as a whole can be separated in field-free problems leaving $3N - 3$ rotations and internal vibrations.

The vibrations and rotations of molecular systems can only be separated approximately.[179,180] The bond lengths and angles change as a molecule rotates, thereby coupling nearby vibrational modes within the same electronic state. In extreme cases, rovibrational coupling can lead to interaction between different electronic

states, causing a breakdown of the Born-Oppenheimer approximation.

In this work, all systems studied are assumed to be in the ground rotational state. For such systems, where the rotational angular momentum, J, is zero, the rovibrational coupling is small. In this work, the rovibrational coupling is assumed to be negligible and set to zero.[181–183] For biomolecular systems this is a reasonable approximation to make.

The removal of translational and rotational motion from the molecular Hamiltonian leaves a system of coupled oscillators. In many studies of vibrational states, the harmonic approximation is made, both as a way of investigating vibrational properties and as a basis for further calculations. This assumption simplifies the potential energy surface allowing a separation of the $3N-6$ coupled vibrations into $3N-6$ independent oscillators. The next section defines and describes the harmonic approximation and discusses its validity.

2.4 The Harmonic Approximation

2.4.1 Introduction

The most widely used technique for studying nuclear oscillations and their effect on spectral, kinetic and other vibrational properties of molecules is the harmonic

approximation. The precise details of this technique are presented in the later sections of this chapter. Briefly, it is the assumption that, near a minimum, the potential has a form that can be well defined by a quadratic function. The benefits of this approximation are:

- Generality. A simple, logical process allows calculation of the vibrational eigenfunction and frequencies for any molecular system.
- Speed. It is a quick way of calculating the vibrational energies and wavefunctions of a system. They can also be performed directly in ab initio calculations.
- Accuracy. The approximation is found to be quite good for many systems, especially for ground states of strongly bound and non-hydrogen containing molecules.
- Orthogonality. The harmonic approximation wavefunctions are orthogonal so the vibrational motion of a N-atom system can be separated into $3N - 6$ oscillators.

Figure 2.1 is a 1D representation of the ground state of a strongly bound system. It is clear that the harmonic approximation is valid only at small displacements from the equilibrium geometry.

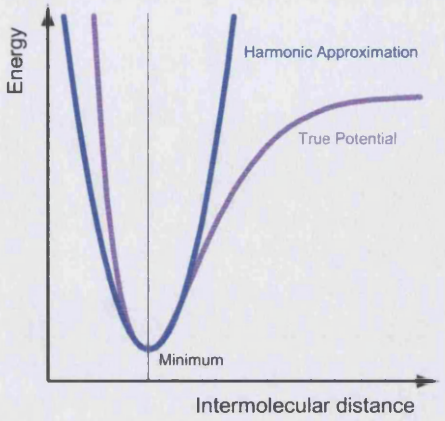


Figure 2.1: Harmonic approximation; at small displacement from the minimum the true potential follows a quadratic function.

2.4.2 The Harmonic Approximation

The separation of kinetic from potential energy in the Schrödinger equation is described in Equation 2.2. The nuclear motion can, therefore, be considered independently from the electronic motion. Classically, the total kinetic energy of the N nuclei system can be described by $3N$ Cartesian coordinates, x_i.

$$\mathbf{T} = \frac{1}{2}\sum_{i=1}^{3N} m_i \left(\frac{\partial x_i}{\partial t}\right)^2 \tag{2.9}$$

where m_i is the mass of atom i. The potential energy can be expanded around a minimum as a Taylor series in powers of x_i.[179,180,184–187] In displacement coordinates i.e. the minimum is at x=0, the potential can be expanded as

$$\mathbf{V}(x_1, x_2, \ldots, x_{3N}) = \mathbf{V}_0 + \sum_{i=1}^{3N} \left(\frac{\partial \mathbf{V}}{\partial x_i}\right)_{x=0} x_i + \\ + \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(\frac{\partial^2 \mathbf{V}}{\partial x_i \partial x_j}\right)_{x=0} x_i x_j \\ + \frac{1}{6} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \sum_{k=1}^{3N} \left(\frac{\partial^3 \mathbf{V}}{\partial x_i \partial x_j \partial x_k}\right)_{x=0} x_i x_j x_k + \ldots \quad (2.10)$$

Cross terms in the expansion such as $\left(\frac{\partial^2 \mathbf{V}}{\partial x_i \partial x_j}\right)_{x=0} x_i x_j$ couple motions in different Cartesian coordinates.

The first term, $\mathbf{V}_0$, is the minimum energy of the potential, an arbitrary constant which can justifiably be set to zero. Furthermore, the coordinates, x_i, are displacement coordinates such that $x = 0$ corresponds to the equilibrium position, the minimum of the potential energy surface

$$\left(\frac{\partial \mathbf{V}}{\partial x_i}\right)_{x=0} = 0 \qquad i = 1, 2, \ldots, 3N \quad (2.11)$$

The potential expansion at vibrational equilibrium, Equation 2.10, is thus reduced to terms of order 2 and higher.

The harmonic approximation is the assumption that, as the displacements from equilibrium are small, the third and higher order terms in the potential expansion are negligible (see Figure 2.1). A simplified multi-dimensional potential results

$$\mathbf{V}(x_1, x_2, \ldots, x_{3N}) = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(\frac{\partial^2 \mathbf{V}}{\partial x_i \partial x_j}\right)_{x=0} x_i x_j \\ = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \mathbf{F}_{ij} x_i x_j \quad (2.12)$$

where $\mathbf{F}_{ij}$ is a symmetric matrix of harmonic force constants.

$$\mathbf{F}_{ij} = \left(\frac{\partial^2 \mathbf{V}}{\partial x_i \partial x_j}\right) = \left(\frac{\partial^2 \mathbf{V}}{\partial x_j \partial x_i}\right) = \mathbf{F}_{ji} \tag{2.13}$$

Thus, Newton's laws of motion, $F = ma$, can be written

$$m\left(\frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2}\right) = -\frac{\partial \mathbf{V}}{\partial x_i} = -\sum_{j=1}^{3N} \mathbf{F}_{ij} x_j \tag{2.14}$$

In general, any, or all, of the elements in $\mathbf{F}$ may be non-zero, resulting in a set of $3N$ coupled differential equations in Cartesian coordinates.

There is, however, another set of coordinates, $q_i, i = 1, 2, \ldots, 3N$, in which Newton's equations are uncoupled, leaving $\mathbf{T}$ and $\mathbf{V}$ as diagonal matrices. The solution for each coordinate can, therefore, be obtained separately:

$$\mathbf{V}(q_1, q_2, \ldots, q_{3N}) = \frac{1}{2}\sum_{i=1}^{3N} \lambda_i q_i^2 \tag{2.15}$$

$$\mathbf{T} = \frac{1}{2}\sum_{i=1}^{3N} \left(\frac{\partial q_i}{\partial t}\right)^2 \tag{2.16}$$

In these so called normal coordinates, q_i, the system is equivalent to a set of $3N$ independent harmonic oscillators. Extending the harmonic approximation to the quantum case, this separation allows the full vibrational wavefunction to be expressed as a product of all the individual normal mode wavefunctions.

$$\Psi = \psi_1 \psi_2 \ldots \psi_{3N-6} \tag{2.17}$$

The normal mode approximation assumes that the potential is roughly harmonic around the minimum. Thus, in order to calculate the normal modes of a system, the first step must be a search for the minimum of interest on the potential energy surface.

2.5 Minimisation of the Potential Energy

Finding minima on complex multi-minima potentials can be extremely difficult both theoretically and computationally. Much research has gone into developing techniques to achieve this. Geometry optimisation methods such as simulated annealing,[188–191] genetic algorithms[192] and the Large-scale Bound-constrained or Unconstrained Optimization, L-BFGS,[193,194] method implemented in the TINKER[195–201] molecular modelling code can be used. These techniques permit a rigorous scan of the conformational landscape, finding conformers of interest in complex multi-minima systems. In this section, an outline is given to the minimisation techniques employed in this project; the simple gradient following steepest descent[202,203] and more efficient Newton-Raphson[202–204] and conjugate gradient[202,203,205,206] minimisations.

Steepest descent, SD, is a standard multi-dimensional minimisation that converges linearly but will eventually converge to a minimum.[202,203] It is an iterative process moving towards the minimum in a series of steps. For example, consider a minimisa-

tion of the potential $V(x)$ with starting position x_0. At each iteration, k, the search direction, is taken as $-\frac{\partial V(x)}{\partial x_k}$, the negative gradient of the function to be minimised at point x_k. This method is simple but can be slow, especially near the solution. The Newton-Raphson and conjugate gradient methods are also iterative derivative techniques, but the search directions and step sizes are calculated in a more complex manner.

Newton methods are based on approximating the potential function locally by a quadratic model and subsequently minimizing.[202–204] The method derives from the Taylor series expansion of the potential.

$$f(x+\delta) \approx f(x) + \left(\frac{\partial f(x)}{\partial x}\right)\delta + \frac{1}{2}\left(\frac{\partial^2 f(x)}{\partial x^2}\right)\delta^2 + \ldots \tag{2.18}$$

In a step analogous to the harmonic approximation, quadratic terms are assumed to dominate the potential as the minimum is approached and terms beyond linear become negligible. Moving to the zero of the function $f(x)$ i.e. $f(x+\delta) = 0$ implies

$$\delta = f(x)/\left(\frac{\partial f(x)}{\partial x}\right) \tag{2.19}$$

The Newton methods find the zero of a function, but the potential is not necessarily zero at the minimum. The zero of the derivative of the potential must be found. Thus, the method requires second order differentials which are harder and take longer to compute than the first order differentials required by steepest descent. However, far fewer iterations are required to find the minimum with the Newton-Raphson technique as it displays quadratic convergence, as shown in Figure 2.2.

The technique is most efficient when close to the minimum. Far from the solution, the higher order terms in Equation 2.18 become important rendering the Newton-Raphson minimisation method grossly inaccurate.

The conjugate gradient, CG, method is the same as steepest descent for the first iteration.[202,203,205,206] However, successive iterations are constructed so they form a set of mutually conjugate vectors with respect to the Hessian, i.e. the force constant matrix, of a general convex quadratic function. Like Newton methods, CG displays quadratic convergence close to a minimum .

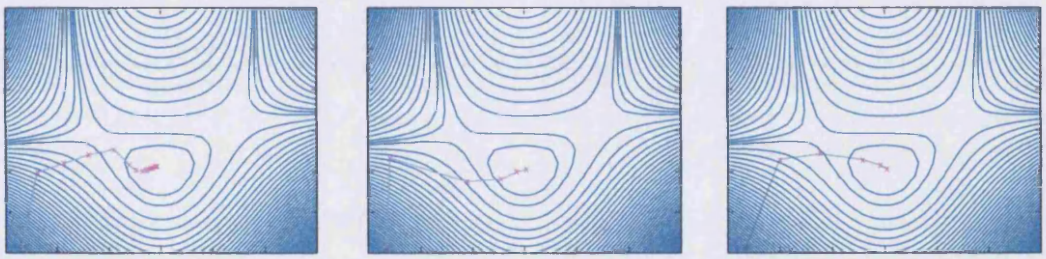


Figure 2.2: Schematic representation of steepest descent, Newton-Raphson and conjugate gradient minimisation techniques on a 3D Hénon-Heiles potential at z=0.

Figure 2.2 is a schematic representation of the efficiency of steepest descent, Newton-Raphson and conjugate gradient minimisations on a Hénon-Heiles potential.[207] Newton-Raphson and CG are more efficient and more accurate than steepest descent when close to the minimum. However, steepest descent is less likely to make

the large errors associated with Newton and CG methods far from root so the techniques can be used together. Steepest descent is often used to find the minimum of interest with a few steps of Newton-Raphson or CG to home in on the root more quickly at the end.

2.6 Normal Mode Analysis

The generalised multi-dimensional extension of the harmonic oscillator model for many body systems is called the normal mode approximation. The aim is to find a coordinate space, q, in which the vibrations of the molecule are reduced to a series of uncoupled harmonic oscillators. There are several techniques for achieving this goal such as the creation-annihilation operator method[184] and the Wilson **FG** matrix method.[179,180,185–187,208] The technique chosen is outlined in this section.

Once the minimum is found, the normal modes can be calculated at that point. To move from Cartesian, x, to normal, q, coordinates requires a simple linear transformation.

$$x_i = \sum_{j=1}^{3N} \mathbf{L}_{ij} q_j \quad \text{or} \quad x = \mathbf{L}q \tag{2.20}$$

$$q_i = \sum_{j=1}^{3N} \mathbf{L}_{ij}^{-1} x_j \quad \text{or} \quad q = \mathbf{L}^{-1}x \tag{2.21}$$

The normal coordinates are just a linear combination of Cartesian coordinates. The transformation matrix $\mathbf{L}$ must be found. It is a $3N$x$3N$ matrix of constant coefficients that transforms from Cartesian to normal coordinate space.

The first stage in determining the matrix $\mathbf{L}$ involves transformation to mass-weighted Cartesian coordinates, $\bar{x}$.

$$\bar{x}_i = m_i^{\frac{1}{2}} x_i \quad \text{or} \quad \bar{x} = \mathbf{M}^{\frac{1}{2}} x \quad \text{where} \quad \mathbf{M}_{ij} = m_i \delta_{ij} \tag{2.22}$$

The potential and kinetic energies become

$$\mathbf{T} = \frac{1}{2} \sum_{i=1}^{3N} \left(\frac{\partial \bar{x}_i}{\partial t} \right)^2 \tag{2.23}$$

$$\mathbf{V} = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \mathbf{F}'_{ij} \bar{x}_i \bar{x}_j \tag{2.24}$$

The mass-weighted force constant matrix, $\mathbf{F}'_{ij}$, is

$$\mathbf{F}'_{ij} = \mathbf{F}_{ij} / \sqrt{m_i m_j} \tag{2.25}$$

Clearly, inputting Equations 2.25 and 2.22 into 2.24 will lead to the simplified harmonic potential described in Equation 2.12.

Diagonalising $\mathbf{F}'_{ij}$ sets off-diagonal force constants in Equation 2.24 that couple the vibrations Cartesian space to zero. The resulting eigenvector matrix is $\mathbf{L}'$, the transformation matrix between mass-weighted Cartesians, $\bar{x}$, and normal coordinates, q.

$$\mathbf{F}'\mathbf{L}' = \mathbf{\Lambda}\mathbf{L}' \quad \text{where} \quad \mathbf{\Lambda}_{ij} = \lambda_i \delta_{ij} \tag{2.26}$$

$$\bar{x} = \mathbf{L}' q \tag{2.27}$$

The square roots of the eigenvalues correspond to the normal mode vibrational frequencies.

$$E = \lambda_i^{\frac{1}{2}} \tag{2.28}$$

To obtain the transformation matrix, **L**, the mass-weighting must be removed from **L'**

$$x = \mathbf{M}^{-\frac{1}{2}}\bar{x} = \mathbf{M}^{-\frac{1}{2}}\mathbf{L}'q = \mathbf{L}q$$

$$\mathbf{L} = \mathbf{M}^{-\frac{1}{2}}\mathbf{L}' \tag{2.29}$$

2.6.1 The Normal Mode Wavefunction

The normal mode wavefunction,[179,184–187] $\psi_v(y)$, can be written in terms of the vibrational quantum number, v.

$$\psi_v(y) = N_v H_v(y) e^{-y^2/2} \qquad y = (\omega/\hbar)^{\frac{1}{2}} q \tag{2.30}$$

The Hermite polynomial $H_v(y)$ is normalised by the constant N_v and ω is the vibrational frequency of the mode. The eigenstate energy for quantum number v is

$$E_v = \left(v + \frac{1}{2}\right)\hbar\omega \tag{2.31}$$

The Hermite polynomials can be found via the recursion relation[184,209]

$$H_0(y) = 1$$

$$H_{v+1}(y) = 2yH_v - 2vH_{v-1} \qquad N_v^2 = 1/(\alpha\pi^{\frac{1}{2}}2^v v!) \tag{2.32}$$

2.7 Breakdown of the Harmonic Approximation

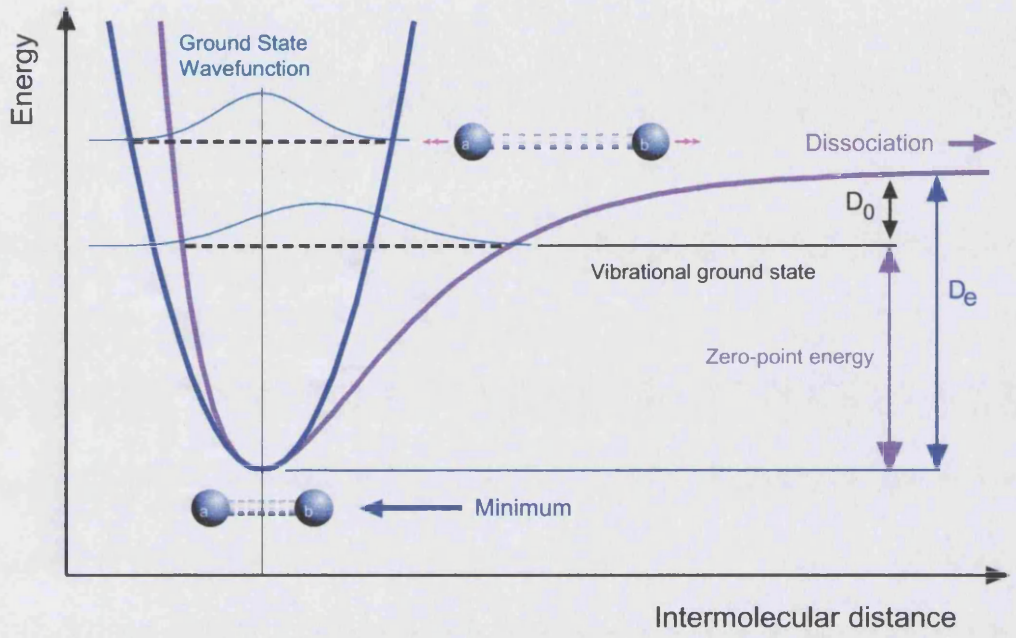


Figure 2.3: Visualisation of the ground state of a weakly bound 1D system. The energy of interaction is labelled D_e and the Dissociation Energy is D_0. The red curve is the true potential. The blue curve shows the potential defined by a quadratic function i.e. that found by normal mode analysis. There can be large differences between harmonic and real zero-point energies and wavefunctions.

While the harmonic approximation holds for small displacements from the minimum, larger displacements cause a breakdown in the model. The difference between the harmonic potential and the real potential can become large. Examining Figure 2.3, it is clear that anharmonicity, deviation from the harmonic model, becomes important both for weakly bound systems, where the interaction energy, D_e, is small and for high vibrational states, where the system is near dissociation.

In these cases, higher order potential terms affect the zero-point energies of systems and, most importantly, the wavefunctions. The wavefunction contains information about all observables. As such, anharmonicity can have a dramatic effect on the accuracy of calculations. For example, the bond lengths are often longer than those predicted by the harmonic model, making rotational constants smaller. Thus, anharmonicity can affect the calculation and fitting of rotational spectra. Furthermore, activation energies for reactions are calculated between the zero-point energies of the ground and transition states. Accurate anharmonic zero-point energies are needed to calculate precise activation energies and reaction rates.

A technique is required to improve on the harmonic eigenenergies and wavefunctions. In the proceeding chapter, a general method for calculating vibrational anharmonicity is developed and discussed.

Chapter 3

Anharmonic Corrections

3.1 Introduction

In order to calculate the normal modes and normal coordinates, it is assumed that the displacement of the atoms from their equilibrium positions is small, rendering the third and higher order terms in expansion 2.10 negligible. While the harmonic approximation works well for small displacements in strongly bound molecules, it breaks down for higher vibrational states and weakly bound systems, such as van der Waals clusters (see Figure 2.3). In such systems, high order and mode-mode coupling terms in the potential expansion become important. A configuration interaction technique is chosen to evaluate their effect on the wavefunctions and vibrational properties.

3.2 Configuration Interaction

Configuration interaction, CI, is a variational technique widely used in electronic structure theory.[184,210] It can improve on self-consistent field methods by including electron correlation effects.

In electronic structure applications, a configuration is, traditionally, a set of orbital occupancies. The CI method generates improvement to a wavefunction by constructing a linear combination of, for example, hydrogenic or Hartree-Fock wavefunctions. This is the CI space. For example, n excited states, ψ_i, can be mixed into the ground state wavefunction, ψ_0.

$$\Psi^{\mathrm{CI}} = c_0\psi_0 + \sum_{i=1}^{n} c_i\psi_i \tag{3.1}$$

The Variational Theorem states that an expectation value for the energy is always an upper bound to the exact, non-relativistic ground state energy. Consequently, the lower the ground state energy of the trial CI wavefunction, the closer it must be to the true ground state energy. Thus, in electronic structure theory, the aim is to find a set of coefficients, c_i $i = 0, \ldots, n$, that minimize the expectation value of the electronic energy with respect to the CI wavefunction. In practice, this minimisation is achieved through diagonalisation of a matrix, $\mathbf{H}$, with elements

$$\mathbf{H}_{ij} = \langle \psi_i | \hat{\mathscr{H}} | \psi_j \rangle \tag{3.2}$$

yielding the coefficients, c_i, as the eigenvectors and the energies as the eigenvalues.

One of the great strengths of CI is its generality. It can be used for any problem that can be defined by a series of basis functions. Its advantage over quantum Monte Carlo, QMC, techniques is its speed, its ease of application to a range of systems and the fact that the calculation produces excited state energies and wavefunctions. Its weakness is poor scaling. As more atoms are included, the size of calculation required to include all excited states rapidly increases. This is a well-known problem in electron correlation studies. Much study has been directed towards making the CI expansion as short as possible.[211–219]

In this study, the CI approach is adapted to calculate anharmonicity in the vibrational motion of biomolecules and clusters. Thus, basis functions, configurations and the expectation value to be minimised must be tailored to these polyatomic problems.

3.3 The Basis Functions

Analogy with electronic structure CI can clarify the difference between configurations and basis functions. In this work, a basis function describes the vibrational wavefunction for an individual vibrational mode. These are analogous to the single orbital hydrogenic AOs in electronic structure theory. The configuration is the full vibrational state of the system, analogous to the full set of orbital occupancies in electronic structure CI. In this study, the configuration is a product of all the

individual basis functions. Note that configurations can be correctly termed the basis functions of a CI calculation but this nomenclature is avoided as it can cause confusion.

The choice of basis function in CI cannot influence the final result of a converged calculation. They can, however, greatly affect the size, and therefore speed, of the calculation required for convergence. For example, in electronic structure CI, different basis functions are routinely used. The obvious choice for the basis would be the hydrogenic Slater type orbitals, STOs.[184] Gauss type orbitals, GTOs, are not such a good description of the actual orbitals so a larger basis is needed to converge the calculation. Nevertheless, GTOs are often chosen to simplify the integrals that must be evaluated. Often, expansion functions, such as pair natural orbitals, are used as the basis. These orbitals are harder to evaluate and non-orthogonal but are optimised, reducing the size of the basis needed to reach convergence.[211–219] Another technique is to perform iterative super-CI calculations.[220] The first iteration is performed with the chosen basis function, successive iterations are then performed substituting the calculated CI wavefunction for the basis. Furthermore, in electronic structure CI, techniques have been developed to aid selection of the best set of excited wavefunctions, the CI space, to achieve results close to full CI with a far smaller expansion. For example, symmetry considerations, the single double CI (CISD)[221–223] method and the frozen core approximation[224] can be used to shrink the CI space.

	Configuration Interaction Technique	
	Electronic Structure CI	Vibrational CI
Basis Functions	Slater/Gauss Atomic Orbitals	Morse Oscillator Wavefunctions
	Molecular Orbitals	Normal Modes
	Pair Natural Orbitals	Local Modes
Configurations	Slater Determinants	Product of Basis Functions

Table 3.1: Illustrative examples of basis functions and configurations for electronic structure and vibrational CI.

In variational and perturbation theory studies of molecular vibrations, various basis functions have been used, such as local modes,[90] normal modes,[225] self consistent field wavefunctions,[79,133,134,137] and functions tailored to the symmetry of the system.[93] Examples of basis functions and configurations in electronic structure and vibrational CI are provided in Table 3.1.

In this study, the basis functions selected are normal modes. Their choice is rationalised by their properties:

- Standard, simple and efficient to calculate, as described in Section 2.6. The technique developed in this chapter must be general and readily applicable to all systems. Normal modes fit this criteria. They can also be applied as a good starting point to calculate the low-lying vibrational states of biomolecules.

- Orthogonal; therefore, there must be a Gauss-type numerical integration technique available to evaluate the CI integrals (Section 3.5). Furthermore, the orthonormal character of normal modes reduces the number of integrals that must be evaluated.

- Widely used in literature as a way of calculating vibrational frequencies. They are a good starting point as CI will improve on the literature values. Furthermore, comparison of basis and CI vibrational energies gives a clear indication of the anharmonicity of the state.

CI calculations in this study will therefore be performed on configurations comprising of a product of normal modes basis functions.

3.3.1 Basis Set Notation

In this section, a notation to enable succinct discussion of the basis and CI configurations is presented. Individual normal mode wavefunctions, defined in Equation 2.30, are denoted by a small ψ with a subscript describing the mode (highest energy first) and a superscript describing the vibrational quantum number. For example, 5 quanta of energy in the bending mode of water will be ψ_3^5 in this notation.

Basis functions describe the state of individual vibrational modes. A configuration, however, is a description of the whole system in a particular vibrational state. Con-

sequently, it is a full vibrational wavefunction of the system, Ψ; a product of the wavefunctions of the individual normal modes.

$$\Psi = \prod_i \psi_i(q_i) \tag{3.3}$$

In the present labelling scheme, the configurations are split into four distinct groups. They are referred to as the ground state, Ψ^0, the singles, Ψ_i^S, the doubles, Ψ_{ij}^D and the triples, Ψ_{ijk}^T. It is important to note that the terms single, double and triple refer to the number of excited modes in the configuration. S, D and T make no reference to the vibrational quantum numbers of the basis functions. For example, a configuration with mode 4 in the sixth vibrationally excited state, but with all other modes in the ground state, is referred to as a single. In general, the SDT configuration types can be defined:

$$\Psi^0 = \prod_i \psi_i^0(q_i) \tag{3.4}$$

$$\Psi_i^S = \sum_{i=1}^{3N-6} \sum_{l=1}^{l_{\max}} \psi_i^l(q_i) \prod_{i' \neq i} \psi_{i'}^0(q_{i'}) \tag{3.5}$$

$$\Psi_{ij}^D = \sum_{i=1}^{3N-7} \sum_{j>i}^{3N-6} \sum_{l=1}^{l_{\max}} \sum_{m=1}^{m_{\max}} \psi_i^l(q_i)\psi_j^m(q_j) \prod_{i' \neq i,j} \psi_{i'}^0(q_{i'}) \tag{3.6}$$

$$\Psi_{ijk}^T = \sum_{i=1}^{3N-8} \sum_{j>i}^{3N-7} \sum_{k>j}^{3N-6} \sum_{l=1}^{l_{\max}} \sum_{m=1}^{m_{\max}} \sum_{n=1}^{n_{\max}} \psi_i^l(q_i)\psi_j^m(q_j)\psi_k^n(q_k) \prod_{i' \neq i,j,k} \psi_{i'}^0(q_{i'}) \tag{3.7}$$

The ground configuration of the system is the product of the ground basis functions of all normal modes. Note that there is only one ground configuration but $l_{\max} \times (3N-6)$ singles and $\frac{1}{2} l_{\max} m_{\max} \times (3N-6)(3N-5)$ doubles. In an accurate expansion of all vibrational states, there are clearly many more possible configurations than

those that fall under S, D and T types. In only including these configurations in the CI calculation, an approximation analogous to single-double-triple electronic structure CI, CISDT, is made. It is assumed that the inclusion of configurations higher in energy than triples is not needed to achieve accurate description of ground and low energy states.[221–223]

When presenting the theory of CI in the next section, configurations are labelled in order of energy, with ψ_i corresponding to the ith excited state. Thus, assuming $M+1$ configurations in total:

$$\begin{aligned}
\psi_0 &= \prod_i \psi_i^0 \\
\psi_1 &= \psi_1^1 \prod_{j \neq 1} \psi_j^0 \\
\psi_2 &= \psi_2^1 \prod_{j \neq 2} \psi_j^0 \\
&\vdots \\
\psi_M &= \psi_i^l \psi_j^m \psi_k^n \prod_{i' \neq i,j,k} \psi_{i'}^0
\end{aligned} \tag{3.8}$$

3.4 The Secular Equations

The configurations describe the vibrational wavefunctions of the system.[184,226] In a CI calculation an attempt is made to improve on the configuration wavefunction by taking a linear combination of $M+1$ ground and excited state configurations. For

example, in the notation described in Equation 3.8, the CI ground state and first excited states are

$$\Psi_0 = c_{00}\psi_0 + c_{01}\psi_1 + \ldots + c_{0M}\psi_M \tag{3.9}$$

$$\Psi_1 = c_{10}\psi_0 + c_{11}\psi_1 + \ldots + c_{1M}\psi_M \tag{3.10}$$

These are the trial CI wavefunctions. In the course of the calculation, the coefficients, c_{ab}, are variables that are optimised such that the linear combination resembles the true wavefunction more closely.

The Schrödinger equation can be solved for each linear combination, resulting in $M+1$ equations.

$$\hat{\mathscr{H}}\psi_i = E_i\psi_i \tag{3.11}$$

$$(\hat{\mathscr{H}} - E_i)\psi_i = 0 \qquad i = 0, 1, \ldots, M \tag{3.12}$$

Solving these simultaneous equations yields $M+1$ eigenvalues, E_i, and $M+1$ eigenvectors, Ψ_i. However, this solution is only non-trivial if the secular determinant disappears i.e.

$$\left|\hat{\mathscr{H}} - E_i\mathbf{1}\right| = 0 \tag{3.13}$$

The eigenvectors form a $(M+1) \times (M+1)$ matrix, $\mathbf{X}$

$$\mathbf{X} = (\Psi_0, \Psi_1, \ldots, \Psi_M) = \begin{pmatrix} c_{00}\psi_0 & c_{10}\psi_0 & \ldots & c_{M0}\psi_0 \\ c_{01}\psi_1 & c_{11}\psi_1 & \ldots & c_{M1}\psi_1 \\ \vdots & \vdots & \ddots & \vdots \\ c_{0M}\psi_M & c_{1M}\psi_M & \ldots & c_{MM}\psi_M \end{pmatrix} \tag{3.14}$$

and a diagonal matrix, **E**, may be formed from the eigenvalues, E_i

$$\mathbf{E} = \sum_{i=0}^{M} E_i \delta_i = \begin{pmatrix} E_0 & 0 & \dots & 0 \\ 0 & E_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & E_M \end{pmatrix} \tag{3.15}$$

Forming the matrix, **H**, where

$$\mathbf{H} = \sum_{i=0}^{M} \sum_{j=0}^{M} \langle \psi_i | \hat{\mathscr{H}} | \psi_j \rangle \tag{3.16}$$

means that the Schrödinger Equations in 3.12 may be written more concisely as

$$\mathbf{HX} = \mathbf{XE} \tag{3.17}$$

Forming the inverse of **X**, $\mathbf{X}^{-1}$, allows a similarity transformation

$$\mathbf{X}^{-1}\mathbf{XE} = \mathbf{E} = \mathbf{X}^{-1}\mathbf{HX} \tag{3.18}$$

which makes **H** diagonal as **E** is diagonal. The matrix **X** that causes $\mathbf{X}^{-1}\mathbf{HX}$ to be diagonal must be found in order to solve the eigenvalue Equation 3.11. This is achieved in practice by diagonalising the CI matrix, **H**.

Variational improvements in the energy yield improvements in the approximate wavefunction, which in turn improves the expectation values of all other properties which are dependent on it. However, these other properties may not converge as fast as the energy with respect to the basis set. Usually, coefficients are found that minimise the energy of the system. It is, however, possible to minimise any expectation value to speed convergence of a given property. In the next section, the choice of expectation value to be minimised in the CI calculation is considered.

