# TOWARDS AN ACCURATE AND EFFICIENT METHOD FOR OBTAINING THE ENERGIES OF LARGE WATER CLUSTERS

**TAN PEIYU, AMELIA**
*(B.Sc.(Hons), UWA)*

**A THESIS SUBMITTED**

**FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**
**DEPARTMENT OF CHEMISTRY**
**NATIONAL UNIVERSITY OF SINGAPORE**

**2015**

# DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety, under the supervision of Assoc. Prof. Ryan P. A. Bettens, in the laboratory S7-04-05, from the Department of Chemistry at the National University of Singapore.

I have duly acknowledged all sources of information which have been used in this thesis. The thesis has also not been submitted for any degree to any universities previously.

The content of this thesis has been partly published in the journal of *CrystEngComm*. **2013**, *15*, 4917-4929.

Tan Peiyu, Amelia
3rd April 2015

# ACKNOWLEDGMENTS

First and foremost, I would like to express my heartfelt gratitude to my supervisor Assoc. Prof. Ryan P. A. Bettens for his continuous guidance and support throughout the duration of my graduate studies and research. I am sincerely grateful for this wonderful and meaningful research project he has offered me. I would also like to thank him for all his constructive advice and the great learning opportunities throughout my Ph.D. research years. His never-ending patience, enthusiasm and motivation have driven me to persevere to the very end. I would also like to thank my examiners for the Qualifying Examination for my Ph.D. candidature – Prof. Wong Ming Wah, Richard, Assoc. Prof. Fan Wai Yip and Dr. Lee Adrian Michael for their immense knowledge and academic guidance. A special note to Dr. Lee Adrian Michael, who always join our research group for the many lunch and dinner occasions, as well as the late nights of karaoke and drinking sessions. I am thankful to all my Science teachers and lecturers who have taught me in the past. Without all of you, I would not have loved Chemistry the way I do right now, and I would not have come this far.

I am filled with boundless gratitude to my fellow working buddies: John Ouyang, Ng Yee Hong, Le Hai Anh Ryo, Michael Patuwo, Krishnan and Ali, and not forgetting all the past and present honors students: Tan Ching, Yi Bei, Simin, Chia Ling, Sok Koon, Shawn Foo and See Kiat, who had been together with me in the same lab. We had many great moments working together and it would have been dull without the uniqueness from all of you. A special shout out to Ali for his programming knowledge and numerous long conversations about the mathematical derivatives and equations for our computational program, written in the Fortran language. I would also like to thank the seniors in our research group, especially Dr. Krishnan, Dr. Michael and Le Hai Anh Ryo who have been sources of inspiration in my early days of research.

I would also like to show my appreciation to the Centre for Computational Science and Engineering (CCSE) for allowing me to use their CPUs for faster calculations, especially Yung Shing Gene for providing the support and resources I needed to do a significant part of my work. I am also thankful to those who had helped me

during the course of my time at the National University of Singapore in various ways, namely Miss Suriawati Binte Sa'ad as well as Miss Agnes for all the help they rendered in non-academic and administrative matters. My gratitude is extended as well to those who had helped me in ways that I may not be aware of or have forgotten. I am honored to be a student at the National University of Singapore (NUS) and am greatly indebted for the graduate research scholarship that was offered to me. Without that, it would have been impossible for me to do my graduate studies at NUS.

Last but certainly not least to my friends, relatives and family for all their love, trust, care, patience, kindness, tremendous support and words of encouragement. I would like to say a very Big thank you to my extremely supportive parents, younger sister Anissa, elder brother Alvin, my sister-in-law, baby nephew Ayden, cousins, uncles, aunties and grandparents. A very special thanks to my mother who drives me to NUS every day and the love of my life Patrick Chu for his boundless support who sticks with me through thick and thin.

Looking back, I do not and will never regret the decision of doing my Ph.D. in NUS. These treasure moments will stay with me forever, whatever the future may be.

# Table of Contents

# SUMMARY

This thesis lays the ground work necessary for the most accurate and detailed simulation of liquid bulk water ever undertaken. Despite considerable effort for more than a century of intense scientific endeavor, this enigma of a material has eluded complete understanding of its properties and behaviour. We examine liquid water in Chapter 1 and note it is very special, yet absolutely essential, anomalous properties and attempts to date to unravel its mysteries. We finish out the chapter considering the specific issues that make modeling water so very difficult and a description of what lies ahead in the rest of the thesis.

Chapter 2 provides the background to all the theoretical methods employed in this thesis. They include the many-body treatment of intermolecular interaction energies, *ab initio* quantum mechanical evaluation of the total electronic energies for a collection of nuclei and the utilization of multipoles in accurately describing interaction energies at moderate to long-range. None of these methods in isolation can provide what is needed to accurately simulate bulk water – indeed, we spend some time pointing out the deficiencies in applying any one of them. However, if these methods could be combined judiciously so as to overcome their inherent limitations a route to understanding water may be within reach and this is the intention of Chapters 3 and 4.

In Chapter 3, we focus our attention on the precise details of exactly what is necessary to accurately simulate bulk water. We discover a requirement that at first sight seems impossibly high – the accurate determination of the total electronic energy or interaction energy of a large spherical cluster of water molecules. However, we show that with a cunning use of energy-based molecular fragmentation the impossible is dragged down to within reach. Nevertheless, difficulties still abound and we examine solutions to these in the next chapter.

While application of the many-body expansion enables accurate evaluation of the interaction energy for a large spherical water cluster, its direct implementation is essentially intractable in a bulk water simulation. This is due to the crushing weight of literally hundreds of thousands to even millions of energy evaluations using perturbation theory and multipoles. In Chapter 4 we show that by carefully considering which specific three-body interactions are significant, and which are not, we are able to remove vast numbers of these energy evaluations without any significant loss of accuracy. We show this towards the end of the chapter where we develop a simple, yet powerful, criterion for selecting out significant interactions. This criterion is not *ad hoc*, but based soundly in the origins of the three-body interaction itself.

The above research concludes our five-year effort in making an accurate bulk water simulation very possible – we leave its practical and detailed implementation for future work. In the final chapter of this thesis, we enter the realm of pure computational chemistry as we describe our contributions to a collaborative endeavor in which the role of aryl-substituents in moderating the nature of hydrogen bonds, N-H···N versus N-H···O, lead to supramolecular chains in the crystal structures of *N*-arylamino 1,2,3-triazole esters.

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1

# Introduction

## 1.1 The Importance of Water and its Anomalous Properties

Water is ubiquitous on Earth and is the third most abundant molecule in the Universe (after $H_2$ and CO). It plays an absolutely central role in living systems and is Earth's natural solvent. By "Earth's natural solvent" we mean that due to the location and size of the planet plus the early chemistry that took place in Earth's 4.6 billion year history liquid water is ubiquitous on the Earth's surface. Consequently virtually all of the chemistry that occurs on the surface in the liquid phase involves water as the solvent. The role of water in biology cannot be overstated and goes beyond merely a "space filler" between biomolecules. Water takes on many active roles in molecular biology[1] – being absolutely crucial for the existence of, and sustaining life.

The pivotal role water plays on Earth is not only due to its great abundance on the planet. Many properties of water are unique and deemed anomalous, and it is these unusual properties that ofttimes bestow essential and favourable behaviour to the system in which it is present. For example, water is the only liquid to expand when cooled – its density passing through a maximum at around 4 °C under atmospheric pressure. This unusual phenomenon in freshwater lakes prevents water from freezing from the bottom up in the lake and hence killing all of the organisms in it. As the surface water cools and approaches 4 °C its density increases, so sinks and is replaced by warmer, less dense, water. This circulation continues and potentially transports more oxygenated water to lower depths. Once the lake is uniformly cool, any water that is then made cooler than 4 °C will float and

eventually freeze, but at the surface of the lake and not the bottom. The process also ensures that while liquid water still exists in a frozen-over lake its bottom remains at least at 4 °C and unfrozen.

Water also exhibits unusual non-monotonic behaviour in its isothermal compressibility, $\beta_T$ and molar heat capacity. In any liquid, the fractional change in volume as pressure is exerted on the liquid at constant temperature[a] is tiny, but measurable. As the temperature is dropped, the fractional change in volume for the same pressure exerted is normally less than that observed at higher temperatures. This is expected because from statistical thermodynamics the isothermal compressibility is directly related to the fluctuations in the molar volume of the liquid which should get smaller as the temperature is dropped. For water, however, the isothermal compressibility passes through a minimum at around 42 °C (see ref. 2).

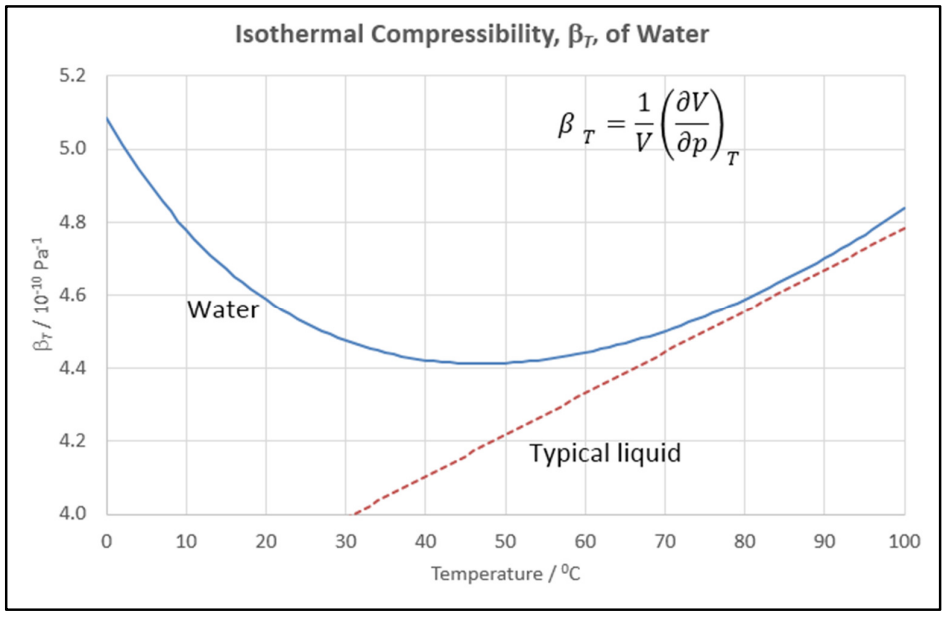



**Figure 1-1 Isothermal compressibility of liquid water as a function of temperature**
Water exhibits non-monotonic behaviour in its isothermal compressibility as it is cooled[3]. A minimum is observed at around 42 °C. However, a typical liquid (red dashed line), decreases monotonically as it is cooled.

[a] This is the definition is isothermal compressibility, $\beta_T = \frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T$.

More odd behaviour from water is observed in its particularly high boiling point for such a simple triatomic and its very large value of dielectric constant. The large value of the dielectric constant has profound consequences in electrostatic interactions present in the proteins of living systems. Understanding these properties, which are atypical compared to other substances of similar size, has been ongoing for many decades. As a consequence, although it is one of the most studied substances in science, complete understanding of water eludes us despite of the extensive work in unravelling its mysteries (see the reviews 4).

## 1.2 What Makes Water Unique?

The unusual properties of bulk water is primarily due to the hydrogen-bonding exhibited by water molecules. Additionally, its relatively small size, almost spherical shape and large dipole moment also make a significant contribution. The notion that water is almost spherical is counter intuitive given its familiar "V" ball-and-stick structural representation. However, this structure of water is specifically referring to the positions of nuclei in the molecule. After all, this is where essentially all of the mass is in any molecule. In this regard, water is far from spherical, as exhibited clearly by the very different three values of its rotational constants ($H_2{}^{16}O$): $A$ = 835.8 GHz, $B$ = 435.4 GHz and $C$ = 278.1 GHz[5].

Nevertheless, the physical extent of the molecule is governed by the surrounding electron cloud. This cloud is responsible for the exchange-repulsion interaction as any other molecule approaches at close range. The surface that best characterises the onset of this exponentially repulsive interaction is the van der Waals surface, which can be represented by an isosurface of electron density surrounding the nuclei of a molecule with a value of 0.002 $a_0^{-3}$. This surface is depicted in Figure

1-2, which well illustrates the almost spherical shape of water. The radius of this approximate sphere is 1.5 Å, which is essentially the van der Waals radius of the oxygen atom itself.

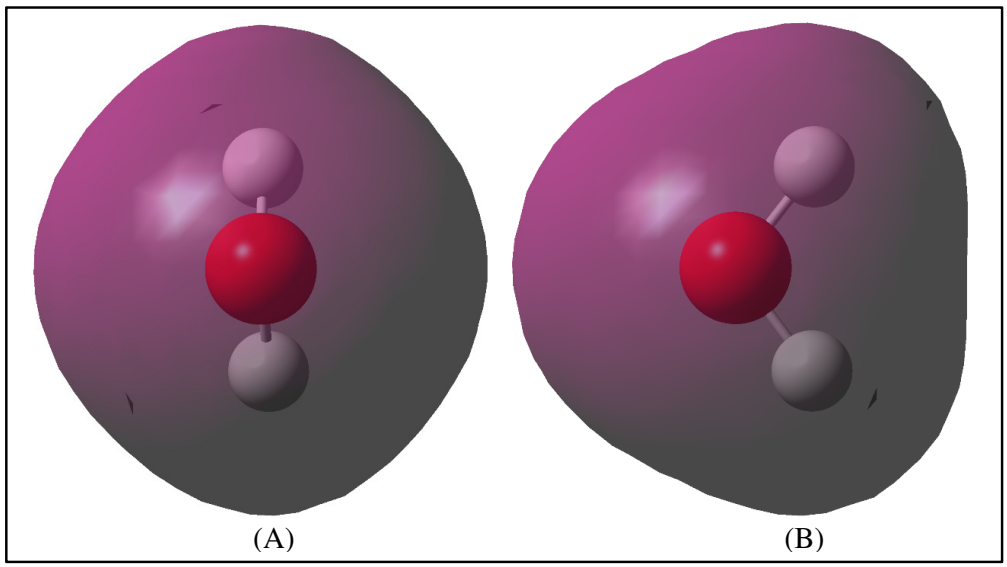



**Figure 1-2 Approximately spherical water**

Panels (A) and (B) illustrate an electron density isosurface ($\rho = 0.002\ a_0^{-3}$), which corresponds to the van der Waals surface, around a water molecule. While not exactly spherical it is clear that the physical shape of water is very nearly so.

Further evidence of the almost spherical shape of water comes from its dipole-dipole polarizability. High level theoretical calculations of the vibrationally averaged $J = 0$ principle dipole-dipole polarizabilities[6] are 1.47, 1.40, 1.55 Å$^3$ which results in an isotropic value of 1.47 Å$^3$ – in excellent agreement with experimental measurements. The three components of this second rank tensor deviate by less than 5% from the isotropic value indicating that, at least as far as electron dipole-dipole polarization is concerned, all three directions in water are almost the same. A consequence of this is that the induction interaction energy in water can be quite accurately described with the isotropic dipole-dipole polarizability.

The polarizability of a single water molecule is not particularly large, nor unusual. E.g., the isotropic polarizabilities of the O atom plus $H_2$ molecule (isotropic polarizabilities are approximately additive)[7] is 1.59 $Å^3$. The polarizabilities of the isoelectronic molecules[7] $CH_4$, $NH_3$, HF and Ne are 2.45, 2.10, 1.28 and 0.38 $Å^3$ respectively. We can see that the polarizability of water is nothing out of the ordinary and well within the observed trend which follows the increase in effective nuclear charge as one moves towards the noble gases – an increased effective nuclear charge means that electrons are more tightly bound to nuclei and are therefore less able to be polarised by an external field.

This typical value of molecule polarizability is unrelated to the extremely large value of the dielectric constant for water – bulk water is highly polarizable. This is largely due to the near spherical shape of water, thus enabling water molecules to almost freely be reoriented in solution, and more importantly, the large dipole of water. Any applied field subjected to a liquid sample of water, from without or within, produces a very large degree of polarization. This occurs because the molecules almost freely, on average, reorient and cooperatively counter the field with their permanent electrostatic dipoles.

While the large dipole of water is not unexpected compared with that of say, $NH_3$ or HF, the fact that a large collection of water molecules (i.e., $10^{23}$) under one bar pressure and at 25 °C forms a liquid in the first place is unexpected. The normal boiling points of $NH_3$ and HF are −33.3 and 19.5 °C respectively[8]. Liquid water appears to be far more stable than it should be. So as such, no other room temperature and pressure liquid possesses the very large value of dielectric constant that water does.

The reason for its anomalous boiling point and other anomalous properties can largely be traced to the unusual close-contact intermolecular interactions taking place in water, summarised in the term – hydrogen bonding (or simply H-bonding). A water molecule can form a maximum of four hydrogen-bonds because it can accept two and donate two hydrogen atoms, as illustrated in Figure 1-3. However for the interaction to be significant, and it is typically of the order of 20 kJ $mol^{-1}$, the O–H···X should be close to linear and the H···X distance should be less than the sum of the van der Waals radii of H and X. Thus in an H-bond, the hydrogen atom penetrates within the electron cloud of the acceptor atom, but rather than give rise to a very large exchange repulsion interaction, there is a substantial degree of stabilization that occurs (see panel (B) of Figure 1-3).

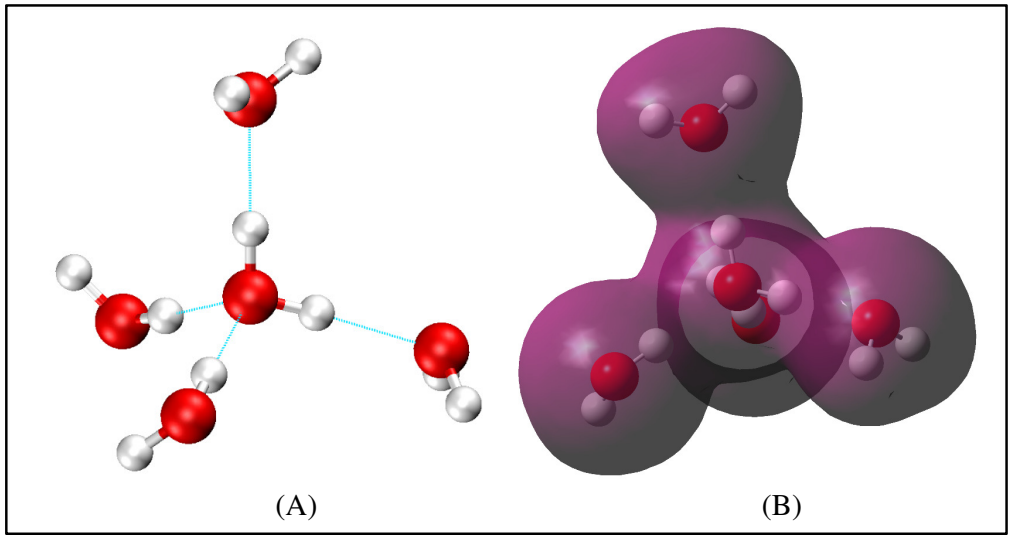



**Figure 1-3 Maximum hydrogen bonding for water**
Panel (A) shows the hydrogen-bonding that may occur around a single water molecule. The maximum number of neighbors is four. Panel (B) is an isosurface ($\rho = 0.002\ a_0^{-3}$) for electron density that closely approximates the van der Waals surface. It is of note that hydrogens in the hydrogen-bonds penetrate into the van der Waals surface of the acceptor (oxygen) atom.

This type of interaction allows water to approach much more closely to specific atoms, and in a specific orientation, than would otherwise be permitted. The

closeness of this approach produces stronger intermolecular interactions in addition to that of the stabilization afforded by the H-bond itself.

The specific interaction occurring due to H-bonding, plus the large dipole of water giving rise to significant induction effects within a collection of water molecules makes water notoriously difficult to model accurately in detail. H-bonding, as seen in the interpenetration of the electron cloud, is fundamentally quantum mechanical in nature. Additionally, the existence of significant induction interactions produces substantial non-additive interaction energies. Once the energy of a system contains a significant contribution from non-additive effects the computational effort required to model it increases substantially because simple pair-wise sums of interaction energies no longer accurately reflects the reality of the system. Nevertheless, a large number of attempts have been made to accurately model water and these attempts are briefly summarised in the next section.

## 1.3 Models of Water

As mentioned earlier, several excellent reviews already exist that well cover all attempts to accurately model water to date[4, 9]. However, it is important to briefly summarise what has been covered in the scientific literature here in order to put the work presented in this thesis in its proper context. As pointed out by Ouyang and Bettens[9], water models can be classified into three categories based upon the overall approach taken. Each approach addresses some of the failings of the previous, but in doing so increases the complexity of the model and the computational expense.

The first approach appends a classical pairwise charge-charge interaction potential to a Lennard-Jones potential. The charges used are enhanced[b] point charges placed on, or near, the atoms of a water molecule in an attempt to account for induction but in an additive (and thus, fundamentally flawed) way. These types of models are computationally the least expensive and are intended to be "transferable" in the sense that the same water parameters can be used in simulations with solutes (also treated in a manner similar to water, i.e., with point charges and Lennard-Jones potential parameters). The small number of parameters in these models are fitted to match as best as possible the bulk properties of water at around 25 – 37 °C. Such models do perform reasonably well at reproducing the bulk water properties at these temperatures, mostly because they have been parameterised to do so. Such models fail hopelessly in reproducing detailed geometries and interaction energies of a small number of water molecules, e.g., dimer, trimer, tetramer etc. or water clustered around some solute. They also fail in accounting for the anomalous properties of water over a wide temperature range. Despite these failings, these models continue to be utilised today as they do provide a qualitative picture of the solvent around large solutes like proteins. The reason being these models are currently the only models cheap enough to perform these enormous calculations. Examples of such models include the transferable intermolecular potential functions (TIP*n*P) family of models[10] and the single point charge (SPC) family of models[11].

In the second approach, the complete neglect of non-additive effects encountered in the former approach is addressed. Additionally, a substantially improved

[b] "Enhanced" compared to, say, what would be needed to reproduce the dipole of water.

treatment of the electrostatic interaction that occurs between water molecules or other solute species is implemented. In this approach, a set of electrostatic multipoles (rather than just a small number of monopoles utilised previously), or a large set of point charges, are added at various sites within the water molecule. The multipoles, usually up to and including quadrupoles, much better reproduce the electrostatic potential that surrounds a water molecule. Unfortunately, however, a multipole approach is only accurate at moderate to long-range, i.e., when two interacting molecules are at least 20 – 50% further away from each other than their respective van der Waals surfaces. Close-contact interactions are not well reproduced utilizing a multipolar approach. The non-additive interaction effects are also handled via multipoles, with the first order treatment requiring dipole-dipole polarizabilities. Again, while very accurate at moderate to long-range, close-contact interactions are not well treated with this approach. For close-contact interactions, either some type of functional form is implemented or the multipoles are "damped" such that numerical instabilities and highly inaccurate results are suppressed. Close-contact interactions not present at moderate to long-range include the H-bonding interaction and short-range exchange-repulsion. Furthermore, such models cannot account for purely quantum mechanical phenomena like hydrogen atom exchange between waters or proton transfer. Examples of this approach include the anisotropic site potential (ASP) family of models[12], the symmetry adapted perturbation theory (SAPT) family of models[13] and the Thole-Type models (TTM)[14].

The final approach abandons the use of multipoles in describing short range interactions due to their failure at such short distances and inevitable messy corrections. However, there is currently no theory for obtaining analytic

expressions for the detailed form of intermolecular close-contact interactions. As such, there is presently no alternative but to assume a flexible functional form for the interaction then fit its parameters to very high quality *ab initio* data. That is, a highly accurate potential energy surface (PES) is constructed. The amount of data typically required is of the order of $10^5$ energy calculations at structurally important water configurations. This approach is by far the most accurate, but computationally very expensive, at least in the PES construction phase. Additionally, this approach must utilise a many-body expansion in order to construct the PESs. There exists a surface for the water monomer (3-dimensional), a surface for the two-body, or dimer, interaction energy (12-dimensional), a surface for the three-body, or trimer, interaction energy (21-dimensional) etc. Each additional body increases the dimensionality of the required PES by nine, thus making construction of ever higher body PESs extremely difficult. Despite these difficulties, this approach is able to fully account for quantum mechanical effects in interactions as well as any chemistry that may take place. The approach is, however, only relatively recent and actively and intensively being pursued at present. Indeed, much of the work in this thesis is focused on following this type of approach. Examples of this approach include the HBB models[15], the CC-pol models[16] and the MB-pol models[17].

## 1.4 Towards Modeling Bulk Water from First Principles

As mentioned in the last section, the work presented in this thesis is largely focused upon developing a method to accurately and cheaply obtain the interaction energy between water molecules. Ultimately for the implementation in a simulation that, once and for all, can account for any and all anomalous properties of this

ubiquitous solvent. Inspection of the years published for the references in the last approach reviewed in the previous section reveals that most of the published work has occurred in the last few years, or even as this thesis is being written. It is important to realise that the work presented in this thesis is entirely independent of that work in this highly competitive field of intensive research.

As we saw in section 1.2 many of the anomalous properties of bulk water can be traced back to the hydrogen-bond – a close-contact interaction. We also saw in the previous section that such interactions cannot possibly be described accurately using monopoles of even multipoles and we shall see precisely why that is in the next chapter. To understand water from first principles we are left with no choice but to adopt a model that accurately reflects the quantum mechanics occurring when water molecules are close to one another. The most obvious and simplest way to handle this is break the interactions between water molecules down into their component monomer, dimer and then trimer etc. contributions – presuming, of course, that such contributions converge rapidly. That is, we shall adopt the many-body expansion for evaluating the energy of interacting waters (described in detail in the next chapter). We saw, however, in the last section that constructing PESs for these types of interactions is extremely costly and difficult. Furthermore, it is unnecessary when a multipole approach successfully and accurately describes interaction energies at moderate and long-range. A natural question arises from this statement and that is: “What is moderate-range?” This question is directly addressed in Chapter 4.

The description of interaction energies with multipoles at moderate and long-range is only accurate when induction and dispersion are also included along with the

electrostatistic interactions. Electrostatic interactions are exactly additive and dispersion is very nearly additive to a high degree of accuracy. Unfortunately, the induction interaction is highly non-additive and it makes a significant contribution to the interaction energy in a cluster of water molecules and hence in bulk water. This seriously complicates accurate modeling. For this reason we examine a method that potentially could minimise the computations necessary to evaluate this induction energy through the use of spherical-shells of water molecules around a central water molecule of interest. Our investigations into this method are described in Chapter 3.

Before any considerations of water and the way it interacts with other water molecules we firstly describe the theoretical methods employed in this thesis. This is done in the following chapter. Finally, Chapter 6 represents a significant collaboration on how aryl-substituents moderate the nature of hydrogen bonds, N–H···N versus N–H···O, leading to supramolecular chains in the crystal structures of *N*-arylamino 1,2,3-triazole esters. We were involved in the computational work which helped elucidate the observed diversity in hydrogen bonding found in the eight crystal structures.

# Chapter 2

# Employed Theoretical Methods

## 2.1 The Energy of a Collection of Molecules

The total energy of a system is the sum of all the kinetic and potential energy associated with that system. In thermodynamic terms, the energy we are referring to is the internal energy of the system[c]. The systems of interest in this thesis are collections of water molecules. The studies in this work are focused entirely on the *electronic* energy of these collections of water molecules. The electronic energy is exclusive of any kinetic energy associated with nuclear motion. That is, nuclear translational, vibrational and rotational kinetic energy. Furthermore, we do not concern ourselves with any kinetic energy associated with nuclear or electronic spin as these energies are extremely small in comparison to the energies mentioned thus far. The work presented in this thesis is therefore not concerned with the total energy of a collection of water molecules, but rather the electronic energy of the same.

The main reason for focusing only on the electronic energy is that the kinetic energy of nuclear motion can readily be obtained through simulation, be it either Monte Carlo or molecular dynamics. These simulations potentially provide all the information necessary to compare theoretical calculations to experimental bulk measurements. These simulations require, first and foremost, an electronic energy

[c] Despite the fact that the free energy, be it either Gibbs or Helmholtz, or enthalpy possess the same units as internal energy, i.e., energy, these latter thermodynamic state functions are not generally conserved in any arbitrary thermodynamic process in an isolated system and do not generally represent the total kinetic and potential energy of a system except at zero Kelvin. Likewise, in a general open or closed system only the internal energy change of the Universe is conserved in a process, while that of the Gibbs, Helmholtz and enthalpy are not generally so.

of the system of interest for a given set of nuclear positions or coordinates (i.e., the potential energy surface or PES). Without the latter the simulations cannot be performed. Furthermore without an accurate estimate of the electronic energy of the system with respect to nuclear positions the results of the simulations will also be void of any accuracy.

The previous paragraphs have made several tactic assumptions. Firstly, it was assumed that the electronic energy of the system can be obtained for a given nuclear configuration and/or without specifying the kinetic energy of the nuclei, which surely all possess non-zero momenta at any given instant of time. This assumption necessarily implies that the electronic energy is independent of the kinetic energy of the nuclei. The presumed independence of nuclear and electronic motion is the basis of almost all quantum mechanical methods today. This assumption is known as the Born-Oppenheimer (BO) approximation.

The BO approximation is justified on the basis that electrons have a much smaller mass than that of the nuclei, hence they move much faster relative to the nuclei for a given amount of kinetic energy. For a given amount of kinetic energy in the system, there is a tendency for this energy to be spread out over all the possible degrees of freedom (the equipartition theorem). As a result, the electrons in any given system will adjust their distribution to provide the energetically most favourable one for a given set of nuclear positions. To adopt a term from modern theatrical parlance, the electrons move in “bullet time” compared to the much slower motion of the nuclei and are thus, able to adopt the energetically most favourable distribution accordingly. To a very high level of accuracy the electrons may be regarded as moving around the fixed and static nuclei, whose distances

with respect to one another are fixed, so the kinetic energy of the nuclei can be completely separated out from the problem and dealt with in detail once having solved for the electronic energy as a function of nuclear positions, i.e., having once solved for the PES.

Another tactic assumption in the first two paragraphs of this section, and indeed the title of this section, is that individual water molecules still exist in the condensed phase. While this may be patently obvious to the reader that this should be the case, it is of note that it was not always so. In fact, it was not until the 19$^{th}$ century did scientists generally begin to accept that matter could be composed of individual molecules[18]. Today, there is no doubt that this is true with numerous experimental studies, e.g., diffraction experiments, microscopy and even the existence of simple Brownian motion, demonstrating the fact. Liquid water is no exception; indeed, it would be odd to call the liquid phase of a collection of H and O atoms in a ratio of 2:1 liquid water if this were not the case. Nevertheless, water molecules may approach each other quite closely so that their identity, or individual properties, may begin to be blurred as indicated in Figure 1-3, panel (B).

The existence of individual molecules of water in the liquid phase implies that the electronic energy of a collection of such molecules may be separated into two parts. The first being the electronic energies associated with the individual molecules and the second being an electronic energy associated with the interaction of such molecules. A jargon as evolved by considering clusters of water molecules in such a manner. The term “water monomer” refers to an individual water molecule and “water dimer” refers to a pair of (usually) interacting water

molecules. Likewise, the terms “trimer”, “tetramer”, “pentamer”, “hexamer”, etc. all refer to different cluster sizes of water molecules. Such clusters possess an electronic energy, but it is also possible to define an interaction energy between the individual waters within each cluster, which is the topic of the next section.

## 2.2 The Interaction Energy of a Collection of Molecules

Let the electronic energy of a collection of *N* water molecules be given by $E(\mathbf{X})$, where $\mathbf{X}$ is a 9*N* dimensional vector of Cartesian coordinates of all 3*N* nuclei in the collection. As discussed in the previous section, we can define this as the electronic energy of the system because we are invoking the BO approximation. Furthermore, we have already established that we are able to differentiate the individual water molecules in this collection, so it also possible to define the individual monomer electronic energies of each water in the collection. Let the electronic energy of an individual isolated water monomer, *i*, be $E(\mathbf{X}_i)$ where $\mathbf{X}_i$ is a nine dimensional vector of Cartesian coordinates of the O, H and H atoms in water molecule *i*. Thus it is also possible to define the interaction energy, $\varepsilon(\mathbf{X})$, of this collection of water molecules as

$$\varepsilon(\mathbf{X}) = E(\mathbf{X}) - \sum_{i=1}^{N} E(\mathbf{X_i}) \qquad (1)$$

Equation (1) is the definition of the interaction energy and can be seen to be the additional electronic energy of the system over and above the sum of the electronic energies of the individual monomers. It should also be noted that the structures of each monomer may not be identical, so that in general $E(\mathbf{X}_i) \neq E(\mathbf{X}_j)$.

It is instructive to consider for the moment the physical meaning of the energy obtained from evaluating equation (1). If a collection of water molecules were

behaving as a perfect gas, then for any configuration, $\mathbf{X}$, of the waters, the $\varepsilon(\mathbf{X}) = 0$. Evidently, the very existence of a condensed phase of water means that, at least for certain $\mathbf{X}$, $\varepsilon(\mathbf{X}) < 0$. The interaction energy between water molecules in the liquid phase must be negative otherwise there would be no reason for the water molecules to remain "condensed" – i.e., in the liquid state – and with the addition of the nuclear kinetic energy (always a positive energy) the collection of water molecules would become unbound and evaporate/boil away.

When $N = 2$ the interaction energy given in equation (1) is also known as the "two-body" interaction energy. For $N = 3$, it is still possible to define a two-body interaction energy, but the system will now contain three possible two-body interactions. If we label the monomers 1, 2 and 3, then the three possible two-body interactions are between monomers 1,2; 1,3 and 2,3. These two-body interactions can be represented as $\varepsilon_{i,j}$ where $i \neq j$ and may be defined in a manner identical to equation (1), i.e.,

$$\varepsilon_{i,j}(\mathbf{X}_{i,j}) = E(\mathbf{X}_{i,j}) - E(\mathbf{X}_i) - E(\mathbf{X}_j) \qquad (2)$$

In equation (2), we have explicitly indicated that the two-body interaction energy $\varepsilon_{i,j}$ depends only on the coordinates of molecules *i* and *j*, i.e., $\boldsymbol{X}_{i,j}$. The electronic energy $E(\mathbf{X}_{i,j})$ is the electronic energy of a dimer formed by extracting the coordinates of molecules *i* and *j* from the trimer.

When $N \geq 2$ there will in general be a total of $\binom{N}{2} = \frac{N(N-1)}{2!}$ different two-body interactions. The sum of all the possible two-body interactions in a given collection of *N* monomers is known as the two-body energy of the system, mathematically represented as

$$\varepsilon^{(2)}(\mathbf{X}) = \sum_{i<j}^{N} \varepsilon_{i,j}(\mathbf{X}_{i,j}) \tag{3}$$

The symbol $i < j$ in the above sum actually represents a double summation over indices *i* and *j* but always ensuring that the index *i* is less than *j*. Thus, there will be $\binom{N}{2} = \frac{N(N-1)}{2!}$ terms in the summation.

These considerations can further be extended to include three-, four-, five-, etc. up to *N*-body interactions. This finite series of interactions is known as the *many-body expansion*, as was mentioned in the first chapter, has played an important role in elucidating the significance of various contributions made to the interaction energy of a collection of water molecules. The many-body expansion is utilised in the work presented in Chapter 4 of this thesis, so we describe this expansion is detail in the next section.

## 2.3 The Many-body Expansion

Perhaps a better name for the many-body expansion might be the many-body energy decomposition. When the energy of a system can be considered to be made up of several "bodies" then that energy can always be decomposed into various contributing orders in the many-body expansion. In general, for an *N*-body system one can decompose the energy into a simple sum of 1-body ($\varepsilon^{(1)}(\mathbf{X})$), 2-body ($\varepsilon^{(2)}(\mathbf{X})$), 3-body ($\varepsilon^{(3)}(\mathbf{X})$), … , *N*-body ($\varepsilon^{(N)}(\mathbf{X})$) contributions.

That is

$$E(\mathbf{X}) = \sum_{i=1}^{N} \varepsilon^{(i)}(\mathbf{X}) \tag{4}$$

The interaction energy of a system, as given in equation (1), can be decomposed into the exact same contributions as the energy of the system except the 1-body

term is missing as it has already been removed from the total energy of the system. That is

$$\epsilon(\mathbf{X}) = E(\mathbf{X}) - \sum_{i=1}^{N} E(\mathbf{X}_i) = \sum_{i=1}^{N} \varepsilon^{(i)}(\mathbf{X}) - \varepsilon^{(1)}(\mathbf{X}) = \sum_{i=2}^{N} \varepsilon^{(i)}(\mathbf{X}) \qquad (5)$$

Here we see that the sum of all the isolated monomer energies is merely the 1-body energy of the system, or $\varepsilon^{(1)}(\mathbf{X}) = \sum_{i=1}^{N} E(\boldsymbol{X}_i)$.

It is important to realise that the 2-body energy (or the *n*-body energy for that matter), is not the same as the energy of 2 bodies (or the energy of *n*-bodies). For example, the energy of *N* interacting bodies is $E(\mathbf{X})$. When $N = 2$, $E(\mathbf{X})$ is the energy of the two bodies and it is given by equation (4) which is not equal to $\varepsilon^{(2)}(\mathbf{X})$ – the 2-body interaction energy – which in the case of $N = 2$ is simply given by equation (1). That is, $E(\mathbf{X})$ – the energy of the two bodies – has had removed from it $\varepsilon^{(1)}(\mathbf{X})$ – the sum of all the energies of the single bodies that make up the pair – to produce $\varepsilon^{(2)}(\mathbf{X})$ – the 2-body energy (see equation (7)).

Higher body interaction energies can be defined in an analogous manner to the 2-body interaction energy. If we consider $N = 3$, $E(\mathbf{X})$ will be the energy of the three bodies. We can remove from it $\varepsilon^{(1)}(\mathbf{X})$ – the sum of all the energies of the single bodies that make up the triple *and* the sum of all the 2-body energies that make up the triple (i.e., $\varepsilon^{(2)}(\mathbf{X})$) to yield $\varepsilon^{(3)}(\mathbf{X})$ – the 3-body energy (see equation (8)). Likewise when $N = 4$ $E(\mathbf{X})$ is the energy of the four bodies. Removing from it $\varepsilon^{(1)}(\mathbf{X})$, $\varepsilon^{(2)}(\mathbf{X})$ and $\varepsilon^{(3)}(\mathbf{X})$ produces $\varepsilon^{(4)}(\mathbf{X})$ – the 4-body energy (see equation (9)). We can therefore generally write the following expressions for *n*-body energies in a system composed of *N*-bodies as:

One-body

$$\epsilon_i(\mathbf{X}_i) = E(\mathbf{X}_i) \tag{6}$$

Two-body

$$\epsilon_{i,j}(\mathbf{X}_{i,j}) = E(\mathbf{X}_{i,j}) - \sum_{s\in\{i,j\}} \epsilon_s(\mathbf{X}_s) \tag{7}$$

Three-body

$$\epsilon_{i,j,k}(\mathbf{X}_{i,j,k}) = E(\mathbf{X}_{i,j,k}) - \sum_{s\in\{i,j,k\}} \epsilon_s(\mathbf{X}_s) - \sum_{\substack{s\in\{i,j,k\}< \\ r\in\{i,j,k\}}} \epsilon_{s,r}(\mathbf{X}_{s,r}) \tag{8}$$

Four-body

$$\epsilon_{i,j,k,l}(\mathbf{X}_{i,j,k,l}) = E(\mathbf{X}_{i,j,k,l}) - \sum_{s\in\{i,j,k,l\}} \epsilon_s(\mathbf{X}_s) - \sum_{\substack{s\in\{i,j,k,l\}< \\ r\in\{i,j,k,l\}}} \epsilon_{s,r}(\mathbf{X}_{s,r}) - \sum_{\substack{s\in\{i,j,k,l\}< \\ r\in\{i,j,k,l\}< \\ t\epsilon\{i,j,k,l\}}} \epsilon_{s,r,t}(\mathbf{X}_{s,r,t}) \tag{9}$$

etc., and continuing recursively all the way up to the *N*-body energy. In the above expressions the indices *i*, *j*, *k* and *l* are unique labels for the bodies (water molecules) selected from the set of *N* bodies (waters) under consideration.

Equation (4) is the many-body decomposition of the total energy of a collection of *N* water molecules. Equations (6) through (9) represent individual one through four-body contributions that are made to the respective $\varepsilon^{(1)}(\mathbf{X})$ through $\varepsilon^{(4)}(\mathbf{X})$ found in equation (4). For a system composed of *N* waters there will be a total of $\binom{N}{1} = N$ 1-body energies of the kind represented by equation (6). The will also be a total of $\binom{N}{2} = \frac{N(N-1)}{2!}$ 2-body energies of the kind represented by equation (7). Similarly there will be $\binom{N}{3} = \frac{N(N-1)(N-2)}{3!}$ and $\binom{N}{4} = \frac{N(N-1)(N-2)(N-3)}{4!}$

individual 3- and 4-body energies represented by equations (8) and (9) respectively. Generally we have

$$\epsilon^{(n)}(\mathbf{X}) = \sum_{\alpha}^{\binom{N}{n}} \epsilon_{\alpha}(\mathbf{X}_{\alpha}) \tag{10}$$

Here $\alpha$ represents a unique collection (a set) of $n$ water monomers – there will be a total of $\binom{N}{n}$ different unique ways of selecting $n$ waters from the original cluster of size $N$. We can see that the number of individual interaction energy contributions grow very rapidly as roughly $N^n$ for $n \ll N$.

An ansatz implicit in the use of the many-body expansion is that the series, equation (4), converges rapidly by $n$ = 3 or 4. For many non-polar systems, the series is satisfactorily convergent by $n$ = 2, i.e., only two-body interactions need to be evaluated in order to accurately estimate the overall interaction energy of the system, equation (1). Put another way, for non-polar systems a simple pair-wise sum of individual interaction energies between monomers in a cluster of $N$ molecules works extremely well.

Unfortunately, water is polar and possesses significant short-range highly anisotropic interactions (H-bonding). These properties of water require that the 3-body and even 4-body interactions be included in the estimate of the interaction energy in order to obtain accurate results. As just seen, the number of such interactions scales as $N^3$ and $N^4$ respectively, where $N$ is the number of waters under consideration. This poor scaling is further exacerbated by the fact that if one is interested in accurately simulating bulk water a reasonable size volume of water molecules needs to be included in the simulation. For the purposes of illustration, if we were considering a spherical volume of water of radius, say, 9 Å then we

would only encompass about three hydration shells. For simulating bulk water, a reasonable number of water molecules is required and three hydration shells are probably not sufficient to accurately quantify all of the anomalous properties of water. Nevertheless, there is about 100 water molecules in such a volume. The number of three- and four-body interactions for this system are 161,700 and 3,921,225 respectively. Perhaps a more appropriate size system would be a sphere of twice the radius. The doubling of the radius increases the volume eight-fold and so is the number of molecules. Now the numbers of three- and four-body interactions for this system are about 85 million and 17 billion respectively.

