

# ENERGY LANDSCAPE EXPLORATION AND GLOBAL OPTIMISATION OF HYDRATED MOLECULAR CLUSTERS

By

JOHN CHRISTOPHER HEY

A thesis submitted to  
the University of Birmingham  
for the degree of  
DOCTOR OF PHILOSOPHY



Physical and Theoretical Chemistry  
School of Chemistry  
College of Engineering and Physical Sciences  
University of Birmingham  
September 2021

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.



## ABSTRACT

The Hofmeister series is a ranking of ions based on their ability to affect various properties of proteins and other biological macro-molecules in aqueous solutions. Initially described by Franz Hofmeister in the late 19<sup>th</sup> century, the series categorises ions according to their effects on the solubility, stability, and activity of proteins. It is often used to understand and predict the behaviour of proteins in different ionic environments, particularly in the context of protein precipitation, folding, and stability. In this thesis I present a systematic study of the stepwise hydration of simple ions within the Hoffmeister series with the aim of elucidating both their structures and their energy landscapes. This study was undertaken using parallel basin hopping global optimisation with a hybrid of potentials, a cheap OPLS-like (Optimised Potentials for Liquid Simulations) potential was used to generate large databases of candidate minima. These putative minima were then reoptimised to a much tighter tolerance using a dispersion-corrected DFT scheme, to yield good minima that are shown to reproduce various experimental results. This technique has allowed us to provide good putative minima to larger sizes of clusters in a systematic manner than would otherwise be available. We establish common structural motifs across a variety of ions from across the Hoffmeister series, and show the effects of interpolating charges between two candidate ions has on the structural motifs present.

We then extend this study to encompass metal ions. We study the stepwise hydration of AuAg nanoclusters as cations, anions and uncharged clusters. This part of the study was conducted using the MEGA/GIGA genetic algorithms, directly at the DFT level. I reproduce some previous experimental infrared spectra for the anionic clusters, and then go on to present novel structures for the cationic system and the neutral system. I then categorise the structural motifs present in the clusters and compare these to those seen in the Hoffmeister ion study and present some computed infrared spectra.

Finally, I present a novel fusion of a parallel pool genetic algorithm with a Thresholding Algorithm I am calling BMPGA. The genetic algorithm is an entirely new implementation of



---

the old BPGA written specifically with the optimisation of molecular clusters in mind. BMPGA implements a modified Deavan and Ho crossover scheme optimised for molecular systems, has massive thread-safe parallelism built-in by design, and implements a number of common and less common mutation schemes.<sup>1</sup> The mutation schemes include the old classics such as single and multiple translations and rotations, as well as some more computationally demanding mutations such as short, high temperature Monte Carlo and molecular dynamics simulations. The hybridisation with the Thresholding scheme comes in to play in two different ways; it is available as an optional mutation scheme in the form of Threshold Monte Carlo. These mutations allow for the system to undergo infinite-temperature MC runs with a total-system energy cap applied the final configuration obtained is then subjected to a number of stochastic quenches. The other way that Thresholding is incorporated into the GA is that it can be applied to the population of minima within the GA itself, this leads to ensemble Thresholding and I show that this method allows for very rapid exploration of the landscape by allowing the population to escape wells that it may otherwise become trapped in. By lowering the energy cap over the course of the optimisation, good performance of this method is seen. I present benchmark results for this new GA using a variety of systems and potentials, from simple Lennard-Jones clusters, through some DFTB results to direct landscape exploration at the DFT level. I show that the parallel performance of this method is extremely good, and scales very well to a large number of CPU cores. This new GA method shows promise for allowing very large-scale optimisations to be performed on a large range of systems, only limited by the availability of HPC resources.

## DEDICATION

This thesis is dedicated to Professor Roy L. Johnston, a great teacher, mentor, and friend.

## ACKNOWLEDGMENTS

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my PhD! My supervisor Professor Roy L. Johnston, without whose guidance and knowledge none of this would have been possible. Dr. Dwaipayan Chakrabarti, who was always available for advice and who kindly took over my supervision when needed especially at those times I was really struggling with my mental health. The various temporary and permanent members of the Johnston and Chakrabarti groups, including James, Helen, Chris, Alejandra, Lewis, Dan, Jack, Heider, Andreas, Edward, Jay, and of course Wesley. The University of Birmingham's School of Chemistry and its members (especially Dr. M. Britton, Dr J. Wilkie, and Dr. M. Tremayne) for their invaluable support. Prof. Christian Schön, of the Max Planck Institute who has been an invaluable mentor. Financially this PhD would not have been possible without the support of the University of Birmingham and the EPSRC. My family have provided constant support throughout and none of this would have been feasible without them. My now-wife Dr. Ruth Maganga has been a great source of comfort and has made the writing of this thesis slightly more bearable.

# Table of Contents

	Page
<b>1 Introduction</b> . . . . .	1
1.1 Hydrated Ions . . . . .	1
1.1.1 Water and Ice . . . . .	1
1.1.2 Hydrated Ions . . . . .	2
1.2 Energy Landscapes . . . . .	5
1.2.1 Features of Energy Landscapes . . . . .	6
1.2.2 Interpreting the Topologies of Energy Landscapes . . . . .	8
1.2.3 Egodicity . . . . .	9
<b>2 Methodologies</b> . . . . .	11
2.1 Energy Evaluation . . . . .	11
2.1.1 Empirical Methods . . . . .	12
2.1.2 Semi-Empirical Methods . . . . .	14
2.1.3 Density Functional Theory (DFT) . . . . .	14
2.1.4 <i>Ab initio</i> Methods . . . . .	15
2.2 Global Optimisation . . . . .	16
2.2.1 Local Optimisation . . . . .	17
2.2.2 Basin-Hopping Monte Carlo . . . . .	18
<b>3 Hydrated Ions</b> . . . . .	20
3.1 Hydrated Ions . . . . .	20
3.2 Introduction . . . . .	20
3.2.1 Water and Ice . . . . .	20
3.2.2 Hydrated Ions . . . . .	21
3.3 Method . . . . .	25
3.3.1 Energy landscapes . . . . .	29
3.4 Results and Discussion . . . . .	31
3.4.1 Sulfite/Chlorate . . . . .	31
3.4.2 Sulfate/Perchlorate . . . . .	39
3.5 Conclusion . . . . .	51

<b>4</b>	<b>Hydrated Metal Clusters</b>	53
4.1	Introduction	53
4.2	Methodology	56
4.2.1	Global Optimisation using GIGA	56
4.2.2	Crossover	56
4.2.3	Mutations	56
4.2.4	Post-genetic step	58
4.2.5	DFT	59
4.2.6	Energetics	59
4.3	Results and Discussion	60
4.3.1	Structures of Hydrated $\text{Ag}_m\text{Au}_n^+$ Clusters	60
4.3.2	Binding	74
4.4	Conclusions	79
<b>5</b>	<b>BMPGA and Thresholding for Finite Clusters</b>	81
5.1	Methodology	81
5.2	Overview	81
5.2.1	BMPGA	82
5.3	Results	106
5.3.1	Bench-marking	106
5.3.2	Global Optimisation of Lennard-Jones Clusters	106
5.4	Conclusions	115
<b>6</b>	<b>Overall Conclusions</b>	116
	<b>References</b>	125

# List of Figures

1.1	Schematic representation of Ih unit cell, showing the regular hexagonal arrangement of the oxygen framework. <sup>2</sup> . . . . .	2
1.2	Current putative global minimum structure for 20 water molecules using the TIP4P potential. The structure exhibits four and five membered rings. . . . .	3
1.3	Two-dimensional representation of a landscape exhibiting (a) a minimum (rank 0 stationary point) and (b) a saddle point (rank 1 stationary point). <sup>3</sup> . . . . .	7
1.4	Two-dimensional representation of (a) a frustrated landscape exhibiting two minima similar in energy but separated by large kinetic barriers, and (b) a funnelled landscape. <sup>3</sup> . . . . .	8
2.1	Schematic of the TIP4P water model, showing the locations of the atomic charge-centres and the pseudo-atom which carries the oxygen partial charge. Note all the particles are in the same plane. . . . .	13
3.1	Schematic representation of Ih unit cell, showing the regular hexagonal arrangement of the oxygen framework. <sup>2</sup> . . . . .	21
3.2	(A) Current putative global minimum structure for 20 water molecules using the TIP4P potential. The structure exhibits four and five membered rings. (B) Four membered ring highlighted in pink, 5 membered ring highlighted in blue. . . . .	22
3.3	Table of lowest energy empirical minima for the $\text{SO}_3^{2-}(\text{H}_2\text{O})_N$ system in the range $3 \leq N \leq 16$ . . . . .	31

3.4	Table of lowest energy empirical minima for the $\text{ClO}_3^-(\text{H}_2\text{O})_N$ system in the range $3 \leq N \leq 16$ . . . . .	32
3.5	Plot showing the fraction of Basin-Hopping runs which converged on the same lowest energy minimum (according to energy and the radius of gyration) for the $\text{ClO}_3^-(\text{H}_2\text{O})_N$ and $\text{SO}_3^{2-}(\text{H}_2\text{O})_N$ systems. . . . .	33
3.6	Total cluster energy for the Sulfite and Chlorate systems calculated using the empirical potential given in Equation 3.1. These are the Boltzmann weighted averages (see Equation 3.3, $T=130\text{K}$ ) of all clusters identified in the combined databases of minima taken from 10 parallel Basin-Hopping runs for each cluster size. . . . .	35
3.7	Energy per water molecule for the Sulfite, Chlorate and bulk TIP4P systems . .	36
3.8	Boltzmann weighted mean number of 'dangling' hydrogen bonds ( $n_d$ ) for $\text{SO}_3^{2-}(\text{H}_2\text{O})_N$ and $\text{ClO}_3^-(\text{H}_2\text{O})_N$ clusters in the range $1 \leq N \leq 50$ . Dangling bonds are defined as those identified as being aligned normal, or near normal to the surface of the cluster. . . . .	37
3.9	Boltzmann Weighted coordination numbers for the hydrated sulfite and chlorate systems. . . . .	37
3.10	Boltzman weighted ion displacement $r'$ from radius of gyration of each cluster per number of water molecules . . . . .	39
3.11	Lowest energy structures identified for the $\text{ClO}_4^-(\text{H}_2\text{O})_N$ system. This Figure shows only the putative global minima identified for global optimisation runs for which all of the independent basin-hopping runs converged on the same minima ( $3 \leq N \leq 16$ ). . . . .	40
3.12	Putative empirical global minima for the hydrated sulfate system as reported by Smeeton et al. <sup>4</sup> . . . . .	41
3.13	Boltzmann weighted ( $T=130\text{K}$ ) empirical energies for the putative global minima of the micro-hydrated sulfate and perchlorate systems, $\text{X}_{(\text{H}_2\text{O})_N}$ , where $3 \leq N \leq 50$	42

3.14 Boltzmann weighted ( $T=130K$ ) empirical energies-per-water molecule for the putative global minima of the micro-hydrated perchlorate system, $ClO_4^- (H_2O)_N$ , where $3 \leq N \leq 50$ . . . . .	43
3.15 Ball and stick models of the perchlorate and pure water systems showing the similarities in structural morphology displayed between certain low energy minima of differing numbers of molecules . . . . .	44
3.16 Ball and stick models of example Sulfate-Water clusters as reported by Smeeton <i>et al.</i> <sup>5</sup> . . . . .	45
3.17 Lowest energy structures observed for the $XO_4^{q-}$ system. (a) $q = -2.0$ , (b) $q = -1.9$ and $q = -1.8$ , (c) $-1.7 \leq q \leq -1.1$ , and (d) $q = -1.0$ . . . . .	47
3.18 Displacement against overall charge ( $q$ ) for the $XO_4^q (H_2O)_{12}$ system in the range $-2.0 \leq q \leq -1.0$ . . . . .	48
4.1 Comparison between our (A) and previously found (B) geometries for $Ag^+H_2O$ and $Au^+H_2O$ . B shows the geometries as reported by Feller, Glendening, and Jong in 1999. <sup>6</sup> . . . . .	60
4.2 $\omega B97X-D$ geometries for $Ag_n^+(H_2O)$ and $Au_n^+(H_2O)$ for $n = 1 - 4$ , found using the aug-cc-pVDZ/aug-cc-pVDZ-pp DFT basis set. . . . .	61
4.3 $\omega B97X-D$ geometries for $Ag_n^+(H_2O)_2$ and $Au_n^+(H_2O)_2$ for $n = 2 - 4$ . . . . .	62
4.4 $\omega B97X-D$ geometries for $[Ag_nAu_m]^+(H_2O)$ for $n + m = 2 - 4$ . . . . .	64
4.5 $\omega B97X-D$ geometries for $[Ag_nAu_m]^+(H_2O)_2$ for $n + m = 3 - 4$ . . . . .	65
4.6 $\omega B97X-D$ structures and energies (relative to the putative global minimum) for the $[Ag_3Au]^+(H_2O)_2$ putative global minimum and other unique minima. . . . .	67
4.7 Energies relative to the global minimum of the lowest 10 minima for $Ag_m^+(H_2O)_2$ for $m = 2 - 4$ and $Ag_4^+(H_2O)_2$ . . . . .	69
4.8 Energies of the lowest 10 minima for $Au_n^+(H_2O)_2$ for $n = 2 - 4$ and $Au_4^+(H_2O)_2$ . . . . .	70
4.9 Energies of the lowest 10 minima for $[Ag_mAu_n]^+(H_2O)$ for $m + n = 2$ . . . . .	71



4.10	Energies of the lowest 10 minima for $[\text{Ag}_m\text{Au}_n]^+(\text{H}_2\text{O})$ for $m + n = 3$ . . . . .	72
4.11	Energies of the lowest 10 minima for $[\text{Ag}_m\text{Au}_n]^+(\text{H}_2\text{O})$ for $m + n = 4$ . . . . .	73
4.12	$\omega\text{B97X-D}$ structures and sequential binding energies for the 2-metal $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_x$ putative global minima. . . . .	74
4.13	$\omega\text{B97X-D}$ structures and sequential binding energies for the 3-metal $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_x$ putative global minima. . . . .	75
4.14	$\omega\text{B97X-D}$ structures and sequential binding energies for the 4-metal atom $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_2$ putative global minima, and a the lowest energy minimum with two water molecules seen to be exhibiting hydrogen bonding. . . . .	76
4.15	$\omega\text{B97X-D}$ structures and relative energies for the planar and tetrahedral, singly and doubly hydrated, $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_x$ system. . . . .	78
5.1	Schematic representation of a GA genome . . . . .	84
5.2	UML diagrams for Atom, Molecule and Cluster objects . . . . .	85
5.3	Schematic example of a 3-way tournament for an 8 member population. The values inside the boxes indicate the fitness of each member of the population, with the highest fitness being 3.0 and the lowest being 1.0 . . . . .	88
5.4	Schematic representation of Roulette Wheel Selection . . . . .	90
5.5	Diagram showing (a) a sample population. (b) Ranks for the population in (a). (c) Those ranks normalised to weights. (d) How the probabilities compare for Rank and Roulette Wheel selection, given the same initial population. . . . .	92
5.6	Schematic representation of the modified Deaven and Ho crossover scheme as implemented in BMPGA . . . . .	94
5.7	Schematic representation of the shake mutation move, all the red particles are translated by small random vectors to give the blue particles. Note the vectors have been exaggerated for illustration. . . . .	96

5.8	Schematic of the Master-Worker model in BMPGA, showing a Master process and three Worker processes, with communication over the public internet . . . .	103
5.9	Lennard-Jones (6-12) potential, highlighting the minimum at $\approx 1.122\sigma$ . . . . .	107
5.10	Energy and Energy per particle for all the putative Lennard-Jones optimisation runs. Icosahedral minima are highlighted using vertical orange lines. . . . .	107
5.11	Plot of the mean energy of the pool during the optimisation of the LJ13 system	110
5.12	Plots for one run of the LJ13 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum . . . . .	110
5.13	Plots for one run of the LJ38 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum . . . . .	111
5.14	Plots for one run of the LJ98 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum . . . . .	112

5.15	Plots for one run of the LJ135 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum . . . . .	113
5.16	Plots for one run of the LJ147 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum . . . . .	114

# List of Tables

2.1	TIP4P Parameters giving the Lennard-Jones parameters, and charges for each site. The angles and bond lengths are also described. Note: all sites are co-planar and the oxygen charge centre (PA) is located on the same side of the Oxygen atom as the Hydrogen atoms. <sup>7</sup> . . . . .	13
3.1	Perchlorate ( $ClO_4^-$ ) and Sulfate ( $SO_4^{2-}$ ) parameters. The Lennard-Jones parameters ( $\epsilon$ , and $\sigma$ ) for Chlorate and Sulfate were taken from work published by Wagner <i>et al.</i> and Cannon <i>et al.</i> respectively. <sup>8,9</sup> The partial atomic charges were calculated for this work using DFT as outlined below. . . . .	27
3.2	Table of Lennard-Jones parameters and partial charges used for sulfite, chlorate and TIP4P water in this study. <sup>9</sup> . . . . .	28
3.3	Charges on the pseudo-sulfur ( $q_X$ ), Oxygen atoms ( $q_O$ ), and overall ionic charge for the $XO_4(H_2O)_{12}$ system under study. . . . .	45
5.1	Potentials in BMPGA . . . . .	104
5.2	Parameters used in the global optimisation of the Lennard-Jones 6-12 system using BMPGA. Note: The mutations selected were uniformly distributed between the methods listed. . . . .	108
5.3	Low lying minima of interest for the LJ 6-12 system . . . . .	109

# Chapter One

## Introduction

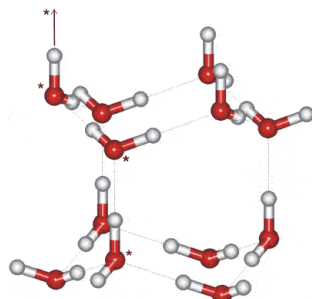
### 1.1 Hydrated Ions

Hydrated ions are found across many fields, including industrial and biological systems. It is well known that the presence and concentration of ions in a system can drastically alter its behaviour.

#### 1.1.1 Water and Ice

Before discussing the previous work on solvated ions it is worth mentioning briefly the reasons that water itself is interesting. Water is the most common molecule in biology, and the majority of the surface of the Earth is covered in it. These facts alone would make water worthy of research and understanding in their own right, however water is also interesting to study in its own right. Water is interesting because of strong intermolecular hydrogen bonding it exhibits with other water molecules. This hydrogen bonding leads water to have a complex phase diagram exhibiting over twelve different identified solid ice phases.<sup>10</sup> These phases exhibit differing morphologies in the networks formed by the intermolecular hydrogen bonds.<sup>10</sup> The ice with which we are most familiar is known as Hexagonal Ice (Ih), as it makes up the snow and ice commonly seen on earth. Ih is formed of a regular hexagonal lattice of water molecules (see Figure 1.1) with hexagonal symmetry. Each water molecule in the bulk has a fully saturated complement of hydrogen bonds (it is accepting and donating two hydrogen bonds). The motif of stacked hexagons exhibited by Ih is extremely common in the structures adopted by water, both in the bulk and in smaller clusters.<sup>11–13</sup> This effect is due to the maximisation of hydrogen

bond acceptor/donor ratio with minimal geometric strain. The local hexagonal symmetry in Ih leads to the formation of ice crystals and snowflakes which exhibit macroscopic hexagonal symmetry.<sup>10</sup>



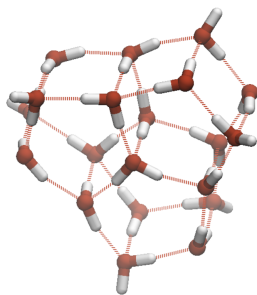
**Figure 1.1** Schematic representation of Ih unit cell, showing the regular hexagonal arrangement of the oxygen framework.<sup>2</sup>

Other common motifs seen in stable water ices and also as local minima in molecular dynamics simulations are three, four and five membered rings.<sup>14,15</sup> See Figure 1.2.

The interactions of water with uncharged solutes has been studied, with some emphasis on methane and other hydrocarbons. Water ices can form under pressure in the presence of small chain hydrocarbons commonly found in natural gas reserves, such as methane. For instance, methane clathrates consist of extended lattices with methane molecules trapped in cage-like voids. The presence of neutral species like methane within these voids is energetically favourable because it maximises the van der Waals interactions between the guest molecules and the clathrate walls. Additionally, this arrangement minimally disrupts the hydrogen bonding network of the surrounding water molecules, making it a more stable configuration compared to the solute being dissolved in water.<sup>16</sup>

### 1.1.2 Hydrated Ions

There have been many experimental investigations into the properties and behaviours of ionic solutions.<sup>13,17</sup> Solvated ions have been shown to confer a variety of different effects to the solutions which they occupy. These effects include modifying the surface tension and viscosity of the solution.<sup>18,19</sup>



**Figure 1.2** Current putative global minimum structure for 20 water molecules using the TIP4P potential. The structure exhibits four and five membered rings.

Ions can effect the properties of other species in solution as well as the behaviour of the solution itself. The experimentalists Lewith and Hofmeister in 1888 published their work on the effects of various ions upon the solubilities of proteins.<sup>17</sup> Hofmeister showed that a range of salts display persistent effects on the solubility of proteins, notably increasing or decreasing their solubilities. The ordering of these ions is now known as the Hofmeister series. The ions which increase protein stability and promote order within the hydrogen bonding framework of water are known as the kosmotropes (Greek, order-makers) and the ions which decrease protein stability and disrupt hydrogen bonding are known as chaotropes (disorder-makers).<sup>20</sup>

## Simple Ions

In biology many of the ions in solution are simple ions with the most common of these being free protons,  $\text{Cl}^-$ ,  $\text{HO}^-$ ,  $\text{K}^+$ , and  $\text{Na}^+$ .<sup>21,22</sup> These ions are involved in a great number of biological processes.<sup>23</sup> They are constantly interacting with proteins, modifying their secondary and tertiary structures. Ions are also actively pumped and channelled in and out of cells, most commonly to modify the osmotic potential of the cell and thus control the flow of water across membranes. Simple ions in solution are also important industrially and in the lab, both as reagents or catalysts, and as charge carriers in electrochemistry.<sup>24</sup>

There have been many previous experimental studies conducted into the effects simple ions have in solution.<sup>21,22,25–29</sup> In this work we are primarily concerned with anionic hydrated clusters in the gas phase as these provide good model systems for bulk behaviour. Clusters are studied for a variety of reasons, often because they provide an accessible test system for studying

behaviors or properties of bulk systems, and often because they are interesting in their own right.<sup>30</sup>

Hydrated halide ion systems have been much studied as model systems, due to their relative simplicity as singly-charged anions.<sup>28</sup> Various studies conducted using Infra-Red Photo-Dissociation (IRPD) spectroscopy have indicated that the larger micro-hydrated halide ions,  $\text{Cl}^-$ ,  $\text{I}^-$ ,  $\text{Br}^-$ , favor non-polar interfacial sites.<sup>19,28,29,31</sup>  $\text{F}^-$  however favours sub-surface sites in these clusters.<sup>19</sup> This behaviour is also shown to manifest in bulk systems.<sup>31</sup>

### More Complex Ions

Highly charged (ions with a charge greater than one) and polyatomic ions are more difficult research targets than monatomic ions. These complex ions are more challenging to simulate due to the increased number of degrees of freedom and reduced symmetry. Experimentally they can be challenging due to their comparative instability,  $\text{SO}_4^{2-}$  for example was shown to decompose if not hydrated by a minimum of three water molecules by Blades and Kebarle.<sup>32</sup>

Much of the modern experimental work on microhydrated anionic clusters is performed using mass spectrometry. Many different experiments can be performed, using a variety of ionisation methods and mass spectrometers.<sup>25</sup> The work by O'Brien *et al.* demonstrated how sulfate ions influence the structural patterning of water clusters. This research was carried out using Fourier-transform ion cyclotron resonance mass spectrometry to analyse these effects. Hydrated sulfate ions  $\text{SO}_4^{2-}(\text{H}_2\text{O})_N$  were produced using nanoelectrospray ionisation. These clusters were then size selected and probed using IRPD spectroscopy. It was found that a single  $\text{SO}_4^{2-}$  ion completely suppresses the appearance of dangling OH bonds normal to the surface in the range  $34 \leq N \leq 43$ .<sup>25</sup>

There have been many recent studies in which polyatomic hydrated ions (both as clusters and in the bulk) have been investigated computationally.<sup>5,33-35</sup> Within the group of Professor Roy L. Johnston, L. Smeeton has recently studied the hydrated sulfate system and successfully reproduced the experimental result mentioned above. This work was accomplished using basin-



hopping Monte Carlo with an empirical potential fitted to *ab initio* calculations.<sup>5</sup>

## Ions and Polypeptides

Recent theoretical studies have shed light on the complex interactions between ions and proteins. In one significant finding, Lund *et al.* utilised molecular dynamics simulations with the AMBER force field to demonstrate that chloride ( $Cl^-$ ) and iodide ( $I^-$ ) ions tend to aggregate at the interfaces between water and the hydrophobic regions of proteins.<sup>36</sup> This research underscores the role of ion clustering in influencing protein structure at molecular interfaces.

Further investigations into how ions impact protein structure, particularly in the gas phase, have also been carried out. Berezovskaya *et al.* explored the specific effects of iodide ions on the conformations of various proteins, including lysozyme, which is the same protein studied in the foundational Hofmeister series. Their experiments, conducted across a range of temperatures, employed drift tube ion mobility mass spectrometry (DT IM MS) to examine these interactions. This technique provided clear evidence of conformational changes, as indicated by variations in the collision cross sections of the proteins when subjected to different ionic charges. This work adds a valuable dimension to our understanding of how environmental factors, such as ion presence and temperature, can alter protein structure in significant ways.<sup>37</sup>

## 1.2 Energy Landscapes

An energy landscape (or potential energy surface (PES)) of a system describes the total potential energy of the system as a function of various physical parameters, such as relative atomic positions, molecular orientations or torsional angles. Energy landscapes can be studied for virtually any chemical system but for the purposes of this literature survey, we will consider only the landscapes of atomic clusters, molecular clusters and small poly-peptides and poly-peptoids.<sup>38</sup>

Many problems in computational chemistry can be reduced to the exploration of potential energy surfaces (PES). Examples of these problems include; the structures and dynamics of

atomic, molecular or colloidal clusters, and the conformations and folding of proteins.<sup>38</sup>

### 1.2.1 Features of Energy Landscapes

The topology of the energy landscape of a system is entirely dependent on the interactions between the species comprising the system.

#### Stationary Points

The primary features of interest in energy landscape theory are the stationary points on the PES  $\mathbf{r}_\alpha$  at which:

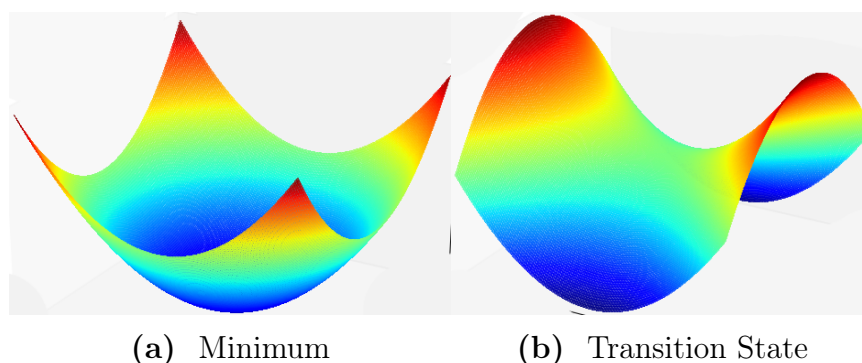
$$\nabla U(\mathbf{r}_i) = \mathbf{0}, \quad \mathbf{r}_i \in \{\mathbf{r}_\alpha\} \quad (1.1)$$

where  $\nabla = (\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N})$  in  $\mathbb{R}_\alpha$  and  $\mathbf{0}$  is the null vector. At a stationary point the net force acting upon each species (atom, molecule, etc.) is zero, as force is proportional to the derivative of  $U(\mathbf{r})$ .<sup>38</sup> It thus follows that the stationary points on in the landscape  $U(\mathbf{r})$ , which are solutions to 1.1, are in mechanical equilibrium.

Of the the stationary points which satisfy 1.1 and are thus in mechanical equilibrium, most are not mechanically *stable*. Mechanical stability is used to describe the way in which a system responds to perturbation. In a mechanically stable stationary point, when subjected to a small perturbation the system will oscillate about its equilibrium position as the equations of motion for this system will seek to return to equilibrium. A non-mechanically stable stationary point on the landscape on the other hand will, when perturbed, not return to its equilibrium position and will instead deviate rapidly from the stationary point (and will eventually oscillate about a remote, mechanically stable stationary point). As chemical systems will spend the majority of their time oscillating about mechanically stable stationary points, these are of the most interest when sampling a novel landscape.

Insight into the mechanical stability of a stationary point is usually afforded by determination of the local curvature and examination of the forces acting on the system in proximity to the equilibrium point. Most methods involve either the explicit calculation or implicit approximation of the Hessian for the system at the point of interest. The Hessian being a matrix of

partial second derivatives of  $U$  with respect to  $\mathbf{r}$ . The number of negative Hessian eigenvalues is commonly known as the Hessian index.<sup>38</sup>



**Figure 1.3** Two-dimensional representation of a landscape exhibiting (a) a minimum (rank 0 stationary point) and (b) a saddle point (rank 1 stationary point).<sup>3</sup>

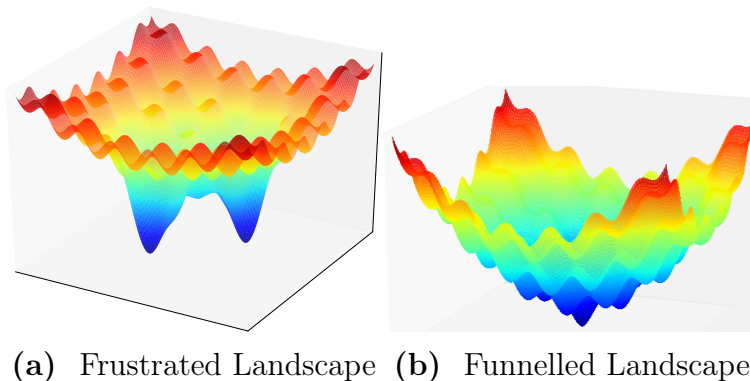
Stationary points which exhibit no negative Hessian eigenvalues, and thus have a rank of 0, are mechanically stable and are therefore minima. Any sufficiently small local perturbations will lead to an increase in potential energy and thus experience a corresponding force which leads to oscillation about the minimum. The importance of finding the minima of a potential energy surface has driven the development of many algorithms designed to find them as efficiently as possible, some of these are discussed in the Methods chapter. Archetypal energy landscapes showing a minimum and a saddle point are shown in Figure 1.3.

Rank 1 stationary points, those which exhibit a single negative Hessian eigenvalue, are saddle points. The simplest saddle points are minima in all but one degree of freedom. Saddle points are interesting as they are the lowest energy point which separates adjacent minima. If a system in equilibrium at a saddle point is perturbed along the axis corresponding to the negative eigenvalue it will begin to oscillate about one of the minima which it separates.

The minimum with the lowest energy on a potential energy surface is known as the *global minimum*. Global minima of systems are important because these structures and surrounding thermally accessible structures contribute highly to the observable equilibrium properties of the system. Much of computational chemistry is concerned with the search for the conformation of a chemical system which corresponds to the global minimum of the PES. These problems

are often known as global geometry optimisation problems. The main problem theorists face in global geometry optimisation is how to efficiently search a complex landscape, with many degrees of freedom. This will be discussed in more detail in the Methods chapter.

### 1.2.2 Interpreting the Topologies of Energy Landscapes



**Figure 1.4** Two-dimensional representation of (a) a frustrated landscape exhibiting two minima similar in energy but separated by large kinetic barriers, and (b) a funnelled landscape.<sup>3</sup>

The topology of an energy landscape is strongly linked to how the system in question responds to stimuli. Some systems are highly funnelled, having a global minimum connected to other minima in the landscape by only small asymmetric barriers.<sup>38</sup> In these systems any random starting configuration has a high chance of finding the global minimum of the system with little chance of becoming kinetically trapped in a distant minimum. This behaviour is often described as the system experiencing a *directing effect* towards a particular conformation. An example of a system which experiences a large directing effect might be the binding of short, complimentary DNA strands.<sup>39</sup> Landscapes can also exhibit frustrated topologies. Frustrated landscapes generally have several minima close in energy to the global minimum but with large kinetic barriers separating them. For systems with frustrated energy landscapes it is likely that a random initial configuration will become kinetically trapped in a region of the energy landscape from which the global minimum is practically inaccessible. Water clusters are a good example of a non-covalent chemical system with a highly frustrated landscape.<sup>14</sup> The frustration in these systems arises because of the highly directional nature of relatively strong

hydrogen bonds, which lead to a large number of favourable structures but which incur large energy penalties when attempting to convert between them as several hydrogen bonds must be broken for a relatively small rearrangement. Landscapes usually lie on a spectrum between funnelled and frustrated and many display overall frustration but with large local funnels. See Figure 1.4 for illustrative examples of funnelled and frustrated landscapes.

### 1.2.3 Ergodicity

Local ergodicity within an energy landscape is essential for understanding the dynamics of complex systems. Ergodicity, a term originally coined by Boltzmann in 1884, refers to the property of a system to explore all accessible states given sufficient time. The Ergodic Principle allows us to recreate a time-averaged property of a system from an ensemble average, weighted over all possible states, using only knowledge of microstates rather than the full dynamics of the system. For many systems, assuming ergodicity is reasonable when considering phase space trajectories of infinite length ( $t_{obs} \rightarrow \infty$ ).<sup>40-42</sup>

Local ergodicity emphasizes behaviour within a confined region of the landscape, where the system explores all accessible microstates within that region before transitioning to another. This is crucial in systems with rough energy landscapes, characterized by numerous local minima and high-energy barriers. For example, in protein folding, local ergodicity ensures thorough exploration of conformations within a local energy basin before escaping to another, thus avoiding kinetic traps and aiding in attaining the native state.<sup>43,44</sup>

Conversely, systems with poor local ergodicity may become kinetically trapped in sub-optimal conformations, as they do not efficiently explore local minima before transitioning. This can be problematic in materials science, where the properties of a material depend on its microstructure, influenced by the local ergodicity during formation. An example is the annealing process in metallurgy, where atomic movements' local ergodicity affects the final grain structure of the metal.<sup>41,45</sup>

Local ergodicity is a fundamental characteristic that influences phase space sampling efficiency, phase transition kinetics, and system stability in various energy landscape regions. Understanding and controlling local ergodicity can lead to significant advancements in fields like biochemistry and materials engineering, optimising processes and developing new technologies.<sup>44</sup>

# Chapter Two

## Methodologies

### 2.1 Energy Evaluation

One of the most fundamental aspects of any theoretical investigation in chemistry is choosing an appropriate energy functional to describe the interactions between species in the system in question. In an ideal world we would investigate every chemical problem from a fully quantum mechanical perspective and find exact solutions. In the real world this is not possible. This is because quantum mechanical problems are very hard and expensive to solve exactly. This shortcoming of reality is why we have to carefully select the level of theory at which to tackle our problems.<sup>46</sup> Generally speaking, the higher the level of theory involved, the more accurate the result will be, but the more computationally expensive the calculation will be.<sup>47</sup> There have been a great number of different energy functionals developed, some of which are appropriate for only a small selection of systems, and some of which are more general.

Basis sets and parameter sets are two different concepts in the field of energy evaluations. Basis sets describe the properties of the atomic orbitals used in the calculations, while parameter sets refer to the values assigned to the parameters used in the calculations. For instance, in simple cases, a parameter set might contain only Lennard-Jones parameters for each atomic site in the system. While in a more complex piece of *ab initio* work a basis set will be used which may include descriptions and occupancy of all the individual orbitals for each atomic site in the system. These basis sets are predetermined before the investigation and vary based on the energy functional being used. On the other hand, parameter sets are determined either exper-

imentally or theoretically and are used in the calculations to assign values to the parameters.<sup>9</sup> Empirical energy evaluations often use parameter sets that are fitted to data from higher level investigations.<sup>5,48</sup> These parameter sets can be adjusted to produce the most accurate results possible within the limitations of the calculation.

### 2.1.1 Empirical Methods

Empirical methods are amongst the simplest and least computationally expensive methods for energy evaluation. Empirical methods are non-quantum and do not involve solving the wave functions derived from the Schrödinger equation like the other methods discussed below. Empirical methods generally attempt to gain the most accurate estimation of the energy of a system using the fewest terms possible. For added simplicity, some empirical models treat molecules or fragments of molecules as being rigid and ignore polarizability, whilst some are more complex.<sup>5,49</sup>

Often the parameters used in empirical energy calculations are very simple. The conformations and long range behaviour of some molecular systems can, for example, be modelled using just Lennard-Jones and Coulombic terms and reproduce experimental data with good accuracy.<sup>4,5</sup> Empirical methods are commonly used in Molecular Dynamics (MD) simulations, which model the behaviour of a chemical system over a period of time, due to their low computational expense and the large nature of the systems.

#### TIP4P water model

The TIP4P potential is a 4-body classical model for water, this pairwise potential takes the form:<sup>7</sup>

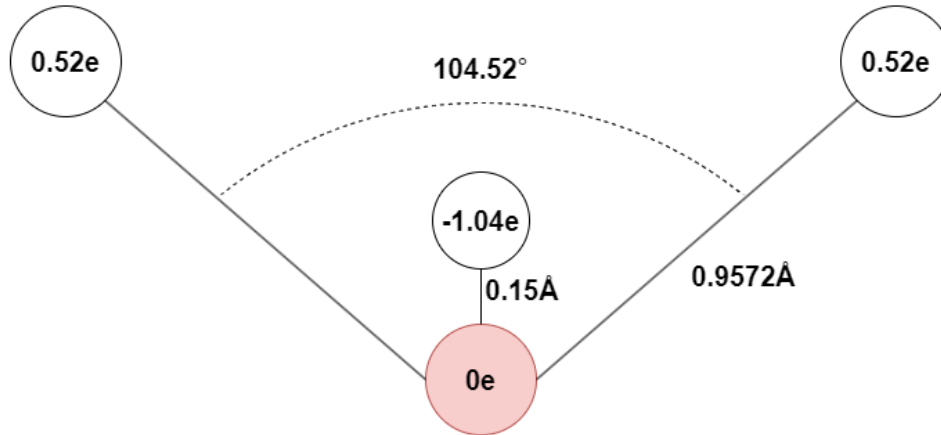
$$\epsilon_{mn} = \sum_i^m \sum_j^n \frac{q_i q_j e^2}{r_{ij}} + \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] \quad (2.1)$$

where:  $r_{ij}$  is the interatomic separation,  $q_i$  and  $q_j$  are the charges of species  $i$  and  $j$ .

In this potential, the Lennard-Jones parameters are centred on each atom, the atomic charges for the Hydrogen atoms are centred on their atomic centres, and the charge centre of the oxygen



atom is moved to a pseudo-atomic site, this is visualised in Figure 2.1.