3.4.1 The Expectation Value to be Minimised

From equations 2.3, 2.4 and 2.16, the general vibrational Hamiltonian of a system with an anharmonic potential, $\hat{\mathbf{V}}$, can be defined:

$$\hat{\mathscr{H}} = -\frac{1}{2}\sum_{i=1}^{3N-6}\frac{\partial^2}{\partial q_i^2} + \hat{\mathbf{V}} \tag{3.19}$$

The expectation value to be minimised can be adapted to the problem at hand. The main interest in this study is the deviation from the harmonic approximation, i.e. the anharmonicity of the system. It is convenient to write the Hamiltonian thus

$$\hat{\mathscr{H}} = \underbrace{\sum_{i=1}^{3N-6}\left(-\frac{1}{2}\frac{\partial^2}{\partial q_i^2} + \frac{1}{2}kq_i^2\right)}_{\hat{\mathbf{B}}} + \underbrace{\hat{\mathbf{V}} - \sum_{i=1}^{3N-6}\frac{1}{2}kq_i^2}_{\hat{\mathbf{A}}} \tag{3.20}$$

The separation of the Hamiltonian into the operators $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ can be thought of as a splitting of the anharmonic effect into two distinct parts. Operator $\hat{\mathbf{B}}$ only acts along the diagonal elements of the matrix, giving a zero point energy. The main work in the calculation is evaluating the matrix elements of $\hat{\mathbf{A}}$. The matrix elements of $\hat{\mathbf{A}}$ show how the harmonic basis functions in the configuration differ from the true anharmonic wavefunctions.

Figure 3.1 is a 3D visualisation of the CI technique for calculating anharmonicity in the vibrational states of the water monomer system studied in Section 5.3. The expectation value of operator $\hat{\mathbf{A}}$ is the anharmonicity, i.e. the difference red-grey surface. In the diagram, the ground state wavefunction is plotted in blue. It is

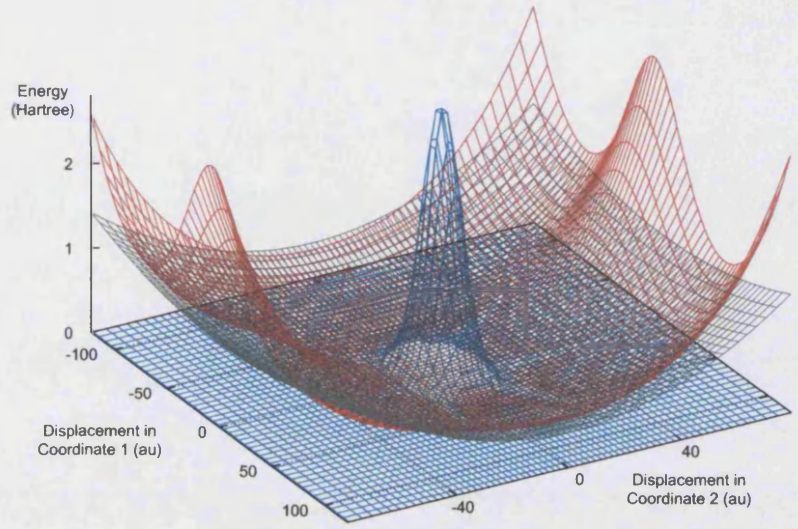


Figure 3.1: 2D visualisation of the CI technique. The red surface the Hoy, Mills and Strey PES.[99] The grey surface is the potential approximated to a harmonic function and the blue surface is the ground state wavefunction.

concentrated near the minimum, so the atoms have a low probability of being found in the highly anharmonic regions, where the difference between the red and grey surfaces is large. The expectation value of $\hat{\mathbf{A}}$ is, therefore, small. Higher energy wavefunctions are more diffuse. The atoms are more likely to be found far from the equilibrium geometry. Therefore, the expectation value of $\hat{\mathbf{A}}$ will be larger.

3.4.2 The Expectation Value of the Operator $\hat{\mathbf{B}}$

The operator $\hat{\mathbf{B}}$ is the Hamiltonian for a simple harmonic oscillator, SHO. As such, the normal mode basis functions are eigenfunctions of $\hat{\mathbf{B}}$.

$$\hat{\mathbf{B}}\psi = E\psi \tag{3.21}$$

In analogy to Equation 3.16, a matrix $\mathbf{B}$ can be defined for an $M+1$ configuration calculation

$$\mathbf{B} = \sum_{i=0}^{M}\sum_{j=0}^{M} \langle \psi_i | \hat{\mathbf{B}} | \psi_j \rangle \tag{3.22}$$

where the individual configurations are defined in Equation 3.8. $\mathbf{B}$ is diagonal, as $\hat{\mathbf{B}}$ is a single particle operator, not mixing modes, and the normal modes are orthogonal.

$$\begin{aligned}\mathbf{B} &= \sum_{i=0}^{M}\sum_{j=0}^{M} \langle \psi_i | \hat{\mathbf{B}} | \psi_j \rangle \\ &= \sum_{i=0}^{M}\sum_{j=0}^{M} \langle \psi_i | \psi_j \rangle E_j \\ &= \sum_{i=0}^{M}\sum_{j=0}^{M} \delta_{i,j} E_j \end{aligned} \tag{3.23}$$

In fact, $\mathbf{B}$ can be evaluated analytically:

$$\langle \psi_i^k | \hat{\mathbf{B}} | \psi_i^l \rangle = \begin{cases} 0 & \text{if } k \neq l, \\ (k+\frac{1}{2})\omega_i & \text{if } k = l \end{cases} \tag{3.24}$$

ω_i being the normal frequency of mode i. Generally, the diagonal matrix elements of the operator $\hat{\mathbf{B}}$ are

$$\langle \psi_i^l \psi_j^m \prod_{k\neq i,j} \psi_k^0 | \hat{\mathbf{B}} | \psi_i^l \psi_j^m \prod_{k\neq i,j} \psi_k^0 \rangle = (l+\frac{1}{2})\nu_i + (m+\frac{1}{2})\nu_j + \sum_{k\neq i,j} \frac{1}{2}\nu_k \tag{3.25}$$

3.4.3 The Expectation Value of the Operator $\hat{\mathbf{A}}$

Operator $\hat{\mathbf{A}}$, defined above in Equation 3.20, contains terms that can couple all configurations, symmetry permitting. Thus, matrix $\mathbf{A}$, where

$$\mathbf{A} = \sum_{i=0}^{M} \sum_{j=0}^{M} \langle \psi_i | \hat{\mathbf{A}} | \psi_j \rangle \tag{3.26}$$

can have non-zero elements on and off the diagonal.

In the general case of a N atom system, the expectation value of $\hat{\mathbf{A}}$ is calculated as a $3N-6$ dimensional integral. It can be expanded as

$$\langle \psi_1^1 \psi_2^0 \cdots \psi_{3N-6}^0 | \hat{\mathbf{A}} | \psi_1^0 \psi_2^1 \cdots \psi_{3N-6}^0 \rangle =$$
$$\int_{3N-6} \cdots \int \psi_1^1 \psi_2^0 \cdots \psi_{3N-6}^0 \left(\hat{\mathbf{V}} - \sum_i \frac{1}{2} k_i q_i^2 \right) \psi_1^0 \psi_2^1 \cdots \psi_{3N-6}^0 \, \mathrm{d}q_1 \mathrm{d}q_2 \ldots \mathrm{d}q_{3N-6}$$

In the next section, the methods used to evaluate these integrals are introduced.

3.5 Numerical Integration

The major computational work in the CI calculation is the evaluation of the integrals of matrix $\mathbf{A}$, defined in Equation 3.26. In simple systems, these can be calculated analytically. In general, however, the integrals have no analytical solution. A technique for calculating the matrix elements numerically is required.

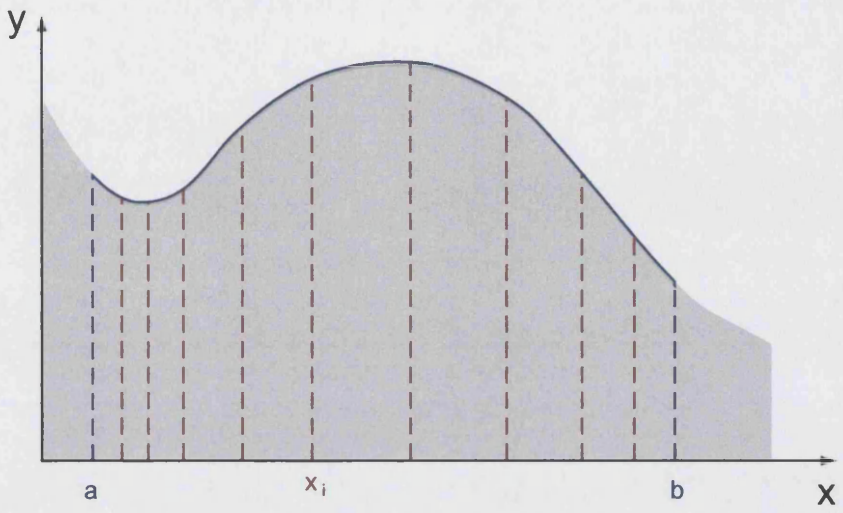


Figure 3.2: Gauss-Hermite Numerical Integration. Note the similarity to Simpson's Rule.[203,227]

Numerical integration,[203,227,228] or quadrature, is needed as the integrals of some functions cannot be evaluated analytically. An example is Simpson's rule, involving the separation of the function into smaller equidistant segments, divided at the abscissae, see Figure 3.2. The aim is to evaluate the integral as accurately as possible, calculating the weighted sum of the integrand at a sequence of abscissae.

$$\int_a^b f(x)\mathrm{d}x \approx \sum_{i=1}^{N} w_i f(x_i) \tag{3.27}$$

The weight of integrand i is w_i and there are N abscissae x_i. For greater efficiency this is done using the smallest number of function evaluations possible, i.e. the number of abscissae, N, is kept to a minimum.

In general, the integration can be closed, where the boundary points of the integrals, a and b, are taken as abscissae, or open, when the function takes some value that is hard to compute at the boundary points (e.g. $f(a)$ zero or infinity). The step size can be constant (as in Simpson's rule) or variable/adaptive.

For techniques with equally spaced abscissae, the only variable is the weighting, w_i, in the sum. By changing this according to the function to be integrated, $f(x)$, the accuracy of the integral can be improved for the same number of abscissae. These integrals can be exact for polynomial functions $f(x) = \sum_i c_i x^i$. The higher the order of $f(x)$, the more abscissae required for the integration to be exact. Quadratures are said to have order dependent on the highest order polynomial for which the technique is exact.

Variable stepsize gives an extra degree of freedom by allowing a choice of the position of the abscissae. The order of such quadratures can be up to twice that of constant stepsize techniques. It should be noted that higher order only means higher accuracy if the integrand can be well approximated by a polynomial.

The CI integrals of interest include wavefunctions of the form

$$\psi_i^l(x_i) = N_l H_l(x_i) e^{-\frac{1}{2}\alpha_i^2 x_i^2} \tag{3.28}$$

The exponential term means they are not well approximated by a polynomial. However, a feature of Gaussian quadrature is that the weights and abscissae can be arranged such that the integral is exact for a polynomial times a specific function, $g(x)$.

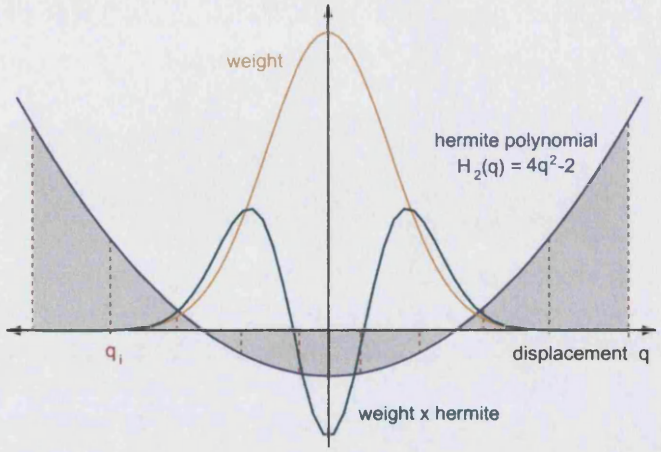


Figure 3.3: Gauss-Hermite Numerical Integration. Second order Hermite polynomial and the weight function multiply together to reproduce the wavefunction.

The form of the vibrational wavefunction of a simple harmonic oscillator is given in Equation 3.28. In this case $g(x) = e^{-x^2}$. The quadrature required is, therefore, a Gauss-Hermite where the approximation

$$\int_a^b e^{-x^2} f(x) \mathrm{d}x \approx \sum_{i=1}^{N} w_i f(x_i) \tag{3.29}$$

can be exact for a given set of weights and abscissae. Assuming $f(x)$ is an order m polynomial, the integral is exact for $\frac{1}{2}(m+1)$ abscissae. Thus, m abscissae allow exact integration of a polynomial function of order $2m+1$.

The weights are of the form e^{-ax^2} which, when multiplied by the Hermite polynomial of the appropriate order, result in a function replicating the wavefunction (see Figure 3.3.) The overlap integral between two doubly excited wavefunctions, like those shown, can be obtained exactly using only 3 abscissae.

In evaluating the CI matrix, **H**, most of the computational effort is expended on the calculation of the expectation value of the anharmonicity operator, $\hat{\mathbf{A}}$. In the following section, the integrals that must be evaluated to calculate **A** for 1D, 3D and general $3N-6$D systems are discussed.

3.5.1 The Expectation Value of $\hat{\mathbf{A}}$ - Model Systems

3.5.1.1 1D System

Consider, first, a 1D oscillator. The matrix elements of the CI matrix **A** are one dimensional. For example

$$\langle \psi_1^0 | \, \hat{\mathbf{A}} \, | \psi_1^1 \rangle = \int_{-\infty}^{\infty} \psi_1^1 \left(\hat{\mathbf{V}} - \frac{1}{2} k_1 q_1^2 \right) \psi_1^0 \, \mathrm{d}q_1 \tag{3.30}$$

3.5.1.2 3D System

Here, the calculation is extended to a 3D system. The matrix elements of **H** are therefore three dimensional. e.g.

$$\langle \psi_1^1 \psi_2^0 \psi_3^0 | \, \hat{\mathbf{A}} \, | \psi_1^0 \psi_2^1 \psi_3^1 \rangle = \iiint \psi_1^1 \psi_2^0 \psi_3^0 \left(\hat{\mathbf{V}} - \sum_{i=1}^{3} \frac{1}{2} k_i q_i^2 \right) \psi_1^0 \psi_2^1 \psi_3^1 \, \mathrm{d}\tau \tag{3.31}$$

3.5.1.3 General (3*N*-6)D System

The matrix elements of the CI matrix will require, in general, $3N - 6$D integrals. Calculation of these multi-dimensional integrals for small systems is feasible. For larger systems such as biomolecules, however, the integrals become too computationally expensive as they must evaluated numerically through the use of nested loops, as described in Section 3.5. As the number of dimensions increases, so do the number of nested loops. The time required for the calculation of a single CI matrix element therefore scales as a^{3N-6} where a is the number of abscissae in the Gauss quadrature. This is, clearly, a very poor scaling. A reduction in the scaling is needed to prevent the calculation from becoming impossibly large for all but the smallest systems. In the following section, a method to achieve such a reduction in integral size is presented.

3.6 The Pairwise Potential Approximation

3.6.1 Introduction

In order to reduce the order of the integrals, it is important to consider the potential energy surface. Consider the expansion of the potential of a general $3N-6$D system in normal coordinate space:

$$\hat{\mathrm{V}} = \sum_i f(q_i) + \sum_i \sum_j f(q_i, q_j) + \sum_i \sum_j \sum_k f(q_i, q_j, q_k) + \ldots + \sum_1 \cdots \sum_{3N-6} f(q_1, \ldots, q_{3N-6}) \tag{3.32}$$

It can be assumed that three (and higher) body terms in the potential are negligible i.e. the potential is approximated to be pairwise. This is the pairwise potential approximation, PPA. Making the PPA is, in effect, the assumption that the coupling of three or more normal modes is negligibly small. The approximation is good for low energy eigenstates and is tested in Section 5.3.3.

The PPA is effective at reducing the time taken to perform a CI calculation as it results in a reduction of the general $3N-6$ dimensional integral into a sum of 1D and 2D integrals. Consider, first, this reduction in the model three mode system described above. Removing three-body terms from the general potential 3.32 yields the pairwise potential:

$$\hat{\mathrm{V}} = \sum_i f(q_i) + \sum_i \sum_j f(q_i, q_j) \tag{3.33}$$

which for a three mode system expands to

$$\hat{\mathbf{V}} = \sum_i b_1^i q_1^i + \sum_i b_2^i q_2^i + \sum_i b_3^i q_3^i + \sum_i \sum_j b_{1,2}^{i,j} q_1^i q_2^j + \sum_i \sum_j b_{1,3}^{i,j} q_1^i q_3^j + \sum_i \sum_j b_{2,3}^{i,j} q_2^i q_3^j \tag{3.34}$$

Note that the notation here is different to that used in describing wavefunctions. In this labelling scheme, q_l^i, subscript, l, defines the normal coordinate as in the labelling of wavefunctions. However, the superscript, i, describes the power to which the coordinate is raised i.e. squared, cubed etc. For the potential coefficients, b_l^i, however, both sub- and superscript are purely labels. $\hat{\mathbf{A}}$, defined as the anharmonicity operator in Equation 3.20, can be expressed in terms of the normal coordinates, in analogy to the expanded potential.

$$\begin{aligned} \hat{\mathbf{A}} &= \hat{\mathbf{V}} - \sum_{i=1}^{3} \frac{1}{2} k_i q_i^2 \\ &= \hat{\mathbf{V}} - \frac{1}{2}\left(k_1 q_1^2 + k_2 q_2^2 + k_3 q_3^2\right) \end{aligned} \tag{3.35}$$

Clearly, the full anharmonicity operator, $\hat{\mathbf{A}}$, can be expressed as a pairwise expansion in normal coordinates.

$$\hat{\mathbf{V}} = \sum_i c_1^i q_1^i + \sum_i c_2^i q_2^i + \sum_i c_3^i q_3^i + \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j + \sum_i \sum_j c_{1,3}^{i,j} q_1^i q_3^j + \sum_i \sum_j c_{2,3}^{i,j} q_2^i q_3^j \tag{3.36}$$

Note the similarity between Equations 3.34 and 3.36. This is due to inclusion of the harmonic expansion in the coefficients c_l^i. In fact, all except the quadratic force constant coefficients of the potential, i.e. b_l^2, are identical to c_l^i:

$$c_l^i = b_l^i \quad ; \quad i \neq 2 \tag{3.37}$$

$$c_l^2 = b_l^2 - k_l \tag{3.38}$$

The CI matrix integrals for the model 3D system can be generalised thus

$$\begin{aligned}
&\langle \psi_1\psi_2\psi_3 | \sum_i c_1^i q_1^i + \sum_i c_2^i q_2^i + \sum_i c_3^i q_3^i + \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j \\
&\quad + \sum_i \sum_j c_{1,3}^{i,j} q_1^i q_3^j + \sum_i \sum_j c_{2,3}^{i,j} q_2^i q_3^j | \psi_1\psi_2\psi_3 \rangle \\
&\quad = \langle \psi_1 | \sum_i c_1^i q_1^i | \psi_1 \rangle \langle \psi_2\psi_3 | \psi_2\psi_3 \rangle \\
&\quad + \langle \psi_2 | \sum_i c_2^i q_2^i | \psi_2 \rangle \langle \psi_1\psi_3 | \psi_1\psi_3 \rangle \\
&\quad + \langle \psi_3 | \sum_i c_3^i q_3^i | \psi_3 \rangle \langle \psi_1\psi_2 | \psi_1\psi_2 \rangle \\
&\quad + \langle \psi_1\psi_2 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j | \psi_1\psi_2 \rangle \langle \psi_3 | \psi_3 \rangle \\
&\quad + \langle \psi_1\psi_3 | \sum_i \sum_j c_{1,3}^{i,j} q_1^i q_3^j | \psi_1\psi_3 \rangle \langle \psi_2 | \psi_2 \rangle \\
&\quad + \langle \psi_2\psi_3 | \sum_i \sum_j c_{2,3}^{i,j} q_2^i q_3^j | \psi_2\psi_3 \rangle \langle \psi_1 | \psi_1 \rangle
\end{aligned} \tag{3.39}$$

Depending on whether the configurations are singles or doubles, some of the overlap integrals above will turn out to be zero, reducing the complexity of the integral. This is covered in greater detail in the following section.

3.6.2 Overcounting Integrals

Four types of configuration are defined in Section 3.3. They are, the ground state, Ψ^0; singles, Ψ^S; doubles, Ψ^D; and triples. There will, therefore, be ten different types of integral. Examples of six of these types are presented in the proceeding sections. The matrix elements are discussed for all possible permutations and combinations of CI matrix elements for the ground, singles and doubles. On making the pairwise potential approximation, there is a possibility of overcounting integrals. Where this occurs, it is highlighted and rectified.

3.6.2.1 Diagonal Elements

The diagonal of the CI matrix, $\mathbf{H}$, is a special case as the expectation value of operator $\hat{\mathbf{B}}$ is non-zero for such elements. Consider the following generalised diagonal CI matrix element.

$$\begin{aligned}
\langle \psi_1^1\psi_2^0\psi_3^0 | \hat{\mathbf{A}} | \psi_1^1\psi_2^0\psi_3^0 \rangle = {} & \langle \psi_1^1 | \sum_i c_1^i q_1^i | \psi_1^1 \rangle \langle \psi_2^0\psi_3^0 | \psi_2^0\psi_3^0 \rangle \\
& + \langle \psi_2^0 | \sum_i c_2^i q_2^i | \psi_2^0 \rangle \langle \psi_1^1\psi_3^0 | \psi_1^1\psi_3^0 \rangle \\
& + \langle \psi_3^0 | \sum_i c_3^i q_3^i | \psi_3^0 \rangle \langle \psi_1^1\psi_2^0 | \psi_1^1\psi_2^0 \rangle \\
& + \langle \psi_1^1\psi_2^0 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j | \psi_1^1\psi_2^0 \rangle \langle \psi_3^0 | \psi_3^0 \rangle \\
& + \langle \psi_1^1\psi_3^0 | \sum_i \sum_j c_{1,3}^{i,j} q_1^i q_3^j | \psi_1^1\psi_3^0 \rangle \langle \psi_2^0 | \psi_2^0 \rangle \\
& + \langle \psi_2^0\psi_3^0 | \sum_i \sum_j c_{2,3}^{i,j} q_2^i q_3^j | \psi_2^0\psi_3^0 \rangle \langle \psi_1^1 | \psi_1^1 \rangle \qquad (3.40)
\end{aligned}$$

The configuration consists of normalised wavefunctions. Integrals such as $\langle\psi_3^0|\psi_3^0\rangle$ are one. The integral simplifies to

$$\begin{aligned}
&\langle\psi_1^1\psi_2^0\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^1\psi_2^0\psi_3^0\rangle \\
&= \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^1\rangle + \langle\psi_2^0|\sum_i c_2^i q_2^i\,|\psi_2^0\rangle + \langle\psi_3^0|\sum_i c_3^i q_3^i\,|\psi_3^0\rangle \\
&+ \langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^1\psi_2^0\rangle + \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^1\psi_3^0\rangle \\
&+ \langle\psi_2^0\psi_3^0|\sum_i\sum_j c_{2,3}^{i,j} q_2^i q_3^j\,|\psi_2^0\psi_3^0\rangle
\end{aligned} \tag{3.41}$$

Note that when $i = 0$ or $j = 0$ in the three 2D integrals above, they reduce to 1D. For example,

$$\begin{aligned}
\langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,0} q_1^i q_2^0\,|\psi_1^1\psi_2^0\rangle &= \langle\psi_1^1|\sum_i\sum_j c_{1,2}^{i,0} q_1^i\,|\psi_1^1\rangle\,\langle\psi_2^0|\psi_2^0\rangle \\
&= \langle\psi_1^1|\sum_i\sum_j c_{1,2}^{i,0} q_1^i\,|\psi_1^1\rangle
\end{aligned} \tag{3.42}$$

This means that, including the actual 1D integrals, the effect of 1D terms in the potential, such as $\sum_i c_1^i q_1^i$, are counted three times, i.e. they are overcounted. Two of these 1D integrals must be taken away for each mode.

$$\begin{aligned}
&\langle\psi_1^1\psi_2^0\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^1\psi_2^0\psi_3^0\rangle \\
&= \langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^1\psi_2^0\rangle + \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^1\psi_3^0\rangle \\
&- \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^1\rangle - \langle\psi_2^0|\sum_i c_2^i q_2^i\,|\psi_2^0\rangle - \langle\psi_3^0|\sum_i c_3^i q_3^i\,|\psi_3^0\rangle \\
&+ \langle\psi_2^0\psi_3^0|\sum_i\sum_j c_{2,3}^{i,j} q_2^i q_3^j\,|\psi_2^0\psi_3^0\rangle
\end{aligned} \tag{3.43}$$

Thus, the 3D diagonal matrix elements reduce to three 2D integrals and three 1D integrals. In a larger system of P modes, there will be $\frac{1}{2}P(P+1)$ 2D integrals and P 1D integrals. This scales roughly $a^2 \times P^2$, far better than the full CI scaling of a^P, with a the number of quadrature points.

Operator $\hat{\mathbf{B}}$ is shown to be diagonal in Equation 3.23. Therefore

$$\mathbf{H}_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij} \quad ; \quad i = j \tag{3.44}$$

$$\mathbf{H}_{ij} = \mathbf{A}_{ij} \quad ; \quad i \neq j \tag{3.45}$$

As such, the following CI matrix elements, $\mathbf{H}$, are all expectation values of $\hat{\mathbf{A}}$.

3.6.2.2 Ground-Single $\langle \Psi^0 | \hat{\mathbf{A}} | \Psi^S \rangle$

Calculation of CI matrix elements of this type requires the evaluation of integrals between singly excited states and the ground state. For example, when the excited vibration is normal mode 1, and this mode has vibrational quantum number 1, the matrix element is

$$\begin{aligned}\langle\psi_1^1\psi_2^0\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^0\psi_3^0\rangle = &\langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle\,\langle\psi_2^0\psi_3^0|\psi_2^0\psi_3^0\rangle \\ &+\langle\psi_2^0|\sum_i c_2^i q_2^i\,|\psi_2^0\rangle\,\langle\psi_1^1\psi_3^0|\psi_1^0\psi_3^0\rangle \\ &+\langle\psi_3^0|\sum_i c_3^i q_3^i\,|\psi_3^0\rangle\,\langle\psi_1^1\psi_2^0|\psi_1^0\psi_2^0\rangle \\ &+\langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^0\rangle\,\langle\psi_3^0|\psi_3^0\rangle \\ &+\langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle\,\langle\psi_2^0|\psi_2^0\rangle \\ &+\langle\psi_2^0\psi_3^0|\sum_i\sum_j c_{2,3}^{i,j} q_2^i q_3^j\,|\psi_2^0\psi_3^0\rangle\,\langle\psi_1^1|\psi_1^0\rangle \end{aligned} \tag{3.46}$$

However, the basis functions, normal mode wavefunctions, are orthonormal. Therefore, all overlap integrals over mode 1 are zero. For example,

$$\langle\psi_1^1\psi_2^0|\psi_1^0\psi_2^0\rangle = 0 \tag{3.47}$$

All other overlap integrals are 1. Therefore, the integral reduces to

$$\begin{aligned}\langle\psi_1^1\psi_2^0\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^0\psi_3^0\rangle = &\langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle + \langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^0\rangle \\ &+\langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle \end{aligned} \tag{3.48}$$

As above, it is clear that when $j = 0$ the 2D integrals reduce to become 1D. The effect of $\sum_i c_1^i q_1^i$ terms in the potential will therefore be included three times so two of these 1D integrals must be subtracted.

$$\begin{aligned}\langle\psi_1^1\psi_2^0\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^0\psi_3^0\rangle &= \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle + \langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^0\rangle \\ &+ \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle - 2\,\langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle \\ &= \langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^0\rangle - \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle \\ &+ \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle \qquad (3.49)\end{aligned}$$

In the general P mode case, there will be $(P-1)$ 2D integrals and one 1D integral. Therefore, the PPA calculation will scale roughly $a^2 \times P$, much better than the full calculation.

3.6.2.3 Ground-Double $\langle\Psi^0|\mathbf{H}|\Psi^D\rangle$

Matrix elements describing the interactions between ground and doubly excited states result in integrals similar to that shown in Equation 3.50. In this example, modes 1 and 2 of the three mode system are excited to the second vibrational eigenstate.

$$\begin{aligned}\langle\psi_1^1\psi_2^1\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^0\psi_3^0\rangle &= \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle\,\langle\psi_2^1\psi_3^0|\psi_2^0\psi_3^0\rangle \\ &+ \langle\psi_2^1|\sum_i c_2^i q_2^i\,|\psi_2^0\rangle\,\langle\psi_1^1\psi_3^0|\psi_1^0\psi_3^0\rangle \\ &+ \langle\psi_3^0|\sum_i c_3^i q_3^i\,|\psi_3^0\rangle\,\langle\psi_1^1\psi_2^1|\psi_1^0\psi_2^0\rangle \\ &+ \langle\psi_1^1\psi_2^1|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^0\rangle\,\langle\psi_3^0|\psi_3^0\rangle \\ &+ \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle\,\langle\psi_2^1|\psi_2^0\rangle \\ &+ \langle\psi_2^1\psi_3^0|\sum_i\sum_j c_{2,3}^{i,j} q_2^i q_3^j\,|\psi_2^0\psi_3^0\rangle\,\langle\psi_1^1|\psi_1^0\rangle \end{aligned} \tag{3.50}$$

As the basis functions are orthonormal, the CI integral is reduced to one 2D integral.

$$\langle\psi_1^1\psi_2^1\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^0\psi_3^0\rangle = \langle\psi_1^1\psi_2^1|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^0\rangle \tag{3.51}$$

There is no overcounting of integrals in this case. In the general $3N-6$D system, matrix elements of this type also reduce to one 2D integral. As there is always only one 2D integral to evaluate there is no increase in computational effort as system size increases. Gauss-Hermite quadrature scaling is reduced from a^P to one for a system of P modes.

3.6.2.4 Single-Single $\langle\Psi^S|\,\hat{\mathbf{A}}\,|\Psi^S\rangle$

When calculating CI matrix elements between two singles, there will be three types of integral. When the excited mode is the same for both configurations, and they have the same vibrational quantum number, the integral will fall under the diagonal type

discussed above. However, when the vibrational quantum numbers are different, the integral will follow the logic shown in the $\langle \Psi^0 | \hat{\mathbf{A}} | \Psi^S \rangle$ case:

$$\langle \psi_1^1 \psi_2^0 \psi_3^0 | \hat{\mathbf{A}} | \psi_1^2 \psi_2^0 \psi_3^0 \rangle = \langle \psi_1^1 \psi_2^0 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j | \psi_1^2 \psi_2^0 \rangle - \langle \psi_1^1 | \sum_i c_1^i q_1^i | \psi_1^2 \rangle + \langle \psi_1^1 \psi_3^0 | \sum_i \sum_j c_{1,3}^{i,j} q_1^i q_3^j | \psi_1^2 \psi_3^0 \rangle \quad (3.52)$$

Another type of integral results when different modes in the configuration are excited. Consider, for example, the matrix element when mode 1 in the configuration and mode 2 in the second configuration are excited. The overlap integrals involving these modes are zero, reducing the matrix element to a simple 2D integral, as in the ground-double case.

$$\langle \psi_1^1 \psi_2^0 \psi_3^0 | \hat{\mathbf{A}} | \psi_1^0 \psi_2^1 \psi_3^0 \rangle = \langle \psi_1^1 \psi_2^0 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j | \psi_1^0 \psi_2^1 \rangle \quad (3.53)$$

3.6.2.5 Single-Double $\langle \Psi^S | \hat{\mathbf{A}} | \Psi^D \rangle$

There are three distinct scenarios in the single-double matrix elements. In the first, there is no overlap in the excited modes, such as when modes 1 and 2 are excited in Ψ^d and mode three is excited in Ψ^S

$$\begin{aligned}
\langle \psi_1^1\psi_2^1\psi_3^0 | \hat{\mathbf{A}} | \psi_1^0\psi_2^0\psi_3^1 \rangle = & \langle \psi_1^1 | \sum_i c_1^i q_1^i | \psi_1^0 \rangle \langle \psi_2^1\psi_3^0 | \psi_2^0\psi_3^1 \rangle \\
& + \langle \psi_2^1 | \sum_i c_2^i q_2^i | \psi_2^0 \rangle \langle \psi_1^1\psi_3^0 | \psi_1^0\psi_3^1 \rangle \\
& + \langle \psi_3^0 | \sum_i c_3^i q_3^i | \psi_3^1 \rangle \langle \psi_1^1\psi_2^1 | \psi_1^0\psi_2^0 \rangle \\
& + \langle \psi_1^1\psi_2^1 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j | \psi_1^0\psi_2^0 \rangle \langle \psi_3^0 | \psi_3^1 \rangle \\
& + \langle \psi_1^1\psi_3^0 | \sum_i \sum_j c_{1,3}^{i,j} q_1^i q_3^j | \psi_1^0\psi_3^1 \rangle \langle \psi_2^1 | \psi_2^0 \rangle \\
& + \langle \psi_2^1\psi_3^0 | \sum_i \sum_j c_{2,3}^{i,j} q_2^i q_3^j | \psi_2^0\psi_3^1 \rangle \langle \psi_1^1 | \psi_1^0 \rangle \qquad (3.54)
\end{aligned}$$

It is clear that all overlap integrals are zero since they involve single and ground state wavefunctions of the same mode. Thus, the overall integral is zero. This is a result of the PPA, the assumption that interaction between three normal modes is negligible.