To place these numbers in context, let us assume that a single four-body energy evaluation could be somehow reduced to 1,000 floating point operations (FLOP). If the above sample of 800 water molecules was utilised in a Monte Carlo simulation, at least 200 million evaluations of all of the above 17 billion four-body interactions would be needed to obtain reasonable statistics from the simulation to compare with experiment. This computation would require at least $3.4 \times 10^{21}$ FLOP. The fastest supercomputer on the planet in 2014 was the Chinese Tianhe-2 demonstrably capable of performing 33.86 P FLOP per second, or FLOPS (P = peta or $10^{15}$). Even on this machine the simulation would require 10 weeks of continuous execution. By comparison, an office desktop PC[d] would require 160,000 years to perform the same simulation.

At this stage it may appear to be madness to pursue a many-body approach in order to approximate the interaction energy of a large enough collection of water molecules so that meaningful bulk water simulations may be performed. Madness

[d] A current typical desktop PC, say the Intel Core i7 3770K @ 3.40 GHz runs at around 41 GFLOPS.

given the sheer number of three- and four-body interactions required, but is it true that so many interaction energies need to be evaluated? This question is addressed in Chapter 4. If pursuing the many-body expansion is madness then what is the alternative? Direct computation of $E(\mathbf{X})$ for $N = 800$? As we shall see in the next section, it is currently not possible to perform highly accurate calculations on such a large system; furthermore this situation will not likely change any time soon.

## 2.4 Calculation of the Electronic Energy

Much of the work presented in this section is a summary of relevant theoretical methods taken from the text "Introduction to Computational Chemistry" by F. Jensen[19]. These methods were employed during the course of my PhD candidature.

### 2.4.1 *Ab Initio*

So far in this chapter we have presumed a method exists to compute $E(\mathbf{X})$. Recall that we are ultimately interested in explaining the anomalous properties of bulk water completely from first principles. The only means by which this can be accomplished is through solution of the stationary state Schrödinger equation.

$$\hat{H}\Psi = E(\mathbf{X})\Psi \tag{11}$$

As already mentioned, we shall adopt the Born-Oppenheimer approximation and neglect any kinetic energy associated with electron and nuclear spin. We seek to solve this equation for the ground-state wavefunction, $\Psi$, for a specified nuclear configuration $\mathbf{X}$. The Hamiltonian operator, $\hat{H}$ can be readily written down exactly and operates directly on $\Psi$ – a mathematical function of the coordinates of the electrons present in the system.

The Hamiltonian operator itself is a differential operator and is the sum of kinetic and potential energy operators.

$$\hat{H} = \hat{T} + \hat{V} \tag{12}$$

The kinetic energy operator $\hat{T}$ is a sum of differential operators:

$$\hat{T} = -\frac{1}{2}\sum_{i}^{n}\left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\right) \tag{13}$$

Note that we have adopted atomic units in this section, whereby the mass of the electron, the electronic charge, $\hbar$ and $(4\pi\epsilon_0)^{-1}$ are all set to unity. The sum in equation (13) is over the *n* electrons present in the system.

The potential energy operator $\hat{V}$ is the coulomb interaction:

$$\hat{V} = \sum_{a<b}^{N}\frac{Z_a Z_b}{|\mathbf{R}_a - \mathbf{R}_b|} + \sum_{i<j}^{n}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{a=1}^{N}\sum_{i=1}^{n}\frac{Z_i}{|\mathbf{R}_a - \mathbf{r}_i|} \tag{14}$$

Where, *N*, now represents the number of nuclei in the system and *n* the number of electrons. The first summation in equation (14) is just a constant and represents the nuclear-nuclear repulsion energy – nothing in this term operates on Ψ. The second summation operates on the electron coordinates of electrons *i* and *j* and represents the electron-electron repulsion energy. The final summation is the nuclear-electron attraction occurring between nucleus *a* and electron *i*. $\mathbf{R}_a$ is the position vector of nucleus *a* and $\mathbf{r}_i$ is the position vector of electron *i*.

### 2.4.2 Obtaining an Approximate Wavefunction

Having specified the equation we wish to solve (11) and the Hamiltonian (12) – (14) we require a means of obtaining Ψ for the ground-state and hence $E(\mathbf{X})$ for a predefined configuration of nuclei **X**. In this thesis, we almost exclusively have

utilised *ab initio* molecular orbital theory to achieve this, so a brief description follows of the employed *ab initio* methods.

The first step in solving equation (11) is to admit that we do not know Ψ at the outset. Without knowledge of this function, the only way forward is to make a guess. The function we shall choose should be as mathematically convenient as possible yet represent as accurately as possible the true and unknown, Ψ. It would be best to use a function that can be modified so that it evolves towards the true Ψ starting from our initial crude guess at it, i.e., $\psi$.[e] However, since we do not know what Ψ is, it does not seem possible to alter the function in any sensible way such that it becomes a better approximation to the true Ψ. Fortunately, a theorem exists that assists us at improving the guessed function $\psi$, and that theorem is the variational theorem.

The variational theorem states that for a time-independent Hamiltonian operator, any trial wavefunction will have an energy expectation value that is greater than or equal to the true ground state energy corresponding to the true wavefunction of the given Hamiltonian. The meaning of "expectation value" is the result of the integration

$$\int \psi^* \hat{H} \psi d\tau = \mathfrak{E}(\mathbf{X}) \qquad (15)$$

where the integral is multidimensional and over all of the coordinates in $\psi$. Additionally we have assumed that $\psi$ is normalised. Note in equation (15) we have temporarily made a distinction between the exact ground state electronic energy, $E(\mathbf{X})$, and that obtained here from the approximate wavefunction, $\mathfrak{E}(\mathbf{X})$. The

[e] We shall use the notation, Ψ, for the true wavefunction, and $\psi$ for its guess, or approximation.

variational theorem states that $E(\mathbf{X}) \leq \mathfrak{E}(\mathbf{X})$. By altering the estimated wavefunction in such a manner so that the expectation value $\mathfrak{E}(\mathbf{X})$ is minimised, we ensure that our estimated wavefunction $\psi$ is the best possible function it can be within the constraints and/or approximations made to obtain it.

Armed with the knowledge that we need to perform the multidimensional integral (15), we can now begin to construct a suitably flexible and mathematically convenient function, $\psi$. Note that $\Psi$ is a function of all the electronic coordinates in our system. For $n$ electrons it is a $3n$ dimensional function. Such a highly dimensional mathematical function seems hopelessly complex to even estimate. Clues as to possible simplifications we may make to this function come from examining the Hamiltonian. Equation (13), the kinetic energy operator, shows that it is a simple sum of $n$ 3-dimensional operators, each one operating only upon the coordinates of a single electron at a time. The potential energy operator is almost the same. The nuclear-nuclear repulsion sum is a constant (for a specific $\mathbf{X}$) and does not operate on the coordinates of any electron. The nuclear-electron attraction operator, like the electron kinetic energy operator, only operates on the coordinates of a single electron at a time. The electron-electron repulsion summation is problematic. Each term in its sum is a 6-dimensional operator and operates on the coordinates of pairs of electrons simultaneously via the distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. Therefore, the form of the Hamiltonian suggests that we may begin by approximating the general $3n$ dimensional $\Psi$ with a function that is the product of $n$ three-dimensional functions – that is by applying an independent particle model.

Although such a model is still $3n$ dimensional, the fact that it is made up of a product of $n$ three-dimensional functions is certainly mathematically convenient

and much simpler than trying to produce a suitably flexible $3n$ dimensional function that incorporates all sorts of coupling terms between the electrons. For example, our trial function will not contain any term like $r_{ij}$, which intimately ties together the coordinates of electrons *i* and *j* simultaneously. However, the serious drawback in using the independent particle model is that we know, at least through the electron-electron repulsion operator, that the motion of pairs of electrons must be correlated. An independent particle model ignores this correlation. Ultimately we need to start somewhere, so we shall utilise this model then correct latter for the, now in-built, error associated with its application. Due to the use of this model we know that our $\psi$ can never be the same as $\Psi$, so that our energy $\mathfrak{E}(\mathbf{X})$ will definitely be larger than $E(\mathbf{X})$, at least until we correct for its application.

At this stage a further complication arises. We cannot simply write $\psi$ as a product of one-electron functions, $\phi_i$ – in doing so we break a law of nature. Electrons are fermions. Fermions have the property that any wavefunction that describes them must change its sign if the coordinates of any pair of them are permuted. Any trial wavefunction that is just a simple product of one electron functions will not possess this fundamental property. It is interesting that even though our Hamiltonian does not contain any terms involving spin of the electrons (nor nuclei), our wavefunction cannot be missing a mathematical function that includes these coordinates. Thus, our trial function now requires the inclusion of electron spin functions. Because the Hamiltonian does not contain any spin terms, we can simply write each of our single electron functions as a product of a spin function $\gamma_i$, a 1-dimensional function, and a 3-dimensional spatial function, $\phi_i$. The new function is called a spin orbital, or spinor, $\chi_i = \phi_i\gamma_i$. To be clear, note that here the subscript *i* is not referring to electron *i*, but mathematical function *i*.

There are only two types of electron spin functions. The possible functions that $\gamma$ may be for any electron are often labeled as $\alpha$ and $\beta$. These functions are orthonormal, i.e.,

$$\int \alpha^* \beta d\tau = \int \beta^* \alpha d\tau = 0 \tag{16}$$

where the integration is taken over the unspecified spin coordinates, and

$$\int \alpha^* \alpha d\tau = \int \beta^* \beta d\tau = 1 \tag{17}$$

Maintaining mathematical convenience we also desire that all of the $\phi_i$ be orthonormal, i.e.,

$$\int \phi_i{}^* \phi_j d\tau = \int \phi_j{}^* \phi_i d\tau = 0 \tag{18}$$

where the three-dimensional integration is taken over all of *xyz* space, and

$$\int \phi_i{}^* \phi_i d\tau = \int \phi_j{}^* \phi_j d\tau = 1 \tag{19}$$

Constructing the simplest possible $\psi$ now from the spinors that satisfies the natural law for fermions is through the use of a determinant.

$$\psi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(1)\chi_2(1) & \cdots & \chi_n(1) \\ \chi_1(2)\chi_2(2) & \ldots & \chi_n(2) \\ \vdots & \ddots & \vdots \\ \chi_1(n)\chi_2(n) & \cdots & \chi_n(n) \end{vmatrix} \tag{20}$$

In quantum chemistry, this determinant is named the Slater determinant. The numbers in the parenthesis represent the coordinates of enumerated electrons. The subscript to the spinor represents that particular mathematical 4-dimensional function. A system consisting of *n* electrons will have *n* different spinors (the *n* columns in the determinant). It should be noted from the Slater determinant that every electron in the system is placed into every possible spinor (the *n* rows in the determinant), thus truly making the electrons indistinguishable. The factor of $(n!)^{-\frac{1}{2}}$ ensures normalization of the $\psi$ because an $n \times n$ determinant produces

every possible occupancy of electrons amongst the $n$ $\chi_i$, whereby there are $n!$ of these products. Each of the $n!$ products of spinors represents a different possible occupancy of electrons amongst the $n$ spinors. The name given to such a product of spinors is called a Hartree product. Permuting the coordinates of any two electrons has the effect of swapping a pair of rows in the determinant. If two rows are swapped in a determinant its value changes sign – consistent with the natural law for fermions. Furthermore, the Slater determinant obeys the Pauli Exclusion Principle. This principle requires that no two electrons can be described with the exact same function. The principle is a consequence of the previously mentioned natural law for fermions. If two electrons were placed in the same orbital then this would lead to a determinant with two identical columns. Any determinant that possess two or more identical columns vanishes.

It may seem nuts to write the trial function, (20), as a linear combination of $n!$ Hartree products. However, because we have ensured that the spinors, $\chi_i$, are all orthonormal when it comes time to perform the integration (15) (discussed in the next subsection) vast numbers of integrals will vanish. We are also only concerned with closed-shell systems in this thesis, so a further simplification can be made to the Slater determinant. We may reuse each spatial function $\phi_i$ once, provided it is multiplied by a different spin function which results in a different spinor. Thus, we may have $\chi_1 = \phi_1\alpha$ and $\chi_2 = \phi_2\beta = \phi_1\beta$. Here an electron may occupy spinor 1 and another electron may occupy spinor 2 without violating the Pauli Exclusion Principle. By reusing each spatial function we are in effect doubly occupying each spatial orbital, but not spinor as this is strictly forbidden. Thus, we only require $\frac{n}{2}$ different spatial functions in our $\psi$.

### 2.4.3 The Electronic Energy of a Slater Determinant

We are now in a position to obtain an expression for $\mathfrak{E}(\mathbf{X})$ through application of equation (15) even though we have not yet specified exactly the form of the functions $\phi_i$. To simplify notation we shall write the following:

$$\hat{V}_{NN} = \sum_{a<b}^{N} \frac{Z_a Z_b}{|\mathbf{R}_a - \mathbf{R}_b|} \tag{21}$$

$$\hat{h}_i = \hat{T}_i - \sum_{a=1}^{N} \frac{Z_a}{|\mathbf{R}_a - \mathbf{r}_i|} \tag{22}$$

$$\hat{g}_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{23}$$

where the subscripts now refer to electrons. Equation (21) is the nuclear-nuclear repulsion operator and does not operate on any electron coordinates. Equation (22) involves only operators that operate on the coordinates of a single electron. Equation (23) operates on the coordinates of two electrons simultaneously. The Hamiltonian operator in equation (12) now becomes

$$\hat{H} = \sum_{i=1}^{n} \hat{h}_i + \sum_{i<j}^{n} \hat{g}_{ij} + \hat{V}_{NN} \tag{24}$$

Additionally we shall use the "bra" and "ket" notation to represent integrals, e.g.

$$\langle \chi_1(1)\chi_2(2)|\hat{g}_{12}|\chi_1(1)\chi_2(2)\rangle = \int \chi_1{}^*(1)\chi_2{}^*(2) \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \chi_1(1)\chi_2(2) d\tau \tag{25}$$

Here electron 1 occupies spinor 1 and electron 2 occupies spinor 2.

Substituting equation (20) into equation (15) it is possible to show that

$$\mathfrak{E}(\mathbf{X}) = \sum_{i=1}^{n} h_i + \sum_{i<j}^{n} \left(J_{ij} - K_{ij}\right) + V_{NN} \tag{26}$$

where $V_{NN}$ is obtained by simply evaluating equation (21) because

$$\langle\psi|\hat{V}_{NN}|\psi\rangle = \langle\psi|\psi\rangle V_{NN} = V_{NN} \tag{27}$$

The remaining terms in equation (26) are

$$h_i = \langle\chi_i(i)|\hat{h}_i|\chi_i(i)\rangle \tag{28}$$

$$J_{ij} = \langle\chi_i(i)\chi_j(j)|\hat{g}_{ij}|\chi_i(i)\chi_j(j)\rangle \tag{29}$$

$$K_{ij} = \langle\chi_i(i)\chi_j(j)|\hat{g}_{ij}|\chi_j(i)\chi_i(j)\rangle \tag{30}$$

The actual identity of the specific electrons in equations (28) – (30) is arbitrary, i.e., electron *i* could be any of the *n* electrons, similarly for electron *j*. The electrons need only be different electrons in equations (29) and (30).

The one-electron integrals, $h_i$, represents the electron-nuclear attraction to all nuclei plus the kinetic energy of an electron associated with an electron being in orbital *i*. The two-electron integral, $J_{ij}$ named the Coulomb integral, is the electron-electron repulsion associated with an electron being in orbital *i* and an electron in orbital *j*. The two-electron integral, $K_{ij}$ named the exchange integral, has no direct classical analogue. However, it is exactly zero when the electron spins are different between spinors *i* and *j*. For example,

$$\begin{aligned}&\langle\phi_i(i)\alpha_i(i)\phi_j(j)\beta_j(j)|\hat{g}_{ij}|\phi_j(i)\beta_j(i)\phi_i(j)\alpha_i(j)\rangle \\ &= \langle\phi_i(i)\phi_j(j)|\hat{g}_{ij}|\phi_j(i)\phi_i(j)\rangle\langle\alpha_i(i)|\beta_i(i)\rangle\langle\beta_j(j)|\alpha_i(j)\rangle = 0\end{aligned} \tag{31}$$

The spin functions can readily be factored out of the integral because $\hat{g}_{ij}$ does not operate on spin coordinates, but only on the spatial coordinates of electrons *i* and *j*. Equation (31) is zero because the spin functions are orthogonal (see equations (18) and (19)).

Examination of equation (26) shows that $K_{ij}$ subtracts an energy from the Coulomb repulsion energy. This subtraction only occurs when the two spinors possess the same spin. A parallel spinning electron pair have their Coulomb

repulsion energy reduced. This is a direct consequence of the fact that electrons are fermions. Parallel spinning electrons are naturally kept further away from each other compared with antiparallel spinning electron pairs, so parallel spinning pairs of electrons have their repulsion energy reduced accordingly. This $K_{ij}$ term would not be present if the wavefunction was written as a single simple Hartree product and the corresponding expectation value of the energy would be considerably higher. This is because it represents a poorer approximation to the true wavefunction compared to the Slater determinant. While the use of the independent particle model in the Slater determinant means that no spatial electron correlation is included in its corresponding expectation value of the energy, the Slater determinant *does* include spin correlation which manifests in the appearance of $K_{ij}$.

### 2.4.4 Improving the Trial Wavefunction and the Hartree-Fock Energy

The energy, $\mathfrak{E}(\mathbf{X})$, obtained from equation (26) is not the best possible energy estimate we can obtain for $E(\mathbf{X})$ using an independent particle model. The spinors, also known as molecular orbitals and henceforth we shall refer to them as such, have not yet been optimised in any way because we have not applied the variational theorem. To apply the variational theorem we need to alter our molecular orbitals in such a manner as to minimise the resulting $\mathfrak{E}(\mathbf{X})$. Unfortunately, we cannot alter the molecular orbitals arbitrarily. This is because as they are changed to lower the expectation value of the energy, they must remain orthonormal. If orthogonality was not maintained equation (26) would no longer be valid as it was derived under this condition. Thus, we are required to conduct our minimization of $\mathfrak{E}(\mathbf{X})$ with respect to changing the molecular orbitals under the constraints of orthonormality of the same.

Usually minimization of a function is done with respect to changing some variables, whether it be under constraints or otherwise. However, in this problem, we are minimizing a function with respect to changing other functions – a trickier proposition (especially since we have not specified in any way as yet the form of those functions!). Nevertheless, a more general version of Lagrange's undetermined multipliers can be brought to bear on the problem. In order to do so, we rewrite the energy expression (26) in terms of two new operators

$$\mathfrak{E}(\mathbf{X}) = \sum_{i=1}^{n} \langle \chi_i(i) | \hat{h}_i | \chi_i(i) \rangle + \sum_{i<j}^{n} \langle \chi_j(j) | \hat{J}_i - \hat{K}_i | \chi_j(j) \rangle + V_{NN} \qquad (32)$$

where

$$\hat{J}_i | \chi_j(j) \rangle = \langle \chi_i(i) | \hat{g}_{ij} | \chi_i(i) \rangle | \chi_j(j) \rangle \qquad (33)$$

$$\hat{K}_i | \chi_j(j) \rangle = \langle \chi_i(i) | \hat{g}_{ij} | \chi_j(i) \rangle | \chi_i(j) \rangle \qquad (34)$$

In equation (33), the $\hat{J}_i$ operator is a "multiply by $\langle \chi_i(i) | \hat{g}_{ij} | \chi_i(i) \rangle$" operator. The integral explicitly involves performing $\left\langle \chi_i(i) \left| \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right| \chi_i(i) \right\rangle$. That is, after the integration we are left with a function of $\frac{1}{\mathbf{r}_j}$ because we have integrated over the coordinates of electron *i* keeping $\mathbf{r}_j$ fixed as we do so. We are perfectly permitted to do this because each of the molecular orbitals depend only upon the coordinates of a single electron – which is our independent particle model we have assumed. The function we are left with when we perform the integration $\langle \chi_i(i) | \hat{g}_{ij} | \chi_i(i) \rangle$ is the average field located at the point $\mathbf{r}_j$ due to an electron in molecular orbital *i*. A second integration $\langle \chi_j(j) | \hat{J}_i | \chi_j(j) \rangle$ is just $J_{ij}$ as before and is the Coulomb electron-electron repulsion felt by an electron in molecular orbital *j* due to the *average* or *mean* field of another electron in molecular orbital *i*. This is why

sometimes the independent particle model applied here is also referred to as the mean-field approximation.

Careful inspection of equation (34) shows that $\hat{K}_i$ is more than just a multiplicative operator. Application of this operator onto some molecular orbital, $\chi_j(j)$ say, ends up swapping, or exchanging, this molecular orbital into the "multiply by" integral and changing the $\chi_j(j)$ into the $\chi_i(j)$. As before (cf. equation (31)), if the spins in the two orbitals, $\chi_i$ and $\chi_j$ are different then when $\hat{K}_i$ operates on $\chi_j$ it will be annihilated.

We are yet to perform the constrained minimization, but after working through the required algebra it is found that the following equation must be satisfied in order to obtain the best possible set of molecular orbitals.

$$\hat{F}_i \chi'_i = \sum_{j=1}^{n} \epsilon_{ij} \chi'_j \tag{35}$$

where the $\epsilon_{ij}$ are the Lagrange's undetermined multipliers which are simple constants that here have the units of energy. The $\chi'_i$ are now the best possible molecular orbitals (within the independent particle model), i.e., they are no longer the same molecular orbitals that we started with, $\chi_i$, because these are the functions that were altered in order to perform the energy minimization. The $\hat{F}_i$ are called Fock operators given by

$$\hat{F}_i = \hat{h}_i + \sum_{j=1}^{n} \left( \hat{J}_i - \hat{K}_\iota \right) \tag{36}$$

Without performing the optimization of the orbitals, then equation (35) would not be satisfied. In this case, application of $\hat{F}_i$ on $\chi_i$ would not give a simple linear combination of $\chi_j$ as shown in equation (35), but a rather complicated mess. The

set of *n* equations given in (35) are called the Hartree-Fock equations. The Hartree-Fock equations can be further simplified by diagonalizing the $n \times n$ matrix of $\epsilon_{ij}$ to yield

$$\hat{F}_i \chi''_i = \epsilon_i \chi''_i \tag{37}$$

The $\epsilon_i$ are identified as the molecular orbital energies. The $\chi''_i$, which are simple linear combinations of the $\chi'_i$, are called the canonical molecular orbitals.

There is one complication that exists before obtaining the best possible energy from our $\psi$, and that is the fact that in order to solve equation (37) we need the Fock operator. The Fock operator, (36), includes the Coulomb and exchange operators, (33) and (34), but these operators require knowledge of the canonical molecular orbitals in the first place! The only way to solve this is to do it self-consistently. That is, first guess the molecular orbitals and use them to obtain the $\hat{J}_i$ and $\hat{K}_i$. Now solve for the canonical molecular orbitals by satisfying the Hartree-Fock equations (37). The new set of molecular orbitals just obtained can now be used to compute a better set of $\hat{J}_i$ and $\hat{K}_i$ operators. Now again solve the Hartree-Fock equations to obtain an even better set of molecular orbitals. The process continues until the molecular orbitals change no further, in which case self-consistency has been reached and the final total energy truly is as low as it can be. The final electronic energy can be obtained from either equation (26), but using the final set of canonical molecular orbitals, or utilizing the molecular orbital energies, thus

$$\mathfrak{E}(\mathbf{X}) = \sum_{i=1}^{n} \epsilon_i - \sum_{i<j}^{n} \left( J_{ij} - K_{ij} \right) + V_{NN} \tag{38}$$

Note that the total energy is not a simple sum of molecular orbital energies (the first summation). The Fock operator containing $\hat{J}_i$ and $\hat{K}_i$ describes the repulsion

to all other electrons, so the sum over all the molecular orbital energies therefore counts the electron-electron repulsion twice, which must be correct for.

Of course, all of the above presumes an ability to actually perform integration, but how can integration be performed if we do not actually have a mathematical form for the $\chi_i$? As yet we have not specified any form for these orbitals, yet it is actually possible to compute all of the required integrals. For very small highly symmetric systems, e.g. atoms and diatoms, the Hartree-Fock equations can be solved numerically[20]. This is done by mapping the molecular orbitals onto a set of grid points. The calculation is very expensive, but essentially yields the Hartree-Fock energy, (38), also described as the Hartree-Fock limit. The use of the latter term will become clear in the next section. Thus, a numerical solution to the Hartree-Fock equations truly yields (for a fine enough grid) the lowest possible energy attainable within the independent particle model. In the next section, we shall see how to specify an actual form for the molecular orbitals so that much larger systems can be studied and all the necessary integrals can be computed rapidly and efficiently. Unfortunately, the price for doing so requires us to introduce yet another approximation.

### 2.4.5 The Basis Set Approximation and the Roothaan-Hall Equations

In order to create functions that may be changed so that the electronic energy of the system can be obtained variationally, a basis set approximation is adopted. This approximation expresses unknown molecular orbitals in terms of a set of known functions. If the basis functions employed formed a complete set of functions there would be no approximation because if the set was complete then any function could be described in terms of them. Unfortunately, for the problem

at hand, a complete set of basis functions requires an infinite number of them. Clearly we shall have to make do with something smaller and most definitely finite. Because we shall employ a finite set of basis functions we are no longer able to solve exactly the Hartree-Fock equations (37). Any electronic energy we now obtain variationally possess sources of error due to the independent particle model *and* the use of a finite set of basis functions.

Of course, it is possible to use a larger and larger set of basis functions, and because of the variational theorem we know that as we do so, our computed electronic energy will approach (from above) more and more closely the true energy of the system. However, we also know that we shall never be able to get any lower in energy than the Hartree-Fock energy because the entire method we are currently employing is based soundly on the independent particle model. Thus, the best (lowest) electronic energy we can obtain variationally using basis sets is approached asymptotically with the number of basis functions used (i.e, size of basis set) and is the Hartree-Fock energy. This energy is most often referred to as the Hartree-Fock limit for the above reasons.

The basis function approximation to a molecular orbital is expressed mathematically in terms of the spatial part of the molecular orbital, $\chi_i$, as

$$\phi_i = \sum_{\beta=1}^{m} c_{\beta i} b_\beta \tag{39}$$

where $b_\beta$ is the $\beta^{\text{th}}$ basis function[f] in the basis set of size $m$. $m \geq \frac{n}{2}$ (closed-shell systems reuse the same spatial molecular orbital as it is doubly occupied with opposite spin electrons), but is usually very much larger than $\frac{n}{2}$. Individual $\chi_i$ are

[f] Note that $\beta$ here is just an index and not a spin function!

trivially generated from (39) by simply multiplying by the appropriate spin function, $\chi_i = \phi_i \gamma_i$.

Any functions may be used as basis functions, but those that best reflect physical reality are expected to perform best, i.e., require fewer functions to obtain the same $\mathfrak{E}(\mathbf{X})$. The functions should decay to zero at long distances from a nucleus in the system and, if the solutions to the H atom are any guide, should also possess a cusp at a nucleus. However, the functions should also be as simple as possible and be computationally efficient to integrate. While simple exponentials placed around each nucleus model the physical reality well, they are computationally expensive to integrate when integrals involve several nuclei simultaneously. In contrast, gaussian functions, while not possessing a cusp at the nucleus like exponentials do, are much simpler to integrate. The other issue with using gaussians as basis functions is that they head much more rapidly towards zero than exponential functions do because gaussians have in their exponents $|\mathbf{R}_a - \mathbf{r}_i|^2$, rather than just $|\mathbf{R}_a - \mathbf{r}_i|$. The effect of the latter is to underestimate electron density at distances far from the nucleus. These latter effects can be accounted for and shall be discussed in later chapters where specific basis sets are employed.

If we now substitute equation (39) into equation (37) (factorizing out the spin functions) we obtain

$$\hat{F}_i \sum_{\beta=1}^{m} c_{\beta i} b_\beta = \epsilon_i \sum_{\beta=1}^{m} c_{\beta i} b_\beta \qquad (40)$$

At this stage we note that equation (40) is only approximately true because our basis set is incomplete. The equations given by (40) are called the Roothaan-Hall

equations for a closed-shell systems[g]. They are the Hartree-Fock equations after application of the basis set approximation. To proceed further we premultiply both sides of the equation by a specific basis function, $b_\alpha$ and integrate.

$$\sum_{\beta=1}^{m} c_{\beta i}\langle b_\alpha|\hat{F}_i|b_\beta\rangle = \epsilon_i \sum_{\beta=1}^{m} c_{\beta i}\langle b_\alpha|b_\beta\rangle \tag{41}$$

Equation (41) can conveniently be written in matrix notation

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{42}$$

where $F_{\alpha\beta} = \langle b_\alpha|\hat{F}_i|b_\beta\rangle$ is called the Fock matrix and formally of dimension $m \times m$, $C_{\beta i} = c_{\beta i}$ is the molecular orbital coefficient matrix and formally of dimension $m \times n$, $S_{\alpha\beta} = \langle b_\alpha|b_\beta\rangle$ is the overlap matrix and formally of dimension $m \times m$, and finally $\epsilon_{ij} = \delta_{ij}\epsilon_i$ is a diagonal matrix of molecular orbital energies formally of dimension $n \times n$. Note that $S_{\alpha\beta} \neq \delta_{\alpha\beta}$, unlike the molecular orbitals which *must* remain orthonormal.

We desire to solve equation (42) by standard techniques, i.e., solve for the $\mathbf{C}$ matrix such that $\boldsymbol{\epsilon}$ is diagonal. This may be accomplished by premultiplying both sides of equation (42) by $\mathbf{S}^{-\frac{1}{2}}$, and inserting $\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{\frac{1}{2}} = \mathbf{I}$ (as $\mathbf{I}$ is the identity matrix) between $\mathbf{F}$ and $\mathbf{C}$.

$$\left(\mathbf{S}^{-\frac{1}{2}}\mathbf{F}\mathbf{S}^{-\frac{1}{2}}\right)\left(\mathbf{S}^{\frac{1}{2}}\mathbf{C}\right) = \left(\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}\right)\left(\mathbf{S}^{\frac{1}{2}}\mathbf{C}\right)\boldsymbol{\epsilon} \tag{43}$$

to obtain

$$\mathbf{F}'\mathbf{C}' = \mathbf{C}'\boldsymbol{\epsilon} \tag{44}$$

[g] Since every spatial molecular orbital is doubly occupied it is most efficient to solve these equations for either just the $\alpha$ or the $\beta$ spinning electrons, as the solution to both sets of molecular orbitals will be identical. The Fock operator only needs to be slightly modified to take this simplification into account by multiplying the Coulomb operator in equation (36) by 2 and summing over $\frac{n}{2}$ (only $\frac{n}{2}$ unique spatial molecular orbitals are occupied by the *n* electrons) instead of *n*. For the purposes of brevity and clarity the discussion here ignores this simplification.

The meaning of $\mathbf{F}'$ and $\mathbf{C}'$ is obvious by comparing equations (43) and (44). This equation appears now as a regular eigenvalue equation except the dimensions of the matrices are troublesome. $\mathbf{F}'$ is $m \times m$ dimensional, $\mathbf{C}'$ is $m \times n$ and $\boldsymbol{\epsilon}$ is $n \times n$. The issue is easily resolved by simply diagonalizing $\mathbf{F}'$ and selecting the lowest energy $n$ eigenvalues and eigenvectors as the molecular orbital energies and columns of $\mathbf{C}'$ respectively. $\mathbf{C}$ is then readily obtained by the back transformation $\mathbf{C} = \mathbf{S}^{-\frac{1}{2}}\mathbf{C}'$. Note that diagonalization of $\mathbf{F}'$ will also yield an additional $m - n$ eigenvalues and vectors known as virtual, or unoccupied, molecular orbitals – but these orbitals do not bare much physical significance and should simply be regarded as an artifact of the numerical procedure used to obtain the occupied molecular orbitals, their corresponding energies and hence the electronic energy[h].

### 2.4.6 The "Hartree-Fock" Energy

As with solving the Hartree-Fock equations, the Roothaan-Hall equations also requires an iterative procedure. This is because obtaining a solution to equation (44), requires setting up the Fock matrix. The Fock matrix requires knowledge of the Coulomb and exchange operators ($\hat{J}_i$ and $\hat{K}_i$ in equation (36)), but in order to obtain $\hat{J}_i$ and $\hat{K}_i$ one requires the solution to equation (44). Thus, the procedure for solving equation (44) is achieved self-consistently and is illustrated in Figure 2-1, and this is known as the SCF procedure.

[h] The reason for the latter is that while the occupied molecular orbital energies are well defined and converge to specific energies as *m* is increased, the virtual orbitals are not well defined. The number of the latter increases with *m* and the energy of the lowest energy unoccupied orbital converges to 0 (being made up of the most diffuse functions in the basis set) as *m* increases. This zero energy solution corresponds to a free, or unbound, electron. All virtual orbitals are contaminated with such solutions, so cannot reliably be interpreted as "anti-bonding orbitals" or anything else physically meaningful.

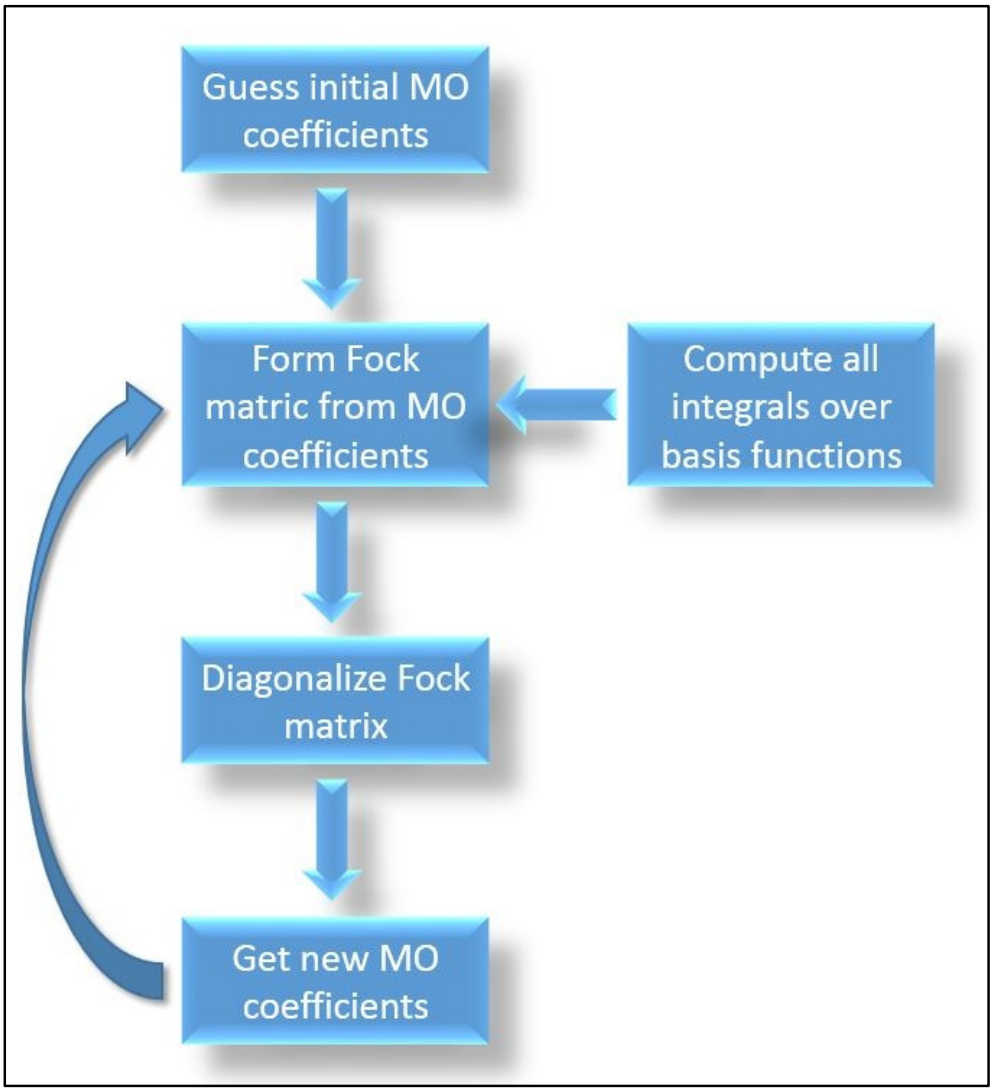



**Figure 2-1 The SCF procedure**

The solution to the Roothaan-Hall equations to obtain the electronic energy of a molecule within the independent particle model and basis set approximation involves the iterative procedure shown here. Once the molecular orbital (MO) coefficients are no longer changing significantly the iterations cease. The final electronic energy can be obtained from equation (38).

The bottle-neck in the SCF procedure is the calculation of the integrals, specifically the two-electron integrals of the type derived from equations (29) and (30). In general, such integrals are required over four completely different basis functions. The computational effort scales, therefore, formally scales as $\mathcal{O}(m^4)$. However, many of the integrals are tiny due to negligible overlap between the exponentially decaying functions. Intelligent screening techniques are used to reduce this computational effort down to as low as $\mathcal{O}(m^2)$, but usually the scaling

is closer to $\mathcal{O}(m^3)$. Nevertheless, these calculations can take many days to complete for systems containing large numbers of electron, as we shall see in the next chapter.

The title of this subsection has Hartree-Fock in quotes because the electronic energy obtained from the SCF procedure described above is not really the Hartree-Fock energy. The latter can only be obtained by solving the Hartree-Fock equations, (37). Nevertheless the basis set approximation is ubiquitous in quantum chemistry so much so that it is presumed to be applied and the energy so obtained from the SCF procedure simply denoted as the Hartree-Fock energy – it being understood that it is not actually this energy, but only an approximation to it. If one wishes to be explicit and clear regarding the electronic energy obtained being the *actual* Hartree-Fock energy, then the terminology to use is the Hartree-Fock limit mentioned previously.

### 2.4.7 Post Hartree-Fock Methods

Apart from Chapter 6 where hybrid-density function theory was used in the computations, no post Hartree-Fock methods were employed in the work described in this thesis. This is because the focus of the present research is to develop and test novel methods for computing the total electronic energies of a large collection of water molecules for use in bulk water simulations. Due to the extreme expense associated with computing accurately these energies, much lower levels of theory were employed in this thesis, i.e., Hartree-Fock (HF) theory, to test the methodologies. The presumption is that if the methodology well reproduces the HF energy then there is no reason to expect it to not also well

reproduce the post HF energies – such energies being entirely inaccessible for large systems, as we shall shortly show.

In the HF method, the spin-correlation exhibited by parallel spinning electrons is fully accounted for by writing the trial wavefunction as a Slater determinant, shown in equation (20). Spin-correlation between parallel spinning electrons gives rise to the exchange energy, $K_{ij}$, in equation (30). However, the use of the independent particle model specifically neglects the necessarily correlated motion of electrons in physical space due to the fact that they repel one another through the Coulomb force.

The HF energy incorporates electron-electron repulsion by allowing electrons to move independently of each other around nuclei then computing the electron repulsion between the charge on each electron to the average electron field produced by the independent motion of all other electrons. This repulsion energy must be solved self-consistently, so there is some feed-back embedded in the SCF procedure to allow for electrons to react to the average field of all the remaining electrons. Nevertheless this "reaction" is always to an average field, which results in too large a repulsion energy than there is in reality. Post HF methods seek to correct for this embedded error in the HF energy.

The missing electron correlation energy can be treated with perturbation theory. The use of perturbation theory is sound because the vast bulk of the electronic energy is accounted for by the HF energy with the small additional effect due to the neglect of explicit electron correlation being a perturbation on top of this HF energy. Møller-Plesset (MP) perturbation theory is the most commonly employed perturbative post Hartree-Fock electron correlation treatment. The first non-zero

correction to the HF energy is at second order, designated as MP2. Higher and more expensive, but more accurate orders of perturbation theory may also be included, i.e., third order MP3, fourth order MP4, or even fifth order MP5 corrections to the HF energy. Use of perturbation theory is not variational, so the corrected energy may overshoot the true energy of the system. Implementation of MP perturbation theory is computationally expensive and scales as $\mathcal{O}(m^4)$ to $\mathcal{O}(m^5)$ for MP2, and $\mathcal{O}(m^{s+3})$ for MP*s*, where $s > 2$. MP2 represents the least expensive *ab initio* post Hartree-Fock method.

Configuration interaction (CI) is another possible way to account for electron correlation, with full-CI being the best one can possibly achieve for a given size basis set. However, for all but the smallest systems (three, or maybe four atoms), full-CI is out-of-the-question as it scales as $\mathcal{O}(m!)$. Including CI with single and double electron excitations (CISD) produces results of about the same quality as MP2 corrections but scales as $\mathcal{O}(m^6)$. CISD has the advantage, however its energy is variational. CI with single, double, triple and quadruple electron excitations scales as $\mathcal{O}(m^{10})$. CI methods are very expensive and it is unclear that the energies obtained from them (except full-CI) are significantly better than MP*s* corrections.

The coupled-cluster (CC) implementation of electron correlation is considered the most accurate for their cost. With single, double and perturbative triple electron excitations taken into account, i.e., CCSD(T), the electron correlation obtained is widely accepted as the "gold-standard" for accuracy. If a problem is amenable to a CCSD(T) calculation with a large basis set, then the electronic energies are expected not to be too far from the true electronic energy of the system and highly accurate (within the BO approximation and also the non-relativistic Hamiltonian).

CCSD(T) calculations are performed iteratively, with the iterations scaling as $\mathcal{O}(m^6)$, and the final non-iterative perturbative triples step scaling as $\mathcal{O}(m^7)$.

As an indication of the computational expense involved in these computations, HF calculations were performed using the Gaussian 09 suite of programs[21] with a reasonably large basis set (aug-cc-pVTZ, explained in later chapters) for the water tetramer, octamer, 12-mer and 16-mer illustrated in Figure 2-2. The calculations were performed on a single core of an IBM HS21XM Bladeserver equipped with 2 Intel Xeon E5450 3.0GHz quad-core Harpertown CPUs. Each of the computations allowed for 2 GB of RAM to be allocated to the job and 20 GB of disk space. The CPU timings for these calculations were 16, 194, 659, 1365 minutes respectively for a single HF energy calculation. The number of basis functions were 368, 736, 1,104 and 1,472 respectively. For comparison, MP2 calculations were performed on the tetramer and octamer which took 33 and 528 minutes of CPU time respectively. A CCSD(T) calculation on the tetramer took 1447 minutes (i.e., approximately one day).

