

**Bangor University**

## **DOCTOR OF PHILOSOPHY**

**Towards a synergy between spectra and theory through case studies:  
vibrational analyses in biocompatible silanes, bioactive peptides and explicit solvent  
algorithms**

Rozhok, Mykola

*Award date:*  
2010

*Awarding institution:*  
Bangor University

[Link to publication](#)

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Towards a synergy between spectra and theory through case studies: vibrational analyses in biocompatible silanes, bioactive peptides and explicit solvent algorithms**

A thesis presented in partial fulfilment

for the

PhD degree

at the

School of Chemistry

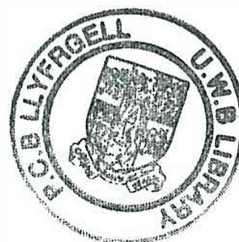
by

Mykola Rozhok



Bangor University • Prifysgol Bangor

© May 2010





## TABLE OF CONTENTS

---

<b>ACKNOWLEDGEMENTS</b> .....	<b>5</b>
<b>HYPOTHESES</b> .....	<b>6</b>
<b>ABSTRACT</b> .....	<b>7</b>
<b>1. PREAMBLE</b> .....	<b>9</b>
<b>2. INTRODUCTION</b> .....	<b>11</b>
2.1 A brief overview <i>Ab initio</i> Molecular Modelling.....	12
2.1.1 History of molecular modelling techniques.....	15
2.1.2 Theoretical foundations and development of <i>ab initio</i> methods.....	20
2.1.3 An illustrated summary of relevant terminology used in the text.....	24
2.2 Computational methods and Artificial Intelligence .....	27
2.2.1 An introduction to computational methods .....	32
2.2.2 Algorithms used in this work .....	43
2.2.3 Programming languages, frameworks and tools.....	51
2.3 Summary .....	53
<b>3. METHODS</b> .....	<b>54</b>
3.1 Experimental .....	55
3.2 G03 and CPMD calculations, AIM analysis .....	56
3.3 Algorithms.....	61
<b>4. RESULTS AND DISCUSSION (Part I)</b> .....	<b>63</b>
4.1 Computational tools for input file generation, network-wide calculation execution and management, and data extraction applications.....	64
4.1.1 Coordinate systems and input structure files .....	65
4.1.2 Queuing systems.....	70
4.1.3 Output tabulation and analysis .....	73
4.1.4 Calculation Management.....	80
4.2 Explicit particle solvation algorithm .....	85

<b>5. RESULTS AND DISCUSSION (Part II)</b> .....	<b>91</b>
5.1 Rational Design of Comb-Like Monolayers of Biocompatible Silanes I. Dynamic Binding Mechanisms and Chain Pre-Organisation Prior to Surface Grafting.....	92
5.1.1 A Background/Introduction.....	92
5.1.2 Methods.....	97
5.1.3 Results and Discussion.....	99
5.1.4 Conclusions .....	105
5.2 Towards the Protonation Limit of Amino Acids and Peptides I: An Algorithm to track the Dynamics of Conformation- and Orbital-Specific Poly-Protonation	107
5.2.1 A Background/Introduction.....	107
5.2.2 Methods.....	110
5.2.3 Results and Discussion.....	111
5.2.4 Conclusions .....	121
<b>6. CONCLUSIONS</b> .....	<b>123</b>
<b>7. Appendix A. Rational design of novel dental adhesives: the role of computational steering in optimising dental silanes</b> .....	<b>127</b>
<b>8. Appendix B. Amino acid protonation: glycine</b> .....	<b>170</b>
<b>9. REFERENCES</b> .....	<b>202</b>
<b>10. INDEX</b> .....	<b>225</b>

## ACKNOWLEDGEMENTS

---

The research in this thesis would have taken far longer to complete without the encouragement from many others. I would especially like to thank my supervisor, Dr. Gregory A. Chass, for his guidance and relaxed, thoughtful insight, and for helping me complete the writing of this dissertation as well as the challenging research that lies behind it.

I am grateful to Professor Sivolob A. V. for getting me interested in molecular sciences. His guidance and support became the foundation of all my future scientific endeavours.

I would also like to thank The Drapers' Company for their funding support.

Finally, I wish to thank my family for their love, support and encouragement, without whom I would never have enjoyed so many opportunities.

## HYPOTHESES

---

Many day-to-day tasks in computational physical chemistry projects, as well as ill-defined problem statements are much more time-consuming than most calculations during the project life-cycle. Development of a standardised approach to input file generation, calculation execution, results collection and presentation is of utmost importance, before one can move on to applying more advanced computer science artificial intelligence based methods to physical and chemical problems.

Grid-based computational chemistry approaches can be combined with molecular dynamics simulations to obtain more meaningful results. Grid-based calculations provide static lowest potential energy points in multidimensional space, while *ab initio* molecular dynamics can be used to connect the points and give an accurate description of molecular pathways. The combination of these two approaches can be used to explain experimental results. Furthermore, it can direct laboratory experiments making them much more space-, resource- and time-efficient.



## ABSTRACT

---

Theoretical calculations, including *ab initio* Car-Parrinello Molecular Dynamics (CPMD), geometry optimisation on the principal grid-based conformers, Infra-red vibrational frequencies, and charge density analysis (Atoms In Molecules, AIM), were performed on three important chemical systems (glycine, 3-isocyanatopropyltriethoxysilane (ICS), 3-styrylethyltrimethoxysilane (STYRX)) under a variety of different pH conditions and solvents (H<sub>2</sub>O, CCl<sub>4</sub>, ethanol and mixed solvents, both using implicit Polarisable Continuum solvation models and explicitly solvated). CPMD studies were carried out at 300.15 K, with 10,000-step trajectories covering the timescale of  $\approx 1$  picosecond. Molecular dynamics, AIM and grid-based approaches were compared and combined to give accurate descriptions of molecular behaviour and preferred geometrical conformations of the aforementioned systems. The aim in characterising these systems is to build a better understanding of protein structure and functioning, as well as to analyse two examples of what we believe are very promising silicon-based coupling agents.

Theoretical calculations were validated by experimental results. A computer application was developed to help in conformer-specific analysis and presentation of Infra-red Spectroscopy data.

A molecular geometry and van der Waals radii-based explicit solvation method was developed, and another more advanced charge distribution based solvation method proposed. Both methods near-optimally solvate molecules using a minimal but sufficient amount of any solvent, under various conditions.

A set of computer applications was developed to aid in creating, executing and doing preliminary analyses of theoretical physical chemistry calculations.

As a result, a complete characterisation of the glycine amino acid behaviour and preferred conformations for gas phase non-protonated and singly-protonated states was accomplished. The successful synergy between experiment and theory for the STYRX and ICS silane coupling agents allowed for a theoretical conformational analysis of these systems in various solvent conditions and quantitative assignment of molecular vibrations to the major peaks present in the Fourier Transform Infrared Spectroscopy (FTIR) results.

## 1. PREAMBLE

---

Quantitatively characterising and predicting the behaviour of molecular and materials systems has remained a central challenge to the sciences over the past few centuries. An enormous number of projects in mathematics, physics, chemistry, biology, medicine and most recently computer science have relied upon a wide range of techniques and approaches from qualitative observation to quantitative analyses and pure mathematical modelling, towards more fully understanding the driving forces behind the assembly of matter's building blocks.

The advent of high-speed computing, combined with recent advances in technology and computer science have made it possible to employ (on a large scale) a powerful tool in the area of modern scientific modelling – *ab initio* molecular dynamics. The method allows chemical processes in condensed phases to be studied in a quantitative and unbiased manner, using electronic structure calculations to account for inter-atomic forces. Despite the variety of different computational techniques built-upon *ab initio* methods, they are all primarily based on the Schrödinger equation which describes how the quantum state of a physical system changes over time. Since solutions to this equation exist only for very specific cases (*e.g.* numerical integration for certain energy values) and it cannot be solved analytically for many-particle systems, approximations must be used. A way to approximate molecular behaviour is to employ various statistical methods on experimental data:

these vary from simple mean-value calculations for “the average behaviour” to computer simulations of the human brain – a very powerful pattern recognition tool.

The transition from “pen & paper” based theoretical studies to super-fast computer calculations has had a great influence on the progress of many scientific research areas, as it allowed new methods to be tested quickly, in-turn increasing the efficiency of experimental determinations through provision of preliminary “exploratory” theoretical predictions. For example, most chemical systems can be initially characterised by one of the available modelling methods prior to commencement of laboratory work, considerably reducing the number of combinatorial and trial-and-error experiments required; not to mention the financial and environmental cost of using extra chemicals nor time wasted. Although most of this work is devoted to results relevant to the physical and molecular sciences, special attention is given to the automation of several time consuming, manual tasks that theoreticians face in each modelling project. These include defining and constructing input structures for molecular systems, in a format that computers can understand, optimising time-scheduling and managing multiple calculations at the same time, extracting and presenting data from calculated results, among many others.

Herein is presented a means to improving the overall efficiency to these ventures, while introducing novel methods in the quest to more fully characterise and understand molecular and materials systems at the atomic level.



## 2. INTRODUCTION

---

One of the most important developments in the area of modern electronic structure theory methods is *ab initio* molecular dynamics (AD) – a scientific tool that comprises finite temperature molecular dynamics with inter-atomic forces obtained from electronic structure calculations.<sup>1</sup> The method greatly extends the list of application areas of the traditional molecular dynamics based on classical mechanics (MD) and electronic structure methods, as it allows the dynamics of chemical processes in condensed phase (*i.e.* molecular system behaviour) to be studied in a time dependent manner. This leads to new paradigms in the clarification of nanoscopic mechanisms, systematisation of experimental data, and prediction of new phenomena.<sup>2</sup>

This work aims to present a set of exemplary case studies highlighting the place of electronic structure methods among the multitude of other scientific approaches in the physical and molecular sciences. It outlines the importance of some recently developed mathematical and computer science methods applied therein. It also introduces aspects where the essential stages of AD are analysed and their performance improved by Artificial Intelligence (AI) techniques.

## 2.1 A brief overview *Ab initio* Molecular Modelling

As this work has a broad scope, it is impossible to provide a complete list of references for all topics covered; in particular no attempt has been made to provide a complete historical attribution of ideas. Instead, the aim has been to provide reference to works that cover the topic in greater detail than is possible herein, with focus to providing introductory points into what, in some cases, is a very extensive literature. Reference is therefore more often made to more recent textbooks and review articles rather than to original historical sources.

A series of review papers has appeared since the early 1990s, primarily dealing with the specificities of *ab initio* molecular dynamics.<sup>2-11</sup> They presented many different viewpoints on the method, as well as its potential uses: calculations of NMR in proteins,<sup>12</sup> structure of nucleic acids,<sup>13</sup> inter-molecular arrangement of water molecules<sup>14, 15</sup> among numerous others. The list of areas where AD could be applied quickly spread through materials science, chemistry, physics, biochemistry, pharmacy and a whole range of interdisciplinary sciences.

The principle reason for AD being applied to solving so many problems in different areas is the method's accuracy in describing atomic interactions and their influence on the resultant dynamics observed. Despite being an acceptable approximation of bulk molecular behaviour, atomistic or Molecular Mechanics methods (MM) which were extensively used in the past and are still often used nowadays for large molecular systems, utilise the classical equations of motion and neglect explicit treatment of electrons. Contrastingly, the foundation of AD methodology is based in

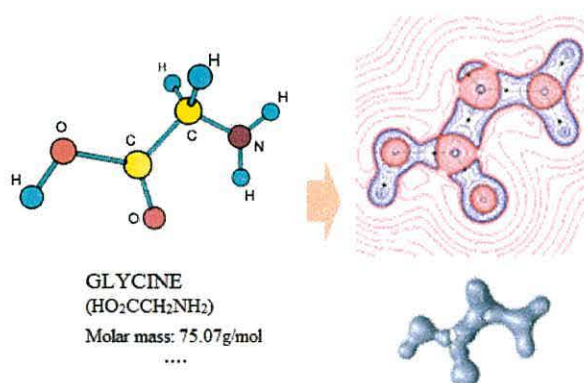
the laws of quantum mechanics and therefore is currently the best mathematical approximation of the actual behaviour of molecular systems. However, this accuracy comes at a relatively high cost, as the treatment of electron distribution requires relatively large computational resources – the number of elementary floating point operations (FLOPs) requisite for completing a computation, increases exponentially with the number of atoms in the system. Another set of methods which tries to handle the “speed-accuracy trade off” is *semi-empirical* molecular dynamics;<sup>16, 17</sup> as the name suggests, these methods exploit some experimental results to directly approximate selected potentials of molecular behaviour (popular examples are methods AM1 and PM3). Although semi-empirical methods generally tend to be inaccurate for problems involving bond formation or chemical transitions,<sup>16, 17</sup> they are extensively used for preliminary theoretical studies of systems which are beyond the scope of *ab initio* calculations. Table 2.1 summarises the general characteristics of the three main groups of methods used in computational chemistry.

The challenges involved in making AD calculations more efficient have been discussed for a long time.<sup>18-20</sup> Although many powerful mathematical and computational methods have been developed and used in the past two decades to approximate and speed up explorations of potential energy hyper surfaces (PEHSs), most dealt with making a better approximation of an idealised force field or specific parts of the PEHS for distinct chemical systems.<sup>21-28</sup> One of the next steps in enhancing these approximations is described in the following chapters – it includes a search for a generalisation in descriptions of atomic interactions, a separation from molecule based to charge distribution based chemistry (Fig. 2.1).



**Table 2.1** Summary of general characteristics of molecular mechanics, semi-empirical, and *ab initio* molecular modelling methods.

Method	Principle	Use cases	Disadvantages	Advantages
<i>Molecular Mechanics</i>	<ul style="list-style-type: none"> <li>• <b>Classical mechanics.</b></li> <li>• Empirical parameters defined in a force-field.</li> </ul>	<ul style="list-style-type: none"> <li>• Large systems (100's of thousands of atoms), no bond breaking / formation.</li> <li>• Basic systems comparison.</li> </ul>	<ul style="list-style-type: none"> <li>• Requires parameters definition.</li> <li>• Cannot be used for electronic structure calculations.</li> <li>• No bond formation or breaking events.</li> </ul>	<ul style="list-style-type: none"> <li>• Fast.</li> <li>• Can give a general overview of a large system's behaviour.</li> </ul>
<i>Semi-empirical</i>	<ul style="list-style-type: none"> <li>• <b>Quantum physics.</b></li> <li>• Experimental parameters and extensive approximations.</li> </ul>	<ul style="list-style-type: none"> <li>• Medium-sized systems (thousands of atoms).</li> <li>• Example use: conformational changes in small protein.</li> </ul>	<ul style="list-style-type: none"> <li>• Parameters are initialized based on <i>ab initio</i> or experimental data.</li> <li>• Less rigorous than <i>ab initio</i> methods.</li> </ul>	<ul style="list-style-type: none"> <li>• Faster than <i>ab initio</i> methods and can calculate transition states semi-quantitatively.</li> </ul>
<i>Ab initio</i>	<ul style="list-style-type: none"> <li>• <b>Quantum mechanics.</b></li> <li>• No empirical parameters.</li> </ul>	<ul style="list-style-type: none"> <li>• Small systems (hundreds to one thousand atoms).</li> <li>• Electronic transition calculations.</li> </ul>	<ul style="list-style-type: none"> <li>• Slow.</li> </ul>	<ul style="list-style-type: none"> <li>• Mathematically rigorous.</li> <li>• Can be used for a broad range of systems, very accurate transition and excited states calculations.</li> </ul>



**Figure 2.1** Molecule based (left) vs. charge distribution based (right) computational physical chemistry. The blue / red colouring scheme in the Isodensity map on the right is used to denote regions of high / low electronic density.

### 2.1.1 History of molecular modelling techniques

**Molecular modelling** is the general term used to describe the use of computers to construct models of molecular or materials systems, and perform a variety of calculations on these towards characterising and predicting their chemical / physical properties and behaviours. Although the term is often used synonymously with the term *computational chemistry*, Dorsett and White<sup>29</sup> define molecular modelling as “*a subset of computational chemistry which concentrates on predicting the behaviour of individual molecules within a chemical system*”. Despite many definitions, molecular modelling is a form of computational re-creation of matter wherein atoms are allowed to interact either time dependently or independently under known laws of physics.

The earliest forms of molecular models date back to the mid-19<sup>th</sup> century, when Archibald Scott Couper, Friedrich August Kekulé von Stradonitz, and Aleksandr Mikhailovich Butlerov independently introduced the general rules of valence for organic chemistry, and the first written structures involving chains of carbons with lines drawn as “bonds” (1858-1861). In 1861, Johann Josef Loschmidt developed a collection of 368 molecular structures (benzene and 120 other aromatic compounds, cyclopropane *etc.*) The first recorded use of a physical molecular model in organic chemistry was in 1865 during a lecture by August Wilhelm von Hofmann, where he used the metaphor of croquet balls joined by sticks to describe methane and chloroform to the Royal Society of Great Britain.<sup>30</sup>

The concept of a **force field** originated at the beginning of the 20<sup>th</sup> century from vibrational spectroscopy, which considered the forces acting between all pairs of atoms in a molecule, or in a lattice of ionic crystals. The first force fields were based on Hooke's law and the Morse potential, yet were not widely used until 1946 when Molecular Mechanics was first proposed. In 1950 Barton's publication on how the geometric conformations of steroids affect their observed chemical properties<sup>31</sup> laid the foundation of *conformational analysis*. In 1953 Watson and Crick disproved the three-chain DNA models and presented an alternate two-chain model, development of which was based mostly on the results of building plastic 'ball and stick' models, with which to visualise geometric structures emerging from mathematical analysis of physical and chemical experimental results and later published.<sup>32</sup> In the same year, the groundwork for computer-based Monte Carlo and simulated annealing methods was laid by a group of scientists from Los Alamos.<sup>33</sup>

The first published use of a computer for empirical force field calculations of molecular structure was in 1961 by Hendrickson,<sup>34</sup> with which to examine the conformational behaviour of medium-sized rings. This was followed by Csizmadia and Slater's works in 1963 at MIT, where they constructed the first ever models of protein components (formamide to model the peptide bond), successfully completing single-point energy calculations. After this, the widely used **steepest descent**<sup>35</sup> method was developed by Wiberg in 1965, whom also published the first algorithms for transforming Cartesian to internal coordinates. Over the second half of the 20<sup>th</sup> century, force-field calculations of molecular structure developed in pace with the development of computing machines. At the beginning of 1970's the major force-fields were published: ECEPP, UNICEPP, CFF, MMI, EAS, Boyd's force field,



MUB and others. In 1971, based on Van der Waals radii calculations, Lee and Richards described the **molecular surface** of a protein structure.<sup>36</sup> In 1972, Wiberg and Boyd developed the broadly used **dihedral driver** method,<sup>37</sup> which explores the conformational space of a molecule. With the development of graphical user interfaces (GUIs) and the World Wide Web by the early 1990's, data visualisation and communication between scientists became much easier. Apart from new types of force fields (Class-2 and Class-3) which contain anharmonic potentials and utilise the off-diagonal terms of the force constant matrix, virtual reality<sup>38,39</sup> and the use of structural data obtained from high-end *ab initio* calculations to parameterise new force fields became some of the most influential trends in molecular modelling.

By the early 1990's chemists started relying heavily upon mathematical descriptions of the fundamental rules of the physical properties of matter which are contained in quantum mechanics and molecular mechanics to study 3D molecular transformations and chemical properties of different systems.<sup>40-42</sup> These molecular properties can be derived from the Schrödinger equation and various approximations thereof.

It is also very important to mention the range of works dealing with searching conformational space, as this work reports results obtained from novel methods of intelligent exploration of potential energy hyper surfaces. References include one of the first ever works in conformational analysis by Eliel *et al.*,<sup>43</sup> subsequent papers,<sup>44-46</sup> as well as a few publications on genetic algorithms in molecular modelling.<sup>47, 48</sup> The end of the 20<sup>th</sup> and start of the 21<sup>st</sup> centuries brought the fundamental question of molecular modelling into focus again: how to more accurately and quantitatively approximate inter-atomic interactions. The traditional route followed in molecular dynamics<sup>49-56</sup> was to predetermine the two-body, three-body and many-body

contributions, long- and short-range interaction terms, among others. Despite overwhelming success in the past, these “static potential” models contained serious drawbacks.<sup>57, 58</sup> Traditional molecular dynamics and electronic structure methods were greatly extended by the family of techniques called *ab initio* molecular dynamics.<sup>18, 59</sup>

More advanced computational techniques have also been used to control and direct electronic structure calculations. Artificial neural networks, an example of artificial intelligence applied to molecular modelling, have been utilised in a whole range of approaches. In 1988, Qian and Sejnowski presented a paper in the *Journal of Molecular Biology* with a new method for predicting the secondary structure of globular proteins based on the non-linearity of neural network models.<sup>60</sup> The performance of the method was heavily dependent on the structure of the proteins that were presented. Although the approach was not very accurate, it illuminated some of the key problems that have to be solved in order to get the neural network based protein structure predicting ‘machine’ working. Around the same time, Bohr *et al.* published a paper on the  $\alpha$ -helices contained in the secondary structure of rhodopsin, with further homology predictions made using neural networks.<sup>61</sup> In 1989, independently of Qian and Sejnowski, Holley and Karplus completed a paper with a similar title and the same objective, with very similar results, but using a differing approach.<sup>62</sup> The main drawback of both methods was that they did not take into consideration any specific internal properties of amino acids, which are greatly influential to the secondary structure of proteins. Additionally, the test sets used by Holley and Karplus included only 14 proteins, therefore they were unable to guarantee good performance of the neural network for other amino acid sequences,



even though the results showed 63% predictive accuracy. The two scientific groups also used a network with only one hidden layer, which is believed to be insufficient to predict structure in such versatile and complex systems as proteins.

A wave of similar projects followed, each one attempting to design a neural network able to predict the secondary structure of proteins.<sup>63-75</sup> Although most of these projects achieved more than 50% accuracy in results, they were generally based on amino acid sequences and did not take into account interactions between non-neighbouring peptide residues. In 1984, Kabsch and Sander showed that identical pentapeptides can have completely different conformations.<sup>76</sup> The publication was followed by Wilson *et al.* providing examples of similar short peptide sequences having different conformations and activity.<sup>77</sup> Cohen *et al.* outlined some possible reasons why identical protein sequences may have totally different conformations.<sup>78</sup> A similar study was done by Sudarsanam in 1998, who showed that identical octapeptides can have dissimilar geometries.<sup>79</sup>

A different approach was introduced by Unger *et al.* – neural network calculations and protein conformation predictions based on  $C_{\alpha}$  Cartesian atomic coordinates.<sup>80</sup> However, this approach suffered in accuracy due to the exclusion of long-range interactions between atoms in larger molecules.

As a short summary, it is noted that the artificial intelligence approaches to molecular modelling described above all share the same problem: they conduct their calculations and predictions based on relative atomic positions and do not make use of the molecular electronic structure, which is the most important factor in molecular behaviour and properties.

### 2.1.2 Theoretical foundations and development of *ab initio* methods

Since a large part of this work is based on *ab initio* calculations, it is only appropriate to give a brief overview of the theoretical foundations of the method. Additional information can be found in other more specialised works.<sup>2-11, 18, 59, 81-83</sup>

The general idea of an *ab initio* molecular dynamics calculation is to solve the time-dependent Schrödinger equation, expressed as follows:

$$i\hbar \frac{\partial \Phi}{\partial t} = \mathcal{H}\Phi$$

wherein  $\Phi$  is the total wave function  $\Phi(\{\mathbf{R}_I\}, \{\mathbf{r}_i\}; t)$  which depends on nuclear  $\{\mathbf{R}_I\}$  and electronic  $\{\mathbf{r}_i\}$  degrees of freedom, as well as time  $t$ ;  $\mathcal{H}$  is the standard Hamiltonian, which is a sum of operators corresponding to the kinetic and potential energies of a system, described as follows:

$$\mathcal{H} = E_K + E_P = -\frac{(-i\hbar\nabla)^2}{2M_I} + V(\{\mathbf{R}_I\}, \{\mathbf{r}_i\}; t)$$

Various derivations of the above equations have been introduced to resolve the dimensionality bottleneck, resulting from the number of degrees of freedom and their couplings:

- Time-dependent self-consistent field (TDSCF) theory introduced as early as 1930 by Dirac,<sup>6, 84</sup> in which both electrons and nuclei are free to move as

dictated by the principles of quantum mechanics, in self-consistently obtained time-dependent average fields (*i.e.* quantum mechanical expectation values for the other class of degrees of freedom – this is the first example of nuclear and electronic wave function calculation separation, which overcomes aforementioned bottleneck at the cost of computational accuracy);

- Ehrenfest addressed the question of how Newtonian classical dynamics can be derived from Schrödinger’s wave equation,<sup>85</sup> and his namesake hybrid approach “Ehrenfest molecular dynamics” is based on a relative nuclear location, subsequently treated as classical particles and electrons – as quantum objects (it is also the oldest approach to “on-the-fly” molecular dynamics);<sup>18, 86-91</sup>
- Born-Oppenheimer proposed another approach and its extensions,<sup>92-98</sup> which also combines classical molecular dynamics with *ab initio* principles:

$$M_I \ddot{R}_I(t) = -\nabla_I \min_{\Psi_0} \{\langle \Psi_0, \mathcal{H}_e, \Psi_0 \rangle\}$$

$$E_0 \Psi_0 = \mathcal{H}_e \Psi_0$$

however, as apparent in the equations above, in a slightly different way – by solving a time-*independent* Schrödinger equation for static electronic structures (*i.e.* a step-by-step quantum problem for fixed nuclear positions);

- One of the most important developments made in molecular modelling was the development of Hohenberg-Kohn-Sham density functional theory, which greatly improved the time / efficiency ratio of electronic structure calculations;<sup>99-103</sup>



- Another crucial development in *ab initio* calculations was the introduction of Car-Parrinello molecular dynamics.<sup>1</sup> It provides a compromise between the Born-Oppenheimer's time scales acceptable for solving the electronic dynamics equations and the Ehrenfest's smooth time-evolution of the dynamically evolving electronic subsystem. For more information on the mathematics of the Car-Parrinello method, the reader is recommended to consult Refs. <sup>1</sup> and <sup>59</sup> (and the updated version – Ref. <sup>18</sup>).

Another important part of any *ab initio* method is the electronic structure (forces  $\langle \Psi_0, \mathcal{H}_e, \Psi_0 \rangle$ ) calculation. Over the years, a variety of different approaches were combined with molecular dynamics methods. Since *ab initio* calculations are not tied to any particular approach, each of them dealt mostly with the aforementioned time / accuracy trade-off:

- Density Functional Theory (DFT)<sup>1, 104-108</sup> simplifies calculations by minimising the Kohn-Sham energy<sup>102, 103</sup> with respect to orthonormal single-particle functions (the Kohn-Sham orbitals), instead of all possible many-body wave functions. For more information on the applications of DFT, see Refs. <sup>58, 95, 104, 105, 107-119</sup>;
- Hartree-Fock Theory,<sup>20, 94, 120-125</sup> Generalised Valence Bond (GVB),<sup>126-130</sup> Complete Active Space Self-Consistent Field (CASSCF),<sup>130, 131</sup> Full Configuration Interaction (FCI),<sup>132</sup> Møller-Plesset Perturbation Theory (MP2),<sup>133</sup> Coupled Cluster,<sup>134-136</sup> semi-empirical<sup>92, 93, 137-140</sup> and other approximation methods<sup>141-146</sup> were also combined with molecular dynamics

to give precise electronic structure models, but at a much higher computational cost.

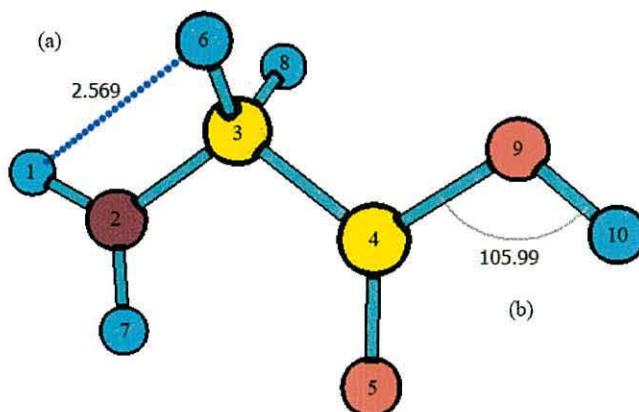
Another important aspect of an *ab initio* calculation is the use of a **basis set** – a set of analytic functions  $f_i$  that represent orbitals  $\psi_i$ :

$$\psi_i(r) = \sum_v c_{iv} f_v(r; \{R_l\})$$

This shows that a linear combination of basis functions is used to represent each orbital. The most widely used basis sets are Slater-type basis functions (STOs) and Gaussian-type basis functions (GTOs).<sup>40</sup>

The first sets of *ab initio* molecular dynamics methods using GTOs were proposed in the mid-1980s.<sup>20, 120-132</sup> Another generation of such approaches that exploit the efficiency of the Car-Parrinello method has been developed in the framework of Car-Parrinello,<sup>133, 147-149</sup> Born-Oppenheimer<sup>97, 98</sup> and Ehrenfest<sup>150</sup> dynamics schemes. For information on other approaches to electron density distribution modelling, consult Ref. <sup>18</sup>.

### 2.1.3 An illustrated summary of relevant terminology used in the text



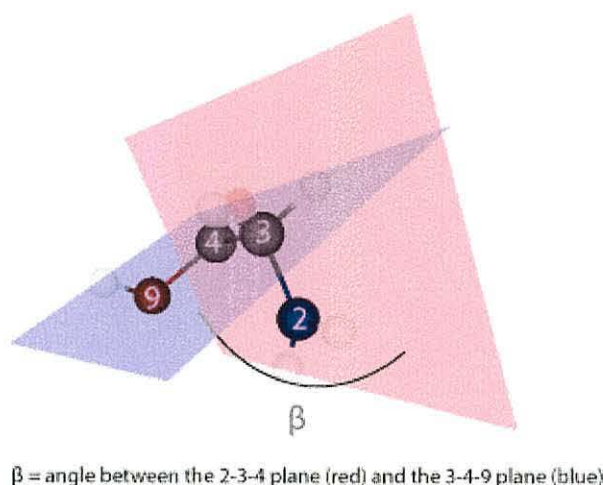
**Figure 2.2** Illustration of bond distance (a), bond angle (b) and a standardised numbering system.

**Bond distance** (Fig. 2.2, a) is the distance between two atoms (not necessarily covalently bonded).

**Bond angle** (Fig. 2.2, b) is the angle between two pairs of atoms. In Figure 2.2, it is the angle between the “red-yellow” and “red-cyan” atom pairs (pairs 9-4 and 9-10).

**Dihedral angle** (Fig. 2.3) is the angle between the planes formed by two groups of three atoms.

**Degree of freedom.** Within this work a molecular degree of freedom refers to the  $3N-6$  vibrations comprised of bond distances, bond angles and dihedral angles.



**Figure 2.3** Illustration of a dihedral angle, specifically that formed by the four atoms 9-4-3-2.

**Internal Coordinates and Z-matrices** (Fig. 2.4) is a convenient way to describe the relative spatial orientation of all atomic constituents of a molecular system. The approach uses bond distances, bond angles and dihedral angles as previously described. The description uses 6 values less than a Cartesian coordinate system – exactly the  $3N-6$  degrees of vibrational freedom (or  $3N-5$  in the few cases of linear molecular systems). The description does not specify any coordinates for the first atom – it is assumed to be at  $(0, 0, 0)$ , uses only one number for the second atom (the distance from atom-2 to atom-1), and two numbers for atom-3 (the distance between atom-3 and atom-2, and the angle between atoms 3-2-1); and a bond, angle, dihedral angle for atoms 4 onwards. Also known as a Z-matrix, the description is extremely useful, as the generation of the molecular geometry is controlled by changing a relatively small number of variables, compared to much more computationally intensive calculations in the XYZ system. Hence, all Cartesian input formats are converted to redundant internal coordinates for computation within the majority of electronic structure program packages.

```

C
H 1 1.089000
H 1 1.089000 2 109.4710
H 1 1.089000 2 109.4710 3 120.0000
H 1 1.089000 2 109.4710 3 240.0000

```

**Figure 2.4** An exemplary internal coordinate definition of a molecular structure; methane (CH<sub>4</sub>) in this case.

**Numbering system** (Fig. 2.2). A standardised, modular and scalable atomic numbering system can be used in conjunction with internal coordinates making selected electronic structure determinations more efficient.<sup>151</sup> Apart from many other very important advantages, this system is useful for polymer or periodic systems such as peptides, lipids, carbohydrates and nucleic acids, among others.



## 2.2 Computational methods and Artificial Intelligence

Artificial computerised agents have been consistently used for various tasks which were previously carried out manually by humans. These range from securely processing online banking orders to computer-aided graphical design through to space exploration. Presently, many major areas of scientific research (mathematics, physics, chemistry, biology and medicine, engineering, among others) have specific branches that deal with using computers to efficiently process either historic or real-time data. Scientific computation, including modelling and simulation have become permanent and essential fixtures in all these areas.

The two big advantages that computer-aided processing has over human-mediated ones are as follows:

1. **Calculation speed.** Computers can presently process information very quickly, with processing speeds reaching up to 1.75 petaflops (1,750,000,000,000,000 floating point operations per second<sup>a</sup>).<sup>152</sup> The most efficient computers are usually composed of arrays of smaller units, also known as super-computers, or ‘clusters’. Even though 1.75 petaflops might seem like a very large number of calculations, current methods of computational chemistry require much more power even for comparatively small molecules consisting of tens of atoms.

---

<sup>a</sup> In computer science, the floating point system is a way to represent numbers which cannot be described as integers (*i.e.* non-whole-numbers). For example: extremely useful constants such as  $\pi$  ( $\approx 3.14159$ ),  $e$  ( $\approx 2.71828$ ),  $K$  (Catalan’s constant,  $\approx 0.91596$ ),  $N_A$  (Avogadro constant,  $\approx 6.022 \cdot 10^{23}$ ) are all examples of floating point numbers.

2. **Data storage space.** Although the first computer with a magnetic disk storage unit (IBM 350 RAMAC; announced September 13, 1956) could save only 5 million 7-bit characters (which works out to about 4.4 megabytes in modern parlance) and was leased for a \$35,000 annual fee, the technological progress has made much larger hard disk drives available for a reasonable price; 1 terabyte<sup>b</sup> of storage is common for most household personal computers in the year 2010.

Another very important part of a computer data storage space is known as Random Access Memory (RAM). RAM has a much smaller capacity than hard disk drives, but the access to information on a RAM chip is much faster, hence the reason for RAM being used by processors the majority of the time during calculation. Computers with 4 to 8 gigabytes of RAM are now (*circa* 2010) widely available for household computer budgets.

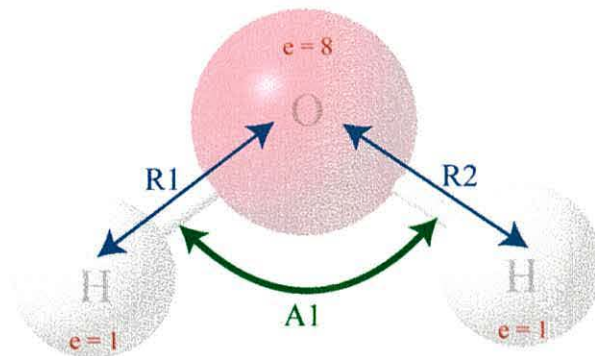
Various parts of computer science deal with using these two principle advantages for solving complex scientific problems. One of the most important notions in computational theory is **growth of functions**.<sup>153</sup> More completely described further on in this chapter, this term describes the situation where for some input of  $N$  numbers one always needs to calculate an approximate time it would take to produce

---

<sup>b</sup> One bit is the smallest amount of information that can be saved on a computer. One bit can either be 'off' or 'on', thus able to contain a value of 0 or 1, respectively. One byte (1 B) = 8 bits. 1 kilobyte (1 KB) = 1024 bytes. 1 megabyte (1 MB) = 1024 KB, or 1,048,576 B. 1 gigabyte (1 GB) = 1024 MB, or 1,073,741,824 B. One terabyte (1 TB) is equal to 1024 GB, or 1,099,511,627,776 B.

an output using a specific algorithm<sup>c</sup>. The function growth analysis always helps in setting approximate deadlines for computational projects, as it can fairly accurately estimate an algorithm runtime in real-time (human time). It also gives a means to mathematically prove that some problems cannot be solved in a reasonable amount of time, even without doing a single calculation for those problems (they are usually referred to as NP-complex problems, which are also defined in the following paragraphs).

The computational sciences abound with such problems. For example, let us take a simple chemical system that is considered to be the most important for life on Earth – water (Fig. 2.4). H<sub>2</sub>O consists of only three comparatively small atoms with a small number of electrons (8 in oxygen, and 1 in each of the two hydrogens).



**Figure 2.4** A three-dimensional visual representation of a water molecule showing its internal structural parameters, where R1 and R2 are the distances between hydrogen atoms and the oxygen, A1 is the angle between bonds; e is the number of electrons in atom.

---

<sup>c</sup> Algorithms are defined and described in more depth further on in this chapter.



Experimental physics and chemistry have generated enough information to make the problem of predicting the molecular behaviour of water intractable even using the most powerful super-computer arrays existing in the year 2010. The reason for this is that the number of possible states that this simple molecule can be in (*i.e.* the number of possible combinations of inter-molecular positions), is near-infinitely large. Moreover, the number of all possible electron distributions, electron-electron and electron-nuclei interactions, is also infinitely large. Considering all possible interactions in a group of water molecules is one of the most challenging and resource-consuming task. That is why, to create a discrete representation of a group of water molecules, the computational sciences use approximations that are sufficiently accurate to keep the theoretical results as close to experimentally observed values as possible. These approximations also allow scientific computations to concurrently make accurate predictions about molecular behaviour in a relatively short time. However, these models fail in when attempting to model the conditions in a biological system, where water mediates the observed phenomena. There exists a large amount of (interesting) work for this and the next generation of scientists, towards more completely understanding the behaviour of water at the molecular level.

Coming to the aid of these scientists is Artificial Intelligence (AI), a subset of computer science that provides a vast range of methods to deal with problems that require approximations. The fundamental concept of AI is to translate and describe a scientific problem to a computer (usually using one of the widely available programming languages like Java, C++, MATLAB, Assembly *etc.*), and subsequently take advantage of its high computational speed and large data storage

capacity to quickly perform calculations. These are usually performed in a serial manner (one calculation at a time), although Artificial Neural Networks (ANNs)<sup>154-177</sup> can model a parallel computation, where multiple calculations run independently at the same time (ANNs are briefly described further on in this chapter).

*As an additional note*, very often AI is directly associated with ANNs, which in practice is not often the case, as AI includes a wide range of computational techniques that are not directly based on neural network parallelism (*e.g.* Genetic algorithms,<sup>178-180</sup> planning,<sup>181</sup> Fuzzy logic,<sup>182</sup> A\* search algorithms *etc.*<sup>157, 183, 184</sup>)

The future of computational power is hard to predict at the moment. The long-term trend in the history of computer hardware described by Moore's law<sup>d</sup> is facing a barrier as the size of transistors is approaching the size of atoms. At the same time, new ideas and developments in the area present possible ways to create faster machines with larger data space capacities (*e.g.* quantum computing,<sup>185-187</sup> probabilistic quantum memories<sup>188</sup> *etc.*), as well as to create "smarter" machines (*e.g. memristors* found recently might be the key to building an artificial human brain).<sup>188-192</sup>

---

<sup>d</sup> Moore's law is an observation made in 1965 by Gordon Moore, co-founder of Intel, that the number of transistors per square inch on integrated circuits had doubled every ~18 months since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future.

## 2.2.1 An introduction to computational methods

A complete coverage of the full spectrum of computational methods with their underlying mathematics is certainly beyond the scope of this work; however, the most important, relevant algorithmic techniques are covered with the most important references also included. Although more specialised texts dealing with specificities of each part of algorithmic analysis exist, the most outstanding and easy-to-read texts on algorithms include “Introduction To Algorithms” by Thomas Cormen<sup>153</sup> and “The Art of Computer Programming” by Donald Knuth.<sup>193</sup>

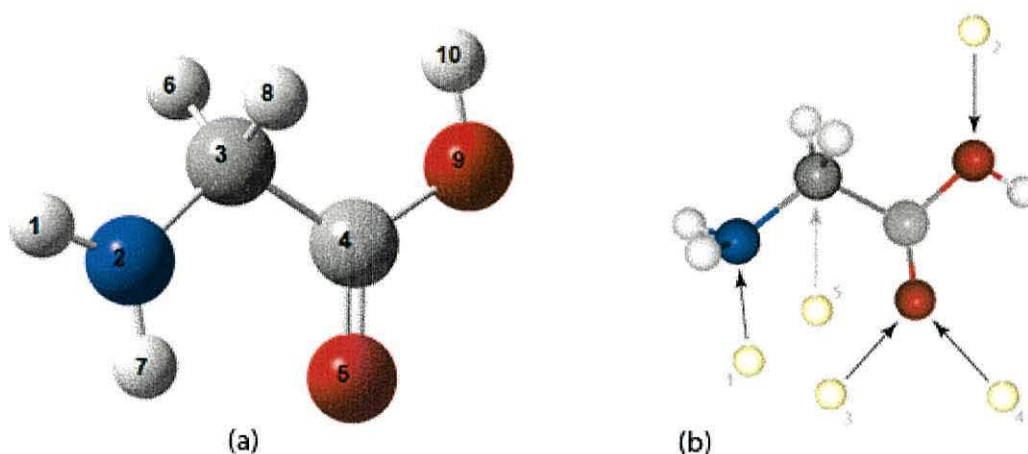
### 2.2.1.1 *Growth of functions.*

Undoubtedly, the most important technique in algorithmic design is the analysis of **growth of functions** for estimation of algorithm runtimes.<sup>193-201</sup> The idea is to calculate an approximate number of basic CPU operations that an algorithm will use for some particular input of  $N$  data elements. The resulting algorithm runtime is typically a function of  $N$ .

Most computer science textbooks use a sorting or a string comparison problem to explain growth of functions. To make this text more adapted to chemistry and physics, a real-life example from a sample computational chemistry project is used here. Specifically, taking a computational representation of a glycine molecule (Fig. 2.5, a), developing an algorithm to explore its conformational space and subsequently perform an automated modelling of its protonation. The latter was



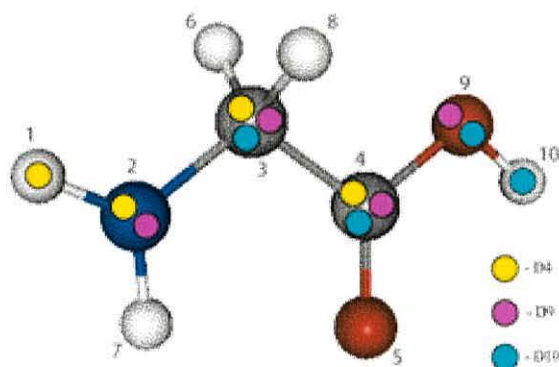
accomplished by placing positively charged hydrogen atoms (*i.e.* protons) at a certain distance from the molecule, in different locations; ‘chemical intuition’ helping guide this step. The number of these locations around glycine is infinite, however, general knowledge of chemistry and physics allows the set of locations to be reduced to a handful of most probable protonation sites (Fig. 2.5, b); the lone pairs on the carboxylic acid oxygens come immediately to mind.



**Figure 2.5** A 3D representation of a computer model of glycine amino acid (a) and its most important protonation sites (b). *Note: the protonation sites are not ranked by energy.*

Apparently simpler than protonation, the quantitative characterisation of the conformational preference of glycine, or many biologically relevant molecules, is in fact a very complex task. There exist a near-infinite number of topologically possible geometries that the molecule can populate on its respective PEHS. However, a generalised model can contain a greatly reduced set able to quantitatively describe glycine’s topologically probable set of conformers. Dihedral angles are usually considered to be the most influential degree of freedom in a molecule’s geometry, as a dihedral change can modify large parts of the structure with minimal energetic

expense. A dihedral can have a value in the range  $[0, 2\pi]$ , or  $[0^\circ, 360^\circ]$ . Selecting the three most influential dihedral angles, comprising the backbone (BB) of glycine (Fig. 2.6), each is predicted by Multi-Dimensional Conformational Analysis (MDCA)<sup>202-205</sup> to have a value of  $60^\circ$ ,  $180^\circ$  or  $300^\circ$  (covering the whole  $360^\circ$  circle with even intervals of 120 degrees, whereas  $300^\circ$  is equal to  $-60^\circ$ ).



**Figure 2.6** A colour-coded structure of glycine, depicting the three most influential dihedral angles.

Combining the two studies, in summary, glycine is predicted to have  $N = 3$  dihedral angles to characterise,  $V = 3$  possible values for each dihedral, and  $P = 5$  protonation sites.

Making the assumption that a basic operation (BO) creates one description of glycine with a set description of dihedral angles (geometry) and protonation sites, in order to generate all conformers of glycine without considering protonations, we need to complete the following number of BOs:

$$n_{BO}(N, V) = \sum_{d=1}^N V_d = N^V = 27$$



If we know that one BO takes  $t_{BO} = 1$  unit of CPU time (*e.g.* 1 second) on average, we can calculate that the execution time of this glycine-conformer generation algorithm will require approximately 27 units of time (*i.e.* 27 seconds).

To explore the effects of protonation on this molecule, one must generate all possible protonation combinations for each of the 27 conformers. For each protonation site one can either place a proton or leave it empty, making a total of  $2^P$  protonation combinations for each conformer, which totals to  $(27 \cdot 2^5 = 864)$  input structure files to be generated. As a final result of our algorithmic analysis, one could predict that this phase of the work would require 864 time units (*i.e.* 864 seconds, or 14 minutes and 24 seconds):

$$T(N, V, P) = n_{BO} t_{BO} = \left( 2^P \sum_{d=1}^N V_d \right) t_{BO}$$

In practice, it usually takes much less than a second to generate 10,000 structures on an average computer, but the algorithm is very simple – it browses through all possible conformers and protonations, and writes the appropriate values in the structure files. An addition to the running time could be done by checking each protonated conformer for degenerate structures / conformers as well as for obvious geometrical errors such as atomic collisions, bond crossing *etc.* For many algorithms, the exact analysis is impossible (for example, when no information is given *a priori*, a search of the PEHS is greatly dependent on the probability of finding a desired lowest-energy state in a certain region). In these and similar cases various scenarios are evaluated and characteristics such as worst-case, best-case and average-case running times are provided by the algorithm.

It is important to mention that some algorithms have exponential running times (*e.g.* the number of different protonation patterns is  $2^n$ , so if the number of protonation sites is only 100, then the number of patterns is 1,267,650,600,228,229,401,496,703,205,376, which would take about 4,000,000,000,000,000 years to process even at a rate of 10,000 structures per second). So, if one needed to find some specific structure among them, it would not be possible using current computational resources. Problems such as these are known as NP-complex problems, where NP stands for **nondeterministic polynomial**.

#### 2.2.1.2 Recurrences.

Recurrences<sup>153, 193, 200</sup> are used in many algorithms described in this work. The basic idea behind a recurrence function is that it keeps calling itself until a certain condition is met (otherwise it would call itself infinitely). The technical definition given by Cormen *et al.*<sup>153</sup> is that recurrence is an equation or inequality that describes a function in terms of its value on smaller inputs. It can be mathematically defined as follows:

$$f(N) = \begin{cases} c & \text{if } C(N) \\ af(n) + b & \text{if not } C(N) \end{cases}$$

where  $c, a, b$  are some constants;  $n$  is an input smaller than  $N$ ; and  $C(N)$  is true if  $N$  meets a specific recurrence relation ‘exit’ condition. For example, in the case of five potential protonation sites, the initial proton addition can occur at any one of these five. For the second protonation (*i.e.* resulting in a doubly-protonated structure), the

input set is reduced as only four protonation sites remain. After the fifth and final protonation, no free protonation sites remain, hence the recursion is exited.

### 2.2.1.3 *Computational geometry.*

Although most of the geometric concepts covered in this subsection are included in high school mathematics books and may seem trivial, they are essential for the understanding of the first version of the explicit particle solvation algorithm presented in the “Results and discussion II” section.

A **Euclidean point** is an object in  $n$ -dimensional space, denoted by  $(d_1, d_2, d_3, \dots, d_n)$ , where  $d_i$  is the  $i^{\text{th}}$  dimension of the point (the distance from  $d_i$  to zero in the  $i^{\text{th}}$  dimension).

The most basic geometric calculation is the **distance between two points**. Although most of this work involves three-dimensional space, multi-dimensional PEHSS involve much higher dimensionality (several thousand for some systems), hence a general formula is included for calculating the distance between two points in an  $n$ -dimensional space:

$$D_n = \sqrt{\sum_{i=1}^n (d_{i_1} - d_{i_2})^2}$$

For example, in 3D space the distance between two points is expressed as follows:

$$D_3 = \sqrt{\sum_{d=1}^3 (d_1 - d_2)^2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

A **Euclidean vector** is a geometric object that has both magnitude and direction. A point  $P = (3, 5, -8)$  in Cartesian space can be viewed as a vector from  $(0, 0, 0)$  to  $P$ , the length thereof equivalent to the distance between its endpoints.

A **cross product of two Euclidean vectors  $\mathbf{a}$  and  $\mathbf{b}$**  is denoted by  $\mathbf{a} \times \mathbf{b}$  and defined by the following formula:

$$\mathbf{a} \times \mathbf{b} = \hat{n}ab\sin\theta$$

where  $\hat{n}$  is a unit vector that is orthogonal (perpendicular) to both  $\mathbf{a}$  and  $\mathbf{b}$ , and  $\theta$  is the angle between  $\mathbf{a}$  and  $\mathbf{b}$ .

A **dot product of two Euclidean vectors  $\mathbf{a}$  and  $\mathbf{b}$**  is denoted by  $\mathbf{a} \cdot \mathbf{b}$  and defined by the following formula:

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta$$

where  $|\mathbf{a}|$  denotes the length of vector  $\mathbf{a}$ .



A **scalar product of a Euclidean vector  $\mathbf{a}$  and a constant  $c$**  is another vector  $\mathbf{b}$ , in which each  $d_i$  from  $\mathbf{a}$  is multiplied by  $c$ . For example, if  $\mathbf{a} = (1, 2, 3, 4, 5)$  and  $c = -5$ , then  $\mathbf{b} = (-5, -10, -15, -20, -25)$ .

Another important concept in computational geometry for which an algorithm has been included is for **determining whether two line segments intersect**. It is described in detail, with pseudo-code, by Cormen *et al.*<sup>153</sup> and is very useful for quickly determining whether two bonds in an automatically generated chemical structure intersect.

#### 2.2.1.4 *Data structures.*

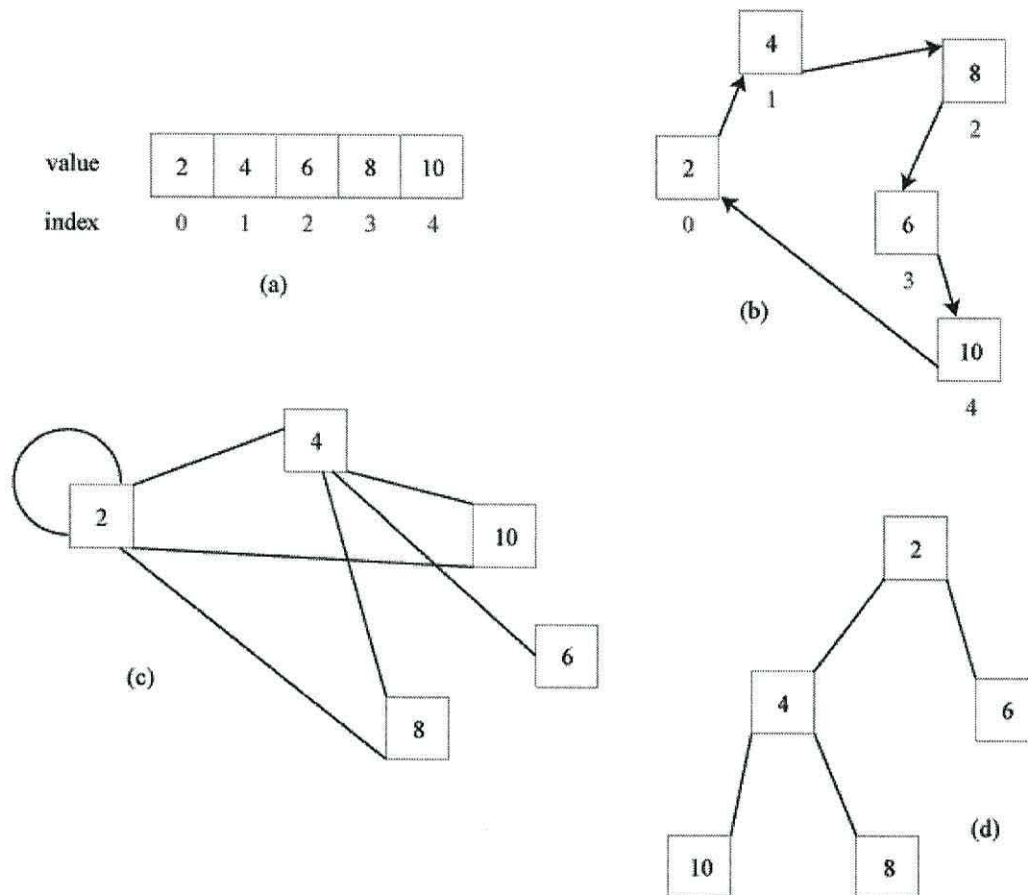
Data structures<sup>153, 193, 200</sup> are covered prior to details on the algorithms, as the understanding of most algorithms is impossible without a preliminary description and understanding of the underlying data hierarchy. Hence, a very brief review of data structures used in this work's algorithms is presented first.

Apart from the data structures designed for use in this work building on basic data entities (*e.g.* CPoint and CVector in the ChemConverter project described in later chapters), the following data structures have also been used:

- **Array (Fig. 2.7, a).** An array is a data structure in which objects are arranged in a linear order and are accessed by their index. Array indexing usually starts from 0. For example, an array of 5 elements could be [2, 4, 6, 8, 10].

'4' in this array is located at index 1; index 0 would contain the number '2'.

Text is usually represented by an array of characters. Queues and stacks are usually coded as arrays with various additional 'pointers'.



**Figure 2.7** A schematic representation of computer data structures: array (a), linked list (b), graph (c) and tree (d).

- **Linked List (Fig. 2.7, b).** In a linked list, objects are also arranged in a linear order, with the order being defined by a pointer to each object (and not by its index, like in array). In a singly-linked list the first element (head) contains a reference (pointer) to the second element, which contains a reference to the third element, and so on. The last element (tail) usually contains a reference

to the head element. In a doubly linked list each element keeps a reference to both the next and the previous nodes.

Linked lists are used for dynamic storage, *i.e.* when there is no way to say how many elements a list will contain (arrays, on the other hand, have a defined size and cannot contain more elements than the programmer decided).

- **Graph (Fig. 2.7, c).** A graph is an abstract representation of a set of objects where some pairs of the objects are connected by links. The objects are usually called nodes, and the connections between them – edges. Graphs can be directed (when edges have orientation) or undirected. Graphs can also be cyclic (where edges form cycles) or acyclic. Graphs usually have two representations: as an adjacency matrix of all connections between nodes (non-existing edges are considered to be zero), or as an adjacency list where each node contains reference only to non-zero edges connected to it.
- **Tree (Fig. 2.7, d).** A tree is an acyclic graph where each node has zero or more ‘children nodes’ and at most one ‘parent node’.

One of the clearest examples of using a computer data structure in theoretical physical chemistry is representing molecules using *graphs*; all molecular structures of ‘balls and sticks’ are in fact graphs. In this case, atoms are nodes and the conventional bonds are represented by edges. Apart from direction, each edge can have a weight (*i.e.* a numerical value). In this simple example, “bond edges” could contain an integer number of electron pairs involved in bond formation, a floating point number representing charge density, or any other numerical characteristic of a bond. Graphs can be extended and changed very quickly – the explicit particle solvation algorithm presented within this work uses this advantage to add multiple

layers of solvent particles around a solute molecule: before adding the  $n^{\text{th}}$  layer, the molecule is extended by the previous  $(n - 1)$  layers, to make the geometric and electronic density calculations much faster. Graph theory algorithms are also used to work with the standardised atomic numbering system previously described: distances, angles and dihedral angles for each atom are defined by traversing a graph of connections (bonds) from the atom to all other atoms.

Molecular dynamics simulations are defined by a certain number of steps that a model molecular or material system ‘makes’ depending on the desired results and different conditions (*i.e.* lowest-energy search, varying temperature, pressure *etc.*). At each step, there may be a few possible next steps. Although only one is selected, explored and the path subsequently characterised by most computational chemistry packages, *a tree* of all possible pathways could be constructed to more fully characterise the associated PEHS.

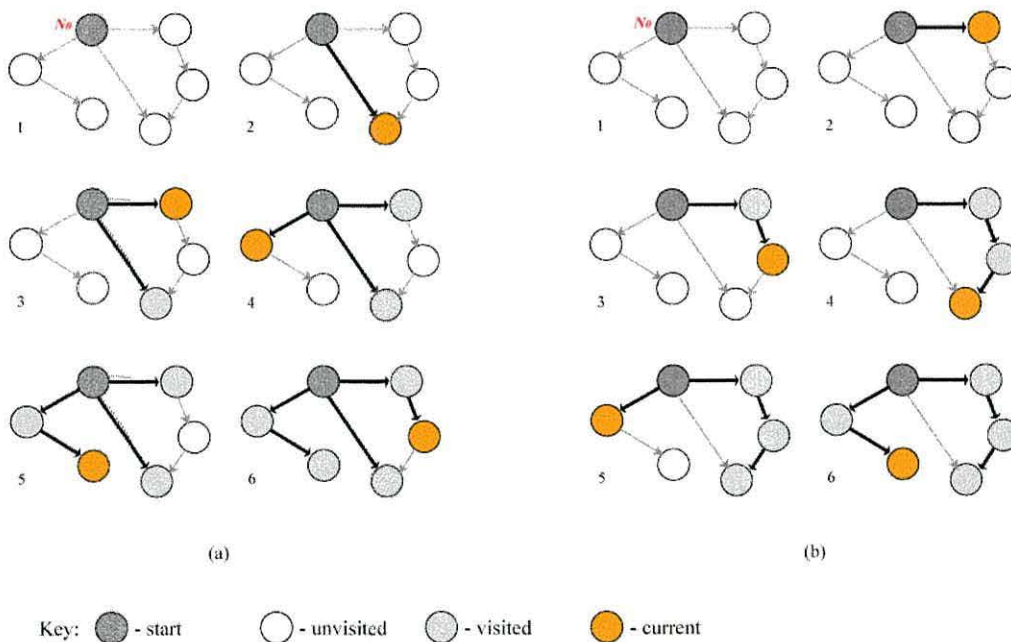
*Arrays* and *linked lists* could be used to keep track of all possible protonation sites and their descriptions, or to detect the lowest-electron-density region on the PEHS during solvation, for example. Multidimensional arrays, nested linked lists and their combinations are also used to represent more complex data structures such as graphs, trees, heaps *etc.*



## 2.2.2 Algorithms used in this work

### 2.2.2.1 BFS and DFS.

The two most widely used graph algorithms are **Breadth-First Search (BFS)** and **Depth-First Search (DFS)**.<sup>153, 193, 200</sup> The number of real-life computational problems solved by these two algorithms quite probably exceeds the number of problems solved by all other graph algorithms taken together. The basic idea of both algorithms could be described as follows: given a node  $N_0$  in a graph (or tree, or any other similar data structure), traverse the graph in search of some other node (or nodes). Figure 2.8 presents a graphical example of a traversal that checks all nodes that can be accessed from  $N_0$ ; some graphs are not connected, which is the case when some nodes might not be accessible from  $N_0$ . In this case, one usually reports that the distance between  $N_0$  and the non-accessible node is equal to infinity.



**Figure 2.8** Schematic representations of the BFS (a) and DFS (b) search algorithms.

The idea behind BFS (Fig. 2.8, a) is to start from  $N_0$  and initially examine all other nodes accessible from  $N_0$  (*i.e.* at a distance of 1 step), then examine all nodes accessible in 2 steps (*i.e.* all nodes accessible from the ‘1-step’ nodes in one step), and so on.

In DFS (Fig. 2.8, b), the idea is to pick any node  $N_1$  accessible from  $N_0$ , mark it as ‘visited’, then pick any unvisited node  $N_2$  accessible from  $N_1$ , mark it as ‘visited’, and so on. Thus, in BFS the search is done in ‘layers’: all nodes at distance 1 from  $N_0$  are initially explored, then nodes at distance 2, then 3 and so on; in DFS, the search initially goes as deep as possible *via*  $N_1$ , then returns a step back, picks another unvisited node accessible from  $N_0$  and goes as deep as possible *via* the node.

#### 2.2.2.2 *Optimisation algorithms.*

In the first version of the explicit particle solvation algorithm, **Greedy Algorithms** are combined with **Computational Geometry Algorithms** to determine solvent molecules’ locations. The basic idea behind a greedy algorithm is to choose the best option available at any step of the optimisation process. Some or all previous steps can be considered, but no prediction of how the choice will influence the following steps is made. We believe using greedy strategies would be a good initial approximation of the real solvation process, as solvent molecules are not given any *a priori* information about the solute system – they solvate the best available spot on the solute surface, *i.e.* they interact with the part of the whole solute system in the best way possible at any given time. **Linear Programming (LP)** and **Dynamic**

**Programming (DP) Algorithms** will be the tools of choice for the second version of the solvation algorithm which will include more precise charge density distribution descriptions, as LP and DP can cover a wide range of combinatorial possibilities of solvent-solute interactions in an optimal way.

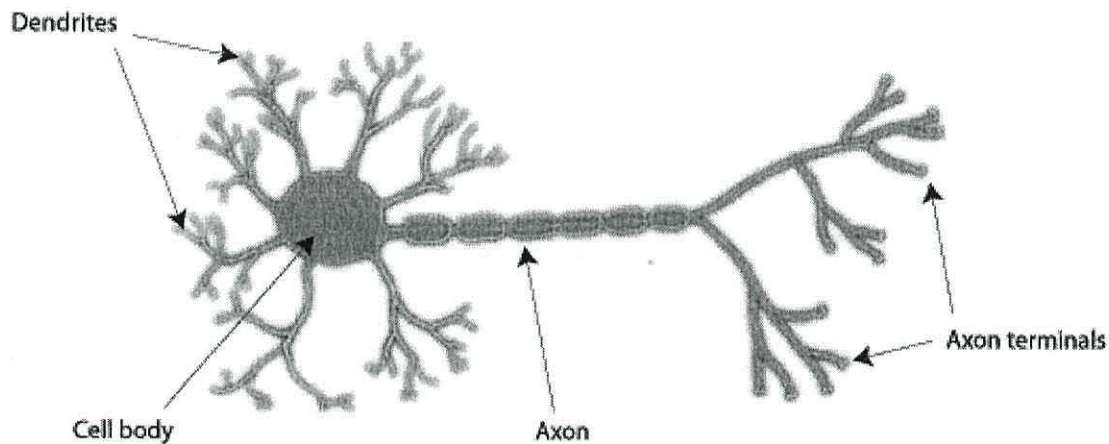
### 2.2.2.3 *Genetic Algorithms and Neural Networks.*

Another two branches of Artificial Intelligence algorithms that were indirectly used in our work are **Genetic Algorithms (GA)**<sup>47, 48, 65, 178, 179</sup> and **Artificial Neural Networks (ANN)**.<sup>24-28, 60-62, 66, 69, 71, 73-75, 154, 156, 158-168, 170-173, 175, 182, 206-210</sup> ANNs and GAs have been used extensively to develop various pattern recognition tools in many areas of science. Since patterns can be observed so often in the behaviour of matter, one of the major long-term goals of this work has been the development of an ANN that would ‘understand’ molecular behaviour to the level of a highly experienced scientist, *i.e.* which could analyse charge density distributions without extensive computations – mostly by recognising, generalising and predicting miscellaneous atomic interactions. At the beginning of this project, ANNs and GAs were of the first priority, but it was realised very soon that in order to generate enough data for the main ANN to learn from, many daily tasks had first to be completely automated. Otherwise, the efficiency of the data generation process would be very low.

Although artificial neural networks were not extensively used in the experimental work outlined in this work, a short introduction to neural networking in artificial intelligence is provided, as a large part of relevant future works will be based on



ANNs. These ANNs<sup>154-177</sup> are designed to simulate the operation of natural neural networks, such as the human brain. ANNs consist of layer(s) of artificial neurons, which simulate the behaviour of biological neurons,<sup>157, 158, 162-164</sup> which in general are fairly simple processing engines (Fig. 2.9).

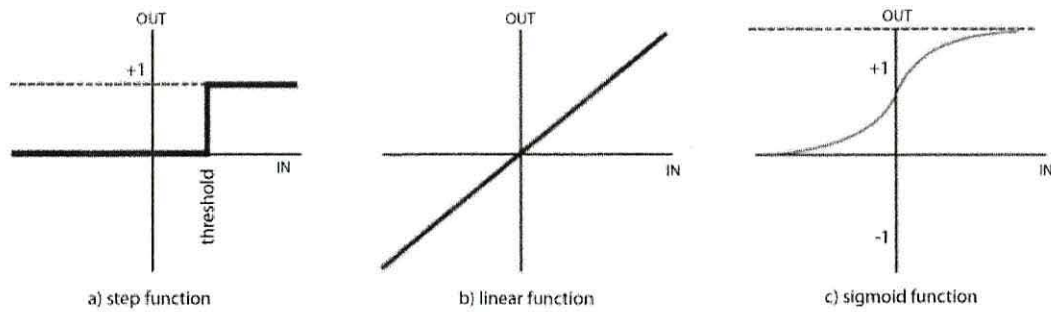


**Figure 2.9** A schematic representation of a biological neuron and its components.

The neuron receives information from other neurons along its dendrites. When the total collected input reaches a certain threshold, the neuron sends an action potential down the axon (*i.e.* an output signal). Artificial neurons (often called “McCulloch & Pitts neurons”,<sup>211</sup> after their namesakes) also collect inputs from other neurons in the artificial neural network (“inner” neurons), or from the outside world (“input” neurons), and fire an output to other neurons or to external acceptors when some activation level has been reached. There are various functions which represent the activation level (also known as the “response condition”).<sup>157-159, 162-164</sup> The three functions on Figure 2.10 represent the most generic of these: **step function** (Fig. 2.10, a) – neuron outputs 0 (or -1) when the cumulative input value is smaller than the *threshold*, or 1 otherwise; **linear function** (Fig. 2.10, b) – similar to the step

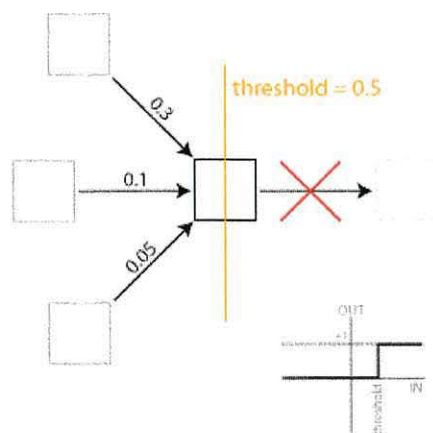


function but the output can be 0 (or -1), 1, or a value between the two (e.g. 0.5, -0.12 etc.); **sigmoid function** (Fig. 2.10, c) can be described as a function which converts any cumulative input from the range  $(-\infty; +\infty)$  into an activation level in the range  $[0; +1]$  or  $[-1; +1]$ .



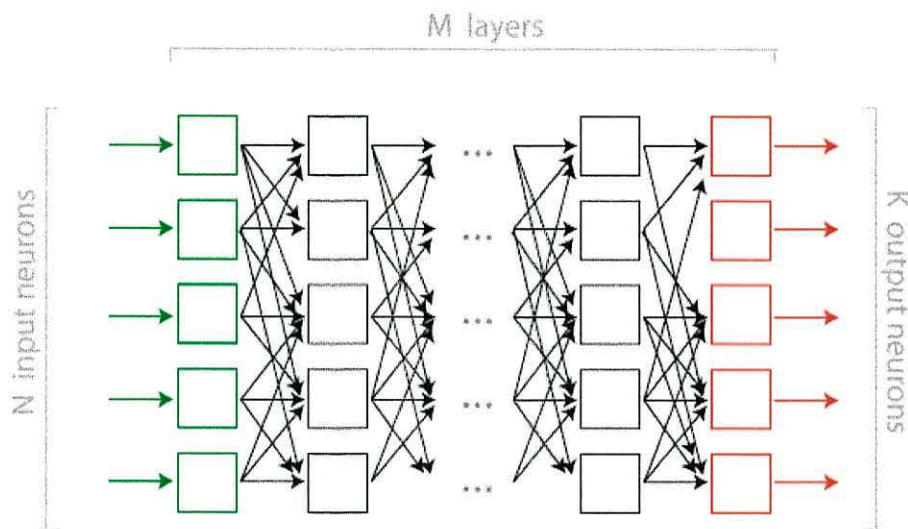
**Figure 2.10** Graphical representation of differing artificial neuron activation functions: Step (a), Linear (b) and Sigmoid (c).

For example, if a neuron with a threshold value of 0.5 uses a step function for output production and receives three inputs (Fig. 2.11) that sum up to 0.45, there will be no output from the neuron.



**Figure 2.11** Exemplary neuron activation for a step function; in this case the total input is below the 0.5 activation level.