However, when the mode excited in Ψ^S is the same as one of the modes excited in Ψ^D, such as mode 2 for example, the integral is non zero once more. If the vibrational quantum number of this mode is the same in both configurations, the integral resembles that from the $\langle \Psi^0 | \hat{\mathbf{A}} | \Psi^S \rangle$ case.

$$\begin{aligned}\langle\psi_1^1\psi_2^1\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^1\psi_3^0\rangle &= \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle\,\langle\psi_2^1\psi_3^0|\psi_2^1\psi_3^0\rangle \\ &+ \langle\psi_2^1|\sum_i c_2^i q_2^i\,|\psi_2^1\rangle\,\langle\psi_1^1\psi_3^0|\psi_1^0\psi_3^0\rangle \\ &+ \langle\psi_3^0|\sum_i c_3^i q_3^i\,|\psi_3^0\rangle\,\langle\psi_1^1\psi_2^1|\psi_1^0\psi_2^1\rangle \\ &+ \langle\psi_1^1\psi_2^1|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^1\rangle\,\langle\psi_3^0|\psi_3^0\rangle \\ &+ \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle\,\langle\psi_2^1|\psi_2^1\rangle \\ &+ \langle\psi_2^1\psi_3^0|\sum_i\sum_j c_{2,3}^{i,j} q_2^i q_3^j\,|\psi_2^1\psi_3^0\rangle\,\langle\psi_1^1|\psi_1^0\rangle \qquad (3.55)\end{aligned}$$

Checking for zero overlap matrix elements and removing extra 1D integrals:

$$\begin{aligned}\langle\psi_1^1\psi_2^1\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^1\psi_3^0\rangle &= \langle\psi_1^1\psi_2^1|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^1\rangle - \langle\psi_1^1|\sum_i c_1^i q_1^i\,|\psi_1^0\rangle \\ &+ \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j} q_1^i q_3^j\,|\psi_1^0\psi_3^0\rangle \qquad (3.56)\end{aligned}$$

The final scenario occurs when vibrational quantum numbers of mode 2 in the configuration are different. The expectation value of the anharmonicity operator, $\hat{\mathbf{A}}$, reduces to a single 2D integral, analogous to the $\langle\Psi^0|\,\hat{\mathbf{A}}\,|\Psi^D\rangle$ case above

$$\langle\psi_1^1\psi_2^1\psi_3^0|\,\hat{\mathbf{A}}\,|\psi_1^0\psi_2^2\psi_3^0\rangle = \langle\psi_1^1\psi_2^1|\sum_i\sum_j c_{1,2}^{i,j} q_1^i q_2^j\,|\psi_1^0\psi_2^2\rangle \qquad (3.57)$$

3.6.2.6 Double-Double $\langle\Psi^D|\,\hat{\mathbf{A}}\,|\Psi^D\rangle$

This type of integral can only be non-zero when the two modes excited are the same in both configurations, otherwise all overlap integrals will be zero. Assuming the

same modes, for example 1 and 2, are excited, there are two possible outcomes. Firstly, when both pairs of excited modes have different vibrational quantum numbers, the integral reduces to a single 2D integral.

$$\int = \langle \psi_1^1 \psi_2^1 \psi_3^0 | \, \hat{\mathbf{A}} \, | \psi_1^2 \psi_2^2 \psi_3^0 \rangle = \langle \psi_1^1 \psi_2^1 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j \, | \psi_1^2 \psi_2^2 \rangle \tag{3.58}$$

However, when only one pair of modes has a different number of quanta, the matrix element is reduced to a sum of 1 and 2D integrals

$$\begin{aligned} \int &= \langle \psi_1^1 \psi_2^1 \psi_3^0 | \, \hat{\mathbf{A}} \, | \psi_1^1 \psi_2^2 \psi_3^0 \rangle \\ &= \langle \psi_1^1 \psi_2^1 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_2^j \, | \psi_1^1 \psi_2^2 \rangle - \langle \psi_2^1 | \sum_i c_1^i q_1^i \, | \psi_2^2 \rangle \\ &+ \langle \psi_2^1 \psi_3^0 | \sum_i \sum_j c_{1,2}^{i,j} q_1^i q_3^j \, | \psi_2^2 \psi_3^0 \rangle \end{aligned} \tag{3.59}$$

Clearly, the integral is of the diagonal type when both pairs of modes have the same vibrational quantum number.

Part III of this report concerns the development of a program to implement the vibrational CI technique. Details of the inner working of the code are provided and the results of benchmark calculations are examined.

Part III

Development and Benchmarking

Chapter 4

Development

A computer program is developed to implement the vibrational CI theory presented in Part II. This chapter concerns the development of this program, and its functionality. In the following chapter the code is tested and calculations compared with results reported for benchmark systems.

4.1 Introduction

The code is developed entirely for this project, except for the sections concerned with finding the potential minimum, written by Wales.[229] It is written in FORTRAN 77 and is developed on a Silicon Graphics platform. A flow chart representation of the code's functionality is given in Figure 4.1.

The project aim is to design a general, adaptable method for calculating anharmonicity in biomolecular systems. Therefore, it is important that the code is as general and flexible as possible. It has been designed to be portable, so that it can be used on other systems, possibly with any FORTRAN compiler, although it has only been tested on SGI machines. With this aim in mind, the use of external subroutine calls has been limited. The only library calls made are to NAG subroutines for evaluation of weights and abscissae for Gauss-Hermite quadrature and for diagonalisation of real symmetric matrices. If required, implementation of non-library routines for these tasks from sources such as Numerical Recipes[203] is trivial, but the NAG libraries are chosen for their speed, having been specifically optimised for the processors used, and their theoretical accuracy, ease of use and consistent error-handling.

To make the program flexible, the code is based around a main routine which calls the separate functions and in which all parameters, such as the number of configurations and number of abscissae used in the Gauss-Hermite integration, may be set. All the subroutines and variables are labelled as clearly as possible and extensive comments are included as instructions and explanations.

The code is made as flexible as possible by a clear and complete separation of the potential from the main program. Separating the potential from the code, with no use of common memory, allows a variety of potential energy surfaces to be used from simple functional potentials such as the Hénon-Heiles surfaces to semi-

empirical molecular modelling potentials like MM3, CHARMM and AMBER. The sole interface is a wrapper subroutine that is adapted to each PES. The only input to the PES wrapper subroutine is an array, x, of length n containing the Cartesian coordinates of the atoms. Its only output is the potential energy at the nuclear configuration described by x. It would also be possible to calculate the surface points "on the fly" using density functional or ab initio methods. However, with the number of function calls presently required to perform a CI calculation, this would be a very slow process for all but the smallest molecules. Methods for reducing the number of calculations of potential points are introduced and implemented later and suggestions are made for further improvements.

Figure 4.1 is a graphical description of the way the code works. All of the operations listed are individual subroutine calls made from the main program so specification of computational variables, changes in running order and extra operations are easy to implement. Optional subroutines are shown in grey and the processes and theories involved in these are discussed in Section 6.3.

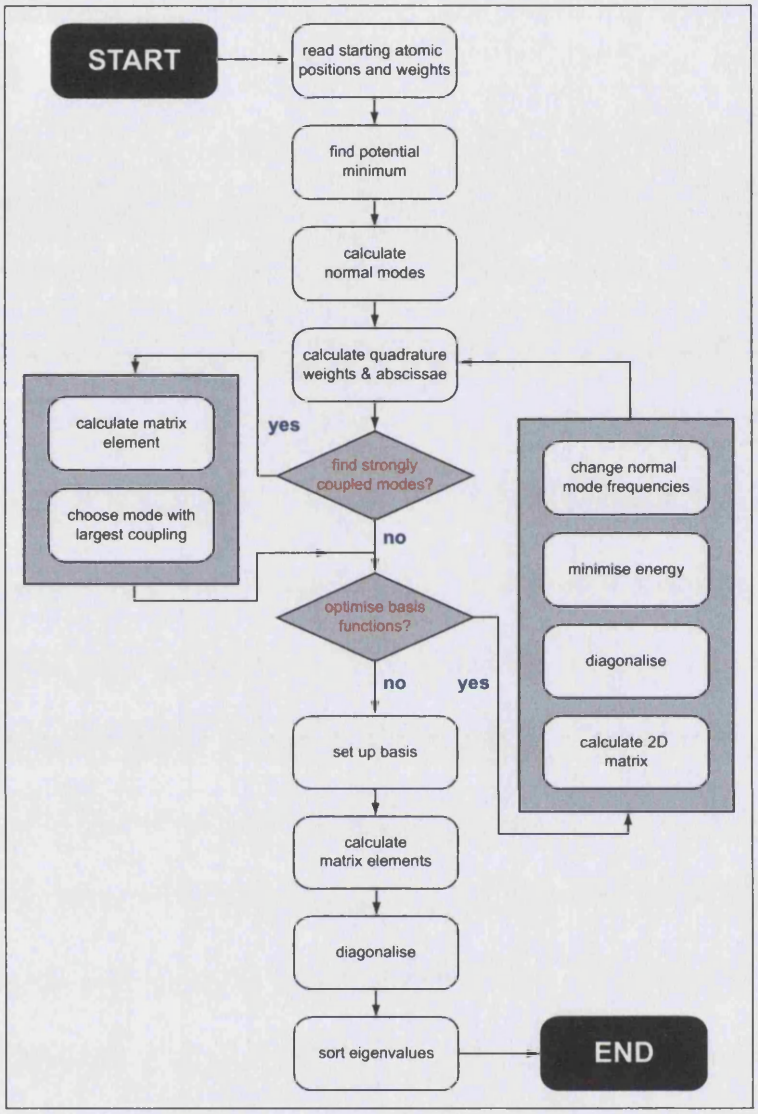


Figure 4.1: Schematic representation of functions of the computer code. Optional subroutines are shown in grey. All but the potential minimisation is developed for this project.

The most complex portions of the code are the search for a minimum in the potential, evaluation of the normal modes and the calculation of the CI matrix. The theory behind this is described in Sections 2.4 and 3.4. The calculation of the CI

matrix is the most time intensive section of the process with many calls made to the potential wrapper subroutine. Performing full CI calculations on many-body systems will quickly become unfeasible as the number of potential calls increases dramatically with the number of vibrational modes under consideration. To overcome this problem, the pairwise potential approximation, described in Section 3.6, is implemented and tested in Section 5.3.3.

The main areas of the program are discussed briefly in the following section. The results of rigorous testing and comparisons with benchmark calculations are detailed.

4.2 The Program

4.2.1 Minimisation

Three methods for finding a minimum on the potential energy surface were considered. The simplest is an iterative simplex method implemented via a NAG subroutine. This method requires many iterations to find the minimum and is therefore slow, but it is robust. Conjugate gradient and Newton-Raphson minimisations have also been included via the code of Wales.[229] These techniques are introduced in Section 2.5.

The accuracy of the minimisation is inherently tested in the calculation of the normal modes. This is achieved through diagonalisation of the mass weighted force constant matrix, defined in Equation 2.25, which will result in six zero and $3N-6$ real eigenvalues at the minimum. If the system has not been minimised correctly, some of the eigenvalues of the mass-weighted force constant matrix will be negative, providing an immediate diagnostic of the quality and accuracy of the minimisation process.

For larger systems, such as tryptamine, there are many minima. Techniques for finding all potential wells and the global minimum have been widely studied and implemented in black-box packages that are simple to use. The TINKER package, developed by Ponder and co-workers,[195–201] implements tools for finding minima of potentials for complex biosystems and allows exploration of minima other than the global minimum. As such, TINKER is used to find minimum geometries of interest in the study of biomolecular systems discussed in Chapter 6. These conformational searches are performed independently to the code developed for this research.

The minimisation method employed by TINKER is a semi-Newton technique, a modified limited-memory large-scale bound unconstrained optimisation, L-BFGS.[193,194] It is designed for problems with large numbers of variables and minimal storage. It finds the minimum over Cartesian coordinates, requiring storage proportional to the number of atoms in the system. The L-BFGS optimisation converges in a slow linear fashion near the minimum so it is a good preliminary

minimisation. It can be used in conjunction with another method, for example conjugate gradient, that displays quadratic convergence.

4.2.2 Calculation of Normal Modes

After minimisation of the structure, normal coordinates must be found. The first step in this is calculation of the Cartesian second derivative, i.e. force constant matrix. The differentiation technique employed is Ridders method,[203] a numerical technique for calculating first derivatives, utilising Neville's algorithm to extrapolate $\delta x \rightarrow 0$. The code can be extended by calculating first-derivatives of first-derivatives to obtain the second-derivatives required.

The force constant matrix is subsequently mass-weighted and diagonalised with a NAG subroutine.[227] The normal mode frequencies are the square roots of the eigenvalues. The eigenvector matrix is mass-weighted, so this must be removed to give the matrix for transforming from Cartesian to normal coordinate space (see Section 2.4).

The code is tested by calculating normal modes for simple, benchmark functional potentials such as the Bartlett *et al.*[230] fit of the Simons, Parr and Finlan water potential[231] and the Hoy, Mills and Strey[99] water potential. Results are compared with literature values in Table 5.3.

4.2.3 Gauss-Hermite Integration

Gauss-Hermite weights and abscissae are calculated using a NAG subroutine. A test of the integration method and the correct calculation of the normal mode basis functions is performed at the start of every calculation by checking the orthonormality of the excited states. The results are printed as an overlap matrix, **O**, of elements O_{lm}

$$O_{lm} = \langle \psi_i^l | \psi_i^m \rangle \tag{4.1}$$

More detailed testing is subsequently performed on the calculation of the matrix elements and the full CI procedure as outlined in the next section.

4.2.4 Matrix Element Calculation

The calculation of matrix elements can be tested by evaluating integrals with analytical solutions. Consider, for example, a 1D system of two particles on an arbitrary potential for which the harmonic frequency is found to be ω. An integral with an analytical solution is chosen. e.g.

$$\langle \Psi_v(q) | \frac{1}{2} q^2 | \Psi_v(q) \rangle \tag{4.2}$$

$$= \int N_v H_v(q) e^{-\frac{1}{2}q^2} \left(\frac{1}{2} q^2 \right) N_v H_v(q) e^{-\frac{1}{2}q^2} \mathrm{d}q \tag{4.3}$$

$$= \int \frac{N_v^2}{2} \left(H_v(q) \right)^2 q^2 e^{-q^2} \mathrm{d}q \tag{4.4}$$

Choosing v to be 1 means

$$(N_1)^2 = \frac{\omega}{2\pi^{\frac{1}{2}}} \tag{4.5}$$

$$(H_1(q))^2 = 4q^2 \tag{4.6}$$

so the integral becomes

$$\int \left(\frac{\omega}{\pi^{\frac{1}{2}}}\right) q^4 e^{-q^2} \mathrm{d}q \tag{4.7}$$

But, q^4 can be expressed in terms of $H_2(q)$, $H_1(q)$ and $H_0(q)$

$$q^4 = \frac{1}{16}\left[(H_2(q))^2 + 4(H_1(q))^2 - 4(H_0(q))^2\right] \tag{4.8}$$

so the integral can be expressed

$$\begin{aligned} &\int \left(\frac{\omega}{2\pi^{\frac{1}{2}}}\right) \left[\frac{1}{8}(H_2(q))^2 + \frac{1}{2}(H_1(q))^2 - \frac{1}{2}(H_0(q))^2\right] e^{-q^2} \mathrm{d}q \\ =&\frac{1}{2}\int N_2^2 (H_2(q))^2 e^{-q^2} \mathrm{d}q + \frac{1}{2}\int N_1^2 (H_1(q))^2 e^{-q^2} \mathrm{d}q \\ &- \frac{1}{4}\int N_0^2 (H_0(q))^2 e^{-q^2} \mathrm{d}q \end{aligned} \tag{4.9}$$

As the wavefunctions are normalised, all the three separate contributing integrals are 1 so the total integral is $\frac{3}{4}$. Calculating such integrals allows a check of the Gauss-Hermite integration procedure.

In the following chapter, the code is applied to benchmark 1D and 3D systems, and the pairwise potential approximation is incorporated into the model.

Chapter 5

Benchmark Systems

5.1 The Morse Oscillator

5.1.1 Introduction

The proceeding section involves testing the code on a one dimensional system to check that the theory has been correctly implemented. The system chosen is the Morse oscillator as implementation of the potential is trivial. The eigenvalues and eigenvectors are known, as the Schrödinger equation for this system can be solved analytically, and the system has been used in previous benchmark calculations.[232]

Morse[233] suggested in 1929 that an empirical potential energy function for diatomics was needed. The true potential energy depends on a complicated function of the interatomic distance, r, and the electronic quantum numbers. Therefore, it is expensive to calculate and difficult to implement. Until then, Taylor series type potentials had been used but Morse proposed that a potential should fulfil four criteria:

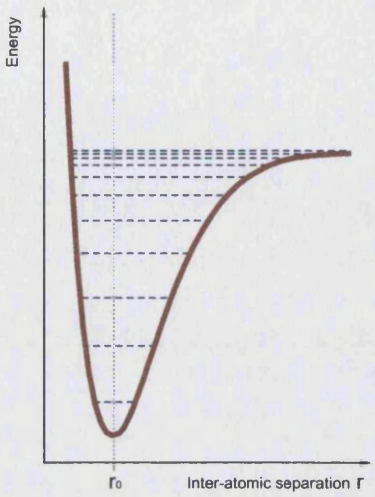


Figure 5.1: The Morse potential with eigenstates.

1. It should come asymptotically to a finite value as $r \rightarrow \infty$.

2. It should have only one minimum at $r = r_0$.

3. It should become infinity (or very large) at $r = 0$.

4. It must give energy levels close to those empirically found to fit IR spectra.

$$E_n = -D + h\omega_0 \left[(n + \frac{1}{2}) - x(n + \frac{1}{2})^2 \right]$$

The Morse potential was proposed as a solution to this problem.

5.1.2 The Potential

The function Morse proposed has the form

$$V(r) = D_e \left(1 - e^{-\alpha(r-r_0)}\right)^2 \tag{5.1}$$

where $\alpha = \sqrt{k/2D_e}$. Solving the Schrödinger equation analytically,[233] it is found that the eigenvalues are[184]

$$E_\nu = \left(\nu + \frac{1}{2}\right)\hbar\omega - \left(\nu + \frac{1}{2}\right)^2 \hbar\omega \left(\frac{\hbar\alpha^2}{2\mu\omega}\right)$$

In the test calculations, D_e is set to 5, $\mu = 1$, $\omega = 1$ so $\alpha = \sqrt{1/10}$. Transforming to atomic units, the eigenvalues are

$$E_\nu = \left(\nu + \frac{1}{2}\right) - \left(\nu + \frac{1}{2}\right)^2 \left(\frac{1}{20}\right) \tag{5.2}$$

5.1.3 Results

The calculated eigenenergies of the Morse oscillator are reproduced in Table 5.1. Results of the analytical solution are provided for comparison. Clearly, the CI eigenvalues converge to the analytical Morse energies. Note, however, the number of configurations required for higher energy eigenvalues to converge increases rapidly. This is a result of the increasing anharmonicity in the Morse potential at higher energies. This can be seen clearly in Figure 5.1.

Mode	Harmon	05 BFs	10 BFs	15 BFs	20 BFs	25 BFs	60 BFs	Morse
1	0.50000	0.48766	0.48750	0.48750	0.48750	0.48750	0.48750	0.48750
2	1.50000	1.39386	1.38766	1.38750	1.38750	1.38750	1.38750	1.38750
3	2.50000	2.30296	2.19422	2.18790	2.18752	2.18750	2.18750	2.18750
4	3.50000	3.33068	2.97030	2.89963	2.88906	2.88769	2.88750	2.88750
5	4.50000	4.76387	3.85410	3.59704	3.51685	3.49457	3.48750	3.48750
6	5.50000	9.30841	4.92804	4.40083	4.16811	4.06280	3.98753	3.98750
7	6.50000		6.24435	5.35603	4.93324	4.69865	4.38976	4.38750

Table 5.1: Vibrational eigenstates of the Morse oscillator. Harmonic values are provided alongside results of CI calculations with varying configuration size. Analytical Morse eigenenergies are reproduced for comparison.

Study of this system allows testing of the evaluation of diagonal and off-diagonal elements and the matrix diagonalisation procedure. However, this is only a 1D system. The code must be extended to perform calculations on real 3D systems and a model 3D system must be found to test this extension.

5.2 Hénon-Heiles Potential

5.2.1 Introduction

Hénon and Heiles[207] introduced their potential in 1964 as a model function for gravitational field effects on the motion of a star. In their original work, periodic orbits are calculated numerically on an arbitrary 2D surface. A new potential function was required that had to be analytically simple, to speed computation, yet complex enough to give non-trivial trajectories. A 2D potential was proposed:

$$V(x,y) = \frac{1}{2}\left(x^2 + y^2 + 2x^2y - \frac{2}{3}y^3\right) \tag{5.3}$$

This potential has been adopted in computational chemistry as a model potential, often used in benchmarking the calculation of vibrational eigenstates, as, in the past, they have been evaluated using several different methods.[80–86,234] It is often extended to three[83–86] and sometimes many dimensions.[234]

5.2.2 The Hénon-Heiles 3D Potential

The potential chosen for this calculation is the Hénon-Heiles 3D oscillator with a potential of the form

$$V = \frac{1}{2}(x^2 + y^2 + z^2) - 0.1(xy^2 + yz^2 + 0.1x^3 + 0.1y^3) \tag{5.4}$$

It is clear from simple analysis of the potential that the normal coordinates are simply the Cartesian coordinates. To enable an understanding of the surface topography, a plot of the potential in normal/Cartesian coordinate space is given in Figure 5.2.

The eigenstates of this system have been studied in depth.[83–86] Light and co-workers adapted a generalised Gaussian Quadrature Discrete Variable Representation, DVR, method[235] to calculate and label the 20 lowest eigenstates of the potential.[85] These are calculated to a higher degree of accuracy by Echave and Clary[86] with a Potential Optimised DVR (PO-DVR) method.

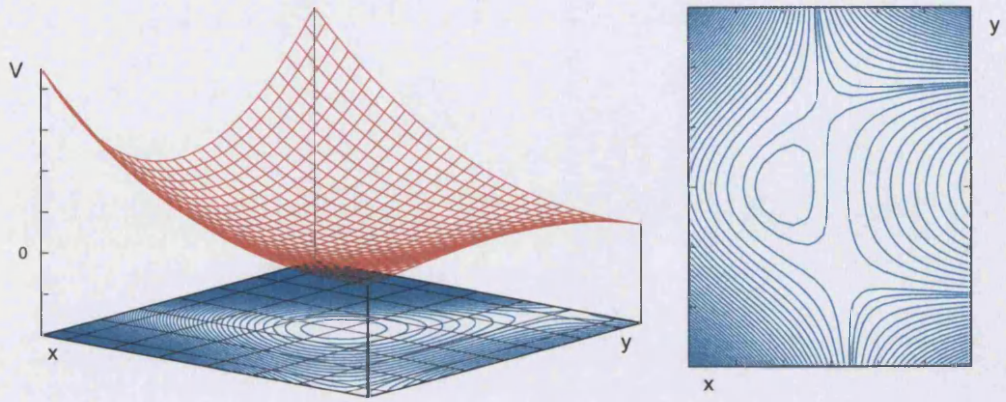


Figure 5.2: 2D representation of Hénon-Heiles potential at $z = 0$ in normal coordinate space.

5.2.3 Results

This single-body potential requires no computational minimisation as the minimum, $(x = 0\, y = 0\, z = 0)$, normal coordinates and eigenvectors can be deduced by analysis of the potential. There is, therefore, no need to employ the areas of the program concerning potential minimisation and normal mode calculation.

Performing a CI calculation with 1942 configurations yields results that are converged for the first 10 eigenstates. Furthermore, the results display excellent agreement with the converged PO-DVR values of Echave and Clary. Note that for a full CI, this size of calculation, even on a simple surface such as the Hénon-Heiles 3D potential studied here, requires 5 days of computer time.

	Absolute Energies (cm^{-1})			
State	Harmonic	Anharmonic	PO-DVR	Configuration
1	1.5000	1.49388975	1.49388975	(0 0 0)
2	2.5000	2.47041786	2.47041786	(0 1 0)
3	2.5000	2.47996173	2.47996173	(0 0 1)
4	2.5000	2.49168459	2.49168459	(1 0 0)
5	3.5000	3.42714282	3.42714282	(0 2 0)
6	3.5000	3.43827646	3.43827646	(0 1 1)
7	3.5000	3.45388584	3.45388584	(0 0 2)
8	3.5000	3.47192175	3.47192175	(1 1 0)
9	3.5000	3.48408013	3.48408013	(1 0 1)
10	3.5000	3.48631241	3.48631241	(2 0 0)
11	4.5000	4.36246634	4.36246633	(0 3 0)
12	4.5000	4.37333157	4.37333155	(0 2 1)
13	4.5000	4.39593488	4.39593489	(0 1 2)
14	4.5000	4.41909346	4.41909342	(0 0 3)
15	4.5000	4.42935671	4.42935669	(0 3 0)
16	4.5000	4.45122547	4.45122546	(2 1 0)
17	4.5000	4.45262411	4.45262411	(0 2 1)
18	4.5000	4.47208896	4.47208896	(1 0 2)
19	4.5000	4.48245731	4.48245730	(3 0 0)
20	4.5000	4.48258834	4.48258834	(2 0 1)

Table 5.2: Comparison of Hénon-Heiles 3D Eigenvalues calculated with the CI method, labelled "Anharmonic", with those reported by Echave and Clary,[86] labelled "PO-DVR". 1942 configurations are used in the CI calculation. 24 abscissae are used in the evaluation of the CI integrals.

Table 5.2 shows the excellent agreement between eigenvalues calculated by Echave's PO-DVR technique and the CI method. This demonstrates an accurate extension of the model to three dimensions. However, the normal modes are all calculated an-

alytically to be one atomic unit. Therefore, the matrix transforming from Cartesian to normal coordinates is clearly the unity matrix. The section of the code concerning the calculation of the normal modes has, therefore, not been tested to three dimensions. Furthermore, the calculation performed on the Hénon-Heiles potential has been a full CI, i.e. the assumption of pairwise interactivity has not been used. The potential, described in Equation 5.4, is in fact pairwise. Application of the pairwise potential approximation to this system, as described in Section 3.6, will, therefore, not affect the results. To test the PPA, another three mode system must be found where the potential contains higher order terms.

5.3 Water Monomer

5.3.1 Introduction

The next stage in benchmarking the CI method is application to a real three mode system. This will allow a full test of the pairwise potential approximation through implementing the technique and benchmarking against full CI results.

Water is a key molecule in biological systems. However, bulk water is an extremely complex system. Initially, water monomer is studied. Subsequent extensions to the models are made to allow for water-water interactions in water clusters, with the final aim of making an accurate model of water's bulk properties.[102,236,237]

As such, water has been studied extensively, both experimentally[237] and theoretically.[90,92,95,96,98–101,103–124,238–246]

In a theoretical study, Watson[247] developed a Hamiltonian for symmetric top triatomic molecules that included both vibrational and rotational motion of the nuclei. Whitehead, Handy, Carter and Sutcliffe[95,238–241] extended this hamiltonian to evaluate rovibrational states of triatomics with general potential functions. These calculations involved linear combinations of configurations, with Morse or harmonic oscillator wavefunctions chosen to model the vibrations and Legendre functions employed to include rotations. Burden and co-workers[96,97] use the Hamiltonian set out by Handy *et al.* but choose numerical wavefunctions as the basis. The Tennyson group have performed extremely detailed calculations on the vibrations of water molecules,[103–124] producing detailed potential energy surfaces[104,115] and including, for example, relativistic effects[112,116] and the breakdown of the Born-Oppenheimer approximation.[106]

The aim of this section is to test the pairwise potential approximation. Therefore, a simple potential is adequate for our needs. The Hoy potential[99] is chosen as it has a simple functional form and it has been used previously in studies of the vibrational states of water.[95,96]

5.3.2 The Hoy, Mills and Strey Potential

In the 1960s and early 1970s, several groups[99,248–253] attempted to incorporate higher order anharmonic terms into empirical and semi-empirical force fields. In 1972, Hoy, Mills and Strey,[99] HMS, introduced a general technique for relating these terms to spectroscopic data. This included a calculation of the force constants for the intramolecular modes of water.

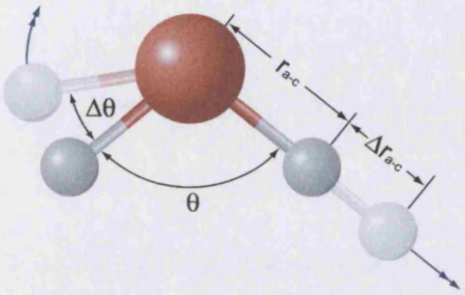
The potential is expanded in internal water coordinates as defined in Figure 5.3

$$V = \sum_i \sum_j \sum_k K_{ijk} \left(\Delta r_{ab}\right)^i \left(\Delta r_{ac}\right)^j \left(\Delta\theta\right)^k \tag{5.5}$$

Values for harmonic and anharmonic force constants, K_{ijk}, and equilibrium geometries, r_{ab}^0, r_{ac}^0 and θ^0, are evaluated by performing a least squares fit of a calculated vibration-rotation spectrum to spectroscopic data.

A plot of the true and harmonic potentials in normal coordinate space is given in Figure 5.4. Clearly the anharmonicity increases greatly at large deviation from the minimum. This potential appears to only be optimised for the low vibrational eigenstates as there does not seem to be a dissociation channel. Consequently, high energy vibrational states will not be well described.

When studying the Hénon-Heiles system, the normal modes were evaluated via a simple analysis of the form of the potential energy function. Consequently, the code for calculation of normal modes has not been tested for systems larger than



(ijk)	K_{ijk}	(ijk)	K_{ijk}
(200)	4.227	(400)	16.0033
(020)	4.227	(040)	16.0033
(110)	-0.101	(310)	-0.8267
(101)	0.219	(130)	-0.8267
(011)	0.219	(220)	0.1425
(002)	0.3485	(301)	0
(300)	-9.8943	(031)	0
(030)	-9.8943	(211)	0
(210)	0.1265	(121)	0
(120)	0.1265	(202)	-0.3525
(201)	0.202	(022)	-0.3525
(021)	0.202	(112)	0.305
(111)	-0.402	(103)	0
(102)	-0.1125	(013)	0
(012)	-0.1125	(004)	-0.0029
(003)	-0.1462		

Figure 5.3: The internal valence coordinates of water monomer. The potential is expanded in terms of the vectors Δr_{ab}, Δr_{ac}, $\Delta\theta$. Hoy, Mills and Strey potential constants in *md*-Å units.[95,99]

1D. The extension to three dimensions of the normal mode code is benchmarked by comparing calculated normal modes with the normal coordinate force constants given by Hoy.