**Figure 2.1** Schematic of the TIP4P water model, showing the locations of the atomic charge-centres and the pseudo-atom which carries the oxygen partial charge. Note all the particles are in the same plane.

The site-wise potential parameters are given in Table 2.1.

**Table 2.1** TIP4P Parameters giving the Lennard-Jones parameters, and charges for each site. The angles and bond lengths are also described. Note: all sites are co-planar and the oxygen charge centre (PA) is located on the same side of the Oxygen atom as the Hydrogen atoms.<sup>7</sup>

Property	Value
$r(\text{O-H})$ , Å	0.9572
$r(\text{O-PA})$ , Å	0.15
$\angle \text{H-O-H}$ , °	104.52
$\angle \text{H-O-PA}$ , °	52.26
$\sigma_O$ , Å	3.2
$\sigma_H$ , Å	0
$\epsilon_O$ , ( $kcal \text{ mol}^{-1}$ )	0.648
$\epsilon_H$ , ( $kcal \text{ mol}^{-1}$ )	0
$q(\text{O})$ , e	0.0
$q(\text{H})$ , e	0.52
$q(\text{PA})$ , e	-1.04

### 2.1.2 Semi-Empirical Methods

DFTB (Density Functional Tight Binding) is a semi-empirical method that builds upon density functional theory (DFT) by incorporating empirical parameters to simplify the calculations and reduce computational costs. In DFTB, the Born-Oppenheimer approximation is also assumed, allowing for the decoupling of nuclear motion from electronic motion. This approximation is essential for treating the nuclei as stationary while solving the electronic structure.

In DFTB, the electron density around each interacting site is approximated using a minimal basis set derived from DFT calculations. The interactions between these electron densities are then described using parameterised tight-binding Hamiltonians, which incorporate approximated exchange-correlation functionals.<sup>50</sup> This approach retains some of the quantum mechanical rigour of DFT while significantly simplifying the computational workload.

The choice of basis sets in DFTB determines the accuracy and computational cost of the calculations. Unlike full DFT, which often requires extensive basis sets to account for polarisation and diffuse orbitals, DFTB typically uses smaller, more efficient basis sets. While this results in less detailed electronic structure descriptions, it allows for the examination of much larger systems and longer timescales than is feasible with full DFT or *ab initio* methods.

DFTB is generally less accurate than DFT, but it is much faster and more practical for large-scale simulations. This makes it an attractive option for exploring the properties of large molecular systems and complex materials where full DFT calculations would be prohibitively expensive.<sup>50,51</sup> DFTB is frequently used to study phenomena such as molecular dynamics, materials properties, and biological systems, providing a good balance between accuracy and computational efficiency.<sup>50,51</sup>

### 2.1.3 Density Functional Theory (DFT)

Density Functional Theory (DFT) is a computational quantum mechanical modelling method that derives wave functions from the Schrödinger equation. This approach allows DFT to describe the electronic structure of atoms, molecules, and condensed matter systems with a

high degree of accuracy, making it a fundamental quantum method in theoretical chemistry and materials science.<sup>52</sup> In density functional theory the Born-Oppenheimer approximation is assumed to be true. The BO approximation accounts for the large difference in mass between nuclei and electrons, and thus allows the decoupling of the nuclear and electronic wave functions. In DFT, the electron density is calculated for the space surrounding each interacting site, and these regions of electron density interact with one another through approximated exchange-correlation functionals.<sup>46</sup> The basis sets used for DFT calculations determine the number and type of functionals applied to approximate the electronic structure of the system. The basis set also specifies whether polarisation and diffuse orbitals are included in the calculations. The inclusion of these diffuse orbitals increases the expense of the calculation and so are often ignored when working with systems in which they are not expected to have a large effect. The more comprehensive the basis set used, the more accurate the calculation is likely to be, but this will make the expense of computation greater too. DFT is generally less accurate than *ab initio* work, but is significantly faster. DFT is frequently used by theorists and experimentalists alike to predict the vibrational spectra of systems for comparison with experimental data.<sup>53</sup> DFT is frequently used to examine significantly larger systems than are feasible when using *ab initio* methods.

#### 2.1.4 *Ab initio* Methods

The highest level of theory in common use are the *ab initio* methods. (*Ab initio*: Latin - meaning "*from the beginning*")<sup>54</sup> *Ab initio* methods seek to either solve or closely approximate a solution to the Schrödinger equations to give an energy for the system. They can not, however, give exact results without a basis set which accounts for all possible electronic orbital configurations. *Ab initio* methods are commonly used when evaluating systems in which charge transfer or proton transfer is expected to occur. There are many different *ab initio* methods used today, two of which will be discussed here:

**Hartree-Fock** (HF) methods include a mean Coulombic repulsion term, rather than calculating it explicitly. These HF methods generally overestimate the energy of the system due to this implicit electron-electron repulsion. This overestimation can be reduced by using a more

comprehensive basis set. In these methods the Born-Oppenheimer approximation along with other assumptions are made, to allow for faster computation, but these assumptions typically reduce the accuracy achievable.<sup>55</sup> In HF the Born-Oppenheimer approximation is also used. Making these approximations allows for significantly faster computation of the energies involved, but also limits the accuracy. Hartree-Fock methods are unsuitable for the analysis of systems in which relativistic effects are strong, such as when performing analysis including gold atoms.

**Post Hartree-Fock** methods are often significantly more accurate than vanilla Hartree-Fock as they include accurate electron correlation terms. Møller-Plesset perturbation theory is an example of a post Hartree-Fock method.<sup>56</sup> These methods are more computationally expensive than other methods, but generally provide a more accurate evaluation of the energy in a system.

In conclusion, very often the parameters used in less computationally expensive (i.e. empirical/parameterised) energy evaluations are fitted to calculations performed at the *ab initio* level or the DFT level. Both *Ab initio* and DFT methods are infrequently used for large global-optimisation problems, due to the large expense they incur. It is more common to use these higher theory methods to locally optimise structures gained from another source such as an empirical global optimisation, experimental data or a selection of likely structures proposed by the researcher.<sup>57</sup>

## 2.2 Global Optimisation

As mentioned in Section 1.2, some of the more common problems tackled by computational chemists are those of global optimisation of the conformations of systems. These problems are all about locating the lowest energy stationary points on the PES for the system. In an ideal world, we would be able to identify all the stationary points on the potential energy surface (PES) for the system in question, solving equation 1.1 to identify all configurations  $\mathbf{r}_0$  with

all-positive Hessian eigenvalues. There are many ways of approaching these problems and we shall examine a few here.

### 2.2.1 Local Optimisation

Before we begin to look at global optimisation we shall examine local optimisation. Local optimisation allows us to find and identify stationary points on the PES for a system. We can look for different types of stationary points during local optimisations, usually we are searching for minima, but often we are interested in higher order stationary points like transition states. Local optimisation is usually achieved by use one of several line search methods:

**Line Search** methods are crucial for refining the steps taken in multidimensional optimisation techniques. These methods aim to find an optimal step size along a given search direction by minimising the objective function in one dimension.<sup>58</sup>

**Bracketing and Sectioning** techniques are commonly used in 1D line searches. Bracketing involves finding an interval where a minimum is known to exist, while sectioning methods like the Golden Section Search and Fibonacci Search iteratively narrow this interval to locate the minimum with increasing precision.<sup>58</sup> These methods are straightforward to implement and are effective for smooth, unimodal functions. However, they may be less efficient for more complex landscapes with multiple local minima.<sup>59</sup>

**Interpolation** methods, such as quadratic and cubic interpolation, use polynomial fits to estimate the minimum within a bracketing interval. These techniques can provide faster convergence than bracketing alone by leveraging curvature information.<sup>60</sup> Quadratic interpolation is simple but may require more iterations to achieve high precision. In contrast, cubic interpolation uses more points to construct the polynomial, potentially leading to more accurate step size determination in fewer iterations.<sup>61</sup>

In practice, line search methods are often integrated with gradient-based techniques to enhance their efficiency. They play a pivotal role in algorithms such as conjugate gradient, quasi-Newton, and Newton methods by optimising step sizes at each iteration.<sup>58</sup> This combination

results in more robust and efficient convergence properties, making line search a foundational component of numerical optimisation strategies.

**Gradient** methods, such as steepest descent, simply make moves along the path with the steepest negative gradient from the current point on the PES.<sup>62</sup> Step sizes are simply scaled to be proportional to the gradient.<sup>63</sup> The gradients can either be calculated explicitly by differentiation or can be approximated. These factors make gradient descent relatively simple to implement and computationally inexpensive to evaluate.<sup>64</sup> Gradient descent methods are sometimes slow to converge on minima, as their step sizes are inversely proportional to the gradient, so can move slowly in wide, shallow basins.<sup>62</sup> These problems can be partially overcome by implementing a step weighting based on momentum for example.

**Newtonian** methods seek stationary points by directly evaluating the energy functional of the system.<sup>63</sup> In fully Newtonian line search methods the full Hessian is calculated for the system. The gradients thus obtained are used to generate likely steps. These methods are generally faster at finding stationary points than gradient descent in terms of the number of steps taken. Calculation of the Hessian is usually very computationally expensive and so these methods are often passed over in favour of a compromise solution.

**Quasi-Newtonian** methods generally use approximations of the Hessian rather than calculating it explicitly to reduce expense.<sup>58</sup> Quasi-Newtonian methods are the most commonly implemented line search methods as they generally converge quickly yet are computationally lower in expense than fully Newtonian methods.<sup>64</sup>

## 2.2.2 Basin-Hopping Monte Carlo

For chapter 1.1 we use basin-hopping as implemented within the pele (**P**ython **E**nergy **L**andscape **E**xplorer) package, published by the Wales group.<sup>65</sup>

Basin-hopping Monte Carlo is a commonly used and effective search strategy for global optimisation problems in science.<sup>5,65–68</sup> Basin-hopping treats every point in a basin as being equal to the local minimum by performing a local minimisation after every step is taken.

The general implementation of basin-hopping Monte Carlo is as follows:

1. The system is in an initial state  $x$ , with potential energy  $U_x$ . This initial state is perturbed in some way.
2. The new configuration is locally optimised, using one of the methods mentioned above, to obtain a minimum  $x'$  with energy  $U_{x'}$ .
3. The new minimum is accepted or rejected in line with the Metropolis criterion:

$$p(x \rightarrow x') = \min \left[ 1, \exp \left( \frac{-\Delta U}{k_B T} \right) \right] \quad (2.2)$$

$$r \in \mathbb{R}, r \in [0, 1] \quad (2.3)$$

where  $\Delta U = U_{x'} - U_x$ ,  $k_B$  is the Boltzmann constant, and  $T$  is a theoretical temperature used to scale the probability of acceptance of unfavourable/uphill steps. If  $r < p(x \rightarrow x')$  the new minimum is accepted, else it is rejected. If the new minimum is accepted, discard the old minimum and the new minimum becomes  $x$ . Else discard the new minimum.

4. If the terminating criterion is not met, return to step 1.

The theoretical temperature used to weight the probability of acceptance of an unfavourable move in the Metropolis criterion (2.2) has little physical meaning.  $T$  can however be adjusted to increase the chance of moving out of a basin and thus increasing the amount of the landscape sampled. The terminating criterion is chosen by the individual, but is typically either a fixed step-count, a known structure or a threshold energy. In pele<sup>65</sup> each unique minimum is added to the database, regardless of whether it is accepted or rejected as a putative move.

# Chapter Three

## Hydrated Ions

### 3.1 Hydrated Ions

Hydrated ions are found across many fields, including industrial and biological systems. It is well known that the presence and concentration of ions in a system can drastically alter its behaviour.

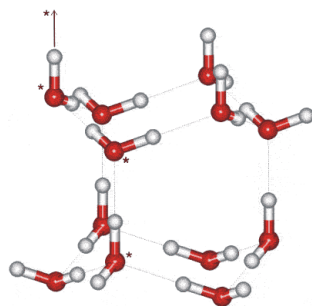
### 3.2 Introduction

#### 3.2.1 Water and Ice

Before discussing the previous work on solvated ions it is worth mentioning briefly the reasons that water itself is interesting. Water is the most common molecule in biology, and the majority of the surface of the earth is covered in it. These facts alone would make water worthy of research and understanding in their own right, however water is also interesting to study in its own right. Water is interesting because of strong intermolecular hydrogen bonding it exhibits with other water molecules. This hydrogen bonding leads water to have a complex phase diagram exhibiting over twelve different identified solid ice phases.<sup>10</sup> These phases exhibit differing morphologies in the networks formed by the intermolecular hydrogen bonds.<sup>10</sup> The ice with which we are most familiar is known as Hexagonal Ice (Ih), as it makes up the snow and ice commonly seen on earth. Ih is formed of a regular hexagonal lattice of water molecules (see Figure 3.1) with hexagonal symmetry. Each water molecule in the bulk has a fully saturated complement of hydrogen bonds (it is accepting and donating two hydrogen bonds). The motif



of stacked hexagons exhibited by Ih is extremely common in the structures adopted by water, both in the bulk and in finite clusters.<sup>11–13</sup> This is due to the maximisation of hydrogen bond acceptor/donor ratio with minimal geometric strain. The local hexagonal symmetry in Ih leads to the formation of ice crystals and snowflakes which exhibit macroscopic hexagonal symmetry.<sup>10</sup>



**Figure 3.1** Schematic representation of Ih unit cell, showing the regular hexagonal arrangement of the oxygen framework.<sup>2</sup>

Other common motifs seen in water ices and also as local minima in molecular dynamics simulations are three, four and five membered rings.<sup>14,15</sup> See Figure 3.2 for examples of these from the literature that have been reproduced for this study.

The interactions of water with uncharged solutes, particularly methane and other hydrocarbons, have been extensively studied.<sup>69,70</sup> Water ices can form under pressure in the presence of small chain hydrocarbons commonly found in natural gas reserves, such as methane. Methane clathrates, for example, have extended lattices with methane trapped in cage-like voids.<sup>71</sup> The presence of neutral species in a void rather than in solution is energetically favourable, as it maximises Van der Waals interactions between the guest and the clathrate walls and minimally disrupts the hydrogen bonding network of the water.<sup>16,72</sup>

### 3.2.2 Hydrated Ions

There have been many experimental investigations into the properties and behaviours of ionic solutions.<sup>13,17</sup> Solvated ions have been shown to confer a variety of different effects to the solutions which they occupy. These effects include modifying the surface tension and viscosity of the solution.<sup>18,19</sup>



**Figure 3.2** (A) Current putative global minimum structure for 20 water molecules using the TIP4P potential. The structure exhibits four and five membered rings. (B) Four membered ring highlighted in pink, 5 membered ring highlighted in blue.

Ions can effect the properties of other species in solution as well as the behaviour of the solution itself. The experimentalists Lewith and Hofmeister in 1888 published their work on the effects of various ions upon the solubilities of proteins.<sup>17</sup> Hofmeister showed that a range of salts display persistent effects on the solubility of proteins, notably increasing or decreasing their solubilities. The ordering of these ions is now known as the Hofmeister series. The ions which increase protein stability and promote order within the hydrogen bonding framework of water are known as the kosmotropes (Greek, order-makers) and the ions which decrease protein stability and disrupt hydrogen bonding are known as chaotropes (disorder-makers).<sup>20</sup>

## Simple Ions

In biology many of the ions in solution are simple ions with the most common of these being  $H^+$  (free protons),  $Cl^-$ ,  $HO^-$ ,  $K^+$ , and  $Na^+$ .<sup>21,22</sup> These ions are involved in a great number of biological processes.<sup>23</sup> They are constantly interacting with proteins, modifying their secondary and tertiary structures. Ions are also actively pumped and channelled in and out of cells, most commonly to modify the osmotic potential of the cell and thus control the flow of water across membranes. Simple ions in solution are also important industrially and in the lab, both as reagents or catalysts, and as charge carriers in electrochemistry.<sup>24</sup>

There have been many previous experimental studies conducted into the effects simple ions have in solution.<sup>21,22,25–29</sup> In this work we are primarily concerned with anionic hydrated clusters in the gas phase. Gas-phase finite clusters make good model systems for bulk behaviour.

Clusters are studied for a variety of reasons, often because they provide an accessible test system for studying behaviours or properties of bulk systems, and often because they are interesting in their own right.<sup>30</sup>

Hydrated halide ion systems have been extensively studied as model systems due to their simplicity as singly-charged anions.<sup>28</sup> Various studies using Infra-Red Photo-Dissociation (IRPD) spectroscopy have shown that larger micro-hydrated halide ions, such as  $\text{Cl}^-$ ,  $\text{I}^-$ , and  $\text{Br}^-$ , which have lower ionic charge densities, tend to favour positions at non-polar interfacial sites, typically interacting with water ring structures at the surface of the cluster, like those highlighted in Figure 3.2(B). This preference is due to the lower energy required for these larger ions to be located at the interface compared to being fully solvated in the bulk water, allowing them to minimise disruption to the hydrogen bonding network of water.<sup>19,28,29,31</sup> The more charge-dense ion  $\text{F}^-$  however favours sub-surface sites in these clusters, that is sites inside the cluster, and generally have coordination to a greater number of water molecules.<sup>19</sup> This behaviour is also shown to manifest in bulk systems.<sup>31</sup>

## More Complex Ions

Multiply charged and polyatomic ions are more difficult research targets than monatomic ions. They are more computationally expensive to simulate due to the increased number of degrees of freedom and reduced symmetry of the ions, leading to more difficult energy landscapes involving more kinetic traps and other complex features. Experimentally they can be challenging due to their comparative instability,  $\text{SO}_4^{2-}$  for example was shown to spontaneously decompose if not hydrated and stabilised by a minimum of three water molecules by Blades and Kebarle.<sup>32</sup>

Much of the modern experimental work on micro-hydrated anionic clusters is performed using mass spectrometry. Many different experiments can be performed, using a variety of ionisation methods and mass spectrometers.<sup>25</sup> The work by Williams *et al.* on the patterning effects sulfate exhibits on water was performed using Fourier-transform ion cyclotron mass spectrometry. Hydrated sulfate ions  $\text{SO}_4^{2-}(\text{H}_2\text{O})_N$  were produced using nanoelectrospray ionisation, these clusters were then size selected and probed using IRPD spectroscopy. It was found that a

single  $SO_4^{2-}$  ion completely suppresses the appearance of dangling OH bonds normal to the surface in the range  $34 \leq N \leq 43$ .<sup>25</sup>

There have been many recent studies in which polyatomic hydrated ions (both as clusters and in the bulk) have been investigated computationally.<sup>5,33-35</sup> Within the group, L. Smeeton has recently studied the hydrated sulfate system and successfully reproduced the experimental result mentioned above. This was accomplished using basin-hopping Monte Carlo with an empirical potential fitted to *ab initio* calculations.<sup>5</sup>

## Ions and Polypeptides

There has been some interesting theoretical work on the interactions of ions with proteins. Lund *et al.* found, using molecular dynamics with the AMBER force field, that  $Cl^-$  and  $I^-$  tend to cluster at the interfaces between water and the non-polar regions.<sup>36</sup> Ions have been shown to affect the tertiary structure of proteins in gas phase clusters. The work conducted by Berezovskaya *et al.* on the effects of iodide on the conformations of three different proteins, including Lysozyme from egg whites as used in the original Hofmeister paper. The experiment was conducted at various temperatures using drift tube ion mobility mass spectrometry (DTIMMS). Evidence for the change in conformation is found in the different collision cross sections of the proteins at different charges.<sup>37</sup>

### 3.3 Method

#### Global Optimisation

Low-energy minima on the potential energy surface for the ion-water clusters were explored using the basin-hopping Monte Carlo algorithm developed by Wales et al. within the pele package.<sup>65,66</sup> Basin-hopping was used to identify a putative lowest-energy structure, (global minimum, GM), to efficiently search the potential energy landscape and to create a database of low energy minima for later mapping of the landscape.<sup>66,73</sup> Ten basin-hopping runs of 500000 steps were performed in parallel for each size of  $X^-(H_2O)_N$  cluster in the range  $3 \leq N \leq 50$ .

The geometry perturbations implemented in this study were performed in blocks of 100 moves of the same type, with minimisation performed after each perturbation. The three move classes implemented are: random translations, random rotations, and cycle inversion moves.

In the random translation moves, random translation vectors are applied to molecules within the cluster. This means that each molecule in the cluster is displaced by a random amount in a random direction. This type of move helps explore different spatial configurations of the molecules by shifting their positions within the cluster, allowing the system to escape from local minima and sample a broader range of configurations.

Random rotation moves apply random rotation vectors to all of the molecules within the cluster. These rotations are in the range of  $\pm\pi$  radians. This move class involves rotating each molecule around a randomly chosen axis by a random angle, which helps in exploring the orientation space of the molecules. Rotations are essential for finding the optimal orientations of molecules relative to each other, which can significantly affect the overall energy and stability of the cluster.

Cycle inversion moves are more complex and involve inverting the positions of a subset of molecules within a cluster. This method was performed using the technique outlined in our previous work<sup>4,5,74</sup>, first proposed by Takeuchi.<sup>75</sup> In a cycle inversion move, a cycle of molecules is identified, and their positions are inverted such that the molecule at position  $i$  moves to

position  $j$ , the molecule at position  $j$  moves to position  $k$ , and so on, eventually cycling back to position  $i$ . This type of move allows for substantial reconfiguration of the cluster, which can help in overcoming larger energy barriers and exploring new regions of the configuration space that might not be accessible through simple translations or rotations.

By implementing these three classes of moves, my study aimed to thoroughly explore the potential energy landscape of the molecular cluster, ensuring that a wide variety of configurations are sampled and that the system can effectively escape from local minima to find the global minimum or other low-energy configurations.

For the landscape exploration of hydrated ion systems, the doubly nudged elastic band (DNEB) method was used to search for transition states between minima found during global optimisation.<sup>73</sup> The DNEB method constructs a series of intermediate states between two minima and refines these states to trace the minimum energy pathway. Initial paths are generated by linearly interpolating the translational coordinates of the rigid-body molecules and using spherical linear quaternion interpolation (SLERP) for the rotational coordinates.<sup>76</sup>

Candidate transition states identified along the pathway are further optimised using the hybrid eigenvector-following method.<sup>77,78</sup> This refinement ensures that the transition state is accurately located. The resultant energy landscapes, with their minima and transition states, are visualised as disconnectivity graphs using the PyConnect package.<sup>79</sup> These graphs provide a comprehensive visualisation of the energy landscape, illustrating the connectivity and hierarchical organisation of the landscape.

## Structural Refinement on the DFT Landscape

The 300 lowest-energy, geometrically unique, minima are re-optimised at the DFT level, using the NWChem<sup>80</sup> DFT package using the B3LYP exchange correlation functional and the 6-311++G\*\* Pople basis set. NWChem implements a quasi-newton optimisation with line searches and approximate energy Hessian updates.<sup>80</sup>

## Potentials

In this section of the work, we use a very simple, very computationally cheap potential during the global optimisation stage. This is a combination of the Lennard-Jones dispersion potential and a simple electrostatic potential, see Equation 3.1.  $q_i$  represents the charge on site  $i$ ,  $\epsilon_{ij}$  and  $\sigma_{ij}$  represent the Lennard-Jones parameters between sites  $i$  and  $j$ , and  $r_{ij}$  represents the inter-atomic distance between sites  $i$  and  $j$ . Lennard-Jones parameters for interactions between unlike atom types were calculated using the Lorentz-Berthelot mixing rules given in Equation 3.2.

**Table 3.1** Perchlorate ( $ClO_4^-$ ) and Sulfate ( $SO_4^{2-}$ ) parameters. The Lennard-Jones parameters ( $\epsilon$ , and  $\sigma$ ) for Chlorate and Sulfate were taken from work published by Wagner *et al.* and Cannon *et al.* respectively.<sup>8,9</sup> The partial atomic charges were calculated for this work using DFT as outlined below.

Atom Type	$\epsilon$ ( $kcal\ mol^{-1}$ )	$\sigma$ ( $\text{\AA}$ )	$q$ ( $e$ )
$Cl_{ClO_4^-}$	0.040	4.9	0.44
$O_{ClO_4^-}$	0.076	3.1	-0.35
$S_{SO_4^{2-}}$	0.25	3.6	2.4
$O_{SO_4^{2-}}$	0.25	3.2	-1.1

$$\begin{aligned}
 E_{Total} &= E_{Electronic} + E_{Lennard-Jones} \\
 E_{Electronic} &= \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \\
 E_{Dispersion} &= 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]
 \end{aligned} \tag{3.1}$$

$$\begin{aligned}
 \sigma_{ij} &= \frac{\sigma_i + \sigma_j}{2} \\
 \epsilon_{ij} &= \sqrt{\epsilon_i \epsilon_j}
 \end{aligned} \tag{3.2}$$

The Lennard-Jones parameters used for sulfite were derived Cannon *et al.* using Moller-Plesset MP4SDTQ ab initio calculations.<sup>9</sup> The sulfite ion is modelled as a rigid pyramidal species with O-S-O bond angles of  $106^\circ$  and S-O bond lengths of  $1.9\text{\AA}$ . The sulfur atom is

0.76Å out of the plane of the three oxygen atoms. See Figure 2.1 for a schematic representation of the TIP4P model. The full table of parameters for TIP4P is given in Table 2.1. The Lennard-Jones parameters used for chlorate were taken from the literature and were determined by *ab initio* LCAO MO SCF calculations.<sup>8,48</sup> The chlorate ion is also modelled as a rigid pyramidal species, the O-Cl-O bond angles are 105.8° and the Cl-O bond lengths are 1.48Å. The chlorine atom is 0.58Å out of the plane of the oxygen atoms. The partial charges for both ions were calculated in this work using the Bader method<sup>81</sup> within the NWChem<sup>80</sup> DFT package using the B3LYP exchange correlation functional and the 6-311++G\*\* Pople basis set, the charges were calculated for a lone ion in the gas-phase, these charges were fixed during global optimisation.

Water molecules in this study are modelled using the TIP4P four-body potential: a rigid four site molecule is defined with a H-O-H bond angle of 104.52° , with an oxygen lone-pair site represented by a pseudoatom which carries the charge of the oxygen atom but has no Lennard-Jones parameters. The TIP4P potential has been shown to replicate many of the properties of bulk water, while also being computationally inexpensive to model.<sup>7,82,83</sup> The Lennard-Jones and partial charges are listed in Table 3.2.

**Table 3.2** Table of Lennard-Jones parameters and partial charges used for sulfite, chlorate and TIP4P water in this study.<sup>9</sup>

Site	$\epsilon(kcal\ mol^{-1})$	$\sigma\ (\text{\AA})$	$q(e)$
$S_{SO_3^{2-}}$	0.25	3.6	0.281
$O_{SO_3^{2-}}$	0.25	3.2	-0.760
$Cl_{ClO_3^-}$	0.040	4.86	0.452
$O_{ClO_3^-}$	0.076	3.10	-0.484
$O_{H_2O}$	0.648	3.2	0
$H_{H_2O}$	0	0	0.52
$LP_{H_2O}$	0	0	-1.04



### 3.3.1 Energy landscapes

For the landscape exploration of the hydrated sulfite and chlorate systems, the doubly nudged elastic band (DNEB) method was employed to search for transition states between minima identified during global optimisation.<sup>73</sup> The DNEB method is a powerful tool in computational chemistry, designed to locate the transition states connecting different local minima on the potential energy surface. This method involves generating a series of intermediate states, or images, between two known minima and refining these images to accurately map the transition state pathway, and is described in.2

To find these pathways, the DNEB method starts by sampling the space between the end points (the two minima) using a linear interpolation of the translational coordinates of the rigid-body molecules. This approach ensures that the initial guess for the path is straightforward and continuous. Additionally, the rotational coordinates of the molecules are interpolated using spherical linear quaternion interpolation (SLERP), which provides a smooth and consistent transition between orientations.<sup>76</sup> SLERP is particularly useful for maintaining the integrity of rotational movements, ensuring that the interpolation does not produce unrealistic or discontinuous rotations.

Once the initial pathway is established, candidate transition states along this path are identified and further optimised using the hybrid eigenvector-following method.<sup>77,78</sup> This method involves refining the transition states by following the eigenvector corresponding to the negative eigenvalue of the Hessian matrix, which represents the direction of the transition state. By doing so, the method ensures that the located transition state is a true saddle point on the potential energy surface, providing an accurate depiction of the energy barrier between minima.

The resultant energy landscapes, including all identified minima and transition states, are visualised using disconnectivity graphs. These graphs are generated with the PyConnect package,<sup>79</sup> a python package designed to illustrate the connectivity between different minima via transition states. Disconnectivity graphs offer a comprehensive visualisation of the potential energy surface, depicting how different local minima are connected and highlighting the hi-

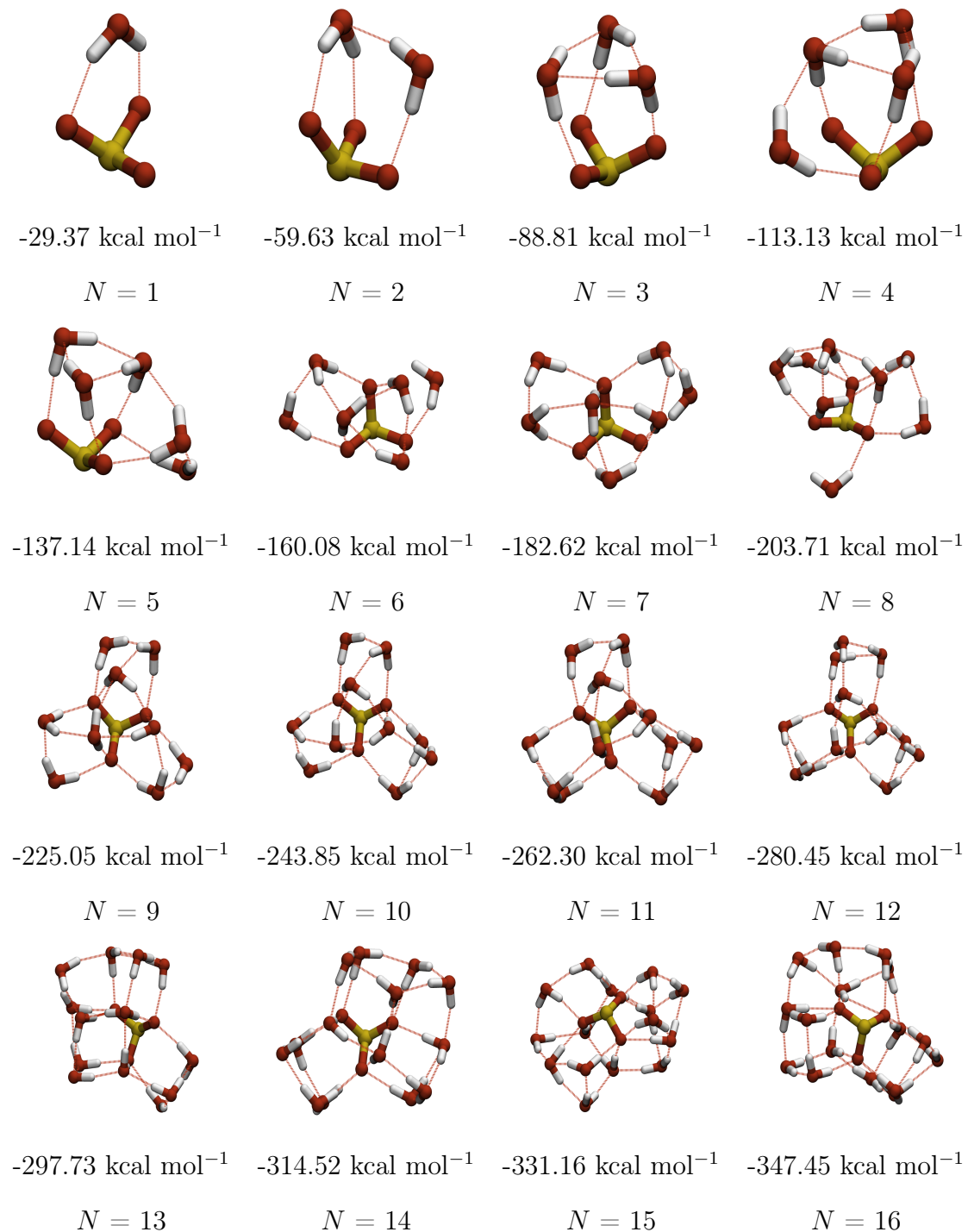
erarchical organisation of the landscape. This visualisation aids in understanding the overall topology of the energy landscape, revealing insights into the stability and dynamics of the hydrated sulfite and chlorate systems.

By utilising these advanced computational techniques, the study provides a detailed and accurate mapping of the potential energy landscapes of hydrated sulfite and chlorate systems. This approach not only identifies the most stable configurations (minima) but also elucidates the pathways and energy barriers connecting these configurations, offering a deeper understanding of the molecular dynamics and stability of these systems.

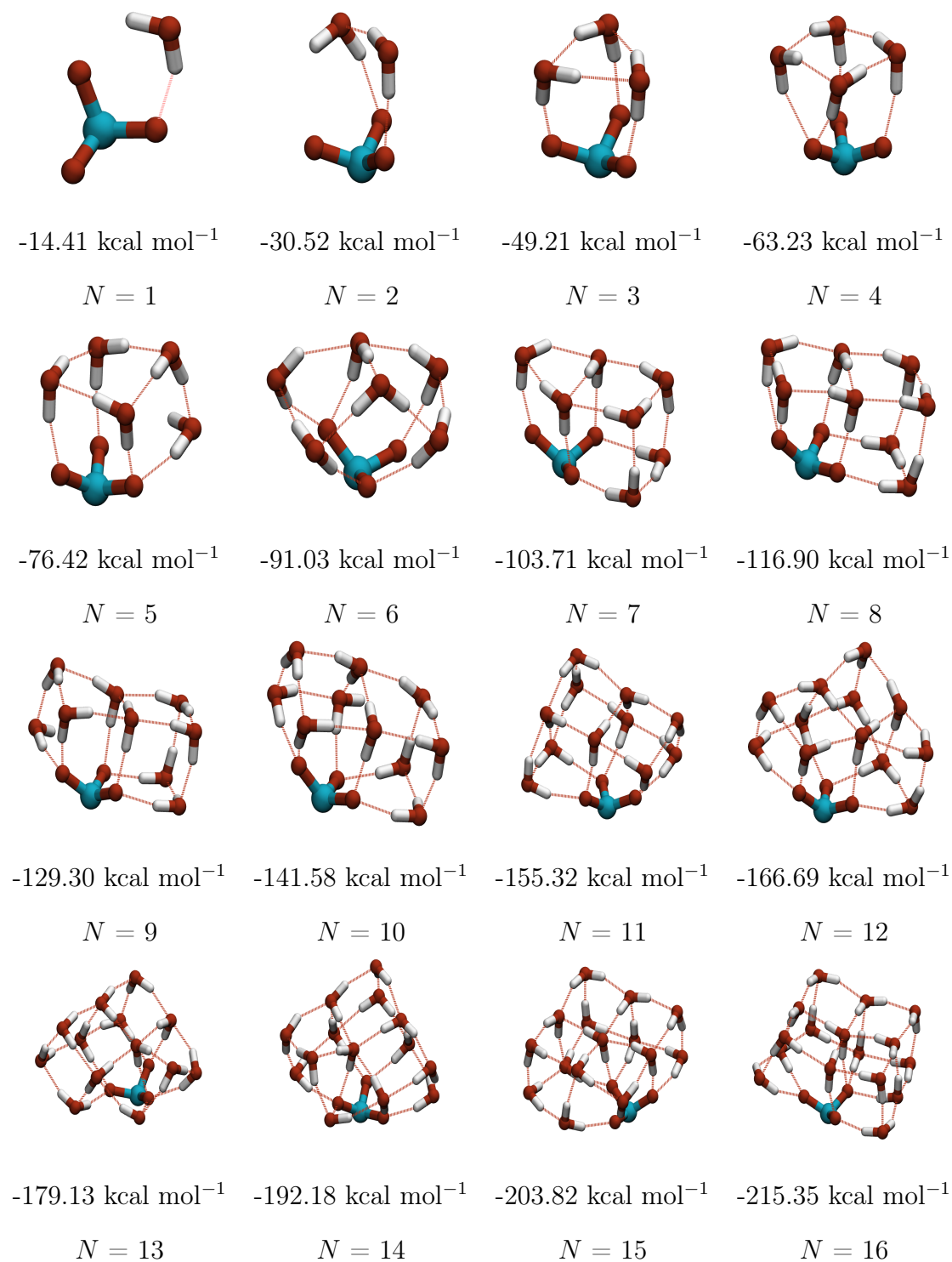
## 3.4 Results and Discussion

### 3.4.1 Sulfite/Chlorate

This section on Sulfite and Chlorate systems is based on work previously published in Hey et al., Isomers and energy landscapes of micro-hydrated sulfite and chlorate clusters 2018.<sup>84</sup>

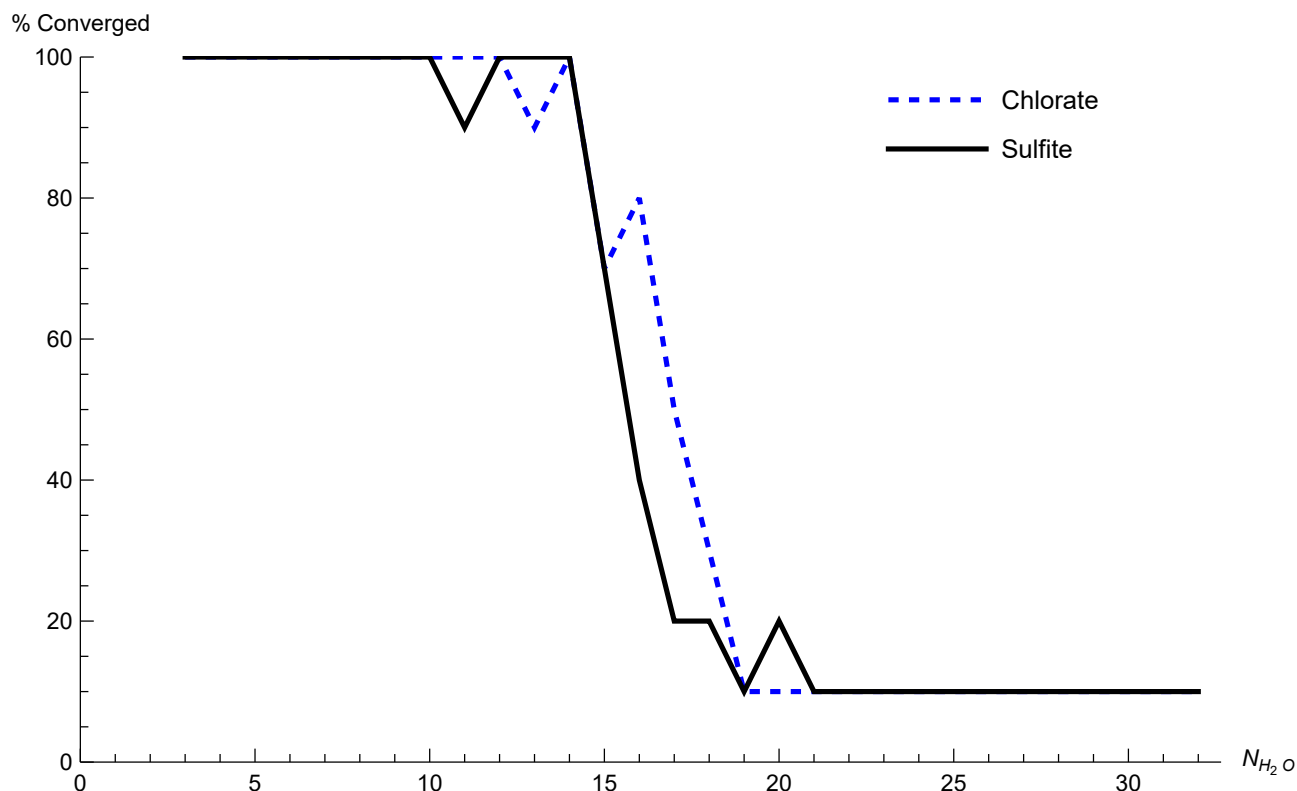


**Figure 3.3** Table of lowest energy empirical minima for the  $\text{SO}_3^{2-}(\text{H}_2\text{O})_N$  system in the range  $3 \leq N \leq 16$ .