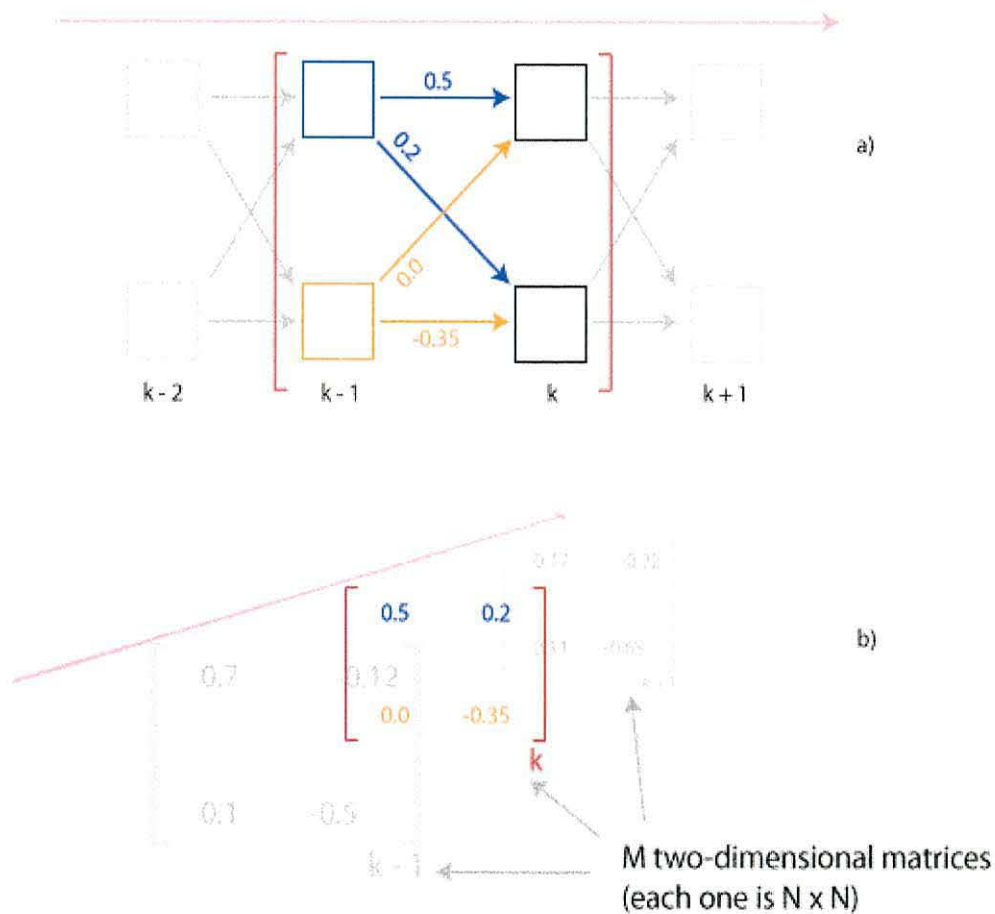
A neural network usually consists of more than 1 neuron. Large natural neural networks such as the human brain contain over 10 billion neurons and approximately 60 trillion connections between them.<sup>157</sup> Modern ANNs are much smaller and contain fewer connections between neurons. A simple neural network can be represented as a three-dimensional matrix (Fig. 2.12).



**Figure 2.12** Schematic representation of a sample artificial neural network with  $M$  layers.

The first layer (highlighted in green) is the input layer; information from the outside world comes into the network through this layer. The last layer (highlighted in red) is the output layer; processed information leaves the network through this layer, and represents the result of the job that was performed by the network. Every other layer is called a hidden layer; the layers of neurons that accomplish the largest part of the work. If every layer has  $N$  neurons (thus,  $K$  is also equal to  $N$ ), then the matrix dimensions are  $N \times N \times M$ . The  $k^{\text{th}}$   $N \times N$  matrix of all  $M$  2-dimensional matrices is the matrix which describes how each neuron in the  $(k - 1)^{\text{th}}$  layer influences every

neuron in the  $k^{\text{th}}$  layer; an example is given in Figure 2.13. As information progresses through the network, the input data goes from layer 1 to layer 2, then to layer 3, and so on (pink arrow on Fig. 2.13, a). It is identical to “passing” an input data vector through matrices 1, 2, 3, ...,  $M$  (or speaking algebraically, multiplying the matrix by the input vector and passing the product further along the 3<sup>rd</sup> dimension of the main matrix (along the pink arrow in Fig. 2.13, b)).



**Figure 2.13** An example of a  $2 \times 2 \times M$  matrix representing a simple neural network. (a) graph representation, (b) 3D matrix representation.

This matrix representation is very convenient due to the vast amount of powerful mathematical frameworks available, which work with matrices and vectors. However, sometimes it can also be very inefficient: when the number of neurons is very large and the number of connections between them is comparatively very small, one has to evaluate a large amount of “ $X \times 0$ ” products while multiplying the matrices with the input vector (where  $X$  is a number). In that case it might be helpful to represent the neural network as an **adjacency list** (a linked list data structure in which every connection is stored in a list joined to the neuron from which the connection starts). This way one can reduce the number of elementary operations used to process data in the ANN, and also reduce the amount of space that is required to store the artificial neural network.

Even for a comparatively small neural network the number of possible ways to assign threshold values to neuron connections is infinitely large, which allows the neural network to contain large amounts of very structured information.



### 2.2.3 Programming languages, frameworks and tools

One of the most efficient ways to communicate with a computer is *via* a programming language. There are many different programming languages with various options available to programmers. The most efficient language is certainly the one that the Central Processing Unit (CPU) understands directly - Machine Language. Although Machine Language allows a programmer to exploit the full power of a CPU, it is not very user-friendly, as each command must be encoded by a binary (or a hexadecimal) number. That is why Assembly language<sup>212</sup> was created, allowing one to explain computational problems to a computer using more user-friendly commands (*e.g.* a Machine Language command to move data between CPU memory locations *esp* and *ebp* would be “89 E5”, whereas in Assembly it looks like “mov %esp %ebp”). Assembly is a very powerful tool in programmer’s hands, but more sophisticated and much more user-friendly programming languages have been developed to make computer programming less language-oriented and more problem-oriented. For a talented programmer it literally takes only a few seconds to write a simple application in languages such as C++, Java and Python, so programmers and scientists can concentrate more on scientific problems and use computational power almost without any special technical knowledge of how CPUs, Random Access Memory or any other components of computer hardware work.

Many frameworks have been developed for languages such as Java, C++, Python, C# *etc.* These frameworks contain the code for solving basic tasks such as creating a piece of text from characters, data structures, basic algorithms, mathematical

functions and more (for example, the *java.util* Java framework contains many useful data management functions that in many cases can be used without modification).

Integrated Development Environments (IDEs) are used to make computer programming even simpler – they offer features like automatic code completion and code snippets, visual application development, database access *etc.* The IDEs used in this work include the following:

- NetBeans (version 6.8) – the best Java development studio available at the time of this writing;
- Microsoft Visual Studio 2008 – we used it for C++ development;
- IDLE – a Python development environment;
- GNU C / Fortran / Assembly compilers and linkers.

These are also described in the “Methods” section of this work.

## 2.3 Summary

The interdisciplinary focus of this work and the fascinating problem of molecular behaviour unravelling require both the attention to the very technical side of computational studies (such as the mathematics behind the efficient approximations of molecular descriptions, the most efficient programming language compiler / CPU matching *etc.*), and the insight acquired from the centuries of experimental physical and chemical research. Therefore, in this work we set the goal of collecting and analysing well-structured data about molecular behaviour not only by following the ‘21<sup>st</sup> century way’, but also by acknowledging the importance of understanding the vast amounts of experimental evidence and experience generated by the scientific community in the past years.

### 3. METHODS

---

In this study the following experimental / theoretical work was performed:

- Elastic and Inelastic Neutron spectroscopy, FTIR spectroscopy;
- Electronic Structure Calculations, using the Gaussian03 program package (g03),<sup>213</sup> including: geometry optimisations, analytical frequency determinations and wave function generation all in various solvents, using different methods and levels of theory / basis sets;
- *Ab initio* Dynamics Calculations, using the Car-Parrinello program package (CPMD),<sup>214</sup> geometry optimisations, dynamics and wavefunction generation;
- Atoms-In-Molecules Wavefunction (AIM)<sup>215-220</sup> analyses of wavefunctions generated from geometry-optimised molecular structures, using *6d* Cartesian molecular orbitals in the place of the *5d* ones traditionally used;
- Algorithm developments: solvation by explicit solvent particles, coordinate system conversion, grid-based conformer generation, peptide protonation patterns, multithreaded parallelised network-wide batch job execution, preliminary output tabulation, IR spectra plotting and peak detection, geometric structure exploration and analyses; a prototype of a neural network and its connection to a database of pre-computed structures.



### 3.1 Experimental

- **Silane Primer Preparation.** All silane primers were prepared as 1.0 vol% primers in a standard solution of 95.0 vol% ethanol, and de-ionised water (milli-Q water) which was first adjusted at 4.5 pH with 1 M acetic acid and then allowed to stabilise for 24 h prior to use. The silane primers were activated (allowed to hydrolyse) for 1 h at room temperature before bonding testing.
- **Infrared Spectroscopy.** The silane monomer hydrolysis was observed analytically up to 60 min using Reflectance-Absorbance Fourier Transform Infrared (RA-FTIR) Perkin Elmer Spectrum One spectrometer (Perkin-Elmer, Beaconsfield, UK) which detects different molecular bending, vibration, wagging and rocking of the functional groups. The surface analysis of a silane primer film layer was conducted throughout the 600-3800  $\text{cm}^{-1}$  spectral range with a specular reflectance monolayer / grazing angle accessory in which the spread silane primer film was on a cleaned, planar, inert Ge crystal.

## 3.2 G03 and CPMD calculations, AIM analysis

The theoretical part of this work was completed using the following list of applications:

1. **Gaussian03.**<sup>213</sup> G03 was the most extensively used tool in this work. It provides the most extensive list of applications for electronic structure determinations. The basic structure of an input file is as follows:

```
route section
title section
molecule specification
```

For example, the following input file was used for calculating the lowest-energy (optimised) geometry of a glycine molecule and its characteristic Infra-red frequencies, using the second Møller-Plesset Perturbation Theory (MP2) with the **6-311++G(d,p)** basis set, 1 CPU, 1 gigabyte or RAM, limited to a maximum of 200 optimisation steps (molecule specification is given in the Z-matrix internal coordinate format):

```
%nproc=1
%mem=1gb
%chk=hn-g-oh-1_paa_m2f_611++dp.chk
%rwf=hn-g-oh-1_paa_m2f_611++dp.rwf
#p mp2(full)/6-311++G(d,p) optcyc=200 opt 6d freq
IOp(1/8=6,2/9=2,2/11=1)

hn-g-oh-1_paa_m2f_611++dp

+0 1
H
N 1 R2
C 2 R3 1 A3
C 3 R4 2 A4 1 D4
O 4 R5 3 A5 2 D5
```

```
H 3 R6 2 A6 1 D6
H 2 R7 3 A7 4 D7
H 3 R8 2 A8 1 D8
O 4 R9 3 A9 2 D9
H 9 R10 4 A10 3 D10
```

```
R2 1.00
R3 1.40
R4 1.54
R5 1.25
R6 1.10
R7 1.00
R8 1.10
R9 1.40
R10 1.00
A3 120.00
A4 109.50
A5 120.00
A6 109.50
A7 120.00
A8 109.50
A9 120.00
A10 109.50
D4 180.00
D5 0.00
D6 -60.00
D7 0.00
D8 60.00
D9 180.00
D10 180.00
```

Herein, the `opt` command was used (perform a geometry optimisation; which can have various options such as the maximum number of steps, molecule output format *etc.*), `freq` (determine analytical vibrational frequencies), and `wfn` (generate a wavefunction from the final geometry). The most widely used methods included: DFT hybrid functional B3LYP (Becke, three-parameter, Lee-Yang-Parr);<sup>221, 222</sup> Restricted Hartree-Fock (RHF); and post-HF including second order Møller-Plesset Perturbation Theory (MP2), so as to include electron correlation. MP2 used principally for smaller systems such as amino acids, water *etc.* There is a large number of basis sets designed for various systems,<sup>213</sup> herein basis sets best suited for each particular molecular system were used.

Both Cartesian (XYZ) and internal coordinate formats were employed for molecular structure specification, dependent on the problem to be solved.

2. **GaussView Molecular Visualisation** was used on a limited basis, primarily for visually examining input / output geometries.
3. Three types of **CPMD**<sup>214</sup> calculations were performed: geometry optimisations, dynamics and wavefunction generation. A sample wave function (WF) calculation input is structured as follows:

```
&INFO
  System: 3-Acryloyloxypropyltrimethoxysilane molecule.
  Job:    Wavefunction optimisation.
  Author: Mykola Rozhok
  Date:   March 2, 2010
&END

&CPMD
  OPTIMIZE WAVEFUNCTION
  PCG MINIMIZE
  TIMESTEP
    20
  CONVERGENCE ORBITALS
    1.0d-7
&END

&SYSTEM
  ANGSTROM
  SYMMETRY
    1
  CELL
    20.0 1.0 1.0 0.0 0.0 0.0
  CUTOFF
    60.0
&END

&DFT
  NEWCODE
  FUNCTIONAL LDA
&END

&ATOMS
  *H_SG_LDA KLEINMAN-BYLANDER
  LMAX=S
    20
    1.87565259 -0.52491254 -0.90472597
    ...
&END
```



The ATOMS section is much longer, as it contains the definition of all atoms in the system. The difference between other calculations is in the dynamics section; for example to run an *ab initio* molecular dynamics simulation, one would use the same input file with the following CPMD section definition:

```
...
&END

&CPMD
MOLECULAR DYNAMICS
RESTART WAVEFUNCTION COORDINATES LATEST
TRAJECTORY XYZ
ISOLATED MOLECULE
TEMPERATURE
  300.15
MAXSTEP
  10000
TIMESTEP
  4.0
&END

&SYSTEM
...
```

For more information about various parameters and a more in-depth explanation of all CPMD capabilities see Ref. <sup>214</sup>.

4. Similarly to GaussView, **VMD**<sup>223-230</sup> was primarily used for manually checking the results of CPMD calculations. It also allows for visualisation of the dynamic changes of molecular geometry over time.
5. **AIM2000**.<sup>215, 220, 231-234</sup> A program package that creates molecular graphs of electronic density. These graphical representations of inter-atomic interactions are based on the program's ability to use a wavefunction to compute density critical points of three kinds: Bond-, Ring-, and Cage-Critical Points (BCP, RCP, and CCP, respectively; an interested reader can

also consult the following website with a very easy-to-follow tutorial on AIM theory and the application: <http://www.aim2000.de/>). For most systems, three types of electron density determinations were employed: Critical Points, Molecular Graph, and Line Plot. All calculations were performed either with default parameters or with the “Stepsize factor for Newton iteration” parameter changed from 1.0 to 0.5, for more precise density calculations.

For Line Plot density calculations’ Contour Interval, the default values were used for most cases, very rarely adding more lines between rows 0 and 2, using an arithmetic progression. Both Grid dimensions were increased from 30 to 300. Point, Line, and Path styles were chosen to make plots more easily interpretable. Page and Plane setups were decided for each system, based on the problem to be solved, and relevant results to be presented.

### 3.3 Algorithms

One of the largest parts of this work was devoted to a concurrent development of novel algorithms and applications / scripts for various manual tasks: from basic ones such as running calculations on multiple machines in a network, to more advanced tasks such as molecular protonation and placement of explicit solvent particles in an optimal and non-combinatorial (trial and error) manner, based on molecular intuition.

All scripts were written in Java and Python programming languages, with additional ‘helper’ scripts written in Bash, Assembly and C/C++. Chapter 4 contains the summary of algorithms, computational techniques and programming frameworks used in each project (common frameworks like *java.io*, *java.util*, *javax.swing*, Python’s *os* and *system* for the sake of efficiency were not included in the descriptions, as these frameworks are used by most applications). However, a few methods that need special attention are outlined below:

- Most applications written in Java used the Java Swing graphical library, as well as some basic AWT routines;
- Both Java and Python applications used multithreading techniques whenever running multiple calculations at the same time was necessary;
- Virtually all file and database access tools were written in low-level programming languages such as C and Assembly to ensure the most efficient time / performance ratio;

The artificial neural network prototype has the following characteristics:

- Most parts are written in Java programming language, with the exception of the database storage scheme and access, which is coded in C / Assembly;
- The network has a connection to a database of *ab initio* pre-computed structures of various molecules;
- It uses unsupervised reinforcement learning the majority of the time, with the exception of the “technical” parts, such as generating approximations of electron distributions for different atoms, bond formation / breaking *etc.*, where a supervised learning scheme is used in its place.



## 4. RESULTS AND DISCUSSION (Part I)

---

As this work presents results at the interface of the physical and computer sciences, the “Results and discussion” section has been split into two sections; one covering all computational / algorithmic and technical aspects of the work (Part I), while the other has focus on the results obtained for molecular systems investigated during the project using ‘in-house’ computational applications in conjunction with software developed by other scientific groups (Part II).

Part I highlights the following two main areas for which we developed our own algorithms, applications and scripts:

1. *Computational tools for input structure generation, network-wide calculation execution and management, and data extraction applications* describes a few easy-to-use tools that simplify input file generation, calculation control and analysis, thus saving much time;
2. *Explicit Particle Solvation algorithm* outlines the results from the “Solvator 1.2” application, used to solvate solute systems with a close-to-optimal amount of explicit solvent particles used. As previously noted in the “Introduction” section, the application has the advantage that it adds one explicit solvent particle at a time, allowing the user to track the influence each particle has on the solute and other solvent molecules.

#### **4.1 Computational tools for input file generation, network-wide calculation execution and management, and data extraction applications**

Combinatorial approaches to problems in theoretical physics, chemistry, biology, medicine, engineering and many other branches of science have been widely used in the past.<sup>235-239</sup> The ‘grid-based’ search approach is, in its simplest form, a brute-force approach: each possible combination of degrees of freedom is evaluated.

The approach can be optimised to make these brute-force grid-based methods much more efficient. For example, using a relatively large molecular system such as a protein fragment, with 50 principle degrees of freedom, each having 3 possible values, the generation of all conformers would require  $\sim 10^{24}$  input files, which becomes unmanageable even on the most powerful computers. Checking all these structures for geometrical errors is an even more time-consuming task. A simple task such as initially checking the structures at each step of the input structure generation process can decrease the total number of calculations by a very large percentage (*i.e.* when the combinations for the first few degrees of freedom have been selected, a check for geometrical errors such as bond crossing and atom overlaps). One could also incorporate experimental data such as general trends and existing knowledge of protein structure to identify highly improbable structures so as to exclude them. These types of optimised grid-based approaches allow for a much smaller subset (relative to the initial one) of input structures to be generated and subsequently characterised. One of the goals of this work was to develop a set of combinatorial

tools that would efficiently explore the conformational space of a molecular system and provide a framework for time-effective calculation creation, execution and management.

#### 4.1.1 Coordinate systems and input structure files

As previously mentioned in the “Introduction” and “Methods” sections, both internal (Z-matrix) and Cartesian coordinate systems were used in this work; the former being preferred for conformational space exploration. For ease of conversion between these – and any other – coordinate sets, a tool known as **ChemConverter** was developed. Although many other converters exist (*e.g.* the Open Babel package)<sup>240, 241</sup> and many computational chemistry applications (GaussView, VMD, ChemCraft<sup>242</sup> *etc.*) have in-built file format conversion capabilities, ChemConverter was created to convert not solely *between* various formats, but *to* and *from* the standardised, modular and scalable atomic numbering system, allowing for effective description of the relative spatial orientation of all constituent atoms in the structure investigated, in turn allowing for more effective automation of results extraction, analyses and trend recognition.<sup>151</sup>

ChemConverter runs using the Java Virtual Machine from a command line. From a programmer’s point of view, it has one interface with the system (*e.g.* command-line argument parser / processor) and one modular and scalable internal interface called *chemconvert.converters.Converter* that allows one to create new converter



specifications (*i.e.* extend the package without any modifications to the core program). At the present time, ChemConverter has one implementation class – *chemconvert.converters.Xyz2IntConverter*, which converts files between Cartesian specification and the established format.<sup>151</sup> The application currently accepts the following four input parameters:

1. **-t** (type); defines coordinate system types. After this parameter a conversion type is entered; for example, the “xyz2int” conversion changes between Cartesian (xyz) and Internal (int) coordinates. ChemConverter can include many other conversion formats – these must be files written in Java 1.1 or higher, that implement the *Converter* interface class that contains only one interface method, as follows:

```
public void convert(String inputFile, String descriptionFile, String
outputFile);
```

2. **-i** (input). As an example, for “xyz2int” the input file must be an XYZ file with the following 4-column format, where columns 1, 2, 3, 4 define the atom type, x-, y-, z-coordinates, respectively, as follows (note that column 1 can either contain the atomic symbol or atomic number):

```
H      -0.611647      0.000000      -2.724666
N      -0.712823      0.000000      -1.727114
C       0.469741      0.000000      -0.904244
...
```

3. **-d** (descriptor). The descriptor input file exploits the full range of possibilities presented by the standardised atomic numbering system.<sup>151</sup> In this file, one can specify the building blocks of the converted system. For



example, one can create a file with the atomic description and numbering of glycine and then write a descriptor file in the following way to create a protein consisting of 5 glycine molecules with an additional N=O group covalently bound between the 2<sup>nd</sup> and 3<sup>rd</sup> residues, as follows:

```
Gly
Gly
N
O
Gly
Gly
Gly
```

The application will then locate the *Gly.inr* file and use the information about glycine structure to recreate the whole molecule. A glycine *.inr* file has the following format:

```
10
0
H
N 1
C 2
C 3
O 4
H 3
H 2
H 3
O 4
H 9
9
```

which means that a glycine molecule has 10 atoms, the previous building block can be connected to atom 0, and the next block – to atom 9 (thus forming a chain). For each atom, its closest neighbour in the numbering system is specified (more distant neighbours for angles and dihedral angles are automatically identified and defined by a DFS algorithm);

4. **-o** (output). The resulting output file for glycine with internal coordinates of the converted system, as follows:

```
H
N 1 R2
C 2 R3 1 A3
C 3 R4 2 A4 1 D4
O 4 R5 3 A5 2 D5
H 3 R6 2 A6 1 D6
H 2 R7 3 A7 4 D7
H 3 R8 2 A8 1 D8
O 4 R9 3 A9 2 D9
H 9 R10 4 A10 3 D10
```

```
R2 1.00
R3 1.40
R4 1.54
R5 1.25
R6 1.10
R7 1.00
R8 1.10
R9 1.40
R10 1.00
A3 120.00
A4 109.50
A5 120.00
A6 109.50
A7 120.00
A8 109.50
A9 120.00
A10 109.50
D4 180.00
D5 -120.00
D6 -60.00
D7 0.00
D8 60.00
D9 60.00
D10 180.00
```

The idea of a grid-based input file **generator** is quite simple. The following Python code or its minor modifications were used as a part of almost all conformer generators for silane systems of interest (covered in more detail, in the “Results and Discussion Part II” section):

```

# D4 GROUP
D4 = [ 60., 180., -60. ]
D6 = [ 180., -60., 60. ]
D8 = [ -60., 60., 180. ]
....
# D23 GROUP
D23 = [ 60., 180., -60. ]
....
# D26 GROUP
D26 = [ 60., 180., -60. ]
....
# D27 GROUP
D27 = [ 60., 180., -60. ]
# D29 GROUP
D29 = [ 60., 180., -60. ]
....
# D30 GROUP
D30 = [ 60., 180., -60. ]
....

# CODE -----
for i4 in [ 0, 1, 2 ]:
    for i23 in [ 0, 1, 2 ]:
        for i26 in [ 0, 1, 2 ]:
            for i27 in [ 0, 1, 2 ]:
                for i29 in [ 0, 1, 2 ]:
                    for i30 in [ 0, 1, 2 ]:
                        ....
                        ....

```

Groups of dihedral angles (*e.g.* # D4 GROUP) are formed because the rotation of one atom or a group of atoms connected to some atom A are not explicit and have other atoms and groups connected to A; labelled within the standard system as ‘related dihedral angles’. The 6 nested loops then cycle over all possible conformers. This conformational gridding can be combined with tasks such as protonation, deprotonation, error checking *etc.* (*e.g.* for each conformer generated in the innermost Python language `for` loop, one can calculate all protonation patterns with 0, 1, 2, ..., *n* protons around the conformer.)

Such generators are of great help when it comes to conformational space exploration; for example, 729 input structures for a silane system of interest were created in

milliseconds. These automation tools can also decrease the influence of human error, where computers can generate a billion identical files that will be truly identical and without typographic errors typical of highly repetitive manual input by humans.

Such automated structure input generators were used for many sub-projects ranging from amino acid and peptide structure analyses, protonation thereof, to dental silanes and spider-silk protein structure, through to DNA and RNA systems, among other molecular and materials systems of interest.

#### **4.1.2 Queuing systems**

When all files are generated, selected physical and chemical properties of interest are explored using the modelling package of choice to answer the questions posed; g03 and CPMD for the case studies described in this work. Manually running any of these calculations from a command line (*e.g.* typing in “g03 input\_file\_name.txt &”) is the most basic type of a queuing system in this case. Once the job has started, one needs to monitor its progress, either through checking the output file for signs that indicate the job must be stopped, or simply waiting until the job has completed, so as to commence another one. Although skilled Linux users can log into a computer machine and check calculations very quickly, when it comes to running a thousand calculations even on a small cluster of 10-15 machines, the job becomes very time-consuming. However, there is still the advantage of human-mediated analyses and decisions ‘on the fly’ (in real-time) that come from this method, allowing projects to



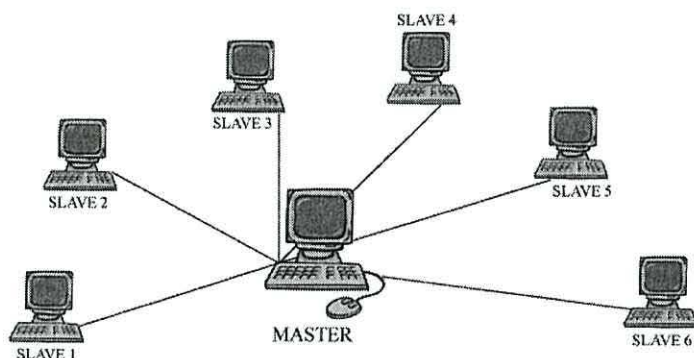
be modified at any sign of error, or more importantly, discovery of novel phenomenon. Automating this latter point is the pan-ultimate goal of AI- and ANN-mediated algorithms.

Another way to run theoretical calculations on a network of computers would be to create an application that would automate the majority of the manual tasks; hence the creation of the **g03\_master** and **g03\_slave** applications. The former (**g03\_master**) is designed to run on a computer that has a list of calculations, while the latter (**g03\_slave**) should run on the computers with Gaussian03 installed (Fig. 4.1).

The protocol is as follows:

- a. Each application has a corresponding *.ini* file which specifies the initialisation parameters. *g03\_master.ini* lists the directories containing input files. *g03\_slave.ini* contain six connection parameters: IP of the master server (e.g. one could write “IP 123.234.43.21” in the *.ini* file), server PORT on the master node, amount of Random Access MEMORY (RAM) that a slave can use on a slave computer, number of CPU cores than can be used, TIMEOUT before each input file request, and life status RUNNING (1 – the slave should keep running, 0 – the slave should stop after finishing its current jobs):

```
IP 123.234.43.21
PORT 50000
MEMORY 4GB
CPU 4
TIMEOUT 60
RUNNING 1
```



**Figure 4.1** Schematic illustration of a network containing six ‘slave’ computers (nodes) connected to one ‘master’ computer (server).

- b. The master node (server) contains a list of jobs and the computational characteristics of each job. Slaves (nodes) can send requests stating “3CPUs and 2GB RAM”, which the master node will try to satisfy by providing either a job for those requirements or a response such as “no files to run” (this is often the case when all jobs require more CPUs than the slave nodes have available);
- c. When a calculation is completed, it is sent back to **g03\_master** for inspection and subsequent storage. Both **g03\_master** and **g03\_slave** are multithreaded applications written in Python programming language, so they can combine tasks such as satisfying three requests at the same time.

The Gaussian queuing system was used throughout this work. Some of the advantages of this system are as follows:

- Optimal matching of computer and network resources to required calculations;
- No time delays between subsequent jobs, and no breaks for cafe / lunch;

- Dynamic queue management (*i.e.* easy removal / addition of calculations while the server is running);
- High availability – the applications are written in Python programming language, so they can be run on almost any UNIX or UNIX-like operating system (*e.g.* LINUX);
- The application is standardised, modular, scalable and extendable. Additional modules can be easily added to **g03\_slave** for checking and directing the running calculations. **g03\_master** is to be extended in order to communicate with the database of pre-computed structures, under continuous development and expansion.<sup>151</sup>

The disadvantages might be summed up in the way that we believe that human-controlled calculations will always be of a higher scientific value, as natural intelligence is so powerful in pattern recognition. On the other hand, the idea behind these computational tools is not to replace human ingenuity, but rather to save time on manual tasks that can be automated without any loss of precision for scientific experiments.

### 4.1.3 Output tabulation and analysis

Output results tabulation, just like calculation execution, is another example of a problem that in practice is usually done by hand, yet can be near-completely automated.



A g03 output file commonly contains blocks of different results and information. For example, a geometry optimisation calculation usually consists of multiple iteration steps. At each step, information such as the total energy, atomic coordinates, distances between all atoms, various minimal energy search parameters and configuration values for algorithms used in the optimisation *etc.*, is written to the output file. As a result, a user can follow what has happened to the system during the geometry optimisation: atomic coordinates at each step can be useful to track the trajectory of the geometry optimisation; energies can be added to this trajectory to track the energetic changes as a function of structural changes *etc.*

**energizer.py**, the application that is a part of the computational toolbox used in this work, is used to quickly extract and plot the energies from a g03 computation, as well as to extract the geometry at any particular step of a geometry optimisation. Many other application packages have the capability to extract this information from g03 output files, but *energizer.py* is different in the way that it can be used without prior installation (*i.e.* it can be used on any machine running a UNIX-like system, or a Windows system with Python interpreter), and it plots the energies directly into the console window (therefore, no special graphics libraries are required – everything is text-based). This is of particular advantage to mobile personal communication devices as well as unreliable internet connections, due to the very compact nature of the data. These tools can be used from these devices to manage or extract results from ongoing computations, from virtually anywhere in the world.

The two tasks that *energizer.py* completes are plotting energies and extracting a desired geometry. An energy plot for a sample system and file names ‘hn-g-oh-



l\_i++\_m2f\_611++dp.log' appears as follows (the line starting with "user@pep:542\$" being the user's command line):

```
user@pep:542$ energizer.py hn-g-oh-l_i++_m2f_611++dp.log
```

```
----- hn-g-oh-l_i++_m2f_611++dp.log:
```

Energies:

1:	x x x x x x x x x x x	-282.877215	
2:	x x x x x x x x x	-282.881970	-0.004755
3:	x x x x x x x	-282.887736	-0.005766
4:	x x x x x x	-282.892368	-0.004632
5:	x x x x x	-282.895851	-0.003483
6:	x x x x	-282.898169	-0.002318
7:	x x x x	-282.900474	-0.002305
8:	x x x	-282.902228	-0.001754
9:	x x x	-282.903077	-0.000849
10:	x x x	-282.903817	-0.000740
11:	x x x	-282.904030	-0.000213
12:	x x	-282.905059	-0.001029
13:	x x	-282.905428	-0.000369
14:	x x	-282.905666	-0.000238
15:	x x	-282.905774	-0.000107
16:	x x	-282.905870	-0.000097
17:	x x	-282.905876	-0.000006
18:	x x	-282.906723	-0.000847
19:	x x	-282.907652	-0.000930
20:	x	-282.909665	-0.002012
21:	x	-282.910362	-0.000697
22:	x	-282.910972	-0.000610
23:	x	-282.911481	-0.000508
24:	x	-282.911623	-0.000142
25:	x	-282.911651	-0.000028*
26:	x	-282.911634	0.000017
27:	x	-282.911627	0.000006
28:	x	-282.911608	0.000019
29:	x	-282.911608	-0.000000
30:	x	-282.911616	-0.000008
31:	x	-282.911625	-0.000009
32:	x	-282.911626	-0.000001
33:	x	-282.911627	-0.000001
34:	x	-282.911627	-0.000000

The first column is the iteration number, the second is a relative graphical representation of the energy of the system, the third is the total energy (in Hartree atomic units), while the fourth is the change in energy (in Hartrees) from the previous iteration. A star is added beside the value in the fourth column, to denote the overall lowest energy.

The extracted geometry of that lowest energy iteration (#25 in the above case, as denoted by the star), for the same file name appears as follows:

```
user@pep:543$ energizer.py hn-g-oh-l_i++_m2f_611++dp.log 25
```

```
----- hn-g-oh-l_i++_m2f_611++dp.log:
```

```
Extracted geometry #25:
```

```
H      1.999382    0.609563    0.771193
N      1.953661   -0.042353   -0.005309
C      0.685200   -0.734580    0.016659
C     -0.554987    0.159252    0.000520
O     -0.526567    1.361136    0.003115
H      0.627645   -1.410639   -0.844186
H      2.022348    0.520155   -0.847132
H      0.631582   -1.363304    0.913434
O     -1.743795   -0.505226   -0.007896
H     -1.574973   -1.454618   -0.020976
```

```
Energies:
```

```
.....
```

Another simple, but very useful tool is **processor.py**, which is a comparatively small Python script written to tabulate the final results of a grid-based conformational space exploration. It searches the output files for the energies of the optimised structures, converts them to energies in  $\text{kJ}\cdot\text{mol}^{-1}$  (or any desired unit of energy), calculates relative energies and outputs conformer / energy values in a format that is required (*i.e.* journal-specific format); other variables can be added, *e.g.* important dihedral angles at any time. For example, a simple version of the script that was used for the glycine protonation studies creates a *.csv* (comma-separated values) file of the following format:

```
code-name,RELATIVE_FREE_ENERGY_KJ
hn-g-oh-l_o-a,10.775050965750
hn-g-oh-l_o-s,5.521425970133
hn-g-oh-l_pa+,30.363904585669
hn-g-oh-l_o+a,10.775050965750
hn-g-oh-l_p+a,10.777676465491
hn-g-oh-l_oaa,16.700803897098
hn-g-oh-l_o-+,5.615943960971
```

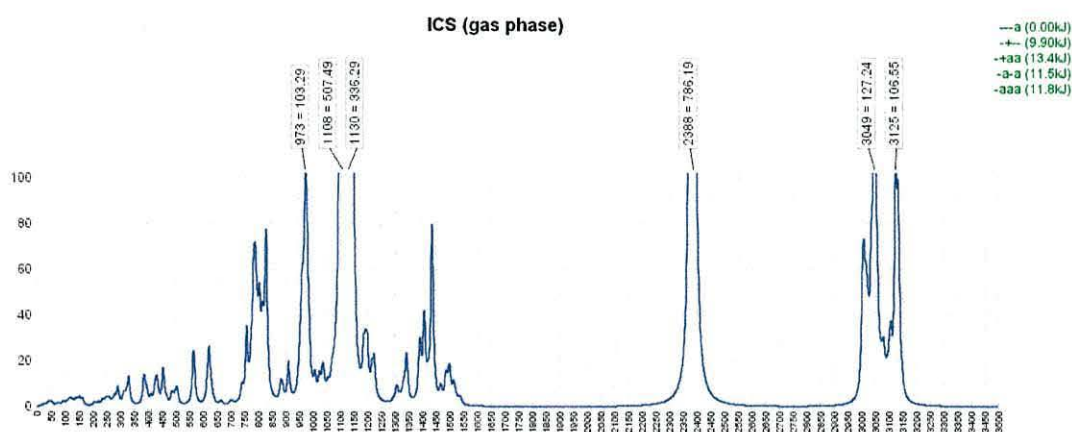
```
hn-g-oh-l_o+-, 5.615943960971
hn-g-oh-l_i++, 22.613429329590
hn-g-oh-l_i--, 22.613429329590
hn-g-oh-l_i-a, 7.167614312002
hn-g-oh-l_ia-, 22.618680329073
hn-g-oh-l_o++, 5.117099008917
hn-g-oh-l_o--, 5.361270485461
hn-g-oh-l_pa-, 30.363904585669
hn-g-oh-l_pas, 44.565232722564
hn-g-oh-l_p-s, 5.639573458793
hn-g-oh-l_i+s, 27.916938820443
hn-g-oh-l_ias, 22.597676330993
hn-g-oh-l_oa+, 30.369155585152
hn-g-oh-l_oas, 46.615748025799
hn-g-oh-l_i-s, 27.916938820443
hn-g-oh-l_p--, 5.505672971535
hn-g-oh-l_iaa, 0.000000000000
hn-g-oh-l_i+a, 7.167614312002
hn-g-oh-l_i-+, 27.914313320702
hn-g-oh-l_p+s, 5.639573458793
hn-g-oh-l_p-a, 10.777676465491
hn-g-oh-l_p+, 5.487294473345
hn-g-oh-l_p+-, 5.487294473345
hn-g-oh-l_ia+, 22.618680329073
hn-g-oh-l_oa-, 30.369155585152
hn-g-oh-l_p++, 5.505672971535
hn-g-oh-l_paa, 22.718449319395
hn-g-oh-l_i+-, 27.914313320702
hn-g-oh-l_o+s, 5.521425970133
```

This file can then be imported into an application like Microsoft Excel and subsequently analysed in the desired fashion.

A similar application (**NiceIR 1.0**, written in Java with a Graphical User Interface Swing) was developed to plot theoretical Infrared intensities, and extract thermodynamic parameters, IR peaks and Cartesian coordinates (xyz) of the geometries from g03 output files. All IR-plots in the appendices were made using **NiceIR 1.0**. An example is presented in Figure 4.2, where the green block on the top right-hand side contains information about conformer classes; in this exemplary case the resultant IR spectrum is averaged over all conformer classes.



At present NiceIR 1.0 peak detection works only for the frequencies  $>20\text{ cm}^{-1}$ ; it is however modular and can be optimised for these ultra low energy vibrations. The IR spectra prediction calculations are based on Lorentzian lineshapes.

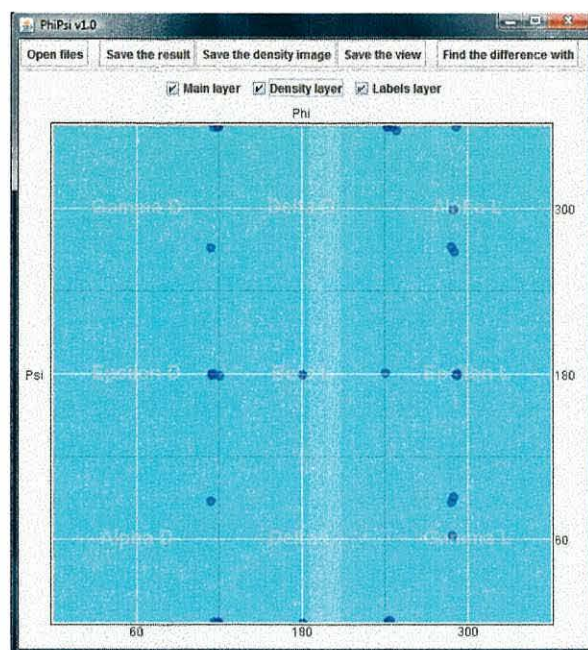


**Figure 4.2** An example IR-plot made by the NiceIR 1.0 application, for the ICS silane system, computed at the B3LYP/6-31G(d,p) level of theory, in gas-phase. The green list in the top-right corner describes various conformational classes of ICS with their corresponding relative energies. The numbers in boxes depict the vibrational frequency ( $\text{cm}^{-1}$ ) and intensity (arbitrary units).

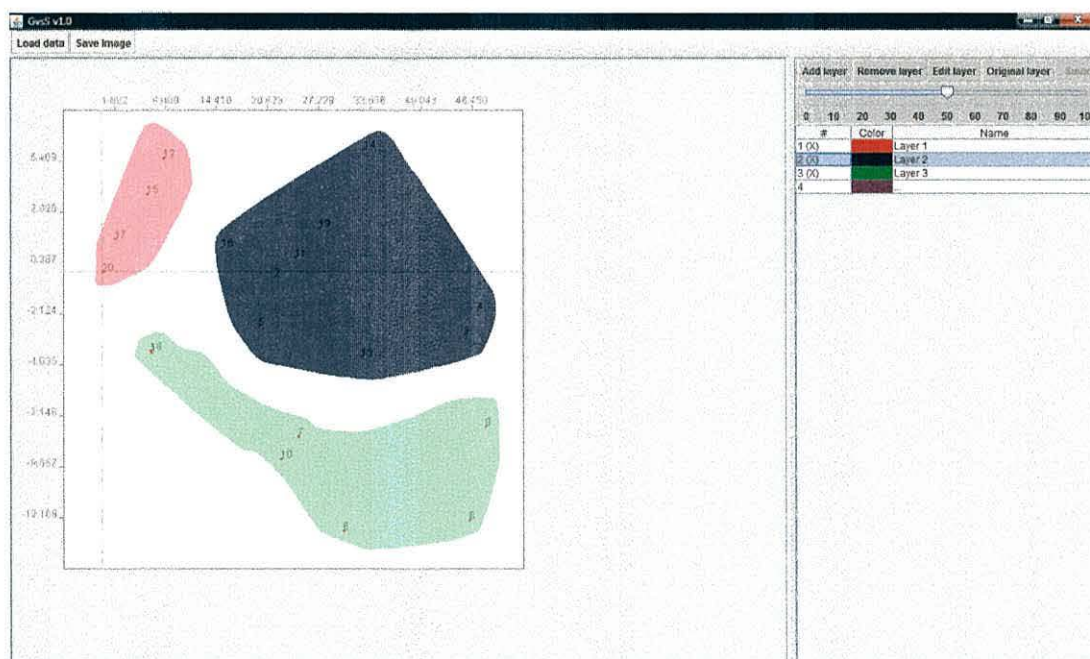
**PhiPsi** (Fig. 4.3) is a simple 2D-graphics tool used to explore the conformational space of peptides. It is similar to a Ramachandran plot,<sup>243</sup> but includes additional capabilities such as the comparison of probabilities of two different Phi/Psi data sets, plotting cumulative probabilities for selected plot areas, among other analyses tasks.

**GvsS** is another 2D-graphics tool (Fig. 4.4) that creates a visual representation of energy plot (*i.e.* Gibbs Energy vs. Entropy) and allows user to manually specify conformer / rotamer classes.





**Figure 4.3** The user interface of the PhiPsi 1.0 program package. The blue dots represent various conformations of two principal degrees of freedom (dihedral angles Phi/Psi) of glycine amino acid.



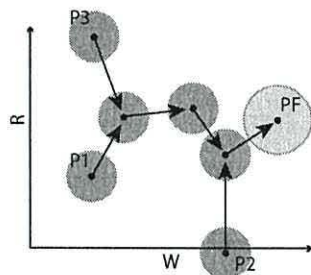
**Figure 4.4** The user interface of the GvsS 1.0 program package. The clear separation of different conformers of glycine using Gibbs Energy and Entropy allows for analytical definition of conformational classes and generalisation / explanation of inter- and intramolecular forces which define various molecular geometries.

#### 4.1.4 Calculation Management

Another application that is important for conformational mapping studies is **optimiser\_simulator.py**, which searches through geometry optimisation calculations for ‘pathways’ (due to geometric changes during the optimisation) which connect different conformers, subsequently building a graph of such pathways (Fig. 4.5). This graph can be used in molecular dynamics simulations to minimise the number of *ab initio* calculations (*i.e.* the graph gives information about which path on the PEHS is preferred and where it leads). For example, using Figure 4.5 we can say that after characterising the geometry optimisation path for P1, inclusion of this information for other starting conformers (P2 and P3) allows for subsequently optimising them more efficiently.

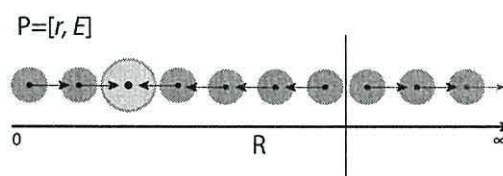
The basic idea behind this approach is to reduce the number of degrees of freedom to explore for a molecule with  $N$  atoms from  $(3N - 6)$  to as few as possible, say  $M$ , and then represent each conformer as a point in an  $M$ -dimensional space (those  $M$  parameters are usually the most chemically influential degrees of freedom). A set of such points with directed inter-connections, calculated from steps of an *ab initio* optimisation calculation, constitutes a graph structure; see “Introduction” (p. 41) for more information on graphs. Sometimes geometric structures are very close to one another on their respective PEHSs, but differ slightly in one or more (non-energetically influential) degrees of freedom. Treating such structures as ‘different points’ would be a waste of computational resources; therefore buffer zones were used to define catchment regions – distances in multidimensional space at which

slightly different structures are still considered the same (these are depicted as filled gray circles on Figures 4.5 and 4.6).



**Figure 4.5** An example of a two-dimensional slice (R, W) of a hyperspace graph. P1, P2 and P3 are the starting conformers of a grid-based conformational search, whereas PF is the lowest energy conformer that all 3 starting conformers optimise to.

One of the simplest examples of using a database of such graphs in order towards increasing the efficiency of molecular geometry optimisations and dynamics would be a theoretical calculation of a molecular system with one degree of freedom (*e.g.* a diatomic molecule). Both bond distance and bond breaking point would be easily identifiable (Fig. 4.6). Moreover, an intelligent artificial neural network (ANN) would be able to browse such a database and notice dependencies that would optimise the database structure and space-usage. For example, it would save only two points on Figure 4.4 – the bond distance point (lighter circle), and the one at the bond-breaking distance (vertical line).



**Figure 4.6** An example of a molecular graph for a system with one degree of freedom (distance between two atoms – R). Bond breaking is denoted by the vertical line, with the equilibrium bond distance at the lighter circle.



This project is still under development, but some preliminary results show high rates of performance increase for molecular geometry optimisation (Table 4.1).

**Table 4.1** Calculation performance increase (as %, relative to standard calculation), for three different molecular systems using the proposed check-point graph computation model

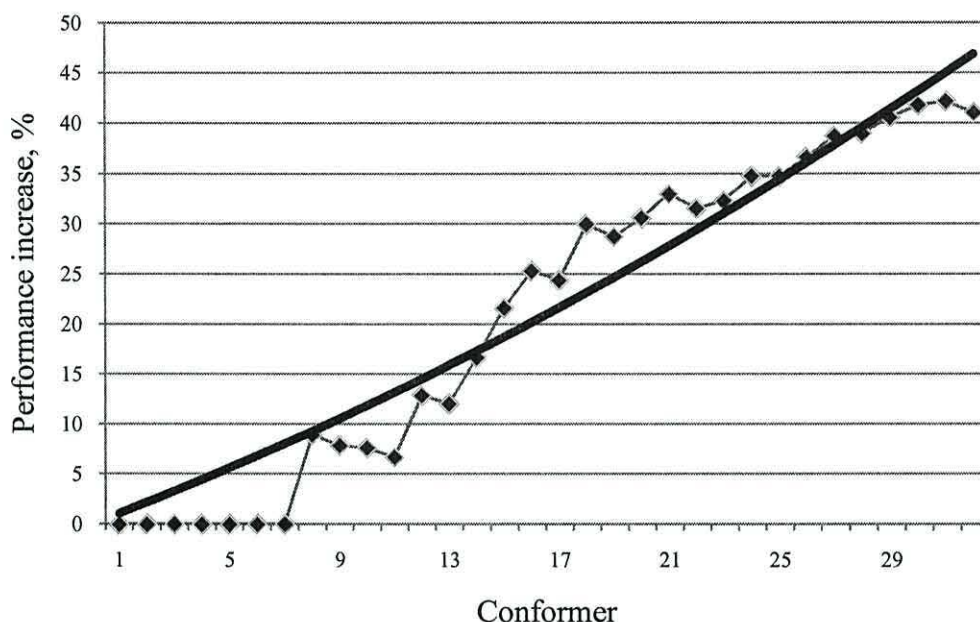
	Number of degrees of freedom explored	Number of conformations analysed	Dihedral angle buffer zone size (in degrees)			
			5	10	15	20
<b>glycine</b> (10 atoms)	3	32	29.25%	34.1%	38.61%	41.04%
<b>silane-1</b> (35 atoms)	6	187	9.71%	12.04%	14.9%	16.24%
<b>silane-2</b> (37 atoms)	4	51	24.92%	26.49%	26.82%	27.59%

Table 4.1 highlights that the larger the number of degrees of freedom the lower the performance increase. This is logical, as the larger number of degrees of freedom relates to a more highly dimensional PEHS, translating to larger set of topologically possible conformations. Table 4.1 also shows that the dihedral angle buffer zone size (BZS) influences the number of calculations for a particular molecular system (the larger the BZS, the fewer calculations are needed to explore the conformational space). But choosing too large a BZS can influence calculation accuracy / precision, since too many distant / different conformers will be treated as one conformer.

It is also logical that the more structures one analyses (*i.e.* the more structures we add to the graph of a molecular system) the higher the performance increase with each subsequently added structure, since most, if not all, of its moves will already be saved in the database. This is shown on the chart below:



Performance increase of Glycine calculations with the number of conformers analysed



As our calculations show, with BZS=20 for dihedral angles, 4 out of 32 glycine conformer optimisation paths were completely computed by calculating other conformers, while 17 were partially calculated. Although larger structures such as the STYRX silane (see “Results and Discussion Part II” section) had only one conformer completely calculated, it is also very time-saving, as one STYRX calculation can take from 15 to 120 hours on modest computational resources; regardless any time-saving is advantageous.

A new molecular dynamics simulation system could work in the following way:

1. Construct the internal coordinates of structure  $X$ ;
2. Check the database for any records of  $X$ . If there is a structure – obtain its respective pathway graph ( $G$ ), otherwise create a new graph ( $G'$ );
3. Create a check-point  $S$  (file containing the history of the optimisation) from the current conformation of  $X$ . Check if there are any points in  $G$  near that

point. If there is a check-point already – add a path to it from a previous check-point file (if available), use a depth-first search (DFS) algorithm to find a path in  $G$  to the closest optimised structure, return the path and the resulting structure (all requiring a few milliseconds on an ordinary computer) and proceed to step 5. If there is no such point, add  $S$  and a connection from a previous check-point (if available), to  $G$ ;

4. Perform one *ab initio* calculation iteration and go to step 3;
5. Process results.

A check-point ‘tuple’ (a point in a multidimensional space) can also be used as an input to an artificial neural network (ANN) learning algorithm. Thus, one could also train a neural network by adding points to a graph of structures to the database.

## 4.2 Explicit particle solvation algorithm

Solvation of a solute is an essential component of a variety of biological and chemical processes including equilibria and kinetics for metabolic reactions, structural changes and intermolecular association.<sup>244-249</sup> The ‘solvent effect’ is especially important in drug discovery, as it is often influential to the bio-affinity and bio-activity of a drug at the site of action.<sup>250-253</sup>

In a molecular and materials modelling calculations,<sup>254-257</sup> one of the two primary ways of accounting for the influence of solvent can be used:

- 1) *Explicit*. This solvation model involves using explicit solvent particles for the construction of consecutive layers of solvent particles around a solute.<sup>49, 258, 259</sup>
- 2) *Implicit* solvation models are computationally much simpler, accounting for the solvent using a continuum with a pre-determined (yet erroneously static!) bulk dielectric constant. Within this system, quantum mechanical approaches may be employed for the ENP (**E**lectronic, **N**uclear, **P**olarisation) effects. For example, one of the earliest approaches (known as the Kirkwood-Onsager model<sup>260-262</sup>) uses a Taylor series<sup>263</sup> to represent the classical multi-polar expansion of the solvent’s electronic structure. Another example could be the more recent continuum approaches that use a generalised Born formalism<sup>264, 265</sup> for the interaction of atomic partial charges with a surrounding dielectric. G03 implements a variety of Self-Consistent Reaction Field (SCRF) implicit solvation models, all of which model the solvent as a continuum of dielectric

constant  $\epsilon$  and include the following models: Onsager,<sup>260</sup> Polarized Continuum Model (PCM),<sup>266, 267</sup> Isodensity PCM (IPCM), Self-Consistent IPCM (SCIPCM)<sup>268</sup> among others.

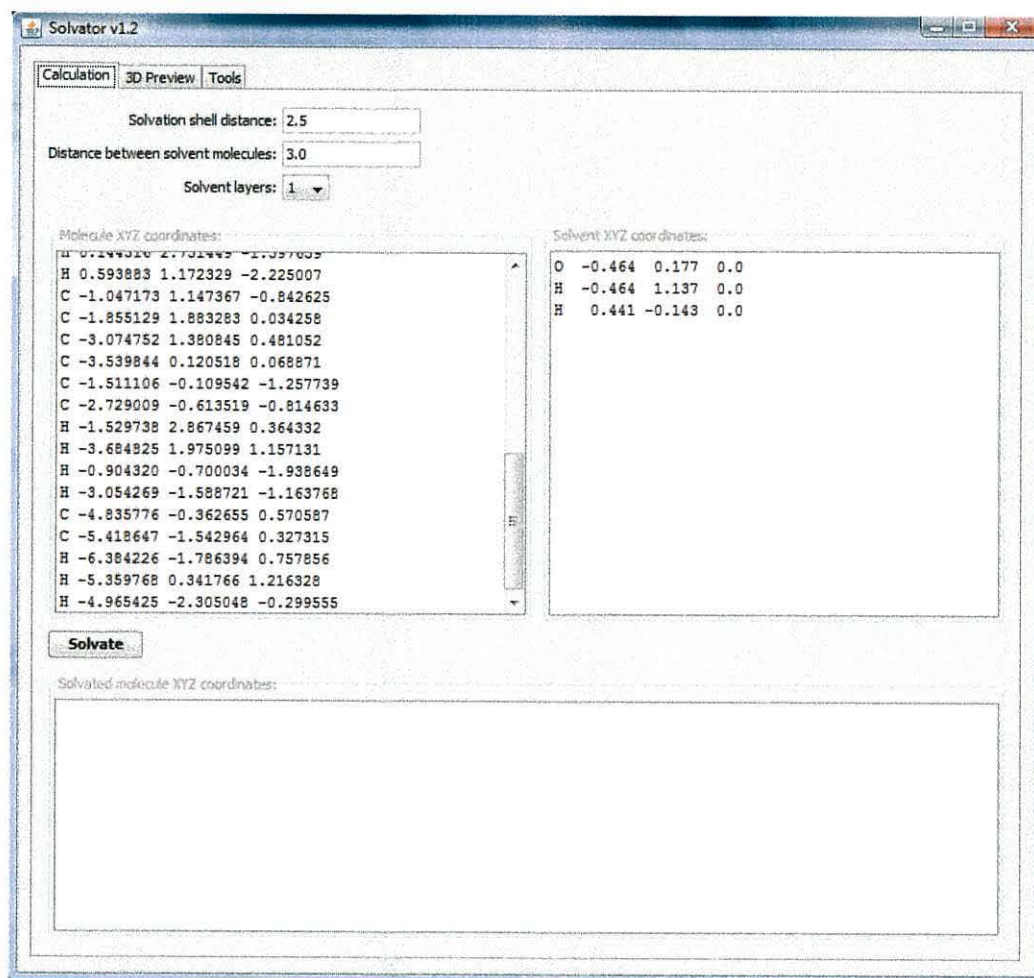
Provided the model accurately accounts for the influence of the solvent in question (this may translate to a sufficiently large number of solvent molecules, or more often proper placement of a relatively small number of solvating particles), the explicit solvent method is undeniably more accurate than the implicit one. The only serious limitation of explicit solvent is the resultant complexity and / or size of the system constructed. As previously mentioned (“Introduction” section), the number of elementary operations in a molecular calculation increases exponentially with the number of atoms in the system. There are three ways of overcoming this limitation:

- (1) increase in computational power;
- (2) development of new algorithms making the calculations more efficient or faster;
- (3) use of the optimal arrangement or amount of solvent.

This work presents two approaches to solution (3). One involves a purely geometric means to fitting a requisite number of solvent molecules around a solute molecule, using Bader atomic radii.<sup>215</sup> Another, (currently under development) uses a similar geometric approach, but is more advanced as it performs ‘on-the-fly’ calculations of electron density of solvent molecules and the solute, adjusting to the influence imposed by each additional solvent particle introduced.



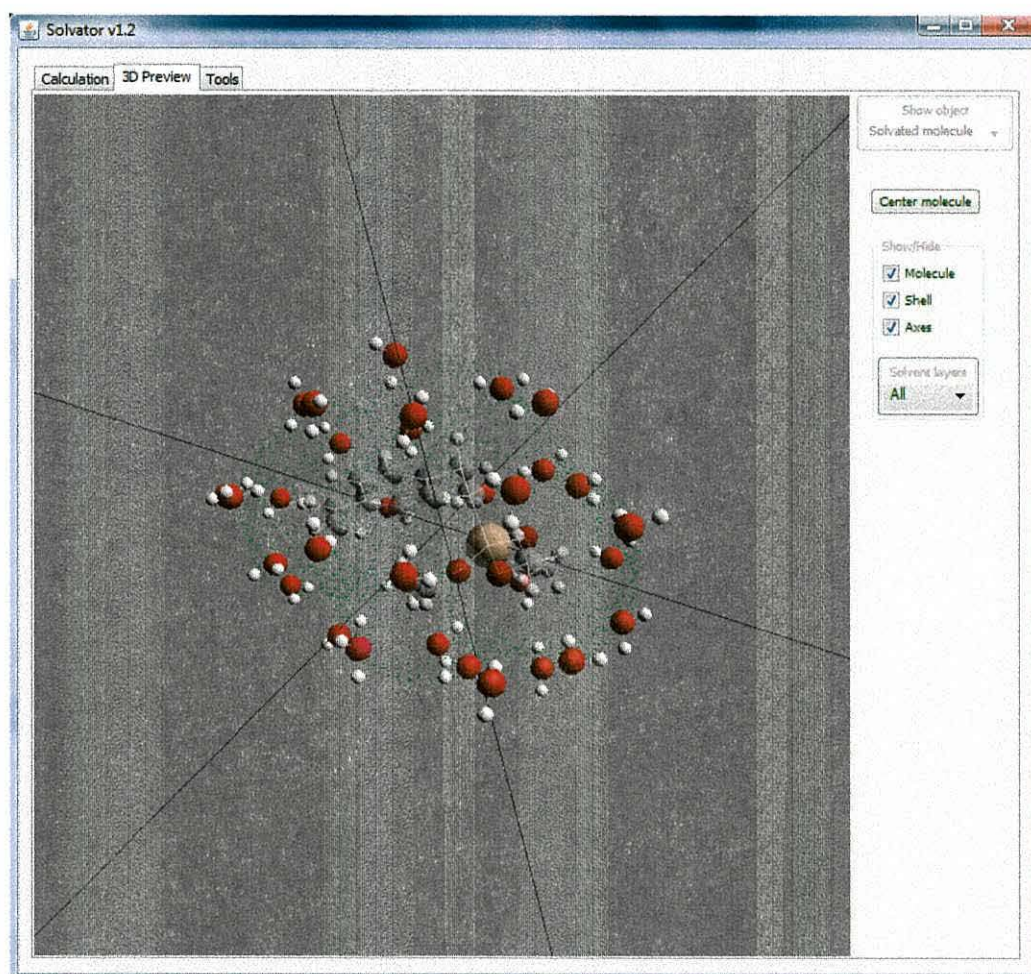
An application was developed, named **Solvator 1.2** that employs a Swing Graphical User Interface (GUI, see Fig. 4.7), to link a user and a very efficient molecule solvation algorithm. The program also uses the Java3D library to present the resultant solvated molecule in 3D (Fig. 4.8).



**Figure 4.7** The “Calculation parameters” window of the Solvator 1.2 program package.

There are two versions of the solvation algorithm: one uses a purely geometric basis for its solvent shell calculations (*i.e.* solvation surfaces for the solute and solvent molecules are generated and then used in geometric calculations to fit the optimal number of solvent particles around the system, so as to cover its entire surface); the second algorithm (currently under development) is based on charge densities more

complex, using an extensively large set of basis functions to represent negative charge distributions around the solute and solvent molecules (thus creating a multilayer solvation surface with various density values defining each layer). A subsequent and relatively rapid (low approximation level) geometric fitting of solvent molecules around the solute is then carried-out.



**Figure 4.8** The “Preview” window of the Solvator 1.2 program package.

It uses a set of common file processing algorithms to help with viewing / editing / saving files. The application can also perform a multiple-layer solvation, using a close-to-optimal amount of solvent molecules to completely solvate the solute. As



solvation surface calculations are exclusive from the solvation process, the former is near-instantaneous in solvating a solute consisting of ~50 atoms.

One of the greatest advantages of the Solvator 1.2 algorithm is that **any solvent** can be used, even systems that are not traditionally considered as solvents can make use of the tool developed. For example, protons ( $H^+$ ) were used as explicit solvent particles in the protonation of amino acids, where a glycine molecule was surrounded by these positively charged particles with subsequent computation of Car-Parrinello Molecular Dynamics trajectories. Another example are the silane systems investigated within this work (see “Results and Discussion Part II” section), where four different solvents were used: ethanol, water,  $CCl_4$ , and a 95% : 5% ethanol-water mixture. Solvent particles can also be added in multiple layers around the solute, or solute + explicit particles. Another advantage of the Solvator 1.2 algorithm is its calculation efficiency (Table 4.2). Preliminary calculations were performed on a very modest and outdated (at time of testing) 32-bit Windows Vista(TM) Service Pack 1 operating system, Intel(R) Core(TM)2 Duo T5800 2.2GHz CPU, 2GB of RAM. Average distance between water molecules was 3 Angstroms ( $\text{\AA}$ ). Average solvent-solute distances were 2.5  $\text{\AA}$ . Each average value in the table is based on 10 trials.

**Table 4.2** Average calculation speed (in seconds) of the Solvator 1.2 explicit particle solvation algorithm in placing explicit solvent molecules around the solute, for various molecular systems.

Solvent layers ( $H_2O$ )	<b>glycine</b> (10 atoms)	<b>silane monomer</b> (37 atoms)	<b>silane dimer</b> (59 atoms)	<b>silane trimer</b> (81 atoms)
1	0.03	0.11	0.16	0.2
2	0.15	0.46	0.86	1.2
3	0.72	3.1	4.6	7.34
4	3.4	13.0	24.18	50.06

Finally, from the successes generated thus far with the Solvator 1.2 algorithm, future work includes building on a foundation of an internal set of results from *ab initio* characterisations of the solute molecule. From these high-level determinations, electronic charge density distributions will be used to more effectively place explicit solvent particles in an iterative manner, keeping the restraint that the average influence of all solvent molecules on the solute must be approximately the same (*i.e.* the charge densities of solvent in layer  $N$  must be ‘equidistant’ to the charge density of the solvated system).



## 5. RESULTS AND DISCUSSION (Part II)

---

The previous “Methods” and “Results and Discussion Part I” sections detailed the methods, both developed ‘in-house’ and by other scientific groups, used in this work. Some preliminary results were also presented, such as those emerging from the *optimiser\_simulator.py* computational tool that tracks the dynamics of various conformers of a molecular system and aids in predicting its behaviour during molecular dynamics and geometry optimisation calculations. These methods and tools were used to increase the efficiency in the modelling of selected short bioactive peptides in addition to characterising the influence of protonation thereof, towards determining their respective protonation limits.

In this second part of the results and discussion section (Part II), the results emerging from experimental and theoretical studies on silane coupling agents (biocompatible adhesives) and selected amino acids are presented.

## 5.1 Rational Design of Comb-Like Monolayers of Biocompatible Silanes I. Dynamic Binding Mechanisms and Chain Pre-Organisation Prior to Surface Grafting

### 5.1.1 A Background/Introduction

Silane coupling agents, organofunctional Si-esters, are synthetic, hybrid organic-inorganic monomeric compounds with at least one direct Si-C bond. They are widely used to form monolayers,<sup>269</sup> O-LEDs,<sup>270</sup> surfaces<sup>271</sup> and coatings.<sup>272, 273</sup> Their inherent biocompatibility has led to use in numerous pharmaceutical applications in the fight against cancer,<sup>274</sup> Alzheimer's and Parkinson's diseases.<sup>275</sup> Silanes also commonly serve as building blocks in polymers<sup>276-278</sup> and biomaterials.<sup>279</sup> Of particular interest and the focus of this work is their efficacy in binding dissimilar materials.<sup>280</sup> This makes them ideal as bioadhesives, with applications in bone implantation<sup>281</sup> and dentistry.<sup>282</sup> Herein we show how rational design of monomers may promote pre-organisation of silanes, prior to surface-binding, towards idealised comb-like monolayers.

Silanes have the general molecular structure  $X_3\text{Si}(\text{CH}_2)_n\text{Y}$ , where  $n = 0 - 3$ ,  $X$  is a hydrolysable group on silicon, and  $Y$  is an organo-functional group selected for desired characteristics or reactivity. Adhesion is promoted by condensation reactions of the silanols deriving from hydrolysis of the original alkoxy groups ( $X$ -groups).<sup>283</sup> Generally applied from dilute aqueous solutions (*i.e.* alcohol-water),<sup>284</sup> silanes effect or promote resin-composite veneering onto mineral

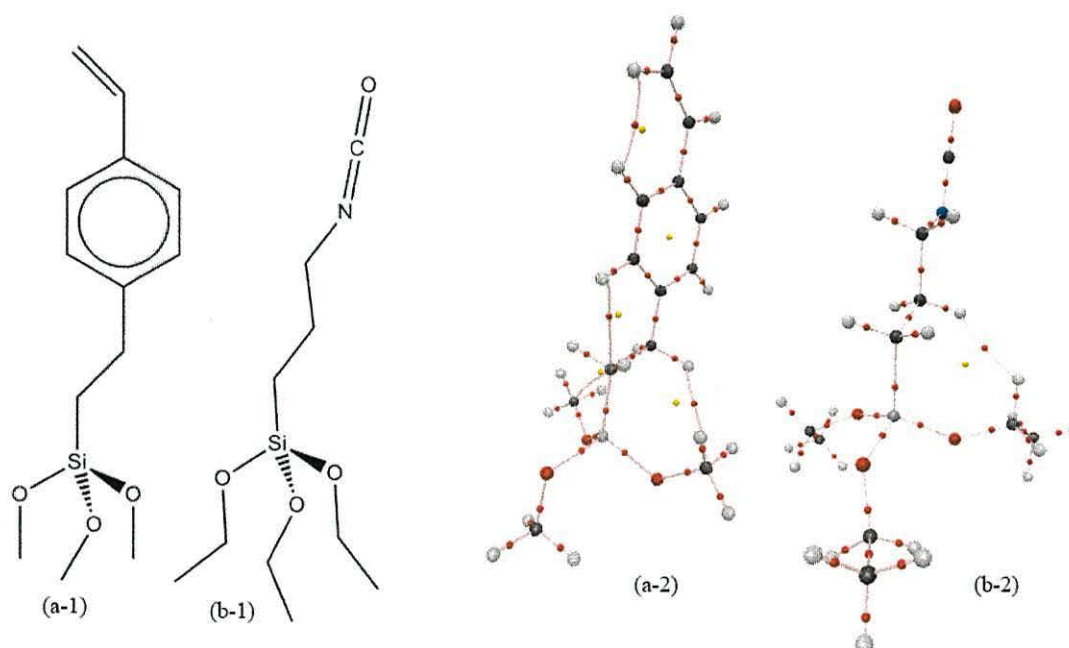
surfaces, including hydroxyapatite, the structural prototype of tooth and bone, silica-coated metals or ceramics, inorganic fillers, or etched glassy porcelain.<sup>282, 284-287</sup>

Surface binding (siloxane bonding) is mediated by condensation reactions of silanol groups derived from the original alkoxygroups post-hydrolysis.<sup>273, 286, 288, 289</sup> The net effect is ‘grafting’ of silane onto the enamel hydroxyapatite of the tooth surface. The desired effect is complete ‘comb-like’ surface coverage, with all silanes standing straight-up on the surface, with the terminal ‘Y-group’ extended away from the surface. The ideal case has all (three) silanol groups bound to surface hydroxyl groups, with no silane chain entangled with any of its neighbours, while retaining maximum reactivity of the organo-functional group. However, the reality is that complete surface-coverage is not always easily attained, complicated by Si-O-Si inter-chain binding;<sup>288</sup> effectively ‘locking’ the silanes into a tangled web and burying the terminal groups. Even with these two aspects optimised, chain ‘floppiness’ and self-binding introduces an additional challenge to grafting an optimal and functional adhesive monolayer.

Clinically silane-based bonding is satisfactory, but the hydrolytic stability and longevity of bonding is a concern.<sup>287, 290</sup> This has been explained by the dynamic balance of breaking and rearrangement of the bonds. Many assumptions in silane-aided bonding are based on ideal monolayer arrangement. The actual siloxane layer has been described as a complex 3D layer, dependent on setting-reaction temperature, silane concentration, pH and the alkoxy Y-group type.<sup>286</sup> Optimising monolayer deposition is of utmost importance in dental materials science, as silanes are used in all resin-composites (cementation, filling, veneering).

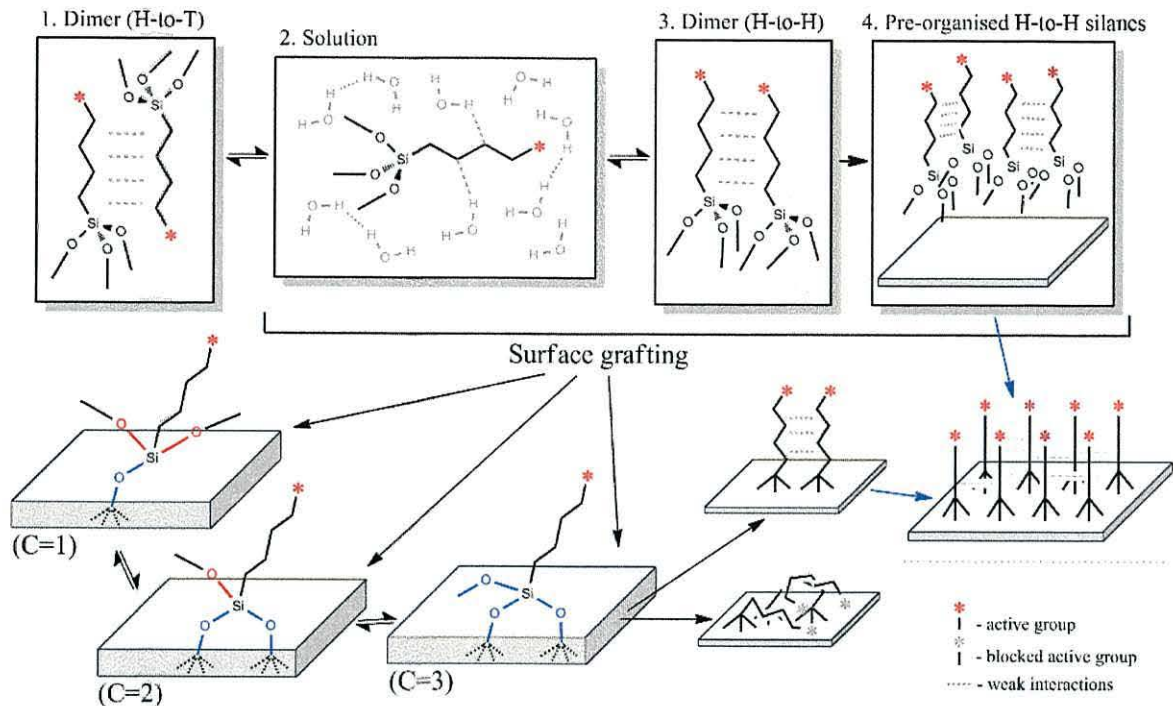


Optimal conditions can be attained through control of the initial layer binding, specifically how the silanes arrange their silanol groups with respect to the substrate surface and subsequently react. Information on how silanes behave in solution prior-to grafting and afterwards as a monolayer is still sparse in the literature. We therefore undertook an in-depth exploratory characterisation of the spatial arrangement of 3-styrylethyltrimethoxysilane (STYRX) and 3-isocyanatopropyltriethoxysilane (ICS) (Fig. 5.1) as both have shown promising adhesion promotion capacity. Their molecular structure studies are not well reported and the information necessary to understand their behaviour might be understood with the help of theoretical methodology (conformational studies and molecular dynamics under various solvent conditions).