Harmonic Frequencies (cm^{-1})	
Hoy	This work
1648.9	1648.82
3832.0	3831.82
3942.5	3942.43

Table 5.3: Normal modes of the HMS potential. Frequencies compared with those of Hoy.[99]

The differences between the normal frequencies calculated in this study and the Hoy normal coordinate force constants are small and acceptable. Differences arise as a result of different accuracies of the calculations, affecting both the minimisation and the calculation of the normal modes. Different values used for the atomic masses of

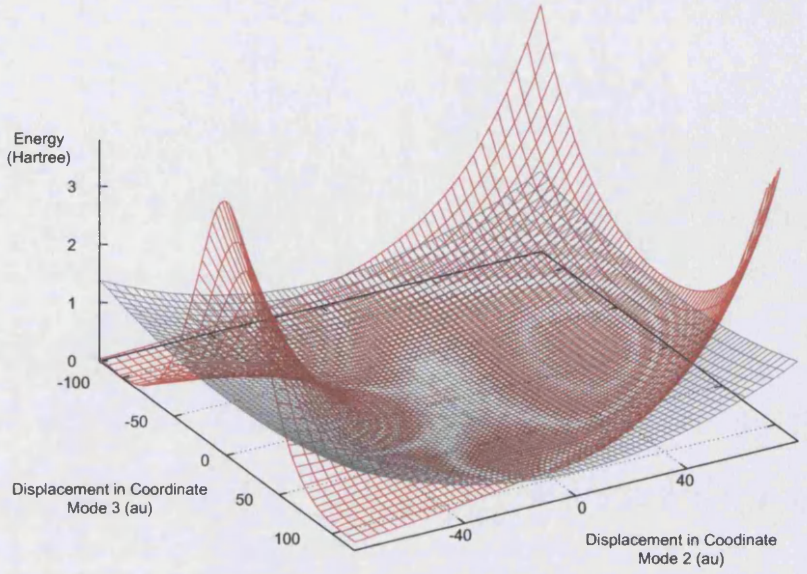


Figure 5.4: Plot of the topography of the Hoy, Mills and Strey[99] water potential in normal coordinate space. The figure depicts the change in energy when moving along normal coordinates. The modes chosen as the coordinate space are the symmetric stretch (Mode 2) and bending (Mode 3) coordinates. The red surface describes the HMS potential and the grey surface is the potential predicted by the harmonic approximation. The plot shows the anharmonicity of the system demonstrating that the harmonic approximation is only valid around the potential minimum.

oxygen and hydrogen and the atomic mass units can have an effect on the normal modes too. The values for the atomic mass unit used in this research are those provided by Cohen and Taylor[254] in 1987, which are clearly different to those used by Hoy *et al.* in 1972. Retro-fitting these constants to values used in 1972 is not possible, since they are not mentioned in the Hoy paper.

5.3.3 The Pairwise Potential Approximation

Calculations on the Morse Oscillator and Hénon-Heiles system have been full CI, making no approximations as to the form of the potentials. Extending to systems of more than three modes will render this technique very slow. A general nD system will require evaluation of n-dimensional integrals in order to calculate the CI matrix elements. The number of potential calls, and therefore the amount of time needed for the calculation, will increase factorially with n.

The pairwise potential approximation, PPA, is introduced in Section 3.5.1.3. The effect of the approximation is to reduce the n-dimensional integrals required to calculated the CI matrix into 1D and 2D integrals. The results of the implementation of this approximation to the calculation of the eigenstates of water monomer are detailed in Table 5.4

Comparing the eigenvalues of the full CI calculation to those obtained making the PPA in Table 5.4 shows the pairwise potential approximation to be quite good. The decrease in calls to the potential subroutine speed up the calculation, taking 525 seconds for the full CI but 248 seconds for the PPA. This speed gain will increase as more configurations are included in the calculation and will have a dramatic effect when moving to larger systems.

	Absolute Energies (cm^{-1})		
Level	Harmonic	Full CI	PPA
(0 0 0)	4711.6	4662.7	4659.4
(0 1 0)	6360.2	6247.6	6239.1
(0 2 0)	8008.8	7805.1	7795.9
(1 0 0)	8543.6	8380.3	8367.5
(0 0 1)	8654.3	8471.8	8450.0
(1 1 0)	10192.2	9978.1	9944.0
(0 1 1)	10302.9	10062.3	10019.3
(1 2 0)	11840.8	11539.1	11492.7
(0 2 1)	11951.5	11598.2	11545.7

Table 5.4: Results of harmonic, full CI and PPA calculations of the absolute energies of vibrational eigenstates of water monomer. Anharmonic eigenstates are calculated with the CI method. The calculation includes 54 configurations and the integrals are evaluated over 24 abscissae.

It is clear that for the lowest modes, the agreement between PPA and full CI calculations is satisfactory. However, as the energy of the state increases, the three body terms in the potential become more important. This is borne out as a greater difference between the higher eigenenergies of the two calculations.

Simple analysis of the potential, defined in Equation 5.5, will not provide an insight into the extent of three body motion. The potential is expanded in terms of force constants in internal coordinates defined in Figure 5.3. Resolution of the two and three body terms is not trivial as any vibration in internal coordinates, when transformed into normal coordinates, is a three body motion. To quantify the extent of

three body terms in the potential, and therefore the accuracy of the approximation, it is possible to plot individual matrix elements of the full CI and PPA calculations. The difference, (full CI - PPA), gives an indication of the effect of the three body terms on the matrix elements, and is plotted in Figure 5.5.

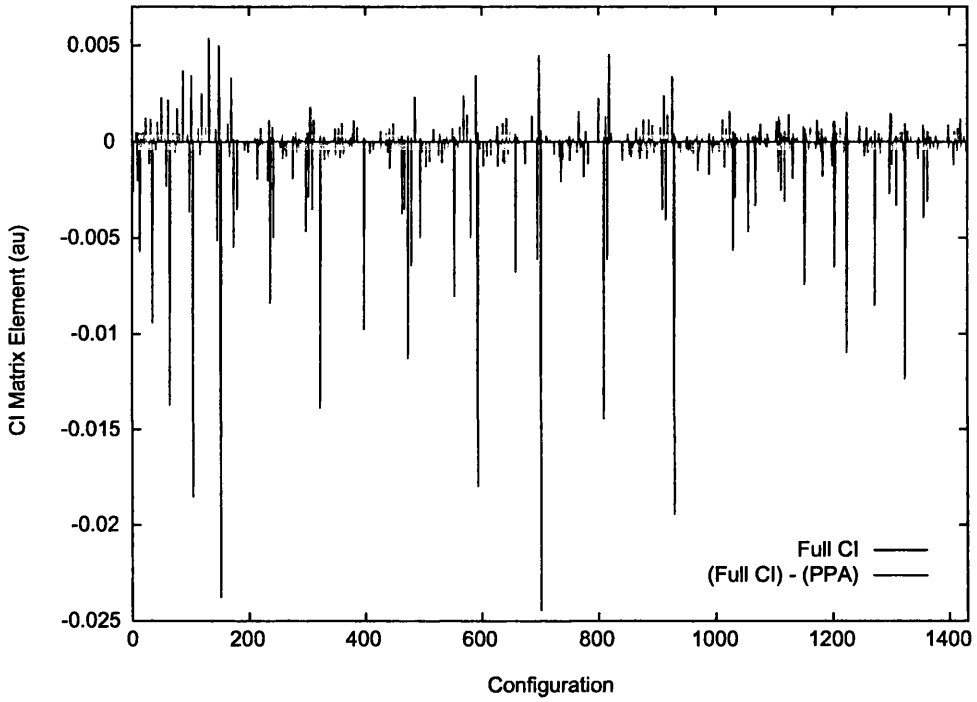


Figure 5.5: Plot to demonstrate the accuracy of the pairwise potential approximation for water monomer. The matrix element of the full CI calculation are plotted in blue. In red is the difference (full CI) - PPA, quantifying the three-body terms in the potential. The higher numbered configurations generally correspond to higher energy terms. The three body terms in the potential become more important at higher energies.

The red line is a plot of the difference between full CI elements and PPA elements. There are two trends to note. Most importantly, the red line is very small in relation to the blue peaks for the lower energy configurations. This is a result of three body terms being relatively small near the potential minimum. However, the red peaks become more pronounced at higher configurations because these are higher in energy. This implies that the pairwise potential approximation is good for lower states but gets worse at higher energies, as expected.

The other clear trend is the positive value of most of the red peaks. As this is the difference (full CI) - (pairwise approximation), it represents the three body interactions. As they are mainly positive, removing them will result in lower absolute energies, as seen in Table 5.4.

5.3.4 Comparison with Literature Values

It is possible to benchmark the pairwise potential approximation by comparing the results of the CI calculation with anharmonic vibrational energies calculated on the same Hoy potential by Whitehead *et al.*[95] and Burden *et al.*[96] Whitehead[95] performs a variational rovibrational calculation for various rotational states. Thus, the CI results must be benchmarked with Whitehead's converged results for J=0 calculations. Burden,[96] uses the Hamiltonian of Whitehead, Handy and Carter.[95,238,239,241] Thus calculations, while performed at J=0, also include rotational effects due to Coriolis coupling terms in the Watson Hamiltonian. Table 5.5 compares the results of a PPA CI calculation with those obtained by Whitehead and Burden.

	Absolute Energies (cm^{-1})			
Level	Harmonic	PPA	Whitehead	Burden
(0 0 0)	4711.6	4659.4	4653.1	4652
(0 1 0)	6360.2	6239.1	6252.8	6249
(0 2 0)	8008.8	7795.9	7820.1	7812
(1 0 0)	8543.6	8367.5	8371.1	8368
(0 0 1)	8654.3	8450.0	8474.5	8472
(1 1 0)	10192.2	9944.0	-	9952
(0 1 1)	10302.9	10019.3	-	10054
(1 2 0)	11840.8	11492.7	-	11494
(0 2 1)	11951.5	11545.7	-	11603

Table 5.5: Results of harmonic and PPA calculations of the absolute energies of vibrational eigenstates of water monomer. The results of calculations by Whitehead[95] and Burden[96] at J=0 are provided for comparison. Anharmonic eigenstates are calculated with the CI method. The calculation includes 54 configurations and the integrals are evaluated over 24 abscissae.

The results show reasonable agreement between some energy levels, especially with the calculation of Burden *et al.*[96] The observed differences between calculations arise as a result of the different configurations and methods used. An important point to note is that the calculations of both Whitehead and Burden have not fully converged. This is demonstrated by a change in absolute energies in going from calculations with the largest configuration to calculations with the second largest configuration in the two studies[95,96] as shown in Table 5.6.

	Basis Set				
Level	$\sum n_i \leqslant 2$	$\sum n_i \leqslant 3$	$\sum n_i \leqslant 4$	$\sum n_i \leqslant 5$	$\sum n_i \leqslant 6$
(000)	4674.51	4654.51	4653.25	4653.20	4653.09
(010)	6281.70	6279.26	6245.23	6252.92	6252.84
(020)	7917.50	7863.54	7855.36	7823.15	7820.12
(100)	8436.34	8421.29	8381.24	8371.34	8371.05
(001)	8538.02	8522.52	8488.22	8474.80	8474.48

	Basis Set	
Level	27 Configurations	64 Configurations
(000)	4652	4652
(010)	6249	6249
(020)	7843	7812
(100)	8369	8368
(001)	8472	8472
(110)	9949	9952
(011)	10055	10054
(120)	11541	11494
(021)	11636	11603

Table 5.6: Tables demonstrating the convergence of the vibrational eigenstates of the Hoy water potential. The left table shows the convergence of J=0 calculations by Whitehead[95] with respect to basis size, defined in terms of $\sum n_i$, where n_i, $i = 1, 2, 3$ is the vibrational quantum number associated with each of the three normal modes. The right hand table shows the convergence of a J=0 calculation by Burden.[96] 27 SCF basis functions is a basis set with up to two quanta in each valence coordinate mode and 64 SCF functions allows up to three quanta in each valence coordinate mode. Note that the two basis set types are different.

An attempt has been made to use a configuration of a similar size to those used by Whitehead and Burden in order to have a better comparison between energy levels. However, the speed at which convergence is reached depends on the type of calculation, along with the number and type of basis functions used. It is almost impossible to reproduce unconverged results with a different Hamiltonian, configuration and basis function. There is also a possibility that Coriolis coupling, non-zero even at J=0, can have an effect.

It is important to consider the differences in the context of the anharmonicity of that state i.e. the drop from calculated harmonic to anharmonic energy, which are as large as 350cm^{-1}. In comparison, differences between the CI energies and those calculated by Burden and Whitehead are small.

In this chapter, the PPA-CI technique and code has been tested against calculations on simple systems. In the next part of this thesis, biomolecular systems of interest are chosen for study. Results of CI calculations are compared with experimental data and harmonic results.

Part IV

Calculations on Biomolecules

Chapter 6

Tryptophan

6.1 Introduction

Accurate calculation of vibrational states and wavefunctions can aid the study of biomolecules in many ways. Their preferred structures in the gas phase and in solution play an essential role in bio-activity and can be affected greatly by the vibrational eigenenergies. For example, one conformer may be the global minimum when comparing the potential minima, D_e, while another conformer may be found to have the lowest zero-point energy, D_0.[164–166] Furthermore, harmonic and anharmonic calculations of zero-point energies can predict different global minima. The zero-point energy may also be large enough to allow free rotation about bonds with a low barrier to torsional motion, as in the case of the benzene-water system.[166]

Calculation of reaction rates with theories such as transition state[47,255] and RRKM[256] requires precise knowledge of activation energies. The activation energy for a reaction is the difference in energy of the ground state and the transition state which must be measured between the ground and transition vibrational states and not the potential minima.

Anharmonic vibrational wavefunctions allow the expectation values for bond lengths to be calculated with greater accuracy, providing a better description of moments of inertia and rotational constants of molecules. Rotational spectra can therefore be fitted more easily. Furthermore, bond lengths also affect molecular dipoles. Exact calculation of the expectation value of bond length can result in more accurate models of dipole moments.[172]

Comparison of experimental spectra with calculated vibrational energies enables testing and parameterisation of potential energy surfaces.[155,158,159] The anharmonic parts of the potential are particularly important as they permit dissociation of the system. Thus testing anharmonicity in force fields is essential in the development of accurate potentials. Since these force fields are used in biomolecular simulations it is important to evaluate their accuracy by comparison with experimental data.

6.1.1 Tryptophan

Of the 20 naturally occurring amino acids, tryptophan is the most complex and is one of eight essential in nutrition for human life. It is a precursor in the biosynthesis of serotonin and melatonin,[7] important neurotransmitters. Serotonin receptors are found throughout the limbic system of the brain and in the gastrointestinal tract. As such, tryptophan is thought to be important in the treatment of many psychological disorders, such as depression,[8,9] seasonal affective disorder[10] and alcoholism[11,12] and in the uptake of food and regulation of appetite.[13,14] Selective serotonin reuptake inhibitors, SSRIs, such as Prozac and Seroxat, are the most commonly prescribed anti-depressants. They function by inhibiting the reuptake of serotonin in the synaptic cleft in order to prolong the serotonin presence but are associated with withdrawal symptoms[15] and other side effects.[16] As a serotonin precursor, L-tryptophan, a naturally occurring amino acid, can also be prescribed in the treatment of psychological and dietary disorders with minimal side effects. It is thought to function by increasing the blood plasma concentrations of tryptophan and aiding the biosynthesis of serotonin.[12,14]

As an amino acid, tryptophan is a building block in the synthesis of peptides and proteins playing a crucial role in many biochemical reactions. For example, this bioactivity is demonstrated in many drugs, such as those designed to prevent the insertion of HIV into cells. As a hydrophobic amino acid, tryptophan is an important constituent of the active site in a drug when binding to the gp41 protein

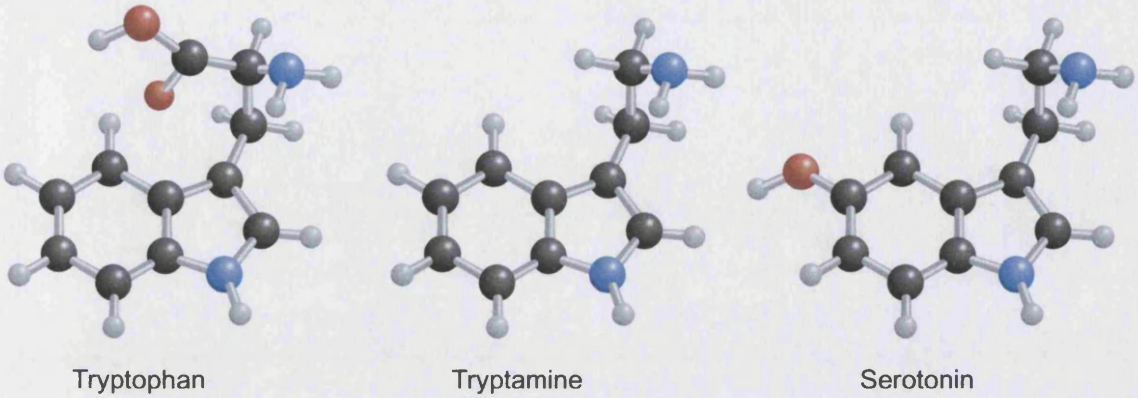


Figure 6.1: The structures of tryptophan, tryptamine and serotonin.

on the surface of the HIV virus.[7] This is shown clearly in Figure 6.2. Tryptophan residues (coloured in the exploded view) in the active site of D10-p1 (white) help bind it to the IQN17 (green), the model used to describe the gp41 protein.[1,2] Once attached, the drug prevents gp41 from folding and fusing viral and human cells.[1–6] Tryptophan has been shown to play a key role in the binding in studies where the binding energy is measured as residues are swapped for inert species.[7] Clearly, understanding and accurately modelling the molecular wavefunctions of tryptamine, an analogue of tryptophan, has scientific and possible medical value. Testing the accuracy, and in particular the anharmonic terms in molecular potentials of tryptophan and similar biomolecules is an important part of this goal. As the majority of the biochemical pathways of these molecules involves solvated species, study of the interaction of tryptophan analogues with water is extremely important.

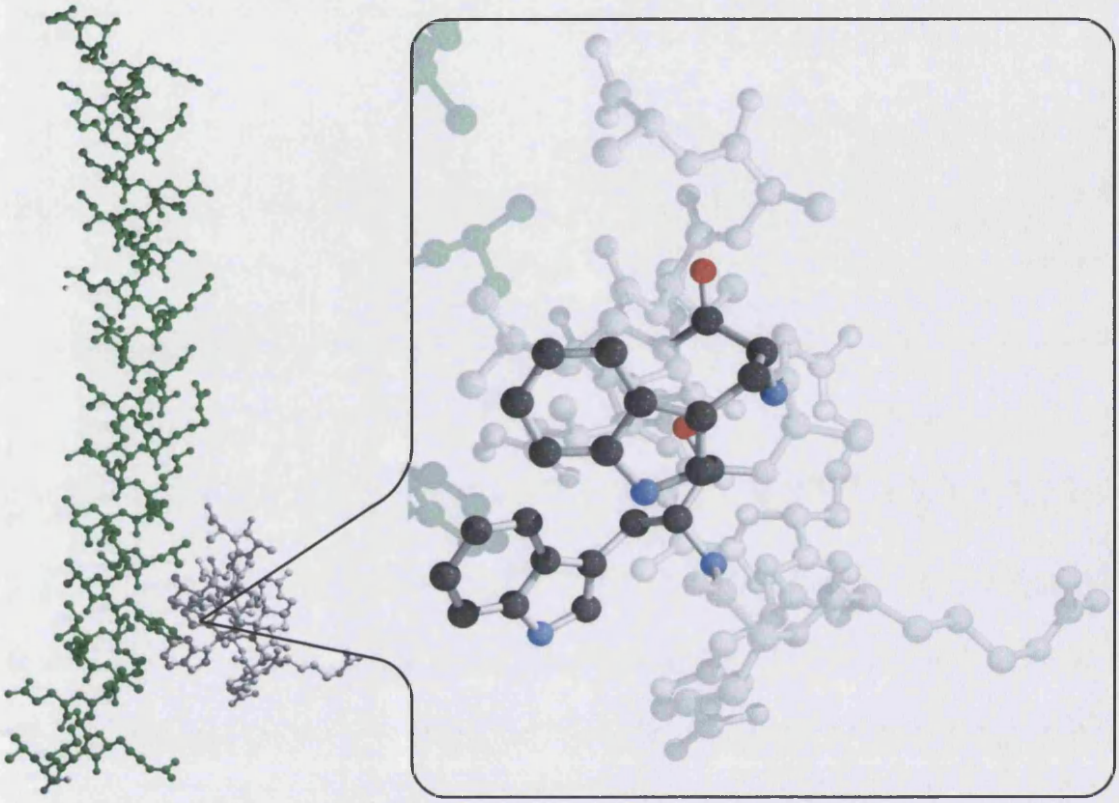


Figure 6.2: Tryptophan (coloured in the exploded view) is crucial to the binding of drugs such as D10-p1 (white molecule) to a cavity in the IQN17 protein (green molecule), blocking the fusion of the HIV to human cells.

Complexation with solvent molecules affects equilibrium geometries and bond strengths. Furthermore, it is thought that the hydrophobic nature of tryptophan residues is a factor in the drug binding activity. Therefore, it is important to test the accuracy of water-amino acid potentials modelling these interactions.

6.1.2 Tryptamine

Tryptamine is the base of the tryptophan amino acid and has been studied extensively.[66–69,72,257–261] The structure is similar to that of both serotonin and its biochemical precursor, tryptophan, though it is not an intermediate in the biochemical pathway (see Figure 6.1). It is chosen for this study as it has relatively few conformational isomers (depicted in Figure 6.3) for a biomolecule, as the only rotational degrees of freedom are the C_α-C_β and C_α-N bonds.[63,64,66,67,74] As a result, tryptamine has been studied extensively, both experimentally and theoretically. Carney, Zwier and co-workers,[66–69] in particular, produced excellent IR spectra in their experiments. These spectra have been assigned to individual conformational isomers. Tryptamine has 24 atoms and 66 vibrational, modes making it an ambitious but technically feasible system to study as an application of the CI techniques discussed in Part II.

Molecular modelling force fields work on the premise that atoms are generally found in a finite number of environments, defined in the potential as atom types. By optimising the parameters of the force field for atoms in various environments, it is possible to predict their behaviour in similar molecules. Many force fields such as CHARMM,[20–27] AMBER[28–33] and MM3Pro[34–42] are created specifically for the study of proteins. As tryptophan is a constituent of many proteins, all the force fields created for their study have atom types optimised for the tryptophan molecule, usually in the zwitterionic form. It is a simple task to remove the COO^- and H^+

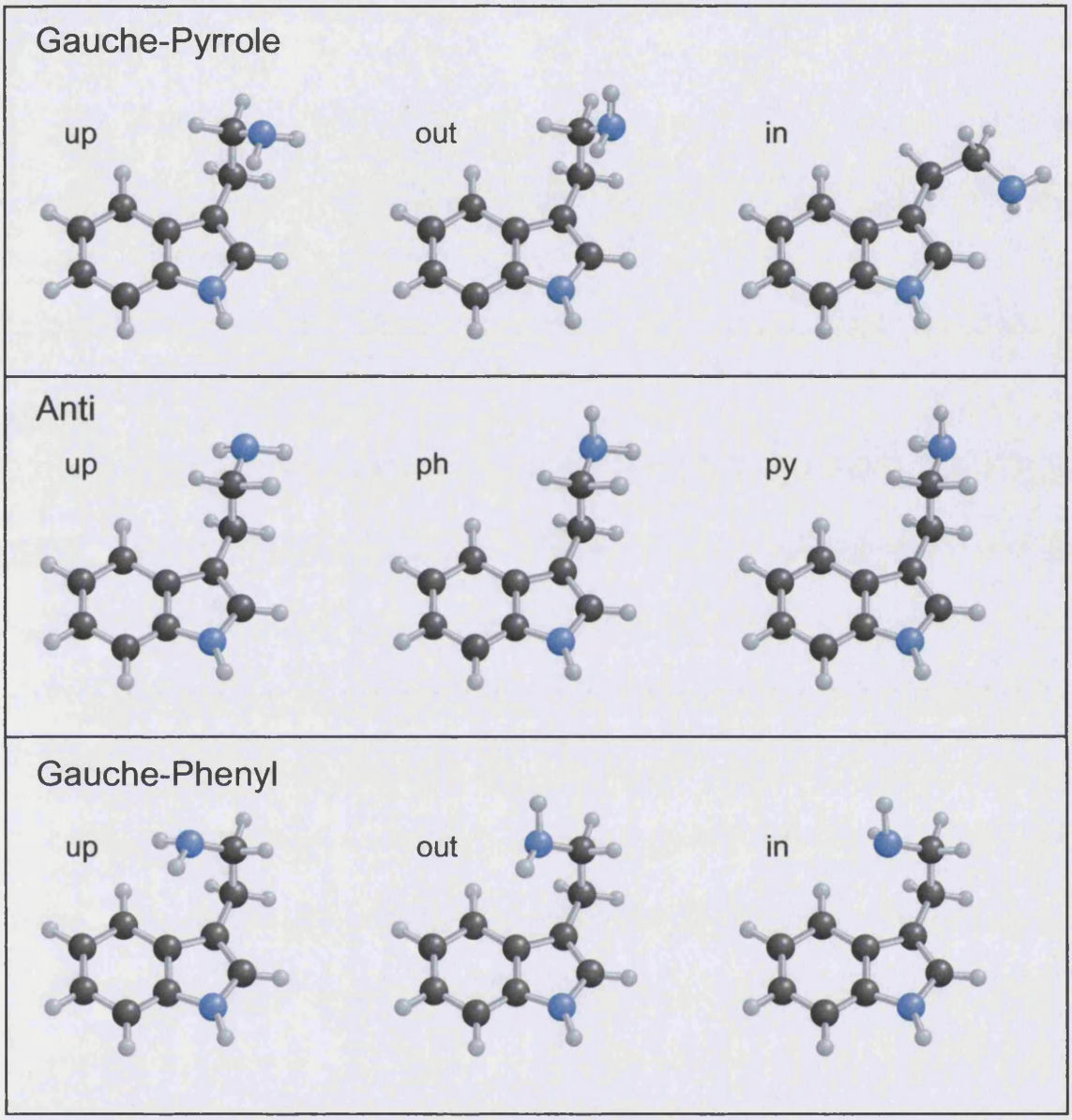


Figure 6.3: The nine possible structural minima for tryptamine.[66,67] They are separated into groups according to the position of the amino group relative to indole. **Gauche Pyrrole** - amino near the pyrrole side of indole. **Anti** - away from the indole. **Gauche Phenyl** - amino near the phenyl side of indole. Inside the three groups, the conformers are labelled according to the position of the amino lone pair relative to the indole ring. The symmetric stretch of the CH_2 group α to the amine displays the most conformational sensitivity, so the calculated frequencies of this mode, typically in the 2840-2940cm^{-1} range, can be used to assign the spectra.

groups to create a tryptamine coordinate file. However, care must be taken to ensure the remaining atoms retain the coordination number and local environment expected in the potential. As a close analogue of tryptophan, the model for tryptamine should be accurate.

6.2 The Potential

The TINKER molecular modelling package by Ponder *et al.*[195–201] is adapted for use as a general potential energy surface.[262,263] The advantage of this is the implementation of many molecular mechanics packages, such as CHARMM,[20–27] OPLS and OPLS-AA,[264–274] AMBER,[28–33] MM3[34–41] and MM3Pro,[42] in TINKER. Once TINKER has been adapted to provide the potential energy at a given molecular orientation, all these force fields are readily available. Another advantage of TINKER is the range of tools that come with the package, including, VIBRATE, a normal mode subroutine to calculate harmonic frequencies. This allows easy comparison of results with the CI code. Various minimisation and surface scanning tools allow easy and complete exploration of the conformational landscape.

The MM3Pro force field is chosen as it is a version of the MM3 potential that has been optimised for the study of proteins. This force field has been widely used in the study of biomolecules.[275–279] The input files for tryptamine and tryptophan include coordinates and atom types which are used by TINKER to assign atom and

bond properties such as bond strengths, equilibrium bond lengths and angles, and dipole moments. Input files can be created by hand, by the TINKER PROTEIN subroutine, or converted from PDB structure files downloaded from the Protein Data Bank.[280] Clearly, while these structures may be in a minimum energy conformation, this cannot be assumed and an energy minimisation should be performed using one of the available TINKER tools. Further subroutines can perform a conformational search in Cartesian or dihedral coordinates allowing the global minimum and other low-energy minima to be found.

Description	Name	Formula
Bond Stretching	E_s	$71.94k_s(l-l_0)^2\left[1-2.55(l-l_0)+(7/12)2.55(l-l_0)^2\right]$
Angle Bending	E_θ	$0.021914(k_\theta)(\theta-\theta_0)^2\left[1-0.014(\theta-\theta_0)+5.6(10^{-5})\right.$ $\left.\times(\theta-\theta_0)^2-7.0(10^{-7})(\theta-\theta_0)^3+9.0(10^{-10})(\theta-\theta_0)^4\right]$
Torsion	E_ω	$(V_1/2)(1+\cos\omega)+(V_2/2)(1-\cos 2\omega)$ $+(V_3/2)(1+\cos 3\omega)$
Stretch-Bend	$E_{s\theta}$	$2.51118K_{s\theta}\left[(l-l_0)+(l'-l'_0)\right](\theta-\theta_0)$
Torsion-Stretch	$E_{\omega s}$	$11.995(K_{\omega s}/2)(l-l_0)(1+\cos 3\omega)$
Torsion-Bend	$E_{\omega\theta}$	0*
Bend-Bend	$E_{\theta\theta'}$	$-0.021914K_{\theta\theta'}(\theta-\theta_0)(\theta'-\theta'_0)$
Van der Waals	E_{vdw}	$\epsilon\left\{-2.25(r_v/r)^6+1.84(10^5)\exp\left[-12.00(r/r_v)\right]\right\}$
Electrostatics	E_{es}	$E_{\mathrm{d-d}}+E_{\mathrm{c-d}}$†‡
Rotational Barriers	E_{rb}	K_{rb}§

*These terms, though present in the MM2 force field are unused in MM3/MM3Pro.

†$E_{\mathrm{d-d}}$ is the dipole-dipole interaction term.

‡$E_{\mathrm{c-d}}$ is the charge-dipole interaction term, non-zero for charged molecules.

§Constant describing the rotational barrier for a four atom linkage.

Table 6.1: The interaction terms included in the MM3Pro force field.[34]

Almost all molecular modelling potentials are semi-empirical, including complex functions to model bending, stretching and torsional motion. The terms included in the MM3Pro potential are listed in Table 6.1. Parameters for all bonds and electrostatic and Van der Waals parameters for individual atoms are fitted to experimental and ab initio data. As MM3Pro and other force-fields include many complex terms, it takes longer to compute the PE than on the simple empirical potentials used in benchmarking the code. Moreover, the system under consideration is much larger than before.