**Figure 3.4** Table of lowest energy empirical minima for the  $\text{ClO}_3^-(\text{H}_2\text{O})_N$  system in the range  $3 \leq N \leq 16$ .

The convergence of the basin-hopping runs for the Chlorate and Sulfite systems is shown in Figure 3.5. Ten basin-hopping runs for each fixed number of water molecules in the range  $3 \leq N \leq 50$  were run. This figure shows the fraction of these runs which converged on the same lowest energy minimum. For both systems the majority of the runs converged on the same minimum for fewer than 16 water molecules, the number of independent optimisations converging on the same minimum drops off rapidly for larger cluster sizes. The confidence in having found the true global minimum for the larger cluster sizes is low, but the low energy minima identified are representative of the low energy sections of the landscapes and the structural analyses will still be valid.



**Figure 3.5** Plot showing the fraction of Basin-Hopping runs which converged on the same lowest energy minimum (according to energy and the radius of gyration) for the  $\text{ClO}_3^-(\text{H}_2\text{O})_N$  and  $\text{SO}_3^{2-}(\text{H}_2\text{O})_N$  systems.

It can be seen from Figure 3.3 that for structures with  $N \geq 19$ , the water molecules complete the first solvation shell around the sulfite ion, which adopts a central position in the cluster. This propensity for the sulfite ion to sit centrally within the cluster can be perhaps most clearly illustrated by Figure 3.10, which shows that the ionic displacement for the larger clusters with more than 19 water molecules plateaus at 0.2. There is a void within the cluster inside a cage-

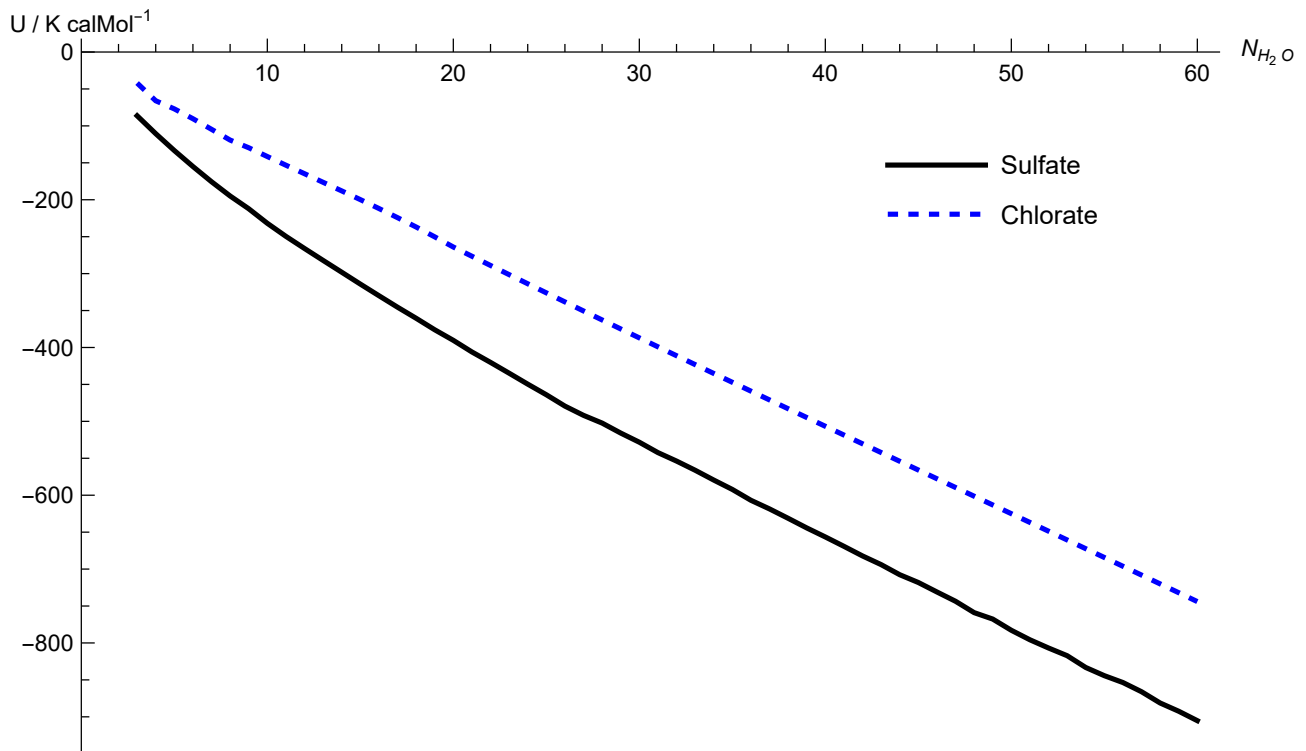
like water structure above the sulfur atom of the sulfite for structures with  $N \geq 19$ , although there is some coulombic attraction between the positively charged S atom and the negative pseudo-atom of the TIP4P water molecules, which carries the partial charge of the oxygen site. The small sulfite clusters ( $N \leq 12$ ) show a high preference for the formation of trimeric hydrogen bond rings, with all of the clusters in this range exhibiting trimeric rings as the only closed loops in their hydrogen bonding networks, which is similar to the behaviour observed for sulfate ions in previous work.<sup>5</sup>

The putative global minima for the chlorate system are presented in Figure 3.4. The hydrated chlorate clusters generally have structural motifs similar to those seen in finite water clusters, displaying a preference for stacked cubes, pentagons and hexagons, which are commonly seen in water clusters.<sup>67</sup> The chlorate ion displays a marked preference for occupying surface sites in the clusters, occupying these sites for all sizes of cluster studied. This preference for water-like structural motifs leads to the chlorate clusters generally having lower symmetry than the corresponding sulfite clusters. As the size of the clusters increases, both systems exhibit larger hydrogen bonding rings, which are common in large pure-water clusters. At these larger sizes, the sulfite continues to occupy sub-surface sites, and the chlorate continues to occupy surface sites.

The evolution of interaction energy per water molecule in the range  $1 \leq N \leq 50$ , shown in Figure 3.6, is obtained by Boltzmann weighting according to equation 3.3:

$$\frac{\bar{U}}{N} = \frac{\sum_i e^{-\Delta U_i \beta} U_i}{\sum_i e^{-\Delta U_i \beta}} \quad (3.3)$$

where:  $\Delta U$  is the interaction energy of cluster  $i$  above the global minimum,  $\beta = 1/k_B T$  and the total cluster energy is denoted by  $U$ . The weighted mean is taken over all structural isomers of unique energy found in all 10 basin-hopping runs for each size of cluster. A theoretical temperature of 130 K is used to maintain consistency with previous computational work<sup>5,74</sup> based on experimental studies of the hydrated sulfate ion.<sup>85</sup> Figure 3.7 shows that the asymptotic limit for the interaction energy for TIP4P water in the bulk ( $\approx -10$  kcal mol<sup>-1</sup>) has not yet been reached at  $N = 50$ . The decreasing magnitude of  $\frac{\bar{U}}{N}$  with increasing  $N$  is due to the dilution effect of the single ion. The deviation from the TIP4P asymptote is greatest for



**Figure 3.6** Total cluster energy for the Sulfite and Chlorate systems calculated using the empirical potential given in Equation 3.1. These are the Boltzmann weighted averages (see Equation 3.3,  $T=130K$ ) of all clusters identified in the combined databases of minima taken from 10 parallel Basin-Hopping runs for each cluster size.

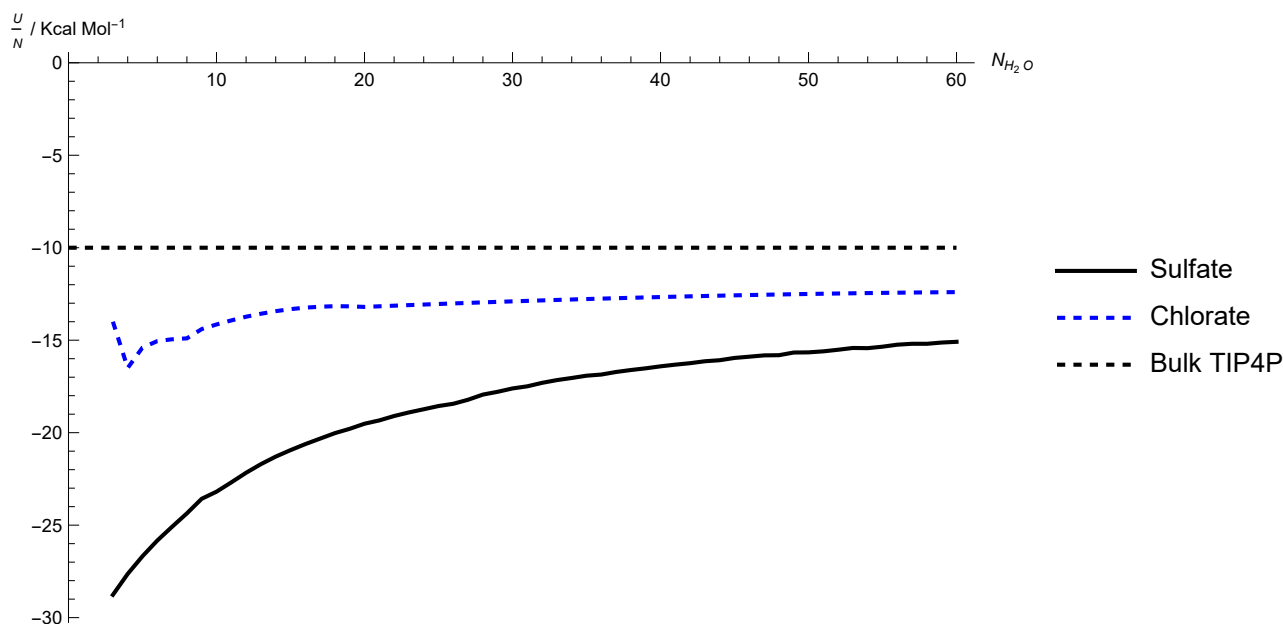
sulfite-water clusters, due to the higher charge of the sulfite ion, leading to a larger coulombic interaction between the TIP4P water molecules and the ion.

It was shown experimentally by Williams *et al.*<sup>85</sup> and then computationally by Smeeton *et al.*<sup>5</sup> that the sulfate ion ( $SO_4^{2-}$ ) promotes long range order within water nano-droplets, with total suppression of hydroxyl groups protruding from the surface of the cluster, (dangling O-H bonds), for all sizes in the range  $18 \leq N \leq 43$ .

The Boltzmann weighted interaction energy per water molecule is given by Equation 3.4:

$$\frac{\bar{U}}{N} = \frac{\sum_i e^{-\Delta U \beta} \left( \frac{U}{N} \right)}{\sum_i e^{-\Delta U \beta}} \quad (3.4)$$

where  $\Delta U$  is the interaction energy of cluster  $i$  above the putative GM,  $\beta = 1/k_B T$  and the average interaction energy per water molecule is denoted by  $\frac{\bar{U}}{N}$ . The weighted mean is taken over all structural isomers of unique energy found in all basin-hopping runs for each size of cluster. A theoretical temperature of 130 K is used to maintain consistency with our other computational studies. The decreasing magnitude of  $\frac{\bar{U}}{N}$  with increasing  $N$  is due to the dilution



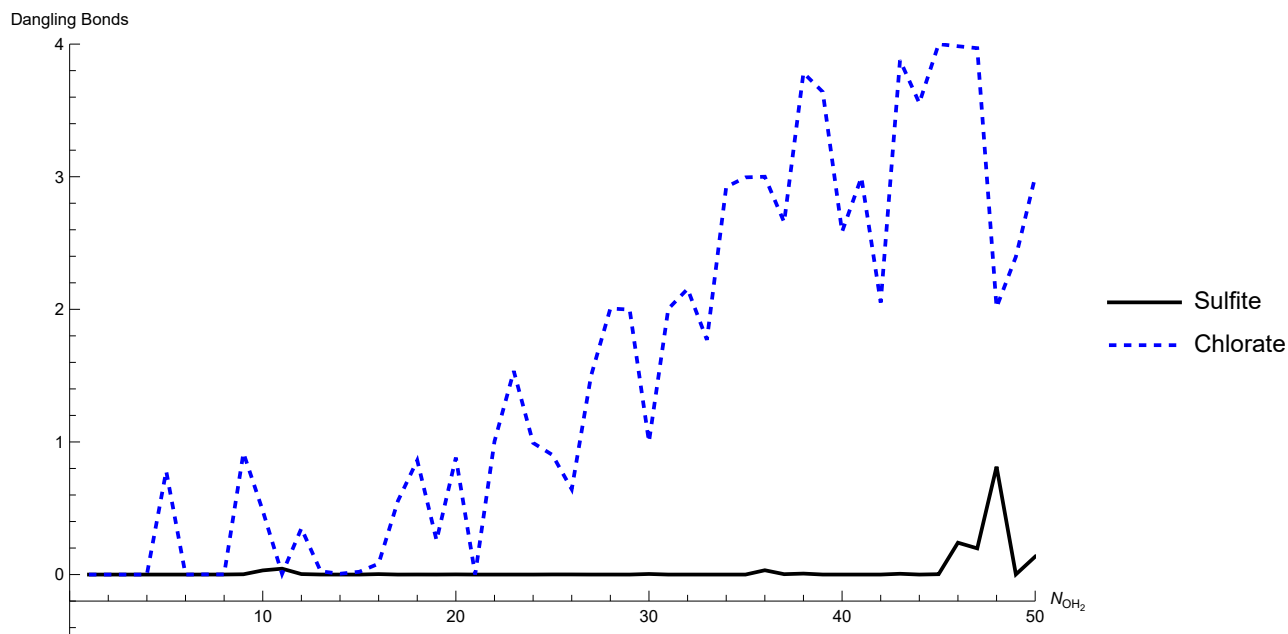
**Figure 3.7** Energy per water molecule for the Sulfite, Chlorate and bulk TIP4P systems

effect of the single ion. The deviation from the TIP4P asymptote is greatest for sulfite-water clusters, due to the higher charge of the sulfite ion, leading to a larger coulombic interaction between the TIP4P water molecules and the ion.

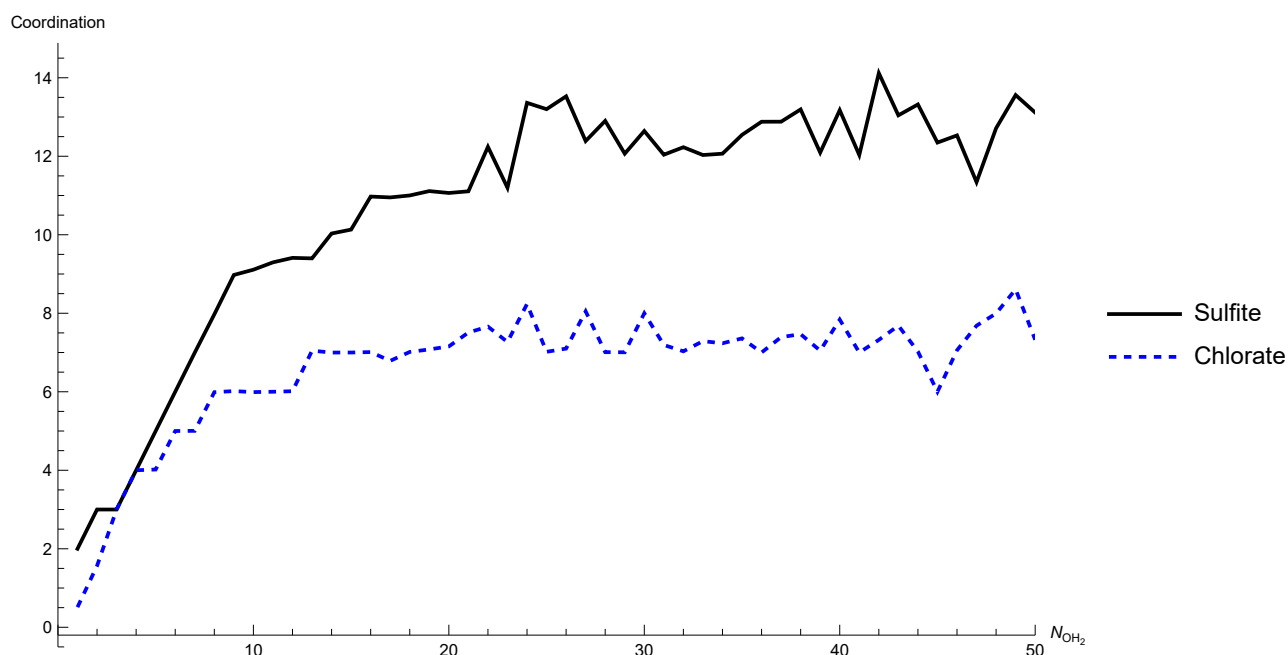
Figure 3.8 shows that the sulfite ion also promotes long-range order within the water clusters by totally suppressing the appearance of dangling O-H bonds for every optimised cluster found with  $\Delta U \leq 5$  kcal mol<sup>-1</sup> in the range  $1 \leq N \leq 50$ . The presence of the sulfite ion induces all the hydrogen atoms of the water molecules in the clusters to interact either with the oxygen atoms of the other water molecules or directly with the oxygen atoms of the sulfite ion. Due to the higher negative charge localised on the oxygen atoms of the sulfite ion, (since the negative charge is only spread over three oxygen atoms), this effect is stronger for sulfite than sulfate, with the effect dominating well into the formation of the third solvation shell. This effect is seen to a lesser degree in the chlorate system, with the chlorate ion providing a greater degree of patterning than the perchlorate system.<sup>74</sup> There is complete suppression of the appearance of dangling O-H bonds in chlorate-water clusters with  $3 \leq N \leq 18$ . In the larger clusters there are fewer unbonded hydroxyl groups than are reported for the equivalent perchlorate clusters (see the following section) of all sizes.

The Boltzmann weighted mean total number of hydrogen bonds donated by the water





**Figure 3.8** Boltzmann weighted mean number of 'dangling' hydrogen bonds ( $n_d$ ) for  $\text{SO}_3^{2-}(\text{H}_2\text{O})_N$  and  $\text{ClO}_3^-(\text{H}_2\text{O})_N$  clusters in the range  $1 \leq N \leq 50$ . Dangling bonds are defined as those identified as being aligned normal, or near normal to the surface of the cluster.



**Figure 3.9** Boltzmann Weighted coordination numbers for the hydrated sulfite and chlorate systems.

molecules of the cluster to the ion is shown in Figure 3.9. Boltzmann weighting is performed across the 1000 lowest energy structurally-unique minima identified by the global optimisation, including the putative global minima. The sulfite ion consistently accepts a greater number of hydrogen bonds than the chlorate ion. This is consistent with the preference of the more highly charged sulfite ion for subsurface sites in contrast to the chlorate which favours surface sites, with the sulfite ion being more accessible to donating water molecules than the chlorate ion at each size. The sulfite ion is observed to coordinate to up to *ca.* 14 water molecules with four or five hydrogen bonds donated to each oxygen atom in the ion. The chlorate ion in contrast coordinates to a maximum of 9 water molecules, with a maximum of three hydrogen bonds donated to each oxygen atom.

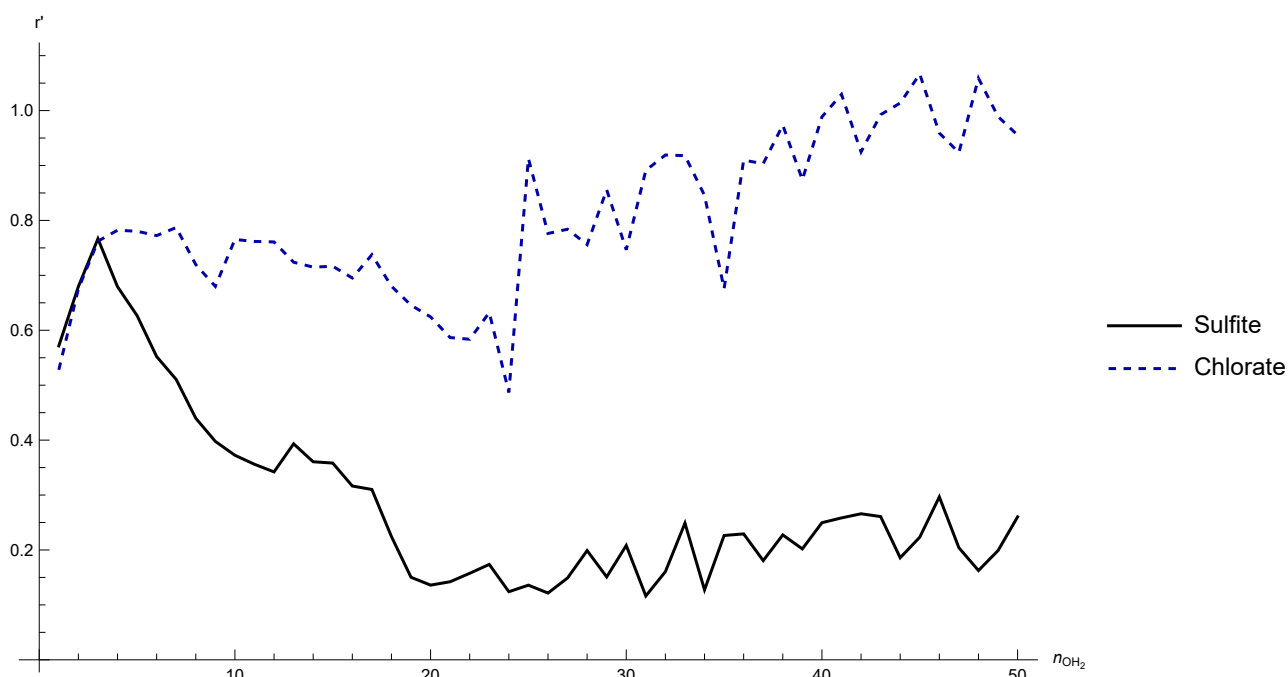
The ratio ( $r'$ ) of the displacement ( $r$ ), of the centre of mass of each ion to the radius of gyration, ( $r_{gyr}$  equation: 3.6), of the cluster for the 1000 lowest energy clusters is Boltzmann weighted according to equation 3.5:

$$r' = \frac{\sum_i e^{-\Delta U \beta} \left( \frac{r}{r_{gyr}} \right)}{\sum_i e^{-\Delta U \beta}} \quad (3.5)$$

$$r_{gyr} = \sqrt{\frac{\sum_{i=1}^n (m_i r_i^2)}{n}} \quad (3.6)$$

Where,  $\Delta U$  is the energy of minimum  $i$  above the global minimum,  $r$  is the displacement of the centre of mass of the sulfite ion from the centre of mass of the water molecules in cluster  $i$ , and  $\beta = 1/k_B T$ .  $m_i$  is the atomic mass of species  $i$ . Figure 3.10 shows that for clusters in the range  $1 \leq N \leq 8$  the sulfite ion lies on the surface of the cluster, i.e. the formation of the water cage around the sulfur site has not yet begun. For clusters with  $12 \leq N \leq 19$  there is a decreasing displacement of the ion from the centre of the cluster as the cage forms around the sulfur atom, and thus the first complete solvation shell is completed. The displacement then remains relatively constant until  $N = 29$ , due to the spherical growth of the second hydration shell. At larger sizes, as the third solvation shell forms, we see the emergence of features very common to pure water clusters, namely stacked cubes, pentamers and hexamers, although we do not see the emergence of the dangling hydrogen bonds which are normally prevalent in those structural features in pure water clusters. The sulfite ion does not remain at the centre of the

cluster, but instead favours subsurface sites, this is visible in Figure 3.3 as well as in this plot with  $r < 0.4$  for all cluster sizes with more than 6 water molecules.



**Figure 3.10** Boltzman weighted ion displacement  $r'$  from radius of gyration of each cluster per number of water molecules

### 3.4.2 Sulfate/Perchlorate

This section on Sulfate and Perchlorate systems is based on work previously published in the Journal Of Physical Chemistry A, under the title: Isomers and Energy Landscapes of Perchlorate-Water Clusters and a Comparison to Pure Water and Sulfate-Water Clusters.<sup>74</sup>

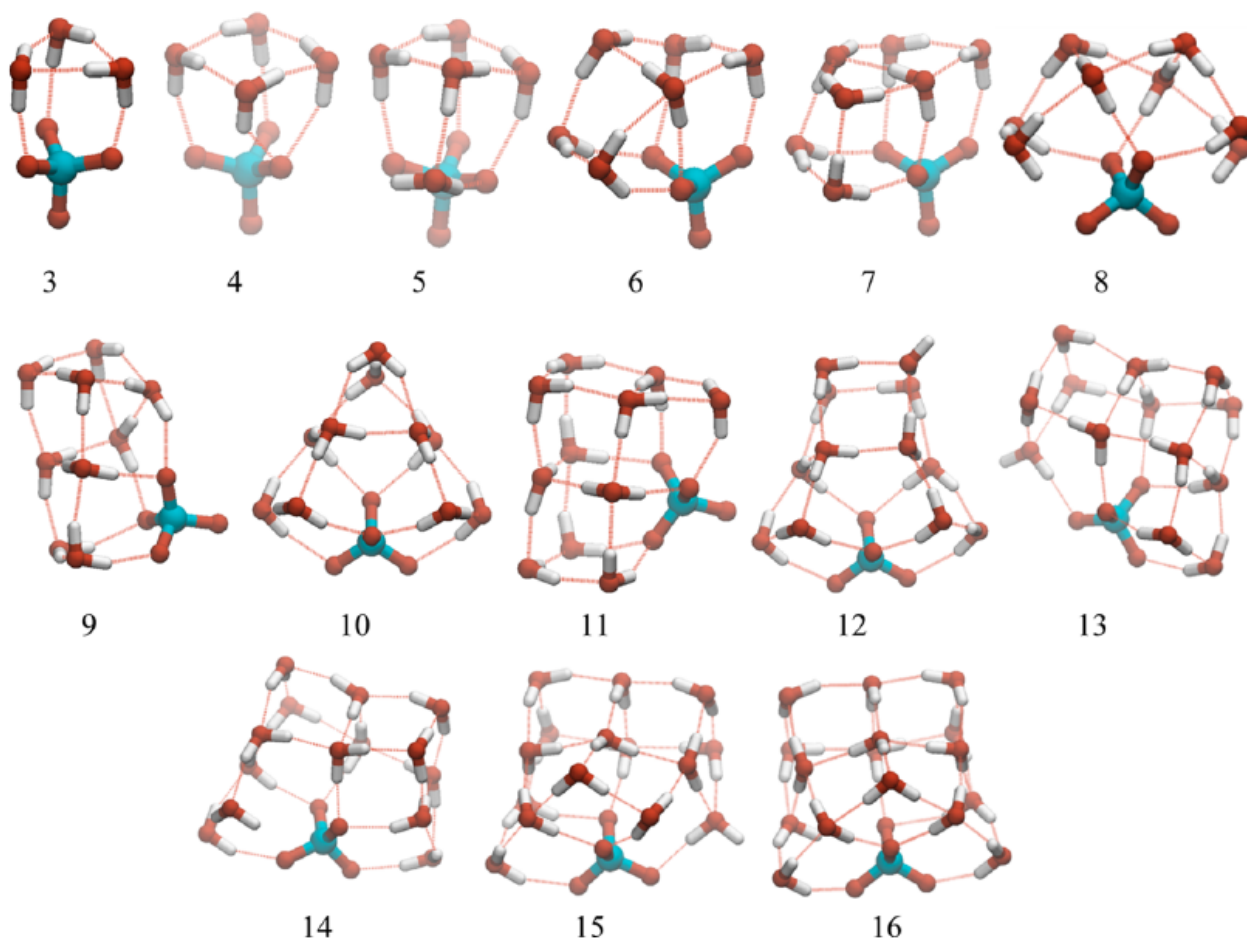
#### Methodology

The methodology employed in this work is the same as that outlined in Section 3.4.

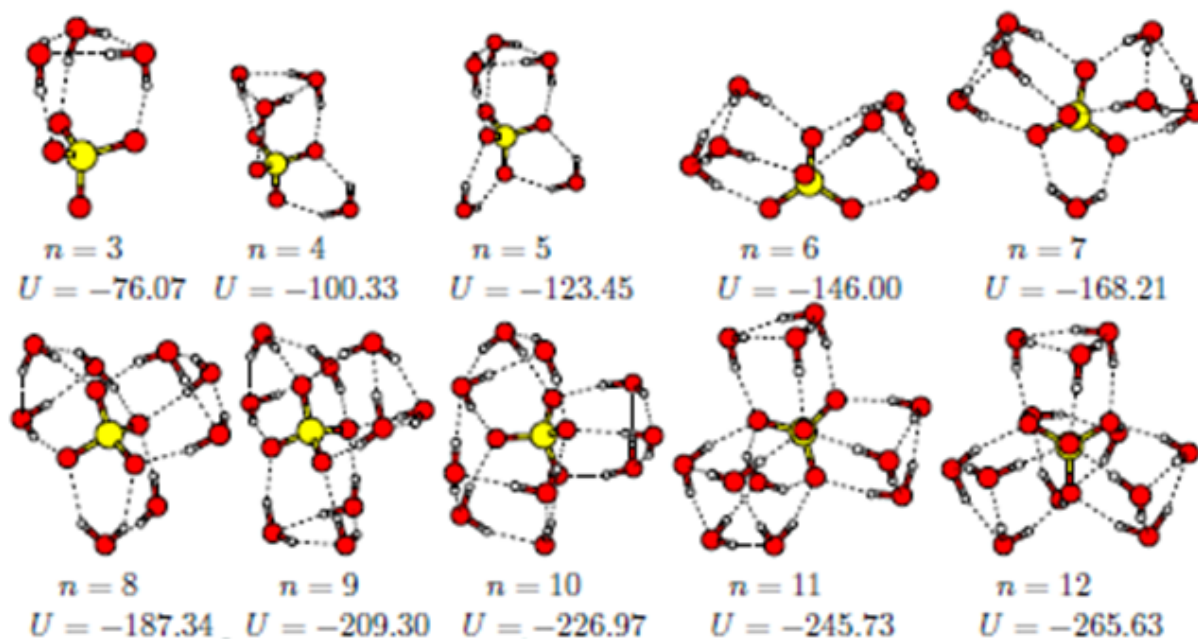
The parameters for the optimisation of the Sulfate and Perchlorate systems are given in Table 3.1.

#### Results

The structures observed for hydrated perchlorate clusters in this study (see Figure 3.11) generally have similar structural motifs to those seen in small pure water clusters, stacked hexagons,



**Figure 3.11** Lowest energy structures identified for the  $\text{ClO}_4^- (\text{H}_2\text{O})_N$  system. This Figure shows only the putative global minima identified for global optimisation runs for which all of the independent basin-hopping runs converged on the same minima ( $3 \leq N \leq 16$ ).



**Figure 3.12** Putative empirical global minima for the hydrated sulfate system as reported by Smeeton et al.<sup>4</sup>

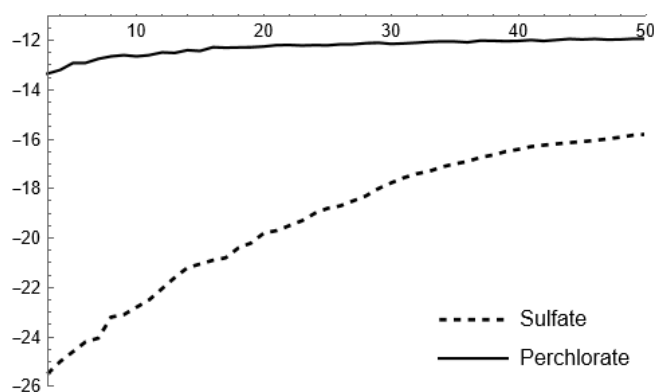
pentagons, and cubes, albeit in somewhat distorted forms. The physical structure of perchlorate-water clusters exhibits unique characteristics influenced by the perchlorate ion's chaotropic nature. Unlike sulfate-water clusters, where the sulfate ion typically resides near the centre of mass, perchlorate ions are often found at the periphery of the cluster. This positioning leads to a more globular and less symmetrical structure for perchlorate-water clusters. The distortion in the structure, compared to pure water clusters, is caused by the smaller effective size of the perchlorate ion relative to the water molecules it replaces.

In hydrated perchlorate clusters, the ion's placement on the cluster surface maximises water-water hydrogen bonding interactions, which are the dominant interactions within these clusters. The perchlorate ion typically acts as a net acceptor of hydrogen bonds from coordinated water molecules, effectively suppressing dangling OH bonds at the surface. This structural motif is evident in clusters up to 16 water molecules, where the lowest-energy configurations are characterized by low-symmetry hydrogen-bonded cages, often seen in small pure water clusters.

So; perchlorate clusters tend to be globular, and display low symmetry. The perchlorate ion is usually found on the periphery of the cluster and is not found in the cluster interior even

at larger cluster sizes. This is indicative of the fact that the water-water hydrogen bonding interactions are the dominant interactions within these clusters.

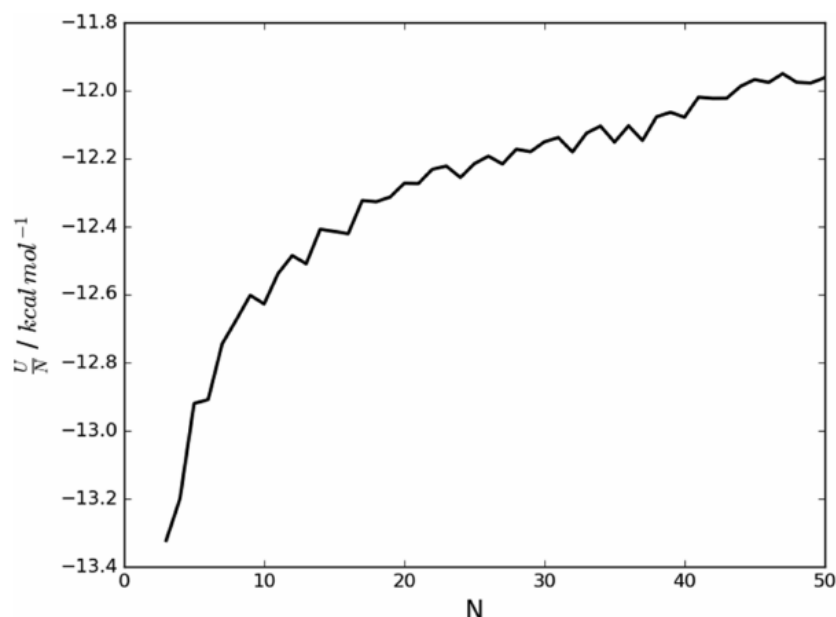
For hydrated sulfate clusters on the other hand (see Figure 3.12) putative global minimum sulfate clusters generally display a much higher degree of symmetry. At smaller sizes from  $3 > n > 12$  sulfate ion sits at the centre of the cluster and favours the formation of three-membered rings of water molecules (water trimers) about each of the faces of the oxygen tetrahedron at  $n = 12$  this process is complete in the putative GM and each face is host to a water trimer. This leads to very spherical clusters with the sulfate close to the centre. At larger sizes, the clusters tend to grow around this core structure, with additional water molecules coordinating to the outside. This leads to the sulfate ion remaining at the centre of the cluster at larger sizes.



**Figure 3.13** Boltzmann weighted ( $T=130K$ ) empirical energies for the putative global minima of the micro-hydrated sulfate and perchlorate systems,  $X_{(H_2O)_N}$ , where  $3 \leq N \leq 50$

Figure 3.13 gives the Boltzmann weighted (according to Equation 3.3) interaction energy per water molecule for the Sulfate and perchlorate systems. The perchlorate system is close to the asymptotic limit for bulk TIP4P water by  $n = 50$ . While the sulfate water clusters still have some way to go before being energetically similar to the bulk.