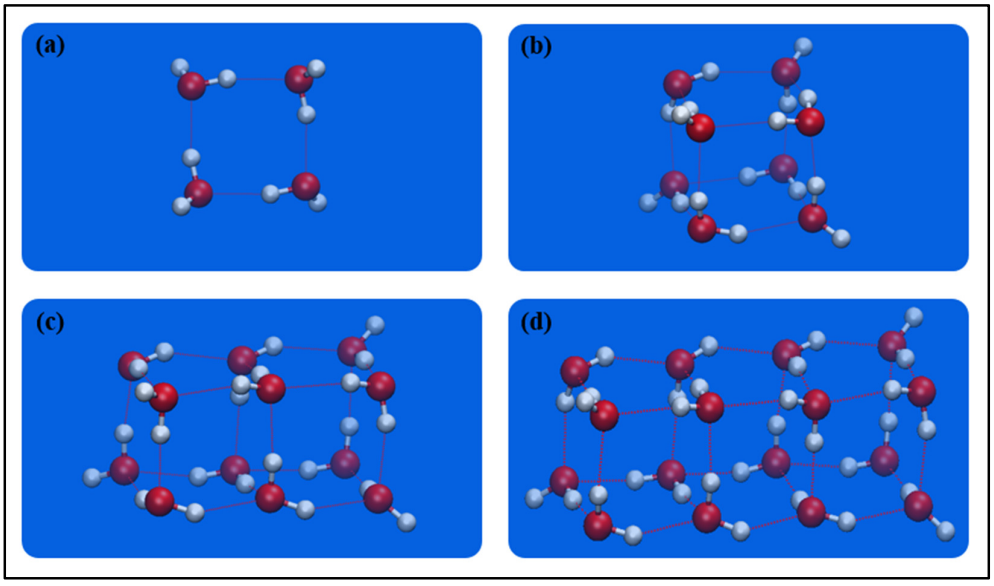



**Figure 2-2 Water 4*n*-mers, *n* = 1 – 4, used for CPU timing tests**
The tetramer (a), octamer (b), 12-mer (c) and 16-mer (d) used in CPU timing tests of HF, MP2 and CCSD(T) *ab initio* methods. See text for details.

In section 2.3 we saw that a sample size of about 800 water molecules was perhaps necessary for a bulk water simulation – this corresponds to 73,600 basis functions needed for a single HF energy. Assuming that enough computational resources were available (unlikely), the above computer would roughly take 900 years to complete a single HF energy calculation and 35 *trillion* years (more than 2,500 times longer than the age of the Universe[i]) for the "gold-standard" CCSD(T) calculation. Also recall that of the order of 200 million such energy evaluations are necessary for statistical averaging of thermodynamic properties to compare with experiment and it should be clear that it is impossible at present as well as any time soon to employ *ab initio* calculations directly in models of bulk water. Accurate, efficient and substantially time (and resource) saving methods must be developed if there is to be any hope of accurately modeling bulk water. Investigating possible methods to achieve this, is the subject of this thesis and described in the following chapters.

Before proceeding to the results chapters of this thesis it is important to point out that a method has existed for a long time that is extremely computationally efficient and is capable of, near exactly, reproducing *ab initio* interaction energies of a collection of molecules. There is one requirement for this treatment to be accurate and that is the molecules need to be far enough from each other so that there is no significant wavefunction overlap between them. This important method is described in the last section of this chapter because a combination of this treatment and some of the methods proposed in this thesis may eventually allow for accurate first principles simulation of bulk water.

[i] Even the Tianhe-2, the current world's fastest supercomputer, would take 42 million years to perform this single energy calculation.

## 2.5 Multipoles and Intermolecular Interactions

Much of the work presented in this section is a summary of relevant theoretical methods taken from the text "The Theory of Intermolecular Forces" by A. J. Stone[22]. These methods were employed during the course of my PhD candidature.

### 2.5.1 Multipole Operators and the Interaction Hamiltonian

We are interested in obtaining an accurate value of the interaction energy, i.e., equation (1), between a collection of molecules without having to resort to performing an extremely expensive *ab initio* calculation. To begin with, we shall consider just two interacting molecules, *A* and *B*. The Hamiltonian corresponding to this interaction energy is purely due to the potential energy associated with all of the charges in molecule *A* (nuclei and electrons) interacting with all of the charges in molecule *B*.

$$\hat{H}' = \sum_{a \in A} \sum_{b \in B} \frac{q_a q_b}{|\mathbf{r}_b - \mathbf{r}_a|} \tag{45}$$

The prime on the Hamiltonian indicates that it is an interaction Hamiltonian. $\mathbf{r}_a$ and $\mathbf{r}_b$ are vectors to the charges $q_a$ and $q_b$ found in molecules *A* and *B* respectively. All of the intramolecular operator terms found in equations (12) – (14), i.e., kinetic energy of the electrons, the intramolecular electron-nuclear attraction and intramolecular electron-electron and nuclear-nuclear repulsions have been subtracted. Only the potential energy of interaction between the charges in *A* and *B* remain. To proceed further, a particularly powerful series expansion of the inverse distance between two points, $|\boldsymbol{r}_1 - \boldsymbol{r}_2|^{-1}$, is utilised.

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{r_<^l}{r_>^{l+1}} (-1)^m C_{l,-m}(\theta_1, \varphi_1) C_{l,m}(\theta_2, \varphi_2) \qquad (46)$$

This expansion is written in spherical polar coordinates. Both *l* and *m* are integers with $l \geq 0$ and *m* ranging over the integers $-l$ to $l$, so for a given value of *l* there are $2m + 1$ possible values of *m*. The functions, $C_{l,m}(\theta, \varphi)$ are renormalised spherical harmonics defined as

$$C_{l,m}(\theta, \varphi) = \left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} Y_{l,m}(\theta, \varphi) \qquad (47)$$

with $Y_{l,m}(\theta, \varphi)$ being the spherical harmonics themselves. The definition of $C_{l,m}(\theta, \varphi)$ is chosen such that $C_{l,0}(0,0) = 1$. These are functions of the direction of a vector only, and not its magnitude, which can generally be non-zero for any arbitrary value of *l*. They do not necessarily get larger, nor smaller, with *l*. What makes the series, (46), converge is the ratio of distances, $\frac{r_<^l}{r_>^{l+1}}$. Most importantly this ratio must be less than one by ensuring that the distance $r_>$ is always the larger of $r_1$ and $r_2$, and $r_<$ the lesser for the series to converge[j]. If this were not so, but the other way around, then the ratio of distances in equation (46) would continue to grow in size with each successive value of *l* and the series would diverge. The latter is a crucial point that we shall visit again shortly.

In order to usefully apply the series expansion to the problem at hand, we choose an origin for molecules *A* and *B* and let them be located at $\mathbf{r}_A$ and $\mathbf{r}_B$ respectively. Next we assign the position of the charge, $q_a$, relative to the origin of *A* as **a**. Thus, $\mathbf{r}_a = \mathbf{r}_A + \mathbf{a}$. We have a similar result for a charge in *B*, i.e., $\mathbf{r}_b = \mathbf{r}_B + \mathbf{b}$. It is

[j] It is for this reason that equation (46) cannot be used in place of the electron-electron repulsion operator (equation (23)). This condition requires us to distinguish between the indistinguishable electrons by ensuring that one electron is *always* further away from its origin than another electron.

important to note that the origins of the molecules depend only on the nuclear coordinates in the respective molecules and not on the coordinates of any of the electrons in each molecule. Now the two charges $q_a$ and $q_b$ interact via Coulomb's law, as seen in equation (45), across the distance

$$|\mathbf{r}_b - \mathbf{r}_a| = |\mathbf{r}_B + \mathbf{b} - \mathbf{r}_A - \mathbf{a}| = |\mathbf{r}_B - \mathbf{r}_A + \mathbf{b} - \mathbf{a}| = |\mathbf{R} + \mathbf{b} - \mathbf{a}| \qquad (48)$$

where **R** is the vector from the origin of molecule *A* to the origin of molecule *B* and is independent of any electron coordinates. We can now write down the potential energy of interaction between all of the charges in *A* and *B* as:

$$\widehat{H}' = \sum_{a\in A}\sum_{b\in B}\frac{q_a q_b}{|\mathbf{r}_b - \mathbf{r}_a|} = \sum_{a\in A}\sum_{b\in B}\frac{q_a q_b}{|\mathbf{R} + \mathbf{b} - \mathbf{a}|} \qquad (49)$$

To usefully apply the expansion (46) to the interaction Hamiltonian we are forced to identify the magnitudes of vectors in (49) with $r_>$ and $r_<$. We take $r_>$ to be $R = |\mathbf{R}| = \mathbf{r}_1$ and $r_<$ to be $|\mathbf{a} - \mathbf{b}| = \mathbf{r}_2$. Having committed to this identification we may rewrite the series expansion (46) as

$$\frac{1}{|\mathbf{R} + \mathbf{b} - \mathbf{a}|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{|\mathbf{a} - \mathbf{b}|^l}{R^{l+1}}(-1)^m C_{l,-m}(\theta_{\mathbf{R}}, \varphi_{\mathbf{R}})C_{l,m}(\theta_{\mathbf{a}-\mathbf{b}}, \varphi_{\mathbf{a}-\mathbf{b}}) \qquad (50)$$

and further note that

$$I_{l,m}(\mathbf{R}) = \frac{C_{l,m}(\theta_{\mathbf{R}}, \varphi_{\mathbf{R}})}{R^{l+1}} \qquad (51)$$

and

$$R_{l,m}(\mathbf{a} - \mathbf{b}) = |\mathbf{a} - \mathbf{b}|^l C_{l,m}(\theta_{\mathbf{a}-\mathbf{b}}, \varphi_{\mathbf{a}-\mathbf{b}}) \qquad (52)$$

The $I_{l,m}$ and $R_{l,m}$ are irregular and regular spherical harmonics respectively

$$\frac{1}{|\mathbf{R}+\mathbf{b}-\mathbf{a}|}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^{m}I_{l,-m}(\mathbf{R})R_{l,m}(\mathbf{a}-\mathbf{b}) \tag{53}$$

The expansion (53) still contains a vector, $\boldsymbol{a}-\mathbf{b}$, that may involve the coordinates of two electrons simultaneously – electrons that reside in different molecules. Fortunately, (53) can be simplified further still using another very powerful series that is exact and most definitely finite. The series is called the regular spherical harmonic addition theorem.

$$\begin{aligned}R_{LM}(\mathbf{x}+\mathbf{y})=&\sum_{l_1 l_2}\sum_{m_1 m_2}\delta_{l_1+l_2,L}(-1)^{L+M}\left[\frac{(2L+1)!}{(2l_1)!\,(2l_2)!}\right]^{\frac{1}{2}}\\&\times R_{l_1 m_1}(\mathbf{x})R_{l_2 m_2}(\mathbf{y})\begin{pmatrix}l_1 & l_2 & L\\ m_1 & m_2 & -M\end{pmatrix}\end{aligned} \tag{54}$$

where $\begin{pmatrix}l_1 & l_2 & L\\ m_1 & m_2 & M\end{pmatrix}$ is a Wigner 3*j* coefficient – just a simple number. The non-zero values of the 3*j* coefficient set the limits in the summations. Application of equation (54) to $R_{l,m}(\mathbf{a}-\mathbf{b})$ in (53) changes what was a function of the coordinates of two particles simultaneously into a product of two functions that act only on the coordinates of each particle separately. Upon utilization of this addition theorem our expansion (53), now only contains terms that:

(i) depend on nuclear coordinates, i.e., the $I_{l,m}(\mathbf{R})$, and

(ii) depend separately of the positions of charged particles in each molecule, i.e., $R_{l_1 m_1}(\mathbf{a})$ and $R_{l_2 m_2}(\mathbf{b})$ relative to the origins of those molecules.

Applying the above simplifications to the series expansion (53) and further taking into account the fact that the vectors **a** and **b** are most conveniently expressed in the molecule-fixed axis systems of *A* and *B* (rather than the lab-fixed axis system), the series can be substituted back into the interaction Hamiltonian (49). Once the

substitution has been made the sum over the charged particles in each molecule can be performed. Thanks to the factorization effected by equation (54), the summations over charged particles appearing in the series expansion of the Hamiltonian involve the terms

$$\sum_{a\in A} q_a R_{l_1 m_1}(\mathbf{a}) = \hat{Q}^A_{l_1 m_2} \tag{55}$$

which are the definitions of multipole operators. E.g., $\hat{Q}^A_{00}$ is the monopole (or charge) operator for molecule *A*, $\hat{Q}^A_{1m_1}$ is the dipole moment operator for component $m_1$ of on *A*, $l = 2$ is a quadrupole, $l = 3$ an octapole, $l = 4$ a hexadecapole, etc.

After a significant amount of algebra, the interaction Hamiltonian can be written very compactly as

$$\widehat{H}' = \sum_{l_1 l_2} \sum_{\kappa_1 \kappa_2} \hat{Q}^A_{l_1\kappa_1}\, \hat{Q}^B_{l_2\kappa_2} T_{l_1\kappa_1, l_2\kappa_2} \tag{56}$$

Here the operators $\hat{Q}^A_{l_1\kappa_1}$ and $\hat{Q}^B_{l_2\kappa_2}$ are purely real functions and are with respect to the molecule fixed axes in each molecule. They are simple linear combinations of the generally complex operators $\hat{Q}^A_{l_1 m_1}$ and $\hat{Q}^B_{l_2 m_2}$. The $T_{l_1\kappa_1, l_2\kappa_2}$ functions are purely functions of the relative distance (*R*) between the two molecules *A* and *B* and their relative orientations. They do not depend on the coordinates of any electrons. All of the relative distance dependence in the $T_{l_1\kappa_1, l_2\kappa_2}$ function is trivially expressed as $R^{-(l_1+l_2+1)}$. For example, a dipole-charge interaction varies as $R^{-2}$.

An important caveat in the use of operator (56), is when we assigned *R* as $r_>$ and $|\mathbf{a} - \mathbf{b}|$ as $r_<$, i.e., $R > |\mathbf{a} - \mathbf{b}|$ in the series expansion (50). This must hold true for all the intermolecular charged particle distances. We shall see later how this

condition can be used to define a "divergence sphere" around each molecule in order to determine if the multipole expansion of the resulting interaction energy will converge.

Finally, the interaction Hamiltonian, (45), can be readily written down for a set, *S*, of *N* molecules.

$$\widehat{H}' = \sum_{A<B\in S} \sum_{a\in A} \sum_{b\in B} \frac{q_a q_b}{|\mathbf{r}_b - \mathbf{r}_a|} \tag{57}$$

Substituting for $|\mathbf{r}_b - \mathbf{r}_a| = |\mathbf{R}_{AB} + \mathbf{b} - \mathbf{a}|$ and following the same arguments as above we obtain,

$$\widehat{H}' = \sum_{A<B\in S} \sum_{l_1 l_2} \sum_{\kappa_1 \kappa_2} \hat{Q}^A_{l_1\kappa_1} \hat{Q}^B_{l_2\kappa_2} T^{AB}_{l_1\kappa_1, l_2\kappa_2} \tag{58}$$

so that the interaction Hamiltonian is simply a pair-wise sum of all the individual unique pairs that can be made from all the molecules is the set *S*. The above caveat with regard to the distance still, of course, is required for the multipole series to converge for any particular pair.

### 2.5.2 Perturbation Theory

To obtain the interaction energy from the Hamiltonian, we need to solve the Schrödinger equation. The interaction energy is only a small fraction of the total electronic energy of the collection of molecules. As such perturbation theory is the perfect tool that can be applied to the problem at hand. Therefore, the interaction Hamiltonian is treated as a perturbing Hamiltonian to the individual isolated molecules electronic energy Hamiltonians.

Again, considering first the interaction of only two molecules, *A* and *B*, let $\Psi^A_n$ be an exact solution for state *n* to the electronic energy Schrödinger equation for molecule *A*. The Hamiltonian used is identical to equation (12) but specifically for

molecule *A* in isolation from all else. Likewise for molecule *B*. Next we introduce the condition that molecules *A* and *B* are far enough apart such that their wavefunctions do not significantly overlap. This condition is entirely consistent with the use of the series expansion invoked in the previous subsection. Because molecules *A* and *B* are at "long-range" we can legitimately use the independent particle model to describe the wavefunction for the super system *AB*. Molecule *A* "owns" its electrons as does molecule "B" – there being no exchange of electrons between them[k].

$$\Psi_{nm} = \Psi_n^A \Psi_m^B \tag{59}$$

we shall also write equation (59) as

$$|nm\rangle = |\Psi_n^A\rangle|\Psi_m^B\rangle \tag{60}$$

where it is understood that the first index, *n*, is referring to the electronic state of molecule *A* and the second index, *m*, refers to that of molecule *B*. Thus, $|00\rangle$ is when both *A* and *B* are both in their ground-state unperturbed wavefunctions.

Perturbation theory yields successively better and better approximations to the actual interaction energy as one includes higher and higher orders of the theory. In the case of long-range intermolecular interactions, only the first order, $W_{00}'$ and second order, $W_{00}''$ energies are necessary to obtain a highly accurate interaction energy. Here the number of primes represent the order of perturbation theory and the subscripts "00" represent a correction to the ground-state energies of molecules

[k] Clearly once *A* and *B* are close enough this statement will no longer be true. One such case would be the formation of an H-bond between two water molecules. Once the molecules are in close-contact the series expansion of $r_{ab}{}^{-1}$ used in the previous subsection is invalid, so that the original Hamiltonian must be applied. Furthermore, account must be made for the fact the electrons can be exchanged between the two molecules. This means that the wavefunction of the super system must change sign upon interchange of the coordinates of a pair of electrons found in different molecules. The wavefunction (59) clearly does not possess this property.

*A* and *B*. Application of Rayleigh-Schrödinger perturbation theory yields the following expressions for these two energies.

$$W'_{00} = \langle 00|\hat{H}'|00\rangle \tag{61}$$

$$W''_{00} = -\sum_{nm} \frac{\langle 00|\hat{H}'|nm\rangle\langle nm|\hat{H}'|00\rangle}{W^0_{nm} - W^0_{00}} \tag{62}$$

In equation (62), *n* and *m* cannot both be zero simultaneously, so that the state $|00\rangle$ is excluded from this double sum. $W^0_{00}$ is the zeroth order ground-state energy of the *AB* super system and is just the sum of the two ground-state energies of molecules *A* and *B* in isolation from each other, i.e., the sum of the two unperturbed ground-state energies. Likewise $W^0_{nm}$ is the sum of the unperturbed $n^{\text{th}}$ excited state energy of molecule *A* and $m^{\text{th}}$ for that of molecule *B*.

The first order energy, given in equation (61), is the *definition* of the electrostatic interaction energy. It is an energy evaluated using the ground-state unperturbed wavefunctions of molecules *A* and *B*. Thus, *A* and *B* are in the presence of each other, but at first order molecule *A* does not have its electron density altered in any way due to the presence of *B* and vice versa. Because this interaction energy is evaluated using the ground-state unperturbed wavefunctions, this interaction is always exactly pair-wise additive.

The second order energy correction does allow for distortion of the electron density of each molecule due to the presence of the other. This interaction energy includes the induction and dispersion interactions between molecules. These two interactions are defined by breaking down the double sum in (62) into three separate summations,

$$U_{\text{ind}}^{A} = -\sum_{n\neq 0} \frac{\langle 00|\hat{H}'|n0\rangle\langle n0|\hat{H}'|00\rangle}{W_n^A - W_0^A} \tag{63}$$

$$U_{\text{ind}}^{B} = -\sum_{m\neq 0} \frac{\langle 00|\hat{H}'|0m\rangle\langle 0m|\hat{H}'|00\rangle}{W_m^B - W_0^B} \tag{64}$$

$$U_{\text{disp}} = -\sum_{\substack{n\neq 0 \\ m\neq 0}} \frac{\langle 00|\hat{H}'|nm\rangle\langle nm|\hat{H}'|00\rangle}{W_n^A - W_0^A + W_m^B - W_0^B} \tag{65}$$

so that $W_{00}'' = U_{\text{ind}}^{A} + U_{\text{ind}}^{B} + U_{\text{disp}}$. These interaction energies are always attractive, unlike the electrostatic interaction, which for consistency we shall label as $U_{\text{es}} = W_{00}'$. Equation (63) is the induction energy of *A* due to the field at *A* arising from molecule *B*. Equation (64) is the induction energy of *B* due to the field at *B* arising from molecule *A*. Equation (65) is the dispersion interaction. Having obtained expressions (61) and (63) – (65) for the interaction energy between two molecules via perturbation theory, only the integrals remain to be evaluated. This we shall perform in the next subsection.

### 2.5.3 Multipoles and the Interaction Energy

The Electrostatic Energy

Substitution of equation (56) into equation (61) yields the electrostatic interaction energy between molecules *A* and *B*. Performing the integration is trivial

$$\begin{aligned} U_{\text{es}} &= \langle 00|\hat{H}'|00\rangle \\ &= \langle \Psi_0^A|\langle \Psi_0^B| \sum_{l_1 l_2}\sum_{\kappa_1\kappa_2} \hat{Q}_{l_1\kappa_1}^{A}\, \hat{Q}_{l_2\kappa_2}^{B} T_{l_1\kappa_1,l_2\kappa_2}\, |\Psi_0^A\rangle|\Psi_0^B\rangle \\ &= \sum_{l_1 l_2}\sum_{\kappa_1\kappa_2} T_{l_1\kappa_1,l_2\kappa_2} \langle \Psi_0^A|\hat{Q}_{l_1\kappa_1}^{A}|\Psi_0^A\rangle \langle \Psi_0^B|\hat{Q}_{l_2\kappa_2}^{B}|\Psi_0^B\rangle \end{aligned}$$

hence

$$U_{\text{es}} = \sum_{l_1 l_2} \sum_{\kappa_1 \kappa_2} T_{l_1\kappa_1, l_2\kappa_2} Q^A_{l_1\kappa_1} Q^B_{l_2\kappa_2} \tag{66}$$

where $Q^A_{l_1\kappa_1} = \langle \Psi^A_0 | \hat{Q}^A_{l_1\kappa_1} | \Psi^A_0 \rangle$ and $Q^B_{l_2\kappa_2} = \langle \Psi^B_0 | \hat{Q}^B_{l_2\kappa_2} | \Psi^B_0 \rangle$ and are simply the actual permanent multipole moments of the two molecules. Equation (66) is then readily evaluated using the interaction functions $T_{l_1\kappa_1, l_2\kappa_2}$conviently listed in Appendix F of ref. 22. The electrostatic interaction energy for a collection of molecules is simply a pair-wise sum of all the unique pairs of molecules in the collection, *S*.

$$U_{\text{es}} = \sum_{A<B \in S} \sum_{l_1 l_2} \sum_{\kappa_1 \kappa_2} T^{AB}_{l_1\kappa_1, l_2\kappa_2} Q^A_{l_1\kappa_1} Q^B_{l_2\kappa_2} \tag{67}$$

The Induction Energy

After substituting equation (56) and integrating both equations (63) and (64), the equation is more involved because the sums are over the excited states of the unperturbed molecules. Simplification of the sum-over-states is achieved once the definition of polarizability is introduced

$$\alpha^A_{l_1'\kappa_1', l_1''\kappa_1''} = \sum_{n \neq 0} \frac{\langle \Psi^A_0 | \hat{Q}^A_{l_1'\kappa_1'} | \Psi^A_n \rangle \langle \Psi^A_n | \hat{Q}^A_{l_1''\kappa_1''} | \Psi^A_0 \rangle + \langle \Psi^A_0 | \hat{Q}^A_{l_1''\kappa_1''} | \Psi^A_n \rangle \langle \Psi^A_0 | \hat{Q}^A_{l_1'\kappa_1'} | \Psi^A_n \rangle}{W^A_n - W^A_0} \tag{68}$$

For example, when $l_1' = l_1'' = 1$ the polarizability is the familiar dipole-dipole polarizability. After substitution for the polarizability, the induction energy of *A* becomes

$$U^A_{\text{ind}} = \frac{1}{2} \sum_{l_1 l_2} \sum_{\kappa_1 \kappa_2} T_{l_1\kappa_1, l_2\kappa_2} \Delta Q^A_{l_1\kappa_1} Q^B_{l_2\kappa_2} \tag{69}$$

The factor of ½ is present to eliminate the double counting of the interaction due to the symmetrised definition of the polarizability (68). The quantity $\Delta Q^A_{l_1\kappa_1}$ is the

induced multipole component $l_1\kappa_1$ of *A* due to the fields at *A* arising from the permanent multipoles of *B*, $Q^B_{l_2\kappa_2}$. The induced multipole of *A* is

$$\Delta Q^A_{l_1\kappa_1} = -\sum_{l_2\kappa_2} \alpha^A_{l_1\kappa_1,l_2\kappa_2} V^A_{l_2\kappa_2} \tag{70}$$

where $V^A_{l_2\kappa_2}$ is the potential gradient of rank $l_2$, component $\kappa_2$at *A* due to *B*. It is given by

$$V^A_{l_2\kappa_2} = \sum_{l_3\kappa_3} T_{l_2\kappa_2,l_3\kappa_3}\left(Q^B_{l_3\kappa_3} + \Delta Q^B_{l_3\kappa_3}\right) \tag{71}$$

Directly analogous equations to (69) – (71) exist for the induction energy of molecule *B* due to the permanent multipoles of molecule *A*. The total induction energy is then the sum of these, i.e., $U_{\text{ind}} = U^A_{\text{ind}} + U^B_{\text{ind}}$.

Examination of equation (70) reveals that the induced multipole at molecule *A* depends on all of the induced multipoles at *B*. The same is true for an induced multipole at molecule *B* – it depends on all of the induced multipoles at molecule *A*. Thus, in practice, the only way to determine the induction energy at *A* is to compute it iteratively, i.e., self-consistently. This is one reason why the induction energy is not exactly additive if instead of two molecules interacting we have a collection of interacting molecules. An additional reason for the non-additivity of the induction interaction is that it depends on the *square* of the permanent multipoles on the other molecules. Equation (70) includes the permanent multipoles of *B*. This equation is then substituted into equation (69) and then multiplied by the same permanent multipoles of *B*. This can be understood by realizing that $T_{l_1\kappa_1,l_2\kappa_2} Q^B_{l_2\kappa_2}$ contributes to the field[1] at *A*. This field when

[1] Strictly speaking, only when $l_2 = 1$ are we talking about an electric field, for $l_2 = 2$ it is an electric field gradient, and $l_2 = 3$ it is a gradient of a field gradient, etc. However, for simplicity we very loosely use the term "field" here.

multiplied by the polarizability produces an induced multipole on *A*. This induced multipole then interacts with the field again to finally give the induction energy.

However, the most significant effect responsible for the non-additivity of the induction interaction is seen when the expression for the induction energy of a molecule in a collection of molecules (*S*) is derived. This expression is

$$U_{\text{ind}}^{A} = \frac{1}{2}\sum_{l_1 l_2}\sum_{\kappa_1\kappa_2} \Delta Q_{l_1\kappa_1}^{A}\left(\sum_{B\neq A\in S} T_{l_1\kappa_1,l_2\kappa_2}^{AB} Q_{l_2\kappa_2}^{B}\right) \tag{72}$$

combined with

$$\Delta Q_{l_1\kappa_1}^{A} = -\sum_{l_2\kappa_2} \alpha_{l_1\kappa_1,l_2\kappa_2}^{A}\left(\sum_{B\neq A\in S} T_{l_1\kappa_1,l_2\kappa_2}^{AB}\left(Q_{l_2\kappa_2}^{B} + \Delta Q_{l_2\kappa_2}^{B}\right)\right) \tag{73}$$

The term in parenthesis in equation (73), shows that the contribution to the induced multipole $l_1\kappa_1$ due to the field $l_2\kappa_2$ at *A* arises from the sum of all of the fields $l_2\kappa_2$ of all the molecules (apart from *A*). For example, it is quite plausible that the field due to one molecule at *A* is cancelled by another molecule. In this case, the relevant induced multipole is zero and so would be the corresponding induction energy. However, if the induction energy was evaluated due to each molecule separately it would be non-zero because the induction energy is always attractive. Having obtained the induction energy of *A*, the total induction energy of a collection of molecules will just be:

$$U_{\text{ind}} = \sum_{A\epsilon S} U_{\text{ind}}^{A} \tag{74}$$

The Dispersion Energy

Equation (65) provides the expression for the dispersion energy. It involves purely a sum over excited states of both molecules. Substituting for the multipole series

Hamiltonian yields complex expressions which can be ingeniously manipulated to finally arrive at

$$U_{\text{disp}} = -\frac{1}{2\pi}\sum_{l_1 l_2}\sum_{\kappa_1\kappa_2} T_{l_1\kappa_1,l_2\kappa_2} T_{l_1'\kappa_1',l_2'\kappa_2'} \int_0^\infty \alpha^A_{l_1\kappa_1,l_1'\kappa_1'}(iv)\alpha^B_{l_2\kappa_2,l_2'\kappa_2'}(iv)dv \qquad (75)$$

The $\alpha_{l\kappa,l'\kappa'}(iv)$ requires some comment. They are dynamic polarizabilities, but at imaginary frequencies and are usually viewed as a mathematical constructs rather than anything physical. They result by applying a mathematical identity known as the Casimir-Polder identity[23]. Such quantities can be computed using response theory available in software suites like Dalton[24] (but not Gaussian[21]). The functions look very much like a half-gaussian when plotted versus imaginary frequency and are trivial to numerically integrate once obtained. Of note is the value of this weird polarizability at $iv = 0$ – it is the regular static polarizabilities given in equation (68).

Equation (75) shows that the interaction involves an integral of a product of two separate functions, one for each molecule. The form of this expression is clearly additive, unlike the induction expression for the interaction of two molecules. When there is a collection of molecules the dispersion interaction simply takes the form

$$U_{\text{disp}} = \sum_{A<B\in S} U^{AB}_{\text{disp}} \qquad (76)$$

where $U^{AB}_{\text{disp}}$ is given by (75). Thus at second order, the dispersion interaction is purely additive.

If the perturbation series is extended to third order, however, a three-body dispersion term appears. For neutral molecules the first non-zero term involves the

dipole operator and is denoted the triple-dipole dispersion interaction. It is a very "close-range" interaction meaning it is only significant when all three molecules are close to each other. This is because the interaction varies as ${R_{AB}}^{-3}{R_{BC}}^{-3}{R_{CA}}^{-3}$. Because of the close-range distance dependence of the triple-dipole dispersion interaction it will not contribute significantly where the multipole expansion is valid and accurate, so can be safely ignored at long-range.

The physical origin of the dispersion interaction is embedded solidly in electron-correlation. For highly accurate work clearly dispersion must be included in the overall interaction energy of a collection of molecules. However, it was already noted in the first chapter of this thesis that to a good level of approximation water is spherical with respect to its electron density and polarizabilities. The first non-zero second-order dispersion interaction terms obtained from (75) for a water dimer only involves a significant contribution from the isotropic term. This term is characterised by the $C_6$ dispersion constant and varies as $R^{-6}$. The non-zero anisotropic terms depend on the relative orientation of the two waters, but contribute negligibly by comparison to the isotropic $-C_6R^{-6}$ term. Thus, at long-range for a cluster of water molecules and in bulk water, the effects of electron correlation can very accurately be taken into account though a simple pair-wise sum of $-C_6R_{AB}^{-6}$. We conclude that for the purposes of methodology development, it is unnecessary to spend CPU time and resources on performing high-level post-HF calculations because the electron correlation at long range can be accounted for trivially. It is for this reason that in the long-range methodology developmental work presented in this thesis, no post-HF calculations were performed nor are they necessary.

### 2.5.4 Accuracy of the Multipole Expansion

There are two approximations made in the above treatment of the long-range interaction energy. The first involves the use of the multipole expansion originating from the infinite series expansion of $r_{ij}{}^{-1}$, equation (46). By specifying that one of two separate distances involved is always greater than the other means that where this is not true the series diverges producing ever greater, on average, "interactions" with each successive term in the series. It has been shown that the multipole expansion fails when molecules approach each other close enough for their "divergence spheres" to overlap[25].

A divergence sphere is a sphere that just encompasses all nuclei in the molecule. For $C_{2v}$ water this is a particularly small sphere of radius about 0.84 Å. Therefore, no two water molecules can be within 1.68 Å or else the multipole expansion will fail. Additionally, water-water distances greater than but close to this value will produce a series that is very slowly convergent. Nevertheless, the divergence of the multipole expansion appears not to be a particularly problematic issue for water[m].

The convergence of the multipole series can be sped up considerably by using a distributed multipole approach. That is, instead of placing one set of multipoles at the origin of each molecule, a set of multipoles can be placed, say, on the different nuclei within the molecule. In doing so, the single divergence sphere around the molecule is replaced by a series of divergence spheres at each multipole site within the molecule with each sphere having a radius of about half a bond length. Distribution of the multipoles out from the origin can be performed using *ab initio*

[m] It most certainly is for larger molecules, e.g., π-stacked complexes of aromatics.

calculations in either basis set space[22, 25a], physical space using the electron density[26] or a combination of both[26]. The process by which this is accomplished is called distributed multipole analysis, or DMA. The new long-range interaction energy expressions given in this section are now only modified by including an extra summation over the multipoles on each molecule. The relevant distances are site-site distances on different molecules rather than molecule-molecule distances between the origins of each molecule. DMA is utilised in this thesis for its accuracy and fast convergence properties.

The second approximation made in obtaining the interaction energy expressions in this section was in writing the wavefunction of a collection of molecules as a simple product of the individual wavefunctions of the separate molecules. This approximation is only valid when the molecules are far enough from each other so that there is no significant electron exchange taking place between them. This only occurs when there is no significant wavefunction overlap between the molecules. When electron exchange does take place, the overall wavefunction for the super system needs to take account of the fact that it must change sign upon interchange of the coordinates of any two electrons – a product wavefunction does not. Additionally, as the wavefunctions overlap, replacing a smeared out charge density with a set of multipoles located at a point somewhere inside the charge cloud is inaccurate. Here these interpenetration effects reduce the accuracy of the multipole expansion which is in addition to the fact that now electrons can be exchanged between molecules.

A natural question therefore arises. At what distance do the wavefunctions of separate molecules begin to significantly overlap? By the time one has moved a

van der Waals radius away from a nucleus the vast majority of electron density has already been accounted for. This density is falling off exponentially as we continue to move further away from any nucleus. Thus, as a guide, atom-atom distances should be no closer than the sum of the van der Waals radii of the two atoms or else significant wavefunction overlap is occurring. The van der Waals radii of O and H are 1.5 and 1.2 Å respectively, thus O–O distances should be greater than 3.0 Å, O–H distances 2.7 Å and H–H distances 2.4 Å. In the worst case, two H atoms could approach each other along their respective O–H bonds. In this case, the O–O atom distance should not be closer than 4.2 Å (1.0 Å per O–H bond plus 2.4 Å for the H-H van der Waals distance).

A typical H-bonded O–O distance is 3.2 Å, whereas van der Waals overlap begins to occur at a distance of about 3.7 Å. There is very clear wavefunction overlap occurring in an H-bond, as is evident in Figure 1-3. Therefore, the treatment in this section of the long-range interaction cannot be used accurately in H-bonds, nor can it be expected to be accurate in close-contact interactions, i.e., in the first hydration shell of waters. Waters in the second hydration shell are about 5 Å away from the central water. This distance does appear far enough away for a long-range treatment to be accurate, hence we shall use this as a working condition to test the accuracy of the multipole long-range interaction energy in later chapters.

# Chapter 3

# Energies of Water Clusters Using Spherical Shells

## 3.1 Introduction

Ultimately we wish to be able to perform an accurate first principles simulation of bulk water. Currently such simulations are conducted using either a Metropolis Monte Carlo algorithm or classical molecular dynamics. A sample of bulk water contains on the order of $10^{23}$ water molecules or more. Obviously simulations that attempt to reproduce the properties of any bulk material cannot include anywhere near this number of particles. The approach currently taken to handle this situation is to:

(i) utilise periodic boundary conditions in the simulation, and

(ii) to restrict the number of water molecules explicitly interacting with one another in a meaningful and physically realistic way.

It is useful to understand the meaning of periodic boundary conditions and how it is utilised in a simulation in order to know just how many water molecules we need to be able to explicitly interact with each other.

Figure 3-1 illustrates periodic conditions in two dimensions. An imaginary box of length $l$ is filled with $N$ water molecules. This box is then surrounded by identical replicas of itself an infinite number of times in all directions. If, during a simulation, a water molecule passes from the box located at (0,0,0) into a neighboring box, say at (1,0,0) then its twin must have passed from the box located

at (-1,0,0) into box (0,0,0), and another identical copy must have passed from the box located at (-2,0,0) into the box (-1,0,0), etc. In fact in this example, every replica of the original box will have a water pass out of it to the right and another water (an image of itself) pass into it from the left. This symmetry exists because all the boxes are identical.

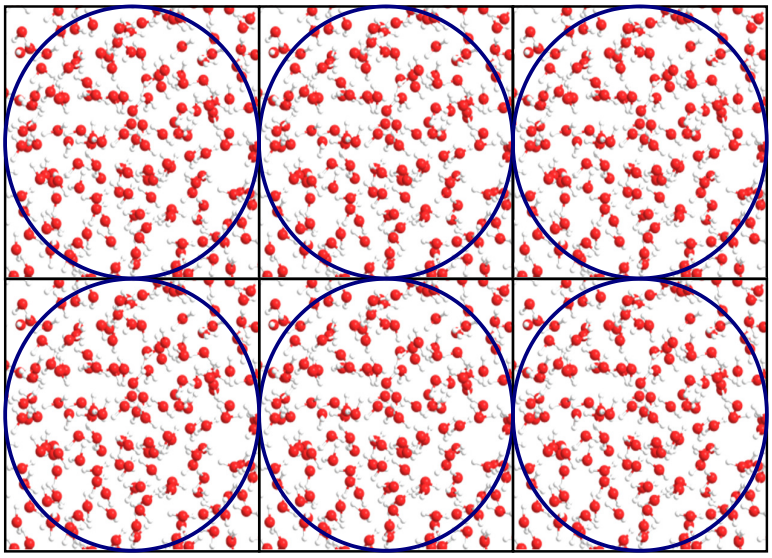


**Figure 3-1 A schematic 2D representation of periodic boundary conditions**
An imaginary box is filled with a sample of water molecules. Identical boxes to the original are placed around the original box in all directions and out to infinity. Six identical boxes in 2D are illustrated above. Water molecules that are explicitly interacted with each other fall within a sphere (circles in the 2D case) of radius less than half the imaginary box length. All the water molecules within the blue circles above illustrate those waters that are explicitly interacting with a water in the center of a box.

Of course a real bulk water sample does not exhibit this repeating unit-cell type property. In order to avoid artifacts, for the simulation resulting from this unphysical symmetry, an additional condition is required. The size of the box, $l$, must be large enough so that the interaction between a water located at the center of the box and a water within distance ½ $l$ or further away must be small. The total interaction between the central water and all other waters within a radius of ½ $l$ can then be computed explicitly. This sphere (blue) is illustrated in Figure 3-1. The interaction between the central water and water molecules further away than

½ $l$ is carried out implicitly whereby the waters in this distant region are treated as a continuum material. The total energy of the unit cell is then just a simple average of the total energies of $N$ spherical water clusters with each sphere centered on every water molecule in the box.

Experiment can be used as a guide to establish what $l$ should be. Figure 3-2 shows the experimental radial distribution functions for water at 298 K and under 1 bar pressure. Evident in the $g_{OO}(r)$ plot are the first, second and third hydration shells around water, with the curve becoming flat at around 9 – 10 Å. Thus, a value of $r$ = 10 Å appears to be the smallest appropriate value for explicit water–water interactions to be computed. This means $r = \frac{1}{2}l \geq 10$ Å so $l \geq 20$ Å. Using a density of 1 g cm$^{-3}$, a box of size $20 \times 20 \times 20$ Å$^3$ contains 268 water molecules (using $r$ = 15 Å produces a box with 903 waters). A sphere with $r$ = 10 Å contains 140 water molecules. As a check, the most favourable dipole-dipole electrostatistic interaction between two water molecules separated by 10 Å corresponds to an interaction energy of $-0.412$ kJ mol$^{-1}$ (at $r$ = 15 Å it will be $-0.122$ kJ mol$^{-1}$). This represents about 5.5% of the kinetic energy in translational and rotational motion combined at 298 K, i.e., 5.5% of 7.433 kJ mol$^{-1}$.

Having established the smallest reasonable size cluster of water molecules necessary for an accurate simulation of bulk water we next need to obtain representative samples of such clusters. To this end we adopted a model of the type described in section 1.3 whereby TIP4P water was used in conjunction with Metropolis Monte Carlo simulation to generate possible water cluster configurations. The Monte Carlo simulation was run with a box large enough to contain about 400 water molecules. This corresponded to an $l \lesssim 23$ Å, slightly

larger than the minimum size necessary as discussed above. The simulation was run until equilibrium was established under 1 bar pressure and 298 K *NpT* ensemble conditions. Random samples of water configurations were extracted during the course of the simulation while water was at equilibrium. The extracted water configurations were then used as a starting point for developing a methodology for rapidly obtaining the total electronic energy of the system.