The calculation of the CI matrix is the slowest step in the CI procedure. Evaluation of the matrix elements requires numerical integration. Computationally expensive potential calls are made at each abscissa. The time taken for a CI calculation has increased dramatically in moving from the simple benchmark systems to tryptamine. An attempt must be made to limit the number of calls made to the potential during the CI procedure. To reduce the time required for a CI calculation, the most commonly used integrals are stored in arrays and re-used, drastically cutting the number of potential points that need to be evaluated. The details of the integrals evaluated are given in Section 3.6.2 where the problem of overcounting integrals is also discussed. There, it is shown that, for a general 3 mode system, the integral $\int \psi_1^1\psi_2^0\psi_3^0\hat{\mathbf{A}}\psi_1^0\psi_2^0\psi_3^0\,\mathrm{d}\tau$ reduces to

$$\langle\psi_1^1\psi_2^0|\sum_i\sum_j c_{1,2}^{i,j}q_1^iq_2^j|\psi_1^0\psi_2^0\rangle - \langle\psi_1^1|\sum_i c_1^iq_1^i|\psi_1^0\rangle + \langle\psi_1^1\psi_3^0|\sum_i\sum_j c_{1,3}^{i,j}q_1^iq_3^j|\psi_1^0\psi_3^0\rangle \tag{6.1}$$

Stored Integrals	
Integrals over 1 Mode	Integrals over 2 Modes
$\langle \psi_i^0 \vert \hat{\mathbf{A}} \vert \psi_i^0 \rangle$	$\langle \psi_l^0 \psi_m^0 \vert \hat{\mathbf{A}} \vert \psi_l^0 \psi_m^0 \rangle$
$\langle \psi_i^1 \vert \hat{\mathbf{A}} \vert \psi_i^0 \rangle$	$\langle \psi_l^1 \psi_m^0 \vert \hat{\mathbf{A}} \vert \psi_l^0 \psi_m^0 \rangle$
$\langle \psi_i^2 \vert \hat{\mathbf{A}} \vert \psi_i^0 \rangle$	$\langle \psi_l^1 \psi_m^1 \vert \hat{\mathbf{A}} \vert \psi_l^0 \psi_m^0 \rangle$
$\langle \psi_i^1 \vert \hat{\mathbf{A}} \vert \psi_i^1 \rangle$	$\langle \psi_l^1 \psi_m^0 \vert \hat{\mathbf{A}} \vert \psi_l^1 \psi_m^0 \rangle$
	$\langle \psi_l^1 \psi_m^1 \vert \hat{\mathbf{A}} \vert \psi_l^1 \psi_m^1 \rangle$

Table 6.2: Commonly calculated integrals stored to reduce the number of calls made to the potential subroutine.

on making the pairwise potential approximation and accounting for overcounting integrals over the single modes. It is clear that one dimensional integrals will be used many times in evaluating matrix elements so all integrals of this type are stored and re-used. Many 2D integrals, such as the one shown in Equation 6.1, are also evaluated more than once and should be stored. Table 6.2 shows the integrals that are most commonly required in setting up the CI matrix. They are stored and re-used to reduce the number of potential calls made.

This is an extremely effective way of reducing the time taken to perform the CI calculation. By far the slowest part of the technique is the calculation of the CI matrix. By reducing the number of integrals that must be evaluated, the speed is dramatically increased. In fact, calculations on 20-30 atoms systems with 2000 configurations required 2-3 days before the integrals were stored. On storing the commonly needed integrals, the same calculation can be performed in 2-4 hours.

6.3 The Basis Set

In this section, the effect of the CI space on the calculation is investigated. To aid discussion, a simple labelling scheme is required that defines the basis and CI configurations as concisely and accurately as possible.

6.3.1 Basis Set Labelling

The notation used to describe the individual wavefunction and configurations is defined in Section 3.3.1. Briefly, the CI space is split into four distinct subgroups, the ground state, Ψ^0, the singles, Ψ_i^S, the doubles, Ψ_{ij}^D and the triples, Ψ_{ijk}^T. The terms single, double and triple refer to the number of modes in the configuration that are excited above the ground state irrespective of the vibrational quantum number of those modes. For example, a configuration with mode 4 excited to vibrational quantum state 6, but all other modes in the ground state is referred to as singly excited. More generally, the configuration types can be defined in terms of the individual normal mode basis functions, ψ_a^y, with y quanta in normal mode a.

$$\Psi^0 = \prod_i \psi_i^0(q_i) \tag{6.2}$$

$$\Psi_i^S = \sum_{i=1}^{3N-6} \sum_{l=1}^{l_{\max}} \psi_i^l(q_i) \prod_{i' \neq i} \psi_{i'}^0(q_{i'}) \tag{6.3}$$

$$\Psi_{ij}^D = \sum_{i=1}^{3N-7} \sum_{j>i}^{3N-6} \sum_{l=1}^{l_{\max}} \sum_{m=1}^{m_{\max}} \psi_i^l(q_i)\psi_j^m(q_j) \prod_{i' \neq i,j} \psi_{i'}^0(q_{i'}) \tag{6.4}$$

$$\Psi_{ijk}^T = \sum_{i=1}^{3N-8} \sum_{j>i}^{3N-7} \sum_{k>j}^{3N-6} \sum_{l=1}^{l_{\max}} \sum_{m=1}^{m_{\max}} \sum_{n=1}^{n_{\max}} \psi_i^l(q_i)\psi_j^m(q_j)\psi_k^n(q_k) \prod_{i' \neq i,j,k} \psi_{i'}^0(q_{i'}) \tag{6.5}$$

The labelling scheme for the CI space can be defined in terms of these four subgroups. As the ground state is always included, this can be ignored, leaving singles, doubles and triples. The only extra information needed is the maximum vibrational quantum number of each of the modes. A full description of the configuration can therefore be achieved with three numbers, *SDT*. In this labelling rationale, each number represents the maximum vibrational quantum number, $v_{\max}$, of each normal mode of the single, double and triple subgroups respectively, such that

$$v_{\max} = l_{\max} = m_{\max} = n_{\max} \tag{6.6}$$

This is referred to as an *SDT* configuration. As an illustrative example, a 112 configuration of a three mode system is shown in Table 6.3

In calculations on large systems, such as tryptamine, it is expensive to perform a CI calculation using more than 2000 configurations as a CI matrix larger than this will exceed the permitted stack size of the computer available to us. In order to decrease the size of the CI space, it is possible to systematically select normal modes to

Individual Configurations				
Ground State	Single	Double	Triple	
$\psi_1^0\psi_2^0\psi_3^0$	$\psi_1^1\psi_2^0\psi_3^0$	$\psi_1^1\psi_2^1\psi_3^0$	$\psi_1^1\psi_2^1\psi_3^1$	$\psi_1^2\psi_2^2\psi_3^1$
	$\psi_1^0\psi_2^1\psi_3^0$	$\psi_1^1\psi_2^0\psi_3^1$	$\psi_1^2\psi_2^1\psi_3^1$	$\psi_1^2\psi_2^1\psi_3^2$
	$\psi_1^0\psi_2^0\psi_3^1$	$\psi_1^0\psi_2^1\psi_3^1$	$\psi_1^1\psi_2^2\psi_3^1$	$\psi_1^1\psi_2^2\psi_3^2$
			$\psi_1^1\psi_2^1\psi_3^2$	$\psi_1^2\psi_2^2\psi_3^2$

Table 6.3: Expansion of a 112 configuration of a three mode system in individual normal mode basis functions.

perform a CI calculation on. For example, in the study of tryptamine IR spectra, only 12 vibrational frequencies fall in the 2800-3800cm^{-1} region of the spectrum measured by Zwier. The normal mode closest in energy to the 12 IR active modes has a much lower vibrational frequency of around 1800cm^{-1}. It is assumed that this large gap in energy results in little coupling between the IR active and IR inactive modes. Therefore only these 12 modes are included in a CI calculation. In such a scenario, where only a subset of the normal modes of the system are included in the CI space, the configuration labelling scheme, defined above, must also list the modes included. A calculation on the highest three energy modes of tryptamine at 112 excitation will involve the individual configurations presented in Table 6.3. A calculation on, for example, the 12 highest energy normal modes of tryptamine, will require a much larger CI space.

In summary, the total configuration of the CI calculation can be described, in full, by the modes chosen for study and three numbers, *SDT*, defining the maximum vibrational quantum number of each normal mode in the single, double and triple subgroups.

An understanding of the effect of the choice of CI space on the calculation is vital. It is important to investigate the effect of varying the configuration on CI results, with a goal of achieving reasonably converged results for certain vibrational states with as short a calculation as possible. Having defined the configuration in terms of the modes included and the *SDT* excitation, these are an obvious choice for the variables to investigate first.

6.3.2 Varying the Normal Modes in the Configuration

Consider the number of modes included in the calculation. Tryptamine has 24 atoms and therefore 66 vibrational modes. Only the 12 modes highest in energy fall into the frequency range of the IR spectra measured, 2800-3800cm^{-1}, so, at the very least, these 12 modes must be included in the configuration. To test the hypothesis that the energy gap between IR active and IR inactive modes is large enough that only IR active modes need to be included in the calculation, more modes can be added.

In Figure 6.4, the theoretical spectrum in the 2800-3600cm^{-1} region is plotted for CI calculations on the 12 highest energy normal modes at 321 excitation. To test the effect of modes outside the IR region, further results are plotted for calculations where the next highest energy mode is added in sequence until the configuration includes the 20 highest energy normal modes. The configuration is kept constant at 321 for these calculations. The results of an harmonic calculation on the MM3Pro surface and the gas phase resonant ion-dip infrared spectra of tryptamine, measured and conformationally assigned by Carney and Zwier,[66] are included for reference.

A clear trend shown in Figure 6.4 is the convergence between experimental and calculated spectra on including more modes in the configuration. In fact, results for the calculations including 20 normal modes agree quite well with experiment. Spectral lines corresponding to empty regions of the spectrum, particularly the NH_2 symmetric and asymmetric stretches at 3315 and 3400cm^{-1} respectively, can be explained by the fact that transition intensities are negligibly small as shown by the calculations of Carney (See Table 6.4 in Section 6.4).

It is to be expected that a variational calculation improves on expanding the CI space. The time taken to perform a CI calculation is important if the technique is to be applicable to large, complex systems in future. The aim, as in electronic structure CI, is to truncate the CI configuration to the maximum extent without greatly affecting

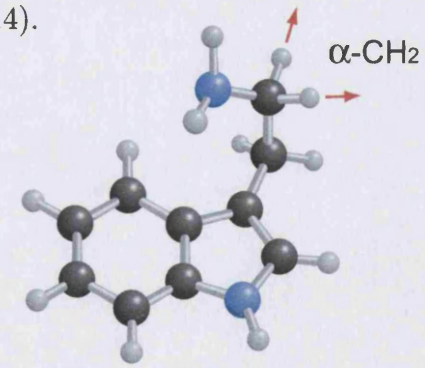


Figure 6.5: Symmetric stretch of the α-CH_2 group of tryptamine.

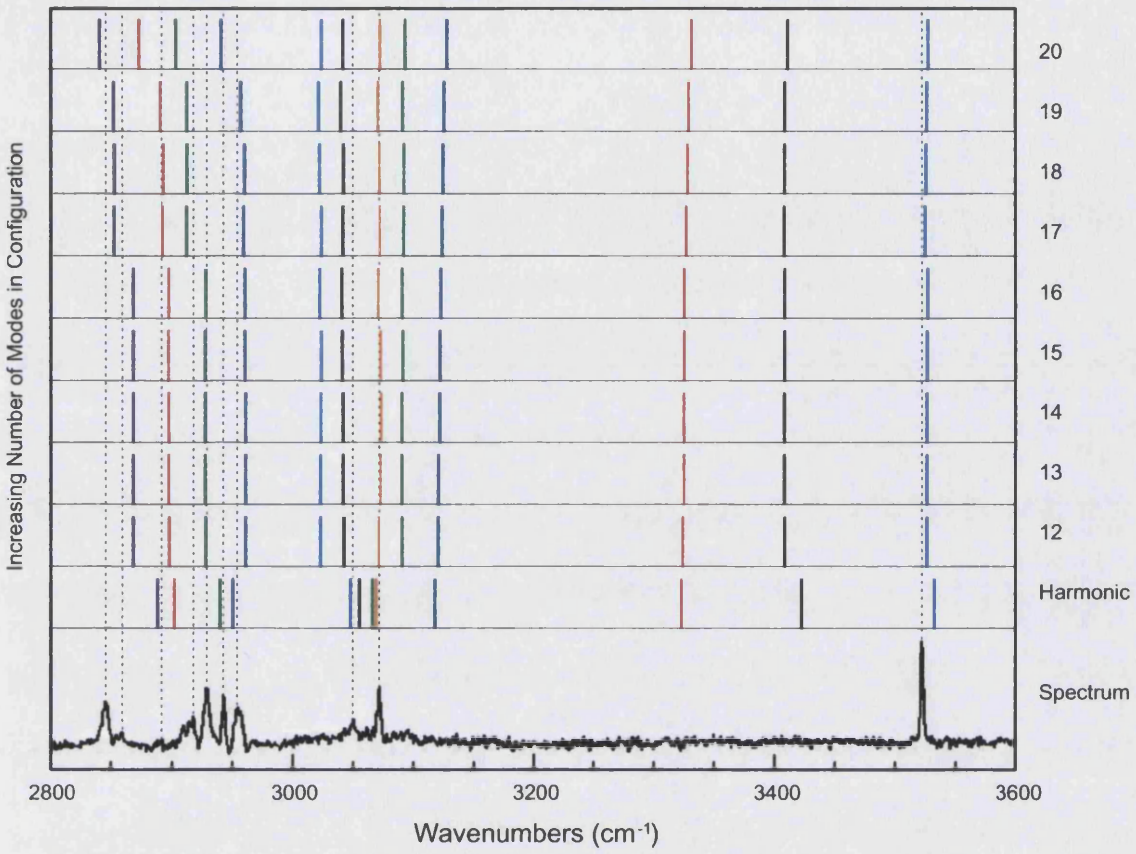


Figure 6.4: CI calculation on the tryptamine gauche-phenyl (out) conformer. Configurations are of the same SDT excitation but include an increasing number of normal modes. Line colours allow the mode to be followed over the range of calculations. Experimental spectrum from Carney and Zwier[66] provided for comparison.

the resulting eigenstates. It is important to determine the smallest configuration, leading to the shortest calculation, that will reproduce experimental results adequately. The harmonic spectrum shows acceptable agreement with experiment for the high energy amine N-H stretch at 3525cm^{-1}. However, the CH_2 group α to the amine (See Figure 6.5) displays the most conformational sensitivity. The vibrational

frequency of the symmetric stretch of this group, which, for the gauche-phenyl (out) conformer, occurs at 2845cm^{-1}, must be calculated accurately in order to assign the spectra. The harmonic approximation breaks down for lower energy vibrations such as these. This breakdown is clearly shown by improved agreement with experiment on including anharmonic effects in the CI calculation. Harmonic calculations on the MM3Pro surface are, therefore, inadequate for comparison with experimental spectra. The spectra predicted by CI calculations show a far greater correlation with experiment, and it is seen that agreement with the experimental spectrum improves as more modes are included in the CI space. The assumption that only the 12 IR active modes are needed in the configuration is shown to be reasonable. The results of the 12 mode CI calculation show a far better correlation with experiment than harmonic calculations. However, as the aim is to assign conformational isomers to experimental spectra, the calculated spectra must be very accurate. It is clear from Figure 6.4 that, in order to model the low energy modes in the spectrum accurately, 17-20 modes must be included in the configuration. Therefore, when attempting to compare experimental and calculated spectra, as in Section 6.4, the configuration should include as many modes as possible.

A re-ordering and splitting in the five modes around 3050cm^{-1} is seen in Figure 6.4 in going from harmonic to CI calculations, with some spectral lines increasing in energy. The normal modes in this region of the spectrum couple strongly, pushing their energies apart. The absolute energies of the 12 spectral modes and the

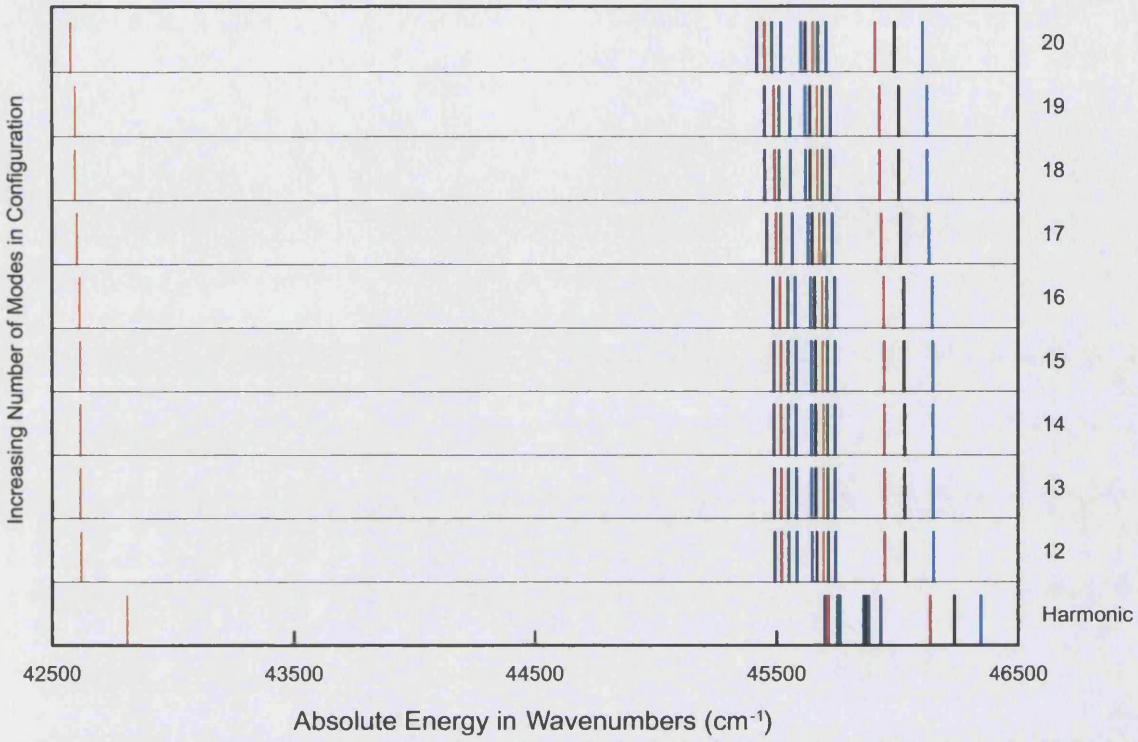


Figure 6.6: Absolute energies of the ground and 12 highest energy vibrational eigenstates of a CI calculation on the tryptamine gauche-phenyl (out) conformer. The configurations are of the same *SDT* excitation but include an increasing number of normal modes. Eigenstate line colours are the same as for Figure 6.4, to enable comparison.

ground state energy are plotted in Figure 6.6. On close examination, it is clear that the variational principle is still obeyed, as the absolute energy of each eigenstate decreases on extending the CI space. However, this decrease in absolute energy is less for the normal modes in the 3100cm^{-1} and 3300-3400cm^{-1} regions than for the ground state, resulting in an increase in the energy gap and therefore the transition energy.

In both Figures 6.4 and 6.6 a change is seen on adding mode 17 (1652cm^{-1}) character to the configuration. The rationale for adding more normal modes is that those closest in energy to the 12 IR active modes will have the greatest effect on them. In both Figures, calculations performed including the spectral modes and adding normal modes 13 (1909cm^{-1}), 14 (1828cm^{-1}), 15 (1805cm^{-1}) and 16 (1732cm^{-1}) show very little change from calculations with the original 12 spectral mode configuration. The change in the calculated spectrum on including mode 17 (1562cm^{-1}) is a breakdown in this assumption.

To understand this breakdown, a visualisation of the actual movement of the atoms involved in the normal mode vibrations is needed. A schematic representation of the molecular motion involved in the 20 highest energy vibrational modes of the gauche phenyl (out) conformer of tryptamine is presented in Figure 6.7. Looking at modes 13-16, it is clear that these involve, on the whole, stretching of bonds inside the indole ring. The 12 IR active modes falling in the 2800-3600cm^{-1} region of the spectrum involve stretching of hydrogen-heteroatom bonds and so are unlikely to couple strongly with modes 13-16. Modes 17-20, however, involve motion of the same atoms as modes 1, 2 and 9-12, so the modes are likely to couple. Thus, it is clear that modes 17-20 should have a greater effect on the spectrum obtained by CI calculations than modes 13-16 as seen in Figures 6.4 and 6.6.

The breakdown in the assumption that modes closest in energy couple most strongly means that many modes must be included in the CI calculation to achieve reasonable

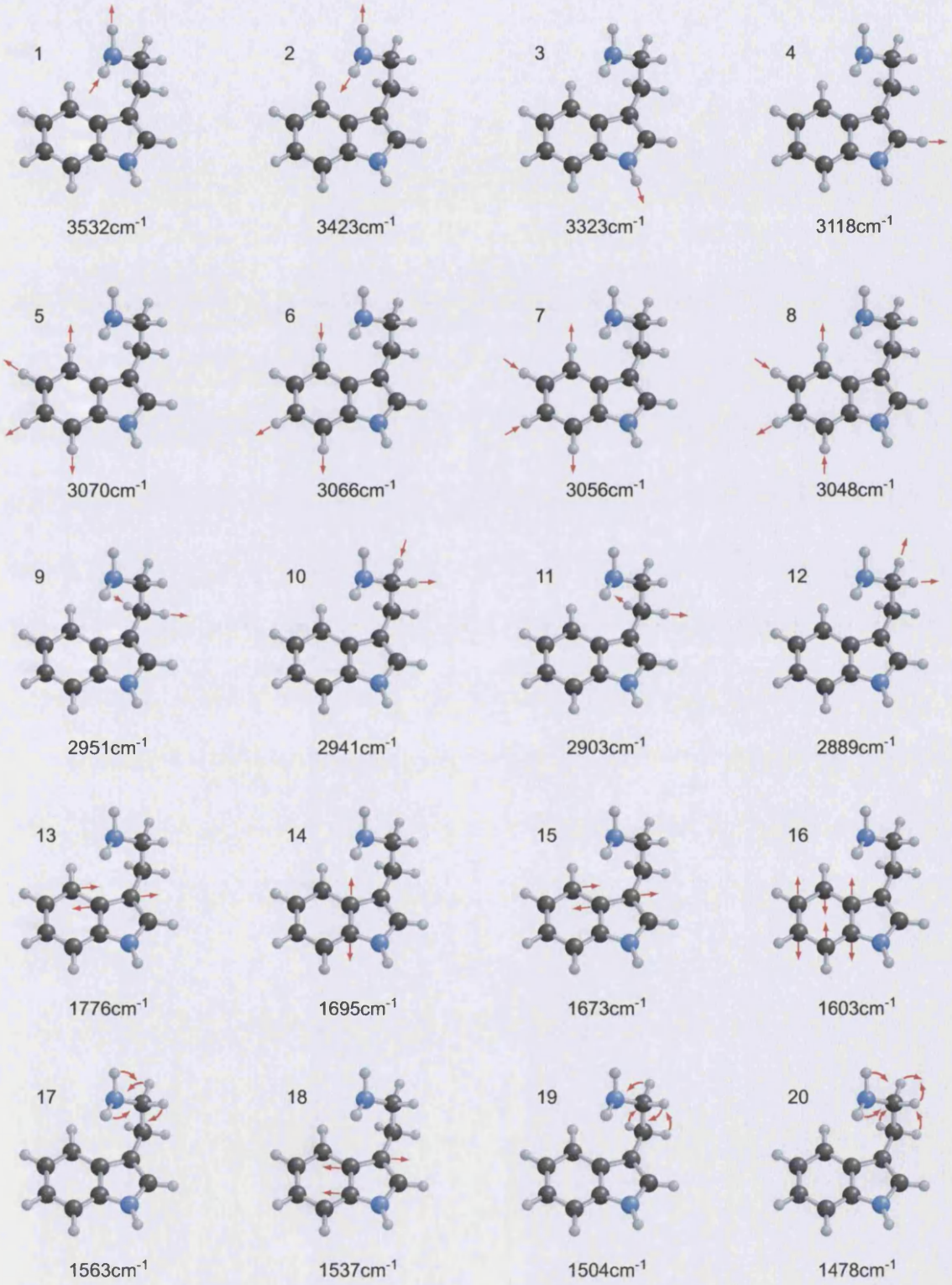


Figure 6.7: The 20 highest energy vibrational normal modes of tryptamine Gauche Phenyl (out). Normal mode frequencies, as calculated on the MM3Pro potential, are given in wavenumber units.

convergence. However, adding extra modes to the configuration greatly increases its size and therefore the length of the calculation. For example, for the current 12 mode problem at 321 excitation there are around 500 configurations. The rate limiting step of the CI calculation is the evaluation of the CI matrix. This is symmetrical so only one half of the off-diagonal elements (but all of the diagonal elements) need be obtained. Therefore, assuming that all elements take roughly the same amount of time to evaluate, the length of calculation scales as $1/2\ n(n+1)$, where n is the size of the configuration. In the case of 500 configurations, a CI calculation on tryptamine on the MM3Pro potential takes roughly 60 minutes. Including 15 modes at 321 excitation requires around 1000 configurations meaning that four times the number of matrix elements must be evaluated. In Section 6.3.1 it is shown that many integrals are stored and not recalculated, so the scaling is, in fact, roughly 2:1 and the calculation actually takes 120 minutes. Including 20 modes in a 321 excitation configuration requires 2000 states. A calculation of this size is expected to take 16 times as long as a 500 configuration calculation but, due to some integrals being reused, the actual time taken is 4 hours. This is still twice as long as a calculation on a 15 mode configuration will take. Though for these systems, the increase in time is only a matter of hours, on more complex potential energy surfaces, the increase is likely to be in days. The number of potential calls made must be kept to a minimum. It will be extremely beneficial to be able to choose the modes with a significant effect on the spectrum with a more rigorous rationale.

6.3.3 Selection of Normal Modes for the Configuration

In the CI calculation, coupling between IR active i.e. modes with vibrational frequencies in the IR range of the spectrum, and inactive modes, i.e. modes with vibrational frequencies outside the IR range of the spectrum, is likely to occur between the first excited state of the IR active modes, ψ_{IR}, and the second excited state of the IR inactive modes, ψ_{inact}. The extent of coupling between these two states can be quantified by evaluating the integrals

$$\langle \psi^1_{\mathrm{IR}} \psi^2_{\mathrm{inact}} \,|\, \mathbf{A} \,|\, \psi^0_{\mathrm{IR}} \psi^0_{\mathrm{inact}} \rangle \tag{6.7}$$

The energy gap is still important so the integrals are adjusted to account for this using the term occurring in second order perturbation theory

$$\frac{\langle \psi^1_{\mathrm{IR}} \psi^2_{\mathrm{inact}} \,|\, \mathbf{A} \,|\, \psi^0_{\mathrm{IR}} \psi^0_{\mathrm{inact}} \rangle^2}{3/2\omega_{\mathrm{IR}} - 5/2\omega_{\mathrm{inact}}} \tag{6.8}$$

The effect of IR inactive modes on the spectrum is quantified in Equation 6.8. A new code was written to evaluate all such coupling terms between all IR active and inactive modes. The three IR inactive modes found to couple most strongly with IR active modes are 20, 17 and 18 in that order. A CI calculation is performed with a configuration of the IR active modes and modes 17, 18 and 20 in 321 excitation. The results are compared with those for calculations with configurations of the highest energy 12, 15 and 20 normal modes in Figure 6.8.

The calculations with configurations of the highest energy 12 and 15 modes produces very similar results. This is explained above with reference to the atoms involved

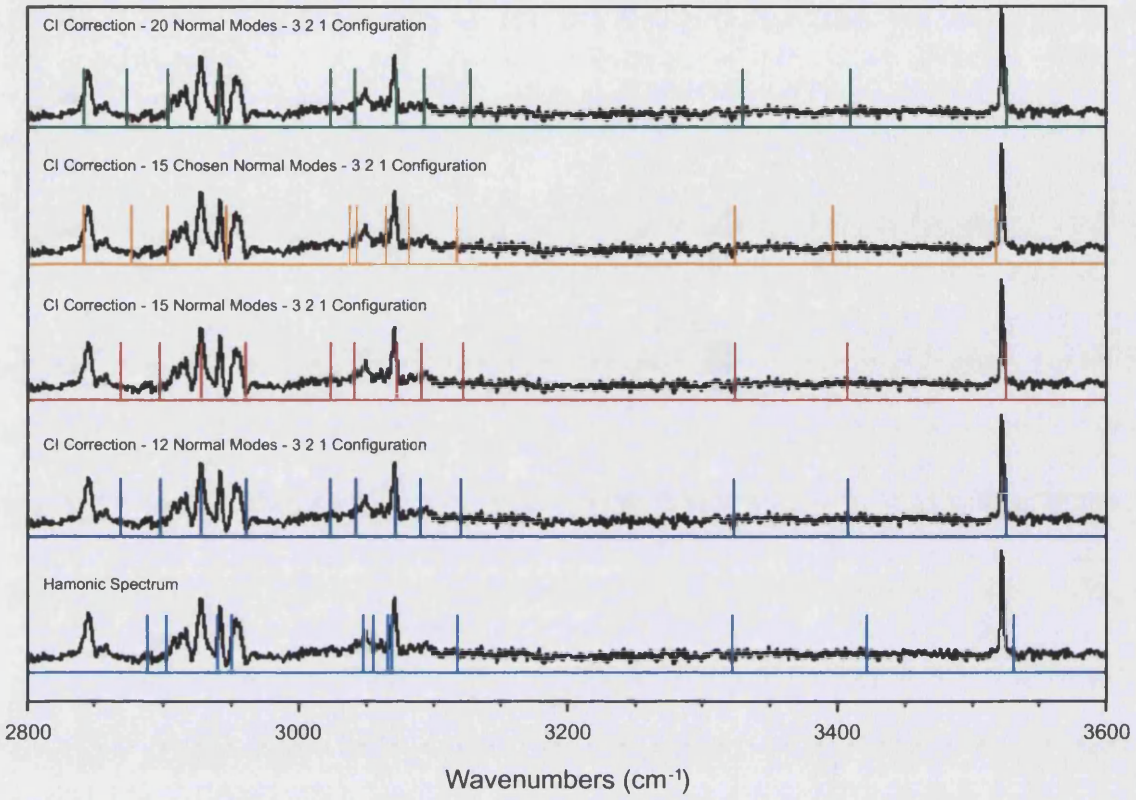


Figure 6.8: Tryptamine gauche-phenyl (out) conformer. Comparison of experimental spectrum[66,67] (corresponding to the C2 line in the one colour resonant two-photon ionisation spectrum of TRA^+) with harmonic and CI calculations. CI configurations include the 12, 15 and 20 modes highest in energy and the 12 highest energy modes with a further three selected according to coupling.

in the vibrations. In fact, in the middle and high energy (3000-3600cm^{-1}) regions of the spectrum, all CI calculations are similar and reproduce experiment well. The differences arise in the low energy (2800-3000cm^{-1}) region that is crucial for assignment of isomers. The configuration including IR active and modes 17,18 and 20 is the same size as that of the highest energy 15 modes and, consequently, results in a CI calculation of the same length. However, the calculation with the

chosen modes shows a greater similarity to experiment than the CI calculation with a configuration of the 15 highest energy modes, especially in the important low energy region. In fact, in this part of the spectrum, the configuration including three strongly coupled modes is effective at reproducing results for the 20 mode calculation despite including only 75% of the number of modes. This is an important finding as a 15 mode configuration at 321 excitation includes 1000 functions whereas a 20 mode configuration at the same excitation includes around 2000 states, a 50% reduction. Halving the CI space results in a calculation that takes roughly a quarter of the time, as explained above. Overall, this technique can be useful as it is certainly an improvement on a calculation of similar size in the important low energy region of the spectrum and can reproduce results of a larger calculation in a quarter of the time.

6.3.4 Varying the *SDT* Excitation of the Configuration

The effect of the modes included in the CI space has been studied. The other term in which a configuration is defined is the mode excitation, the maximum number of quanta in each mode of the subgroup S, D or T. In Figure 6.9, the spectrum is calculated with a configuration of the 12 highest energy normal modes, varying the vibrational quantum numbers for each mode. The configurations are labelled with the SDT notation described above.