There are morphological similarities observed between some of the proposed  $ClO_4^-(H_2O)_N$  minima and the structures of pure water clusters reported by Wales et al.<sup>66</sup> The perchlorate anion tends to take the place of 2-3 water sites within the structure, that is two or three of the oxygen atoms of the perchlorate ion take the place of water oxygen atoms in the equivalent sized cluster. This leads to visual and structural similarities with the equivalent sized pure

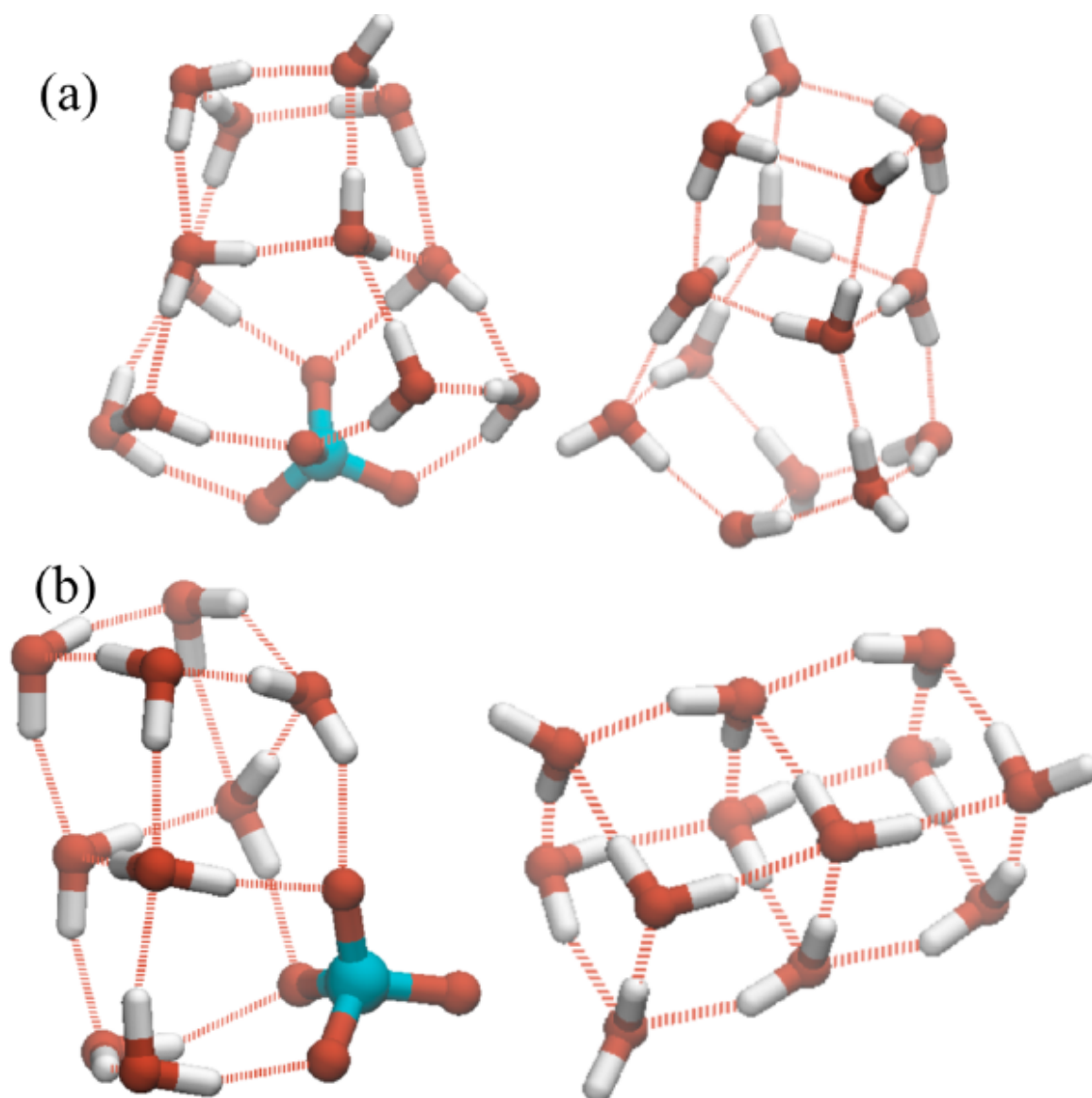


**Figure 3.14** Boltzmann weighted ( $T=130\text{K}$ ) empirical energies-per-water molecule for the putative global minima of the micro-hydrated perchlorate system,  $\text{ClO}_4^- (\text{H}_2\text{O})_N$ , where  $3 \leq N \leq 50$

water cluster. The presence of the perchlorate ion suppresses dangling OH bonds at the surface by acting as a net acceptor of hydrogen bonds from the coordinated water molecules. This effect can be seen clearly in Figure 3.15. The global minima proposed by Smeeton et al. for the hydrated sulfate system share few morphological similarities with the minima proposed for the perchlorate clusters in this study.<sup>5</sup> For sulfate-containing clusters, the ion is almost exclusively located at or near the centre of mass of the cluster, which leads to structures with high symmetry being formed, especially at lower numbers of water molecules ( $N < 25$ ), with examples shown in Figure 3.16. These high-symmetry structures are almost entirely absent in the minima found for small perchlorate clusters, with the exception of the  $\text{ClO}_4^- (\text{H}_2\text{O})_3$  cluster which has the same structure as  $\text{SO}_4^{2-} (\text{H}_2\text{O})_3$  (since the  $(\text{H}_2\text{O})_3$  global minimum (for TIP4P) is a trimeric ring, this is not surprising).

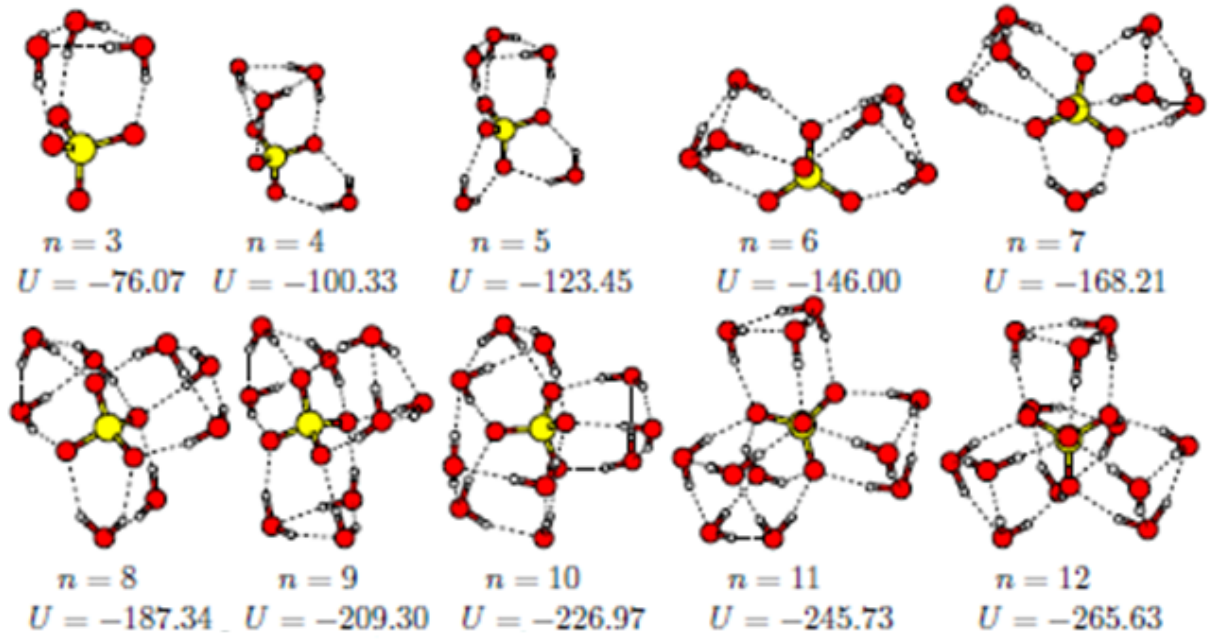
### Investigation into effects of ionic charge on cluster morphology

Since the discussed ion-pairs (sulfite-chlorate and sulfate-perchlorate) are so conformationally similar and yet display clusters with radically different morphologies when solvated, I chose to investigate the extent to which the ionic charge drives the conformational differences observed.



**Figure 3.15** Ball and stick models of the perchlorate and pure water systems showing the similarities in structural morphology displayed between certain low energy minima of differing numbers of molecules





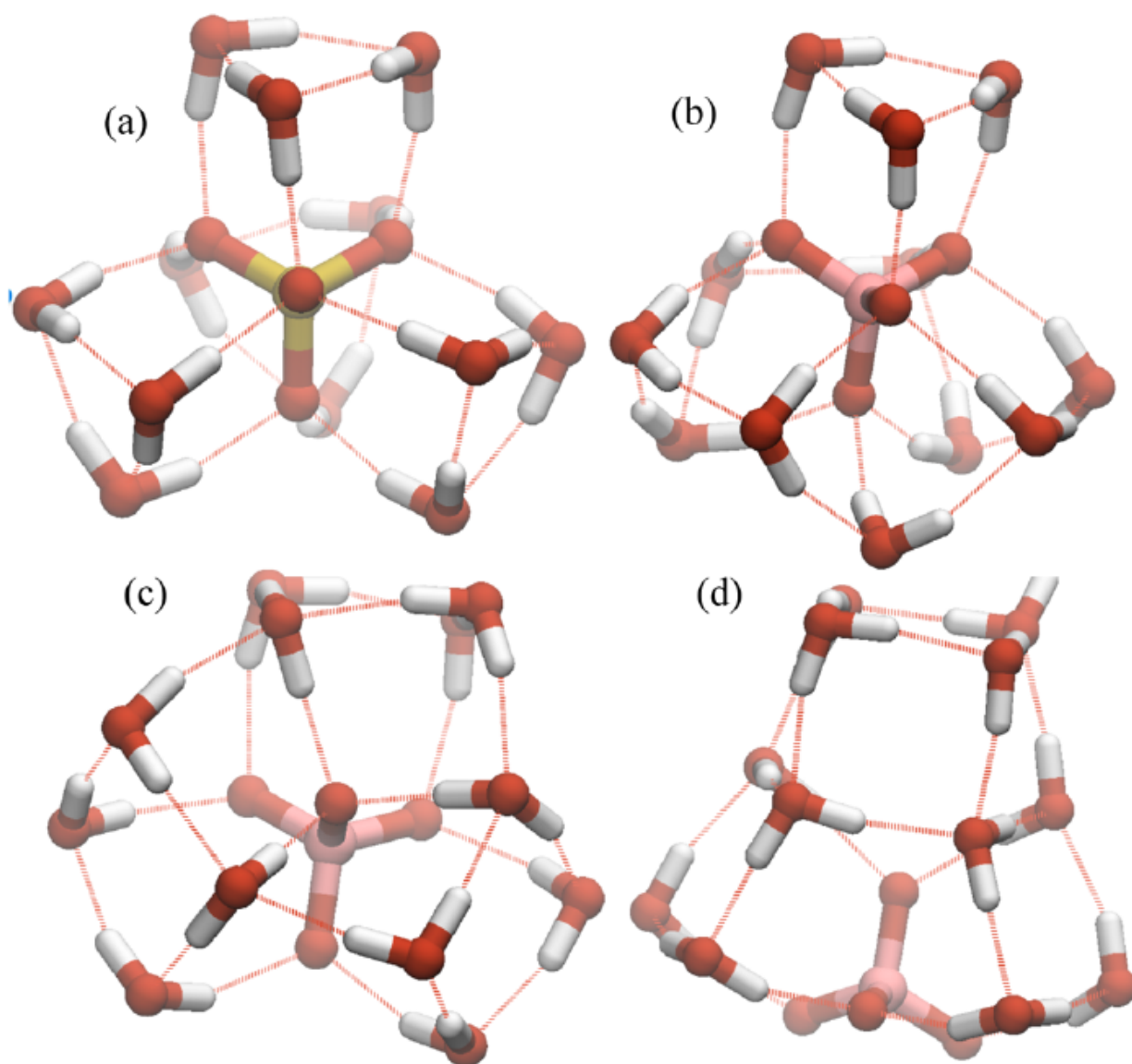
**Figure 3.16** Ball and stick models of example Sulfate-Water clusters as reported by Smeeton *et al.*<sup>5</sup>

**Table 3.3** Charges on the pseudo-sulfur ( $q_X$ ), Oxygen atoms ( $q_O$ ), and overall ionic charge for the  $XO_4(H_2O)_{12}$  system under study.

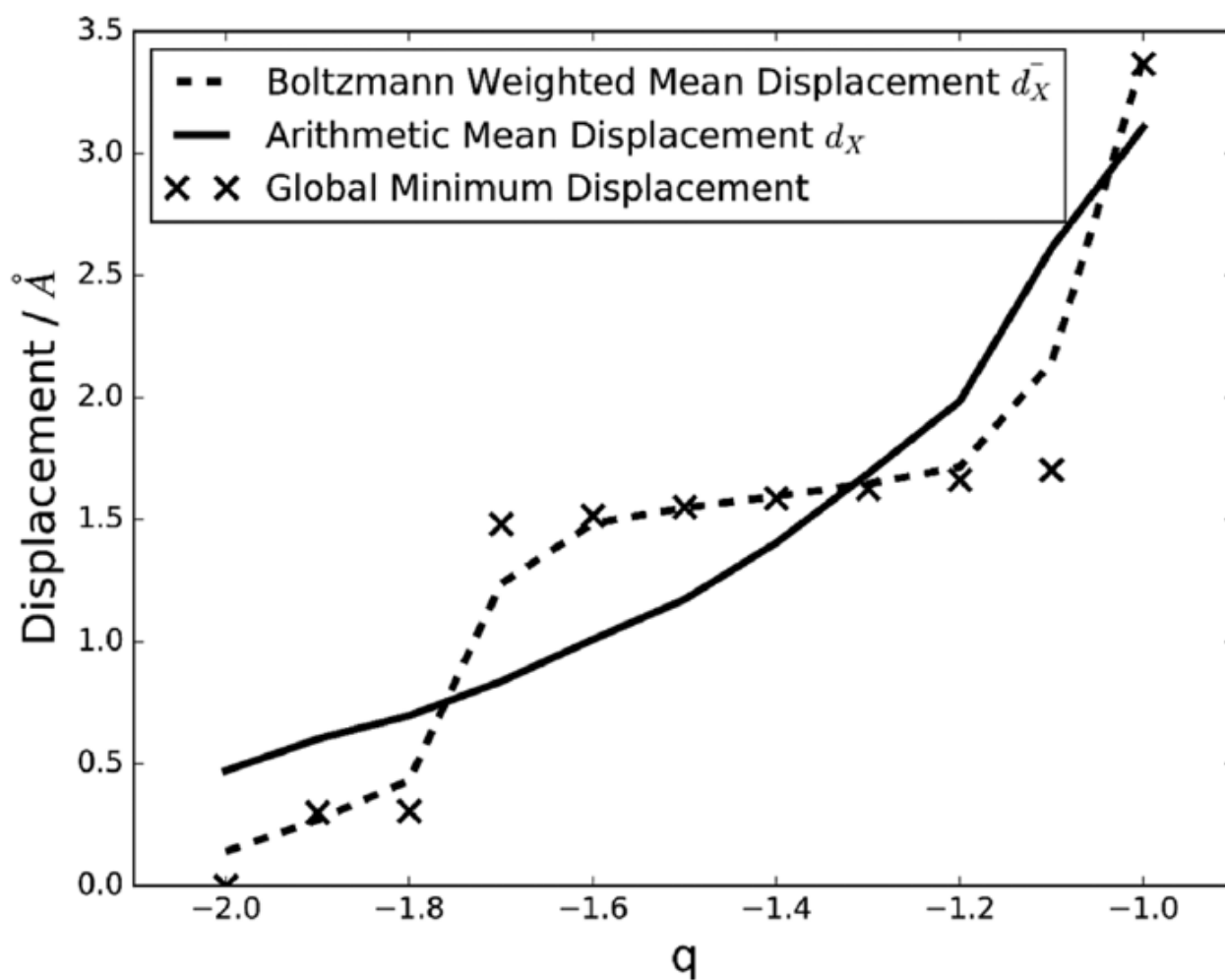
$q_X$	$q_O$	$q$
+2.4	-1.10	-2.0
+2.28	-1.045	-1.9
+2.16	-0.99	-1.8
+2.04	-0.935	-1.7
+1.92	-0.88	-1.6
+1.80	-0.825	-1.5
+1.68	-0.77	-1.4
+1.56	-0.715	-1.3
+1.44	-0.66	-1.2
+1.32	-0.605	-1.1
+1.20	-0.55	-1.0

In order to determine to what extent this is due to the charge on the  $XO_4^q$  oxyanion, the sulfate interaction energy potential<sup>9</sup> was taken and the overall charge ( $q$ ) varied in steps of 0.1e from -2.0 (corresponding to sulfate) to -1.0 (perchlorate), while maintaining a constant ratio between the charge on the central atom and on the oxygen, as listed in Table 3.3. The ion structure and the Lennard-Jones parameters are fixed to the same values as those listed above in Table 3.1 for the Sulfate ion.

The system chosen for this study is  $XO_4^q(H_2O)_{12}$ . For both perchlorate and sulfate systems of this size ( $N_{H_2O} = 12$ ) the global minima are well defined and visually distinct from one another. The same methods as applied above were applied to this system, with 10 parallel basin-hopping runs applied to each charge system. Figure 3.17 shows the four distinct lowest energy structures found with varying  $q$ . Minimum (a) is observed to be lowest in energy for only  $q = -2.0$  and is the same minimum observed for the  $SO_4^{-2}(H_2O)_{12}$  system. This minimum has three hydrogen-bonded trimer rings sitting symmetrically over the faces of the tetrahedral sulfate ion. The sulfate ion lies at the centre of the cluster, and each oxygen atom is coordinated to three water molecules. Similarly, minimum (d) is only seen for  $q = -1.0$  and has the same backbone structure as that previously found as the lowest energy minimum seen for  $ClO_4^-(H_2O)_{12}$  system, consisting of three fused pentagonal prisms. The ion lies at the surface of the cluster, and has a much lower coordination than that seen in minimum (a), with 2 oxygen atoms coordinated to two water molecules and two coordinated only to one, for a total coordination of 6. In the interval between  $q = -2.0$  and  $-1.0$ , there are two intermediate structures, which become more “perchlorate-like” as the total charge approaches  $-1.0$ . Minimum (b) is the minimum for  $q = -1.9$  and  $-1.8$ , while minimum (c) is the lowest energy minimum over a wide range of charges from  $q = -1.7$  to  $-1.1$ . Structures (b) and (c) were also observed as competitive structures for  $SO_4^{-2}(H_2O)_{12}$ . Structure (b) was found to be the second lowest energy structure of  $SO_4^{-2}(H_2O)_{12}$  with  $\Delta U = 0.6 kcal mol^{-1}$  and structure (c) being the 59<sup>th</sup> lowest energy structure found with  $\Delta U = 3.1 kcal mol^{-1}$ . Minimum (b) is a rearrangement of minimum (a) with a 9-membered ring taking the place of three of the trimer rings. As seen in minimum (a) the oxygen atoms of the ion are each coordinated to three water molecules, thus giving a total coordination of 12.



**Figure 3.17** Lowest energy structures observed for the  $XO_4^{q-}$  system. (a)  $q = -2.0$ , (b)  $q = -1.9$  and  $q = -1.8$ , (c)  $-1.7 \leq q \leq -1.1$ , and (d)  $q = -1.0$ .



**Figure 3.18** Displacement against overall charge ( $q$ ) for the  $XO_4^q(H_2O)_{12}$  system in the range  $-2.0 \leq q \leq -1.0$ .

The displacement of the heavy pseudo-atom (X) from the geometric centre of the cluster ( $d_x$ ) is used as a metric to distinguish between sulfate-like and perchlorate-like clusters. In sulfate-water clusters the ion is located in a very central, subsurface location, whereas in perchlorate-water clusters the ion adopts a surface position and is located further from the centre of the cluster. Values for displacement are shown in Figure 3.18 for the Boltzmann weighted mean displacement across the 2000 lowest energy structures, the Arithmetic mean across the 100 lowest energy structures, and for the lowest energy structure. The  $q = -2.0$  minimum is the same minimum as observed for the Sulfate system at  $N_{H_2O} = 12$ , this is a pseudo-spherical cluster with the Sulfur lying at the centre, and thus the displacement in of the putative GM is very low, other low-energy minima skew the arithmetic and Boltzmann weighted means to have slightly higher displacements. Instances of minima (b) and (c) are seen as relatively high energy minima. Minimum (b) is observed for  $q = -1.9$  and  $-1.8$ , this minimum sees the ion offset slightly from the centre of the cluster. The nonomer ring has a larger H-O-H angle than in minimum (a) and thus the reduced strain makes this energetically favourable when the OH-OX interaction's coulombic component is weakened. Minimum (c) has larger OH-O water-water hydrogen bond angles again. This also allows for the ion to lie further from the centre. Minimum (d) shares a structure with the Perchlorate putative GM for the twelve water system.

Across all the charges from  $q = -2.0$  to  $q = -1.0$ , I observe an increasing displacement of the ion, even in minima that appear visually similar and share the same hydrogen bonding network. This trend is evident in the GM displacement, BW displacement, and mean displacement. As the difference in hydrogen-ion charge decreases, the clusters become increasingly dominated by water-water hydrogen bonding interactions due to the reduced ion charge. Consequently, the ion is gradually pushed out from the centre of the cluster, as the dominance of water-water hydrogen bonding promotes a higher number of these hydrogen bonds.

In conclusion, this investigation into the effects of ionic charge on cluster morphology reveals that the position and coordination of the ion within the cluster are significantly influenced by the ionic charge. By varying the charge on the  $XO_4^q$  oxyanion from  $-2.0$  (sulfate) to  $-1.0$  (perchlorate), it was observed that sulfate clusters exhibit a central ion position and high

coordination numbers, forming highly symmetrical structures. In contrast, perchlorate clusters tend to have the ion at the surface with lower coordination numbers and less symmetry. The displacement of the ion from the cluster centre increases as the ionic charge decreases, indicating that reduced Coulombic interaction leads to a dominance of water-water hydrogen bonding interactions. This results in the ion being gradually excluded from the cluster

### 3.5 Conclusion

This study highlights the distinct morphological differences between sulfate and perchlorate micro-hydrated clusters, driven primarily by the ionic charge of the oxyanion. For sulfate clusters, the ion typically resides near the centre of mass, resulting in highly symmetrical structures, especially for smaller clusters ( $N < 25$ ). In contrast, perchlorate clusters display less symmetry and more surface-localised ions, with significant structural differences even in small clusters (e.g.,  $X(\text{H}_2\text{O})_3$ ). These differences underscore the role of the ionic environment in determining cluster morphology.

The investigation also included sulfite-chlorate pairs, which, despite their conformational similarity, exhibited drastically different morphologies when hydrated. Sulfite clusters showed a similar central ion positioning to sulfate, while chlorate clusters resembled perchlorate in having ions at the surface. This further supports the hypothesis that the ionic charge and the resulting electrostatic interactions significantly influence cluster structure.

To understand the influence of ionic charge on cluster morphology, the charge on the  $\text{XO}_4$  oxyanion was varied from  $-2.0$  (sulfate) to  $-1.0$  (perchlorate). For both perchlorate and sulfate systems of size  $X(\text{H}_2\text{O})_{12}$ , the global minima were well-defined and visually distinct from one another. It was observed that sulfate clusters maintained a central ion position and high coordination numbers, forming pseudo-spherical structures with the ion at the cluster centre. In contrast, as the charge decreased, I show that the clusters exhibit more perchlorate-like characteristics, with the ion increasingly displaced from the centre and lower coordination numbers. Specifically, the study identified four distinct lowest-energy structures as the ionic charge was varied: (a) a central ion configuration for  $q = -2.0$ , (b) an offset ion configuration for  $q = -1.9$  and  $q = -1.8$ , (c) intermediate structures for  $q$  ranging from  $-1.7$  to  $-1.1$ , and (d) a surface ion configuration for  $q = -1.0$ .

The displacement I observe of the ion from the geometric centre of the cluster ( $d_x$ ) was used as a metric to distinguish between sulfate-like and perchlorate-like clusters. In sulfate-water clusters, the ion is located in a very central, subsurface location, whereas in perchlorate-water

clusters, the ion adopts a surface position and is located further from the centre. The displacement trend was evident across the Boltzmann-weighted mean displacement, the arithmetic mean displacement of the lowest energy structures, and the lowest energy structure itself. For  $q = -2.0$ , the minimum displacement corresponded to the sulfate system's pseudo-spherical cluster. As the charge decreased, the ion displacement increased, reflecting a shift from central to surface positioning.

My analysis showed a progressive displacement of the ion from the cluster centre as the ionic charge decreased, regardless of the visual similarity in hydrogen bonding networks. This displacement trend is evident in the lowest energy structures and across the Boltzmann-weighted mean displacement data. This study concludes that the shift in ion position is due to the reduced Coulombic interaction between the ion and water molecules, resulting in water-water hydrogen bonding becoming the dominant interaction. This leads to the ion being gradually expelled from the centre of the cluster to maximise the number of favourable water-water hydrogen bonds.

Overall, my findings in this chapter underscore the significant role of ionic charge in determining the structural characteristics of hydrated ion clusters. By elucidating the interplay between ionic interactions and hydrogen bonding, this study provides valuable insights into the behaviour of micro-hydrated systems, which can inform future research in the field of cluster chemistry and related applications.



# Chapter Four

## Hydrated Metal Clusters

### 4.1 Introduction

Nano-alloys (NAs) are nano-scale metal clusters comprised of two or more metallic elements. The combination of multiple metallic elements results in properties which are different from the equivalent monometallic clusters and dependent not only on the size and shape of the cluster, but on the composition and the ordering of the elements within the cluster.<sup>86</sup>

Hydrated gold-silver nanoclusters (Au-Ag NAs), indeed, present a remarkable platform for catalytic applications, stemming from their unique atomic structure, electron distribution, and surface morphology.<sup>87,88</sup>

As is known, the catalytic performance of a material is primarily dependent on its active surface area and the electronic state of the active sites. In the case of Au-Ag NAs, the nano-scale dimension offers an enormous surface-to-volume ratio, providing a high density of active sites. Meanwhile, the introduction of water molecules (as implied by 'hydrated') possibly results in hydrophilic surfaces, which can facilitate catalytic reactions, particularly in aqueous environments.<sup>86,89</sup>

Another noteworthy aspect is the alloying effect of gold and silver. Alloying typically leads to the alteration of the electronic structure of the individual metals, thus influencing the chemisorption and activation of reactant molecules - key steps in catalytic reactions. By adjusting the Au:Ag ratio and their spatial arrangement, we can potentially modulate the charge

distribution over the active sites and the overall electronic state of the nanoclusters, thereby tuning their catalytic behaviour.<sup>86,90</sup>

Additionally, the complimentary effect between gold and silver atoms might significantly contribute to the catalytic enhancement of these nanoclusters. While silver is known for its high electronic conductivity and gold for its high resistance to oxidation, the combination of both can result in improved catalyst durability and reactivity. For instance, these nanoclusters can serve as excellent catalysts in oxidation reactions, fuel cell technology, and  $CO_2$  reduction, to name a few.<sup>86,91</sup>

The quantum size effect also plays a vital role in determining the catalytic properties of Au-Ag NAs. As the cluster size decreases to a few nanometres or less, quantum confinement can lead to discrete electronic states, which can induce unique reactivity compared to bulk materials. This phenomenon can be exploited to design catalysts with enhanced selectivity and efficiency.<sup>86,89</sup>

The work presented in this chapter builds upon some previous work conducted by Shayeghi *et al.* in 2014.<sup>92</sup> In this published work, the study investigates the structures of small Ag-Au nanoclusters, specifically  $Ag_nAu_m^+$  ( $n + m = 1 - 4$ ), using both experimental and theoretical approaches. The Birmingham Cluster Genetic Algorithm (BCGA) coupled with density functional theory (DFT) was employed to determine the lowest energy configurations. The results revealed three primary structural motifs: the rhombus, distorted tetrahedron, and Y-shaped structures. For  $Ag_3Au^+$ , the Y-shaped structure with the gold atom centrally located is energetically favoured, aligning with experimental ion mobility measurements. The  $Ag_2Au_2^+$  clusters exhibited a mix of rhombus, tetrahedral, and Y-shaped configurations, indicating close competition among these isomers. For  $AgAu_3^+$ , the rhombus structure was predominantly observed, with significant stabilisation due to heterometallic bonding. Overall, the study provides insights into the preferred geometries and electronic distributions in small Ag-Au clusters, highlighting the role of charge transfer and atomic arrangement in determining cluster stability and properties.<sup>92</sup> We extend this work to include the hydration of these cationic clusters and compare

the observed configurations with those that are stabilised in the presence of water molecules.

## 4.2 Methodology

### 4.2.1 Global Optimisation using GIGA

In this study, low-energy minima for the  $Ag_mAu_n(H_2O)_X$  system are found using the latest iteration of the BPGA (Birmingham Parallel Genetic Algorithm) code referred to as GIGA, which allows direct DFT global optimisation of nanoalloy clusters in the presence of ligands directly at the DFT level.<sup>93–96</sup>

This Genetic Algorithm (GA) employs a pool methodology instead of the more traditional generation-based approach, allowing us to minimise and evaluate structures in parallel, with several GA instances operating on the same database of minima at the same time.

The specifics of how Genetic Algorithms operate will be discussed in greater depth in the following chapter, so this section will be kept concise to avoid redundancy.

### 4.2.2 Crossover

BPGA/GIGA implement<sup>94,96</sup> the Deaven and Ho Cut and Splice method.<sup>1</sup> This scheme is explained in detail in section 5.2.1 starting on page 86, but I will cover it briefly here.

The Deaven-Ho cut and splice crossover is a genetic algorithm technique used to generate new offspring by recombining genetic material from two parent structures. In this method, each parent structure is divided at randomly selected points, and the segments from different parents are then spliced together to form new offspring. This approach is particularly useful for problems involving spatial structures, such as molecular configurations, as it allows for the creation of new configurations by combining substructures from each parent. The Deaven-Ho cut and splice method helps maintain diversity in the population by producing offspring that inherit distinct structural features from each parent, potentially leading to novel and improved solutions.

### 4.2.3 Mutations

The subset of mutations within GIGA<sup>96</sup> used for this study are:

### Single atom translations

Single atom translations are the simplest mutations employed in this work. A single atom is selected at random from the metal cluster. A random vector is applied to the selected atom, and the new cluster is checked for overlap. Overlap here is defined as two atomic centres being within 0.6 Å. It is possible to configure GIGA to allow for multiple independent single atom translations to occur for each mutation, but for this study only a single atomic translation was allowed.

### Multi-atom translations

In multi-atom translations, a random number of atoms are chosen at random (up to N atoms, where N is the total number of atoms in the metal cluster). A single random vector is applied to all of the selected atoms, so they are moved as a group.

### Cluster rotations

In the cluster rotation scheme a random axis is selected that passes through the geometric centre of the metal cluster. A random rotation of  $0-2\pi$  is then applied to the entire metal cluster around that axis. This rotation is applied to only the cluster atoms, not to the ligands.

### Ligand translations

In ligand translation mutations a random number of ligand molecules is selected from 1-n where n is the total number of ligands. Rigid body translations are applied to each ligand molecule, the molecule's conformation is frozen and the same translation vector is applied to all the atoms in the ligand molecule. Different translation vectors are applied to each ligand molecule selected.

These Ligand Translation mutations are intended to allow the ligands to explore the surface of the metal cluster. As the translations applied are relatively small compared to the total size of the cluster, sometimes ligands can become trapped in local minima. To offset this problem, if no minima with better energy than the global minima has been found for a number of steps, the step-size is increased.

### Ligand rotations

Similar to the ligand translation moves described above, random number of ligand molecules is selected from 1-n where n is the total number of ligands. Acting independently on each chosen ligand molecule, a random axis is selected that passes through the geometric centre of the ligand molecule. A random rotation of 0-2 $\Pi$  is then applied to the entire ligand molecule about this axis. If more than one ligand molecule is to be mutated, a new axis and random rotation is chosen and applied to the next molecule.

### Ligand inversion

In ligand inversion mutations, one ligand molecule is selected. For each atom in the ligand molecule the vector between the atom and the geometric centre of the metal cluster is calculated. This vector is doubled and applied to the atom's position. This process is then repeated for the next atom in the ligand.

The intention of the ligand inversion mutation is to allow the ligands to rapidly explore the entire surface of the metal cluster. In monometallic or alloy clusters this can allow the the ligand to escape kinetic traps that small translations or perturbations may not escape. In alloy clusters these moves increase the likelihood of the ligand interacting with different metal atom types.

## 4.2.4 Post-genetic step

After Crossover and Mutation steps some sanity-checks are performed to ensure that there are no overlapping atoms (any pairs of atomic centres having a separation of  $\leq 0.6$  Å), with the output of the step being discarded and the original cluster/clusters being re-perturbed if this is detected. As BMPGA/GIGA is a Lamarckian GA, local optimisation is conducted after each genetic move (crossover or mutation) and the locally minimised conformation is used as the resulting child.

### 4.2.5 DFT

During global optimisation using GIGA, the PBE0 exchange correlation functional as implemented in NWChem is used with Dunning’s correlation consistent cc-pVDZ basis set for the light atoms (H and O) and cc-pVDZ-PP for the heavy atoms (Ag and Au) to account for the additional polarisation on the heavier atoms.<sup>56,97,98</sup> The structures obtained using this method are then re-optimised using the  $\omega$ B97X-D long-range corrected hybrid functional which includes empirical dispersion corrections using a damped version of Grimme’s D2 dispersion correction.<sup>98</sup> All energies and structures reported have been reoptimised using this potential.

### 4.2.6 Energetics

The hydration energies per water molecule reported in this chapter are calculated according to Equation 4.1:

$$E_{hyd(Ag_mAu_n(H_2O)_X)} = E_{Ag_mAu_n(H_2O)_{X-1}} + E_{H_2O} - E_{Ag_mAu_n(H_2O)_X} \quad (4.1)$$

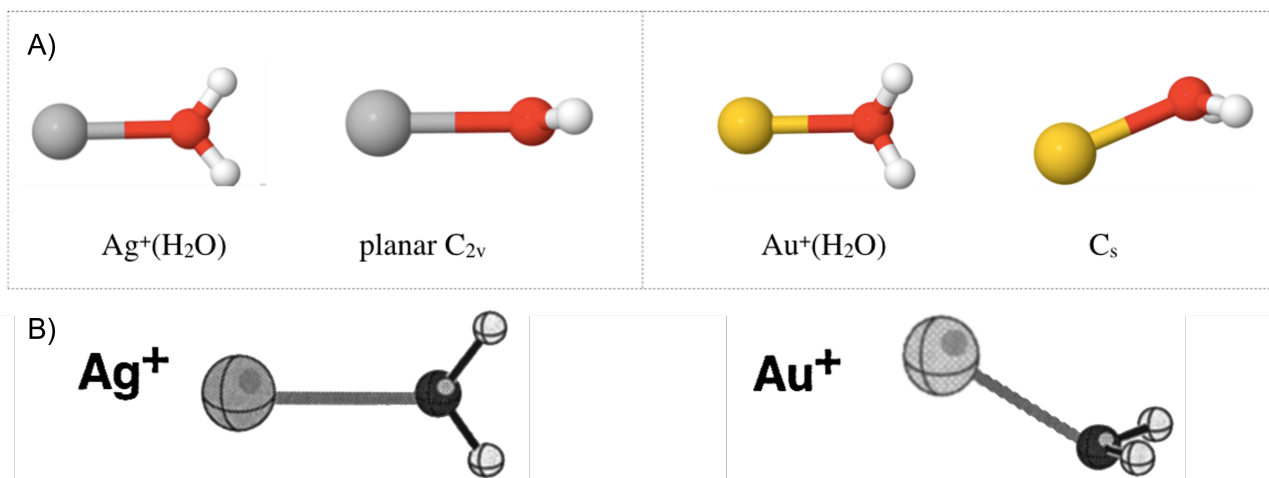
Where:  $E_{Ag_mAu_n(H_2O)_X}$ , and  $E_{Ag_mAu_n(H_2O)_{X-1}}$  are the total energy (EV) of  $Ag_mAu_n(H_2O)_X$  and  $Ag_mAu_n(H_2O)_{X-1}$  clusters respectively.  $E_{H_2O}$  is the energy calculated for a single water molecule in vacuum.

## 4.3 Results and Discussion

### 4.3.1 Structures of Hydrated $\text{Ag}_m\text{Au}_n^+$ Clusters

Unless stated otherwise, all the minima, and putative global minima given in this section are those obtained using  $\omega\text{B97X-D}$ . Re-optimisation of the global minima using the  $\omega\text{B97X-D}$  functional from those produced within GIGA using aug-cc-pVDZ/aug-cc-pVDZ-pp yielded no changes in overall geometry or energy ordering.

It can be seen in Figure 4.1 that using the aug-cc-pVDZ/aug-cc-pVDZ-pp selected for this study we reproduce the same geometries as previously reported by Feller *et al.* for the Silver and Gold cations with single water molecules. Notably we observe the same planar  $\text{C}_{2v}$  geometry for the  $\text{Ag}^+(\text{H}_2\text{O})$  interaction. We also find the previously reported  $\text{C}_s$  geometry for  $\text{Au}^+\text{H}_2\text{O}$ . It should be noted that when re-optimising using the  $\omega\text{B97X-D}$  functional, we see no change in geometry.

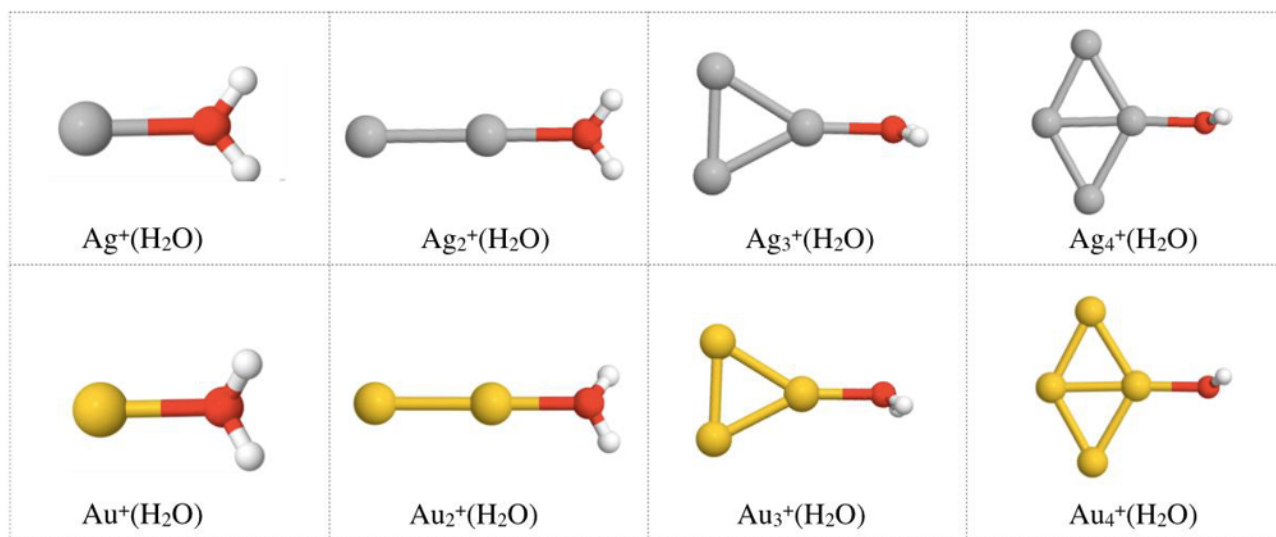


**Figure 4.1** Comparison between our (A) and previously found (B) geometries for  $\text{Ag}^+\text{H}_2\text{O}$  and  $\text{Au}^+\text{H}_2\text{O}$ . B shows the geometries as reported by Feller, Glendening, and Jong in 1999.<sup>6</sup>

Figure 4.2 shows the geometries of the  $\text{Ag}_n^+(\text{H}_2\text{O})$  and  $\text{Au}_n^+(\text{H}_2\text{O})$  clusters for 1-4 metal atoms. Notably the Au-O interaction is always observed occurring through one of the Oxygen lone-pair sites, whereas the Ag-O interaction is consistently occurring equally through both oxygen lone-pair sites. This leads to the previously noted planar  $\text{C}_{2v}$  geometry being observed for the  $\text{Ag}^+(\text{H}_2\text{O})$  and the  $\text{C}_s$  geometry for  $\text{Au}^+\text{H}_2\text{O}$  being maintained for all the monometallic clusters.



It is also notable that the normally tetrahedral  $\text{Au}_4^+$  cluster is not the observed minimum when stabilised by a water interaction, though it is observed as a meta-stable state.