**Figure 5.1** Molecular structures of the 3-styrylethyltrimethoxysilane (STYRX, a-1) and 3-isocyanatopropyltriethoxysilane (ICS, b-1) silanes characterised in this work, and their AIM analyses (a-2, b-2).



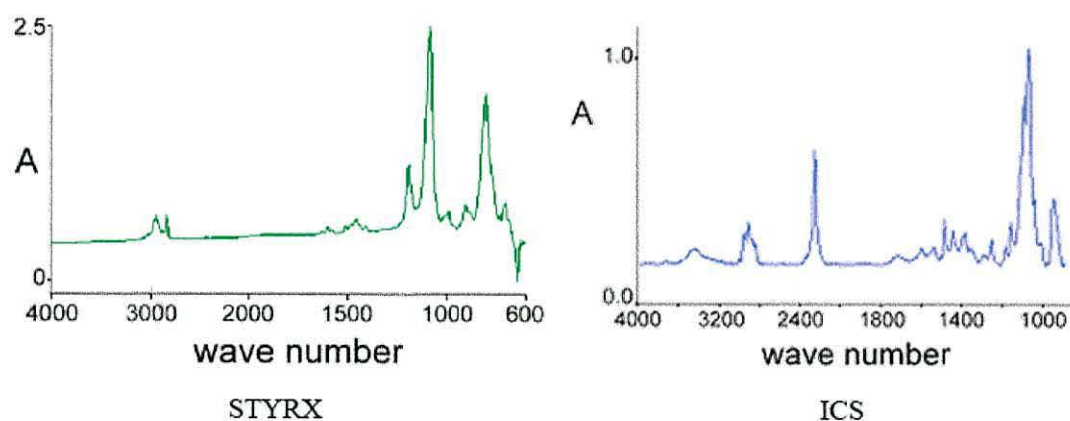


**Figure 5.2** Simplified illustration of the idealised process of grafting silanes to surfaces (*i.e.* tooth hydroxyl apatite). The stepwise process is controlled by thermodynamic and entropic forces derived from dimer  $\leftrightarrow$  monomer equilibrium in solution (panels 1-3), differing head-to-tail (H-to-T – panel 1) and head-to-head (H-to-H – panel 3) dimer arrangements, prior to surface binding during setting reaction (bottom). The ideal case has three silanol groups bound to the surface (C=3), although high entropic ‘costs’ and chain flexibility may result in lower saturation (C=1, C=2). Chain flexibility also introduces problems in the saturated case (C=3), with the bulk of the silane falling onto the surface and potentially blocking binding sites, effectively blocking and deactivating the terminal organofunctional groups to polymerisation (bottom right). Rational design and optimisation of setting reaction conditions may lead to pre-organisation of silanes prior to surface grafting (panel 4, top right); albeit at high entropic contributions to the free-energy.

The goal is to have a material that ‘stands up’, or exclusively populates (and remains in) the extended (all-*anti*) conformation (Fig. 5.2, bottom-right). This would correspond to a deep, steep ‘well’ on the material’s respective potential energy hyper surface. This is the ideal, and rational design initiates from this

goal, building on a foundation of reductionist structural and energetic resolution. However, solvent conditions play an important role in all aspects of surface-binding, molecular flexibility and silane reactivity (in subsequent polymerisation). The aqueous conditions of the oral cavity would only further complicate things, not to mention saliva chemistry that would presently render this ideal design near-impossible.

At present, FTIR spectroscopy provides a convenient method for monitoring silane dynamics, especially during hydrolysis; Figure 5.3 displays spectra of the non-activated silanes determined in conjunction with this work.



**Figure 5.3** FTIR spectra in  $\text{CCl}_4$ , of the 3-isocyanato-propyl-triethoxy silane (STYRX) and 3-styryl-ethyl-trimethoxy silane (ICS) systems characterised in conjunction with this work.

## 5.1.2 Methods

**5.1.2.1. Silane Primer Preparation.** All silane primers were prepared as 1.0 vol% primers in a standard solution of 95.0 vol% ethanol, and de-ionized water (milli-Q water) which was first adjusted at 4.5 pH with 1M acetic acid and then allowed to stabilise for 24 hours prior to use. The silane primers were activated (allowed to hydrolyze) for 1 hour at room temperature before bonding testing.

**5.1.2.2. Infrared Spectroscopy.** The silane monomer hydrolysis was observed analytically up to 60 minutes using a Reflectance-Absorbance Fourier Transform Infrared (RA-FTIR) Perkin Elmer Spectrum One spectrometer (Perkin-Elmer, Beaconsfield, UK), which detects different molecular bending, vibration, wagging and rocking of the functional groups. The surface analysis of a silane primer film layer was conducted throughout the 600-3800  $\text{cm}^{-1}$  spectral range with a specular reflectance monolayer / grazing angle accessory in which the spread silane primer film was on a cleaned, planar, inert Ge crystal.

**5.1.2.3. DFT Experiments.** Theoretical investigations were conducted through STYRX and ICS model-construction and subsequent characterisation using the Becke-3-Lee-Yang-Parr (B3LYP) Density Functional Theory (DFT) method, employing the 6-31G(d,p) Pople basis set. All computations were carried out using Gaussian03. To ensure quantitative characterisation of all intra-molecular interactions, AIM analyses were conducted on the wavefunctions generated from the B3LYP/6-31G(d,p) geometry-optimised structures.



**5.1.2.4. Car Parrinello Molecular Dynamics (CPMD) Experiments.** CPMD simulation of the STYRX and ICS structures, initially in their of ‘all-*anti*’ conformations, were performed in two solvents, ethanol and CCl<sub>4</sub>, using the explicit particle solvation algorithm, with 3 solvent molecules per system. The timestep was equal to 4 a.u. ( $\approx 0.1$  femtoseconds) with a trajectory generated for 10,000 steps, with a total trajectory time  $\approx 1$  picosecond. CPMD DFT parameters: FUNCTIONAL localised-density approximation (LDA; the exchange-correlation functional), NEWCODE. Temperature = 300.15 K. CPMD SYSTEM parameters: ANGSTROM, SYMMETRY = 1, CELL = 20.0 1.0 1.0 0.0 0.0 0.0, CUTOFF = 60.0.

**5.1.2.5. Atoms-In-Molecules Wavefunction Analyses.** Bader’s Atoms-In-Molecules analyses<sup>215</sup> were performed using the AIM2000 program package,<sup>232</sup> using default values. The topological properties of the electron density distribution  $\rho(r)$  of a molecule are based on the gradient vector field of the electron density  $\nabla\rho(r)$ , and on the Laplacian of the electron density  $\nabla^2\rho(r)$ , where  $r$  is the positional vector in three-dimensional (3D) space. In view of the AIM approach, critical points (CPs) of rank 3 were identified in the electron densities, which include bond, ring and cage critical points (BCPs, RCPs and CCPs, respectively). The pairs of gradient paths that originate at a BCP and terminate at neighbouring nuclei define a line through which  $\rho(r)$  is a maximum with respect to any lateral displacement. Relevant bond and interaction strengths are directly comparable through their respective  $\rho_b$ -values, defined as the number of electrons,  $N_e$ , per spherical Bohr-volume,  $V_B$ ,  $\rho_b = N_e V_B^{-3}$ . The  $\rho_b$ -values are a comparable quantitative measure for identical pairs of atoms. A region of space



$\Omega$  is bound by a surface satisfying the condition of zero-flux in the gradient vector field of the charge density, with  $\nabla\rho(r)\cdot n(r)=0$  being the normal to surface unit vector. The region of space  $\Omega$  is employed for integrating the electron population and energy of atom in a molecule,  $N(\Omega)=\int_{\Omega}\rho(r)d\tau$  and  $E_e(\Omega)=\int_{\Omega}E_e(r)d\tau$ .

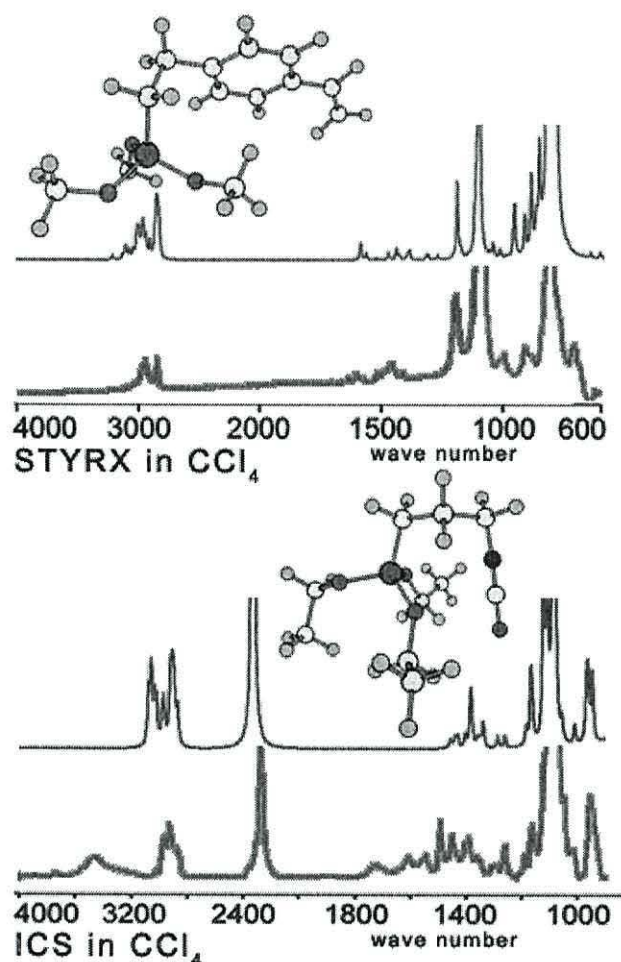
### 5.1.3 Results and Discussion

**5.1.3.1. Dimerisation and solvation.** For ICS, the dimerised state (diethoxy-silane---O---diethoxy-silane) is more stable than monomeric ICS ( $\Delta G_{298} \approx 15.1$  kJ·mol<sup>-1</sup>). The formation of silanols (solvated silanes) proceeds with the overall change in energy of 80.9 kJ·mol<sup>-1</sup> (one monomer forming one silanol) and 126.2 kJ·mol<sup>-1</sup> (one dimer forming two silanols). Blum *et al.*<sup>288</sup> says that “the first methoxy groups hydrolyse most slowly”. In case of ethoxy groups, the energy change is as follows: first ethoxy group hydrolyses with the overall energy change of 21.9 kJ·mol<sup>-1</sup>, the second one with 34.7 kJ·mol<sup>-1</sup>, and the third one 24.3 kJ·mol<sup>-1</sup>. Since the first one proceeds with the smallest change in energy, it must also hydrolyse most quickly. The theoretical ICS hydrolysis calculation clearly shows much lower peaks near the 3000-3100 cm<sup>-1</sup> region after solvation (the region corresponds to C–H bond stretches), since much fewer C–H bonds are present. Also, the IR after hydrolysis shows another set of peaks around the 3800-3900 cm<sup>-1</sup> region, which corresponds to O–H stretches in silanols. Another

newly-formed group of peaks is around 400-500  $\text{cm}^{-1}$ , which are due to the “H–O–Si–C” dihedral angle changes after hydrolysis.

**5.1.3.2. Gas-phase and solvated conformers.** Both gas-phase and implicitly solvated (in  $\text{H}_2\text{O}$  and  $\text{CCl}_4$ ) STYRX silanes converged to one major and two minor conformational classes with small energy differences (all less than 10  $\text{kJ}\cdot\text{mol}^{-1}$ ). The two minor classes under all conditions had a very small energy difference (less than 2.5  $\text{kJ}\cdot\text{mol}^{-1}$ ) and the structures were not stable (jumped between the classes). All ICS conformers converged to five major conformational classes with energy differences ranging from 0.4 to 13.4  $\text{kJ}\cdot\text{mol}^{-1}$ . The theoretical IR spectrum of ICS shows the characteristic peak around 2350-2400 $\text{cm}^{-1}$ , which corresponds to the N=C bond stretch of the organic functional part – the isocyanate group “–N=C=O”.

**5.1.3.3.  $\text{CCl}_4$  Solvated conformers.** Both  $\text{CCl}_4$ - and water-solvated ICS silanes have confirmed that the  $\sim 3450 \text{ cm}^{-1}$  peak observed in experimental results during the solvation process does not correspond to a structural parameter of silanes, but rather corresponds to bond stretches in water molecules. Theoretical calculations of  $\text{CCl}_4$ -solvated conformers show little difference in IR spectra.

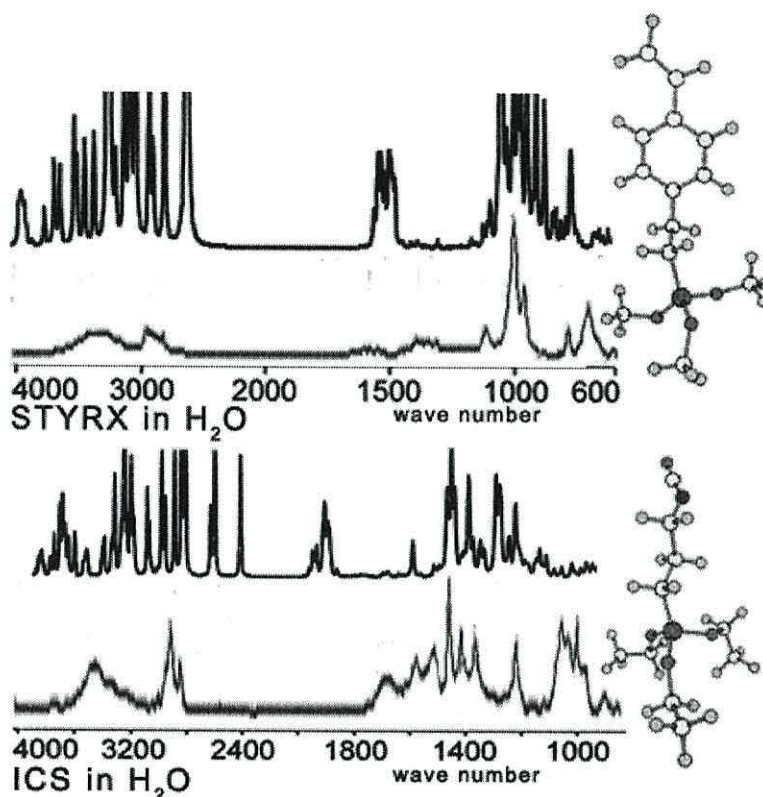


**Figure 5.4** Comparison of computed (top on each part) and experimental (bottom on each part) infrared spectra for STYRX and ICS. The theoretical match was attained using explicit  $\text{CCl}_4$  solvent particles. See peaks descriptions in the text.

**5.1.3.4.  $\text{H}_2\text{O}$  Solvated conformers.** Water-solvated ICS silanes have shown that the experimental peak at  $3451\text{ cm}^{-1}$  has corresponding theoretical peaks in the range around  $3420\text{-}3490\text{ cm}^{-1}$  (water molecules O-H stretches). In water solvated STYRX the experimental peak at  $3352\text{ cm}^{-1}$  corresponds to the theoretically predicted peaks in the range  $3328\text{-}3388\text{ cm}^{-1}$ , which are present in both  $\text{CCl}_4$  and  $\text{H}_2\text{O}$  explicitly solvated structures, but **absent** in implicitly solvated and gas-phase silanes. These peaks correspond to silane structural C-H stretches, clearly



influenced by H<sub>2</sub>O solvent particles. The experimental peaks around ~2950 cm<sup>-1</sup> in both ICS and STYRX correspond to water molecule O-H stretches. In water-solvated ICS silanes, the peaks corresponding to most C-H stretches (see 5.1.3.1) are shifted towards the 3200-3290 cm<sup>-1</sup> region.

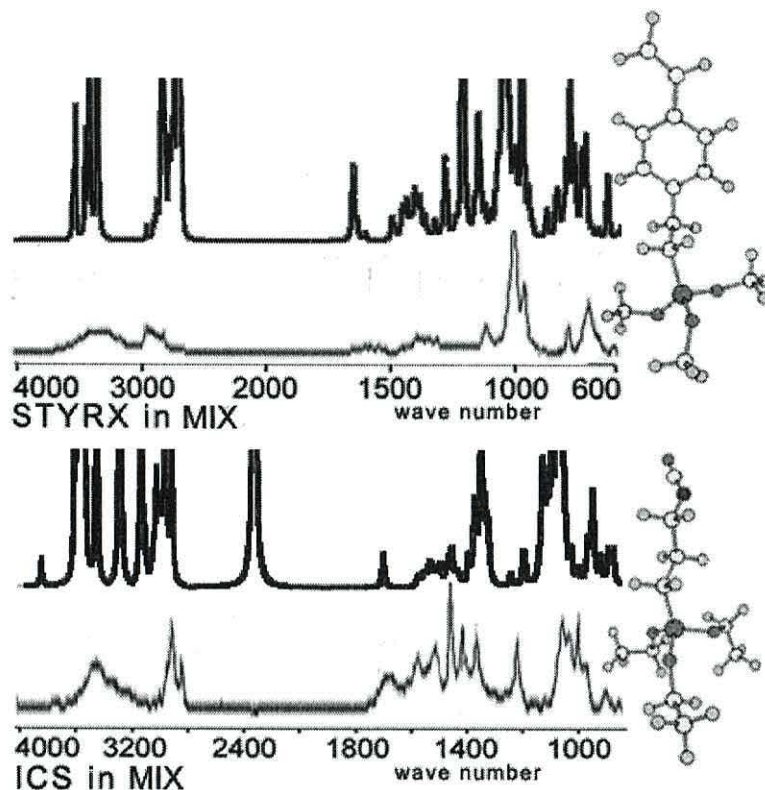


**Figure 5.5** Comparison of computed (top on each part) and experimental (bottom on each part) infrared spectra for STYRX and ICS. The theoretical part was attained using explicit H<sub>2</sub>O solvent particles, the experimental part was attained using 95 : 5 ethanol : H<sub>2</sub>O solvent.

**5.1.3.4. 95 : 5 ethanol : H<sub>2</sub>O Solvated conformers.** As Figure 5.5 above shows, water-solvated silane calculations resulted in a comparatively poor matching between theoretical and experimental results. This once again illustrates how important solvent is for molecular behaviour. Figure 5.6 shows a much better match. Please note that even though the peak around 2400 cm<sup>-1</sup> in ICS is very low



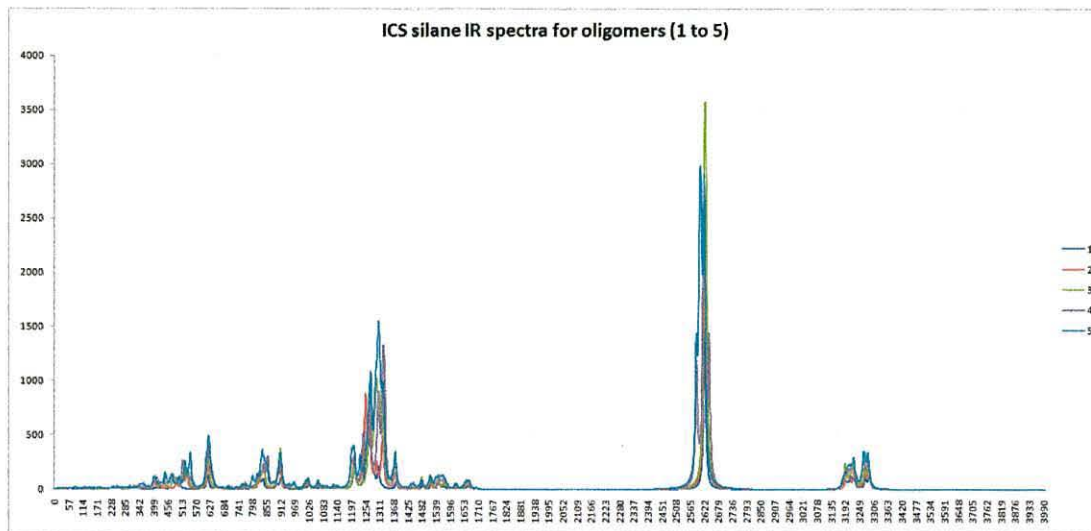
in the experimental data, it is one of the identifying peaks for 3-isocyanatopropyltriethoxysilane, as it corresponds to the double bond stretches in the terminal N=C=O group.



**Figure 5.6** Comparison of computed (top on each part) and experimental (bottom on each part) infrared spectra for STYRX and ICS. The theoretical match was attained using MIX – explicit 6 : 1 ethanol : H<sub>2</sub>O solvent.

**5.1.3.5. Spectra.** Some of the major IR peaks have been described in previous sections. Additionally, IR analysis confirms that the main change during hydrolysis is the replacement of methoxy / ethoxy groups by OH-groups. 1180-1200 cm<sup>-1</sup> region corresponds to CH<sub>2</sub> twisting. Both hydrolysed and non-hydrolysed ICSs have peaks in the 950-1000 cm<sup>-1</sup> region, the only difference is that in hydrolysed silanes the region is also intensified by Si-O-H dihedral angle

changes. The ICS oligomers' spectra does not show any major differences from the monomer's IR (except for the relative intensity of key peaks, see Figure 5.7).



**Figure 5.7** Five overlaid ICS monomer and oligomer (2-, 3-, 4-, 5-) theoretical spectra.

**5.1.3.6. AIM analysis of interactions.** In STYRX, the most preferable conformation (24 of 36 conformers converged to it) is stabilised by interactions between carbon atoms from the phenol ring and the methoxy group. The interactions are present in gas-phase and in implicit  $\text{CCl}_4$ . In ICS, a very common long-range interaction is found between the  $\text{CH}_2$  group (closest to the “ $-\text{N}=\text{C}=\text{O}$ ” organofunctional group) and one of the ethoxy groups. Another common pair of interactions in ICS is between the nitrogen and oxygen atoms from the “ $-\text{N}=\text{C}=\text{O}$ ” organofunctional group and two of the three ethoxy groups. These interactions (especially the first between  $\text{CH}_2$  and an ethoxy group) are present in the gas-phase, and both implicit solvents ( $\text{H}_2\text{O}$  and  $\text{CCl}_4$ ). A very weak interaction ( $\rho_b \approx 0.0084$ ) in STYRX in almost all conditions exists between the terminal  $\text{CH}_2$  group and a hydrogen from the phenol ring. It does not exist in some explicitly

solvated conformers (the most preferable ones), in which the CH<sub>2</sub> group and the phenol ring are not coplanar.

**5.1.3.7. Preliminary CPMD results.** Both STYRX and ICS were noticeably much more flexible in CCl<sub>4</sub> than in ethanol, but remained in their fully-extended, ‘all-*anti*’ conformations. For STYRX: although the predicted difference in energy between two ‘all-*anti*’ conformers is very low (<5.0 kJ·mol<sup>-1</sup> in each solvent), the entire CPMD calculation in ethanol did not show any transitions between the two conformations, whereas the molecular dynamics in CCl<sub>4</sub> showed a relatively low-energy barrier, manifested as conformational transitions between the ‘all-*anti*’ space.

#### 5.1.4 Conclusions

As a result of the successful synergy between experiment (FTIR) and theory for the STYRX and ICS silane coupling agents, the results of isolated single molecule calculations were indeed validated. From these, extrapolations could therefore be confidently made into more in-depth characterisations of their geometric and electronic structures, energetic trends and dynamic behaviours in both the gas-phase and solvated environs.

Theoretical conformational analysis showed the existence of several lowest-energy conformer classes (see also Appendix A). The most populated conformer classes are not in the preferred ‘all-*anti*’ geometry in the non-hydrolysed state. However, the

CPMD results showed that in the two solvents (CCl<sub>4</sub> and ethanol) the ‘all-*anti*’ conformations are comparatively stable (especially in ethanol).

The validated synergy also allowed for quantitative assignment of molecular vibrations to the major peaks present in the FTIR results. Although beyond the scope of the main body of this work, these specificities are visually and graphically detailed in the appendices. The most important aspect to note is the proof-of-concept for the theoretical-design of novel biocompatible silane adhesives using the novel algorithms and explicit particle solvation algorithm, with resultant quantitative prediction of the IR spectra and expected peaks, prior to commencement of synthesis. Determination and subsequent prediction of mechanical properties thereof, being the end-goal of this work.



## **5.2 Towards the Protonation Limit of Amino Acids and Peptides I: An Algorithm to track the Dynamics of Conformation- and Orbital- Specific Poly-Protonation**

### **5.2.1 A Background/Introduction**

Proteins are essential bioactive molecules, found in every living organism on Earth. They participate in almost every task that is essential for life, and their malfunctioning, lack or abnormal organisation can cause serious illnesses such as cancer,<sup>291</sup> neurodegenerative disorders (Huntington's, Alzheimer's, *etc.*),<sup>292, 293</sup> among others. Proteins also serve as a perfect example of an immense multitude of fine machinery, working as a system to display functional properties beyond the sum-total of their molecular components.

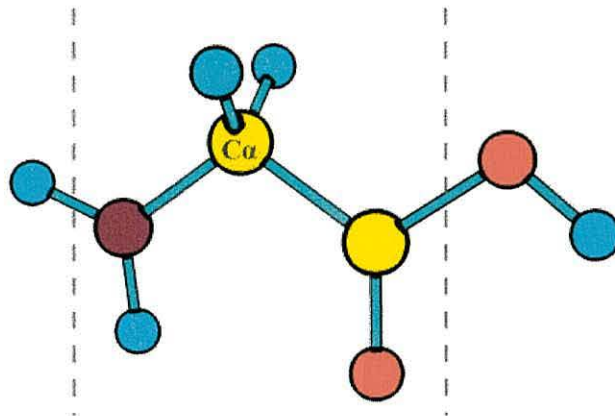
Protein structure stability and functioning, as well as the behaviour of other macromolecules, are highly dependent on the concentration of hydrogen ions (pH),<sup>294</sup> which varies in different parts of human body.<sup>295</sup> Theoretical and experimental studies of pH-dependent properties of macromolecules (stability, activity, sensitivity, solubility *etc.*) show the fine-tuning of the solvent environment (and pH thereof) as being essential to proper metabolic function.<sup>296-302</sup>

A protein can exchange protons with its environment, which changes both the total charge and the distribution of charge around the molecule. These changes, in turn, influence the overall three-dimensional structure of the polypeptide chain as a

function of its constituent peptide components (amino acid residues). Several works on the pH-dependence of protein denaturation are covered in the literature.<sup>296, 303-306</sup>

A protein's **pH optimum** is defined as the concentration of hydrogen ions ( $H^+$ ) at which the protein optimally operates.<sup>295</sup> The value of the optimal concentration depends on many factors (in general: protein structure and conformation, environmental conditions, *etc.*), with protein primary structure (amino acid composition) and the organisation (location within the protein with respect to the surface) of the titrable side-chain groups within the protein being the most influential.<sup>295</sup> Titration curves and effective pKa values of ionisable groups in proteins have been shown as being sufficient to calculate the pH-dependence of the denaturation free energy with respect to some reference pH value, as well as proving the influence of ionisable groups on protein destabilising as being strongly pH-dependent.<sup>307, 308</sup>

Most works on the pH-dependence of the structural stability and activity of proteins were carried-out on polypeptide chains of ten or more amino acids (*e.g.* see Refs. <sup>296, 298-300, 309-312</sup>), however some considered much shorter chains (*e.g.* see Ref. <sup>313</sup> for an experimental study on intramolecular interactions in protonated tripeptides, or Ref. <sup>314</sup> for a theoretical study to identify the conformational preferences of short lysine-based oligopeptides); even single amino acids have been so-characterised (pay special attention to Ref. <sup>315</sup> – a study on mono-, di-, tri-, and tetra-protonation of Guanidine, and Ref. <sup>316</sup> about the effects of ionic strength on the protonation of methionine, leucine, threonine and cysteine). As a continuation of these works, we present a theoretical protonation and conformational analysis of glycine.



**Figure 5.8** Schematic representation of the free amino acid glycine with the N- and C-terminal ends arranged on the left and right, respectively.

Glycine (Fig. 5.8) is the smallest of all amino acids and the only one that does not display point chirality (since the side chain of this hydrophilic amino acid consists of just one hydrogen atom). Although most proteins contain comparatively small quantities of glycine, some structural proteins (*e.g.* collagen, elastin) can contain up to 35% glycine content.<sup>317</sup>

Glycine is physiologically very important, as in higher eukaryotes it takes part in the biosynthesis of the key precursor of porphyrins – *D*-aminolevulinic acid. It also provides the C<sub>2</sub>N subunit of all purines.<sup>317</sup> Glycine also acts as an inhibitory neurotransmitter in the Central Nervous System, as well as a co-agonist (along with glutamate) for *N*-methyl *D*-aspartate receptors (the predominant molecular devices for controlling synaptic plasticity and memory function),<sup>318</sup> also serving as an intermediate in the synthesis of a variety of chemical products.



## 5.2.2 Methods

**5.2.2.1. MP2 calculations.** Theoretical investigations of the conformational space of glycine under various protonation conditions were conducted through construction of models characterised using the post-Hartree-Fock second order Møller-Plesset Perturbation Theory (full) method, employing the 6-311++G(d,p) Pople basis set. All computations were carried out using g03. To ensure quantitative characterisation of all intra-molecular interactions, atoms-in-molecules (AIM) analyses were conducted on the wavefunctions generated from the MP2(full)/6-311++G(d,p) geometry-optimised structures.

For the naming notation and protonation site descriptions, see Appendix B. For each protonation (including the non-protonated glycine), 36 conformers were generated, geometry-optimised and split into classes depending on their optimised conformations. Frequency calculations at the same level as the geometry-optimisations, were used to confirm that each class representative resided at a minimum on its respective PEHS.

**5.2.2.2. Car Parrinello Molecular Dynamics (CPMD).** CPMD simulations of a non-protonated, and five single-protonated glycines were performed, with the timestep of 4 a.u. ( $\approx 0.1$  femtoseconds) and a trajectory generated for 10,000 steps, for a total trajectory time  $\approx 1$  picosecond. CPMD DFT parameters: FUNCTIONAL BLYP (the exchange-corellation functional), NEWCODE. Temperature = 300.15 K. CPMD SYSTEM parameters: ANGSTROM, SYMMETRY = simple cubic, CELL = 10.0 1.0 1.0 0.0 0.0 0.0, CUTOFF = 60.0.



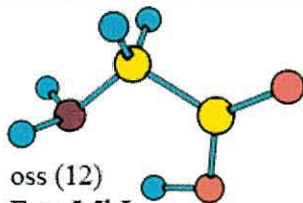
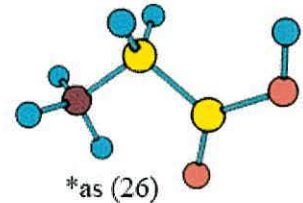
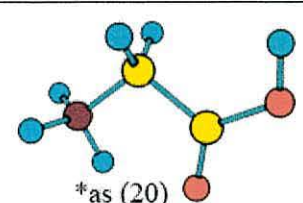
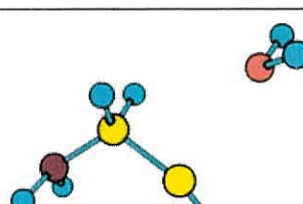
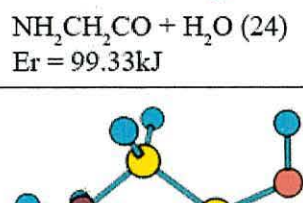
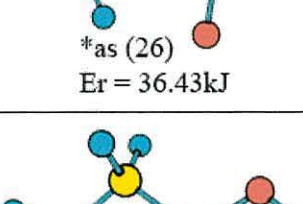
**5.2.2.3. Atoms-In-Molecules Wavefunction Analyses.** Bader's Atoms-In-Molecules analyses<sup>215</sup> were performed using the AIM2000 program package,<sup>232</sup> using default values except for the "Critical Points Calculation" where the "Stepsize factor for Newton iteration" parameter was changed from 1.0 to 0.5 for more precise density critical point calculations.

### 5.2.3 Results and Discussion

Prior to the in-depth characterisation of all protonation conformer classes, the list in Table 5.1 allows for a general outline of all conformational space mapping for the *ab initio* calculations carried-out; each protonation (as well as the non-protonated state) having a total of 36 conformers.

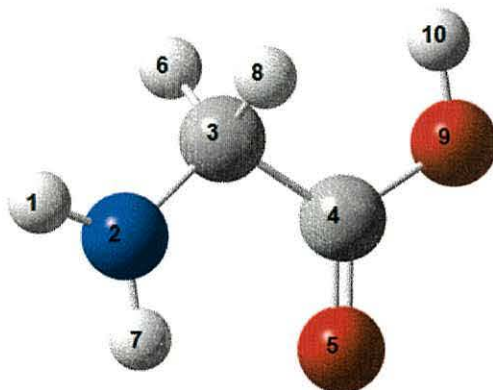
The general trend is that a non-protonated glycine has more flexibility, and in general – a larger number of stable low-energy conformation classes (CCs), whereas single-protonated glycines tend to have fewer CCs. The only exception among the protonated glycines is "hn-gh2-oh" (last row of Table 5.1), which quite surprisingly has seven CCs. However, the large energy difference between the lowest-energy (most-populated) class and the other 6 classes (with relative energies higher than 110 kJ·mol<sup>-1</sup>) suggests very low flexibility of "hn-gh2-oh". Hence, the other six CCs are rarely visited local minima, whereas the lowest-energy class is by-far the most populated.

**Table 5.1** Glycine protonation and conformational space mapping: a general summary of results.

	Number of conformer classes (CCs)	Range of difference in energy between CCs, kJ·mol <sup>-1</sup>	Most populated CC
<b>hn-g-oh</b>	7	30.34	 oss (12) Er = 5.5kJ
<b>h2n-g-oh</b>	3	36.07	 *as (26) Er = 36.07kJ
<b>hn-hg-oh</b>	3	23.7	 *as (20) Er = 109.85kJ
<b>hn-g-oh2</b>	2	99.33	 NH <sub>2</sub> CH <sub>2</sub> CO + H <sub>2</sub> O (24) Er = 99.33kJ
<b>hn-gh1-oh</b>	4	115.21	 *as (26) Er = 36.43kJ
<b>hn-gh2-oh</b>	7	133.68	 *aa (16) Er = 0kJ

*Notes:* (1) each one of the six structures is a separate system and relative energies are given **separately** for each system, not for all systems together; (2) “oss (12)” means that 12 out of 36 structures converged to a conformational class named “oss”; (3) the number of conformer classes is given for each system separately.

For reference the atomic numbering used for glycine is presented, where the additional proton is always #11. The numbers shown in Figure 5.9 are used in atomic-nomenclature from here-on (*e.g.* O5, N2, H10, *etc.*)



**Figure 5.9** The atomic-numbering and nomenclature system used for glycine.

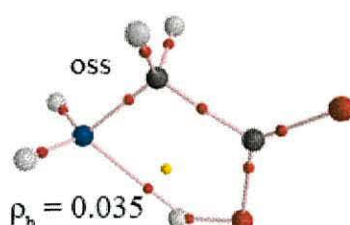
**5.2.3.1. Non-protonated glycine.** For the neutral, parent (non-protonated) glycine:

- of 36 topologically-possible conformers, seven represent stable geometry-optimised conformer classes;
- free energy difference between conformational classes is not very high (maximum of  $\sim 30.34 \text{ kJ}\cdot\text{mol}^{-1}$ ) and 4 classes have relative free-energies lower than  $15 \text{ kJ}\cdot\text{mol}^{-1}$ ;
- conformer convergence to different classes is comparatively even – the class occupancies are 1, 2, 4, 4, 4, 5, and 12 (four did not converge).

The three points outlined above imply high flexibility of neutral glycine, and

show that the global minimum is not a well-defined single geometry, but rather a resonance between closely energetically-spaced local minima on the respective PEHS, in near-equilibrium.

The most populated and lowest-energy class, is stabilised by a hydrogen bond between the H10 and N2, with  $\rho_b = 0.0350$  (Fig. 5.10).

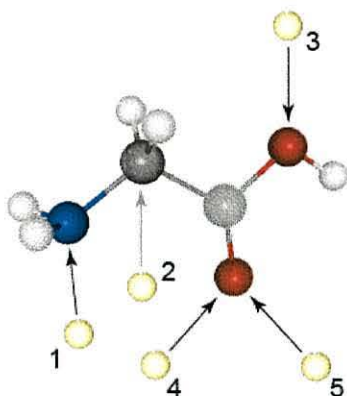


**Figure 5.10** One of the lowest energy and most populated classes (12 conformers out of 36) of neutral glycine is stabilised by an intra-molecular interaction ( $\rho_b = 0.0350$ ) between the C-terminus Hydrogen and the Nitrogen.

Only the most geometrically-proximate conformers converged to this stable optimised class (see “Calculation results” table, Appendix B), further supporting the finding that there is a resonance and equilibrium with other stable, albeit slightly higher in energy, minima on the neutral glycine PEHS.

**5.2.3.2. Protonations: general characteristics, conformer classes and their analysis.** The five singly-protonated glycines were categorised as follows: *h2n-g-oh* (Fig. 5.11, 1; protonated at N2), *hn-hg-oh* (Fig. 5.11, 2; an attempt to protonate the  $\alpha$ -Carbon), *hn-g-oh2* (Fig. 5.11, 3; protonated at O9), *hn-gh1-oh* and *hn-gh2-oh* (Fig. 5.11, 4 & 5; the latter two protonated at the lone pairs of O5).



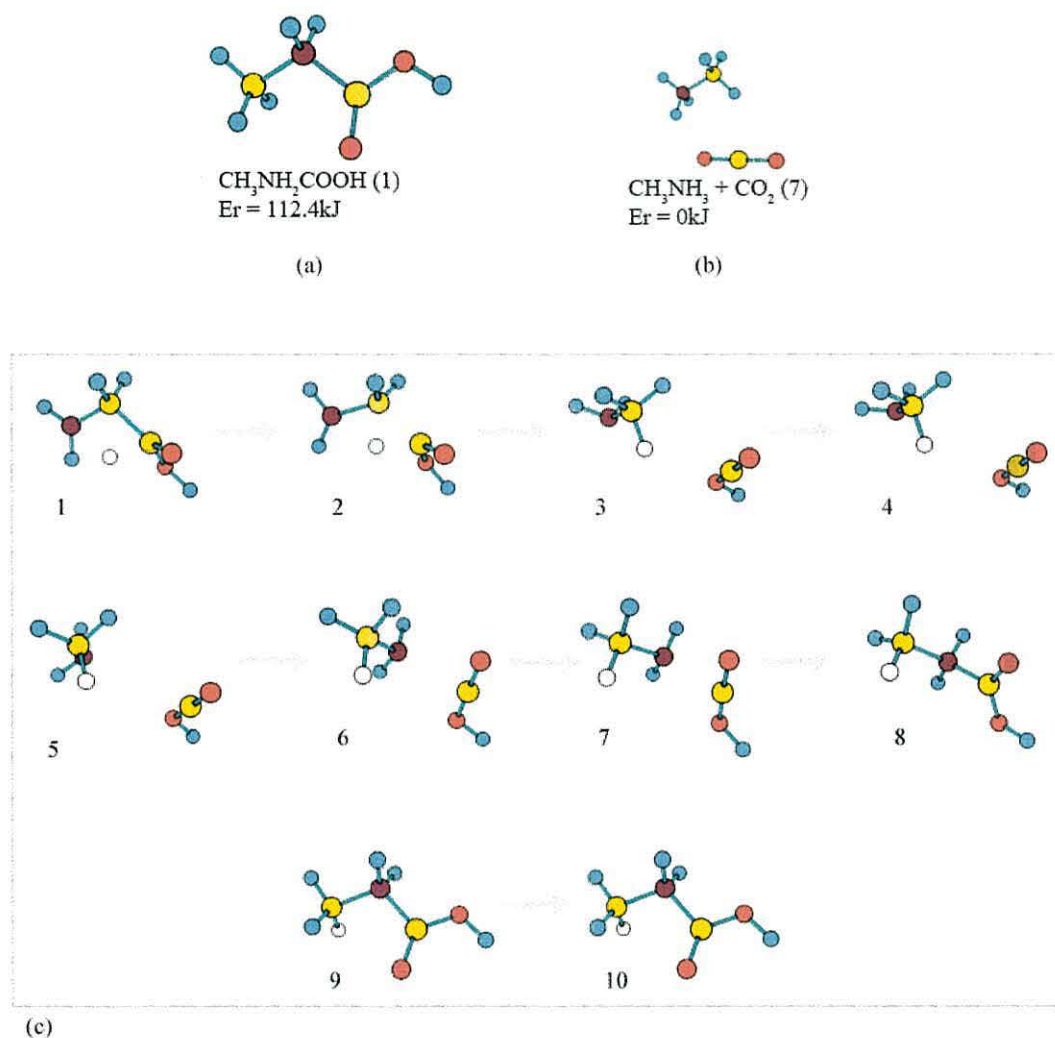


**Figure 5.11** The 5-differing potential protonation sites on the glycine amino acid; 4 and 5 represent protonation at the two lone pairs of the carbonyl oxygen.

***h2n-g-oh***: the nitrogen-protonated glycine conformational analysis showed only three principle conformational classes as being stable, with a relatively small relative energy range of  $\sim 36.07 \text{ kJ}\cdot\text{mol}^{-1}$ . In all three conformer classes, the added proton remained bound to the nitrogen, avoiding proton transfer as observed in other singly-protonated systems (*i.e.* between N2 and O5, see below). The lowest energy as well as most highly-populated conformer classes showed very similar geometries, with the exception of the hydroxyl rotor (H10 dihedral). The conformers in this class are stabilised by an interaction between O5 and one of the N-terminus hydrogens.

*Note*: the decrease in the number of conformer classes is also caused by the fact that once H11 is added to the system and connected to N2, there is no distinction between the conformational classes having labels starting with ‘o’, ‘i’, or in ‘p’, as there is no difference between H1, H7 and the newly-connected H11 (‘p’ is in fact never present as  $\text{NH}_3$  always maintains a near trigonal pyramidal geometry; see Appendix B).

**hn-hg-oh:** Expectations for this system were that H11 would not stably bind the  $\alpha$ -Carbon (C3), however results were quite novel in several ways. First, there are three stable optimised conformer classes, one being a fragmented system consisting of CO<sub>2</sub> and methyl-amine (NH<sub>3</sub>CH<sub>3</sub>) and **none being the lowest-energy state** for this protonation pattern (Fig. 5.12, b). In fact, the system re-arranges to form CH<sub>3</sub>NH<sub>2</sub>CO<sub>2</sub>H, through the stepwise process shown in Figure 5.12, c.

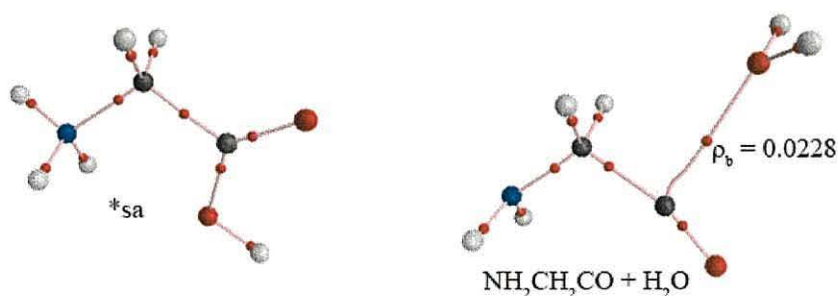


**Figure 5.12** *hn-hg-oh* conformational class representatives ((a) and (b)). (c) shows the optimisation process of a glycine conformer protonated at the  $\alpha$ -Carbon. *Note:* step 10 is only slightly different from step 9, as those are some of the last optimisation steps where the new structure found a minimum and was slowly optimising towards it.

The additional proton does initially bind the  $\alpha$ -Carbon, fragmenting the molecule. It is hard to predict what would happen to the OCOH fragment in solution, as it would have great difficulty in re-binding to the  $\text{NH}_2\text{CH}_3$  fragment (Figure 5.12, c). This also illustrates that in order to completely understand the full dynamic behaviour of any molecular system, it is essential to initially complete gas-phase calculations.

A similar reaction happened to form the conformer shown in Figure 5.12, b; however, in that case **the OCOH part did not connect back, but lost a proton to  $\text{NH}_2\text{CH}_3$**  (this will also be covered in the ‘Proton transfer examples’ section below).

*hn-g-oh2*: only two conformer classes were found for O9 protonation (see Fig. 5.13 for AIM analyses). The first class is very similar to one of the *h2n-g-oh* classes, showing that nitrogen will ‘steal the spare proton’, either the one added to the system (H11), or the one already in the system (H10). The second class is more interesting in the way that neutral glycine is fragmented, with retention of an inter-molecular interaction between the newly formed water molecule and the remaining glycine parent fragment ( $\rho_b = 0.0228$ , Figure 5.13); the stability of this complex has yet to be determined in ‘real-world’ solution.



**Figure 5.13** AIM analyses of *hn-g-oh2* conformer classes.



***hn-gh1-oh***: with four conformer classes, this protonation featured two structural types: (1) in 30 of 36 conformers, the newly added H11 proton attached to the Nitrogen, forming a geometry that is the same as that of most ***h2n-g-oh*** lowest-energy glycines; (2) 6 other conformers had H11 connected to O5, thus forming two OH groups at the C-terminus of glycine, but their energy was more than  $110 \text{ kJ}\cdot\text{mol}^{-1}$  higher than that of the representatives of the first type. This once again leads to the conclusion that provided sufficient time the nitrogen will pull H11 from the oxygen and form a more stable system.

***hn-gh2-oh***: this system had a similar behaviour to the previous one, with only one exception – many more conformers (20 out of 36) had two OH groups (all other conformers formed a stable glycine with H11 connected to N2). These twenty conformers converged to 6 classes with different geometries, but very close to each other in terms of energy (with relative values ranging between  $\sim 110\text{-}133 \text{ kJ}\cdot\text{mol}^{-1}$ ).

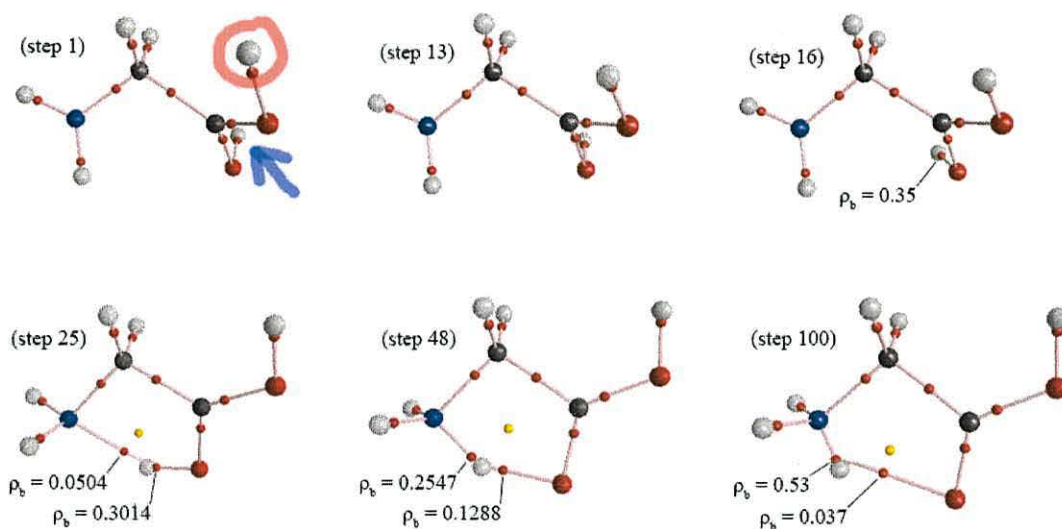
**5.2.3.3. Proton transfer examples.** A few *intra*-molecular proton transfers were observed during geometry optimisations and molecular dynamics of the single protonation of glycine. Despite the fact that the proton H11 is more stable at N2 than O5, molecular dynamics results show that when the nitrogen has 3 hydrogens, it does not have enough density to retain all three all the time; once again resonance between closely energetically-space structures provides the equilibrium.

*Inter*-molecular proton transfers also took place during geometry optimisation of selected ***hn-hg-oh*** conformers: Figure 5.12, c shows that when neutral glycine is fragmented to  $\text{NH}_2\text{CH}_3$  and  $\text{OCOH}$ , the latter can reconnect with the former by forming a bond between C4 and N2, stabilising nitrogen. A different situation can



occur, where the terminal hydrogen of OCOH connects to the nitrogen, with OCO leaving as a separate molecule.

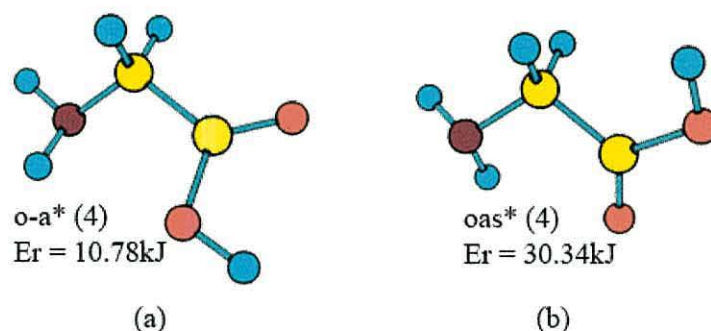
Finally, for in the *hn-gh1-oh* protonated system, proton approach made a significant change in the negative charge distribution of glycine, making H10 much more active, promoting its transfer to nitrogen (Figure 5.14), where bonding seems to be more preferable.



**Figure 5.14** AIM analysis of an *hn-gh1-oh* proton transfer. Up to step 16, the OH-group Hydrogen (marked by blue arrow) was almost inert, but after the proton (circled in red) approached Oxygen #5, Hydrogen #10 became much more active, subsequently transferring to the Nitrogen.

**5.2.3.4. Molecular Dynamics experiments.** CPMD results for neutral glycine showed an extremely high stability of the conformer shown in Figure 5.15, a. The structure never left the conformation over the whole trajectory of 10,000 steps ( $\approx 1$  picosecond), even though the temperature was high enough (300.15 K) to make other conformers explore large areas of the potential energy hyper surface.

The conformer shown in Figure 5.15, b explored almost every possible configuration of both the C-terminus and the N-terminus, whereas the dihedral angle N2-C3-C4-O5 was very stable.



**Figure 5.15** A stable conformer of neutral glycine (a), and an active conformer of the same molecular system (b). Notice that the structures are almost the same and the difference is only the conformation of the terminal carboxylic acid moiety.

CPMD calculations of *h2n-g-oh* reproduced the predicted *grid-based conformational space search* protonation calculations. Specifically, that Nitrogen captures H11 and retains it most of the time, in a near trigonal pyramidal geometry, with possibility of repeated proton transfer between N2 to O5. A slight increase in the flexibility of the C-terminus was also observed in *h2n-g-oh* calculations.

Molecular Dynamics of the protonation at oxygen #9 (*hn-g-oh2*) showed that the molecule is not only fragmented to *two* pieces –  $\text{NH}_2\text{CH}_2\text{CO}$  and  $\text{H}_2\text{O}$  – but actually in three, as  $\text{NH}_2\text{CH}_2\text{CO}$  is not stable and is further divided to form  $\text{NH}_2\text{CH}_2$  and  $\text{CO}$ .

*hn-gh1-oh* calculation showed a very high level of flexibility of glycine and multiple examples of N2---O5 proton transfer. Actually, this O-protonated glycine was so flexible that the otherwise rigid dihedral angle N2-C3-C4-O5 easily twisted by 180 degrees.

Results of the *hn-gh2-oh* molecular dynamics calculation showed di-hydroxy glycine being very stable (as predicted earlier by conformational analysis), but not stable enough to prevent one of the OH groups from turning to the Nitrogen and eventually losing its Hydrogen, through an H-transfer to N2.

#### 5.2.4 Conclusions

A combination of grid-based conformational search and Car-Parrinello Molecular Dynamics (CPMD) based analyses has allowed for the characterisation of the stability and preferable geometries of neutral and singly-protonated glycine. Additionally, some proton-mediated fragmentation patterns have also been characterised, which might help shed light on related phenomena in mass-spectrometric investigations of peptide systems.<sup>313, 319, 320</sup>

It was shown that in order to fully characterise the intra-molecular dynamics of a molecular system, initial investigations should be carried-out in gas-phase conditions, with subsequent comparison to results in solvent.

Neutral glycine is a flexible molecule with multiple local minima which are closely energetically-spaced on the PEHS. Singly-protonated glycine has a much smaller number of structures residing at minima on its respective PEHS, with only one well-defined global minimum, all other minima being energetically much less favourable.

During protonation, the N-atom most often binds the first proton, proton transfer between atoms N2 and O5 is quite probable, making initial O-capture, followed by re-arrangement another likely pathway to a stable singly-protonated structure.



## 6. CONCLUSIONS

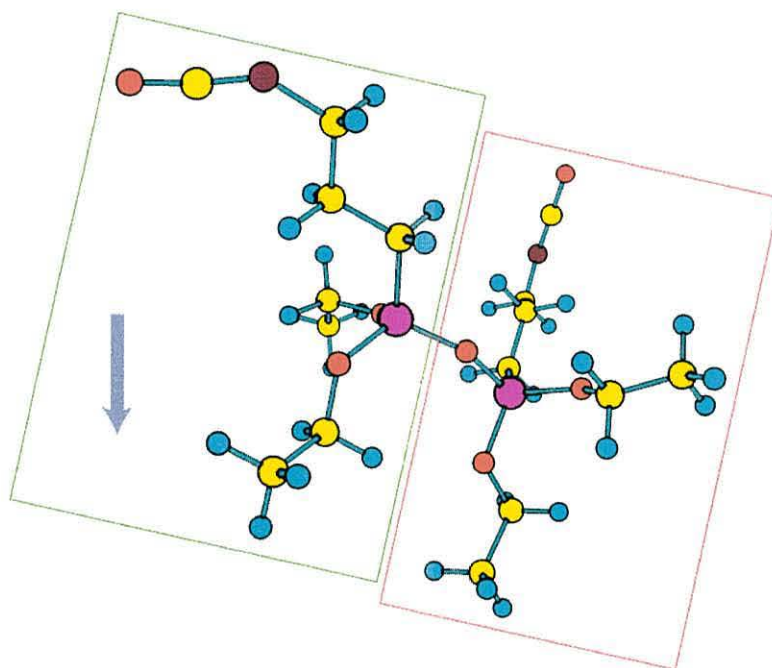
---

The development of an automated computerised approach to conformational search calculations and analysis provided for a complete description of three bioactive chemical systems: 3-styrylethyltrimethoxysilane (STYRX), 3-isocyanatopropyltriethoxysilane (ICS), and glycine amino acid. The theoretical studies were performed both in gas-phase and under various solvent conditions (implicit and explicit models; CCl<sub>4</sub>, H<sub>2</sub>O, ethanol and a mix of 95% : 5% ethanol with H<sub>2</sub>O). Silane theoretical findings were confirmed by experiment.

**On the computational side**, the most time-consuming tasks were (1) input file generation, and (2) results tabulation. These two tasks were completely automated with algorithm execution times on an average computer not exceeding 3-4 seconds even for comparatively large systems consisting of a few thousand conformers.

Special attention was paid to explicit solvent model generation, as solvent effect is one of the defining factors in molecular behaviour. An algorithm based on Bader atomic radii<sup>215</sup> was developed to facilitate in optimal solvation of biochemical systems by any solvent. Another algorithm was also proposed, which is based on in-built *ab initio* charge density calculations. In both cases the defining factor is the optimal matching of charge density concentrations and depletions on the surfaces of the solute and solvent particles, as well as solvent-solvent interactions.

**STYRX and ICS theoretical computations** were matched with experimental data. These calculations show that the most populated conformational classes do not have the preferred ‘standing up’ geometry (which is referable for both the strong surface attachment *via* hydrolysis of silicon-connected groups on one end, and the polymerisation of the organo-functional group on the other). However, a few low-energy conformers do have the required geometry, and Car Parrinello Molecular Dynamics<sup>1</sup> experiments show that in two solvents (CCl<sub>4</sub> and ethanol) these conformers are comparatively stable (especially in ethanol).



**Figure 6.1** The surface setting of one ICS molecule (left rectangle, the arrow shows the setting direction) is partially blocked by another ICS molecule (right rectangle) in the dimerised silane.

The validated synergy between experiment and theory allowed for quantitative assignment of molecular vibrations to the major peaks in the Fourier Transform Infrared Spectroscopy results. It also partially answered the question about polymerisation of silanes *via* formation of Si—O—Si bonds. Theoretical calculations

of di-, tri-, tetra- and pentasilanes show that the structures keep the ‘standing up’ geometry, but are not in the most convenient conformation for the setting reaction (see Figure 6.1 for a dimer example).

A complete conformational analysis of the non-protonated and five most probable singly-protonated **glycine** molecules was performed using the grid-based conformational search method and Car-Parrinello Molecular Dynamics. A good matching between grid-based conformer geometry optimisation and molecular dynamics results was obtained. It was shown that the full characterisation of a molecular system should necessarily start with the initial investigation in gas-phase conditions with subsequent comparison to results in solvent.

Protonation of glycine in gas-phase conditions reduces conformer flexibility; this also supports the finding that the potential energy hyper surface of the protonated amino acid usually features only one well-defined global minimum. On the other hand, for non-protonated glycine there exist multiple minima which are closely energetically spaced on the potential energy hyper surface – this explains the higher flexibility of the molecular system.

The calculations helped to answer the question about the preference of protonation sites of glycine and its molecular behaviour (including some proton-mediated fragmentation patterns) under various protonation scenarios. It was shown that the nitrogen is the most favourable protonation location. Moreover, protonation at the nitrogen is the only one that produces stable low-energy conformers in a singly-protonated glycine. Protonation at the oxygen (atom #5, see Fig. 5.9, p. 116) produces stable conformers which are much higher in energy ( $>110 \text{ kJ}\cdot\text{mol}^{-1}$ ).



Protonation at other sites (*e.g.*  $\alpha$ -Carbon or the C-terminus OH-group oxygen) usually leads to the destruction of glycine in gas-phase conditions.

The proton-mediated fragmentation patterns and proton transfer examples might help shed light on related phenomena in experimental mass-spectrometric investigations of peptides.

**Future works** in this and related areas include (1) di-, tri-, and polyprotonation of glycine and other amino acids both in gas-phase and in solvent, towards finding the protonation limit of peptides; (2) study of the solvent effect on the molecular behaviour of chemical systems; (3) search for a silane coupling agent with the required geometry and solvent conditions best for its functioning; and as a necessary prerequisite for all these projects – (4) development of the *ab initio* based solvation algorithm.

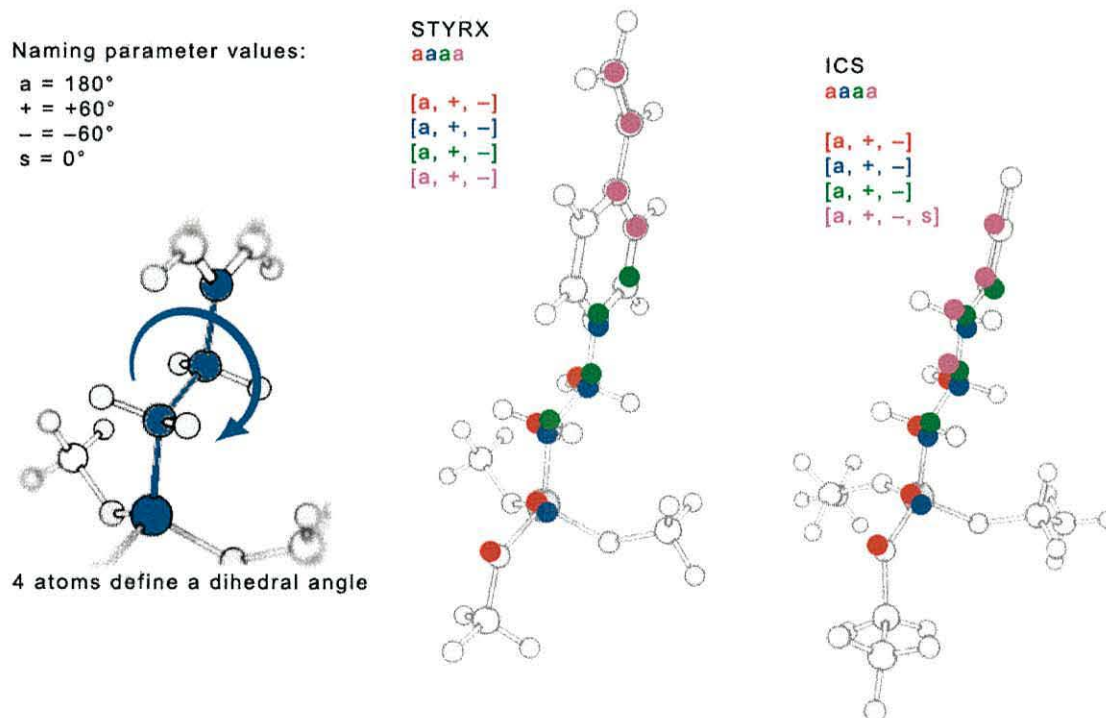
Another important aspect of our future research will be (5) the theory based design of experiments to validate all theoretical results (*i.e.* NMR of peptide protonation in super-acid, in real-time).



## 7. Appendix A. Rational design of novel dental adhesives: the role of computational steering in optimising dental silanes

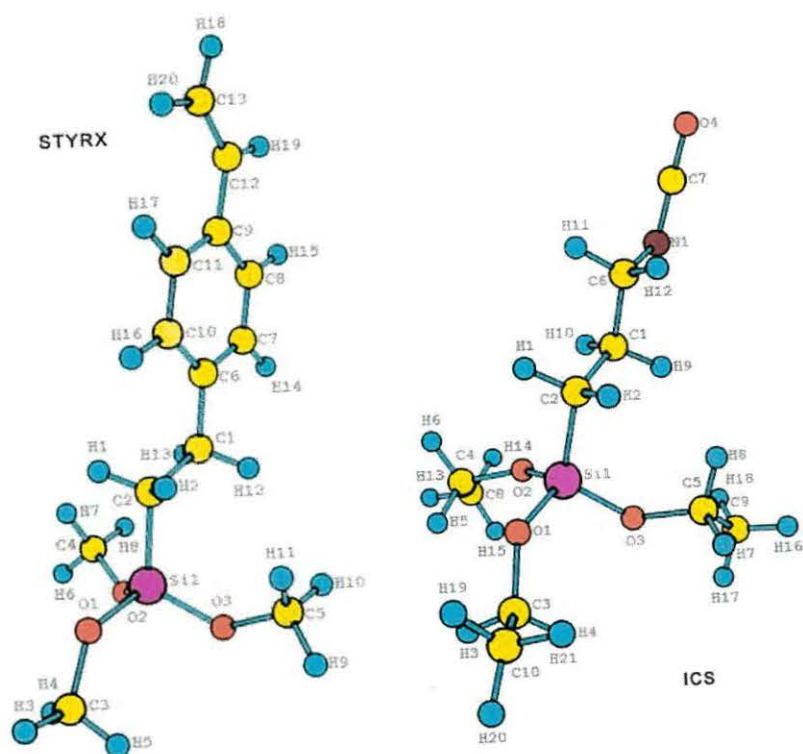
---

**Naming notations.** Parameter values in brackets represent possible values for each dihedral angle.

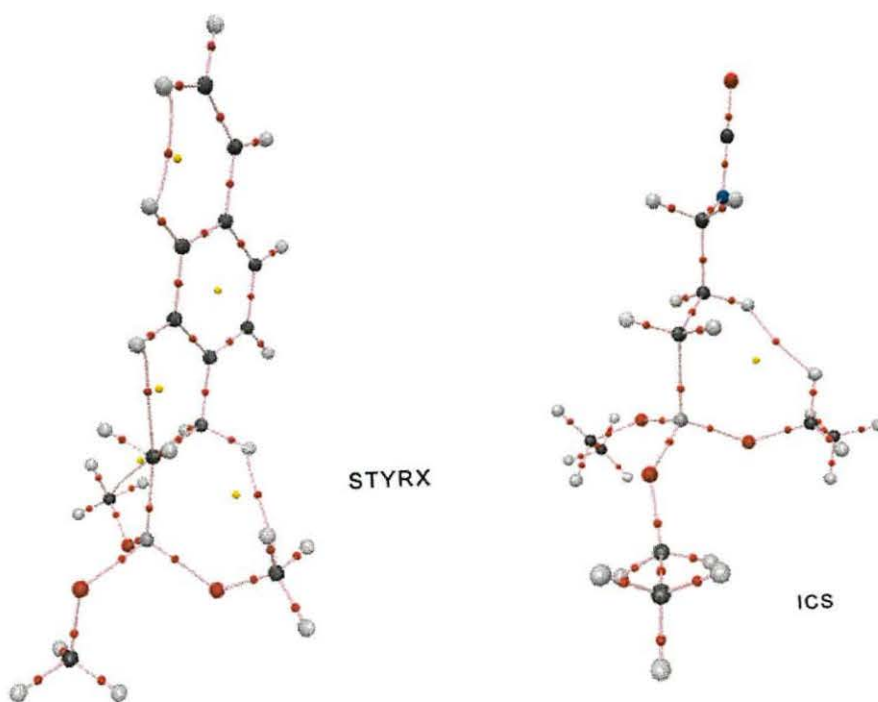


All energies (denoted as kJ for simplicity) are in  $\text{kJ}\cdot\text{mol}^{-1}$ . Energies in the “Thermodynamics” sections are in Hartrees. In the “Normal modes of vibration” sections, the negative frequencies indicate transition states of different orders.

## Structures.



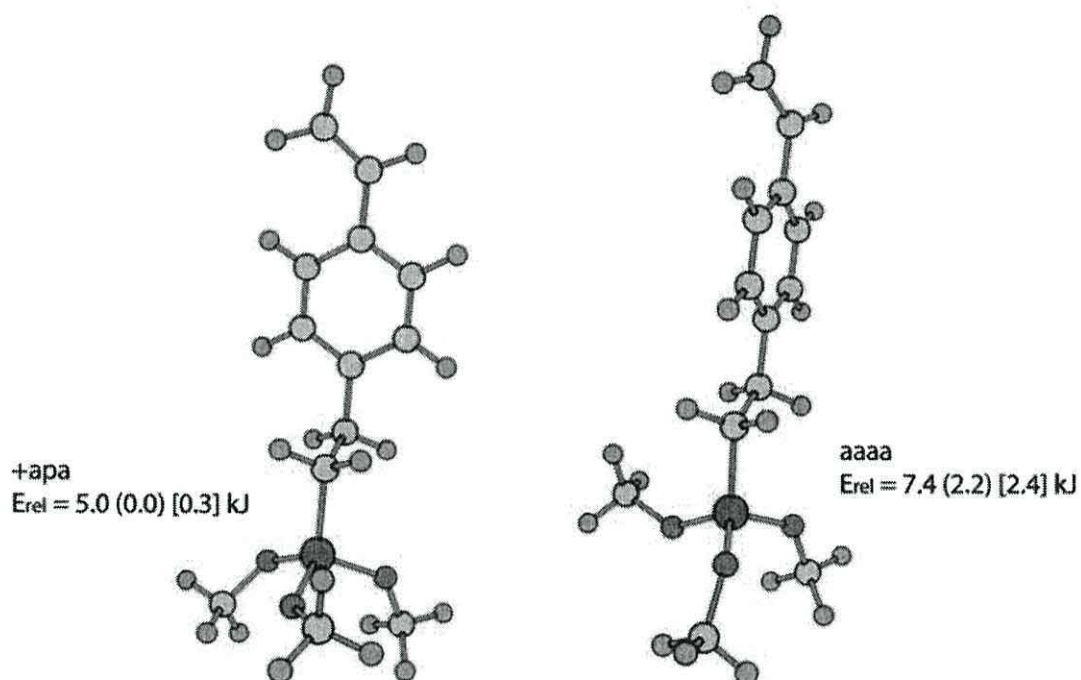
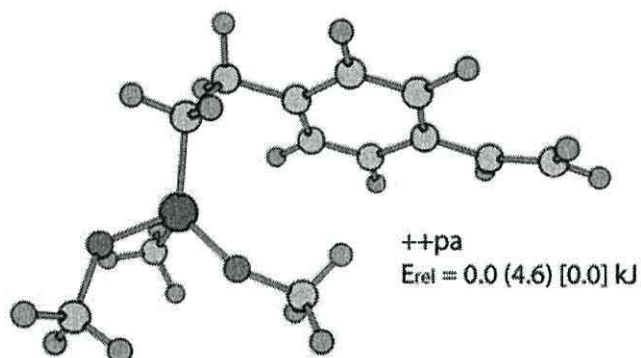
## AIM analysis.