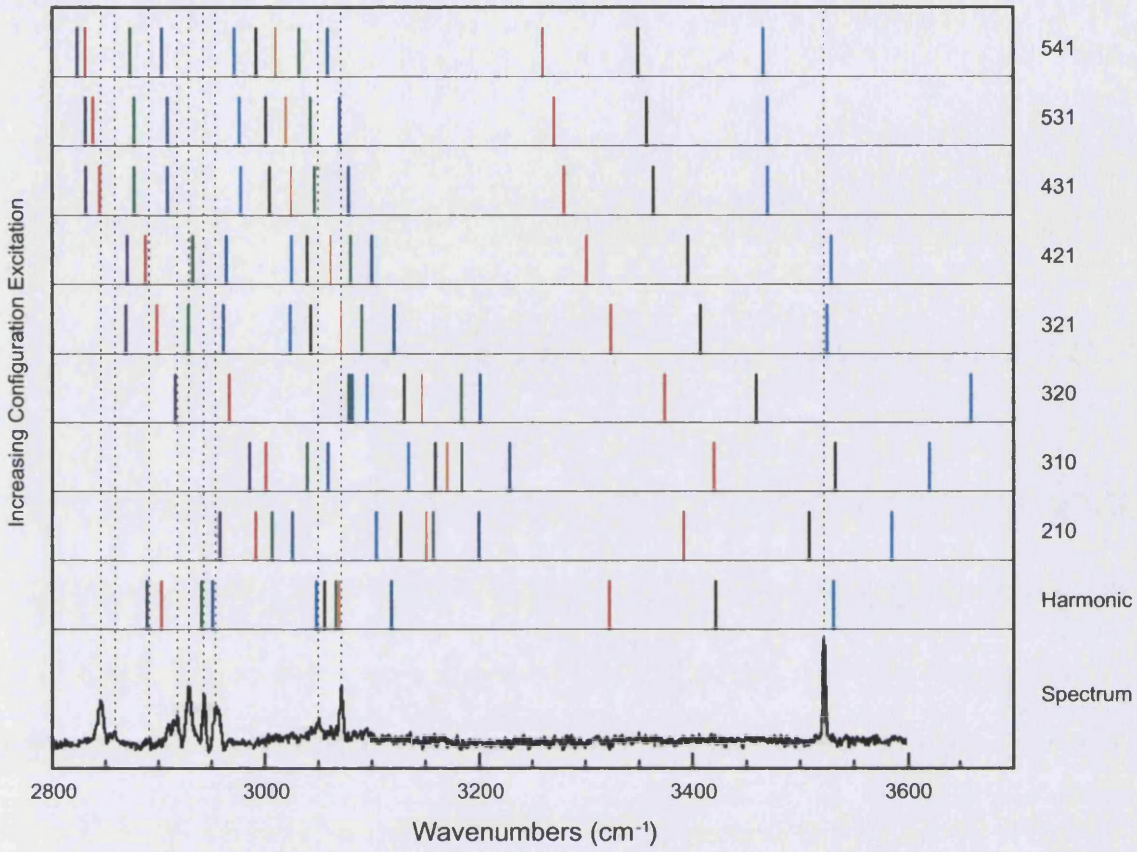


Figure 6.9: CI calculations on the tryptamine gauche-phenyl (out) conformer. Configurations include the 12 spectral modes with increasing SDT excitation. Line colours allow the mode to be followed over the range of calculations. Experimental spectrum from Carney and Zwier[66]

The transition energies increase from the harmonic spectrum up to the CI calculation with a 320 configuration, an apparent contravention of the variational principle. However, plotting the absolute energies of the vibrational eigenstates in Figure 6.10 clearly shows that the variational principle is followed. It is the decrease in energy of the ground state relative to the excited states that results in the increase in transition energies. Including triple excitations, i.e. one quanta of excitation in three modes, in the CI space reduces the energy of the excited states relative to the ground state.

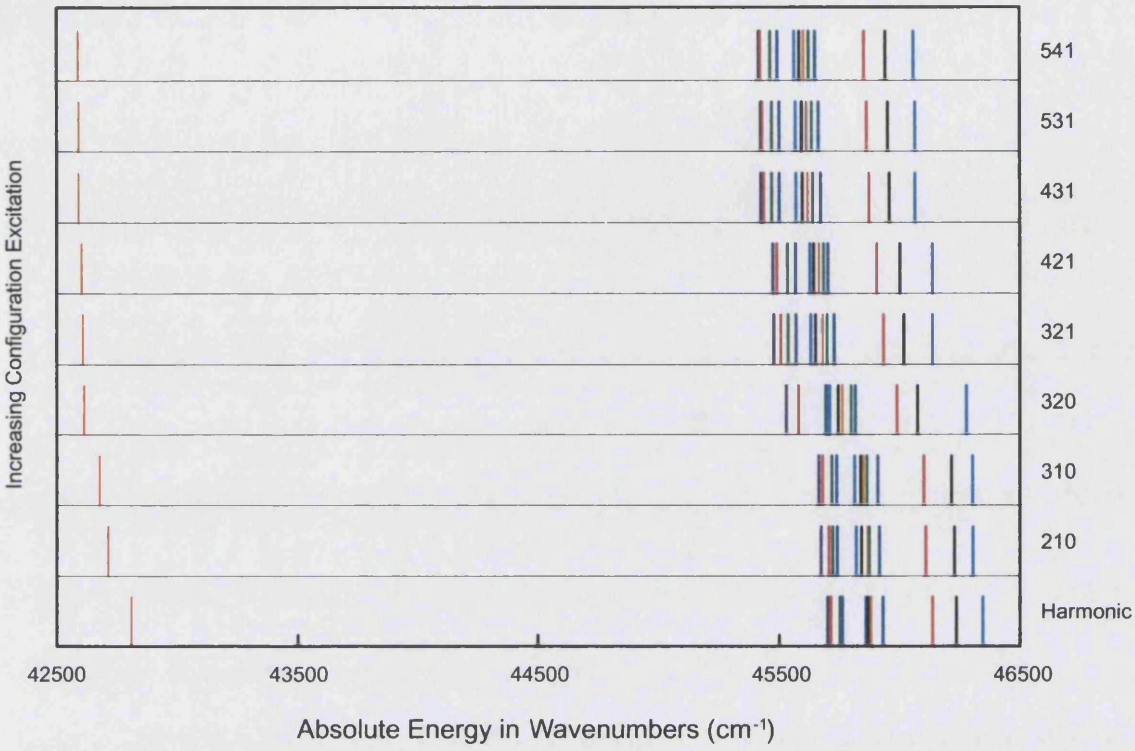


Figure 6.10: Absolute energies calculated for the tryptamine gauche-phenyl (out) conformer. Configurations include the 12 spectral modes with increasing SDT excitation. This demonstrates the variational behaviour of the calculation. Line colours are the same as for Figure 6.9 allowing a comparison.

The reverse effect is seen to be true in moving from 421 excitation to a 541. Here, the excited states are stabilised relative to the ground state resulting in a decrease in transition energies. This is probably due to the higher energy eigenstates in the single and double subgroups interacting more strongly with the first excited states than the ground state, as they are closer in energy. Such a problem may be overcome by a more careful choice of configuration excitation. Including quadruply excited states of the type Ψ^{Q}_{ijkl}

$$\Psi^{Q}_{ijkl} = \sum_{i=1}^{3N-9} \sum_{j>i}^{3N-8} \sum_{k>j}^{3N-7} \sum_{l>k}^{3N-6} \sum_{a=1}^{a_{\max}} \sum_{b=1}^{b_{\max}} \sum_{c=1}^{c_{\max}} \sum_{d=1}^{d_{\max}} \psi_i^a(q_i)\psi_j^b(q_j)\psi_k^c(q_k)\psi_l^d(q_l) \prod_{i' \neq i,j,k,l} \psi_{i'}^0(q_{i'})$$

or introducing two quanta of energy into each of the triples, in a 632 configuration, for example, may improve the correlation between results of CI calculation and experiment. However, the size of calculation required to probe these theories prohibits further research on the current computer hardware.

In the following sections, results of the CI calculation will be compared with experiment in order to check assignment of spectra to individual conformers and attempt assignation of more complex spectra. A configuration must be found that will produce results which compare well with experiment. It is clear from Figure 6.4 that the CI calculation correlates better with experiment as more modes are included in the configuration. The *SDT* excitation chosen must be small enough to allow the inclusion of as many modes as possible. However, including too few excited states yields results far from convergence. The Figure 6.9 suggests that triples must be included, so the smallest configuration that produces meaningful results is 321. With this configuration, it is possible to perform a calculation on up to 20 modes - a calculation on more modes will make the CI matrix larger than the stack size of the computer used to perform these simulations. All further calculations are performed with a 321 excited configuration.

In Section 6.3.3, an attempt is made to reduce the number of modes that must be included in the CI space in order to achieve convergence. The aim is to reduce the time taken for the CI calculation by reducing the total number of configurations and therefore the number of CI integrals that must be evaluated. The technique introduced previously is to reduce the number of normal modes included so, working systematically, an attempt should be made to reduce the SDT excitation. In a variational calculation, the closer the resemblance of the trial wavefunction to the true wavefunction, the fewer excited trial functions needed in the configuration to achieve converged results. Therefore, in the next section, an attempt is made to improve the initial trial wavefunctions.

6.3.5 Optimising the Basis Functions

The configuration can be improved by adapting the trial basis wavefunction to show a greater resemblance to the true wavefunction. Currently, the trial wavefunction is that of an harmonic oscillator. Associated with the basis functions is a potential energy surface. The harmonic oscillator wavefunction is a solution of the vibrational Hamiltonian with a potential of the form $a(x - x_e)^2$. Figure 6.11 shows the deviation of

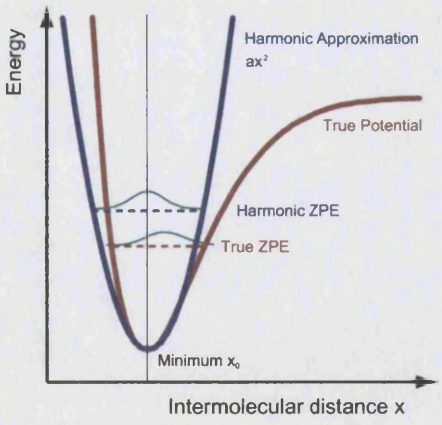


Figure 6.11: Harmonic and true potential energies.

this trial harmonic potential, in blue, from the true potential, in red. The wavefunctions are plotted in green. The process of calculating harmonic wavefunctions involves assuming terms of order 3 or more are negligible in the Taylor expansion of the true potential near the minimum i.e. at low energy, the potential scales roughly $(x - x_e)^2$

The Wilson **GF**[179] method of making the harmonic approximation and solving the vibrational Hamiltonian is described in Section 2.6. Briefly, diagonalising the mass weighted force constant matrix yields the normal modes. The associated wavefunctions are[179,184]

$$\psi_v = N_v H_v(y) e^{-y^2/2} \qquad y = (\omega/\hbar)^{\frac{1}{2}} q \tag{6.9}$$

where v is the vibrational quantum number. The vibrational frequency of the normal mode, ω, is linked to the force constant of the potential, k, and the reduced mass of the system, μ.

$$\omega = \sqrt{k/\mu} \tag{6.10}$$

The harmonic approximation assumes that a real potential is roughly quadratic at the bottom of the well. Effectively, this is a minimisation of the difference between the true and trial potentials at the minimum. As the zero-point energy of this system is higher in energy than the minimum, there is a substantial deviation between trial and true potentials at this point, as seen in Figure 6.11. The true potential is more attractive than the trial harmonic potential at $x < x_0$ and less attractive at $x > x_0$. On average however, the true potential is not as attractive, implying a smaller force

constant and consequently a lower vibrational frequency. This is shown in the lower zero-point energy and the broader, more diffuse nature of the true wavefunction.

The zero-point energy is the first to converge in the CI calculation. In order to reduce the *SDT* excitation needed to reach convergence, the trial wavefunctions should resemble the true wavefunction as much as possible at this point. This can be achieved by minimising the difference between the trial and true potentials, i.e. the anharmonicity, of the ground state. This is, in effect, an optimisation of the basis functions. As the CI technique exploits properties such as the orthonormality of harmonic wavefunctions, the new trial functions must be of the same form, as described in Equation 6.9. Therefore, the variable used in the minimisation must be the exponential term y. This corresponds to the force constant of the bond and defines the curvature of the harmonic potential. The force constant may be adjusted to minimise the anharmonicity of the ground state, calculated by evaluating the integral

$$\begin{aligned} &\langle \psi_0 | \sum_{i=1}^{3N-6} \left(-\frac{1}{2}\frac{\partial^2}{\partial q_i^2} + \frac{1}{2}kq_i^2 \right) + \hat{\mathbf{V}} - \sum_{i=1}^{3N-6} \frac{1}{2}kq_i^2 \, | \psi_0 \rangle \\ = \quad &\langle \psi_i | \, \hat{\mathbf{A}} \, | \psi_i \rangle + \frac{1}{2}\omega_i \end{aligned} \tag{6.11}$$

A new section of code is added to the program in order to perform this optimisation. A Newton-Raphson minimisation technique is chosen for this as the starting point, the harmonic force constant, is close to the minimum for the strong, intramolecular bonds under study. The technique is tested and the results are shown in Figure 6.12. This provides a comparison of CI calculations with a configuration of the 12 IR active

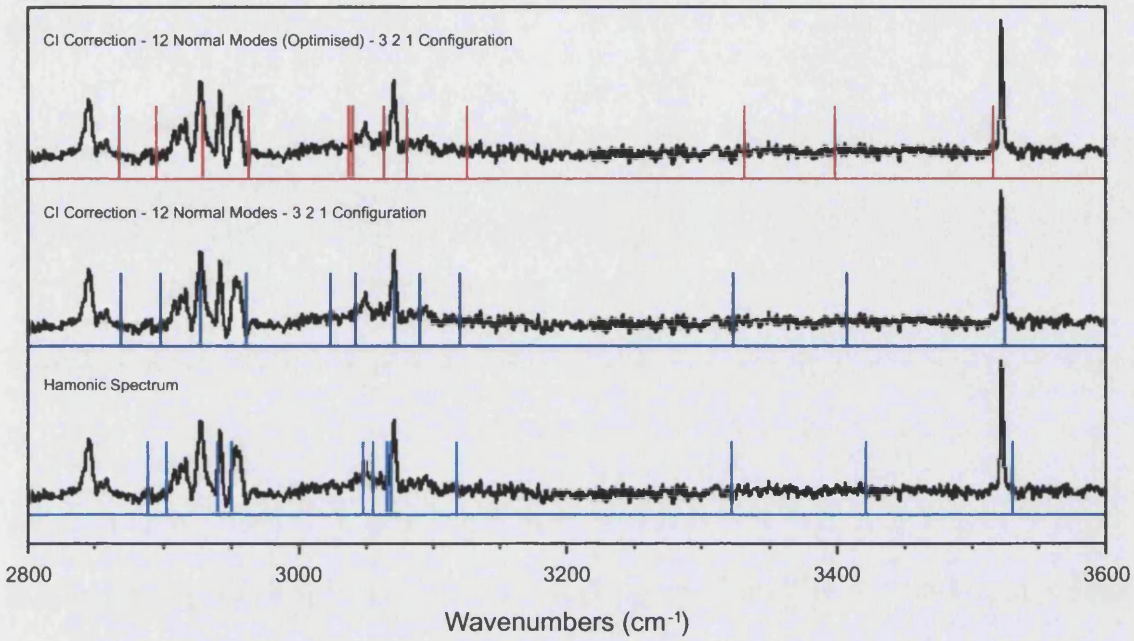


Figure 6.12: Tryptamine gauche-phenyl (out) conformer. Comparison of experimental spectrum[66,67] (corresponding to the C2 line in the one colour resonant two-photon ionisation spectrum of TRA^+) with an harmonic, a 12 mode CI and a 12 mode optimised α calculation.

normal modes at 321 excitation both before and after optimisation. The harmonic and experimental spectra are also provided in order to gauge the performance of the optimisation.

The results of the CI calculation with an optimised basis shows little difference to the calculation with a normal mode basis, particulary in the low energy region of the spectrum. Further investigation into optimisation of the first excited state wavefunctions is undertaken but again, no significant improvements are seen. In order for this technique to work successfully, the form of the basis functions must

be adaptable. A basis of Morse or a linear combination of harmonic oscillator wavefunctions will show a greater resemblance to the true wavefunctions. However, the CI method makes use of the orthonormality of harmonic oscillator wavefunctions and the analytical solutions of the harmonic oscillator Hamiltonian in order to speed up the calculation dramatically. Switching to more complex basis functions will allow a calculation with a smaller basis to reach somewhat converged results but this may not in fact result in a shorter calculation. The speed gains made will most likely be greatly outweighed by the losses due to a more complex calculation. Another major problem with the use of more complex basis functions will be the loss of generality of the technique which is a cornerstone of this project.

The effect of the basis set and the configuration on the CI calculation have been studied in depth. In the next section, the results of the CI calculation with the best possible configuration are compared with experimental spectra for some conformational isomers of tryptamine.

6.4 Results

Zwier and co-workers[66–69] measured gas phase resonant ion-dip infrared spectra of tryptamine and the other tryptophan analogue 3-indole-propionic acid, IPA, and their water clusters. Supersonic expansion cooling is used to remove vibrational hot bands and reduce the rotational temperature in order to simplify the spectra. In

the supersonic expansion, the gaseous tryptamine molecules are collisionally cooled to very low temperatures. Of the nine low-lying conformers of tryptamine, shown in Figure 6.3, seven are thermally accessible at the pre-expansion temperature.[67] These are cooled in the expansion into their associated zero-point levels. Zwier *et al.* provide a spectrum for all conformers, as double resonance photo-ionisation leads to different conformers appearing in the same mass channel. UV-UV hole burning is used to selectively ionise individual isomers for the spectra which are subsequently identified using UV spectroscopy, calculated IR spectra and rotational constants. These spectra are compared with the results obtained by the CI technique using the MM3Pro potential.

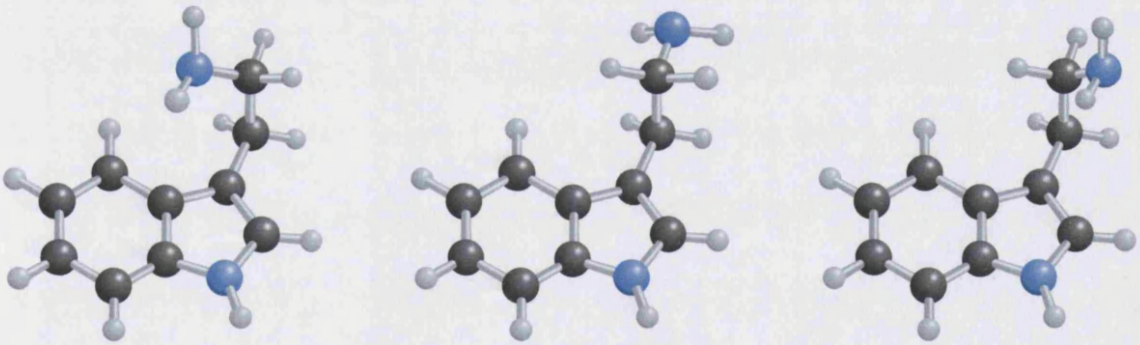


Figure 6.13: The three low energy conformers of tryptamine chosen for study.

Of the seven conformers for which spectra are available, three, gauche-phenyl (out), anti (up) and gauche-pyrrole (out), are chosen for study. As shown above, in order to obtain the best results, the configuration must be chosen in such a way as to include

as many excited vibrational states as possible at 321 excitation. Consequently, the results of CI calculations for these three isomers performed with a 321 excited configuration on the 20 highest energy normal modes is given in Table 6.4. Results of harmonic calculations on the MM3Pro potential and on an ab initio density functional potential at a Becke3LYP/6-31+G*(5d) level of theory, as presented by Carney and Zwier,[66] are provided for comparison.

Table 6.4: Comparison of harmonic and CI spectra with those calculated by Carney *et al.*[66] Energies in wavenumbers (cm^{-1}.

Anti(up)			Gpy(out)			Gph(out)		
Harmonic*	CI†	Carney‡	Harmonic	CI	Carney	Harmonic	CI	Carney
2892	2858	2892	2889	2857	2856	2889	2842	2841
2903	2876	2926	2901	2890	2912	2903	2874	2912
2943	2914	2933	2941	2920	2947	2941	2904	2942
2952	2946	2975	2947	2953	2961	2951	2941	2954
3048	3027	3051	3043	3035	3050	3048	3025	3052
3056	3039	3057	3052	3040	3058	3056	3042	3062
3069	3064	3068	3066	3064	3069	3070	3073	3075
3067	3085	3079	3064	3080	3080	3066	3094	3084
3118	3121	3129	3111	3119	3136	3118	3128	3129
3323	3327	3336	3320	3325	3341	3323	3331	3343
3429	3414	3418	3421	3407	3424	3423	3411	3423
3541	3533	3524	3527	3520	3524	3532	3527	3523

*Harmonic calculation on MM3Pro potential.

†Configuration interaction anharmonic corrections calculated on MM3Pro potential.

‡Harmonic calculation of IR frequencies on ab initio potential using Becke3LYP/6-31+G*(5d) level of theory.

In general, a decrease in transition frequency is seen on comparing harmonic with CI calculations on the MM3Pro potential. The drop in energy is seen to be more marked

in the low energy end of the spectrum. This is due to the greater anharmonicity of the low energy vibrations. There tends to be good correlation between the CI calculations and the harmonic ab initio calculations of Carney *et al.* . These trends are better illustrated visually and the results are plotted as line spectra below. Carney provides line intensities alongside the transition frequencies so these are also plotted.

6.4.1 Comparison with Experimental Spectra

In Figures 6.14-6.16, the harmonic and CI spectra for a 12 and 20 mode configurations are plotted against the assigned experimental spectra. Ab initio harmonic frequencies calculated by Carney are included to allow a comparison between MM3Pro and ab initio surfaces. Note that the structure of tryptophan is highly dependent on Van der Waals forces which are not well described by the DFT method employed by Carney.

6.4.2 Tryptamine Gauche-Phenyl (Out) Conformer

It is clear from Figure 6.14 that the harmonic calculation on the MM3Pro surface is a poor match with the experimental spectrum in the low energy, 2800-3200cm^{-1}, region. Similarly, the results of a CI calculation with a configuration of the 12 highest

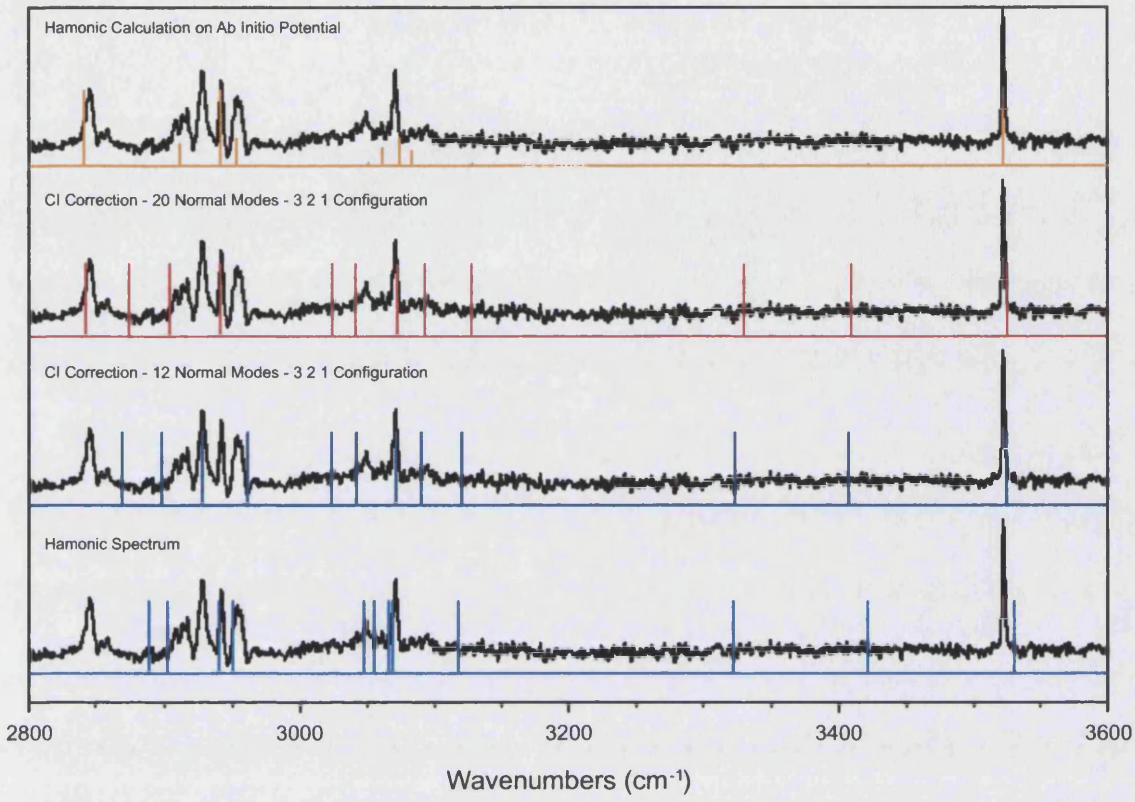


Figure 6.14: Tryptamine gauche-phenyl (out) conformer. Comparison of experimental spectrum[66,67] (corresponding to the C2 line in the one colour resonant two-photon ionisation spectrum of TRA^+) with an harmonic and two CI calculations on the MM3Pro potential. Results of an harmonic ab initio calculation by Carney, using Becke3LYP/6-31+G*(5d) level of theory,[66] are included for reference.

energy modes do not correlate well with experiment in this part of the spectrum. The calculation with a configuration including the 20 highest energy normal modes, however, compares well with ab initio results and experiment. This suggests the vibrations in the 2800-3200cm^{-1} region are anharmonic and that anharmonic terms in the MM3Pro potential are significant.

However, the harmonic calculation by Carney on an ab initio surface also compare quite well to experiment. It is important to note that the results of Carney are scaled by 0.9603 to fit the intense peaks at 3520cm^{-1} and 3065cm^{-1} so this part of the Carney spectrum reproduces experiment well by default. However, the lower energy modes also compare well to experiment. Only one scaling factor has been used for all the spectral lines. Use of only one scaling factor lowers the energy of high energy modes more than low energy modes on an absolute scale.

6.4.3 Tryptamine Anti (Up) Conformer

A similar pattern is seen in this spectrum, whereby the harmonic and 12 mode calculations on the MM3Pro surface fail to reproduce the true spectrum accurately. For this system however, the ab initio calculation also fails to model the lower energy modes. If the modes chosen for fitting (3520cm^{-1} and 3065cm^{-1}) are excluded, there is poor agreement between experimental and calculated peaks. The simulation with the 20 highest energy modes in the CI space displays a far better agreement with experiment.

As in the spectrum of tryptamine gauche-phenyl (out), TRA-GPHO, there appear to be more absorption peaks in the experimental spectrum than the calculated spectrum. This could be a result of overtones or combination bands, especially in the case of TRA-GPHO where there are many peaks of low intensity. Overtones and

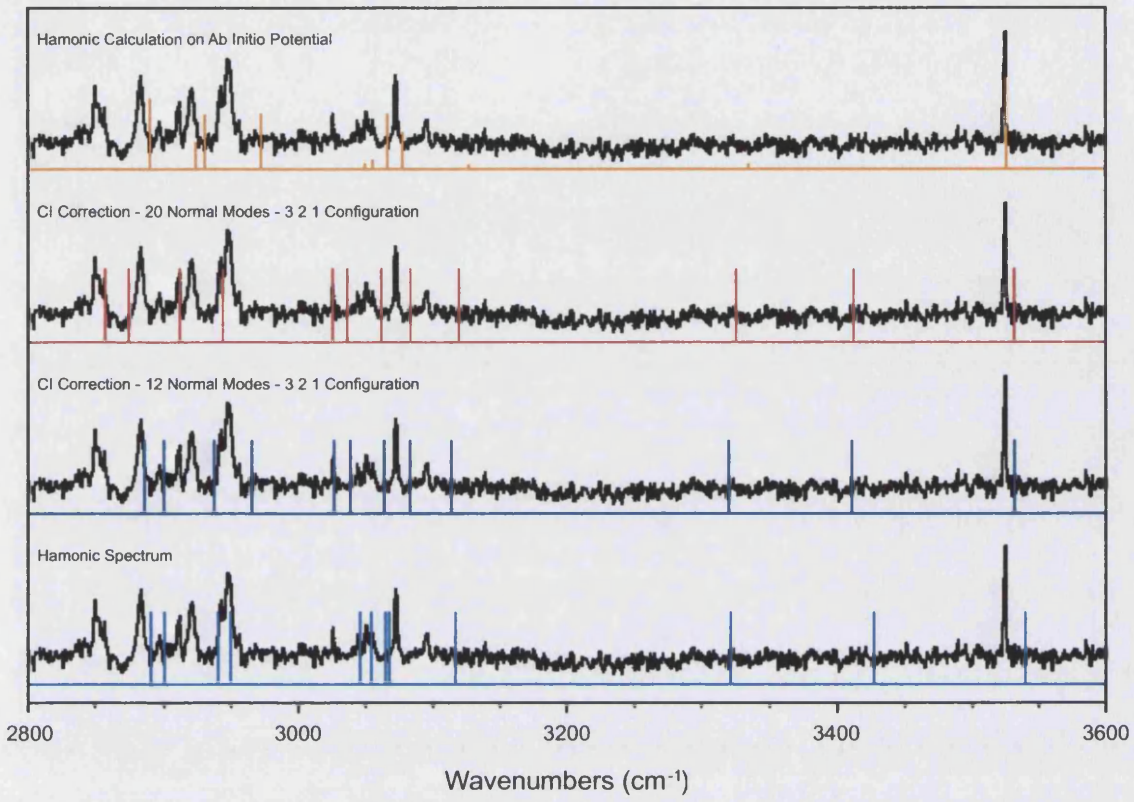


Figure 6.15: Tryptamine anti (up) conformer. Comparison of experimental spectrum[66,67] with an harmonic and two CI calculations on the MM3Pro potential. Results of an harmonic ab initio calculation by Carney, using Becke3LYP/6-31+G*(5d) level of theory,[66] are included for reference.

combination band are forbidden in the harmonic approximation. However, the anharmonic nature of this system means that weak absorptions due to transitions of $\Delta v = 2$ or the excitation of two fundamentals by one photon are possible, though weak. The presence of these lines makes it very hard to fit spectra. It may appear that the calculated CI or ab initio spectral lines have shown good agreement with experiment but, in fact, they can be inaccurate as they are corresponding with over-

tones or combination bands and not the fundamental lines. Calculating transition intensities can help rectify this problem but not in all cases as weak fundamental lines can look identical to overtones and combination bands as these exhibit weak absorption in most cases.

6.4.4 Tryptamine Gauche-Pyrrole (Out) Conformer

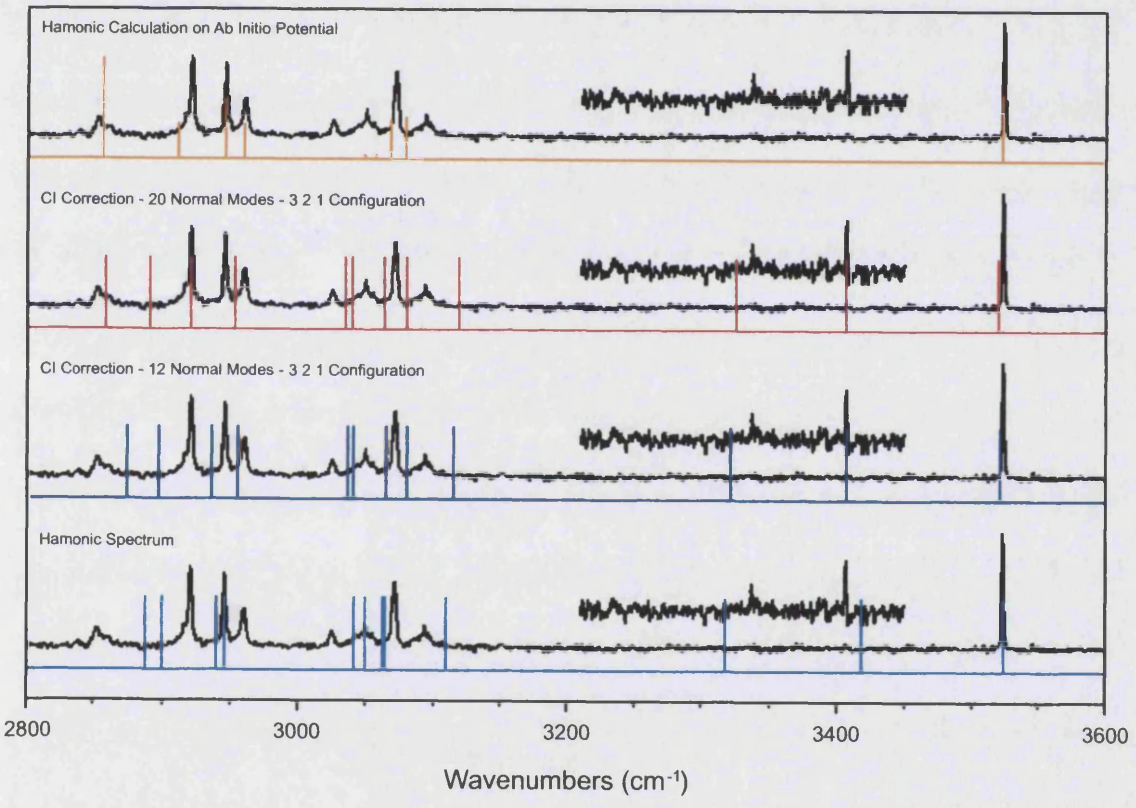


Figure 6.16: Tryptamine gauche-pyrrole (out) conformer. Comparison of experimental spectrum[66,67] with an harmonic and two CI calculations on the MM3Pro potential. Results of an harmonic ab initio calculation by Carney, using Becke3LYP/6-31+G*(5d) level of theory,[66] are included for reference.