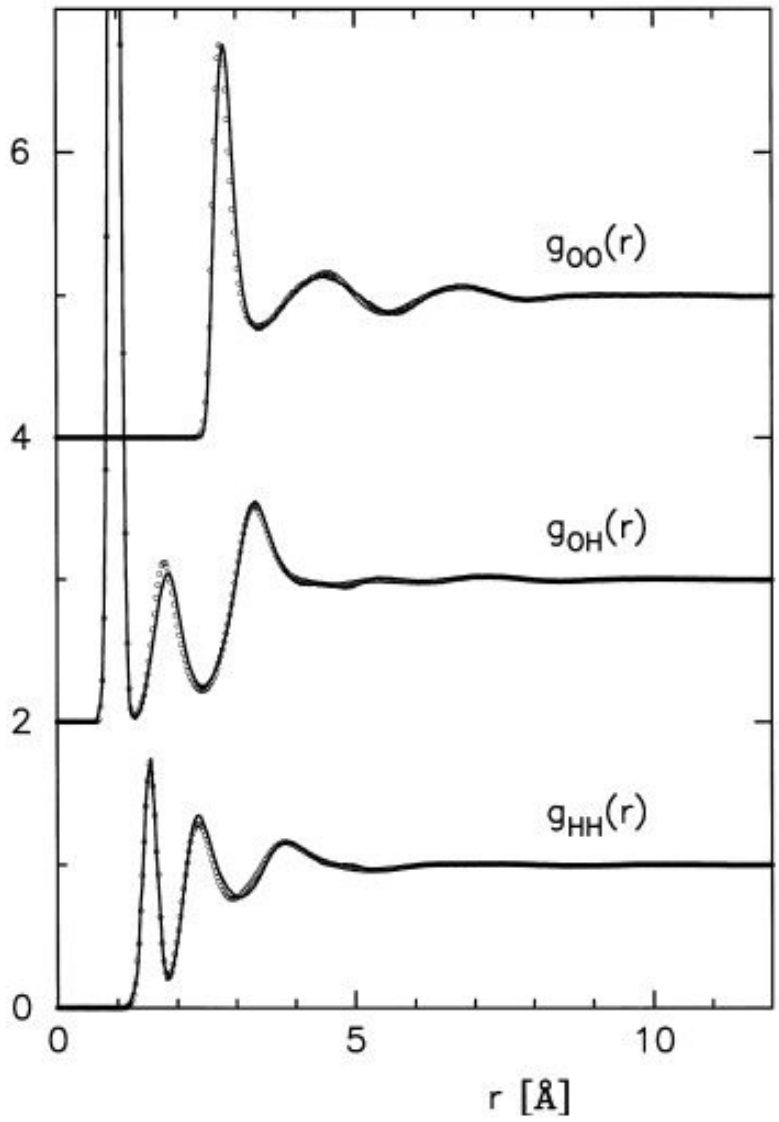



**Figure 3-2 Experimental O–O radial distribution function for water**
The experimental radial distribution functions for water under 1 bar and 298 K taken from ref. 27.

## 3.2 Computational Methods and Basis Set

All calculations reported in this chapter were performed at the HF level of theory as implemented in the Gaussian90 software suite[21]. The basis set employed was the 6-31G*. This basis set consists of a single linear combination of six gaussian functions (the "6" in the 6-31G* designation) available for the 1s atomic orbital (AO) of the O atom. There are additionally two mathematical functions: one a linear combination of three gaussians (the "3" in the 6-31G* designation) and another a single primitive gaussian (the "1" in the 6-31G* designation), available to describe each of the 2s, $2p_x$, $2p_y$ and $2p_z$ AO's of O. Note that the *p* gaussian functions are multiplied by *x*, *y* or *z* depending on whether the basis function is a $p_x$, $p_y$ or $p_z$ function respectively. Similarly there are two functions available to describe 1s AO on each H atom – a function which is a linear combination of three gaussians, and another which is a single primitive gaussian. Finally, the O atom also has a set of six Cartesian *d* functions. Each function is given by the product of two Cartesian coordinates (i.e., $x^2$, $y^2$, $z^2$, $xy$, $xz$, $yz$) and a single gaussian. These functions are commonly referred to as "polarization" functions and their presence in the basis set notation is indicated by the "*" in the 6-31G* designation. Thus, each water is represented with a total of 19 basis functions. Such a basis set is considered "small".

## 3.3 Selecting Test Spherical Water Clusters

Five randomly selected spherical water clusters of radius 12 Å were extracted from the above Monte Carlo simulations and labeled as *A* through *E*. The clusters were chosen such that at the center of each sphere was an oxygen of a random water molecule. This oxygen was given the origin coordinate (0,0,0). These clusters

represent snap-shots of possible water configurations within TIP4P bulk water at 298 K and 1 bar, i.e., the clusters are not energy minimised in any way. All water molecules were rigid and possessed the TIP4P geometry ($C_{2v}$ symmetry) of $r$(OH) = 0.9572 Å and $a$(HOH) = 104.52°. The five clusters with the number of waters in each and total electronic energies are given in Table 3-1. The corresponding geometries of the five water clusters can be found in Appendix A.

**Table 3-1 HF/6-31G* Energies of Water Clusters Studied**

| Cluster | No. of Waters | Eh / Hartree |
|---|---|---|
| *A* | 235 | −17865.0428203 |
| *B* | 229 | −17408.7295156 |
| *C* | 234 | −17789.0532464 |
| *D* | 246 | −18701.3342562 |
| *E* | 233 | −17713.0565957 |

## 3.4 Fragmenting Spherical Water Clusters

### 3.4.1 Fragmentation

One alternative means to accurately and efficiently obtaining the total electronic energy of a large chemical system not easily amenable to a single *ab initio* calculation is to utilise a fragment-based method. Molecular fragmentation is a relatively new field in theoretical and computational chemistry. It is been recently well reviewed by Gordon *et al.*[28]. The type of fragmentation selected here for testing on spherical water clusters was reviewed by Collins *et al.*[29] (specifically the precursor to the CFM method described therein), therefore only a brief description will be provided below for the relevant features.

In energy-based fragmentation methods, the total electronic energy of a chemical system is a simple linear combination of electronic energies of smaller fragments of the whole water cluster.

That is,

$$E(\mathbf{X}) \approx \sum_{i=1}^{N_f} f_i E_i(\mathbf{X}_i) \tag{77}$$

where $f_i$ is a fragmentation coefficient – a simple integer often $+1$ or $-1$. $N_f$ is the number of fragments and $E(\mathbf{X}_i)$ is the electronic energy of fragment $i$ with geometry $\mathbf{X}_i$. Note that we are now using $E_i(\mathbf{X}_i)$ for the approximate electronic energy of the system $\mathbf{X}_i$ rather than $\mathfrak{E}_i(\mathbf{X}_i)$ (as should be used based on the discussion in Chapter 2) because it is understood at this point that we cannot obtain the exact electronic energy of the system. The means of obtaining fragments is, in principle, simple and yet systematic.

Fragments are obtained by first identifying "groups" of atoms within the large molecule that will be combined together to form the fragments. In a typical valence-bonded system, these groups correspond to the functional groups in organic chemistry, e.g., –$CH_2$–, C=C, –OH, –COOH, C=O, etc. In the case of water clusters, a "group" might naturally be considered to be a single water molecule. In valence-bonded systems, fragments are then formed based on valence-bonded connectivity. For water clusters one naturally would consider fragments to be built up based on H-bond connectivity.

The formation of fragment molecules, the $\mathbf{X}_i$ in equation (77), from the previously established groups follows a prescription which is hierarchical and denoted by a level. Level 1 fragments are the smallest, i.e., possess the least number of atoms. Because the fragments are small their electronic energies can be computed very rapidly. However, the error between the energy obtained from equation (77) and the actual electronic energy of the system when computed without any

fragmentation (we shall call this error hereafter the fragmentation error) is largest. Level 2 fragments are larger than level 1 fragments, thus it takes longer to compute their electronic energies, but the fragmentation error is smaller than level 1. Level 3 fragments are larger still, take longer to compute, but produce smaller again fragmentation error. The hierarchy can continue to ever higher fragmentation levels until a level is reached where the "fragment" is so large it is the same as the chemical system being fragmented.

The prescription followed to obtain the fragments, once a level of fragmentation has been chosen, is based on connectivity of the groups. For a water cluster this means that the waters (single groups) are formed into fragments based on the H-bonding taking place in the cluster. In fragmenting a normal valence-bonded system, the number of connections between groups is typically quite small. For example, in a straight chain fully saturated hydrocarbon each group (–$CH_2$–) is only connected to its directly adjacent neighbors. This low degree of connectivity makes forming fragments near trivial in many cases with some complications arising when cycles of groups exist. Unfortunately, in the case of water clusters low connectivity is not something that is at all common with up to four groups being connected to a single group. Furthermore, water clusters possess very large numbers of interconnecting cyclic and branched H-bonded networks. This situation made fragmenting the water clusters at any level other than the highly approximate, inaccurate and essentially illustrative level 1, extremely arduous and ultimately unsuccessful despite many, *many* long and trying attempts which shall not be expounded upon here[n,30].

[n] It was later discovered that a fairly simple procedure can be followed to generate the secondary fragment molecules once the primary fragment molecules have been readily determined from the level of fragmentation. The procedure utilises the "inclusion-exclusion" principle of set theory, as explained in the work of Gadre (ref. above). Despite this, the approach still neglects important non-bonded interactions which we shall later see cannot be disregarded.

### 3.4.2 Fragmenting Spherical Water Clusters: Stage 1

Due to the difficulties encountered in fragmenting water clusters, an alternative approach was adopted. Rather than single water molecules forming groups, spherical shells of water molecules were defined as groups instead. Each shell of water molecules was approximately half a water thick. After some preliminary tests, a shell thickness of 1.5 Å was selected and illustrated in Figure 3-3. Thus, a 12 Å spherical water cluster can readily be divided into eight regions, with a central 1.5 Å radius sphere containing a single water molecule, along with seven spherical concentric shells 1.5 Å thick which are located around the central water molecule (deliberately coloured in yellow, shown on the right in Figure 3-3).

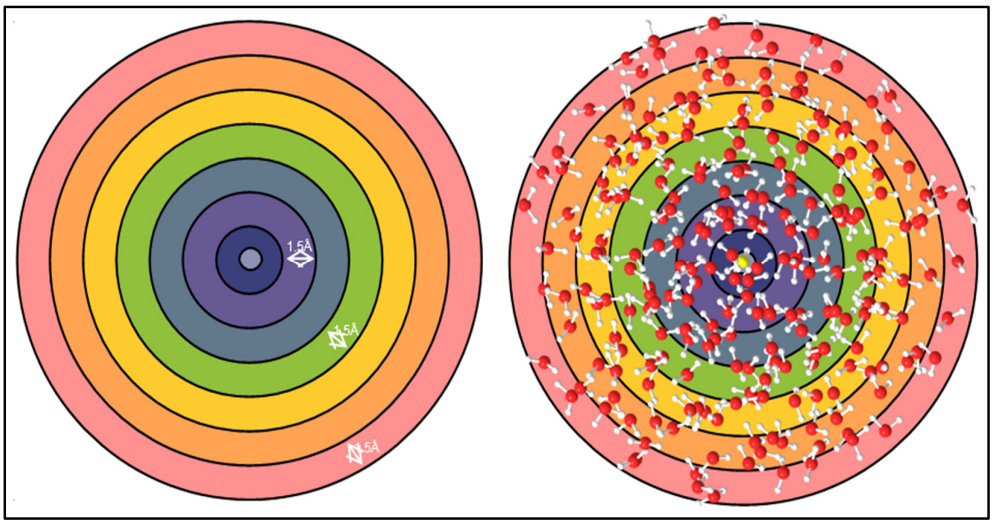



**Figure 3-3 Illustration of spherical shells in a water cluster**
A spherical water cluster of 12 Å radius is divided into 7 spherical shells and a small central sphere of radius 1.5 Å. Each shell is 1.5 Å thick. Each of these regions can be labeled 0 through 7, with 0 being the central sphere of radius 1.5 Å and containing a single water molecule. The remaining regions are labeled 1 through 7 with progressively greater radii for each consecutive shell.

There are advantages and disadvantages for this choice of groups. The main advantage is that a water cluster can now be readily and trivially fragmented, which we shall describe shortly. The second advantage is the high symmetry associated with the spherical shells tends to result in significant cancellation of

interactions between water molecules in outer shells with those in inner shells. The main disadvantage of this choice of groups is the quadratic increase in the number of water molecules in each group as we move outwards from the central water to the outermost shell. Table 3-2 indicates the number of water molecules in each spherical shell. This issue can be handled, however, with further fragmentation (see next subsection).

**Table 3-2 Number of Waters in Each Shell in Each Cluster**

| Cluster | Shell 1 | Shell 2 | Shell 3 | Shell 4 | Shell 5 | Shell 6 | Shell 7 |
|---|---|---|---|---|---|---|---|
| *A* | 3 | 7 | 15 | 36 | 37 | 65 | 71 |
| *B* | 4 | 7 | 20 | 25 | 38 | 64 | 70 |
| *C* | 5 | 7 | 16 | 26 | 43 | 65 | 71 |
| *D* | 3 | 9 | 20 | 26 | 49 | 50 | 88 |
| *E* | 3 | 8 | 17 | 29 | 47 | 51 | 77 |

A water cluster grouped in the above manner is fragmented at level 1 as follows. Regions that are adjacent to one another are considered as connected. At level 1, the primary fragments are formed from all possible pairs of connected groups. The secondary fragments are simply the groups that are double counted. The primary and secondary fragments possess fragmentation coefficients of $+1$ and $-1$ respectively. The primary fragments are hence given by groups $G_i G_{i+1}, i = 0 - 7$ and the secondary fragments are simply groups 1 – 7.

The level 2 primary fragments are formed by considering each group and adding to it all other groups that are connected to it. The secondary fragments are the segments of these fragments that are double counted. The primary and secondary fragments possess fragmentation coefficients of $+1$ and $-1$ respectively. The primary fragments are therefore given by groups $G_{i-1} G_i G_{i+1}, i = 1 - 6$ and the secondary fragments are simply $G_i G_{i+1}, i = 1 - 5$.

The levels 3, 5, 7 etc. primary fragments are generated from the previous odd level primary fragments by adding all groups connect to them. Similarly for levels 4, 6, 8 etc., primary fragments are generated from the previous even level primary fragments by adding all groups connect to them. Secondary fragments are always formed from segments of the primary fragments that are double counted. The primary and secondary fragments possess fragmentation coefficients of $+1$ and $-1$ respectively. In general the primary fragments at level *n* is given by $G_i G_{i+1} G_{i+2} \dots G_{i+n}, i = 0 - (7 - n)$ and the secondary are given by $G_i G_{i+1} G_{i+2} \dots G_{i+n-1}, i = 1 - (7 - n)$.

### 3.4.3 Fragmenting Spherical Shell Water Clusters: Stage 2

Given that stage one fragmentation can provide an accurate enough total energy at a low enough level of fragmentation, then a second and next stage of fragmentation is required. This is necessary in order to handle the quadratically increasing number of water molecules as the shell radius is increased. For fragments composed of shells with large radii, each of these fragments can be considered a new system for which we require an accurate total energy. Fragmentation can be applied to these large spherical systems in a second stage of fragmentation. In this second stage, groups can be formed by "zones" as illustrated in Figure 3-4. Fragments can then be formed from these groups following the same prescription as stage one fragmentation. The stage two fragments will be either spherical caps, or large zones being composed of several smaller zones.

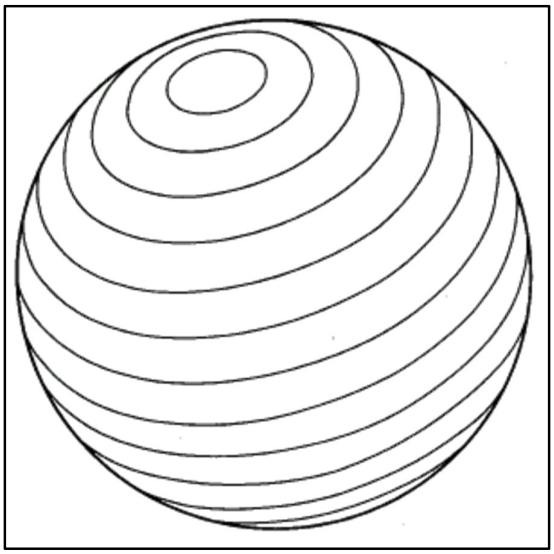


**Figure 3-4 Illustration of latitudinal zones in a spherical shell fragment**
A large spherical fragment from stage one fragmentation may be fragmented again. Groups are formed using latitudinal zones[o] from which a new set of fragments may be generated. This is denoted as stage two fragmentation.

### 3.4.4 Fragmenting Latitudinal Zone Water Clusters: Stage 3

Given that stage two fragmentation can provide an accurate enough total energy at a low enough level of fragmentation, then a third and final stage of fragmentation is required. This is necessary in order to handle the still quite large fragments expected in the equatorial zones. For latitudinal zones containing, or close to, the equatorial zone each of these fragments can be considered a new system for which we require an accurate total energy. Fragmentation can be applied to these large latitudinal zone systems in a third stage of fragmentation. In this third stage, groups can be formed by "quadrangles" as illustrated in Figure 3-5. Fragments then can be formed from these groups following the same prescription as stage one fragmentation. The stage three fragments will be large quadrangles being composed of several smaller quadrangles.

[o] Figure taken from http://facweb.bhc.edu/academics/science/harwoodr/GEOG101/Study/LongLat.htm (accessed on 12-Feb-2015).

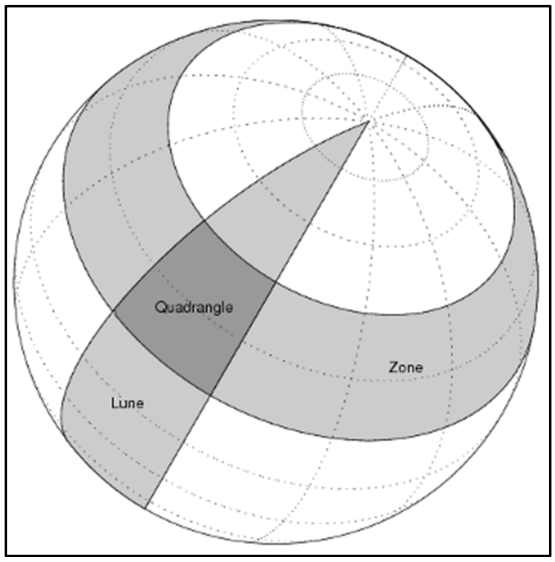



**Figure 3-5 Illustration of quadrangles in a zonal fragment**

A large latitudinal zone from stage two fragmentation may be fragmented again. Groups are formed using quadrangles[p] from which a new set of fragments may be generated. This is denoted as stage three fragmentation.

## 3.5 Stage 1 Fragmentation Energies of Water Clusters

### 3.5.1 Fragmentation Energies using Isolated Fragments

Figure 3-6 illustrates the mean absolute deviation (MAD) in μ-Eh per water monomer between the fragmentation energies of the spherical water clusters *A* – *E* and their total electronic energies. Clearly evident is the expected rapid convergence to the exact total electronic energy with fragmentation level. Note that 1000 μ-Eh per water monomer ≡ 2.625 kJ mol$^{-1}$. At room temperature, the kinetic energy in translation and rotation is 7.433 kJ mol$^{-1}$. Thus, an error of 1000 μ-Eh or more is unacceptably high. Much more acceptable are errors approximately to a tenth or less than this.

[p] Figure taken from http://www.mathworks.com/help/map/ref/areaquad.html (accessed on 12-Feb-2015).

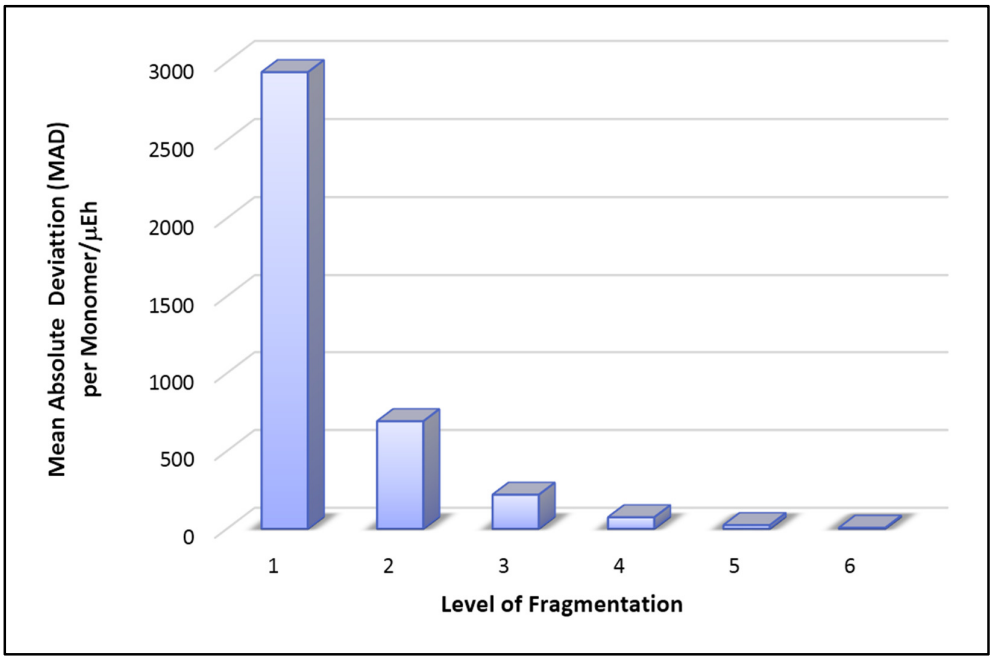



**Figure 3-6 MAD for isolated spherical shell fragments**
The mean absolute deviation (MAD) between fragmentation energies and total electronic energies at stage one fragmentation. The fragment energies are computed and using isolated fragments.

Table 3-3 provides the errors in reproducing the total energies of the individual clusters given in Table 3-1. We note from Table 3-3 that an error of 100 µ-Eh per water monomer or less does not occur until fragmentation level 4. Unfortunately, at level 4 the primary fragments contain five groups with the largest fragment being composed of shells (groups) 3 – 7. Referring to Table 3-2, we note that this single fragment comprises about 95% of the entire cluster, thanks to the quadratic increase in group size as we move further outward from the central water. Requiring this level of fragmentation for adequately low errors is unacceptable because the systems being fragmented are still essentially fully intact. Therefore, further reduction in the error is necessary before proceeding further onto stage two and three fragmentation.

**Table 3-3 Error/μ-Eh per Monomer in the Total Energy of Water Clusters Using Isolated Fragments**

| Cluster | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 |
|---|---|---|---|---|---|---|
| *A* | −2877 | −844 | −228 | −87 | −14 | 1 |
| *B* | −3118 | −1106 | −438 | −126 | −55 | −11 |
| *C* | −2639 | −570 | −189 | −89 | −38 | −9 |
| *D* | −2913 | −265 | −63 | −19 | −4 | 7 |
| *E* | −3127 | −669 | −174 | −48 | −5 | 4 |
| MAD* | 2935 | 691 | 219 | 74 | 23 | 6 |

* Mean Absolute Deviation.

### 3.5.2 Fragmentation Energies using Fragments Embedded in a Charge Field

One of the issues that plagues all fragmentation methods of the type described in section 3.4.1 is the fact that each fragment has its total electronic energy computed in isolation from the rest of the molecule. Each fragment represents part of a whole and for this representation to be authentic the fragment should be embedded in a Coulomb polarizable field. Such a field would then mimic the environment that the fragment is located in when it is part of a larger molecule.

One means to crudely approximate the field each fragment is located in is to augment the fragmentation prescription by placing simple point charges down around each fragment molecule at atomic sites present in the greater system, but not present in the fragment. For example, if a system was composed of three groups, $G_1G_2G_3$, which we shall simply write as 123, then its level 1 fragments would be 12, 23 and 2. The fragmentation energy of the system is just $E(12) + E(23) - E(2)$. The fragment 12 is computed in isolation. However, a charge field may be introduced into this calculation such that the point charges are placed on all the atoms belong to group 3. We will represent this type of fragment as $12\overline{3}$, where the bar above the "3" means point charges have been placed at the nuclei of

the atoms contained within group 3 when computing the electronic energy of fragment 12. The fragmentation energy of the whole system is now $E(12\bar{3}) + E(\bar{1}23) - E(\bar{1}2\bar{3})$.

Introducing a charge field in this way does not significantly increase the CPU time of the electronic structure calculations. Most fragmentation methods described in the Gordon review[28] (mentioned in section 3.4.1 ) incorporate "embedded charged" in the fragment calculations. The approach is highly approximate but it has been noted to produce better fragmentation energies. The results have also been noted to be fairly robust with respect to which specific type of charges are used. As such, for simplicity, we chose the TIP3P charges (TIP3P because the charges are located on the O and two H's) to place on the absent water nuclei in each fragment. The TIP3P charges[10e] are O, –0.834*e*; H, +0.417*e*, where *e* is the magnitude of the charge of an electron.

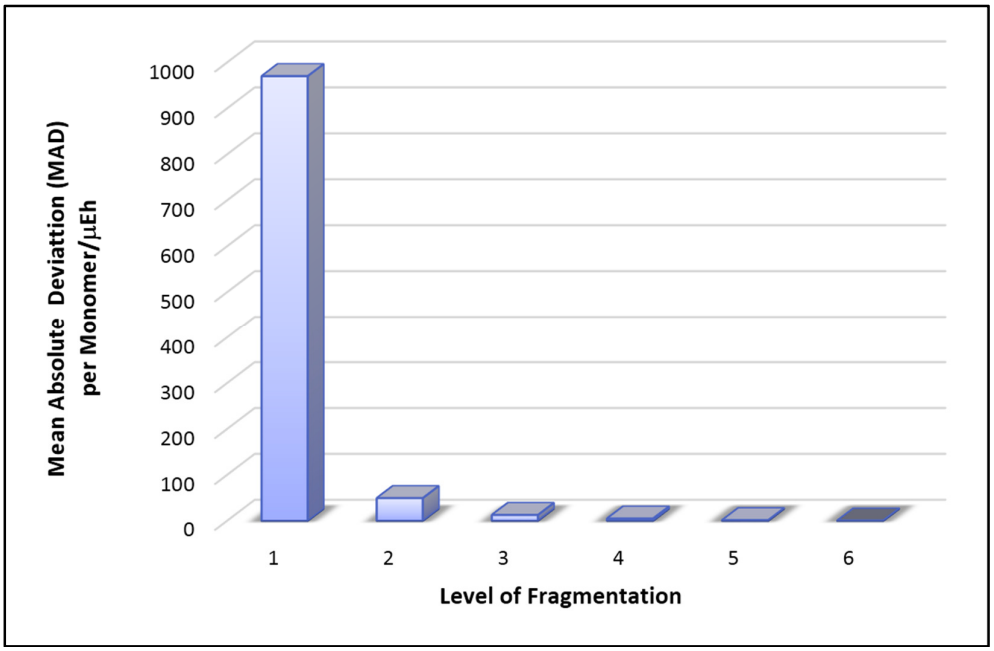



**Figure 3-7 MAD for spherical shell fragments in an embedded charge field**
The mean absolute deviation (MAD) between fragmentation energies and total electronic energies at stage one fragmentation. The fragment energies are computed using fragments embedded in a charge field of waters that are missing from the fragment.

Figure 3-7 is similar to Figure 3-6 except the fragments are now all computed in the presence of embedded point charges. The difference between the two figures is striking. Firstly note the full-scale on the vertical axis in Figure 3-7, which is a factor of three smaller than that seen in Figure 3-6. Secondly, convergence to the exact result is much more rapid. Level 2 fragmentation is also seen to be in error by less than 100 μ-Eh per water monomer. This is clearly illustrated in Table 3-4 below.

**Table 3-4 Error/μ-Eh per Monomer in the Total Energy of Water Clusters Using Fragments in a Charge Field**

| Cluster | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 |
|---|---|---|---|---|---|---|
| *A* | 857 | −65 | −15 | −7 | −1 | 0 |
| *B* | 981 | −67 | −19 | −6 | −3 | 0 |
| *C* | 748 | −49 | −8 | −4 | −2 | 0 |
| *D* | 1329 | −12 | −7 | −3 | 0 | 1 |
| *E* | 936 | −54 | −14 | −4 | −1 | 0 |
| MAD* | 970 | 49 | 13 | 5 | 1 | 0 |

* Mean Absolute Deviation.

Despite the crudeness of representing the Coulomb field due to the rest of the cluster missing from a fragment by simple TIP3P point charges, it appears that we are able to attain the required accuracy in the total energy by level 2 fragmentation. At level 2 the largest spherical-shell fragment contains shells 5-6-7. Referring to Table 3-2 we see that this one fragment constitutes, on average, 75% ± 2% of the entire cluster. It is very clear that if one wished to pursue this spherical-shell-fragmentation-method to accurately compute the total energy of a large water cluster rapidly and efficiently then stage two and probably stage three fragmentation would at least need to be applied to the largest fragments.

While fragmenting fragments is certainly an option, it is clear that if a method could be utilised that avoided stage two and three fragmentation it would be beneficial. One possibility is to separate the monomer energies from the total energy as indicated in equation (1). Isolated monomer energies can be quickly and

easily computed accurately, so that all of the computational effort can be transferred to computing the interaction energy. The interaction energy of a cluster is the sum of the individual monomer-cluster interaction energies less all interactions that are counted more than once. By considering the interaction between each monomer individually and their associated spherical clusters about them, only the close-contact interactions need to be computed using *ab initio* calculations. This is because multipoles and perturbation theory can be used to accurately compute all of the necessary long-range interactions as described in section 2.5 In following this approach, we are no longer concerned with the individual interactions between waters in outer spherical shells – we are only concerned with the interactions between the waters in the outer spherical shells and the central water. Thus, a significant speed-up in accurately determining the interaction energy of all the water molecules in a unit cell could be achieved if this approach were successful. We consider the possibility of accurately determining monomer-spherical cluster interaction energies in the next section.

## 3.6 Fragmentation Interaction Energies of Water Clusters

### 3.6.1 Interaction Energies via Fragmentation

Equation (77) gives the fragmentation energy of a chemical system. To obtain an approximate interaction energy between two chemical systems, *A* and *B* one first fragments each of these systems separately to obtain approximate total electronic energies of each in isolation from the other. That is,

$$E(\mathbf{X}^A) \approx \sum_{i=1}^{N_f^A} f_i^A E_i^A(\mathbf{X}_i^A) \tag{78}$$

and

$$E(\mathbf{X}^B) \approx \sum_{i=1}^{N_f^B} f_i^B E_i^B(\mathbf{X}_i^B) \tag{79}$$

An approximate "two-body" interaction energy (where a "body" is a fragment) can we written as

$$\varepsilon(\mathbf{X}^A \cup \mathbf{X}^B) \approx \sum_{i=1}^{N_f^A} \sum_{j=1}^{N_f^B} f_i^A f_j^B \varepsilon_{i,j}(\mathbf{X}_i^A \cup \mathbf{X}_j^B) \tag{80}$$

where $\varepsilon_{i,j}$ is the interaction energy of fragments *i* in system *A* and *j* in system *B* and is just $\varepsilon_{i,j} = E\left(\boldsymbol{X}_i^A \cup \boldsymbol{X}_j^B\right) - E\left(\boldsymbol{X}_i^A\right) - E\left(\boldsymbol{X}_j^B\right)$. Equation (80) is not exact as it does not include higher body interactions between the fragments, but with large enough fragments equation (80) will yield a closer and closer approximation to the true interaction energy between systems *A* and *B*.

In the case of our spherical water clusters, we have specifically chosen a water to lie at the center of the sphere. When the cluster is broken into spherical shells to form the groups, the central water is in the innermost sphere. This inner sphere contains only the central water, hence we consider this to be chemical system *A*. As this system contains only a single water molecule, there is no fragmentation to be performed on it. Applying equation (80) to determine the interaction energy between this single water and the rest of the spherical cluster yields

$$\varepsilon\left(\mathbf{X}^{(j)} \cup \mathbf{X}^{(i)}\right) \approx \sum_{i=1}^{N_f} f_i \varepsilon_{i,j}(\mathbf{X}^{(i)} \cup \mathbf{X}^{(j)}) \tag{81}$$

where *i* is a fragment formed from the spherical shells which does not include the central sphere, *j*. In general, the primary fragments of level *n* are given by $G_i G_{i+1} G_{i+2} \ldots G_{i+n}, i = 1 - (7 - n)$ and the secondary are given by $G_i G_{i+1} G_{i+2} \ldots G_{i+n-1}, i = 2 - (7 - n)$. These expressions are identical to the

total energy fragmentation formulae, except that the index $i$ starts one group further outward. An important difference here, however, is that we are interested in $\varepsilon_{i,j}$ given by

$$\varepsilon_{i,j}\left(\mathbf{X}^{(i)} \cup \mathbf{X}^{(j)}\right) = E\left(\mathbf{X}^{(i)} \cup \mathbf{X}^{(j)}\right) - E\left(\mathbf{X}^{(i)}\right) - E\left(\mathbf{X}^{(j)}\right) \qquad (82)$$

Thus, the interactions between all the waters found in fragment $i$ are irrelevant and can be ignored – only the interaction between the central water, $j$, and the fragment, $i$, is needed. In doing so, it should not be necessary to compute via *ab initio* methods the interaction energy between the central water and fragments at long range from it – perturbation theory and multipole methods can be used to compute these interactions, as described in section 2.5

### 3.6.2 Interaction Energies using Isolated Fragments

As pointed out in subsection 3.5.1 we consider an acceptable error for a snapshot single spherical shell configuration of water to be 100 μ-Eh per monomer. The total energy per water monomer for all of the $N$ waters in the periodic box is an average of all $N$ spherical shells total energies. Thus, the error in this average total energy per water monomer of all $N$ waters in the periodic box will be 100 μ-Eh/$\sqrt{N}$. To put this on equal footing with the interaction energy per monomer for all the $N$ waters in the periodic box, we first sum up all the interaction energies with each water in the periodic box, then we must remove any over counting of interactions. Interactions will have at least been double counted, thus our total interaction energy must at least be halved. The total interaction, after having been halved, will need to be divided by $N$ to obtain the interaction energy per monomer for waters within the periodic box. It is this factor of two that allows our tolerance

in the error in an individual interaction energy of a single water molecule with its surrounding spherical cluster to be 200 μ-Eh.

**Table 3-5 Error/μ-Eh in the Interaction Energy of a Central Water with the Remaining Waters in the Clusters Using Isolated Fragments**

| Cluster | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---|---|---|---|---|---|
| *A* | −6160 | −256 | −491 | 46 | −5 |
| *B* | −8892 | −3836 | −2508 | −1527 | −557 |
| *C* | −10461 | −2608 | −2009 | −1850 | −804 |
| *D* | −1172 | −1015 | −32 | 420 | 490 |
| *E* | −2451 | −455 | −515 | −311 | 165 |
| MAD* | 5827 | 1634 | 1111 | 831 | 404 |

* Mean Absolute Deviation.

Table 3-5 shows the errors in the interaction energy between the central water in each of the spherical clusters *A* – *E* relative to the rest of the water molecules. It is very clear from this table and Figure 3-8, which shows the mean absolute deviation in the errors, that these errors are completely unacceptable. Even at level 5 the MAD is fully a factor of two larger than acceptable. While this is a particularly disappointing result, perhaps the inclusion of an embedded charge field about each of the fragments may improve matters as occurred with errors in the total energies of spherical clusters. Some improvement is certainly expected due to the presence of a strong Coulomb field in the vicinity of the central water and spherical shell fragments – the influence of which is not included in the calculations given in this subsection.

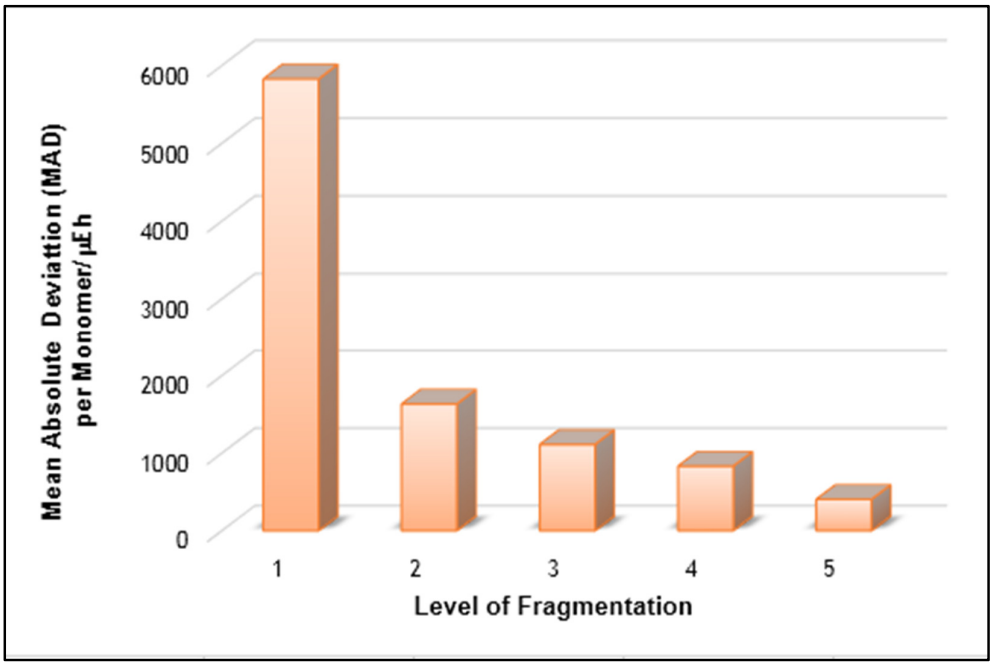



**Figure 3-8 MAD using isolated fragments in computing interaction energy**
The mean absolute deviation (MAD) between fragmentation interaction energies and actual interaction energies. The interaction energy is between a central water and the rest of the spherical cluster of water molecules. The fragment interaction energies are computed and using isolated fragments.

### 3.6.3 Interaction Energies using Fragments Embedded in a Charge Field

Table 3-6 shows the errors in the interaction energy between the central water in each of the spherical clusters *A* – *E* to the rest of the water molecules. Here each of the fragments are embedded in a charge field of water molecules not present in the fragment. Examination of Table 3-6 and Figure 3-9 clearly demonstrate an approximate ten-fold reduction in the errors in the interaction energies compared to errors associated with using isolated fragments found in the last subsection.

**Table 3-6 Error/µ-Eh in the Interaction Energy of a Central Water with the Remaining Waters in the Clusters Using Fragments in a Charge Field**

| Cluster | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---|---|---|---|---|---|
| *A* | −595 | −3 | −107 | −4 | −2 |
| *B* | −319 | −280 | −230 | −152 | −25 |
| *C* | −1025 | −177 | −124 | −187 | −58 |
| *D* | −260 | −179 | −35 | 60 | 76 |
| *E* | −230 | −52 | −54 | −44 | 1 |
| MAD* | 486 | 138 | 110 | 89 | 33 |

* Mean Absolute Deviation.

Even though Figure 3-9 shows that there is a definite downward trend in the errors with the level of fragmentation, it is nowhere near as distinct as compared with the total energies illustrated in Figure 3-7. This is very probably due to the fact that the total energies already contain a significant degree of "noise reduction" due to the averaging that is occurring in the total energies over all the water molecules within a cluster. Here the reported interaction energies are all only with a single water molecule to the rest of the cluster and not an average over such all such interactions within a cluster.

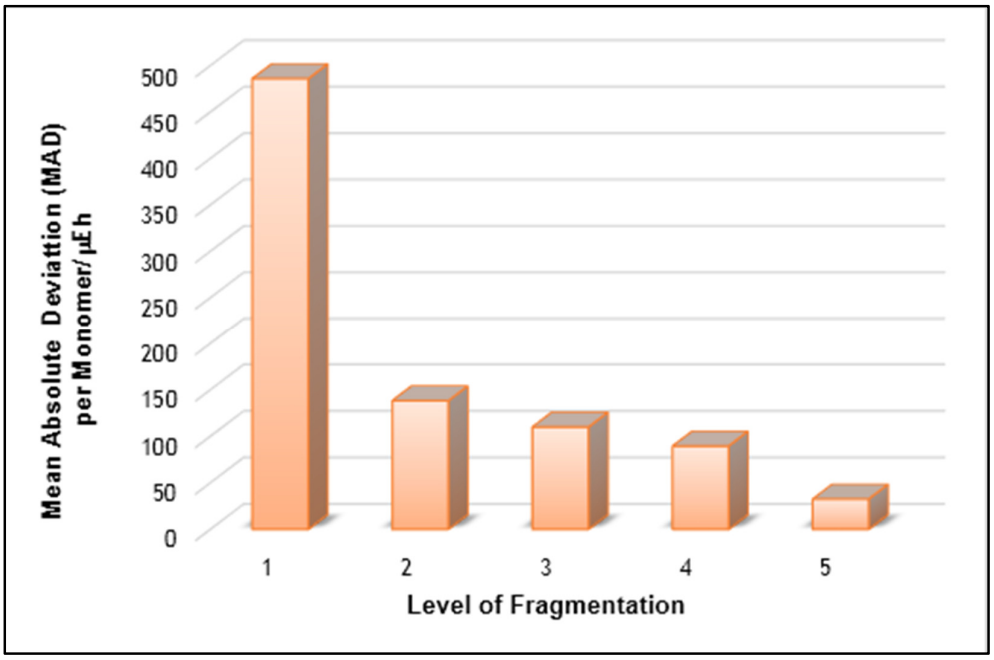



**Figure 3-9 MAD using fragments in a charge field in the interaction energy**
The mean absolute deviation (MAD) between fragmentation interaction energies and actual interaction energies. The interaction energy is between a central water and the rest of the spherical cluster of water molecules. The fragment interaction energies are computed and using fragments embedded in a charge field.

We note that the errors are now reduced to below 200 μ-Eh at level 2 – a level consistent with the conclusions drawn from the total energy results in subsection 3.5.2 As previously mentioned, the use of multipoles with perturbation theory should be able to very accurately reproduce interactions between the central water and waters further away than about 3.5 – 4.0 Å. Spherical shell two is located at a

distance of 3.0 – 4.5 Å from the oxygen atom of the central water in the spherical cluster. At level 2, the fragments which contain shell two are 123, 234 and 23. All other fragments, i.e., 345, 456, 567, 34, 45 and 56 lie further than 4.5 Å from the central oxygen atom. Thus, these interactions should be readily and accurately amenable to computation through the use of multipoles and perturbation theory.

The largest fragment at level 2 which contains shell two is fragment 234. Referring to Table 3-2 we note that on average this fragment possess 54 ± 3 water molecules and represents about 23% of an entire spherical cluster. It seems wasteful to have to include shells three and four in this fragment. When interacting the central water with fragment 234, it seems more efficient to only include shell two (containing on average 8 waters) in the *ab initio* calculation and leaving the interaction of the central water with shells 34 to multipoles. Shells three and four together contain, on average, 46 water molecules. It should be quite possible to break the interaction of the central water with fragment 234 down into an *ab initio* calculation of the interaction with shell two and then use multipoles to compute the interaction with shells three and four. In doing so, we would have reduced the accurate calculation of the interaction energy for a large spherical cluster to the *ab initio* calculation of many 9-mers, along with large numbers of fast and efficient multipole interaction energy calculations.

## 3.7 Summary

In order to perform an accurate bulk water simulation, the total energy or interaction energy of water molecules in a spherical cluster of radius about 12 Å needs to be evaluated. Such a spherical cluster contains about 235 water molecules, which is entirely too large to be amenable to accurate electronic structure methods. In this chapter, we explored the possibility of authentically reproducing the total electronic energy of these clusters. For testing purposes, the *ab initio* energy of the clusters were obtained from a highly approximate, yet representative, electronic structure computation. In order to authentically reproduce this electronic energy, a unique method of fragmenting the cluster was devised. The method involved breaking the cluster down into seven disjoint spherical shells about a small core sphere enclosing a single water molecule. Spherical shell fragments were then constructed from these thinner spherical shells and the core sphere. It was found that by dividing the large cluster up in this manner we were able to acceptably reproduce the total energy of the clusters, but not without having to embed each spherical fragment in a Coulomb field of point charges located at atomic sites not present in each fragment.