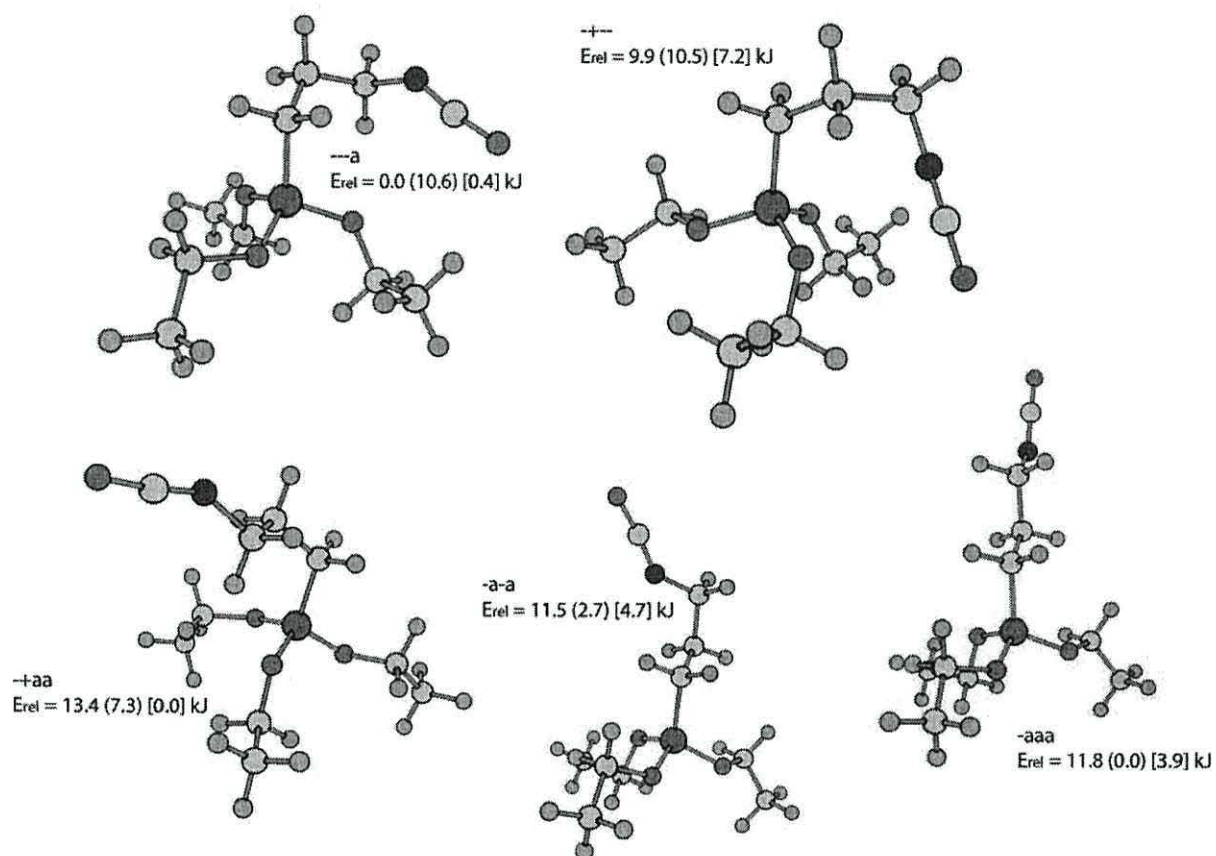
## STYRX and ICS lowest-energy conformer classes.

*STYRX* gas phase, PCM- $H_2O$ ,  $CCl_4$ .

(3 classes, relative energies are as follows: GAS\_PHASE ( $H_2O$ ) [ $CCl_4$ ]  $\text{kJ}\cdot\text{mol}^{-1}$ )



Five classes, relative energies are as follows: GAS\_PHASE (H<sub>2</sub>O) [CCl<sub>4</sub>] kJ·mol<sup>-1</sup>

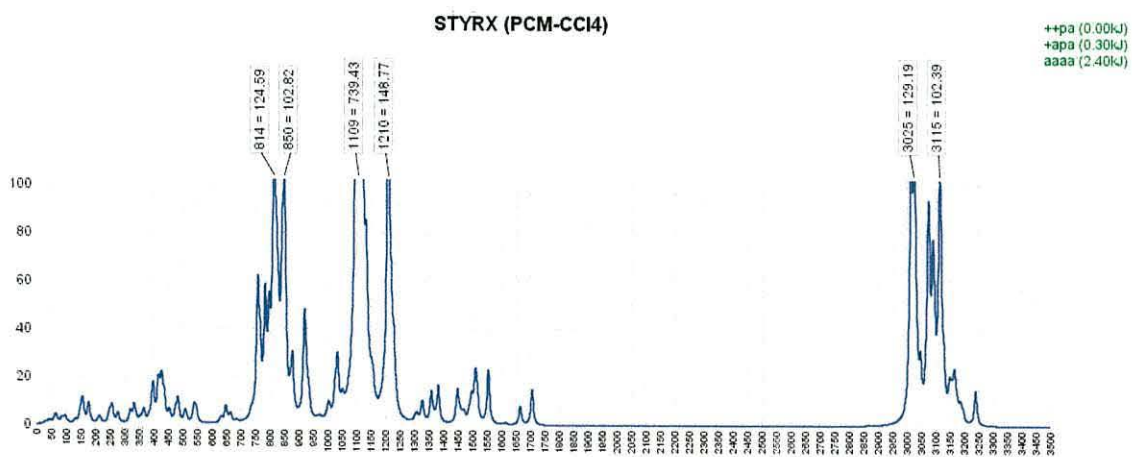
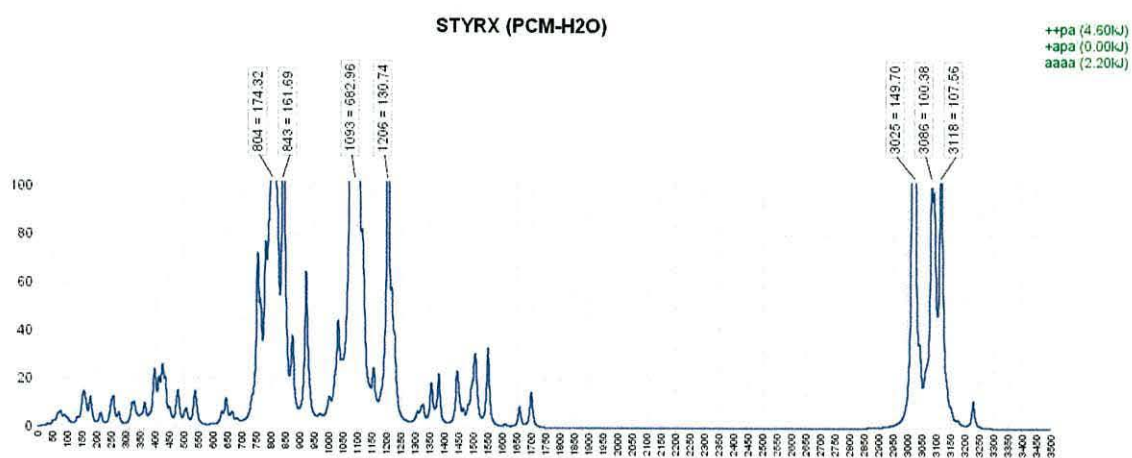
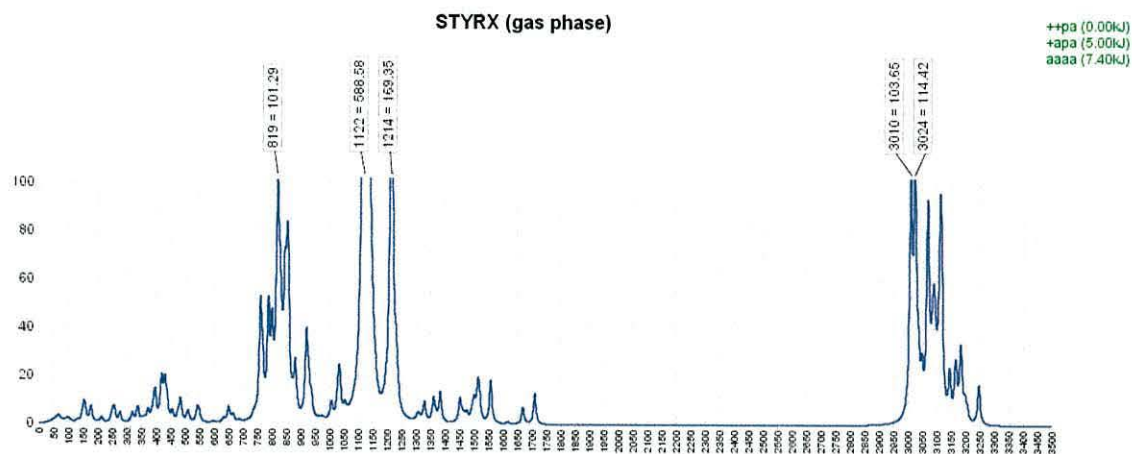


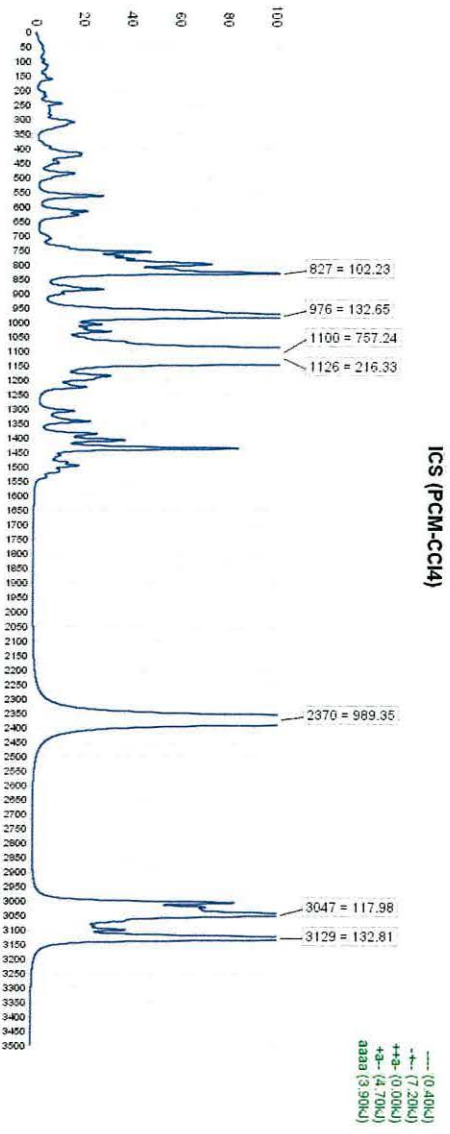
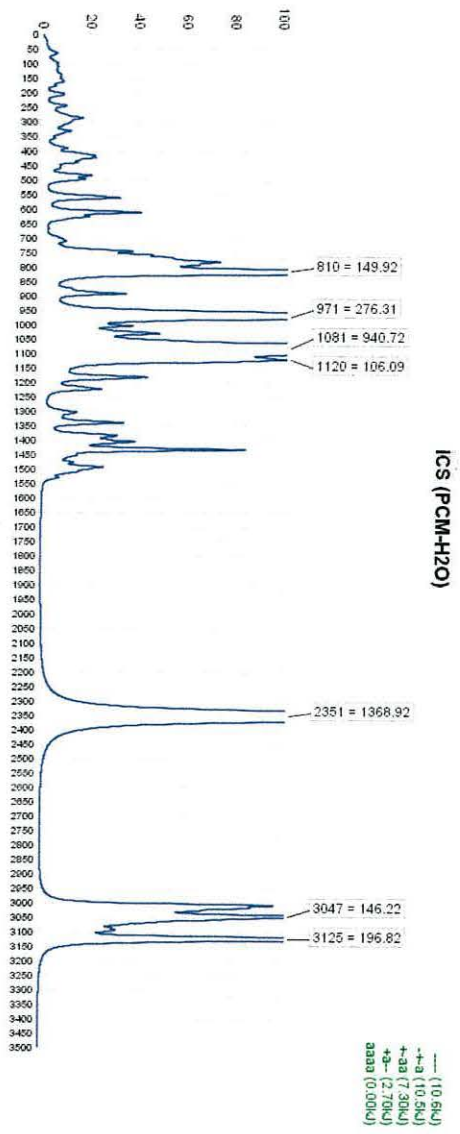
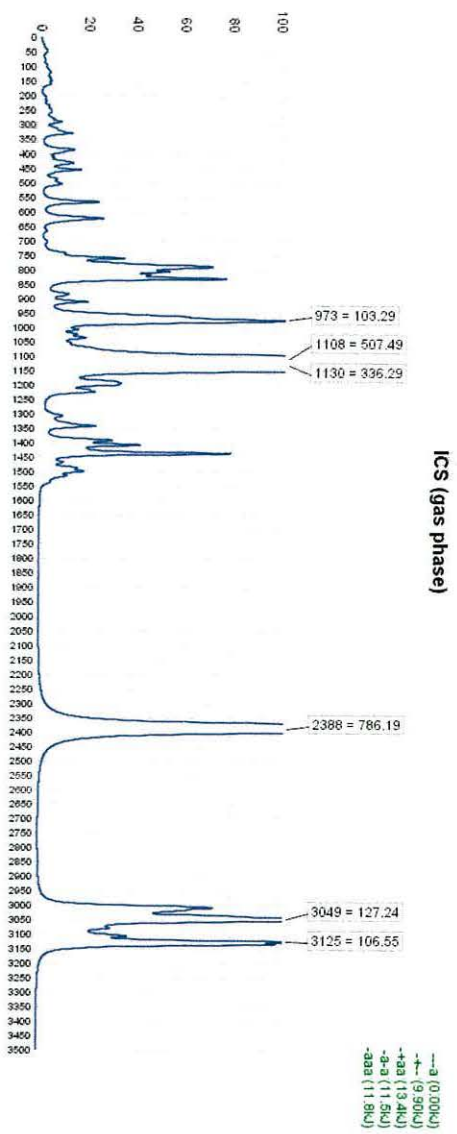


## Calculation results.

### AIM and Spectra.

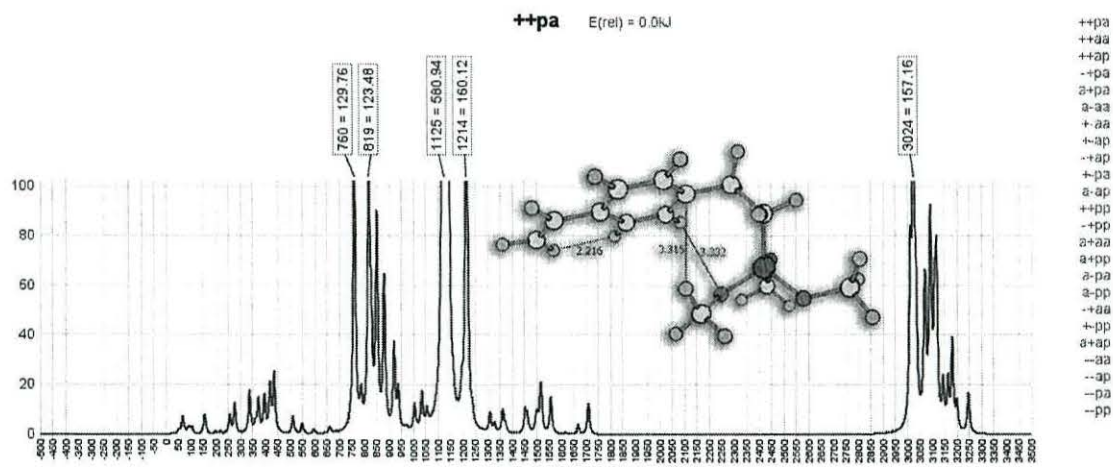
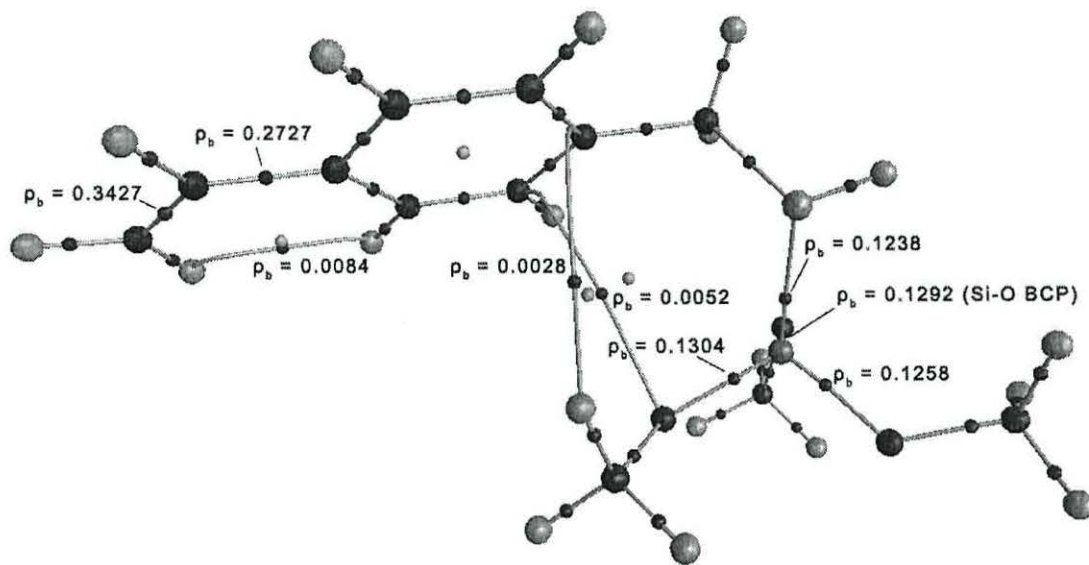
Spectra averaged over conformer classes. The green list in the top-right corner describes various conformational classes with their corresponding relative energies.



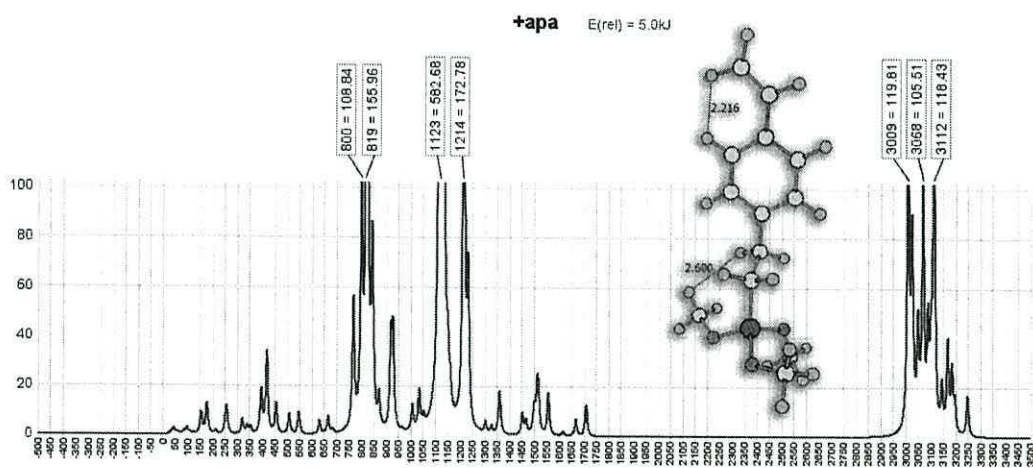
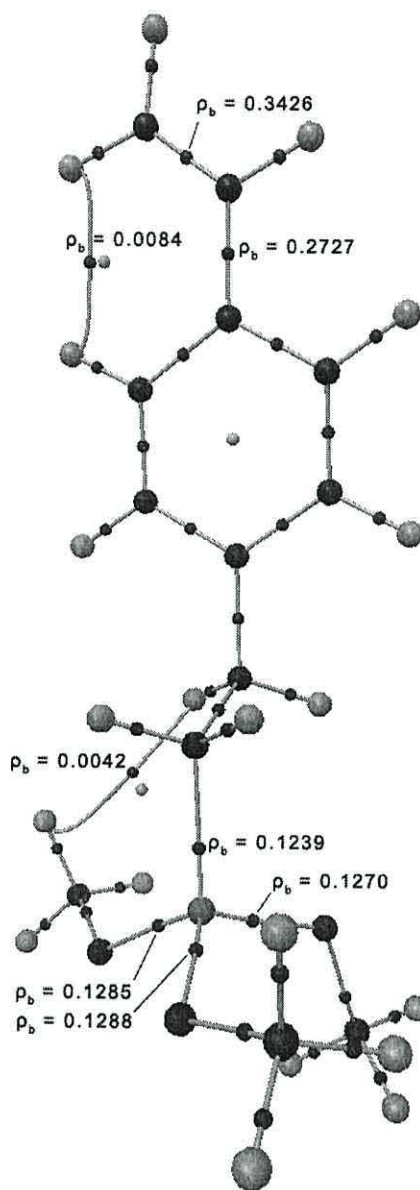


# STYRX

## STYRX, gas phase “++pa”



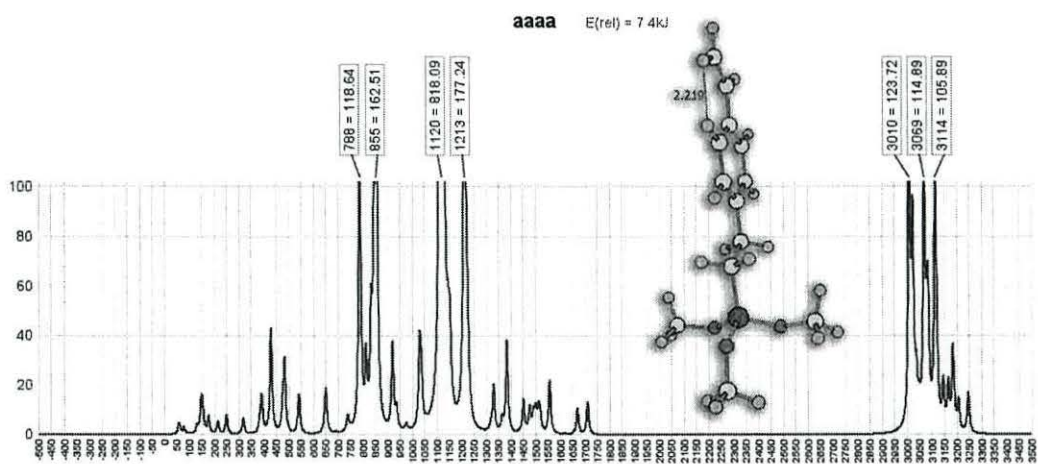
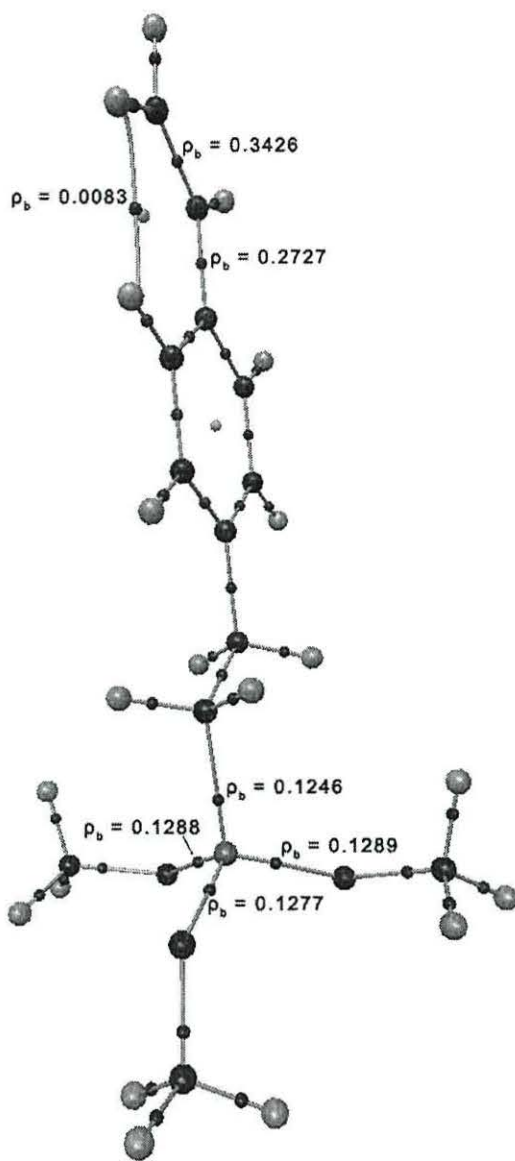
STYRX, gas phase "+apa"



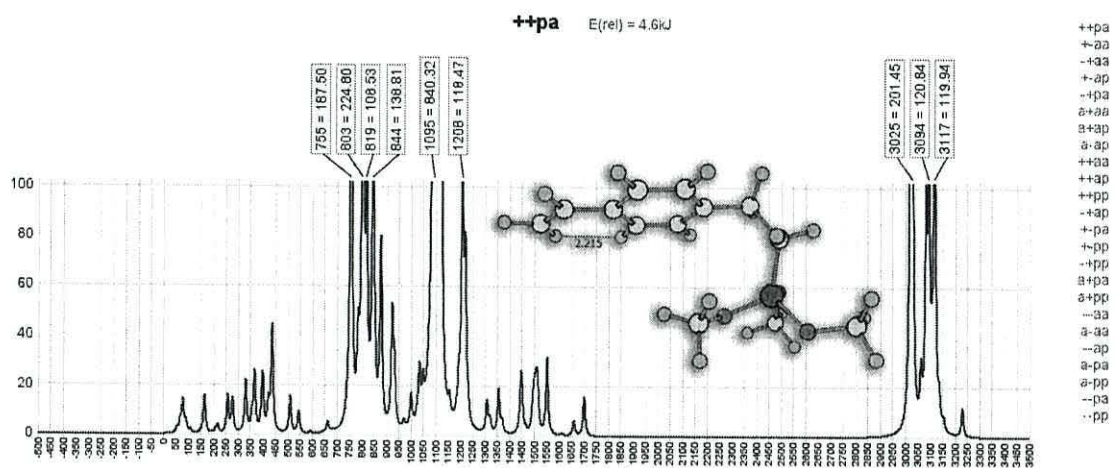
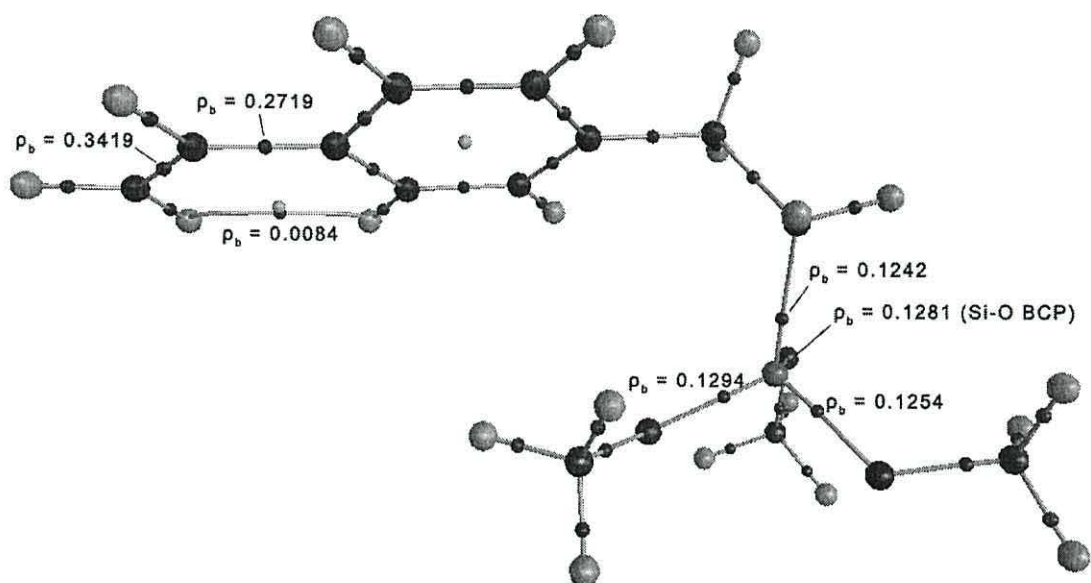
+apa  
+aaa  
-apa  
+app  
-aaa  
-pap  
-app  
+asp  
asp  
app  
aaa



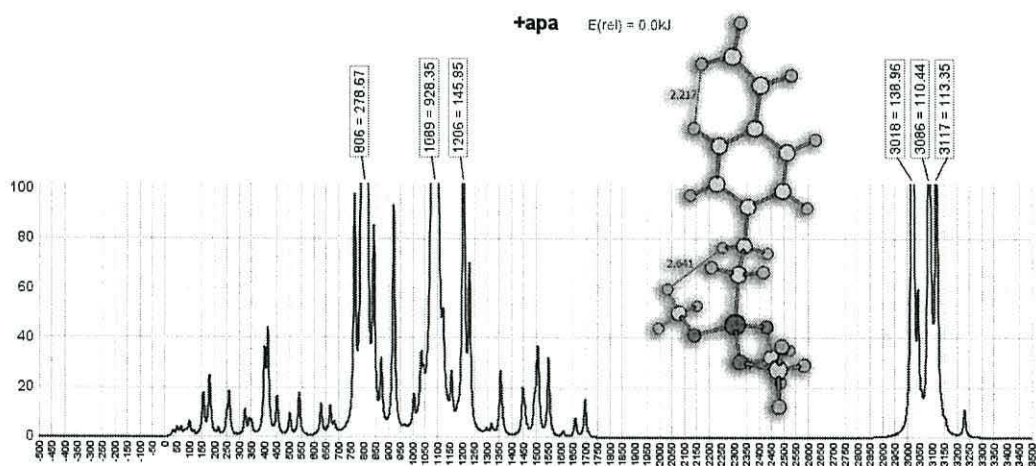
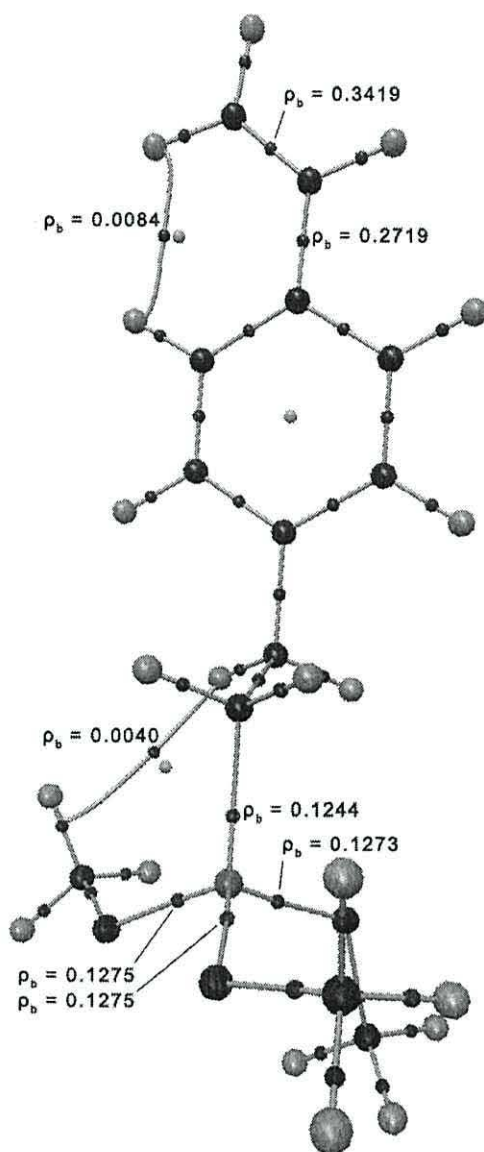
STYRX, gas phase "aaa"



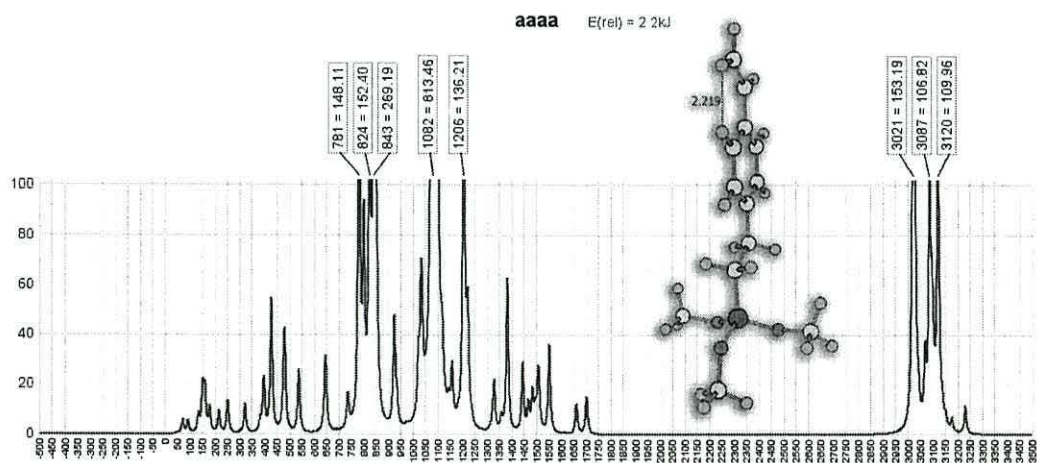
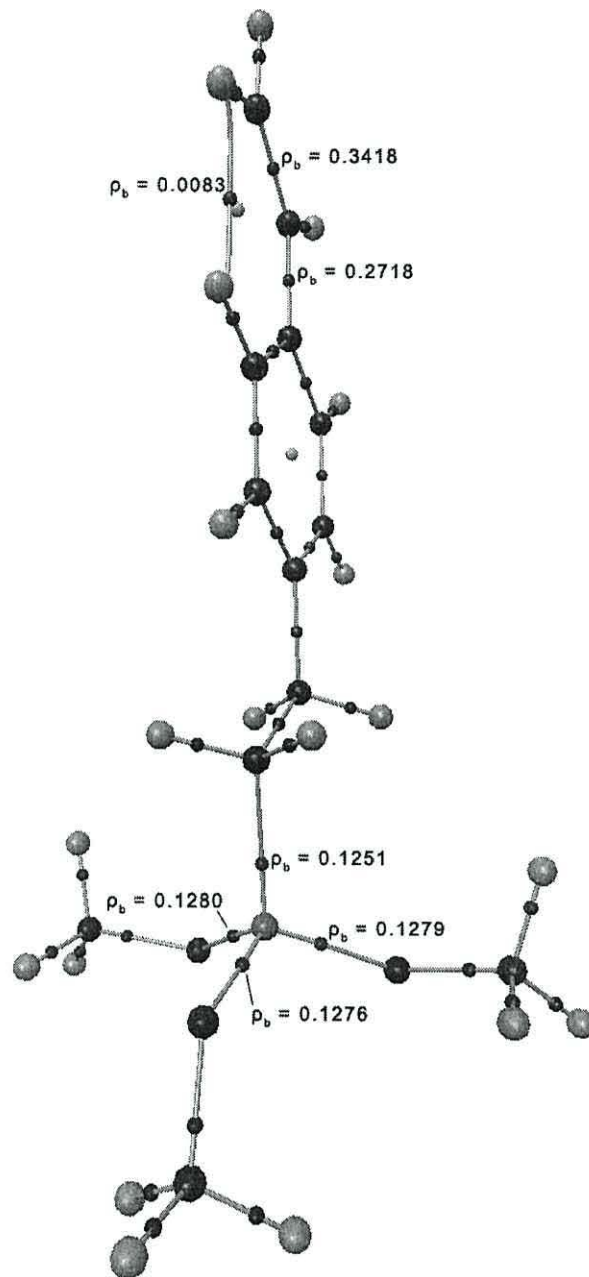
STYRX, PCM-H<sub>2</sub>O “++pa”



STYRX, PCM-H<sub>2</sub>O “+apa”



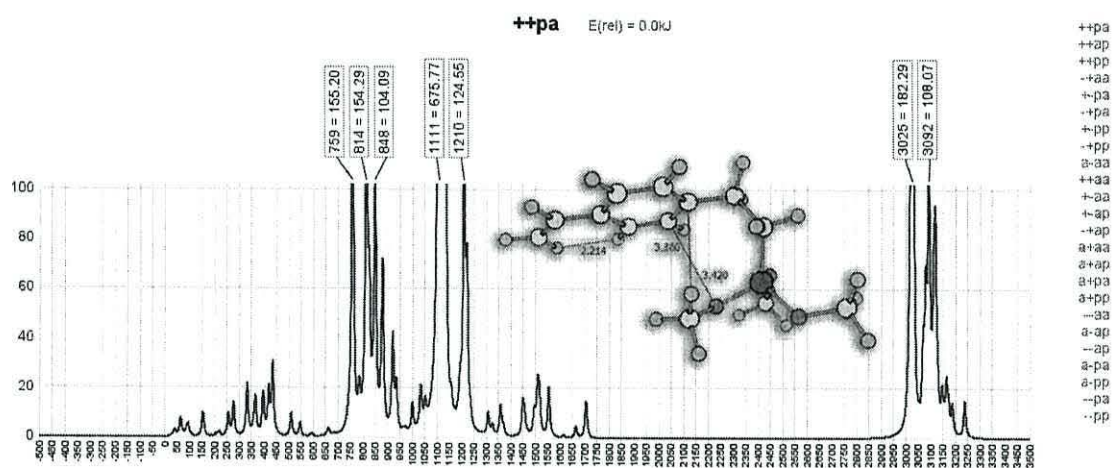
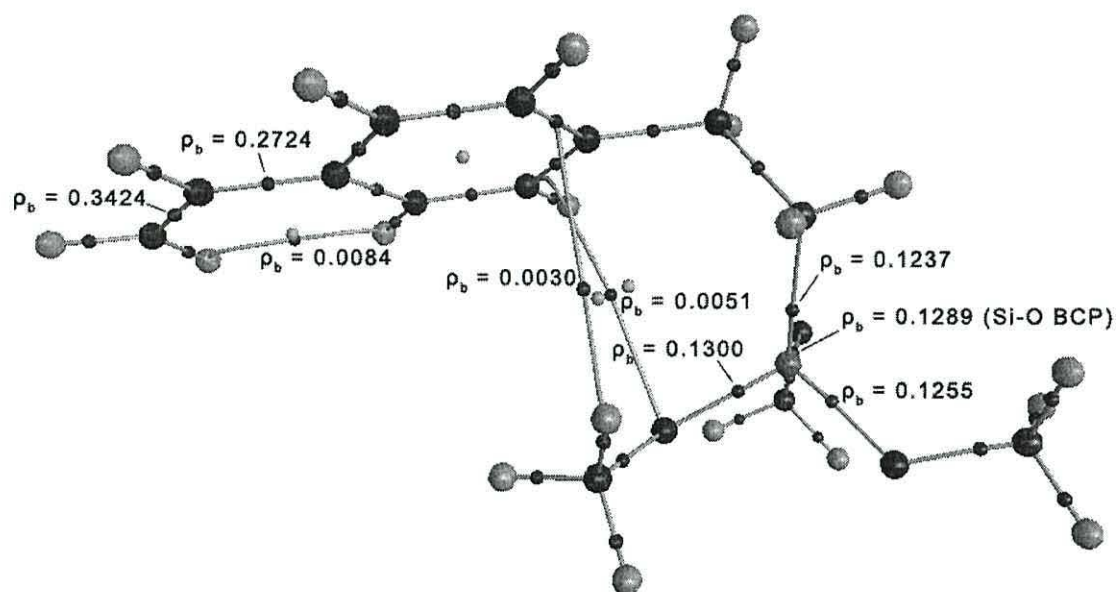
+apa  
 +app  
 -app  
 app  
 app  
 -apa  
 -app

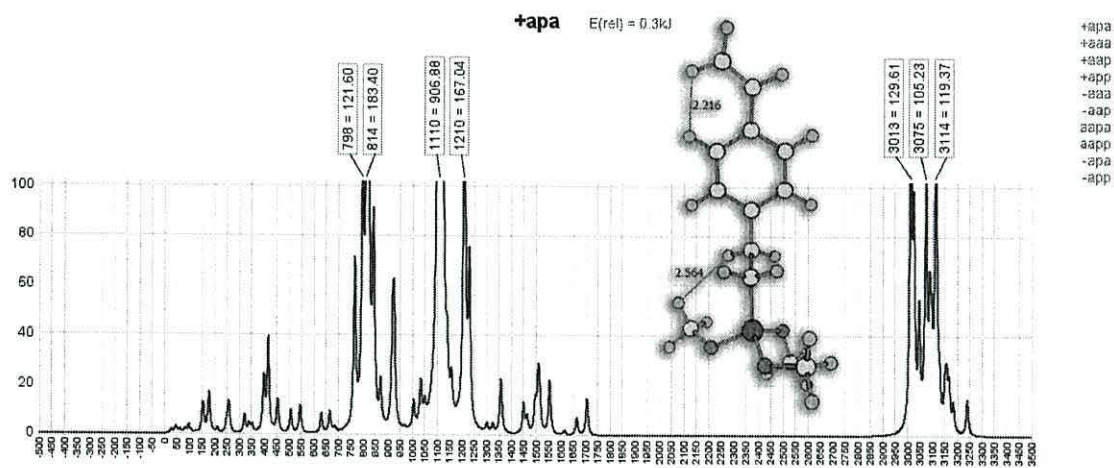
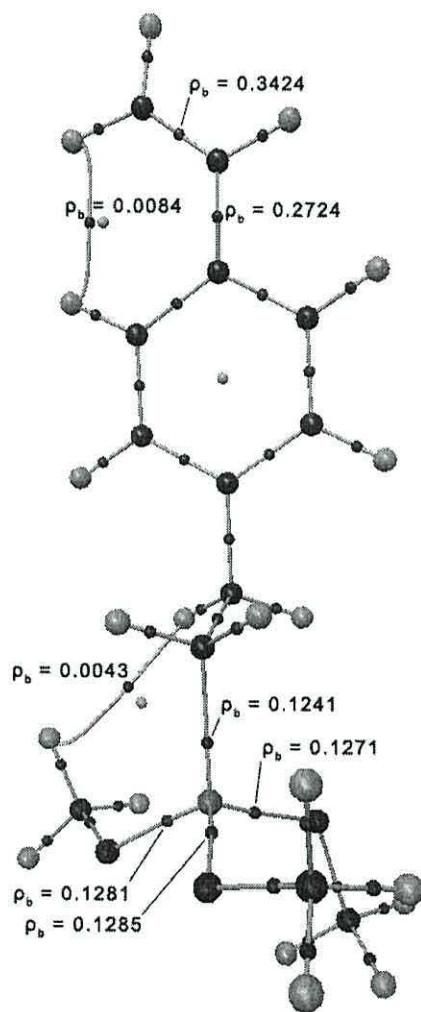


aaaa  
+aaa  
+aap  
-aaa  
aaap

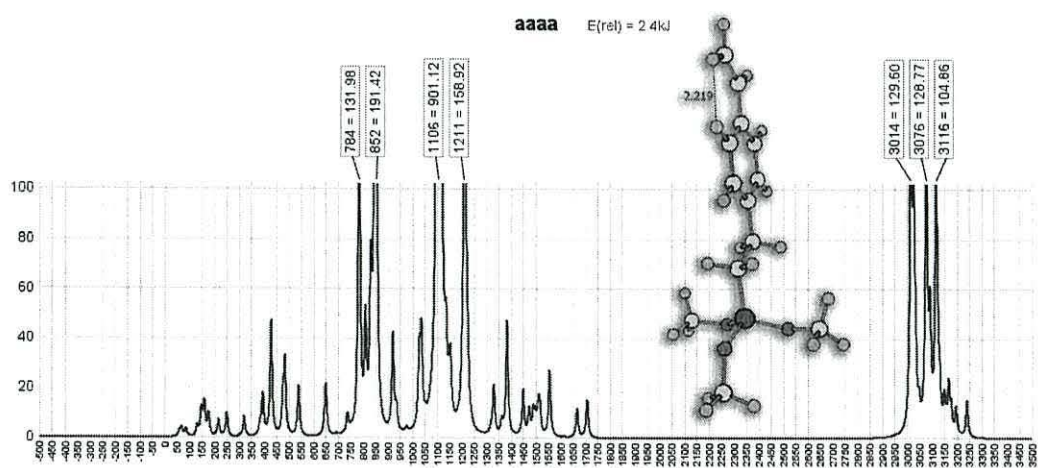
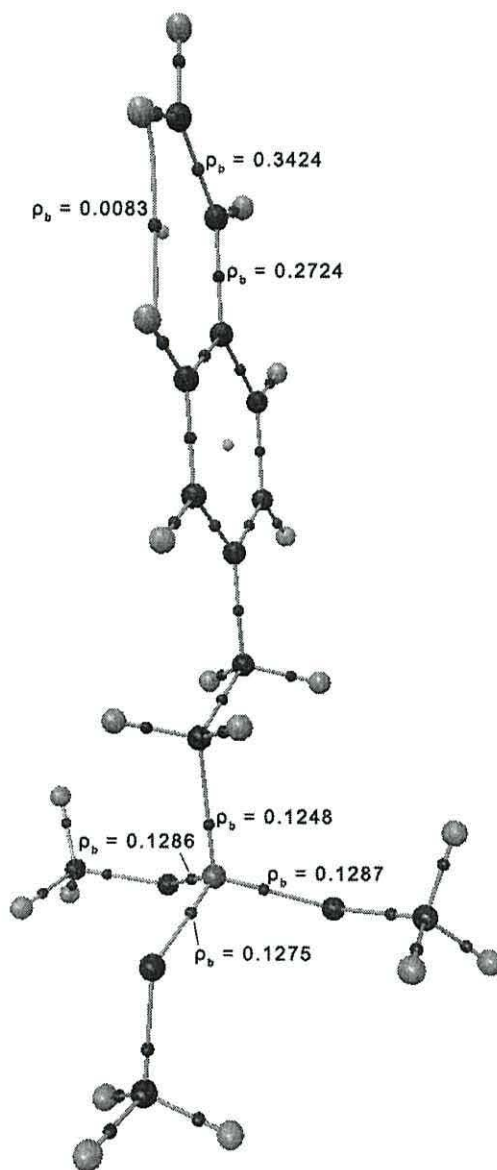


STYRX, PCM-CCl<sub>4</sub> “++pa”

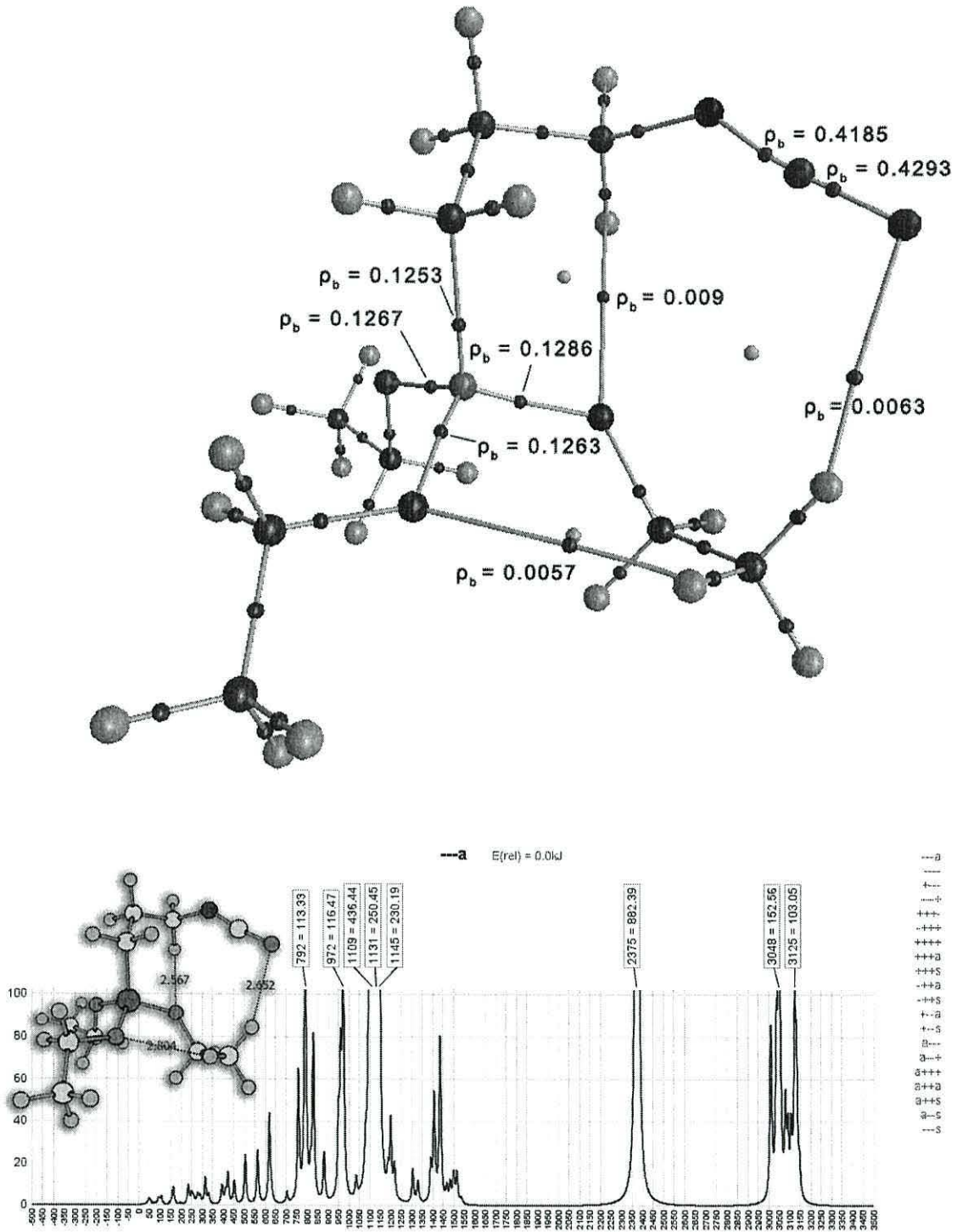




STYRX, PCM-CCl<sub>4</sub> "aaa"



ICS, gas phase “---a”

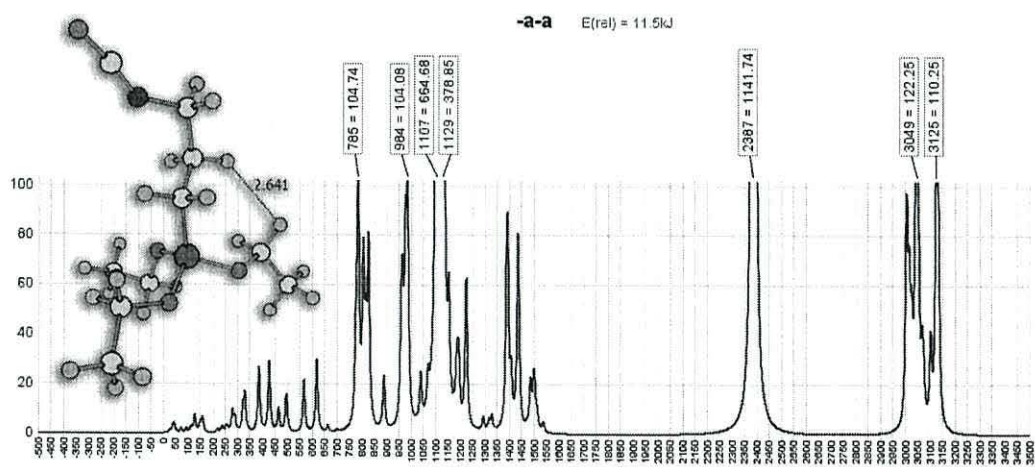
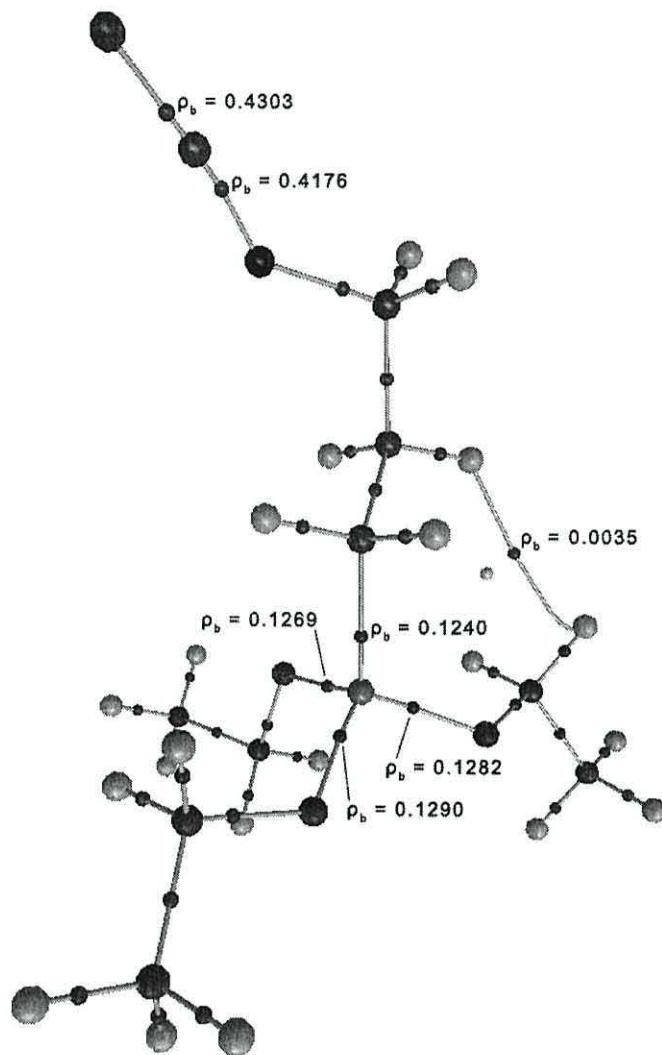






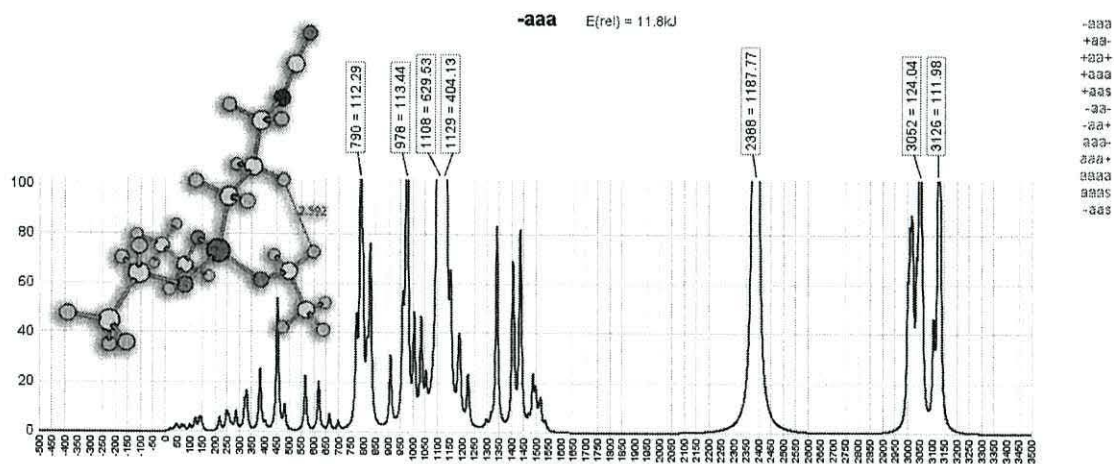
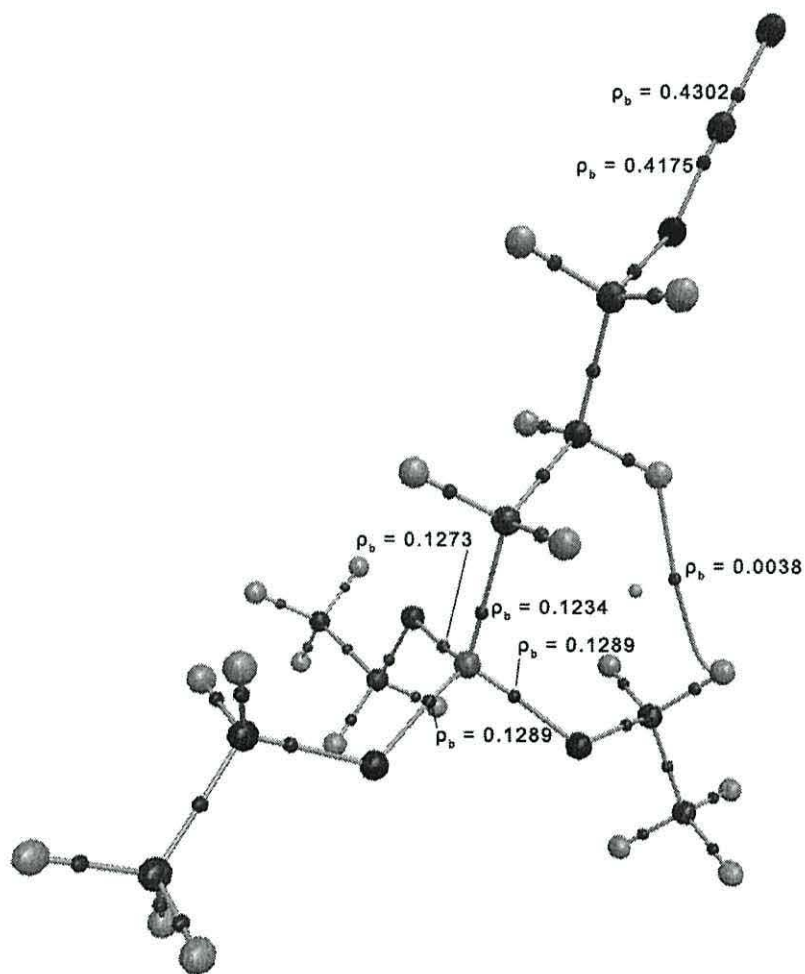


ICS, gas phase “-a-a”

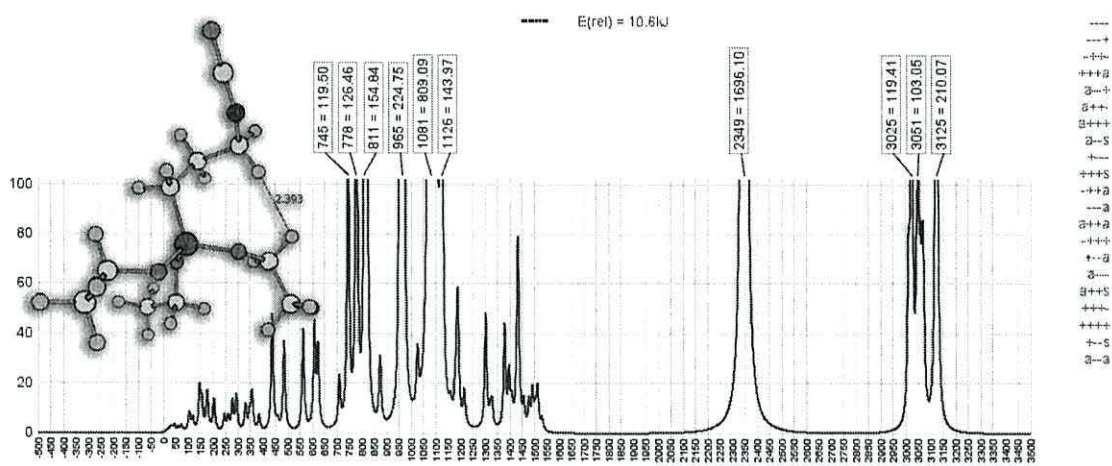
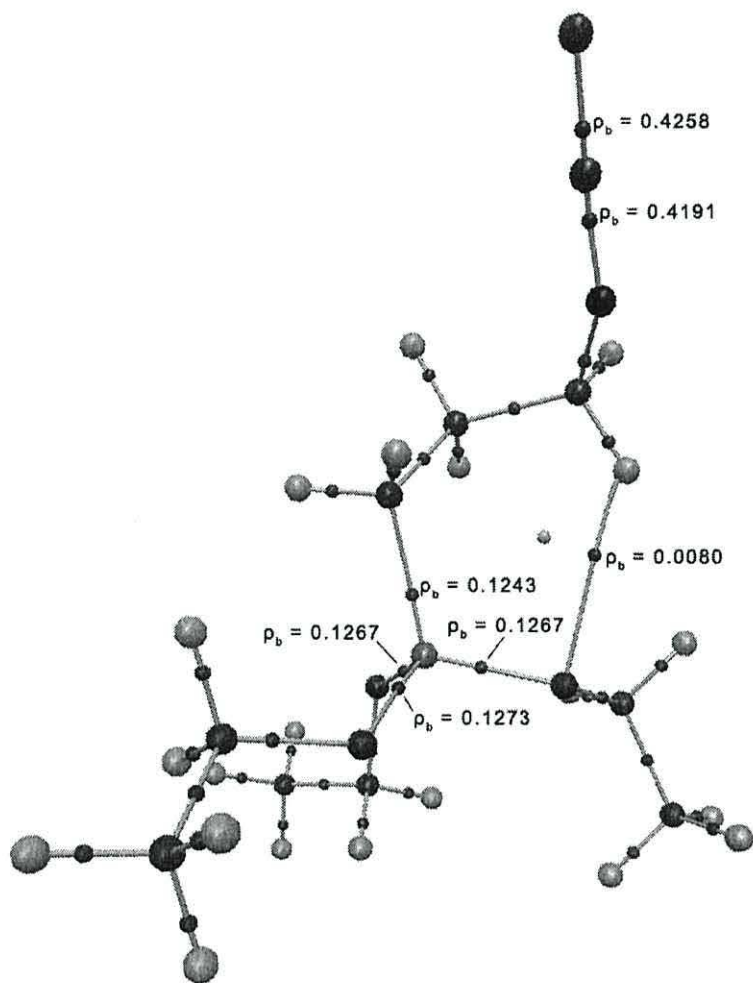


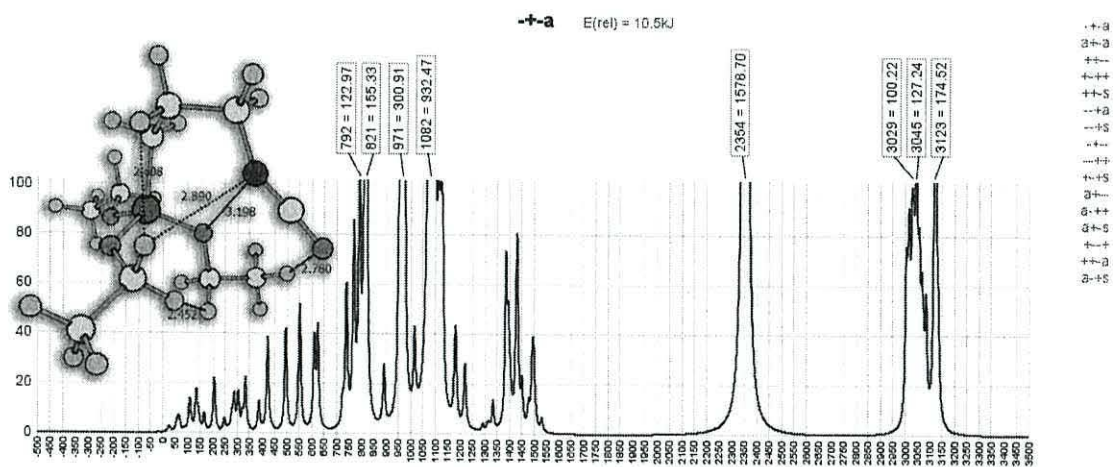
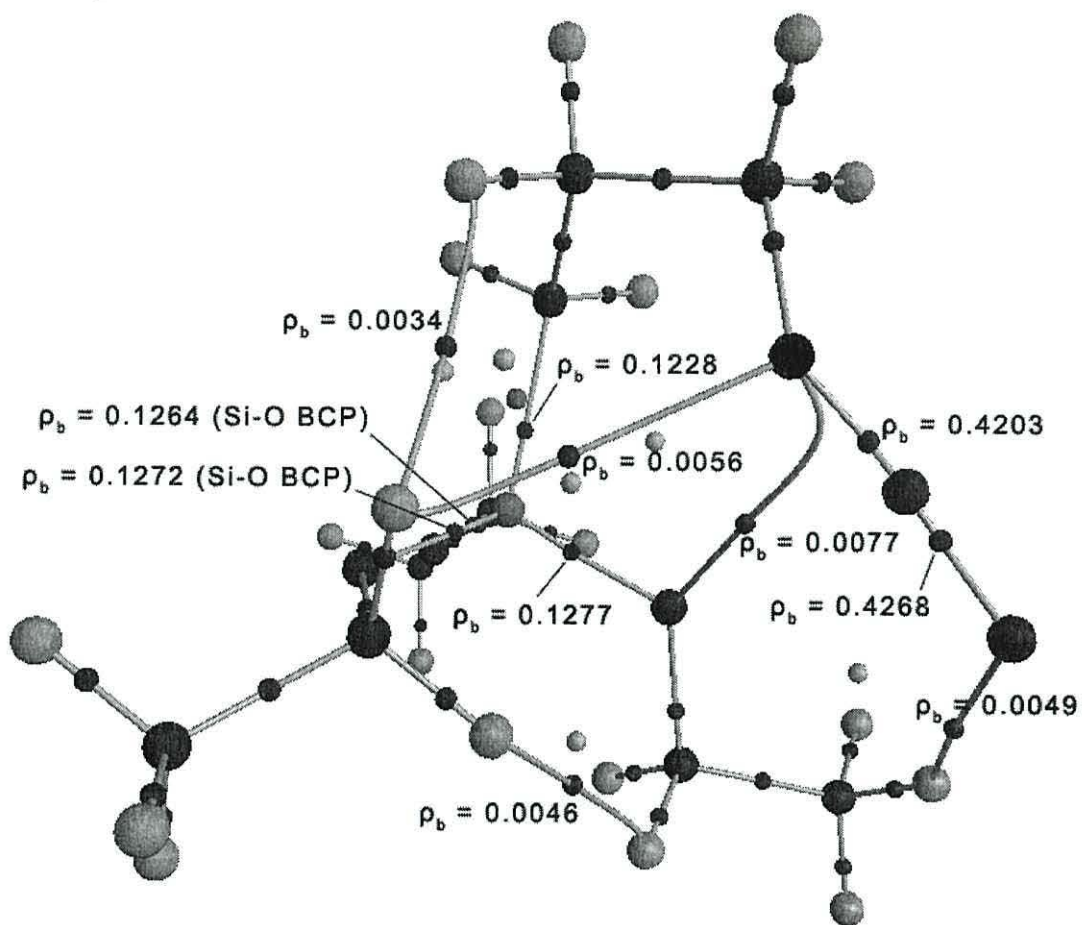
-a-a  
 +a--  
 +a+-  
 +a++  
 +a+a  
 +a+s  
 +a-s  
 -a-  
 -a+-  
 -a++  
 -a+a  
 -a+s  
 -a-s  
 aa--  
 aa+-  
 aa++  
 aa+a  
 aa+s  
 aa-s  
 -a-s

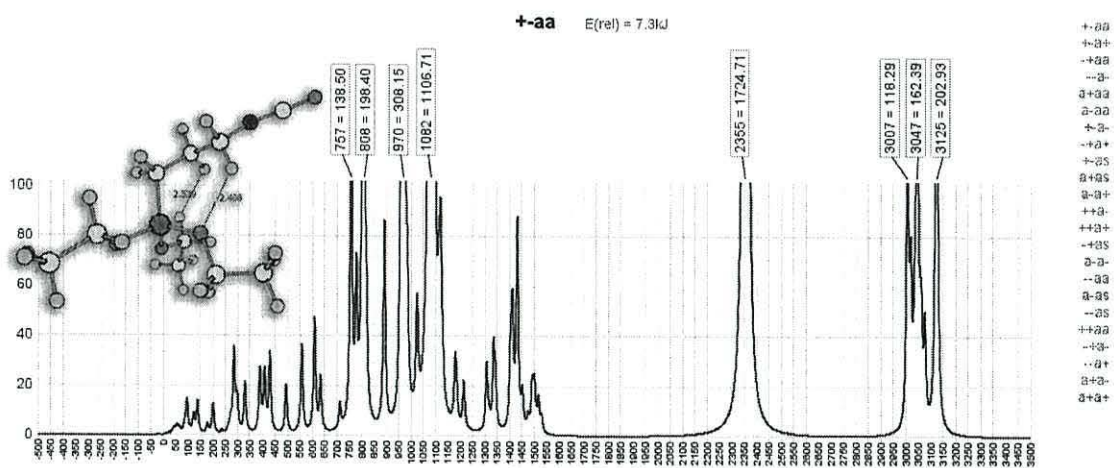
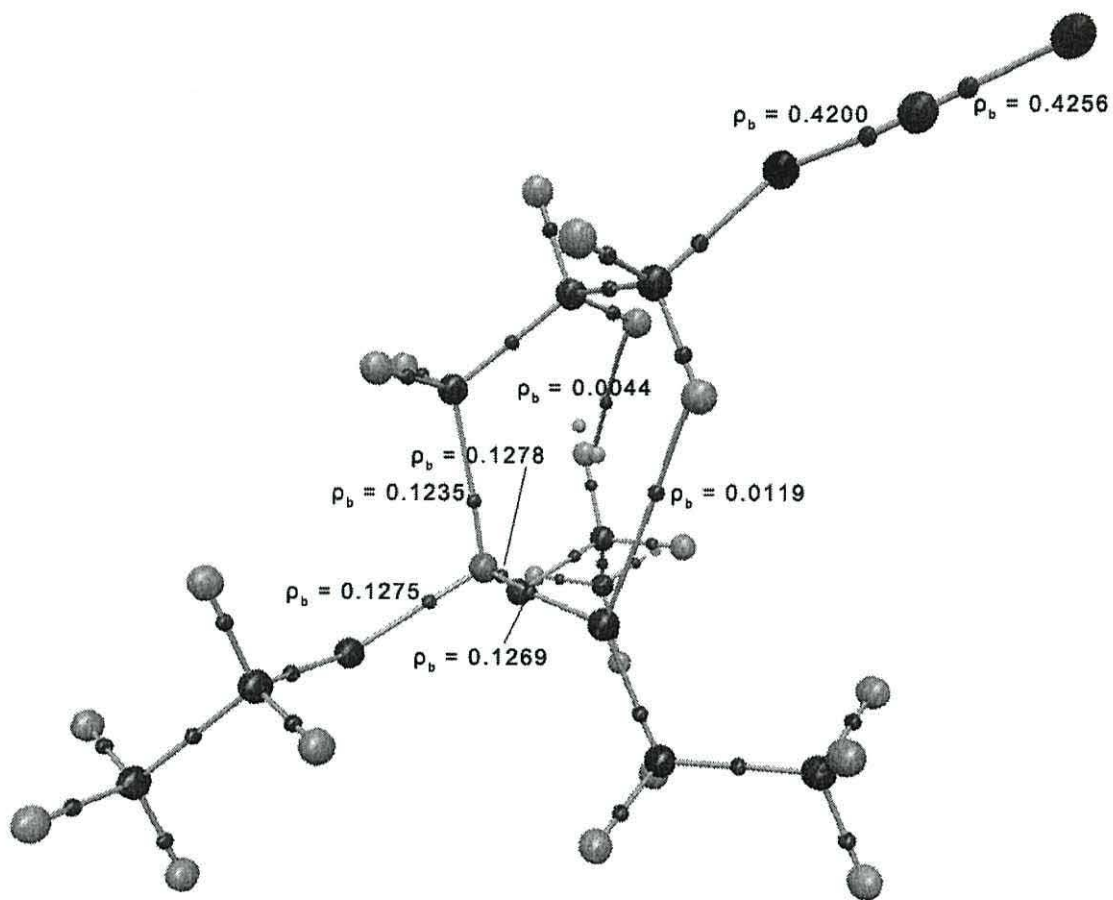
ICS, gas phase “-aaa”