**Figure 4.2**  $\omega\text{B97X-D}$  geometries for  $\text{Ag}_n^+(\text{H}_2\text{O})$  and  $\text{Au}_n^+(\text{H}_2\text{O})$  for  $n = 1 - 4$ , found using the aug-cc-pVDZ/aug-cc-pVDZ-pp DFT basis set.

Figure 4.3 shows the aug-cc-pVDZ putative global minima found using GIGA for the doubly hydrated monometallic  $n = 2 - 4$   $\text{Ag}_n^+(\text{H}_2\text{O})_2$  and  $\text{Au}_n^+(\text{H}_2\text{O})_2$  clusters. One water molecule coordinates to each of the silver atoms in the  $\text{Ag}_2^+$  cluster. Like with all of the water molecules interacting with these cationic clusters, these water molecules coordinate to the metal cluster via the oxygen atom. The hydrogen atoms of the two water molecules are in a staggered conformation compared to one another, this maximises the distance between them.

In the  $\text{Ag}_3^+$  cluster putative global minimum, the two water molecules interact with adjacent Ag atoms. The two oxygen atoms are close to planar with the silver atoms, with an O-Ag-Ag-O dihedral angle of  $-1.9^\circ$ . The Ag-Ag-O angle is  $150^\circ$  for both. Again, the hydrogen atoms of the water molecules are angled away from each other to maximise distance.

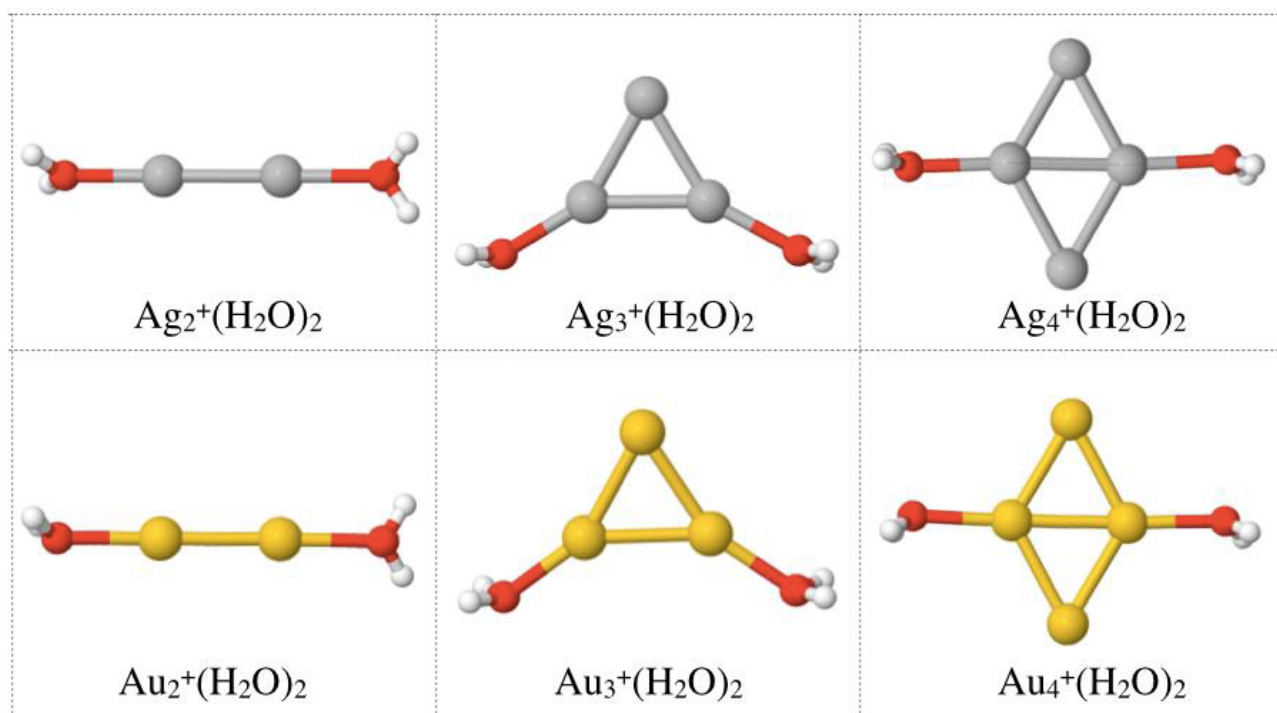
The metal atoms in the  $\text{Ag}_4^+$  cluster take a planar rhombus shape. The two water molecules coordinate to the two three-coordinate Ag sites, via the water oxygen atoms. Again, the hydrogen atoms of the water molecules are angled in opposite directions compared to one another. Both Ag-Ag-O angles are  $176.8^\circ$  (using the 3-coordinate Ag atoms), the O-Ag-Ag-O dihedral is  $164.4^\circ$ , and the H-O-O-H dihedral is  $-47.8^\circ$ .

The putative global minima for the 2 and 3 metal atom cationic gold clusters adopt the same

conformations as seen in the silver clusters. In the three atom gold cluster, the O-Au-Au-O dihedral angle is  $-0.7^\circ$ . The Au-Au-O angle is  $144.3^\circ$  for both. The Au-O bond length is slightly shorter at  $2.2\text{\AA}$  compared to  $2.25\text{\AA}$  in the silver cluster.

Like in the silver cluster, the metal atoms in the  $\text{Au}_4^+$  cluster take a planar rhombus shape. Both of the water molecules are bound to the two 3-coordinate water gold atoms. Interestingly, in this cationic gold cluster, the hydrogen atoms of the two water molecules are both angled in the same direction, towards the 'bottom' of Figure 4.3. Both Au-Au-O angles are  $178.2^\circ$  (using the 3-coordinate Ag atoms), the O-Au-Au-O dihedral is  $3.0^\circ$ , and the H-O-O-H dihedral is  $-4.8^\circ$ .

Higher coordinate metal sites are generally favoured in the putative global minima as they are better able to stabilise the positive charge of the cationic cluster. This allows for a larger partial charge on these sites, and thus a stronger interaction with the oxygen lone-pair sites.



**Figure 4.3**  $\omega\text{B97X-D}$  geometries for  $\text{Ag}_n^+(\text{H}_2\text{O})_2$  and  $\text{Au}_n^+(\text{H}_2\text{O})_2$  for  $n = 2 - 4$

Figure 4.4 shows the putative global minima for the singly hydrated mixed metal AgAu NAs with 1 to 4 metal atoms.

The water molecule coordinates to the silver atom in  $[\text{AgAu}]^+(\text{H}_2\text{O})$  via its oxygen atom.

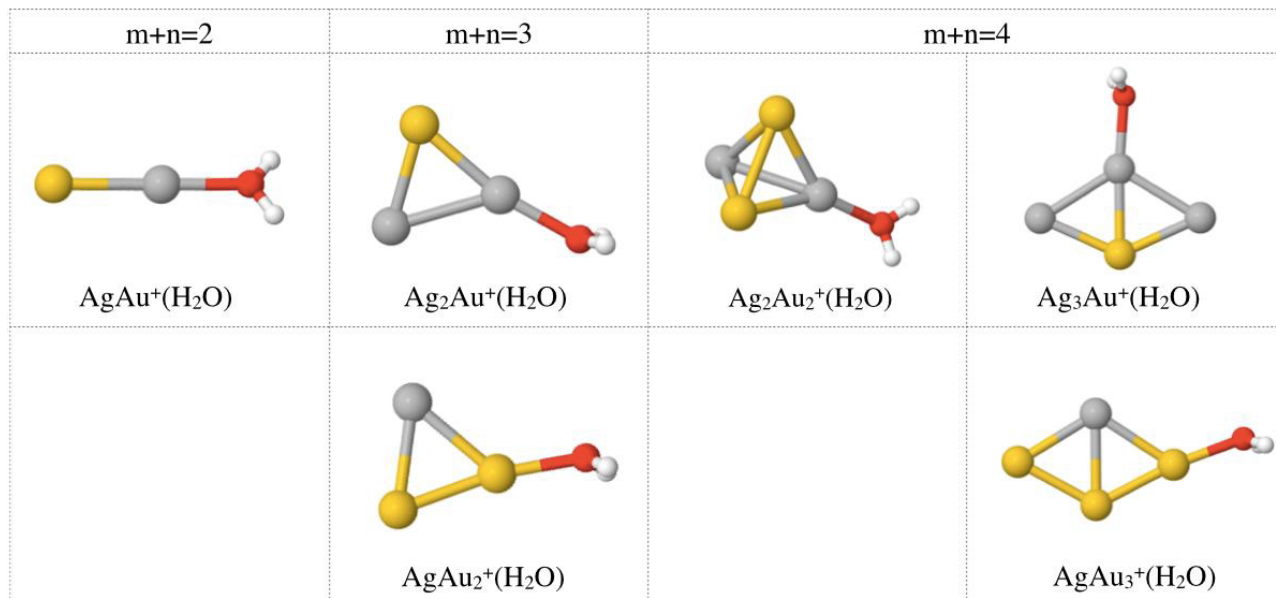
In  $[\text{Ag}_2\text{Au}]^+(\text{H}_2\text{O})$  the water molecule coordinates to one of the silver sites on the triangular

metal cluster. The hydrogen atoms of the water molecule are angled due to the interaction occurring primarily through one of the oxygen lone pairs.

$[\text{AgAu}_2]^+(\text{H}_2\text{O})$  also adopts a planar structure; however, in this cluster, the water molecule coordinates to a gold atom rather than silver. The Au-Au-O angle in this cluster is nearly planar at  $170^\circ$ , compared to the Ag-Ag-O angle in the previous cluster. Additionally, the Ag-Au-O angle in this cluster is very similar to the Ag-Ag-O angle in the majority-silver cluster ( $\text{Ag-Au-O}\angle = 130^\circ$  compared to  $\text{Ag-Ag-O}\angle = 134.2^\circ$ ).

In the three clusters with  $n + m = 4$ ,  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})$  and  $[\text{AgAu}_3]^+(\text{H}_2\text{O})$  have a planar rhombus conformation, with  $[\text{Ag}_2\text{Au}_2]^+(\text{H}_2\text{O})$  adopting a tetrahedral shape. In the  $\text{Ag}_3\text{Au}$  cluster, the water molecule binds to the 3-coordinate silver atom site. This contrasts starkly with the  $\text{AgAu}_3$  cluster, where the most preferential binding site is one of the two-coordinate gold atoms. The water molecule binds to one of the silver sites in the  $\text{Ag}_2\text{Au}_2$  cluster.

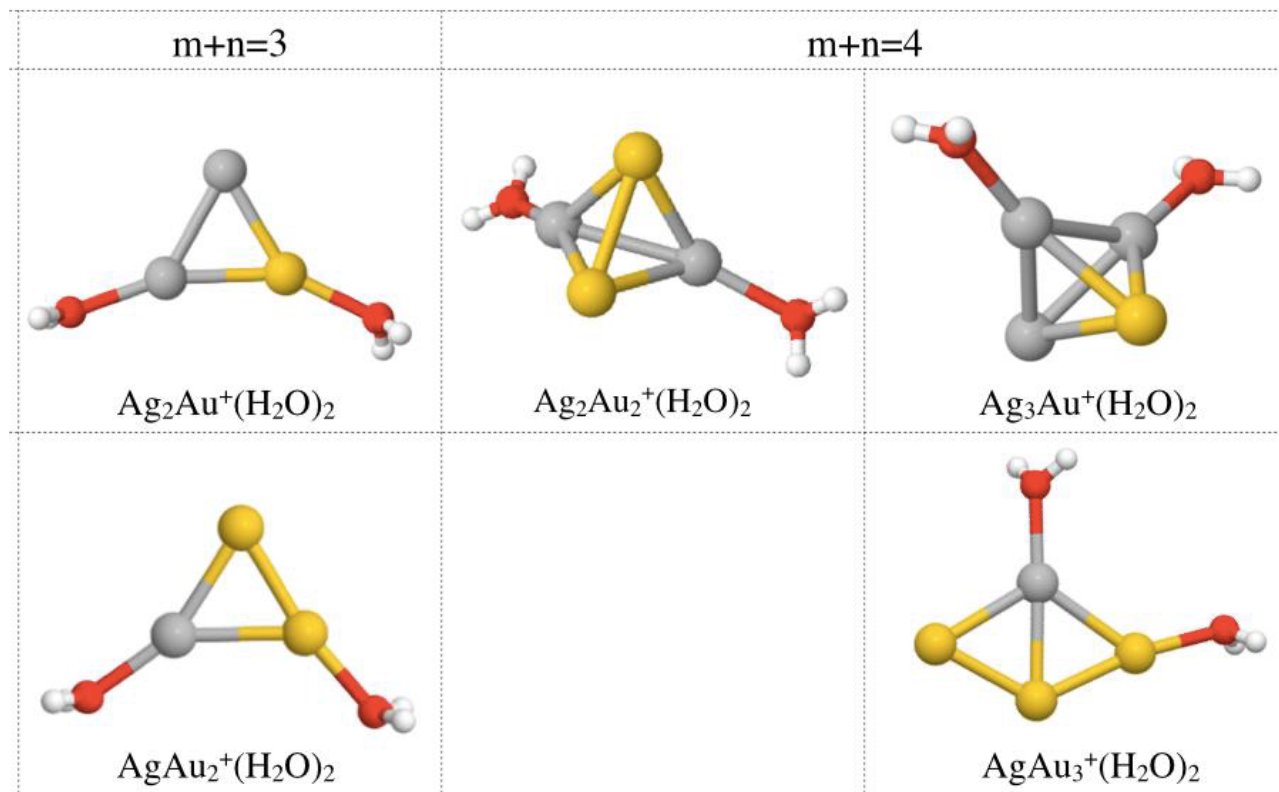
It can be seen from the putative global minima presented in Figure 4.4 that silver sites are the preferred first binding site for water molecules in all clusters containing 50% or more silver. This preference is attributed to the higher positive charge being stabilised on the Ag atom, as observed in previous studies.<sup>92</sup> The likely reason for this is the significantly higher Pauling electronegativity of gold compared to silver (2.5 vs. 1.9).<sup>92</sup>



**Figure 4.4**  $\omega$ B97X-D geometries for  $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})$  for  $n + m = 2 - 4$

Finally, the doubly hydrated NA clusters are presented in Figure 4.5. The two metal atom clusters are not represented here as there were no stable global minima found. In all cases, one or both of the water molecules disassociated. The metal alloy for all of the clusters takes the same conformation as in the singly hydrated clusters, except the putative global minimum for the doubly hydrated  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_2$  cluster adopts a tetrahedral structure with both the water molecules binding to AG sites. This shows us that the second water molecule is stabilising this tetrahedral conformation over the planar rhombus conformation that is seen in the singly hydrated system.

In each of the planar doubly hydrated clusters, the two water molecules are binding to one Ag and one Au site. In the tetrahedral putative global minimum clusters, the water molecules both bind to Ag atom sites, indicating that the positive charges are more stabilised on these sites. I will discuss this further in the Binding section below.



**Figure 4.5**  $\omega$ B97X-D geometries for  $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_2$  for  $n + m = 3 - 4$

### In-depth examination of $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_2$ minima

Shayeghi *et al.* in their 2014 paper, reported finding four minima for the  $[\text{Ag}_3\text{Au}]^+$  system.<sup>92</sup> I reproduce all the metal cluster putative minima and show that they are all stable in the presence of two bound water molecules.

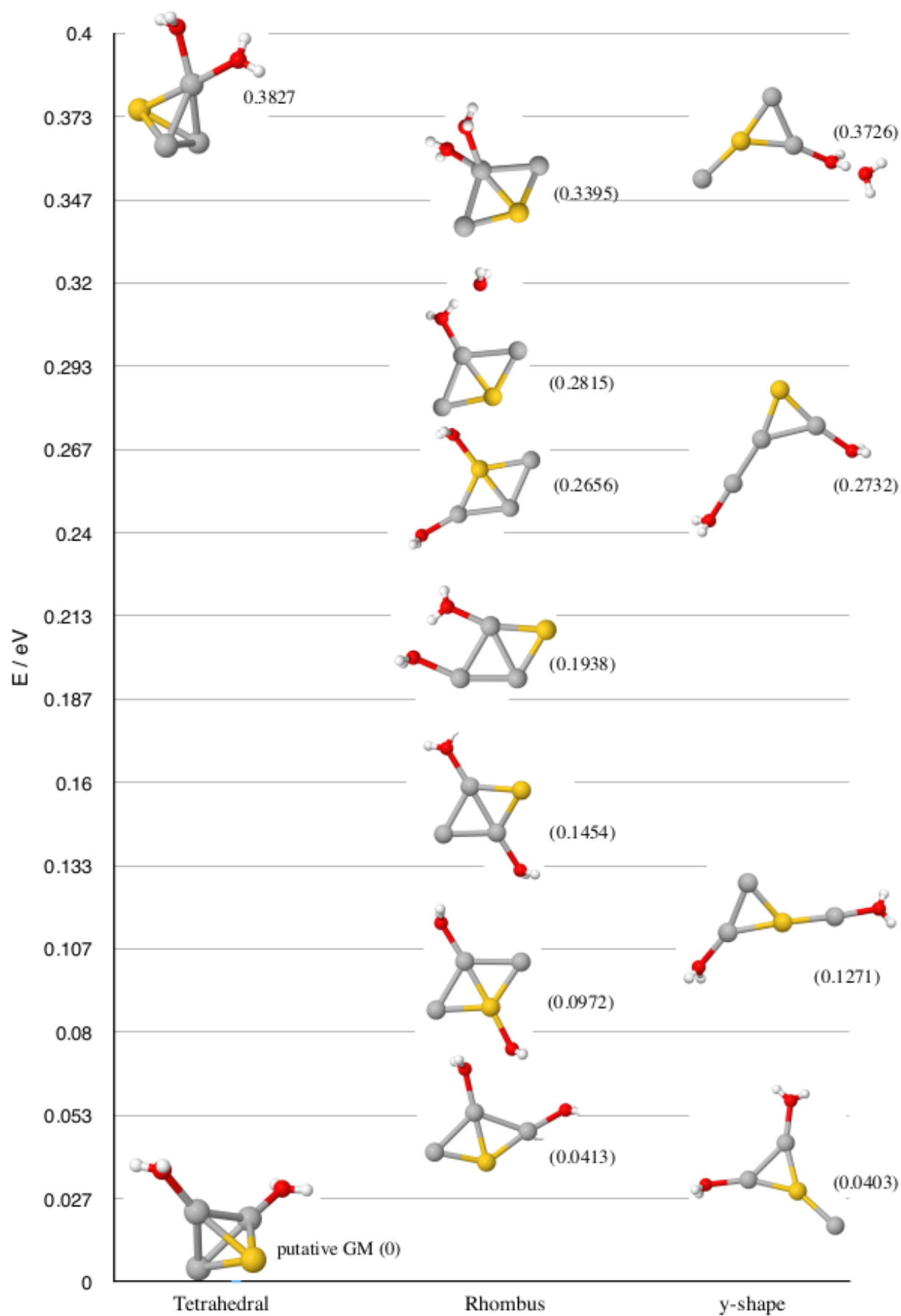
Figure 4.6 shows the thirteen, lowest energy, distinct minima found for the  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_2$  system, within 0.4 eV of the putative global minimum. These minima were identified as unique by the GIGA genetic algorithm.<sup>96</sup> The energies in eV above the putative global minimum energy are given in brackets.

We find the same four metal cluster structures as reported by Shayeghi *et al.*<sup>92</sup> The putative global minimum in this doubly hydrated system is the tetrahedral structure (labelled as Iso-IV in the Shayeghi paper), with the two water molecules bound to different Ag sites. The next lowest energy tetrahedral isomer is 0.3827 eV higher in energy than the putative global minimum. This tetrahedron has the two water molecules bound to the same Ag site. This

minimum is the 13<sup>th</sup> lowest energy minimum found by the GA.

I find the 'Y' shaped minimum (Iso-II) as the second lowest energy structure. In this low-lying minimum, (0.0403 eV above the putative GM), the two water molecules are bound to the two adjacent Ag atoms, this is likely because these two sites are relatively highly positively charged (both carry a partial charge of 0.13 e). The next lowest 'Y' shaped minimum is at 0.1271 eV above the GM, and has the two water molecules coordinated to the two different Ag sites, rather than the two equivalent ones. The highest energy 'Y' minimum shown here lies at 0.3726 eV above the putative GM. This minimum has one water molecule bound to an Ag site, and the other water molecule forming a hydrogen bond to it. The final Y type minimum is an isomer of the others, with the third silver atom bound to another silver atom rather than the gold atom. The water molecules are bound to the 1 and two coordinate silver atoms. This arrangement of the metal atoms is not reported by Shayeghi, and may lie significantly higher on the energy landscape when not stabilised by water molecules.

Finally there are the two variants of the planar, rhombus metal cluster, Iso-I which has the gold atom in a 2-coordinate site and Iso-III which has it in a 3-coordinate site. The two lowest energy rhombus minima are type Iso-III, and have energies of 0.0413 and 0.0972 eV respectively. The lower energy minimum has water bound to the 3-coordinate Ag site, and one of the 2 coordinate Ag sites. The higher energy of the two has water molecules bound to the two 3-coordinate metal sites, one silver and one gold. The first Iso-I type minimum appears at 0.1454 eV above the putative GM, and has water bound to the two 3-coordinate Ag sites. Higher energy rhombus minima have both of the water molecules interacting with the same Ag site, or forming hydrogen bonds.



**Figure 4.6**  $\omega$ B97X-D structures and energies (relative to the putative global minimum) for the  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_2$  putative global minimum and other unique minima.

## Energies

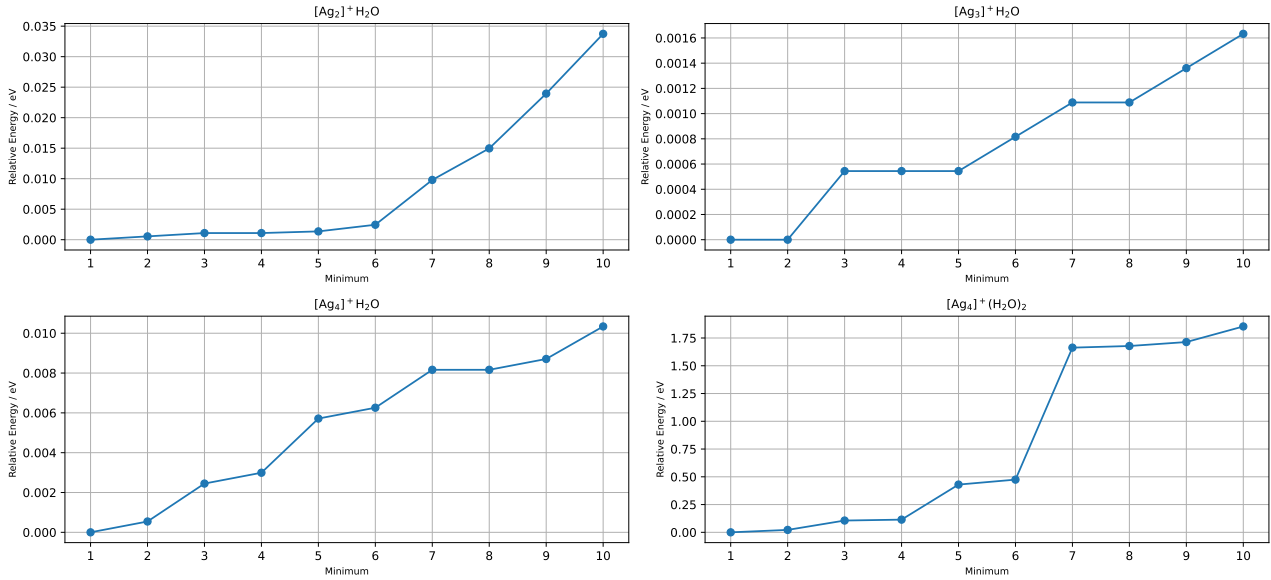
All energies reported in this section are calculated using the aug-cc-pVDZ/aug-cc-pVDZ-pp basis set and method outlined above.

The energies above the global minimum for the first ten minima for the clusters containing only silver and water are given in Figure 4.7. The 10 lowest  $\text{Ag}_{2-3}^+(\text{H}_2\text{O})$  clusters minima are within 0.035 eV of the global minimum, and represent minor reorientation of the water molecule, with all observed binding being to the same Ag atom for the global minima and the associated higher energy minima.

For  $\text{Ag}_4^+(\text{H}_2\text{O})_2$  the first 5 lowest energy minima have the water interacting with one of the most highly connected silver atoms (three coordinate), the higher energy minima have the water molecule interacting with the other silver atom site.

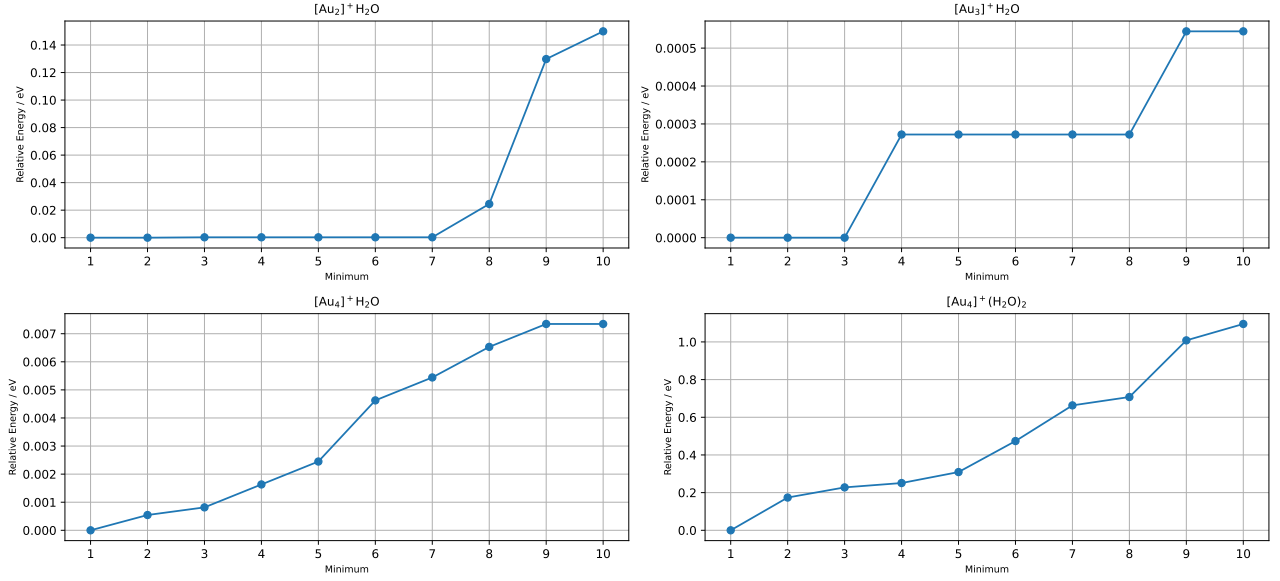
For  $\text{Ag}_4^+(\text{H}_2\text{O})_2$  minima 1-6 have the two water molecules interacting with the two 3-coordinate Ag atoms, on 'opposite' sides of the silver cluster. Minima 7-10 have one water molecule bound to a three coordinate Ag site and the other bound to one of the 2 coordinate Ag sites, it should be noted that there is no indication of hydrogen bonding between the water molecules. There is a difference in energy of 1.79 eV between the putative global minimum and the tenth lowest energy minimum, and a difference of 1.1 eV between minima 6 and 7, where the rearrangement of sites occurs.





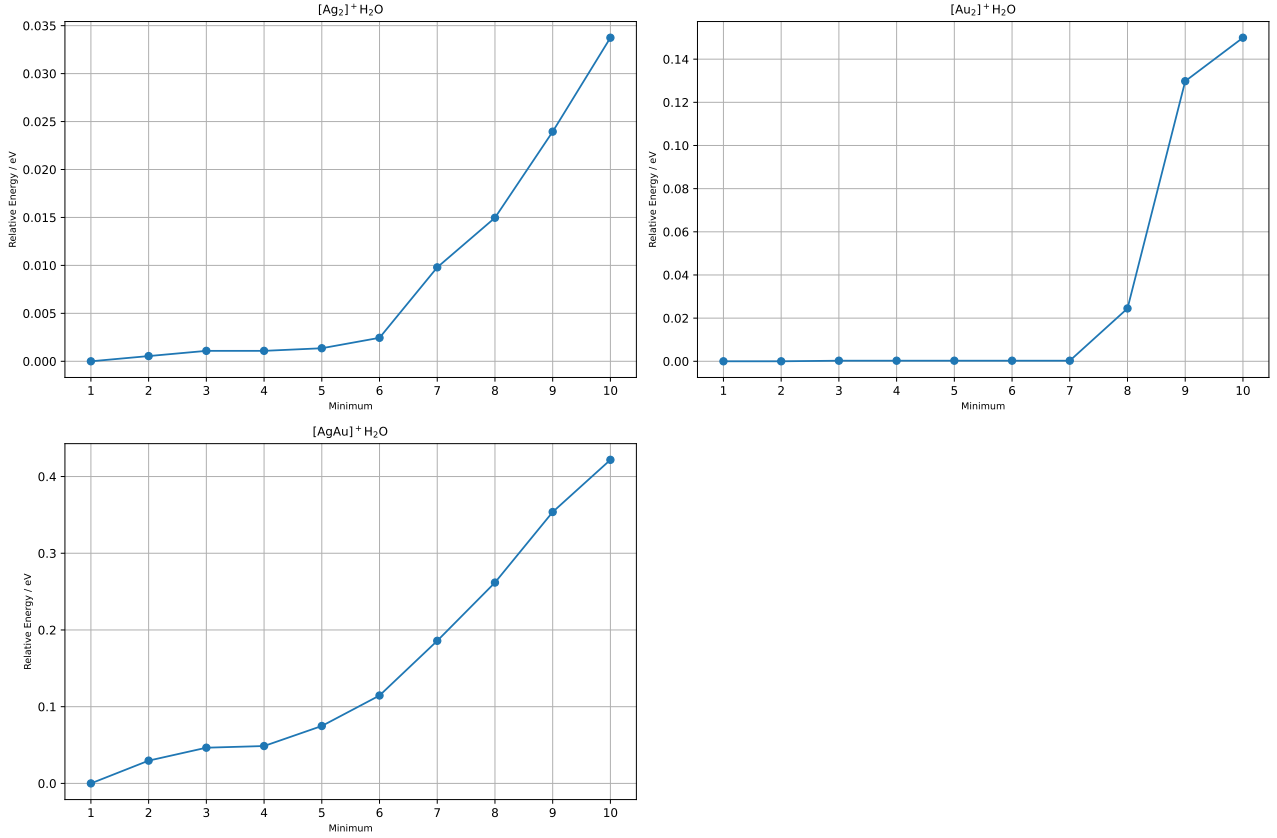
**Figure 4.7** Energies relative to the global minimum of the lowest 10 minima for  $\text{Ag}_m^+(\text{H}_2\text{O})_2$  for  $m = 2 - 4$  and  $\text{Ag}_4^+(\text{H}_2\text{O})_2$

The energies above the global minima for the hydrated gold clusters for the lowest lying ten minima are shown in Figure 4.8. The  $\text{Au}_2$  cluster displays four distinct minima in the lowest ten, with a difference of 0.14 eV between 1 and 10. The singly hydrated  $\text{Au}_3$  cation cluster shows three conformations which are close in energy amongst the lowest ten conformations 0.0015 eV. These are minor reorientation of the water molecule. The singly and doubly hydrated  $\text{Au}_4$  clusters both display planar conformations of gold atoms, with the tetrahedral gold clusters seen in the neutral clusters being observed at higher energies.



**Figure 4.8** Energies of the lowest 10 minima for  $\text{Au}_n^+(\text{H}_2\text{O})_2$  for  $n = 2 - 4$  and  $\text{Au}_4^+(\text{H}_2\text{O})_2$

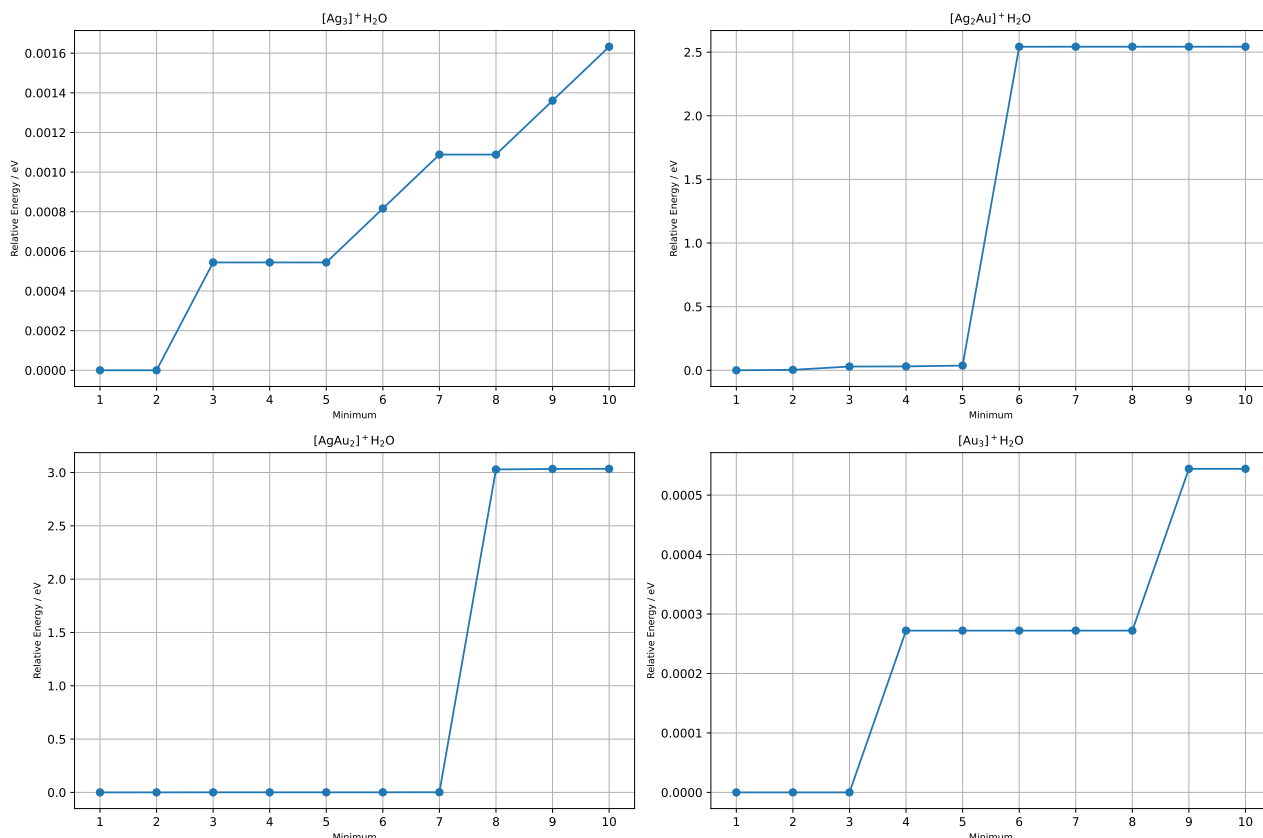
The energies above the global minima for the first ten minima found for the  $[\text{Ag}_m\text{Au}_n]^+\text{H}_2\text{O}$  clusters for  $m + n = 2$  are shown in Figure 4.9. We observe a number of orientations of the water molecule in  $[\text{AgAu}]_n^+(\text{H}_2\text{O})$ . There is a difference of 0.4 eV observed between the global minimum and the 10<sup>th</sup> lowest energy minimum. The first 4 minima are rotations of the water molecule, with the water molecule bound to the, with minimum 5 binding between the water molecule and the Ag and the Au atoms. For minima 9 and 10, the water molecule interacts primarily with the Ag atom.



**Figure 4.9** Energies of the lowest 10 minima for  $[\text{Ag}_m\text{Au}_n]^+(\text{H}_2\text{O})$  for  $m + n = 2$

Figure 4.10 shows the energies above the global minima for the first ten minima for the  $[\text{Ag}_m\text{Au}_n]^+\text{H}_2\text{O}$  clusters for  $m + n = 3$ . The metal cluster adopts a triangular structure arrangement in all of these 3 metal-atom clusters. For the  $[\text{Ag}_2\text{Au}]^+\text{H}_2\text{O}$  clusters, the first 6 minima correspond to clusters in which the water molecule interacts with an Ag atom, with the H-H axis of the water molecule almost perpendicular to that of the Ag-Ag metal bond. In minima 6-10, the water is shifted to interact with the Au atom instead. There is an increase of 2.5 eV between minima 5 and 6.

For the  $[\text{AgAu}_2]^+\text{H}_2\text{O}$  clusters, in minima 1-7 the water molecule is coordinated to one of the Au atoms, for the remaining three minima the water molecule coordinates to the Ag atom. There is a difference of 3.05 eV between minima 7 and 8.

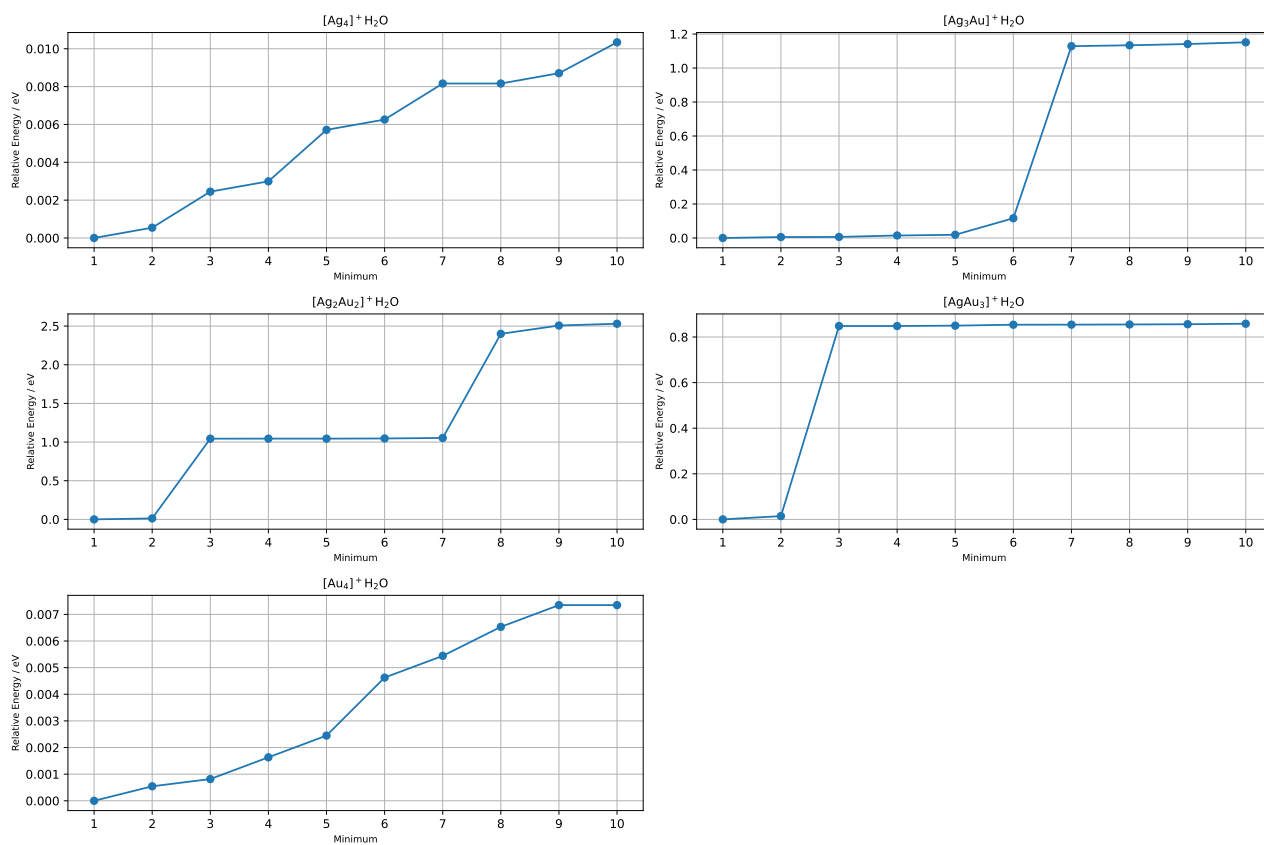


**Figure 4.10** Energies of the lowest 10 minima for  $[\text{Ag}_m\text{Au}_n]^+(\text{H}_2\text{O})$  for  $m + n = 3$

The energies above the global minima for the first ten minima for the  $[\text{Ag}_m\text{Au}_n]^+\text{H}_2\text{O}$  clusters for  $m + n = 4$  are given in Figure 4.11. All the global minima for the  $m + n = 4$  metal clusters adopt a planar structure, with the exception of the  $\text{Ag}_2\text{Au}_2$  cluster which adopts a tetrahedral structure. The first 5 minima for the  $\text{Ag}_3\text{Au}^+\text{H}_2\text{O}$  cluster have the water molecule coordinated to the 3-coordinate Ag atom. The 6<sup>th</sup> minimum shows the water molecule coordinated to one of the 2-coordinate Ag atoms, and minima 7-10 have the water molecule coordinated to the Au atom.