Unfortunately, this method of fragmentation generates a distribution of spherical shell fragments of quadratically increasing size, with the largest fragment still being too large for highly accurate electronic structure methods. To remedy this, continued fragmentation of the largest fragments was suggested. Alternatively, the interaction energy of the central core water with the rest of the cluster was considered. When such an interaction is summed up over all the waters in the cluster (taking proper account of multiple counting of interactions) then added to the individual isolated water monomer energies, the total energy of the water cluster may be recovered. The advantage of focusing on the interaction energy is

that interactions between the central water and water molecules at long-range from it may be very accurately determined via the use of multipoles and perturbation theory. In this case, the quadratic scaling in size of the fragment as one moves out from the central water becomes irrelevant because it is the interaction of the central water with these outer spherical shell fragments where multipoles may be utilised.

It was found that the interaction energy between the central water and the rest of the spherical cluster could be adequately reproduced, but again, only when the interacting waters were embedded in a Coulomb field of point charges. In this case, the largest size fragments requiring accurate electronic structure calculations were reduced to about 50 – still too large for practical computation. Careful examination of the largest fragment revealed that only a small portion of it really necessitated the use of an *ab initio* calculation. It should be possible to reduce the necessary 50 water cluster interaction energy to an *ab initio* calculation involving only about 9 waters – a size that is well within the reach of the most accurate electronic structure methods. The remaining interaction between the central water and rest of the 41 water cluster could be determined via multipoles and perturbation theory.

In fact, we have stated throughout this chapter that the use of multipoles and perturbation theory "can" and "should" be able to accurately reproduce interaction energies at long-range. We have even provided an estimate of what is meant by "long-range" without very much verification or testing. Indeed, the interaction energy of a water cluster can be broken down into a series of fragment interaction energies denoted as "two-body" (dimer fragments), "three-body" (trimer fragments), "four-body" (tetramer fragments), etc., as described in section 2.3 Of particular importance is knowing exactly under what conditions multipoles may

be used accurately to describe three-body and higher interactions. We note that little work has been done in establishing the above conditions. In the following chapter, we investigate the ability of perturbation theory with multipoles to accurately reproduce interaction energies. Furthermore, we attempt to discover *a priori* criteria for when multipoles should be accurate and, just as importantly, when individual interactions will be negligible.

# Chapter 4

# Energies of Water Clusters Using the Many-body Expansion

## 4.1 Introduction

In the previous chapter, we saw that it is possible to obtain sufficiently accurate total and interaction energies of large spherical clusters of water molecules by breaking the spheres down into spherical shells[q] then constructing spherical fragments from them. These spherical fragments, however, still contained too many water molecules to be able to accurately determine, i.e., using a high level of *ab initio* theory, their total energies on a practical time scale. We noted that by focusing on the individual interaction energies of a central water in a spherical cluster with the remaining cluster we could, in principle, eliminate the need to directly compute the total energies of the larger spherical clusters. Instead, the interactions between the central water and all the waters in a large spherical fragment could be computed via perturbation theory and multipoles.

For smaller spherical fragments, located closer to the central water, there is little recourse but to perform *ab initio* calculations because they are in close-contact with the central water. That is for waters in the first hydration shell. Those water molecules that are, say, in the second hydration shell and beyond should be sufficiently far enough away from the central water for perturbation theory and multipoles to be accurately applied. In this chapter, we will investigate how well

[q] Provided, or course, that the fragments derived from the shells are embedded in a charge field.

the latter treatment can describe these interactions. To approach this task we shall call upon the many-body expansion, described in section 2.3 By breaking down interactions into two-, three-, four-body etc. interactions, we immediately limit the size of fragments under consideration to dimers, trimer, tetramers, etc. Unfortunately, however, this comes at a cost.

The many-body expansion produces an explosively large number of such tiny fragment systems as the number of waters in the fragment is increased from one. Even though we may have considerably reduced the number of waters we need to consider at any one time, there are still a very large number of higher-body interactions we need to account for. For example, consider the largest and furthest spherical fragment at level 2. In the last chapter, we noted that sufficient accuracy is obtained for the interaction energy at this level of fragmentation. This spherical fragment contains shells 5, 6 and 7 and constitutes, on average, about 180 water molecules. The number of two-body interactions involving the central water and each of the waters in this fragment is 180. The number of three-body interactions is 16,110. The number of four-body interactions is 955,860. Obviously we need a means of reducing these numbers of calculations. For example, we expect that almost all, if not all, of the three and higher body interactions between the central water and all 180 waters in this far-away fragment to be negligible, and hence can be ignored. This we set out to prove in the sections that follow in which our attention is focused on the three-body interactions – the expected next largest and most significant interaction occurring in bulk water after two-body interactions.

Two-body interactions between a central water and the remaining waters are the farthest reaching and include all electrostatic, induction and dispersion effects. The number of such interactions is the least of all the many-body interactions possibly present in the system. In fact, the numbers involved are quite manageable and

entirely acceptable since this is the usual way of modeling water at present, i.e., include all significant two-body interactions. However, for high accuracy, at the very least, three-body interactions are required.

## 4.2 Trimers Characterised by Intermolecular Distances

At a fundamental level, three-body interactions are required to correct for the over/under counting of the always attractive two-body induction interactions. In a trimer, the sum of the three two-body interactions present sometimes may not be stable enough to account for electric field re-enforcement that may be occurring at each monomer due to the remaining two monomers. At times, the sum of the three two-body interactions present may be too stable due to electric field cancellation that may be occurring at each monomer due to the remaining two monomers.

In order to study three-body interactions, we require a means of classifying them so that some classes of interactions can be disregarded as negligible. The most obvious approach is to use the terms "close" and "far", where "close" means the waters are separated by a distance too short to be accurately described with perturbation theory and multipoles. That is, when two waters are close they are considered in "close-contact" and their interactions can only be accurately described by directly performing *ab initio* calculations. For two-body interactions, an extensive study on what "close" is, has appeared in the literature[30], so we shall use the conclusions of this work here. Figure 4-1 illustrates the definition of "close". If two water molecules are not close, they are therefore considered "far" from each other. Our definition of "close" is shown in Figure 4-1, illustrates that when two molecules are far from each other, no two atoms from each molecule, can be closer than 1.5 times the sum of their van der Waals radii. At "far" distances, dimer interactions should be nearly exactly[r] described with perturbation theory and multipoles[30].

[r] "exactly" here means the computed interaction energy using perturbation theory and multipoles agrees to within a few μ-Eh of the same interaction energy computed using an *ab initio* calculation.

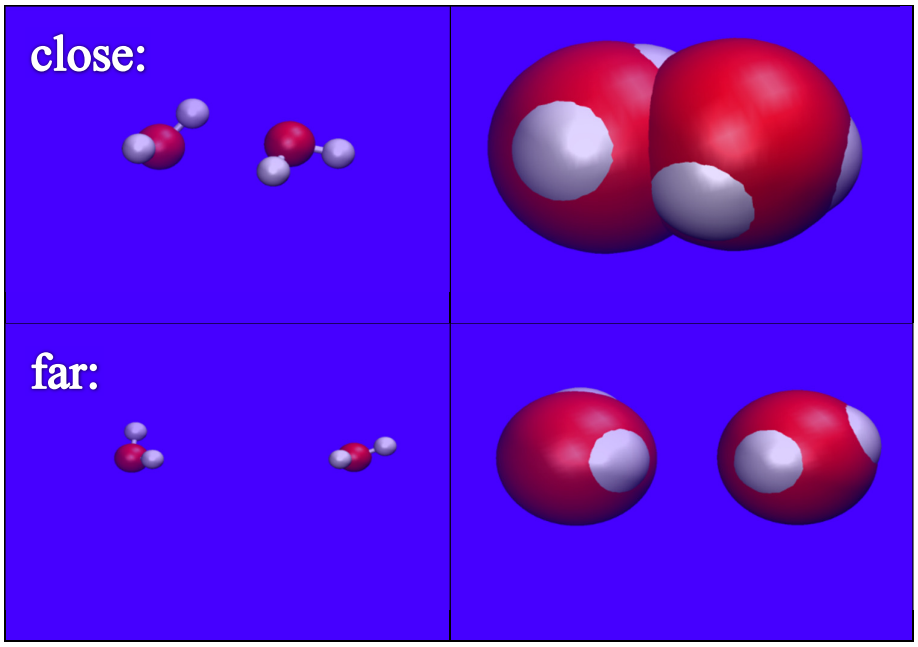



**Figure 4-1 Illustration of the definition of close and far distances**
The left-hand panels are the water configurations for the van der Waals space filling structures shown in the right-hand panels. However, the van der Waals radii used in the right-hand panels are 50% greater than the standard van der Waals radii (i.e., 50% greater than 1.2 and 1.5 Å for hydrogen and oxygen respectively). Water molecules are deemed close when the above space-filling models overlap (upper right panel), or far otherwise (lower right panel).

Where three-body interactions (or trimer) are concerned, there are three (and not one) distances that uniquely specify the configuration of their centers, i.e., $r_{12}$, $r_{13}$ and $r_{23}$ where we have labeled the three waters as 1, 2 and 3. As we are only concerned with close and far distances, i.e., whether $r_{ij}$ is close or far, we have a total of four possible close/far trimer configurations, as illustrated in Figure 4-2. Henceforth we shall assign a close distance between two waters with a "C" and a far distance with an "F". The four possible configurations that may be adopted by a trimer are: CCC, CCF, CFF and FFF – as the order of "C" and "F" does not matter. In the subsections that follow, we shall consider each type of configuration in turn and evaluate how well perturbation theory and multipoles reproduce the *ab initio* three-body interaction energy.

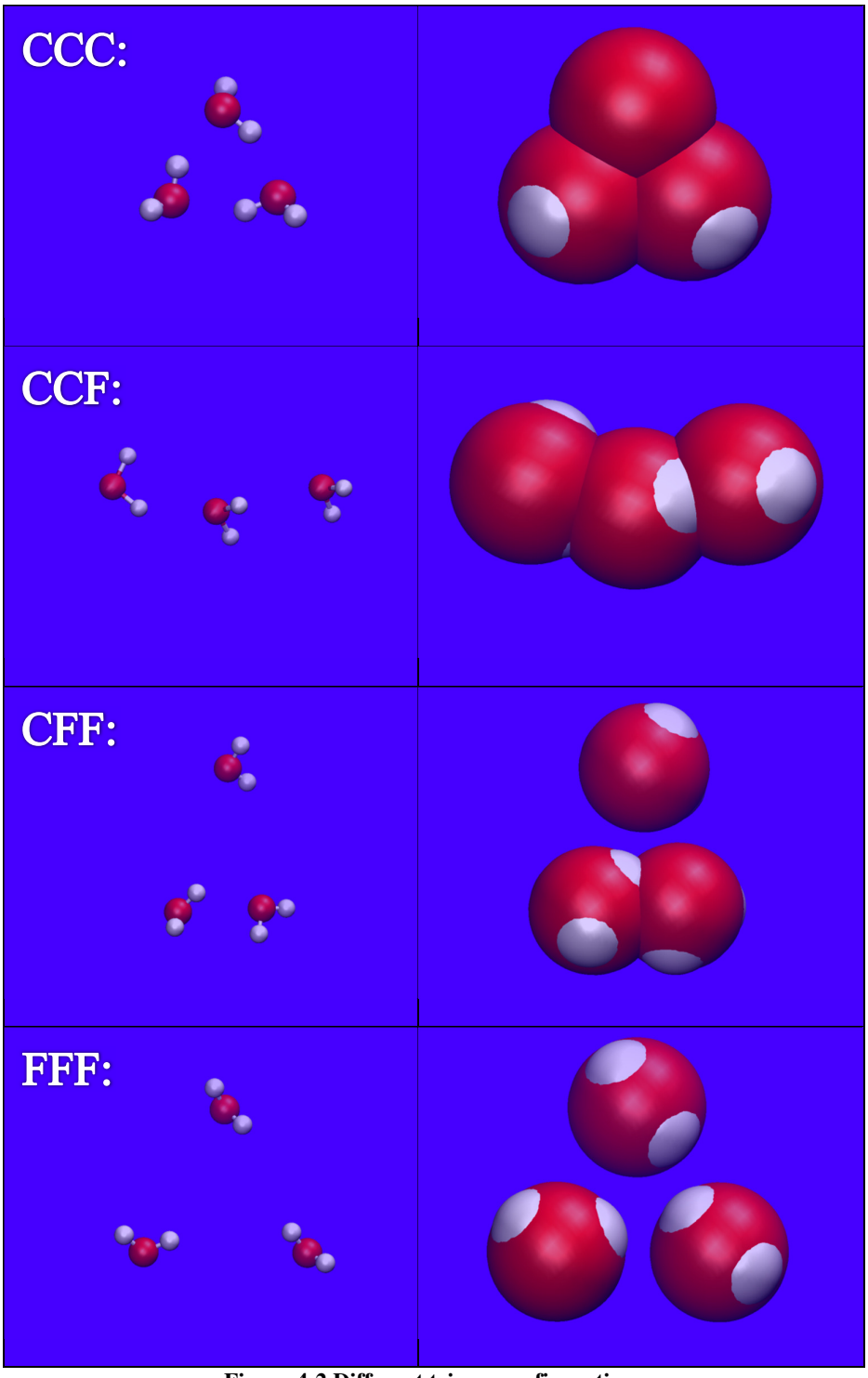



**Figure 4-2 Different trimer configurations**

The above four panels illustrate all the possible configurations of a trimer with regard to whether or not waters are “close” or “far” from one another. See Figure 4-1 for the definition of close and far. “C” represents “close” and “F” represents “far” and these designations refer to the three possible OO atom distances that exist in a trimer. The order of “C” and “F” is unimportant.

### 4.2.1 Computational Details

Large sets of trimer configurations were obtained from snapshots of water configurations from the same Monte Carlo simulations that were used to obtain the sample spherical clusters studied in Chapter 3. Trimers were selected by first designating any water in the snapshot as a central water, then choosing any other two water molecules that lie within a distance of ½ *l* (recall that *l* is the box-length in the Monte Carlo simulation) from the central water. Furthermore, the two additional waters selected should not be further from each other than ½ *l*. Trimers were selected in this way because it is only these interactions that require explicit calculations in any bulk water simulation, and water molecules further apart than ½ *l* are considered to interact negligibly in said simulations. Based on the classification scheme describe above, each trimer was classified as one of CCC, CCF, CFF or FFF types.

*Ab initio* calculations were performed at the HF/aug-cc-pVTZ level using the MOLPRO suite of programs[32]. This basis set is considerably larger than that used in the previous chapter (6-31G*). The present basis set consists of 4 s gaussian type functions, 1 to represent the 1s atomic orbital (AO) of O, and 3 to represent the 2s AO. There are also 3 p gaussian type functions to represent each of the $p_x$, $p_y$ and $p_z$ (AO) of O, 2 d functions for each of the 5 d AO of O, and 1 f-type function for each of the f AO of O. For each H atom there are 3 s functions, 2 p functions and a set of d functions. Additionally, this basis set is "augmented" (the "aug" in the designation of the basis set). This means that on O atom, there are additional s, p, d and f functions which are quite "diffuse" (small gaussian exponent). For H there are additional s, p and d functions. Thus, the total number of basis functions representing each water is now 74. This basis set is considered

a larger basis set. Such a basis set provides for a much more realistic representation of the polarizability of water. As a consequence, induction plays a much more significant role in the three-body interactions than with the 6-31G* basis set (19 functions per water).

Three-body interactions computed via HF theory were compared with the same computed using perturbation theory and multipoles. The permanent multipoles (to rank 5) and polarizabilities (to rank 2) necessary for this comparison were generated using A. J. Stone's distributed multipole analysis[24a,25] as implemented in the program GDMA[33]. The perturbative/multipole calculations were performed using ORIENT[34].

Additionally, after considerable effort, it was realised that the three-body energies needed to have the effects of BSSE removed from them in order to provide useful comparisons to predictions of the three-body energies obtained from perturbation theory and multipoles. BSSE, or the basis set superposition error, arises because the wavefunction describing each monomer is not complete, i.e., at the Hartree-Fock limit. As such when an interaction energy is computed via equation (1), the energy of the complex is "contaminated" because some of the basis functions on each monomer in the complex are utilised in the electronic description of other monomers in the complex, rather than purely being used to describe the *interaction* between each monomer in the complex. The result is that the energies of the monomers in the complex are more stable than they would otherwise be when in isolation. Thus, BSSE leads to too stable interaction energies. The effect can largely be compensated for through the use of error cancellation. "Ghost" basis functions are purposefully added at the locations of missing monomers when the

monomer energies are computed and used in equation (1). This approach is known as the "counterpoise correction"[35]. It can be extended when computing a three body interaction. The three-body interaction for system *ABC* can be written in terms of the electronic energies of the trimer, dimers and monomers as

$$\varepsilon_{A,B,C} = E(ABC) - E(AB) - E(AC) - E(BC) + E(A) + E(B) + E(C) \qquad (83)$$

However to remove the effects of BSSE, the energy of a dimer needs to be computed in the presence of ghost basis functions located at the position of the monomer present in the timer, but not present in the dimer. Similarly, the energies of the monomers need to be computed in the presence of ghost basis functions located at the positions of the two other missing monomers. Without correcting for the effects of BSSE it was found the *ab initio* computed three-body interaction energies were wildly in error.

### 4.2.2 CCC Configurations

As indicated in Figure 4-2, these configurations are the very definition of "close-contact". As such, there is no possibility of a multipole description of the interaction to be accurate. Therefore, we did not compute these three-body interactions nor did we compare them to what would be estimated from a perturbative/multipole description. Such configurations occur between the central water and waters in the first hydration shell. In evaluating the interaction energy of a central water with the remaining waters in a spherical clusters, there is little recourse for these types of interactions but to compute them via *ab initio* calculations.

### 4.2.3 CCF Configurations

Figure 4-2 also shows a typical configuration of this type. This configuration should also probably be considered as "close-contact" because it represents an approximate linear string of three monomers. Thus, the central monomer is close to the two terminal monomers (the CC in CCF) in the string, with only the two terminal monomers far from each other. When computing the three-body interaction with multipoles, we notice from equation (8) that two of the three two-body interactions involve a distance between pairs of monomers that are close. The computed two-body interactions from these close pairs are not expected to be accurate, and as such the three-body interactions is likely to be the least accurate of all the configuration types studied here.

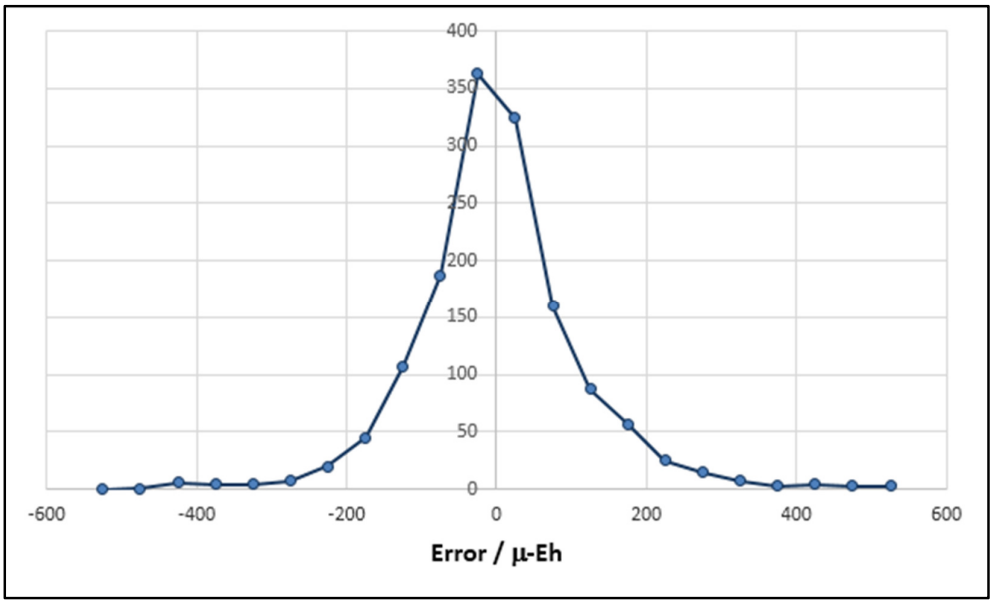



**Figure 4-3 Histogram of errors for CCF configurations**
Distribution of errors between counterpoise correct HF/aug-cc-pVTZ three-body interaction energies and that predicted using distributed multipoles (to rank 5) and central polarizabilities (to rank 2) including terms in the multipole expansion interaction energy up to $R^{-6}$. The vertical axis is the number of configurations observed with the corresponding error indicated on the horizontal axis. The total number of configurations studied was 1437.

This is illustrated in Figure 4-3 and Table 4-1. The ratio of the standard deviations of the three-body interaction energy to the error in the prediction of the interaction

energy is only 2.4, which implies that we are almost 95% confident at being able to predict the interaction at all. The discrepancy between the actual interaction and the prediction is too large. We conclude that along with the CCC configurations, the CCF configurations should also be considered as close-contact and be computed directly with *ab initio* calculations.

**Table 4-1 Mean Absolute Interaction Energy and Deviation for Different Three-body Configurations**

| | **CCF / μ-Eh** | **CFF / μ-Eh** | **FFF / μ-Eh** |
|---|---|---|---|
| Mean Absolute Interaction Energy | 246 | 105 | 38 |
| $\sigma_{\mathrm{MAIE}}$ | 205 | 89 | 29 |
| Mean Absolute Deviation | 80 | 26 | 3 |
| $\sigma_{\mathrm{MAD}}$ | 86 | 26 | 2 |
| Number in sample | 1437 | 1024 | 293 |

### 4.2.4 CFF Configurations

CFF configurations necessarily require one monomer to be located far from a close-contact dimer, as shown in Figure 4-2. The existence of the close-contact dimer means that the two-body interaction computed for it using perturbation theory and multipoles will not be particularly accurate. However, the remaining two two-body interactions should be well predicted. Of course, in the three-body interaction itself there exists a close-contact pair so there will be some error cancellation when applying equation (8), which should be more effective than in the CCF configurations. All considered we expect that errors in the predicted three-body interaction to be smaller than that of the CCF configurations.

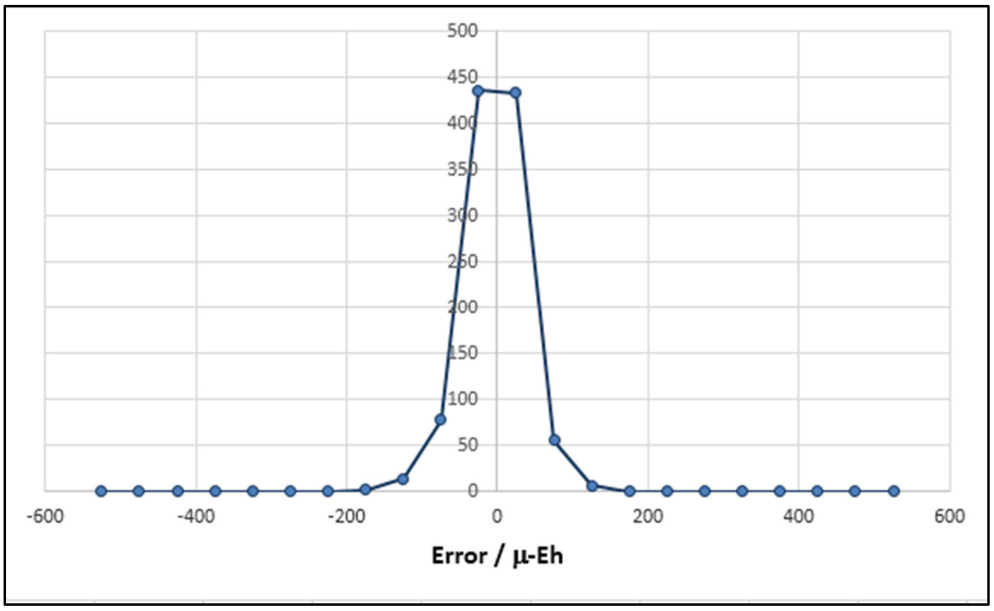



**Figure 4-4 Histogram of errors for CFF configurations**
Distribution of errors between counterpoise correct HF/aug-cc-pVTZ three-body interaction energies and that predicted using distributed multipoles (to rank 5) and central polarizabilities (to rank 2) including terms in the multipole expansion interaction energy up to $R^{-6}$. The vertical axis is the number of configurations observed with the corresponding error indicated on the horizontal axis. The total number of configurations studied was 1024.

The distribution of errors is indicated in Figure 4-4, and when considered in combination with the results in Table 4-1, it is clear that significantly better predictions can be made of the three-body interaction compared to CCF. The ratio of the standard deviations of the three-body interaction energy to the error in the prediction of the interaction energy now is 3.4 clearly indicating a statistically significant improvement in the three-body interaction energy prediction. While better predicted than CCF configurations, the errors here are still somewhat large with a MAD of 25 μ-Eh.

### 4.2.5 FFF Configurations

No pair of water molecules in this configuration are close. Perturbation theory plus multipoles are expected to predict very well the three-body interaction energy, and the results do not disappoint. Figure 4-5 shows what is essentially a delta function centered at zero error. The MAD is a mere 3 μ-Eh, as indicated in Table 4-1. The

ratio of the standard deviations of the three-body interaction energy to the error in the prediction of the interaction energy is 14.5 indicating complete reliability of the predicted interaction energies. The prediction are so accurate, in fact, that they are essentially exact. There is a tiny systematic error in the predicted interactions because the counterpoise correction tends to slightly over compensate for BSSE. Only a relatively small number of FFF configurations (293) were studied in comparison to the CFF (1024) and CCF (1437) configurations because it became very clear during the calculations that perturbation theory plus multipoles were fully capable to predicting the interactions with high accuracy.

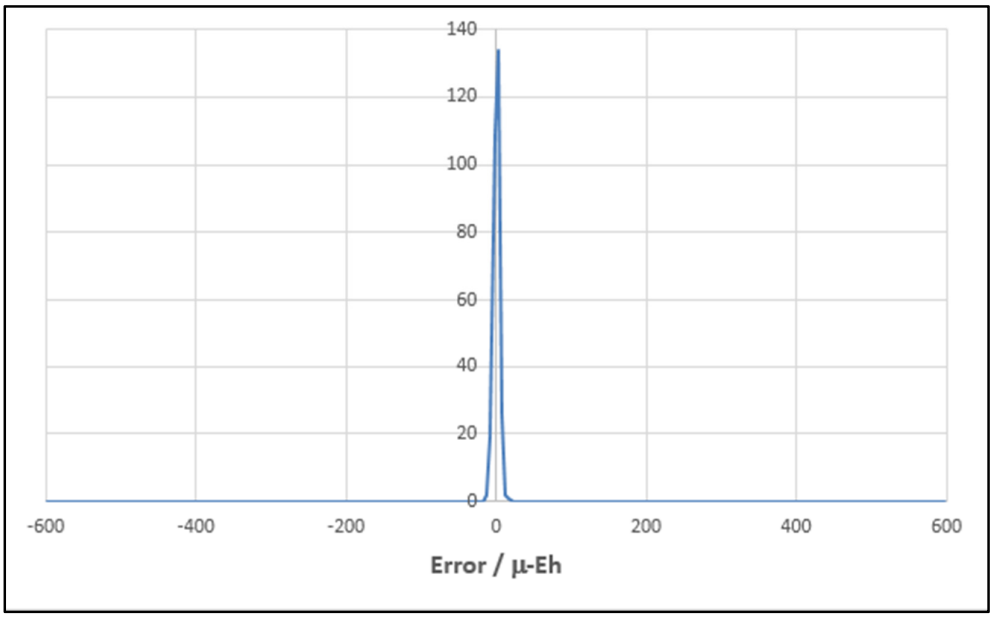



**Figure 4-5 Histogram of errors for FFF configurations**
Distribution of errors between counterpoise correct HF/aug-cc-pVTZ three-body interaction energies and that predicted using distributed multipoles (to rank 5) and central polarizabilities (to rank 2) including terms in the multipole expansion interaction energy up to $R^{-6}$. The vertical axis is the number of configurations observed with the corresponding error indicated on the horizontal axis. The total number of configurations studied was 293. Unlike the previous two figures, individual data points are not shown due to their high density.

## 4.3 Trimers Characterised with Shell Sums

During the course of this work a paper appeared in the literature in which water dimers through to the pentamer, 13-mer and 17-mer were studied using hybrid

DFT. While such works are not at all unusual, in this particular study the authors suggested using "shell sums" as a means of characterizing three-body interactions[36]. Their original suggestion was to determine all three O–O atom distances in a trimer and then select the shortest two. A "shell" is assigned to each of these distances then these "shells" are added. If an O–O distance was less than 3.1 Å (the approximate first hydration shell), then the distance was assigned "shell" 1. If the distance was longer than 3.1 Å then it was assigned "shell" 2. The authors considered timers with a shell sum (SS) of 2 to be close-contact. This designation is identical to both our CCC and CCF configurations in the previous section if we had of used an O–O atom distance criterion of 3.1 Å. Indeed, inspection of Figure 4-1 and Figure 4-2 shows that each monomer can be essentially represented by a single sphere of radius 50% larger than the van der Waals radius of oxygen, which is 2.25 Å. A designation of "C" was given if the spheres of two monomers overlapped, i.e., if the O–O atom distance was < 4.5 Å. A SS of 3 corresponds to our CFF configurations and a SS of 4, an FFF configuration.

Rather than use the above SS definition which is essentially identical to our close-far designations except that a considerably smaller distance criterion was used, we decided to assign a distance between monomers to a specific hydration shell and then sum these. Thus, shells can be any number from 1 to infinity (rather than just 1 or 2), and the SS can be anything from 2 to infinity. The distances we choose for the assignment were: O–O atom distances $\leq 3.50$ Å, shell 1. Then for every 2.25 Å further distant an additional shell was added. We note here that while the SS does allow examining three-body interactions as a function of SS beyond 4, it is

not as selective for the particular interactions as the CCC, CCF, CFF and FFF designation.

In light of this alternative characterization of the three-body interactions, the data set of the previous section was reanalyzed and distances between oxygen atoms used to classify an interaction according to its SS value. Note that we did not perform any calculations on CCC trimers because there is no chance that multipoles can accurately describe these interactions. Thus for our SS = 2 all CCC configurations were excluded, and only CCF were considered. The results for the root-mean-square (RMS) three-body interaction energy (counterpoise corrected) as a function of SS is given in Figure 4-6, while the RMS deviation (RMSD) between the *ab initio* and multipole computed three-body interaction energies are presented in Table 4-2.

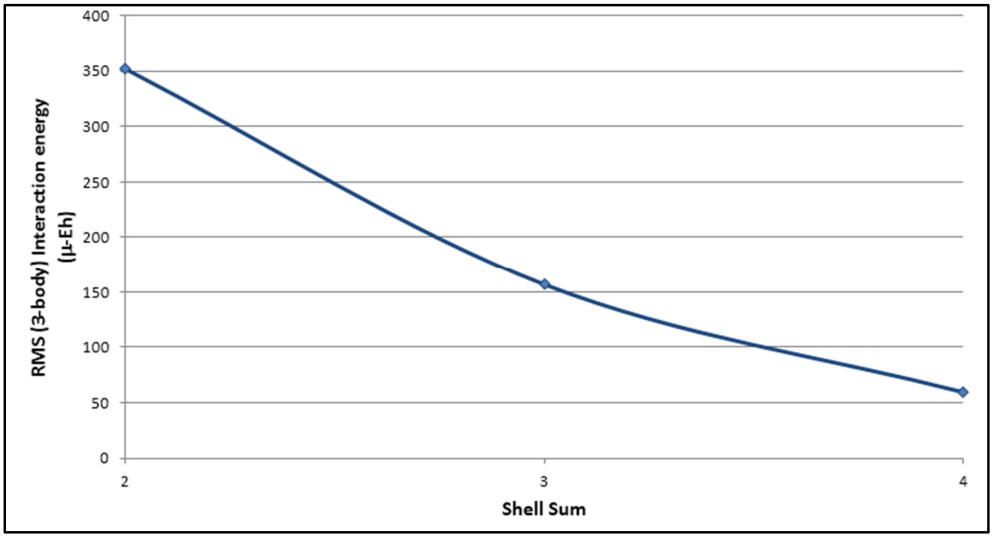



**Figure 4-6 RMS three-body interaction energy versus shell sum**
The root-mean-square (RMS) *ab initio* three-body interaction energy versus shell sum (SS). Note that SS = 2 does not include any CCC configurations, but only CCF configurations.

**Table 4-2 RMS Interaction Energy and Deviation for Different Shell Sums**

| SS | N | RMSI / μ-Eh | RMSD / μ-Eh |
|---|---|---|---|
| 2 | 1091 | 353 | 132 |
| 3 | 1232 | 157 | 45 |
| 4 | 431 | 59 | 7 |

It is apparent from Table 4-2 that a perturbative and multipole treatment for SS greater than or equal to 4 is essentially exact. Errors are slightly higher due to the shorter distances used in the SS treatment compared to the C/F treatment of the previous section. Nevertheless, our overall conclusions are that they are fairly similar. SS = 2 requires *ab initio* evaluation. SS = 3, due to the closer distances involved here are also too high, so a longer distance would be necessary in order to ensure sufficient accuracy in the multipole treatment. At SS = 4 and beyond, the three-body interaction energies are reproduced almost exactly.

As an aside, which does deserve mentioning at this point, we thought it was prudent to test the convergence of the three-body interaction as a function of basis set size. While we expected that the aug-cc-pVTZ basis set to most definitely have reached convergence in this interaction energy (i.e., we are essentially at the Hartree-Fock limit, cf. section 2.4.5 ), it is interesting to see whether a smaller basis set (and therefore less expensive) could have also yielded similar results. To this end we examined the RMS three-body interaction as a function of SS for different basis sets ranging from the smallest, STO-3G, or minimal basis set (7 basis functions per water), to the very expensive, aug-cc-pVQZ (172 basis functions per water). The results are shown in Figure 4-7.

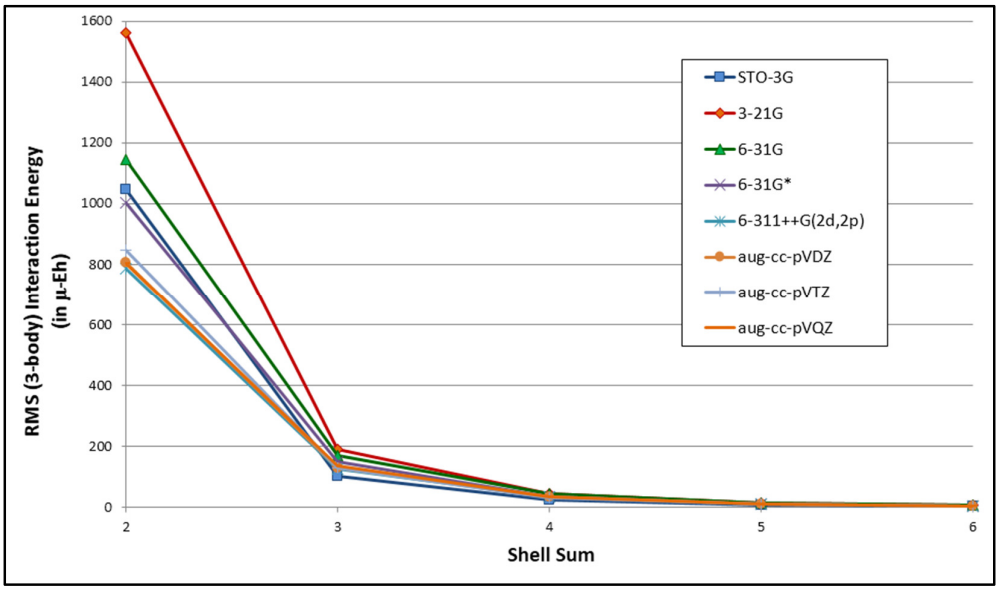



**Figure 4-7 Effect of basis set on the three-body interactions**
Basis sets, from the smallest possible (STO-3G, top of legend), to the very large and expensive (aug-cc-pVQZ, bottom of legend) were used to compute the same set of three-body interaction energies as a function of shell sum. This was performed to test for convergence in the three-body interaction as a function of basis set. See text for details.

It is noted that the RMS three-body interaction settles down to a consistent result using the aug-cc-pVDZ (41 basis functions per water) or the "triple zeta" basis set 6-311++G(2d,2p) (47 basis functions per water). We also note that at SS $\geq$ 5 the RMS three-body interaction is extremely small and can effectively be considered negligible.

## 4.4 Trimers Characterised by a Single Distance Measure

We have shown in the previous two sections that a perturbation/multipole treatment can provide virtually exact three-body interaction energies compared to an *ab initio* calculation provided the waters are far enough from each other. Is there a more rigorous condition that we can use to establish just when a three-body interaction is significant? It turns out that it is possible, by considering the expressions for the induction interaction found in equations (69) to (71). We shall consider specifically:

(i) only the first order (and therefore most significant) induction energy,

(ii) only the dipole-dipole polarizability,

(iii) only the permanent dipoles on molecules *A* and *B* which we shall take as equilibrium water so that

(iv) the only non-zero component of the dipole is $Q_{10}$ which we shall represent as $\mu$.

(v) Finally, we note that for equilibrium water in its usual axis system, the polarizability tensor is diagonal, and we shall approximate it with an isotropic $\alpha$.

In this case, the first order induction energy at water *A* due to water *B* is

$$U_{\text{ind}}^{A(1)B} = -\frac{\alpha}{2}\sum_{\kappa} V_{1\kappa}^{A}(B)V_{1\kappa}^{A}(B) \tag{84}$$

Here $V_{1\kappa}^{A}(B)$ means component $\kappa$ of the gradient of the electrostatic potential *at* water *A* *due to* water *B*.

Upon introducing a third water molecule, *C*, the potential gradient at *A* now due to *B* and *C* is

$$V_{1\kappa}^{A}(B,C) = V_{1\kappa}^{A}(B) + V_{1\kappa}^{A}(C) \tag{85}$$

which we can directly substitute into equation (84) to obtain the first order induction energy at *A* due to waters *B* and *C*

$$U_{\text{ind}}^{A(1)\{B,C\}} = -\frac{\alpha}{2}\sum_{\kappa}[V_{1\kappa}^{A}(B) + V_{1\kappa}^{A}(C)][V_{1\kappa}^{A}(B) + V_{1\kappa}^{A}(C)] \tag{86}$$

Because we are interested in the three-body interaction, we notice upon expanding equation (86) we have the additional terms,

$$-\alpha \sum_{\kappa} V_{1\kappa}^{A}(B) V_{1\kappa}^{A}(C) \tag{87}$$

over and above the two-body terms, $U_{\text{ind}}^{A(1)B}$ and $U_{\text{ind}}^{A(1)C}$. These additional terms, must literally be the first order three-body interaction energy.

As we note from equation (71), the potential gradient at *A* due to any water *X* is

$$V_{1\kappa}^{A}(X) = T_{1\kappa 10}^{A}(X)\mu \tag{88}$$

which possess a distance dependence between *A* and *X* of

$$V_{1\kappa}^{A}(X) = \frac{T'^{A}_{1\kappa 10}(X)\mu}{R_{AX}^{3}} \tag{89}$$

Upon substitution of equation (89) into (87) we find the three-body interaction energy at *A* due to *B* and *C* to be

$$-\frac{\alpha\mu^{2}}{R_{AB}^{3}R_{AC}^{3}} \sum_{\kappa} T'^{A}_{1\kappa 10}(B) T'^{A}_{1\kappa 10}(C) \tag{90}$$

Directly analogous expressions exist for the three-body induction energy at water *B* due to waters *A* and *C* and at water *C* due to waters *A* and *B*.

The largest in magnitude value of the three-body interaction energy here is when all three dipoles are aligned along their respective $\mathbf{R}_{AB}$, $\mathbf{R}_{AC}$ and $\mathbf{R}_{BC}$ vectors. In this case, $T'^{A}_{1\kappa 10}(X)$, $T'^{B}_{1\kappa 10}(X)$ and $T'^{C}_{1\kappa 10}(X)$ are all equal to 2. Thus the maximum absolute three-body interaction between waters *A*, *B* and *C* is

$$\left|\varepsilon_{A,B,C}\right|_{\max} = 4\alpha\mu^{2}\left[\frac{1}{R_{AB}^{3}R_{AC}^{3}} + \frac{1}{R_{AB}^{3}R_{BC}^{3}} + \frac{1}{R_{AC}^{3}R_{BC}^{3}}\right] \tag{91}$$

or

$$\left|\varepsilon_{A,B,C}\right|_{\max} = \frac{12\alpha\mu^{2}}{L^{6}} \tag{92}$$

where, *L*, is given by

$$L = 3^{1/6}\left[\frac{1}{R_{AB}^3 R_{AC}^3} + \frac{1}{R_{AB}^3 R_{BC}^3} + \frac{1}{R_{AC}^3 R_{BC}^3}\right]^{-1/6} \quad (93)$$

The factor of $3^{1/6}$ is present simply to ensure that when all the $R_{ij}$ distances are equal, $L = R_{ij}$. With this definition of *L*, $\left|\varepsilon_{A,B,C}\right|_{\max} \propto L^{-6}$, and we note that small values of *L* indicate that a three-body interaction could be very significant, whereas configurations with large values of *L* are expected to produce tiny three-body interactions.

In order to test this analysis, we extracted *all* possible trimers from a 57-mer water cluster published in ref. 37. There was a total of $\binom{57}{3} = 29{,}260$ such configurations. We computed all three-body interactions at the HF/aug-cc-pVDZ level of theory using the Gaussian09 suite of programs[21]. The cluster possesses slightly different geometries for each water monomer, so the $\alpha$ and $\mu$ used in the above analysis were for the equilibrium structure of water at this level of theory. The optimised parameters are given in Table 4-3. No counterpoise correction was applied to the computed three-body energies, so as such there is expected to be significant BSSE present in close-contact configurations.