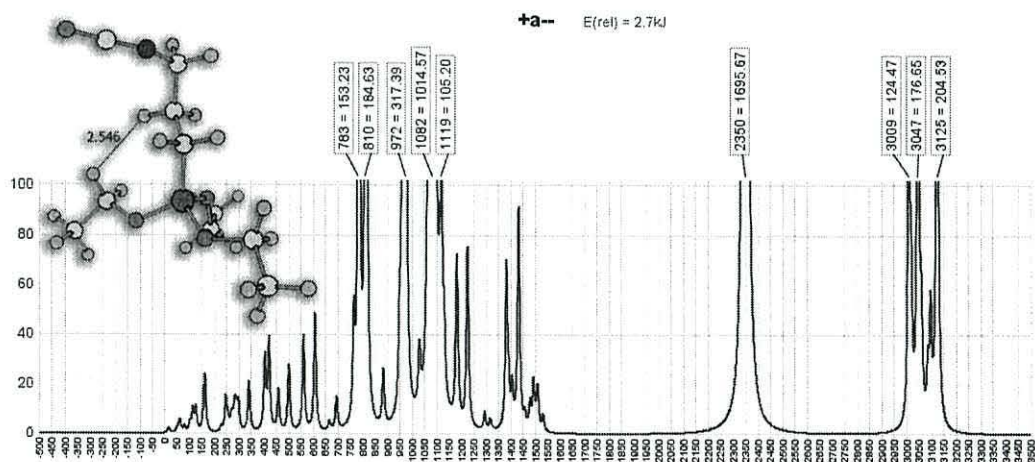
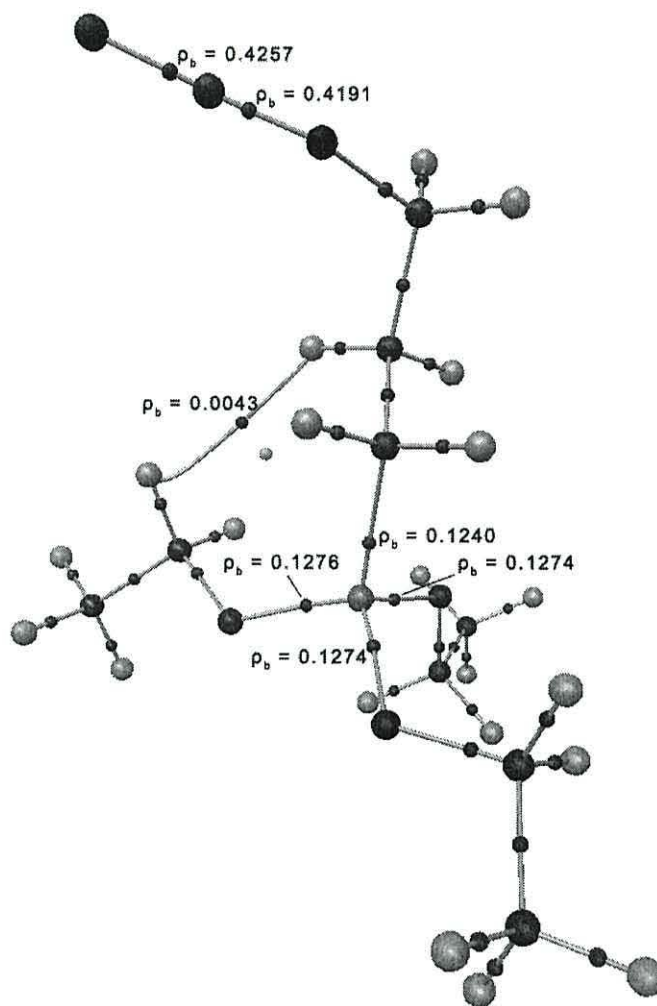






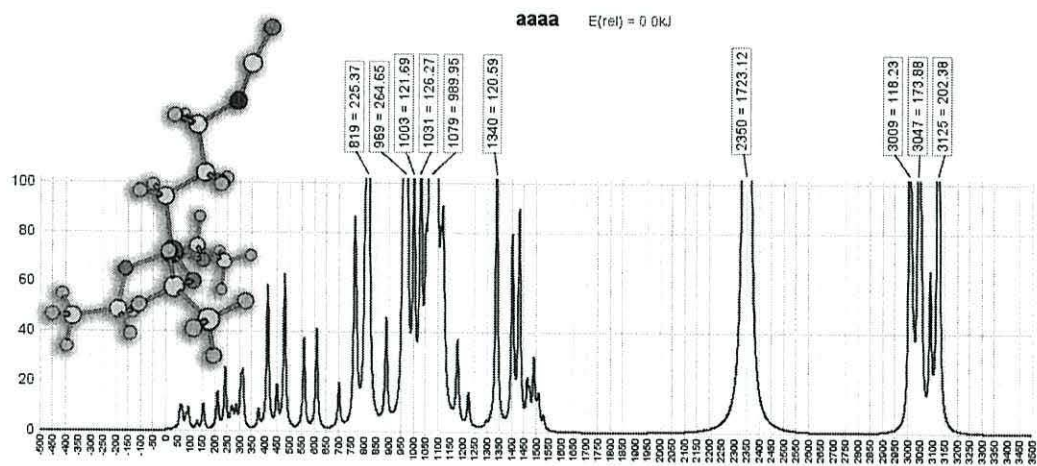
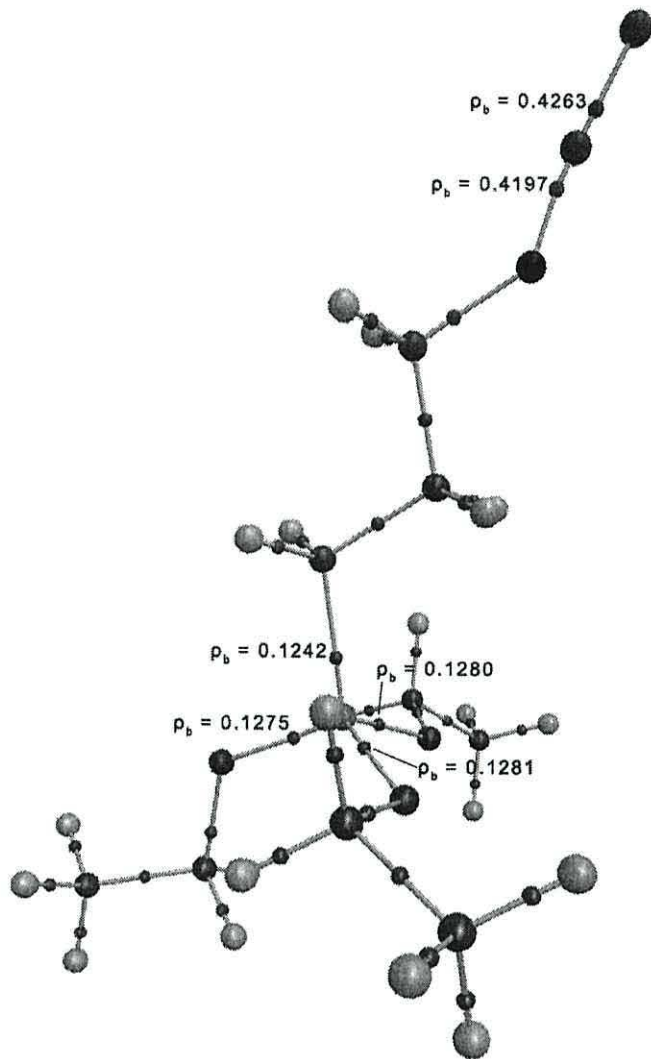




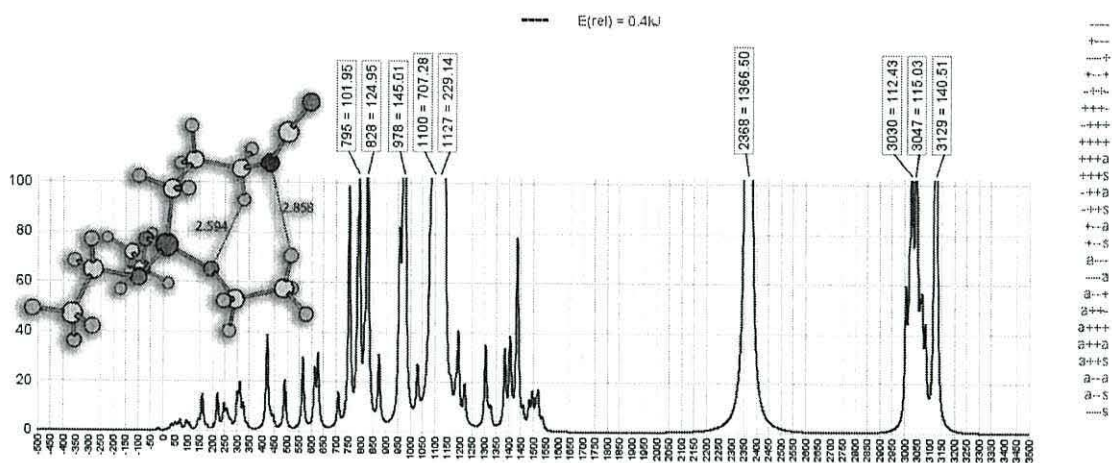
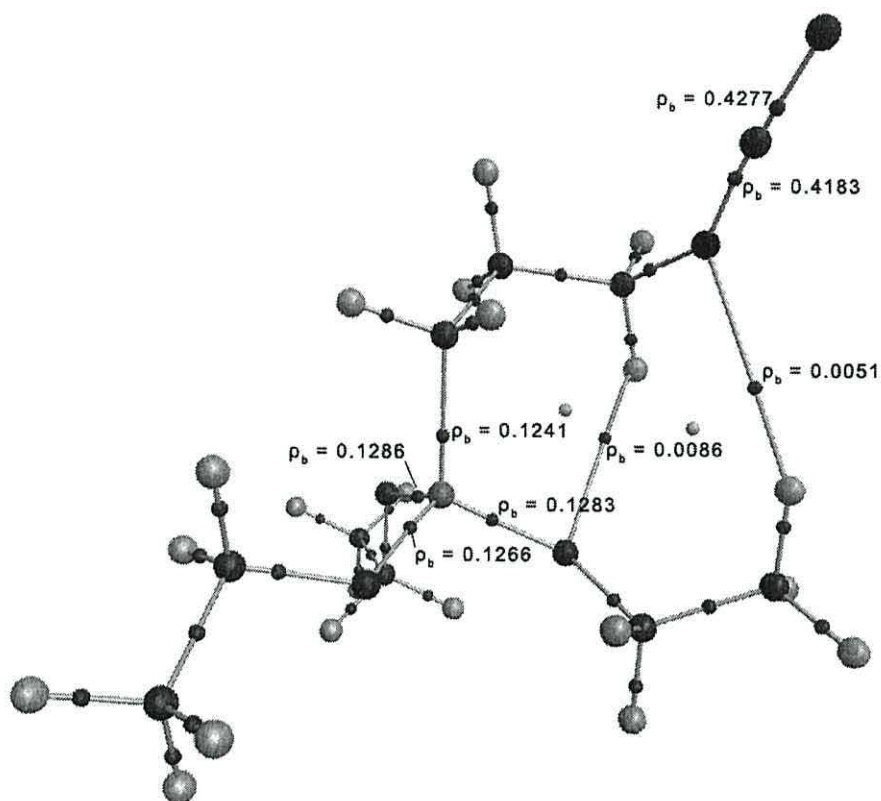


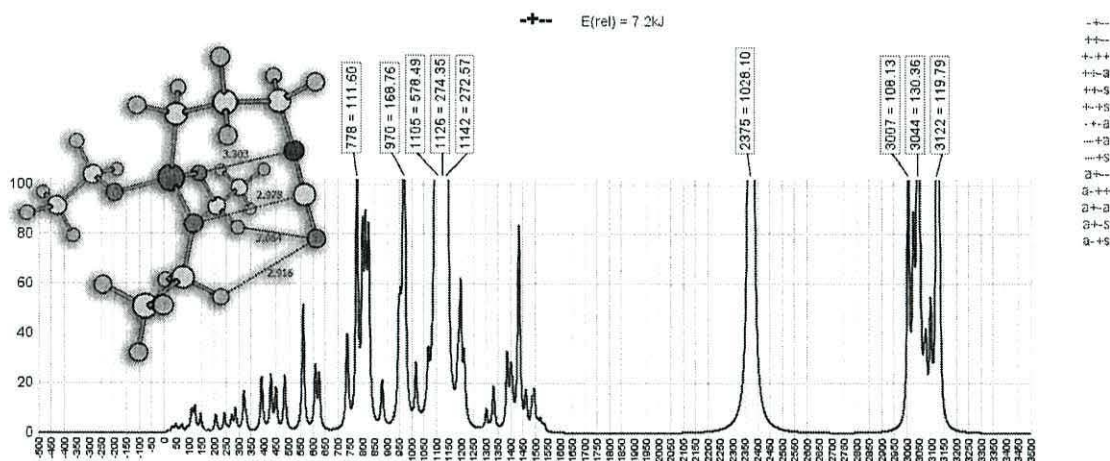
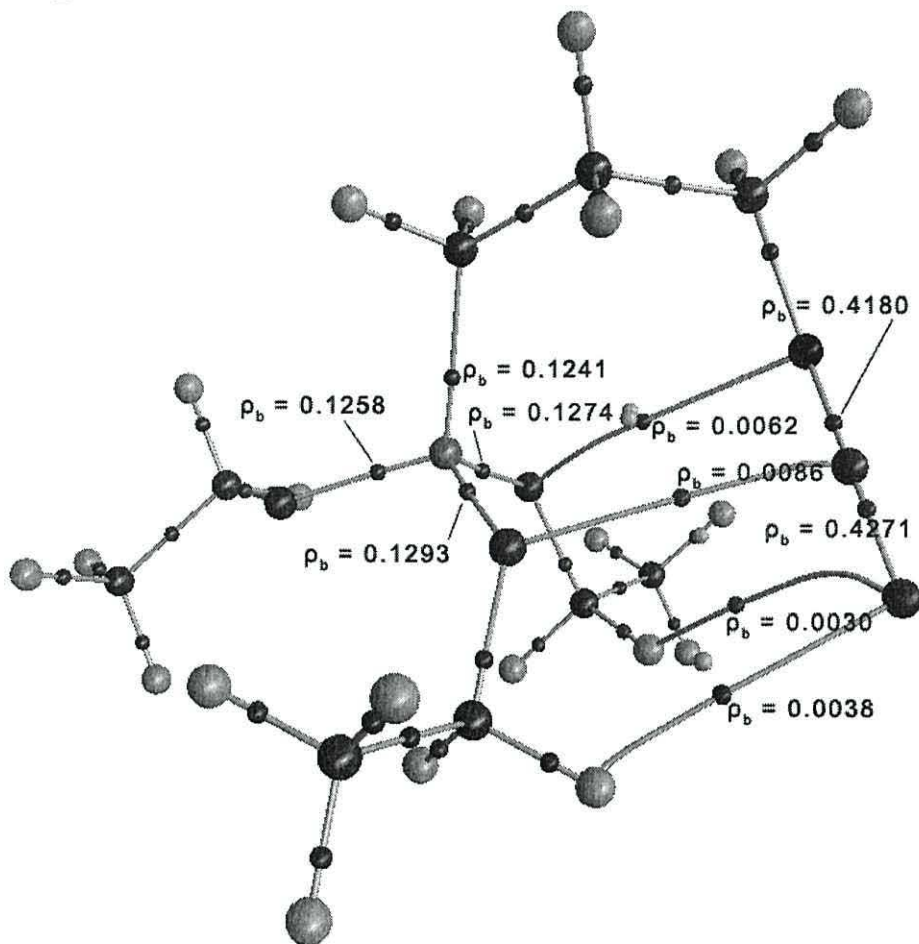
+a--  
 +a+S  
 +a-S  
 -a+S  
 -a-S  
 aa+  
 aa+  
 -a+  
 -a-  
 -a-  
 aa+S  
 -a-S  
 +a+  
 +a-  
 -a+  
 +a+  
 +a+  
 -a-  
 -a+  
 aa+  
 aa+  
 aa-

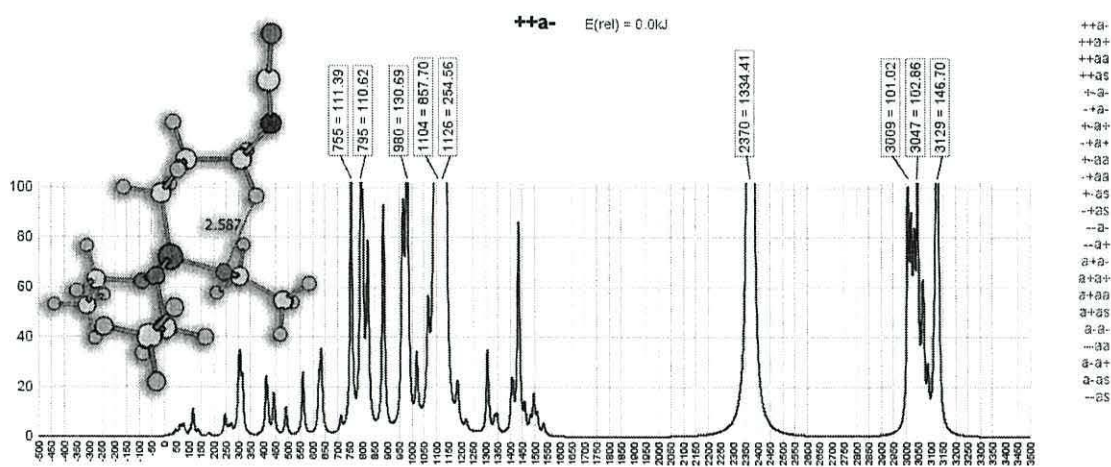
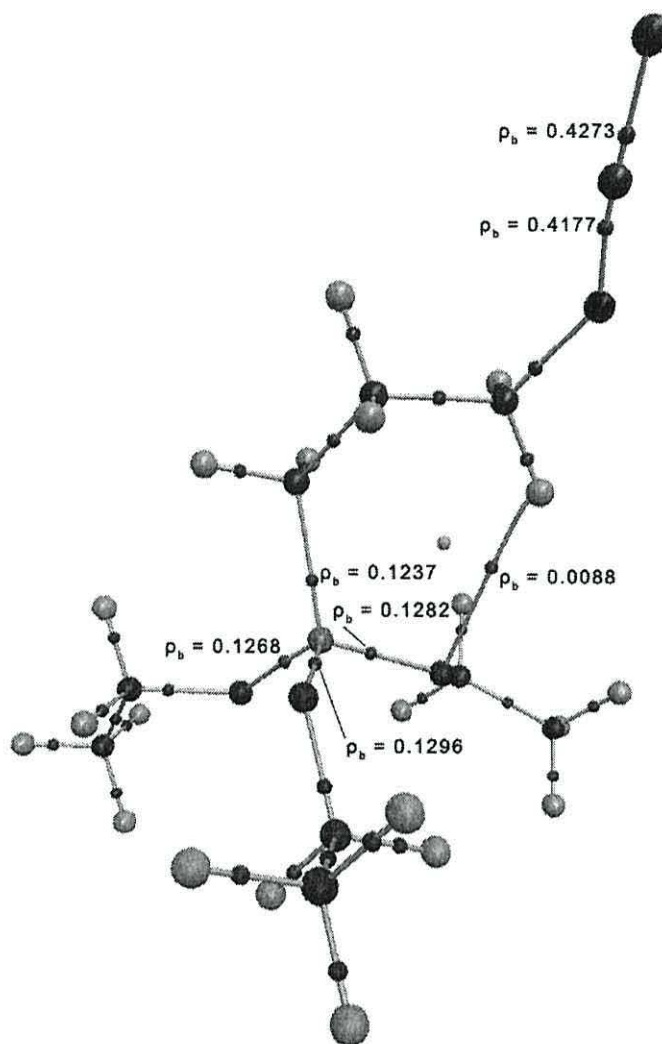




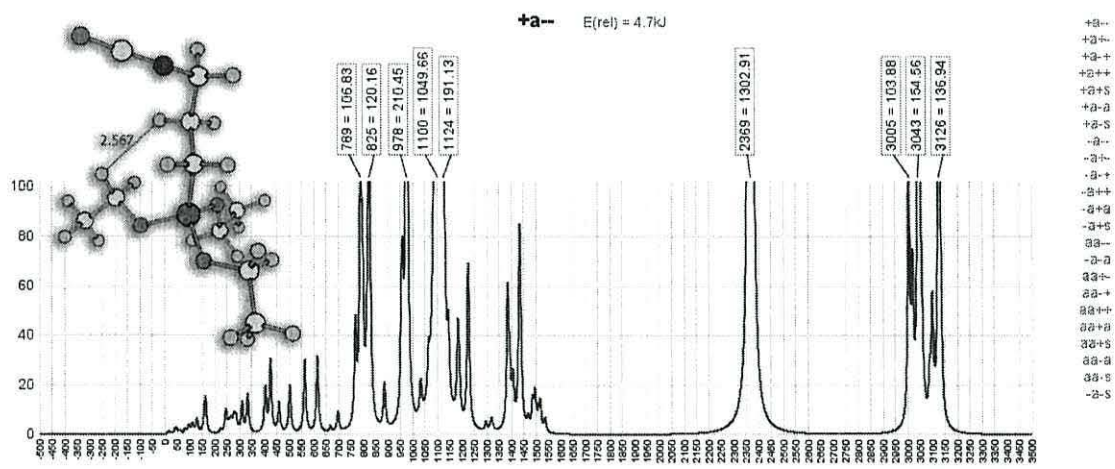
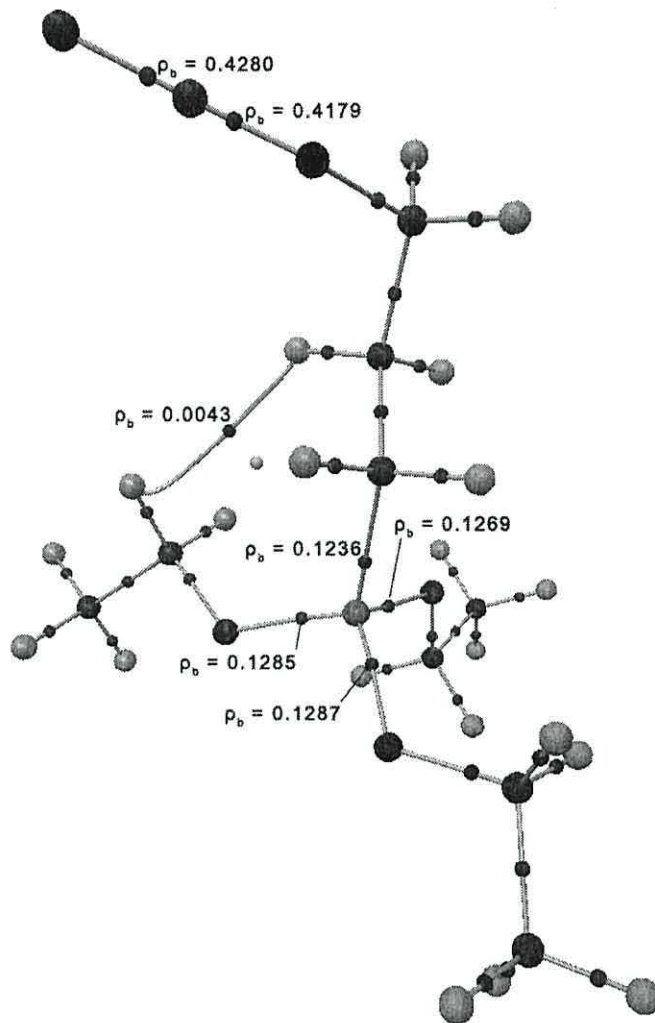
6666  
 -33-  
 8888  
 +33-  
 +88+  
 -88+  
 888+  
 +885  
 -885

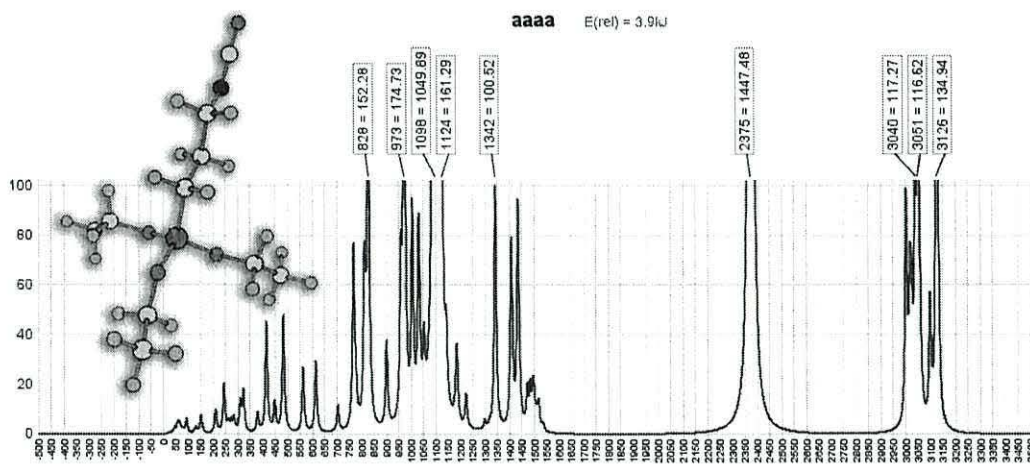
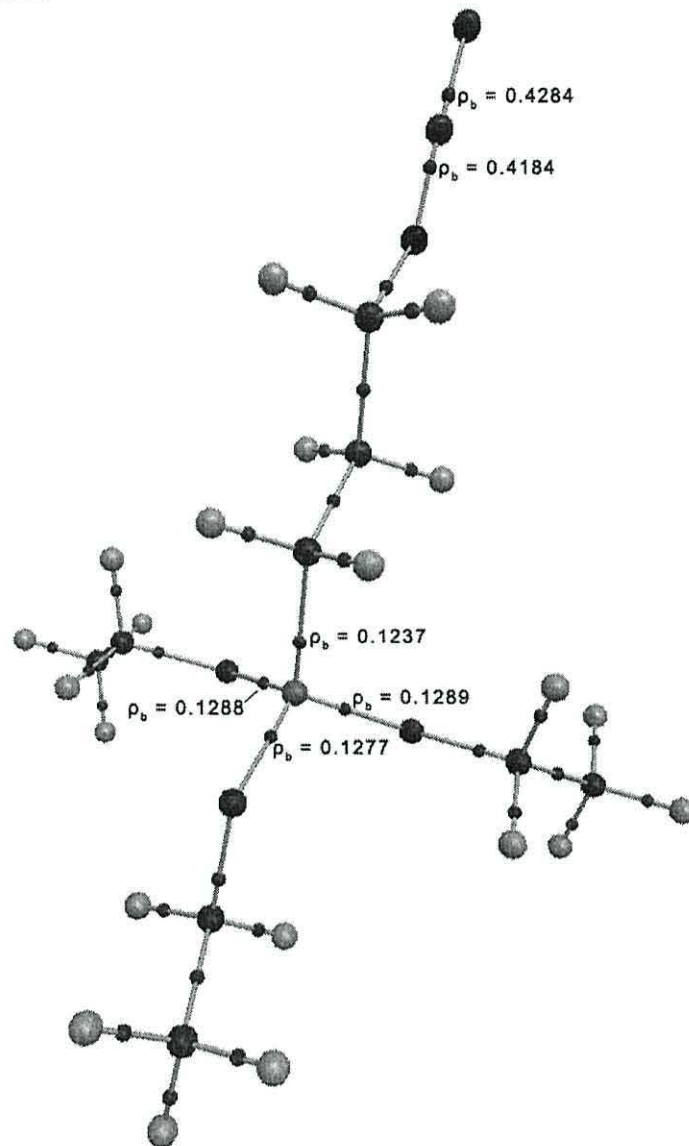












aaaa  
 +aa-  
 +aaa+  
 +aaa-  
 -aa-  
 -aa+  
 -aa-  
 -aa-  
 aaaa+  
 aaaa-  
 -aa-

## Thermodynamic Parameters.

### STYRX

#### STYRX, gas phase “++pa”

Zero-point correction= 0.309295 (Hartree/Particle)  
Thermal correction to Energy= 0.330170  
Thermal correction to Enthalpy= 0.331114  
Thermal correction to Gibbs Free Energy= 0.255924  
Sum of electronic and zero-point Energies= -1022.411835  
Sum of electronic and thermal Energies= -1022.390960  
Sum of electronic and thermal Enthalpies= -1022.390016  
Sum of electronic and thermal Free Energies= -1022.465206

#### STYRX, gas phase “+apa”

Zero-point correction= 0.308815 (Hartree/Particle)  
Thermal correction to Energy= 0.330003  
Thermal correction to Enthalpy= 0.330947  
Thermal correction to Gibbs Free Energy= 0.253747  
Sum of electronic and zero-point Energies= -1022.410393  
Sum of electronic and thermal Energies= -1022.389205  
Sum of electronic and thermal Enthalpies= -1022.388260  
Sum of electronic and thermal Free Energies= -1022.465461

#### STYRX, gas phase “aaaa”

Zero-point correction= 0.308298 (Hartree/Particle)  
Thermal correction to Energy= 0.327061  
Thermal correction to Enthalpy= 0.328005  
Thermal correction to Gibbs Free Energy= 0.257989  
Sum of electronic and zero-point Energies= -1022.410028  
Sum of electronic and thermal Energies= -1022.391264  
Sum of electronic and thermal Enthalpies= -1022.390320  
Sum of electronic and thermal Free Energies= -1022.460337

#### STYRX, PCM-H<sub>2</sub>O “++pa”

Zero-point correction= 0.307795 (Hartree/Particle)  
Thermal correction to Energy= 0.327897  
Thermal correction to Enthalpy= 0.328841  
Thermal correction to Gibbs Free Energy= 0.255807  
Sum of electronic and zero-point Energies= -1022.422503  
Sum of electronic and thermal Energies= -1022.402402  
Sum of electronic and thermal Enthalpies= -1022.401457  
Sum of electronic and thermal Free Energies= -1022.474492

#### STYRX, PCM-H<sub>2</sub>O “+apa”

Zero-point correction= 0.307502 (Hartree/Particle)  
Thermal correction to Energy= 0.327850  
Thermal correction to Enthalpy= 0.328794  
Thermal correction to Gibbs Free Energy= 0.254536  
Sum of electronic and zero-point Energies= -1022.424531  
Sum of electronic and thermal Energies= -1022.404183  
Sum of electronic and thermal Enthalpies= -1022.403239  
Sum of electronic and thermal Free Energies= -1022.477497

#### STYRX, PCM-H<sub>2</sub>O “aaaa”

Zero-point correction= 0.307218 (Hartree/Particle)  
Thermal correction to Energy= 0.326867  
Thermal correction to Enthalpy= 0.327812  
Thermal correction to Gibbs Free Energy= 0.254662  
Sum of electronic and zero-point Energies= -1022.423988  
Sum of electronic and thermal Energies= -1022.404338  
Sum of electronic and thermal Enthalpies= -1022.403394  
Sum of electronic and thermal Free Energies= -1022.476543

#### STYRX, PCM-CCl<sub>4</sub> “++pa”

Zero-point correction= 0.308828 (Hartree/Particle)  
Thermal correction to Energy= 0.329719  
Thermal correction to Enthalpy= 0.330663  
Thermal correction to Gibbs Free Energy= 0.255368  
Sum of electronic and zero-point Energies= -1022.414879  
Sum of electronic and thermal Energies= -1022.393988  
Sum of electronic and thermal Enthalpies= -1022.393044  
Sum of electronic and thermal Free Energies= -1022.468339

#### STYRX, PCM-CCl<sub>4</sub> “+apa”

Zero-point correction= 0.308306 (Hartree/Particle)  
Thermal correction to Energy= 0.329566  
Thermal correction to Enthalpy= 0.330510  
Thermal correction to Gibbs Free Energy= 0.252712  
Sum of electronic and zero-point Energies= -1022.415289  
Sum of electronic and thermal Energies= -1022.394029  
Sum of electronic and thermal Enthalpies= -1022.393085  
Sum of electronic and thermal Free Energies= -1022.470883

#### STYRX, PCM-CCl<sub>4</sub> “aaaa”

Zero-point correction= 0.308140 (Hartree/Particle)  
Thermal correction to Energy= 0.329476  
Thermal correction to Enthalpy= 0.330420  
Thermal correction to Gibbs Free Energy= 0.252310  
Sum of electronic and zero-point Energies= -1022.414652  
Sum of electronic and thermal Energies= -1022.393316  
Sum of electronic and thermal Enthalpies= -1022.392372  
Sum of electronic and thermal Free Energies= -1022.470481

### ICS

#### ICS, gas phase “---a”

Zero-point correction= 0.313881 (Hartree/Particle)  
Thermal correction to Energy= 0.335190  
Thermal correction to Enthalpy= 0.336135  
Thermal correction to Gibbs Free Energy= 0.259035  
Sum of electronic and zero-point Energies= -1038.708628  
Sum of electronic and thermal Energies= -1038.687319  
Sum of electronic and thermal Enthalpies= -1038.686375  
Sum of electronic and thermal Free Energies= -1038.763475

#### ICS, gas phase “-+--”

Zero-point correction= 0.313659 (Hartree/Particle)  
Thermal correction to Energy= 0.335103  
Thermal correction to Enthalpy= 0.336047  
Thermal correction to Gibbs Free Energy= 0.258607  
Sum of electronic and zero-point Energies= -1038.705079  
Sum of electronic and thermal Energies= -1038.683634  
Sum of electronic and thermal Enthalpies= -1038.682690  
Sum of electronic and thermal Free Energies= -1038.760131

#### ICS, gas phase “-+aa”

Zero-point correction= 0.313096 (Hartree/Particle)  
Thermal correction to Energy= 0.334908  
Thermal correction to Enthalpy= 0.335852  
Thermal correction to Gibbs Free Energy= 0.254494  
Sum of electronic and zero-point Energies= -1038.704298  
Sum of electronic and thermal Energies= -1038.682486  
Sum of electronic and thermal Enthalpies= -1038.681542



Sum of electronic and thermal Free Energies= -1038.762900

#### ICS, gas phase “-a-a”

Zero-point correction= 0.312897 (Hartree/Particle)  
Thermal correction to Energy= 0.334767  
Thermal correction to Enthalpy= 0.335711  
Thermal correction to Gibbs Free Energy= 0.254325  
Sum of electronic and zero-point Energies= -1038.705220  
Sum of electronic and thermal Energies= -1038.683350  
Sum of electronic and thermal Enthalpies= -1038.682406  
Sum of electronic and thermal Free Energies= -1038.763792

#### ICS, gas phase “-aaa”

Zero-point correction= 0.312739 (Hartree/Particle)  
Thermal correction to Energy= 0.334708  
Thermal correction to Enthalpy= 0.335653  
Thermal correction to Gibbs Free Energy= 0.254740  
Sum of electronic and zero-point Energies= -1038.705292  
Sum of electronic and thermal Energies= -1038.683323  
Sum of electronic and thermal Enthalpies= -1038.682378  
Sum of electronic and thermal Free Energies= -1038.763291

#### ICS, PCM-H<sub>2</sub>O “----”

Zero-point correction= 0.312545 (Hartree/Particle)  
Thermal correction to Energy= 0.334081  
Thermal correction to Enthalpy= 0.335025  
Thermal correction to Gibbs Free Energy= 0.256493  
Sum of electronic and zero-point Energies= -1038.714851  
Sum of electronic and thermal Energies= -1038.693316  
Sum of electronic and thermal Enthalpies= -1038.692372  
Sum of electronic and thermal Free Energies= -1038.770903

#### ICS, PCM-H<sub>2</sub>O “-+-a”

Zero-point correction= 0.312226 (Hartree/Particle)  
Thermal correction to Energy= 0.332865  
Thermal correction to Enthalpy= 0.333809  
Thermal correction to Gibbs Free Energy= 0.259738  
Sum of electronic and zero-point Energies= -1038.715200  
Sum of electronic and thermal Energies= -1038.694560  
Sum of electronic and thermal Enthalpies= -1038.693616  
Sum of electronic and thermal Free Energies= -1038.767687

#### ICS, PCM-H<sub>2</sub>O “+-aa”

Zero-point correction= 0.312005 (Hartree/Particle)  
Thermal correction to Energy= 0.332941  
Thermal correction to Enthalpy= 0.333885  
Thermal correction to Gibbs Free Energy= 0.257497  
Sum of electronic and zero-point Energies= -1038.716643  
Sum of electronic and thermal Energies= -1038.695708  
Sum of electronic and thermal Enthalpies= -1038.694763  
Sum of electronic and thermal Free Energies= -1038.771152

#### ICS, PCM-H<sub>2</sub>O “+a--”

Zero-point correction= 0.312223 (Hartree/Particle)  
Thermal correction to Energy= 0.333900  
Thermal correction to Enthalpy= 0.334844  
Thermal correction to Gibbs Free Energy= 0.255650  
Sum of electronic and zero-point Energies= -1038.718166  
Sum of electronic and thermal Energies= -1038.696489  
Sum of electronic and thermal Enthalpies= -1038.695545  
Sum of electronic and thermal Free Energies= -1038.774739

#### ICS, PCM-H<sub>2</sub>O “aaaa”

Zero-point correction= 0.311732 (Hartree/Particle)  
Thermal correction to Energy= 0.333720  
Thermal correction to Enthalpy= 0.334664  
Thermal correction to Gibbs Free Energy= 0.254058  
Sum of electronic and zero-point Energies= -1038.719695  
Sum of electronic and thermal Energies= -1038.697707  
Sum of electronic and thermal Enthalpies= -1038.696763  
Sum of electronic and thermal Free Energies= -1038.777369

#### ICS, PCM-CCl<sub>4</sub> “----”

Zero-point correction= 0.313201 (Hartree/Particle)  
Thermal correction to Energy= 0.333852  
Thermal correction to Enthalpy= 0.334796  
Thermal correction to Gibbs Free Energy= 0.259150  
Sum of electronic and zero-point Energies= -1038.710662  
Sum of electronic and thermal Energies= -1038.690011  
Sum of electronic and thermal Enthalpies= -1038.689067  
Sum of electronic and thermal Free Energies= -1038.764713

#### ICS, PCM-CCl<sub>4</sub> “-+--”

Zero-point correction= 0.313095 (Hartree/Particle)  
Thermal correction to Energy= 0.334626  
Thermal correction to Enthalpy= 0.335570  
Thermal correction to Gibbs Free Energy= 0.257322  
Sum of electronic and zero-point Energies= -1038.708195  
Sum of electronic and thermal Energies= -1038.686664  
Sum of electronic and thermal Enthalpies= -1038.685720  
Sum of electronic and thermal Free Energies= -1038.763968

#### ICS, PCM-CCl<sub>4</sub> “++a-”

Zero-point correction= 0.312685 (Hartree/Particle)  
Thermal correction to Energy= 0.334588  
Thermal correction to Enthalpy= 0.335532  
Thermal correction to Gibbs Free Energy= 0.252854  
Sum of electronic and zero-point Energies= -1038.711331  
Sum of electronic and thermal Energies= -1038.689428  
Sum of electronic and thermal Enthalpies= -1038.688484  
Sum of electronic and thermal Free Energies= -1038.771162

#### ICS, PCM-CCl<sub>4</sub> “+a--”

Zero-point correction= 0.312838 (Hartree/Particle)  
Thermal correction to Energy= 0.334552  
Thermal correction to Enthalpy= 0.335496  
Thermal correction to Gibbs Free Energy= 0.255776  
Sum of electronic and zero-point Energies= -1038.709375  
Sum of electronic and thermal Energies= -1038.687662  
Sum of electronic and thermal Enthalpies= -1038.686717  
Sum of electronic and thermal Free Energies= -1038.766437

#### ICS, PCM-CCl<sub>4</sub> “aaaa”

Zero-point correction= 0.312127 (Hartree/Particle)  
Thermal correction to Energy= 0.333352  
Thermal correction to Enthalpy= 0.334296  
Thermal correction to Gibbs Free Energy= 0.254091  
Sum of electronic and zero-point Energies= -1038.710386  
Sum of electronic and thermal Energies= -1038.689161  
Sum of electronic and thermal Enthalpies= -1038.688217  
Sum of electronic and thermal Free Energies= -1038.768422



*Normal Modes of Vibration (within computational error of  $\pm 25 \text{ cm}^{-1}$ ).*

**STYRX**

Frequencies -- 3163.6113 3165.4537 3183.6793  
 Frequencies -- 3187.7073 3206.4795 3245.9727

STYRX, gas phase “+apa”

Frequencies -- 17.7003 27.6530 30.4408  
 Frequencies -- 38.1561 45.8859 54.5334  
 Frequencies -- 58.2662 64.7909 73.1680  
 Frequencies -- 85.7035 97.9445 110.1667  
 Frequencies -- 125.3505 154.9090 176.4077  
 Frequencies -- 180.8589 211.5427 251.8024  
 Frequencies -- 257.2921 320.5001 340.3913  
 Frequencies -- 353.3271 396.0873 414.5817  
 Frequencies -- 417.8182 455.4548 508.0931  
 Frequencies -- 547.7339 632.2097 656.7771  
 Frequencies -- 666.0938 685.7636 734.4241  
 Frequencies -- 767.5086 799.9743 818.5455  
 Frequencies -- 825.3576 842.7546 850.3108  
 Frequencies -- 871.8639 919.0496 927.6330  
 Frequencies -- 963.8161 965.4252 1007.1312  
 Frequencies -- 1034.1832 1034.6496 1052.8233  
 Frequencies -- 1117.6020 1122.7342 1126.4761  
 Frequencies -- 1149.3765 1159.5554 1184.0293  
 Frequencies -- 1185.9926 1186.8545 1210.6130  
 Frequencies -- 1210.8464 1214.6614 1216.9920  
 Frequencies -- 1231.1014 1235.2082 1238.2092  
 Frequencies -- 1303.4112 1327.3590 1359.6893  
 Frequencies -- 1361.6155 1365.0855 1452.5681  
 Frequencies -- 1466.6884 1472.9329 1497.3399  
 Frequencies -- 1500.7761 1503.5546 1509.3433  
 Frequencies -- 1511.6337 1512.6935 1514.4178  
 Frequencies -- 1515.5807 1518.7550 1518.9828  
 Frequencies -- 1556.6062 1616.5699 1667.7431  
 Frequencies -- 1709.2955 3008.2674 3010.3576  
 Frequencies -- 3023.9931 3030.5217 3048.1553  
 Frequencies -- 3067.0408 3067.9493 3070.1686  
 Frequencies -- 3088.6639 3100.6876 3110.4145  
 Frequencies -- 3112.6492 3115.0162 3144.9420  
 Frequencies -- 3164.7432 3167.6455 3167.9945  
 Frequencies -- 3185.0708 3196.7133 3247.1582

STYRX, PCM-H<sub>2</sub>O “++pa”

Frequencies -- -90.0077 18.1521 31.3137  
 Frequencies -- 39.3471 54.5266 57.7473  
 Frequencies -- 68.6469 77.8383 78.6958  
 Frequencies -- 89.9574 92.9299 108.3972  
 Frequencies -- 131.5534 163.9401 189.5710  
 Frequencies -- 209.4595 216.7787 257.7534  
 Frequencies -- 277.6032 331.1316 355.9051  
 Frequencies -- 366.2780 398.2945 415.1906  
 Frequencies -- 421.8747 436.3740 510.2385  
 Frequencies -- 545.4992 593.7012 648.5110  
 Frequencies -- 662.4584 682.6506 741.7263  
 Frequencies -- 755.2282 784.8513 802.5983  
 Frequencies -- 818.8281 843.8321 853.0408  
 Frequencies -- 876.5536 922.9394 931.9968  
 Frequencies -- 970.7483 973.7847 1000.6605  
 Frequencies -- 1025.6339 1034.1026 1048.4051  
 Frequencies -- 1093.7110 1098.9258 1117.5684  
 Frequencies -- 1123.0553 1155.4925 1181.5234  
 Frequencies -- 1182.7906 1183.3054 1192.8747  
 Frequencies -- 1205.0664 1207.0670 1207.9734  
 Frequencies -- 1219.2810 1231.0730 1234.9641  
 Frequencies -- 1309.2261 1321.3784 1354.8133  
 Frequencies -- 1356.5624 1370.4373 1445.9277  
 Frequencies -- 1451.4207 1466.6033 1493.1367  
 Frequencies -- 1496.6992 1499.5598 1500.9150  
 Frequencies -- 1502.6360 1505.8171 1507.9476  
 Frequencies -- 1510.5821 1511.7296 1514.7373  
 Frequencies -- 1551.6451 1611.1549 1660.3624  
 Frequencies -- 1701.8593 3016.6872 3018.1215  
 Frequencies -- 3024.7396 3026.2986 3027.5590  
 Frequencies -- 3062.1513 3082.1973 3082.9215  
 Frequencies -- 3093.9188 3094.5977 3105.3506  
 Frequencies -- 3113.5677 3117.5150 3118.1879  
 Frequencies -- 3123.1297 3135.5084 3138.0215  
 Frequencies -- 3144.8074 3154.2667 3229.8651

STYRX, gas phase “aaaa”

Frequencies -- -36.8717 -16.9644 -6.5410  
 Frequencies -- 21.5766 36.7375 46.6188  
 Frequencies -- 50.6433 57.9179 61.2307  
 Frequencies -- 79.1732 94.4880 102.6342  
 Frequencies -- 131.1071 148.3854 152.9946  
 Frequencies -- 176.0966 213.6900 248.5442  
 Frequencies -- 249.3522 301.6043 317.4650  
 Frequencies -- 383.5393 391.3335 413.3552  
 Frequencies -- 428.9788 476.0595 484.1679  
 Frequencies -- 541.8730 649.4001 651.7798  
 Frequencies -- 655.6297 691.8868 738.6343  
 Frequencies -- 750.2525 787.5304 812.5270  
 Frequencies -- 833.2033 844.6683 853.2435  
 Frequencies -- 856.1475 919.6457 935.6902  
 Frequencies -- 966.7499 975.1281 1029.8880  
 Frequencies -- 1035.7945 1036.7993 1056.0691  
 Frequencies -- 1120.0668 1120.4837 1144.9515  
 Frequencies -- 1148.6955 1152.7973 1184.8728  
 Frequencies -- 1186.2990 1188.2301 1200.5498  
 Frequencies -- 1210.0605 1213.5677 1214.2982  
 Frequencies -- 1221.9557 1226.4992 1240.0664  
 Frequencies -- 1318.5077 1329.5872 1330.9484  
 Frequencies -- 1364.4058 1383.4567 1450.5628  
 Frequencies -- 1472.5014 1475.5150 1490.3119  
 Frequencies -- 1498.4965 1500.9140 1504.8422  
 Frequencies -- 1509.9415 1512.7133 1513.8672  
 Frequencies -- 1514.5224 1517.6883 1519.9250  
 Frequencies -- 1556.7777 1616.2093 1668.6951  
 Frequencies -- 1709.8370 3009.7410 3011.5386  
 Frequencies -- 3023.0427 3024.8597 3037.8115  
 Frequencies -- 3049.2248 3068.5147 3069.2755  
 Frequencies -- 3080.1338 3085.6836 3113.4747  
 Frequencies -- 3113.7498 3117.7913 3145.4683

STYRX, PCM-H<sub>2</sub>O “+apa”

Frequencies -- -29.0016 23.2150 31.7267  
 Frequencies -- 35.2650 51.3892 53.5994  
 Frequencies -- 61.6945 65.7068 67.7321  
 Frequencies -- 82.3940 89.8689 101.6710  
 Frequencies -- 119.7966 156.0275 180.2998  
 Frequencies -- 184.2928 215.0826 250.7812  
 Frequencies -- 259.7658 324.0612 341.3582  
 Frequencies -- 350.0460 401.9978 414.4882  
 Frequencies -- 415.9453 451.9560 503.5816  
 Frequencies -- 541.2769 630.6372 656.0359  
 Frequencies -- 668.1729 684.5613 733.0756  
 Frequencies -- 765.1811 791.5926 805.5530  
 Frequencies -- 813.0073 841.7428 851.8675  
 Frequencies -- 872.6203 922.6740 923.0729  
 Frequencies -- 968.0188 971.7469 1005.3865  
 Frequencies -- 1029.5185 1035.4998 1045.0466  
 Frequencies -- 1087.7454 1093.3545 1123.0077  
 Frequencies -- 1124.5957 1157.3044 1179.9312  
 Frequencies -- 1180.5302 1182.8105 1203.9662  
 Frequencies -- 1204.0831 1206.7608 1209.6443  
 Frequencies -- 1229.3461 1232.1787 1234.3575  
 Frequencies -- 1299.3832 1320.0585 1354.5263  
 Frequencies -- 1356.8221 1363.8525 1446.7504  
 Frequencies -- 1454.8468 1466.6089 1492.9661  
 Frequencies -- 1493.8397 1496.9945 1501.8313  
 Frequencies -- 1504.9933 1505.3145 1506.6644  
 Frequencies -- 1508.7300 1510.0308 1511.3674  
 Frequencies -- 1551.1614 1610.9975 1660.0072  
 Frequencies -- 1699.7312 3016.8095 3018.2190  
 Frequencies -- 3019.7961 3026.7682 3044.0465  
 Frequencies -- 3059.2984 3083.1855 3086.1560  
 Frequencies -- 3091.9142 3095.4897 3103.1431



Frequencies --	3115.7521	3117.0681	3121.8983
Frequencies --	3125.0146	3126.9179	3139.2505
Frequencies --	3145.1937	3154.1378	3230.7341

STYRX, PCM-H<sub>2</sub>O “aaaa”

Frequencies --	-32.0996	-10.9739	15.1077
Frequencies --	21.3873	41.8808	48.5184
Frequencies --	52.2607	60.5563	68.2041
Frequencies --	71.2919	92.5071	102.7782
Frequencies --	134.6433	151.1164	159.0752
Frequencies --	178.3629	215.2948	250.5092
Frequencies --	251.7393	302.3362	321.8449
Frequencies --	382.4778	396.6431	413.4090
Frequencies --	426.9913	476.8465	481.4204
Frequencies --	537.8839	644.7474	647.8507
Frequencies --	652.8048	690.8393	735.8171
Frequencies --	751.0755	781.3599	800.0061
Frequencies --	824.1648	842.1295	843.8219
Frequencies --	854.2106	923.7247	934.4342
Frequencies --	972.9059	976.0275	1022.1310
Frequencies --	1032.8962	1036.2099	1049.2383
Frequencies --	1081.8219	1093.4815	1118.1145
Frequencies --	1144.2512	1157.5677	1180.0122
Frequencies --	1181.9028	1182.9840	1198.2573
Frequencies --	1203.2804	1206.1908	1208.5214
Frequencies --	1218.3141	1221.7182	1237.2766
Frequencies --	1318.0866	1323.9122	1328.6608
Frequencies --	1358.1223	1382.2934	1444.1897
Frequencies --	1466.0046	1466.9472	1482.9553
Frequencies --	1492.7092	1494.7900	1497.5808
Frequencies --	1504.5080	1505.9429	1506.8880
Frequencies --	1507.8019	1508.8276	1511.4583
Frequencies --	1551.0617	1610.5440	1660.9417
Frequencies --	1701.8427	3015.1303	3019.4194
Frequencies --	3021.3685	3025.3332	3028.1462
Frequencies --	3041.7621	3068.1774	3086.0303
Frequencies --	3087.6217	3097.0905	3104.3541
Frequencies --	3119.6890	3119.7202	3123.0223
Frequencies --	3124.5294	3137.9866	3144.3025
Frequencies --	3151.7769	3176.1095	3230.1623

STYRX, PCM-CCl<sub>4</sub> “++pa”

Frequencies --	17.7286	29.1758	35.8835
Frequencies --	41.2196	54.6790	64.7198
Frequencies --	66.3251	88.6432	91.5349
Frequencies --	92.7713	96.7488	104.9437
Frequencies --	133.1136	155.0583	188.5043
Frequencies --	209.2696	218.9079	256.8249
Frequencies --	277.6713	333.5993	358.6764
Frequencies --	366.1735	397.5361	414.8328
Frequencies --	420.4049	436.3675	511.9518
Frequencies --	547.9813	595.1402	654.5549
Frequencies --	662.0872	685.0986	742.5042
Frequencies --	758.7922	786.6850	814.0562
Frequencies --	825.6527	847.6328	851.5264
Frequencies --	878.8800	921.0750	935.1365
Frequencies --	964.5325	973.2633	1001.2787
Frequencies --	1031.4744	1035.3144	1052.7668
Frequencies --	1108.1377	1112.9149	1123.7833
Frequencies --	1133.9050	1158.5709	1183.3151
Frequencies --	1183.8619	1184.3977	1193.4628
Frequencies --	1209.0954	1209.6572	1210.5829
Frequencies --	1221.1887	1235.0129	1237.5787
Frequencies --	1308.7851	1325.9536	1358.6617
Frequencies --	1360.2150	1370.8021	1450.0636
Frequencies --	1456.7981	1470.9134	1494.1838
Frequencies --	1497.1080	1499.2609	1508.9583
Frequencies --	1509.3979	1509.7347	1511.5158
Frequencies --	1514.5535	1516.7100	1518.0069
Frequencies --	1555.3909	1614.7686	1665.1249
Frequencies --	1706.7779	3016.1084	3024.3205
Frequencies --	3025.6525	3026.2755	3033.6550
Frequencies --	3069.5706	3079.6488	3090.5442
Frequencies --	3092.5037	3095.4162	3111.7917
Frequencies --	3117.7717	3119.5643	3128.8121
Frequencies --	3148.0049	3155.8030	3164.8577
Frequencies --	3173.2807	3189.3177	3239.2406

STYRX, PCM-CCl<sub>4</sub> “+apa”

Frequencies --	16.8242	25.8259	28.3718
Frequencies --	33.6310	46.2891	53.3339
Frequencies --	54.0503	61.8962	63.6114
Frequencies --	82.4960	97.6733	108.8878
Frequencies --	127.4293	154.4749	177.7717
Frequencies --	181.8613	211.2656	250.0358
Frequencies --	258.0830	321.7831	304.4849
Frequencies --	352.2005	399.0614	414.6376
Frequencies --	416.3003	454.4945	507.9352
Frequencies --	546.3231	631.4895	653.9371
Frequencies --	666.2052	685.9094	733.5259
Frequencies --	766.5036	797.6164	813.3757
Frequencies --	821.6743	841.2845	850.4633
Frequencies --	871.4416	919.7247	925.6198
Frequencies --	964.9490	967.4980	1006.0110
Frequencies --	1030.2833	1034.9749	1050.2220
Frequencies --	1109.1619	1110.3721	1123.2750
Frequencies --	1139.8725	1158.3593	1181.6960
Frequencies --	1183.5093	1184.7782	1208.0704
Frequencies --	1208.3749	1210.9359	1213.7627
Frequencies --	1229.9482	1233.5335	1237.0864
Frequencies --	1301.6485	1325.2726	1358.4170
Frequencies --	1359.4198	1364.1416	1450.2214
Frequencies --	1463.8170	1470.5761	1496.2347
Frequencies --	1497.8304	1500.1885	1506.7720
Frequencies --	1508.7822	1509.7847	1510.4599
Frequencies --	1512.8574	1516.0308	1516.5515
Frequencies --	1554.8454	1614.5639	1664.9543
Frequencies --	1706.4446	3012.2410	3013.9326
Frequencies --	3024.7805	3028.0835	3047.5437
Frequencies --	3067.4241	3074.2891	3074.6338
Frequencies --	3090.1395	3097.7461	3113.1870
Frequencies --	3113.8630	3117.9973	3129.9424
Frequencies --	3151.9907	3157.1772	3159.1676
Frequencies --	3168.0127	3184.9566	3241.2013

STYRX, PCM-CCl<sub>4</sub> “aaaa”

Frequencies --	11.7410	20.9803	38.1895
Frequencies --	40.0397	49.4754	51.6737
Frequencies --	60.6334	64.6781	68.7259
Frequencies --	73.1675	85.7219	103.0986
Frequencies --	131.5036	148.2613	158.9840
Frequencies --	176.0385	215.2079	248.7217
Frequencies --	248.9297	301.8592	319.4543
Frequencies --	383.1181	394.0185	413.8275
Frequencies --	427.7612	476.0734	484.0959
Frequencies --	539.6137	648.0115	652.3598
Frequencies --	652.9106	691.4564	737.2280
Frequencies --	750.9816	783.6980	809.1007
Frequencies --	830.5318	843.7526	851.6460
Frequencies --	854.2333	919.9733	934.3619
Frequencies --	969.2476	974.8705	1027.2245
Frequencies --	1035.0379	1035.6007	1053.1843
Frequencies --	1104.7777	1107.3111	1134.5210
Frequencies --	1146.2954	1153.0030	1183.9228
Frequencies --	1184.5606	1187.7570	1199.3664
Frequencies --	1207.7546	1211.0918	1212.2182
Frequencies --	1219.0830	1224.5204	1238.7093
Frequencies --	1316.9884	1327.4246	1329.7622
Frequencies --	1361.8873	1382.8018	1447.8446
Frequencies --	1469.7325	1472.1488	1488.6137
Frequencies --	1497.0558	1498.6821	1502.8089
Frequencies --	1508.9008	1510.2897	1511.3762
Frequencies --	1512.9509	1514.9176	1517.2014
Frequencies --	1554.7642	1613.9591	1665.6146
Frequencies --	1706.6667	3013.5305	3016.0747
Frequencies --	3021.5495	3025.7588	3032.3839
Frequencies --	3047.3119	3074.5690	3075.8097
Frequencies --	3077.4739	3091.3616	3114.0999
Frequencies --	3116.1129	3119.9212	3129.9825
Frequencies --	3149.0490	3155.9031	3166.2042
Frequencies --	3176.4773	3196.4238	3239.5161



**ICS**

**ICS, gas phase “---a”**

Frequencies --	17.1428	24.4854	34.4754
Frequencies --	37.4072	44.1026	48.6148
Frequencies --	52.7418	56.3133	75.6435
Frequencies --	93.0790	97.4586	106.3958
Frequencies --	149.6475	162.1565	195.1221
Frequencies --	233.7622	249.4938	257.7081
Frequencies --	267.5303	278.8923	291.0739
Frequencies --	314.7474	331.6046	394.2814
Frequencies --	410.8083	422.5677	453.1811
Frequencies --	505.5649	564.6404	620.7209
Frequencies --	632.0484	703.6138	758.9363
Frequencies --	788.8990	793.5680	816.2204
Frequencies --	818.2928	819.2869	829.9917
Frequencies --	881.2574	950.0569	960.7525
Frequencies --	971.5856	974.4783	1034.9416
Frequencies --	1079.9768	1100.4803	1108.3248
Frequencies --	1112.3708	1128.3637	1131.3897
Frequencies --	1145.1465	1166.3395	1185.6398
Frequencies --	1186.9479	1199.7923	1217.3945
Frequencies --	1304.9253	1314.1461	1315.9350
Frequencies --	1327.8977	1329.9784	1390.4861
Frequencies --	1406.5530	1406.5858	1407.6656
Frequencies --	1408.5374	1433.9083	1435.3389
Frequencies --	1436.7733	1466.3846	1482.7546
Frequencies --	1498.0294	1498.3490	1498.9620
Frequencies --	1502.4088	1512.4255	1513.7624
Frequencies --	1515.2960	1516.5324	1534.0251
Frequencies --	1534.8554	1538.6045	2375.0827
Frequencies --	3009.1382	3010.1964	3035.8732
Frequencies --	3036.9172	3044.7387	3046.7248
Frequencies --	3048.6554	3049.2563	3050.7228
Frequencies --	3050.8414	3056.6252	3079.6149
Frequencies --	3079.6665	3091.8227	3107.1968
Frequencies --	3123.0595	3125.0660	3125.1447
Frequencies --	3132.8896	3133.2403	3146.8602

**ICS, gas phase “-+--”**

Frequencies --	19.7776	22.5445	25.7389
Frequencies --	35.4523	42.4719	51.4680
Frequencies --	53.1705	64.0760	79.4956
Frequencies --	90.7776	110.4431	112.3695
Frequencies --	120.1057	149.8087	177.7016
Frequencies --	208.5504	243.4070	256.9042
Frequencies --	269.8280	276.1309	292.2774
Frequencies --	310.5917	329.9329	394.0939
Frequencies --	397.2567	430.0495	453.1500
Frequencies --	487.8534	563.8527	615.0645
Frequencies --	626.0683	738.7603	741.3384
Frequencies --	781.9447	804.1513	814.7625
Frequencies --	817.9257	821.5546	828.9839
Frequencies --	882.8192	952.4833	956.0411
Frequencies --	970.5919	973.3523	1021.6754
Frequencies --	1068.9136	1105.8759	1111.2517
Frequencies --	1113.8239	1130.1006	1132.8481
Frequencies --	1147.8807	1186.7240	1187.7638
Frequencies --	1194.9815	1200.0153	1215.5588
Frequencies --	1304.9507	1315.9512	1319.5674
Frequencies --	1331.2659	1333.4770	1389.8886
Frequencies --	1404.7894	1407.9452	1408.2150
Frequencies --	1409.0824	1434.5764	1436.4414
Frequencies --	1438.1056	1467.8514	1491.7981
Frequencies --	1497.8966	1498.4811	1499.0057
Frequencies --	1500.8201	1511.1603	1514.0305
Frequencies --	1515.8851	1525.2345	1529.5195
Frequencies --	1537.6535	1541.2039	2383.5510
Frequencies --	3005.0142	3007.9616	3025.8487
Frequencies --	3033.9180	3034.9340	3042.7647
Frequencies --	3047.5322	3049.6268	3050.2085
Frequencies --	3050.4495	3057.1048	3075.0974
Frequencies --	3082.0338	3084.8529	3099.7759
Frequencies --	3120.4043	3123.9309	3124.1036
Frequencies --	3131.9159	3135.5760	3136.2386

**ICS, gas phase “-+aa”**

Frequencies --	3.8873	19.7546	22.0469
Frequencies --	25.7904	32.6205	38.2666
Frequencies --	44.1141	53.3506	72.3195
Frequencies --	81.3195	96.4423	118.5211
Frequencies --	138.1370	143.8570	159.0198
Frequencies --	202.9094	238.4909	257.3557
Frequencies --	263.8621	269.8826	288.6309
Frequencies --	291.7468	329.7854	331.0001
Frequencies --	386.1897	409.6610	433.3795
Frequencies --	502.3246	563.6791	622.9178
Frequencies --	632.4186	717.0458	758.0864
Frequencies --	785.7397	804.2856	813.1524
Frequencies --	818.5300	819.0377	826.5434
Frequencies --	910.1830	958.3820	968.8538
Frequencies --	977.3689	984.9728	1024.2869
Frequencies --	1074.8264	1100.4980	1107.3914
Frequencies --	1116.4969	1126.5427	1129.8291
Frequencies --	1152.6882	1184.6905	1186.0804
Frequencies --	1188.7374	1194.8160	1216.3104
Frequencies --	1308.9715	1316.6023	1319.7542
Frequencies --	1322.2810	1338.0020	1345.2613
Frequencies --	1407.6258	1408.5919	1410.4277
Frequencies --	1420.3310	1434.6851	1435.8858
Frequencies --	1439.0443	1466.7794	1487.3889
Frequencies --	1496.5817	1496.7984	1499.1161
Frequencies --	1510.2205	1511.6815	1513.0557
Frequencies --	1514.1247	1525.7946	1534.4505
Frequencies --	1536.8795	1540.9966	2388.5566
Frequencies --	3001.7991	3016.0801	3020.4305
Frequencies --	3026.8118	3034.3186	3037.9622
Frequencies --	3045.5604	3049.5709	3050.4921
Frequencies --	3051.9564	3052.7404	3060.1140
Frequencies --	3071.9634	3080.8271	3106.9812
Frequencies --	3123.4689	3124.9397	3126.3407
Frequencies --	3131.6516	3134.4306	3135.7416

**ICS, gas phase “-a-a”**

Frequencies --	10.3754	17.1063	22.7664
Frequencies --	24.8578	27.7111	38.0988
Frequencies --	43.5256	45.4833	54.2686
Frequencies --	72.5829	94.3285	114.0970
Frequencies --	128.6675	149.8943	158.6803
Frequencies --	219.2283	237.2593	253.5392
Frequencies --	259.1230	267.9545	279.5940
Frequencies --	289.7761	322.5376	330.8109
Frequencies --	386.2565	427.6878	465.0279
Frequencies --	497.0975	567.6584	619.1303
Frequencies --	663.6303	700.4278	776.9742
Frequencies --	784.9804	805.2565	814.9175
Frequencies --	817.4152	818.1908	826.9369
Frequencies --	888.6834	960.2803	965.1435
Frequencies --	976.4197	984.8562	1038.9519
Frequencies --	1068.9274	1105.9259	1107.7354
Frequencies --	1117.2852	1128.6657	1129.7188
Frequencies --	1153.4460	1182.9190	1185.6793
Frequencies --	1187.7291	1193.4594	1224.0477
Frequencies --	1293.9450	1314.7813	1317.0666
Frequencies --	1318.4381	1328.8921	1388.9202
Frequencies --	1393.8867	1407.2424	1407.9575
Frequencies --	1410.2675	1434.4663	1435.2456
Frequencies --	1438.6768	1470.9212	1485.4114
Frequencies --	1496.6262	1497.2874	1497.6001
Frequencies --	1499.5013	1502.7871	1511.6027
Frequencies --	1512.6086	1513.5576	1534.4093
Frequencies --	1536.7220	1540.3242	2387.3219
Frequencies --	3003.6662	3006.8577	3015.6776
Frequencies --	3024.8168	3028.8745	3039.9823
Frequencies --	3043.0371	3047.5441	3049.9149
Frequencies --	3050.0936	3051.5428	3053.9594
Frequencies --	3067.8325	3074.2986	3102.6679
Frequencies --	3123.7102	3124.0633	3125.8973
Frequencies --	3131.7630	3134.4986	3135.0742



ICS, gas phase “-aaa”

Frequencies --	9.8467	22.4714	26.6429
Frequencies --	30.4031	34.4762	38.8218
Frequencies --	46.1691	51.2434	69.0571
Frequencies --	77.7105	80.4446	101.7008
Frequencies --	122.1792	139.6324	147.4099
Frequencies --	212.2790	219.7357	249.0204
Frequencies --	257.7700	261.8548	270.2482
Frequencies --	287.6086	322.2757	330.1318
Frequencies --	383.8198	404.6490	453.7589
Frequencies --	483.3656	567.5363	621.4896
Frequencies --	663.9690	699.0578	773.1474
Frequencies --	789.5172	798.4830	815.0886
Frequencies --	818.6711	819.5854	828.9912
Frequencies --	910.8705	960.5085	972.4661
Frequencies --	978.9222	1006.6215	1035.0439
Frequencies --	1053.9398	1106.2892	1109.0337
Frequencies --	1117.6939	1128.5054	1130.1101
Frequencies --	1154.3833	1184.2079	1186.4627
Frequencies --	1187.8900	1191.6029	1224.4535
Frequencies --	1299.0530	1317.4384	1318.5938
Frequencies --	1319.8651	1333.2111	1341.5430
Frequencies --	1406.2637	1408.2917	1409.5600
Frequencies --	1410.3407	1434.9872	1436.2022
Frequencies --	1439.0996	1469.3658	1487.0120
Frequencies --	1496.9303	1497.3658	1498.1572
Frequencies --	1502.6690	1511.7021	1512.3997
Frequencies --	1513.6072	1518.2867	1534.1414
Frequencies --	1537.0418	1540.3852	2388.4754
Frequencies --	3001.7349	3010.5961	3017.4568
Frequencies --	3018.6209	3022.5486	3037.4641
Frequencies --	3046.6202	3050.3470	3050.5312
Frequencies --	3051.6469	3054.7461	3055.2477
Frequencies --	3060.9856	3062.1174	3107.1097
Frequencies --	3124.5486	3124.6212	3126.0102
Frequencies --	3132.1629	3134.7697	3135.6121

ICS, PCM-H<sub>2</sub>O “----”

Frequencies --	14.5595	24.1014	28.1913
Frequencies --	32.4318	35.7284	42.6453
Frequencies --	45.6571	60.3965	72.9193
Frequencies --	86.7674	105.3991	118.7328
Frequencies --	147.2536	155.2045	175.9732
Frequencies --	202.2267	243.6530	257.6580
Frequencies --	261.5713	276.8399	293.7679
Frequencies --	330.4270	346.8858	356.7366
Frequencies --	385.2447	438.0661	459.8071
Frequencies --	485.8582	562.8235	608.5044
Frequencies --	623.4020	709.4968	745.1008
Frequencies --	776.7572	782.1589	810.1001
Frequencies --	816.9385	817.8280	823.0036
Frequencies --	873.5464	951.8196	953.0693
Frequencies --	963.7634	970.6068	1026.0512
Frequencies --	1075.1319	1080.5472	1086.1374
Frequencies --	1105.0243	1116.9919	1119.4315
Frequencies --	1126.6852	1158.6784	1178.5289
Frequencies --	1185.0355	1188.9106	1215.0106
Frequencies --	1302.3686	1319.3995	1322.8539
Frequencies --	1326.5415	1328.2377	1378.6088
Frequencies --	1396.8818	1407.0357	1408.1821
Frequencies --	1408.7228	1430.6953	1433.0226
Frequencies --	1434.7546	1456.0570	1477.5229
Frequencies --	1491.0824	1491.3106	1491.5796
Frequencies --	1501.1469	1502.1820	1503.6230
Frequencies --	1504.4371	1511.6147	1528.0364
Frequencies --	1529.8011	1531.2240	2348.5797
Frequencies --	3011.2646	3019.1100	3025.5589
Frequencies --	3026.1627	3029.6095	3044.1500
Frequencies --	3045.1292	3046.6107	3051.0120
Frequencies --	3054.1372	3060.2076	3067.3081
Frequencies --	3068.0504	3073.0193	3118.8175
Frequencies --	3120.8078	3122.8287	3125.2250
Frequencies --	3125.9687	3126.0948	3129.1904

ICS, PCM-H<sub>2</sub>O “-+a”

Frequencies --	-17.4288	24.3769	29.9778
Frequencies --	34.7394	49.8872	57.6957
Frequencies --	62.4290	66.5839	70.8104
Frequencies --	87.7818	107.6973	110.7743
Frequencies --	135.0199	144.4494	165.6664
Frequencies --	206.3432	248.4281	265.1643
Frequencies --	277.2108	279.1896	287.7842
Frequencies --	303.9988	320.5529	333.3522
Frequencies --	387.3385	422.1430	439.8965
Frequencies --	495.8803	552.7026	611.0517
Frequencies --	625.3180	724.8692	740.0668
Frequencies --	769.7771	792.0849	809.1711
Frequencies --	813.8517	816.6142	821.2945
Frequencies --	892.4147	952.3265	958.1030
Frequencies --	970.6750	971.8921	1017.1838
Frequencies --	1076.5002	1081.5324	1089.3633
Frequencies --	1112.0844	1118.7128	1123.4161
Frequencies --	1129.6626	1179.5036	1180.4634
Frequencies --	1181.8015	1183.6439	1220.3184
Frequencies --	1292.2016	1310.8099	1315.6568
Frequencies --	1318.4043	1333.4909	1386.5760
Frequencies --	1396.1939	1402.6114	1404.6880
Frequencies --	1406.9711	1428.1478	1430.8817
Frequencies --	1433.1066	1450.9790	1477.8503
Frequencies --	1491.4916	1491.6659	1493.4805
Frequencies --	1494.2301	1498.5893	1502.7984
Frequencies --	1503.8549	1508.5018	1530.2092
Frequencies --	1531.1423	1534.8118	2354.2337
Frequencies --	3004.6894	3010.9144	3016.4831
Frequencies --	3027.7313	3030.8451	3034.5612
Frequencies --	3043.6776	3045.1465	3045.7713
Frequencies --	3047.3990	3053.4531	3058.1477
Frequencies --	3068.0458	3075.0302	3085.9214
Frequencies --	3119.8623	3120.6229	3121.7304
Frequencies --	3125.0230	3125.2892	3135.8103

ICS, PCM-H<sub>2</sub>O “+aa”

Frequencies --	-19.3871	16.5840	22.7503
Frequencies --	32.6911	40.2274	44.9204
Frequencies --	54.8305	60.6326	66.5303
Frequencies --	84.0527	90.1449	97.5917
Frequencies --	122.9993	139.1090	177.5550
Frequencies --	200.4078	235.9291	262.2814
Frequencies --	270.0569	273.5416	285.7981
Frequencies --	290.0178	298.6300	330.3655
Frequencies --	390.3960	408.9389	430.1288
Frequencies --	495.3563	558.5423	610.1895
Frequencies --	634.7012	712.3666	756.6447
Frequencies --	778.5580	799.0617	808.3186
Frequencies --	815.1657	818.5253	819.2993
Frequencies --	891.4508	957.3667	969.4813
Frequencies --	970.9200	984.9941	1025.2621
Frequencies --	1076.5467	1081.2648	1084.7542
Frequencies --	1113.0622	1119.1580	1120.2819
Frequencies --	1128.8170	1178.1099	1179.5817
Frequencies --	1181.7115	1186.3036	1215.2026
Frequencies --	1307.9675	1315.8104	1317.5931
Frequencies --	1318.6944	1336.0381	1343.6138
Frequencies --	1404.4824	1405.6300	1406.4525
Frequencies --	1412.2691	1430.9622	1432.3386
Frequencies --	1432.7355	1452.8430	1475.4337
Frequencies --	1491.8467	1492.1285	1492.8256
Frequencies --	1500.2489	1503.2145	1503.2447
Frequencies --	1504.3155	1517.2228	1528.8619
Frequencies --	1529.8392	1531.9510	2355.3895
Frequencies --	3006.2405	3007.0874	3014.0782
Frequencies --	3020.6206	3032.2117	3038.2149
Frequencies --	3045.4378	3046.6224	3046.6975
Frequencies --	3046.7757	3047.4379	3057.6077
Frequencies --	3062.6603	3076.9208	3118.7078
Frequencies --	3120.8374	3122.5752	3122.9314
Frequencies --	3125.7546	3126.2913	3126.6590



ICS, PCM-H<sub>2</sub>O “+a--”

Frequencies --	16.3491	20.8727	23.6273
Frequencies --	26.1353	40.5242	44.0582
Frequencies --	52.5667	62.6426	64.8228
Frequencies --	82.0011	103.8271	114.9991
Frequencies --	127.3756	156.7061	162.1385
Frequencies --	227.0614	246.7408	255.5403
Frequencies --	268.1131	273.1279	282.9131
Frequencies --	289.7902	298.1148	341.2029
Frequencies --	404.4257	420.9508	459.0942
Frequencies --	501.8543	560.8636	607.3994
Frequencies --	665.1796	694.3654	764.4779
Frequencies --	782.0017	788.1675	809.5959
Frequencies --	815.0419	818.1608	820.3391
Frequencies --	882.8570	948.6257	958.5091
Frequencies --	971.2702	974.2248	1029.2286
Frequencies --	1061.7205	1081.0392	1088.0011
Frequencies --	1112.8651	1119.2029	1120.0141
Frequencies --	1129.5886	1178.9420	1180.3315
Frequencies --	1180.8560	1183.0049	1223.8742
Frequencies --	1293.7829	1314.9696	1317.4350
Frequencies --	1318.4316	1319.0529	1382.1012
Frequencies --	1390.2341	1405.3122	1406.0365
Frequencies --	1406.9964	1430.7657	1432.3160
Frequencies --	1433.4178	1457.1259	1477.4905
Frequencies --	1491.4871	1491.8771	1492.0711
Frequencies --	1492.4662	1503.0598	1503.0952
Frequencies --	1504.2490	1508.2684	1529.2283
Frequencies --	1530.4063	1531.7043	2349.9247
Frequencies --	3007.1606	3009.3963	3013.4649
Frequencies --	3018.1501	3041.9678	3044.5789
Frequencies --	3045.5946	3046.4530	3046.5812
Frequencies --	3048.0824	3049.1842	3059.4844
Frequencies --	3060.1452	3086.3686	3097.8052
Frequencies --	3121.2834	3122.4099	3122.8420
Frequencies --	3125.8778	3126.2274	3126.3978

ICS, PCM-H<sub>2</sub>O “aaaa”

Frequencies --	12.2085	18.8676	24.8079
Frequencies --	29.7679	37.7188	38.3257
Frequencies --	60.9399	62.1692	68.3986
Frequencies --	70.1390	85.9228	95.3310
Frequencies --	126.6611	130.8870	155.7827
Frequencies --	211.7729	226.6529	242.7854
Frequencies --	263.0795	267.1150	273.8829
Frequencies --	286.7040	306.2685	314.0273
Frequencies --	377.4718	413.5923	450.9340
Frequencies --	481.8454	559.8801	611.3375
Frequencies --	670.6198	701.9955	764.9222
Frequencies --	772.4144	805.1561	813.9192
Frequencies --	814.9272	816.4086	819.7421
Frequencies --	891.2399	958.4894	968.6359
Frequencies --	973.3350	1003.1182	1031.2513
Frequencies --	1052.0664	1078.4309	1087.1786
Frequencies --	1112.1485	1119.3399	1120.7446
Frequencies --	1127.4131	1178.4103	1179.8029
Frequencies --	1180.2309	1181.4460	1224.9828
Frequencies --	1297.7063	1315.3716	1317.5602
Frequencies --	1318.6114	1333.2795	1339.7649
Frequencies --	1403.7515	1404.2669	1405.9151
Frequencies --	1407.7740	1429.9955	1432.0776
Frequencies --	1433.7527	1464.3325	1473.3913
Frequencies --	1489.5602	1491.6062	1491.7154
Frequencies --	1492.0073	1503.0355	1503.3743
Frequencies --	1503.8024	1511.0236	1529.0618
Frequencies --	1530.6224	1533.4521	2350.4926
Frequencies --	3007.7990	3008.7893	3014.2098
Frequencies --	3016.7443	3026.2815	3043.4246
Frequencies --	3046.1370	3046.8422	3047.1612
Frequencies --	3047.9310	3048.9392	3052.6524
Frequencies --	3057.3068	3068.3320	3093.0098
Frequencies --	3122.1120	3122.8781	3123.2050
Frequencies --	3125.8595	3126.6825	3127.0939

ICS, PCM-CCl<sub>4</sub> “----”

Frequencies --	-18.7398	19.5585	23.3158
Frequencies --	33.3747	36.8626	37.4303
Frequencies --	50.6981	59.0886	68.9795
Frequencies --	71.2646	96.8102	107.4225
Frequencies --	145.2616	161.3033	180.1474
Frequencies --	220.6931	247.2526	258.1463
Frequencies --	266.6036	276.6074	302.4446
Frequencies --	311.8986	325.7550	342.6578
Frequencies --	421.6332	436.3447	446.3924
Frequencies --	492.6892	566.3715	614.6603
Frequencies --	628.3012	709.9746	754.2482
Frequencies --	788.2778	795.0392	815.3715
Frequencies --	816.5256	820.3498	828.4767
Frequencies --	874.0104	956.0325	958.8381
Frequencies --	971.3007	978.9930	1030.7411
Frequencies --	1079.6387	1098.1880	1104.0384
Frequencies --	1112.6759	1124.6805	1127.4793
Frequencies --	1137.5472	1165.5265	1183.9510
Frequencies --	1184.6329	1194.8092	1220.0857
Frequencies --	1306.4941	1313.8310	1319.1699
Frequencies --	1320.5902	1327.1572	1384.4736
Frequencies --	1404.8438	1406.5227	1408.0999
Frequencies --	1409.4409	1433.1707	1433.9004
Frequencies --	1436.5990	1458.6303	1483.1755
Frequencies --	1495.4482	1496.1418	1498.0165
Frequencies --	1498.5830	1509.7536	1510.4762
Frequencies --	1512.8393	1519.3617	1533.6037
Frequencies --	1536.3997	1539.4306	2367.7653
Frequencies --	3006.2875	3019.7865	3027.7208
Frequencies --	3030.4660	3035.6048	3044.5493
Frequencies --	3045.4158	3048.1612	3049.4029
Frequencies --	3052.2159	3063.7455	3071.2340
Frequencies --	3072.4220	3084.5338	3119.0031
Frequencies --	3122.5962	3124.2739	3125.7407
Frequencies --	3129.0963	3130.4883	3131.0828

ICS, PCM-CCl<sub>4</sub> “-+--”

Frequencies --	14.1405	17.1132	27.9367
Frequencies --	34.1778	40.5286	48.6033
Frequencies --	52.0575	63.2343	71.8715
Frequencies --	92.7654	109.4195	112.8327
Frequencies --	123.7536	146.8446	174.9218
Frequencies --	206.7450	242.1210	255.3807
Frequencies --	268.6039	271.7481	287.3371
Frequencies --	320.6598	328.8390	392.1536
Frequencies --	397.1114	429.9330	450.3812
Frequencies --	486.1132	560.5618	609.7991
Frequencies --	625.7407	735.3945	738.7953
Frequencies --	778.2634	800.4997	812.2733
Frequencies --	817.6890	819.9821	826.0521
Frequencies --	880.7515	948.1368	954.7584
Frequencies --	968.2360	971.4334	1018.5135
Frequencies --	1067.5072	1102.0630	1105.7047
Frequencies --	1110.7031	1124.9473	1127.3030
Frequencies --	1142.1046	1183.7338	1186.9759
Frequencies --	1192.0101	1198.8030	1213.2233
Frequencies --	1302.4440	1315.8200	1319.3249
Frequencies --	1330.0192	1332.8252	1387.1837
Frequencies --	1401.3529	1406.7236	1407.3286
Frequencies --	1407.5884	1432.1427	1434.3634
Frequencies --	1435.5344	1461.6972	1486.8933
Frequencies --	1494.8826	1495.1832	1495.8179
Frequencies --	1496.6764	1508.5929	1510.5110
Frequencies --	1511.7085	1520.1535	1529.8816
Frequencies --	1534.8873	1537.6453	2375.0163
Frequencies --	3006.0909	3007.2034	3024.5313
Frequencies --	3027.4562	3031.5525	3042.1942
Frequencies --	3045.8837	3047.1740	3048.5652
Frequencies --	3051.5228	3055.4512	3070.1595
Frequencies --	3077.5584	3080.9645	3097.6224
Frequencies --	3118.8450	3121.5254	3122.7790
Frequencies --	3128.3447	3132.0522	3133.8902



ICS, PCM-CCl<sub>4</sub> “++a-”

Frequencies --	1.7997	15.9206	23.4418
Frequencies --	27.8862	28.8914	43.6844
Frequencies --	46.3810	52.7417	64.4688
Frequencies --	74.2644	80.7886	105.6195
Frequencies --	117.3102	140.2397	178.7421
Frequencies --	188.0354	245.1459	259.5020
Frequencies --	266.7354	270.2363	300.7140
Frequencies --	303.9743	307.7646	316.6932
Frequencies --	412.1092	418.1197	443.3866
Frequencies --	491.8372	560.5210	625.5850
Frequencies --	634.7368	713.6187	755.0303
Frequencies --	793.9573	802.2669	816.1391
Frequencies --	816.6739	818.3242	823.1744
Frequencies --	885.1064	960.6041	966.4186
Frequencies --	979.0644	983.9445	1020.9269
Frequencies --	1066.6300	1102.0477	1105.4557
Frequencies --	1115.0275	1125.2289	1126.5268
Frequencies --	1137.4886	1182.1939	1183.8674
Frequencies --	1185.4555	1187.8380	1220.8955
Frequencies --	1307.8309	1313.7234	1316.5086
Frequencies --	1317.5305	1336.6404	1346.5247
Frequencies --	1405.6322	1406.1782	1407.5714
Frequencies --	1411.4623	1432.7546	1434.3284
Frequencies --	1435.5692	1459.3450	1481.7610
Frequencies --	1495.1395	1495.7978	1495.9062
Frequencies --	1506.9146	1509.0143	1510.0272
Frequencies --	1510.5617	1529.4710	1533.4704
Frequencies --	1536.9446	1537.3450	2370.4703
Frequencies --	3007.2075	3010.2710	3020.9935
Frequencies --	3021.9253	3031.5959	3035.0699
Frequencies --	3045.7137	3048.4984	3048.8438
Frequencies --	3049.3911	3064.6850	3068.0240
Frequencies --	3071.0210	3088.6266	3113.0452
Frequencies --	3123.1357	3123.3290	3124.3586
Frequencies --	3130.1814	3130.2786	3130.9634

ICS, PCM-CCl<sub>4</sub> “+a--”

Frequencies --	16.6583	17.7709	20.2072
Frequencies --	28.8939	39.6792	41.6611
Frequencies --	48.0357	58.4385	63.0795
Frequencies --	81.7252	96.6112	111.1015
Frequencies --	128.9014	154.8864	161.8475
Frequencies --	224.4360	245.5434	256.7943
Frequencies --	265.2884	270.1852	278.3480
Frequencies --	286.9182	309.9936	333.1961
Frequencies --	404.0746	425.2423	459.5346
Frequencies --	503.3208	563.7467	614.9551
Frequencies --	668.1537	697.6156	769.0084
Frequencies --	788.9203	796.9693	812.6744
Frequencies --	818.0488	821.6681	826.6691
Frequencies --	885.3805	955.0457	960.3722
Frequencies --	975.9708	978.7070	1032.8204
Frequencies --	1064.8238	1099.5942	1100.2852
Frequencies --	1116.4739	1124.0601	1125.5958
Frequencies --	1144.5879	1182.0965	1184.0015
Frequencies --	1184.9508	1188.3589	1225.1874

Frequencies --	1296.7609	1311.3184	1318.0263
Frequencies --	1320.2112	1322.9006	1385.0962
Frequencies --	1393.0570	1405.1761	1407.9452
Frequencies --	1409.6933	1432.1885	1435.1779
Frequencies --	1437.3184	1466.2367	1484.9126
Frequencies --	1494.4560	1494.8738	1495.5756
Frequencies --	1499.5102	1508.6892	1509.4144
Frequencies --	1510.4022	1515.9133	1533.2981
Frequencies --	1536.2491	1538.5907	2369.0785
Frequencies --	3003.1370	3004.9142	3019.0886
Frequencies --	3020.2611	3040.1890	3042.1495
Frequencies --	3042.5939	3046.9775	3048.5082
Frequencies --	3049.8603	3050.0266	3056.9679
Frequencies --	3068.9154	3088.7581	3099.9654
Frequencies --	3123.4750	3124.4254	3124.7480
Frequencies --	3129.0168	3132.1674	3132.6985

ICS, PCM-CCl<sub>4</sub> “aaaa”

Frequencies --	-2.7043	3.6473	13.3793
Frequencies --	22.4294	28.4452	37.5844
Frequencies --	46.7567	57.4999	62.1981
Frequencies --	68.8145	78.3426	94.0600
Frequencies --	124.4521	130.7101	151.8366
Frequencies --	210.1608	225.5404	244.3183
Frequencies --	260.6574	262.2124	272.7948
Frequencies --	284.8960	309.8952	322.8112
Frequencies --	379.1463	415.0281	448.7960
Frequencies --	484.1405	563.0962	614.4239
Frequencies --	671.2327	703.6133	767.6305
Frequencies --	777.0160	811.4769	814.3755
Frequencies --	815.8504	817.2083	827.7898
Frequencies --	900.8743	959.7122	972.3102
Frequencies --	975.2967	1006.1015	1032.1918
Frequencies --	1052.6912	1098.0647	1099.3938
Frequencies --	1116.3754	1123.5578	1125.6833
Frequencies --	1141.9614	1181.9406	1183.6137
Frequencies --	1184.9147	1186.0673	1223.1693
Frequencies --	1299.1445	1315.4428	1317.2606
Frequencies --	1317.5917	1333.9875	1342.0989
Frequencies --	1406.7789	1407.4698	1407.7927
Frequencies --	1409.5506	1434.2325	1434.6474
Frequencies --	1436.7791	1471.3290	1483.6574
Frequencies --	1494.5447	1494.9290	1495.4673
Frequencies --	1499.1200	1508.2052	1509.2512
Frequencies --	1509.6862	1516.8464	1532.8267
Frequencies --	1535.3582	1538.5065	2375.3148
Frequencies --	3001.8281	3003.0644	3018.2766
Frequencies --	3021.3674	3024.7700	3038.7339
Frequencies --	3039.6907	3048.0794	3049.4606
Frequencies --	3050.0772	3053.0759	3056.4220
Frequencies --	3057.7976	3065.1610	3099.4299
Frequencies --	3122.8041	3124.0825	3124.8807
Frequencies --	3128.6035	3131.8694	3132.3626



**STYRX**

STYRX, gas phase “++pa”

C	0.280936	1.699000	-1.317696
C	1.414684	1.611623	-0.263757
Si	2.079828	-0.101811	0.112933
O	2.567550	-0.726784	-1.335955
C	2.963545	-2.082392	-1.522467
O	3.314039	-0.006972	1.221656
C	4.484062	0.781665	1.068946
O	1.001478	-1.168685	0.754405
C	0.409388	-1.061522	2.042526
H	1.085915	2.077151	0.674690
H	2.265262	2.214173	-0.611462
H	3.104251	-2.249942	-2.594016
H	3.909250	-2.294875	-1.008818
H	2.199780	-2.773404	-1.150219
H	5.145163	0.581382	1.917090
H	5.023185	0.536365	0.145045
H	4.252470	1.855053	1.061554
H	-0.133716	-1.988809	2.244463
H	1.168145	-0.920838	2.820952
H	-0.304535	-0.229529	2.079943
H	0.144316	2.751449	-1.597639
H	0.593883	1.172329	-2.225007
C	-1.047173	1.147367	-0.842625
C	-1.855129	1.883283	0.034258
C	-3.074752	1.380845	0.481052
C	-3.539844	0.120518	0.068871
C	-1.511106	-0.109542	-1.257739
C	-2.729009	-0.613519	-0.814633
H	-1.529738	2.867459	0.364332
H	-3.684825	1.975099	1.157131
H	-0.904320	-0.700034	-1.938649
H	-3.054269	-1.588721	-1.163768
C	-4.835776	-0.362655	0.570587
C	-5.418647	-1.542964	0.327315
H	-6.384226	-1.786394	0.757856
H	-5.359768	0.341766	1.216328
H	-4.965425	-2.305048	-0.299555

STYRX, gas phase “+apa”

C	-0.073442	0.379314	0.723636
C	0.763694	-0.130951	-0.476437
Si	2.626397	0.019879	-0.310394
O	2.969187	-0.434087	1.247966
C	4.279251	-0.408732	1.806733
O	3.429789	-0.914356	-1.413067
C	3.349112	-2.329988	-1.460929
O	3.224339	1.523258	-0.649146
C	2.707964	2.742086	-0.141411
H	0.455453	0.384375	-1.395603
H	0.527090	-1.190248	-0.645087
H	4.189031	-0.493309	2.893410
H	4.882629	-1.247807	1.439877
H	4.802432	0.524371	1.567467
H	4.085885	-2.688226	-2.185324
H	3.568021	-2.788418	-0.487810
H	2.356130	-2.667251	-1.786853
H	3.371059	3.551126	-0.460846
H	1.703280	2.949492	-0.532818
H	2.661857	2.747953	0.955647
H	0.260705	-0.132538	1.631743
H	0.121221	1.446848	0.881720
C	-1.559858	0.169153	0.532151
C	-2.190682	-0.995600	0.983924
C	-3.551978	-1.202672	0.772718
C	-4.340707	-0.253677	0.101930
C	-2.346549	1.121088	-0.135074
C	-3.704950	0.918156	-0.347477
H	-1.609330	-1.747770	1.511984

H	-4.018230	-2.114764	1.137449
H	-1.883560	2.039471	-0.489370
H	-4.279376	1.681756	-0.863137
C	-5.774188	-0.522575	-0.091599
C	-6.674598	0.247025	-0.715292
H	-7.711464	-0.062004	-0.794430
H	-6.115659	-1.467198	0.331157
H	-6.421672	1.201553	-1.166927

STYRX, gas phase “aaaa”

C	-0.119744	0.080147	-0.757262
C	0.816972	-0.067579	0.454197
Si	2.634291	0.001004	0.006316
O	3.445627	-0.153976	1.442848
C	4.865307	-0.157748	1.552012
O	3.073450	-1.167357	-1.078720
C	2.925890	-2.558875	-0.839421
O	3.054677	1.387031	-0.791177
C	2.854981	2.689113	-0.262477
H	0.631566	-1.018721	0.970457
H	0.611832	0.713826	1.196973
H	5.125282	-0.296019	2.605004
H	5.311755	-0.972916	0.969935
H	5.296741	0.788674	1.204609
H	3.383307	-3.097321	-1.674075
H	3.422019	-2.871979	0.088237
H	1.869096	-2.851796	-0.781650
H	3.309874	3.408334	-0.949171
H	1.787635	2.930436	-0.170450
H	3.322312	2.809057	0.723509
H	0.111376	1.018650	-1.278656
H	0.110783	-0.707748	-1.486810
C	-1.613091	0.048750	-0.470043
C	-2.515405	0.217203	-1.530614
C	-3.889199	0.196390	-1.318826
C	-4.428268	0.006360	-0.034015
C	-2.147373	-0.141576	0.810817
C	-3.523416	-0.162468	1.025964
H	-2.132099	0.367621	-2.537432
H	-4.563464	0.330595	-2.161236
H	-1.482694	-0.275092	1.658048
H	-3.894768	-0.311608	2.035378
C	-5.889718	-0.004999	0.135015
C	-6.581686	-0.171666	1.268759
H	-7.666495	-0.160010	1.265742
H	-6.450490	0.142054	-0.787735
H	-6.105182	-0.324003	2.232649

STYRX, PCM-H<sub>2</sub>O “++pa”

C	0.266225	1.529843	-1.460917
C	1.408638	1.553347	-0.411897
Si	2.099713	-0.099663	0.134068
O	2.655797	-0.847298	-1.235555
C	3.228467	-2.155644	-1.240503
O	3.318264	0.116244	1.248039
C	4.480000	0.907082	1.001625
O	1.023498	-1.127315	0.848425
C	0.238467	-0.811157	1.997238
H	1.089733	2.118962	0.473417
H	2.255070	2.117503	-0.828890
H	3.549464	-2.382149	-2.260956
H	4.101106	-2.214661	-0.579119
H	2.496457	-2.907218	-0.925303
H	5.197479	0.719065	1.805225
H	4.953146	0.647452	0.046909
H	4.239896	1.977535	0.995579
H	-0.765848	-1.218907	1.853740
H	0.683458	-1.260183	2.892312
H	0.145226	0.268914	2.157734
H	0.130462	2.549995	-1.843168
H	0.569796	0.915180	-2.315235
C	-1.062582	1.031071	-0.929047

C	-1.829137	1.822133	-0.061352
C	-3.052664	1.370633	0.429391
C	-3.563569	0.109975	0.070511
C	-1.571308	-0.225660	-1.289792
C	-2.793514	-0.680519	-0.802249
H	-1.467100	2.808526	0.228699
H	-3.631416	2.006702	1.098646
H	-1.000639	-0.857037	-1.969440
H	-3.151221	-1.660957	-1.110191
C	-4.866962	-0.311522	0.611950
C	-5.519974	-1.455482	0.367304
H	-6.482718	-1.655516	0.828235
H	-5.331411	0.411910	1.285136
H	-5.129913	-2.227482	-0.291128

STYRX, PCM-H<sub>2</sub>O “+apa”

C	-0.082295	0.396540	0.724090
C	0.758695	-0.148746	-0.458230
Si	2.618533	0.015217	-0.306899
O	2.995322	-0.404946	1.252815
C	4.329952	-0.382125	1.767168
O	3.426832	-0.930225	-1.405100
C	3.336257	-2.355329	-1.413501
O	3.215719	1.511028	-0.700217
C	2.727820	2.732680	-0.147048
H	0.451059	0.339775	-1.392953
H	0.528927	-1.214924	-0.593716
H	4.289570	-0.650652	2.825978
H	4.969408	-1.103539	1.245652
H	4.775697	0.614353	1.672454
H	4.169203	-2.745445	-2.004649
H	3.400814	-2.775416	-0.402737
H	2.397392	-2.686837	-1.873382
H	3.430544	3.528094	-0.409915
H	1.744324	2.988218	-0.559587
H	2.651818	2.686640	0.946290
H	0.242368	-0.094458	1.647769
H	0.111023	1.467947	0.853865
C	-1.568327	0.180324	0.530084
C	-2.192113	-0.991659	0.977043
C	-3.554158	-1.203879	0.767238
C	-4.348057	-0.252089	0.102910
C	-2.359252	1.134438	-0.130728
C	-3.718986	0.926865	-0.341560
H	-1.605278	-1.746837	1.499605
H	-4.017378	-2.122002	1.127795
H	-1.900999	2.059371	-0.480812
H	-4.296398	1.695011	-0.852043
C	-5.782304	-0.528686	-0.087385
C	-6.685492	0.240613	-0.710104
H	-7.722407	-0.072522	-0.787936
H	-6.118413	-1.477464	0.336224
H	-6.434159	1.197111	-1.161175

STYRX, PCM-H<sub>2</sub>O “aaaa”

C	0.124718	0.039982	-0.766586
C	-0.813045	-0.032075	0.450702
Si	-2.628913	0.000497	0.007826
O	-3.436406	-0.094305	1.452933
O	-4.864373	-0.093880	1.550589
C	-3.073966	1.340087	-0.862603
C	-2.891840	2.668033	-0.366235
O	-3.080339	-1.209375	-1.032629
C	-2.922365	-2.592947	-0.711411
H	-0.614355	0.801970	1.137448
H	-0.621148	-0.944481	1.031577
H	-5.130859	-0.157968	2.608515
H	-5.290711	0.825857	1.134656
H	-5.299055	-0.951599	1.025143
H	-3.444845	3.351687	-1.015664
H	-3.272462	2.776778	0.656297
H	-1.833195	2.954133	-0.380203



H	-3.498624	-3.176768	-1.433944
H	-1.870984	-2.897505	-0.777795
H	-3.292562	-2.822931	0.294617
H	-0.099721	-0.792759	-1.447380
H	-0.100998	0.945566	-1.346225
C	1.617743	0.026216	-0.474220
C	2.520926	0.129786	-1.543982
C	3.895922	0.119838	-1.330426
C	4.433337	0.005710	-0.034239
C	2.150211	-0.086832	0.817644
C	3.527610	-0.097371	1.035287
H	2.138178	0.220459	-2.560625
H	4.572873	0.202539	-2.180329
H	1.487486	-0.168067	1.674968
H	3.895243	-0.186675	2.055487
C	5.896750	0.000878	0.133297
C	6.586163	-0.101081	1.277671
H	7.672089	-0.092721	1.274566
H	6.457523	0.092112	-0.799276
H	6.106065	-0.195381	2.248455

### STYRX, PCM-CCl<sub>4</sub> “++pa”

C	0.305752	1.633426	-1.395856
C	1.438058	1.590177	-0.336320
Si	2.080007	-0.108080	0.137362
O	2.519668	-0.845106	-1.275773
C	2.902716	-2.216697	-1.358743
O	3.354724	0.025844	1.198774
C	4.521992	0.800715	0.947225
O	0.991665	-1.100548	0.876673
C	0.290117	-0.803852	2.077669
H	1.114168	2.113559	0.572642
H	2.296360	2.160858	-0.715700
H	3.054498	-2.462806	-2.413510
H	3.839245	-2.402398	-0.818889
H	2.125475	-2.870302	-0.949010
H	5.232453	0.618342	1.758530
H	4.998321	0.521057	-0.000454
H	4.296014	1.874105	0.923214
H	-0.652396	-1.357559	2.067516
H	0.876435	-1.108800	2.952004
H	0.054072	0.263240	2.165514
H	0.173887	2.673268	-1.721939
H	0.618898	1.065963	-2.278074
C	-1.025891	1.106689	-0.901830
C	-1.825522	1.874654	-0.044603
C	-3.044610	1.391903	0.425856
C	-3.516884	0.120198	0.058084
C	-1.497560	-0.161466	-1.272976
C	-2.714749	-0.646455	-0.805873
H	-1.494365	2.868369	0.252436
H	-3.648997	2.010437	1.086712
H	-0.898705	-0.776833	-1.939415
H	-3.045715	-1.631763	-1.121812
C	-4.812644	-0.340092	0.583362
C	-5.408202	-1.520241	0.369465
H	-6.371985	-1.745972	0.814936
H	-5.324066	0.383814	1.219099
H	-4.967578	-2.298734	-0.247033

### STYRX, PCM-CCl<sub>4</sub> “+apa”

C	-0.076368	0.385865	0.716394
C	0.761948	-0.146221	-0.473595
Si	2.623446	0.016598	-0.313686
O	2.977655	-0.412269	1.248689
C	4.295659	-0.396118	1.792997
O	3.432898	-0.925931	-1.407390
C	3.338747	-2.344595	-1.438942
O	3.212430	1.518708	-0.682842
C	2.734099	2.734191	-0.122618
H	0.453166	0.351481	-1.402403
H	0.529035	-1.209493	-0.621889
H	4.222313	-0.580859	2.868181
H	4.921001	-1.177210	1.344685
H	4.783982	0.572300	1.633836
H	4.124112	-2.722863	-2.099181
H	3.477548	-2.788808	-0.445167
H	2.368819	-2.672146	-1.834729

H	3.395177	3.541516	-0.449842
H	1.717493	2.961972	-0.467498
H	2.734316	2.710585	0.974606
H	0.254623	-0.110682	1.634386
H	0.118658	1.455822	0.855613
C	-1.563099	0.174032	0.526648
C	-2.195675	-0.982753	0.997428
C	-3.558055	-1.190582	0.790272
C	-4.345266	-0.250042	0.104691
C	-2.348058	1.117012	-0.155907
C	-3.707434	0.913361	-0.364536
H	-1.615564	-1.728312	1.537872
H	-4.026282	-2.096496	1.170253
H	-1.884481	2.029490	-0.525839
H	-4.279524	1.670850	-0.893044
C	-5.780182	-0.519208	-0.082529
C	-6.679799	0.245548	-0.714255
H	-7.717882	-0.062318	-0.787541
H	-6.122840	-1.458461	0.353220
H	-6.424877	1.194245	-1.177841

### STYRX, PCM-CCl<sub>4</sub> “aaaa”

C	0.125597	0.175253	-0.742253
C	-0.812976	-0.118064	0.440207
Si	-2.629024	-0.006416	-0.000407
O	-3.439387	-0.323457	1.411277
C	-4.863229	-0.329107	1.510980
O	-3.057811	1.458785	-0.640682
C	-2.921334	2.683431	0.071590
O	-3.067341	-1.045368	-1.211629
C	-2.918595	-2.457180	-1.111424
H	-0.614998	0.574325	1.269005
H	-0.624103	-1.121364	0.844641
H	-5.129879	-0.551532	2.547618
H	-5.287271	0.644555	1.239231
H	-5.306359	-1.093484	0.862109
H	-3.487193	3.452000	-0.461803
H	-3.313251	2.611257	1.093852
H	-1.871866	2.999884	0.122975
H	-3.522931	-2.920110	-1.896463
H	-1.873353	-2.758566	-1.257377
H	-3.259091	-2.840327	-0.141171
H	-0.096872	-0.524085	-1.559940
H	-0.109229	1.166517	-1.152360
C	1.618444	0.118720	-0.457816
C	2.519948	0.470545	-1.473595
C	3.894504	0.425170	-1.266389
C	4.434453	0.025623	-0.030390
C	2.153767	-0.277599	0.775054
C	3.530374	-0.323954	0.985932
H	2.136034	0.785488	-2.442478
H	4.568563	0.704142	-2.073980
H	1.491529	-0.555929	1.588946
H	3.901384	-0.636936	1.957994
C	5.896628	-0.005106	0.135179
C	6.588878	-0.361615	1.224623
H	7.674076	-0.343141	1.224839
H	6.456430	0.302814	-0.748704
H	6.112277	-0.682072	2.146836

## ICS

### ICS, gas phase “---a”

C	-1.399174	1.891333	-1.514436
C	-0.549739	0.618361	-1.692717
Si	0.504708	0.101530	-0.235444
O	1.412578	1.433478	0.159195
C	2.192114	1.573077	1.349674
O	-0.462502	-0.372364	1.015906
C	-0.230622	-1.437269	1.950711
O	1.479534	-1.201026	-0.571494
C	2.564403	-1.187082	-1.502580
H	0.118137	0.759087	-2.552741
H	-1.197794	-0.227346	-1.953636
H	1.573233	1.342631	2.227331
H	3.025594	0.856847	1.335594
H	0.843011	-1.541161	2.149490
H	-0.715054	-1.137268	2.886505
H	3.161598	-0.273999	-1.375097
H	2.167670	-1.181302	-2.527618
H	-0.744860	2.748173	-1.316267
H	-1.935953	2.107486	-2.446099
C	-2.419944	1.829713	-0.370235
H	-1.912766	1.702391	0.591151
H	-2.973747	2.773751	-0.337088
N	-3.386799	0.762170	-0.556088
C	-3.637202	-0.312526	-0.068965
O	-4.004788	-1.370309	0.311603
C	-0.805145	-2.754131	1.449994
H	-0.680510	-3.533049	2.210846
H	-1.870880	-2.648764	1.230445
H	-0.286496	-3.073473	0.541916
C	3.426155	-2.419561	-1.282888
H	4.254370	-2.440701	-1.998873
H	3.841953	-2.423206	-0.270968
H	2.832601	-3.329256	-1.412042
C	2.721919	2.995062	1.434482
H	1.895764	3.711279	1.466313
H	3.328959	3.124744	2.336572
H	3.341684	3.226437	0.563167

### ICS, gas phase “-+---”

C	1.545575	-0.762005	2.337183
C	0.143024	-0.138776	2.174586
Si	-0.536440	0.012254	0.431733
O	-0.164864	1.523406	-0.132211
C	0.059211	1.893946	-1.494374
O	0.111394	-1.208463	-0.464899
C	-0.413043	-1.793217	-1.662063
O	-2.183337	-0.212319	0.382750
C	-3.155002	0.763706	0.760899
H	0.116671	0.848556	2.654268
H	-0.572337	-0.761582	2.766990
H	0.695731	1.146832	-1.985354
H	-0.898387	1.921350	-2.034026
H	-1.197487	-1.157910	-2.091970
H	0.414190	-1.841902	-2.378256
H	-2.901488	1.737861	0.322687
H	-3.152082	0.885063	1.854046
H	1.717358	-0.977624	3.400307
H	1.591958	-1.721845	1.811344
C	2.727028	0.110992	1.888869
H	2.653387	1.098730	2.355484
H	3.662316	-0.340064	2.243546
N	2.828345	0.321032	0.461842
C	2.928705	-0.288232	-0.572195
O	3.042241	-0.738499	-1.660574
C	-0.966590	-3.181504	-1.378917
H	-1.311525	-3.652134	-2.306277
H	-0.194430	-3.817819	-0.936229
H	-1.809686	-3.121989	-0.684726
C	-4.527586	0.308355	0.292341
H	-5.293192	1.034864	0.584344
H	-4.546525	0.203016	-0.796505
H	-4.781704	-0.659850	0.733515
C	0.724848	3.259593	-1.533473
H	1.685723	3.226994	-1.012391
H	0.898855	3.571104	-2.568815
H	0.093910	4.009411	-1.046791



## ICS, gas phase “-+aa”

C	1.583226	-0.796124	-1.629537
C	0.079302	-1.119783	-1.518607
Si	-0.979656	0.000761	-0.447251
O	-0.379724	-0.082819	1.101672
C	-1.043029	0.470519	2.246499
O	-1.012738	1.556376	-0.997160
C	-0.095768	2.602051	-0.674715
O	-2.569951	-0.447825	-0.504765
C	-3.054612	-1.774097	-0.293731
H	-0.050124	-2.160144	-1.189044
H	-0.367480	-1.071786	-2.520256
H	-0.863242	1.553385	2.283353
H	-2.126619	0.321254	2.158541
H	0.475983	2.350251	0.228683
H	0.622732	2.706826	-1.498662
H	-2.694681	-2.162867	0.669857
H	-2.675404	-2.442677	-1.079569
H	2.027991	-1.363488	-2.455577
H	1.732307	0.262491	-1.872925
C	2.362079	-1.125838	-0.352524
H	1.960182	-0.566642	0.497716
H	2.248919	-2.193222	-0.122882
N	3.773097	-0.838284	-0.530985
C	4.739485	-0.540993	0.124899
O	5.759300	-0.246490	0.645839
C	-0.855384	3.904511	-0.475475
H	-0.159983	4.725137	-0.268999
H	-1.428667	4.152113	-1.373236
H	-1.554374	3.821250	0.362191
C	-4.574109	-1.754004	-0.316036
H	-4.972755	-2.762188	-0.162124
H	-4.960886	-1.102322	0.472730
H	-4.936438	-1.378487	-1.277125
C	-0.511114	-0.195499	3.504405
H	0.568220	-0.041594	3.595665
H	-0.995370	0.224111	4.392325
H	-0.702354	-1.272452	3.481999

## ICS, gas phase “-aaa”

C	1.861631	-0.172129	-0.104948
C	0.630530	-0.904804	-0.664814
Si	-1.017800	-0.068275	-0.344811
O	-1.076106	0.175281	1.294472
C	-2.210251	0.709693	1.987384
O	-1.226311	1.339153	-1.182463
C	-0.614879	2.594038	-0.883166
O	-2.284234	-0.985920	-0.880803
C	-2.514888	-2.344251	-0.508066
H	0.582472	-1.916896	-0.238022
H	0.728844	-1.039597	-1.750847
H	-2.243324	1.798535	1.846833
H	-3.137020	0.297149	1.567629
H	-0.397891	2.672253	0.191050
H	0.343000	2.664438	-1.416777
H	-2.478244	-2.452698	0.585408
H	-1.727656	-2.986587	-0.928035
H	1.950711	0.826264	-0.548815
H	1.756440	-0.026641	0.975239
C	3.157855	-0.393177	-0.378688
H	3.103589	-1.934177	0.079769
H	3.285471	-1.082324	-1.459378
N	4.301687	-0.224320	0.153416
C	5.501979	-0.322313	0.208210
O	6.676763	-0.295201	0.334001
C	-1.538465	3.720135	-1.320039
H	-1.073484	4.692942	-1.127957
H	-1.754225	3.642541	-2.389374
H	-2.486968	3.673618	-0.776862
C	-3.874035	-2.777193	-1.032993
H	-4.071687	-3.821521	-0.769452
H	-4.665970	-2.154353	-0.607156
H	-3.911128	-2.679076	-2.121607
C	-2.096769	0.372288	3.464718
H	-1.172248	0.782146	3.881823
H	-2.943045	0.791391	4.019064
H	-2.087377	-0.711542	3.613659

ICS, PCM-H<sub>2</sub>O “-+a”

C	-1.200106	-1.832688	-1.905807
C	0.192283	-1.197819	-1.713059
Si	0.686900	-0.423956	-0.075306
O	0.106579	1.127216	-0.015095
C	0.497013	2.098429	0.969316
O	0.264653	-1.323983	1.250511
C	-0.971758	-1.262690	1.976381
O	2.345796	-0.426730	0.059123
C	3.213002	0.096399	-0.956822
H	0.372791	-0.461993	-2.509695
H	0.942646	-1.985861	-1.870406
H	-0.025433	1.890632	1.912605
H	1.574438	2.021292	1.163074
H	-1.258245	-0.216616	2.142438
H	-1.768849	-1.729627	1.385612
H	2.919829	1.124518	-1.210536
H	3.123958	-0.508226	-1.870030
H	-1.181306	-2.448211	-2.814253
H	-1.434024	-2.513779	-1.079031
C	-2.361093	-0.851126	-2.068107
H	-2.123177	-0.120310	-2.868162
H	-3.250325	-1.405756	-2.406774
N	-2.698889	-0.157273	-0.854252
C	-2.941611	0.944680	-0.420268
O	-3.211294	1.967717	0.094238
C	-0.803690	-1.982454	3.304064
H	-1.742726	-1.960532	3.866915
H	-0.522390	-3.027925	3.143807
H	-0.026436	-1.505142	3.908909
C	4.646189	0.070836	-0.452425
H	5.324064	0.464516	-1.216826
H	4.749487	0.681849	0.449606
H	4.952589	-0.952151	-0.212808
C	0.139196	3.487490	0.468969
H	-0.935427	3.553452	0.277694
H	0.408909	4.241198	1.216301
H	0.672825	3.714859	-0.459268

## ICS, gas phase “-a-a”

C	-1.630311	-1.392566	0.006971
C	-0.971939	-0.188965	-0.687505
Si	0.858626	0.035813	-0.352716
O	1.015876	0.008483	1.298578
C	2.230958	0.296245	1.999011
O	1.810546	-1.109723	-1.070550
C	1.985417	-2.450276	-0.615024
O	1.439749	1.436470	-1.009814
C	0.863003	2.727665	-0.812242
H	-1.495258	0.724299	-0.375653
H	-1.107254	-0.261212	-1.775722
H	2.877569	-0.591866	1.987326
H	2.775898	1.104247	1.493682
H	1.819188	-2.515798	0.469178
H	1.239911	-3.095922	-1.100418
H	0.734598	2.928903	0.260887
H	-0.133214	2.768185	-1.274760
H	-1.165185	-2.328957	-0.325554
H	-1.474899	-1.328621	1.089160
C	-3.131535	-1.511119	-0.271213
H	-3.305431	-1.609035	-1.350631
H	-3.528918	-2.412385	0.211778
N	-3.847124	-0.352341	0.229683
C	-4.951418	0.130021	0.188358
O	-5.981976	0.706847	0.229099
C	3.387746	-2.923836	-0.964572
H	3.529283	-3.964958	-0.655183
H	3.554642	-2.855585	-2.043252
H	4.139478	-2.306342	-0.464138
C	1.770357	3.774527	-1.437498
H	1.345768	4.775599	-1.308355
H	2.759854	3.753058	-0.971641
H	1.892855	3.582980	-2.507279
C	1.901603	0.688460	3.429916
H	1.356814	-0.116165	3.932462
H	2.819088	0.890233	3.992758
H	1.277475	1.586781	3.449892

ICS, PCM-H<sub>2</sub>O “-aaa”

C	-1.858017	-0.416493	1.487094
C	-0.950667	0.672592	0.877436
Si	0.602376	0.105071	-0.005671
O	1.647935	-0.407828	1.176625
C	3.045957	-0.672105	0.966243
O	0.329904	-1.078566	-1.133075
C	0.604626	-2.481177	-0.996493
O	1.251169	1.334962	-0.913386
C	1.499834	2.651906	-0.395776
H	-0.620620	1.349473	1.676545
H	-1.538169	1.292673	0.187682
H	3.176522	-1.743776	0.768003
H	3.407277	-0.123718	0.086986
H	0.989611	-2.700575	0.006711
H	-0.344635	-3.019966	-1.109119
H	2.265106	2.602002	0.390441
H	0.585001	3.057389	0.057102
H	-1.260902	-1.170516	2.016004
H	-2.522836	0.029583	2.237409
C	-2.726918	-1.158228	0.462821
H	-2.101725	-1.603491	-0.314053
H	-3.281050	-1.966484	0.953290
N	-3.672750	-0.280581	-0.207783
C	-4.670541	0.357173	-0.008752
O	-5.650833	1.024902	0.043100
C	1.592179	-2.938159	-2.059047
H	1.751851	-4.019251	-1.984971
H	1.211266	-2.713795	-3.060028
H	2.558307	-2.437960	-1.938912
C	1.965909	3.555416	-1.524650
H	2.168860	4.561382	-1.143283
H	2.882665	3.166222	-1.978191
H	1.199838	3.628401	-2.302710
C	3.834030	-0.267595	2.201153
H	3.468076	-0.801561	3.083569
H	4.894104	-0.507843	2.068177
H	3.743681	0.807137	2.386934

ICS, PCM-H<sub>2</sub>O “+aa”

C	-1.553044	0.043773	-1.667725
C	-0.076957	-0.400190	-1.743775
Si	1.082431	0.011869	-0.326702
O	0.279348	-0.288105	1.093994
C	0.830508	-0.003627	2.392601
O	2.494666	-0.852985	-0.434790
C	2.564989	-2.278762	-0.284441
O	1.668399	1.561659	-0.378962
C	0.850938	2.737983	-0.468803
H	0.383838	0.047711	-2.634835
H	-0.035104	-1.485563	-1.916265
H	1.504221	-0.820801	2.681119
H	1.423665	0.918640	2.351685
H	2.119585	-2.576264	0.674290
H	1.987819	-2.765350	-1.083060
H	0.108226	2.743001	0.340666
H	0.305941	2.736843	-1.422728
H	-1.635048	1.100017	-1.385361
H	-2.012591	-0.045970	-2.659717
C	-2.371947	-0.794482	-0.683548
H	-2.327044	-1.852961	-0.969683
H	-1.966349	-0.699630	0.327184
N	-3.764178	-0.372713	-0.693191
C	-4.792373	-0.498286	-0.087732
O	-5.868781	-0.533760	0.412544
C	4.018894	-2.714526	-0.349732
H	4.093750	-3.801634	-0.243593
H	4.464246	-2.429348	-1.307806
H	4.598348	-2.247859	0.452779
C	1.737767	3.968142	-0.376511
H	1.132205	4.877209	-0.451036
H	2.274856	3.987182	0.576769
H	2.472579	3.974751	-1.187370
C	-0.298703	0.135862	3.398876
H	-0.891197	-0.783192	3.445105
H	0.106140	0.334893	4.396669
H	-0.962345	0.961358	3.123807



ICS, PCM-H<sub>2</sub>O “+a--”

C	-1.674619	-0.079511	1.276347
C	-0.830453	-0.892053	0.279200
Si	0.827045	-0.159745	-0.194286
O	1.522906	0.357528	1.217859
C	2.789433	1.034257	1.290393
O	1.776494	-1.257818	-1.000029
C	2.284284	-2.453647	-0.388904
O	0.747007	1.054531	-1.320374
C	-0.028131	2.252022	-1.167776
H	-1.397565	-1.061639	-0.645805
H	-0.641662	-1.890444	0.700618
H	3.593441	0.286851	1.307501
H	2.934727	1.658093	0.399294
H	2.840959	-2.200231	0.523229
H	1.449016	-3.105465	-0.097470
H	0.119449	2.675332	-0.165038
H	-1.095151	2.013406	-1.274447
H	-1.103101	0.093013	2.195805
H	-1.916182	0.909664	0.868372
C	-2.982883	-0.769825	1.683807
H	-2.764232	-1.747433	2.125896
H	-3.508065	-0.173116	2.438500
N	-3.872934	-0.998325	0.557603
C	-4.650061	-0.432177	-0.161344
O	-5.440409	-0.003852	-0.937060
C	3.188191	-3.172353	-1.375912
H	3.581965	-4.090919	-0.928787
H	2.635356	-3.438704	-2.282023
H	4.032368	-2.536904	-1.660711
C	0.389825	3.256596	-2.228349
H	-0.202614	4.172654	-2.136185
H	1.447550	3.515448	-2.120056
H	0.235783	2.844975	-2.320553
C	2.832243	1.887763	2.546510
H	2.688740	1.269713	3.438183
H	3.800479	2.392589	2.628792
H	2.045685	2.648277	2.524679

ICS, PCM-H<sub>2</sub>O “aaaa”

C	-1.860430	0.001321	-0.134760
C	-0.585938	0.009740	-0.996584
Si	1.012907	0.015872	-0.024715
O	2.216396	0.002634	-1.166271
C	3.617555	0.017449	-0.838697
O	1.146401	1.306246	1.008570
C	1.128570	2.670844	0.560220
O	1.143743	-1.250174	1.037159
C	1.098175	-2.627843	0.634284
H	-0.584507	0.889801	-1.655594
H	-0.578954	-0.860230	-1.668413
H	3.883315	0.997387	-0.421391
H	3.828039	-0.738368	-0.070845
H	1.853456	2.808869	-0.253341
H	0.133605	2.916118	0.163930
H	1.941318	-2.847556	-0.034046
H	0.172386	-2.824783	0.076225
H	-1.878129	-0.881637	0.515110
H	-1.881237	0.876135	0.526136
C	-3.122468	0.004545	-0.999281
H	-3.137423	0.893673	-1.641542
H	-3.135556	-0.878232	-1.650358
N	-4.318370	-0.001981	-0.171678
C	-5.516040	0.019052	-0.257902
O	-6.701084	0.035791	-0.196591
C	1.467806	3.582863	1.726827
H	1.445423	4.629808	1.407430
H	0.746212	3.454075	2.539371
H	2.466949	3.360141	2.113558
C	1.161180	-3.509838	1.869605
H	1.133681	-4.565771	1.581834
H	2.084943	-3.327414	2.427201
H	0.312909	-3.309028	2.531026
C	4.427450	-0.261771	-2.092932
H	4.228910	0.495275	-2.857902
H	5.497746	-0.247088	-1.862743
H	4.175998	-1.243653	-2.505517

ICS, PCM-CCl<sub>4</sub> “----”

C	-1.459857	1.201205	-1.474155
C	-0.557320	-0.045435	-1.394890
Si	0.794441	-0.020397	-0.095012
O	1.681374	1.341806	-0.400508
C	2.717263	1.833016	0.461958
O	0.237411	0.077090	1.457488
C	-0.026238	-1.018993	2.344017
O	1.695784	-1.412624	-0.151720
C	2.363788	-1.916739	-1.311333
H	-0.068027	-0.187037	-2.367800
H	-1.172908	-0.940821	-1.238965
H	2.359663	1.842309	1.498932
H	3.583067	1.158930	0.413734
H	-0.018424	-1.967919	1.795146
H	0.786944	-1.064100	3.079324
H	2.997904	-1.134717	-1.751327
H	1.624447	-2.208942	-2.070181
H	-0.842480	2.102649	-1.565643
H	-2.083582	1.153903	-2.376311
C	-2.380600	1.402708	-0.262690
H	-1.790277	1.479860	0.652986
H	-2.948084	2.333507	-0.376697
N	-3.310913	0.301721	-0.074174
C	-4.292197	-0.184904	-0.572365
O	-5.254197	-0.762989	-0.952273
C	-1.361834	-0.811764	3.040709
H	-1.549610	-1.626207	3.748933
H	-1.364559	0.132065	3.594752
H	-2.178973	-0.787674	2.313445
C	3.208476	-3.116772	-0.915851
H	3.724324	-3.525913	-1.790786
H	3.958378	-2.830439	-0.172581
H	2.580775	-3.902217	-0.484488
C	3.111455	3.231673	0.018342
H	2.255797	3.910828	0.079551
H	3.910495	3.622759	0.656897
H	3.468750	3.222446	-1.015883

ICS, PCM-CCl<sub>4</sub> “+--”

C	1.677643	-0.750905	2.288450
C	0.247808	-0.178531	2.188806
Si	-0.512156	-0.027007	0.481094
O	-0.213514	1.503970	-0.078479
C	-0.105821	1.897631	-1.449241
O	0.135187	-1.217368	-0.456843
C	-0.408071	-1.775486	-1.658726
O	-2.151698	-0.301188	0.495255
C	-3.145506	0.650491	0.882192
H	0.202872	0.799504	2.686477
H	-0.421400	-0.835835	2.759414
H	0.555302	1.204710	-1.985320
H	-1.093763	1.854688	-1.930046
H	-1.220037	-1.147488	-2.048175
H	0.400299	-1.780265	-2.398666
H	-2.846791	1.658902	0.567594
H	-3.233333	0.659343	1.978135
H	1.897458	-0.971672	3.341502
H	1.740351	-1.702656	1.749614
C	2.809532	0.169216	1.809257
H	2.713193	1.152706	2.280816
H	3.772843	-0.246523	2.129596
N	2.860213	0.384511	0.379335
C	2.924899	-0.198225	-0.669885
O	3.003575	-0.624636	-1.773369
C	-0.920804	-3.185292	-1.407200
H	-1.278474	-3.634073	-2.340586
H	-0.122436	-3.816393	-1.005038
H	-1.746560	-3.168739	-0.690115
C	-4.477481	0.270053	0.255701
H	-5.258816	0.974291	0.560717
H	-4.407435	0.281353	-0.836269
H	-4.775959	-0.734949	0.568810
C	0.449836	3.310457	-1.517438
H	1.440397	3.351253	-1.055465
H	0.536609	3.636995	-2.559032
H	-0.206232	4.008701	-0.988646

ICS, PCM-CCl<sub>4</sub> “++a-”

C	-1.881163	-0.416718	-1.148818
C	-0.426992	-0.119747	-1.567079
Si	0.846103	-0.038629	-0.190790
O	0.707297	-1.458957	0.638098
C	1.353444	-1.738788	1.888263
O	2.366198	0.217750	-0.803600
C	3.032980	-0.634275	-1.739457
O	0.594998	1.194783	0.882243
C	1.031498	2.549149	0.719376
H	-0.383412	0.818112	-2.138963
H	-0.099607	-0.901451	-2.265382
H	2.442076	-1.754703	1.743481
H	1.125093	-0.940235	2.604213
H	3.053038	-1.665097	-1.360427
H	2.485321	-0.643391	-2.692438
H	0.409578	3.045207	-0.039458
H	2.067132	2.563599	0.360781
H	-2.494073	-0.596058	-2.041873
H	-1.917691	-1.331688	-0.547642
C	-2.520858	0.727048	-0.351104
H	-1.958250	0.920181	0.566054
H	-2.498301	1.655169	-0.937318
N	-3.895067	0.460348	0.021975
C	-4.868221	-0.135719	-0.358474
O	-5.894136	-0.673310	-0.610422
C	4.448323	-0.124621	-1.955092
H	4.977469	-0.759277	-2.673539
H	4.433129	0.898739	-2.341776
H	5.005022	-0.127302	-1.013360
C	0.911913	3.272972	2.049953
H	1.230136	4.315644	1.946184
H	-0.122722	3.260896	2.405753
H	1.539944	2.792452	2.806044
C	0.868057	-3.081439	2.607169
H	1.100135	-3.878475	1.694299
H	1.351971	-3.318100	3.360546
H	-0.214682	-3.065148	2.563150

ICS, PCM-CCl<sub>4</sub> “+a--”

C	-1.678592	-0.128273	1.237925
C	-0.831982	-0.895019	0.208286
Si	0.817646	-0.125593	-0.243774
O	1.499696	0.331507	1.198495
C	2.774727	0.978114	1.308382
O	1.779524	-1.180004	-1.081038
C	2.232465	-2.430397	-0.555168
O	0.713685	1.146388	-1.296794
C	-0.057086	2.328234	-1.065477
H	-1.404727	-1.043292	-0.716092
H	-0.627546	-1.904031	0.595113
H	3.570941	0.261031	1.068499
H	2.842858	1.797489	0.580955
H	2.632035	-2.295509	0.459775
H	1.388230	-3.131003	-0.485911
H	0.052506	2.661126	-0.023636
H	-1.121236	2.110670	-1.232681
H	-1.099834	0.023591	2.156191
H	-1.934462	0.870370	0.862734
C	-2.977177	-0.846877	1.628645
H	-2.742185	-1.836011	2.035238
H	-3.500401	-0.285559	2.411675
N	-3.869527	-1.047883	0.501466
C	-4.693639	-0.456042	-0.146145
O	-5.521481	0.002165	-0.858516
C	3.305000	-2.997287	-1.470333
H	3.655352	-3.963999	-1.093861
H	2.909924	-3.140308	-2.480320
H	4.159383	-2.316671	-1.530122
C	0.406065	3.420771	-2.015078
H	-0.185907	4.330146	-1.868074
H	1.459720	3.659590	-1.842985
H	0.293441	3.097137	-3.053828
C	2.945592	1.506496	2.722633
H	2.886635	0.689885	3.448492
H	3.918461	1.996921	2.831867
H	2.162832	2.233131	2.960347

ICS, PCM-CCl<sub>4</sub> “aaaa”

C	-1.866674	-0.009364	-0.130511
C	-0.596101	-0.005945	-0.997954
Si	1.003740	0.030780	-0.022092
O	2.208192	0.035486	-1.161277
C	3.604236	0.055806	-0.831381
O	1.103473	1.335480	0.991326
C	1.091864	2.688556	0.527208
O	1.149579	-1.237185	1.030310
C	1.083275	-2.608915	0.630908
H	-0.601898	0.863949	-1.669765
H	-0.583964	-0.883574	-1.659106
H	3.892360	1.074558	-0.540221
H	3.800037	-0.598491	0.028281
H	1.732587	2.795728	-0.359422
H	0.070793	2.967246	0.229792
H	1.885000	-2.831057	-0.086971
H	0.127348	-2.810549	0.126311
H	-1.881575	-0.886364	0.526808
H	-1.882519	0.866694	0.527958
C	-3.136382	-0.009757	-0.984279
H	-3.151819	0.875730	-1.632275
H	-3.150436	-0.893611	-1.634628
N	-4.322966	-0.013345	-0.149667
C	-5.523112	0.010620	-0.229129
O	-6.705348	0.030274	-0.164638
C	1.579163	3.602119	1.639624
H	1.559430	4.646904	1.312357
H	0.942256	3.502628	2.523336
H	2.603669	3.346272	1.925924
C	1.218919	-3.492016	1.859820
H	1.172621	-4.548628	1.576341
H	2.173543	-3.308278	2.361550
H	0.413363	-3.286614	2.570766
C	4.410637	-0.398383	-2.036415
H	4.217500	0.250349	-2.896091
H	5.482014	-0.366146	-1.812385
H	4.145961	-1.423030	-2.314714



## 8. Appendix B. Amino acid protonation: glycine

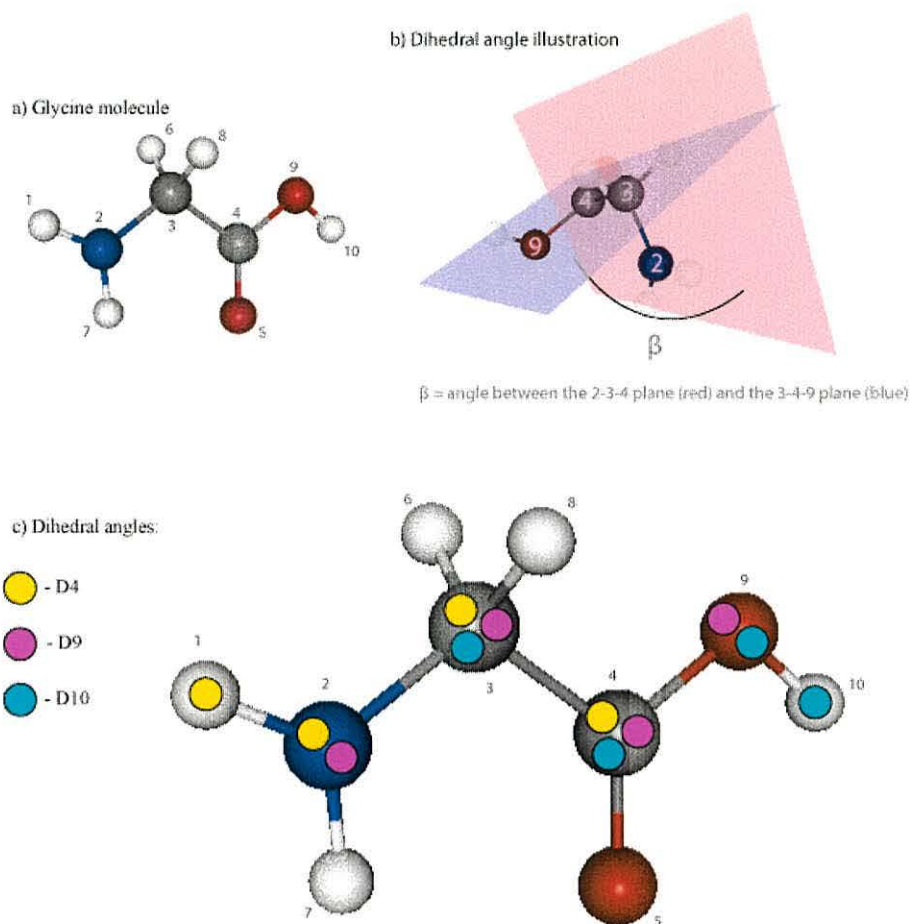
---

**Naming notations.** Three main dihedral angles of glycine: D4, D9 and D10 (Fig. B.1). D4 and D9 had three possible values: +60, 180, -60, where D4 was generated using the rule depicted in Figure B.2. D10 had four possible values: 0, +60, 180, -60. Protons were placed at a distance of 1.5 Å from the molecule.

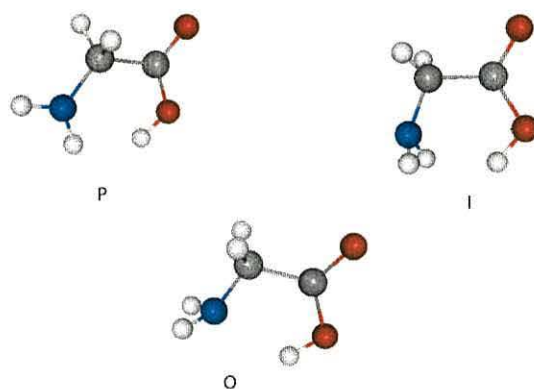
The glycine code-names consist of a *conformation* and a *protonation* part. The conformation part consists of three letters describing dihedral angles D4, D9 and D10 ('a' for 'anti' (180), 'p' for 'plus' (+60), 'm' for 'minus' (-60), and 's' for 'syn' (0)). The protonation part consists of a description of the molecular structure:

- *hn-g-oh*: non-protonated glycine;
- *h2n-g-oh*: glycine with one proton added close to the nitrogen (#2);
- *hn-hg-oh*: a proton near the  $\alpha$ -carbon;
- *hn-g-oh2*: protonated at the oxygen (#10);
- *hn-gh1-oh*: a proton on one side of the oxygen (#5);
- *hn-gh2-oh*: a proton on the other side of the oxygen (#5).

All energies (denoted as kJ for simplicity) are in  $\text{kJ}\cdot\text{mol}^{-1}$ . Energies in the “Thermodynamics” sections are in Hartrees. In the “Normal modes of vibration” sections, the negative frequencies indicate transition states of different orders.



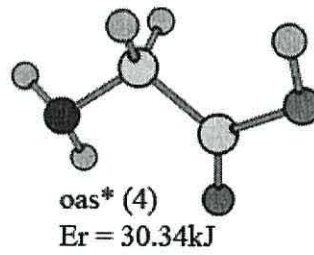
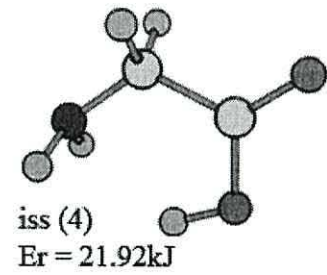
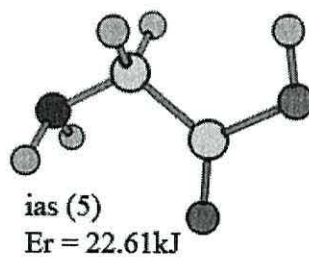
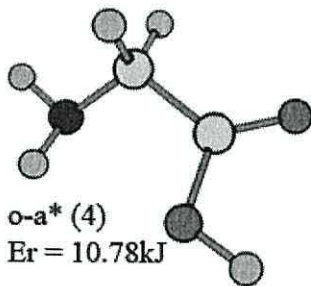
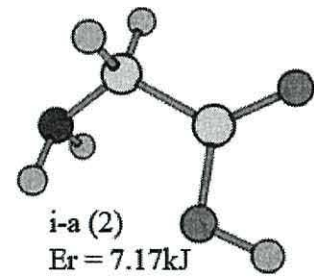
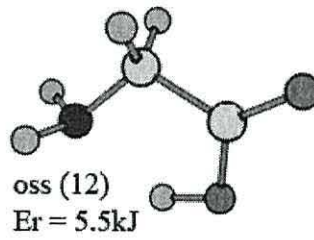
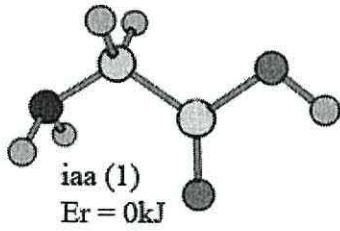
**Figure B.1** A computer representation of a glycine molecule with atom numbering (a), illustration of a dihedral angle (b), and three colours (yellow, magenta and cyan) which mark the atoms that define the three dihedral angles which were used to create 36 glycine structures used in the calculations (c) – three angles for D4 and D9, and four angles for D10.



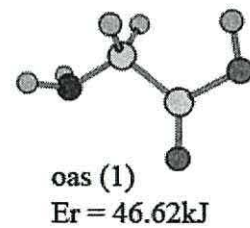
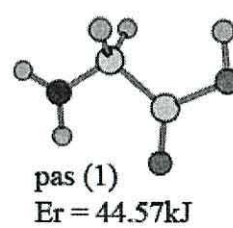
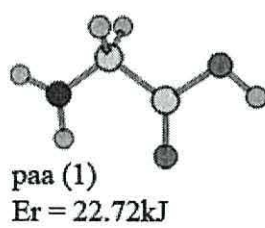
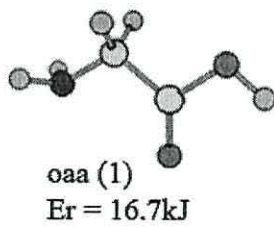
**Figure B.2** An illustration of the “planar” (P), “inner” (I), and “outer” (O) configurations of the amino-end of glycine.

# hn-g-oh

Conformer classes.

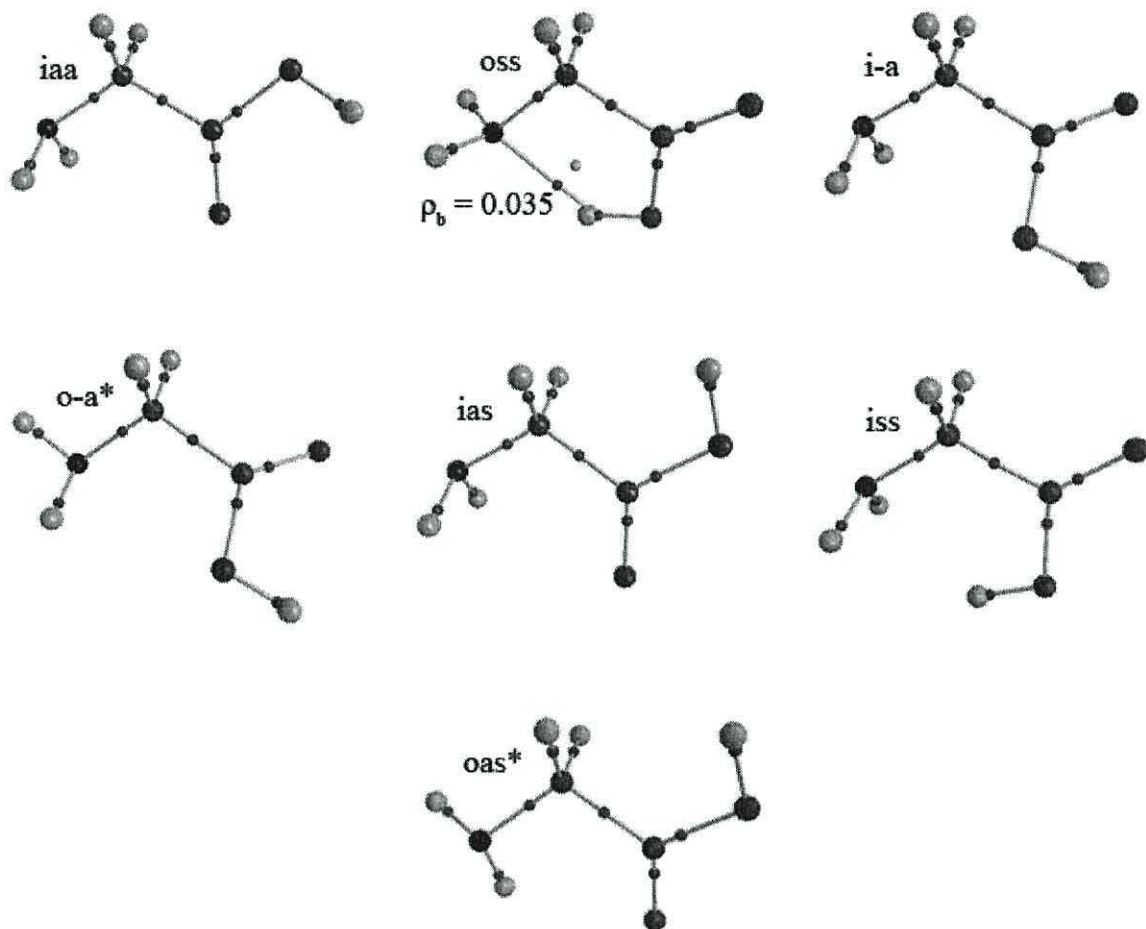


Non-minimised conformers:

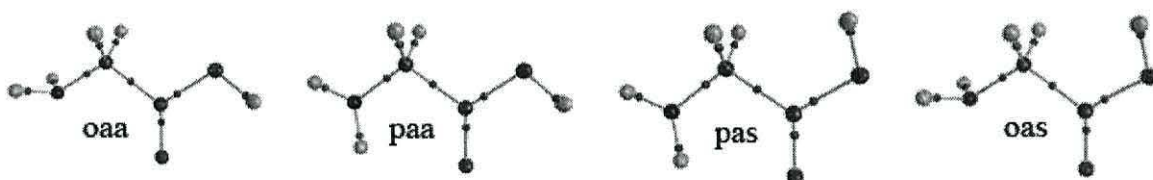




*AIM analysis.*



Non-minimised conformers:



Calculation results.

Code name	Finished as	Relative free energy, kJ·mol <sup>-1</sup>
hn-g-oh-l_iaa	iaa	0
hn-g-oh-l_o++	oss	5.117099009
hn-g-oh-l_o--	oss	5.361270485
hn-g-oh-l_p+	oss	5.487294473
hn-g-oh-l_p+-	oss	5.487294473
hn-g-oh-l_p--	oss	5.505672972
hn-g-oh-l_p++	oss	5.505672972
hn-g-oh-l_o-s	oss	5.52142597
hn-g-oh-l_o+s	oss	5.52142597
hn-g-oh-l_o++	oss	5.615943961
hn-g-oh-l_o+-	oss	5.615943961
hn-g-oh-l_p-s	oss	5.639573459
hn-g-oh-l_p+s	oss	5.639573459
hn-g-oh-l_i-a	i-a	7.167614312
hn-g-oh-l_i+a	i-a	7.167614312
hn-g-oh-l_o-a	o-a*	10.77505097
hn-g-oh-l_o+a	o-a*	10.77505097
hn-g-oh-l_p+a	o-a*	10.77767647
hn-g-oh-l_p-a	o-a*	10.77767647
hn-g-oh-l_oaa	oaa	16.7008039
hn-g-oh-l_ias	ias	22.59767633
hn-g-oh-l_i++	ias	22.61342933
hn-g-oh-l_i--	ias	22.61342933
hn-g-oh-l_ia-	ias	22.61868033
hn-g-oh-l_ia+	ias	22.61868033
hn-g-oh-l_paa	paa	22.71844932
hn-g-oh-l_i-+	iss	27.91431332
hn-g-oh-l_i+-	iss	27.91431332
hn-g-oh-l_i+s	iss	27.91693882
hn-g-oh-l_i-s	iss	27.91693882
hn-g-oh-l_pa+	oas*	30.36390459
hn-g-oh-l_pa-	oas*	30.36390459
hn-g-oh-l_oa+	oas*	30.36915559
hn-g-oh-l_oa-	oas*	30.36915559
hn-g-oh-l_pas	pas	44.56523272
hn-g-oh-l_oas	oas	46.61574803
		conformer class
		non-minimised conformer

## Thermodynamic Parameters.

### iaa

Zero-point correction= 0.080669 (Hartree/Particle)  
Thermal correction to Energy= 0.086318  
Thermal correction to Enthalpy= 0.087262  
Thermal correction to Gibbs Free Energy= 0.051451  
Sum of electronic and zero-point Energies= -283.821284  
Sum of electronic and thermal Energies= -283.815635  
Sum of electronic and thermal Enthalpies= -283.814691  
Sum of electronic and thermal Free Energies= -283.850502

### oss

Zero-point correction= 0.081259 (Hartree/Particle)  
Thermal correction to Energy= 0.086595  
Thermal correction to Enthalpy= 0.087539  
Thermal correction to Gibbs Free Energy= 0.052546  
Sum of electronic and zero-point Energies= -283.819840  
Sum of electronic and thermal Energies= -283.814505  
Sum of electronic and thermal Enthalpies= -283.813560  
Sum of electronic and thermal Free Energies= -283.848553

### i-a

Zero-point correction= 0.080979 (Hartree/Particle)  
Thermal correction to Energy= 0.086506  
Thermal correction to Enthalpy= 0.087450  
Thermal correction to Gibbs Free Energy= 0.051905  
Sum of electronic and zero-point Energies= -283.818698  
Sum of electronic and thermal Energies= -283.813172  
Sum of electronic and thermal Enthalpies= -283.812228  
Sum of electronic and thermal Free Energies= -283.847772

### o-a\*

Zero-point correction= 0.080858 (Hartree/Particle)  
Thermal correction to Energy= 0.086366  
Thermal correction to Enthalpy= 0.087310  
Thermal correction to Gibbs Free Energy= 0.052069  
Sum of electronic and zero-point Energies= -283.817610  
Sum of electronic and thermal Energies= -283.812102  
Sum of electronic and thermal Enthalpies= -283.811158  
Sum of electronic and thermal Free Energies= -283.846398

### ias

Zero-point correction= 0.080203 (Hartree/Particle)  
Thermal correction to Energy= 0.085997  
Thermal correction to Enthalpy= 0.086941  
Thermal correction to Gibbs Free Energy= 0.050933  
Sum of electronic and zero-point Energies= -283.812625  
Sum of electronic and thermal Energies= -283.806831  
Sum of electronic and thermal Enthalpies= -283.805887  
Sum of electronic and thermal Free Energies= -283.841895

### iss

Zero-point correction= 0.080127 (Hartree/Particle)  
Thermal correction to Energy= 0.085779  
Thermal correction to Enthalpy= 0.086724  
Thermal correction to Gibbs Free Energy= 0.050954  
Sum of electronic and zero-point Energies= -283.810698  
Sum of electronic and thermal Energies= -283.805045  
Sum of electronic and thermal Enthalpies= -283.804101  
Sum of electronic and thermal Free Energies= -283.839870

### oas\*

Zero-point correction= 0.080412 (Hartree/Particle)  
Thermal correction to Energy= 0.086011  
Thermal correction to Enthalpy= 0.086955  
Thermal correction to Gibbs Free Energy= 0.051712  
Sum of electronic and zero-point Energies= -283.810237  
Sum of electronic and thermal Energies= -283.804638  
Sum of electronic and thermal Enthalpies= -283.803694  
Sum of electronic and thermal Free Energies= -283.838937

### oaa

Zero-point correction= 0.079692 (Hartree/Particle)  
Thermal correction to Energy= 0.084939  
Thermal correction to Enthalpy= 0.085884  
Thermal correction to Gibbs Free Energy= 0.050356  
Sum of electronic and zero-point Energies= -283.814806  
Sum of electronic and thermal Energies= -283.809558  
Sum of electronic and thermal Enthalpies= -283.808614  
Sum of electronic and thermal Free Energies= -283.844141

### paa

Zero-point correction= 0.078496 (Hartree/Particle)  
Thermal correction to Energy= 0.083427  
Thermal correction to Enthalpy= 0.084371  
Thermal correction to Gibbs Free Energy= 0.050407  
Sum of electronic and zero-point Energies= -283.813759  
Sum of electronic and thermal Energies= -283.808829  
Sum of electronic and thermal Enthalpies= -283.807884  
Sum of electronic and thermal Free Energies= -283.841849

### pas

Zero-point correction= 0.077879 (Hartree/Particle)  
Thermal correction to Energy= 0.083003  
Thermal correction to Enthalpy= 0.083947  
Thermal correction to Gibbs Free Energy= 0.049619  
Sum of electronic and zero-point Energies= -283.805267  
Sum of electronic and thermal Energies= -283.800144  
Sum of electronic and thermal Enthalpies= -283.799200  
Sum of electronic and thermal Free Energies= -283.833528

### oas

Zero-point correction= 0.078962 (Hartree/Particle)  
Thermal correction to Energy= 0.083540  
Thermal correction to Enthalpy= 0.084484  
Thermal correction to Gibbs Free Energy= 0.051350  
Sum of electronic and zero-point Energies= -283.805136  
Sum of electronic and thermal Energies= -283.800558  
Sum of electronic and thermal Enthalpies= -283.799613  
Sum of electronic and thermal Free Energies= -283.832747



*Normal Modes of Vibration (within computational error of  $\pm 25 \text{ cm}^{-1}$ ).*

iaa

Frequencies --	51.1320	238.2370	260.5778
Frequencies --	471.2061	494.2727	623.9057
Frequencies --	641.8367	845.0927	927.3383
Frequencies --	951.9284	1145.5670	1189.8740
Frequencies --	1191.2014	1318.0520	1400.3353
Frequencies --	1421.4785	1476.2430	1679.6918
Frequencies --	1818.7388	3101.5790	3153.8139
Frequencies --	3555.2302	3647.5243	3804.6511

OSS

Frequencies --	69.0609	269.3171	321.0680
Frequencies --	510.4986	547.0298	654.9037
Frequencies --	848.1846	866.8765	911.8326
Frequencies --	957.0072	1101.6857	1170.1815
Frequencies --	1240.6391	1338.0123	1379.3161
Frequencies --	1436.8308	1484.3550	1659.4842
Frequencies --	1842.3067	3106.5903	3167.2376
Frequencies --	3543.0939	3574.3625	3668.8154

i-a

Frequencies --	51.8564	268.0087	293.9114
Frequencies --	470.0761	536.8686	608.6561
Frequencies --	686.4469	829.4018	920.0080
Frequencies --	937.5650	1156.2346	1166.7730
Frequencies --	1225.0011	1357.3200	1384.0418
Frequencies --	1405.9878	1479.5627	1673.8233
Frequencies --	1813.8945	3096.9902	3170.0532
Frequencies --	3556.4527	3652.8792	3803.8852

O-a\*

Frequencies --	75.8206	251.1113	292.2743
Frequencies --	451.3394	549.5374	592.6588
Frequencies --	708.0712	846.0615	867.8723
Frequencies --	1035.9242	1096.2788	1152.5363
Frequencies --	1245.4363	1312.3598	1362.2343
Frequencies --	1449.7987	1512.1871	1653.4273
Frequencies --	1826.4312	3023.1260	3173.7184
Frequencies --	3552.4256	3654.0571	3807.8489

ias

Frequencies --	55.3568	231.3713	263.6624
Frequencies --	365.6611	472.0003	560.2286
Frequencies --	650.8614	850.1506	929.1311
Frequencies --	950.5231	1138.5054	1183.0605
Frequencies --	1188.3917	1297.6871	1398.8813
Frequencies --	1406.8579	1485.8834	1680.3943
Frequencies --	1843.6037	3073.9310	3127.1591
Frequencies --	3554.1339	3645.4265	3852.1406

iss

Frequencies --	46.5876	282.5248	335.1559
Frequencies --	503.3615	506.3681	518.7477
Frequencies --	584.7983	662.6494	851.5012
Frequencies --	908.8809	1140.8063	1166.9916
Frequencies --	1199.9272	1347.2851	1372.4975
Frequencies --	1381.6849	1466.5739	1659.8034

Frequencies --	1839.5053	3113.2756	3169.2593
Frequencies --	3593.1256	3713.7058	3806.5001

oas\*

Frequencies --	99.6940	223.9473	289.7522
Frequencies --	435.9470	470.4981	573.4717
Frequencies --	662.3617	863.1313	872.2536
Frequencies --	1035.4529	1134.2231	1156.1246
Frequencies --	1231.9323	1285.9125	1317.9064
Frequencies --	1455.2828	1514.0693	1642.5555
Frequencies --	1847.5729	3016.9134	3098.6762
Frequencies --	3549.8375	3655.1820	3864.0964

oaa

Frequencies --	-246.0188	30.0694	272.3256
Frequencies --	469.8452	481.0598	604.3659
Frequencies --	636.4912	741.4210	882.0973
Frequencies --	925.7081	1131.3250	1166.1633
Frequencies --	1190.0016	1311.9656	1368.9423
Frequencies --	1426.3046	1476.4873	1669.0220
Frequencies --	1842.4506	3081.0045	3135.3625
Frequencies --	3610.5075	3714.1489	3813.6064

paa

Frequencies --	-790.1083	-110.5970	125.9770
Frequencies --	283.3146	479.8142	484.3957
Frequencies --	611.9248	626.5839	844.5329
Frequencies --	1036.6456	1094.8767	1163.3694
Frequencies --	1235.0606	1248.3729	1330.7679
Frequencies --	1459.4249	1528.0463	1617.3889
Frequencies --	1823.4401	3041.2576	3069.0249
Frequencies --	3696.9832	3811.7084	3843.0241

pas

Frequencies --	-775.6531	-71.8160	122.0514
Frequencies --	289.1918	312.4710	485.2871
Frequencies --	559.1115	634.8930	847.7198
Frequencies --	1028.5527	1093.7542	1157.6792
Frequencies --	1235.0163	1251.6802	1306.3280
Frequencies --	1437.5124	1535.2461	1612.5721
Frequencies --	1849.5414	3007.7733	3031.3354
Frequencies --	3690.2292	3839.2891	3857.8486

oas

Frequencies --	-261.3925	-40.0307	275.6182
Frequencies --	306.4545	482.0274	539.1888
Frequencies --	640.1021	722.8805	876.9133
Frequencies --	925.4770	1116.7860	1165.0692
Frequencies --	1179.0116	1289.5464	1373.8809
Frequencies --	1402.9882	1487.7817	1669.1521
Frequencies --	1865.7226	3050.8945	3106.9169
Frequencies --	3612.4434	3718.0528	3853.2455

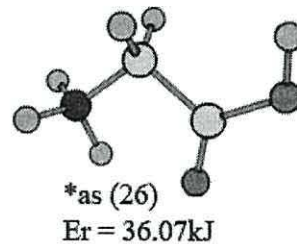
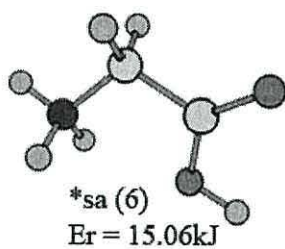
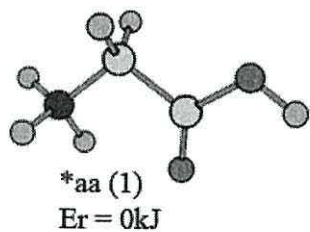
*Cartesian Coordinates of Geometry-Optimised Structures.*

<u>iaa</u>			H 0.854362 -0.678192 1.419570	C -0.482447 -0.868170 0.000000
H 1.020987 -1.834604 0.810652			O -0.649684 1.234072 0.122477	C 0.000000 0.574201 0.000000
N 0.416965 -1.922014 0.000000			H -1.562155 1.495717 -0.068285	O 1.136717 0.969304 0.000000
C -0.563750 -0.857218 0.000000				H -1.135117 -0.974182 -0.873238
C 0.000000 0.552099 0.000000			<u>ias</u>	H 0.600201 -2.411734 -0.821040
O 1.173201 0.843186 0.000000			H 1.098462 -1.775940 0.810161	H -1.135117 -0.974182 0.873238
H -1.214184 -0.951433 -0.874652			N 0.497833 -1.889909 0.000000	O -1.080660 1.405680 0.000000
H 1.020987 -1.834604 -0.810652			C -0.517475 -0.860770 0.000000	H -0.723185 2.305007 0.000000
H -1.214184 -0.951433 0.874652			C 0.000000 0.577462 0.000000	
O -0.996411 1.471647 0.000000			O 1.163524 0.880852 0.000000	<u>paa</u>
H -0.564181 2.338226 0.000000			H -1.160253 -0.987486 -0.879015	H -0.260378 -2.795723 0.000000
			H 1.098462 -1.775940 -0.810161	N -0.462108 -1.815703 0.000000
<u>oss</u>			H -1.160253 -0.987486 0.879015	C 0.586777 -0.847265 0.000000
H -2.174001 0.072057 -0.994101			O -0.966475 1.537018 0.000000	C 0.000000 0.542826 0.000000
N -1.775617 0.121482 -0.063051			H -1.832787 1.113103 0.000000	O -1.178927 0.809076 0.000000
C -0.639630 -0.791844 0.087483				H 1.241325 -0.906889 0.880926
C 0.690630 -0.036660 0.002292			<u>iss</u>	H -1.415298 -1.503576 0.000000
O 1.753360 -0.602498 -0.069111			H 2.070067 0.527122 0.838376	H 1.241325 -0.906889 -0.880926
H -0.662072 -1.261120 1.074246			N 1.894507 -0.008902 -0.000053	O 0.976715 1.480108 0.000000
H -2.515305 -0.086997 0.596236			C 0.667705 -0.777240 0.000044	H 0.524819 2.336161 0.000000
H -0.609763 -1.594462 -0.653655			C -0.682225 -0.048736 0.000008	
O 0.558655 1.296863 0.043976			O -1.728016 -0.644913 -0.000037	<u>pas</u>
H -0.411660 1.436256 0.081061			H 0.655040 -1.433158 -0.873886	H -0.385411 -2.767329 0.000000
			H 2.069933 0.527202 -0.838457	N -0.545382 -1.779467 0.000000
<u>i-a</u>			H 0.655092 -1.432961 0.874116	C 0.536993 -0.854523 0.000000
H 1.896253 0.986459 0.306143			O -0.637151 1.304450 0.000020	C 0.000000 0.567883 0.000000
N 1.907324 0.085196 -0.158806			H 0.296777 1.553671 0.000049	O -1.167701 0.852809 0.000000
C 0.754195 -0.701551 0.242205				H 1.187429 -0.948072 0.884571
C -0.624775 -0.105767 0.016910			<u>oas*</u>	H -1.483014 -1.421548 0.000000
O -1.634178 -0.744915 -0.168971			H -2.700446 -0.351383 0.174640	H 1.187429 -0.948072 -0.884571
H 0.766088 -1.666422 -0.267779			N -1.861458 0.042314 -0.235651	O 0.950932 1.538767 0.000000
H 1.876750 0.270935 -1.155800			C -0.689337 -0.716241 0.162230	H 1.823437 1.128523 0.000000
H 0.824394 -0.911454 1.315194			C 0.551781 0.148094 0.016491	
O -0.610893 1.249713 0.091921			O 0.531726 1.349817 0.056083	<u>oas</u>
H -1.530713 1.529627 -0.024413			H -0.689853 -1.052520 1.212122	H 0.690544 -2.373183 0.822933
			H -1.763192 0.998070 0.094016	N 0.668359 -1.788945 0.000000
<u>o-a*</u>			H -0.594181 -1.609398 -0.465558	C -0.443308 -0.865804 0.000000
H 1.770334 1.100651 -0.004566			O 1.726300 -0.525591 -0.087603	C 0.000000 0.601740 0.000000
N 1.785990 0.147334 -0.351479			H 1.539005 -1.465899 -0.185827	O 1.129059 0.999506 0.000000
C 0.764104 -0.648694 0.321846				H -1.088063 -1.000432 -0.878175
C -0.617960 -0.119018 0.025851			<u>oaa</u>	H 0.690544 -2.373183 -0.822933
O -1.593235 -0.788178 -0.219084			H 0.600201 -2.411734 0.821040	H -1.088063 -1.000432 0.878175
H 0.797619 -1.674389 -0.045764			N 0.605605 -1.824175 0.000000	O -1.047400 1.483298 0.000000
H 2.704404 -0.235997 -0.153930				H -1.876897 0.991803 0.000000

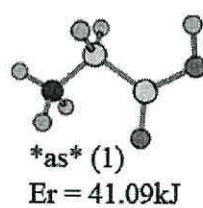
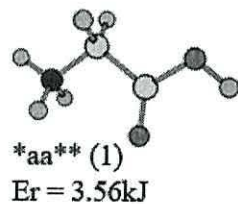
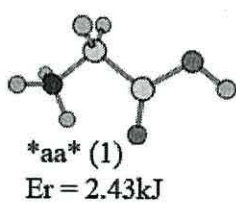


## *h2n-g-oh*

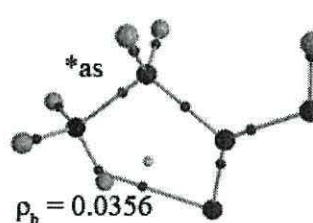
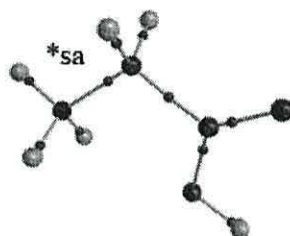
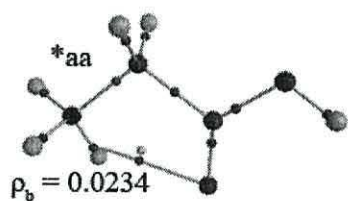
Conformer classes.



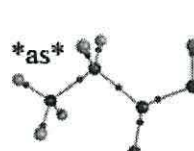
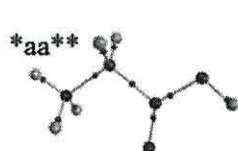
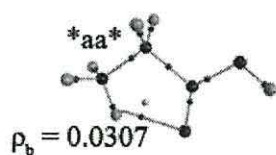
Non-minimised conformers:



AIM analysis.



Non-minimised conformers:





Calculation results.

Code name	Finished as	Relative free energy, kJ·mol <sup>-1</sup>
h2n-g-oh-l_paa	*aa	0
h2n-g-oh-l_oaa	*aa*	2.433838266
h2n-g-oh-l_iaa	*aa**	3.554926659
h2n-g-oh-l_o+a	*sa	15.02048406
h2n-g-oh-l_o-a	*sa	15.02048406
h2n-g-oh-l_i-a	*sa	15.06511755
h2n-g-oh-l_i+a	*sa	15.06511755
h2n-g-oh-l_p+a	*sa	15.08087055
h2n-g-oh-l_p-a	*sa	15.08612155
h2n-g-oh-l_pa+	*as	34.99266064
h2n-g-oh-l_o++	*as	35.33135011
h2n-g-oh-l_o--	*as	35.33397561
h2n-g-oh-l_p+s	*as	35.36023061
h2n-g-oh-l_pas	*as	35.3838601
h2n-g-oh-l_i++	*as	35.4127406
h2n-g-oh-l_i--	*as	35.4127406
h2n-g-oh-l_p+-	*as	35.4389956
h2n-g-oh-l_p--	*as	35.4494976
h2n-g-oh-l_p++	*as	35.51513509
h2n-g-oh-l_p-s	*as	35.51776059
h2n-g-oh-l_p+-	*as	35.54401559
h2n-g-oh-l_pa-	*as	35.56239409
h2n-g-oh-l_i+s	*as	35.65428658
h2n-g-oh-l_i-s	*as	35.65428658
h2n-g-oh-l_o+-	*as	35.72780057
h2n-g-oh-l_o+-	*as	35.72780057
h2n-g-oh-l_ia-	*as	35.76193207
h2n-g-oh-l_ia+	*as	35.76193207
h2n-g-oh-l_i+-	*as	35.78293607
h2n-g-oh-l_i+-	*as	35.89320705
h2n-g-oh-l_oas	*as	38.29028832
h2n-g-oh-l_oa-	*as	38.29291382
h2n-g-oh-l_oa+	*as	38.29291382
h2n-g-oh-l_o-s	*as	38.30866682
h2n-g-oh-l_o+s	*as	38.31654332
h2n-g-oh-l_ias	*as*	41.09169656
		conformer class
		non-minimised conformer

## *Thermodynamic Parameters.*

### \*aa

Zero-point correction= 0.095373 (Hartree/Particle)  
Thermal correction to Energy= 0.101033  
Thermal correction to Enthalpy= 0.101977  
Thermal correction to Gibbs Free Energy= 0.066348  
Sum of electronic and zero-point Energies= -284.156806  
Sum of electronic and thermal Energies= -284.151146  
Sum of electronic and thermal Enthalpies= -284.150202  
Sum of electronic and thermal Free Energies= -284.185831

### \*sa

Zero-point correction= 0.095461 (Hartree/Particle)  
Thermal correction to Energy= 0.101100  
Thermal correction to Enthalpy= 0.102044  
Thermal correction to Gibbs Free Energy= 0.066369  
Sum of electronic and zero-point Energies= -284.151018  
Sum of electronic and thermal Energies= -284.145379  
Sum of electronic and thermal Enthalpies= -284.144434  
Sum of electronic and thermal Free Energies= -284.180110

### \*as

Zero-point correction= 0.094485 (Hartree/Particle)  
Thermal correction to Energy= 0.100244  
Thermal correction to Enthalpy= 0.101188  
Thermal correction to Gibbs Free Energy= 0.065145  
Sum of electronic and zero-point Energies= -284.143163  
Sum of electronic and thermal Energies= -284.137404  
Sum of electronic and thermal Enthalpies= -284.136460  
Sum of electronic and thermal Free Energies= -284.172503

### \*aa\*

Zero-point correction= 0.094969 (Hartree/Particle)  
Thermal correction to Energy= 0.099794  
Thermal correction to Enthalpy= 0.100739  
Thermal correction to Gibbs Free Energy= 0.067089  
Sum of electronic and zero-point Energies= -284.157023  
Sum of electronic and thermal Energies= -284.152198  
Sum of electronic and thermal Enthalpies= -284.151254  
Sum of electronic and thermal Free Energies= -284.184904

### \*aa\*\*

Zero-point correction= 0.095348 (Hartree/Particle)  
Thermal correction to Energy= 0.100258  
Thermal correction to Enthalpy= 0.101202  
Thermal correction to Gibbs Free Energy= 0.067329  
Sum of electronic and zero-point Energies= -284.156458  
Sum of electronic and thermal Energies= -284.151547  
Sum of electronic and thermal Enthalpies= -284.150603  
Sum of electronic and thermal Free Energies= -284.184477

### \*as\*

Zero-point correction= 0.094766 (Hartree/Particle)  
Thermal correction to Energy= 0.099843  
Thermal correction to Enthalpy= 0.100787  
Thermal correction to Gibbs Free Energy= 0.066634  
Sum of electronic and zero-point Energies= -284.142048  
Sum of electronic and thermal Energies= -284.136971  
Sum of electronic and thermal Enthalpies= -284.136027  
Sum of electronic and thermal Free Energies= -284.170180

*Normal Modes of Vibration (within computational error of  $\pm 25 \text{ cm}^{-1}$ ).*

\*aa

Frequencies --	101.6908	121.8780	293.4923
Frequencies --	471.5818	523.8688	640.5166
Frequencies --	660.8283	852.0047	919.8628
Frequencies --	1027.3815	1102.6072	1116.5836
Frequencies --	1207.2886	1328.2768	1357.3714
Frequencies --	1465.9226	1496.7551	1521.6964
Frequencies --	1639.0541	1678.5865	1825.4479
Frequencies --	3144.0659	3218.8958	3350.6899
Frequencies --	3490.0742	3548.5089	3758.9364

\*sa

Frequencies --	65.6698	200.5458	281.2085
Frequencies --	477.9979	532.8862	595.3263
Frequencies --	665.4308	832.7797	895.3014
Frequencies --	997.2297	1086.3315	1117.9488
Frequencies --	1198.5062	1343.7980	1354.0392
Frequencies --	1423.1883	1495.3838	1544.1450
Frequencies --	1653.0262	1669.6247	1870.0456
Frequencies --	3142.5311	3217.5518	3425.5465
Frequencies --	3520.2942	3534.0231	3762.1953

\*as

Frequencies --	50.3735	196.8236	319.6264
Frequencies --	371.5504	479.3024	576.3664
Frequencies --	670.9084	851.4353	931.3631
Frequencies --	1031.7472	1099.4515	1111.0588
Frequencies --	1201.5499	1318.3862	1332.7773
Frequencies --	1413.8861	1477.5918	1512.1391
Frequencies --	1644.4414	1673.3489	1837.8007
Frequencies --	3132.5019	3163.5265	3205.6999
Frequencies --	3488.9498	3557.0620	3824.3286

\*aa\*

Frequencies --	-80.7218	175.2703	301.4655
Frequencies --	482.2742	507.8443	645.7937
Frequencies --	656.2439	854.8901	937.2897
Frequencies --	1031.9267	1103.6054	1111.5909
Frequencies --	1206.8556	1312.6311	1343.3238
Frequencies --	1456.2785	1486.1994	1515.5748
Frequencies --	1653.9588	1670.4401	1823.4682
Frequencies --	3148.5436	3219.1421	3228.3821
Frequencies --	3495.3808	3558.8302	3759.3854

\*aa\*\*

Frequencies --	-99.7062	150.7187	279.0389
Frequencies --	474.7119	520.1440	633.3213
Frequencies --	644.6605	848.3087	904.7440
Frequencies --	1022.7366	1106.9418	1111.6487
Frequencies --	1204.7410	1344.7019	1354.8999
Frequencies --	1471.6737	1498.6310	1550.4122
Frequencies --	1641.2122	1657.5448	1828.9135
Frequencies --	3142.0126	3213.3606	3427.8399
Frequencies --	3516.0025	3542.2068	3761.7573

\*as\*

Frequencies --	-121.5008	147.7417	298.6789
Frequencies --	346.7760	475.1605	581.3037
Frequencies --	645.9253	847.5545	904.8504
Frequencies --	1017.2296	1100.0359	1109.1691
Frequencies --	1195.9866	1321.0770	1354.8624
Frequencies --	1437.8647	1506.2803	1551.9130
Frequencies --	1633.3961	1654.8029	1841.6728
Frequencies --	3126.2736	3198.1442	3423.4606
Frequencies --	3511.4415	3536.4203	3829.4035



*Cartesian Coordinates of Geometry-Optimised Structures.*

\*aa

H 2.632240 -0.296838 -0.409961  
N 1.804162 0.065838 0.068983  
C 0.604136 -0.816346 -0.092885  
C -0.606810 0.106575 -0.014240  
O -0.476264 1.308833 -0.048651  
H 0.646839 -1.291030 -1.074527  
H 1.554622 0.997076 -0.303894  
H 0.595924 -1.581572 0.682515  
O -1.724225 -0.590606 0.065873  
H -2.480714 0.020654 0.072171  
H 2.041909 0.203658 1.055790

\*as

H 2.499248 -0.179497 -0.663672  
N 1.790165 0.028001 0.043180  
C 0.576691 -0.842349 -0.058147  
C -0.614084 0.133731 -0.006940  
O -0.395284 1.321325 -0.029129  
H 0.595359 -1.371737 -1.012524  
H 1.421512 0.998374 -0.086786  
H 0.565727 -1.562774 0.760623  
O -1.829297 -0.379324 0.036206  
H -1.824889 -1.345825 0.069498  
H 2.232893 -0.018849 0.964512

\*aa\*\*

H 0.995146 -1.653921 0.816945  
N 0.385963 -1.770373 0.000000  
C -0.695491 -0.734849 0.000000  
C 0.000000 0.618830 0.000000  
O 1.205435 0.704283 0.000000  
H -1.311151 -0.867216 -0.890656  
H 0.995146 -1.653921 -0.816945  
H -1.311151 -0.867216 0.890656  
O -0.890416 1.596125 0.000000  
H -0.424185 2.449456 0.000000  
H 0.007245 -2.721721 0.000000

\*sa

H -1.792450 0.448993 -1.068760  
N -1.781535 0.109916 -0.101626  
C -0.620269 -0.814301 0.141528  
C 0.710894 -0.085758 0.006216  
O 1.753981 -0.657453 -0.105201  
H -0.713050 -1.208240 1.155337  
H -2.676968 -0.353406 0.081639  
H -0.670608 -1.636132 -0.572074  
O 0.510566 1.250130 0.070334  
H 1.371021 1.698813 0.010263  
H -1.707324 0.939494 0.497455

\*aa\*

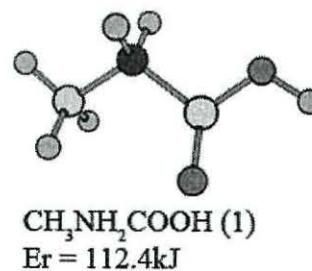
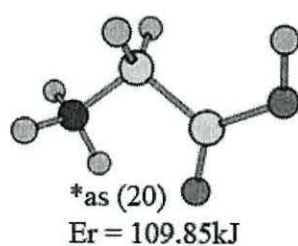
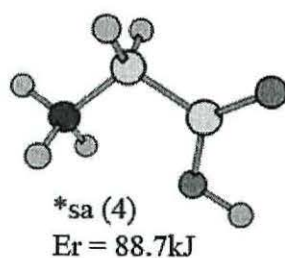
H 0.351539 -2.368113 0.827117  
N 0.375407 -1.767287 0.000000  
C -0.713090 -0.734785 0.000000  
C 0.000000 0.615342 0.000000  
O 1.209304 0.680467 0.000000  
H -1.333818 -0.846732 -0.888930  
H 0.351539 -2.368113 -0.827117  
H -1.333818 -0.846732 0.888930  
O -0.867436 1.606489 0.000000  
H -0.388012 2.452690 0.000000  
H 1.268313 -1.230988 0.000000

\*as\*

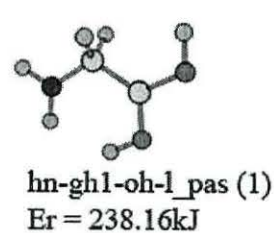
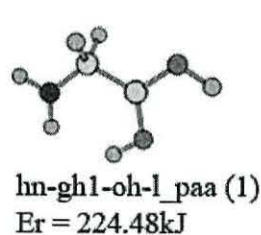
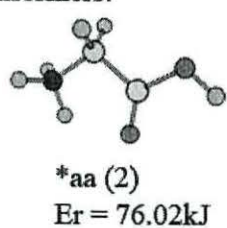
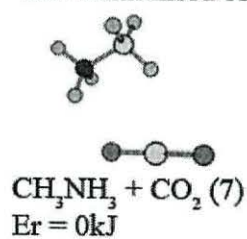
H 1.026760 -1.612306 0.815217  
N 0.419321 -1.750282 0.000000  
C -0.683824 -0.740121 0.000000  
C 0.000000 0.633782 0.000000  
O 1.203371 0.689120 0.000000  
H -1.292903 -0.895911 -0.892844  
H 1.026760 -1.612306 -0.815217  
H -1.292903 -0.895911 0.892844  
O -0.783592 1.703208 0.000000  
H -1.724114 1.479880 0.000000  
H 0.065869 -2.712058 0.000000

# hn-hg-oh

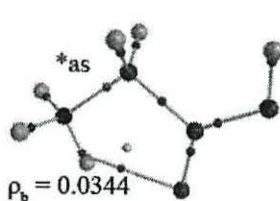
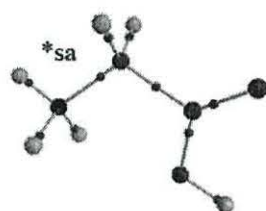
Conformer classes.



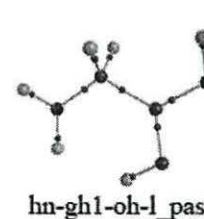
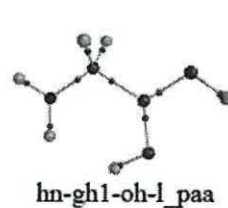
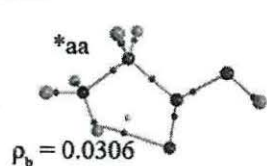
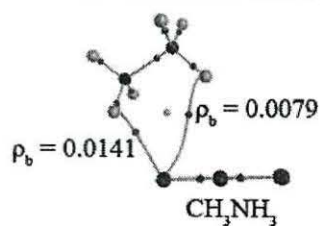
Non-minimised conformers:



AIM analysis.



Non-minimised conformers:



Calculation results.

Code name	Finished as	Relative free energy, kJ·mol <sup>-1</sup>
hn-hg-oh-l_p+	CH3NH3 + CO2	0
hn-hg-oh-l_p-a	CH3NH3 + CO2	0
hn-hg-oh-l_p+s	CH3NH3 + CO2	0
hn-hg-oh-l_p--	CH3NH3 + CO2	0
hn-hg-oh-l_p+-	CH3NH3 + CO2	0
hn-hg-oh-l_p++	CH3NH3 + CO2	0
hn-hg-oh-l_p-s	CH3NH3 + CO2	0
hn-hg-oh-l_iaa	*aa	76.0134687
hn-hg-oh-l_oaa	*aa	76.0187197
hn-hg-oh-l_i-a	*sa	88.63424599
hn-hg-oh-l_i+a	*sa	88.68938149
hn-hg-oh-l_o-a	*sa	88.74714248
hn-hg-oh-l_o+a	*sa	88.74714248
hn-hg-oh-l_pa+	*as	108.7429486
hn-hg-oh-l_pa-	*as	108.7429486
hn-hg-oh-l_o+	*as	108.8821
hn-hg-oh-l_i+	*as	108.950363
hn-hg-oh-l_i++	*as	108.9529885
hn-hg-oh-l_i--	*as	108.9529885
hn-hg-oh-l_i+-	*as	108.9582395
hn-hg-oh-l_oa-	*as	108.9634905
hn-hg-oh-l_o+-	*as	109.0317535
hn-hg-oh-l_o++	*as	109.4570845
hn-hg-oh-l_o--	*as	109.45971
hn-hg-oh-l_o-s	*as	109.590985
hn-hg-oh-l_o+s	*as	109.590985
hn-hg-oh-l_i-s	*as	109.6303675
hn-hg-oh-l_i+s	*as	109.6303675
hn-hg-oh-l_ia-	*as	111.8594168
hn-hg-oh-l_ias	*as	111.8725443
hn-hg-oh-l_ia+	*as	111.8725443
hn-hg-oh-l_oas	*as	111.8751698
hn-hg-oh-l_oa+	*as	111.8804208
hn-hg-oh-l_p+a	CH3NH2COOH	112.4002697
hn-hg-oh-l_paa	hn-gh1-oh-l_paa	224.477603
hn-hg-oh-l_pas	hn-gh1-oh-l_pas	238.1590821
		conformer class
		non-minimised conformer



## Thermodynamic Parameters.

### \*sa

Zero-point correction= 0.095465 (Hartree/Particle)  
Thermal correction to Energy= 0.101102  
Thermal correction to Enthalpy= 0.102046  
Thermal correction to Gibbs Free Energy= 0.066383  
Sum of electronic and zero-point Energies= -284.151013  
Sum of electronic and thermal Energies= -284.145376  
Sum of electronic and thermal Enthalpies= -284.144432  
Sum of electronic and thermal Free Energies= -284.180095

### \*as

Zero-point correction= 0.094520 (Hartree/Particle)  
Thermal correction to Energy= 0.100286  
Thermal correction to Enthalpy= 0.101231  
Thermal correction to Gibbs Free Energy= 0.065216  
Sum of electronic and zero-point Energies= -284.143132  
Sum of electronic and thermal Energies= -284.137365  
Sum of electronic and thermal Enthalpies= -284.136421  
Sum of electronic and thermal Free Energies= -284.172436

### CH<sub>3</sub>NH<sub>2</sub>COOH

Zero-point correction= 0.094573 (Hartree/Particle)  
Thermal correction to Energy= 0.100257  
Thermal correction to Enthalpy= 0.101201  
Thermal correction to Gibbs Free Energy= 0.064831  
Sum of electronic and zero-point Energies= -284.141300  
Sum of electronic and thermal Energies= -284.135617  
Sum of electronic and thermal Enthalpies= -284.134673  
Sum of electronic and thermal Free Energies= -284.171043

### CH<sub>3</sub>NH<sub>3</sub> + CO<sub>2</sub>

Zero-point correction= 0.092671 (Hartree/Particle)  
Thermal correction to Energy= 0.099317  
Thermal correction to Enthalpy= 0.100261

Thermal correction to Gibbs Free Energy= 0.060682  
Sum of electronic and zero-point Energies= -284.181865  
Sum of electronic and thermal Energies= -284.175219  
Sum of electronic and thermal Enthalpies= -284.174275  
Sum of electronic and thermal Free Energies= -284.213854

### \*aa

Zero-point correction= 0.094971 (Hartree/Particle)  
Thermal correction to Energy= 0.099796  
Thermal correction to Enthalpy= 0.100740  
Thermal correction to Gibbs Free Energy= 0.067091  
Sum of electronic and zero-point Energies= -284.157022  
Sum of electronic and thermal Energies= -284.152197  
Sum of electronic and thermal Enthalpies= -284.151252  
Sum of electronic and thermal Free Energies= -284.184902

### hn-hg-oh-l paa

Zero-point correction= 0.090651 (Hartree/Particle)  
Thermal correction to Energy= 0.095087  
Thermal correction to Enthalpy= 0.096031  
Thermal correction to Gibbs Free Energy= 0.063165  
Sum of electronic and zero-point Energies= -284.100870  
Sum of electronic and thermal Energies= -284.096433  
Sum of electronic and thermal Enthalpies= -284.095489  
Sum of electronic and thermal Free Energies= -284.128355

### hn-hg-oh-l pas

Zero-point correction= 0.090706 (Hartree/Particle)  
Thermal correction to Energy= 0.095165  
Thermal correction to Enthalpy= 0.096109  
Thermal correction to Gibbs Free Energy= 0.063206  
Sum of electronic and zero-point Energies= -284.095645  
Sum of electronic and thermal Energies= -284.091185  
Sum of electronic and thermal Enthalpies= -284.090241  
Sum of electronic and thermal Free Energies= -284.123144

*Normal Modes of Vibration (within computational error of  $\pm 25 \text{ cm}^{-1}$ ).*

\*sa

Frequencies --	66.4321	199.8865	282.7182
Frequencies --	477.9884	532.5644	595.3808
Frequencies --	664.9215	832.7322	895.7924
Frequencies --	998.2459	1086.2016	1118.2986
Frequencies --	1198.4944	1343.4774	1353.7224
Frequencies --	1422.8325	1495.5365	1543.4083
Frequencies --	1653.0410	1670.2843	1870.3516
Frequencies --	3142.5491	3217.7440	3425.1553
Frequencies --	3519.8598	3533.9278	3762.7537

\*as

Frequencies --	54.5672	187.3119	318.3828
Frequencies --	372.9445	478.3602	576.9009
Frequencies --	669.2484	850.9629	930.4261
Frequencies --	1031.3462	1098.8875	1111.3081
Frequencies --	1201.3591	1318.6944	1334.6369
Frequencies --	1415.5312	1479.6993	1512.4014
Frequencies --	1643.0813	1674.4599	1837.8284
Frequencies --	3132.6959	3186.0294	3205.4977
Frequencies --	3487.2324	3554.7667	3824.7082

CH<sub>3</sub>NH<sub>2</sub>COOH

Frequencies --	28.4163	219.5892	275.4133
Frequencies --	470.5935	554.9919	617.9477
Frequencies --	683.7981	772.8982	868.1359
Frequencies --	998.7400	1141.5111	1170.8237
Frequencies --	1205.7333	1325.6353	1363.0883
Frequencies --	1410.7848	1484.6777	1521.8959
Frequencies --	1525.1135	1627.6371	1946.5661
Frequencies --	3128.4041	3246.2467	3253.6928
Frequencies --	3426.3669	3495.1505	3749.0782

CH<sub>3</sub>NH<sub>3</sub> + CO<sub>2</sub>

Frequencies --	-107.3432	33.0372	63.6710
Frequencies --	133.6606	173.5483	350.6262
Frequencies --	608.9467	620.5870	921.2192
Frequencies --	940.1235	955.8540	1287.2743

Frequencies --	1298.8779	1337.6775	1474.2008
Frequencies --	1518.6773	1526.5849	1549.4986
Frequencies --	1663.4317	1688.2767	2423.0716
Frequencies --	3163.5425	3274.4670	3336.6852
Frequencies --	3373.0764	3461.8299	3499.2662

\*aa

Frequencies --	-79.6460	175.5232	301.1809
Frequencies --	482.3751	508.0312	646.1257
Frequencies --	656.1816	855.1539	937.4413
Frequencies --	1031.7022	1103.5556	1111.6877
Frequencies --	1206.9055	1312.6304	1343.2802
Frequencies --	1456.3432	1486.3398	1515.7104
Frequencies --	1654.0004	1670.5501	1823.4589
Frequencies --	3148.8125	3219.4603	3229.7663
Frequencies --	3494.4310	3557.9571	3758.9331

hn-hg-oh-l paa

Frequencies --	-858.0956	-476.9852	-114.4057
Frequencies --	289.9466	455.9486	506.3296
Frequencies --	547.7028	576.9711	686.4980
Frequencies --	847.1625	999.5473	1090.0426
Frequencies --	1170.0340	1212.9053	1251.2414
Frequencies --	1270.8333	1407.4240	1474.8673
Frequencies --	1601.0170	1651.4081	1688.6617
Frequencies --	3027.0298	3052.0934	3689.2102
Frequencies --	3702.7003	3780.1427	3811.3536

hn-hg-oh-l pas

Frequencies --	-845.4246	-464.3405	-109.6345
Frequencies --	293.8018	463.2707	496.4605
Frequencies --	503.5039	594.2336	661.3236
Frequencies --	860.5539	1005.7715	1094.9465
Frequencies --	1177.9507	1200.2834	1250.8436
Frequencies --	1282.2883	1406.7264	1492.6646
Frequencies --	1593.1264	1650.8234	1696.0698
Frequencies --	3005.8537	3030.7690	3685.8137
Frequencies --	3764.1253	3794.5210	3809.3913

*Cartesian Coordinates of Geometry-Optimised Structures.*

\*sa

H	-1.807371	0.424395	-1.078160
N	-1.781044	0.109941	-0.103059
C	-0.620670	-0.813861	0.143625
C	0.710803	-0.085786	0.006567
O	1.753340	-0.657984	-0.106844
H	-0.713341	-1.203634	1.159046
H	-2.675765	-0.343237	0.107022
H	-0.670463	-1.638525	-0.566701
O	0.510774	1.250113	0.070773
H	1.371201	1.699082	0.013207
H	-1.690653	0.953174	0.474416

\*as

H	2.516617	-0.202901	-0.637333
N	1.791034	0.028029	0.045300
C	0.576925	-0.840748	-0.061237
C	-0.614464	0.134051	-0.007015
O	-0.397737	1.321803	-0.030525
H	0.597017	-1.364764	-1.018580
H	2.210377	0.003295	0.978647
H	0.566465	-1.565792	0.753500
O	-1.829090	-0.381174	0.037820
H	-1.822840	-1.347510	0.074588
H	1.434976	0.996619	-0.116767

CH<sub>3</sub>NH<sub>2</sub>COOH

H	-0.547284	-1.551785	0.234735
N	-0.650028	-0.650281	-0.245476
C	-1.885117	0.077692	0.205926

C	0.622167	0.163017	-0.040507
O	0.596746	1.345532	-0.086327
H	-1.914204	1.035054	-0.308610
H	-0.716261	-0.863306	-1.248987
H	-2.747934	-0.534382	-0.049684
O	1.583095	-0.714649	0.155274
H	2.431632	-0.249837	0.264993
H	-1.816775	0.224909	1.281801

CH<sub>3</sub>NH<sub>3</sub> + CO<sub>2</sub>

H	-2.905890	-0.669879	0.392613
N	-2.051036	-0.451409	-0.132638
C	-1.587171	0.967346	0.122225
C	1.583870	-0.182329	0.009948
O	2.301955	0.723985	-0.114631
H	-2.305001	1.651885	-0.313714
H	-2.272062	-0.623707	-1.123603
H	-1.514695	1.092094	1.200028
O	0.858922	-1.097305	0.135718
H	-1.299711	-1.106963	0.130154
H	-0.612598	1.092894	-0.318749

\*aa

H	0.352297	-2.368057	0.827315
N	0.375810	-1.767344	0.000000
C	-0.712839	-0.734827	0.000000
C	0.000000	0.615111	0.000000
O	1.209210	0.681431	0.000000
H	-1.333421	-0.847052	-0.888959
H	0.352297	-2.368057	-0.827315
H	-1.333421	-0.847052	0.888959

O	-0.868113	1.605927	0.000000
H	-0.388840	2.452254	0.000000
H	1.268675	-1.231196	0.000000

hn-hg-oh-l paa

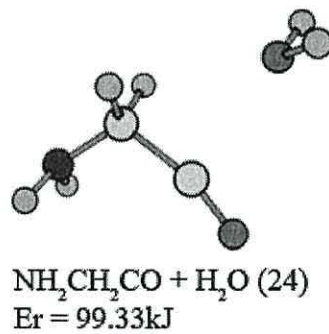
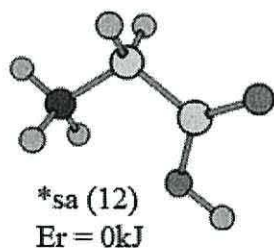
H	0.396624	-2.805822	0.000000
N	-0.066224	-1.913906	0.000000
C	0.739492	-0.741888	0.000000
C	0.000000	0.544951	0.000000
O	-1.273826	0.692321	0.000000
H	1.400666	-0.663451	0.878441
H	-1.067264	-1.952944	0.000000
H	1.400666	-0.663451	-0.878441
O	0.709874	1.607535	0.000000
H	0.162905	2.416713	0.000000
H	-1.755361	-0.150926	0.000000

hn-hg-oh-l pas

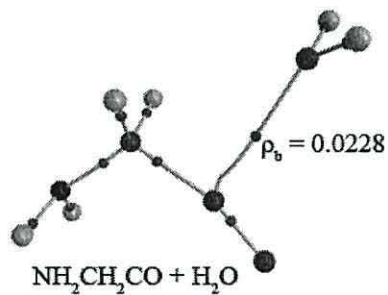
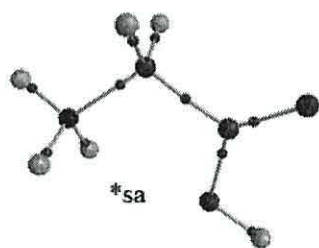
H	0.287857	-2.810370	0.000000
N	-0.144212	-1.902820	0.000000
C	0.697001	-0.758786	0.000000
C	0.000000	0.560804	0.000000
O	-1.263034	0.738462	0.000000
H	1.356478	-0.710831	0.883902
H	-1.146267	-1.905943	0.000000
H	1.356478	-0.710831	-0.883902
O	0.652495	1.663221	0.000000
H	1.615154	1.528099	0.000000
H	-1.757908	-0.095960	0.000000



*Conformer classes.*



*AIM analysis.*



Calculation results.

Code name	Finished as	Relative free energy, kJ·mol <sup>-1</sup>
hn-g-oh2-l_o+s	*sa	0
hn-g-oh2-l_p+s	*sa	0.047258995
hn-g-oh2-l_o-s	*sa	0.049884495
hn-g-oh2-l_o-a	*sa	0.10239449
hn-g-oh2-l_o+	*sa	0.10764549
hn-g-oh2-l_p-s	*sa	0.115521989
hn-g-oh2-l_p-a	*sa	0.120772988
hn-g-oh2-l_p+a	*sa	0.120772988
hn-g-oh2-l_o+a	*sa	0.133900487
hn-g-oh2-l_o+-	*sa	0.144402486
hn-g-oh2-l_p+-	*sa	0.183784982
hn-g-oh2-l_p++	*sa	0.194286981
hn-g-oh2-l_pa-	NH2CH2CO+H2O	92.45172263
hn-g-oh2-l_pa+	NH2CH2CO+H2O	92.45172263
hn-g-oh2-l_oa+	NH2CH2CO+H2O	93.7224645
hn-g-oh2-l_oa-	NH2CH2CO+H2O	93.730341
hn-g-oh2-l_oaa	NH2CH2CO+H2O	93.97188698
hn-g-oh2-l_oas	NH2CH2CO+H2O	93.97976348
hn-g-oh2-l_ias	NH2CH2CO+H2O	95.75985231
hn-g-oh2-l_iaa	NH2CH2CO+H2O	95.8359918
hn-g-oh2-l_ia+	NH2CH2CO+H2O	95.8569958
hn-g-oh2-l_ia-	NH2CH2CO+H2O	95.8569958
hn-g-oh2-l_o--	NH2CH2CO+H2O	97.02796869
hn-g-oh2-l_i+s	NH2CH2CO+H2O	97.03584519
hn-g-oh2-l_i-s	NH2CH2CO+H2O	97.03584519
hn-g-oh2-l_o++	NH2CH2CO+H2O	97.04634719
hn-g-oh2-l_i+a	NH2CH2CO+H2O	97.04897269
hn-g-oh2-l_i+-	NH2CH2CO+H2O	97.04897269
hn-g-oh2-l_i+	NH2CH2CO+H2O	97.04897269
hn-g-oh2-l_i-a	NH2CH2CO+H2O	97.04897269
hn-g-oh2-l_p--	NH2CH2CO+H2O	97.09360618
hn-g-oh2-l_p++	NH2CH2CO+H2O	97.09360618
hn-g-oh2-l_i--	NH2CH2CO+H2O	97.09885718
hn-g-oh2-l_i++	NH2CH2CO+H2O	97.09885718
hn-g-oh2-l_pas	NH2CH2CO+H2O	137.7179638
hn-g-oh2-l_paa	NH2CH2CO+H2O	137.7310913
		conformer class
		non-minimised conformer

## Thermodynamic Parameters.

\*sa

Zero-point correction= 0.095454 (Hartree/Particle)  
 Thermal correction to Energy= 0.101097  
 Thermal correction to Enthalpy= 0.102042  
 Thermal correction to Gibbs Free Energy= 0.066343  
 Sum of electronic and zero-point Energies= -284.151025  
 Sum of electronic and thermal Energies= -284.145381  
 Sum of electronic and thermal Enthalpies= -284.144437  
 Sum of electronic and thermal Free Energies= -284.180135

NH<sub>2</sub>CH<sub>2</sub>CO+H<sub>2</sub>O

Zero-point correction= 0.087565 (Hartree/Particle)  
 Thermal correction to Energy= 0.095764  
 Thermal correction to Enthalpy= 0.096708  
 Thermal correction to Gibbs Free Energy= 0.055284  
 Sum of electronic and zero-point Energies= -284.112641  
 Sum of electronic and thermal Energies= -284.104442  
 Sum of electronic and thermal Enthalpies= -284.103498  
 Sum of electronic and thermal Free Energies= -284.1449

## Normal Modes of Vibration (within computational error of $\pm 25$ cm<sup>-1</sup>).

\*sa

Frequencies -- 64.4505 198.4608 282.3681  
 Frequencies -- 478.8174 532.3090 594.7280  
 Frequencies -- 664.6237 832.4105 895.5248  
 Frequencies -- 997.9172 1086.2432 1118.0238  
 Frequencies -- 1198.2806 1343.7038 1353.6326  
 Frequencies -- 1422.6612 1495.3302 1543.9677  
 Frequencies -- 1652.8662 1669.9586 1870.6796  
 Frequencies -- 3142.7455 3217.4526 3425.7077  
 Frequencies -- 3520.5089 3534.0386 3761.8663

NH<sub>2</sub>CH<sub>2</sub>CO+H<sub>2</sub>O

Frequencies -- 51.5478 86.1317 130.4498  
 Frequencies -- 185.3093 215.4619 284.3343  
 Frequencies -- 345.6479 389.8251 473.8139  
 Frequencies -- 494.8401 600.0152 735.5047  
 Frequencies -- 896.0378 1193.1620 1205.2163  
 Frequencies -- 1353.3904 1375.3127 1430.9635  
 Frequencies -- 1670.1915 1673.9363 2229.0669  
 Frequencies -- 3126.9120 3187.3285 3601.2709  
 Frequencies -- 3721.2319 3839.0477 3940.7656

## Cartesian Coordinates of Geometry-Optimised Structures.

\*sa

H -1.804904 0.442802 -1.066652  
 N -1.781641 0.109840 -0.097658  
 C -0.619869 -0.815728 0.136003  
 C 0.711276 -0.085462 0.005914  
 O 1.755199 -0.656189 -0.101313  
 H -0.710707 -1.218810 1.146340  
 H -2.675945 -0.349659 0.100158  
 H -0.670326 -1.631472 -0.584573

O 0.508895 1.250241 0.067377  
 H -1.696801 0.942577 0.495560  
 H 1.368971 1.700403 0.012766

NH<sub>2</sub>CH<sub>2</sub>CO+H<sub>2</sub>O

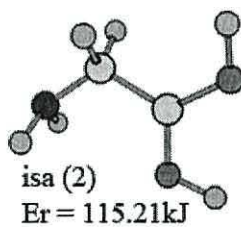
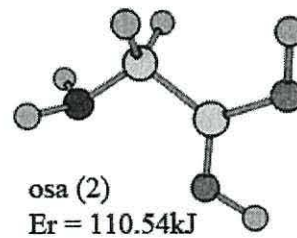
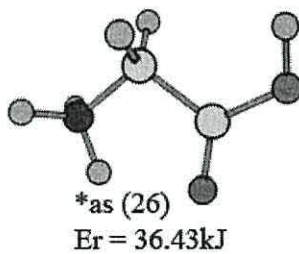
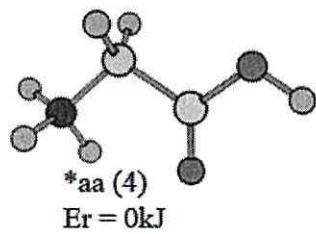
H 2.528066 -0.065028 0.870169  
 N 2.142405 -0.444347 0.016488  
 C 0.753005 -0.680922 -0.004894

C -0.053711 0.608247 -0.015368  
 O -0.382593 1.688425 -0.015332  
 H 0.420745 -1.220695 -0.895255  
 H 2.553739 -0.053889 -0.820012  
 H 0.389349 -1.224846 0.870195  
 O -2.088167 -0.706279 0.013499  
 H -2.567176 -1.121117 -0.712669  
 H -2.751245 -0.625114 0.708390

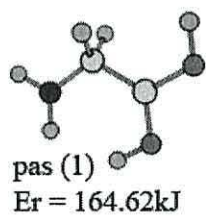
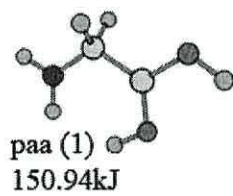


# *hn-gh1-oh*

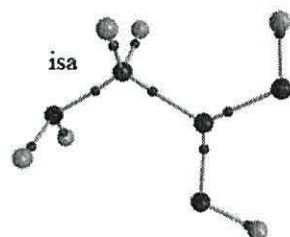
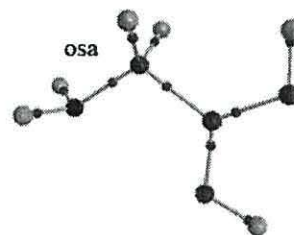
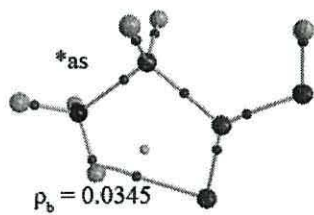
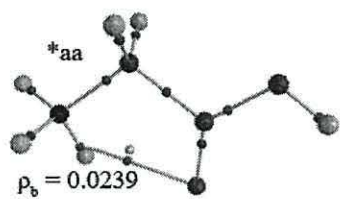
Conformer classes.



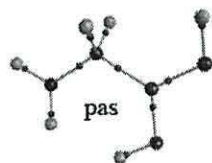
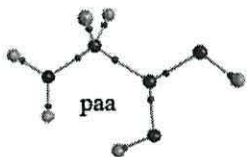
Non-optimised conformers:



*AIM analysis.*



Non-optimised conformers:



Calculation results.

Code name	Finished as	Relative free energy, kJ·mol <sup>-1</sup>
hn-gh1-oh-l_o+a	*aa	0
hn-gh1-oh-l_o-a	*aa	0.042007996
hn-gh1-oh-l_oaa	*aa	2.460093264
hn-gh1-oh-l_iaa	*aa	2.494224761
hn-gh1-oh-l_oa-	*as	34.89026615
hn-gh1-oh-l_i--	*as	35.13443763
hn-gh1-oh-l_o--	*as	35.27358911
hn-gh1-oh-l_o++	*as	35.29984411
hn-gh1-oh-l_oa+	*as	35.34185211
hn-gh1-oh-l_o+s	*as	35.4337446
hn-gh1-oh-l_o-s	*as	35.4416211
hn-gh1-oh-l_i++	*as	35.4547486
hn-gh1-oh-l_p-s	*as	35.48362909
hn-gh1-oh-l_pa+	*as	35.53351359
hn-gh1-oh-l_pa-	*as	35.53351359
hn-gh1-oh-l_p+	*as	35.54139009
hn-gh1-oh-l_p+-	*as	35.54139009
hn-gh1-oh-l_i-s	*as	35.54401559
hn-gh1-oh-l_i+s	*as	35.59390008
hn-gh1-oh-l_o+-	*as	35.67003958
hn-gh1-oh-l_o+-	*as	35.67003958
hn-gh1-oh-l_oas	*as	38.30079032
hn-gh1-oh-l_p--	*as	38.30866682
hn-gh1-oh-l_p++	*as	38.30866682
hn-gh1-oh-l_ia+	*as	38.31129232
hn-gh1-oh-l_ia-	*as	38.31129232
hn-gh1-oh-l_ias	*as	38.32704532
hn-gh1-oh-l_p+s	*as	38.33492182
hn-gh1-oh-l_i+-	*as	38.33754732
hn-gh1-oh-l_i-+	*as	38.33754732
hn-gh1-oh-l_p-a	osa	110.4731529
hn-gh1-oh-l_p+a	osa	110.6018024
hn-gh1-oh-l_i+a	isa	115.2069289
hn-gh1-oh-l_i-a	isa	115.2200564
hn-gh1-oh-l_paa	paa	150.937355
hn-gh1-oh-l_pas	pas	164.6162087
		conformer class
		non-minimised conformer



## Thermodynamic Parameters.

### \*aa

Zero-point correction= 0.095362 (Hartree/Particle)  
Thermal correction to Energy= 0.101018  
Thermal correction to Enthalpy= 0.101962  
Thermal correction to Gibbs Free Energy= 0.066341  
Sum of electronic and zero-point Energies= -284.156818  
Sum of electronic and thermal Energies= -284.151162  
Sum of electronic and thermal Enthalpies= -284.150218  
Sum of electronic and thermal Free Energies= -284.185839

### \*as

Zero-point correction= 0.094493 (Hartree/Particle)  
Thermal correction to Energy= 0.100272  
Thermal correction to Enthalpy= 0.101217  
Thermal correction to Gibbs Free Energy= 0.065102  
Sum of electronic and zero-point Energies= -284.143159  
Sum of electronic and thermal Energies= -284.137379  
Sum of electronic and thermal Enthalpies= -284.136435  
Sum of electronic and thermal Free Energies= -284.172550

### osa

Zero-point correction= 0.092791 (Hartree/Particle)  
Thermal correction to Energy= 0.098824  
Thermal correction to Enthalpy= 0.099768  
Thermal correction to Gibbs Free Energy= 0.062906  
Sum of electronic and zero-point Energies= -284.113878  
Sum of electronic and thermal Energies= -284.107845  
Sum of electronic and thermal Enthalpies= -284.106901  
Sum of electronic and thermal Free Energies= -284.143762

### isa

Zero-point correction= 0.093024 (Hartree/Particle)  
Thermal correction to Energy= 0.098813  
Thermal correction to Enthalpy= 0.099757  
Thermal correction to Gibbs Free Energy= 0.063676  
Sum of electronic and zero-point Energies= -284.112611  
Sum of electronic and thermal Energies= -284.106822  
Sum of electronic and thermal Enthalpies= -284.105878  
Sum of electronic and thermal Free Energies= -284.141959

### paa

Zero-point correction= 0.090655 (Hartree/Particle)  
Thermal correction to Energy= 0.095093  
Thermal correction to Enthalpy= 0.096037  
Thermal correction to Gibbs Free Energy= 0.063169  
Sum of electronic and zero-point Energies= -284.100864  
Sum of electronic and thermal Energies= -284.096426  
Sum of electronic and thermal Enthalpies= -284.095482  
Sum of electronic and thermal Free Energies= -284.128350

### pas

Zero-point correction= 0.090711 (Hartree/Particle)  
Thermal correction to Energy= 0.095172  
Thermal correction to Enthalpy= 0.096117  
Thermal correction to Gibbs Free Energy= 0.063210  
Sum of electronic and zero-point Energies= -284.095638  
Sum of electronic and thermal Energies= -284.091177  
Sum of electronic and thermal Enthalpies= -284.090233  
Sum of electronic and thermal Free Energies= -284.123140

## Normal Modes of Vibration (within computational error of $\pm 25$ $\text{cm}^{-1}$ ).

### \*aa

Frequencies --	98.9109	125.9752	294.1748
Frequencies --	471.7731	523.5218	641.1267
Frequencies --	661.0034	852.2237	920.6953
Frequencies --	1027.7437	1102.5614	1116.5156
Frequencies --	1207.3469	1327.3686	1357.0431
Frequencies --	1465.4617	1496.3012	1520.6804
Frequencies --	1639.5688	1678.7233	1825.2891
Frequencies --	3144.2574	3219.0633	3343.9627
Frequencies --	3489.8715	3549.0650	3758.8512

### \*as

Frequencies --	49.0510	188.6089	317.3282
Frequencies --	371.3152	478.4058	576.3742
Frequencies --	669.4486	850.8039	930.5958
Frequencies --	1031.3647	1099.0073	1110.9920
Frequencies --	1201.4359	1318.6349	1333.9983
Frequencies --	1414.9742	1479.2730	1512.3963
Frequencies --	1643.5704	1673.9877	1837.9634
Frequencies --	3132.6749	3181.2238	3205.3486
Frequencies --	3488.1334	3555.8370	3824.7491

### osa

Frequencies --	46.5797	103.3925	278.8888
Frequencies --	473.9234	485.6595	598.8214
Frequencies --	616.3705	732.5180	771.0293
Frequencies --	878.1036	931.0802	1178.5362

Frequencies --	1193.2814	1207.4342	1217.2513
Frequencies --	1367.1351	1387.2379	1450.1496
Frequencies --	1566.5101	1668.1658	1695.2704
Frequencies --	3039.0927	3100.3002	3603.2021
Frequencies --	3683.7998	3705.1250	3751.6598

### isa

Frequencies --	45.4582	254.9738	291.9082
Frequencies --	469.1610	492.3060	612.6460
Frequencies --	613.3919	691.6533	761.0526
Frequencies --	847.6357	920.8710	1192.9874
Frequencies --	1205.1126	1206.2200	1225.5943
Frequencies --	1371.3311	1394.0460	1437.8730
Frequencies --	1558.6378	1666.6993	1680.4681
Frequencies --	3058.0279	3107.8721	3594.5995
Frequencies --	3682.1788	3705.7318	3744.4669

### paa

Frequencies --	-858.5988	-476.8914	-114.9839
Frequencies --	289.5866	455.4347	506.2827
Frequencies --	547.3387	576.8159	686.4320
Frequencies --	846.8619	999.6662	1090.1969
Frequencies --	1170.0387	1212.9923	1251.4395
Frequencies --	1270.9686	1407.5475	1475.0130
Frequencies --	1601.4626	1651.6504	1688.7605
Frequencies --	3027.0395	3052.1380	3689.3681
Frequencies --	3702.9420	3781.0833	3811.9016

pas

Frequencies --	-846.4456	-464.7604	-110.7389
Frequencies --	293.7063	461.1465	496.4221
Frequencies --	503.5172	593.9637	660.7916
Frequencies --	860.3929	1005.8195	1095.0871

Frequencies --	1177.3020	1200.5788	1250.8018
Frequencies --	1282.2694	1406.8377	1492.7273
Frequencies --	1593.2429	1651.1538	1696.3015
Frequencies --	3006.0886	3031.0815	3686.2992
Frequencies --	3763.8626	3797.4038	3810.7348

*Cartesian Coordinates of Geometry-Optimised Structures.*

\*aa

H	2.053641	0.188770	1.054862
N	1.803940	0.065911	0.069106
C	0.604219	-0.816755	-0.092940
C	-0.606713	0.106423	-0.014260
O	-0.475733	1.308713	-0.048732
H	0.646533	-1.291163	-1.074706
H	2.627219	-0.285871	-0.425861
H	0.595328	-1.582294	0.682109
O	-1.724286	-0.590204	0.065955
H	-2.480626	0.021252	0.072070
H	1.545443	1.001855	-0.286798

osa

H	-2.584653	0.007158	0.677077
N	-1.885194	0.187897	-0.029431
C	-0.778187	-0.713369	0.045851
C	0.543928	0.010039	0.006948
O	1.675404	-0.606564	-0.034421
H	-0.764130	-1.264311	0.996578
H	-2.321597	0.215145	-0.941448
H	-0.726752	-1.473858	-0.748633
O	0.567256	1.278139	0.029521
H	1.482898	1.621636	0.010387
H	1.574866	-1.573669	-0.065532

paa

H	0.395476	-2.805969	0.000000
N	-0.067352	-1.914058	0.000000
C	0.738828	-0.742421	0.000000
C	0.000000	0.545146	0.000000
O	-1.273424	0.693464	0.000000
H	1.400002	-0.664377	0.878470
H	-1.068380	-1.952202	0.000000
H	1.400002	-0.664377	-0.878470
O	0.711152	1.606750	0.000000
H	0.165190	2.416586	0.000000
H	-1.755619	-0.149320	0.000000

\*as

H	-2.509989	-0.195313	-0.649749
N	-1.791938	0.028243	0.043120
C	-0.577203	-0.840544	-0.058185
C	0.614738	0.134059	-0.006840
O	0.398914	1.322019	-0.029002
H	-0.567302	-1.561894	0.759835
H	-2.221735	-0.003438	0.971363
H	-0.596353	-1.369380	-1.012888
O	1.828974	-0.381771	0.036045
H	1.822313	-1.348159	0.071184
H	-1.431680	0.997410	-0.107773

isa

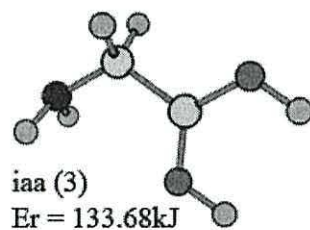
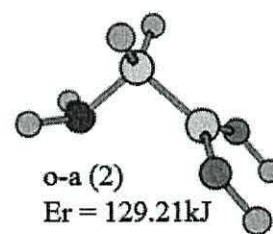
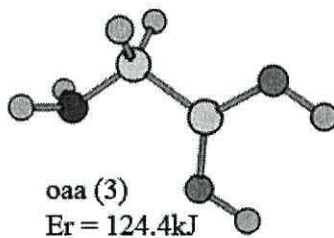
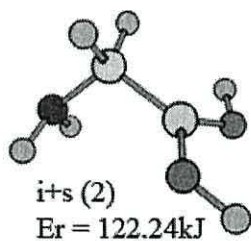
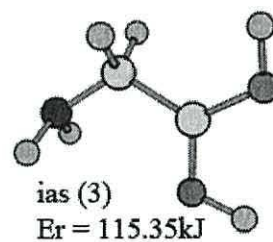
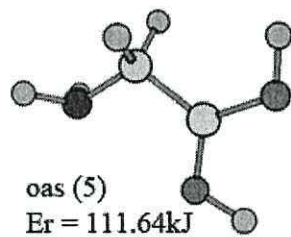
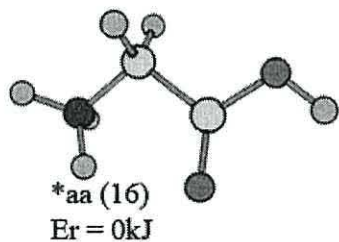
H	-2.112139	0.583885	-0.844519
N	-1.970522	0.038115	-0.004948
C	-0.775582	-0.730771	0.006558
C	0.544327	0.006052	0.000784
O	1.689481	-0.581446	-0.004077
H	-0.742584	-1.379340	0.893366
H	-2.118256	0.599195	0.823388
H	-0.739754	-1.402676	-0.862379
O	0.539902	1.279092	0.002435
H	1.444553	1.651236	-0.000313
H	1.614290	-1.551958	-0.005820

pas

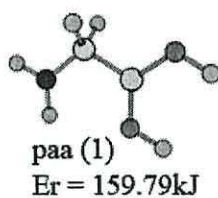
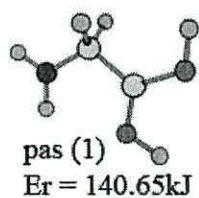
H	0.287678	-2.810254	0.000000
N	-0.144476	-1.902788	0.000000
C	0.696952	-0.758894	0.000000
C	0.000000	0.560930	0.000000
O	-1.262916	0.738628	0.000000
H	1.356332	-0.710994	0.883939
H	-1.146443	-1.904726	0.000000
H	1.356332	-0.710994	-0.883939
O	0.652701	1.662917	0.000000
H	1.615323	1.527369	0.000000
H	-1.757878	-0.095458	0.000000

## hn-gh2-oh

Conformer classes.