The MM3Pro results are seen to show a greater correlation with experiment on going from an harmonic to a 20 mode CI calculation. In this spectrum, the low intensity peaks in the 3300-3400cm^{-1} part of the spectrum have been amplified. The calculated transition at 3410cm^{-1} demonstrates particularly good agreement with experiment. However, modes in the 3000-3100cm^{-1} region do not correlate well with the experimental spectrum. For this conformer, the scaled harmonic ab initio calculation performs quite well, even for lower energy modes.

For the three conformers studied, results of CI calculations consistently display greater correlation with experiment than harmonic results. It is clear that in all cases, more than 12 modes must be included in the CI space to achieve good results. Generally, the performance of CI is comparable to that of the ab initio calculations but calculations take 1-2 hours to complete compared to 20-30 hours for an ab initio calculation of this level, and this time can be further reduced. It is also important to note that the CI technique requires no fitting/scaling so there is no need for experimental spectra. The results show clearly that the anharmonic terms in the MM3Pro potential are significant for calculating vibrational frequencies of these biomolecules. The results also suggest that parameterising force-fields for biomolecules must be done with care if vibrational frequencies are used without anharmonic modifications.

Chapter 7

Tryptamine Water Clusters

In biological systems, tryptophan and the proteins of which it is a constituent, are found in solvated environments. It is important to study how the structure and bonding is affected in moving from gaseous to real environments. The biochemical reactions of amino acids take place in aqueous media. The hydrophobic nature of the large indole group of tryptophan is thought to play a key role in drug binding sites and protein folding. The amine and acid groups of the tryptophan are involved in the polymerisation that creates the protein backbone, leaving the indole ring as the functional group exposed on the surface. As the largest amino acid, it is one of the most hydrophobic. Protein folding can be compared to formation of micelles, with hydrophobic residues on the protein interior and polar residues on the surface.[1,2] In the monomeric form, tryptophan has three potential binding sites. Studies can help in finding the site preferred by solvent molecules.

Water is the key biological solvent, so studies of water-tryptamine complexes are extremely important. Zwier and Carney obtained spectra of tryptamine-water$_n$ $n = 1 - 3$ complexes by introducing solvent vapour into the supersonic expansion along with the tryptamine.[68,69] The cluster size, i.e. the value of n, is controlled by the relative concentration of solute to solvent and the absolute pressure of the molecular beam.

In the case of tryptamine monomer, seven isomers are found in the supersonic expansion cooled mixture. The molecules in the expansion are vibrationally and rotationally cold but the presence of significant quantities of different conformers in the pre-expansion nozzle results in a mixture of conformers in the cooled jet. In the case of tryptamine-water$_n$, however, only a single conformer is found for each cluster of size, n. Zwier identifies the minimum energy conformations of these clusters on a DFT potential.[68,69] However, no attempt is made to assign the spectra. In this chapter, an attempt is made to identify the conformers by comparison of calculated and experimental spectra.

7.1 Tryptamine-Water

The MM3Pro potential has 21 conformational isomers of the tryptamine-water$_1$ complex. In Figure 7.1, five low energy conformers chosen for study are depicted, including the minimum energy structure, **A**, as predicted by Zwier and confirmed

by the MM3Pro potential and, **C**, a structure very similar to it.

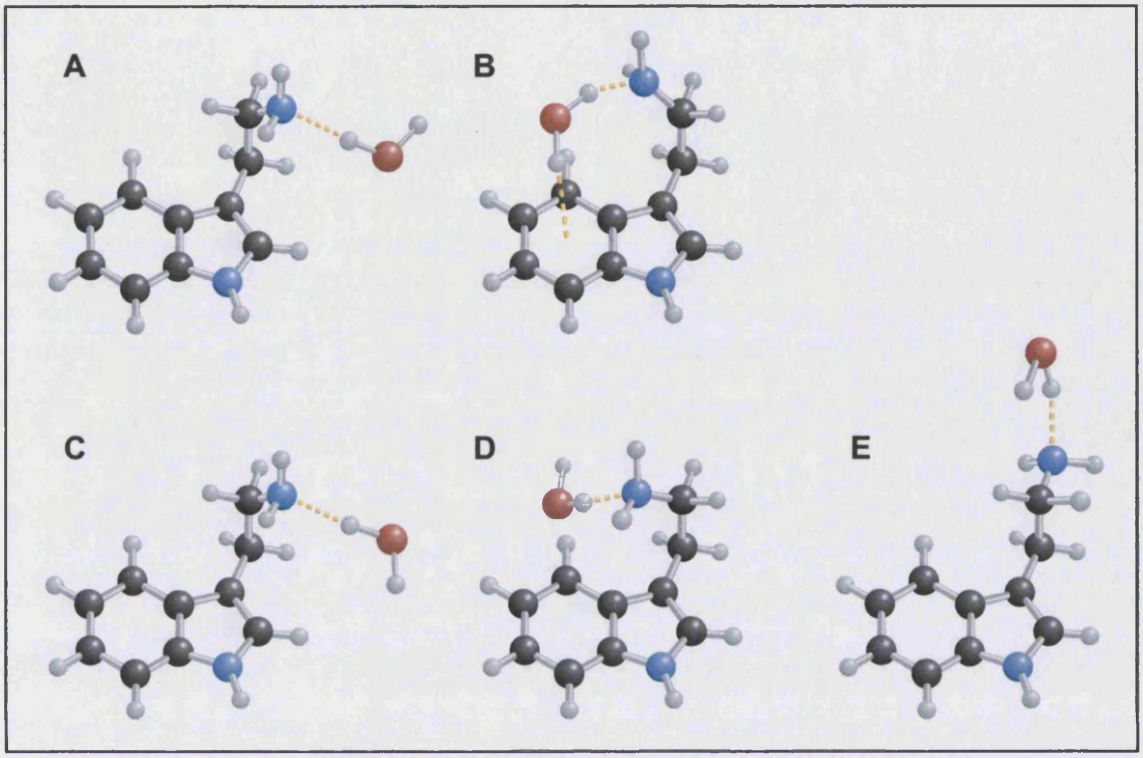


Figure 7.1: Tryptamine-water$_1$ conformations chosen for calculation.

There are 27 atoms in the system with 75 normal modes. 14 of these fall in the 2800-3800cm^{-1} region of the electromagnetic spectrum so a minimum of 14 modes must be included in the calculation. The best possible results are needed so 20 normal modes are included in the configuration and CI calculations are performed for all 5 conformers depicted above. These are compared with each other and experiment in Figure 7.2.

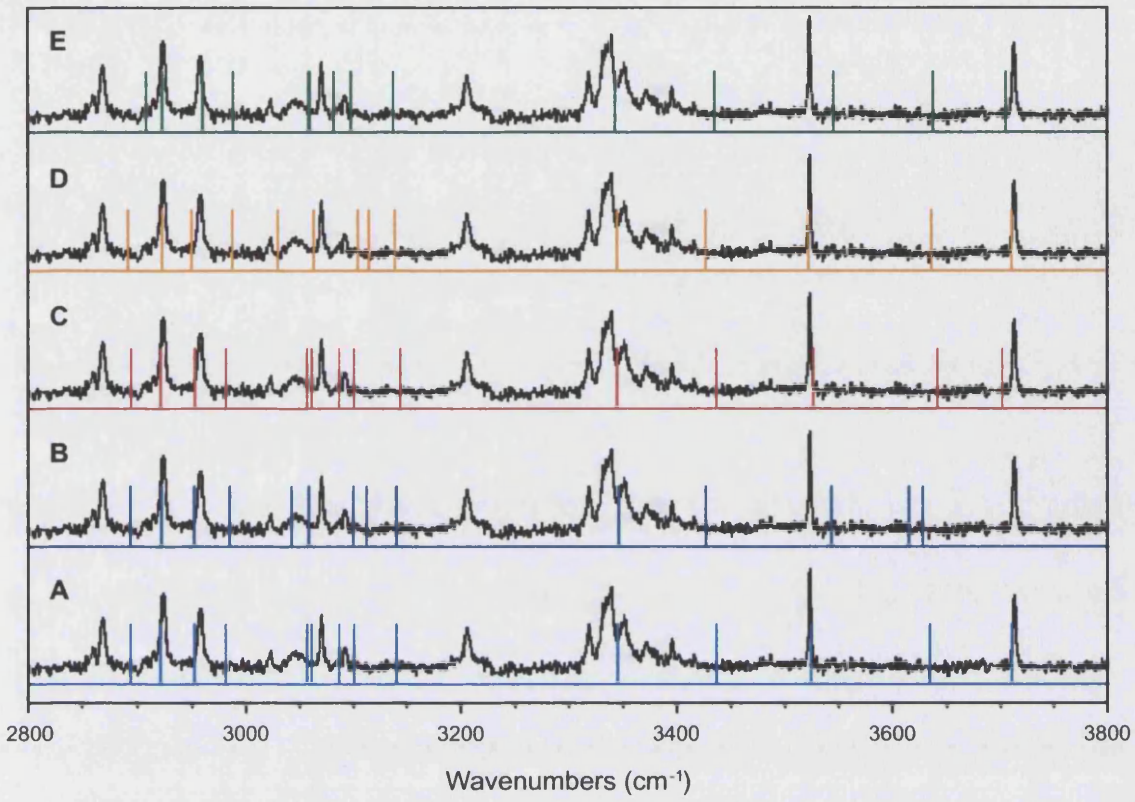


Figure 7.2: Tryptamine-water$_1$ spectra calculated with a 321 configuration of the last 20 normal modes for the conformers **A-E**. The highest three energy modes are assigned by Zwier[68,69] to the Free OH stretch (3715cm^{-1}), Indole N-H stretch (3522cm^{-1}) and H-Bond OH stretch (broad 3340cm^{-1}).

Five of the 21 conformational isomers found on the MM3Pro surface are similar to conformer **B** in that the solvent molecule acts as a hydrogen donor in two hydrogen bonds. Zwier assigns the three highest energy modes to an O-H stretch (3715cm^{-1}), indole N-H stretch (3522cm^{-1}) and another O-H stretch (broad 3340cm^{-1}). Comparing calculated and experimental spectra it can be concluded that the complex found in the molecular beam cannot contain two hydrogen bonds. The system must

have a free water hydrogen as the calculated vibrational frequency for one O-H stretch is significantly lower for conformer **B** than for the experimental spectrum. Isomers **C** and **E** are also predicted to have the free O-H stretch at a slightly lower energy than shown by the experimental spectrum. Isomers **A** and **C** differ only in the orientation of the free hydrogen in the water molecule. The calculated vibrational frequency of this bond stretch is lower for system **C** than that of **A** and the experimental value presumably due to some small attractive interaction between the free hydrogen and the indole nitrogen. This also shows up in a small effect on the position of the indole N-H stretch. This effect may be seen in the true spectrum of the conformer **C** or it may be a failing of the potential as seen later in Section 7.2.

The indole N-H stretch ($3522cm^{-1}$) in this spectrum appears in an identical position to the tryptamine monomer spectra. This implies that neither the indole nitrogen and hydrogen atoms nor the conjugated indole π electrons are involved in any H-bonding to water. The calculated spectra for all isomers show this behaviour with the exception of **B** and **E**. It is clear that the increase in N-H bond stretch frequency in conformer **B** is due to an interaction between the indole π electrons and a hydrogen atom on the solvent molecule. This leads to an increase in the bond strength of the indole N-H bond and a resultant increase in its vibrational frequency. The frequency shift of this bond in isomer **E** is not so easily explained. It may be linked to the decrease in vibrational frequency of the free O-H stretch. However, this seems unlikely given the conformation of the molecule.

The calculated spectra of conformers **A** and **D** bear the closest resemblance to the experimental spectrum. **A** and **D** differ in the conformation of the tryptamine component, being based on the gauche-pyrrole (out) and gauche-phenyl (out) conformers respectively. The CH_2 group α to amine on tryptamine are in different environments in these two isomers. The stretching frequency of this group falls in the 2800-3000cm^{-1} region. In this part of the spectrum, there is poor agreement between the calculated and experimental spectra. It is difficult to assign the experimental spectrum conclusively to conformer **A** or **D** as it could be either or one of the other isomers not studied.

7.2 Tryptamine-(Water)$_2$

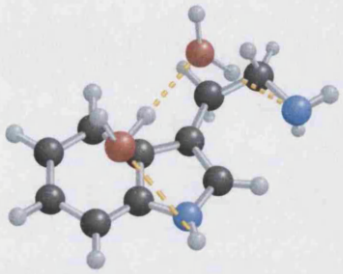


Figure 7.3: Tryptamine-Water$_2$ global minimum predicted by Zwier.[68,69]

As with the tryptamine-water$_1$ complex, only one conformational isomer is found in the supersonic expansion cooled molecular beam. This is not assigned to a structure though Zwier gives the minimum energy structure on the ab initio potential, reproduced in Figure 7.3. The minimum energy structure predicted by Zwier involves the indole NH acting as a hydrogen donor and the amine group acting as a hydrogen-acceptor.

However, while 121 minima corresponding to conformational isomers, are found on the MM3Pro surface, none of those found resemble the structure predicted by Zwier. The MM3Pro surface predicts indole NH will act as an hydrogen acceptor. Despite adding bonding parameters and adjusting dipole moments and Van der Waals coefficients, the minimum found by Zwier cannot be reproduced without changing the MM3Pro potential to the extent that the parameters become meaningless. As a result six low energy minima of interest, depicted in Figure 7.4, are chosen for study with the aim of testing the hydrogen bonding scheme predicted by MM3Pro. An attempt is made to assign the structure of the species in the experimental spectrum and comparing the different bonding rationales predicted by ab initio and molecular modelling potentials.

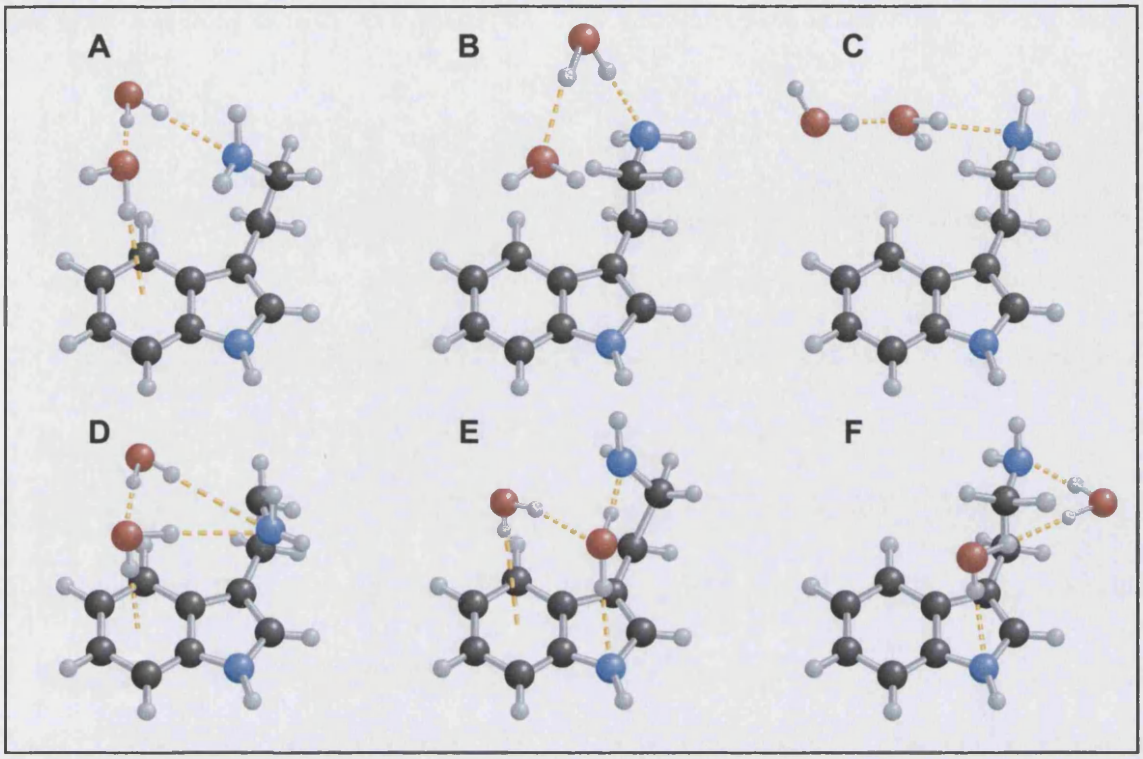


Figure 7.4: Clusters **A-F** are tryptamine-water$_2$ conformations chosen for study. Note the Indole NH acts as a hydrogen acceptor in hydrogen bonding in the MM3Pro potentials but Zwier suggests it is a hydrogen donor as shown in Figure 7.3

The conformers are chosen as a representative sample of the various hydrogen bonding schemes predicted by the MM3Pro potential. The 30 atoms in the system have 84 vibrational modes, of which 16 fall into the 2800-3800cm^{-1} region of the spectrum. Results for CI calculations of the spectra of the conformers with a configuration of 20 normal modes are compared with the experimental spectrum in Figure 7.5.

Zwier assigns the two highest frequency lines (3710cm^{-1} and 3715cm^{-1}) to free O-H stretches. The two intense broad peaks at 3490cm^{-1} and 3175cm^{-1} are assigned to

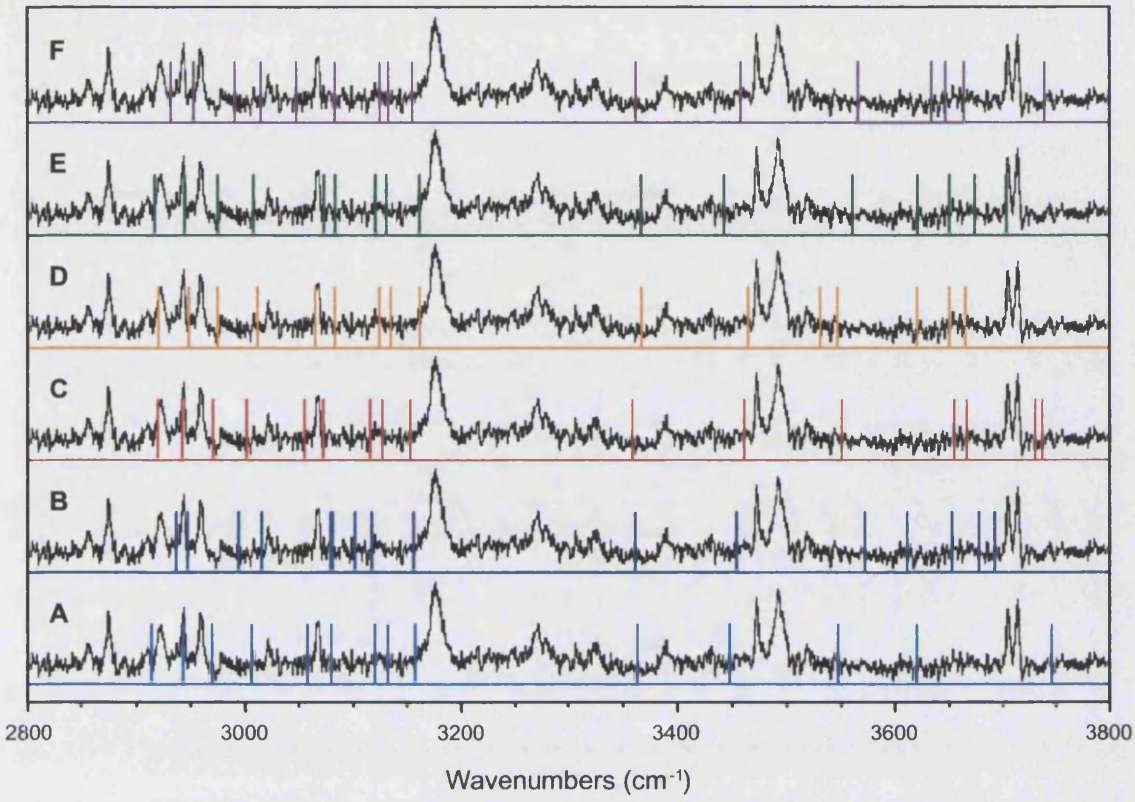


Figure 7.5: Tryptamine-water$_2$ spectra calculated with a 321 configuration including the last 20 normal modes for the conformers **A-F**.

H-Bond O-H stretches and the intense peak at 3470cm^{-1} is assigned to the indole N-H stretch. Clearly, none of the calculated spectra match these peaks and it can be seen that the bonding schemes predicted by the MM3Pro potential are incorrect. It is not possible, therefore, to suggest a structure for the species studied by Zwier. Close inspection of the MM3Pro potential shows that there is no hydrogen bonding parameters for interaction between indole H and water O. Many hydrogen bond strengths and lengths for this interaction were tested with no success. Attempts at changing Van der Waals parameters and dipole moments were made. However,

no minimum energy structures of the type found by Zwier were found. Note that such anomalous solvent-biomolecule behaviour in molecular modelling potentials has been reported elsewhere.[49]

Zwier also measured the IR spectrum of a tryptamine-water$_3$ complex but the MM3Pro potential is not able to predict an accurate structure for this cluster and attempts at identifying the conformer in the molecular beam would prove futile.

Part V

Conclusions

Chapter 8

Conclusions

In the last two Chapters, the CI technique for calculating anharmonic effects has been applied to real systems of biological importance. The results have shown promising agreement with experimental spectra suggesting that the technique works well. The improved agreement with experiment, especially for lower energy modes, on moving from harmonic to anharmonic energies suggests a degree of anharmonic behaviour in these systems and that the MM3Pro potential includes terms that allow these anharmonic effects to be accurately modelled. The CI technique is seen to be a useful technique to model such anharmonic effects.

The CI technique is applied effectively to tryptamine monomer and the complex tryptamine-water$_1$. It has been shown to be a useful tool to aid in assigning spectra,

requiring no scaling or fitting, producing results similar to scaled ab initio calculations of Carney in a fraction of the time. On moving to the tryptamine-water$_2$ cluster, the MM3Pro potential breaks down, unable to predict the correct hydrogen bonding schemes and preventing the identification of the species responsible for the experimental spectra. This is attributed to be a failing of the potential as no minima similar to those found on ab initio surfaces can be found. For such hydrogen bonded systems more accurate potentials will be needed.

A general method and computer program has been developed for calculating the vibrational states of biomolecules. A particular feature of this technique is the systematic treatment of anharmonicity using a variational method. The method is first tested on simple benchmark systems such as the Hénon-Heiles potential and the water molecule. It is then applied to biomolecules chosen for their significant medical interest. Promising comparisons with experimental data are made. The results show that there are significant anharmonic effects on the vibrational states in these systems in the potential energy surface used. The method will have useful future use in testing potential energy surfaces in biomolecular systems and predicting vibrational spectra.

Chapter 9

Further Computational Work

Calculation of vibrational eigenstates for larger systems is desirable. As such, the CI program should be adapted to use multiple arrays to hold and diagonalise the CI matrix.

The rate limiting part of the calculation is the calculation of the CI matrix. Numerical evaluation of integrals when calculating the expectation value of the anharmonicity, $\hat{\mathbf{A}}$, requires many calculations of the potential at the abscissae. As more complex potential energy surfaces and larger systems are studied, the time needed for each evaluation of the potential energy increases. Improvements on the speed of calculation have been made by storing some commonly used integrals in the calculation of the CI matrix. This reduces the number of calls to the potential. Further

speed increases by reducing the number of potential calls would be important.

The quadrature abscissae, q_i, and weights, w_i, are evaluated at the start of the CI program, and they remain constant throughout the calculation. In calculating 1D integrals, the potential is evaluated at each abscissa.

$$\begin{aligned}\langle\psi_1^0|\,\hat{\mathbf{A}}\,|\psi_1^1\rangle &= \int_a^b \psi_1^0\left(V(q)-\frac{1}{2}k_1q_1^2\right)\psi_1^1\mathrm{d}q \\ &= \int_a^b N_1^0H_1^0(q)e^{-y^2/2}\left(V(q)-\frac{1}{2}k_1q_1^2\right)N_1^1H_1^0(q)e^{-y^2/2}\mathrm{d}q \\ &\approx N_1^0N_1^1\sum_{i=1}^{N}w_iH_1^0(q)H_1^1(q)\left(V(q_i)-\sum_{j=1}^{3N-6}\frac{1}{2}kq_j^2\right) \qquad (9.1)\end{aligned}$$

Where $y = (\omega/\hbar)^{\frac{1}{2}}q$. To aid visualisation of this technique, a 2D plot of the two surfaces $V(q)$ and $\frac{1}{2}k_iq_i^2$ for the HMS water monomer potential is given in Figure 5.4. In the expansion of the integral, the term $V(q)-\frac{1}{2}k_iq_i^2$ describes the anharmonicity and must be evaluated at each abscis point. As the abscissae are constant throughout the calculation, the value of this term can be stored as an "anharmonicity grid". An example 2D grid is shown for the water monomer in Figure 9.1. Clearly, the 1D grid for calculation of one-dimensional matrix elements is the central line in the 2D grid.

Once this grid is stored, the calculation of all the CI matrix elements involving these two modes can be calculated by performing a sum analogous to that described in Equation 9.1. A 2D sum is needed to evaluate matrix elements describing the interaction between the modes. Therefore, once the grids have been stored, the will be no need to make any more potential calls. The numerical integration would

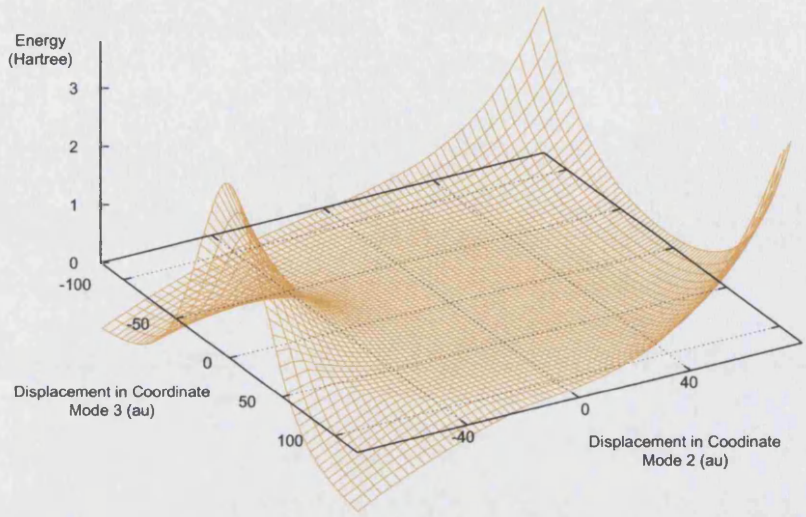


Figure 9.1: Plot of the anharmonicity grid of the Hoy Mills and Strey[99] water potential in normal coordinate space. The modes chosen as the coordinate space are the symmetric stretch (Mode 2) and bending (Mode 3) coordinates. The orange surface describes the anharmonicity of the HMS potential. The plot demonstrates that the harmonic approximation is only valid near the potential minimum.

require only a calculation of the relevant normalised hermites at the abscis points. It was shown in Figure 3.3 that the hermite polynomial and the quadrature weight multiply to model the wavefunction. This fact is shown in two dimensions in Figure 9.2. The blue surface is the weight multiplied by the normalised hermite for the integral $\langle \psi_2^1 \psi_3^1 | \, \hat{\mathbf{A}} \, | \psi_2^0 \psi_3^0 \rangle$. The orange anharmonicity grid stays constant for all integrals. Plotted together, these two surfaces depict the numerical integration process. Clearly, the anharmonicity is constant for all integrals. It is the wavefunctions of the normal mode basis functions that change between matrix elements. Low energy wavefunctions are concentrated near the potential minimum. There-

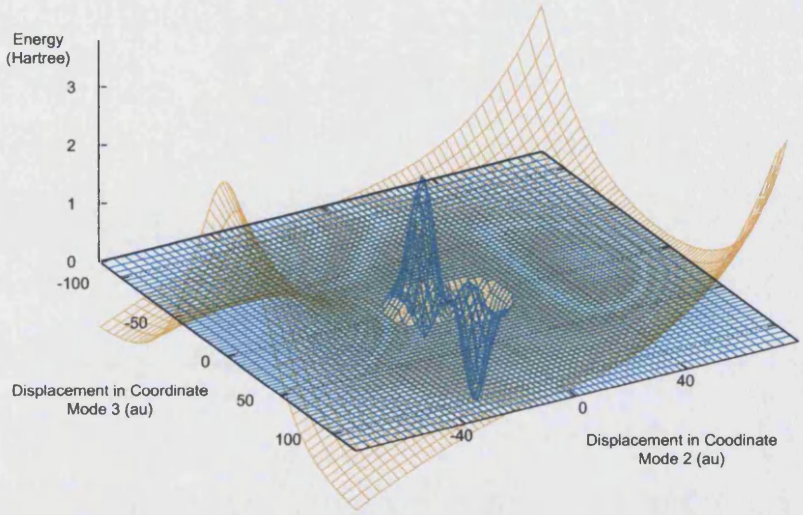


Figure 9.2: Depiction of the numerical integration method for evaluation of the matrix element $\langle \psi_2^1 \psi_3^1 | \, \hat{\mathbf{A}} \, | \psi_2^0 \psi_3^0 \rangle$. The orange surface describes the anharmonicity of the HMS potential. The blue surface describes the wavefunction.

fore, in low vibrational states, the system experiences little anharmonicity. It is only higher energy wavefunctions that are more diffuse and can sample regions of greater anharmonicity.

When calculating the grid, care must be taken to ensure it is constructed from an odd number of abscissae in each coordinate so that there is a central line at $q_i = 0$. This will allow calculation of 1 and 2D integrals using a single grid. There should also be enough abscissae to permit the evaluation of the integral at the required accuracy. For example, consider a calculation on a configuration of 632 excitation. The highest order hermites will be found in matrix elements of the type $\langle \psi_i^6 \psi_j^0 | \, \hat{\mathbf{A}} \, | \psi_i^0 \psi_j^6 \rangle$. Assuming that terms of order 5 or higher in the potential are negligible, the maximum order of the polynomial in the Gauss-Hermite quadrature

is 16. For Gauss quadrature techniques m abscissae enable exact integration of a polynomial function of order $2m + 1$. Therefore, a 9×9 anharmonicity grid would allow accurate calculation of the matrix elements in the 632 excited configuration calculation.

Further research into techniques for optimising the basis functions could prove valuable. Vibrational studies by Bowman,[87] Gerber[79,129,133–135,137] and Carter and Handy[154–159] make improvements to the normal mode basis functions by performing a SCF calculation. The SCF basis functions are then used as the basis for the CI configurations. Implementing a similar algorithm might prove valuable by speeding the convergence of the CI calculation.

After implementation of the above techniques for speeding calculation, it should be possible to use potential points calculated ab initio. This would enable direct comparison with the results obtained by calculation on molecular modelling potentials. Such comparisons allow an in depth check of the anharmonic terms in molecular modelling force fields.

To permit calculations on larger molecules, the CI code should be implemented on a machine with a larger stack size. The maximum CI matrix size can also be increased by storing matrix elements in more arrays. Once all elements are calculated, these arrays can be diagonalised as one.

Bibliography

[1] D M Eckert, V N Malashkevich, L H Hong, P A Carr, and P S Kim. *Cell*, 99:103, 1999.

[2] J P Moore, and T Dragic. *Nature*, 401:759, 1999.

[3] S Rowland-Jones. *Lancet*, 354:5, 1999.

[4] C Wild, T Oas, and C McDanal. *Proc. Natl. Acad Sci. USA*, 89:10537, 1992.

[5] S Jiang, K Lin, N Strick, and A R Neurath. *Nature*, 365:113, 1993.