The  $\text{Ag}_2\text{Au}_2$  cluster putative global minimum and the lowest-energy 10 minima have a tetrahedral structure. In the Global minimum and the second lowest energy minimum, the water molecule is coordinated to one of the silver atoms, in minima 3-7 the water is coordinated to one of the gold atoms.

The  $\text{AgAu}_3$  cluster putative global minimum and the lowest-energy 10 minima have a planar structure.

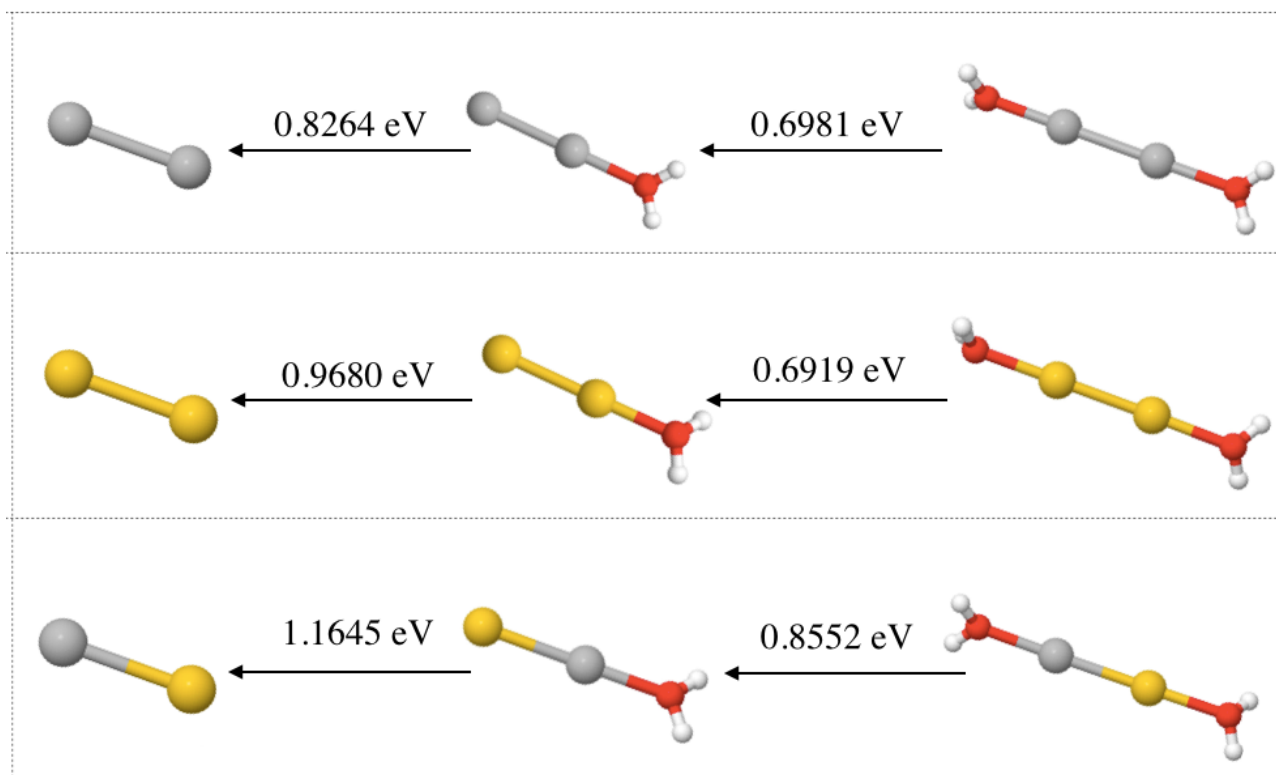


**Figure 4.11** Energies of the lowest 10 minima for  $[\text{Ag}_m\text{Au}_n]^+(\text{H}_2\text{O})$  for  $m + n = 4$

### 4.3.2 Binding

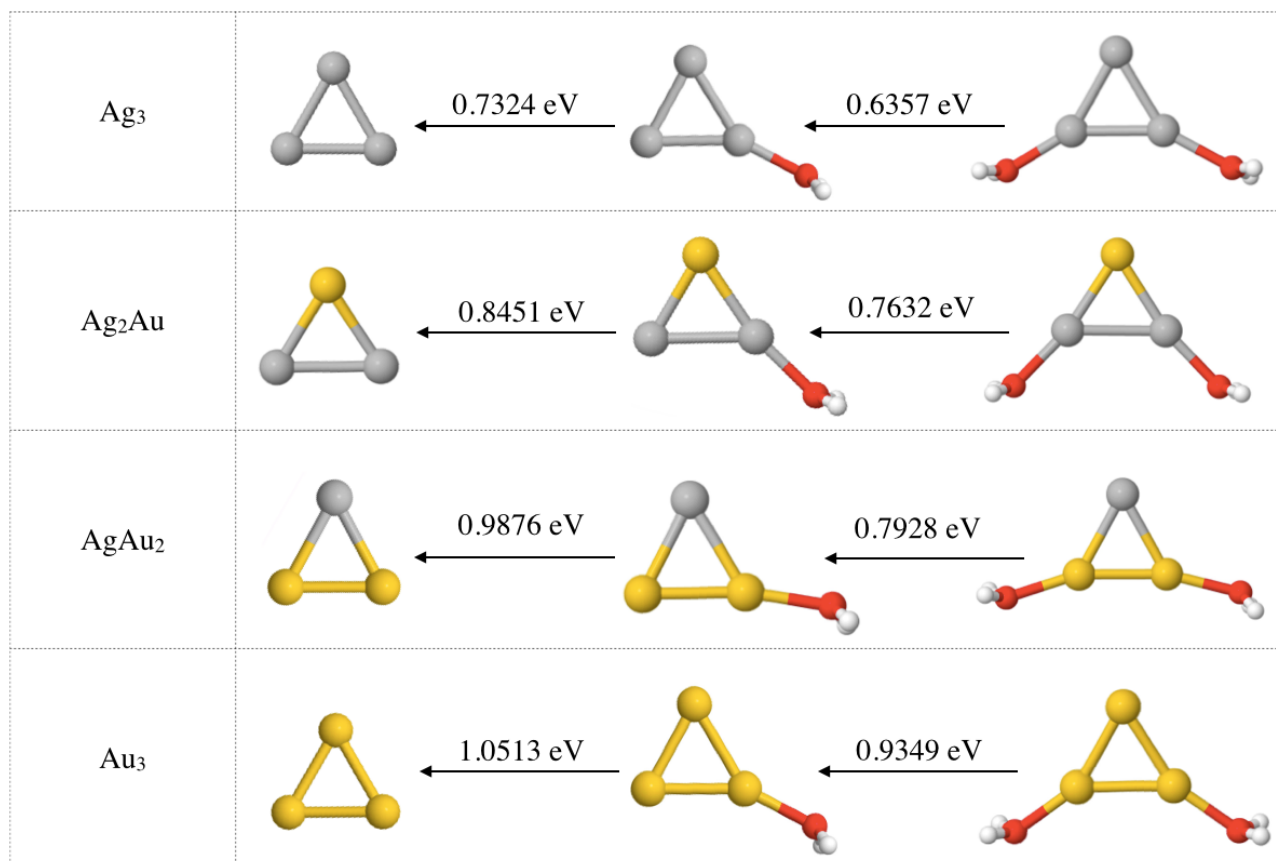
In this section I will examine the stepwise binding energies of the water molecules to the metal nanoalloy clusters. The binding, (or hydration), energies reported here are calculated according to Equation 4.1, and thus will be positive. The energies of the metal clusters and gas-phase water molecule were calculated using  $\omega$ B97X-D.

Figure 4.12 shows the sequential binding energies for the 2-metal atom clusters. For both the  $[\text{Ag}_2]^+$  and  $[\text{Au}_2]^+$  systems, the binding energy of the second water molecule is approximately equal at 0.6981 and 0.6819 eV respectively. The binding energy of the first water molecule, however, is 0.14 eV higher for the  $[\text{Au}_2]^+$  system, due to its higher polarizability. For the mixed  $[\text{AgAu}]^+$  system, the first hydration energy is higher than for either of the other two systems, due to the relatively more electronegative gold withdrawing electron density from the silver, leading to a greater electrostatic interaction with the water molecule.



**Figure 4.12**  $\omega$ B97X-D structures and sequential binding energies for the 2-metal  $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_x$  putative global minima.

Figure 4.13 shows the sequential binding energies for the 3-metal atom clusters. Much like in the 2-metal clusters, the binding energy for the second water molecule is less than for the first. Water molecules bond to the metal atom that makes up the majority of the cluster for preference, e.g.  $[\text{Ag}_2\text{Au}]^+(\text{H}_2\text{O})_2$  both the water molecules bind to the silver sites, and in the  $[\text{AgAu}_2]^+(\text{H}_2\text{O})_2$  cluster both water molecules bond to the gold sites. Hydration energies increase as the percentage of gold atoms in the metal cluster increases. These same trends continue for the 4 metal clusters.



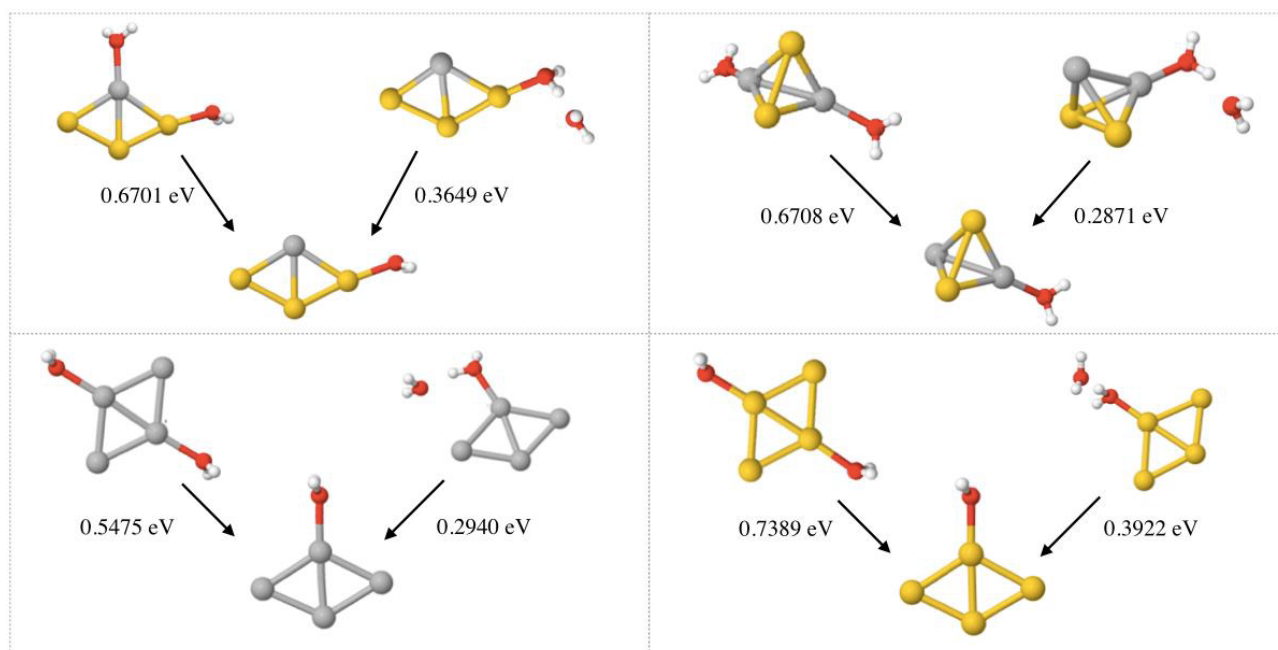
**Figure 4.13**  $\omega\text{B97X-D}$  structures and sequential binding energies for the 3-metal  $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_x$  putative global minima.

### Effects of metal cluster composition on hydrogen bonding

In Figure 4.14 I show the effect that the nanoalloy composition has on hydrogen bonding. To accomplish this, we take compare the binding energy of the  $[\text{M}_4]^+(\text{H}_2\text{O})_2$  putative global minimum, with that of the lowest energy minimum exhibiting hydrogen bonded water molecules, this approximates the loss of the second water and the dehydration to the  $[\text{M}_4]^+(\text{H}_2\text{O})$  minimum. Note that I have excluded the  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_x$  system, as this is the only cluster which exhibits

a structural change upon loss of water.

As noted previously, the binding energy increases as the percentage of gold in the cluster increases. There is a difference of 0.1914 eV between the water-metal binding energy in the Ag-only cluster and the Au-only cluster. It is interesting that this effect is still present for the hydrogen bonded water molecules, with a difference of 0.1051 eV between the Ag cluster and the Au cluster. The difference in binding energy for the hydrogen bonded water molecule is less pronounced, likely because the hydrogen bonded water molecule is further from the metal atoms. Gold has a higher Pauling electronegativity than silver and thus draws more charge from the oxygen atom of the directly bound water molecule.<sup>92</sup> This effect likely leads to a greater relative positive charge on the hydrogen atoms, and thus a greater electrostatic interaction with the second water molecule, leading to a higher binding energy. It is worth noting that the  $\omega$ B97X-D functional includes explicit dispersion in the form of Grimme's D2 dispersion model, which is why it was chosen. Thus dispersion is accounted for in this effect.



**Figure 4.14**  $\omega$ B97X-D structures and sequential binding energies for the 4-metal atom  $[\text{Ag}_n\text{Au}_m]^+(\text{H}_2\text{O})_2$  putative global minima, and a the lowest energy minimum with two water molecules seen to be exhibiting hydrogen bonding.



## Change in structural ordering due to water stabilisation effect

As promised above, Figure 4.15 aims to clarify the stabilisation effect hydration has on the tetrahedral  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})$  system. This figure provides a clear visual representation of how hydration influences the stability and conformation of the metal clusters.

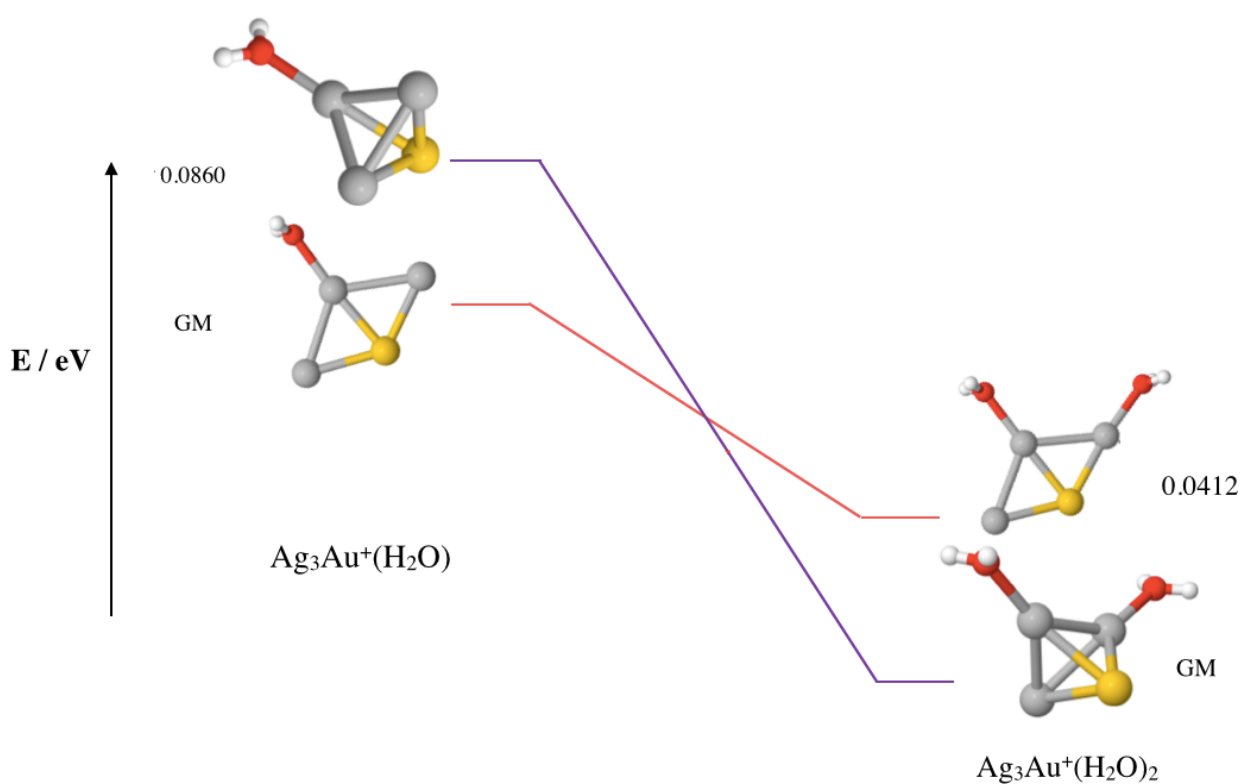
For the singly hydrated  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})$  system, the planar rhombus configuration of metal atoms is the putative global minimum, with the single water molecule bound to the 3-coordinate Ag site. The tetrahedral cluster, though stable, is found to be 0.086 eV above the global minimum, with the water molecule bound to one of the Ag sites. This small energy difference indicates the tetrahedral structure is less favourable but still significantly stable.

In the doubly hydrated system, the tetrahedral configuration becomes the putative global minimum, with the two water molecules both bound to Ag sites. The planar rhombus cluster remains a stable minimum but is now 0.0412 eV above the putative global minimum. This cluster features water molecules bound to both the 3-coordinate Ag site and one of the other Ag sites, showing a distinct binding pattern compared to the singly hydrated system.

The binding energy of the second water molecule is 0.6708 eV on the tetrahedral cluster, whereas it is only 0.5436 eV on the planar rhombus cluster. This significant difference in binding energy is the crucial factor leading to the rearrangement in the energetic ordering of the different metal geometries. The additional binding energy provided by the second water molecule enhances the stability of the tetrahedral structure over the planar one.

To understand this phenomenon, I calculated charges using the Bader method, as implemented within NWChem.<sup>80</sup> In the planar cluster, the three-coordinate Ag site has a partial charge of 0.16 e, while the two-coordinate site has a partial charge of 0.06 e. In contrast, in the tetrahedral cluster, all three Ag sites possess partial charges of 0.13 e. This result is consistent with the findings of Shayeghi *et al.*<sup>92</sup> The additional stabilisation provided by the second water molecule binding to an Ag site with a partial charge of 0.13 e, compared to one with 0.06 e, leads to the observed change in the structure of the most favourable minimum.

I did not observe this effect for any of the other clusters, with all of the other putative global minima keeping the same metal cluster conformation. This effect is only seen in the  $[\text{Ag}_n\text{Au}_m]^+$  clusters due to the combination of the closeness in energy of the tetrahedral and rhombus clusters, and the comparatively large charge difference between the two different Ag sites. These combined factors allow for the binding energy of the second water molecule to make the tetrahedron more favourable. This is an interesting effect that would be a good target to study experimentally using infrared photodissociation spectroscopy, or computationally at a higher level of theory.



**Figure 4.15**  $\omega\text{B97X-D}$  structures and relative energies for the planar and tetrahedral, singly and doubly hydrated,  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_x$  system.

## 4.4 Conclusions

My study on the structures of hydrated  $\text{Ag}_m\text{Au}_n^+$  clusters revealed several key points. The global minima and putative global minima geometries of these clusters were found using the GIGA algorithm with the aug-cc-pVDZ/aug-cc-pVDZ-pp basis set, then re-optimised using the  $\omega\text{B97X-D}$  functional. Importantly, re-optimisation with the  $\omega\text{B97X-D}$  functional did not result in any changes in overall geometry or energy ordering, dispute including explicit dispersion terms. The observed planar  $\text{C}_{2v}$  geometry for  $\text{Ag}^+(\text{H}_2\text{O})$  and Cs geometry for  $\text{Au}^+(\text{H}_2\text{O})$  were consistent with previously reported structures, indicating the robustness of these configurations. I also observe the same minima as previously reported experimentally and computationally for the 4-metal atom clusters.<sup>92</sup>

For  $\text{Ag}_n^+(\text{H}_2\text{O})$  and  $\text{Au}_n^+(\text{H}_2\text{O})$  clusters with  $n = 1-4$ , it was noted that the Ag-O interactions occur equally through both oxygen lone-pair sites, whereas the Au-O interactions occur through only one lone-pair site. This interaction pattern led to the maintenance of the planar  $\text{C}_{2v}$  geometry for  $\text{Ag}^+(\text{H}_2\text{O})$  and the Cs geometry for  $\text{Au}^+(\text{H}_2\text{O})$  across all monometallic clusters. Interestingly, the typically tetrahedral  $\text{Au}_4^+$  cluster was not observed as the minimum structure when stabilised by water, although it existed as a meta-stable state.

In doubly hydrated monometallic clusters, water molecules coordinated to the metal atoms via oxygen atoms, with the hydrogen atoms staggered to maximise the distance between them. Higher coordinate metal sites were favoured in the putative global minima due to their ability to better stabilise the positive charge of the cationic cluster, leading to stronger interactions with oxygen lone-pair sites.

The binding energies of water molecules to the metal clusters were calculated, showing a preference for silver sites in mixed metal clusters containing 50% or more silver. This preference is attributed to the higher positive charge on the silver atoms. Sequential binding energy calculations for 2-metal, 3-metal and 4-metal atom clusters show that hydration energies increased with a higher percentage of gold atoms in the cluster. This effect is due to the higher electronegativity of gold compared to silver, this leads to higher electrostatic interactions between

the metal and the water molecule.

A significant change in the structural ordering of the  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_2$  cluster was observed upon hydration, with the tetrahedral structure becoming more stable than the planar rhombus configuration. The binding energy of the second water molecule was higher on the tetrahedral cluster compared to the planar rhombus cluster, driving the rearrangement in energetic ordering. This structural change was attributed to the partial charges on the silver sites, with the second water molecule providing additional stabilisation.

The effect of hydration on structural ordering was unique to the  $[\text{Ag}_3\text{Au}]^+$  system due to the combination of closeness in energy of the tetrahedral and rhombus clusters and the comparatively large charge difference between the two different Ag sites. This phenomenon was not observed in other clusters, and is distinct to the  $[\text{Ag}_3\text{Au}]^+$  system in my work.

These findings suggest that further experimental studies using infrared photodissociation spectroscopy or computational studies at a higher level of theory could provide deeper insights into the effects of hydration on metal cluster structures. Understanding the binding energies and structural changes due to hydration can offer valuable information about the stability and reactivity of metal nanoalloy clusters in various chemical environments.

# Chapter Five

## BMPGA and Thresholding for Finite Clusters

### 5.1 Methodology

### 5.2 Overview

I wrote the bulk of BMPGA at the Max-Planck Institute for Solid State Systems Research in Stuttgart during a collaboration with Professor Christian Schön funded through the HPC Europa 3 grant.

BMPGA consists of 8000 lines of Python (Version 3.6+) code and 200 lines of FORTRAN(2008) and 200 lines of C++(14). The genetic logic, program flow control, and networking/distribution of work are implemented in python. The built-in empirical potentials and some of the more computationally expensive mutation moves outlined below. I am aware that python code often executes slowly when compared to FORTRAN or C++, however python code is significantly faster to develop/write. Python also has an advantage in that adding new modules/packages to an existing codebase does not require linking or recompilation, so adding new features/potentials/extensions is easier. I made the decision deliberately not to over-optimize my code while I was writing it, as the primary use was envisioned as being in global optimisation using expensive potentials, and thus a relatively small proportion of time would be spent executing the code written in python. Given that I had limited time to develop

the code, and the trade-offs outlined above, I made the decision to write the core of BMPGA in Python3 to maximise development speed. I wrote the empirical potentials, and line-search in fairly modern C++(14) and FORTRAN(2008) as writing these potentials was a relatively small investment of time compared to writing the rest of the code, and would benefit from the additional speed of computation. These empirical potentials were used extensively by me during testing, so the test-driven development cycle was improved by making this decision.

In my opinion, there is a lot of badly written scientific code in the world. Having spent a large amount of time trying to read, understand, and extend code of this nature, I tried not to produce more of it during this work. I wanted to create high quality code inline with modern software engineering principles when I created BMPGA. In order to do this I implemented many of the python coding standards outlined in PEP8, along with type hinting as outlined in PEP484.<sup>99–101</sup> I developed the codebase using Test Driven Development principles, with > 95% unit test coverage, and a significant level of integration test coverage. I used a modern Object Oriented programming approach when developing in python, with the idea that this would make it easier for others to develop new or extended functionality in the future.

It should be noted that in this Methods section, when I say "object", I am referring to the concept of an object in the sense of object oriented programming (OOP), that being a memory construct that contains both data and methods/functions. Specifically I am referring to python objects. Describing OOP goes beyond the scope of this thesis.<sup>102</sup>

### 5.2.1 BMPGA

#### Overview

Genetic Algorithms (GAs) are another method employed in global optimisation. GAs are one instance of the larger class of biologically-inspired optimisation methods.<sup>94,103,104</sup> Biologically inspired methods employ the same strategies for optimisation which are seen in nature. Inspiration has been drawn from a large number of different sources seen in nature. Some well known examples include; Artificial Bee Colonies (ABCs), Particle Swarm Optimisation (PSO), Genetic Algorithms, Evolutionary Algorithms etc.<sup>95,105,106</sup>

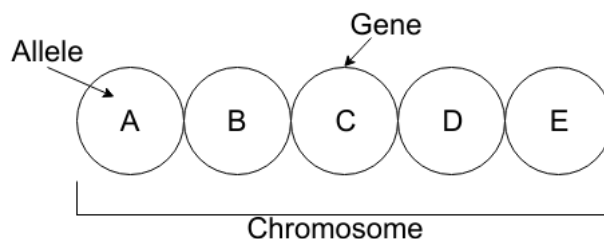
In this work I detail the development of a Genetic Algorithm we are calling the Birmingham Molecular Pool Genetic Algorithm (BMPGA). This is a parallel pool-GA based loosely on the BPGA method previously published by the group.<sup>94–96</sup> The GA implements a Lamarckian evolutionary scheme by default, and thus a local optimisation is performed after each crossover/mutation step, before the child is added to the pool.

In a pool GA, rather than performing genetic operations on the whole of a static population before beginning a new generation, one simply picks members from a non-fixed population and performs crossover and mutation operations, before adding the newly formed child to the pool. In our implementation the pool is periodically culled of the least fit individuals in order to increase the average genomic fitness. Periodically, (by default, every time the pool reaches the user-defined maximum pool size), the size of the pool is reduced down to the defined minimum pool size. How this reduction in pool-size is accomplished is able to be defined by the user, but the default behaviour, (which is used for all runs in this thesis), is to remove the highest energy minima in order until the pool is of the desired size. These minima are removed from the current pool, but are still stored within the database of all minima found during the run. If these removed minima are found again, then they will be re-added to the pool. Other options for culling minima are implemented within the code, such as stochastic, and roulette wheel removal methods.

## Genetic encoding

Genetic algorithms encode information about the system to be optimised using a series of artificial gene-analogues which are collectively referred to as an artificial genome. These genes must be capable of encoding the properties of the system. In early work on evolutionarily inspired algorithms a very simple method for encoding the genome was used; a simple binary string. This simple method had the advantages of being cheap in terms of computation and memory which made it attractive for researchers using early computer systems.<sup>103–106</sup> The crossover and mutation schemes employed on these genomes are also inexpensive, if somewhat crude by modern standards.<sup>1,103</sup> In the case of chemical systems, this encoding must include the atomic and molecular positions, as well as potentially the electronic state of the system.

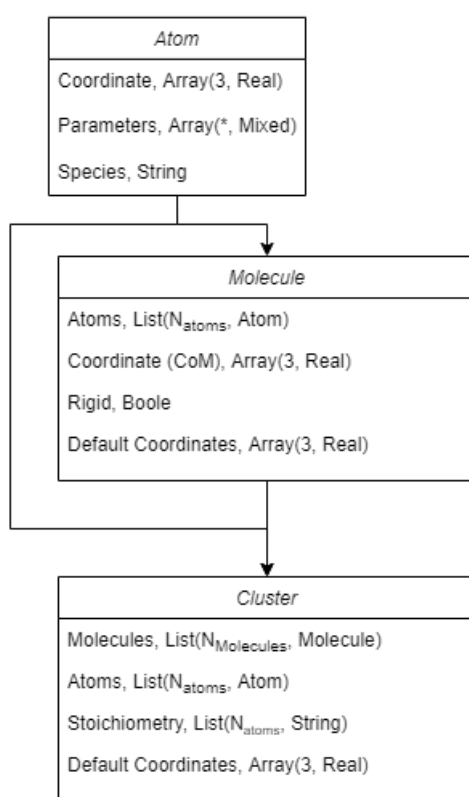
The atomic positions are crucial as they define the geometry and structure of the molecule, which are essential for understanding and predicting chemical properties and reactivity. The electronic state, including information about the distribution of electrons and energy levels, is equally important for accurately representing the system's behaviour under various conditions. Throughout this thesis, the genome encoding the state of a given configuration  $i$  will be referred to as  $\mathbf{r}_i$ .



**Figure 5.1** Schematic representation of a GA genome

Within BMPGA the genome is encoded in a somewhat more complex way. Starting from the bottom, there is an Atom object. This is named for the broader meaning of the word Atomic (*Uncuttable*, Greek) rather than the more familiar chemical meaning. We are describing the smallest elements of our system, these are often atoms if investigating atomic or molecular clusters, but can also be more generic particles if investigating clusters of other types, e.g. colloidal clusters. Within BMPGA, Atom objects encode the coordinates and data on the species they represent (atom type, particle labels, interaction parameters etc.) Molecule objects are made up of one or more Atom objects. They provide methods to calculate the centre of mass of the molecule, as well as methods for rigid body translations and rotations. The cluster object provides a view onto the minima database. This is the object that tracks the calculated fitness and is passed around between the main GA process and the minimisers. Mutations and selections accept cluster objects as their inputs. There are methods implemented to calculate the centre of mass (CoM), transform and align clusters, generate new random clusters, and detect particle overlap.





**Figure 5.2** UML diagrams for Atom, Molecule and Cluster objects

## Crossover

Crossover is the name given to the process most analogous to sexual reproduction. It involves the mixing of genetic material from two or more individuals (parents) to form new individual or individuals (children). This process is central to evolutionary algorithms, as it allows for the combination of traits from different parents, potentially leading to offspring with improved performance or novel characteristics.<sup>107,108</sup>

Many different mating schemes have been conceptualised over the course of research into evolutionary computation, ranging from simple methods to more complex strategies. One of the simplest methods is bitwise crossover, where genomes encoded in binary are combined using operations such as XOR. This method is computationally inexpensive and straightforward to implement, making it popular in early research.<sup>109,110</sup>

As evolutionary computation has evolved, more elaborate crossover schemes have been developed, especially for applications in chemistry and other fields that require maintaining specific structural properties. One significant challenge in implementing crossover for chemical applications is maintaining the stoichiometry of the resulting offspring. Geometric crossover, which involves combining structural elements of molecules, can disrupt the balance of atomic or molecular species, leading to offspring that do not adhere to the desired stoichiometry. In BMPGA this is managed by grouping particle/atom objects together into molecule objects. Crossover is then performed on these molecule objects which makes the bookkeeping of stoichiometry significantly easier.<sup>104,111</sup>

Overall, crossover is a vital mechanism in evolutionary algorithms that fosters diversity and innovation within the population. It enables the recombination of genetic material, allowing for the creation of offspring that may inherit beneficial traits from multiple parents. This process is essential for the ongoing adaptation and improvement of solutions, driving the evolutionary search towards optimal or near-optimal solutions in complex problem spaces.<sup>112</sup>

## Selection of parents

Selection of parents is an extremely important part of the GA process. This is the primary location where the evolutionary pressure is applied to the population as by selecting an appropriate selection method and tuning the parameters of that selection method, one can favour the selection of parents with higher or lower average fitness, and thus the direction the overall average fitness of the population moves in and thus the balance of exploration to exploitation of the system as a whole. As selection is arguably the most important aspect of a GA, I have implemented a number of different methods of selection into BMPGA, to allow the user to choose the most appropriate selection method for the system of interest.

**Tournament Selection:** I implemented tournament selection within BMPGA. In tournament selection a portion of the population is selected randomly, this subsection of the population then compete against one another and the fittest member goes on to mate. The only heuristic that needs to be controlled in Tournament Selection is the number of parents selected to take part in the tournament, i.e. the tournament size. A larger tournament size leads to increased evolutionary pressure, as it makes it more likely that the fittest members of the population will be competing in the tournament, and thus will lead to parents with higher average fitness. As the tournament size approaches 1, the average fitness of tournament winners approaches the average fitness of the population, thus a tournament size of 1 is analogous to stochastic selection, as the randomly selected parent will always be selected.

**Algorithm 1** Tournament Selection

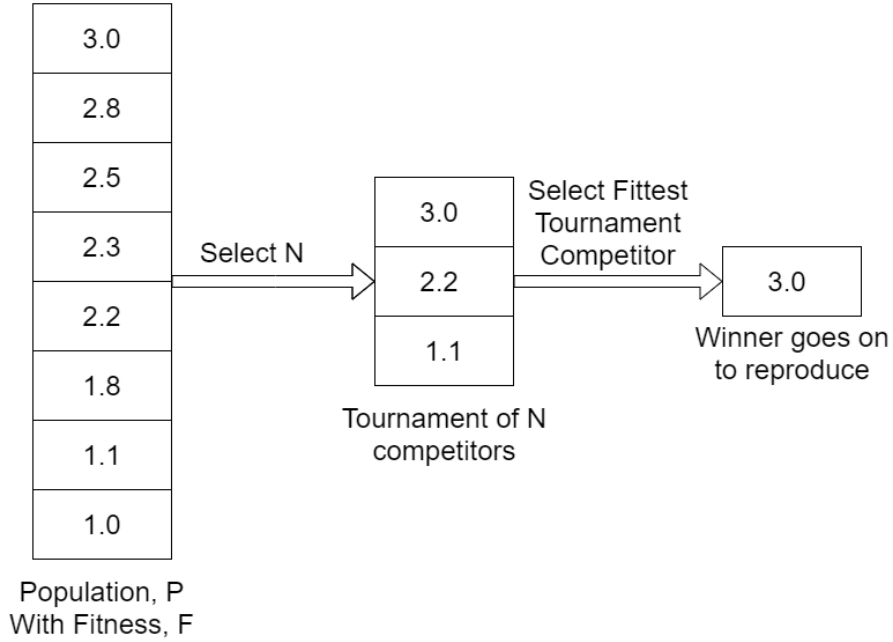
---

```

1: function TOURNAMENT( $k, P_{total}, S, F$ )           ▷ Where  $k_{max}$  is the tournament size,  $P$  is
   the total population,  $S$  is the selected parent,  $F_i$  is the fitness of a given member,  $i$ , of the
   population.
2:   while  $k < k_{max}$  do
3:     Select from  $P_{total}$  with probability =  $\mathbb{R} \cup [0, 1]$ 
4:   end while
        $S_1 = k_1$                                      ▷ Select the first member of the tournament
5:   for  $i = 2$ , to  $i = k_{total}$  do
6:      $S_2 = k_i$ 
7:     if  $F_S < F_C$  then State  $S_1 = S_2$ 
8:     end if
9:   end for
       return  $S_1$ 
10: end function Tournament

```

---



**Figure 5.3** Schematic example of a 3-way tournament for an 8 member population. The values inside the boxes indicate the fitness of each member of the population, with the highest fitness being 3.0 and the lowest being 1.0

**Stochastic Selection:** There are a great many methods employed to select parents for crossover. The simplest (and arguably least effective) method employed routinely is that of fully stochastic selection. In this way parents are selected randomly from the population without reference to the fitness of the individuals. This is akin to a Tournament selection Algorithm:1 with a fixed tournament size of 1. The advantage of Stochastic Selection is that it allows for rapid exploration of the wider energy landscape, as it allows the population to ignore barriers to crossing between funnels. The disadvantages of Stochastic Selection are that it removes one of the strongest sources of evolutionary pressure on the population as a whole as there is no advantage to a particular configuration having a high fitness in its ability to reproduce. In BMPGA we implement Stochastic Selection as a special case of Tournament Selection with the tournament size explicitly fixed at 1.

**Roulette Wheel Selection:** In Roulette Wheel Selection, parents are selected with probability explicitly proportionate to their fitness (hence the alternate name: Fitness Proportionate Selection). The probability  $p_i$  of selecting a given configuration,  $i$  is given by Equation 5.1 where  $f_i$  is the fitness of configuration  $i$ , and  $\mathbf{F}$  is the total population fitness (Equation 5.2). Roulette wheel selection is an elitist selection method that provides a high degree of evolutionary pressure, which promotes the propagation of advantageous traits through successive generations.