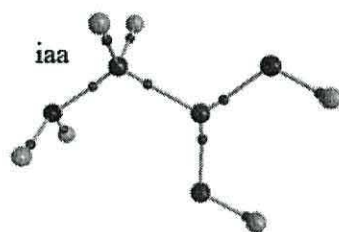
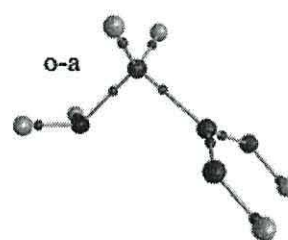
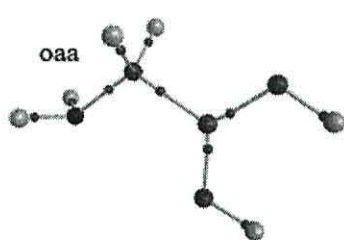
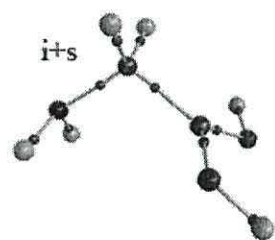
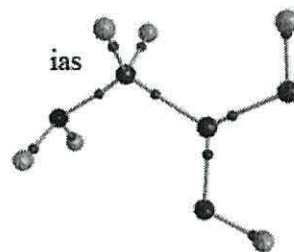
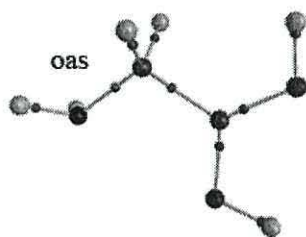
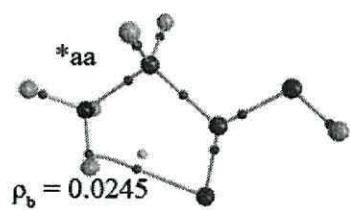


Non-minimised conformers:

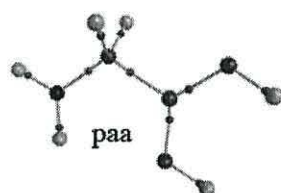
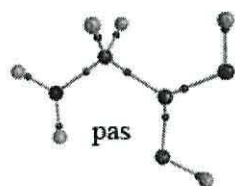




*AIM analysis.*



Non-minimised conformers:



Calculation results.

Code name	Finished as	Relative free energy, kJ·mol <sup>-1</sup>
hn-gh2-oh-l_p+s	*aa	0
hn-gh2-oh-l_p-s	*aa	0
hn-gh2-oh-l_p++	*aa	0.044633496
hn-gh2-oh-l_o++	*aa	0.131274987
hn-gh2-oh-l_o++	*aa	0.147027986
hn-gh2-oh-l_p--	*aa	0.165406484
hn-gh2-oh-l_o+-	*aa	0.173282983
hn-gh2-oh-l_i-s	*aa	0.178533983
hn-gh2-oh-l_o--	*aa	0.178533983
hn-gh2-oh-l_o-s	*aa	0.181159483
hn-gh2-oh-l_i+s	*aa	0.191661482
hn-gh2-oh-l_p++	*aa	0.194286981
hn-gh2-oh-l_o+s	*aa	0.202163481
hn-gh2-oh-l_p+-	*aa	0.228418478
hn-gh2-oh-l_i+-	*aa	2.646503746
hn-gh2-oh-l_i+-	*aa	2.646503746
hn-gh2-oh-l_pa+	oas	110.6805674
hn-gh2-oh-l_pa-	oas	110.6805674
hn-gh2-oh-l_oa+	oas	112.2611182
hn-gh2-oh-l_oa-	oas	112.2611182
hn-gh2-oh-l_oas	oas	112.3267557
hn-gh2-oh-l_ia+	ias	115.3119489
hn-gh2-oh-l_ia-	ias	115.3460804
hn-gh2-oh-l_ias	ias	115.3959649
hn-gh2-oh-l_i++	i+s	122.2380173
hn-gh2-oh-l_i--	i+s	122.2380173
hn-gh2-oh-l_oaa	oaa	124.3541701
hn-gh2-oh-l_p+a	oaa	124.4198076
hn-gh2-oh-l_p-a	oaa	124.4198076
hn-gh2-oh-l_o-a	o-a	129.2060936
hn-gh2-oh-l_o+a	o-a	129.2060936
hn-gh2-oh-l_i-a	iaa	133.6589412
hn-gh2-oh-l_i+a	iaa	133.6615667
hn-gh2-oh-l_iaa	iaa	133.7219532
hn-gh2-oh-l_pas	pas	140.645396
hn-gh2-oh-l_paa	paa	159.7852892
		conformer class
		non-minimised conformer

## Thermodynamic Parameters.

### \*aa

Zero-point correction= 0.095336 (Hartree/Particle)  
Thermal correction to Energy= 0.100997  
Thermal correction to Enthalpy= 0.101942  
Thermal correction to Gibbs Free Energy= 0.066276  
Sum of electronic and zero-point Energies= -284.156841  
Sum of electronic and thermal Energies= -284.151180  
Sum of electronic and thermal Enthalpies= -284.150236  
Sum of electronic and thermal Free Energies= -284.185901

### oas

Zero-point correction= 0.092795 (Hartree/Particle)  
Thermal correction to Energy= 0.098824  
Thermal correction to Enthalpy= 0.099768  
Thermal correction to Gibbs Free Energy= 0.062924  
Sum of electronic and zero-point Energies= -284.113874  
Sum of electronic and thermal Energies= -284.107845  
Sum of electronic and thermal Enthalpies= -284.106900  
Sum of electronic and thermal Free Energies= -284.143745

### ias

Zero-point correction= 0.093016 (Hartree/Particle)  
Thermal correction to Energy= 0.098809  
Thermal correction to Enthalpy= 0.099753  
Thermal correction to Gibbs Free Energy= 0.063654  
Sum of electronic and zero-point Energies= -284.112619  
Sum of electronic and thermal Energies= -284.106826  
Sum of electronic and thermal Enthalpies= -284.105882  
Sum of electronic and thermal Free Energies= -284.141981

### i+s

Zero-point correction= 0.093475 (Hartree/Particle)  
Thermal correction to Energy= 0.099234  
Thermal correction to Enthalpy= 0.100178  
Thermal correction to Gibbs Free Energy= 0.064174  
Sum of electronic and zero-point Energies= -284.110041  
Sum of electronic and thermal Energies= -284.104283  
Sum of electronic and thermal Enthalpies= -284.103339  
Sum of electronic and thermal Free Energies= -284.139343

### oaa

Zero-point correction= 0.092297 (Hartree/Particle)  
Thermal correction to Energy= 0.098496  
Thermal correction to Enthalpy= 0.099440

Thermal correction to Gibbs Free Energy= 0.061441  
Sum of electronic and zero-point Energies= -284.107681  
Sum of electronic and thermal Energies= -284.101482  
Sum of electronic and thermal Enthalpies= -284.100538  
Sum of electronic and thermal Free Energies= -284.138537

### o-a

Zero-point correction= 0.093341 (Hartree/Particle)  
Thermal correction to Energy= 0.099160  
Thermal correction to Enthalpy= 0.100104  
Thermal correction to Gibbs Free Energy= 0.064411  
Sum of electronic and zero-point Energies= -284.107760  
Sum of electronic and thermal Energies= -284.101941  
Sum of electronic and thermal Enthalpies= -284.100997  
Sum of electronic and thermal Free Energies= -284.136689

### iaa

Zero-point correction= 0.092524 (Hartree/Particle)  
Thermal correction to Energy= 0.098402  
Thermal correction to Enthalpy= 0.099346  
Thermal correction to Gibbs Free Energy= 0.062797  
Sum of electronic and zero-point Energies= -284.105266  
Sum of electronic and thermal Energies= -284.099388  
Sum of electronic and thermal Enthalpies= -284.098443  
Sum of electronic and thermal Free Energies= -284.134993

### pas

Zero-point correction= 0.090597 (Hartree/Particle)  
Thermal correction to Energy= 0.094995  
Thermal correction to Enthalpy= 0.095939  
Thermal correction to Gibbs Free Energy= 0.063148  
Sum of electronic and zero-point Energies= -284.104883  
Sum of electronic and thermal Energies= -284.100485  
Sum of electronic and thermal Enthalpies= -284.099541  
Sum of electronic and thermal Free Energies= -284.132332

### paa

Zero-point correction= 0.090116 (Hartree/Particle)  
Thermal correction to Energy= 0.094612  
Thermal correction to Enthalpy= 0.095556  
Thermal correction to Gibbs Free Energy= 0.062619  
Sum of electronic and zero-point Energies= -284.097544  
Sum of electronic and thermal Energies= -284.093049  
Sum of electronic and thermal Enthalpies= -284.092105  
Sum of electronic and thermal Free Energies= -284.125042



*Normal Modes of Vibration (within computational error of  $\pm 25 \text{ cm}^{-1}$ ).*

\*aa

Frequencies -- 89.6089 132.3391 294.6312  
 Frequencies -- 472.1468 522.8314 642.2554  
 Frequencies -- 661.1076 852.5662 922.1219  
 Frequencies -- 1028.7685 1102.2421 1116.3243  
 Frequencies -- 1207.7435 1326.0285 1356.4003  
 Frequencies -- 1464.9494 1495.5595 1519.4783  
 Frequencies -- 1640.5636 1678.6869 1825.0350  
 Frequencies -- 3144.6189 3219.0077 3333.4046  
 Frequencies -- 3490.3910 3550.5710 3758.2427

oas

Frequencies -- 46.6852 105.0211 278.9192  
 Frequencies -- 474.0765 485.6989 598.8474  
 Frequencies -- 616.3070 732.7975 771.0188  
 Frequencies -- 877.9602 931.3435 1178.4288  
 Frequencies -- 1193.3295 1207.4392 1217.2904  
 Frequencies -- 1366.9583 1387.3318 1450.2092  
 Frequencies -- 1566.5555 1668.1882 1695.3120  
 Frequencies -- 3038.8369 3100.3964 3602.9903  
 Frequencies -- 3683.8032 3704.9482 3751.5515

ias

Frequencies -- 44.8147 254.9072 291.0788  
 Frequencies -- 469.1265 492.2202 612.3770  
 Frequencies -- 613.8999 688.2634 760.9354  
 Frequencies -- 847.2022 920.8216 1193.0825  
 Frequencies -- 1205.0016 1206.5535 1225.8562  
 Frequencies -- 1371.4111 1393.9502 1437.9556  
 Frequencies -- 1558.5620 1666.7300 1680.3338  
 Frequencies -- 3057.9681 3107.7927 3595.6726  
 Frequencies -- 3682.2975 3707.0499 3743.5519

i+s

Frequencies -- 50.0968 225.9589 298.9303  
 Frequencies -- 443.6271 542.6454 636.5381  
 Frequencies -- 650.3069 693.7427 769.8348  
 Frequencies -- 904.3319 913.0107 1133.8505  
 Frequencies -- 1178.3604 1200.9343 1224.5034  
 Frequencies -- 1375.5184 1387.4440 1482.3239  
 Frequencies -- 1563.2529 1663.5298 1666.6024  
 Frequencies -- 3114.9170 3186.9890 3580.8433  
 Frequencies -- 3688.8876 3698.8655 3755.1210

oaa

Frequencies -- 21.3373 72.1223 284.7788  
 Frequencies -- 479.0602 499.4268 526.4561  
 Frequencies -- 622.4189 696.4837 744.0757  
 Frequencies -- 888.3472 915.0596 1135.8273

Frequencies -- 1180.4457 1191.4917 1209.3200  
 Frequencies -- 1365.3277 1385.9896 1429.9756  
 Frequencies -- 1571.7229 1643.9311 1683.8997  
 Frequencies -- 3067.1695 3117.3651 3608.4930  
 Frequencies -- 3707.5877 3722.7732 3743.0105

o-a

Frequencies -- 97.0707 217.2914 221.9581  
 Frequencies -- 410.3836 563.6086 565.1178  
 Frequencies -- 667.2116 727.0103 758.3575  
 Frequencies -- 921.6540 938.4473 1109.1326  
 Frequencies -- 1148.6265 1195.8606 1212.1794  
 Frequencies -- 1361.9210 1392.5290 1491.1145  
 Frequencies -- 1574.0063 1647.3636 1668.0216  
 Frequencies -- 3131.6958 3208.9128 3586.0835  
 Frequencies -- 3692.7735 3724.3682 3739.0534

iaa

Frequencies -- 29.7119 261.9094 295.7112  
 Frequencies -- 483.8480 500.5930 533.1820  
 Frequencies -- 617.7108 670.5581 691.2663  
 Frequencies -- 850.3418 913.0799 1134.0530  
 Frequencies -- 1189.3694 1205.3920 1217.1458  
 Frequencies -- 1376.4002 1386.4490 1418.6373  
 Frequencies -- 1559.0912 1635.1865 1680.1189  
 Frequencies -- 3076.1901 3123.5539 3600.4238  
 Frequencies -- 3713.0871 3717.1523 3733.1017

pas

Frequencies -- -742.8954 -349.0261 -28.9162  
 Frequencies -- 284.7786 469.3236 494.7608  
 Frequencies -- 566.1630 597.1128 723.7726  
 Frequencies -- 856.8035 996.3488 1110.3704  
 Frequencies -- 1190.8432 1219.0280 1251.7620  
 Frequencies -- 1258.0675 1401.2141 1485.0491  
 Frequencies -- 1603.0433 1634.7753 1680.8347  
 Frequencies -- 2987.4120 3004.3726 3684.2345  
 Frequencies -- 3691.5630 3750.6710 3825.0681

paa

Frequencies -- -766.0344 -358.9618 -69.3924  
 Frequencies -- 290.0988 458.2135 471.3632  
 Frequencies -- 509.7548 602.3749 642.2019  
 Frequencies -- 861.0737 990.7735 1103.2381  
 Frequencies -- 1140.2781 1187.5868 1251.1069  
 Frequencies -- 1252.9889 1401.3183 1466.2634  
 Frequencies -- 1601.7131 1626.1730 1663.2015  
 Frequencies -- 3009.4336 3027.5284 3696.8313  
 Frequencies -- 3724.6271 3748.3256 3830.0583

## Cartesian Coordinates of Geometry-Optimised Structures.

### \*aa

H	-2.074473	0.165629	1.050979
N	-1.804031	0.066250	0.068265
C	-0.604532	-0.817139	-0.091776
C	0.606670	0.106033	-0.014133
O	1.724430	-0.589748	0.065203
H	-0.595010	-1.580684	0.685241
H	-2.617968	-0.268408	-0.453089
H	-0.645708	-1.294724	-1.071979
O	0.475630	1.308464	-0.048171
H	-1.532448	1.009250	-0.260364
H	2.480523	0.022094	0.070549

### i+s

H	1.865361	0.986802	-0.768274
N	1.830989	0.037318	-0.418671
C	0.886597	-0.174627	0.650330
C	-0.522895	-0.005636	0.148836
O	-0.958360	1.184367	-0.011704
H	1.003469	-1.178869	1.065775
H	1.796322	-0.637588	-1.172200
H	1.049521	0.555355	1.445550
O	-1.287533	-0.967812	-0.233318
H	-0.905049	-1.842105	-0.044849
H	-1.841609	1.204322	-0.430129

### iaa

H	2.121058	0.603497	0.828462
N	1.969547	0.048537	-0.002563
C	0.781060	-0.730294	0.003715
C	-0.543411	-0.008049	0.000408
O	-1.597883	-0.740537	-0.002414
H	0.741869	-1.401581	-0.863782
H	2.119664	0.591893	-0.841485
H	0.743310	-1.388954	0.881072
O	-0.536639	1.272303	0.001406
H	-1.411596	1.701493	-0.000532
H	-2.450863	-0.270182	-0.002468

### oas

H	2.585764	0.004416	0.675884
N	1.885148	0.187932	-0.028775
C	0.778146	-0.713502	0.045250
C	-0.543886	0.010027	0.006964
O	-0.567162	1.278114	0.029550
H	0.726663	-1.473394	-0.749852
H	2.320036	0.218347	-0.941440
H	0.764243	-1.265072	0.995591
O	-1.675410	-0.606489	-0.034215
H	-1.574929	-1.573607	-0.065403
H	-1.482799	1.621640	0.010679

### oaa

H	0.099753	-2.464643	0.825039
N	0.183379	-1.887090	0.000000
C	-0.720600	-0.777779	0.000000
C	0.000000	0.543620	0.000000
O	1.275359	0.551636	0.000000
H	-1.385372	-0.735634	-0.872942
H	0.099753	-2.464643	-0.825039
H	-1.385372	-0.735634	0.872942
O	-0.748770	1.589982	0.000000
H	-0.286243	2.446377	0.000000
H	1.684719	1.435817	0.000000

### pas

H	0.303990	-2.805381	0.000000
N	-0.106386	-1.889597	0.000000
C	0.739452	-0.756709	0.000000
C	0.000000	0.529528	0.000000
O	-1.273642	0.513933	0.000000
H	1.402441	-0.689356	0.882580
H	-1.106214	-1.814793	0.000000
H	1.402441	-0.689356	-0.882580
O	0.575587	1.680452	0.000000
H	1.546817	1.620349	0.000000
H	-1.657046	1.413721	0.000000

### ias

H	-2.115000	0.592264	-0.833912
N	-1.970163	0.038306	-0.000402
C	-0.775742	-0.730839	0.000094
C	0.544391	0.006084	0.000001
O	0.539862	1.279121	0.000193
H	-0.741211	-1.390517	0.878505
H	-2.117218	0.589530	0.834531
H	-0.740991	-1.391765	-0.877386
O	1.689503	-0.581439	-0.000111
H	1.614310	-1.552021	-0.000204
H	1.444434	1.651438	0.000056

### o-a

H	2.203998	0.833562	-0.681191
N	1.641269	0.000961	-0.560343
C	0.896527	-0.004167	0.692275
C	-0.502873	-0.000380	0.162821
O	-1.040567	-1.137836	-0.084176
H	1.055414	-0.902215	1.290269
H	2.213064	-0.825204	-0.682584
H	1.057253	0.886348	1.301096
O	-1.035610	1.140349	-0.080014
H	-1.883268	1.117298	-0.558855
H	-1.887846	-1.109341	-0.563395

### paa

H	0.373743	-2.797371	0.000000
N	-0.061422	-1.893381	0.000000
C	0.759026	-0.739696	0.000000
C	0.000000	0.529806	0.000000
O	-1.280258	0.473077	0.000000
H	1.424069	-0.652754	0.877048
H	-1.062470	-1.849301	0.000000
H	1.424069	-0.652754	-0.877048
O	0.690510	1.613679	0.000000
H	0.181460	2.442934	0.000000
H	-1.747096	1.328214	0.000000



## 9. REFERENCES

---

1. R. Car and M. Parrinello, *Physical Review Letters*, 1985, **55**, 2471-2474.
2. R. Iftimie, P. Minary and M. E. Tuckerman, *PNAS*, 2005, **102**, 6654-6659.
3. J. Kohanoff and J. P. Hansen, *Physical Review Letters*, 1995, **74**, 626-629.
4. R. B. Capaz and J. D. Joannopoulos, *Physical Review B*, 1996, **54**, 13402-13405.
5. J. Sarnthein, A. Pasquarello and R. Car, *Physical Review B*, 1995, **52**, 12690-12695.
6. E. Deumens, A. Diz, R. Longo and Y. Öhrn, *Reviews of Modern Physics*, 1994, **66**, 917-983.
7. D. Wei, H. Guo and D. R. Salahub, *Physical Review E*, 2001, **64**, 4.
8. P. L. Silvestrelli, *Physical Review B*, 1999, **59**, 9703-9706.
9. B. Schäfer, M. Perić and B. Engels, *Journal of Chemical Physics*, 1999, **110**, 7802-7810.
10. Y. Liu, D. A. Yarne and M. E. Tuckerman, *Physical Review B*, 2003, **68**, 8.
11. M. M. G. Alemany, L. J. Gallego and D. J. González, *Physical Review B*, 2004, **70**, 6.
12. S. Piana, D. Sebastiani, P. Carloni and M. Parrinello, *Journal of American Chemical Society*, 2001, **123**, 8730-8737.
13. J. Hutter, P. Carloni and M. Parrinello, *Journal of American Chemical Society*, 1996, **118**, 8710-8712.



14. P. L. Silvestrelli and M. Parrinello, *Journal of Chemical Physics*, 1999, **111**, 3572-3580.
15. P. L. Silvestrelli, M. Bernasconi and M. Parrinello, *Chemical Physics Letters*, 1997, **277**, 478-482.
16. C. J. Cramer, *Essentials of Computational Chemistry*, 2 edn., Wiley, 2004.
17. I. N. Levine, *Quantum Chemistry*, 6 edn., Prentice Hall, 2008.
18. D. Marx, *Computational Nanoscience: Do It Yourself!*, 2006, **31**, 195-244.
19. R. Ahlrichs, S. Elliot and U. Huniar, *Modern Methods and Algorithms of Quantum Chemistry*, 2000, **3**, 7-25.
20. J. C. Greer, R. Ahlrichs and I. V. Hertel, *Zeitschrift für Physik D Atoms, Molecules and Clusters*, 1991, **18**, 413-426.
21. N. L. Allinger, *Adv. Phys. Org. Chem.*, 1976, **13**, 1-82.
22. J. R. Maple, M. J. Hwang, T. P. Stockfish, U. Dinur, M. Waldman, C. S. Ewig and A. T. Halger, *Journal of Computational Chemistry*, 1993, **15**, 162-172.
23. P.-O. Norrby and P. Brandt, *Coordination Chemistry Reviews*, 2001, **212**, 79-109.
24. B. G. Sumpter and D. W. Noid, *Chemical Physics Letters*, 1992, **192**, 455-462.
25. T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. J. Doren, *Journal of Chemical Physics*, 1995, **103**, 4129-4137.
26. T. B. Blank and S. D. Brown, *Analytical Chemistry*, 1993, **65**, 3081.
27. D. F. R. Brown, M. N. Gibbs and D. C. J. Clary, *Journal of Chemical Physics*, 1996, **105**.
28. E. Tafeit, W. Estelberger, R. Horejsi, R. Moeller, K. Oettl, K. Vrecko and G. J. Reibnegger, *Journal of Molecular Graphics*, 1996, **14**, 12-18.

29. H. Dorsett and A. White, *Overview of Molecular Modelling and Ab initio Molecular Orbital Methods Suitable for Use with Energetic Materials*, General Document, DSTO Aeronautical and Maritime Research Laboratory, Salisbury, 2000.
30. A. W. von Hofmann, *Proceedings of the Royal Institution*, 1865.
31. D. H. R. Barton, *Experientia*, 1950, **VI**/8.
32. J. D. Watson and F. H. C. Crick, *Nature*, 1969, **224**, 470-471.
33. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *The Journal of Chemical Physics*, 1953, **21**, 1087-1092.
34. J. B. Hendrickson, *Journal of American Chemical Society*, 1961, **83**, 4537-4547.
35. K. B. Wiberg, *Journal of American Chemical Society*, 1965, **75**, 1070-1078.
36. B. Lee and F. M. Richards, *Journal of Molecular Biology*, 1971, **55**, 379-400.
37. K. B. Wiberg and R. H. Boyd, *Journal of American Chemical Society*, 1972, **94**, 8426-8430.
38. R. A. Davies, W. J. Nigel, N. M. John and H. H. Keith, in *Proceedings of the tenth international conference on 3D Web technology*, ACM, Bangor, United Kingdom, Editon edn., 2005.
39. R. A. Davies, S. M. James and W. J. Nigel, in *Proceedings of the 14th International Conference on 3D Web Technology*, ACM, Darmstadt, Germany, Editon edn., 2009.
40. W. J. Henre, L. Radom, P. v.R. Schleyer and P. J. A., *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, 1986.
41. J. J. P. Stewart, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyd, VCH Publishers, Editon edn., 1990, vol. 1, pp. 45-81.

42. T. P. Lybrand, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyd, VCH Publishers, Editon edn., 1990, vol. 1, pp. 295-320.
43. E. Eliel and e. al, *Conformational Analysis*, Wiley Interscience, 1965.
44. J. Cioslowski, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyd, VCH Publishers, Editon edn., 1991, vol. 2, pp. 1-55.
45. G. R. Marshall, C. D. Barry, H. E. Bosshard, R. A. Dammkoehler and D. A. Dunn, *ACS Symposium*, 1979, **112**.
46. R. A. Dammkoehler, S. F. Karasek, E. F. B. Shands and G. R. Marshall, *Journal of Computer-Aided Molecular Design*, 1989, **3**.
47. S. Forrest, *Science*, 1993, **261**, 872-878.
48. A. W. R. Payne and R. C. Glen, *Journal of Molecular Graphics*, 1993, **11**, 74-91.
49. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
50. G. Ciccotti, D. Frenkel and I. R. McDonald, *Simulation of Liquids and Solids*, North-Holland, Amsterdam, 1987.
51. D. Frenkel and B. Smit, *Understanding Molecular Simulation – From Algorithms to Applications*, Academic Press, San Diego, 1996.
52. K. Binder and G. Ciccotti, *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*, Italian Physical Society SIF, Bologna, 1996.
53. B. J. Berne, G. Ciccotti and D. F. Coker, *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, Singapore, 1998.
54. R. Esser, P. Grassberger, J. Grotendorst and M. Lewerenz, *Molecular Dynamics on Parallel Computers*, World Scientific, Singapore, 2000.
55. M. E. Tuckerman and G. J. Martyna, *Journal of Physical Chemistry*, 2000, **104**.



56. D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, Cambridge, 2001.
57. D. K. Remler and P. A. Madden, *Molecular Physics*, 1990, **70**, 921-966.
58. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias and J. D. Joannopoulos, *Reviews of Modern Physics*, 1992, **64**, 1045-1097.
59. D. Marx and J. Hutter, *Modern Methods and Algorithms of Quantum Chemistry*, 2000, **1**, 301-449.
60. N. Qian and T. J. Sejnowski, *Journal of Molecular Biology*, 1988, **202**, 865-884.
61. H. Bohr, B. J. B. S., R. M. J. Cotterill, B. Lautrup, L. Norskov, O. H. Olsen and S. B. Petersen, *FEBS Letters*, 1988, **241**, 223-228.
62. L. H. Holley and M. Karplus, *The Proceedings of the National Academy of Sciences*, 1989, **86**, 152-156.
63. S. Anzali, J. Gasteiger, U. Holzgrabe, J. Polanski, J. Sadowski, A. Teckentrup and M. Wagener, *Perspectives in Drug Discovery and Design*, 1998, **9-11**, 273-299.
64. G. Bohm, *Biophysical Chemistry*, 1996, **59**, 1-32.
65. J.-M. Chandonia and M. Karplus, *Proteins: Structure, Function, and Genetics*, 1999, **35**, 293-306.
66. M. Compiani, P. Fariselli and R. Casadio, 2nd Italian Workshop on Parallel Architectures and Neural Networks, Singapore, 1990.
67. M. Compiani, P. Fariselli and R. Casadio, SPIE, Orlando USA, 1996.
68. J. A. Cuff and G. J. Barton, *Proteins: Structure, Function, and Genetics*, 2000, **40**, 502-511.
69. P. Fariselli, M. Compiani and R. Casadio, *European Biophysics Journal*, 1993, **22**, 41-51.

70. I. Jacoboni, P. L. Martelli, P. Fariselli, M. Compiani and R. Casadio, *Proteins: Structure, Function, and Genetics*, 2000, **41**, 535 - 544.
71. D. T. Jones, *Journal of Molecular Biology*, 1999, **292**, 195-202.
72. R. Maclin and J. W. Shavlik, in *Computational Learning Theory and Natural Learning Systems*, MIT Press, Editon edn., 1994, vol. 1, pp. 249-286.
73. M. J. McGregor, T. P. Flores and M. J. E. Sternberg, *Protein Engineering*, 1989, **2**, 521-526.
74. S. Pascarella and F. Bossa, *Bioinformatics*, 1989, **5**, 319-320.
75. P. Stolorz, A. Lapedes and Y. Xia, *Journal of Molecular Biology*, 1992, **225**, 363-377.
76. W. Kabsch and C. Sander, *Proceedings of the National Academy of Sciences*, 1984, **81**, 1075-1078.
77. I. A. Wilson, D. H. Haft, E. D. Getzoff, J. A. Tainer, R. A. Lerner and S. Brenner, *Proceedings of the National Academy of Sciences*, 1985, **82**, 5255-5259.
78. B. I. Cohen, S. R. Presnell and F. E. Cohen, *Protein Sci*, 1993, **2**, 2134-2145.
79. S. Sudarsanam, *Proteins: Structure, Function, and Genetics*, 1998, **30**, 228-231.
80. R. Unger, D. Harel, S. Wherland and J. L. Sussman, *Proteins*, 1989, **5**, 355-373.
81. J. C. Tully, in *Modern Methods for Multidimensional Dynamics Computations in Chemistry*, ed. D. L. Thompson, World Scientific, Singapore, Editon edn., 1998.
82. J. C. Tully, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, eds. B. J. Berne, G. Ciccotti and D. F. Coker, World Scientific, Singapore, Editon edn., 1998, p. 498.
83. G. Pastore, E. Smargiassi and F. Buda, *Physical Review A*, 1991, **44**, 6334.

84. P. A. M. Dirac, *Mathematical Proceedings of the Cambridge Philosophical Society*, 1930, **26**, 361-375.
85. P. Ehrenfest, *Zeitschrift für Physik A Hadrons and Nuclei*, 1927, **45**, 455-457.
86. J. B. Delos, W. R. Thorson and S. K. Knudson, *Physical Review A*, 1972, **6**, 709.
87. Z. Deng, G. J. Martyna and M. L. Klein, *Physical Review Letters*, 1992, **68**, 2496.
88. Z. Deng, G. J. Martyna and M. L. Klein, *Physical Review Letters*, 1993, **71**, 267.
89. J. C. Tully, in *Modern Theoretical Chemistry: Dynamics of Molecular Collisions*, ed. W. H. Miller, New York, Plenum Press, Editon edn., 1976, vol. B, p. 217.
90. H.-D. Meyer and W. H. Miller, *The Journal of Chemical Physics*, 1979, **70**, 3214-3223.
91. U. Saalman and R. Schmidt, *Zeitschrift für Physik D Atoms, Molecules and Clusters*, 1996, **38**, 153-163.
92. I. S. Y. Wang and M. Karplus, *Journal of the American Chemical Society*, 1973, **95**, 8160-8164.
93. A. Warshel and M. Karplus, *Chemical Physics Letters*, 1975, **32**, 11-17.
94. C. Leforestier, *The Journal of Chemical Physics*, 1978, **68**, 4406-4410.
95. G. Kresse and J. Furthmüller, *Physical Review B*, 1996, **54**, 11169.
96. G. Kresse and J. Furthmüller, *Computational Materials Science*, 1996, **6**, 15-50.
97. J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Computer Physics Communications*, 2005, **167**, 103-128.
98. J. M. Herbert and M. Head-Gordon, *Physical Chemistry Chemical Physics*, 2005, **7**, 3269-3275.



99. E. Wimmer, *Science*, 1995, **269**, 1397-1398.
100. W. Kohn, *Reviews of Modern Physics*, 1999, **71**, 1253.
101. A. P. John, *Angewandte Chemie International Edition*, 1999, **38**, 1894-1902.
102. P. Hohenberg and W. Kohn, *Physical Review*, 1964, **136**, B864.
103. W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, A1133.
104. R. M. Wentzcovitch and J. Martins, *Solid State Communications*, 1991, **78**, 831-834.
105. R. N. Barnett, U. Landman, A. Nitzan and G. Rajagopal, *The Journal of Chemical Physics*, 1991, **94**, 608-616.
106. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias and J. D. Joannopoulos, *Reviews of Modern Physics*, 1992, **64**, 1045.
107. G. Kresse and J. Hafner, *Physical Review B*, 1993, **47**, 558.
108. R. N. Barnett and U. Landman, *Physical Review B*, 1993, **48**, 2081.
109. M. J. Gillan, *Journal of Physics: Condensed Matter*, 1989, **1**, 689.
110. M. P. Teter, M. C. Payne and D. C. Allan, *Physical Review B*, 1989, **40**, 12255.
111. I. Åtich, R. Car, M. Parrinello and S. Baroni, *Physical Review B*, 1989, **39**, 4997.
112. M. Sprik, in *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*, eds. K. Binder and G. Ciccotti, Italian Physical Society SIF, Bologna, Editon edn., 1996.
113. B. G. Johnson, P. M. W. Gill and J. A. Pople, *The Journal of Chemical Physics*, 1993, **98**, 5612-5626.
114. M. Ernzerhof and G. E. Scuseria, *The Journal of Chemical Physics*, 1999, **110**, 5029-5036.

115. F. Sim, A. St. Amant, I. Papai and D. R. Salahub, *Journal of the American Chemical Society*, 1992, **114**, 4391-4400.
116. Y.-M. Juan and E. Kaxiras, *Physical Review B*, 1993, **48**, 14944.
117. A. J. Cohen and N. C. Handy, *Chemical Physics Letters*, 2000, **316**, 160-166.
118. V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *The Journal of Chemical Physics*, 2004, **121**, 11507-11507.
119. V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *Physical Review B*, 2004, **69**, 075102.
120. M. J. Field, *The Journal of Physical Chemistry*, 1991, **95**, 5104-5108.
121. S. A. Maluendes and M. Dupuis, *International Journal of Quantum Chemistry*, 1992, **42**, 1327-1338.
122. B. Hartke and E. A. Carter, *Chemical Physics Letters*, 1992, **189**, 358-362.
123. H. Bernd, A. G. Douglas and A. C. Emily, *International Journal of Quantum Chemistry*, 1993, **45**, 59-70.
124. J. Jellinek, V. Bonacic-Koutecky, P. Fantucci and M. Wiechert, *The Journal of Chemical Physics*, 1994, **101**, 10092-10100.
125. A. Heidenreich and J. Sauer, *Zeitschrift für Physik D Atoms, Molecules and Clusters*, 1995, **35**, 279-283.
126. B. Hartke and E. A. Carter, *The Journal of Chemical Physics*, 1992, **97**, 6569-6578.
127. B. Hartke and E. A. Carter, *Chemical Physics Letters*, 1993, **216**, 324-328.
128. D. A. Gibson and E. A. Carter, *The Journal of Physical Chemistry*, 1993, **97**, 13429-13434.
129. D. A. Gibson, I. V. Ionova and E. A. Carter, *Chemical Physics Letters*, 1995, **240**, 261-267.

130. A. J. R. da Silva, M. R. Radeke and E. A. Carter, *Surface Science*, 1997, **381**, L628-L635.
131. A. J. R. da Silva, J. W. Pang, E. A. Carter and D. Neuhauser, *The Journal of Physical Chemistry A*, 1998, **102**, 881-885.
132. Z. Liu, L. E. Carter and E. A. Carter, *The Journal of Physical Chemistry*, 1995, **99**, 4355-4359.
133. H. B. Schlegel, S. S. Iyengar, X. Li, J. M. Millam, G. A. Voth, G. E. Scuseria and M. J. Frisch, *The Journal of Chemical Physics*, 2002, **117**, 8694-8704.
134. J. Cizek, *The Journal of Chemical Physics*, 1966, **45**, 4256-4266.
135. Ji, rcaron, í, Ccaron, zcaron and ek, in *Advances in Chemical Physics*, ed. C. M. R. LeFebvre, Editon edn., 2007, pp. 35-89.
136. J. Cizek and J. Paldus, *International Journal of Quantum Chemistry*, 1971, **5**, 359-379.
137. J. Broughton and F. Khan, *Physical Review B*, 1989, **40**, 12098.
138. M. J. Field, *Chemical Physics Letters*, 1990, **172**, 83-88.
139. A. Caro, S. Ramos de Debiaggi and M. Victoria, *Physical Review B*, 1990, **41**, 913.
140. D. S. Wallace and *et al.*, *Journal of Physics: Condensed Matter*, 1991, **3**, 3879.
141. M. Pearson and *et al.*, *Journal of Physics: Condensed Matter*, 1993, **5**, 3221.
142. J. Ortega, J. P. Lewis and O. F. Sankey, *Physical Review B*, 1994, **50**, 10516.
143. G. Seifert, D. Porezag and F. Th, *International Journal of Quantum Chemistry*, 1996, **58**, 185-192.
144. J. Ortega, J. P. Lewis and O. F. Sankey, *The Journal of Chemical Physics*, 1997, **106**, 3696-3702.
145. I. D. Brett, *International Journal of Quantum Chemistry*, 1998, **69**, 317-325.



146. E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García and J. M. Soler, *physica status solidi (b)*, 1999, **215**, 809-817.
147. H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria and M. J. Frisch, *The Journal of Chemical Physics*, 2001, **114**, 9758-9763.
148. C. Raynaud, L. Maron, J.-P. Daudey and F. Jolibois, *Physical Chemistry Chemical Physics*, 2004, **6**, 4226-4232.
149. J. M. Herbert and M. Head-Gordon, *The Journal of Chemical Physics*, 2004, **121**, 11542-11556.
150. X. Li, J. C. Tully, H. B. Schlegel and M. J. Frisch, *The Journal of Chemical Physics*, 2005, **123**, 084106-084107.
151. G. A. Chass, M. A. Sahai, M. S. J. Law, S. Lovas, O. Farkas, A. Perczel, J.-L. Rivail and I. G. Csizmadia, *International Journal of Quantum Chemistry*, 2002, **90**, 933-968.
152. A. S. Bland, R. A. Kendall, D. B. Kothe, J. H. Rogers and G. M. Shipman, *CUG 2009 Proceedings*, 2009.
153. T. H. Cormen, C. E. Leiserson, R. L. Rivest and S. C., *Introduction to Algorithms*, 2 edn., The MIT Press, 2001.
154. S. Behnke, *Hierarchical Neural Networks for Image Interpretation*, 2003.
155. E. Brown, J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat and J. D. Cohen, *International Journal of Bifurcation and Chaos*, 2005, **15**, 803–826.
156. H. M. Cartwright and C. Jones, EUROCON 2003. Computer as a Tool. The IEEE Region 8, 2003.
157. B. Coppin, *Artificial Intelligence Illuminated*, 1 edn., JONES AND BARTLETT PUBLISHERS, 2004.
158. H. Cruse, *Neural Networks as Cybernetic Systems*, Brains, Minds & Media, 2006.

159. H. Demuth and M. Beale, Editon edn., 1998.
160. A. M. Di Giulio, E. Germani, E. Lesma, E. Muller, A. Gorio, M. A. Kon and L. Plaskota, *Neural Networks*, 2000, **13**, 365-375.
161. P. F. Dominey and F. Ramus, *Language and Cognitive Processes*, 2000, **15**, 87-127.
162. A. I. Galushkin, *Neural Networks Theory*, Springer-Verlag Berlin Heidelberg, 2007.
163. D. Graupe, *Principles Of Artificial Neural Networks*, 2 edn., World Scientific Publishing Co. Pte. Ltd., 2007.
164. B. Kröse and P. van der Smagt, *An introduction to Neural Networks*, 8 edn., 1996.
165. P. D. McNelis, *Neural networks in finance: gaining predictive edge in the market*, Elsevier Academic Press, 2005.
166. N. Nedjah, A. Abraham and L. d. M. Mourelle, *Neural Computing and Applications*, 2007, **16**, 207-208.
167. S. Nolfi and D. Parisi, *Handbook of brain theory and neural networks*, 2002, 418-421.
168. A. G. Pedersen and H. Nielsen, International Conference on Intelligent Systems for Molecular Biology, 1997.
169. S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.
170. R. Simpson, R. Williams, R. Ellis and P. F. Culverhouse, *Marine Ecology Progress Series*, 1992, **79**, 303–308.
171. L. Spaanenburg, R. Alberts, C. H. Slump and B. J. vanderZwaag, *Bioengineered and Bioinspired Systems*, 2003, **5119**, 273-284.
172. K. O. Stanley, B. D. Bryant and R. Miikkulainen, Evolutionary Computation, 2003. CEC '03. The 2003 Congress on, 2003.

173. A. G. Tijsseling and L. Berthouze, 8th International Conference on Neural Information Processing, Shanghai (China), 2001.
174. E. Tyugu, *Algorithms and Architectures of Artificial Intelligence*, 2007.
175. J. Vohradský, *Federation of American Societies for Experimental Biology*, 2001, **15**, 846-854.
176. H. Wang, H. Zheng and F. Azuaje, IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 2007.
177. O. L. White, D. D. Lee and H. Sompolinsky, *Physical Review Letters*, 2004, **92**, 4.
178. M. Melanie, *An introduction to Genetic Algorithms*, The MIT Press, Cambridge, Massachusetts; London, England, 1999.
179. D. E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning*, Addison Wesley, 1989.
180. D. B. Fogel, *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*, IEEE Press, New York, 2000.
181. G. Phillips-Wren, N. Ichalkaranje and L. C. Jain, *Intelligent Decision Making: An AI-Based Approach*, Springer, Berlin Heidelberg, 2008.
182. N. K. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, The MIT Press, Cambridge, Massachusetts; London, England, 1998.
183. N. J. Nilsson, *Principles of Artificial Intelligence*, Tioga Publishing Company, Palo Alto, California, 1980.
184. D. Rina and P. Judea, *J. ACM*, 1985, **32**, 505-536.
185. D. P. DiVincenzo, *Science*, 1995, **270**, 255-261.



186. L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin and R. J. Schoelkopf, *Nature*, 2009, **460**, 240-244.
187. D. P. DiVincenzo, *Fortschritte der Physik*, 2000, **48**, 771-783.
188. C. A. Trugenberger, *Physical Review Letters*, 2001, **87**.
189. J. M. Tour and H. T., *Nature*, 2008, **453**, 42-43.
190. L. Chua, *Circuit Theory, IEEE Transactions on Circuit Theory*, 1971, **18**, 507-519.
191. B. Mouttet, *Cybernetics and Information Technologies, Systems and Applications*, 2009.
192. L. O. Chua and S. M. Kang, *Proceedings of IEEE*, 1976, **64**, 209-223.
193. D. E. Knuth, *The Art of Computer Programming*, 2 edn., Addison-Wesley Professional, 1998.
194. C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, 1994.
195. S. Arora and B. Barak, *Computational Complexity: A Modern Approach*, Cambridge University Press, 2009.
196. O. Goldreich, *Computational Complexity: A Conceptual Perspective*, 1 edn., Cambridge University Press, 2008.
197. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
198. A. V. Aho, J. E. Hopcroft and J. D. Ullman, *The Design and Analysis of Computer Algorithms*, Addison-Wesley, 1974.
199. R. G. Dromey, *How to Solve It by Computer*, Prentice Hall, 1982.
200. R. Sedgewick, *Algorithms*, 2 edn., Addison-Wesley Pub, 1988.
201. G. J. Chaitin, *Algorithmic Information Theory*, Cambridge University Press, 2008.

202. *New Theoretical Concepts for Understanding Organic Reactions*, 1989.
203. H. A. Baldoni, R. D. Enriz, E. A. Jáuregui and I. G. Csizmadia, *Journal of Molecular Structure: THEOCHEM*, 1996, **363**, 167-178.
204. H. A. Baldoni, R. D. Enriz, E. A. Jáuregui and I. G. Csizmadia, *Journal of Molecular Structure: THEOCHEM*, 1997, **391**, 27-38.
205. H. A. Baldoni, R. D. Enriz and I. G. Csizmadia, *Journal of Molecular Structure: THEOCHEM*, 1999, **463**, 251-270.
206. Y. Bengio, N. Le Roux, P. Vincent, O. Delalleau and P. Marcotte, *Neural Information Processing Systems*, Vancouver, British Columbia, Canada, 2005.
207. R.-M. Memmesheimer and M. Timme, *Physical Review Letters*, 2006, **97**, 4.
208. A. S. d'Avila Garcez and D. M. Gabbay, *American Association for Artificial Intelligence*, 2004, 6.
209. V. A. Ptitchkin, *The 2nd International Conference on Neural Networks and Artificial Intelligence*, 2001.
210. G. Conde, P. Ramos and G. Vasconcelos, in *Proceedings of the VI Brazilian Symposium on Neural Networks (SBRN'00)*, IEEE Computer Society, Editon edn., 2000.
211. W. S. McCulloch and W. Pitts, *Mathematical Biophysics*, 1943, **5**, 18-27.
212. R. Blum, *Professional Assembly Language*, Wiley Publishing, Inc., Indianapolis, Indiana, USA, 2005.
213. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, J. Vreven, T., K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev,

- A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, Editon edn., 2004.
214. CPMD, <http://www.cpmc.org/>, Accessed May 5, 2010.
215. R. F. W. Bader, *Atoms in molecules: a Quantum Theory*, Oxford University Press, Oxford, UK, 1990.
216. R. F. W. Bader, *Physical Review B*, 1994, **49**, 13348.
217. R. F. W. Bader, J. R. Cheeseman, K. E. Laidig, K. B. Wiberg and C. Breneman, *Journal of the American Chemical Society*, 1990, **112**, 6530-6536.
218. R. F. W. Bader and P. F. Zou, *Chemical Physics Letters*, 1992, **191**, 54-58.
219. R. Glaser and G. S. C. Choy, *Journal of the American Chemical Society*, 1993, **115**, 2340-2347.
220. W. B.-k. Friedrich, F. W. B. Richard and T. Ting-Hua, *Journal of Computational Chemistry*, 1982, **3**, 317-328.
221. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623-11627.
222. K. Kim and K. D. Jordan, *The Journal of Physical Chemistry*, 1994, **98**, 10089-10094.
223. J. Eargle, D. Wright and Z. Luthey-Schulten, *Bioinformatics*, 2006, **22**, 504-506.
224. D. Frishman and P. Argos, *Proteins: structure, function and genetics*, 1995, **23**, 566-579.



225. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33-38.
226. M. Sanner, A. Olsen and J.-C. Spehner, *Proceedings of the 11th ACM Symposium on Computational Geometry*, 1995, C6-C7.
227. R. Sharma, M. Zeller, V. I. Pavlovic, T. S. Huang, Z. Lo, S. Chu, Y. Zhao, J. C. Phillips and K. Schulten, 2000, **20**, 29-37.
228. J. Stone, Computer Science Department, University of Missouri-Rolla, 1998.
229. J. Stone, J. Gullingsrud, P. Grayson and K. Schulten, *2001 ACM Symposium on Interactive 3D Graphics*, 2001, 191-194.
230. A. Varshney, F. P. Brooks and W. V. Wright, *IEEE Computer Graphics and Applications*, 1994, **14**, 19-25.
231. B.-K. Friedrich, *Journal of Computational Chemistry*, 2000, **21**, 1040-1048.
232. F. Biegler-König, J. Schönbohm and D. Bayles, *Journal of Computational Chemistry*, 2001, **22**, 545-559.
233. B.-K. Friedrich and S. Jens, *Journal of Computational Chemistry*, 2002, **23**, 1489-1494.
234. D. C. Scott and C. H. Alan, *Comput. Phys.*, 1996, **10**, 138-143.
235. X. D. Xiang, X. Sun, G. Briceno, Y. Lou, K.-A. Wang, H. Chang, W. G. Wallace-Freedman, S.-W. Chen and P. G. Schultz, *Science*, 1995, **268**, 1738-1740.
236. C. F. Barbas, A. S. Kang, R. A. Lerner and S. J. Benkovic, *Proceedings of the National Academy of Sciences of the United States of America*, 1991, **88**, 7978-7982.
237. B. A. Bunin, M. J. Plunkett and J. A. Ellman, *Proceedings of the National Academy of Sciences of the United States of America*, 1994, **91**, 4708-4712.
238. M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett and M. A. Gallop, *Proceedings of the National*

- Academy of Sciences of the United States of America*, 1993, **90**, 10700-10704.
239. A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818-822.
240. R. Apodaca and e. al., Editon edn., 2005, p. The Open Babel Package.
241. R. Guha, M. T. Howard, G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. r. Wegner and E. L. Willighagen, *Journal of Chemical Information and Modeling*, 2006, **46**, 991-998.
242. G. A. Zhurko, Editon edn., p. Chemcraft.
243. G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *Journal of Molecular Biology*, 1963, **7**, 95-99.
244. B. Honig and A. Nicholls, *Science*, 1995, **268**, 1144-1149.
245. C. J. Cramer and D. G. Truhlar, *Chemical Reviews*, 1999, **99**, 2161-2200.
246. P. Jianfeng, W. Qi, Z. Jiaju and L. Luhua, *Proteins: Structure, Function, and Bioinformatics*, 2004, **57**, 651-664.
247. X. Zou, Yaxiong and I. D. Kuntz, *Journal of the American Chemical Society*, 1999, **121**, 8033-8043.
248. K. S. Brian, R. L. Andrew and D. K. Irwin, *Proteins: Structure, Function, and Genetics*, 1999, **34**, 4-16.
249. H. Y. Samuel and C. V. Shri, *Journal of Pharmaceutical Sciences*, 1980, **69**, 912-922.
250. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Advanced Drug Delivery Reviews*, 2001, **46**, 3-26.
251. H. Wang and A. Ben-Naim, *Journal of Medicinal Chemistry*, 1996, **39**, 1531-1539.
252. W. L. Jorgensen and E. M. Duffy, *Advanced Drug Delivery Reviews*, 2002, **54**, 355-366.

253. G. Pèpe, G. Guiliani, S. Loustalet and P. Halfon, *European Journal of Medicinal Chemistry*, 2002, **37**, 865-872.
254. W. L. Jorgensen, *Accounts of Chemical Research*, 1989, **22**, 184-189.
255. D. L. Beveridge and F. M. DiCapua, *Annu Rev Biophys Biophys Chem.*, 1989, **18**, 431-492.
256. P. A. Kollman and K. M. Merz, *Accounts of Chemical Research*, 1990, **23**, 246-252.
257. T. P. Straatsma and J. A. McCammon, *The Journal of Chemical Physics*, 1991, **95**, 1175-1188.
258. J. M. Haile, *Molecular Dynamics Simulation*, Wiley-Interscience, New York, 1992.
259. A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley-Interscience, New York, 1991.
260. L. Onsager, *Journal of the American Chemical Society*, 1936, **58**, 1486-1493.
261. J. G. Kirkwood, *The Journal of Chemical Physics*, 1939, **7**, 911-919.
262. O. Tapia and O. Goscinski, *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 1975, **29**, 1653 - 1661.
263. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9 edn., Dover, New York, 1972.
264. M. Born, *Zeitschrift für Physik A Hadrons and Nuclei*, 1920, **1**, 45-48.
265. A. A. Rashin and B. Honig, *The Journal of Physical Chemistry*, 1985, **89**, 5588-5593.
266. R. Cammi, B. Mennucci and J. Tomasi, *The Journal of Physical Chemistry A*, 1999, **103**, 9100-9108.
267. B. Mennucci, R. Cammi and J. Tomasi, *International Journal of Quantum Chemistry*, 1999, **75**, 767-781.



268. J. B. Foresman, T. A. Keith, K. B. Wiberg, J. Snoonian and M. J. Frisch, *The Journal of Physical Chemistry*, 1996, **100**, 16098-16104.
269. M. Wang, K. M. Liechti, Q. Wang and J. M. White, *Langmuir*, 2005, **21**, 1848-1857.
270. M.-H. Kim, S.-Y. Kim, B.-H. Kim, H.-G. Woo, S.-H. Park and Y.-J. Kim, *Journal of Nanoscience and Nanotechnology*, 2008, **8**, 4834-4837.
271. J. P. Matinlinna, L. V. J. Lassila, M. Özcan, A. Yli-Urpo and P. K. V., *The International Journal of Prosthodontics*, 2002, **17**, 155-164.
272. H. A. Clark and E. P. Plueddemann, *Modern Plastics*, 1963, **40**, 133-135, 137-138, 195-196.
273. M. W. Daniels and L. F. Francis, *Journal of Colloid and Interface Science*, 1998, **205**, 191-200.
274. E. K. U. Larsen, T. Nielsen, T. Wittenborn, H. Birkedal, T. Vorup-Jensen, M. H. Jakobsen, L. Astergaard, M. R. Horsman, F. Besenbacher, K. A. Howard and J. Kjems, *ACS Nano*, 2009, **3**, 1947-1951.
275. M. L. Cosnier, F. Martin, A. Bouamrani, F. Berger and P. Caillat, *Biomedical Engineering, IEEE Transactions on*, 2009, **56**, 2898-2904.
276. K. Kanta and H. L. Morton, *Journal of Polymer Science Part C: Polymer Letters*, 1988, **26**, 25-32.
277. V. Sambhy, B. R. Peterson and A. Sen, *Langmuir*, 2008, **24**, 7549-7558.
278. J. R. Damewood and R. West, *Macromolecules*, 1985, **18**, 159-164.
279. K. L. Mittal, *Silanes and Other Coupling Agents*, 2007.
280. S. Sterman and J. G. Marsden, *Industrial & Engineering Chemistry*, 1966, **58**, 33-37.
281. S. Shuichi, N. Takashi, K. Tadashi and K. Yoshiro, *Journal of Biomedical Materials Research*, 2001, **55**, 277-284.

282. M. B. Blatz, A. Sadan and M. Kern, *The Journal of Prosthetic Dentistry*, 2003, **89**, 268-274.
283. K. J. M. Soderholm and S. W. Shang, *Journal of Dental Research*, 1993, **72**, 1050-1054.
284. M.-C. Brochier Salon, M. Abdelmouleh, S. Boufi, M. N. Belgacem and A. Gandini, *Journal of Colloid and Interface Science*, 2005, **289**, 249-261.
285. J. P. Matinlinna, L. V. J. Lassila and P. K. Vallittu, *Journal of Dentistry*, 2006, **34**, 721-726.
286. J. P. Matinlinna, L. V. J. Lassila, M. Özcan, A. Yli-Urpo and P. K. V., *The International Journal of Prosthodontics*, 2004, **17**, 155-164.
287. T. Hooshmand, R. van Noort and A. Keshvad, *Dental Materials*, 2002, **18**, 179-188.
288. F. D. Blum, W. Meesiri, H.-J. Kang and J. E. Gambogi, *Journal of Adhesion Science and Technology*, 1991, **5**, 479-496.
289. F. Beari, M. Brand, P. Jenkner, R. Lehnert, H. J. Metternich, J. Monkiewicz and H. W. Siesler, *Journal of Organometallic Chemistry*, 2001, **625**, 208-216.
290. J. P. Matinlinna, L. V. J. Lassila, I. Kangasniemi, A. Yli-Urpo and P. K. Vallittu, *Dental Materials*, 2005, **21**, 287-296.
291. Y. Cho, S. Gorina, P. D. Jeffrey and N. P. Pavletich, *Science*, 1994, **265**, 346-355.
292. E. H. Koo, P. T. Lansbury and J. W. Kelly, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 9989-9990.
293. C. M. Dobson, *Nature*, 2002, **418**, 729-730.
294. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. Watson, *Molecular Biology of the Cell*, 3 edn., Garland Publishing, New York, 1994.
295. E. Alexov, *European Journal of Biochemistry*, 2004, **271**, 173-185.

296. S. T. Whitten and B. Garcia-Moreno, *Biochemistry*, 2000, **39**, 14292-14304.
297. C. Scharnagl, R. Raupp-Kossmann and S. F. Fischer, *Biophys. J.*, 1999, **77**, 1839-1857.
298. M. Tollinger, K. A. Crowhurst, L. E. Kay and J. D. Forman-Kay, *Proceedings of the National Academy of Sciences*, 2003, **100**, 4545-4550.
299. A. M. Pots, H. H. J. de Jongh, H. Gruppen, M. Hessing and A. G. J. Voragen, *Journal of Agricultural and Food Chemistry*, 1998, **46**, 2546-2553.
300. R. Khurana, A. T. Hate, U. Nath and J. B. Udgaonkar, *Protein Sci*, 1995, **4**, 1133-1144.
301. T. Brink and T. E. Exner, *Chemistry Central Journal*, 2008, **2**, P12.
302. M. Długosz and J. M. Antosiewicz, *Journal of Physics: Condensed Matter*, 2005, **17**, S1607-S1616.
303. C. N. Pace, D. V. Laurents and R. E. Erickson, *Biochemistry*, 1992, **31**, 2728-2743.
304. C. N. Pace, D. V. Laurents and J. A. Thomson, *Biochemistry*, 1990, **29**, 2564-2572.
305. S. T. Whitten, J. O. Wooll, R. Razeghifard, E. B. Garcia-Moreno and V. J. Hilser, *Journal of Molecular Biology*, 2001, **309**, 1165-1175.
306. P. Del Vecchio, G. Graziano, V. Granata, G. Barone, L. Mandrich, M. Rossi and G. Manco, *Biochem. J.*, 2002, **367**, 857-863.
307. A.-S. Yang and B. Honig, *Journal of Molecular Biology*, 1993, **231**, 459-474.
308. A.-S. Yang and B. Honig, *Journal of Molecular Biology*, 1994, **237**, 602-614.
309. C. M. Soares, V. H. Teixeira and A. M. Baptista, *Biophysical Journal*, 2003, **84**, 1628-1641.
310. Y. Xiao and M. Braiman, *The Journal of Physical Chemistry B*, 2005, **109**, 16953-16958.



311. A. Makoto, Y. Takeshi, J. W. Ronald, C. P. Julio, T. C. Brian, P. Anthony, R. M. Lois and M. M. James, *Protein Science*, 2005, **14**, 1458-1471.
312. T. Ashutosh, F. Micaela, S. Francesca, M. Andrea, C. Pietro and E. K. Glen *Chemistry & Biodiversity*, 2007, **4**, 2564-2577.
313. T. D. Vaden, T. S. J. A. d. Boer, J. P. Simons and L. C. Snoek, *Physical Chemistry Chemical Physics*, 2008, **10**, 1443-1447.
314. J. A. Vila, D. R. Ripoll, M. E. Villegas, Y. N. Vorobjev and H. A. Scheraga, *Biophysical Journal*, 1998, **75**, 2637-2646.
315. G. A. Olah, A. Burrichter, G. Rasul and G. K. S. Prakash, *Journal of the American Chemical Society*, 1997, **119**, 12929-12933.
316. Vilarino, T, Fiol, S, Armesto, L. X, Brandariz, I, V. Sastre De and E. M, *Effect of ionic strength on the protonation of various amino acids analysed by the mean spherical approximation*, Royal Society of Chemistry, Cambridge, ROYAUME-UNI, 1997.
317. D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman and Company, New York, 2005.
318. F. Li and J. Z. Tsien, *N Engl J Med*, 2009, **361**, 302-303.
319. J. U. Andersen, H. Cederquist, J. S. Forster, B. A. Huber, P. Hvelplund, J. Jensen, B. Liu, B. Manil, L. Maunoury, S. B. Nielsen, U. V. Pedersen, J. Rangama, H. T. Schmidt, S. Tomita and H. Zettergren, *Physical Chemistry Chemical Physics*, 2004, **6**, 2676-2681.
320. G. Grégoire, B. Lucas, M. Barat, J. A. Fayeton, C. Dedonder-Lardeux and C. Jouvet, *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 2009, **51**, 109-116.

## 10. INDEX

---

- 3-isocyanatopropyltriethoxysilane .*See*  
ICS
- 3-styrylethyltrimethoxysilane.....*See*  
STYRX
- ab initio* molecular dynamics...11, 12,  
18, 20, 59
- action potential ..... 46
- AD ..*See ab initio* molecular dynamics
- AIM ..... *See* Atoms-In-Molecules  
Contour Interval ..... 60  
Critical Point..... 60  
Line Plot ..... 60  
Molecular Graph..... 60  
theory ..... 60
- AIM2000 ..... 59, 98, 111
- algorithm ..... 29, 54, 61  
A\* search..... 31  
Breadth-First Search..... 43  
Depth-First Search..... 43  
Dynamic Programming ..... 45  
Genetic..... 17, 31, 45  
Graph ..... 43  
Greedy ..... 44  
Linear Programming..... 44  
optimisation ..... 44  
runtime..... 29  
solvation ..... 41, 85, 87
- algorithm runtime ..... 32  
average-case ..... 35  
best-case ..... 35  
exponential ..... 36  
worst-case..... 35
- amino acid ..... 108  
Cysteine ..... 108  
glycine ..... 109  
Guanidine ..... 108  
Leucine..... 108  
Methionine ..... 108  
Threonine ..... 108
- ANN .... *See* Artificial Neural Network
- Artificial Intelligence ..... 27, 30
- Artificial Neural Network.....18, 31,  
45, 48, 81  
model..... 18
- Assembly..... 51
- Atoms-In-Molecules ..... 54, 98, 111
- Avogadro constant ..... 27
- axon..... 46
- B3LYP ..... 57
- Bader atomic radii ..... 86
- basis set ..... 23, 110  
Gaussian-type..... 23  
Slater-type ..... 23
- BCP ..... *See* Bond-Critical Point
- BFS.....*See* algorithm : Breadth-  
First Search
- bioadhesives ..... 92
- bond  
angle ..... 24  
distance..... 24
- Bond-Critical Point ..... 59
- brute-force approach ..... 64
- buffer zone size ..... 82
- byte..... 28
- BZS ..... *See* buffer zone size
- Cage-Critical Point..... 59

Car-Parrinello Molecular Dynamics .....	89, 98, 110, 125	DFS..... <i>See</i> algorithm : Depth-First Search	
Cartesian coordinates ....	25, 58, 65, 77	DFT .... <i>See</i> Density Functional Theory	
CASSCF..... <i>See</i> Complete Active Space Self-Consistent Field		dihedral angle.....	24, 33
Catalan's constant.....	27	related.....	69
CCP .....	<i>See</i> Cage-Critical Point	dihedral driver method .....	17
Central Processing Unit.....	51	dimensionality bottleneck .....	20
ChemConverter .....	65	discrete representation.....	30
ChemCraft .....	65	distance between two points .....	37
Complete Active Space Self-Consistent Field.....	22	dot product .....	38
computational		<i>e</i> .....	27
chemistry .....	15	electronic structure .....	11, 22, 56
geometry.....	37	energizer.py.....	74
method .....	27	energy	
computer cluster .....	27	kinetic.....	20
conformational		Kohn-Sham .....	22
analysis .....	16	potential.....	20
space .....	32	ENP .....	85
conformer .....	33	Euclidean	
Coupled Cluster.....	22	point .....	37
CPMD.....	54, 58, 59, 98, 110, 119	vector.....	38
CPU .....	<i>See</i> Central Processing Unit	explicit solvent method .....	86
cross product.....	38	FCI <i>See</i> Full Configuration Interaction	
data structure .....	39	floating point	
acyclic graph.....	41	number.....	27
adjacency list .....	41	operation.....	13
adjacency matrix.....	41	FLOP .....	<i>See</i> floating point operation
array.....	39	force field .....	16
doubly linked list.....	41	Boyd's .....	16
graph .....	41	CFF.....	16
linked list .....	40	Class-2.....	17
pointer.....	40	Class-3.....	17
queue .....	40	EAS .....	16
stack.....	40	ECEPP.....	16
tree .....	41	MMI .....	16
degree of freedom.....	24	MUB.....	17
dendrite.....	46	UNICEPP .....	16
Density Functional Theory ..	21, 22, 97	FTIR spectroscopy .....	96
dentistry .....	92	Full Configuration Interaction.....	22
		Fuzzy logic .....	31
		g03_master .....	71



g03_slave.....	71	MDCA.....	<i>See</i> Multi-Dimensional Conformational Analysis
Gaussian directive		memristor .....	31
freq.....	57	MM.....	<i>See</i> Molecular Mechanics
opt.....	57	molecular	
wfn.....	57	model.....	15
Gaussian input file.....	56	modelling .....	15
Gaussian03 .....	54	pathway .....	80
GaussView.....	58	surface .....	17
Generalised Valence Bond .....	22	molecular behaviour.....	30
glycine.....	111, 125, <i>See</i>	molecular dynamics	
amino acid : glycine		ab initio .....	12, 14
h2n-g-oh .....	115	Born-Oppenheimer.....	21, 23
hn-gh1-oh .....	118	Car-Parrinello .....	22, 23
hn-gh2-oh .....	118	Ehrenfest .....	21, 23
hn-g-oh2 .....	117	Molecular Mechanics.....	12, 14
hn-hg-oh .....	116	semi-empirical.....	13, 14, 22
non-protonated.....	113	Møller-Plesset Perturbation Theory	22, 57, 110
grid-based method .....	64	Monte Carlo method .....	16
growth of functions .....	28, 32	Moore's law .....	31
GTO.....	<i>See</i> basis set : Gaussian-type	MP2.....	<i>See</i> Møller-Plesset Perturbation Theory
GVB....	<i>See</i> Generalised Valence Bond	Multi-Dimensional Conformational Analysis.....	34
GvsS .....	78	neuron.....	46
Hamiltonian .....	20	neuron activation function	
hard disk drive .....	28	linear.....	46
Hartree-Fock Theory .....	22, 57	sigmoid.....	47
hydroxyapatite .....	93	step .....	46
ICS.....	94, 104, 105, 123	neuron activation threshold.....	46
IDE.....	<i>See</i> Integrated Development Environment	NiceIR .....	77
input file generator .....	68	NP-complex problem .....	29, 36
Integrated Development Environment .....	52	numbering system .....	26, 65, 66
intermolecular association.....	85	O-LED.....	92
Internal coordinates .....	25, 58, 66	optimiser_simulator.py .....	80
IPCM .....	<i>See</i> Isodensity PCM	pattern recognition .....	45
Isodensity PCM .....	86	PCM <i>See</i> Polarized Continuum Model	
Kirkwood-Onsager model .....	85	PEHS.....	33, 35, 37, 80, 110, 122, <i>See</i> potential energy hyper surface
line segment intersection .....	39	pH.....	107
Machine Language .....	51		

optimum.....	108	Schrödinger equation .....	17, 20
PhiPsi.....	78	SCIPCM....	<i>See</i> Self-Consistent IPCM
planning.....	31	SCRf.....	<i>See</i> Self-Consistent Reaction Field
Polarized Continuum Model .....	86	Self-Consistent IPCM .....	86
potential energy hyper surface..13, 17, 95		Self-Consistent Reaction Field.....	85
probabilistic quantum memories .....	31	silane	
processor.py.....	76	condensation reaction.....	92
programming		coupling agent.....	92
framework .....	51, 61	surface binding.....	93
language.....	30, 51	simulated annealing method.....	16
protein.....	107	solvation surface.....	87
structure stability .....	107	Solvator .....	87
proton transfer .....	118	solvent effect .....	85
inter-molecular .....	118	steepest descent method .....	16
intra-molecular .....	118	STO .....	<i>See</i> basis set : Slater-type
protonation.....	32	STYRX.....	83, 104, 123
site .....	33, 110	Taylor series .....	85
quantum computing.....	31	TDSCF.....	<i>See</i> Time-dependend self-consistent field
queuing system .....	70	Time-dependend self-consistent field	20
RAM.....	<i>See</i> Random Access Memory	vibrational spectroscopy.....	16
Ramachandran plot.....	78	VMD .....	59
Random Access Memory .....	28	water.....	29
RCP .....	<i>See</i> Ring-Critical Point	wave function.....	20, 58
recurrence .....	36	Z-matrix.....	25, 65
Restricted HF.....	57	$\pi$ .....	27
Ring-Critical Point .....	59		
scalar product .....	39		