**Table 4-3 Equilibrium Water Properties at the HF/aug-cc-pVDZ Level**

| Parameter | Value |
|---|---|
| $r_{OH}$ / Å | 0.9437 |
| ∠HOH / deg | 105.99 |
| $\alpha$ / Å$^3$ | 1.18 |
| $\mu$ / D | 1.96 |
| *E* / Eh | –76.0418435 |

Rather than use distances measured between O atoms, we chose to measure distances between the nuclear centres of charge for each monomer as this is the default origin in the *ab initio* calculations at which all central multipoles and polarizabilities are computed. The location of the nuclear centre-of-charge is

slightly displaced off the O nucleus in the direction of the two hydrogens, but makes little difference to our definitions of close-contact discussed in section 4.2 If we choose a 4.5 Å centre-of-charge separation between two monomers as "close" then the number of CCC, CCF, CFF and FFF configurations found in the 57-mer is summarised in Table 4-4.

**Table 4-4 Summary of Configurations Found in the 57-mer**

| Configuration | Number | Total 3B / m-Eh |
|---|---|---|
| CCC | 356 | –36.1 |
| CCF | 1,218 | –51.1 |
| CFF | 10,576 | –6.5 |
| FFF | 17,110 | 2.0 |
| Total | 29,260 | –91.7 |

We can see from Table 4-4 that there are 1,574 close-contact configurations (CCC and CCF configurations) in the 57-mer, for which perturbation theory and multipoles have no chance of predicting the three-body interaction energy accurately. This constitutes only 5.4% of the total number of three-body configurations, yet apparently accounts for the bulk of the three-body interaction energy. Of course, we do expect that the close-contact configurations to possess the most significant three-body interactions, but caution should be followed in placing too much weight on just how significant that is, based on the above numbers, because BSSE does play a significant role in these particular interaction energies. BSSE will not substantially contribute to the three-body interaction energy of the more distant CFF configurations, and it will make virtually no contribution at all in the FFF configurations.

Figure 4-8 illustrates all the 29,260 three-body interactions present in the 57-mer as a function of *L*. Note that the interaction energy units are in μ-Eh. Included in

the figure are all the close-contact configurations. The orange curves are derived from equation (92), i.e., no point should lie above the positive going curve, nor should any point be below the negative going curve (which is just the negative of $\left|\varepsilon_{A,B,C}\right|_{\max}$). Evidently there are such points, but this is entirely due to the fact that these recalcitrant configurations are close-contact and their interaction energies are contaminated by BSSE. Re-plotting Figure 4-8 but with only CFF and FFF configurations and changing the scale appropriately produces Figure 4-9.

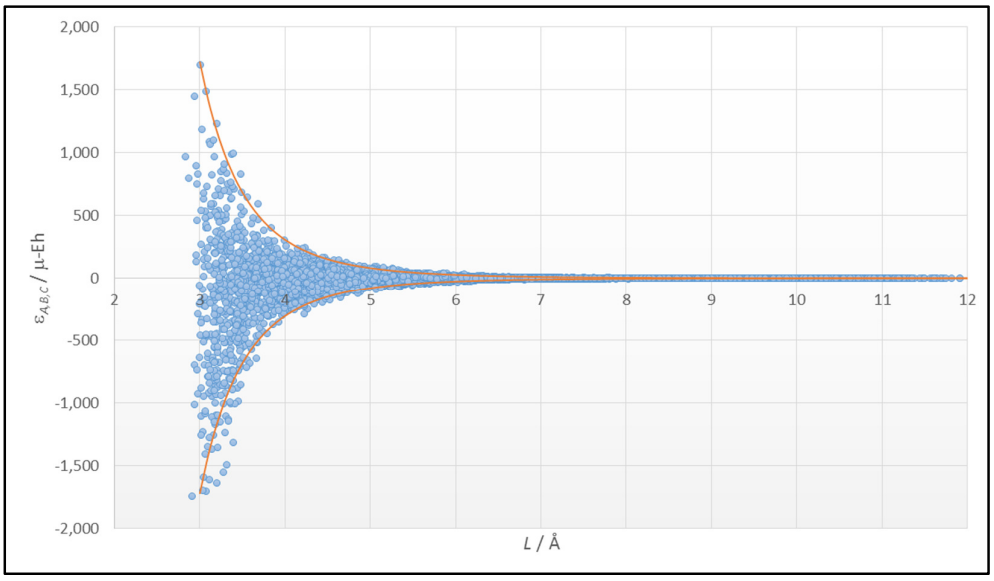



**Figure 4-8 All 3B interactions in the 57-mer as a function of *L***
Each point in the above figure is one of the 29,260 three-body interactions present in the 57-mer. The orange curves were obtained from equation (92) – no points should lie above the positive going curve nor any point below the negative going curve. Evidently some points do, but for good reason. See the text for the discussion.

Examination of Figure 4-9 immediately reveals that not a single three-body interaction in all of the 27,686 CFF and FFF configurations exceeds $\left|\varepsilon_{A,B,C}\right|_{\max}$ or is less than $-\left|\varepsilon_{A,B,C}\right|_{\max}$. Additionally, the data Figure 4-8 and Figure 4-9 can be re-expressed as plots of $\left|\varepsilon_{A,B,C}\right|$ versus $\left|\varepsilon_{A,B,C}\right|_{\max}$. These graphs are shown in Figure 4-10 and Figure 4-11 respectively. Again, after removing the close-contact configurations, Figure 4-11 illustrates just how well $\left|\varepsilon_{A,B,C}\right|_{\max}$ can be used as a

criterion to eliminate with absolute certainty three-body, FFF and CFF interactions which can at the very most only contribute $\left|\varepsilon_{A,B,C}\right|_{\max}$ to the total three-body interaction energy.

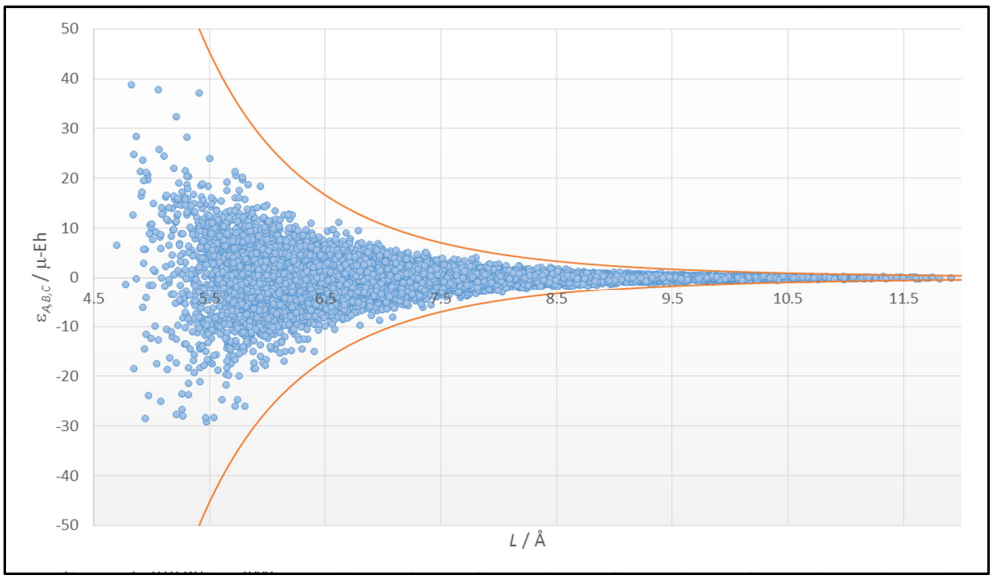



**Figure 4-9 CFF and FFF 3B interactions in the 57-mer as a function of *L***
Each point in the above figure is one of the 27,686 CFF and FFF three-body interactions present in the 57-mer. The orange curves were obtained from equation (92) – no points should lie above the positive going curve nor any point below the negative going curve. Note also that no CFF nor FFF configuration contributes more than 100 μ-Eh to a three-body configuration.

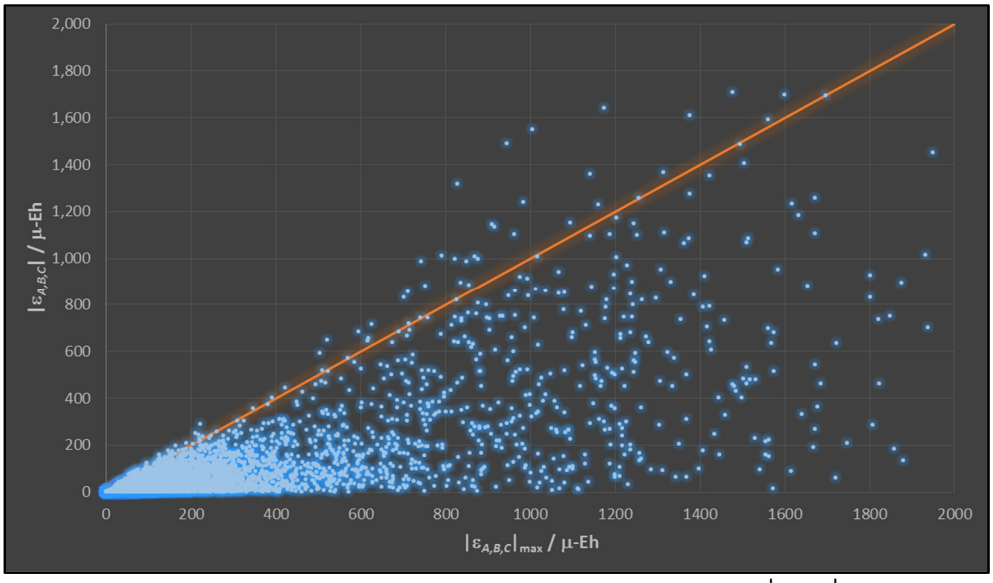



**Figure 4-10 All 3B interactions in the 57-mer as a function of $\left|\varepsilon_{A,B,C}\right|_{\max}$**
Each point in the above figure is one of the 29,260 three-body interactions present in the 57-mer. The orange line was obtained from equation (92) – no points should lie above it. Evidently some points do, but for good reason. See the text for the discussion.

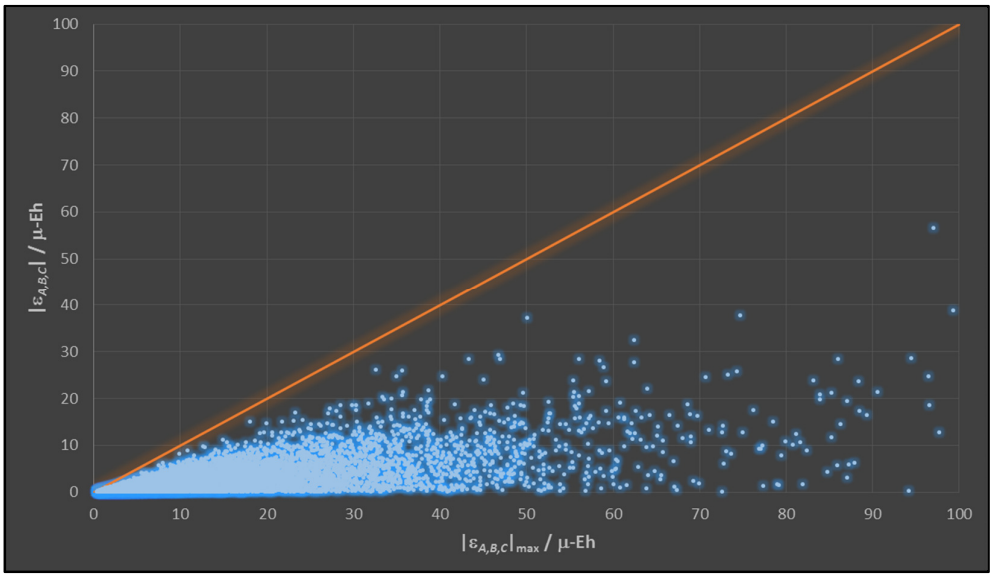



**Figure 4-11 CFF and FFF 3B interactions in the 57-mer as a function of $|\varepsilon_{A,B,C}|_{\max}$**
Each point in the above figure is one of the 27,686 CFF and FFF three-body interactions present in the 57-mer. The orange line was obtained from equation (92) – no points should lie above it. Note also that no CFF nor FFF configuration contributes more than 100 μ-Eh to a three-body configuration.

To this end, Table 4-5 illustrates the error per monomer associated with ignoring small three-body energy contributions based on equation (92). We can see that, while there is a fortuitous cancellation of opposite going on in the three-body interactions at an excluded energy of <200 μ-Eh, excluding all three-body contributions with $|\varepsilon_{A,B,C}|_{\max}$ predicted to be less than 50 μ-Eh which removes 22,421 three-body evaluations without introducing significant error. The remaining configurations are close-contact and 5,265 CFF and FFF (mostly CFF) configurations that may be evaluated efficiently and rapidly using perturbation theory coupled with multipoles, as described in the previous sections. Thus we have achieved all the three-body interactions (excluding CCF) without any significant loss in accuracy, by eliminating to evaluate 77% of the numerous 22,421 configurations.

**Table 4-5 Summary of Error Incurred when Excluding Various Three-body Interactions**

| Exclude 3B < / μ-Eh | $\varepsilon_{3B}$ / m-Eh | Error per monomer / μ-Eh | *N* excluded | *N* included |
|---|---|---|---|---|
| 0.1 | –91.73 | 0 | 0 | 29,260 |
| 1 | –91.73 | 0 | 296 | 28,964 |
| 10 | –91.83 | –2 | 11,709 | 17,551 |
| 20 | –91.98 | –5 | 16,564 | 12,696 |
| 40 | –93.09 | –24 | 21,090 | 8,170 |
| 50 | –93.10 | –24 | 22,421 | 6,839 |
| 100 | –91.73 | –56 | 25,493 | 3,767 |
| 200 | –91.73 | 0 | 27,463 | 1,797 |
| 400 | –88.56 | 56 | 28,451 | 809 |
| 500 | –87.70 | 71 | 28,604 | 656 |
| 1000 | –43.45 | 847 | 29,034 | 226 |
| 2000 | 0.03 | 1610 | 29,257 | 3 |
| 4000 | 0.00 | 1609 | 29,260 | 0 |

## 4.5 Summary

In this chapter, we have shown that three-body interactions can be divided into two categories, close-contact, which must be evaluated directly via *ab initio* calculations, and interactions that may be evaluated with reasonable accuracy using perturbation theory and multipoles. We further categorised the latter interactions into configurations labeled as CFF (or shell sum 3) or FFF (shell sum 4+). CFF configurations were of the kind depicted in Figure 4-2 and posed the most challenge for perturbation theory and multipoles at accurate prediction. We showed that for the FFF configurations, the three-body interaction energies can essentially be exactly reproduced. An alternative approach to using perturbation theory and multipoles for evaluating the CFF three-body interaction is to treat the interaction as a pseudo-two-body interaction, i.e., an interaction of a dimer, computed via *ab initio* calculations, and a monomer. Because the monomer is located far from the dimer, perturbation theory and multipoles should essentially

be able to reproduce this interaction exactly. However, this possibility remains open for further study.

In the last section of the chapter, we developed a simple treatment that can be used as a criterion to determine precisely whether a three-body interaction could be important. While simple, the approach proved to be particularly powerful whereby in the study of a 57-mer water cluster we were able to eliminate 77% of all three-body interaction evaluations, i.e., some 22,421 of them, without any significant loss in accuracy. In a larger cluster, like those studied in Chapter 3, even more significant savings could be obtained because vast numbers of three-body interactions involve small FFF configuration interactions. Lastly we note for future work that the simple theoretical treatment developed here to obtain an expression for $\left|\varepsilon_{A,B,C}\right|_{\max}$ could readily be extended to four and perhaps higher body interactions.

# Chapter 5

# Conclusion and Future Work

In the first part of this thesis, we examined what was necessary to produce a highly accurate and computationally efficient model of bulk water from first principles. We saw that such models utilise periodic boundary conditions, which require a central box containing water to be at least several hydration spheres in size. We estimated that such a box would need to have the physical dimensions of around 20 Å in length on each side. In models of bulk water using this cube under periodic boundary conditions, explicit water interactions need to be computed between all waters within the central box. Thus we need to be able to determine accurately and efficiently the total energy, or interaction energy, of all waters within a spherical cluster of radius $r \approx \frac{20}{2} = 10$ Å.

A spherical cluster of radius 10 Å in a sample of bulk water at 25°C contains about 140 water molecules. Such a water cluster is considered huge from the point-of-view of modern *ab initio* calculations. Put plainly, it is simply impossible now and in the near (or even distant) future to be able to compute the energy of such a cluster with the "gold standard" *ab initio* computational method, i.e., CCSD(T)/aug-cc-pVTZ. Nevertheless, we have undertaken in this thesis ways of obtaining accurate approximations to the *ab initio* total energies, or interaction energies, of such large clusters within a practical time-frame.

In Chapter 3, we applied the method of energy-based molecular fragmentation in order to estimate both the total energy and interaction energy of large water clusters. This method has the advantage of being a linear-scaling technique so, in

principle, fragmenting into smaller systems can be utilised to overcome the steep scaling in computational expense with system size when high level *ab initio* calculations are required. However, in this work we deviated from the more usual approach to the fragmentation methodology and fragmented the spherical water cluster into spherical shells. We found that once we incorporated a charge-field into the calculations of the fragment shells, we were able to achieve highly satisfactory agreement between the *ab initio* energy, or interaction energy, of the full cluster, and the same but using the fragmentation energy formula.

Unfortunately, the number of molecules within a spherical-shell fragment increases quadratically with distance from the centre of the cluster. The consequence of this is that for spherical fragments involving the outer spherical shells far too many water molecules are contained within the fragment. That is, these large fragments are still too large for high-level *ab initio* calculations. To remedy this issue, we proposed a three-stage fragmentation of the spherical cluster, with the first stage being fragmentation into spherical shells. The second stage involves fragmentation of the spherical shells into latitudinal zones, and the final stage involves fragmentation of the largest of these zones into quadrangles. This three-stage fragmentation potentially solves the large fragment issue, but we left its detailed testing for future work.

In Chapter 4, we examined an alternative approach to spherical-shell fragmentation. In this chapter, we considered utilisation of the many-body expansion in determining accurately the total energy, or interaction energy, of a spherical water cluster. The main drawback with applying the many-body expansion is that the system does not have to be particularly large before the

number of required, but small, calculations necessary becomes overwhelming. To help alleviate this issue, we focused in on which of the vast numbers of three-body interactions do we expect to be important (i.e., significant).

We considered two approaches to determine which three-body interactions are important. First of all, we classified water-water distances into "close" and "far" regimes. In a three-body interaction, or trimer, there are only three possible water-water distances, so any particular configuration of water can be classified as one and only one of the following types: (a) close-close-close (CCC), (b) close-close-far (CCF), (c) close-far-far and (CFF), and (d) far-far-far (FFF). These designations proved useful in determining whether perturbation theory plus a multipolar treatment (PTM) of three-body interactions could accurately reproduce the full *ab initio* three-body BSSE corrected interaction energy. The later point is important because if a PTM treatment can replace an *ab initio* calculation we have achieved a tremendous speed-up in computer time needed to eventually determine the total energy of the entire spherical cluster.

We found that only a relatively small number of calculations involved configurations of the type CCC and CCF. For these configurations a PTM treatment failed at accurately reproducing the three-body interaction. As such we considered these configurations to be "close-contact", and there is no alternative but to compute these three-body interactions using *ab initio* theory. On the other hand, the vast majority of three-body interactions arose from the CFF and FFF configurations. Here, PTM satisfactorily reproduced the three-body interactions.

Secondly, we considered in establishing which three-body interactions were significant, which can be readily generalised to four- and higher body interactions.

The above classification scheme using the "close" and "far" designations is appropriate for three-body interactions – only three inter-water distances exist. However for a four-body system, there now exists six possible inter-water distances. The number of such distances increasing quadratically with the number of waters under consideration. This makes *n*-body classification of significant configurations awkward. Furthermore, this classification does not differentiate amongst the different configurations within a particular classification, e.g., in which of the several FFF configurations are expected to be significant, compared with the other FFF configurations. Thus we introduced the use of the "*L*" single distance parameter in determining whether a three-body interaction could be significant. We found that after taking account of close-contact configurations, the *L* parameter well predicated potentially significant three-body interactions. These results enabled us to further reduce the computational effort involved in accurately reproducing an *ab* initio total energy, or interaction energy, of a large spherical water cluster by removing many PTM evaluations predicted to be insignificant. Future work in this particular line of research would involve extension of the "*L*" parameter to four-body interactions and then its evaluation and testing.

Thus, overall, we found that the one serious bottle-neck to obtaining the total three-body interaction energy for a spherical water cluster is the evaluation of the three-body interactions for the close-contact configurations. Although not investigated in this thesis, future work should involve obtaining accurate three-body interaction surfaces for close-contact water trimers. One possible approach here would be to use the method of Shepard interpolation to accurately describe this 21-dimensional surface. In doing so, the need to perform an *ab initio* calculations during a simulation can be completely removed. With this achieved, an accurate bulk water

simulation could be performed with results equal in quality to the "gold-standard" of *ab initio* calculations.

# Chapter 6

# A Collaboration Investigating H-bonding in Crystals

In this final chapter, we describe our contribution to the collaborative work published as ref. 38 entitled "Aryl-substituents moderate the nature of hydrogen bonds, N–H···N versus N–H···O, leading to supramolecular chains in the crystal structures of *N*-arylamino 1,2,3-triazole esters". The full publication can be found in the Supporting Publication in this thesis. Apart from the first four paragraphs of the introduction, which introduces supramolecular chemistry, much of the remainder of the chapter is extracted from the above cited work – work originally conducted by us.

## 6.1 Introduction

Supramolecular chemistry is the "chemistry beyond the molecule" and is the study of non-covalent interactions which is crucial to understanding many biological processes and systems. While traditional chemistry focuses on the bonds that hold atoms together in a molecule, supramolecular chemistry examines the weaker interactions that hold groups of molecules together. Important concepts that have been demonstrated by supramolecular chemistry include molecular self-assembly, folding, molecular recognition, host-guest chemistry, mechanically-interlocked molecular architectures and dynamic covalent chemistry.

The importance of supramolecular chemistry was underscored by the 1987 Nobel Prize for Chemistry, which was awarded to Donald J. Cram, Jean-Marie Lehn and

Charles J. Pedersen in recognition for their development and use of molecules with structure-specific interactions of high selectivity[s].

In 1978, Jean-Marie Lehn introduced the term "supramolecular chemistry" to generalise the early developments and layout future concepts and visions that resulted from an enhanced understanding and application of the non-covalent bond. He defined supramolecular chemistry as the chemistry beyond the molecule, bearing on the organised entities of higher complexity that result from the association of two or more chemical species held together by intermolecular forces. Today, a large fraction of papers (≈30%) published in the leading general chemistry journals such as *Angew. Chem.*, *Chem. Comm.*, *Chem. Eur. J.*, or *J. Am. Chem. Soc.* report on the practical realization of the concepts and visions in supramolecular chemistry which span from the core of chemistry to the interfaces of biology, physics, advanced materials and nanosciences. The impact on journals in the areas of advanced materials and nanomaterials is indeed larger.

The past decade has seen dramatic developments in the field, with supramolecular chemistry leaving its roots in classical host guest chemistry and expanding into exciting areas of materials chemistry and nanoscience with many real and potential applications. Supramolecular findings are evolving our understanding of the way chemical concepts at the molecular level build up into materials and systems with fascinating, emergent properties on the nanoscale.

The robust and directional nature of hydrogen bonding interactions makes these favourite supramolecular synthons prominent in the crystal engineer's toolbox.

[s] "The Nobel Prize in Chemistry 1987". Nobelprize.org. Nobel Media AB 2014. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1987/ (accessed on 2-Feb-2015).

However, not all molecular systems have hydrogen bonding functionality or, even if they do, hydrogen bonding may not extend in three dimensions. In these circumstances, weak intermolecular interactions naturally come to the fore, e.g., π···π, C–H···π, halogen bonding, etc. Such considerations make it imperative to study and understand these "second tier" supramolecular synthons to enable their control for supramolecular assembly. Desiraju identifies the study of intermolecular interactions as the "what" of crystal engineering and the first stage of the continuum leading to evaluating the influence of rational changes in molecular packing upon functional crystalline materials (the "why")[39]. Clearly, in order to determine the importance and prevalence of specific supramolecular synthons, systematic structural studies of closely related chemical species are required. Herein, an evaluation of a series of closely related structures is made (Figure 6-1 and Figure 6-2). Specifically, the influence of differences in the nature of the aryl-bound substituents upon intermolecular hydrogen bonding patterns is evaluated.

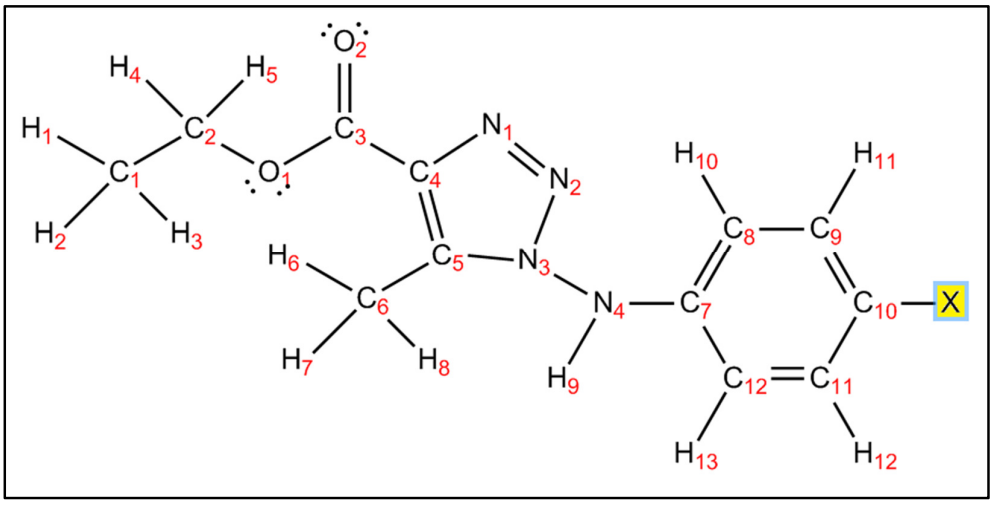



**Figure 6-1 The structures studied in this chapter**
The structures studied in this work is when X is replaced with H (**1**), F (**2**), Cl (**3**), Br (**4**) and I (**5**), O–$CH_3$ (**6**) and $NO_2$ (**7**). In addition, the structure found in Figure 6-2 was also studied.

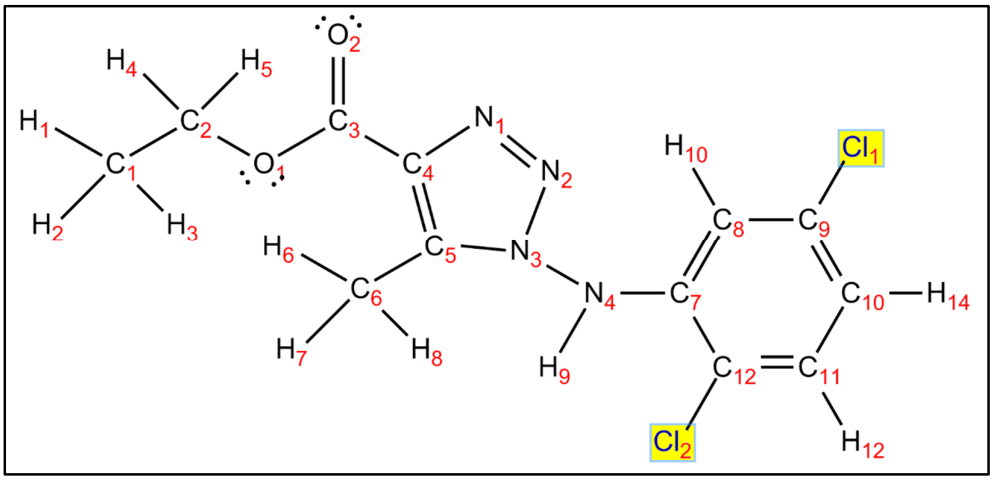



**Figure 6-2 Final structure studied in this chapter**

In addition to the structures given in Figure 6-1, the above structure (**8**) was also studied in this work.

Early success in rationalizing supramolecular interactions based on systematic variation of the electronic profile of the substituents was found during the investigation of the quandary of how carboxylic acids associate in the solid-state – dimer versus catemer[40]. Here, catemer formation was shown to be favored when a proximate C–H group was sufficiently activated, by judicious substitution at adjacent sites, to form an intramolecular C–H···O interaction[40].

Substituent effects based on steric considerations have also proven vital in the control of the manner by which aromatic rings interact, e.g., edge-to-face, in the condensed phase[41]. Other studies have revealed varying conclusions. For example, having electron-donating or -withdrawing substituents directly influenced the conformation observed in diarylacetone derivatives and, hence, supramolecular synthon formation[42]. In another study, the systematic variation of halides in benzyl derivatives exerted little influence upon the crystal structure[43]. It was in this context and in continuation of related studies[44], that a series of seven closely related *N*-arylamino-1,2,3-triazoles (differing only in the nature of the substituent in the 4-position of the aryl ring), as well as an eighth derivative with di-

substitution in the ring, have been investigated by crystallographic and computational methods (Figure 6-1 and Figure 6-2). The influence of substitution is shown to be marked in terms of the ways in which the molecules aggregate to form supramolecular chains, i.e. via N–H···O or N–H···N hydrogen bonds.

In molecules **1** – **8** (Figure 6-1 and Figure 6-2), there are three likely acceptor sites for the acidic N–H atom (atom $N_4$), with either of the two formally doubly bonded nitrogen atoms or the carbonyl-O as the acceptor atom. Our contribution to this collaboration was to perform hybrid DFT/*ab initio* calculation on these molecules and analyze the Mulliken and Natural charges localised on each nuclei. This assisted in the interpretation of the observed supramolecular synthon formation.

## 6.2 Computational Method

Structures **1** – **8** were investigated employing B3LYP using the 6-311+G(d,p) basis set and the gaussian suite of programs[21]. Crystal structures were obtained from our collaborators (see Supporting Publication) and single-point energy calculations performed. Additionally all structures were optimised, starting from the crystal structures and the resulting equilibrium geometries can be found in Appendix B. A frequency calculation confirmed that each optimised structure was a true minimum.

## 6.3 Results and Discussion

As anticipated from their molecular compositions, significant hydrogen bonding exists in the crystal structures of **1** – **8** with three distinct motifs being evident, based on the presence of N–H···N hydrogen bonding only, N–H···O interactions

only or having both N–H··N and N–H···O. Figure 6-3 collects representative diagrams of the three motifs, **I** to **III**, observed.

A qualitative explanation for the diversity in the hydrogen bonding in **1** – **8** is based on electronegativity arguments. In those structures with the least electronegative substituents, i.e., H (**1**), Cl (**3**), Br (**4**), I (**5**) and OMe (**6**), hydrogen bonding involves the amine–H interacting with the ring $N_3$ atom. When the most electronegative substituent is present, i.e., $NO_2$ in **7**, the hydrogen bonding occurs via the amine–H and carbonyl-$O_1$ atoms exclusively. When the electronegativity of the substituents lies between these extremes, i.e., F in **2** and 2 × Cl in **8**, both N–H···N and N–H···O hydrogen bonding occurs.

Support for this explanation is found by comparing the hydrogen bonding patterns in **3** and **8**, which differ by an additional Cl in the ring in **8**. The result of increasing the electronegativity by having two Cl substituents in **8** as opposed to one Cl in **3** is the formation of both N–H···N and N–H···O hydrogen bonds (**8**) rather than N–H···N hydrogen bonding alone (**3**), suggesting that the energy of stabilization of each type of interaction is similar and is finely tuned to electronic effects. The implication of the foregoing is that increasing the electronegativity reduces the hydrogen bonding ability of the $N_3$ atom, thereby promoting the formation of N–H···O hydrogen bonds with the carbonyl–$O_1$ atom. To determine the validity of this qualitative argument, an evaluation of the Natural Population Analysis (NPA), calculated for the geometry optimised molecular structures, was undertaken.

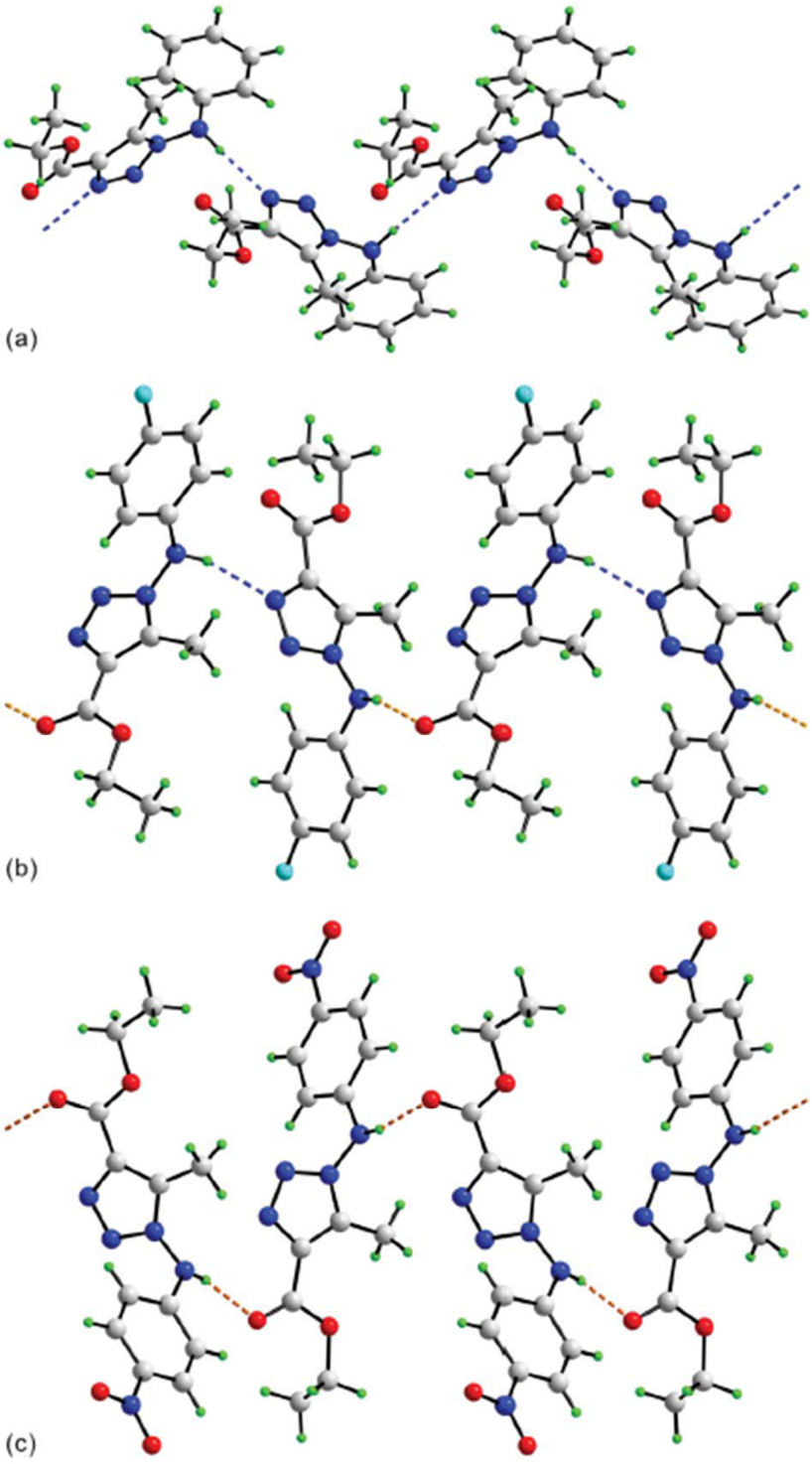



**Figure 6-3 The three distinct motifs of H-bonding observed**
Exemplars of supramolecular chains found in the crystal structures of **1** – **8**. (a) chain mediated by N–H···N hydrogen bonding, motif **I**. (b) Chain sustained by a combination of N–H···N (blue dashed lines) and N–H···O (orange dashed lines) hydrogen bonding, motif **II** and (c) chain mediated N–H···O hydrogen bonding, motif **III**.

**Table 6-1 Natural Charges of Molecules 1 – 8**

| Molecule | Motif | $O_1$ | $N_1$ | $N_2$ | $N_3$ | $N_4$ |
|---|---|---|---|---|---|---|
| 1 | **I** | –0.576 | –0.202 | –0.070 | –0.047 | –0.441 |
| 3 | **I** | –0.578 | –0.200 | –0.070 | –0.049 | –0.440 |
| 4 | **I** | –0.578 | –0.200 | –0.070 | –0.049 | –0.439 |
| 5 | **I** | –0.578 | –0.200 | –0.069 | –0.050 | –0.438 |
| 6 | **I** | –0.578 | –0.205 | –0.075 | –0.044 | –0.437 |
| 2 | **II** | –0.576 | –0.201 | –0.072 | –0.048 | –0.442 |
| 8 | **II** | –0.577 | –0.196 | –0.066 | –0.053 | –0.435 |
| 7 | **III** | –0.578 | –0.195 | –0.067 | –0.054 | –0.429 |

We examined the Mulliken and natural charges on all nuclei for geometries corresponding to the crystal and equilibrium structures. Natural population analysis was developed to calculate atomic charges and orbital populations of the molecular wave functions in a general atomic orbital basis sets. The natural analysis is an alternative to the conventional Mulliken population analysis and exhibits improved numerical stability so as to better describe the electron distribution in compounds of high ionic character, such as those containing metal atoms. Relevant natural charges can be found in Table 6-1. For a full list of charges, see Appendix B.

No significant differences in the NPA charges were ascertained for the $O_1$ atom, but systematic variations were noted in the charges for the nitrogen atoms. Leaving the structure of the F derivative (**2**) to one side, the populations fall in two classes. In **7** and **8**, the charges on the ring–$N_1$, –$N_2$ and –$N_3$ atoms are marginally but systematically less negative, less negative and more negative, respectively, than those on the equivalent atoms in the remaining structures. Further, the charge on the $N_4$ atom in **7** is less negative than in the remaining structures, with a similar but less pronounced trend for **8**. These observations are not consistent with the qualitative arguments above, as the NPA indicates that the basic character of the $N_3$ atom is enhanced with increasing electronegativity of the substituent(s) in the

aryl ring. The analysis of the NPA charges calculated for 2 correlates closely with those structures exhibiting N–H…O hydrogen bonding only. Calculations were also performed on the experimentally determined structures, i.e., without geometry optimization. The NPA analysis, summarised in Appendix B, mirrors that for the geometry optimised structures.

A consideration of the Hammett values[45], $\sigma_p$, which take into account both inductive and resonance contributions, i.e., $\sigma_p = 0.78$ ($NO_2$), 0.23 (Cl and Br), 0.06 (F), 0.00 (H) and −0.27 (OMe), does not assist in formulating a correlation between electronic effects and the observed hydrogen bonding. However, it is worth reiterating that the changes in the NPA are minimal and it may not be worthwhile seeking strict correlations, in consideration of the fact that the molecular structures are subject to the requirements of global crystal packing.

In summary, while a simple interpretation of the experimental observations is not forthcoming from the NPA, it is perhaps not surprising that this is the case given the highly qualitative nature of these quantities and the fact that they are computed for isolated molecules in a vacuum.

# SUPPORTING PUBLICATION

CrystEngComm

RSC Publishing

PAPER

# Aryl-substituents moderate the nature of hydrogen bonds, N–H···N *versus* N–H···O, leading to supramolecular chains in the crystal structures of *N*-arylamino 1,2,3-triazole esters†

Cite this: *CrystEngComm*, 2013, **15**, 4917

Anna C. Cunha,[a] Vitor F. Ferreira,[a] Alessandro K. Jordão,[a] Maria C. B. V. de Souza,[a] Solange M. S. V. Wardell,*[b] James L. Wardell,[c] Peiyu Amelia Tan,[d] Ryan P. A. Bettens,[d] Saikat Kumar Seth[e] and Edward R. T. Tiekink*[f]

Structural analysis reveals the presence of supramolecular chains in a series of eight *N*-arylamino 1,2,3-triazole esters, which differ only in the nature of the substituent (Y) of the terminal aryl ring. In each of **1** (Y = 4-H), **3** (4-Cl), **4** (4-Br), **5** (4-I) and **6** (4-OMe), the chains are sustained by N–H···N hydrogen bonding. In **2** (Y = 4-F) and **8** (Y = 2,5-$Cl_2$), the chains are mediated by alternating N–H···N and N–H···O hydrogen bonding, whereas in **7** (Y = 4-$NO_2$) the chain is sustained by N–H···O hydrogen bonding only. While the differences in the adopted supramolecular motifs are qualitatively correlated with the electronegativity of the Y substituents, no quantitative correlations could be made with the electronic structures of the theoretical gas-phase molecules. Two distinct patterns of crystal packing are observed, with the first of these being based on the inter-digitation of layers, comprised of supramolecular chains and connections of the type C–X···π(aryl) between them for **3–5** and **8**; only weak off-set edge-to-edge π···π interactions were noted in the case of **1**. A common feature of the zigzag chains in these crystal structures was a *syn*-disposition of successive aryl rings along the axis of propagation. The remaining structures adopted three-dimensional architectures where the Y substituents of the *anti*-disposed aryl rings participated in F···H (**2**) or C–H···O (**6** and **7**) interactions. A detailed analysis of the Hirshfeld surfaces and fingerprint plots for **1–8** enabled a comparison of the intermolecular interactions involved in constructing the disparate supramolecular architectures. In the structures featuring N–H···N hydrogen bonding leading to the supramolecular chain, the maximum contribution to the overall crystal packing was less than 20%. This increased to over 25% in the case where there was exclusive N–H···O hydrogen bonding in the chain.