[6] J M Kilby. *Nature Med.*, 4:1302, 1998.

[7] S Wang, J York, W Shu, M O Stoller, J H Nunberg, and M Lu. *Biochem.*, 41:7283, 2002.

[8] V Lucini, A Lucca, M Catalano, and E Smeraldi. *J. Affect. Disord.*, 36:129, 1996.

[9] M K Spillman, A J Van der Does, M A Rankin, R D Vuolo, and J E Alpert. *Psychopharmocology*, 155:123, 2001.

[10] D Attar-Levy, P J Schwartz, N E Rosenthal, and T A Wehr. *Therapie*, 53:489, 1998.

[11] D A Nielsen, M Virkkunen, J Lappalainen, M Eggert, G L Brown, J C Long, D Goldman, and M Linnoila. *Arch. Gen. Psychiatry*, 55:593, 1998.

[12] A C Swann, B A Johnson, C R Cloninger, and Y R Chen. *Psychopharmocology*, 143:380, 1999.

[13] J E Blundell, C L Lawton, and J C Halford. *Obes. Res.*, 4:471, 1995.

[14] K A Smith, C G Fairburn, and P J Cowen. *Arch. Gen. Psychiatry*, 56:171, 1999.

[15] T Trenque, D Piednoir, C Frances, H Millart, and M L Germain. *Pharmacoepidimiological Drug Safety*, 11:281, 2002.

[16] H G Nurnberg, J Lauriello, P L Hensley, L M Parker, and S J Keith. *J. Clin. Psychiatry*, 60:33, 1999.

[17] J L Finney. *Faraday Disc.*, 103:1, 1996.

[18] C L Brooks, and M Karplus. *Adv. Chem. Phys.*, 239:11, 1988.

[19] J M Goodfellow, M Knaggs, M A Williams, and J M Thornton. *Faraday Disc.*, 103:339, 1996.

[20] B R Brooks, R E Bruccoleri, and B D Olafson. *J. Comput. Chem.*, 4:187, 1983.

[21] L Nilsson, and M Karplus. *J. Comput. Chem.*, 7:591, 1986.

[22] A D MacKerrel Jr., J Wiorkeiwicz-Kuczera, and M Karplus. *J. Am. Chem. Soc.*, 117:11946, 1995.

[23] E Neria, S Fischer, and M Karplus. *J. Chem. Phys.*, 105:1902, 1996.

[24] S E Feller, D Yin, R W Pastorm, and A D MacKerrel Jr. *Biophysical Journal*, 73:2269, 1997.

[25] A D MacKerrel Jnr., B R Brooks, C L Brooks, L Nilsson, B Roux, Y Won, and M Karplus. The Encyclopedia of Computational Chemistry. John Wiley and Sons: Chichester, 1998.

[26] N Foloppe, and A D MacKerrel Jr. *J. Comput. Chem.*, 21:86, 2000.

[27] N Banavali, and A D MacKerrel Jr. *J. Comput. Chem.*, 21:105, 2000.

[28] J Aqvist. *J. Phys. Chem.*, 94:8021, 1990.

[29] W S Ross, and C C Hardin. *J. Am. Chem. Soc.*, 116:4363, 1994.

[30] D A Pearlman, D A Case, J W Caldwell, W S Ross, T E Cheatham III, S DeBolt, D Ferguson, G Seibel, and P Kollman. *Comp. Phys. Commun.*, 91:1, 1995.

[31] W D Cornell, P Cieplak, C I Bayly, I R Gould, K M Merz, D M Ferguson, D C Spellmeyer, T Fox, J W Caldwell, and P A Kollman. *J. Am. Chem. Soc.*, 117:5179, 1995.

[32] P Kollman, R Dixon, W Cornell, T Fox, C Chipot, and A Pohorille. Computer simulation of biomolecular systems, 1997.

[33] G Moyna, H J Williams, R J Nachman, and A I Scott. *Biopolymers*, 49:403, 1999.

[34] N L Allinger, Y H Yuh, and J H Lii. *J. Am. Chem. Soc.*, 111:8551, 1989.

[35] J H Lii, and N L Allinger. *J. Am. Chem. Soc.*, 111:8556, 1989.

[36] J H Lii, and N L Allinger. *J. Am. Chem. Soc.*, 111:8576, 1989.

[37] N L Allinger, H J Geise, W Pyckhout, L A Paquette, and J C Gallucci. *J. Am. Chem. Soc.*, 111:1106, 1989.

[38] N L Allinger, F Li, and L Yan. *J. Comput. Chem.*, 11:848, 1990.

[39] N L Allinger, F Li, L Yan, and J C Tai. *J. Comput. Chem.*, 11:868, 1990.

[40] J H Lii, and N L Allinger. *J. Phys. Org. Chem.*, 7:591, 1994.

[41] J H Lii, and N L Allinger. *J. Comput. Chem.*, 19:1001, 1998.

[42] J H Lii, and N L Allinger. *J. Comput. Chem.*, 12:186, 1991.

[43] M Y Okamura, and G Feher. *Ann. Rev. Biochem.*, 61:861, 1992.

[44] A J Farebrother, A J H M Meijer, D C Clary, and A J Fisher. *Chem. Phys. Lett.*, 319:308, 2000.

[45] A J H M Meijer, A J Farebrother, D C Clary, and A J Fisher. *J. Phys. Chem. A*, 105:2173, 2001.

[46] A J H M Meijer, A J Farebrother, and D C Clary. *J. Phys. Chem. A*, 106:8996, 2002.

[47] J K Gregory, D J Wales, and D C Clary. *J. Chem. Phys.*, 102:1592, 1995.

[48] J M Sorenson, J K Gregory, and D C Clary. *J. Chem. Phys.*, 106:849, 1997.

[49] G M Chaban, R B Gerber, and K C Janda. *J. Phys. Chem. A*, 105:8323, 2001.

[50] G M Chaban, J Lundell, and R B Gerber. *Chem. Phys. Lett.*, 364:628, 2002.

[51] G M Chaban, J Lundell, and R B Gerber. *J. Chem. Phys.*, 115:7341, 2001.

[52] Z Bihary, G M Chaban, and R B Gerber. *J. Chem. Phys.*, 117:5105, 2002.

[53] J Lundell, G M Chaban, and R B Gerber. *J. Phys. Chem. A*, 104:7944, 2000.

[54] J Lundell, M Petterson, L Kriachtchev, M Räsänen, G M Chaban, and R B Gerber. *Chem. Phys. Lett.*, 322:389, 2000.

[55] J Lundell, G M Chaban, and R B Gerber. *Chem. Phys. Lett.*, 331:308, 2000.

[56] G M Chaban, and R B Gerber. *Spectrochimica Acta Part A*, 58:887, 2002.

[57] Z Bihary, G M Chaban, and R B Gerber. *J. Chem. Phys.*, 116:13, 2002.

[58] R B Gerber, G M Chaban, S K Gregurick, and B Brauer. *Biopolymers*, 68:370, 2002.

[59] C J Gruenloh, J R Carney, F C Hagemeister, T S Zwier, J T Wood III, and K D Jordan. *J. Chem. Phys.*, 113:2290, 2000.

[60] G M Florio, C J Gruenloh, R C Quimpo, and T S Zwier. *J. Chem. Phys.*, 113:11143, 2000.

[61] L C Snoek, E G Robertson, R T Kroemer, and J P Simons. *Chem. Phys. Lett.*, 321:49, 2000.

[62] T S Zwier. *Ann. Rev. Phys. Chem.*, 47:205, 1996.

[63] J A Dickinson, P W Joireman, R W Randall, E G Robertson, and J P Simons. *J. Phys. Chem. A*, 101:513, 1997.

[64] R T Kroemer, K R Liedl, J A Dickinson, E G Robertson, J P Simons, D R Borst, and D W Pratt. *J. Am. Chem. Soc.*, 120:12573, 1998.

[65] J R Carney, and T S Zwier. *J. Phys. Chem. A*, 103:9943, 1999.

[66] J R Carney, and T S Zwier. *J. Phys. Chem. A*, 104:8677, 2000.

[67] J R Carney, and T S Zwier. *Chem. Phys. Lett.*, 341:77, 2001.

[68] T S Zwier. *J. Phys. Chem. A*, 105:8827, 2001.

[69] J R Carney, B C Dian, G M Florio, and T S Zwier. *J. Am. Chem. Soc.*, 123:5596, 2001.

[70] J R Carney, A V Federov, J R Cable, and T S Zwier. *J. Phys. Chem. A*, 105:3487, 2001.

[71] A V Federov, J R Cable, J R Carney, and T S Zwier. *J. Phys. Chem. A*, 105:8162, 2001.

[72] L L Connell, T C Corcoran, P W Joireman, and P M Felker. *J. Phys. Chem.*, 94:1229, 1990.

[73] E G Robertson, and J P Simons. *Phys. Chem. Chem. Phys.*, 3:1, 2001.

[74] L C Snoek, R T Kroemer, M R Hockridge, and J P Simons.

[75] L C Snoek, R T Kroemer, and J P Simons. *Phys. Chem. Chem. Phys.*, 4:2130, 2002.

[76] C Plützer, E Nir, M S de Vries, and K Kleinermanns. *Phys. Chem. Chem. Phys.*, 3:5466, 2001.

[77] E Nir, C Janzen, P Imhof, K Kleinermanns, and M S de Vries. *Phys. Chem. Chem. Phys.*, 4:732, 2002.

[78] E Nir, C Janzen, P Imhof, K Kleinermanns, and M S de Vries. *Phys. Chem. Chem. Phys.*, 4:740, 2002.

[79] G M Chaban, J O Jung, and R B Gerber. *J. Chem. Phys.*, 111:1823, 1999.

[80] D W Noid, and R A Marcus. *J. Chem. Phys.*, 67:559, 1977.

[81] M J Davis, and E J Heller. *J. Chem. Phys.*, 71:3383, 1979.

[82] D W Noid, M L Koszykowski, M Tabor, and R A Marcus. *J. Chem. Phys.*, 72:6169, 1980.

[83] J W Neuberger, and D W Noid. *Chem. Phys. Lett.*, 104:1, 1984.

[84] J W Neuberger, and D W Noid. *Chem. Phys. Lett.*, 112:393, 1984.

[85] I P Hamilton, and J C Light. *J. Chem. Phys.*, 84:306, 1986.

[86] J Echave, and D C Clary. *Chem. Phys. Lett.*, 190:225, 1992.

[87] J M Bowman. *J. Chem. Phys.*, 68:608, 1978.

[88] J R Reimers, and R O Watts. *Chem. Phys.*, 85:83, 1984.

[89] D F Coker, R E Miller, and R O Watts. *J. Chem. Phys.*, 82:3554, 1985.

[90] J R Reimers, and R O Watts. *Mol. Phys.*, 52:357, 1984.

[91] D F Coker, and R O Watts. *J. Phys. Chem.*, 91:2513, 1987.

[92] E J Shipsey, and D A Kohl. *J. Chem. Phys.*, 89:5486, 1988.

[93] S C Althorpe, and D C Clary. *J. Chem. Phys.*, 101:3603, 1994.

[94] J M Bowman, A Wierzbicki, and J Zúñiga. *Chem. Phys. Lett.*, 150:269, 1988.

[95] R J Whitehead, and N C Handy. *J. Mol. Spec.*, 55:356, 1975.

[96] F R Burden, and H M Quiney. *Mol. Phys.*, 53:917, 1984.

[97] F R Burden, and A Cuno. *Mol. Phys.*, 62:33, 1987.

[98] K M Christoffel, and J M Bowman. *Chem. Phys. Lett.*, 85:220, 1982.

[99] A R Hoy, I M Mills, and G Strey. *Mol. Phys.*, 24:1265, 1972.

[100] R O Watts. *Chem. Phys.*, 26:367, 1977.

[101] J R Reimers, R O Watts, and M L Klein. *Chem. Phys.*, 64:95, 1982.

[102] J R Reimers, and R O Watts. *Chem. Phys.*, 91:201, 1984.

[103] J A Fernley, S Miller, and J Tennyson. *J. Mol. Spec.*, 150:597, 1991.

[104] C D Paulse, and J Tennyson. *J. Mol. Spec.*, 168:313, 1994.

[105] A E Lynas-Gray, S Miller, and J Tennyson. *J. Mol. Spec.*, 169:458, 1995.

[106] N F Zobov, O L Polyansky, C R Le Seur, and J Tennyson. *Chem. Phys. Lett.*, 260:381, 1996.

[107] O L Polyansky, N F Zobov, S Viti, J Tennyson, P F Bernath, and L Wallace. *Science*, 277:346, 1997.

[108] O L Polyansky, N F Zobov, S Viti, J Tennyson, J A Lotoski, and P F Bernath. *J. Mol. Spec.*, 184:35, 1997.

[109] O L Polyansky, J Tennyson, and P F Bernath. *J. Mol. Spec.*, 186:213, 1997.

[110] O L Polyansky, N F Zobov, S Viti, J Tennyson, P F Bernath, and L Wallace. *J. Mol. Spec.*, 186:422, 1997.

[111] O L Polyansky, N F Zobov, S Viti, and J Tennyson. *J. Mol. Spec.*, 189:291, 1998.

[112] A G Császár, J S Kain, O L Polyansky, N F Zobov, and J Tennyson, *Chem. Phys. Lett.*, 293:317, 1998.

[113] H Y Mussa, and J Tennyson. *J. Chem. Phys.*, 109:10885, 1998.

[114] O L Polyansky, J Tennyson, and N F Zobov. *Spectrochim. Acta*, 55A:659, 1999.

[115] J S Kain, O L Polyansky, and J Tennyson. *Chem. Phys. Lett.*, 317:365, 2000.

[116] H M Quiney, P Barletta, G Tarczay, A G Császár, O L Polysansky, and J Tennyson. *Chem. Phys. Lett.*, 344:413, 2001.

[117] R Schermaul, R C M Learner, A A D Canas, J W Brault, O L Polyansky, D Belmiloud, N F Zobov, and J Tennyson. *J. Mol. Spec.*, 211:169, 2001.

[118] R Schermaul, R C M Learner, D A Newnham, R G Williams, J Ballard, N F Zobov, D Belmiloud, and J Tennyson. *J. Mol. Spec.*, 208:32, 2001.

[119] R Schermaul, R C M Learner, D A Newnham, J Ballard, N F Zobov, D Belmiloud, and J Tennyson. *J. Mol. Spec.*, 208:43, 2001.

[120] R N Tolchenov, J Tennyson, J W Brault, A A D Canas, and R Schermaul. *J. Mol. Spec.*, 215:269, 2002.

[121] M Tanaka, J W Brault, and J Tennyson. *J. Mol. Spec.*, 216:77, 2002.

[122] O L Polyansky, A G Császár, S V Shirin, N F Zobov, P Barletta, J Tennyson, D W Schwenke, and P J Knowles. *Science*, 299:5606, 2003.

[123] S V Shirin, O L Polyansky, N F Zobov, P Barletta, and J Tennyson. *J. Chem. Phys.*, 118:2124, 2003.

[124] R N Tolchenov, J Tennyson, S V Shirin, N F Zobov, O L Polyansky, and A N Maurellis. *J. Mol. Spec.*, 221:99, 2003.

[125] G Brocks, and J Tennyson. *J. Mol. Spec.*, 99:263, 1983.

[126] J Tennyson, and B T Sutcliffe. *J. Mol. Spec.*, 101:71, 1983.

[127] J Tennyson, and A van der Avoird. *Chem. Phys. Lett.*, 105:49, 1984.

[128] L Neale, S Miller, and J Tennyson. *Astrophys. J.*, 464:516, 1996.

[129] L S Norris, M A Ratner, A E Roitberg, and R B Gerber. *J. Chem. Phys.*, 105:11261, 1996.

[130] R B Gerber, B Brauer, S K Gregurick, and G M Chaban. *Phys. Chem. Comm.*, 5:142, 2002.

[131] N Matsunaga, G M Chaban, and R B Gerber. *J. Chem. Phys.*, 117:3541, 2002.

[132] S K Gregurick, G M Chaban, and R B Gerber. *J. Phys. Chem. A*, 106:8696, 2002.

[133] J O Jung, and R B Gerber. *J. Chem. Phys.*, 105:10332, 1996.

[134] G M Chaban, J O Jung, and R B Gerber. *J. Phys. Chem. A*, 104:2772, 2000.

[135] N J Wright, and R B Gerber. *J. Chem. Phys.*, 112:2598, 2000.

[136] N J Wright, R B Gerber, and D J Tozer. *Chem. Phys. Lett.*, 324:206, 2000.

[137] J O Jung, and R B Gerber. *J. Chem. Phys.*, 105:10682, 1996.

[138] S Broude, J O Jung, and R B Gerber. *Chem. Phys. Lett.*, 299:437, 1999.

[139] G M Chaban, J O Jung, and R B Gerber. *J. Phys. Chem. A*, 104:10035, 2000.

[140] G M Chaban, and R B Gerber. *J. Chem. Phys.*, 115:1340, 2001.

[141] N J Wright, and R B Gerber. *J. Chem. Phys.*, 114:8763, 2000.

[142] Z Bihary, R B Gerber, and V A Apkarian. *J. Chem. Phys.*, 115:2695, 2001.

[143] G M Chaban, S S Xantheas, and R B Gerber. *J. Phys. Chem. A*, 107:4952, 2003.

[144] S Carter, I M Mills, and N C Handy. *J. Chem. Phys.*, 99:4379, 1993.

[145] P Palmieri, R Tarroni, M M Huhn, N C Handy, and A Willetts. *Chem. Phys.*, 190:327, 1995.

[146] G Klatt, A Willetts, N C Handy, and C D Esposito. *Chem. Phys. Lett.*, 237:273, 1995.

[147] S Carter, and N C Handy. *J. Mol. Spec.*, 179:65, 1996.

[148] G Klatt, A Willetts, N C Handy, R Tarroni, and P Palmieri. *J. Mol. Spec.*, 176:64, 1996.

[149] G Klatt, A Willetts, and N C Handy. *Chem. Phys. Lett.*, 249:272, 1996.

[150] D K W Mok, N C Handy, and R D Amos. *Mol. Phys.*, 92:667, 1997.

[151] N C Handy, and A Willetts. *Spectrochim. Acta*, 53:1169, 1997.

[152] S Carter, N Pinnavaia, and N C Handy. *Chem. Phys. Lett.*, 240:400, 1995.

[153] M P Jigato, K Somasundram, V Termath, and N C Handy. *Surface Science*, 380:83, 1997.

[154] S Carter, S J Culik, and J M Bowman. *J. Chem. Phys.*, 107:10458, 1997.

[155] S Carter, J M Bowman, and L B Harding. *Spectrochimica Acta*, 53:1179, 1997.

[156] S Carter, J M Bowman, and N C Handy. *Th. Chem. Acc.*, 100:191, 1998.

[157] M-L Senent, P Palmieri, S Carter, and N C Handy. *Chem. Phys. Lett.*, 354:1, 2002.

[158] S Carter, and N C Handy. *Chem. Phys. Lett.*, 352:1, 2002.

[159] R Burcl, S Carter, and N C Handy. *Chem. Phys. Lett.*, 373:357, 2003.

[160] J B Anderson. *J. Chem. Phys.*, 63:1499, 1975.

[161] D F Coker, and R O Watts. *Mol. Phys.*, 58:1113, 1986.

[162] J K Gregory, and D C Clary. *Chem. Phys. Lett.*, 228:547, 1994.

[163] J K Gregory, and D C Clary. *J. Chem. Phys.*, 102:7817, 1995.

[164] J C Gregory, and D C Clary. *J. Phys. Chem.*, 100:18014, 1996.

[165] K Liu, M G Brown, C Carter, R J Saykally, J K Gregory, and D C Clary. *Nature*, 381:501, 1996.

[166] D C Clary, D M Benoit, and T van Mourik. *Acc. Chem. Res.*, 33:441, 2000.

[167] D C Clary, and D M Benoit. *J. Chem. Phys.*, 111:10559, 1999.

[168] J K Gregory, and D C Clary. *J. Chem. Phys.*, 103:8924, 1995.

[169] J K Gregory, and D C Clary. *Chem. Phys. Lett.*, 237:39, 1995.

[170] J K Gregory, and D C Clary. *J. Chem. Phys.*, 105:6626, 1996.

[171] J M Sorenson, J K Gregory, and D C Clary. *Chem. Phys. Lett.*, 263:680, 1996.

[172] J K Gregory, D C Clary, K Liu, M G Brown, and R J Saykally. *Science*, 275:814, 1997.

[173] J K Gregory, and D C Clary. *J. Phys. Chem. A*, 101:6813, 1997.

[174] J C Gregory, and D C Clary. *Mol. Phys.*, 88:33, 1996.

[175] D M Benoit, and Clary. *J. Phys. Chem. A*, 104:5590, 2000.

[176] J K Gregory, and D C Clary. In J.M. Bowman, and Z. Bačić, editors, *Advances in Molecular Vibrations and Collision Dynamics*, volume 3, page 311. JAI Press, London, 1998.

[177] D M Benoit, and D C Clary. *J. Chem. Phys.*, 113:5193, 2000.

[178] D M Benoit, D Sebastiani, and M Parrinello. *Phys. Rev. Lett.*, 87:Art. No. 226401, 2001.

[179] D C Wilson, J C Decius, and P C Cross. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra.* Dover, 1954.

[180] Philip R Bunker. *Molecular Symmetry and Spectroscopy.* Academic Press, London, 1979.

[181] C Eckart. *Phys. Rev.*, 47:552, 1935.

[182] J H Van Vleck. *Phys. Rev.*, 47:487, 1935.

[183] D M Dennison, and M Johnson. *Phys. Rev.*, 47:93, 1935.

[184] Peter W Atkins. *Molecular Quantum Mechanics.* Oxford University Press, Oxford, 1987.

[185] S Califano. *Vibrational States.* John Wiley and Sons, first edition, 1976.

[186] Leonard I Schiff. *Quantum Mechanics.* McGraw-Hill., 1968.

[187] K Kuczera. Normal mode analysis. http://pekoe.chem.ukans.edu/ kuczera/Public/web/html/lect/nma.

[188] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller, and E Teller. *J. Chem. Phys.*, 21:1087, 1958.

[189] M Pincus. *Oper. Res.*, 18:1225, 1970.

[190] S Kirkpatrick, C D Gelatt Jr., and M P Vechi. *Science*, 220:671, 1983.

[191] H Szu, and R Hartley. *Phys. Lett. A*, 122:157, 1987.

[192] J H Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, Michigan, 1975.

[193] J Nocedal. *Math. Comp.*, 24:773, 1980.

[194] D C Liu, and J Nocedal. *Math. Programming*, 45:503, 1989.

[195] J W Ponder, and F M Richards. *J. Comput. Chem.*, 8:1016, 1987.

[196] C E Kundrot, Ponder J W, and F M Richards. *J. Comput. Chem.*, 12:402, 1991.

[197] M J Dudek, and J W Ponder. *J. Comput. Chem.*, 16:781, 1995.

[198] M E Hodsdon, J W Ponder, and D P Cistolam. *J. Mol. Biol.*, 264:585, 1996.

[199] Y Kong, and J W Ponder. *J. Chem. Phys.*, 107:481, 1997.

[200] R V Pappu, R K Hart, and J W Ponder. *J. Phys. Chem. B*, 102:9725, 1998.

[201] P Ren, and J W Ponder. *J. Comput. Chem.*, 23:1497, 2002.

[202] Mathematical optimisation. http://csep1.phy.ornl.gov/mo/mo.html.

[203] W A Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes in FORTRAN.* Cambridge University Press, Cambridge, 1994.

[204] D G Luenberger. *Linear and Non-Linear Programming.* Addison-Wesley, Reading, Massachusetts, 1984.

[205] G H Golub, and C F Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 1983.

[206] M R Hestenes. *Conjugate Direction Methods in Optimisation.* Springer-Verlag, New York, 1980.

[207] M Hénon, and C Heiles. *Astron. J.*, 69:73, 1964.

[208] G Herzberg. *Molecular Spectra and Molecular Structure: Spectra of Diatomic Molecules.* Van Nostrand, New York, second edition, 1950.

[209] M Abramowitz, and I A Stegun. *Handbook of Mathematical Functions.* U.S. National Bureau of Standards, Washington, 1964.

[210] David M Hirst. *A Computational Approach to Chemistry.* Blackwell Scientific Publications, Oxford, 1990.

[211] P-O Löwdin. *Phys. Rev.*, 97:1474, 1955.

[212] R Ahlrichs, H Lischka, V Staemmler, and W Kutzelnigg. *J. Chem. Phys.*, 62:1125, 1975.

[213] R Ahlrichs, H Lischka, V Staemmler, and W Kutzelnigg. *J. Chem. Phys.*, 62:1235, 1975.

[214] R Ahlrichs, H Lischka, V Staemmler, and W Kutzelnigg. *J. Chem. Phys.*, 63:455, 1975.

[215] R Ahlrichs, H Lischka, V Staemmler, and W Kutzelnigg. *J. Chem. Phys.*, 63:4658, 1975.

[216] H J Werner, and W Meyer. *J. Chem. Phys.*, 73:2342, 1980.

[217] H J Wener, and W Meyer. *J. Chem. Phys.*, 74:5794, 1981.

[218] H J Werner, and P J Knowles. *J. Chem. Phys.*, 82:5053, 1985.

[219] P J Knowles, and H J Werner. *Chem. Phys. Lett.*, 115:259, 1985.

[220] K Ruedenberg, L M Chueng, and S T Elbert. *Int. J. Quantum. Chem.*, 16:1069, 1979.

[221] R J Harrison, and N C Handy. *Chem. Phys. Lett.*, 95:386, 1983.

[222] T J Lee, R B Remington, Y Yamaguchi, and H F Schaefer. *J. Chem. Phys.*, 89:408, 1988.

[223] J T Fermann, C D Sherrill, T D Crawford, and H F Schaefer. *J. Chem. Phys.*, 100:8132, 1994.

[224] R P Hosteny, T H Dunning, R R Gilman, A Pipano, and I Shavitt. *J. Chem. Phys.*, 62:4764, 1975.

[225] D F Coker, J R Reimers, and R O Watts. *Aust. J. Phys.*, 35:623, 1982.

[226] A Szabo, and N S Ostlund. *Modern Quantum Mechanics: An Introduction to Electronic Structure Theory.* McGraw-Hill, New York, 1989.

[227] *NAG Library Mark18.*

[228] P J Davis, and P Rabinowitz. *Numerical Integration.* Blaisdell Publishing Group, 1967.

[229] Orient - a program for studying interactions between molecules. http://www-wales.ch.cam.ac.uk/software.html.

[230] R J Bartlett, S J Cole, G D Purvis, W C Ermler, H C Hsieh, and I Shavitt. *J. Chem. Phys.*, 87:6579, 1987.

[231] G Simons, R G Parr, and J M Finlan. *J. Chem. Phys.*, 59:3229, 1973.

[232] W Yang, and A C Peet. *Chem. Phys. Lett.*, 153:98, 1988.

[233] P M Morse. *Phys. Rev.*, 34:57, 1929.

[234] M Nest, and H-D Meyer. *J. Chem. Phys.*, 117:10499, 2002.

[235] J C Light, I P Hamilton, and J V Lill. *J. Chem. Phys.*, 82:1400, 1985.

[236] K Watanabe, and M L Klein. *Chem. Phys.*, 131:157, 1989.

[237] Y Maréchal. *Faraday Disc.*, 103:349, 1996.

[238] S Carter, and N C Handy. *Mol. Phys.*, 47:1445, 1982.

[239] S Carter, N C Handy, and B T Sutcliffe. *Mol. Phys.*, 49:745, 1983.

[240] B T Sutcliffe. *Mol. Phys.*, 48:561, 1983.

[241] S Carter, and N C Handy. *J. Chem. Phys.*, 87:4294, 1987.

[242] P Botschwina. *Chem. Phys.*, 40:33, 1979.

[243] J R Reimers, and R O Watts. *Chem. Phys. Lett.*, 94:222, 1983.

[244] L Halonen, and T Carrington Jr. *J. Chem. Phys.*, 88:1988, 1988.

[245] Z Bačić, D Watt, and J C Light. *J. Chem. Phys.*, 89:957, 1988.

[246] E Kauppi, and L Halonen. *J. Phys. Chem.*, 94:5779, 1990.

[247] J K G Watson. *Mol. Phys.*, 15:479, 1968.

[248] K Kuchitsu, and L S Bartell. *J. Chem. Phys.*, 36:2460, 1962.

[249] K Kuchitsu, and Y Morino. *J. Chem. Soc. (Japan)*, 38:814, 1965.

[250] K Kutchitsu, and Y Morino. *Spectrochim. Acta*, 22:33, 1966.

[251] M A Pariseau, I Suzuki, and J Overend. *J. Chem. Phys.*, 42:2335, 1965.

[252] K Machida. *J. Chem. Phys.*, 44:4186, 1966.

[253] D F Smith, and J Overend. *Spectrochim. Acta*, 28A:471, 1972.

[254] E R Cohen, and B N Taylor. *Physics Today*, August:BG11, 1987.

[255] G Mills, G K Schenter, D E Makarov, and H Jónsson. *Chem. Phys. Lett.*, 278:91, 1997.

[256] H W Schranz, S Nordholm, and B C Freasier. *Chem. Phys.*, 108:69, 1986.

[257] Y D Park, T R Rizzo, L A Peteanu, and D H Levy. *J. Chem. Phys.*, 84:6539, 1986.

[258] L A Philips, and D H Levy. *J. Chem. Phys.*, 89:85, 1988.

[259] L A Peteanu, and D H Levy. *J. Phys. Chem.*, 92:6554, 1988.

[260] Y R Wu, and D H Levy. *J. Chem. Phys.*, 91:5278, 1989.

[261] L L Connell, T C Corcoran, P W Joireman, and P M Felker. *Chem. Phys. Lett.*, 166:510, 1990.

[262] T F Miller, and D C Clary. *J. Chem. Phys.*, 116:8262, 2002.

[263] T F Miller, and D C Clary. *J. Chem. Phys.*, 119:68, 2003.

[264] W L Jorgensen, J Chandrasekhar, J D Madura, R W Impey, and M L Klein. *J. Chem. Phys.*, 79:926, 1983.

[265] J Chandrasekhar, D Spellmeyer, and W L Jorgensen. *J. Am. Chem. Soc.*, 106:903, 1984.

[266] J Pranata, S G Wierschke, and W L Jorgensen. *J. Am. Chem. Soc.*, 113:2810, 1991.

[267] G Kaminski, E M Duffy, T Matsui, and W L Jorgensen. *J. Phys. Chem.*, 98:13077, 1994.

[268] W L Jorgensen, D S Maxwell, and J Tirado-Rives. *J. Am. Chem. Soc.*, 118:11225, 1996.

[269] W Damm, A Frontera, J Tirado-Rives, and W L Jorgensen. *J. Comput. Chem.*, 18:1955, 1997.

[270] W L Jorgensen, and N A McDonald. *J. Mol. Structure Theochem*, 424:145, 1998.

[271] N A McDonald, and W L Jorgensen. *J. Phys. Chem. B*, 102:8049, 1998.

[272] N A McDonald, E M Duffy, and W L Jorgensen. *J. Am. Chem. Soc.*, 120:5104, 1998.

[273] R C Rizzo, and W L Jorgensen. *J. Am. Chem. Soc.*, 121:4827, 1999.

[274] M W Mahoney, and W L Jorgensen. *J. Chem. Phys.*, 112:8910, 2000.

[275] E S Vermuelen, A W Schmidt, J S Sprouse, H V Wikstrom, and C J Grol. *J. Med. Chem.*, 46:5365, 2003.

[276] J Becker, F Becher, O Hucke, A Labahn, and T Koslowski. *J. Phys. Chem. B*, 107:12878, 2003.

[277] M J Clement, A Imberty, A Phalipon, S Perez, C Siminel, L A Mulard, and M Delepierre. *J. Biolog. Chem.*, 278:47928, 2003.

[278] J L Klepeis, H D Schafroth, K M Westerberg, and C A Floudas. *Adv. Chem. Phys.*, 120:265, 2002.

[279] S Fort, P M Coutinho, M Schulein, R Nardin, S Cottaz, and H Driguez. *Tetrahedron Lett.*, 42:3443, 2001.

[280] Protein data bank. http://www.pdb.org.