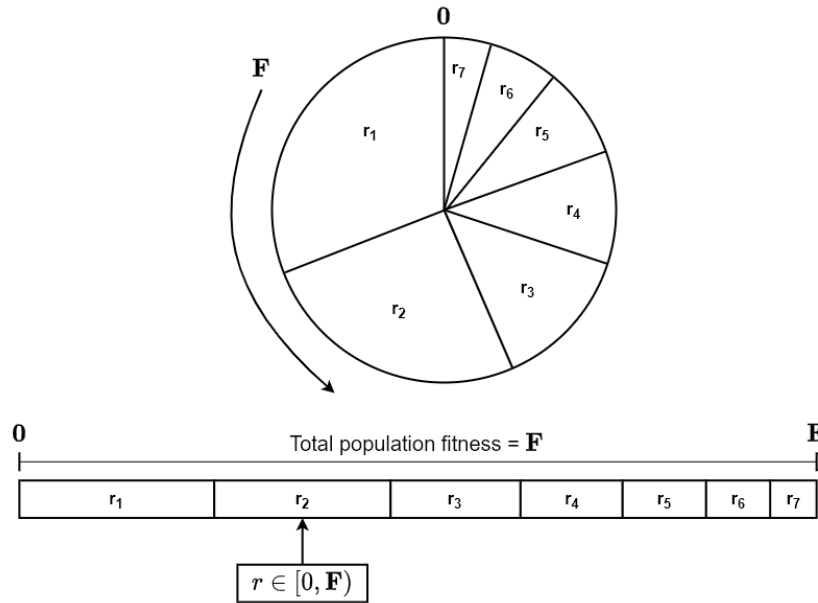
One of the key benefits of roulette wheel selection is that it strikes a balance between elitism and maintaining genetic diversity within the population. Elitism ensures that the best-performing individuals are more likely to reproduce, thereby driving the population towards optimal solutions more quickly. However, preserving genetic diversity is equally crucial for the long-term success of the evolutionary algorithm.

Removing diversity from the population can be detrimental for several reasons. This increases the risk of premature convergence, where the population may become stuck in a local optimum and fail to explore other potentially superior solutions. Genetic diversity introduces variability, allowing the algorithm to explore a broader search space and avoid getting trapped

in suboptimal regions. Without sufficient diversity, the population may lack the necessary variability to respond effectively to such changes, leading to stagnation or failure in dynamic or complex problem spaces. Therefore, roulette wheel selection's ability to balance elitism with maintaining genetic diversity makes it a powerful tool in evolutionary algorithms such as this, promoting both the exploitation of known good solutions and the exploration of new possibilities. This balance is essential for ensuring the algorithm's robustness, adaptability, and overall effectiveness in solving complex optimisation problems such as those in cluster optimisation.<sup>103,105</sup>

$$p_i = \frac{f_i}{F} \quad (5.1)$$

$$F = \sum f \quad (5.2)$$



**Figure 5.4** Schematic representation of Roulette Wheel Selection

**Boltzmann Selection:** In Boltzmann Selection the parents are selected with probability proportionate to their fitness, weighted by the Boltzmann formula, (Equation 5.4). This allows the incorporation of a theoretical temperature term and thus provides some physical relevance to the relative fitnesses of the pool members. The temperature can be varied during the course

**Algorithm 2** Roulette Wheel Selection

---

```

1: function ROULETTEWHEELSELECTION( $\mathbf{P}, \mathbf{f}, N$ )  ▷  $\mathbf{P}$  is the population vector (sorted
   by fitness),  $N$  = Current pool size,  $\mathbf{f}$  is the sorted population fitness vector
2:   for  $i = 0, i < N, i++$  do
3:      $p_i = \frac{f_i}{\sum \mathbf{f}}$ 
4:   end for
5:    $\bar{\mathbf{p}} = \frac{\mathbf{p}}{\sum \mathbf{p}}$   ▷ Normalise probability vector:  $\mathbf{p}$ 
6:    $r = \bigcup \in \mathbb{R} [0, 1)$   ▷ Draw uniform random number
    $p_{sum} = 0.0$ 
7:   for  $i = 0, 1 < N, i++$  do
8:      $p_{sum} += p_i$ 
9:     if  $r \leq p_{sum}$  then return  $P_i$ 
10:    end if
11:  end for
12: end function RouletteWheelSelection

```

---

of the optimisation to modify the acceptance rate of offspring, with an increased temperature reducing the selection pressure and thus allowing for more rapid exploration of the landscape, and reduced temperature encouraging exploitation of low energy regions. We can also use this theoretical temperature to analogue the process of simulated annealing. The  $T$  heuristic can be started high to allow for rapid exploration of the landscape, this can then be decreased over the course of several generations. This reduction in the 'Temperature' increases the evolutionary pressure, causing the population to have a strong pressure towards maximising fitness, with a sufficiently low temperature there will be no parents selected other than the fittest, and thus all the offspring will be children of iso-energetic minima. Within the context of the GA, when the population has converged on a particular energetic funnel, we can increase the temperature again and explore the wider landscape.

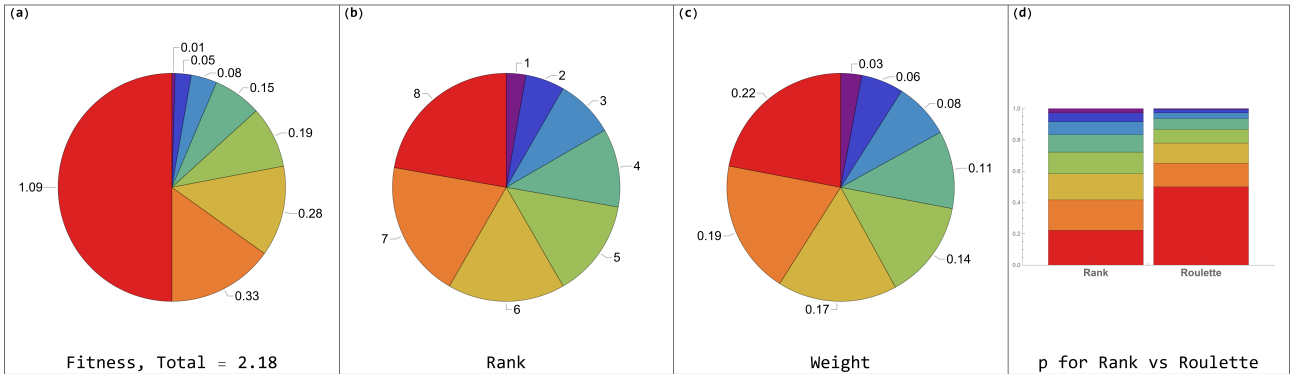
The probability of selecting any given configuration  $r_i$  with energy  $E$  is proportional to Equation 5.3, where  $T$  is the theoretical temperature,  $E$  is the fitness of the configuration in question and  $k_B$  is the Boltzmann constant. The probability of selection is then given by

Equation 5.4.

$$\exp\left(\frac{-E}{k_B T}\right) \quad (5.3)$$

$$p_i = \frac{\exp\left(\frac{-E_i}{k_B T}\right)}{\sum_j \exp\left(\frac{-E_j}{k_B T}\right)} \quad (5.4)$$

**Rank Selection:** In Rank Selection, the members of the population are ranked in order of fitness. The least fit member of the population is given a rank of 1, with members being given successive integer ranks in order of fitness above this. Parents are selected based on these ranks with probabilities given by Equation 5.5, where:  $p_i$  is the probability of selecting population member  $i$ ,  $\vec{R}$  is the vector of ranks  $R_{0,i}$ . This selection method applies less evolutionary pressure than exerted by Roulette wheel selection. Figure 5.5 (a-c) demonstrates this, with the fitness of members of a population being shown in (a), their assigned ranks being shown in (b), and their probabilities of selection being shown in (c). The difference in selective pressure can be seen most clearly in Figure 5.5 (d) where the computed probabilities of selection for the same example population are shown graphically for Rank and Roulette selection.



**Figure 5.5** Diagram showing (a) a sample population. (b) Ranks for the population in (a). (c) Those ranks normalised to weights. (d) How the probabilities compare for Rank and Roulette Wheel selection, given the same initial population.

$$p_i = \frac{R_i}{\sum \vec{R}} \quad (5.5)$$



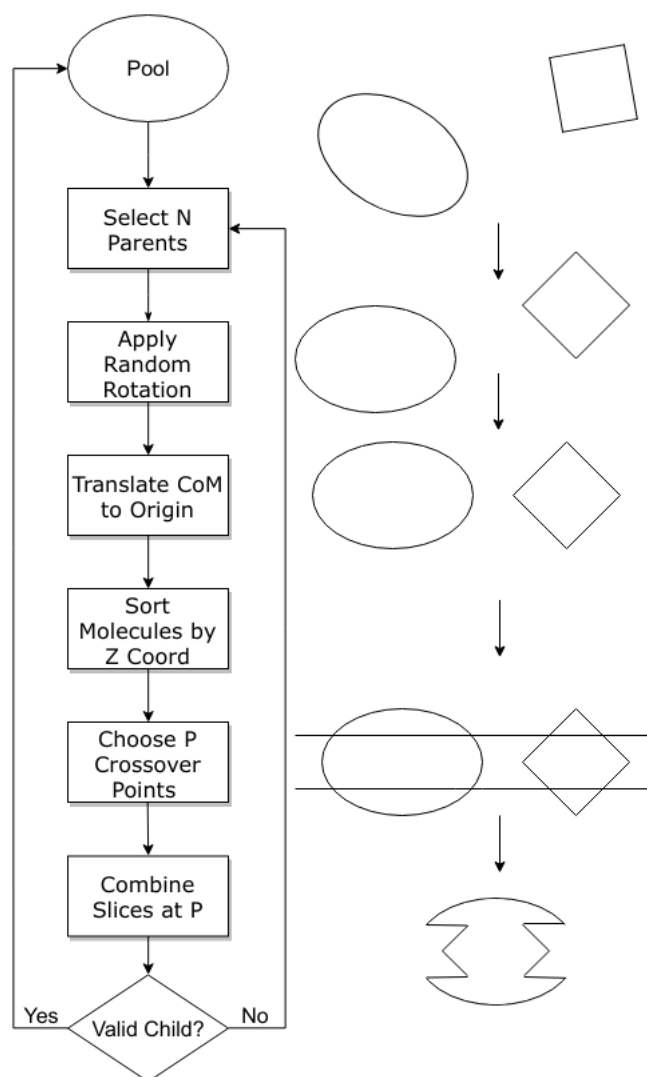
## Crossover

As touched on in at the top of this section (Section: 5.2.1 - Crossover), crossover is the process within genetic computation that is analogous to sexual reproduction. In biological evolution, sexual reproduction involves the combination of genetic material from two parents to produce offspring that inherit a mixture of their parents' traits. This genetic mixing, achieved through the process of meiosis and fertilisation in biology, enhances genetic diversity within the population. In the context of genetic algorithms, this concept is mirrored through crossover techniques where binary strings or other representations of individual solutions (akin to genomes) from two selected 'parent' trail solutions are combined. This recombination results in new 'offspring' solutions, potentially incorporating and propagating beneficial traits from each parent, thereby driving the evolutionary process forward in search of optimal solutions.

Crossover is implemented using a version of the Deaven and Ho atomic crossover scheme modified to operate on molecular cluster systems.<sup>1</sup> In traditional Deaven and Ho crossover parents are selected for crossover using one of the normal selection methods (e.g. Roulette Wheel Selection). For each parent cluster a random axis of the cluster is chosen, and both parents are rotated to align these axes. The atoms of each parent are sorted according to their distance along this axis. A crossover site  $P \leq N_{total} - 1$  is then chosen, and atoms  $[1, P]$  are taken from the first parent and atoms  $[P + 1, N_{total}]$  are taken from the second parent cluster. The atoms from both parents are then combined to form a child cluster which is added to the next generation.

I implemented a modified version of the Deaven and Ho crossover scheme which is somewhat more flexible, and works for both molecular and atomic systems. This is shown schematically in Figure 5.6. In the modified crossover scheme we select a number of parents ( $N_{parents} \geq 2$ ) and a number of crossover points  $P_N \leq N_{parents} - 1$  at random. The scheme is implemented as outlined above, with the primary difference being that the objects being arranged are Molecule objects as outlined in Section 5.2.1. We also allow for more than two parent clusters to be involved in crossover, taking Molecules from each in turn to form the child cluster. The molecule objects are treated as rigid bodies, centred at the centre of mass of the molecule for the purposes of

crossover. The clusters are translated so their centres of mass are in the same position, then a random axis is chosen and extending along that axis molecules are taken from the first parent cluster, according to their position along the axis until the first crossover point is reached. From this point, molecules are taken from the next parent until the next crossover point is reached etc.



**Figure 5.6** Schematic representation of the modified Deaven and Ho crossover scheme as implemented in BMPGA

## Mutation

Mutation is the process which allows for increased genetic diversity to be introduced into the population.

Most mutation schemes must be tuned for the systems being investigated, but there are some which are general enough that they can be used with almost any system. For example, one common mutation added in to many GAs is to replace a population member with an entirely new member with a freshly randomised genome. This has a biological analogue when one considers the case of a sub-population of a species that has become separated from another group through some geological or environmental event. For example a thawing ice age has led to two or more populations becoming isolated from one another when retreating glaciation makes previously passable bodies of water into impassable barriers. If there is then a temporary, cold, glacial period after a prolonged period of separation, a small number of individuals could cross between the two populations and this injects new genetic information into the meta-genome of the population.

**Single Particle Translations:** Single particle translations are arguably the simplest mutations implemented within BMPGA. In a single particle translation mutation, a particle is chosen at random (this can be either an Atom, or a Molecule object), and a Uniform or Normal random vector is applied to it. A check is then performed to ensure there is no overlap between particles, with the mutation being rejected if there is an overlap found. The user must select the maximum and minimum magnitudes of the applied vector, and can select the distribution the random vector is drawn from. Options for the distribution are: Uniform or Normal with a Normal distribution being the default.

**Multiple Particle Translations:** In Multiple Particle Translation mutations, multiple particles numbering up to the total cluster size are selected from the cluster and have Uniform or Normal random vectors added to their coordinates. The user can configure the number of particles selected for each step using, a fixed number of particles, or a range of numbers of particles drawn from a chosen distribution (Uniform by default).

**Algorithm 3** Multiple Particle Translation

---

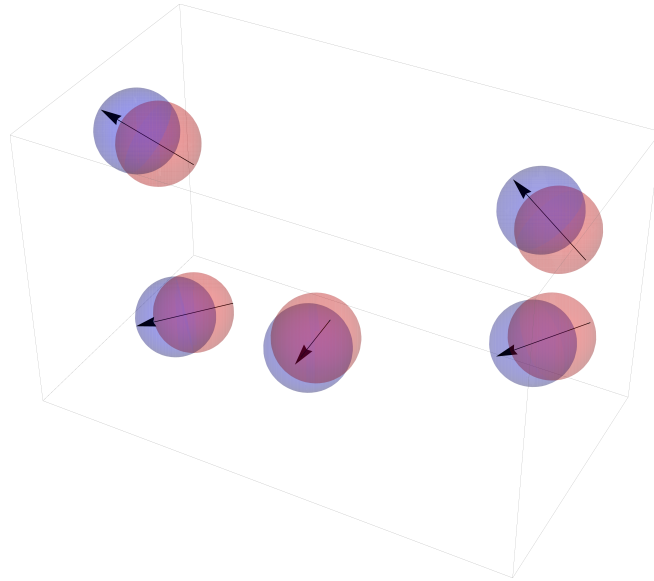
```

1: function MULTIPLEPARTICLETRANSLATION( $Particles, v_r$ )      ▷ Where; Particles is all
   candidate particles,
2:   Select Particles  $p_{\mathbf{n}}$  with coordinates  $r_{\mathbf{n}}$  from Particles
3:   for  $i, 0, \mathbf{n}$  do
4:      $v_r = f[3]$       ▷ Draw a length(3) vector from pseudo-random distribution  $f()$  (e.g.
       Normal or Uniform)
5:      $r_i += v_r$ 
6:   end for
7: end function MultipleParticleTranslation

```

---

**Shake:** In Shake moves every particle in the cluster has a small uniform random vector applied to its' coordinate before the whole cluster is locally optimised. These whole-cluster mutation moves can aid the system in escaping local kinetic traps. Small random vectors with a significantly smaller magnitude than used for the other translation moves. These moves are used as this move-class is designed to overcome kinetic traps. See Figure 5.7 for a visual representation of this mutation.



**Figure 5.7** Schematic representation of the shake mutation move, all the red particles are translated by small random vectors to give the blue particles. Note the vectors have been exaggerated for illustration.

**Algorithm 4** Shake

---

```

1: function SHAKE( $r_i, f(), r_{ij}$ ) ▷ Where;  $r_i$  is the coordinate of
   particle  $i$ ,  $f(3)$  is a length 3 vector drawn from a given pseudo-random distribution,  $r_{ij}$  is
   the maximum inter-particle distance.

2:   for  $i, 0, i_{max}$  do

3:      $v_r = f(3)$  ▷  $v_r \ll \max r_{ij}$ 

4:      $r_i+ = v_r$ 

5:   end for

6: end functionShake

```

---

**Single Particle Rotations:** In single particle rotation moves, a particle is selected at random from the cluster and has a random rotation around a random axis applied to it. These moves are not suitable or useful in systems with spherically isotropic particles as they result in no net change to the system. For the mutation move, a random vector is drawn from the chosen distribution and a random rotation in the interval  $\pm\pi$  is chosen. The corresponding rotation matrix is then calculated and applied to the particle about the chosen axis.

---

**Algorithm 5** Single Particle Rotation

---

```

1: function SINGLEPARTICLEROTATION( $p, \mathbf{c}, v_r$ ) ▷ Where;  $p$  is all candidate particles,  $\mathbf{c}_i$  is
   the coordinate of particle  $i$ ,  $v_r$  is the random vector.

2:   Select Particle  $p_i$  with centre of mass  $\mathbf{c}_i$  from Particles

3:    $v_r = f[3]$  ▷ Draw a length(3) vector from pseudo-random distribution  $f()$  (e.g. Normal
   or Uniform)

4:    $r_i+ = v_r$ 

5: end functionSingleParticleRotation

```

---

**Multiple Particle Rotations:** In Multiple Particle Rotation mutation moves, different random rotations are individually applied to a number of particles within the cluster. It is possible to select a fixed or uniform random number of particles from the clusters to operate on. The selected particles then each have a uniform random rotation applied to them in the

same method as described above.

---

**Algorithm 6** Multiple Particle Rotation

---

```

1: function MULTIPLEPARTICLETRANSLATION( $Particles, v_r$ )      ▷ Where; Particles is all
   candidate particles,
2:   Select Particles  $p_n$  with coordinates  $r_n$  from Particles
3:   for  $i, 0, n$  do
4:        $v_r = f[3]$       ▷ Draw a length(3) vector from pseudo-random distribution  $f()$  (e.g.
   Normal or Uniform)
5:        $r_i += v_r$ 
6:   end for
7: end function MultipleParticleTranslation
    
```

---

**Rattle:** In the Rattle mutation scheme a small, unique, uniform random rotation about an axis originating from the centre of mass is applied to each particle in the cluster. This is the rotational equivalent of the Shake mutation outlined above. This mutation scheme is designed to help the system overcome kinetic traps, and is particularly useful in clusters with highly directional inter-particle bonds, for example hydrogen bonding.

---

**Algorithm 7** Rattle

---

```

1: function RATTLE( $\mathbf{R}, i_{max}$ ) ▷ Where;  $\mathbf{R}$  is the configuration of the cluster, and  $i_{max}$  is the
   total number of particles in the cluster.  $U(0, x)$  is a uniform distribution between 0 and  $x$ 
2:   for  $i = 1, i = i_{max}, i += 1$  do
3:        $\vec{u} = \|U(0, 1)[3]\|$       ▷ Uniform random vector  $u_x^2 + u_y^2 + u_z^2 = 1$ 
4:        $\theta = U(0, 2\pi)$ 
5:        $\mathbf{R}_i = \mathbf{u}(\mathbf{u} \cdot \mathbf{R}_i) + \cos(\theta)(\mathbf{u} \times \mathbf{R}_i) \times \mathbf{u} + \sin(\theta)(\mathbf{u} \times \mathbf{R}_i)$ 
6:   end for
7: end function Rattle
    
```

---

**New Random Cluster:** There is a method implemented to generate new random clusters. This was originally written to generate the initial pool as described above. In this mutation,

there is no genetic information passed on. Instead an entirely new cluster is generated, locally optimised, and added to the pool. The advantage of this method is that it injects a significant amount of new genetic diversity into the population, however this comes at the cost of slowing down convergence.

**Monte Carlo:** I implement short, finite-temperature, Monte Carlo (MC) runs as a form of mutation. In these mutation moves we start with the child-cluster as the initial configuration, and run a (usually) relatively high temperature Monte Carlo simulation for a pre-determined number of steps, (configured by the user at instantiation time). The user can configure all the required parameters before the start of the optimisation. The temperature parameter used in the the MC runs can either be fixed giving the NVT ensemble or there is an adaptive temperature scheme implemented, targeting a specified acceptance ratio. There is also an optional quenching scheme which can be applied at the end of the MC run, and also an annealing scheme, though this tends to lead to the runs being longer than is reasonable as a mutation. The energy evaluations are performed by the worker nodes and the moves are implemented on the master node. As this is a Lamakian GA, the final configuration is locally minimised at the end of each MC run.

**Threshold Monte Carlo:** Threshold Monte Carlo is implemented as a special case of the Monte Carlo mutation move class. Threshold MC sees the chosen configuration undergo a short, infinite-temperature Monte Carlo run, subject to a total-energy cap for the system. Any configuration will be accepted provided it does not breach this energy cap. This permissiveness tends to see the system choose many states close to the energy cap as this area of the landscape is the richest in terms of the density of states. Thus the system is allowed to rapidly explore the landscape. The final configuration of the system is then stochastically quenched a number of times, and all novel minima are added to the database. Like with the standard MC runs, the user can configure all the required parameters before the start of the optimisation and the energy evaluations are again performed by the worker nodes and the moves are implemented on the master node.

## Local optimisation

Since BMPGA is a Lamarckian GA there is a local optimisation step performed after each genetic operation. Lamarckian evolution is a theory of natural selection that was proposed as an alternative to Darwinian evolution. Lamarck's theory had it that individuals would accumulate small changes to their physiology over the course of their lives, and that these individual physical adaptations were then heritable and passed on to the offspring. The most famous thought experiment associated with this theory is probably the one about the development of Giraffes. The thinking goes that the proto-giraffe was some kind of horse/gazelle -like creature that lived on the plains of sub-Saharan Africa. Food was scarce for these creatures at ground level, and what little food there was was under much competition from other animals in similar niches. This led to the proto-giraffes seeking a new food source that was under less competition, the food source they saw was leaves on the local trees. But since the proto-giraffes were fairly short they could only reach the leaves on the lower branches. Through repeated stretching upwards to strain for the tantalising leaves that were on the slightly higher branches, Lamarck proposed that the necks of the proto-giraffes became slightly elongated. This adaption through personal effort, he argued, was then passed on to their children, who were born with slightly longer necks than their parents had at birth. Through repeated applications of this cycle of soft inheritance, Lamarck's theory had it that the stubby-necked proto-giraffe gradually morphed over many generations into the majestically benecked creatures we know today.

The main problem with Lamarck's theory is that it is *almost*<sup>1</sup> totally wrong. Soft heritability is not a primary driver for evolution by natural selection, however it were a real phenomenon it would lead to extremely fast evolution. It is for this reason that we borrow a little from the not-entirely-correct theory of Lamarckian evolution for our GA and decide to locally optimise at each step in our evolution before we go on to create our offspring. This is analogous to us allowing our proto-giraffes to do that little extra bit of neck-growth and pass it down to their children.

---

<sup>1</sup>I say almost here because there is a relatively new and fascinating science called epigenetics which actually does allow for the passing of phenotypic and or other traits through extra-chromosomal methods which effect gene expression rather than the genes themselves, i.e. methylation of DNA or translocation events.



## Data Storage / Database

BMPGA stores minima in an SQL database. Every minimum found during the optimisation is stored. The database is implemented in a thread-safe manner by using a FIFO queue data structure, this allows minima to be inserted in any order as they are returned by the worker nodes, thus allowing for the inherent asynchronicity in the system. I have written an Object-Relational Mapping (ORM) using the SQLAlchemy package, that directly maps Cluster objects onto a database table.<sup>113</sup> This allows for fast access to the clusters for analysis, as SQL allows for fast retrieval of clusters using the native Query language. For the sake of efficiency we compare each cluster to clusters already in the database before inserting it. If the cluster already exists in storage, then we instead increment a counter, and append the step at which the cluster was found to one of the columns in the database. Using an SQL database with automated UUIDs as the keys to the clusters table, we gain a unique identifier for each cluster, this again makes the analysis easier, as merging databases from multiple different optimisations automatically removes duplicate minima. Using a SQL database also allows me to define one of the columns as an arbitrary binary datatype, which allows the storage of any additional properties that are not initially expected to be calculated.

## Model of Parallelism

BMPGA implements a master-worker, distributed parallelism model. Under this model we have a controlling, (Master), process communicating with an arbitrary number of Worker processes. In the case of BMPGA, the master process is responsible for all of the genetic operations and controls the database of minima. This Master is implemented in a multi-threaded manner, and spawns a number of separate threads to perform the genetic operations, as well as the more mundane bookkeeping tasks such as controlling the worker nodes and managing the database.

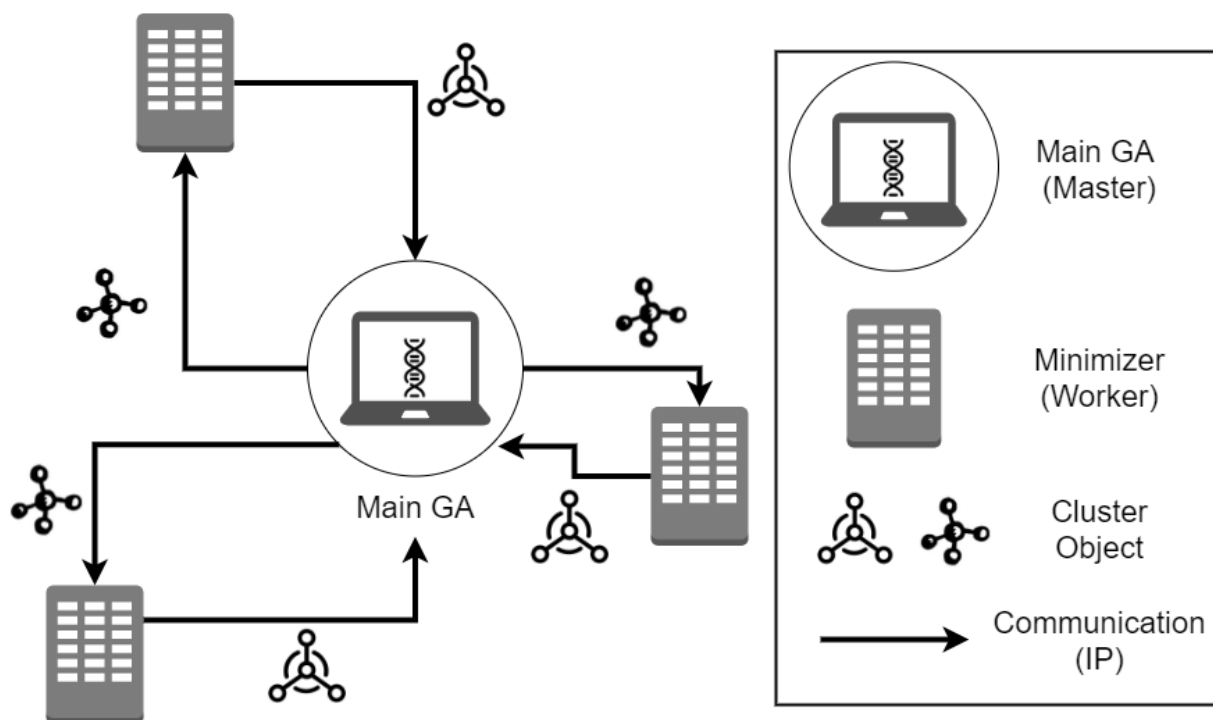
The Master node is responsible for all of the genetic operations, and coordinates all the minimisation and energy evaluations performed by the worker nodes. The selection of parents, crossover and all mutation types are implemented in a thread-safe manner making them suitable for accelerated multi-threaded computation.

The Worker nodes are used solely for performing local optimisations. As BMPGA is implemented as a Lamarkian GA, there is a local optimisation performed after each genetic operation.

The communication between the Master and worker nodes is implemented using web-sockets communicating via HTTPS provided by the Pyro4 python package.<sup>114</sup> This is represented schematically in Figure 5.8. Using this method for communication was chosen because of its flexibility. The user can spawn worker nodes on any HPC system they have access to and can register these workers by providing them with the URI (Unique Resource Identifier) of the master node. This initiates communication between the worker and master. The master makes use of FIFO (First-In-First-Out) queues to asynchronously distribute and collect work. As trial members of the population are created, by crossover, mutation etc. they are placed into a queue. When a worker registers that it is ready to accept a new job it sends a message to the master node. The master will then send the first cluster object in the queue to the worker to be locally optimised or calculate the current energy. Once the cluster reaches geometric convergence to within the required tolerance, the cluster is sent back to the master node and this is placed in a separate FIFO queue until there is a thread available on the master to add it to the pool.

## Pool Methodology

Due to the inherently asynchronous nature of the Master-Worker parallelism model employed in BMPGA it would be inefficient to keep the pool size fixed throughout the optimisation. As the local optimisations are performed by separate compute kernels, potentially on different computers to the genetic processes, and the work is distributed and collected via Internet Protocol (IP), there can be large networking overheads to the distribution and collection of work. When using expensive potentials such as DFT or *ab initio* methods or when optimising large clusters the time the workers take to complete local optimisations is highly dependent on initial configuration. This makes the traditional generational method of performing all of the genetic operations on the population, then performing all the optimisations and forming a new population of the same size unsuitable. Instead, children are added to the pool without culling or replacing previous configurations. This makes keeping a constant pool size by waiting for all



**Figure 5.8** Schematic of the Master-Worker model in BMPGA, showing a Master process and three Worker processes, with communication over the public internet

the members of the current generation to finish optimising an inefficient prospect.

### Initial Pool

There are a number of methods for populating the initial pool of minima. The one we use most commonly in this work is to generate random clusters based on a template. The template can be provided as a cluster object as outlined above, simply defining a list of molecule objects. This reference cluster is copied and random displacement and rotational vectors applied to each molecule object. A check for molecular overlap is performed, canonical atom pairs are assumed to have radii 5% larger than their radii as quoted by Slater.<sup>115</sup> Custom 'atom' types (e.g. Lennard-Jones particles or colloids) are assumed to have a radius of 1.05 unless this is overridden by the user. If overlap is detected, the overlapping molecule objects have small translations applied to them with a bias applied in the opposite direction from the centre of mass of the other molecule, until the overlap is fixed. The initial random displacement vectors are normally distributed around the coordinate origin, with a user-defined maximum displacement. Another option is to pass in a pre-existing database of minima, the GA will then either pick minima using a uniform random distribution, or can use a user defined function to

select minima, (i.e. using a weighted distribution, or minima selected by geometric diversity etc.) The user can also pass in a specific list of minima, for example if they wish to continue from the output of an earlier optimisation attempt that is believed not to have converged, or aborted for some other reason.

## Implemented Potentials

There are a number of potentials implemented in BMPGA, for use in a variety of scenarios, these are listed in Table Table 5.1. Different potentials are useful for studying different systems, and at different levels of precision. There are a number of purely empirical potentials implemented in BMPGA. The implemented Newtonian methods are the Generalised Lennard-Jones potential, TIP4P water potential (and by extension the OPLS force field) and the AMBER potential as implemented in the AmberTools 18 package.<sup>116</sup> The advantage of these Newtonian methods over the more accurate methods such as DFTB, DFT and *Ab Initio* is that they are comparatively cheap to compute. This allows for the rapid exploration of the landscape. The putative minima can then be further refined using one of the more accurate methods.

**Table 5.1** Potentials in BMPGA

Type	Potential	Form
Newtonian Methods	Lennard Jones	$V_{LJ} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{2N} - \left( \frac{\sigma}{r} \right)^N \right]$
	TIP4P water	$V = \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_0 r^2} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$
	AMBER	See equation 5.6
Quasi-Newtonian Methods	DFTB	
Non-Newtonian Methods	Gaussian DFT	Gaussian
		NWChem
	Plane-wave DFT	VASP
		Quantum Espresso

The AMBER potential is given by:

$$\begin{aligned} E_{Total} &= E_{Bonds} + E_{Angles} + E_{Dihedrals} + E_{Non-bonded} \\ E_{Bonds} &= \sum_{i \in bonds} k_{bi} (l_i - l_i^0)^2 \\ E_{Angles} &= \sum_{i \in angles} a_i (\theta_i - \theta_i^0)^2 \\ E_{Dihedrals} &= \sum_{i \in dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ E_{Non-bonded} &= \sum_{i \in non-bonded} \frac{q_i^2}{4\pi\epsilon_0 r_i} + 4\epsilon_i \left[ \left( \frac{\sigma_i}{r_i} \right)^{12} - \left( \frac{\sigma_i}{r_i} \right)^6 \right] \end{aligned} \tag{5.6}$$

The total AMBER energy is the sum of the intra-molecular and inter-molecular energy contributions. The intra-molecular forces are the sum of the deviation from the ideal inter-atomic bond lengths and angles as approximated by a harmonic spring potential, and the difference from the parameterised ideal dihedral angle. The inter-molecular interaction energy is given by a simple OPLS-like force field combining a coulombic and a dispersion term.

## 5.3 Results

### 5.3.1 Bench-marking

BMPGA was bench-marked for both performance as an optimiser and speed/parallelism. The implemented potentials were successfully tested against known/published systems during development, and these results are not presented here as they are boring.

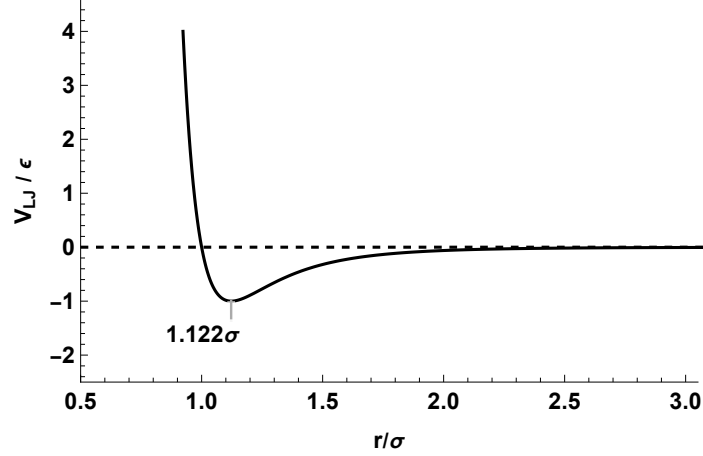
#### Speed/Parallelism Bench-marking

There is very little point in developing new global optimisation algorithms/implementations if they provide no benefit over earlier methods. In the case of BMPGA, we were attempting to improve over prior work in a number of ways, namely; ease of use for the optimisation of gas-phase molecular systems, inclusion of novel mutation schemes, simple extensibility, and improved parallelism.

In the parallelism tests below, the compute workers were run on the tier 1 Hazel Hen HPC resource at HLRS in Stuttgart, Germany. On this system each compute node comprises two 12-core Intel CPUs (2.5GHz) for a total of 24 real cores and 48 threads. There are 128GB RAM per node and unless otherwise stated this was allocated at 2.5GB per thread. There are a total of 7712 compute nodes on the system, but due to this being a shared compute resource and a limited amount of resource being allocated for this project, I did not attempt any really large runs.

### 5.3.2 Global Optimisation of Lennard-Jones Clusters

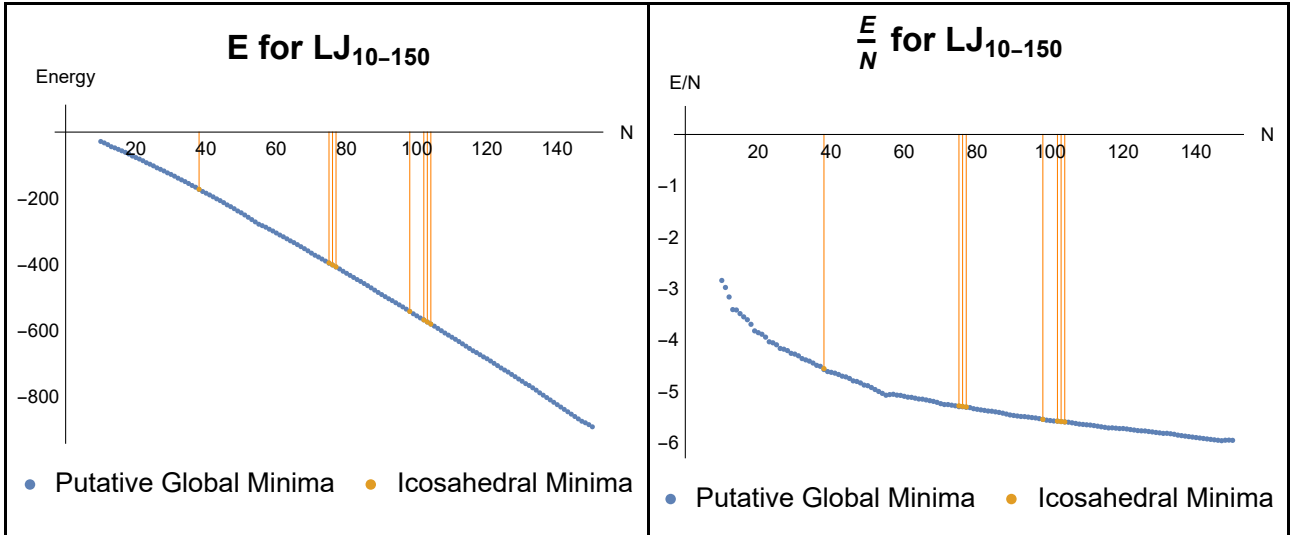
The Lennard-Jones system has been extensively studied over the years, and the low-lying minima and underlying energy landscapes are very well known. Due to the well-understood nature of the Lennard-Jones system it was chosen as one of the test candidates. The interaction energy  $V_{LJ}$ , for the Lennard-Jones (LJ) potential is described by  $V_{LJ} = 4\epsilon \left[ \left(\frac{\sigma}{r}\right)^{2N} - \left(\frac{\sigma}{r}\right)^N \right]$ . Where  $\epsilon$  denotes the interaction strength between the two particles in question,  $\sigma$  is the ideal interaction length at which the energy is 0, and  $r$  is the interparticle distance.



**Figure 5.9** Lennard-Jones (6-12) potential, highlighting the minimum at  $\approx 1.122\sigma$

Using BMPGA as reported above, I reproduce the published global minima for the Lennard-Jones  $N$ -particle system in the range 10-150. I also capture the icosahedral local minima close to iso-energetic with the global minimum of the  $LJ_{38,98,103,104}$ , and 148 systems.

The parameters for the GA and the Lennard-Jones potential used in the global optimisation of these clusters are given in Table 5.2. We use a small minimum pool size (10).



**Figure 5.10** Energy and Energy per particle for all the putative Lennard-Jones optimisation runs. Icosahedral minima are highlighted using vertical orange lines.

As can be seen in Figure 5.10 and Table 5.3 the previously known global minima for all the studied cluster sizes are reproduced, along with the corresponding important low-lying minima.

**Table 5.2** Parameters used in the global optimisation of the Lennard-Jones 6-12 system using BMPGA. Note: The mutations selected were uniformly distributed between the methods listed.

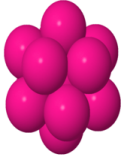
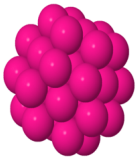
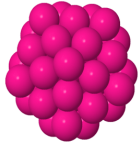
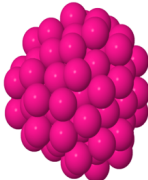
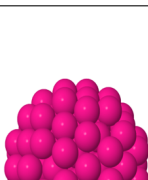
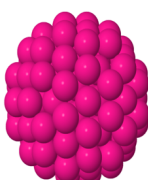
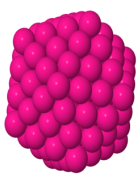
GA Parameters	
$PoolSize$	30
$Gen_{Max}$	100
Initial Pool Method	Random Cluster
<b>Selection Method</b>	Tournament
Tournament Size	3
Mutation Schemes	
Mutation	$p_{rel}$
Random Translations	1
Shake	1
New Random Cluster	1
Lennard-Jones Parameters	
$N$	6
$\sigma_{ij}$	1
$\epsilon_{ij}$	1

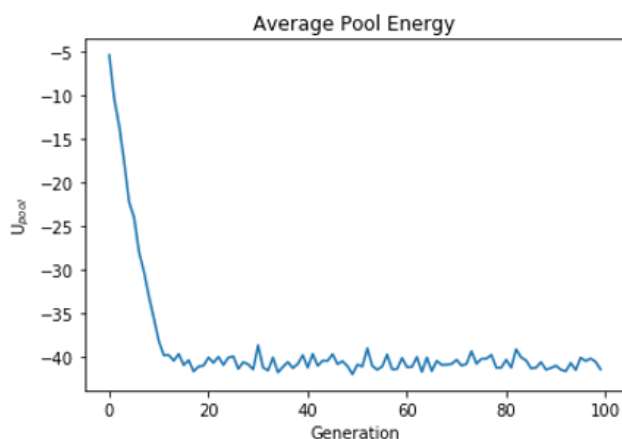
Looking specifically at the LJ13 system which was optimised using the parameters outlined in Table 5.2, Figure 5.12 shows the mean energy of minima in the pool over successive generations. There is a rapid drop in average energy within the pool over the first 10 generations ( $\approx 300$  genetic operations) after which the average pool energy remains low as shown in Figure 5.11. As can be seen in Figure 5.12 (b)), the lowest energy minimum is found after 850 genetic operations.

For the LJ38 system, Figure 5.13, the global minimum is found in 1100 local optimisations.

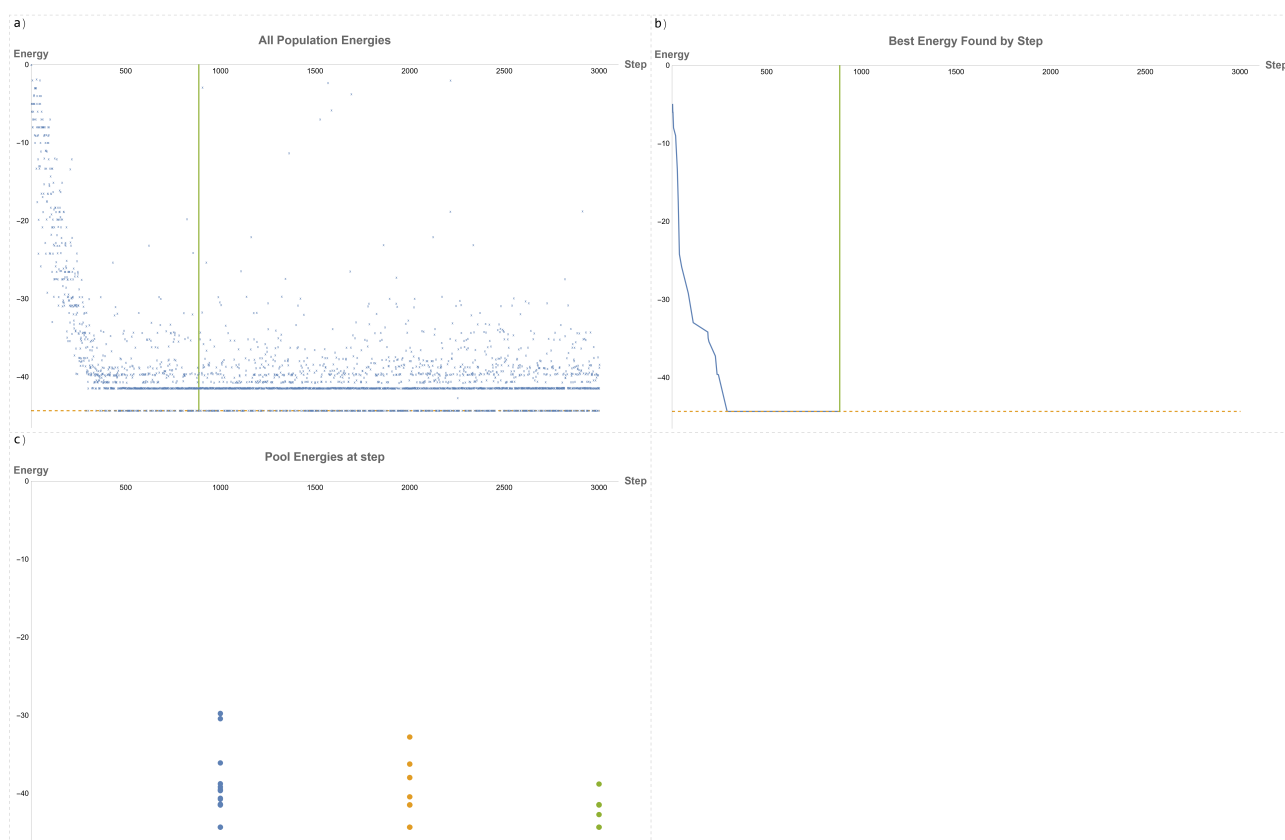


**Table 5.3** Low lying minima of interest for the LJ 6-12 system

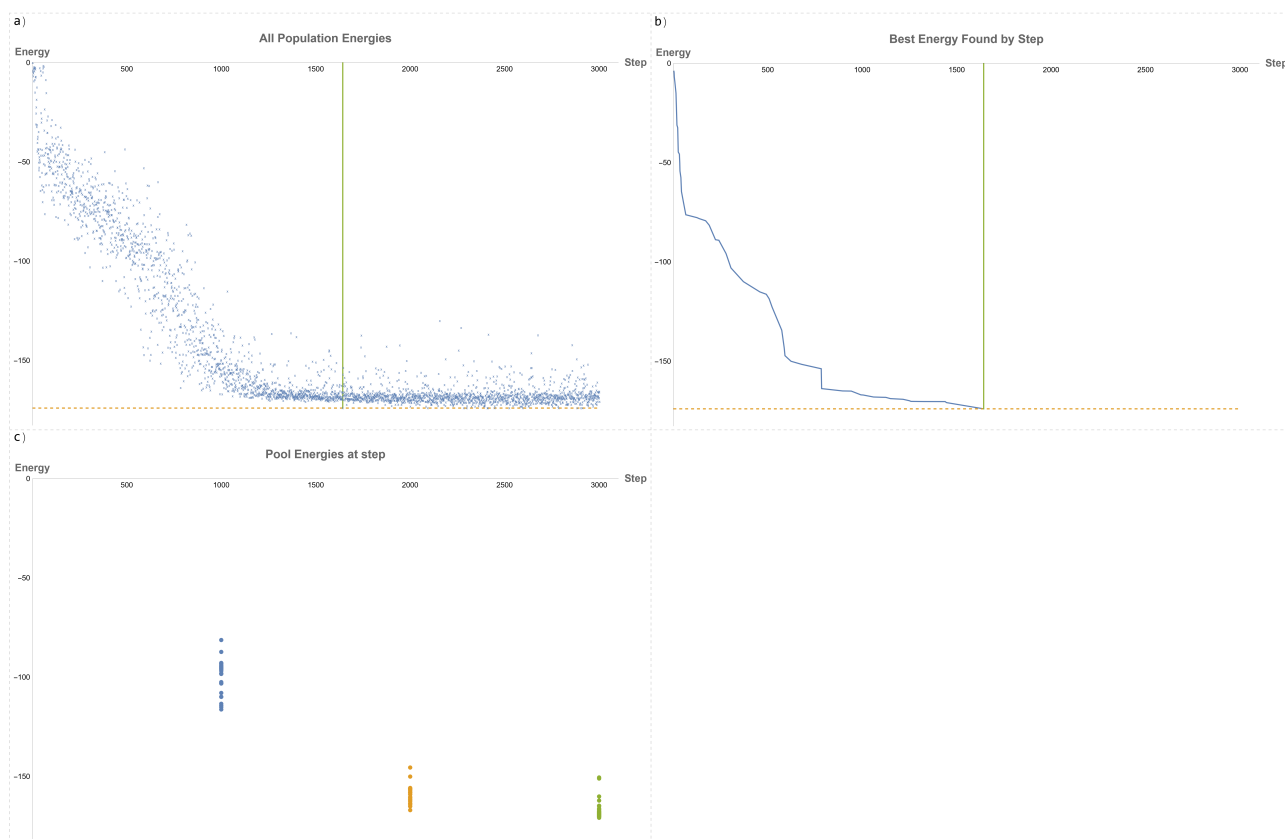
$N$	Energy ( $E$ )	$EN^{-1}$	Image
13	-44.326801	-3.40975	
38	-173.928427	-4.57706	
$38C5_v$	-173.252378	-4.55927	
98	-543.665361	-5.54761	
$98C5_v$	-543.643061	-5.54738	
135	-790.278120	-5.85391	
147	-876.461207	-5.96232	



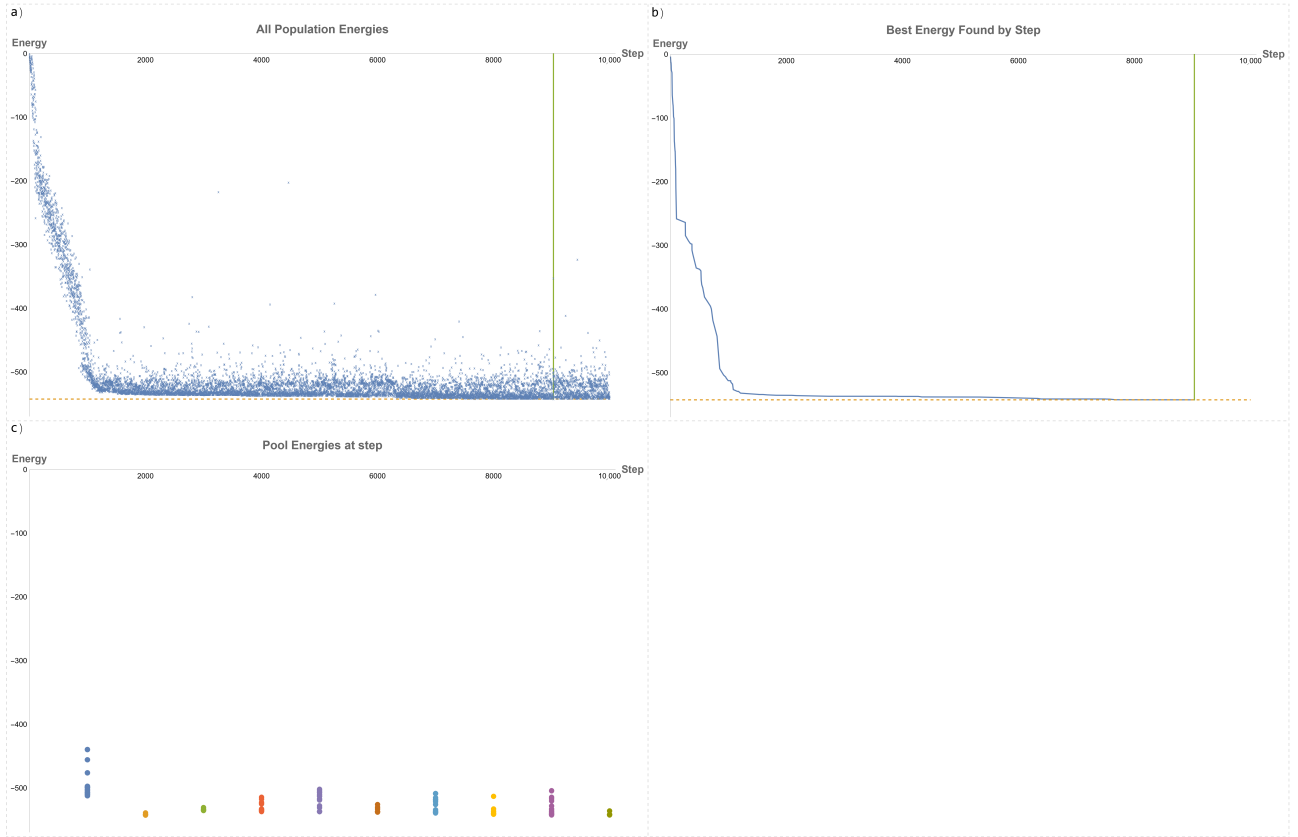
**Figure 5.11** Plot of the mean energy of the pool during the optimisation of the LJ13 system



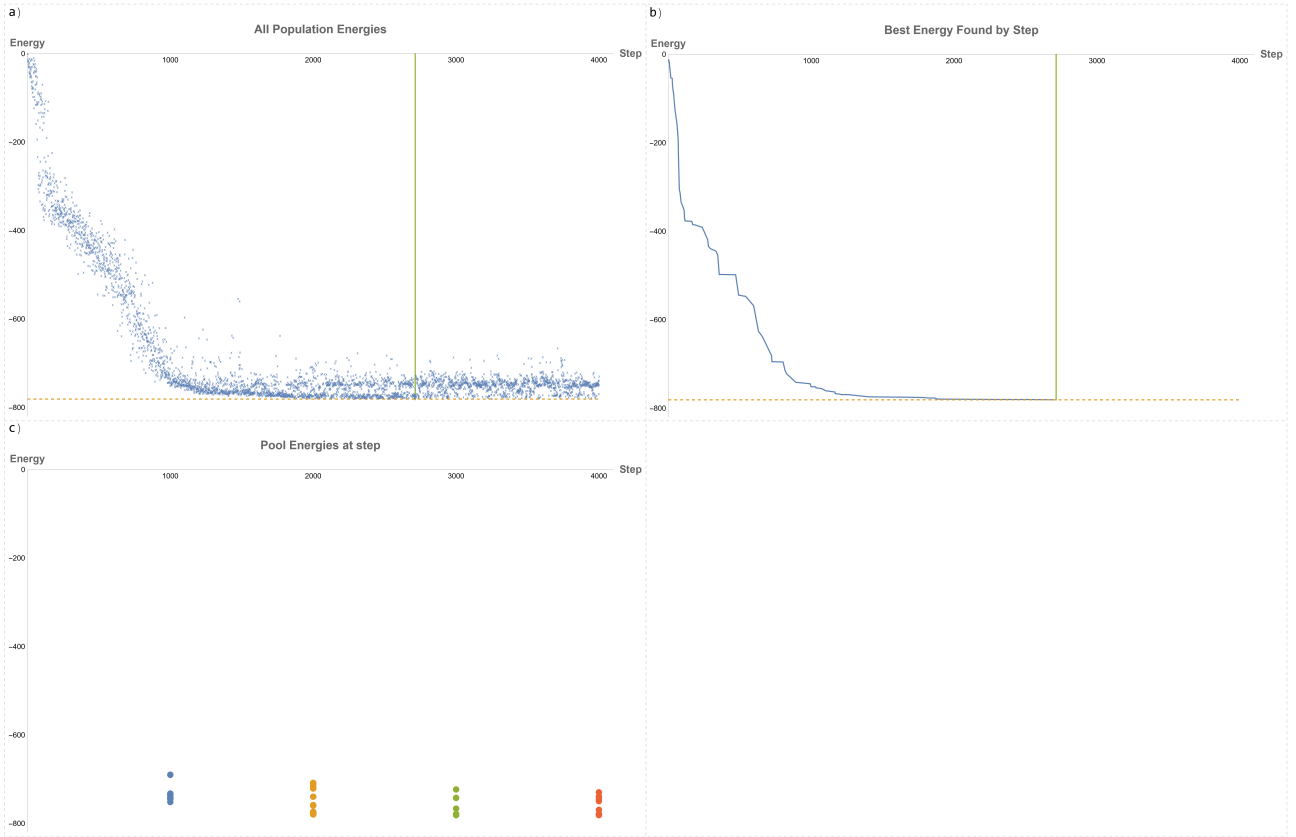
**Figure 5.12** Plots for one run of the LJ13 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum



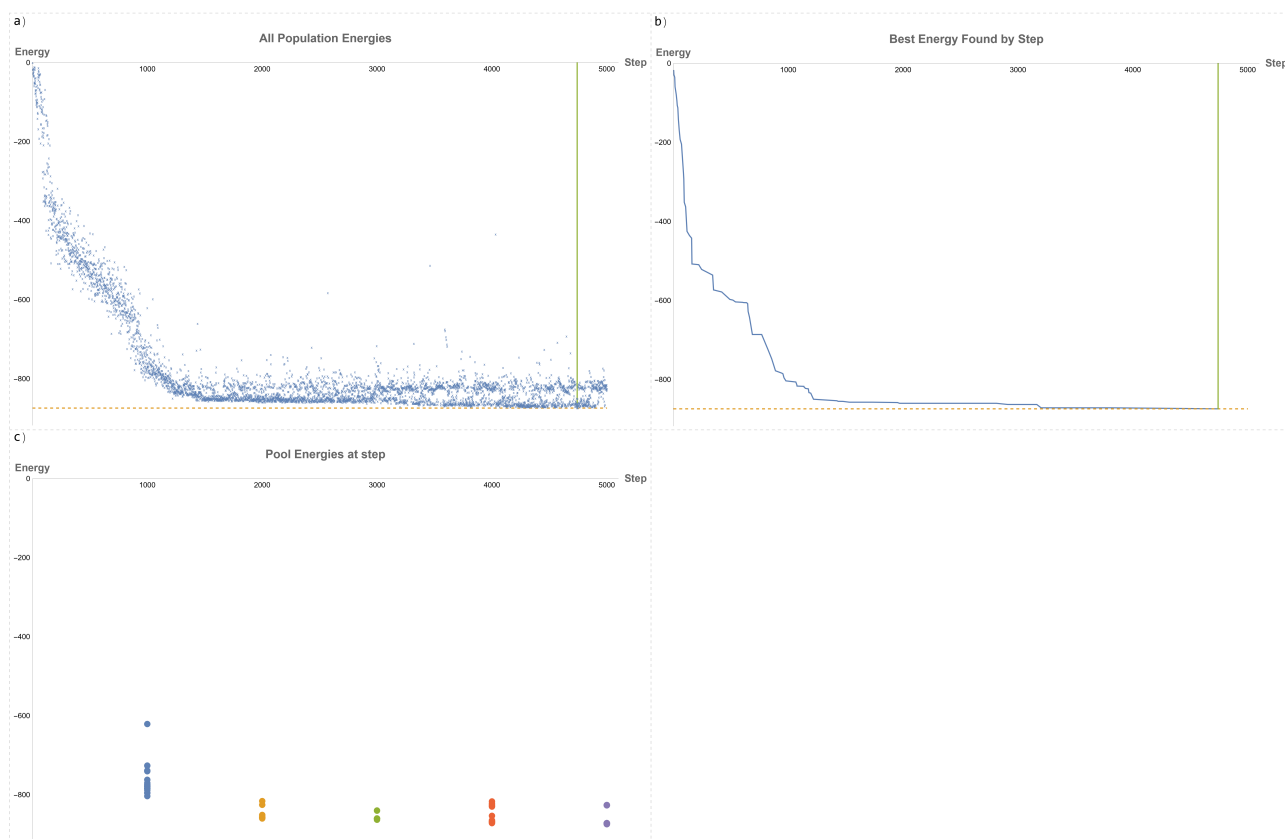
**Figure 5.13** Plots for one run of the LJ38 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum



**Figure 5.14** Plots for one run of the LJ98 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum



**Figure 5.15** Plots for one run of the LJ135 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum



**Figure 5.16** Plots for one run of the LJ147 system showing a.) All minima energies found at each step in the GA run, b.) The best energy found so far at each step in the optimisation, c.) The energy of all minima currently in the pool every 1000 steps during the optimisation. The solid green line in a+b shows the step at which the putative GM was first found. The dashed line shows the energy of the putative global minimum

## 5.4 Conclusions

I have presented benchmarking results for this new GA using the Lennard-Jones potential. I show that the parallel performance of this method is extremely good, and scales very well to a large number of cores. This new GA method shows promise for allowing very large-scale optimisations to be performed on a large range of systems, only limited by the availability of HPC resources. The inclusion of the Threshold method into the GA in two different capacities has been successful and allows for rapid energy landscape exploration. There is also the potential benefit of allowing this hybrid method to perform automatic transition state searches using the Threshold method, though this may require a little extra development time. BMPGA is a successful implementation of a parallel pool GA geared towards optimisations of molecular and atomic clusters at scale and using a number of different potentials.

# Chapter Six

## Overall Conclusions

I believe that I have provided comprehensive insights into the structural and energetic properties of various hydrated ion and metal clusters. My investigation into sulfate and perchlorate micro-hydrated clusters highlights the significant influence of ionic charge on cluster morphology. Sulfate and sulfite clusters tend to have centrally positioned ions, forming highly symmetrical structures, whereas perchlorate and chlorate clusters exhibit less symmetry with surface-localised ions. This distinction underscores the role of ionic environment and charge in determining cluster configurations, with sulfate-like clusters maintaining central ion positions and higher coordination numbers, and perchlorate-like clusters showing increased ion displacement from the centre as ionic charge decreases.

My study on hydrated  $\text{Ag}_m\text{Au}_n^+$  clusters revealed that global minima geometries, determined using the GIGA algorithm and re-optimised with the  $\omega\text{B97X-D}$  functional, remained consistent with previously reported structures. The interaction patterns differed for Ag-O and Au-O bonds, influencing the overall geometry and stability of the clusters. Notably, the tetrahedral  $\text{Au}_4^+$  cluster was not the minimum structure when stabilised by water. Binding energies indicated a preference for silver sites in mixed clusters, attributed to the higher positive charge on silver atoms. A unique structural rearrangement was observed in the  $[\text{Ag}_3\text{Au}]^+(\text{H}_2\text{O})_2$  cluster upon hydration, driven by higher binding energy of the second water molecule on the tetrahedral cluster. Differences in interaction strength for non-direct bonds is also observed, with water-water hydrogen bond interactions becoming stronger in clusters with higher gold content.



Finally, the benchmarking of a new Genetic Algorithm (GA) using the Lennard-Jones potential demonstrated excellent parallel performance and scalability, promising for large-scale optimisations across diverse systems. The inclusion of the Threshold method facilitated rapid energy landscape exploration, indicating potential for automated transition state searches.

Overall, these findings emphasise the crucial role of ionic charge, composition, and hydration in determining the structure and stability of ion and metal clusters. This research provides insights for future experimental and computational studies aimed at exploring the stability and reactivity of hydrated clusters in various chemical environments.

# References

- [1] D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.*, 1995, **75**, 288–291.
- [2] M. Chaplin, *Hexagonal Ice*, [http://www1.lsbu.ac.uk/water/water\\_lh.html](http://www1.lsbu.ac.uk/water/water_lh.html), 2016, Accessed: 25/01/16.
- [3] L. C. Smeeton, *Ph.D. thesis*, University of Birmingham, 2015.
- [4] L. C. Smeeton, J. C. Hey and R. L. Johnston, *Inorganics*, 2017, 20.
- [5] L. C. Smeeton, J. D. Farrell, M. T. Oakley, D. J. Wales and R. L. Johnston, *Journal of Chemical Theory and Computation*, 2015, **11**, 2377–2384.
- [6] D. Feller, E. D. Glendening and W. A. de Jong, *The Journal of Chemical Physics*, 1999, **110**, 1475–1491.
- [7] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926.
- [8] E. L. Wagner, *The Journal of Chemical Physics*, 1962, **37**, 751.
- [9] W. R. Cannon, B. M. Pettitt and J. A. Mccammont, *J. Phys. Chem.*, 1994, **98**, 6225–6230.
- [10] M. Cogoni, B. Aguanno, L. N. Kuleshova, D. W. M. Hofmann, B. Daguanno, L. N. Kuleshova and D. W. M. Hofmann, *J. Chem. Phys.*, 2011, **134**, 1–10.
- [11] J. D. Farrell and D. J. Wales, *The journal of physical chemistry. A*, 2014, **118**, 7338–48.
- [12] D. J. Wales and M. P. Hodges, *Chemical Physics Letters*, 1998, **286**, 65–72.
- [13] W. J. Xie and Y. Q. Gao, *Journal of Physical Chemistry Letters*, 2013, **4**, 4247–4252.

- 
- [14] D. J. Wales and M. P. Hodges, *Chemical Physics Letters*, 1998, **286**, 65–72.
- [15] S. Kazachenko and A. J. Thakkar, *Chemical Physics Letters*, 2009, **476**, 120–124.
- [16] T. M. Chang, R. J. Cooper and E. R. Williams, *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, 2013, **135**, 14821–14830.
- [17] F. Hofmeister, *Arch. Exp. Pathol. Pharmacol. (Leipzig)*, 1888, **24**, 247–260.
- [18] P. Jungwirth and D. J. Tobias, *Chemical Reviews*, 2006, **106**, 1259–1281.
- [19] A. P. dos Santos and Y. Levin, *Faraday Discussions*, 2013, **160**, 75–87.
- [20] M. Chaplin, *Kosmotropes and Chaotropes*, [http://www1.lsbu.ac.uk/water/kosmotropes\\_chaotropes.html](http://www1.lsbu.ac.uk/water/kosmotropes_chaotropes.html), Accessed: 12/01/16.
- [21] J. Pokorná, J. Heyda and J. Konvalinka, *Faraday Discussions*, 2013, **160**, 359–370.
- [22] K. D. Collins, *Proceedings of the National Academy of Sciences of the United States of America*, 1995, **92**, 5553–5557.
- [23] W. Kunz, *Current Opinion in Colloid and Interface Science*, 2010, **15**, 34–39.
- [24] E. F. Douglass, P. F. Driscoll, D. Liu, N. a. Burnham, C. R. Lambert and W. G. McGimpsey, *Analytical Chemistry*, 2008, **80**, 7670–7677.
- [25] J. T. O'Brien, J. S. Prell, M. F. Bush and E. R. Williams, *Journal of the American Chemical Society*, 2010, **132**, 8248–8249.
- [26] G. Niedner-Schatteburg and V. E. Bondybey, *Chemical Reviews*, 2000, **100**, 4059–4086.
- [27] W. Kunz, *Current Opinion in Colloid and Interface Science*, 2010, **15**, 34–39.
- [28] G. M. Brown and B. Gu, *The chemistry of perchlorate in the environment*, Springer, 2006, pp. 17–47.
- [29] M. D. Baer, I.-F. W. Kuo, H. Bluhm and S. Ghosal, *The journal of physical chemistry. B*, 2009, **113**, 15843–15850.

- 
- [30] R. G. Keesee, R. Sievert and A. W. Castleman, in *Chemistry of Multiphase Atmospheric Systems*, ed. W. Jaeschke, Springer Berlin Heidelberg, Berlin, Heidelberg, 1986, ch. The Intera, pp. 501–505.
- [31] K. D. Collins, *Proceedings of the National Academy of Sciences of the United States of America*, 1995, **92**, 5553–5557.
- [32] A. T. Blades and P. Kebarle, *Journal of the American Chemical Society*, 1994, **116**, 10761–10766.
- [33] M. Lalitha and L. Senthilkumar, *Journal of Molecular Graphics and Modelling*, 2014, **54**, 148–163.
- [34] U. Baul and S. Vemparala, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2015, **91**, 1–6.
- [35] L. J. A. Siqueira, S. M. Urahata and M. C. C. Ribeiro, *The Journal of Chemical Physics*, 2008, **19**, 8002–8012.
- [36] M. Lund, L. Vrbka and P. Jungwirth, *Communication*, 2008, **130**, 11582–11583.
- [37] Y. Berezovskaya, M. Porrini and P. E. Barran, *International Journal of Mass Spectrometry*, 2013, **345-347**, 18.
- [38] D. J. Wales, *Energy Landscapes*, Cambridge University Press, 2003.
- [39] T. van Mourik, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 2004, **362**, 2653–2670.
- [40] N. Goldenfeld, *Lectures On Phase Transitions And The Renormalization Group*, CRC Press, 1992.
- [41] R. Mrugala, J. D. Nulton, J. C. Schön and P. Salamon, *Phys. Rev. A*, 1990, **41**, 3156–3160.
- [42] P. Sibani, J. C. Schön, P. Salamon and J.-O. Andersson, *EPL*, 1993, **22**, 479–485.
- [43] K. A. Dill and J. L. MacCallum, *Science*, 2012, **338**, 1042–1046.

- 
- [44] J. C. Schön, H. Putz and M. Jansen, *J. Phys. Condens. Matter*, 1996, **8**, 143–156.
- [45] G. Krauss, *Steels: Processing, Structure, and Performance*, ASM International, 2015.
- [46] A. D. Becke, *Journal of Chemical Physics*, 2014, **140**, 18A301.
- [47] S. Yoo, E. Aprà, X. C. Zeng and S. S. Xantheas, *Journal of Physical Chemistry Letters*, 2010, **1**, 3122–3127.
- [48] G. Heinje, W. P. Luck and K. Heinzinger, *J. Phys. Chem.*, 1987, **91**, 331–338.
- [49] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *Journal of Computational Chemistry*, 2003, **24**, 1999–2012.
- [50] P. Koskinen and V. Mäkinen, *Computational Materials Science*, 2009, **47**, 237–253.
- [51] M. Elstner and G. Seifert, *Philos. Trans. A Math. Phys. Eng. Sci.*, 2014, **372**, 20120483.
- [52] G. Murdachaew, C. J. Mundy and G. K. Schenter, *Journal of Chemical Physics*, 2010, **132**, 164102.
- [53] S. Neelamraju, A. Bach, J. C. Schön, D. Fischer, M. Jansen, S. Neelamraju, A. Bach, J. C. Schön, D. Fischer and M. Jansen, *The Journal of Chemical Physics*, 2012, 1–10.
- [54] L. Page and S. Brin, *Google Translate*, [www.google.com/translate](http://www.google.com/translate), Accessed: 26/01/16.
- [55] N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas and M. Head-Gordon, *Journal of Chemical Theory and Computation*, 2013, **9**, 1368–1380.
- [56] M. Goldey, A. Dutoi and M. Head-Gordon, *Physical chemistry chemical physics : PCCP*, 2013, **15**, 15869–75.
- [57] D. S. Lambrecht, G. N. I. Clark, T. Head-Gordon and M. Head-Gordon, *The journal of physical chemistry. A*, 2011, **115**, 11438–54.
- [58] J. Nocedal and S. Wright J., *Numerical Optimization*, Springer, 2006, vol. 2.
- [59] J. Kiefer, *Proc. Am. Math. Soc.*, 1953, **4**, 502.

- 
- [60] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, Chichester, England, 2nd edn., 2000.
- [61] R. P. Brent, *Algorithms for Minimization Without Derivatives*, Prentice-Hall, 1973.
- [62] V. M. Madorski, *DOKLADY AKADEMII NAUK BELARUSI*, 1972, **16**, 884–&.
- [63] D. C. Liu and J. Nocedal, *Mathematical Programming*, 1989, **45**, 503–528.
- [64] G. Liu, Z. Zhou, H. Zhong and S. Xie, *IET COMPUTER VISION*, 2014, **8**, 287–298.
- [65] e. a. Wales, D, *PELE, the Python Energy Landscape Explorer*, <https://github.com/pele-python/pele>.
- [66] D. J. Wales and J. Doye, *The Journal of Physical Chemistry A*, 1997, **93**, 5111–5116.
- [67] J. D. Farrell and D. J. Wales, *The journal of physical chemistry. A*, 2014, **118**, 7338–48.
- [68] V. Rühle, H. Kusumaatmaja, D. Chakrabarti and D. J. Wales, *Journal of Chemical Theory and Computation*, 2013, **9**, 4026–4034.
- [69] M. F. Chaplin, *Journal of the Chemical Society, Faraday Transactions*, 1998, **94**, 1921–1931.
- [70] B. Guillot and Y. Guissani, *Journal of Chemical Physics*, 2003, **119**, 11740–11748.
- [71] E. D. Sloan and C. Koh, *Clathrate Hydrates of Natural Gases*, CRC Press, 3rd edn., 2007.
- [72] A. K. Sum and P. R. Bishnoi, *Fluid Phase Equilibria*, 2005, **228-229**, 489–495.
- [73] S. A. Trygubenko and D. J. Wales, *Journal of Chemical Physics*, 2004, **120**, 2082–2094.
- [74] J. C. Hey, L. C. Smeeton, M. T. Oakley and R. L. Johnston, *Journal of Physical Chemistry A*, 2016, **120**, 4008–4015.
- [75] H. Takeuchi, *JOURNAL OF CHEMICAL INFORMATION AND MODELING*, 2008, **48**, 2226–2233.
- [76] K. Shoemake, *ACM SIGGRAPH Computer Graphics*, 1985, **19**, 245–254.

- 
- [77] G. Henkelman and H. Jonsson, *Journal of Chemical Physics*, 1999, **111**, 7010–7022.
- [78] L. J. Munro and D. J. Wales, *Physical Review B*, 1999, **59**, 3969–3980.
- [79] L. C. Smeeton, M. T. Oakley and R. L. Johnston, *J. Comp. Chem.*, 2014, **35**, 1481–1490.
- [80] M. Valiev, E. Bylaska, G. N., K. Kowalski, T. Straatsma, H. van Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus and W. de Jong, *Computer Physics Communications*, 2010, **181**, 1477.
- [81] R. Bader, *Atoms in molecules : a quantum theory*, Clarendon Press Oxford University Press, Oxford England New York, 1994.
- [82] S. Kazachenko and A. J. Thakkar, *Chemical Physics Letters*, 2009, **476**, 120–124.
- [83] E. Sanz, C. Vega, J. L. F. Abascal and L. G. MacDowell, *Journal of Chemical Physics*, 2004, **121**, 1165–1166.
- [84] J. C. Hey, E. J. Doyle, Y. Chen and R. L. Johnston, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018, **376**, 20170154.
- [85] J. T. O’Brien, J. S. Prell, M. F. Bush and E. R. Williams, *Journal of the American Chemical Society*, 2010, **132**, 8248–8249.
- [86] R. L. Johnston, *Atomic and Molecular Clusters*, Taylor and Francis: London, 2002.
- [87] S. Alayoglu, P. Zavalij, B. Eichhorn, Q. Wang, A. I. Frenkel and P. Chupas, *ACS Nano*, 2009, **3**, 3127–3137.
- [88] Y. Han, D. S. He, Y. Liu, S. Xie, T. Tsukuda and Z. Y. Li, *Small*, 2012, **8**, 2361–2364.
- [89] M. Grzelczak, J. Pérez-Juste, P. Mulvaney and L. M. Liz-Marzán, *Chemical Society Reviews*, 2008, **37**, 1783.
- [90] R. Ferrando, J. Jellinek and R. L. Johnston, *Chemical Reviews*, 2008, **108**, 845–910.
- [91] M. Valden, X. Lai and D. W. Goodman, *Science*, 1998, **281**, 1647–1650.

- 
- [92] A. Shayeghi, C. Heard, R. Johnston and R. Schaefer, *The Journal of chemical physics*, 2014, **140**, 054312.
- [93] A. Shayeghi, D. Götz, J. B. A. Davis, R. Schäfer and R. L. Johnston, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2104–2112.
- [94] J. B. A. Davis, A. Shayeghi, S. L. Horswell and R. L. Johnston, *Nanoscale*, 2015, **7**, 14032–14038.
- [95] J. B. A. Davis, S. L. Horswell and R. L. Johnston, *The Journal of Physical Chemistry C*, 2016, **120**, 3759–3765.
- [96] M. Jäger, R. Schäfer and R. L. Johnston, *Nanoscale*, 2019, **11**, 9042–9052.
- [97] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. De Jong, *Computer Physics Communications*, 2010, **181**, 1477–1489.
- [98] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *Journal of Chemical Physics*, 2010, **132**, 154104.
- [99] G. van Rossum, B. Warsaw and N. Coghlan, *Style Guide for Python Code*, Python Software Foundation PEP 8, 2001.
- [100] G. van Rossum, J. Lehtosalo and Łukasz Langa, *Type Hints*, Python Software Foundation PEP 484, 2014.
- [101] J. Hey, *BMPGA Git Repository*, <https://bitbucket.org/john-hey/bmpga/>, 2021, [Accessed 20-05-2021].
- [102] P. Wegner, in *Object-oriented programming (OOP)*, John Wiley and Sons Ltd., GBR, 2003, p. 1279–1284.
- [103] S. M. Woodley, in *Prediction of Crystal Structures Using Evolutionary Algorithms and Related Techniques*, ed. R. L. Johnston, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 95–132.



- 
- [104] S. M. Woodley and R. Catlow, *Nat. Mater.*, 2008, **7**, 937–946.
- [105] S. M. Woodley, A. A. Sokol and C. R. A. Catlow, *Z. Anorg. Allg. Chem.*, 2004, **630**, 2343–2353.
- [106] S. M. Woodley, G. M. Day and R. Catlow, *Philos. Trans. A Math. Phys. Eng. Sci.*, 2020, **378**, 20190600.
- [107] F. Hayes-Roth, *ACM SIGART Bulletin*, 1975, 15–15.
- [108] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley Pub. Co, Reading, Mass, 1989.
- [109] M. Mitchell, *An introduction to genetic algorithms*, MIT Press, Cambridge, Mass., 7th edn., 2001.
- [110] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.
- [111] A. E. Eiben and J. E. Smith, *Introduction to evolutionary computing*, Springer, Berlin Heidelberg New York Dordrecht London, Second Edition edn., 2015.
- [112] K. A. De Jong, *Evolutionary computation: a unified approach*, MIT Press, Cambridge, Mass, 2006.
- [113] M. Bayer, in *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks*, ed. A. Brown and G. Wilson, aosabook.org, 2012, ch. 1, p. 1.
- [114] A. Harpole, M. Zingale, I. Hawke and T. Chegini, *Journal of Open Source Software*, 2019, **4**, 1265.
- [115] J. C. Slater, *The Journal of Chemical Physics*, 1964, **41**, 3199–3204.
- [116] D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.