Received 18th January 2013,
Accepted 10th April 2013

DOI: 10.1039/c3ce40102b

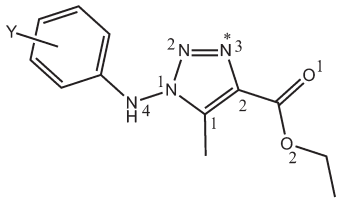
**www.rsc.org/crystengcomm**

## Introduction

The robust and directional nature of hydrogen bonding interactions makes these favourite supramolecular synthons present in the crystal engineer's toolbox. However, not all molecular systems have hydrogen bonding functionality or, even if they do, hydrogen bonding may not extend in three dimensions. In these circumstances weak intermolecular interactions naturally come to the fore, *e.g.* π···π, C–H···π, halogen bonding, *etc*. Such considerations make it imperative to study and understand these ''second tier'' supramolecular synthons to enable their control for supramolecular assembly. Desiraju identifies the study of intermolecular interactions as the ''what'' of crystal engineering and the first stage of the continuum leading to evaluating the influence of rational changes in molecular packing upon functional crystalline materials (the ''why'').[1] Clearly, in order to determine the importance and prevalence of specific supramolecular synthons, systematic structural studies of closely related chemical

[a]*Departamento de Química Orgânica, Instituto de Química, Universidade Federal Fluminense, Outeiro de São João Baptista, 24020-141 Niterói, RJ, Brazil*
[b]*CHEMSOL, 1 Harcourt Road, Aberdeen, AB15 5NY, Scotland. E-mail: solangewardell@gmail.com*
[c]*Centro de Desenvolvimento Tecnológico em Saúde (CDTS), Fundação Oswaldo Cruz, (FIOCRUZ), Casa Amarela, Campus de Manguinhos, Av. Brasil 4365, 21040-900, Rio de Janeiro, RJ, Brazil*
[d]*Department of Chemistry, National University of Singapore, Singapore 117543, Singapore*
[e]*Department of Physics, M. G. Mahavidyalaya, Bhupatinagar, Purba Medinipur, West Bengal-721425, India*
[f]*Department of Chemistry, University of Malaya, 50603 Kuala Lumpur, Malaysia. E-mail: Edward.Tiekink@um.edu.my; Fax: +60 3 7967 4193; Tel: +60 3 7697 6775*

† Electronic supplementary information (ESI) available: Diagrams illustrating molecular structures and supramolecular aggregation, data describing intermolecular interactions and NPA charges. CCDC 845654–845660 and 672062. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c3ce40102b




Y:
(**1**) 4-H (**2**) 4-F (**7**) 4-$NO_2$
(**3**) 4-Cl (**8**) 4-H; 2,5-$Cl_2$
(**4**) 4-Br
(**5**) 4-I
(**6**) 4-OMe

**Scheme 1** The chemical structures of the *N*-arylamino-1,2,3-triazole derivatives (**1–8**) investigated herein. These are grouped in accordance with the supramolecular motif they adopt. For species **1**, **2**, **5** and **8**, in which two molecules comprise the crystallographic asymmetric unit, the second molecule is labelled so that N1 becomes N5, N2 becomes N6, O1 becomes O3, *etc*. For **7**, which also has two independent molecules, the N6–N10 atoms correspond to the N1–N4 atoms.

species are required. Herein, an evaluation of a series of closely related structures is made (Scheme 1). Specifically, the influence of differences in the nature of the aryl-bound substituents upon hydrogen bonding patterns is evaluated.

Early success in rationalising supramolecular interactions based on systematic variation of the electronic profile of the substituents was found during the investigation of the quandary of how carboxylic acids associate in the solid-state – dimer *versus* catemer.[2] Here, catemer formation was shown to be favoured when a proximate C–H group was sufficiently activated, by judicious substitution at adjacent sites, to form an intramolecular C–H···O interaction.[2] Substituent effects based on steric considerations have also proven vital in the control of the manner by which aromatic rings interact, *e.g.* edge-to-face, in the condensed phase.[3] Other studies have revealed varying conclusions. For example, having electron-donating or -withdrawing substituents directly influenced the conformation observed in diarylacetone derivatives and, hence, supramolecular synthon formation.[4] In another study, the systematic variation of halides in benzyl derivatives exerted little influence upon the crystal structure.[5] It was in this context, and in continuation of related studies,[6] that a series of seven closely related *N*-arylamino-1,2,3-triazoles, differing only in the nature of the substituent in the 4-position of the aryl ring, as well as an eighth derivative with di-substitution in the ring, have been investigated by crystallographic and computational methods (Scheme 1). The influence of substitution is shown to be marked in terms of the ways in which the molecules aggregate to form supramolecular chains, *i.e. via* N–H···O or N–H···N hydrogen bonds.

Molecules **1–8** became available as a consequence of an anti-cantagalo virus replication study.[7a] 1,2,3-Triazoles are known to exhibit a wide range of biological activities ranging from cytostatic,[8a] anti-neoplastic,[8b] anti-HIV[8c] and anti-microbial[8d] to anti-inflammatory agents,[8e] and they are potassium channel activators.[8f] Over the past decade, Ferreira *et al.* have evaluated various biological applications of triazoles,[7] during which time the basic 1,2,3-triazole structure has been embellished to include an amino group.[7a]

The novelty of the molecules reported herein is borne out by the observation that there is only one crystal structure determination[7a] included in the Cambridge Crystallographic Database (CSD)[9] of a neutral molecule containing the C–N(H)–$N_3$ sequence, with the terminal three nitrogen atoms within a ring. This is a hydrazine derivative whereby the –C(=O)OEt group in **2**, Scheme 1, is replaced by –$NHNH_2$.[7a] A very recent report describes related structures whereby the –C(=O)OEt group in **1**, **2** and **3** is substituted by –$CH_2OH$.[10] In **1–8**, there are three likely acceptor sites for the acidic N–H atom, with either of the two formally doubly bonded nitrogen atoms or the carbonyl–O as the acceptor atom. Herein, supramolecular chains are formed for all eight structures based on N–H hydrogen bonds involving the nitrogen atom indicated with an asterisk in Scheme 1 or, less frequently, involving the carbonyl–O atom. Intriguingly, structures **2** and **8** exhibit both types of hydrogen bonding interactions within a single chain.

## Results and discussion

The structural analyses of seven *p*-substituted derivatives of ethyl 1-(arylamino)-5-methyl-1*H*-[1,2,3]-triazole-4-carboxylate (**1–7**) along with a 2,5-disubstituted analogue (**8**) have been accomplished (Scheme 1). The availability of eight closely related structures has enabled an investigation of the influence of systematic substitution in the aryl ring upon supramolecular aggregation patterns based on hydrogen bonding.

## Molecular structures

Owing to the presence of two molecules in the asymmetric unit in each of **1**, **2**, **5**, **7** and **8**, there are a total of 13 distinct molecules in the present series; molecular structures and overlay diagrams are illustrated in the ESI,† Fig. S(1)–S(8). In each of **1** and **2** the two independent molecules are related by a pseudo centre of inversion, whereas the two molecules are approximately super-imposable in the cases of **5**, **7** and **8**. Each molecular structure comprises a central five-membered 1,2,3-triazole ring connected *via* an N–N bond to a secondary amine that carries an aryl ring, and *via* a C–C bond to the ester functionality; salient geometric parameters are collected in Table 1. As seen from the overlay diagram in Fig. 1, the ester–carbonyl atom is oriented away from the ring-bound methyl group and is almost co-planar with the five-membered ring in all structures, as seen in the range of N(3,7)–C–C–O(1,3) torsion angles of 0.7(3)° in **4** to 7.4(3)° in **2**. The terminal ester group has considerably more flexibility, with the C(4,14)–O–C–C(6,18) torsion angles indicating that dispositions range from co-planar, *e.g.* −176.4(2)° in **1**, to orthogonal, *e.g.* −86.7(2)° in

**Table 1** Selected geometric parameters (Å, °) for **1–8**

| Parameter | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| N3–C2–C4–O1 | 4.0(4) | −0.8(3) | 0.9(3) | 0.7(3) | 5.7(8) | −3.8(3) | 6.0(4) | −4.5(4) |
| C4–O2–C5–C6 | −86.8(3) | −86.7(2) | −171.53(18) | 172.61(17) | 169.1(5) | −176.25(16) | 169.6(2) | −173.2(3) |
| N7–C14–C15–O3 | −3.1(4) | 7.4(3) | — | — | −0.6(8) | — | 2.0(4)[a] | 0.8(4) |
| C16–O4–C17–C18 | −176.4(2) | −170.55(19) | — | — | 170.8(5) | — | 89.6(3) | 88.4(3) |
| N2–N1–N4–C7 | −73.8(3) | 69.5(3) | 55.7(2) | −57.0(2) | −68.0(6) | −81.3(2) | −69.5(4) | 78.1(3) |
| N6–N5–N8–C19 | 74.5(3) | −90.5(2) | — | — | −60.6(7) | — | −64.2(4)[b] | −99.1(3) |
| N1–N4–C7–C8 | −13.6(3) | 20.7(3) | 36.7(3) | 146.72(17) | 147.2(5) | −10.7(3) | −12.9(4) | −159.6(2) |
| N5–N8–C19–C20 | 13.6(3) | 9.3(3) | — | — | 149.3(5) | — | −22.9(4)[c] | 177.8(2) |
| Motif[d] | **I** | **II** | **I** | **I** | **I** | **I** | **III** | **II** |
| Graph set symbol | *C*(5) | *C*(5)*C*(7) | *C*(5) | *C*(5) | *C*(5) | *C*(5) | *C*(7) | *C*(5)*C*(7) |
| Pitch[e] | 110 | 104 | 125 | 125 | 120 | 103 | 136(O) | 109 |
| | 106 | 92(O) | — | — | 125 | — | 122(O) | 106(O) |
| Repeat distance[f] | 10.4 | 10.1 | 11.1 | 11.1 | 11.0 (×2) | 9.9 | 10.2 | 11.2 |

[a] The torsion angle is N8–C14–C15–O5. [b] The torsion angle is N7–N6–N9–C19. [c] The torsion angle is N6–N9–C19–C20. [d] The motif refers to the supramolecular chain. [e] The pitch refers to the angle subtended at the N(O) atom by the adjacent atoms participating in the hydrogen bonding, *i.e.* N···N(H)···N for motif **I**. [f] The repeat distance is the distance between successive molecules comprising the chain.

**2**. Similarly, the aryl groups occupy a range of positions, approximately orthogonal to the five-membered ring, with the N(2,6)–N–N–C(7,19) torsion angles lying in the range 55.7(2) (**3**) to −99.1(3)° (**8**), presumably to minimise interactions between the ring-methyl and aryl rings. The aryl ring is also twisted with respect to the N(amine)–C(ipso) bond, as seen in the N(1,5)–N–C–C(8,20) torsion angles, which range from 9.3(3) (**2**) to 146.72(17)° (**4**).

The molecular structures of **1–8** were also investigated employing B3LYP theory using the 6-311+G(d,p) basis set within Gaussian09.[11] From Fig. 2 it is evident that the great disparity observed in the experimental structures (Fig. 1) no longer persists in the geometry optimised structures. This observation is borne out by the narrower ranges of torsion angles. For example, the maximum deviation of the carbonyl group from co-planarity with the ring is less than 2.2° (for **2**). The terminal ethyl groups adopt two conformations, *i.e.* co-planar with the $CO_2$ residue (torsion angles: 178.3 to 180°) or approximately normal (85.3 and 85.6° for **1** and **2**, respectively). The aryl ring is close to being perpendicular to the triazole ring in all cases, with the range of N–N–N(amine)–C torsion angles being 74.7° (**8**) to 80.2° (**6**). Finally, the twist of the aryl ring from the N–N(amine) bond is less pronounced, with the range of N–N(amine)–C–C torsion angles being narrow at 5.6° (**6**) to 17.5° (**7**).

From the foregoing, there is no systematic influence upon the molecular structure that can be correlated to the nature of the aryl-bound substituents.

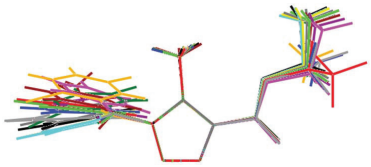


**Fig. 1** Overlay diagrams highlighting the relative orientations of the terminal substituents with respect to the central and superimposed 1,2,3-triazole ring in the experimental structures of **1–8**. Colour codes: first independent molecule of **1**, red; second independent molecule of **1** (inverted), green; first independent molecule of **2**, blue; second independent molecule of **2** (inverted), pink; **3**, light-blue; **4**, yellow; first independent molecule of **5**, grey; second independent molecule of **5**, black; **6**, brown; first independent molecule of **7**, purple; second independent molecule of **7**, dark-grey; first independent molecule of **8**, orange; second independent molecule of **8**, dark-green.

## Supramolecular structures based on hydrogen bonding

As anticipated from their molecular compositions, significant hydrogen bonding exists in the crystal structures of **1–8** with three distinct motifs being evident, based on the presence of N–H···N hydrogen bonding only, N–H···O interactions only or having both N–H···N and N–H···O. Fig. 3 collects representative diagrams of the three motifs and Table 2 collates their geometric data. No evidence was found for bifurcated interactions despite the *syn* relationship of the putative hydrogen bonding acceptors N3 and O1, and they being separated by only two carbon atoms. For example, in **1**, where

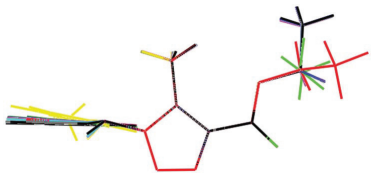


**Fig. 2** Overlay diagrams highlighting the relative orientations of the terminal substituents with respect to the central and superimposed 1,2,3-triazole ring in the theoretical structures of **1–8**. Colour codes: **1**, red; **2**, green; **3**, blue; **4**, pink; **5**, light-blue; **6**, yellow; **7**, grey; and **8**, black.

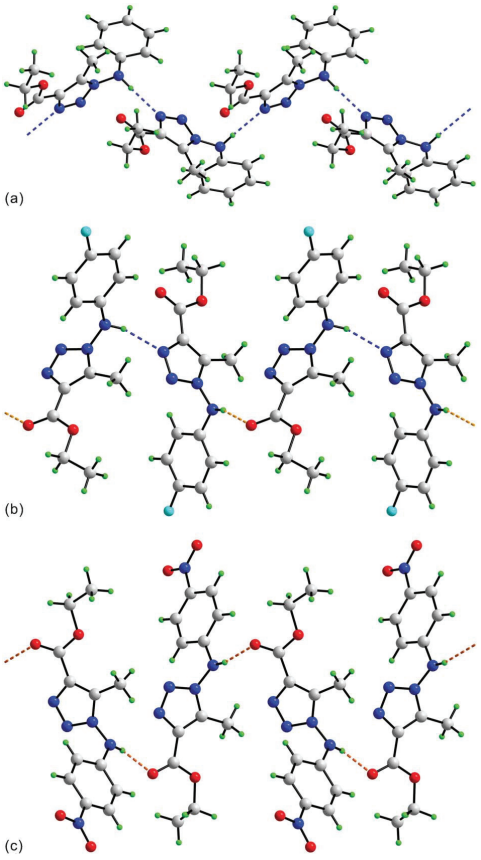



**Fig. 3** Exemplars of supramolecular chains found in the crystal structures of **1–8**: (a) chain mediated by N–H···N hydrogen bonding exclusively in **1**. (b) Chain sustained by a combination of N–H···N (blue dashed lines) and N–H···O (orange dashed lines) hydrogen bonding in **2**. (c) Chain mediated by N–H···O hydrogen bonding exclusively in **7**.

N–H···N hydrogen bonding is observed, the two H···O separations are 2.66 and 2.73 Å, respectively, and the two N–H···O angles are 110 and 111°, respectively. In **7**, with N–H···O hydrogen bonding only, the H···N separations are 2.63 and 2.98 Å, respectively, and the N–H···N angles are 119 and 123°, respectively.

In **1**, the two independent molecules comprising the asymmetric unit associate into a zigzag supramolecular chain sustained by N–H···N hydrogen bonds, Fig. 3a, and it has Graph Set symbol $C(5)$.[12] This motif, *i.e.* **I**, is adopted by the majority of the structures, see ESI,† Fig. S(9), although the topologies of the chains vary. In each of **3**, **4** and **5**, the topology of the chain is zigzag, being propagated by glide symmetry, but in **6**, the chain is helical, being propagated by $2_1$ screw symmetry along the *b*-axis; in **5** each of the independent

**Table 2** Summary of hydrogen bonding interactions (A–H···B; Å, °) operating in the crystal structures of **1–8**[a]

| A | H | B | A–H | H···B | A···B | A–H···B | Symmetry operation |
|---|---|---|---|---|---|---|---|
| **1** | | | | | | | |
| N4 | H4n | N7 | 0.88(2) | 2.16(2) | 3.035(3) | 177(2) | $1-x$, $1-y$, $\frac{1}{2}+z$ |
| N8 | H8n | N3 | 0.88(2) | 2.10(2) | 2.984(3) | 178(2) | $1-x$, $-y$, $-\frac{1}{2}+z$ |
| **2** | | | | | | | |
| N4 | H4n | O3 | 0.88(2) | 2.07(2) | 2.915(3) | 161(2) | $x$, $\frac{1}{2}-y$, $\frac{1}{2}+z$ |
| N8 | H8n | N3 | 0.880(15) | 2.177(18) | 3.031(3) | 163(2) | $1+x$, $\frac{1}{2}-y$, $-\frac{1}{2}+z$ |
| **3** | | | | | | | |
| N4 | H4n | N3 | 0.881(17) | 2.10(2) | 2.931(2) | 157.9(17) | $x$, $-\frac{1}{2}-y$, $-\frac{1}{2}+z$ |
| **4** | | | | | | | |
| N4 | H4n | N3 | 0.871(17) | 2.12(2) | 2.945(2) | 156.9(17) | $x$, $\frac{1}{2}-y$, $-\frac{1}{2}+z$ |
| **5** | | | | | | | |
| N4 | H4n | N3 | 0.88(4) | 2.15(4) | 2.989(6) | 158(4) | $-\frac{1}{2}+x$, $1-y$, $z$ |
| N8 | H8n | N7 | 0.88(4) | 2.13(5) | 2.918(6) | 149(5) | $\frac{1}{2}+x$, $-y$, $z$ |
| **6** | | | | | | | |
| N4 | H4n | N3 | 0.92(2) | 2.19(2) | 3.072(3) | 160.5(19) | $2-x$, $\frac{1}{2}+y$, $1\frac{1}{2}-z$ |
| **7** | | | | | | | |
| N4 | H4n | O5 | 0.89(3) | 2.05(3) | 2.900(3) | 162(3) | $1+x$, $y$, $1+z$ |
| N9 | H9n | O1 | 0.89(2) | 1.98(3) | 2.822(3) | 159(3) | $x$, $y$, $-1+z$ |
| **8** | | | | | | | |
| N4 | H4n | N7 | 0.88(2) | 2.22(2) | 3.081(3) | 166(2) | $x$, $y$, $z$ |
| N8 | H8n | O1 | 0.88(2) | 2.29(3) | 2.921(3) | 129(2) | $x$, $1+y$, $z$ |
| N4 | H4n | Cl1 | 0.88(2) | 2.61(3) | 2.983(2) | 107(2) | $x$, $y$, $z$ |
| N8 | H8n | Cl3 | 0.88(2) | 2.63(3) | 2.939(2) | 102(2) | $x$, $y$, $z$ |

[a] For each of **1**, **2**, **5** and **8**, having two molecules in the asymmetric unit, the N3 atom of the first independent molecule corresponds to the N7 atom of the second, N4 with N8, and the O3 atom to the O1 atom. For **7**, also with two independent molecules, the N4 corresponds to N9 atom, and O1 with O5.

molecules self-associates into a zigzag chain. The second motif, **II**, is found for both **2**, Fig. 3b, and **8** (for the latter intramolecular N–H···Cl hydrogen bonding is also noted, Table 2). In each of **2** and **8**, the two independent molecules comprising the asymmetric unit are connected into a zigzag supramolecular chain *via* alternating N–H···N and N–H···O hydrogen bonds, *i.e.* with Graph Set symbol $C(5)C(7)$. The third motif, **III**, is found in one example only, namely **7**, Fig. 3c. Here, a helical chain propagated by $2_1$ screw symmetry along the *b*-axis and with a repeat unit of two molecular entities is sustained solely by N–H···O hydrogen bonding, with Graph Set symbol $C(7)$. Despite the different modes of association between the molecules comprising the supramolecular chains in **1–8**, no systematic correlations describing the nature of the chains, *e.g.* topology, pitch and repeat distance (Table 1), are evident.

A qualitative explanation for the diversity in the hydrogen bonding in **1–8** is based on electronegativity arguments. In those structures with the least electronegative substituents, *i.e.* H (**1**), Cl (**3**), Br (**4**), I (**5**) and OMe (**6**), hydrogen bonding

**Table 3** Selected values from the Natural Population Analysis data for **1–8** arranged in order of the supramolecular motif they adopt

| Compound | Motif | O1 | N1 | N2 | N3 | N4 |
|---|---|---|---|---|---|---|
| **1** | **I** | −0.576 | −0.202 | −0.070 | −0.047 | −0.441 |
| **3** | **I** | −0.578 | −0.200 | −0.070 | −0.049 | −0.440 |
| **4** | **I** | −0.578 | −0.200 | −0.070 | −0.049 | −0.439 |
| **5** | **I** | −0.578 | −0.200 | −0.069 | −0.050 | −0.438 |
| **6** | **I** | −0.578 | −0.205 | −0.075 | −0.044 | −0.447 |
| **2** | **II** | −0.576 | −0.201 | −0.072 | −0.048 | −0.442 |
| **8** | **II** | −0.577 | −0.196 | −0.066 | −0.053 | −0.435 |
| **7** | **III** | −0.578 | −0.195 | −0.067 | −0.054 | −0.429 |

involves the amine–H interacting with the ring N3 atom. When the most electronegative substituent is present, *i.e.* $NO_2$ in **7**, the hydrogen bonding occurs *via* the amine–H and carbonyl–O1 atoms exclusively. When the electronegativity of the substituents lies between these extremes, *i.e.* F in **2** and 2 × Cl in **8**, both N–H···N and N–H···O hydrogen bonding occurs. Support for this explanation is found by comparing the hydrogen bonding patterns in **3** and **8**, which differ by an additional Cl in the ring in **8**. The result of increasing the electronegativity by having two Cl substituents in **8** as opposed to one Cl in **3** is the formation of both N–H···N and N–H···O hydrogen bonds (**8**) rather than N–H···N hydrogen bonding alone (**3**), suggesting that the energy of stabilisation of each type of interaction is similar and is finely tuned to electronic effects. The implication of the foregoing is that increasing the electronegativity reduces the hydrogen bonding ability of the N3 atom, thereby promoting the formation of N–H···O hydrogen bonds with the carbonyl–O1 atom. To determine the validity of this qualitative argument, an evaluation of the Natural Population Analysis (NPA), calculated for the geometry optimised molecular structures, was undertaken.

Table 3 collates the NPA charges for the oxygen and nitrogen atoms in **1–8**. No significant differences were ascertained for the O1 atom but systematic variations were noted in the charges for the nitrogen atoms. Leaving the structure of the F derivative (**2**) to one side, the populations fall in two classes. In **7** and **8**, the charges on the ring-N1, -N2 and -N3 atoms are marginally but systematically less negative, less negative and more negative, respectively, than those on the equivalent atoms in the remaining structures. Further, the charge on the N4 atom in **7** is less negative than in the remaining structures, with a similar but less pronounced trend for **8**. These observations are not consistent with the qualitative arguments above, as the NPA indicates that the basic character of the N3 atom is enhanced with increasing electronegativity of the substituent(s) in the aryl ring. The analysis of the NPA charges calculated for **2** correlates closely with those structures exhibiting N–H···O hydrogen bonding only. Calculations were also performed on the experimentally determined structures, *i.e.* without geometry optimisation. The NPA analysis, summarised in ESI,† Table S(1), mirrors that for the geometry optimised structures.

A consideration of the Hammett values,[13] $\sigma_p$, which take into account both inductive and resonance contributions, *i.e.* $\sigma_p$ = 0.78 ($NO_2$), 0.23 (Cl and Br), 0.06 (F), 0.00 (H) and −0.27 (OMe), does not assist in formulating a correlation between electronic effects and the observed hydrogen bonding. However, it is worth reiterating that the changes in the NPA are minimal and it may not be worthwhile seeking strict correlations, in consideration of the fact that the molecular structures are subject to the requirements of global crystal packing.

In summary, while a simple interpretation of the experimental observations is not forthcoming from the NPA, it is perhaps not surprising that this is the case given the highly qualitative nature of these quantities and the fact that they are computed for isolated molecules in a vacuum.

## Crystal packing

The ensuing description of the crystal packing patterns in **1–8** is based on the standard significance criteria established in PLATON.[14] The common feature of the crystal structures is the arrangement of supramolecular chains into layers, with or without specific interactions between the adjacent chains, and, with the exception of **1**, they are connected into a three-dimensional architecture by interactions of varying types but always involving the Y substituents. Geometric parameters describing these and other intermolecular interactions discussed in this section are listed in ESI,† Table S(2).

The exceptional supramolecular architecture is found in **1**, where chains are linked into layers in the *bc*-plane by C–H···π [occurring between the methyl– and methylene–H hydrogen atoms and aryl rings] interactions, which are classified as type III and type I interactions, following Malone *et al.*[15] As shown in Fig. 4a, the layers, having aryl rings on either side, interdigitate along the *a*-axis with no specific interactions between them. The closest approach of the rings derived from neighbouring layers are off-set edge-to-edge contacts with the shortest separation between these being 3.605(3) Å for C11···C20 with a dihedral angle of 14.48(7)° between the rings (symmetry operation ½ + *x*, 1 − *y*, *z*). From the foregoing, it is apparent that the carbonyl–O atoms do not participate in the stabilisation of the crystal structure. However, they do provide stability to the supramolecular chain by the apparent formation of C=O···π($C_2N_3$) interactions (as detailed in ESI,† Table S(2)). Increasingly, O···π interactions, first recognised in macromolecular crystallography,[16a] are being recognised as being important in stabilizing crystal structures.[16b–d] In the case of the O1 atom, the contact appears to be semi-localised,[17] being directed towards the C1–N1 bond with separations of 2.922(3) and 2.934(3) Å, respectively, with the other contacts ranging from 3.455(3) (C4) to 3.756(6) Å (N3) (symmetry operation 1 − *x*, −*y*, ½ + *z*). Under these circumstances, the interactions with the five-membered ring are best represented as C=O···π(C–N). A similar situation pertains for the second carbonyl–O atom (ESI,† Table S(2)).

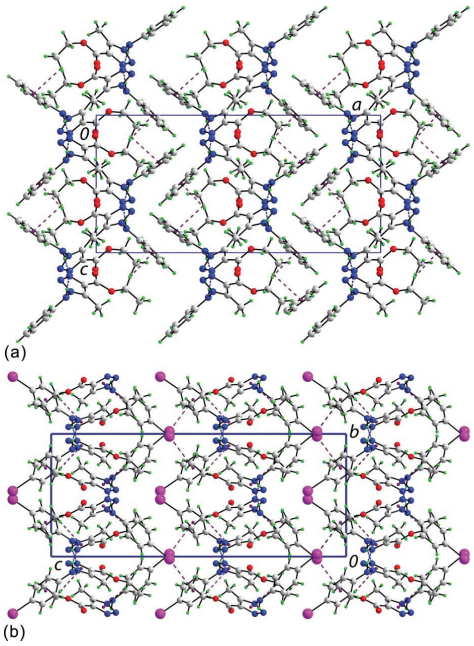



**Fig. 4** (a) View in projection down the *b*-axis of the unit cell contents of **1**, highlighting the inter-digitation of layers. C–H···π interactions are shown as purple dashed lines. (b) View in projection down the *a*-axis of the unit cell contents of **5**, highlighting the inter-digitation of layers and the I···π(aryl) interactions (pink dashed lines) connecting them.

Very similar crystal packing is found in each of isostructural **3** and **4**, in **5**, and in disubstituted **8**.

In **3**, the layers are consolidated by C–H···O [where the bifurcated carbonyl–O1 atom accepts interactions from ring-methyl– and aryl–H] and C–H···N [methylene–H···central N2 atom of the ring] interactions. Further stabilisation to the layer is provided by semi-localised C=O···π(C–N) interactions. In the case of the Y = Br (**4**) derivative, the intra-layer interactions are based on C–H···O (aryl–H···O(carbonyl)) and semi-localised C=O···π(C–N) interactions to the C1–N1 bond. In **5**, with two independent molecules, the layers are stabilised by type III[15] C–H···π($C_2N_3$) (involving methylene–H from each independent molecule) and semi-localised C=O···π(C–N) interactions. As for **1**, the layers in **2–5** inter-digitate along the stacking axis but differ in that there are connections between the layers, of the type X···π(aryl), as exemplified in Fig. 4b for the structure of **5** (ESI,† Fig. S(10) includes the other packing diagrams). Ongoing research into X···π(aryl) interactions, which have been known for some time[18a–c] and are known to be to be directional in their mode of interaction,[18d] continues to illustrate the importance of these in crystal engineering.[18e–g]

Despite the presence of O–H···O hydrogen bonding, the crystal structure of **8** closely resembles the foregoing. Here,

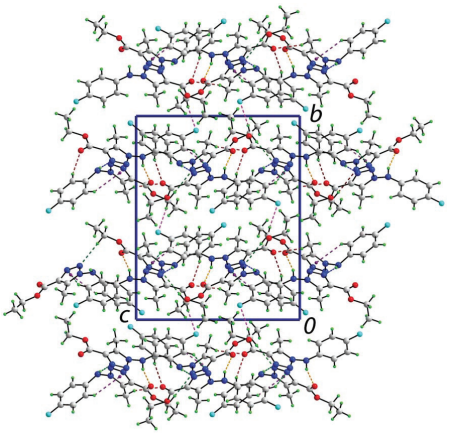



**Fig. 5** (a) View in projection down the *a*-axis of the unit cell contents of **2** highlighting the C–F···H interactions (pink dashed lines) that serve to link layers along the *b*-axis. The C–H···O and C–F···N interactions are shown as brown and green dashed lines, respectively.

supramolecular chains are stabilised by C–H···O (ring-methyl–H···O1) and C=O···π(C–N) interactions and connected into a somewhat jagged layer by type V[15] methyl–C–H···π(aryl) contacts. The layers stack along the *c*-axis, being connected by Cl···π(aryl) interactions (ESI,† Fig. S(10)).

The remaining crystal structures feature specific intermolecular interactions leading to three-dimensional architectures, with each of the Y substituents playing a key role. In **2**, layers may be discerned in the *ac*-plane that are stabilised by a network of C–H···O (bifurcated carbonyl–O1 atom accepts interactions from ring-methyl– and aryl–H), C–H···N (where the N3 atom of the second independent molecule not involved in hydrogen bonding forms an interaction with aryl–H, and the central N2 atom of the same ring interacts with methylene–H) and type V[15] aryl–C–H···π($C_2N_3$) contacts. The layers are connected by C–H···F interactions, Fig. 5.

The consolidation of the supramolecular chains in **6** is dominated by type V[15] C–H···π($C_2N_3$) (involving methylene–H) and methoxy–C–H···π(aryl) interactions (ESI,† Fig. S(10)). Finally, the description of the crystal structure of **7** can be simplified by considering it in terms of arbitrary layers. Thus, layers comprising supramolecular chains are formed through the agency of C–H···O (involving methylene–, ring-methyl– and methyl–H interacting with carbonyl– and nitro–O) and C–H···N (involving methylene– and aryl–H interacting with the N2 and N3 atoms of one ring) interactions. Layers are connected along the *b*-axis by C–H···O (methylene–H and aryl–H interacting with nitro–O atoms derived from both independent molecules) and delocalised nitro–O···π($C_2N_3$) interactions;[16] see ESI,† Fig. S(10). It is noteworthy that the N3 atom of the second independent molecule does not participate in a significant

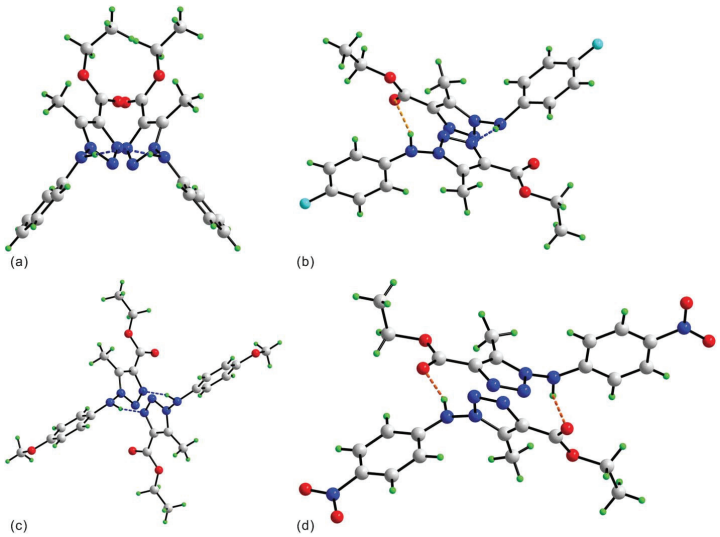



**Fig. 6** End-on views of the supramolecular chains in (a) **1**, (b) **2**, (c) **6** and (c) **7**. The chain illustrated for **1** is representative of **3–5** and **8**.

intramolecular interaction, arguably owing to the close approach of the aforementioned nitro–O atom.

To a first approximation, the crystal structures of **1–8** fall in two classes. Those of **1**, **3–5** and **8** comprise layers of zigzag supramolecular chains with minimal (**1**) or C–X···π(aryl) (**3–5** and **8**) interactions between the inter-digitating layers. Such modes of interaction require a *syn*-disposition of adjacent aryl rings and this is indeed borne out in the end-on view of the supramolecular chain for **1**, as shown in Fig. 6a (see ESI,† Fig. S(9) for analogous views for **3–5** and **8**). By contrast, the three remaining crystal structures do not feature C–X···π(aryl) interactions and achieve a three-dimensional architecture *via* different interactions involving the Y substituents, *i.e.* F···H in **2** and C–H···O in the cases of **6** (C–H donor) and **7** (C–H acceptor), occurring between arbitrarily defined layers. In **2**, **6** and **7**, successive aryl rings within the chain are *anti*-, as seen from Fig. 6b–d. There is no correlation between the motif of the supramolecular chain and crystal packing, as both N–H···N hydrogen bonding (**1**, **3–5** and **8**) and N–H···N and N–H···O hydrogen bonding (**8**) are found in the layered crystal structures. For the second type of crystal packing, all three motifs are represented. With these considerations in mind, it is apparent that no systematic correlation of supramolecular chain motif with crystal structure is found in **1–8**.

## Hirshfeld surfaces

The Hirshfeld surfaces[19–21] of **1–8** are illustrated in Fig. 7, showing surfaces that have been mapped over a $d_{norm}$ range of −0.5 to 1.5 Å. Since compounds **1**, **2**, **5**, **7** and **8** have two independent molecules in the asymmetric unit, the individual molecules have been designated by ***A*** and ***B***. Referring to Fig. 7, the dominant interactions between amine N–H with triazole–N atoms and/or carbonyl O atoms in **1–8** can be seen in the Hirshfeld surfaces as the bright-red areas marked with encircled '***a***' and '***b***', respectively. The light-red spots are due to C–H···O interactions and other visible spots on the $d_{norm}$ surfaces correspond to H···H contacts. The small extents of visible area and very light-coloured regions on the surfaces indicate weaker and longer contacts other than hydrogen bonds. The dominant N–H···N and N–H···O hydrogen bonding interactions appear as two distinct spikes in the two-dimensional fingerprint plots,[22] shown in Fig. 8, labelled as N···H/H···N and O···H/H···O.

For the N···H/H···N interactions, complementary regions are visible in the fingerprint plots where one molecule acts as a donor ($d_e > d_i$) and the other as an acceptor ($d_e < d_i$). Prominent pairs of sharp spikes of nearly equal lengths in the region 1.962 Å $< (d_e + d_i) <$ 2.623 Å are characteristic of nearly equal N(donor)···N(acceptor) distances (2.99 ± 0.09 Å). The upper spikes correspond to the donor spike (amine–H interacting with triazole–N atoms), with the lower spike being an acceptor spike (triazole–N atoms interacting with the H atoms of amine groups). Compound **7** does not feature N–H···N interactions but it does exhibit intermolecular C–H···N interactions, which have a clear signature in the Hirshfeld surface as the light-red spots marked with an encircled '***c***' in Fig. S(11) in the ESI.† The Hirshfeld surface does not show similar proportions of N···H interactions for each molecule, ranging from 15.4% in moiety **5A** to 19.1% in **2B**, where the proportions of N···H interactions have more variety than its

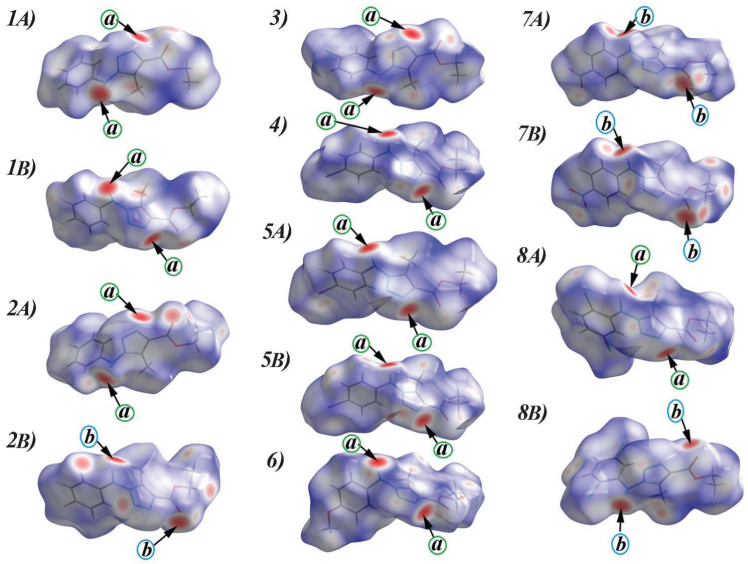



**Fig. 7** Hirshfeld surfaces mapped with $d_{norm}$ for compounds **1–8**. For compounds that have two moieties in the asymmetric unit these are denoted by '***A***' and '***B***'.

H···N, *i.e.* from 6.8% in **2A** to 8.8% in **2B**. The decomposition of fingerprint plots, which enables the separation of contributions from N···H/H···N interactions, is depicted in Fig. S(12) in the ESI.†

The N–H···O hydrogen bonds also appear as two distinct spikes in the two-dimensional fingerprint plots (Fig. 8). The spikes in the region 1.848 Å $<$ $(d_e + d_i)$ $<$ 2.568 Å are characteristic of a nearly equal N(donor)···O(acceptor) distance (2.96 $\pm$ 0.07 Å). The upper spike denotes that the amine–H atoms are interacting with the carbonyl–O atoms and the lower spikes indicate that the O atoms are interacting with the H-atoms of NH groups. The decomposition of the fingerprint plot due to O···H contributions is depicted in Fig. S(12) in the ESI,† which clearly shows that compounds **1** and **3–6** do not exhibit dominant N–H···O hydrogen bonding. The O···H/H···O contribution to the total Hirshfeld surface varies from 8.5% in **8A** to 32.2% in **7A**.

The aryl-substituents in **1**, **2**, **5**, **6** and **8** provide geometric conditions to enable C–H···π interactions in their crystal structures. These C–H···π contacts are represented by each $d_e$ surface showing a significant bright orange spot (ESI,† Fig. S(13)), which is also viewed by a distinct pattern of a pair of ''wings'' in the two-dimensional fingerprint plots (Fig. 8). At the top left and bottom right of the plots, these ''wings'' illustrate the characteristic features of C–H···π interactions.[20a,23] The shape of the ''wings'' in the breakdown of the fingerprint plot, Fig. S(12), ESI,† and the sums of $d_e$ and $d_i$ highlight the importance of these interactions. The decomposition of the fingerprint plots shows that C···H/H···C contact comprises 21.7, 20.9, 12.8, 12.8, 15.0, 16.5 and 13.1% for **1A**, **1B**, **2B**, **5A**, **5B**, **6** and **8B**, respectively. No significant C–H···π interactions were observed for **2A**, **3**, **4**, **7A**, **7B** and **8A**, with C···H close contacts varying from 11.7% in **4** to 15.2% in **2A**. These C–H contacts are mainly due to C–H···N/C–H···O interactions. A significant difference between the molecular interactions in **1–8** in terms of H···H contacts is reflected in the distribution of scattered points in the fingerprint plots, which are $d_i = d_e$ = 1.192 Å in **1A**, 1.197 Å in **1B**, 1.124 Å in **2A**, 1.121 Å in **2B**, 1.128 Å in **3**, 1.151 Å in **4**, 1.015 Å in **5A**, 1.011 Å in **5B**, 1.132 Å in **6**, 1.106 Å in **7A**, 1.087 Å in **7B**, 1.111 Å in **8A** and 1.197 Å in **8B**, and which contributes 47.8%, 48.8%, 37.4%, 39.4%, 37.4%, 36.8%, 37.0%, 31.6%, 45.9%, 30.0%, 31.9%, 26.7% and 24.9% of the total Hirshfeld surface area. The relative contributions for the variety of contacts calculated by Hirshfeld surface analysis are summarised in Fig. 9. The contribution of N–H···N *versus* N–H···O hydrogen bonding has been attributed to the electronegativity of the aryl-bound substituents, which in turn facilitate the formation of different supramolecular chains, leading to diverse crystal packing arrangements.

## Conclusions

Supramolecular chains are formed in each of **1–8** that are sustained by N–H···N hydrogen bonding only (**1** and **3–6**), a combination of N–H···N and N–H···O hydrogen bonding (**2** and **8**) and N–H···O hydrogen bonding only (**7**). The crystal packing falls in two distinct classes, an observation correlated with the *syn*-disposition of successive aryl rings in the supramolecular

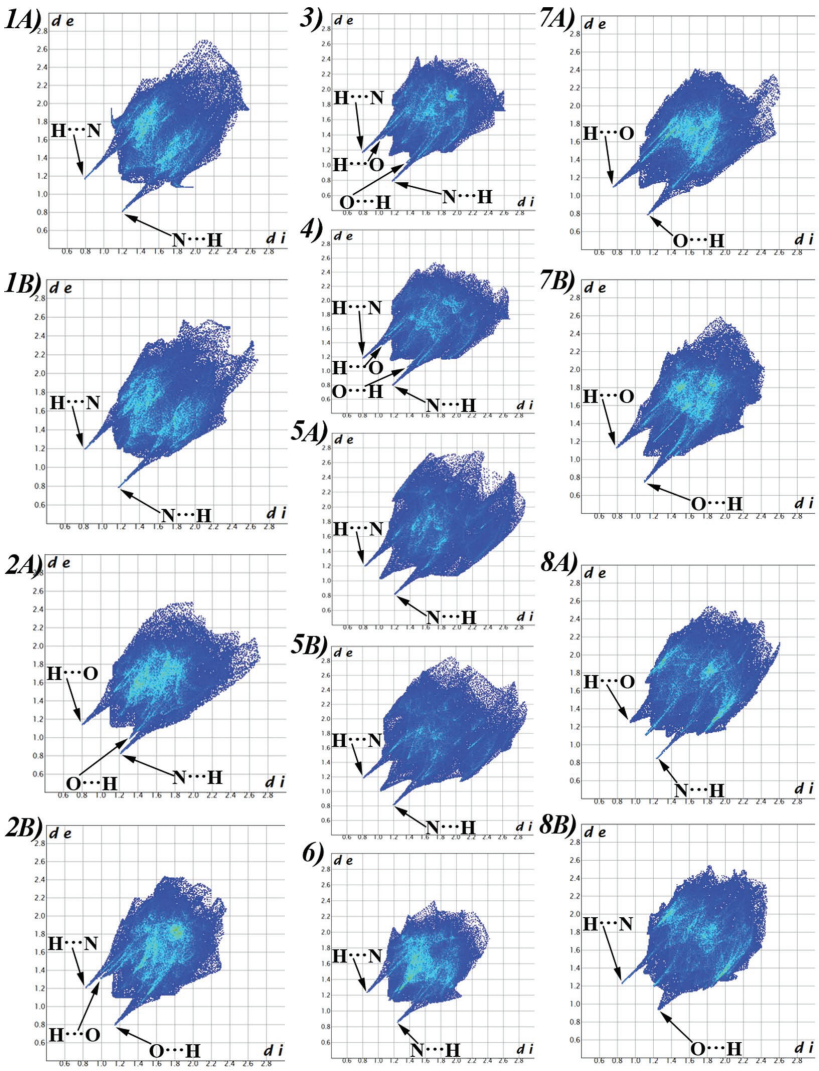



**Fig. 8** Two dimensional fingerprint plots of compounds **1–8**. For compounds that have two moieties in the asymmetric unit these are denoted by '***A***' and '***B***'.

chain when viewed down the propagating axis (**1**, **3–5** and **8**), by contrast to an *anti*-disposition (**2**, **6** and **7**). The former disposition allows for the inter-digitation of layers comprising supramolecular chains and connections of the type C–X···$\pi$(aryl) between them for **3–5** and **8.** Layers can be discerned in the remaining crystal structures and connections between them also involving the *anti*-disposed Y substituents, *i.e.* F···H in **2** and C–H···O in each of **6** and **7**. No correlation between the supramolecular chain motif and crystal packing is evident.

While no quantitative correlations could be made between the adopted supramolecular motif (*i.e.* sustained by either N–H···N or N–H···O interactions or a combination of these) and the theoretical electronic structures of the optimised gas-phase molecules, a qualitative trend based on the relative electronegativity of the Y substituent was found in that N–H···N hydrogen bonding was found exclusively for structures having the least electronegative substituents and N–H···O

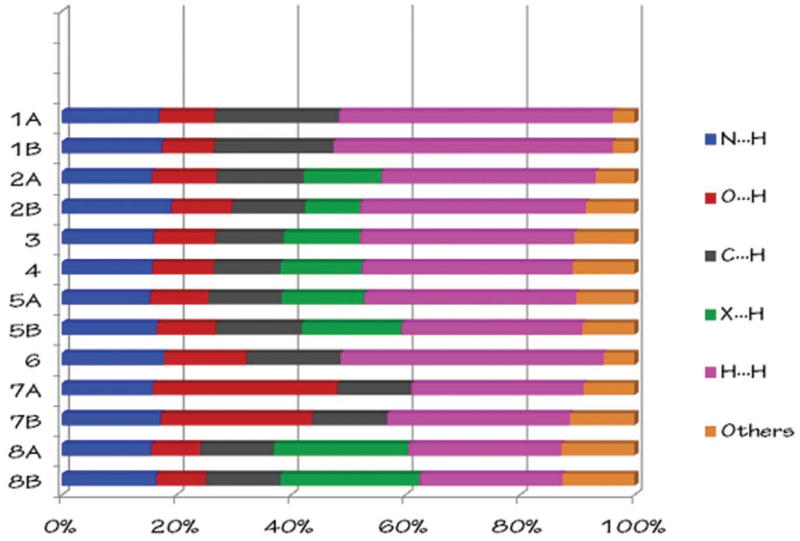



**Fig. 9** The relative contributions of various intermolecular interactions to the Hirshfeld surface area in compounds **1–8**.

hydrogen bonding occurred once the electronegativity of Y increased.

## Experimental

### Synthesis

The compounds were obtained from reactions of substituted phenylhydrazines and ethyl 2-diazoacetoacetate, as previously reported.[7a] For the structural study, the compounds were re-crystallized from their respective EtOH solutions.

### X-ray crystallography

Data for **1**, **3** and **4** were measured at 120 K on a Bruker-Nonius FR591 diffractometer equipped with a 95 mm CCD camera on a $\kappa$-goniostat, employing Mo K$\alpha$ radiation ($\lambda$ = 0.71073 Å) at the EPSRC National crystallographic service at the University of Southampton, UK.[24] Data collection, data processing and cell refinement and absorption correction were accomplished with COLLECT,[25a] the COLLECT and DENZO software combination,[25b] and SADABS,[25c] respectively. Intensity data for **2** ($\lambda$ = 0.6911 Å), **5** and **8** ($\lambda$ = 0.6893 Å) were also collected at 120 K but on a Bruker SMART APEX2 CCD using synchrotron radiation. The data sets were reduced using standard methods,[25a,b] and corrected for absorption in the case of **5** based on multiple scans.[25c] Intensity data for **6** and **7** were measured at 98 K on a Rigaku AFC12/Saturn724 CCD fitted with Mo K$\alpha$ radiation. Data processing and absorption correction were accomplished with Crystal Clear[26a] and ABSCOR,[26b] respectively. The structures were solved by direct-methods with SHELXS-97[27a] and refinement (anisotropic displacement parameters, hydrogen atoms in the riding model approximation and a weighting scheme of the form $w = 1/[\sigma^2(F_o^2) + (aP)^2 + bP]$ for $P = (F_o^2 + 2F_c^2)/3$) was on $F^2$ by means of SHELXL-97.[27a] In the refinement of **2**, three reflections were omitted from the final refinement, *i.e.* (214), (516) and (2012), owing to poor agreement. In the refinement of **4**, two reflections apparently affected by the beam-stop, *i.e.* (100) and (110), were removed from the final refinement. The absolute structure of **1** could not be determined and, hence, 2525 Friedel pairs included in the data set were merged in the final refinement. The absolute structure of **5** was determined on the basis of differences in 2449 Friedel pairs included in the data set (Flack parameter[27b] = 0.10(3)). Crystallographic data and final refinement details are given in Table 4. Fig. S(1)–S(8), ESI,† were drawn with ORTEP-3 for Windows[27c] at the 50% probability level, overlap diagrams were generated with QMol[27d] and the remaining crystallographic figures were drawn with DIAMOND using arbitrary spheres.[27e] Data manipulation and interpretation were with WinGX[27c] and PLATON.[14]

### Computational study

Geometry optimisation was performed starting from the experimentally determined fractional atomic coordinates using Gaussian09.[11] A frequency calculation confirmed that each optimised structure was a true minimum. A combination of Becke's three parameters exchange functional (B3)[28a] with the exchange functional (LYP)[28b] makes up the B3LYP hybrid density functional theory (DFT) method employed in this study. The 6-311+G(d,p) basis set was employed.[28c]

### Hirshfeld surface analysis

Molecular Hirshfeld surfaces[19–21] in a crystal structure are constructed based on the electron distribution calculated as

**Table 4** Crystallographic data and refinement details for **1–8**

| Compound | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| Formula | $C_{12}H_{14}N_4O_2$ | $C_{12}H_{13}FN_4O_2$ | $C_{12}H_{13}ClN_4O_2$ | $C_{12}H_{13}BrN_4O_2$ |
| Formula weight | 246.27 | 264.26 | 280.71 | 325.17 |
| Temperature/K | 120 | 120 | 120 | 120 |
| Crystal colour | Brown | Colourless | Colourless | Light-brown |
| Crystal size/mm$^3$ | 0.20 × 0.40 × 0.70 | 0.05 × 0.10 × 0.15 | 0.18 × 0.25 × 0.40 | 0.14 × 0.30 × 0.36 |
| Crystal system | Orthorhombic | Monoclinic | Monoclinic | Monoclinic |
| Space group | $Pca2_1$ | $P2_1/c$ | $P2_1/c$ | $P2_1/c$ |
| $a$/Å | 22.1366(6) | 10.057(3) | 12.6207(7) | 12.7435(3) |
| $b$/Å | 10.4274(3) | 17.510(4) | 9.8187(5) | 9.9927(2) |
| $c$/Å | 10.7186(3) | 14.102(4) | 11.0587(7) | 11.1150(3) |
| α (°) | 90 | 90 | 90 | 90 |
| $\beta$ (°) | 90 | 95.722(3) | 105.846(3) | 105.3049(13) |
| $\gamma$ (°) | 90 | 90 | 90 | 90 |
| $V$/Å$^3$ | 2474.14(12) | 2471.0(12) | 1318.31(13) | 1365.17(6) |
| $Z$ | 8 | 8 | 4 | 4 |
| $D_c$/g cm$^{-3}$ | 1.322 | 1.421 | 1.414 | 1.582 |
| $F$(000) | 1040 | 1104 | 584 | 656 |
| $\mu$(Mo Kα)/mm$^{-1}$ | 0.094 | 0.065 | 0.294 | 3.015 |
| Measured data | 19 967 | 21 852 | 14 912 | 20 395 |
| $\theta$ range (°) | 3.3–27.5 | 2.6–27.5 | 3.0–27.5 | 2.8–27.5 |
| Unique data | 2976 | 6059 | 3031 | 3126 |
| Observed data ($I \geqslant 2.0\sigma(I)$) | 2461 | 4711 | 2127 | 2583 |
| No. parameters | 335 | 353 | 177 | 177 |
| $R$, obs. data; all data | 0.039; 0.054 | 0.064; 0.081 | 0.047; 0.080 | 0.028; 0.041 |
| $a$; $b$ in weighting scheme | 0.050; 0.289 | 0.094; 1.369 | 0.071; 0.270 | 0.036; 0.669 |
| GoF | 1.04 | 1.11 | 1.03 | 1.02 |
| $R_w$, obs. data; all data | 0.085; 0.092 | 0.174; 0.187 | 0.117; 0.134 | 0.066; 0.071 |
| Range of residual electron density peaks/e Å$^{-3}$ | −0.25–0.18 | −0.25–0.36 | −0.41–0.23 | −0.54–0.28 |
| Compound | **5** | **6** | **7** | **8** |
| Formula | $C_{12}H_{13}IN_4O_2$ | $C_{13}H_{16}N_4O_3$ | $C_{12}H_{13}N_5O_4$ | $C_{12}H_{12}Cl_2N_4O_2$ |
| Formula weight | 372.16 | 276.30 | 291.27 | 315.16 |
| Temperature/K | 120 | 98 | 98 | 120 |
| Crystal colour | Colourless | Orange | Yellow | Colourless |
| Crystal size/mm$^3$ | 0.02 × 0.03 × 0.04 | 0.02 × 0.18 × 0.20 | 0.02 × 0.06 × 0.30 | 0.01 × 0.02 × 0.02 |
| Crystal system | Orthorhombic | Monoclinic | Monoclinic | Triclinic |
| Space group | $Pca2_1$ | $P2_1/c$ | $P2_1/c$ | $P\bar{1}$ |
| $a$/Å | 11.033(2) | 8.965(4) | 10.1622(19) | 7.2818(11) |
| $b$/Å | 10.411(2) | 9.911(4) | 17.127(3) | 11.242(3) |
| $c$/Å | 24.957(5) | 15.067(7) | 15.487(3) | 16.916(4) |
| α (°) | 90 | 90 | 90 | 87.87(2) |
| $\beta$ (°) | 90 | 96.587(14) | 92.476(7) | 88.86(4) |
| $\gamma$ (°) | 90 | 90 | 90 | 88.29(2) |
| $V$/Å$^3$ | 2866.7(10) | 1329.9(10) | 2693.0(9) | 1383.0(5) |
| $Z$ | 8 | 4 | 8 | 4 |
| $D_c$/g cm$^{-3}$ | 1.725 | 1.380 | 1.437 | 1.514 |
| $F$(000) | 1456 | 584 | 1216 | 648 |
| $\mu$(Mo Kα)/mm$^{-1}$ | 2.238 | 0.101 | 0.111 | 0.476 |
| Measured data | 20 053 | 13 169 | 21 138 | 11 827 |
| $\theta$ range (°) | 2.5–24.2 | 2.3–26.5 | 2.0–25.0 | 2.2–26.5 |
| Unique data | 5041 | 2753 | 4750 | 6043 |
| No. parameters | 353 | 187 | 389 | 371 |
| Observed data ($I \geqslant 2.0\sigma(I)$) | 4113 | 2427 | 3964 | 5182 |
| $R$, obs. data; all data | 0.036; 0.050 | 0.056; 0.066 | 0.070; 0.088 | 0.051; 0.059 |
| $a$; $b$ in weighting scheme | 0.028; 2.972 | 0.047; 0.743 | 0.064; 1.643 | 0.039; 1.856 |
| GoF | 1.01 | 1.17 | 1.18 | 1.11 |
| $R_w$, obs. data; all data | 0.074; 0.079 | 0. 123; 0.129 | 0.150; 0.159 | 0.131; 0.136 |
| Range of residual electron density peaks/e Å$^{-3}$ | −0.52–0.55 | −0.23–0.24 | −0.27–0.27 | −0.30–0.47 |

the sum of spherical atom electron densities.[29] For a given crystal structure and set of spherical atomic electron densities, the Hirshfeld surface is unique.[30] The normalized contact distance ($d_{norm}$) based on both $d_e$ and $d_i$, and the van der Waals radii of the atom, given by eqn (1) enables identification of the regions of particular importance to intermolecular interactions.[19] The value of the $d_{norm}$ is negative or positive when intermolecular contacts are shorter or longer than van der Waals separations, respectively. The combination of $d_e$ and $d_i$ in the form of a two dimensional fingerprint plot[22] provides a summary of the intermolecular contacts in the crystal.[19] The Hirshfeld surfaces, mapped with $d_{norm}$, and two-dimensional

fingerprint plots presented in this paper were generated using CrystalExplorer 2.1.[31]

$$d_{\text{norm}} = \frac{d_i - r_i^{\text{vdw}}}{r_i^{\text{vdw}}} + \frac{d_e - r_e^{\text{vdw}}}{r_e^{\text{vdw}}} \quad (1)$$

## Acknowledgements

The X-ray data sets for **1**, **3** and **4** were collected at the EPSRC X-ray Crystallographic Service, University of Southampton, England; the authors thank the staff of the Service for help and advice. We acknowledge Prof. Bill Clegg and the synchrotron component, at Daresbury, of the EPSRC National Crystallography Service, based at the University of Southampton, England, for data collections of **2**, **5** and **8**. The authors also thank CNPq and CAPES (Brazil) for financial support. The financial support from University Grants Commission, Government of India, Major Research Project (42-830/2013 (SR)) is gratefully acknowledged. The support from the Ministry of Higher Education, Malaysia, High-Impact Research scheme (UM.C/HIR-MOHE/SC/03) is similarly gratefully acknowledged.

## References

1 G. R. Desiraju, *Angew. Chem., Int. Ed.*, 2007, **46**, 8342.
2 D. Das, R. K. R. Jetti, R. Boese and G. R. Desiraju, *Cryst. Growth Des.*, 2003, **3**, 675.
3 H. Adams, P. L. Bernad Jr, D. S. Eggleston, R. C. Haltiwanger, K. D. M. Harris, G. A. Hembury, C. A. Hunter, D. J. Livingstone, B. M. Kariuki and J. F. McCabe, *Chem. Commun.*, 2001, 1500.
4 S. Varughese and S. M. Draper, *Cryst. Growth Des.*, 2010, **10**, 2298.
5 M. Felsmann, F. Eissmann, A. Schwarzer and E. Weber, *Cryst. Growth Des.*, 2011, **11**, 982.
6 (*a*) R. A. Howie, G. M. de Lima, D. C. Menezes, J. L. Wardell, S. M. S. V. Wardell, D. J. Young and E. R. T. Tiekink, *CrystEngComm*, 2008, **10**, 1626; (*b*) C. R. Kaiser, K. C. Pais, M. V. N. De Souza, J. L. Wardell, S. M. S. V. Wardell and E. R. T. Tiekink, *CrystEngComm*, 2009, **11**, 1133.
7 (*a*) A. K. Jordaõ, P. P. Afonso, V. F. Ferreira, M. C. B. V. de Souza, M. C. B. Almeida, C. O. Beltrame, D. P. Paiva, S. M. S. V. Wardell, J. L. Wardell, E. R. T. Tiekink, C. R. Damaso and A. C. Cunha, *Eur. J. Med. Chem.*, 2009, **44**, 3777; (*b*) S. B. Ferreira, M. S. Costa, N. Boechat, R. J. Bezerra, M. S. Genestra, M. M. Canto-Cavalheiro, W. B. Kover and V. F. Ferreira, *Eur. J. Med. Chem.*, 2007, **42**, 1388; (*c*) M. S. Costa, N. Boechat, É. A. Rangel, F. de C. da Silva, A. M. T. de Souza, C. R. Rodrigues, H. C. Castro, I. N. Junior, M. C. S. Lourenço, S. M. S. V. Wardell and V. F. Ferreira, *Bioorg. Med. Chem.*, 2006, **14**, 8644; (*d*) A. C. Cunha, J. M. Figueiredo, J. L. M. Tributino, A. L. P. Miranda, H. C. Castro, R. B. Zingali, C. A. M. Fraga, M. C. B. V. De Souza, V. F. Ferreira and E. J. Barreiro, *Bioorg. Med. Chem.*, 2003, **11**, 2051.
8 (*a*) Y. S. Sanghvi, B. K. Bhattacharya, G. D. Kini, S. S. Matsumoto, S. B. Larson, W. B. Jolley, R. K. Robins and G. R. Revankar, *J. Med. Chem.*, 1990, **33**, 336; (*b*) N. A. Al-Masoudi and Y. A. Al-Soud, *Tetrahedron Lett.*, 2002, **43**, 4021; (*c*) R. Alvarez, S. Velázquez, A. San-Félix, S. Aquaro, E. De Clercq, C. F. Perno, A. Karlsson, J. Balzarini and M. J. Camarasa, *J. Med. Chem.*, 1994, **37**, 4185; (*d*) M. J. Genin, D. A. Allwine, D. J. Anderson, M. R. Barbachyn, D. E. Emmert, S. A. Garmon, D. R. Graber, K. C. Grega, J. B. Hester, D. K. Hutchinson, J. Morris, R. J. Reischer, C. W. Ford, G. E. Zurenko, J. C. Hamel, R. D. Schaadt, D. Stapert and B. H. Yagi, *J. Med. Chem.*, 2000, **43**, 953; (*e*) C. Hager, R. Miethchen and H. Reinke, *J. Fluorine Chem.*, 2000, **104**, 135; (*f*) V. Calderone, I. Giorgi, O. Livi, E. Martinotti, E. Mantuano, A. Martelli and A. Nardi, *Eur. J. Med. Chem.*, 2005, **40**, 521.
9 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380.
10 A. K. Jordão, V. F. Ferreira, A. C. Cunha, J. L. Wardell, S. M. S. V. Wardell and E. R. T. Tiekink, *CrystEngComm*, 2012, **14**, 6534.
11 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUSSIAN 09 (Revision A.1)*, Gaussian, Inc., Wallingford, CT, 2009.
12 (*a*) M. C. Etter, *Acc. Chem. Res.*, 1990, **23**, 120; (*b*) M. C. Etter, J. C. MacDonald and J. Bernstein, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1990, **46**, 256.
13 C. Hansch, A. Leo and R. W. Taft, *Chem. Rev.*, 1991, **91**, 165.
14 A. L. Spek, *J. Appl. Crystallogr.*, 2003, **36**, 7.
15 J. F. Malone, C. M. Murray, M. H. Charlton, R. Docherty and A. J. Lavery, *J. Chem. Soc., Faraday Trans.*, 1997, **93**, 3429.
16 (*a*) M. Egli and R. V. Gessner, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 180; (*b*) T. J. Mooibroek, P. Gamez and J. Reedijk, *CrystEngComm*, 2008, **10**, 1501; (*c*) J. Zukerman-Schpector, I. Haiduc and E. R. T. Tiekink, *Chem. Commun.*, 2011, **47**, 12682; (*d*) J. Zukerman-Schpector, I. Haiduc and E. R. T. Tiekink, *Adv. Organomet. Chem.*, 2012, **60**, 49.
17 (*a*) D. Schollmeyer, O. V. Shishkin, T. Rühl and M. O. Vysotsky, *CrystEngComm*, 2008, **10**, 715; (*b*) O. V. Shishkin, *Chem. Phys. Lett.*, 2008, **458**, 96.
18 (*a*) D. S. Reddy, D. C. Craig and G. R. Desiraju, *J. Am. Chem. Soc.*, 1996, **118**, 4090; (*b*) I. Csoregh, E. Weber, T. Hens and M. Czugler, *J. Chem. Soc., Perkin Trans. 2*, 1996, 2733; (*c*) M. D. Prasanna and T. N. Guru Row, *Cryst. Eng.*, 2000, **3**, 135;

(*d*) T. J. Mooibroek and P. Gamez, *CrystEngComm*, 2012, **14**, 1027; (*e*) R. K. R. Jetti, A. Nangia, F. Xue and T. C. W. Mak, *Chem. Commun.*, 2001, 919; (*f*) Z. Yin, W. Wang, M. Du, X. Wang and J. Guo, *CrystEngComm*, 2009, **11**, 2441; (*g*) A. S. Paton, A. J. Lough and T. P. Bender, *CrystEngComm*, 2011, **13**, 3653.

19 M. A. Spackman and J. J. McKinnon, *CrystEngComm*, 2002, **4**, 378.

20 (*a*) J. J. McKinnon, D. Jayatilaka and M. A. Spackman, *Chem. Commun.*, 2007, 3814; (*b*) M. A. Spackman, J. J. McKinnon and D. Jayatilaka, *CrystEngComm*, 2008, **10**, 377; (*c*) M. A. Spackman and D. Jayatilaka, *CrystEngComm*, 2009, **11**, 19; (*d*) F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129; (*e*) H. F. Clausen, M. S. Chevallier, M. A. Spackman and B. B. Iversen, *New J. Chem.*, 2010, **34**, 193.

21 (*a*) S. K. Seth, *CrystEngComm*, 2013, **15**, 1772; (*b*) S. K. Seth, I. Saha, C. Estarellas, A. Frontera, T. Kar and S. Mukhopadhyay, *Cryst. Growth Des.*, 2011, **11**, 3250; (*c*) S. K. Seth, D. Sarkar, A. D. Jana and T. Kar, *Cryst. Growth Des.*, 2011, **11**, 4837; (*d*) S. K. Seth, D. Sarkar and T. Kar, *CrystEngComm*, 2011, **13**, 4528; (*e*) S. K. Seth, D. Sarkar, A. Roy and T. Kar, *CrystEngComm*, 2011, **13**, 6728.

22 (*a*) A. L. Rohl, M. Moret, W. Kaminsky, K. Claborn, J. J. Mckinnon and B. Kahr, *Cryst. Growth Des.*, 2008, **8**, 4517; (*b*) A. Parkin, G. Barr, W. Dong, C. J. Gilmore, D. Jayatilaka, J. J. Mckinnon, M. A. Spackman and C. C. Wilson, *CrystEngComm*, 2007, **9**, 648.

23 (*a*) M. A. Spackman and P. G. Byrom, *Chem. Phys. Lett.*, 1997, **267**, 215; (*b*) J. J. McKinnon, A. S. Mitchell and M. A. Spackman, *Chem.–Eur. J.*, 1998, **4**, 2136.

24 S. J. Coles and P. A. Gale, *Chem. Sci.*, 2012, **3**, 683.

25 (*a*) R. W. W. Hooft, COLLECT, Nonius BV, Delft, The Netherlands, 1998; (*b*) Z. Otwinowski and W. Minor, *Methods in Enzymology, Macromolecular Crystallography, Part A*, ed. C. W. Carter Jr and R. M. Sweet, Academic Press, New York, 1997, vol. 276, pp. 307–326; (*c*) G. M. Sheldrick, SADABS, Version 2.10, Bruker AXS Inc., Madison, Wisconsin, USA, 2003.

26 (*a*) *CrystalClear*, User Manual, Rigaku/MSC Inc., Rigaku Corporation, The Woodlands, TX, 2005; (*b*) T. Higashi, ABSCOR, Rigaku Corporation, Tokyo, Japan, 1995.

27 (*a*) G. M. Sheldrick, *Acta Crystallogr., Sect A: Found Crystallogr.*, 2008, **64**, 112; (*b*) H. D. Flack, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1983, **39**, 876; (*c*) L. J. Farrugia, *J. Appl. Crystallogr.*, 2012, **45**, 849; (*d*) J. Gans and D. Shalloway, *J. Mol. Graphics Modell.*, 2001, **19**, 557; (*e*) DIAMOND, Visual Crystal Structure Information System, Version 3.1, CRYSTAL IMPACT, Postfach 1251, D-53002 Bonn, Germany, 2006.

28 (*a*) A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098; (*b*) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785; (*c*) J. B. Foresman and A. E. Frisch, *Exploring Chemistry with Electronic Structural Methods: A Guide to Using Gaussian*, Gaussian, Pittsburgh, Pa, 2nd edn, 1996.

29 (*a*) C. A. Hunter, J. Singh and J. M. Thornton, *J. Mol. Biol.*, 1991, **218**, 837; (*b*) J. Singh and J. M. Thornton, *J. Mol. Biol.*, 1990, **211**, 595.

30 J. J. McKinnon, M. A. Spackman and A. S. Mitchell, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2004, **60**, 627.

31 S. K. Wolff, D. J. Grimwood, J. J. McKinnon, D. Jayatilaka and M. A. Spackman, *Crystal Explorer 2.1*, University of Western Australia, Perth, Australia, 2007.

# BIBLIOGRAPHY


1. (a) Ball, P., Water as an active constituent in cell biology. *Chem. Rev.* **2008,** *108* (1), 74-108;

   (b) Chaplin, M., Opinion - Do we underestimate the importance of water in cell biology? *Nat. Rev. Mol. Cell Biol.* **2006,** *7* (11), 861-866;

   (c) Cheung, M. S.; Garcia, A. E.; Onuchic, J. N., Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (2), 685-690;

   (d) Fuxreiter, M.; Mezei, M.; Simon, I.; Osman, R., Interfacial water as a "Hydration fingerprint" in the noncognate complex of BamHI. *Biophys. J.* **2005,** *89* (2), 903-911;

   (e) Tait, M. J.; Franks, F., Water in biological systems. *Nature* **1971,** *230* (5289), 91-94.

2. Millero, F. J.; Curry, R. W.; Drosthan.W, Isothermal compressibility of water at various temperatures. *Journal of Chemical and Engineering Data* **1969,** *14* (4), 422-425.

3. Kell, G. S., Density, thermal expansivity, and compressibility of liquid water from 0 degrees to 150 degreesc - correlations and tables for atmospheric-pressure and saturation reviewed and expressed on 1968 temperature scale. *Journal of Chemical and Engineering Data* **1975,** *20* (1), 97-105.

4. (a) Demerdash, O.; Yap, E. H.; Head-Gordon, T., Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu. Rev. Phys. Chem.* **2014,** *65*, 149-174;

   (b) Finney, J. L., The water molecule and its interactions: the interaction between theory, modelling, and experiment. *J. Mol. Liq.* **2001,** *90* (1-3), 303-312;

   (c) Guillot, B., A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **2002,** *101* (1-3), 219-260;

   (d) Halgren, T. A.; Damm, W., Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001,** *11* (2), 236-242;

(e) Szalewicz, K.; Leforestier, C.; van der Avoird, A., Towards the complete understanding of water by a first-principles computational approach. *Chem. Phys. Lett.* **2009,** *482* (1-3), 1-14.

5. DeLucia, F. C.; Helminger, P.; Kirchhoff, W. H., *Journal of Physical and Chemical Reference Data* **1974,** *3*, 211.

6. Avila, G., *Ab initio* dipole polarizability surfaces of water molecule: Static and dynamic at 514.5 nm. *J. Chem. Phys.* **2005,** *122* (14), 10.

7. Johnson III, R. D., NIST Computational Chemistry Comparison and Benchmark Database. August 2013 ed.; NIST: 2013.

8. Weast, R. C., *CRC Handbook of Chemistry and Physics*. 64 ed.; Boca Raton: Florida, 1983.

9. Ouyang, J. F.; Bettens, R. P. A., Modelling Water: A Lifetime Enigma. *Chimia* **2015,** *Submitted*.

10. (a) Abascal, J. L. F.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005,** *123* (23);

(b) Habershon, S.; Markland, T. E.; Manolopoulos, D. E., Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* **2009,** *131* (2);

(c) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120* (20), 9665-9678;

(d) Jorgensen, W. L., Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers - application to liquid water. *Journal of the American Chemical Society* **1981,** *103* (2), 335-340;

(e) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935;

(f) Mahoney, M. W.; Jorgensen, W. L., A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000,** *112* (20), 8910-8922;

(g) Mahoney, M. W.; Jorgensen, W. L., Quantum, intramolecular flexibility, and polarizability effects on the reproduction of the density anomaly of liquid water by simple potential functions. *J. Chem. Phys.* **2001,** *115* (23), 10758-10768;

(h) Rick, S. W.; Stuart, S. J.; Berne, B. J., Dynamical fluctuating charge force-fields - application to liquid water. *J. Chem. Phys.* **1994,** *101* (7), 6141-6156.

11. (a) Berendsen, H. J. C.; M., P. J. P.; von Gunstaren, W. F.; Hermans, J., *Intermolecular Forces*. Reidel: Dordrecht, Holland., 1981;

(b) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271;

(c) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, T. E.; Voth, G. A., An accurate and simple quantum model for liquid water. *J. Chem. Phys.* **2006,** *125* (18);

(d) Wu, Y. J.; Tepper, H. L.; Voth, G. A., Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006,** *124* (2).

12. (a) Fellers, R. S.; Leforestier, C.; Braly, L. B.; Brown, M. G.; Saykally, R. J., Spectroscopic determination the water pair potential. *Science* **1999,** *284* (5416), 945-948;

(b) Goldman, N.; Fellers, R. S.; Brown, M. G.; Braly, L. B.; Keoshian, C. J.; Leforestier, C.; Saykally, R. J., Spectroscopic determination of the water dimer intermolecular potential-energy surface. *J. Chem. Phys.* **2002,** *116* (23), 10148-10163;

(c) Millot, C.; Soetens, J. C.; Costa, M.; Hodges, M. P.; Stone, A. J., Revised anisotropic site potentials for the water dimer and calculated properties. *J. Phys. Chem. A* **1998,** *102* (4), 754-770;

(d) Millot, C.; Stone, A. J., Towards an accurate intermolecular potential for water. *Mol. Phys.* **1992,** *77* (3), 439-462.

13. (a) Bukowski, R.; Szalewicz, K.; Groenenboom, G.; van der Avoird, A., Interaction potential for water dimer from symmetry-adapted perturbation theory based on density functional description of monomers. *J. Chem. Phys.* **2006,** *125* (4);
(b) Groenenboom, G. C.; Wormer, P. E. S.; van der Avoird, A.; Mas, E. M.; Bukowski, R.; Szalewicz, K., Water pair potential of near spectroscopic accuracy. II. Vibration-rotation-tunneling levels of the water dimer. *J. Chem. Phys.* **2000,** *113* (16), 6702-6715;
(c) Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation-theory approach to intermolecular potential-energy surfaces of van-der-waals complexes. *Chem. Rev.* **1994,** *94* (7), 1887-1930;
(d) Mas, E. M.; Bukowski, R.; Szalewicz, K., *Ab initio* three-body interactions for water. I. Potential and structure of water trimer. *J. Chem. Phys.* **2003,** *118* (10), 4386-4403;
(e) Mas, E. M.; Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Wormer, P. E. S.; van der Avoird, A., Water pair potential of near spectroscopic accuracy. I. Analysis of potential surface and virial coefficients. *J. Chem. Phys.* **2000,** *113* (16), 6687-6701;
(f) Mas, E. M.; Szalewicz, K.; Bukowski, R.; Jeziorski, B., Pair potential for water from symmetry-adapted perturbation theory. *J. Chem. Phys.* **1997,** *107* (11), 4207-4218.

14. (a) Burnham, C. J.; Li, J. C.; Xantheas, S. S.; Leslie, M., The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters (n=2-21) and the phonon spectrum of ice Ih. *J. Chem. Phys.* **1999,** *110* (9), 4566-4581;
(b) Thole, B. T., Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981,** *59* (3), 341-350;
(c) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F., The vibrational proton potential in bulk liquid water and ice. *J. Chem. Phys.* **2008,** *128* (15);
(d) Burnham, C. J.; Xantheas, S. S., Development of transferable interaction models for water. IV. A flexible, all-atom polarizable potential (TTM2-F) based on geometry dependent charges derived from an *ab initio* monomer dipole moment surface. *J. Chem. Phys.* **2002,** *116* (12), 5115-5124;

(e) Burnham, C. J.; Xantheas, S. S., Development of transferable interaction models for water. III. Reparametrization of an all-atom polarizable rigid model (TTM2-R) from first principles. *J. Chem. Phys.* **2002,** *116* (4), 1500-1510;

(f) Fanourgakis, G. S.; Xantheas, S. S., The flexible, polarizable, thole-type interaction potential for water (TTM2-F) revisited. *J. Phys. Chem. A* **2006,** *110* (11), 4100-4106;

(g) Fanourgakis, G. S.; Xantheas, S. S., Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable, Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *J. Chem. Phys.* **2008,** *128* (7);

(h) Partridge, H.; Schwenke, D. W., The determination of an accurate isotope dependent potential energy surface for water from extensive *ab initio* calculations and experimental data. *J. Chem. Phys.* **1997,** *106* (11), 4618-4639;

(i) Ren, P. Y.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003,** *107* (24), 5933-5947.

15. (a) Huang, X.; Braams, B. J.; Bowman, J. M.; Kelly, R. E. A.; Tennyson, J.; Groenenboom, G. C.; van der Avoird, A., New *ab initio* potential energy surface and the vibration-rotation-tunneling levels of $(H_2O)_2$ and $(D_2O)_2$. *J. Chem. Phys.* **2008,** *128* (3);

(b) Huang, X. C.; Braams, B. J.; Bowman, J. M., *Ab initio* potential energy and dipole moment surfaces of (H2O)(2). *J. Phys. Chem. A* **2006,** *110* (2), 445-451;

(c) Shank, A.; Wang, Y.; Kaledin, A.; Braams, B. J.; Bowman, J. M., Accurate *ab initio* and "hybrid" potential energy surfaces, intramolecular vibrational energies, and classical ir spectrum of the water dimer. *J. Chem. Phys.* **2009,** *130* (14);

(d) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M., Flexible, *ab initio* potential, and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer. *J. Chem. Phys.* **2011,** *134* (9).

16. (a) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A., Polarizable interaction potential for water from coupled cluster calculations. I. Analysis of dimer potential energy surface. *J. Chem. Phys.* **2008,** *128* (9);

(b) Cencek, W.; Szalewicz, K.; Leforestier, C.; van Harrevelt, R.; van der Avoird, A., An accurate analytic representation of the water pair potential. *Phys. Chem. Chem. Phys.* **2008,** *10* (32), 4716-4731;

(c) Gora, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K., Predictions for water clusters from a first-principles two- and three-body orce field. *J. Chem. Phys.* **2014,** *140* (19);

(d) Leforestier, C.; Szalewicz, K.; van der Avoird, A., Spectra of water dimer from a new *ab initio* potential with flexible monomers. *J. Chem. Phys.* **2012,** *137* (1).

17. (a) Babin, V.; Leforestier, C.; Paesani, F., Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.* **2013,** *9* (12), 5395-5403;

(b) Babin, V.; Medders, G. R.; Paesani, F., Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *J. Chem. Theory Comput.* **2014,** *10* (4), 1599-1607;

(c) Medders, G. R.; Babin, V.; Paesani, F., A Critical Assessment of Two-Body and Three-Body Interactions in Water. *J. Chem. Theory Comput.* **2013,** *9* (2), 1103-1114;

(d) Medders, G. R.; Babin, V.; Paesani, F., Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.* **2014,** *10* (8), 2906-2910.

18. Leicester, H. M., *The Historical Background of Chemistry*. Dover Inc.: New York, 1956.

19. Jensen, F., *Introduction to Computational Chemistry*. 2nd ed.; John Wiley & Sons, Ltd.: West Sussex, 2007; p 599.

20. Kobus, J., Diatomic molecules: Exact solutions of HF equations. In *Advances in Quantum Chemistry, Vol. 28: Recent Advances in Computational Chemistry*,

Lowdin, P. O.; Sabin, J. R.; Zerner, M. C.; Karwowski, J.; Karelson, M., Eds. 1997; Vol. 28, pp 1-14.

21. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian*, 09, Revision C.01; Gaussian, Inc.: 340 Quinnipiac St., Bldg. 40, Wallingford CT 06492, 2010.

22. Stone, A. J., *The Theory of Intermolecular Forces*. Clarendon Press: Oxford, 2002; p 264.

23. Casimir, H. B. G.; Polder, D., The influence of retardation on the london-vanderwaals forces. *Physical Review* **1948,** *73* (4), 360-372.

24. Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansik, B.; Jensen, H. J. A.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.;

Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V.; Salek, P.; Samson, C. C. M.; Merás, A. S. d.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. *DALTON2013*, WIREs Comput. Mol. Sci.: 2013.

25. (a) Stone, A. J.; Alderton, M., Distributed multipole analysis - methods and applications. *Mol. Phys.* **1985,** *56* (5), 1047-1064;
(b) Vignemaeder, F.; Claverie, P., The exact multicenter multipolar part of a molecular charge-distribution and its simplified representations. *J. Chem. Phys.* **1988,** *88* (8), 4934-4948.

26. Stone, A. J., Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005,** *1* (6), 1128-1132.

27. Soper, A. K., The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics* **2000,** *258* (2-3), 121-137.

28. Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V., Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012,** *112* (1), 632-672.

29. Collins, M. A.; Cvitkovic, M. W.; Bettens, R. P. A., The Combined Fragmentation and Systematic Molecular Fragmentation Methods. *Acc. Chem. Res.* **2014**.

30. Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R., Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. *J. Chem. Phys.* **2006,** *125* (10).

31. Bettens, R. P. A.; Lee, A. M., On the accurate reproduction of *ab initio* interaction energies between an enzyme and substrate. *Chem. Phys. Lett.* **2007,** *449* (4-6), 341-346.

32. Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schutz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Koppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pfluger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO*, 2012.1; 2012.

33. Stone, A. J. *GDMA*, 2.2.01; University of Cambridge: Cambridge, 2006.

34. Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; A., P. P. L.; Wales, D. J. *Orient*, 4.5; University of Cambridge: Cambridge, 2002.

35. (a) Boys, S. F.; Bernardi, F., Calculation of small molecular interactions by differences of separate total energies - some procedures with reduced errors. *Mol. Phys.* **1970,** *19* (4), 553-&;
(b) Simon, S.; Duran, M.; Dannenberg, J. J., How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *J. Chem. Phys.* **1996,** *105* (24), 11024-11031.

36. Cobar, E. A.; Horn, P. R.; Bergman, R. G.; Head-Gordon, M., Examination of the hydrogen-bonding networks in small water clusters (n=2-5, 13, 17) using absolutely localized molecular orbital energy decomposition analysis. *Phys. Chem. Chem. Phys.* **2012,** *14* (44), 15328-15339.

37. Richard, R. M.; Herbert, J. M., A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.* **2012,** *137* (6), 17.

38. Cunha, A. C.; Ferreira, V. F.; Jordao, A. K.; de Souza, M.; Wardell, S.; Wardell, J. L.; Tan, P. A.; Bettens, R. P. A.; Seth, S. K.; Tiekink, E. R. T., Aryl-substituents

moderate the nature of hydrogen bonds, N-H center dot center dot center dot N versus N-H center dot center dot center dot O, leading to supramolecular chains in the crystal structures of N-arylamino 1,2,3-triazole esters. *Crystengcomm* **2013,** *15* (24), 4917-4929.

39. Desiraju, G. R., Crystal engineering: A holistic view. *Angewandte Chemie-International Edition* **2007,** *46* (44), 8342-8356.

40. Das, D.; Jetti, R. K. R.; Boese, R.; Desiraju, G. R., Stereoelectronic effects of substituent groups in the solid state. Crystal chemistry of some cubanecarboxylic and phenylpropiolic acids. *Crystal Growth & Design* **2003,** *3* (5), 675-681.

41. Adams, H.; Bernad, P. L.; Eggleston, D. S.; Haltiwanger, R. C.; Harris, K. D. M.; Hembury, G. A.; Hunter, C. A.; Livingstone, D. J.; Kariuki, B. M.; McCabe, J. F., Substituent effects on aromatic interactions in the solid state. *Chemical Communications* **2001,** (16), 1500-1501.

42. Varughese, S.; Draper, S. M., Solid State Conformational Preferences of a Flexible Molecular Backbone Derived from Acetone: Dependence on Electron Donating/Withdrawing Ability of Substitutions. *Crystal Growth & Design* **2010,** *10* (5), 2298-2305.

43. Felsmann, M.; Eissmann, F.; Schwarzer, A.; Weber, E., Competitive Interactions in the Crystal Structures of Benzils Effected by Different Halogen Substitution. *Crystal Growth & Design* **2011,** *11* (4), 982-989.

44. (a) Howie, R. A.; de Lima, G. M.; Menezes, D. C.; Wardell, J. L.; Wardell, S.; Young, D. J.; Tiekink, E. R. T., The influence of cation upon the supramolecular aggregation patterns of dithiocarbamate anions functionalised with hydrogen bonding capacity-the prevalence of charge-assisted O-H center dot center dot center dot S interactions. *Crystengcomm* **2008,** *10* (11), 1626-1637;

(b) Kaiser, C. R.; Pais, K. C.; de Souza, M. V. N.; Wardell, J. L.; Wardell, S.; Tiekink, E. R. T., Assessing the persistence of the N-H center dot center dot

center dot N hydrogen bonding leading to supramolecular chains in molecules related to the anti-malarial drug, chloroquine. *Crystengcomm* **2009,** *11* (6), 1133-1140.

45. Hansch, C.; Leo, A.; Taft, R. W., A survey of hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991,** *91* (2), 165-195.

# Appendix A

# Geometry of Water Clusters

The geometries of water clusters $A - E$ studied in Chapter 3 can be found in the CD accompanying this thesis. The file is entitled "Appendix A.docx", and consists of 41 pages.

# Appendix B

# Charges and Geometries of *N*-arylamino Compounds

The optimised geometries, Mulliken charges and natural charges of the *N*-arylamino 1,2,3-triazole esters studied in Chapter 6, i.e., compounds 1 – 8 illustrated in Figure 6-1 and Figure 6-2 can be found on the CD accompanying this thesis contained in the file entitled “Appendix B.docx”.

# CD-ROM

The attached CD-ROM contains this copy of the thesis, Appendices A and B: