Hydrate formation in pharmaceutically relevant salts

by
Alwyn Bernard Dippenaar

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Chemistry at the University of Stellenbosch

Supervisors: Prof. Catharine Esterhuysen & Prof. Delia Haynes
Faculty of Science
Department of Chemistry & Polymer Science

December 2014

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2014

Abstract

A theoretical and experimental study was performed in order to identify factors that influence the propensity of compounds containing anionic functional groups that are commonly found on pharmaceutical drug compounds to form hydrates. A Cambridge Structural Database (CSD) survey was initially undertaken to determine the propensity of different pharmaceutically acceptable anions to form hydrates. The results showed that hydrate formation will take place more regularly when the polarity of the functional group increases. Furthermore, if the charge distribution is very concentrated over the polar groups, hydrate formation will occur more readily. This observation was further investigated by performing a series of potential energy surface (PES) scans for the hydrogen bond (H-bond) in the structure of N-(aminoiminomethyl)-N-methylglycine monohydrate (creatine monohydrate) with various Density Functional Theory (DFT) and Wave Functional Theory (WFT) methods. WFT is often also referred to as *ab initio*, which refers to the construction of the wave function from first principles when this theory is applied. The scans revealed that several strong and directional H-bonds with different geometrical parameters between the carboxylate group and the water molecule are possible, which suggests that the H-bond plays an important role in driving the formation of pharmaceutical hydrates.

A total of 44 hydrate structures were identified that have pharmaceutically acceptable functional groups. Optimisations in the gas phase and in an implicit solvent polarisable continuum solvent model with a variety of solvents showed that there is a significant dependence of the H-bond interaction energy on the anionic group as well as the steric density of surrounding substituents. It was found that the M06-2X method utilising the 6-311++G(d,p) basis set outperformed the other methods that were tested when compared to optimisations performed with the benchmark MP2/aug-cc-pVTZ level of theory. Furthermore, the strength of the H-bond was measured in the 44 experimentally determined structures by using a total of five generalized gradient approximation (GGA) methods, of which two methods contained the DFT-D3 correction. The results of these DFT methods were subsequently compared to results obtained at the benchmark MP2/aug-cc-pVTZ level of theory. The M06-2X method was identified as the most economical method to calculate H-bond energies. It was also found that the H-bond interaction energy shows a substantial dependence on the electrostatic environment. This was observed by a significant decrease in H-bond strength as the relative permittivity of the solvent increases.

The effect of steric density on the H-bond interaction energy was investigated by performing hydrogen bond propensity calculations. These values were then compared to the interaction energies of each structure and the results showed that the presence of large bulky substituents can lead to an increase in bond energy by forcing the anionic functional group closer to the water molecule. Contrastingly, the bulky group can also push the anionic group away from the water molecule and result in a decrease in bond energy. Approximate values for the amount of stabilisation offered to the H-bonding system by the surrounding crystalline environment were calculated by optimising the H-bond geometrical parameters of selected compounds with a combination of the M06-2X and MP2 methods utilising the

6-311++G(d,p) basis set. The H-bond interaction energies were then calculated at the M06-2X/6-311++G(d,p) level of theory and compared to the H-bond interaction energies in geometries that have been fully optimised. After these energies were compared and the crystal packing of each structure was investigated, it was found that the packing of some structures within the crystalline environment limits the number of H-bonds that can be formed between the water and the compound of interest. Full optimisation calculations result in structures with cooperative stabilisation, such that more than one H-bond is found between the two fragments.

The effect of substituents on H-bond interaction energy was investigated by the addition of six electron-donating and electron-withdrawing groups on four aromatic compounds with different anionic functional groups, namely carboxylate, nitrogen dioxide, sulfonate and phosphonate. It should also be mentioned that the nitrogen dioxide is not an anionic functional group, but it was included as it is a neutral radical that often forms hydrogen bonds.

A total of 80 structures were optimised with a combination of the M06-2X and MP2 methods utilising the 6-311++G(d,p) basis set. This was followed by counterpoise corrected single point calculations at the M06-2X/6-311++G(d,p) level of theory. The results showed that the H-bond interaction energy bears no relationship to the inductive strength or the inductive ability of the substituents, but rather the ability of these substituents to rotate the anionic functional group and allow cooperative stabilisation of the H-bond. Furthermore, AIM analysis was performed for the substituted H-bonded aromatic structure. The results showed that electron-donating groups that are placed at the *para* position yield stronger H-bonds, which is once again accompanied by cooperative stabilisation. Electron-withdrawing groups with sufficient inductive effects can result in a weaker H-bond when placed at the *meta* position.

The effect of water activity (a_w) on the hydrate crystal formation was investigated experimentally by performing a series of crystallisations in various solvent mixtures. These mixtures consisted of water mixed with acetone, ethanol and ethyl acetate. A total of three organic acids were used in crystal formation, namely pyridine-4-carboxylic acid (isonicotinic acid), N-amino-iminomethyl-N-methylglycine (creatine) and benzene-1,3,5-tricarboxylic acid. It was found that water activity affects the formation of the hydrate as well as the anhydrous product. Additionally, nucleation and super saturation plays a large role in crystal formation and can serve as an effective technique when the formation of crystals of an appropriate shape and size is required for further analysis.

Uittreksel

'n Teoretiese en eksperimentele studie was uitgevoer om faktore te identifiseer wat die geneigdheid van verbindings met anioniese funksionele groepe wat algemeen gevind word op farmaseutiese dwelm verbindings om die hidraat produk te vorm, affekteer. 'n Opname van strukture in die Cambridge Strukturele Databasis (CSD) is onderneem om die geneigdheid van verskillende farmaseutiese aanvaarbare anione om hidrate te vorm te bepaal. Die resultate het getoon dat hidraatvorming meer gereeld plaasvind indien die polariteit van die funksionele groepe toeneem. Verder is daar ook opgemerk dat 'n gekonsentreerde ladingsverspreiding op die polêre groepe ook tot 'n toename in hidraat vorming sal lei. Hierdie waarneming is verder ondersoek deur 'n reeks potensiële energie oppervlak (PES) skanderings van die waterstof binding (H-binding) vir die struktuur van N-amino-iminometiel-N-metielglisien monohidraat (kreatien monohidraat) met verskeie Digtheids-Funksionele Teorie (DFT) en Golffunksie Teorie (WFT) metodes uit te voer. Die skanderings het getoon dat verskeie sterk, gerigte H-bindings met verskillende geometriese parameters tussen die karboksilaatgroep en die watermolekule kan vorm. Hierdie bevindinge lê klem op die belangrike rol wat H-bindings in die vorming van farmaseutiese koolhidrate speel.

'n Totaal van 44 hidraat strukture met farmaseutiese aanvaarbare funksionele groepe was geïdentifiseer. Optimaliserings is in die gas fase asook in 'n implisiete kontinuum polariseerbare oplosmiddel model met 'n verskeidenheid oplosmiddels uitgevoer. Die resultate het 'n beduidende afhanklikheid van die H-binding interaksie-energie op die anioniese groep asook die steriese afkskerming van omringende groepe getoon. Daar is bepaal dat die M06-2X metode wat saam met die 6-311++G(d,p) basisstel die mees akkuraatste resultate gelewer het in vergelyking met die ander DFT metodes asook die MP2/aug-cc-pVTZ maatstaf. Die H-binding se sterkte is vir hierdie strukture bereken deur vyf GGA metodes te gebruik, waarvan twee metodes van die DFT-D3 korreksie gebruik maak. Die resultate van die berekeninge met hierdie DFT metodes is daarna vergelyk met resultate verkry met die MP2/aug-cc-pVTZ maatstaf. Daar is gevolglik bepaal dat die M06-2X metode die mees ekonomiese metode is om H-binding energië te bereken. Die H-binding interaksie energie toon 'n aansienlike afhanklikheid op die diëlektriese konstante van die oplosmiddel aan. Hierdie waarneming is op grond van 'n beduidende afname in die H-binding interaksie-energie indien die relatiewe permittiwiteit van die oplosmiddel verhoog word gemaak.

Die effek van steriese digtheid is ondersoek deur waterstofbindinggeneigdheid waardes te bereken. Hierdie waardes is met die interaksie-energië van elke struktuur vergelyk. Die resultate dui daarop dat steries digte groepe tot 'n toename in interaksie energie kan lei wanneer die anioniese funksionele groep nader aan die water molekule gestoot word. Verder is dit ook moontlik vir hierdie steries digte groepe om die anioniese groep weg van die water molekule te stoot en gevolglik 'n afname in interaksie energie te veroorsaak. Benaderde waardes vir die hoeveelheid stabilisering wat die omringende kristallyne omgewing aan die H-binding bied is bereken deur die H-binding geometriese parameters van geselekteerde

verbindings met die M06-2X en MP2 metodes en die 6-311++G (d,p) basisstel te optimaliseer. Die H-binding interaksie-energië is gevolglik by die M06-2X/6-311++G(d,p) vlak van teorie bereken en met die H-binding energië in strukture wat volledige optimaliseer is vergelyk. Nadat hierdie waardes vergelyk is, is daar gevind dat die pakking van strukture in the kristallyne omgewing verhoed dat sekere H-bindings tussen die water molekule en die verbinding van belang kan vorm. Strukture wat volledig optimaliseer is, lei tot strukture wat in staat is om koöperatiewe stabilisering te ondergaan. Koöperatiewe stabilisering word gekenmerk deur die vorming van meer as een H-binding tussen twee fragmente.

Die effek van substituente op die H-binding interaksie energie is ondersoek deur die bevoeging van ses elektrondonor- en elektronontrekkendegroepe op vier aromatiese verbindings, naamlike die karboksilaatgroep , stikstofdioksied , sulfonaat en fosfonaat. Dit moet ook genoem word dat stikstofdioksied nie 'n anioniese funksionele groep is nie, maar dit was wel ingesluit omdat dit ‘n neutrale radikaal groep is wat dikwels waterstofbindings vorm.

'n Totaal van 80 strukture optimiserings was uitgevoer met 'n kombinasie van die M06-2X en MP2 metodes wat gebruik maak van die 6-311++G(d,p) basisstel. Dit is gevolg deur interaksie-energie berekeninge op die M06-2X/6-311++G(d,p) vlak van teorie. Die resultate het getoon dat daar geen verband tussen die induktiewe vermoë van die substituente en die sterkte van die H-binding is nie, dit is eerder die vermoë van hierdie substituente om die anioniese funksionele groep te laat roteer wat toelaat dat koöperatiewe stabilisering van die H-binding kan geskied. Die AIM analise is op 'n gesubstitueerde H-binding struktuur toegepas. Die resultate het getoon dat elektrondonorgroepe wat by die *para* posisie geplaas word tot sterker H-bindings sal lei, wat weereens met koöperatiewe stabilisering vergesel word. Elektrononttrekkendegroepe met sterk induktiewe effekte kan tot 'n swakker H-binding lei indien hulle by die *meta* posisie geplaas word.

Die effek van water aktiwiteit (a_w) op hidraatkristalvorming is deur die uitvoering van 'n reeks kristallisasies in verskeie oplosmiddelmengsels ondersoek. Hierdie oplosmiddel mengsels bestaan uit water met asetoon, etanol of etielasetaat gemeng. Kristallisasies is vir drie organiese sure, naamlik piridien-4-karboksielsuur, N-amino-iminometiel-N-metielglisien monohidraat en 1,3,5-benseen tri-karboksielsuur uitgevoer. Daar is gevind dat water aktiwiteit 'n invloed op die vorming van die hidraat en watervrye produkte kan hê. Daarbenewens, speel water aktiwiteit 'n belangrike rol in die nukleasie fase van kristalvorming en kan as 'n effektiewe tegniek dien om kristalle van 'n toepaslike vorm en grootte vir verdere analise te verkry.

Acknowledgements

I would like to thank my two highly skilled supervisors, Prof. Catharine Esterhuysen and Prof. Delia Haynes for their continuous guidance and support during the tenure of my Masters. Thank you for always having an open door policy when require assistance with my work. I would also like to thank you for providing me with the necessary tools to perform the experimental and computational studies in this dissertation. Thank you for always having an open mind when it comes to studies I wanted to perform.

I would also like to thank my family and friends for their continuous support throughout the past two years. Thank you for always being interested in my work and the support you have provided me whenever I needed it most. I would also like to extend my thanks to the Supramolecular Materials Chemistry group at the University of Stellenbosch for their support during my MSc.

Finally, I would like to thank the University of Stellenbosch for providing me with the necessary facilities and the NRF for funding.

Publications and Posters

This work was presented as a poster at IYCr2014Africa held in Bloemfontein (13-17 Oct) entitled: *Hydrate formation in pharmaceutically relevant salts.*

Abbreviations

AIM	Atoms in Molecules
BCP	Bond Critical Point
BSSE	Basis set superposition error
cc	Correlation consistent
CC	Coupled Cluster
CP	Critical Points
CSD	Cambridge Structural Database
DDB	Dortmund Data Bank
DFT	Density Functional Theory
DSC	Differential Scanning Calorimetry
E_{el}	electrostatic energy contribution
E_{er}	exchange repulsion energy contribution
E_{ct}	charge transfer energy contribution
ECP	Electrostatic Core Potential
E_{disp}	dispersion energy contribution
E_{INT}	Hydrogen bond interaction energy
E_{pol}	polarization energy contribution
E_X	Exchange energy
GGA	Generalised Gradient Approximation
GTO	Gaussian Type Orbitals
HF	Hartree-Fock
H-bond	Hydrogen bond
IEF-PCM	Integral Equation Formalism Polarisable Continuum Model
MP2	Møller-Plesset 2nd order perturbation theory
ND	Neutron diffraction
NRTL	Non-Random Two-Liquid
LHP	Logit hydrogen-bonding propensity
LSD	Local-spin-density
LTR	Left-to-right
PES	Potential energy Surface
PCM	polarisable continuum model

QM	Quantum Mechanics
RTL	right-to-left
STO	Slater Type Orbitals
TGA	Thermogravimetric Analysis
TS	Transition state
WFT	Wave Function Theory
1D	1-dimensional
2D	2 dimensional
3D	3-dimensional

Table of Contents

1. Introduction 1
1.1. Hydrate formation of pharmaceutical salts 2
1.2. The Hydrogen Bond 3
1.3. Computational analysis 5
1.3.1 The Schrödinger equation in quantum mechanics 5
1.3.2. The Variational Principle 6
1.3.3. The Hartree-Fock (HF) approximation 7
1.3.4. Electron correlation 8
1.3.5. The electron density 8
1.3.6. Density Functional Theory 9
1.3.7. Basis sets 10
1.3.8. Implicit solvent model 11
1.3.9. Performance 12
1.3.10. Atoms in molecules 12
1.4. Aims 13
2. Methodology 16
2.1. Software 17
2.1.1. Cambridge Structure Database (CSD) searches 17
2.1.2. Neutron diffraction structures 18
2.1.3. Structure optimisation 19
2.1.4. Modelling in a solvent model 20
2.1.5. Visualisation of structures 23
2.1.6. H-bond strength in a solvent model 23
2.1.7. Stabilisation of the H-bond by the surrounding crystalline environment 24
2.1.8. Potential energy scans on CREATH04 25
2.1.9. Hydrogen bond propensities 27
2.1.10. Atoms in molecules (AIM) analysis 28
2.2. Experimental analytical techniques 28
2.2.1. Hydrate formation 28
2.2.2. Water activity 28
2.2.3. Crystallisations 30
2.2.4. Single-Crystal X-ray Diffraction Analysis 32
2.2.5. Thermogravimetric Analysis (TGA) 32
2.2.6. Differential Scanning Calorimetry (DSC) analysis 32
3. A CSD survey of the hydrate formation in salts with pharmaceutically acceptable anions and cations containing an NH group 38
3.1. Pharmaceutical salts in the CSD 39
3.2. CSD searches for hydrate occurrence in pharmaceutical anions 40
3.3. Other pharmaceutical anions 50
3.3.1. Nitrites 50
3.3.2. Phosphonate 50

3.3.3. Nitrogen dioxide 50
3.3.4. Nitroxide 51
3.4. CSD searches for hydrate occurrence in other pharmaceutical anions 51
4. Potential energy surfaces of creatine monohydrate 58
4.1. The properties and structure and of creatine monohydrate 59
4.2. The potential energy surface (PES) 61
4.3. PES scans of creatine monohydrate (PES) 61
4.4. Potential energy surfaces (PESs) curves of CREATH04 63
4.5. The horizontal, left-to-right diagonal and right-to-left diagonal PES scan 66
4.6. Comparing the accuracy of various methods 71
4.7. Optimising the minimum geometries of each scan 73
5. Modelling of pharmaceutical salts in the gas phase and in an implicit solvent model 77
5.1. Method for geometry optimisation 78
5.2. CSD searches for H-bonds in different functional groups 81
5.3. Geometry optimisation 82
5.4. The H-bond strength with different functional groups 83
5.5. Accuracy of DFT methods reproducing WFT results 87
5.6. The effect of steric density on H-bond interaction strength 88
5.7. The effect of the crystalline environment on the H-bond interaction energy 92
5.8. Interaction energy calculations in implicit polarisable continuum 95
6. The effect of aromatic ring substituents on hydrogen bond strength 103
6.1. Geometry optimisation calculations of four aromatic structures 104
6.2. The inductive effect of substituents on H-bond energies 106
The effect of aromatic substituents on the H-bond in compounds that contain:
6.2.1. The carboxylate functional group 107
6.2.2. The nitrogen dioxide functional group 109
6.2.3. The sulfonate functional group 111
6.2.4. The phosphonate functional group 113
6.3. AIM analysis of structure C7 116
7. The effect of water activity on hydrate formation 122
7.1. Water activity 123
7.2. Crystal formation from organic solvent mixtures 125
7.2.1. Creatine monohydrate 125
7.2.2. Anhydrous creatine 127
7.2.3. Isonicotinic acid 128
7.2.4. Trimesic acid dihydrate 130
8. Conclusions and future work 134
8.1. Conclusions 135
8.2. Future work 138
Appendix A – Supplementary information

CHAPTER 1

Introduction

1.1. Hydrate formation in pharmaceutical salts

Pharmaceutical compounds can exist in both anhydrous and hydrated forms, where the presence of water within the crystal lattice generally results in a completely different packing arrangement for the hydrates. The formation of hydrates is of particular interest for pharmaceutical compounds as these may exhibit improved stability, solubility, hygroscopicity, crystallinity and bioavailability as opposed to the dehydrated form[1-4].

The modern development of valuable drug compounds has been posed with the problem of finding an appropriate method for administering these medicines. A solution to the physico-chemical and solid-state stability of certain drugs is the formation of pharmaceutical salts. Pharmaceutical salts generally consist of an active pharmaceutical ingredient (API), which is the component of the drug that is active. Salt formation offers several other advantages that have led to an estimated 50% of currently available drugs being administered as salts[5]. Firstly, they form more reliable products that are chemically stable and easy to recrystallize[2]. Secondly, problems with polymorphism can be avoided, and working from a synthetic point of view with salts simplifies the separation of intermediates from by-products. Finally, salts are of particular use in therapeutic applications by providing improved physico-chemical properties, dissolution rate and solubility in water. The solubility of pharmaceutical compounds in water is often very poor for pharmaceuticals in the neutral state. The formation of a pharmaceutical salt that is very soluble in water will increase a compound's chance of reaching the biological target[6]. This was shown in a study by Stephenson, who reported a 2000-fold increase in solubility of a salt product, which is most probably due to hydration energy of the charged species[7].

It is therefore important to understand the role of hydrate formation in pharmaceutical salts. A number of factors affecting hydrate formation have been identified[8-11], namely water activity (a_w), different pharmaceutically acceptable anion groups, the number of donor and acceptor atoms in a chemical system and the polar surface in the chemical system. These factors will be discussed below.

Herman *et al.* investigated the role of water activity (a_w), relative humidity (RH) and temperature on the hydration state of theophylline[8]. The results from their study showed that the hydrated or anhydrous product could be obtained at different water activity values. In a more recent study, Haynes *et al.* performed a CSD survey to investigate the propensity of pharmaceutically active ingredients to form hydrates[9]. It was found that many such compounds can be found in the hydrated form, but it was still unclear what factors influenced the actual process of introducing a water molecule into the crystal structure. In a related study, Briggs *et al.* investigated how the number of donor and acceptor atoms in a chemical system affects hydrate formation. This was done by taking various structures with a different number of donor and acceptor atoms and measuring the propensity of these structures to form hydrates. Their results showed that hydrate formation is likely when there is a high number of potential hydrogen bond acceptor and donor atoms, as well as a low ratio of potential hydrogen bond donor to acceptor atoms[11]. Contrary to these findings, Infantes *et al.*

suggested that it was not the donor-acceptor ratio, but the increase in polar surface area that can be correlated to increasing hydrate formation[10]. In all these studies a strong, directional H-bond was present in the hydrates that were studied. A further study by Infantes *et al.* found that hydrate formation in the absence of an H-bond is very rare[10]. Despite this large amount of experimental work that has been done to determine the factors that influence hydrate formation, no theoretical investigation has been conducted to establish how all of these factors influence the strength of the H-bond, and hence the propensity for pharmaceutically relevant salts to form hydrates.

1.2. The Hydrogen Bond

The hydrogen bond (H-bond) has been a topic of scientific research for more than a hundred years[12]. This prolonged interest is mainly due to the importance of H-bonds for structure, function, and dynamics of various chemical systems. These systems fall into a diverse number of scientific branches, which include mineralogy, materials science, general inorganic and organic chemistry, supramolecular chemistry, biochemistry, molecular medicine, and pharmacy. Studying the role of the H-bond in hydrate formation has been made easier by advances in crystallographic software and crystallographic databases containing vast amounts of structural information, such as the Cambridge Structural Database (CSD)[13]. The CSD is the largest source of crystallographic data on small organic and organometallic crystal structures obtained by X-ray and neutron diffraction, which provides the means to study intermolecular interactions such as the H-bond.

The classical view of the H-bond can be explained by the directional interaction between two water molecules.

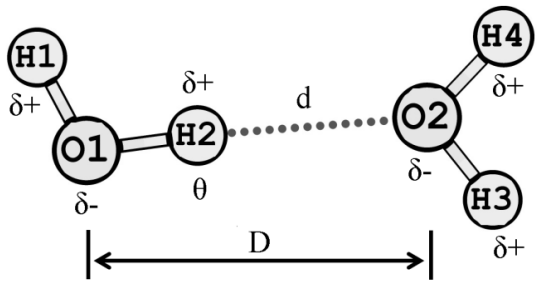

Figure 1.1. Schematic of the H-bond between the two molecules in a water dimer. It should be noted that the two water molecules are not coplanar.

Figure 1.1 illustrates the geometrical parameters of the H-bond between two water molecules, which are: d = H2•••O2 distance, D = O1•••O2 distance and θ = O1–H2•••O2 angle. The large difference in electronegativity between the H and O atoms results in inherently polar water molecules, with partial charges of +0.4 on each H atom and -0.8 on each O atom[14]. This causes the local dipoles of neighbouring water molecules to align in a specific way that allows overlap of the electron orbitals to form a three-centre four-electron bond, which is 1 Å shorter than the sum of the van der Waals radii for the H and O atoms[15]. This definition of the H-bond between water molecules can be extended with a few modifications to the

analogous interactions X–H•••A. Here there are strongly polar groups $X^{\delta-}$–$H^{\delta+}$ on one side and typically electronegative atoms $A^{\delta-}$ on the other (X = O, N, halogen; A = O, N, S, halide, etc.)[14]. The H-bond can have long-range interactions, which allows a single donor to interact with up to three acceptors simultaneously (Figure 1.2).

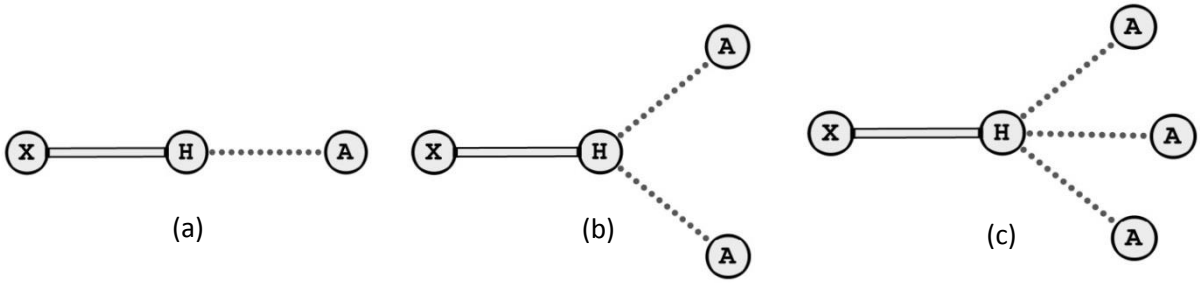

Figure 1.2. Illustration of the different types of hydrogen bonding: a single hydrogen bond (a), bifurcated hydrogen bond (b) and trifurcated hydrogen bond (c).

Modern studies have suggested that the H-bond is much more complex than described above. For instance, strong H-bonds with properties similar to covalent bonds exist, and yet there are very weak H-bonds that can hardly be distinguished from van der Waals interactions. The hydrogen bond is a complex interaction that consists of several different constituents[16].

The large variation in H-bond properties can be explained by the different contributions to the total energy of the H-bond, namely electrostatic (E_{el}), polarization (E_{pol}), charge transfer (E_{ct}), dispersion (E_{disp}), and exchange repulsion (E_{er})[17,18]. These components all have different characteristics, which lead to intermolecular interactions having certain geometric characteristics depending on the largest contributors. For instance, the electrostatic term is of a directional and long range nature and will decrease at rates of r^{-3} for dipole-dipole and r^{-2} for dipole-monopole interactions, where the r symbol represents the distance between the H-bond donor and acceptor atoms. The polarization term decreases much faster (r^{-4}), while the charge-transfer term decreases even faster (e^{-r})[19]. The dispersion term decreases as r^{-6}, while the exchange-repulsion term increases sharply as the distance decreases (as r^{-12}). The dispersion and exchange-repulsion terms are often combined into a "van der Waals (vdW)" contribution. This contribution can be described by the Lennard-Jones potential with a distance dependence of $E_{vdW} \sim A\ r^{-12} - B\ r^{-6}$, where A and B are empirically determined constants[14].

A CSD survey performed by Desiraju and Steiner to investigate the difference between strong and weak H-bonds[20] suggested that the fundamental difference between a weak and strong H-bond is that a strong H-bond is very directional and tends to favour a linear geometry ($\theta = \pm 180°$), while a weak H-bond is fairly isotropic and independent of the angle θ. The strengths of H-bonds range from 0.2 kcal/mol to 40 kcal/mol[21,22], which is largely influenced by the environment that surrounds the interacting molecules[14].

The factors that resulting in the formation of a strong or weak H-bond may be determined with the aid of a wide variety of analytical and computational techniques.

1.3. Computational analysis

Various computational software packages exist that allow electronic structure modelling. Gaussian 09 is a powerful computational tool that was used in this study in order to calculate energies, molecular structures, vibrational frequencies and molecular properties for molecules and ions in a wide variety of chemical environments using a range of computational methods[23]. In particular, the Density Functional Theory (DFT) and Wave Function Theory (WFT) methods were used in this study. These methods will be briefly discussed here based on the more in-depth description in reference[24].

1.3.1. The Schrödinger equation in quantum mechanics

The quantum chemical approach to explaining a chemical system involves approximating a solution for the time-independent, non-relativistic Schrodinger equation

$$\hat{H}\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_N,\vec{R}_1,\vec{R}_2,\ldots,\vec{R}_M) = E_i\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_N,\vec{R}_1,\vec{R}_2,\ldots,\vec{R}_M) \qquad \text{(eq. 1)}$$

where $\hat{H}$ is the Hamiltonian operator for a molecular systems consisting of M nuclei and N electrons in the absence of a magnetic or electronic field. Furthermore, $\hat{H}$ can be described by the sum of the kinetic and potential energies of all particles in the system:

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \quad \text{(eq. 2)}$$

Kinetic energy component of electrons (*N*) and nuclei (*M*)

Potential energy component consisting of electrostatic electron-nucleus (*N-M*), electron-electron (*N-N*) and nuclei-nuclei (*M-M*) interactions

Here, A and B cover nuclei *M*, while *i* and *j* represent the *N* electrons in the system. M_A is the mass of nucleus A (described as a product of the mass of an electron), defined in terms of atomic units to simplify the system. The ∇_q^2 (where $q = i$ or A) symbol in the two kinetic components represents the Laplacian operator, which can be defined as the sum of differential operators in terms of Cartesian coordinates

$$\nabla_q^2 = \frac{\partial^2}{\partial x_q^2} + \frac{\partial^2}{\partial y_q^2} + \frac{\partial^2}{\partial z_q^2}. \qquad \text{(eq. 3)}$$

The remaining three terms describe the attractive and repulsive potential of the Hamiltonian. r_{ij} (similarly R_{ij}) is the distance between two particles, A and B.

$\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_N,\vec{R}_1,\vec{R}_2,\ldots,\vec{R}_M)$ represents the wave function of the i'th state, which is dependent on the 3N spatial coordinates $\{\vec{r}_i\}$, and the N spin coordinates $\{s_i\}$ of all electrons, which are collectively termed $\{\vec{x}_i\}$ and the 3M spatial coordinates of the nuclei, $\{\vec{R}_I\}$. Lastly, E_i is the numerical value of the energy of a system that is described by Ψ_i. From the

explanation above it is evident that the wave function Ψ_i will be able to describe all of the information about a quantum system.

We can further simplify the system by applying the Born-Oppenheimer or clamped-nuclei approximation. Simply put, the approximation fixes the positions of the nuclei. This results in a nuclear kinetic energy of zero, while the potential energy due to nucleus-nucleus repulsion reduces to a constant. This reduces equation 2 to the electronic Hamiltonian:

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee}. \quad \text{(eq. 4)}$$

The solution to the Schrodinger equation with the $\hat{H}_{elec}$ operator is the electronic wave function and the electronic energy E_{elec}. The nuclear repulsion term is added to E_{elec} to give E_{Tot}:

$$\hat{H}_{elec}\Psi_{\text{elec}} = E_{\text{elec}}\Psi_{\text{elec}} \quad \text{(eq. 5)}$$

and

$$E_{\text{tot}} = E_{\text{elec}} + E_{\text{nuc}}. \quad \text{(eq. 6)}$$

The wave function Ψ itself is not an observable. Therefore a physical interpretation can only be achieved with the square of the wave function so that

$$\int|\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_N)|^2 d\vec{x}_1, d\vec{x}_2,\ldots,d\vec{x}_N \quad \text{(eq. 7)}$$

represents the probability that electrons 1, 2, … , N are found simultaneously in volume elements $d\vec{x}_1, d\vec{x}_2,\ldots,d\vec{x}_N$. Furthermore, the probability of finding electrons must not change if the coordinates of two electrons (i and j) are switched, namely

$$\left|\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_i,\vec{x}_j,\ldots,\vec{x}_N)\right|^2 = \left|\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_j,\vec{x}_i,\ldots,\vec{x}_N)\right|^2. \quad \text{(eq. 8)}$$

The result of the probability interpretation of the wave function is that the integral of equation 7 over the full range of all variables will equal 1:

$$\int\ldots\int\left|\Psi_i(\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_i,\vec{x}_j,\ldots,\vec{x}_N)\right|^2 d\vec{x}_1, d\vec{x}_2,\ldots,d\vec{x}_N = 1. \quad \text{(eq. 9)}$$

If a wave function satisfies equation 9 it is said to be normalised. For the systems in our study we deal exclusively with normalised wave functions.

1.3.2. The variational principle

The first step to finding a solution to the Schrödinger equation (eq. 5) for any arbitrary molecule is to know the parts of the Hamiltonian operator that are specific for the system at hand. It is evident from equation 4 that the only information that is dependent on the molecule is the number of electrons (N) and the $\hat{V}_{Ne}$ term, which is also termed the external potential (V_{ext}) in DFT. All of the other parts of the Hamiltonian, such as the operators representing the kinetic energy or the electron-electron repulsion, are independent of the

molecule. The second step to solving the Schrödinger equation is to obtain the eigenfunctions Ψ_i and corresponding eigenvalues E_i of $\hat{H}$. All of the properties of a particular system can be obtained by applying appropriate operators on the eigenfunctions Ψ_i. Determining the wave function of the ground state Ψ_0 (state with lowest energy E_0) can be systematically approached by applying the variational principle. This method makes use of a normalised trial wave function Ψ_{trial} and an appropriate operator $\hat{O}$ to yield energy E_{trial}. This implies that $E_{trial} > E_0$. This method allows us to approach the ground state wave function, resulting in the lowest energy for a molecule, which is mathematically described as:

$$\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle = E_{trial} \geq E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle. \qquad \text{(eq. 10)}$$

1.3.3. The Hartree-Fock (HF) approximation

The HF approximation forms an integral part of wave-function based quantum chemical methods. The HF method approximates the N-electron wave functions by taking the antisymmetrized product of N one-electron wave functions $\chi_i(\vec{x}_i)$. This product is referred to as the Slater determinant, Φ_{SD}:

$$\Psi_0 \approx \Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{bmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & \cdots & \chi_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & \cdots & \chi_N(\vec{x}_N) \end{bmatrix}. \qquad \text{(eq. 11)}$$

Equation 11 can be reduced to a short-hand notation, where only the diagonal elements are given:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} det\{\chi_1(\vec{x}_1)\, \chi_2(\vec{x}_1) \cdots \chi_N(\vec{x}_N)\}. \qquad \text{(eq. 12)}$$

The one-electron functions $\chi_i(\vec{x}_i)$ are called *spin orbitals*, which are composed of a spatial orbital $\phi_i(\vec{r})$ and one of the two spin functions, α(s) or β(s):

$$\chi(\vec{x}) = \phi_i(\vec{r})\, \sigma(s), \sigma = \alpha, \beta. \qquad \text{(eq. 13)}$$

The spin functions are orthonormal, for example $\langle\alpha|\alpha\rangle = \langle\beta|\beta\rangle = 1$ and $\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle = 0$.

In the Hartree-Fock approach the spin orbitals must be orthonormal in order to obtain a minimum energy from the corresponding Slater determinant:

$$E_{HF=} \min_{\Phi_{SD \to N}} E[\Phi_{SD}]. \qquad \text{(eq. 14)}$$

The Hamiltonian operator with a Slater determinant can be derived by expanding equation 14 to the individual terms with respect to the various parts in the Hamiltonian.

The HF energy is given by

$$E_{HF=}\langle\Phi_{SD}|\hat{H}|\Phi_{SD}\rangle = \sum_i^N(\,\mathrm{i}|\,\hat{h}\,|\mathrm{i}) + \frac{1}{2}\sum_i^N\sum_j^N(ii\,|jj) - (ij\,|ji) \qquad \text{(eq. 15)}$$

which is a function of the spin orbitals, $E_{HF} = E[\chi_i]$. In order for the function to remain orthonormal, Lagrangian multipliers ε_i are introduced.

$$\hat{f}\chi_i = \varepsilon_i\chi_i,\ \mathrm{i} = 1,2,\ldots,N. \qquad \text{(eq. 16)}$$

The Fock one-electron operator, $\hat{f}$, is defined as

$$\hat{f} = -\frac{1}{2}\nabla_i^2 - \sum_A^M\frac{Z_A}{r_{iA}} + V_{HF}(i) \qquad \text{(eq. 17)}$$

where the first two terms are the kinetic energy and the potential energy and $V_{HF}(i)$ is the *Hartree-Fock potential*. This equation is defined as a pseudo-eigenvalue problem that has to be worked out iteratively. The technique used to achieve this is called the self-consistent field procedure (SCF), due to the orbitals being derived from their own effective potential. This technique makes use of a 'guessed' set of orbitals, which is used to solve the HF equations. The resulting orbitals are then used in the next iteration and this cycle is repeated.

1.3.4. Electron correlation

The Slater determinant can describe a significant portion of the physics of a many electron system, but it never corresponds to the exact wave function. This results in the E_{HF} always being larger than the exact ground energy E_0. The difference between these two energies is defined as the correlation energy

$$E_C^{HF} = E_0 - E_{HF}. \qquad \text{(eq. 18)}$$

E_C^{HF} is a negative quantity because E_0 and $E_{HF} < 0$ and $|E_0| > |E_{HF}|$. It is also a measure of the error of the HF approximation. This may be large, as electron correlation is mainly due to the repulsion of electrons, which is not explained by the effective HF potential.

1.3.5. The electron density

The value obtained using equation 7 of finding any of the N electrons within a volume can be defined as electron density $\rho(\vec{r})$. Unlike the wave function, the electron density can be measured by certain experimental techniques, for example X-ray diffraction.

1.3.6. Density Functional Theory (DFT)

The conventional approach to quantum chemistry is to use the wave function to calculate certain information regarding a particular system. The Hartree-Fock approximation is an approach to calculating the wave function, from which the required information can be obtained. The problem with the wave function is that it is a very complicated quantity that depends on 4N variables: three spatial and one spin orbital for each of the N electrons. Some systems contain a very large number of atoms and many more electrons, which must be described by a wave function of an enormous size. The operators in the Hamiltonian of the Schrödinger equation can be rewritten in terms of reduced one and two-particle density matrices. This results in an equation that depends on 8, rather than 4N variables. The electron density $\rho(\vec{r})$ can now be used to reach a solution to the Schrödinger equation and thus obtain the energy and properties of the system. This approach was described by the Thomas-Fermi model (T_{TF}) and is based on a quantum statistical model of electrons, which only takes kinetic energy into account while treating the nuclear-electron and electron-electron contributions in a classical way, namely,

$$T_{TF}\,[\rho(\vec{r})] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r})d\vec{r}, \qquad \text{(eq. 19)}$$

which, combined with the classical expression for the nuclear-electron attractive potential and electron-electron repulsive potential, gives the Thomas-Fermi expression for the energy of an atom,

$$E_{TF}\,[\rho(\vec{r})] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho^{5/3}(\vec{r})d\vec{r} - Z \int \frac{\rho(\vec{r})}{r} + \frac{1}{2}\int\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}}\, d\vec{r}_1 d\vec{r}_2. \quad \text{(eq. 20)}$$

This equation is capable of giving the energy of an atom in terms of electron density $\rho(\vec{r})$. Furthermore, although equation 20 gives a reasonable approximation of the kinetic energy, the exchange and correlation effects are completely neglected. More modern density functional theory methods make use of the Hohenberg-Kohn theory and the Kohn-Sham theory. To quote the landmark paper in 1964 by Hohenberg and Kohn[25]: '*the external potential $V_{ext}(\vec{r})$ is (to within a constant) a unique functional of $\rho(\vec{r})$; since, in turn $V_{ext}(\vec{r})$ fixes $\hat{H}$ we see that the full many-particle ground state is a unique functional of $\rho(\vec{r})$*."

The Kohn-Sham theory uses self-consistent equations that are analogous to the Hartree-Fock equations. In these equations the exchange and correlation portions of the chemical potential appear as additional effective potentials. The energy from the exchange portion is defined as the Fermi hole function, which is the hole in the probability density of electrons and applies to electrons with the same spin. The exchange energy accounts for interactions that are electrostatic in nature. The correlation energy is defined by the Coulomb hole functions. The Coulomb hole refers to electrons that interact with antiparallel spin, which results in the Coulomb hole bearing no charge. The correlation energy accounts for interactions that are of a weaker nature, such as van der Waals type interactions.

When using DFT to perform calculations, the exchange and correlation parts of the functional are chosen independently of each other, for example BLYP, a very popular DFT method where the B method describes the exchange energy and the LYP method describes the correlation energy. Several additional variations of the DFT method have been developed, each with their own benefits to specific systems, however these will not be discussed here.

1.3.7. Basis sets

Sets of functions, known as basis sets, are used to describe the orbitals, χ_i, of atoms, which in turn are combined to construct the appropriate wave function. Very large basis sets are required if high quality wave functions that take electron correlation into account are the target, since they provide an accurate description of the system. However, large basis sets can dramatically increase the resources required to perform calculations. In this section we will give an overview of the different kinds of basis sets that are used in modern computational chemistry.

For the Hartree-Fock wave function approach, the so-called Cartesian Gaussian-type-orbitals (GTO) are used in the general form

$$\eta^{GTO} = N\, x^l\, y^m\, z^n \exp[-\alpha r^2]. \qquad \text{(eq. 20)}$$

N is a normalization factor, α represents the orbital exponent which determines how compact (large α) or diffuse (small α) the resulting function is. The exponents $L = l + m + n$ are used to classify the GTO as *s* functions ($L = 0$), *p* functions ($L = 1$), *d* functions ($L = 2$), etc. Slater type orbitals (STO) contain simple exponentials that represent the eigenfunctions of the hydrogen atom. Unlike the GTO functions, STOs exhibit accurate results for $r \to 0$ and when $r \to \infty$. The STO can be expressed as

$$\eta^{STO} = N\, r^{n-1} \exp[-\zeta r]\, Y_{lm}(\Theta, \phi). \qquad \text{(eq. 21)}$$

Here, *n* corresponds to the principal quantum number, the orbital exponent is represented by ζ and Y_{lm} are the usual spherical harmonics that describe the angular part of the function. One usually needs about three times as many GTO as STO functions to achieve the same accuracy. The smallest number of functions that is required to describe the electrons of each atom is defined as the minimum basis set, which consists of a single basis function for each orbital, for example, one s-function will be used for a hydrogen atom. An improvement on this is the double-zeta basis set, which will improve the accuracy by having two basis functions that represent each orbital. An improvement on the double-zeta is the triple-zeta, which will further increase the accuracy by having three basis functions that represent each orbital.

Pople-style basis sets were employed in this study. The notation for Pople-style basis sets of the so-called split-valence type is given in the X-YZg style[26]. For split-valence type orbitals, the valence orbitals are represented by more than one basis function. Here the X symbol is the the number of Gaussian-type orbitals that represent the core electrons. The Y and Z symbols indicate that the valence orbitals are composed of two basis functions each, the first

one composed of a linear combination of Y primitive Gaussian functions, the other composed of a linear combination of Z primitive Gaussian functions. Two numbers after the hyphen implies that the basis set is a split-valence double-zeta basis set. Three or four numbers after the hyphen are split-valence triple- and quadruple-zeta basis sets respectively. Diffuse functions can also be added as indicated by the plus sign, +. Two plus signs indicate that diffuse functions are also added to light atoms (hydrogen and helium). If atoms larger than krypton are considered, effective core potential (ECP) or pseudopotential-type basis sets are employed in order to decrease the computational cost. Pseudopotential-type and ECP basis sets can be defined as basis sets that that make use of an approximation for the simplified description of complex systems by replacing the core electrons with an approximate potential.

Correlation consistent (cc) basis sets for larger atoms were designed by Dunning in 2003[27]. These basis sets contain pseudopotential functionals and have ECPs for larger atoms to recover a large portion of the correlation energy of the valence electrons. Dunning type basis sets can also be augmented with additional diffuse functions (indicated with the aug- prefix) and are very important for anions that have loosely bound electrons. The pseudopotential functional is given by:

$$V_{PP}(r) = -\frac{Q}{r} + \sum_{ljk} B_{lj}^{k} \exp(-\beta_{lj}^{k} r^2)\, P_{lj} \qquad \text{(eq. 22)}$$

where Q is the inner-core charge and the sum is over the Gaussian expansion (index *k*) of semilocal short-range radial potentials, which are different for different orbital angular momentum quantum numbers *l*. For the given value of *l*, the *j* value is $\pm\frac{1}{2}$. The P_{lj} symbol is defined as the projector onto the complete space of functions with angular symmetry *l*,*j* around a core under investigation. B_{lj}^{k} and β_{lj}^{k} are two constants that are adjusted by a least-squares fit so that V_{PP} yields results as close as possible to the all-electron reference data[28].

1.3.8. Implicit solvent model

There are two main types of solvent models employed in computational calculations, which are the continuum solvent model and the discrete solvent model. In this study the continuum solvent model was employed. This model simulates a solvent medium by placing the solute in a cavity surrounded by a uniform non-directional distribution of point charges. The polarisability of each solvent is indicated by the relative permittivity (ε). Furthermore, there are several variations in the size and reaction field of the cavity. In the current study the Integral Equation Formalism Polarisable Continuum Model (IEF-PCM) was used, which describes the cavity using overlapping spheres[29].

1.3.9. Performance

In our study of pharmaceutically relevant salts, we are particularly interested in the formation of H-bonds in hydrated structures. The performance of several DFT and *ab initio* methods was investigated by measuring their ability to perform geometry optimisation as well as calculate intermolecular interaction energies. All of these calculations were initially performed at a very high level of theory that would serve as a benchmark for the particular study, as more expensive methods will provide more accurate results at the expense of a significant increase in the requirement of computer resources. It is therefore important to select an appropriate method that can be used within the scope of this study. The coupled cluster with single, double and triple excitations [CCSD(T)] method is often called the "gold standard" of computational chemistry[30]. Performing all of the calculation in this study at the CCSD(T) level of theory was however not within the scope of this study and the MP2/aug-cc-pVTZ level of theory often served as the benchmark level of theory for our study.

1.3.10. Atoms in Molecules

The theory of Atoms in Molecules (AIM) was described in a paper published by Bader in 1985[31]. This model is based on the ability of a quantum mechanical state function to provide the energy of a molecular structure as well as describe areas where charge density ρ is locally concentrated and depleted. The gradient vectors of ρ provide both qualitative and quantitative understanding of the topology of ρ, where the gradient vector (∇ρ) can be defined as the first derivative of the charge density and describes the change in charge density. Where the scalar product of a gradient vector and a unit vector of the electron density equals zero, it defines a zero flux surface[32]. The zero flux surface partitions the charge distribution of the system (molecules) into atoms (subsystems). Any point on the zero flux surface where ∇ρ = 0 is defined as a critical point (3, -1), where the 3 indicates the rank/dimension of the critical point [i.e. 3-dimensional (3D)] and -1 indicates that the point is at a 2-dimensional (2D) maximum and 1-dimensional (1D) minimum. The -1 can be explained by the minima and maxima in each dimension. This specific example has two minima and one maximum, corresponding to the simple calculation: -1 + -1 + 1 = -1. There are four different 3D critical points[31], namely:

(3, -3)	All curvatures of the electron density are negative indicating a local 3D maximum;
(3, -1)	Two curvatures are negative, indicating a 2D-maximum, while one is positive and showing a 1D minimum perpendicular to the associated plane defined by the negative curvatures;
(3, +1)	Two curvatures are positive, *i.e.* a 2D-minimum while one curvature is negative at a 1D maximum;
(3, +3)	All three curvatures are positive at a local 3D-minimum.

Nuclei can be described as 3D-maxima (3, -3), as the electron density decreases as one moves away from the nucleus. A bond is defined as a pathway of maximum electron density

between two nuclei is called a bond path. A local minimum between the bonded nuclei is defined as a bond critical point (BCP). The characterisation of the nature by which atoms are bonded atoms can be done by investigating the electronic properties at the BCPs[33].

The second derivative of the charge density ($\nabla^2\rho$) is defined as the Laplacian. The Laplacian indicates the regions where electron density is locally concentrated and depleted, and is given by:

$$L(\boldsymbol{r}) = \nabla^2\rho(\boldsymbol{r}).$$

This equation is powerful since at $\nabla^2\rho(\boldsymbol{r}) < 0$, the charge density is concentrated at a point *r*, and when $\nabla^2\rho(\boldsymbol{r}) > 0$, the charge density is depleted. The reaction of a nucleophile with an electrophile can be described as a centre of charge density concentration on one atom that reacts with the centre of charge depletion of a second atom. This reaction can also be explained as the reaction of a region with excess potential energy with another of excess kinetic energy. These charge density studies can be used to link the electronic distribution with the chemical properties of a molecule. For our particular study, the properties of the H-bond were investigated by studying the position of the BCPs and bond paths linking the nuclei involved in the H-bond[34].

1.4. Aims

Although several techniques have been employed in order to determine what drives hydrate formation, a systematic study that specifically focuses on the role of the H-bond in hydrate formation has, to the extent of our knowledge, not been performed. Therefore the aim of this study is to understand the role of the H-bond in hydrate formation. A further aim is to understand what factors influence the strength of H-bonds and if they can be used to fine-tune the strength of H-bonds when a pharmaceutical hydrate compound is the target compound.

A CSD survey of the hydrate propensity of pharmaceutically relevant salts is described in Chapter 2. In Chapter 3 the various methodologies will be described. The results from a series of different potential energy surface scans that investigate the chemical environment surrounding the H-bond in N-carbamimidoyl-N-methylglycine will be discussed in Chapter 4.

The modelling of various pharmaceutically relevant anions utilising various methods and basis sets in the gas phase and in several solvents will be presented in Chapter 5. Furthermore, the effect of the lattice and the role of steric density will also be presented. A suitable level of theory was identified by comparing the accuracy of several methods to results obtained at the MP2/aug-cc-pVTZ benchmark. The effect of adding substituents of varying inductive strengths to aromatic structures will be discussed in Chapter 6. AIM analysis of the aromatic substituent effect will also be presented in Chapter 6. The effect of water activity (a_w) on the crystal hydrate formation of three organic acids in a series of water-organic solvent mixtures is discussed in Chapter 7. In the final chapter, we summarise our results and possible suggestions for future work.

References

(1) Vippagunta, S. R., Brittain, H. G., Grant, D. J. W.: Crystalline solids. *Advanced Drug Delivery Reviews* **2001**, *48*, 3-26.

(2) Stahl, P. H., Wermuth, C. G.: Handbook of Pharmaceutical Salts: Properties, Selection and Use. Zürich, *Wiley* **2002**.

(3) Aaltonen, J., Allesø, M., Mirza, S., Koradia, V., Gordon, K. C., Rantanen, J.: Solid form screening–a review. *European Journal of Pharmaceutics and Biopharmaceutics* **2009**, *71*, 23-37.

(4) Shimanovich, R., Cooke, M., Peterson, M. L.: A rapid approach to the preliminary assessment of the physical stability of pharmaceutical hydrates. *Journal of Pharmaceutical Science* **2012**, *10*, 4013-4017.

(5) Sarma, B., Chen, J., Hsi, H.-Y., Myerson, A.: Solid forms of pharmaceuticals: Polymorphs, salts and cocrystals. *Korean Journal of Chemical Engineering* **2011**, *28*, 315-322.

(6) Tilborg, A., Norberg, B., Wouters, J.: Pharmaceutical salts and cocrystals involving amino acids: A brief structural overview of the state-of-art. *European Journal of Medicinal Chemistry* **2014**, *74*, 411-426.

(7) Stephenson, G. A., Aburub, A., Woods, T. A.: Physical stability of salts of weak bases in the solid-state. *Journal of Pharmaceutical Sciences* **2011**, *100*, 1607-1617.

(8) Herman, J., Remon, J. P., Visavarungroj, N., Schwartz, J. B., Klinger, G. H.: Formation of theophylline monohydrate during the pelletisation of microcrystalline cellulose-anhydrous theophylline blends. *International Journal of Pharmaceutics* **1988**, *42*, 15-18.

(9) Haynes, D. A., Jones, W., Motherwell, W. D. S.: Occurrence of pharmaceutically acceptable anions and cations in the Cambridge Structural Database. *CrystEngComm* **2005**, *7*, 342-345.

(10) Infantes, L., Fabian, L., Motherwell, W. D. S.: Organic crystal hydrates: what are the important factors for formation. *CrystEngComm* **2007**, *9*, 65-71.

(11) Briggs, N. E. B., Kennedy, A. R., Morrison, C. A.: 42 salt forms of tyramine: structural comparison and the occurrence of hydrate formation. *Acta Crystallographica Section B* **2012**, *68*, 453-464.

(12) L. Pauling: The Nature of the Chemical Bond and the Structure of Molecules and Crystals - An Introduction to Modern Structural Chemistry. Oxford, *Oxford University Press* **1940.**

(13) Allen, F.: The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B* **2002**, *58*, 380-388.

(14) Steiner, T.: The Hydrogen Bond in the Solid State. Weinheim, *Angewandte Chemie International Edition* **2002**, *41*, 48-76.

(15) Rowland, R. S., Taylor, R.: Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *The Journal of Physical Chemistry* **1996**, *100*, 7384-7391.

(16) Morokuma, K.: Why do molecules interact? The origin of electron donor-acceptor complexes, hydrogen bonding and proton affinity. *Accounts of Chemical Research* **1977**, *10*, 294-300.

(17) Scheiner, S.: Hydrogen Bonding. A Theoretical Perspective. Oxford, *Oxford University Press* **1997**.

(18) Jeffrey, G. A.: An Introduction to Hydrogen Bonding. *Oxford University Press* **1997**.

(19) Reed, A. E., Curtiss, L. A., Weinhold, F.: Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chemical Reviews* **1988**, *88*, 899-926.

(20) Steiner, T. R., Desiraju, G.: Distinction between the weak hydrogen bond and the van der Waals interaction. *Chemical Communications* **1998**, 891-892.

(21) Gronert, S.: Theoretical studies of proton transfers. 1. The potential energy surfaces of the identity reactions of the first- and second-row non-metal hydrides with their conjugate bases. *Journal of the American Chemical Society* **1993**, *115*, 10258-10266.

(22) Nieschalk, J., Batsanov, A. S., O'Hagan, D., Howard, J.: Synthesis of monofluoro- and difluoro- methylenephosphonate analogues of sn-glycerol-3-phosphate as substrates for glycerol-3-phosphate dehydrogenase and the X-ray structure of the fluoromethylenephosphonate moiety. *Tetrahedron* **1996**, *52*, 165-176.

(23) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, M. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

(24) Koch W. H. M.: A Chemist's guide to Density Functional Theory. Weinheim, *Wiley* **2000**.

(25) Hohenberg, P., Kohn, W.: Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864-B871.

(26) Ditchfield, R., Hehre, W. J., Pople, J. A.: Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **1971**, *54*, 724-728.

(27) Dunning, T. H.: Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007-1023.

(28) Peterson, K. A., Shepler, B. C., Figgen, D., Stoll, H.: On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions. *The Journal of Physical Chemistry A* **2006**, *110*, 13877-13883.

(29) Marenich, A. V., Olson, R. M., Kelly, C. P., Cramer, C. J., Truhlar, D. G.: Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011-2033.

(30) Řezáč, J., Šimová, L., Hobza, P.: CCSD[T] Describes Noncovalent Interactions Better than the CCSD(T), CCSD(TQ), and CCSDT Methods. *Journal of Chemical Theory and Computation* **2012**, *9*, 364-369.

(31) Bader, R. F. W.: Atoms in molecules. *Accounts of Chemical Research* **1985**, *18*, 9-15.
(32) Bader, R. F. W.: Molecular fragments or chemical bonds. *Accounts of Chemical Research* **1975**, *8*, 34-40.
(33) Kirby, I. L., Brightwell, M., Pitak, M. B., Wilson, C., Coles, S. J., Gale, P. A.: Systematic experimental charge density analysis of anion receptor complexes. *Physical Chemistry Chemical Physics* **2014**, *16*, 10943-10958.
(34) Koch, U., Popelier, P. L. A.: Characterization of C-H-O Hydrogen Bonds on the Basis of the Charge Density. *The Journal of Physical Chemistry* **1995**, *99*, 9747-9754.

CHAPTER 2

Methodology

2.1. Software

2.1.1. Cambridge Structure Database (CSD) searches

The Cambridge Structure Database (CSD)[1] was established in 1965 and is the world's largest repository for crystal structures of small organic and metal organic molecules. Each crystal structure in the CSD can be identified by a six letter refcode, a unique identifier assigned to every new entry in the CSD. In addition to raw three-dimensional structural data, the database also contains bibliographic, chemical and physical property information for over half a million crystal structures analysed using X-ray and neutron diffraction. Each crystal structure undergoes a series of extensive validation and cross-checking tests by expert chemists and crystallographers to ensure the validity of the data. All of the crystals that pass these tests can be accessed by various software programs that are associated with the database and can be subdivided into several interfaces:

Conquest: Searches for crystal structures[2]
Mercury: Visualization and analysis of crystal structures[3]
Vista: Analysis and display of data retrieved from the CSD[4]
Prequest: Validation of crystal structure and chemical information for entry to the CSD[5]
Isostar: Intermolecular interaction occurrence studies[6]
Mogul: Retrieval of molecular geometry data from the CSD[7].

All CSD searches in this study were performed on CSD v5.35 with the November 2013 update utilising the ConQuest V1.16 search program[2]. The search for all salts was done according to the queries described in Chapter 3. The salts consist of a cationic and an anionic group, where the cationic group may be a primary amine, secondary amine, tertiary amine or cyclic amine (excluding pyridinium), while the anionic groups were identified on compounds that are included in the list of pharmaceutically acceptable molecules published by Stahl and Wermuth[8]. The anionic group can be further divided into carboxylates and carbonates, halides, sulfates and sulfonates, inorganic phosphates, nitrates and thiocyanates. Some molecules contain both anionic and cationic functional groups and are defined as zwitterions. All zwitterions were excluded from the study as they lead to an inaccurate representation of hydrate formation. We therefore defined zwitterions in this study as any structure that contained two charged atoms, X1 and X2, with an intramolecular contact between them. This contact can be separated by 1–999 bonds, with a contact between 0.001 and 40 Å[9].

Searches can be refined by using subsets that have been developed specifically for the CSD to ensure high quality results. These subsets were created by removing crystal structures with low quality data or are repeat structures. All of the searches in this study were performed using the best polymorph subset developed by Jaco van de Streek[10].

In addition to the best polymorph set, other additional constraints to the searches were applied that specify that 3D coordinates must be available and transition metals should be excluded from the results. Searches for hydrates were performed using the sample of structures that fulfil all of the criteria mentioned above. Hydrates were identified as any structure that contains one molecule consisting of an oxygen atom with a minimum of two hydrogen atoms, *i.e.* H_2O or H_3O^+.

2.1.2. Neutron diffraction structures

An atom will only produce a scattering pattern if radiation is scattered by its nuclei or electrons. The type of radiation determines if the nuclei or electrons will produce the scattering pattern. In X-ray diffraction (XRD), a beam of X-rays is scattered by the electrons in a crystal. Since small atoms with a low atomic number (Z) value have few electrons, they will not produce intense scattering patterns[11]. It is for this reason that XRD does not yield accurate positions for the hydrogen atoms in crystal structures. This was confirmed by Chatra *et al.* who investigated the properties of hydrogen atoms in carboxylic acids by using different diffraction techniques[12]. Their study showed that H-bond parameters determined by XRD tend to be 0.1 Å shorter than the true mean internuclear distance, which is why the hydrogen atom positions have to be optimised when performing computational calculations. In neutron diffraction (ND), on the other hand, the beam of neutron radiation is scattered by the atomic nuclei in a crystal instead of the electrons. This has several beneficial consequences:

- It is possible with neutron diffraction to determine the geometrical parameters of light atoms, such as hydrogen, in the presence of heavy atoms.
- Only ND (not XRD) enables H-atom anisotropic thermal parameters to be determined.
- Elements with similar Z-values have different scattering patterns.

The second point is of particular importance for the computational study of H-bonding in hydrate structures, since it means that the hydrogen atom positions are known accurately, unlike with XRD. Fisher *et al.*[13] performed a study showing that neutron diffraction plays a crucial role in determining how the structures of pharmaceutical compounds affect their properties. In their study the position of a hydrogen atom involved in the catalytic activity of a drug compound could be accurately determined by neutron diffraction, which enabled them to study and optimise the binding to the target compound. Since we are particularly interested in the properties of the H-bond within the crystal, we specifically utilised the crystal structures that have been determined using ND where available to enable us to perform single point calculations that would yield sensible results.

2.1.3. Structure optimisation

As the aim of this study was to determine the influence of H-bonds on hydrate formation in pharmaceutically acceptable salts, only the portion of the crystal structure involving the H-bond was analysed. Consider for example the dicyclohexyl-ammonium 4-hydroxy-benzenesulfonate monohydrate (refcode: ABERAB), which consists of three fragments. There is however only one H-bond between 4-hydroxybenzene-sulfonate and the water fragment. Therefore we can exclude the dicyclohexylammonium fragment (see Figure 2.1) when our study focuses on the H-bond in this structure and we only perform calculations on the two fragments that are involved in the H-bond. All other fragments were removed as their presence means that a significantly larger amount of computational power and time would be required. It should also be mentioned that frequency calculations were performed for all structure optimisation calculations in this study in order to confirm that the structures are at a minimum energy conformation.

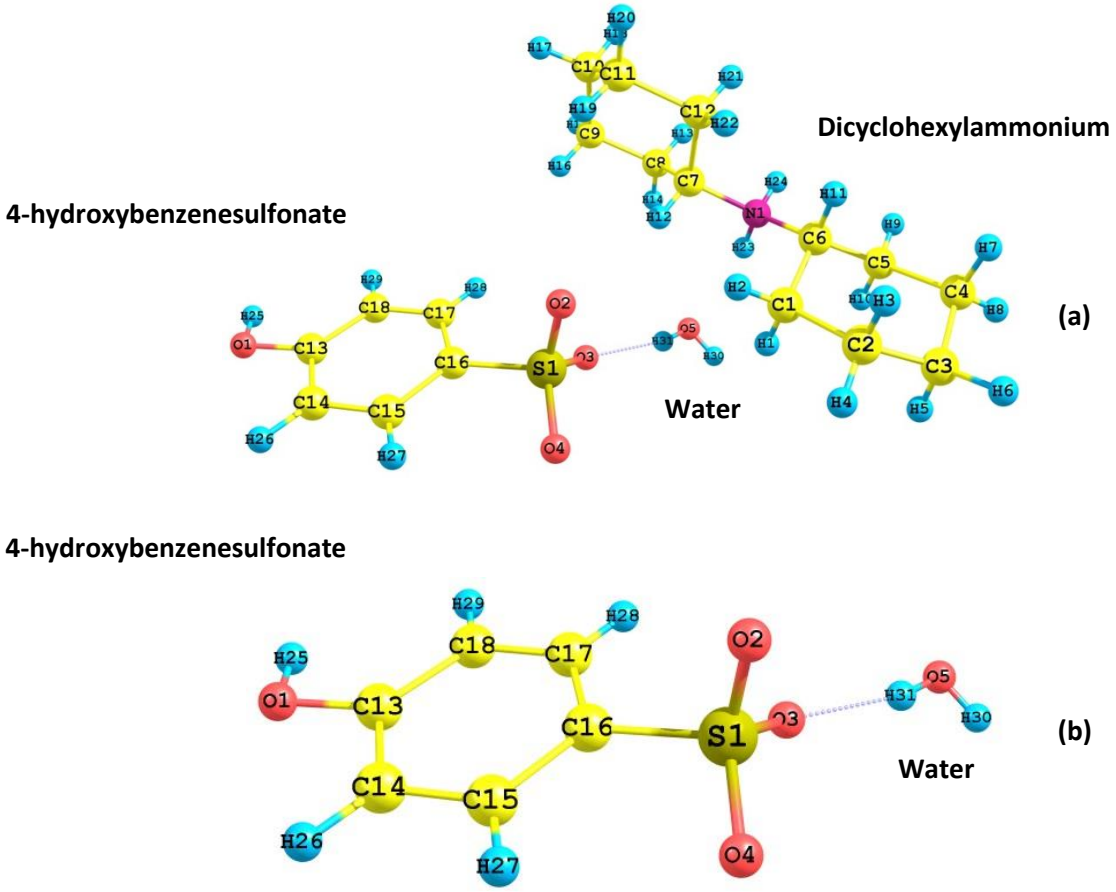

Figure 2.1. The structure of dicyclohexylammonium 4-hydroxy-benzenesulfonate monohydrate with refcode ABERAB, showing (a) the full asymmetric unit consisting of three components; (b) the two fragments involved in H-bonding.

2.1.4. Modelling of hydrate salts in the gas phase

Infantes *et al.* described the absence of an H-bond in hydrates to be a very rare occurrence[14]. Therefore we investigated the H-bond in the hydrates of pharmaceutical salts by performing calculations with the Gaussian 09 rev. B.01 and rev. D.01 packages[15]. Suitable DFT and WFT methods were chosen for all geometry optimisation calculations. The DFT method was initially used to optimise a structure and the coordinates were then subsequently used to perform a WFT geometry optimisation. The optimised geometries of DFT and WFT methods will correspond to the same minimum, but the DFT method will be less accurate in describing the minimum as illustrated in Figure 2.2. Both methods approach the minimum with several steps from the starting geometry. By initially using a less computationally expensive DFT method, we will decrease the number of steps that are required for a WFT method to approach the minimum energy geometry.

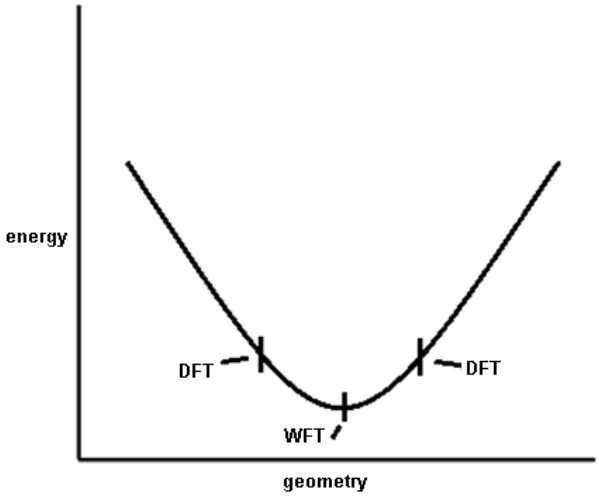

Figure 2.2. Shows a plot of energy versus geometry and illustrates the difference in minimum energy geometries obtained with DFT and WFT methods.

In this study we tested the ability of various DFT and WFT methods to reproduce optimised geometries and interaction energies that have been calculated with the Møller-Plesset Perturbation theory (MP2)[16,17] WFT method. DFT methods approximate the energy of a system by using electron density functionals, whereas WFT methods follow a systematic route towards the exact electronic wave function and hence highly accurate energies[18]. The MP2 method has been benchmarked to give accurate results within ± 0.01 Å for structures with a similar composition as the structures in this study[17,19].

Optimisation calculations were performed with the HF[20], B3LYP[21,22], M06-HF, M06-2X, M06-L[23], PBEPBE[24,25], B97D[26] and B2PLYPD[27,28] methods in combination with the 6-311++G(d,p)[29,30], cc-pVTZ[31,32] and aug-cc-pVTZ[31-33] basis sets. It was found that the M06-2X method in combination with the 6-311++G(d,p) basis set provides results that are comparable to calculations performed at the MP2/aug-cc-pVTZ level of theory. These results are tabulated in later chapters in order to show how this was achieved. All geometry optimisation calculations were therefore performed at the M06-2X/6-311++G(d,p) level of theory. The coordinates from these geometries were then subsequently used for geometry

optimisation calculations with the MP2 method in combination with the 6-311++G(d,p), cc-pVTZ and aug-cc-pVTZ basis sets. These are large basis sets that have been reported to provide accurate results for calculations with M06-2X and MP2 methods[34]. Geometry optimisation calculations with the MP2 method and a large basis set have been reported to provide results comparable to calculations performed with the CCSD(T) method[35], one of the most accurate computationally expensive WFT methods.

After all structures were fully optimised, we calculated the H-bond interaction energy with various DFT and WFT methods. The DFT methods used to calculate interaction energies consisted of the M06-2X, M06-2X-D3[36], B971[37], B97D3 and TPSS[38] methods in combination with the 6-311++G(d,p), cc-pVTZ and aug-cc-pVTZ basis sets, while the WFT MP2 method was used in combination with the 6-311++G(d,p) and aug-pVTZ basis sets.

The addition of D3 in the DFT methods given above indicates that a dispersion correction, developed and published by Grimme[39] in 2006, was applied. This correction is applied to the total energy as given in equation 1:

$$E_{DFT\text{-}D} = E_{KS\text{-}DFT} + E_{disp}. \qquad \text{(eq. 1)}$$

The $E_{KS\text{-}DFT}$ component consists of the Kohn-Sham energy obtained using the chosen DFT method. Kohn and Sham[40] developed a set of equations that are analogous to the Hartree-Fock (HF) equations, although they also include correlation effects. The HF method is a very basic method that approximates the wave function and energy of a many-body system. This method assumes that the exact, N-body wave function of the system can be approximated by a single Slater determinant, which is an expression that describes the wave function of a multi-electron system. The variational method is then applied in order to generate N-coupled equations for the N spin orbitals. The variational method approximates the ground and excited states. A solution to these approximated equations yields Hartree-Fock wave functions from which the energy of the system can be calculated[20]. The Kohn-Sham equations replace the Hartree-Fock equations with an exchange-correlation functional, which includes both exchange and electron correlation effects. The E_{disp} in equation 2 can be defined as given in the following equation:

$$E_{disp} = -s_6 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{C_6^{ij}}{R_{ij}^6} f_{damp}(R_{ij}). \qquad \text{(eq. 2)}$$

The s_6 symbol represents a scaling factor that depends on the specific density functional used, N represents the number of atoms, C_6^{ij} represents the dispersion coefficient for an atom pair *ij* given by the equation $C_6^{ij} = \sqrt{C_6^i C_6^j}$, and R_{ij} represents the interatomic distance.

One of the key ingredients of all DFT-D methods is the damping function f_{damp}, which determines the short-range behaviour of the dispersion corrections to avoid errors for electrons of atoms in close proximity to one another:

$$f_{damp}\left(\mathrm{R}_{ij}\right) = \frac{1}{1+ e^{a\left(\frac{\mathrm{R}_{ij}}{\mathrm{R}_r}-1\right)}}. \qquad \text{(eq. 3)}$$

The R_r symbol represents the sum of the atomic radii, while a is a constant equal to 20.

The counterpoise correction was applied to all interaction energy calculations in order to minimise the basis set superposition error (BSSE)[41,42]. Calculations that are performed with small basis sets are susceptible to BSSE, which occurs when the basis set functions of fragments start to overlap as they approach one another. The total energy for the overlapping fragments consists of contributions from short-range mixed basis sets and long-range unmixed basis sets, which introduces the BSSE[43]. The counterpoise correction was developed by Boys and Bernardi to correct for BSSE[41] by calculating the energy of each fragment with the same basis set as the total system. The uncorrected interaction energy between fragments A and B can be calculated as:

$$\Delta E_{int}\,(AB) = E_{AB}^{AB}(AB) - E_{A}^{A}(A) - E_{B}^{B}(B), \qquad \text{(eq. 4)}$$

where the superscripts represent the basis set used and the subscripts represent the geometry of the fragment, while the symbol in parentheses represents the chemical system. Thus, $E_{AB}^{AB}(AB)$ represent the energy of the complex of the two fragments A and B and the combination of the basis sets of A and B, calculated at the geometry of the complex. Likewise, the energies of monomers A and B are calculated at their own geometries and basis sets. We can correct equation 4 for BSSE by calculating the amount by which monomer A is stabilised by the basis set of monomer B (and vice versa). This can be calculated as:

$$\begin{aligned}\Delta E_{BSSE}\,(A) &= E_{A}^{AB}(A) - E_{A}^{A}(A),\\ \Delta E_{BSSE}\,(B) &= E_{B}^{AB}(B) - E_{B}^{B}(B),\end{aligned} \qquad \text{(eq. 5)}$$

where we subtract the energy of monomer A with its monomer basis set from the energy of monomer A with the basis set of the complex AB (and likewise for monomer B). If we subtract the error in equation 5 from the interaction energy in equation 4, we can rewrite the equation for interaction energy as:

$$\Delta E_{int}^{CP}\,(AB) = E_{AB}^{AB} - E_{A}^{AB} - E_{B}^{AB} \qquad \text{(eq. 6)}$$

From equation 6 we see that the counterpoise correction minimises the BSSE by using the basis set of complex AB for both monomers A and B[41].

2.1.5. Visualisation of structures

All structure coordinates were obtained from the CSD, converted to Cartesian coordinates and saved in xyz Cartesian coordinates format using the Mercury v3.3 suite[44]. All output files from Gaussian were visualised and rendered with the Chemcraft[45] visualisation suite. The .xyz files were converted to .zmat files using the Molden program[46].

2.1.6. H-bond strength in a solvent model

There are two types of models that can be used to simulate an actual solvent environment[47], namely the continuum solvent model and the discrete solvent model. In this study we employed a continuum solvent model, which consists of a discrete number of point charges surrounding the solute to form a cavity. Various models can be used to describe the distribution of the charges, however, in this study we used the Polarisable Continuum Model (PCM). The polarisability of a solvent is determined by the relative permittivity (Ɛ), which is a property unique to each solvent. Structures were optimised in a range of solvents with increasing Ɛ values, namely heptane (Ɛ=1.911), n-octanol (Ɛ=9.863), 1-propanol (Ɛ=20.524), dimethyl-sulfoxide (Ɛ=46.826) and water (Ɛ=78.355). Solvents with different Ɛ values were chosen to test how each solvent affects the final geometry.

The interaction energies were calculated for optimised geometries at the M06-2X/ 6-311++G(d,p) level of theory. The counterpoise correction could not be applied to interaction energy calculations in solvent systems, as the theory for these types of calculations is much more complicated than for gas phase calculations. Without the counterpoise correction, equation 6 yielding the interaction energy can be adapted as follows:

$$\Delta E_{int}\,(AB) = E(AB) - \big(E(A) + E(B)\big)\,. \qquad \text{(eq. 7)}$$

The *E(A)* and *E(B)* components represent the energies of the two monomers A and B respectively, while *E(AB)* represents the energy of the interacting assembly (a dimer of both monomers). The energies of the dimer and the two monomers were calculating at the M06-2X/6-311++G(d,p) level of theory. We can alternatively correct the BSSE by calculating the BSSE for a compound in the absence of a solvent and adding this value to the energy obtained for the compound. This was however found to be unnecessary as the BSSE error was very small, with the largest BSSE found for the structure Dimethylaminomethylene-bis(phosphonic acid) hydrate (refcode: FEBZER) by performing a single point calculation at the M06-2X/6-311++G(d,p) level of theory. The BSSE was found to be equal to 1.74 kcal/mol, which is only 8.27 % of the H-bond interaction energy. For this reason, the BSSE energy was not considered to have a significant effect on the H-bond interaction energy and therefore no further studies were performed to investigate the BSSE of each system.

2.1.7. Stabilisation of the H-bond by the surrounding crystalline environment

An approximation of the amount of stabilisation that is provided by the crystal environment was obtained by calculating the H-bond interaction energy for the compound in the same orientation as the crystal structure and then comparing this to the H-bond interaction energy of fully optimised structures. It is assumed that the stability of the crystal structure is affected by the penalty associated with the distorted conformation in the crystal, which must be balanced by stabilising interactions with the surrounding crystalline environment. However, as discussed earlier, hydrogen atom positions are poorly determined. Internal coordinates describing the H-bond were optimised and are illustrated in Figure 2.3. The internal coordinates consist of three distances, R1, R2 and R3, an angle, θ, and a torsion angle Φ as specified in Table 2.1.

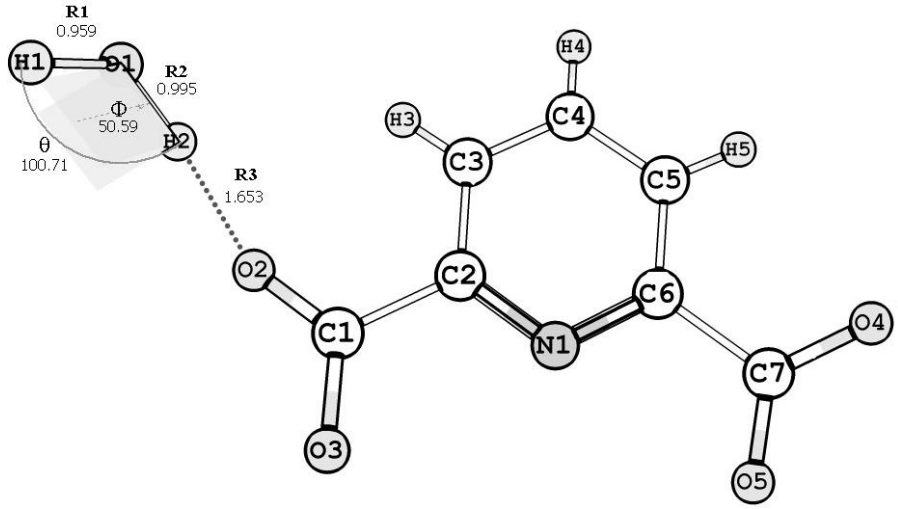

Figure 2.3. The structure of 2,6-dicarboxybenzene monohydrate indicating the H-bond internal coordinates.

Table 2.1. List of internal coordinates used to describe the H-bond.

Symbol	Type	Value	Atoms
R1	Distance	0.959	H1-O1
R2	Distance	0.995	O1-H2
R3	Distance	1.653	H2-O2
θ	Angle	100.7	H1-O1-H2
Φ	Torsion Angle	50.59	H1-O1-H2-O2

The angle between atoms H2, O2 and C1 could not be included as one of the parameters, as it is close to 180°, leading to errors in the internal coordinate system. Geometry optimisations were performed with a combination of the M06-2X and MP2 methods utilising the 6-311++G(d,p) basis set.

2.1.8. Potential energy scans on CREATH04

In order to study the variation in H-bond strength for different positions of a water molecule relative to a carboxylate group, scans of the Potential Energy Surface (PES) were performed on creatine monohydrate. This specific structure was chosen as it has neutron diffraction coordinates available in the CSD (refcode: CREATH04). In a PES scan values of selected internal coordinates within a molecule are varied systematically with the energy being calculated at each point. Single point calculations were performed at each step with DFT and WFT methods. The DFT methods consisted of M06-2X, B3LYP, B97D and B3LYPD3 all in combination with the 6-311++G(d,p) basis set. The WFT methods consisted of HF, CCSD(T) and MP2 in combination with the 6-311++G(d,p) basis set.

For the PES scans in our study the variable θ, as defined in Figure 2.4, was varied by a fixed stepsize of 5°. Values were obtained by performing single point calculations at each point. From these values equation 7 was used to calculate the interaction energies, with monomers A and B represented by the water and creatine molecules respectively.

Dummy atoms were used in order to correctly define the motion of the water molecule in the scans. These dummy atoms are signified by the X symbol and are placeholders that do not contain basis functions or electrons and will thus not influence the energy of the system, but help us define certain parameters used to construct the PES scans. Three dummy atoms, X1, X2 and X3, were placed in a straight line that runs parallel to the oxygen atoms in the carboxylate group, namely O1, C1 and O2. Dummy atom X2 was placed at a distance of 0.8 Å from C1. Dummy atoms X1 and X3 were placed 2.0 Å from dummy atom X2 and 1.0 Å from O2 and O1 respectively. The angle between dummy atoms X1, X2 and X3 is defined as θ. The movement of the water molecule was linked to the movement of the dummy atom X3. In addition to the association between H11 and X3, there is also a link between H11 and O1 which keeps the interaction distance constant. The range of motion of the dummy atom X3 and the water molecule that is associated with it is moved from 0 to 180 degrees as illustrated in Figure 2.4.

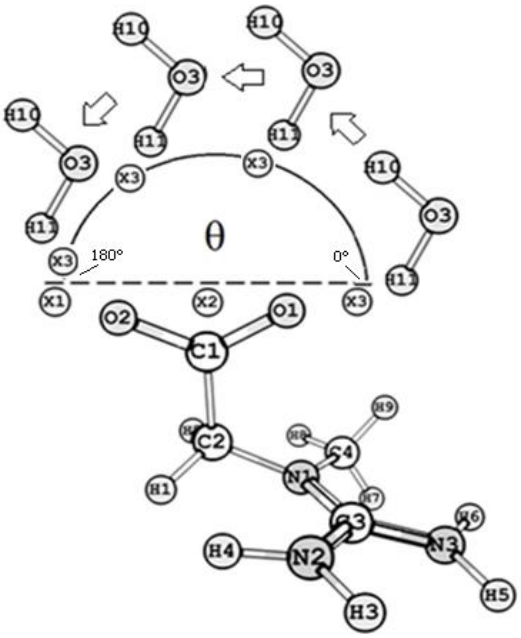

Figure 2.4. Illustration of the vertical scan performed such that the value of θ varies in increments of 5° from 0-180°. The geometry of the creatine molecule was obtained from the neutron diffraction data available for the CREATH04 structure.

Three additional scans were performed. The movement of the water molecule in all four scans is illustrated in Figure 2.5, with the first scan shown by red arrows between dummy atoms X1 and X3; blue arrows between dummy atoms X4 and X5 for the second scan; green arrows between dummy atoms X6 and X7 for the third scan and purple arrows between dummy atoms X8 and X9 for the fourth scan.

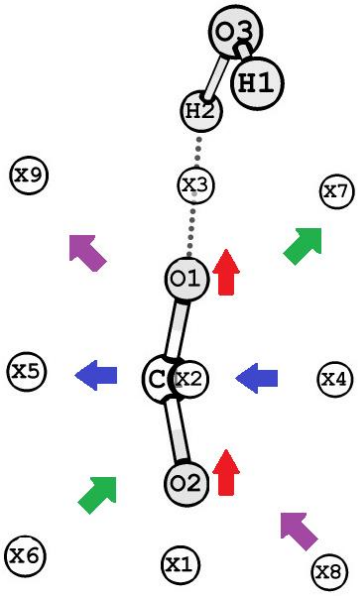

Figure 2.5. Definition of the four angle ranges used for the scans. The vertical scan (X1-X3, indicated by red arrows); the horizontal scan (X4-X5, indicated by the blue arrows); the left-to-right diagonal scan(X6-X7, indicated by the green arrows) and the right-to-left diagonal scan (X8-X9, indicated by the purple arrows).

The placement of the dummy atoms in these scans can be explained by the Cartesian system in Figure 2.6. The dummy atoms in the first scan are positioned at the following Cartesian coordinates:

X1 (0, -2, 0), X2 (0, 0, 0), X3 (0, 2, 0). (Scan 1)

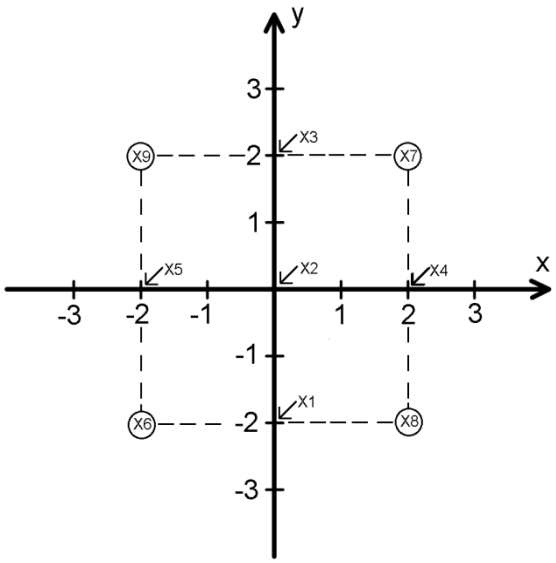

Figure 2.6. Cartesian representation of dummy atoms X1-X9 that are used to construct the PES scans.

For the second scan, dummy atoms X4, X2 and X5 are arranged in a line perpendicular to the atoms in the carboxylate group. The motion of the water molecule in the third and fourth scans can be described as diagonal and intermediary to the paths of the first two scans. The Cartesian coordinates of the dummy atoms in the last three scans are:

X5 (2, 0, 0), X2 (0, 0, 0), X6 (-2, 0, 0) (Scan 2)

X6 (-2, -2, 0), X2 (0, 0, 0), X7 (2, 2, 0) (Scan 3)

X8 (2, -2, 0), X2 (0, 0, 0), X9 (-2, 2, 0). (Scan 4)

The water molecule was linked to dummy atoms X4, X7 and X9. In each scan the θ angle was varied in 36 steps of 5° until it completed a full 180°.

2.1.9. Hydrogen bond propensities

All possible H-bond acceptors and donors on different fragments within a particular salt are in constant competition to form various possible interactions that can occur between two fragments. The possibility of an H-bond existing between two fragments can be predicted by the logit hydrogen-bonding propensity (LHP) model, which is part of Mercury v3.3 suite. This model calculates propensity values from a statistical model that describes all likely and unlikely H-bonds for each donor and acceptor pair. The statistical model is composed of

several molecules extracted from the CSD that contain the same functional groups as those present in the target system.

There are several variables that form part of the statistical model. These variables help describe the environment of each functional group. Amongst these variables are steric effects, aromaticity and competition between donor and acceptor groups. These variables are recorded for each functional group and incorporated into the database used for the statistical model. Hydrogen bond propensity values are then calculated by applying the LHP model to each functional group of the target compound. The propensity values take on a number between 0 and 1, where 0 will represent no likelihood of H-bond formation and 1 will indicate that an H-bond will always be found[48].

2.1.10. Atoms in Molecules (AIM) analysis

The wave function files (*.wfx files) used for AIM analysis were generated from fully optimised structures, which consisted of aromatic compounds with various electron withdrawing and donating substituents. The AIM analysis was performed utilising the AIMALL (v.11.04.03) software package[49], generating figures that show electron density (r_p) values at each bond critical point (BCP). The effect of these substituents was investigated by comparing the electron density values at the BCPs in substituted compounds. All wave function integrations were performed using the default software settings.

2.2. Experimental analytical techniques

2.2.1. Hydrate formation

The hydrate formation of three acids was investigated experimentally, namely benzene-1,3,5-tricarboxylic acid (trimesic acid), 4-pyridinecarboxylic acid (isonicotinic acid) and N-carbamimidoyl-N-methylglycine (creatine). Carboxylic acids were specifically chosen as the list of pharmaceutically acceptable acids published by Stahl and Wermuth[8] consists of mostly carboxylic acids. The effect of water activity, where water activity can be defined as the chemical potential of water molecules within a solution, was subsequently investigated by growing crystals from different mixed-solvent systems. Water activity is a thermodynamic property that has been proven to influence the propensity of compounds to form hydrates[50]. The properties of the crystals thus obtained were then investigated by the various analytical techniques that will be described in this section.

2.2.2. Water activity

To calculate water activity (a_w) values we require the mole fraction (x_w) and activity coefficient (γ_i) of water. Water activity is calculated from the product of these two values in equation 8:

$$a_w = \gamma_w . x_w \, . \qquad \text{(eq. 8)}$$

Several models can be used to calculate activity coefficients. The Non-Random Two-Liquid (NRTL) model equation developed by Renon *et al.* was used in our study as it can accurately describe mixed-solvent systems[51]. The NRTL model consists of equations that can be used to describe the correlation between the mole fraction, x_i, and activity coefficients, γ_i, in a specific liquid phase. The mole fraction is defined as the amount of a constituent (expressed in moles), n_i, divided by the total amount of all constituents in a mixture, n_{tot}, while the activity coefficient can be defined as a factor that accounts for deviations from ideal behaviour in a mixture of constituents[52]. The NRTL model is based on the interaction energy differences between the central molecules interacting with molecules of a similar chemical composition as compared to those of a different composition.

In the NRTL model[53], two dimensionless interaction parameters, τ_{12} and τ_{21} can be calculated from the interaction energy parameters Δg_{12} $(g_{12} - g_{21})$, and Δg_{21} $(g_{21} - g_{11})$, temperature (T) and the gas constant (R):

$$\tau_{12} = \frac{g_{12}-g_{21}}{RT}; \tau_{21} = \frac{g_{21}-g_{11}}{RT}. \qquad \text{(eq. 9)}$$

The subscript 12 implies an interaction between the central molecules of solvent 1 with molecules of solvent 2 (and vice versa for subscript 21). Solvent 1 represents water in our study and solvent 2 represents any of the other organic solvents that were used. The values for Δg_{12} and Δg_{21} were obtained from data tables published by Gmehling and Onken[54], which contain constants determined from experimental data. The values of τ_{12} and τ_{21} can then be used to calculate the Gibbs energy values, $G_{12}$and G_{21}

$$G_{12} = e^{\alpha_{12}\tau_{12}}; G_{21} = e^{\alpha_{12}\tau_{21}} \qquad \text{(eq. 10)}$$

where $\propto_{12}$ represents the non-randomness parameter, which depends on the distribution of molecules around the central molecule. $\propto_{12}$ can also be obtained from the published data tables of Gmehling and Onken. The values of τ_{12}, τ_{21}, $G_{12}$and G_{21}can then be used to calculate the activity coefficients of solvents 1 and 2 using equations 11 and 12:

$$ln\gamma_1 = x_2^2 \left(\left(\frac{G_{21}}{x_1 + x_2 G_{21}}\right)\right)^2 \tau_{21} + \frac{G_{12}\tau_{12}}{(x_2 + x_1 G_{12})^2} \qquad \text{(eq. 11)}$$

$$ln\gamma_2 = x_1^2 \left(\left(\frac{G_{12}}{x_2 + x_1 G_{12}}\right)\right)^2 \tau_{12} + \frac{G_{21}\tau_{21}}{(x_1 + x_2 G_{21})^2} \qquad \text{(eq. 12)}$$

where γ_1 and γ_2 are the activity coefficients of solvents 1 and 2 that can be substituted into equation 8, along with the mole fraction of water (x_w), to calculate water activity (a_w).

2.2.3. Crystallisations

The propensity of trimesic acid, isonicotinic acid and creatine to form crystalline hydrates in various solvent mixtures was investigated. These mixtures consist of water and one additional organic solvent, as shown in Tables 2.2-2.4, where the amount of water present will result in different water activity values. The water activity values and composition of each solvent mixture are summarised in Tables 2.2-2.4, along with descriptions of the crystals obtained from each mixture. The formation of hydrate crystals at the different water activity values are indicated by a yes.

Table 2.2. Proportions of water and ethyl acetate used in various solvent mixtures to test the role of a_w on hydrate formation.

Solvents		Water activity	Crystal formation		
Water (mL)	**Ethyl acetate (mL)**	**(a_w)**	**Isonicotinic acid**	**Creatine**	**Trimesic acid**
0.0	1.0	0.00	No	No	No
0.1	0.9	0.59	No	Yes	No
0.2	0.8	0.72	Yes	Yes	No
0.3	0.7	0.79	Yes	Yes	Yes
0.4	0.6	0.84	Yes	Yes	Yes
0.5	0.5	0.88	Yes	Yes	Yes
0.6	0.4	0.91	Yes	Yes	Yes
0.7	0.3	0.93	Yes	Yes	Yes
0.8	0.2	0.96	Yes	Yes	Yes
0.9	0.1	0.98	Yes	Yes	Yes
1.0	0.0	1.00	Yes	Yes	Yes

Table 2.3. Proportions of water and ethanol used in various solvent mixtures to test the role of water activity on hydrate formation.

Solvents		Water activity	Crystal formation		
Water (mL)	**Ethanol (mL)**	**(a_w)**	**Isonicotinic acid**	**Creatine**	**Trimesic acid**
0.0	1.0	0.00	No	No	No
0.1	0.9	0.34	No	No	No
0.2	0.8	0.52	No	Yes	No
0.3	0.7	0.64	Yes	Yes	No
0.4	0.6	0.72	Yes	Yes	No
0.5	0.5	0.79	Yes	Yes	No
0.6	0.4	0.84	Yes	Yes	Yes
0.7	0.3	0.89	Yes	Yes	Yes
0.8	0.2	0.93	Yes	Yes	Yes
0.9	0.1	0.97	Yes	Yes	Yes
1.0	0.0	1.00	Yes	Yes	Yes

Table 2.4. Proportions of water and acetone used in various solvent mixtures to test the role of water activity on hydrate formation.

Solvents		Water activity	Crystal formation		
Water (mL)	**Acetone (mL)**	**(a_w)**	**Isonicotinic acid**	**Creatine**	**Trimesic acid**
0.0	1.0	0.00	No	No	No
0.1	0.9	0.43	No	No	No
0.2	0.8	0.60	Yes	Yes	No
0.3	0.7	0.70	Yes	Yes	No
0.4	0.6	0.77	Yes	Yes	Yes
0.5	0.5	0.83	Yes	Yes	Yes
0.6	0.4	0.87	Yes	Yes	Yes
0.7	0.3	0.91	Yes	Yes	Yes
0.8	0.2	0.94	Yes	Yes	Yes
0.9	0.1	0.97	Yes	Yes	Yes
1.0	0.0	1.00	Yes	Yes	Yes

All of the crystallisations were prepared by adding approximately 0.01 g of isonicotinic acid, creatine or trimesic acid to 1.0 mL of water-organic solvent mixture. All mixtures were placed in small polytop vials and left to stand to allow crystal formation to proceed at room temperature. All compounds and solvents, other than distilled water, were purchased from Sigma-Aldrich and used without any further purification. Table 2.5 summarises the properties of the compounds that were used in the crystallisation procedure.

Table 2.5. Physical properties of starting materials used for the preparation of pharmaceutical hydrates

Compounds Molecular	Formula	Molar mass (Mr) (g/mol)	Physical Appearance	Densities (kg/m^3)*
Benzoic acid	$C_7H_6O_2$	122.12	White powder	1.27
Isonicotinic acid	$C_6H_5NO_2$	123.11	White powder	
Creatine	$C_4H_9N_3O_2$	131.13	White powder	1.40
Water (distilled)	H_2O	18	Colourless liquid	1.00
Ethanol	C_2H_6O	46.07	Colourless liquid	0.789
Ethyl acetate	$C_4H_8O_2$	88.105	Colourless liquid	0.897
Acetone	C_3H_6O	58.08	Colourless liquid	0.791

* Densities were obtained from Sigma-Aldrich catalogue 2008-2009

2.2.4. Single-Crystal X-ray Diffraction Analysis

The crystals obtained in this study were analysed by Single-Crystal X-ray diffraction (SCD). Single-Crystal X-ray (SCD) is a popular analytical technique that provides information about the internal lattice of crystalline substances, which includes unit cell dimensions and molecular structures, including bond lengths and bond angles of molecules within these crystals. All crystals that were analysed in this study were inspected under a microscope to select those of an appropriate size and uniform shape. Each crystal chosen for analysis was then mounted on a MiTeGen mount. This was then placed onto the goniometer head of the single-crystal diffractometer. All X-ray data was collected on a Bruker-Nonius Smart Apex diffractometer equipped with a fine-focus sealed tube and a 0.5 mm Monocap collimator (monochromated Mo-Kα radiation, $\lambda = 0.71073$ Å). The data were collected with a Charge-Coupled Device (CCD) area-detector with the generator powered at 40 kV and 30 mA. The diffractometer was equipped with a 700 Series Cryostream Plus, produced by Oxford Cryogenic Cryostat. This provided a constant stream of nitrogen for low temperature data collection at 100 K.

All data reduction, absorption corrections and unit cell determinations were carried out using the APEXII diffractometer software developed by Bruker[55]. All structure solution and refinement was carried out using the SHELX[56] suite of programs under the X-Seed interface[57]. The X-seed interface permits the analysis of data that has been obtained from SCD experiments, including visualisation of the crystal structure. POV-RAY[58] is additional software package, integrated into the X-seed interface for generating figures from the single-crystal data.

2.2.5. Thermogravimetric Analysis (TGA)

TGA analysis was used for thermal analysis of crystals. TGA is an analytical technique that is often used in supramolecular chemistry to measure the mass loss or gain of materials as a function of temperature.

This loss or gain in molar weight can be attributed to drying, structural water release, structural decomposition, gas evolution or re-hydration. TGA was carried out using a TA Instruments Q500 analyser under nitrogen gas (N_2) with a flow rate of 50 ml/min which was used to purge the furnace. The run started at an initial temperature of 20 °C and was set at a ramp rate of 10 °C/min until it reached 550 °C. The resulting thermogram was analysed using the TA Instruments Universal analysis program.

2.2.6. Differential Scanning Calorimetry (DSC) analysis

Differential Scanning Calorimetry (DSC) is used to determine the changes in heat flow between a sample and a reference material (often an empty pan). It is a powerful and versatile thermal technique that allows the analysis of compounds from -150 to 600 °C at a specific heating rate[59]. The heating rate of the system is controlled by creating a temperature

program, which is designed to linearly increase the temperature as a function of time. Differences in the temperature between the sample and reference material will constitute a thermal event, which can be related to phase transitions (solid-solid transitions, eutectics, melting, sublimation, decomposition etc.). All of this information can subsequently be displayed as a flow-versus-temperature scan, commonly referred to as a thermogram. Thermograms give information regarding the sample by measuring the total number of peaks, the temperature at the individual maxima, the quantity of absorbed or released energy, and the change of the heat capacity at each phase transition. In this study, DSC was used to determine the onset temperature of melting or phase transitions.

DSC analysis was carried out using a TA Instruments Q100 system under a N_2 gas purge (flow rate of 50.0 mL/min) coupled to a cooling unit. Sample preparation was done by placing a powdered sample of approximately 2-5 mg in an aluminium pan, non-hermetically sealed with a vented aluminium lid. Reference pans were prepared in a similar manner. The sample started at an initial temperature of 20 °C and was cooled to a temperature of -50 °C and then heated at a set ramp rate of 10 °C/min until it reached 200 °C and then cooled back down to 20 °C. The procedure was repeated so that the sample completed two thermal cycles per run.

References

(1) Allen, F.: The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B* **2002**, *58*, 380-388.

(2) Bruno, I. J., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J., Taylor, R., Cole, J. C.: Designing a New Multi-Component API Form Based on a Known Structure. *Acta Crystallographica Section B* **2002**, *58*.

(3) Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M., van de Streek, J.: Mercury: visualization and analysis of crystal structures. *Journal of Applied Crystallography* **2006**, *39*, 453-457.

(4) CCDC (1994). Vista - A Program for the Analysis and Display of Data Retrieved from the CSD. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

(5) CCDC (1994). PreQuest - A program for the validation of crystal structure and chemical information for entry to the CSD Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

(6) Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R., Verdonk, M. L.: IsoStar: A library of information about nonbonded interactions. *Journal of computer-aided molecular design* **1997**, *11*, 525-537.

(7) Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E., Orpen, A. G.: Retrieval of Crystallographically-Derived Molecular Geometry Information. *Journal of Chemical Information and Computer Sciences* **2004**, *44*, 2133-2144.

(8) Stahl, P. H., Wermuth, C. G.: Handbook of Pharmaceutical Salts: Properties, Selection and Use. Weinheim, *Wiley* **2002**.

(9) Haynes, D. A., Jones, W., Motherwell, W. D. S.: Occurrence of pharmaceutically acceptable anions and cations in the Cambridge Structural Database. *CrystEngComm* **2005**, *7*, 342-345.

(10) van de Streek J.: *Acta Crystallographica Section B* **2005**, *61*, 504–510.

(11) Wilson, C. C., Henry, P. F., Schmidtmann, M., Ting, V. P., Williams, E., Weller, M. T.: Neutron powder diffraction – new opportunities in hydrogen location in molecular and materials structure. *Crystallography Reviews* **2014**, *20*, 162-206.

(12) Chitra, R., Das, A., Choudhury, R. R., Ramanadham, M., Chidambaram, R.: Hydrogen bonding in oxalic acid and its complexes: A database study of neutron structures. *Journal of Physics* **2004**, *63*, 263-269.

(13) Fisher, S. Z., Aggarwal, M., Kovalevsky, A. Y., Silverman, D. N., McKenna, R.: Neutron Diffraction of Acetazolamide-Bound Human Carbonic Anhydrase II Reveals Atomic Details of Drug Binding. *Journal of the American Chemical Society* **2012**, *134*, 14726-14729.

(14) Infantes, L., Chisholm, J., Motherwell, S.: Extended motifs from water and chemical functional groups in organic molecular crystals. *CrystEngComm* **2003**, *5*, 480-486.
(15) Frisch, M. J., et al.: Gaussian 09, Revision B.01. *Wallingford CT* **2009**.
(16) Binkley, J. S., Pople, J. A.: Møller–Plesset theory for atomic ground state energies. *International Journal of Quantum Chemistry* **1975**, *9*, 229-236.
(17) Møller, C., Plesset, M. S.: Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618-622.
(18) Grabowski, I., Teale, A. M., Śmiga, S., Bartlett, R. J.: Comparing ab initio density-functional and wave function theories: The impact of correlation on the electronic density and the role of the correlation potential. *The Journal of Chemical Physics* **2011**, *135*.
(19) Helgaker, T., Gauss, J., Jorgensen, P., Olsen, J.: The prediction of molecular equilibrium structures by the standard electronic wave functions. *The Journal of Chemical Physics* **1997**, *106*, 6430-6440.
(20) Frank, N. H.: Note on the Hartree and Hartree-Fock Methods. *Physical Review* **1937**, *51*, 577-583.
(21) Becke, A. D.: Density-functional thermochemistry. V. Systemic optimization of exchange-correlation functionals. *Journal of Chemical Physics* **1997**, *107*, 8554.
(22) Hamprecht, F. A., Cohen, A. J.: Development and assessment of new exchange-correlation functionals. *Journal of Chemical Physics* **1998**, *109*, 6264.
(23) Zhao, Y., Truhlar, D.: The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemical Accounts* **2008**, *120*, 215-241.
(24) Perdew, J. P., Burke, K., Ernzerhof, M.: Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865-3868.
(25) Perdew, J. P., Burke, K., Ernzerhof, M.: Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1997**, *78*, 1396-1397.
(26) Grimme, S.: Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27* 1787-1799.
(27) Grimme, S.: Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of Chemical Physics* **2006**, *124*-125
(28) Schwabe, T., Grimme, S.: Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Physical Chemistry Chemical Physics* **2007**, *9*, 3397-3406.
(29) Krishnan, R., Binkley, J. S., Seeger, R., Pople, J. A.: Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980**, *72*, 650-654.

(30) McLean, A. D., Chandler, G. S.: Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18. *The Journal of Chemical Physics* **1980**, *72*, 5639-5648.

(31) Dunning, T. H.: Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007-1023.

(32) Woon, D. E., Dunning, T. H.: Calculation of the electron affinities of the second row atoms: Al–Cl. *The Journal of Chemical Physics* **1993**, *99*, 3730-3737.

(33) Kendall, R. A., Dunning, T. H., Harrison, R. J.: Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992**, *96*, 6796-6806.

(34) Bhavaraju, M., Gwaltney, S. R.: A theoretical analysis of substituted aromatic compounds. *International Journal of Quantum Chemistry* **2013**, *113*, 1171-1179.

(35) Dąbkowska, I., Jurečka, P., Hobza, P.: On geometries of stacked and H-bonded nucleic acid base pairs determined at various DFT, MP2, and CCSD(T) levels up to the CCSD(T)/complete basis set limit level. *The Journal of Chemical Physics* **2005**, *122*.

(36) Grimme, S., Ehrlich, S., Goerigk, L.: Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456-1465.

(37) Hamprecht, F. A., Cohen, A. J., Tozer, D. J., Handy, N. C.: Development and assessment of new exchange-correlation functionals. *The Journal of Chemical Physics* **1998**, *109*, 6264-6271.

(38) Tao, J. M., Perdew, J. P., Staroverov, V. N., Scuseria, G. E.: Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Physical Review Letters* **2003**, *91*

(39) Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate ab initio parameterization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*.

(40) Kohn, W., Sham, L. J.: Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133-A1138.

(41) Boys, S. F., Bernardi, F.: The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970**, *19*, 553-566.

(42) Simon, S., Duran, M., Dannenberg, J. J.: How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *The Journal of Chemical Physics* **1996**, *105*, 11024-11031.

(43) Belikov, V. V., Bokhan, D. A., Trubnikov, D. N.: Estimating the basis set superposition error in the CCSD(T)(F12) explicitly correlated method using the example of a water dimer. *Russian Journal of Physical Chemistry* **2014**, *88*, 629-633.

(44) Allen, F.: The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B* **2002**, *58*, 380-388.
(45) http://www.chemcraftprog.com. accessed on 20 March 2013.
(46) Schaftenaar, G., Noordik, J. H.: Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design* **2000**, *14* 123-134.
(47) Marenich, A. V., Olson, R. M., Kelly, C. P., Cramer, C. J., Truhlar, D. G.: Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011-2033.
(48) Majumder, M., Buckton, G., Rawlinson-Malone, C. F., Williams, A. C., Spillman, M. J., Pidcock, E., Shankland, K.: Application of hydrogen-bond propensity calculations to an indomethacin-nicotinamide (1 : 1) co-crystal. *CrystEngComm* **2013**, *15*, 4041-4044.
(49) Keith, T. A.: AIMAll *TK Gristmill Software* **2012**
(50) Cazier, J.-B., Gekas, V.: Water activity and its predictions: A review. *International Journal of Food Properties* **2001**, *4*, 35.
(51) Renon, H., Prausnitz, J. M.: Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE Journal* **1968**, *14*, 135-144.
(52) IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Oxford, *Blackwell Scientific Publications*, 1997.
(53) Weidlich, U., Gmehling, J.: A Modified UNIFAC Model. 1. Prediction of VLE, hE, and gamma Infinite. *Industrial and Engineering Chemistry Research* **1987**, *26*, 1372-1381.
(54) Gmehling, J., Onken, U., Arlt, W., Grenzheuser, P., Weidlich, U., Kolbe, B., Rarey, J.: DECHEMA Chemistry Data Series, Vapor-Liquid Equilibrium Data Collection. Part 1a: Aqueous-Organic Systems, *Supplement 1* **1998**, *1.1a*, 1-750.
(55) Bruker (2007).APEXII. Bruker AXS Inc.Wisconsin, USA.
(56) Sheldrick, G.: A short history of SHELX. *Acta Crystallographica Section A* **2008**, *64*, 112-122.
(57) Barbour, L. J.: *Journal of Supramolecular Chemistry* **2001**, *1*, 189-191.
(58) POV-Ray for Windows, Version 3.6.1a.icl8.win32, Persistence of Vision Team, Persistence of Vision Pty. Ltd., 2003-2004.
(59) O'Neill, M. J.: The Analysis of a Temperature-Controlled Scanning Calorimeter. *Analytical Chemistry* **1964**, *36*, 1238-1245.

CHAPTER 3

A CSD survey of the hydrate formation in salts with pharmaceutically acceptable anions and cations containing an NH group

Introduction

In this chapter we will focus on salt structures obtained within the Cambridge Structural Database (CSD) and their propensity to form hydrates[1]. The CSD is the world's largest repository for crystal structures of small organic and metal organic molecules. In 2005, Haynes *et al.*[2] subdivided a list of pharmaceutically acceptable acids that was published by Stahl and Wermuth[3] into those containing different anionic functional groups. These groups are carboxylates and carbonates, halides, sulfates and sulfonates, phosphates, nitrates and thiocyanates. Their study consisted of CSD searches for combinations of N-bearing cations and anions from these groups. Furthermore, they investigated the trends in hydration across the groups. The most pronounced trend in their results showed a very low occurrence of hydrate formation in pyridinium carboxylate salts: only 9.1% as compared to 16.7 and 24.6% for carboxylate salts of other N-bearing cations. These results show the difficulty in forming hydrates of compounds that contain the carboxylate anionic group. This is problematic for the pharmaceutical industry as salts of potential drug molecules are often prepared as it is known that this leads to the modification of their properties. For example, an increase or decrease in solubility, reduce toxicity or improve stability[4]. Salts are ionic compounds that are composed of cations (positively charged ions) and anions (negative ions) so that the product is neutral (no net charge)[5]. Studies have shown that the propensity for hydrate formation of specific anionic and cationic groups is higher than for other groups[2,4]. We aim to investigate the properties of these groups and identify what factors lead to increased hydrate formation.

3.1. Pharmaceutical salts in the CSD

The CSD was systematically searched for compounds that contain cationic and anionic functional groups that commonly occur in pharmaceutically acceptable ions. The searches were performed by creating specific queries in the CSD and making use of specialised subsets, which restrict the searches to contain only structures that meet certain criteria. An anion can be classified as pharmaceutically acceptable if it is biologically non-toxic and does not interfere with the efficacy of the drug[6]. Table 3.1 contains the list of all pharmaceutically acceptable acids that was published by Stahl and Wermuth[3]. These acids all form the anionic component of pharmaceutical salts and are known to form strong H-bonds in hydrate structures. The cations were chosen to be nitrogen-based as these are often pharmaceutically acceptable in salt design. The propensity of these anions to crystallise with specific groups of cations was screened as well as the occurrence of hydrates within these specific groups. The analysis performed in this survey focuses on a limited number of anionic and cationic functional groups and their effect on crystal structure properties.

The results from the study by Haynes *et al.* mentioned previously also showed a frequent occurrence of particular H-bonding motifs within different salts[7]. More recent studies have identified the crucial role that a strong H-bond plays in the formation of hydrates of a specific group of salts and in many chemical and biological processes[8,9]. For example, in biological macromolecules such as proteins, DNA and RNA, the presence of a large number of H-bonds

will determine the structure and dynamics of these molecules[10-12]. With the development of modern techniques that can study the properties of the H-bond, we aim to identify various pharmaceutical hydrate salts with an H-bond between the pharmaceutically acceptable anion and water molecule[13]. Furthermore, we wish to establish trends for the hydrate formation of different anionic and cationic components of salts.

Table 3.1. List of pharmaceutically acceptable acids taken from reference 3.

1-hydroxy-2-naphthoic acid	ethanesulfonic acid	methanesulfonic acid
2,2-dichloroacetic acid	formic acid	naphthalene-1,5-disulfonic acid
2-hydroxyethanesulfonic acid	fumaric acid	naphthalene-2-sulfonic acid
2-oxoglutaric acid	galactaric acid	nicotinic acid
4-acetamidobenzoic acid	gentisic acid	nitric acid
4-aminosalicylic acid	glucoheptonic acid (D)	oleic acid
acetic acid	gluconic acid (D)	oxalic acid
adipic acid	glucuronic acid (D)	palmitic acid
ascorbic acid (L)	glutamic acid	pamoic acid
aspartic acid (L)	glutaric acid	phosphoric acid
benzenesulfonic acid	glycerophosphoric acid	proprionic acid
benzoic acid	glycolic acid	pyroglutamic acid (- L)
camphoric acid (+)	hippuric acid	salicylic acid
camphor-10-sulfonic acid (+)	hydrobromic acid	sebacic acid
capric acid (decanoic acid)	hydrochloric acid	stearic acid
caproic acid (hexanoic acid)	isobutyric acid	succinic acid
caprylic acid (octanoic acid)	lactic acid (DL)	sulfuric acid
carbonic acid	lactobionic acid	tartaric acid (+ L)
cinnamic acid	lauric acid	thiocyanic acid
citric acid	maleic acid	toluenesulfonic acid (*p*)
cyclamic acid	malic acid (- L)	undecylenic acid
dodecylsulfuric acid	malonic acid	
ethane-1,2-disulfonic acid	mandelic acid (DL)	

3.2. CSD searches for hydrate occurrence in pharmaceutical anions

The CSD survey performed by Haynes *et al.* on the hydrate formation in pharmaceutically relevant groups of anions was performed on the CSD version 5.25 (November 2003), which consisted of 270815 crystal structures at the time. The searches were repeated in this study using the same queries and restrictions when performing searches on the CSD v5.35 (Nov 2013 + 1 update), which has a total of 673859 crystal structures[14]. We aimed to investigate if the trends that were identified by Haynes *et al.* in 2003 are still observed when the number of crystals has more than doubled in 2013[2]. The search conditions were adapted from the study of Haynes *et al.* and are described below.

Firstly, all salt structures must consist of an anionic and cationic component that commonly found on pharmaceutical compounds. Furthermore, searches were done in order to determine the propensity to form hydrates for particular pharmaceutically acceptable anions. As mentioned previously, these groups of anions consist of carboxylates and carbonates, halides, sulfates and sulfonates, phosphates, nitrate and thiocyanates. A total of five N-bearing cationic groups were investigated, namely pyridinium, primary amine, secondary amine, tertiary amine and cyclic amines that exclude pyridinium. The propensity of compounds to form hydrates was determined by creating a query that searches for all structures that contain H_2O and H_3O^+ molecules. The search criteria for the respective ions are graphically illustrated in Figure 3.1 with two additional filters in place, namely hits must have their 3-D coordinates determined and have no transition metals present. Furthermore, it should also be mentioned that the carboxylate functional group in the Figure 3.1 will also provide search results that contain carbonate functional groups. This also holds for the sulfate functional group.

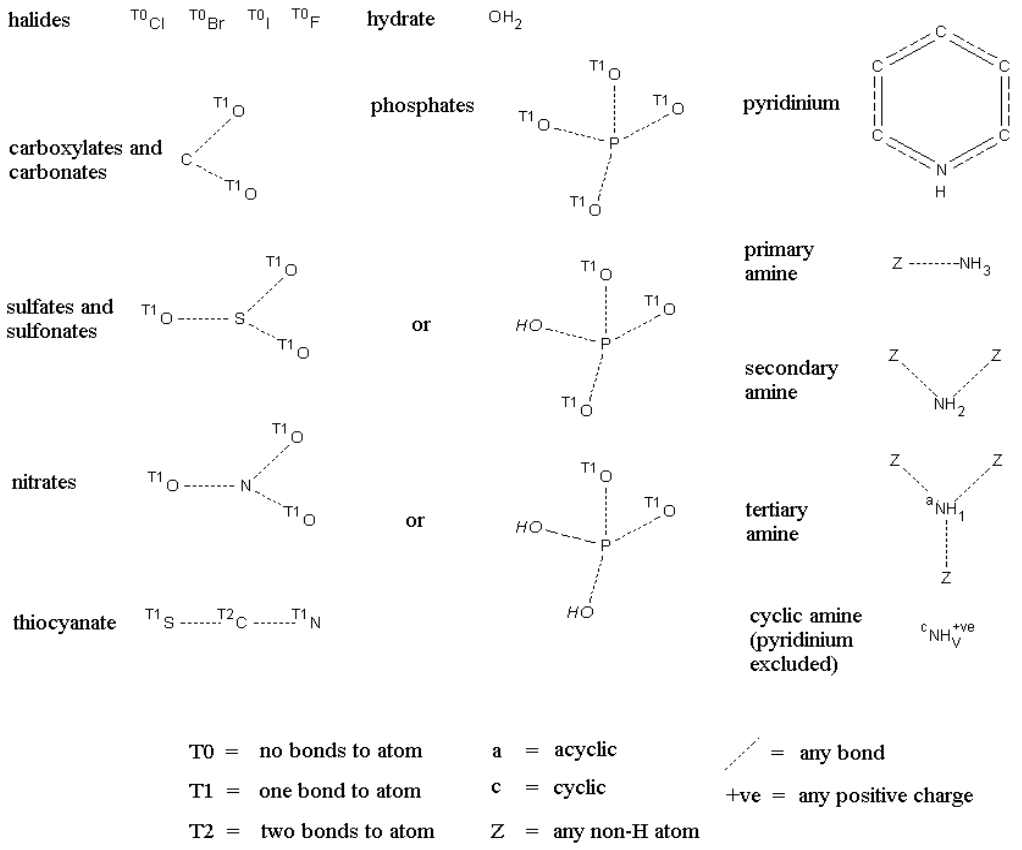

Figure 3.1. Fragments with restrictions used in ConQuest searches of the CSD as adapted from reference 2.

It was observed that several zwitterions are present in the search results. A zwitterion can be defined as a molecule that contains both cationic and anionic groups. From Figure 3.2 it is clear that there is a high propensity for zwitterions to form hydrates, with 38.16 % of zwitterions and salts forming hydrates and 24.84 % hydrate formation when zwitterions are excluded. The thiocyanate anion group has been excluded from Figure 3.2 as there are only a

limited number of crystal structures with this specific functional group in the CSD. It would thus not be possible to accurately determine the propensity of thiocynate anions to form hydrates. We observe a much higher hydrate occurrence for all of the anionic functional groups within the zwitterions. The most pronounced trend for the anionic groups was for the phosphate group, with 56.25% hydrate formation in zwitterions as opposed to 29.31% for salts with zwitterions excluded. This effect is clearly evident in all anionic groups chosen for this study and is the reason why 7395 zwitterions were excluded from our searches, leaving us with 19017 non-zwitterionic salts as the sample used in this study.

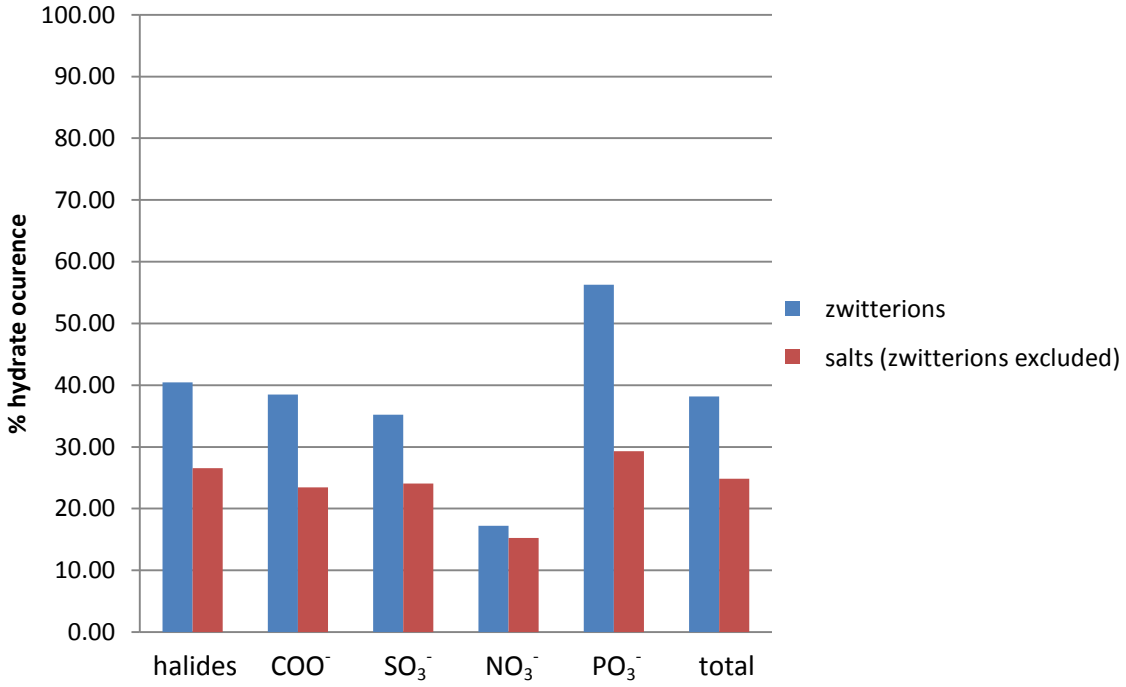

Figure 3.2. Percentage occurrence of hydrates in salts with only zwitterions (7395 hits) and zwitterions excluded (19 017 hits) and in the CSD.

Following the exclusion of zwitterions, all searches were performed on the sample of 19017 non-zwitterionic salts and are summarised in Table 3.2. The results from Haynes *et al.* are summarised in Table 3.3 to enable comparison to the results obtained in this study.

Table 3.2. Results of the number of salts that form hydrates for each N-bearing cationic group from searches performed on CSD v5.35 (Nov 2013 + 1 update).

Cation		Halide	Carboxylate or carbonate	Sulfate or sulfonate	Phosphate	Nitrate	SCN-	Other	Total
Pyridinium	$N_{compounds}$	638	809	255	46	135	8	926	4048
	$N_{hydrates}$	238	180	118	9	19	2	264	1169
	% hydrates	28.1	30.3	38.6	26.8	10.4	0.0	26.8	28.88
Primary Amine	$N_{compounds}$	1158	1165	440	164	125	11	985	3460
	$N_{hydrates}$	325	353	170	44	13	0	264	912
	% hydrates	27.5	25.6	35.2	60.2	28.6	6.7	19.5	26.4
Secondary Amine	$N_{compounds}$	1284	843	264	83	105	15	866	3460
	$N_{hydrates}$	353	216	93	50	30	1	169	912
	% hydrates	27.5	25.6	35.2	60.2	28.6	6.7	19.5	26.4
Tertiary Amine	$N_{compounds}$	493	218	78	14	22	4	502	1331
	$N_{hydrates}$	101	40	23	8	2	0	47	221
	% hydrates	20.5	8.1	10.6	10.3	14.3	0.0	1175.0	16.6
Cyclic Amine (Excluding Pyridinium)	$N_{compounds}$	2617	1653	524	122	221	24	1577	6738
	$N_{hydrates}$	854	485	212	57	57	4	324	1993
	% hydrates	32.6	29.3	40.5	46.7	25.8	16.7	20.5	29.6
Total	$N_{compounds}$	6190	4688	1561	429	608	62	4856	19037
	$N_{hydrates}$	2271	1903	753	205	237	13	1730	5207
	% hydrates	36.7	40.6	48.2	47.8	39.0	21.0	35.6	27.4

Table 3.3. Results of the number of salts that form hydrates for each N-bearing cationic group from reference 2.

Cation		Halide	Carboxylate or carbonate	Sulfate or sulfonate	Phosphate	Nitrate	SCN-	Other	Total
Pyridinium	$N_{compounds}$	210	143	44	11	35	3	321	767
	$N_{hydrates}$	73	13	15	4	6	1	58	170
	% hydrates	34.76	9.09	34.09	36.36	17.14	33.33	18.07	22.16
Primary Amine	$N_{compounds}$	553	514	131	90	42	3	382	1715
	$N_{hydrates}$	146	114	58	32	3	0	100	453
	% hydrates	26.4	22.2	44.3	35.6	7.1	0.0	26.2	26.4
Secondary Amine	$N_{compounds}$	272	172	41	25	8	2	207	727
	$N_{hydrates}$	36	30	12	17	1	0	21	117
	% hydrates	13.2	17.4	29.3	68.0	12.5	0.0	10.1	16.1
Tertiary Amine	$N_{compounds}$	272	72	20	6	10	1	268	649
	$N_{hydrates}$	51	12	5	6	1	0	19	94
	% hydrates	18.8	16.7	25.0	100.0	10.0	0.0	7.1	14.5
Cyclic Amine (Excluding Pyridinium)	$N_{compounds}$	1374	513	183	24	76	10	703	2883
	$N_{hydrates}$	412	126	80	10	15	3	133	779
	% hydrates	30.0	24.6	43.7	41.7	19.7	30.0	18.9	27.0
Total	$N_{compounds}$	2681	1414	419	156	171	19	1881	6741
	$N_{hydrates}$	718	295	170	69	26	4	331	1613
	% hydrates	26.8	20.9	40.6	44.2	15.2	21.1	17.6	23.9

From the results of the searches in Tables 3.2 and 3.3, we observe a significant increase in the number of salts for each of the functional groups. Furthermore, the trends in hydration for some groups have also changed considerably. On inspecting the results of these searches it was noticed that the results included various repetitions of crystal structures and also crystal structures that lack experimental data. There are structures with no unit cell information or missing experimental conditions, which complicates the process of establishing trends and drawing conclusions from these results. The exponential growth in the number of crystal structures in the CSD has made it a nearly impossible task to manually inspect every crystal structure. The lack of experimental data for crystal structure in the CSD is a direct result of it being published before the development of the uniform crystallographic information file (CIF)[15]. Before CIF files were implemented, several other formats were used at once with different data exchange and data were regularly lost in translation. CIF files provide a general, flexible, rapidly extensible and universal file format protocol that allows the incorporation of new data without the need for modification. This has led to the development of a method that can refine searches in order to exclude crystal structures that lack experimental data.

All crystal structures deposited into the CSD before the introduction of CIF files had to be individually inspected for errors. This proved to be a time-consuming task and led to development of the *best representative polymorph set* by Jaco van de Streek[16]. He identified *republication (rerefinement), redetermination, reinterpretation* and *polymorph* as four different criteria that result in a bias in the statistical analysis of data from CSD searches. A crystal structure can be defined as a republication if the same structure is published without any additional crystallographic data. A redetermination is the publication of a known structure with different experimental data and usually different authors, while a reinterpretation involves publishing the same structure by adding corrections to an available structure. A polymorph is the publication of a crystal structure that differs from other published crystal structures in its packing arrangement. He developed a computer program that addresses these potential sources of bias in three stages:

(i) Eliminate crystal structures that are very suspicious or downright wrong;
(ii) Distinguish between redeterminations and polymorphs to cluster all redeterminations per unique polymorph;
(iii) Per cluster, select the ‘best’ representative.

These stages were completed by analysing space groups, disorder, 3D coordinates, R-factors, matched entries, chemical and structural formulae, unit-cell volume, intermolecular contacts, aromatic rings, bond lengths, reinterpretations, hard-coded list exclusions and polymorphs. These criteria will not be discussed in this chapter, but they are fully discussed in the 2006 publication by van de Streek. This elimination process then results in an unbiased subset of the CSD, with only 70% of the crystal structures passing the quality tests. van de Streek estimated that roughly 10% of the crystal structures were rejected due to human error and a further 20% owing to disorder. After all of these criteria were applied, the resulting list contained the most reliable crystal structures in the CSD. These can still be broken down into

clusters of unique polymorphs. These clusters consist of structures that have been grouped according to specific experimental parameters, which are:

(i) Lowest R-factor
(ii) All H atoms present, and if everything else equal, neutron study or best hydrogen list
(iii) Lowest temperature
(iv) Room temperature.

For the searches conducted in this study, the best hydrogen list was chosen as we are particularly interested in H-bonds that are often present in hydrates, which have been reported to be involved in the stability of pharmaceutical compounds[17]. All of the searches described in Figure 3.1 were repeated on the *best hydrogen list polymorph set* and the results are summarised in Table 3.4. The data were then compared to the results in Table 3.3. The most pronounced observation was that the hydrate occurrence for pyridinium salts changed from 9.09% to 22.0%. This increase was the largest change and warranted some additional investigation.

Table 3.4. Results of the number of salts that form hydrates for each N-bearing cationic group from searches performed on CSD v5.35 (Nov 2013 + 1 update) utilising the best hydrogen list polymorph set.

Cation		Halide	Carboxylate or carbonate	Sulfate or sulfonate	Phosphate	Nitrate	SCN-	Other	Total
Pyridinium	$N_{compounds}$	495	617	184	38	107	6	641	2088
	$N_{hydrates}$	185	136	66	7	17	1	97	509
	% hydrates	37.4	22.0	35.9	18.4	15.9	16.7	15.1	24.38
Primary Amine	$N_{compounds}$	823	1240	301	120	93	9	707	3293
	$N_{hydrates}$	224	256	110	26	10	0	189	815
	% hydrates	27.2	20.6	36.5	21.7	10.8	0.0	26.7	24.7
Secondary Amine	$N_{compounds}$	957	634	175	52	73	13	551	2455
	$N_{hydrates}$	232	160	49	29	16	1	94	581
	% hydrates	24.2	25.2	28.0	55.8	21.9	7.7	17.1	23.7
Tertiary Amine	$N_{compounds}$	354	170	49	11	15	4	341	944
	$N_{hydrates}$	66	29	10	6	1	0	25	137
	% hydrates	18.6	17.1	20.4	54.5	6.7	0.0	7.3	14.5
Cyclic Amine (Excluding Pyridinium)	$N_{compounds}$	1994	1178	371	83	164	22	1035	4847
	$N_{hydrates}$	607	328	141	34	34	4	207	1355
	% hydrates	30.4	27.8	38.0	41.0	20.7	18.2	20.0	28.0
Total	$N_{compounds}$	4623	3839	1080	304	452	54	3275	13627
	$N_{hydrates}$	1314	909	376	102	78	6	612	3397
	% hydrates	28.4	23.7	34.8	33.6	17.3	11.1	18.7	24.9

On inspecting the 143 pyridinium carboxylate salts in Table 3.3, it was observed that several crystal structures were deposited into the CSD by the same authors. Several authors investigated the change in crystal packing with slight variations of one specific structure, *i.e.* by adding/removing substituents or by changing the experimental conditions of one structure, i.e. increasing the temperature or pressure. Consequently, all of these structures are deposited into the CSD. The number of crystal structures that have a common author are summarised in Table 3.5. One specific example is the work done by Smith *et al.*[18] on anhydrous nitro-based benzoates, where a total of 16 out of the 143 hits consisted of slight variations of these structures.

Table 3.5. List of authors and the total number of anhydrous pyridinium carboxylate salts they had in CSD version 5.25 (November 2003).

Author	Crystals	% (Total=143)
Smith	16	11.2
Ferguson	13	9.09
Biradha	9	6.29
Farrell	8	5.59
Lynch	6	4.20
Sugiyama	4	2.80
Total	56	39.2

By the year 2003, the authors in Table 3.5 also published several anhydrous crystal structures that are slight variations of one structure, which amount to just below 40% of all the pyridinium carboxylate salts. Therefore, the small number of pyridinium carboxylate salts in Table 3.3 led to a misrepresentation of the propensity of this group to form hydrates. Once the number of pyridinium carboxylate salts increased to a total of 617 hits as seen in Table 3.4, we have a larger distribution of different structures and the hydrate occurrence has increased to 22.0%. The increase in hydrate formation of pyridinium carboxylate salts from 2003 to 2013 is illustrated in Figure 3.3.

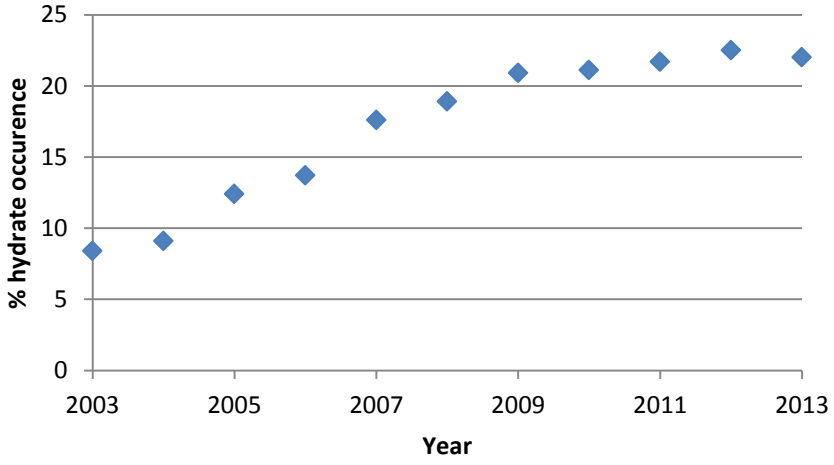

Figure 3.3. Increase of hydrate occurrence of pyridinium carboxylate salts from 2003 to 2013.

From Figure 3.3 we observe the gradual increase in hydrate formation for pyridinium carboxylate salts as the number of structures in the CSD increased every year. This trend suggests that there were simply not enough pyridinium carboxylate salts in 2003 to describe the propensity of pyridinium carboxylate salts to form hydrates accuratly.

Another reason for the low hydrate occurrence for pyridinium carboxylate salts in Table 3.4 could be due to the fact that the *best representative subset* was also only developed in 2006 and was not available when Haynes *et al.* performed their initial study[2]. We suggest that a combination of these two factors led to the low hydrate occurrence for pyridinium carboxylate salts in Table 3.3.

The number of salts in Table 3.4 increased considerably when comparing it to Table 3.3. As with any statistical study, a larger data set will enable us to make general observations and establish trends much more confidently. In Table 3.4 we see a high incidence of halide and carboxylate or carbonate salts of nitrogen based cations, and a much lower occurrence of the other anionic groups. The reason for the large number of carboxylate and carbonate salts is that these functional groups are often accompanied by easy synthesis, low cost, good redox activity, high thermal stability, high charge mobility and increased biological activity[19]. This is also evident from the fact that more than 80% of the salt formers in Table 3.1 have carboxylate and carbonate groups. Halides are also very often single counterion fragments in pharmaceutical salts that increase the stability and function of the compound with halogen bonds[20]. The number of structures of the remaining anionic groups in Table 3.1 is much lower.

From the results in Table 3.4 we observe that some anionic groups form hydrates much more readily than other anionic groups, which is due to the number of polar groups on the anionic group and the electronegativity values of each atom. These electronegative atoms consist of H-bond acceptors atoms. We observe a lower hydrate occurrence for the halide, carboxylate and carbonate, nitrate and thiocyanate salts. It has been previously reported that the hydrate occurrence in halide salts decreases as the size of the ion increases, i.e. F^- shows the highest hydrate occurrence and I^- shows the lowest as illustrated in Figure 3.4[2]. The low hydrate occurrence in the carboxylate and carbonate group can be explained by the carboxylate anionic group only having two polar oxygen atoms as opposed to three or four electronegative atoms in sulfates, sulfonates and phosphates. Fewer electronegative atoms will correspondingly result in few available sites for H-bond formation. If the H-bond plays a crucial role in the formation of hydrates, this will result in lower hydrate formation in these anionic groups.

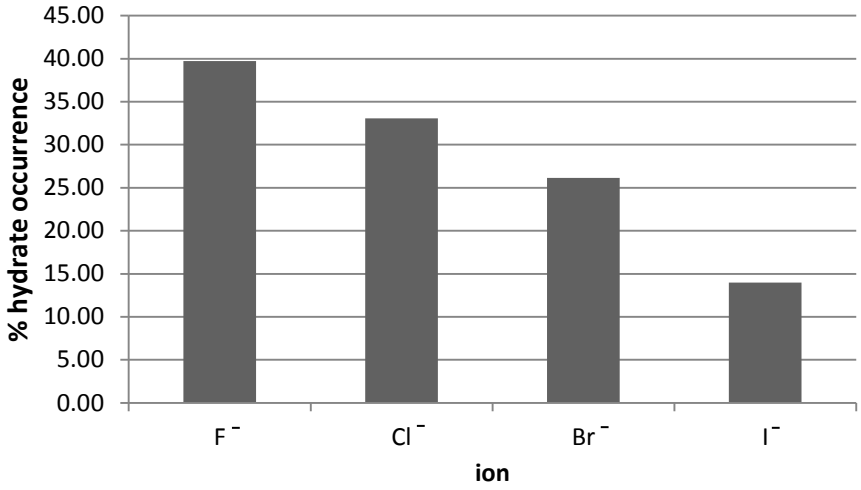

Figure 3.4. Decrease in hydrate occurrence in halide salts as the size of the halide ion increases.

The differences in hydrate occurrence for these groups of anions can also be explained by the differences in their electronegativity values[21]. Electronegativity describes the ability of atoms to attract electrons or electron density. The electronegativity values of phosphorus (P), carbon (C), sulfur (S) nitrogen (N) relative to the oxygen (O) atoms are summarised in Table 3.6.

Table 3.6. Electronegativities and electronegativity differences relative to O for atoms P, C, S, N and O.

Atom	Electronegativity (χ)	$\Delta\chi$ ($\chi_O - \chi_x$)
P	2.1	1.3
C	2.5	1.0
S	2.6	0.9
N	3.0	0.5
O	3.5	0.0

$\Delta\chi$ represents the difference in electronegativity between two atoms. χ_x represents the electronegativity of the central atom in the anionic functional group.

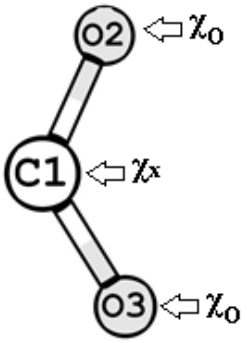

Figure 3.5. Carboxylate group showing the C-atom as the χ_x group and the O-atom as the χ_O group.

A study by Infantes *et al.* on the hydrate formation in organic crystals reported two important findings[9]. Firstly, they found that an increase in the number of polar groups will lead to higher water affinity. Secondly, hydrate formation in the absence of an H-bond is very rare. They described the polarity of a group by using electronegativity and noted that the overall polarity of a compound will also increase with an increase in the number of polar groups. For example, a carboxylate group only has two polar oxygen atoms, while a phosphate group has four oxygen atoms. A larger electronegativity difference between atoms in a bond will also result in a more polar bond.

If we arrange the anions according to electronegativity and number of polar groups, we have the following order:

Phosphate > Sulfate > Carbonate > Sulfonate > Nitrate > Carboxylate > Thiocyanate

The order of anions corresponds to the hydrate occurrence trends in Table 3.4 if we consider that carbonates are grouped with carboxylates and sulfates are grouped with sulfonates in the table. It should also be mentioned that a mere 81 out of the 3839 carboxylate and carbonate structures are carbonates (2.11%), which means the % hydrates mainly describe the hydrate occurrence in carboxylate structures. Furthermore, the highest propensity is observed for the sulfate and sulfonate group, as several phosphate groups have one or two hydroxyl groups as illustrated in the queries in Figure 3.1, which affects the polarity, electronegativity and charge distribution of the functional group. The thiocyanate anion has the lowest polarity as it only has two weakly electronegative atoms. The $\Delta\chi$ for these two groups are $0.1(\chi_S-\chi_C)$ and $0.5\ (\chi_N-\chi_C)$. The second weakest group is the nitrate anion, which has three polar groups with $\Delta\chi$ of 0.5 for all three groups.

The highest propensity to form hydrates is observed for the sulphate and phosphate anions and the lowest for the nitrate and thiocyanate anions. Groups with low polarity will be weaker H-bond acceptors, and consequently the H-bond strength will decrease.

The total propensity of secondary amines to form hydrates has increased from 16.1% (CSD version 5.25 Nov 2003) to 23.7% (CSD v5.35 (Nov 2013 + 1 update + Best H-List), whilst the propensity of the tertiary amines to form hydrates has remained at 14.5%. This can be explained by the propensity of different nitrogen based cations to form hydrates as a result of different H-bond donor strengths. We can identify trends in amine hydration by comparing the entries in the total column and the hydrate rows for every nitrogen-based cation in Table 3.4. This process can be explained by a study that was performed by Kellend on the effect of different amines on hydrate formation[22]. He found that large bulky amine groups such as *n*-butyl, *n*-pentyl or *iso*-pentyl have the best ability to inhibit the formation of hydrates. This was confirmed by replacing the butyl group with a smaller amine group, which led to a significantly lower inhibition of hydrate formation. This corresponds to our findings, where we see a clear trend of decreasing percentages of hydrates formed from primary amines to the bulkier tertiary amines. However, Kellend's study did not include cyclic amines. A study by Graton *et al.* looked at the calculated enthalpy values of 68 primary, secondary, tertiary and cyclic amines[23]. They calculated the electronic energy of the H-bonds between members of a series of optimised structures. The results showed a decrease in bond strength on going from methylamine (6.91 kcal/mol, primary amine) to trimethylamine (6.44 kcal/mol, tertiary amine), while they calculated a bond strength of 6.83 kcal/mol for piperidine, a six membered cyclic amine. There is a clear correlation between the strength of the H-bond to these amines and their propensity to form hydrates. This is

confirmed by arranging the amines in Table 3.4 according to increasing values of hydrate formation:

Cyclic Amine > Pyridinium > Primary Amine > Secondary Amine > Tertiary Amine

This trend suggests the presence of stronger H-bonds in primary and cyclic amines, which correspond to a greater tendency to form hydrates, while secondary and tertiary amines have weaker H-bonds and lower tendency to form hydrates.

3.3. Other pharmaceutically acceptable anions

On inspection of the hits in the "other" anion column in Table 3.4, we found that there are several nitrogen-, phosphorus- and sulfur-bearing anions amongst these structures. These anionic groups were mainly excluded from the list of acceptable acids by Wermuth and Stahl[3] as they were considered as biologically toxic groups that are not stable enough for processing. Three noticeable examples from the "other" anions are nitrite, nitroxide and phosphonate. Most of these anions were previously thought to have mainly adverse effects on the biological system, but recent pharmaceutical studies have explored the beneficial functions of these anions and their potential as novel pharmaceutical compounds. We will describe the beneficial uses of each anionic group and also investigate their propensity to form hydrates as revealed by CSD searches.

3.3.1. Nitrites

There has been a very negative image of nitrites, since they have been known to be responsible for potential carcinogenic effects, blue baby syndrome, as well as the suspected health risks related to fertilizer overuse[24]. Griess, however, showed that the endogenous production of nitrites is part of the biological immune response to fever[25]. More recent studies on the body have found that nitrites are reduced to $NO^{\bullet}$, which plays a crucial role in vascular smooth muscle relaxation, the control of blood pressure and flow, and protecting organs against ischemia-related tissue injury[26-29]. These effects ultimately led to therapeutic treatments for coronary artery disease (CAD), angina and heart failure.

3.3.2. Phosphonate

The phosphonate moiety has found renewed interest as a pharmaceutical salt that can treat various diseases. Zoschke *et al.* have very recently introduced a novel guanosine-analog phosphonate for the treatment of non-melanoma skin cancer[30]. Phosphonate analogues have also been used in the treatment of Alzheimer's disease and the human papilloma viruses (HPV)[30-32].

3.3.3. Nitrogen dioxide

Nitrogen dioxide is known to have a radical resonance form with a delocalised negative charge as illustrated in Figure 3.6, which is responsible for the strong and directional H-bonds in structures with the nitrogen dioxide group. The radical has been detected by EPR in salt matrices at low temperatures[33]. It exists as a radical ($NO_2^{\bullet}$) in the biological system and can be formed *in vivo* as peroxynitrite-derived radical intermediates. Nitrogen dioxide has been found to be selective towards biomolecules and can be produced by several physiological routes, including nitric oxide autoxidation and peroxidase-catalysed reactions[34]. The participation of nitrogen dioxide in protein nitration mechanisms *in vivo* has also been extensively investigated as the full effect of protein nitration in cell and organ function is not yet fully understood. Some evidence suggests that the radical contributes to cell and tissue injury associated with several pathologies such as inflammatory conditions, vascular diseases, and neurodegeneration[24]. Other studies have suggested beneficial uses of the radical. Kono *et al.*[35] found that the radical can kill bacteria at controlled pH levels.

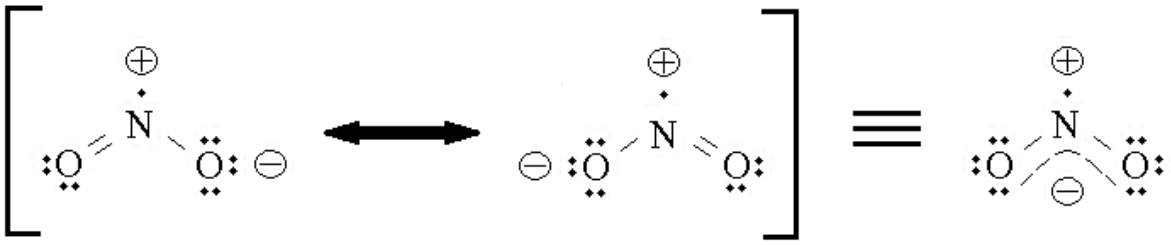

Figure 3.6. Resonance forms of nitrogen dioxide.

These findings have aroused renewed interest in the use of nitrogen dioxide as a pharmaceutical compound as bacteria are becoming increasingly resistant to common antibiotics[34]. Other studies suggest that the radical plays a specific role in homeostasis as it is synthesized by mammalian cells[34,36].

3.3.4. Nitroxide

Nitroxide is a radical anion known to enhance the production of oxygen and nutrient delivery to active muscles. This leads to an improved tolerance to physical exercise and recovery mechanisms[37]. These beneficial uses have led to mass production of $NO^{\bullet}$-based supplements aimed at enhancing training performance. Other studies also suggest that $NO^{\bullet}$ regulates specific functions involved in memory, synaptic plasticity and neuroprotection[38].

3.4. CSD searches for hydrate occurrence in other pharmaceutical anions

The searches for nitrogen-bearing structures with nitrite, nitrogen dioxide, nitroxide and phosphonate anions as well as hydration trends were performed on CSD v5.35 (Nov 2013 + 1 update), using ConQuest version 1.6[14]. The nitrogen dioxide group was specified not to

contain any charge, but it should be part of a non-zwitterionic salt. Zwitterions were excluded from these searches and the best H-list was used. The search criteria for the respective groups are illustrated in Figure 3.7, with two filters in place: hits must have their 3-D coordinates determined, and have no transition metals present. Hydrates were once again defined as all structures that contain the H_2O and H_3O^+ molecules. The results of the searches are summarised in Table 3.7.

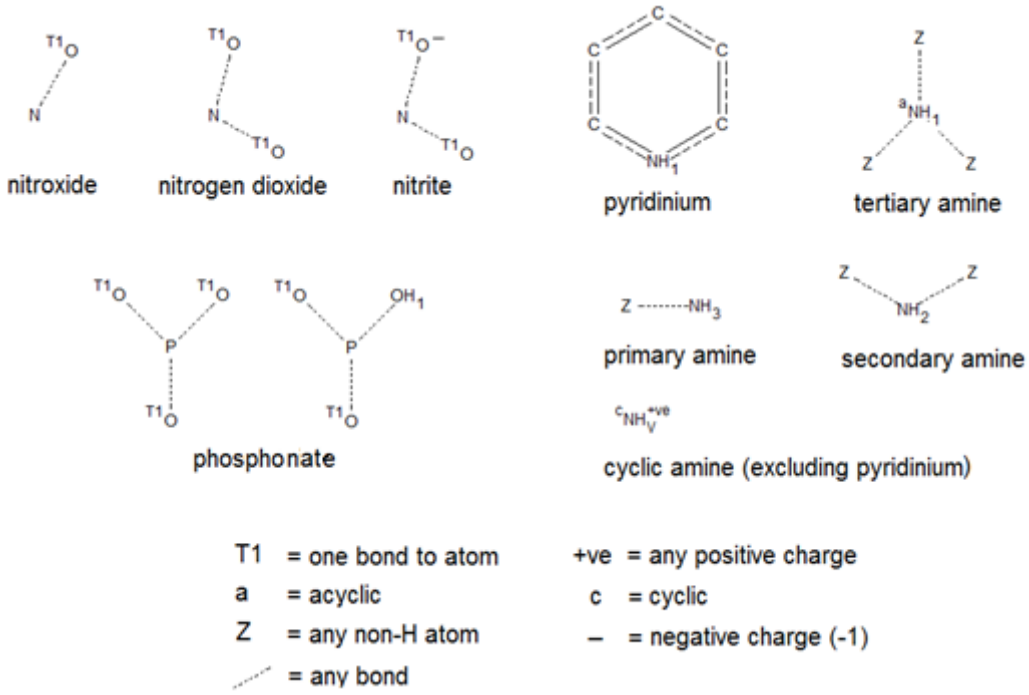

Figure 3.7. Fragments with restrictions used in ConQuest searches of the CSD.

Firstly, we observe few structures in the CSD that contain the selected groups in Figure 3.7. Most of these functional groups have only recently resurfaced to be used as potential pharmaceutical compounds. The investigation of various anionic groups is important for the development of novel pharmaceutical compounds.

According to the World Health Organisation[39], a specialised agency concerned with international public health, the global pharmaceuticals market is worth US$300 billion a year and this figure is expected to rise to US$400 billion within the next three years. The 10 largest drugs companies control over one-third of this market. Several of these companies have sales of more than US$10 billion a year and profit margins of about 30%. This has driven the development of improved screening and development methods of novel pharmaceutical compounds. We expect the number of cationic groups with the functional groups in Table 3.7 to increase as we have previously observed with the results in Table 3.2 and 3.4. We can however still make some general observations from the structures summarised in Table 3.7. If we look at the total column, we see the familiar trend where the cyclic and primary amines have the highest hydrate occurrence, while the secondary and tertiary amines have the lowest. The phosphonate group has the highest hydrate occurrence as is expected for the most polar group. We obtained only a few nitrite structures, which

makes it difficult to comment on the hydrate formation for this specific anion. By not subdividing the cations into the N-based cation groups, we produce Table 3.8, which shows at least one hydrate structure for each of the groups in Table 3.7, which confirms the propensity of these groups to form hydrates.

Table 3.7. Results of the number of "other" anion compounds that form hydrates for each N-bearing cationic group from searches performed on CSD v5.35 (Nov 2013 + 1 update) utilising the best hydrogen list polymorph set.

Cation		Nitroxide	Nitrogen dioxide	Nitrite	Phosphonate	Other	Total
Pyridimium	$N_{compounds}$	9	213	1	18	1694	1935
	$N_{hydrates}$	2	21	0	3	452	478
	% hydrates	22.22	9.86	0.00	16.67	26.68	24.70
Primary Amine	$N_{compounds}$	4	255	0	54	2710	3023
	$N_{hydrates}$	0	57	0	11	694	762
	% hydrates	0.0	22.4	0.0	20.4	25.6	25.2
Secondary Amine	$N_{compounds}$	13	145	4	19	2118	2299
	$N_{hydrates}$	3	20	0	4	489	516
	% hydrates	23.1	13.8	0.0	21.1	23.1	22.4
Tertiary Amine	$N_{compounds}$	2	113	1	2	796	914
	$N_{hydrates}$	0	11	0	0	117	128
	% hydrates	0.0	9.7	0.0	0.0	14.7	14.0
Cyclic Amine (Excluding Pyridinium)	$N_{compounds}$	19	407	3	29	4098	4556
	$N_{hydrates}$	6	74	0	12	1167	1259
	% hydrates	31.6	18.2	0.0	41.4	28.5	27.6
Total	$N_{compounds}$	47	1133	9	122	11416	12727
	$N_{hydrates}$	11	183	0	30	2919	3143
	% hydrates	23.4	16.2	0.0	24.6	25.6	24.7

Table 3.8. Summary of other anions in N-based salts (zwitterions excluded).

		Nitroxide	Nitrogen dioxide	Nitrite	Phosponate	Total
All N-Based salts (no-Zwitterions)	Ncompounds	69	1456	15	153	1693
	Nhydrates	12	234	1	44	291
	% hydrates	17.4	16.1	6.7	28.8	17.2

Chapter Summary

We have established that the low hydrate occurrence in pyridinium carboxylate salts observed in previous studies was due to the lack of crystal structures at the time the study was performed by Haynes *et al.*[2]. We performed similar searches on CSD with the best hydrogen list and found a clear correlation between the hydration of the pharmaceutical anions and the polarity and electronegativity of each anionic group. Furthermore, we found that cyclic amines and primary amines have the highest hydrate occurrence, while secondary and tertiary amines have lower hydrate occurrence.

We have also identified four additional anion and radical groups with the ability to form hydrates. Several studies in the literature have tried to identify a single factor that influences hydrate formation. The most dominant factors are the H-bonds, chemical functional group polarity, donor and acceptor ratios, sterics effects and geometrical parameters[40-42]. The influence of the H-bond has been described by Infantes *et al.* to play a crucial role in hydrate formation[9]. Many studies have been done on H-bonds, but only a few on the H-bond in hydrate structures. We will focus on the electronic nature of the H-bond in this dissertation and describe several studies in the following chapters to investigate the strength and nature of the H-bond in hydrate structures.

References

(1) Allen, F.: The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B* **2002**, *58*, 380-388.

(2) Haynes, D. A., Jones, W., Motherwell, W. D. S.: Occurrence of pharmaceutically acceptable anions and cations in the Cambridge Structural Database. *CrystEngComm* **2005**, *7*, 342-345.

(3) Stahl, P. H., Wermuth, C. G.: Handbook of Pharmaceutical Salts: Properties, Selection and Use. Weinheim, *Wiley* **2002**.

(4) Serajuddin, A. T. M.: Salt formation to improve drug solubility. *Advanced Drug Delivery Reviews* **2007**, *59*, 603-616.

(5) Quéré, J. W. L.: Pharmaceutical Salts and Co-crystals. *RSC Drug Discovery Series No. 16, RSC,* Cambridge, **2012**.

(6) Bing, M., Xiaoyi, G., Yi, S.: Quantitation of Ammonium Cations in Pharmaceutical Samples by Nonaqueous Capillary Electrophoresis with Indirect UV Detection. *Journal of Liquid Chromatography and Related Technologies* **2004**, *27*, 661-675.

(7) Haynes, D. A., Chisholm, J. A., Jones, W., Motherwell, W. D. S.: Supramolecular synthon competition in organic sulfonates: A CSD survey. *CrystEngComm* **2004**, *6*, 584-588.

(8) Aakeroy, C. B., Forbes, S., Desper, J.: The effect of water molecules in stabilizing co-crystals of active pharmaceutical ingredients. *CrystEngComm* **2012**, *14*, 2435-2443.

(9) Infantes, L., Fabian, L., Motherwell, W. D. S.: Organic crystal hydrates: what are the important factors for formation. *CrystEngComm* **2007**, *9*, 65-71.

(10) Shirhatti, P. R., Maity, D. K., Bhattacharyya, S., Wategaonkar, S.: C-H···N Hydrogen-Bonding Interaction in 7-Azaindole:CHX3 (X=F, Cl) Complexes. *ChemPhysChem* **2014**, *15*, 109 – 117.

(11) Desiraju, G. R., T. Steiner: The Weak Hydrogen Bond in Structural Chemistry and Biology. Oxford, *Oxford University Press,* **1999**.

(12) Schuster, P., Zundel, G., Sandorfy, C.: The Hydrogen Bond. Recent Developments in Theory and Experiments. Weinheim, *Wiley* **2003**.

(13) Steiner, T.: The Hydrogen Bond in the Solid State. Weinheim, *Angewandte Chemie International Edition in English* **2002**, *41*, 48 - 76.

(14) Bruno, I. J., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J., Taylor, R., Cole, J. C.: Designing a New Multi-Component API Form Based on a Known Structure. *Acta Crystallographica* **2002**, *58*.

(15) Hall, S. R., Allen, F. H., Brown, I. D.: The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A* **1991**, *47*, 655-685.

(16) Van De Streek, J.: Searching the Cambridge Structural Database for the `best' representative of each unique polymorph. *Acta Crystallographica Section B* **2006**, *62*, 567-579.

(17) Datta, S., Grant, D. J. W.: Crystal structures of drugs: Advances in determination, prediction and engineering. *Nature Reviews Drug Discovery* **2004**, *3*, 42-57.

(18) Bott, R. C., Smith, G., Wermuth, U. D., Dwyer, N. C.: Molecular Cocrystals of Aromatic Carboxylic Acids with Unsymmetrically Substituted Ureas. The Structures of Phenylurea and the 1 : 1 Adducts of Phenylurea with a Series of Nitro-Substituted Acids. *Australian Journal of Chemistry* **2000**, *53*, 767-777.

(19) Zhang, L., Wen, Y., Yao, Y., Xu, J., Duan, X., Zhang, G.: Synthesis and Characterization of PEDOT Derivative with Carboxyl Group and Its Chemo/Bio Sensing Application as Nanocomposite, Immobilized Biological and Enhanced Optical Materials. *Electrochimica Acta* **2014**, *116*, 343-354.

(20) Viger-Gravel, J., Leclerc, S., Korobkov, I., Bryce, D. L.: Direct Investigation of Halogen Bonds by Solid-State Multinuclear Magnetic Resonance Spectroscopy and Molecular Orbital Analysis. *Journal of the American Chemical Society* **2014**, *136*, 6929-6942.

(21) Obanijesu, E. O., Gubner, R., Barifcani, A., Pareek, V., Tade, M. O.: The influence of corrosion inhibitors on hydrate formation temperature along the subsea natural gas pipelines. *Journal of Petroleum Science and Engineering* **2014**, *120*, 239-252.

(22) Kelland, M. A.: Tetrahydrofuran hydrate crystal growth inhibition by bis-and tris-amine oxides. *Chemical Engineering Science* **2013**, *98*, 1-6.

(23) Graton, J., Berthelot, M., Besseau, F., Laurence, C.: An Enthalpic Scale of Hydrogen-Bond Basicity. 3. Ammonia, Primary, Secondary, and Tertiary Amines. *The Journal of Organic Chemistry* **2005**, *70*, 7892-7901.

(24) Butler, A. R., Feelisch, M.: Therapeutic Uses of Inorganic Nitrite and Nitrate: From the Past to the Future. *Circulation* **2008**, *117*, 2151-2159.

(25) Griess, P.: Über Metadiamidobenzol als Regens auf Saltpetrige Säure. *Chemische Berichte* **1878**, *11*, 624 – 627.

(26) Gladwin, M. T., Shelhammer, J. H., Schechter, A. N., Pease-Fye, M. E., Waclawiw, M. A., Panza, J. A., Ognibene, F. P., Cannon, R. O.: Role of circulating nitrite and S-nitrosohemoglobin in the regulation of regional blood flow in humans. *Proceedings of the National Academy of Sciences* **2000**, *97*, 11482–11487.

(27) Dejam, A., Hunter, C. J., Tremonti, C., Pluta, R. M., Hon, Y. Y., Grimes, G., Partovi, K., Pelletier, M. M., Oldfield, E. H., Cannon, R. O., Schechter, A. N., Gladwin, M. T.: Nitrite infusion in humans and nonhuman primates: endocrine effects, pharmacokinetics, and tolerance formation. *Circulation* **2007**, *116*, 1821–1831.

(28) Lundberg, J. O., Weitzberg, E., Gladwin., M. T.: The nitrate-nitrite-nitric oxide pathway in physiology and therapeutics. *Nature Reviews Drug Discovery* **2008**, *7*, 156–167.

(29) Klemenska, E., Beresewicz, A.: Bioactivation of organic nitrates and the mechanism of nitrate tolerance. *Journal of Cardiology* **2009**, *16* 11–19.

(30) Ali-von Laue, C., Zoschke, C., Do, N., Lehnen, D., Küchler, S., Mehnert, W., Blaschke, T., Kramer, K. D., Plendl, J., Weindl, G., Korting, H. C., Hoeller Obrigkeit, D., Merk, H. F., Schäfer-Korting, M.: Improving Topical Non-Melanoma Skin Cancer Treatment: In vitro Efficacy of a Novel Guanosine-Analog Phosphonate. *Skin Pharmacology and Physiology* **2014**, *27*, 173-173.

(31) Valasani, K. R., Hu, G., Chaney, M. O., Yan, S. S.: Structure-Based Design and Synthesis of Benzothiazole Phosphonate Analogues with Inhibitors of Human ABAD-Aβ for Treatment of Alzheimer's Disease. *Chemical Biology & Drug Design* **2013**, *81*, 238-249.

(32) Coremans, G., Snoeck, R.: Cidofovir: clinical experience and future perspectives on an acyclic nucleoside phosphonate analog of cytosine in the treatment of refractory and premalignant HPV-associated anal lesions. *Expert Opinion on Pharmacotherapy* **2009**, *10*, 1343-1352.

(33) Luz, Z. R., A., Holmberg, R. W., Silver, B. L. : ESR of ^{17}O-Labeled nitrogen dioxide trapped in a single crystal of sodium nitrite. *The Journal of Chemical Physics* **1969**, *51*, 4017–4024.

(34) Ohara, A., Marcelo, G. B., Angélica, M. A., Edlaine, L., Célio, C. X. S., Sílvia, L. M.: Nitrogen dioxide and carbonate radical anion: Two emerging radicals in biology. *Free Radical Biology & Medicine* **2002**, *32*, 841–859.

(35) Kono, Y., Shibata, H., Adachi, K., Tanaka, K.: Lactate-Dependent Killing of Escherichia coli by Nitrite plus Hydrogen-Peroxide: A Possible Role of Nitrogen Dioxide. *Archives of Biochemistry and Biophysics* **1994**, *311*, 153-159.

(36) Ignarro, L. J.: Biosynthesis and metabolism of endothelium-derived nitric oxide. *Annual Review of Pharmacology and Toxicology* **1990**, *30*, 535-60.

(37) Bescós, R., Sureda, A., Tur, J. A., Pons, A.: The Effect of Nitric-Oxide-Related Supplements on Human Performance. *Sports Medicine* **2012**, *42*, 99-117.

(38) Epstein, F. H. M. D., Moncada, S., Higgs, A.: The L-arginine-nitric oxide pathway. *The New England Journal of Medicine* **1993**, *329*, 2002-2012.

(39) http://www.who.int/ accessed on 22 August 2014.

(40) Briggs, N. E. B., Kennedy, A. R., Morrison, C. A.: 42 salt forms of tyramine: structural comparison and the occurrence of hydrate formation. *Acta Crystallographica Section B* **2012**, *B68*, 453-464.

(41) Zhu, H., Grant, D. J. W.: Influence of water activity in organic solvent + water mixtures on the nature of the crystallizing drug phase. 2. Ampicillin. *International Journal of Pharmaceutics* **1996**, *139*, 33-43.

(42) Ticehurst, M. D., Storey, R. A., Watt, C.: Application of slurry bridging experiments at controlled water activities to predict the solid-state conversion between anhydrous and hydrated forms using theophylline as a model drug. *International Journal of Pharmaceutics* **2002**, *247*, 1-10.

CHAPTER 4

Potential energy surface of creatine monohydrate

Introduction

In the past there have been several attempts to identify a particular factor that influences hydrate formation in pharmaceutical compounds. For example, Briggs *et al.* performed a study to investigate the influence of the total number of potential H-bond donor and acceptor atoms on hydrate formation. Their results showed an increase in hydrate formation with an increase in the number of H-bond acceptor and donor atoms[1]. A study by Zhu *et al.* investigating the effect of water activity on hydrate formation suggested that the water activity in the crystallization medium directly influences the level of hydration of a particular system[2]. In another study Fabbiani *et al.* managed to obtain the monohydrate form of γ-aminobutyric acid (GABA) by keeping the temperature low and increasing the pressure to 0.4–0.8 GPa, suggesting that an increase in pressure can increase the propensity for hydrate formation[3]. The only common factor in all of these studies was the presence of an H-bond when the hydrated product was obtained. The role of the H-bond in hydrate formation is not yet fully understood, but a review by Datta *et al.* identified it as a major intermolecular force in the formation of most pharmaceutical crystals[4]. Rather than one single factor influencing the hydration of compounds, it has been suggested that a combination of the H-bond, water activity in the crystallization medium, donor/acceptor ratio, pressure and temperature determines if hydrate formation will take place[5]. In this chapter we will focus on the role of the H-bond, by studying the variation in strength of the H-bond in creatine monohydrate by performing a series of Potential Energy Surface (PES) scans on the structure of creatine monohydrate. From this study we wish to identify if the H-bond interaction is strong enough to drive hydrate formation.

4.1. The properties and structure of creatine monohydrate

Creatine monohydrate is a common pharmaceutical compound that promotes the supply of energy to cells of the body, particularly muscle tissue. The effect of chronic supplementation with creatine monohydrate has been shown to promote increases in total intramuscular creatine, phosphocreatine, skeletal muscle mass, lean body mass and muscle fibre size. Several studies have provided clear evidence that creatine monohydrate supplementation will lead to a significant increase in muscular strength and power[6]. More recent studies have also explored the ability of creatine monohydrate to quench reactive oxygen and nitrogen species, such as the H_2O_2 and $ONOO^-$ species. A study performed by Stefani *et al.* showed that rats receiving creatine monohydrate supplementation along with an 8-week resistance training program experienced decrease in oxidative stress[7]. Some creatine analogues have also proven to be potent anticancer agents that act synergistically with currently-used chemotherapeutics[8]. There are, however, some adverse effects if creatine monohydrate supplements are not introduced according to a recommended daily dosage or if the supplements contain impurities. Possible adverse effects consist of acute tubular necrosis (ATN), interstitial nephritis and ultimately, complete renal failure[9].

The physicochemical, mechanical, and biological properties of creatine monohydrate are dependent on the stability of the crystalline structure of the drug, which many studies suggest is determined by H-bonding[10]. Therefore we will investigate the strength of the H-bond in creatine monohydrate and how it benefits the stability of the structure. Additionally, we would like to identify other intermolecular interactions present in the creatine monohydrate crystal structure. If these prove to be stronger than the known H-bond in creatine monohydrate, they could possibly be manipulated to improve the stability of the drug. Creatine monohydrate was specifically chosen as it is small and therefore relatively computationally inexpensive and the crystal structure listed in the CSD as refcode CREATH04, was been determined by neutron diffraction. As mentioned in Chapter 2, neutron diffraction provides accurate positions for the hydrogen atoms, which are involved in the H-bond under investigation in this study.

The crystal structure of creatine monohydrate obtained by neutron diffraction analysis is characterised by a strongly directional H-bond between the water molecule and the carboxylate functional group[11]. Furthermore, creatine monohydrate is a zwitterionic compound with the negative charge localised on the carboxylate functional group and the positive charge on the double bonded amine group. The structure of creatine monohydrate is illustrated in Figure 4.1.

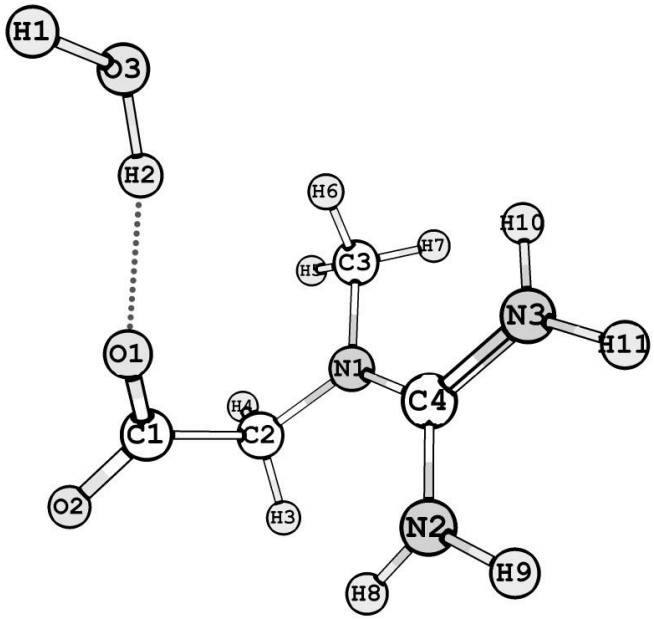

Figure 4.1. Illustration of the structure of creatine monohydrate.

4.2. The potential energy surface (PES)

Various methods can be used to determine the electronic properties of structures. Scans are a computational technique where one or more internal coordinates are varied in order to investigate the potential energy surface (PES). Scans are commonly used for the analysis of chemical processes and the stabilities of structures[12]. The result of a scan is a two-dimensional plot of energy versus the chosen internal coordinate. We can identify two classes of stationary points (point with a gradient of zero) on a PES, namely minimum and maximum stationary points. For the purpose of this study, all minimum energy stationary points correspond to stable structures and maximum stationary points correspond to transition states (TSs) or saddle points. These TSs lie on the highest point of the reaction pathway and connect the minima corresponding to the reactants and products[13]. It should be noted that the scans done in this study are only in one dimension, where the minima and maxima in 1D scans could correspond to saddle points, which are positions on the PES that correspond to a minimum in all directions except one.

A number of different coordinate systems can be used when calculating PESs. The Cartesian coordinate system is usually avoided as it cannot allow the variation of single parameters. Therefore a system with internal coordinates, more commonly referred to as a Z-matrix, is preferred. This set of internal coordinates defines a structure in terms of bond lengths, angles and dihedral angles. The use of a Z-matrix enables the variation of a single parameter to generate a PES[14].

4.3. Construction of PES scans of creatine monohydrate

The structure of creatine monohydrate was exported from the CSD in Cartesian coordinates with the Mercury suite and subsequently converted to a Z-matrix. In this study we performed the four PES scans that were described in Chapter 2, namely the vertical, horizontal, left-to-right diagonal and right-to-left diagonal scans. The water molecule is linked to the movement of the dummy atom, as illustrated in Figure 4.2 for each scan.

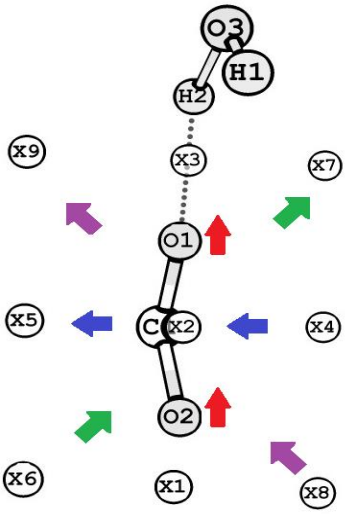

Figure 4.2. Definition of the four angle ranges used for the scans. The vertical scan (X1-X3, indicated by red arrows); the horizontal scan (X4-X5, indicated by the blue arrows); the left-to-right diagonal scan (X6-X7, indicated by the green arrows) and the right-to-left diagonal scan (X8-X9, indicated by the purple arrows).

Each scan was further constructed to enable the arched movement of the dummy atoms from 0 to 180° (θ) as illustrated in Figure 4.3. This particular motion keeps the interaction distance between the H2 and O1/O2 atoms below the sum of the van der Waals radii, while varying the other geometrical parameters. The PES scans will describe the geometrical dependence of the H-bond interaction, which could be used in order to improve the stability of creatine monohydrate. It should be noted that 1D fixed scans were performed instead of relaxed scans. As mentioned previously, the neutron structure of creatine monohydrate does not require geometry optimisation. By keeping the creatine fragment fixed, we can determine the direct relationship between the position of the water molecule and the strength of the H-bond, whereas relaxed scans could possible provide yield more accurate energies, they were beyond the scope of this study.

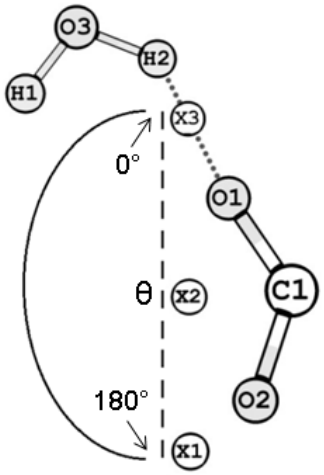

Figure 4.3. The arched movement of the dummy atoms for the vertical scan.

The angle θ was increased in step sizes of 5° starting at 0°. Single point calculations were performed at each step by defining water and creatine fragments A and B respectively. The interaction energy was then calculated according to equation 7 discussed in Chapter 2 with the water and creatine fragments defined as A and B, respectively:

$$\Delta E_{INT}\,(AB) = E(AB) - \big(E(A) + E(B)\big).$$

Counterpoise corrections were applied to all interaction energy calculations. Furthermore, all positive E_{INT} values represent a destabilising effect on the system, which is mainly due to the steric density of the surrounding environment having a stronger destabilising effect than the stabilising effect of the H-bond interaction. All negative E_{INT} values indicate that the H-bond interaction energy results in an overall stabilising effect of the system. All DFT and WFT calculations were performed with the Gaussian 09 rev. B.01 and rev. D.01 packages[15]. The DFT methods consisted of M062X, B3LYP, B97D and B3LYPD3 utilising the 6-311++G(d,p) basis set. Larger basis sets were not utilised as they proved to be too computationally expensive for this study. The WFT methods were performed at the HF/6-311++G(d,p), MP2/6-311++G(d,p) and CCSD(T)/6-311++G(d,p) level of theory. MP2 and

CCSD(T) are known to describe van der Waals interactions by incorporating correlation effects. The CCSD(T) method is generally accepted to be the gold standard in quantum chemistry and it often serves to provide benchmark data to be used for the validation of other methods[16]. It has been the method of choice for over 20 years to obtain accurate bond energies and molecular properties. The computational cost of using this method is too expensive for large molecules, as it scales to the seventh power (N^7) of the size of the system[17]. The accuracies of all methods were therefore investigated by employing various WFT and DFT methods and comparing them to calculations performed at the CCSD(T) benchmark level of theory. From this we hoped to identify a cost-effective method that could be used for geometry optimisation calculations.

In particular, DFT methods have been designed to estimate exchange-correlation effects, with each method parameterised to provide accurate results for certain systems. DFT methods are parameterised by calculating interaction energies for various systems at the CCSD(T) benchmark level of theory and scaling the DFT method to reproduce benchmark results[18,19]. Meanwhile, the MP2 method is the most economical WFT electronic structure method that provides a description of electrostatics and van der Waals interactions, such as induction and dispersion. However, the greatest shortcoming of the MP2 theory is an overestimation of the dispersion interaction energy[20].

4.4. Potential energy surface (PES) curves of creatine monohydrate

The variation in potential energy with changes in the geometric orientation of the water molecule relative to creatine for the vertical scan is shown in Figure 4.4. The PES curve is a graphical representation of energy in kcal/mol plotted against angle θ for seven different computational methods all using the 6-311++G(d,p) basis set. Table 4.1 provides an example of energy values corresponding to the θ values that were used to generate the CCSD(T) curve. Interaction energies at θ = 0° and 180° could not be calculated as PES scans that are described by a Z-matrix are unable to define systems at these two angles. The aim is to find geometries that are more stable than the neutron structure, by identifying minima in the PES curves.

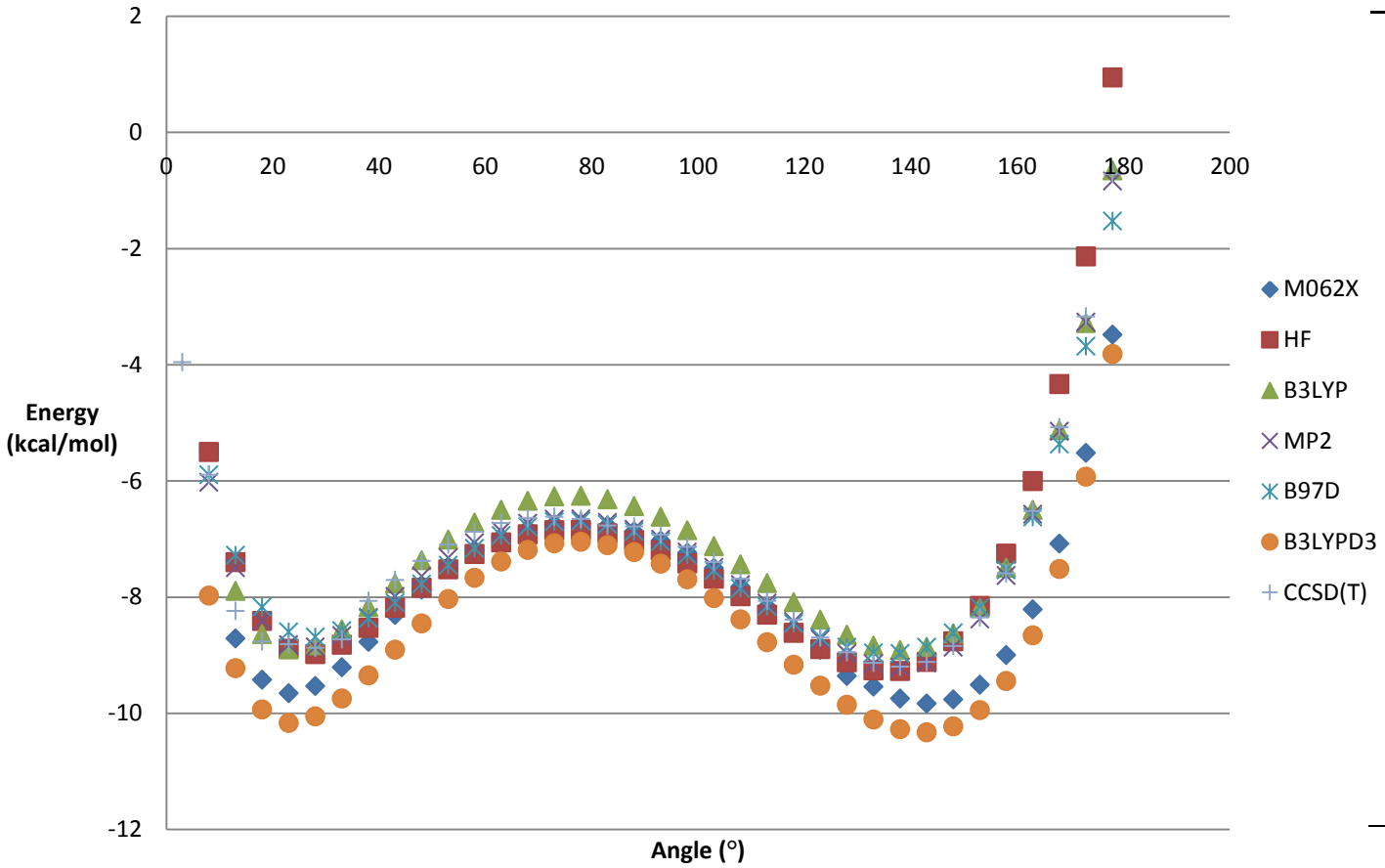

Figure 4.4 Results of vertical scan with energy (kcal/mol) plotted against angle (°) plotted for the M062X, HF, B3LYP, MP2, B97D, B3LYPD3 and CCSD(T) methods utilising the 6-311++G(d,p) basis set.

Table 4.1. Values used to plot PES Scan 1 at the CCSD(T)/ level of theory.

Scan 1 CCSD(T)/6-311++G(d,p)			
Angle (°)	Energy (kcal/mol)	Angle (°)	Energy (kcal/mol)
0	-	95	-7.14
5	-5.90	100	-7.40
10	-8.24	105	-7.69
15	-8.77	110	-8.07
20	-8.81	115	-8.39
25	-8.88	120	-8.70
30	-8.73	125	-8.96
35	-8.07	130	-9.13
40	-7.71	135	-9.20
45	-7.38	140	-9.12
50	-7.10	145	-8.85
55	-6.88	150	-8.35
60	-6.73	155	-7.60
65	-6.64	160	-6.52
70	-6.62	165	-5.08
75	-6.66	170	-3.18
80	-6.77	175	-0.72
85	-6.79	180	-
90	-6.93		

Upon inspection of PES curve obtained with the CCSD(T) method, we are able to identify two minima and one maximum that occur at θ = 28°, 78° and 138°.

The minima that were obtained at the CCSD(T) level of theory served as the benchmarks when we evaluate the computational methods by comparison to the PES curves. The bond length, bond angle and bond energy parameters of H-bond interaction found at the two minima are summarised in Table 4.2.

Table 4.2. Structural parameters of the H-bonded minimum energy conformations shown in Figure 4.4.

Minimum	1	2	Maximum	1
bond lengths [Å]	2.03	2.09	bond lengths [Å]	3.072
bond angles [°]	158	156	bond angles [°]	173.94
bond energy [kcal/mol]	-8.88	-9.20	bond energy [kcal/mol]	-6.66

Table 4.2 lists the geometrical parameters of the two minima and one maximum in the CCSD(T) curve shown in Figure 4.4. We observe similar geometrical parameters for the two minima, as might be expected from the symmetrical shape of the curve in Figure 4.4, which is mainly due to the movement of the water molecule in the vertical scan. The similarity between the structural parameters of the two minima is graphically illustrated in Figure 4.5.

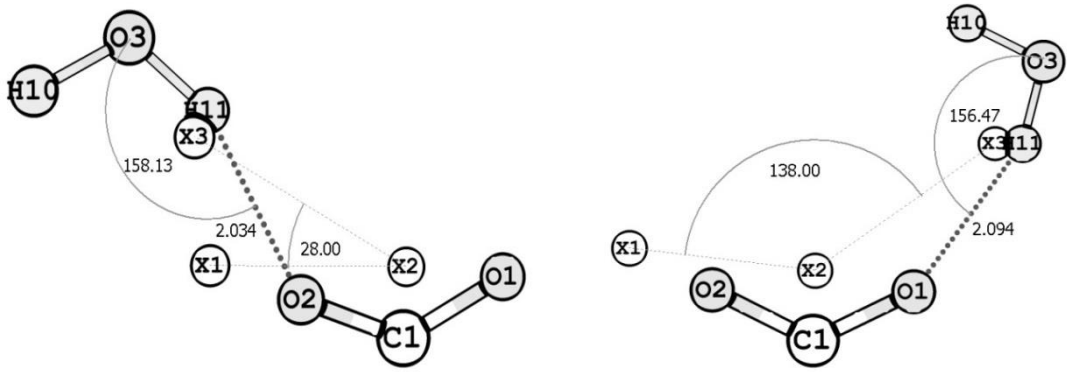

Figure 4.5. Graphical illustration of the structural parameters where θ = 28°, 138°.

The maximum could correspond to a saddle point as it corresponds to an attractive interaction energy, however, the nature of this maximum was not investigated further as it is beyond the scope of this study. The rapid increase in H-bond interaction energy below θ values of 20 and above 140 is due to the collision of the water molecule with the atoms in creatine (Figure 4.1). We can compare these energies to that of the creatine-water complex in the experimental neutron structure, where the bond length (H---A) and bond angle are 1.756 Å and 145.48° respectively. Calculation of the bond strength at the CCSD(T)/6-311++G(d,p) level of theory delivered an interaction energy of 11.40 kcal/mol. Several studies have classified a strong H-bond to have an interaction energy in the range of 15 – 40 kcal/mol and to have linear interaction orientation of the interacting moeities[21,22]. From the parameters listed in Table 4.2 and the calculated strength of the H-bond in the experimental neutron structure we cannot define one of the H-bonds in the minimum energy conformations as a strong H-bond.

4.5. The horizontal, left-to-right diagonal and right-to-left diagonal PES scan

In order to identify all possible minimum energy conformations of the water molecule relative to creatine, three additional scans were performed. PES curves for the horizontal, left-to-right diagonal and right-to-left diagonal scan (see Chapter 2) are shown in Figures 4.6, 4.9 and 4.11 for various levels of theory with CCSD(T) serving as the benchmark method. The description of the movement of the water molecule relative to creatine for each of these scans can be found by referring back to Figure 4.2. These three additional scans were chosen to search for additional H-bond interactions by covering the largest surface area.

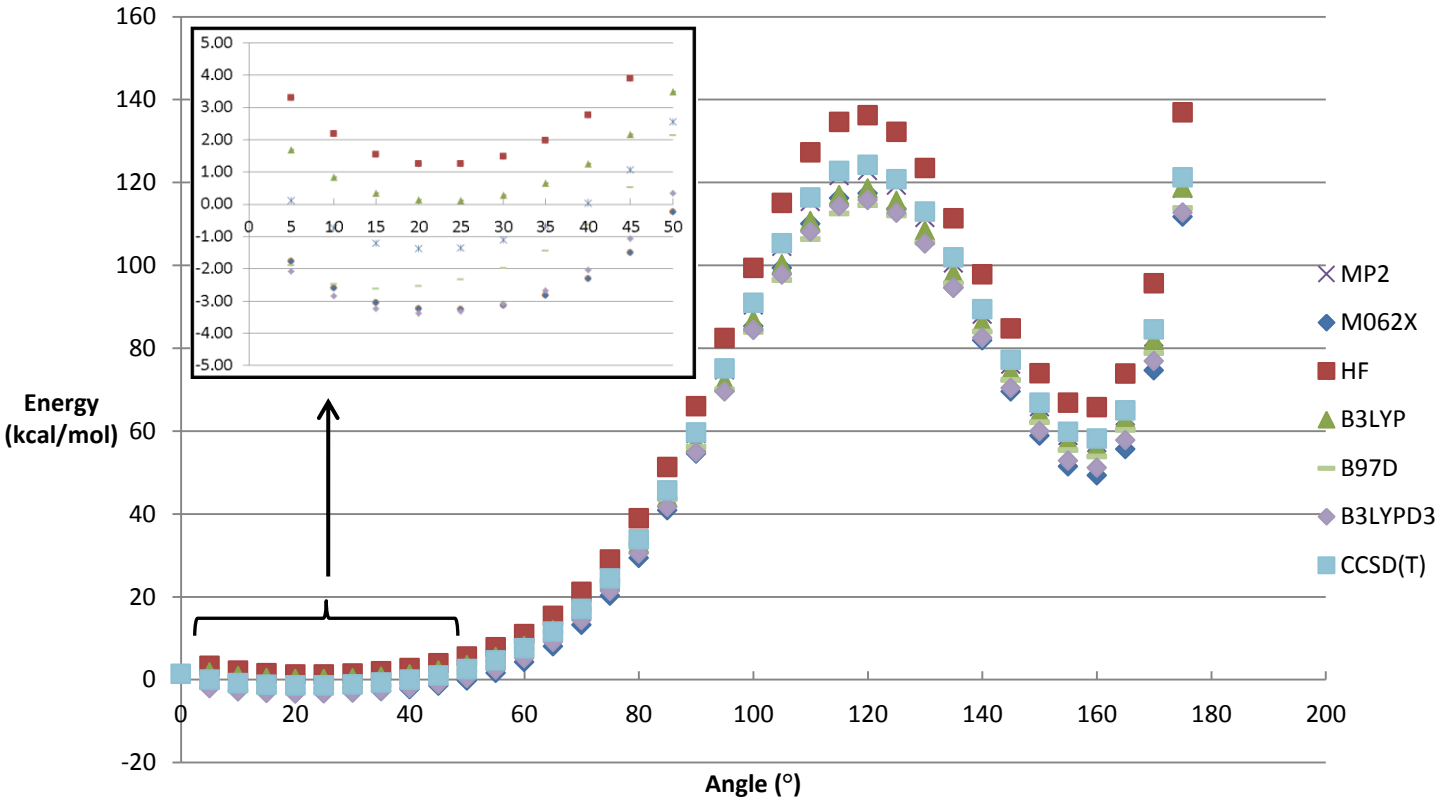

Figure 4.6. Results of horizontal scan with energy (kcal/mol) plotted against angle (°) for the M062X, HF, B3LYP, MP2, B97D, B3LYPD3 and CCSD(T) methods utilising the 6-311++G(d,p) basis set.

In Figure 4.6 we observe two minima and one maximum for the horizontal scan, of which only one minimum exhibits an attractive interaction energy. Various PES curves overlap in the area where this minimum is located, which is why the 0° to 50° region was expanded to show the minimum found in this range of θ values. The repulsive minimum and maximum H-bond result from the collision of the water molecule with atoms of creatine. Furthermore, even though very strong H-bonds are characterised by a short bond length, Wang *et al.* have suggested that there are two types of contributions to the overall strength of the H-bond, namely an attractive and a repulsive contribution[23].

The strength of each type will vary according to how well the orbitals overlap. The strongly repulsive interaction in the horizontal scan would suggest that at a very short distance, the repulsive contribution to the interaction energy is dominant, which results in an overall repulsive interaction. The geometrical parameters of the maximum energy structure are shown in Figure 4.7.

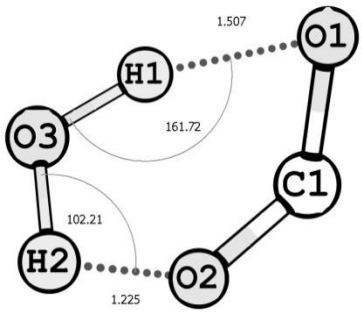

Figure 4.7. Schematic of the H-bond geometry found for the maximum energy conformation in the horizontal scan, where θ = 120°.

The geometrical parameters of the two minimum energy conformation H-bonds for the horizontal scan are shown in Figure 4.8.

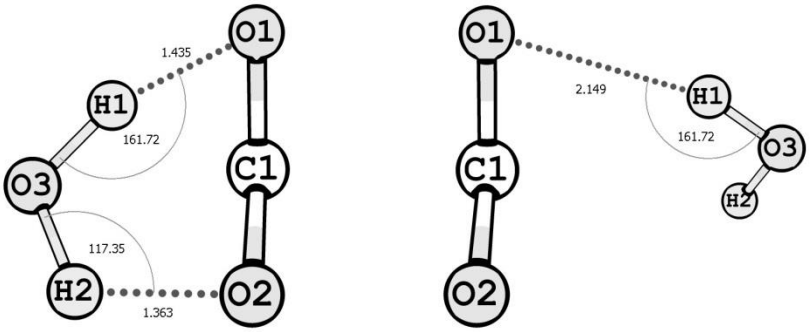

Figure 4.8. Schematic of the H-bond geometries in the horizontal scan, where θ = 20° (left) and 160° (right).

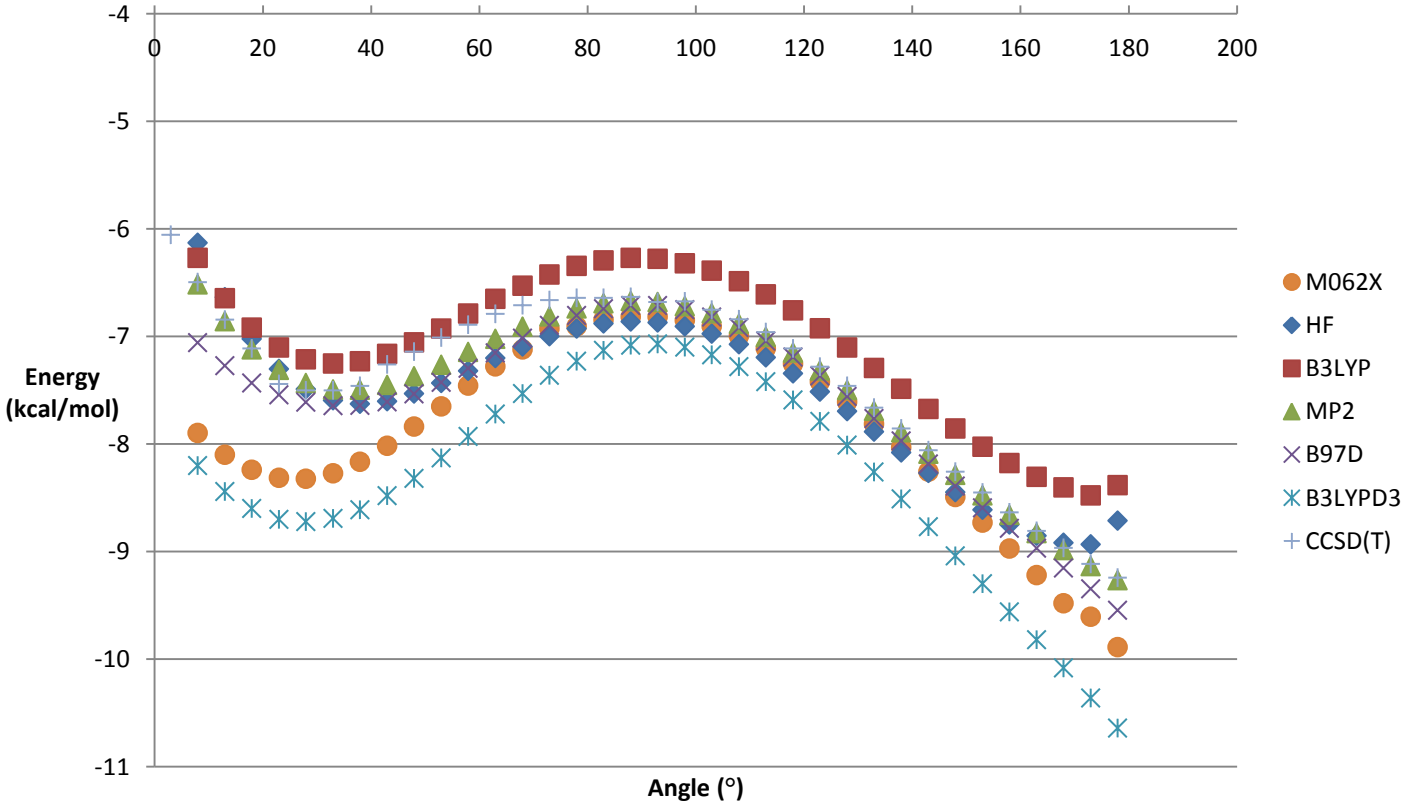

Figure 4.9. Results of left-to right diagonal scan with energy (kcal/mol) plotted against angle (°) plotted for the M062X, HF, B3LYP, MP2, B97D, B3LYPD3 and CCSD(T) methods utilising the 6-311++G(d,p) basis set.

The left-to-right diagonal scan shown in Figure 4.9 exhibits two minima and one saddle point. This figure shows a poor agreement between PES curves for this scan, showing the inability of some DFT methods to accurately reproduce the CCSD(T) benchmark results for the PES scan, which could possibly be due to the overestimation of dispersion by some of the methods in Figure 4.9. All of the values in Figure 4.9 describe attractive interaction energies, which suggests that the geometrical parameters of the water molecule relative to creatine for this particular scan will always be attractive in this region and slightly changing these parameters could even result in a stronger H-bond. This was confirmed by the fact that the experimental neutron structure in Figure 4.1 shows an alternative H-bond in this region, suggesting that steric repulsion from creatine is low in this region. The geometrical parameters of the two minimum energy conformation H-bonds for the left-to-right scan are shown in Figure 4.10.

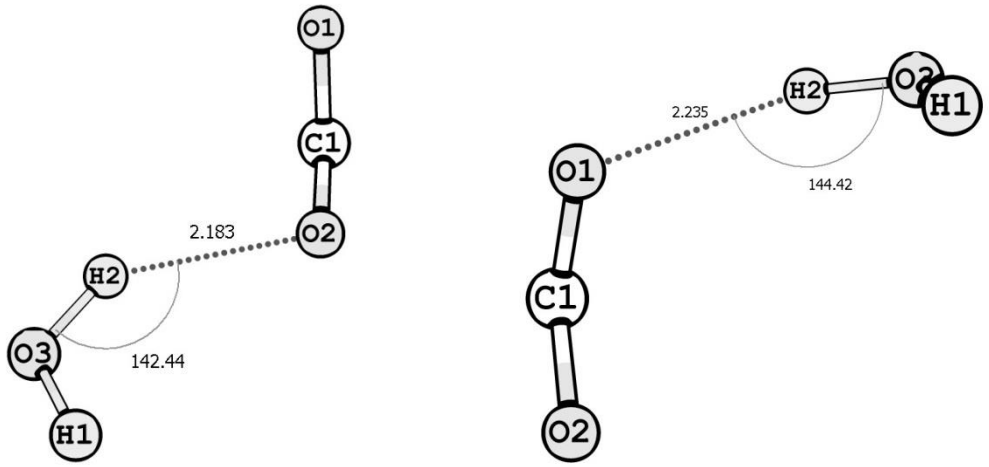

Figure 4.10. Schematic of the H-bond geometries in the left-to right diagonal scan, where θ = 28° (left) and 178° (right).

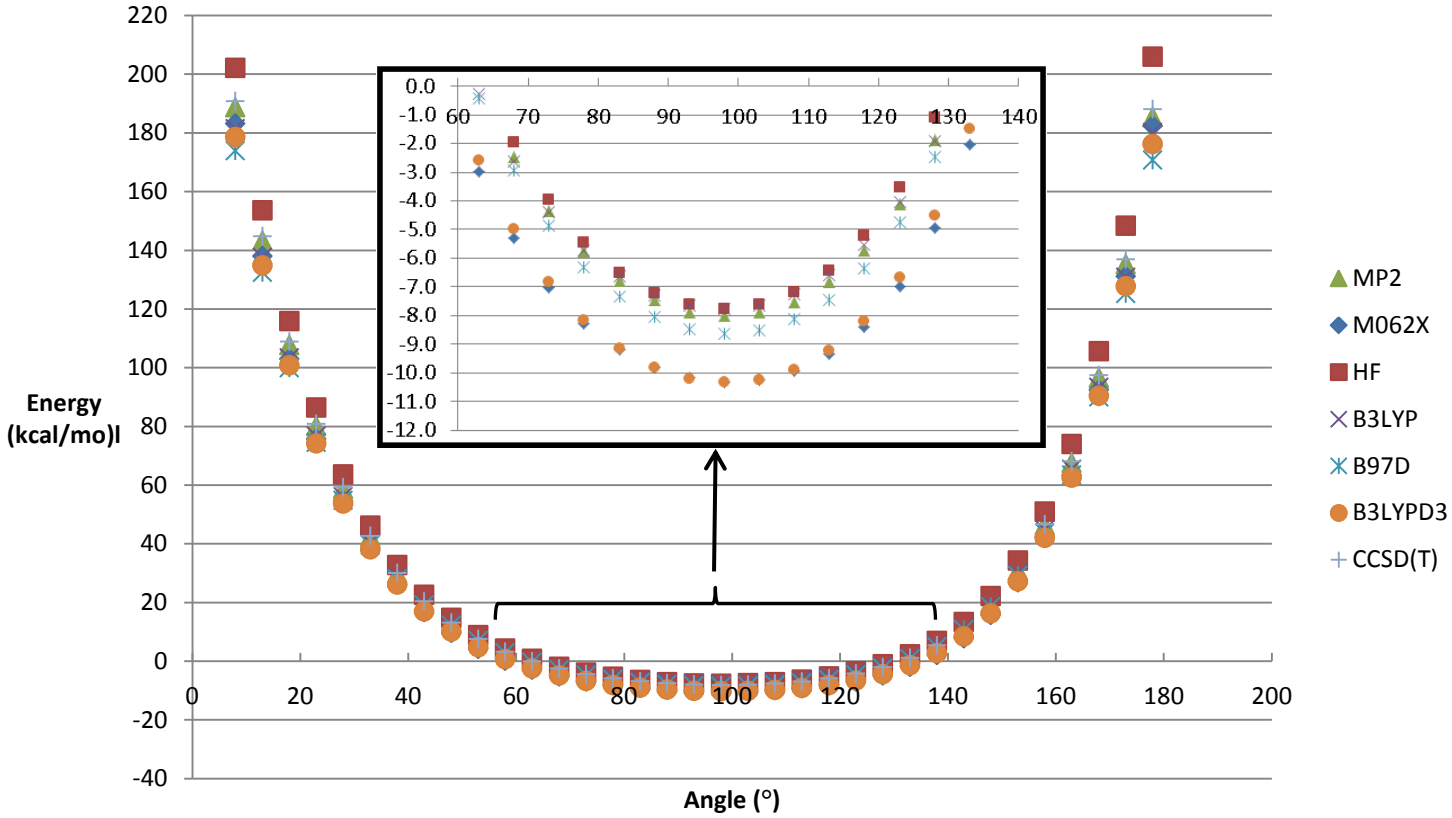

Figure 4.11. Results of right-to-left diagonal scan with energy (kcal/mol) plotted against angle (°) plotted for the M062X, HF, B3LYP, MP2, B97D, B3LYPD3 and CCSD(T) methods utilising the 6-311++G(d,p) basis set.

The right-to-left diagonal scan shown in Figure 4.11 shows a clear minimum and similar shape for all methods. The interaction energy values between 60° and 140° were expanded for clarity. The repulsive interaction energies found at θ values below 60° and above 140° result from the steric repulsion between the water molecule and the creatine. When comparing the shape of the curves in Figure 4.8 and Figure 4.11, it is clear that the steric density of neighbouring substituents play a much larger role in the repulsive interaction energies found in Figure 4.11.

The geometrical parameters of the minimum energy conformation H-bond for the right-to-left scan are shown in Figure 4.12.

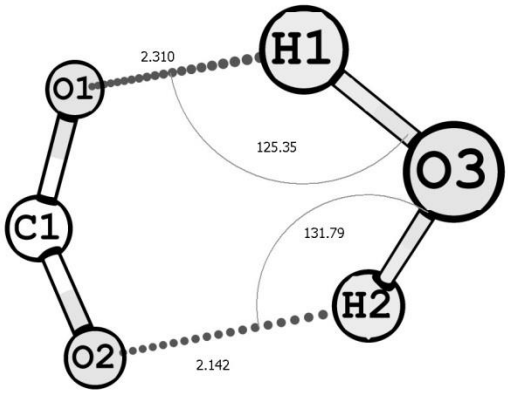

Figure 4.12. Schematic of the H-bond geometries in the right-to-left diagonal scan, where θ = 98°.

The H-bonds at different θ values that correspond to minimum energy conformations obtained at the CCSD(T)/ 6-311++G(d,p) level of theory can be compared to the parameters in Table 4.2. We chose to only compare the attractive H-bond interactions of each scan.

The minima obtained with each scan can be compared to their respective bond length, angle and strength parameters. It was also found in some of the schematics of the H-bond geometries (see Figure 4.8, Figure 4.10 and Figure 4.12) that the position of the water molecule relative to creatine can be identified as having more than one H-bond. The parameters of both of these H-bonds will be considered where applicable. Only attractive minima were considered and attractive maxima will also not be considered as the true identity of the interactions at these points can only be determined by performing PES scans in more than one dimension, which is beyond the scope of this study.

Table 4.3. Bond length, angle and strength parameters of the minima in the vertical, horizontal, left-to right diagonal and right-to-left diagonal scans obtained at the the CCSD(T)/ 6-311++G(d,p) level of theory.

		Vertical		Horizontal	LTR[a] diagonal		RTL[b] diagonal
		θ [°]		θ [°]	θ [°]		θ [°]
H-bond #	Parameters	28	138	20	28	178	98
1	Bond length [Å]	2.03	*	*	2.18	2.23	2.14
	Angle [°]	158	*	*	142.4	144.4	108.7
2	Bond length [Å]	*	2.09	2.14	*	*	2.31
	Angle [°]	*	156.4	161.7	*	*	101.85
	Bond strength [kcal/mol]	-8.9	-9.2	-1.4	-7.5	-9.2	-8.2

* The interactions were omitted as they are weak, repulsive or do not occur.

[a] LTR represents left-to-right

[b] RTL represents right-to-left

The bond lengths, angles and H-bond strengths for the minima in the vertical, horizontal, left-to right diagonal and right-to-left diagonal scan are summarised in Table 4.3. Not one of the bond strengths in Table 4.3 were found to be stronger than the H-bond in the experimental neutron structure. We did however notice that the experimental neutron structure has similar parameters to the H-bonds identified in the left-to-right scan is due to the use of fixed scans, rather than relaxed scans, so that the optimum geometries were not identified in order to obtain the minimum energy conformations. Full optimisation of the minimum energy conformations identified from the fixed scans was undertaken. Choosing the appropriate method to perform these calculations was done by comparing the minimum energy obtained in the four scans with those obtained at the CCSD(T)/6-311++G(d,p) benchmark level of theory as is described in 4.6.

4.6. Comparing the accuracy of various methods

In order to determine the most effective computational method, we need to identify the parameter that provides the best description of the accuracy of results obtained with each method. From the scans shown in Figures 4.4, 4.6, 4.9 and 4.11 it is clear that small changes in the geometrical parameters of the H-bond can result in significant changes in the strength of the H-bond. Therefore we identified the bond strength as the parameter that will be used to compare the accuracy of each method. Only the attractive interaction energy minima from each method that are similar to the values in Table 4.3 will be considered when identifying the most effective computational method. Furthermore, the study is simplified by only comparing the bond strength energies at each of the PES curve minima. It is important to mention that the minima obtained with each method are found at slightly different θ values and are thus included in the table below.

Table 4.4. H-bond interaction energies (kcal/mol) for the minimum energy conformations at different angle θ values (°) of each scan calculated at various levels of theory.

	Vertical		Horizontal	LTR diagonal		RTL diagonal
Angle (°)	28	138	20	28	178	98
CCSD(T)	-8.88	-9.20	-1.43	-7.50	-9.24	-8.17
Angle (°)	28	138	20	38	178	98
MP2	-8.86	-9.22	-1.38	-7.50	-9.27	-8.02
Angle (°)	28	138	15	33	178	98
B97D	-8.69	-8.97	-2.62	-7.64	-9.55	-8.63
Angle (°)	23	143	20	28	178	98
B3LYPD3	-10.17	-10.33	-3.37	-8.72	-10.64	-10.33
Angle (°)	23	143	25	28	178	98
M062X	-9.84	-9.97	-3.26	-8.32	-9.89	-10.31
Angle (°)	23	138	25	33	173	98
B3LYP	-8.91	-8.92	0.120	-7.25	-8.48	-7.77
Angle (°)	28	138	25	38	173	98
HF	-9.84	-9.28	1.24	-7.63	-8.93	-7.76
Average	-9.31	-9.41	-1.88	-7.79	-9.43	-8.71

Table 4.4 contains the interaction energies corresponding to each minimum energy conformation calculated at various levels of theory. There are notable differences in the interaction energies, but the two most pronounced results are found for the horizontal scan, where the B3LYP and HF methods yield repulsive interaction energies of 0.120 and 1.24 kcal/mol, respectively for the weak H-bond. These two methods are unable to identify the minimum in the PES scan as an attractive interaction, which is due them being unable to describe the exchange repulsion force at the short intermolecular distance in the horizontal scan[24]. The effectiveness of each method can be measured by calculating the root mean square deviations (RMSD) values of each method for the four scans.

Table 4.5. RMSD values (kcal/mol) of each method when compared values obtained at CCSD(T)/ 6-311++G(d,p).

Method	RMDS (kcal/mol)
MP2	0.067
B97D	0.553
B3LYPD3	1.572
M062X	1.325
B3LYP	0.740
HF	1.179

The RMSD values of each method are summarised in Table 4.5, which shows that the MP2 method provides the most accurate results when compared to interaction energy values that were obtained at the CCSD(T)/6-311++G(d,p) level of theory. As mentioned previously, the MP2 method is capable of accurately describing van der Waals interactions, electrostatics and induction, but tends to overestimate dispersion interactions. Furthermore, Table 4.5 shows that the most accurate DFT method is the B97D method. This observation correlates well with a study performed by Grimme *et al.*, which showed that dispersion-corrected DFT methods are capable of recovering a large portion of the non-covalent interaction energy, which allows them to accurately describe H-bond energies[22]. Interestingly, the B3LYP method provides more accurate results than the dispersion corrected B3LYPD3 method, possibly as a result of the D3 correction for B3LYP overestimating the dispersion of this particular system.

We have therefore identified MP2 and B97D as the two most effective methods to yield energies that are comparable to those obtained at the benchmark CCSD(T)/ 6-311++G(d,p) level of theory. The various minima obtained with the four scans have slightly weaker H-bonds compared to the neutron structure, which can be attributed to the rigid geometry in the scans. We can correct for this by performing geometry optimisation calculations with the coordinates of each of the minimum energy geometries.

4.7. Optimising the minimum energy geometries obtained from each scan

The structures corresponding to the lowest energy conformations listed in Table 4.3 were optimised first with B97D and then the MP2 method utilising the 6-311++G(d,p) basis set. Counterpoise-corrected single point calculations were then performed at the CCSD(T)/6-311++G(d,p) level of theory to be compared with the -11.4 kcal/mol interaction energy obtained for the experimental neutron structure at the same level of theory.

Table 4.6. E_{INT} (kcal/mol) values calculated at the CCSD(T)/6-311++G(d,p) level of theory for the optimised geometries of each scan.

Methods	Neutron	Vertical	Horizontal	LTR Diag.	RTL Diag.
θ (°)		138	20	178	98
E_{INT}	-11.4	-16.0	-18.3	-18.3	-13.1

The interaction energies calculated from the optimised geometries are summarised in Table 4.6. The differences between these values and the experimental neutron structure interaction energy are much larger than anticipated. The lower energy in the neutron structure could possibly be due to the effect of the surrounding crystalline environment, which was suggested by Brown *et al.* who found that crystalline environment leads to a pronounced lowering in the total electronic energy of a system[25].

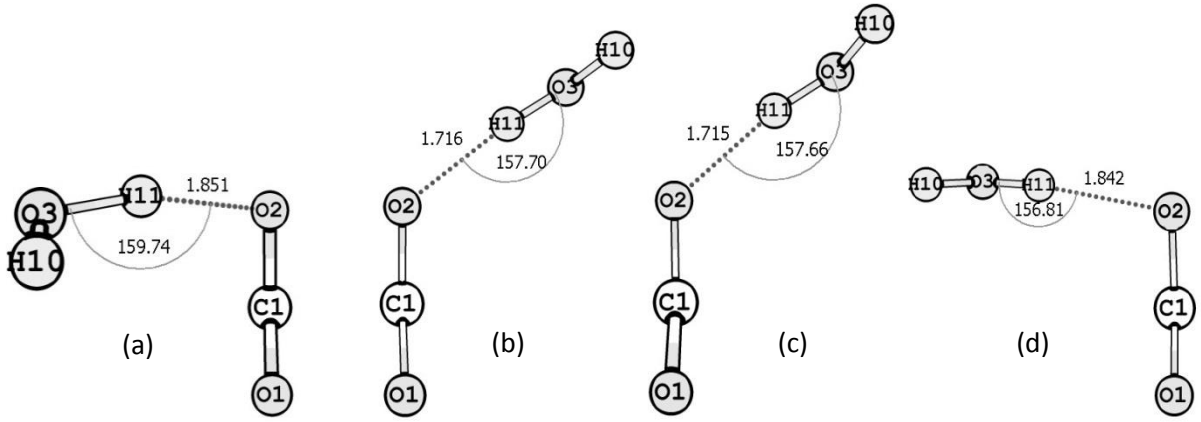

Figure 4.13. Optimised geometries from vertical (a), horizontal (b), left-to-right (c) and right-to-left scan (d).

The relative positions of the water molecules for the various optimised geometries are shown in Figure 4.6. The optimised geometries of the structures identified in the vertical and right-to-left scan are very similar. This is also the case for the optimised structures of the horizontal and left-to-right scan. When comparing the geometrical parameters of these optimised structures and comparing them to the values in Table 4.3, we observe that the bond angles have not changed significantly, whereas changes as large as 0.42 Å are observed for the bond lengths. It should also be mentioned that all of the atoms in creatine were also allowed to optimise, which allows the water molecule to form the strong H-bonds, as indicated by short distances, are shown in Figure 4.6.

The results in Table 4.6 suggest that H-bonds can be significantly stronger stronger than those in the experimental neutron structure if the geometrical parameters are changed. If factors can be identified that result in these geometrical parameters, we could fine-tune the strength of the H-bond in pharmaceutical compounds, which will result in different or improved properties. A study by Bingham *et al.* found that the strength of a H-bond strength is closely linked to the affinity of specific ligands for their target compound[26]. These results suggest that if we can fine-tune the strength of the H-bond to increase from a weak to a strong H-bond, we could significantly increase the affinity of water for a target compound. The presence of a strong H-bond will also result in more stable structures and may even yield new conformations or a complete phase transition. These new phases could possibly have new or improved properties. In subsequent chapters experiments probing the factors that influence H-bond strength will be described. With this knowledge we can fine-tune the H-bond and subsequently the properties of pharmaceutically acceptable compounds.

References

(1) Briggs, N. E., Kennedy, A. R. Morrison, C. A.: *Acta Crystallographica Section B* **2012**, *B68*, 453-464.

(2) Zhu, H., Grant, D. J. W.: Influence of water activity in organic solvent + water mixtures on the nature of the crystallizing drug phase. 2. Ampicillin. *International Journal of Pharmaceutics* **1996**, *139*, 33-43.

(3) Fabbiani, F. P. A., Buth, G., Levendis, D. C., Cruz-Cabeza, A. J.: Pharmaceutical hydrates under ambient conditions from high-pressure seeds: a case study of GABA monohydrate. *Chemical Communications* **2014**, *50*, 1817-1819.

(4) Datta, S., Grant, D. J. W.: Crystal structures of drugs: Advances in determination, prediction and engineering. *Nature Reviews Drug Discovery* **2004**, *3*, 42-57.

(5) Wouters, J., Quere, L.: Pharmaceutical Salts and Co-crystals. Cambridge, *RSC Drug Discovery*, **2012**.

(6) Antonio, J., Ciccone, V.: The effects of pre versus post workout supplementation of creatine monohydrate on body composition and strength. *Journal of the International Society of Sports Nutrition* **2013**, *10*, 36.

(7) Ivchenko, O., Bachert, P., Imhof, P.: Umbrella sampling of proton transfer in a creatine–water system. *Chemical Physics Letters* **2014**, *600*, 51-55.

(8) Wyss, M., Kaddurah-Daouk, R.: Creatine and Creatinine Metabolism. *Physiological Reviews*, *American Physiological Society* 2000, Vol. 80.

(9) Taner, B., Aysim, O., Abdulkadir, U.: The effects of the recommended dose of creatine monohydrate on kidney function. Oxford, *NDT Plus* **2011**, *4*, 23-24.

(10) Sakata, Y., Shiraishi, S., Otsuka, M.: Effect of tablet geometrical structure on the dehydration of creatine monohydrate tablets, and their pharmaceutical properties. *AAPS PharmSciTech* **2005**, *6*, E527-E535.

(11) S. Frampton, C., C. Wilson, C., Shankland, N., J. Florence, A.: Single-crystal neutron refinement of creatine monohydrate at 20 K and 123 K. *Journal of the Chemical Society, Faraday Transactions* **1997**, *93*, 1875-1879.

(12) Alexander, W. A.: On the accuracy of analytical potentials: comment on 'Accurate ab initio calculation of the Ar–CF4 intermolecular potential energy surface'. *Molecular Simulation* **2014**, 1-3.

(13) Dykstra, C. E. F., G. Kim, K.S. Scuseria, G.E. : Theory and Applications of Computational Chemistry The First Forty Years. Amsterdam, *Elsevier* **2005**.

(14) Kamarchik, E., Jasper, A. W.: Anharmonic state counts and partition functions for molecules via classical phase space integrals in curvilinear coordinates. *The Journal of Chemical Physics* **2013**, *138*.

(15) Frisch, M. J., et al.: Gaussian 09, Revision B.01. *Wallingford CT* **2009**.

(16) Ramabhadran, R. O., Raghavachari, K.: Extrapolation to the Gold-Standard in Quantum Chemistry: Computationally Efficient and Accurate CCSD(T) Energies for Large Molecules Using an Automated Thermochemical Hierarchy. *Journal of Chemical Theory and Computation* **2013**, *9*, 3986-3994.

(17) Řezáč, J., Hobza, P.: Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *Journal of Chemical Theory and Computation* **2013**, *9*, 2151-2155.

(18) Huix-Rotllant, M., Filatov, M., Gozem, S., Schapiro, I., Olivucci, M., Ferré, N.: Assessment of Density Functional Theory for Describing the Correlation Effects on the Ground and Excited State Potential Energy Surfaces of a Retinal Chromophore Model. *Journal of Chemical Theory and Computation* **2013**, *9*, 3917-3932.

(19) Tao, J., Perdew, J. P., Staroverov, V. N., Scuseria, G. E.: Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Physical Review Letters* **2003**, *91*, 146401.

(20) Tkatchenko, A., DiStasio, R. A., Head-Gordon, M., Scheffler, M.: Dispersion-corrected Møller–Plesset second-order perturbation theory. *The Journal of Chemical Physics* **2009**, *131*.

(21) Jeffrey, G. A.: An Introduction to Hydrogen Bonding. Oxford, *Oxford University Press* **1997**.

(22) Hujo, W., Grimme, S.: Comparison of the performance of dispersion-corrected density functional theory for weak hydrogen bonds. *Physical Chemistry Chemical Physics* **2011**, *13*, 13942-13950.

(23) Wang, S.-C., Sahu, P. K., Lee, S.-L.: Intermolecular orbital repulsion effect on the blue-shifted hydrogen bond. *Chemical Physics Letters* **2005**, *406*, 143-147.

(24) Zhang, I. Y., Wu, J., Xu, X.: Extending the reliability and applicability of B3LYP. *Chemical Communications* **2010**, *46*, 3057-3070.

(25) Brown, R. D., O'Dwyer, M. F., Roby, K. R.: Effect of ionic lattices on electronic structures of polyatomic ions. *Theoretica Chimica Acta* **1968**, *11*, 1-7.

(26) Bingham, A. H., Davenport, R. J., Gowers, L., Knight, R. L., Lowe, C., Owen,D. A., Parry, D. M., Pitt, W. R.: A novel series of potent and selective IKK2 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 409-412.

CHAPTER 5

Modelling of pharmaceutical salts in the gas phase and in an implicit solvent model

Introduction

The role of the H-bond in biological and chemical systems has been the topic of research for well over a hundred years[1]. The development of increasingly powerful analytical techniques has helped to identify the various properties of H-bonds. One example is the role of H-bonds in molecular recognition patterns and the use of suitably robust motifs for the construction of crystalline architectures with desired properties[2,3]. These properties can be further investigated by computational methods that are capable of calculating H-bond energies. In this particular study we have investigated how the functional groups as well as the chemical environment affect the strength of the H-bond in pharmaceutically acceptable compounds. These studies were performed by utilising various WFT and DFT methods in combination with various computational tools for the analysis of different properties of electronic systems.

5.1. Method for geometry optimisation

The process of geometry optimisation can be accelerated by using a combination of DFT and WFT methods. In this study, the structure was initially fully optimised with a suitable DFT method, from which the coordinates were subsequently used as a basis to optimise with a WFT method. This procedure provides results much faster than immediately optimising at the highest level of theory.

To choose a suitable DFT method to optimise structures, we optimised the H-Bond of creatine monohydrate with a number of different levels of theory and basis sets using the Gaussian 09 rev. B.01 and rev. D.01 packages. The HF, B3LYP, M06HF, M062X, M06L, PBEPBE, B97D, B2PLYPD and MP2 method utilising the 6-311++G(d,p), cc-pVTZ and aug-cc-pVTZ basis sets were tested for their ability to reproduce H-bond distances in the experimental neutron structure. As mentioned in previous chapters, the structure of creatine monohydrate has neutron data available, which provides accurate hydrogen atom positions. The Cartesian coordinates of the creatine monohydrate structure were exported from the CSD utilising the Mercury software package. The Cartesian coordinates were then subsequently converted to internal coordinates in a Z-matrix by using the Molden software package[4]. Instead of performing a full geometry optimisation, only the internal coordinates of the atoms O1 and H11 were optimised. The specific bond distance, angle and torsion angle parameters of the experimental neutron structure are illustrated in Figure 5.1. These internal coordinates were specifically chosen to allow the water molecule to rotate freely. Optimising only a selected number of internal coordinates dramatically reduced the computational time and resources required for an optimisation calculation.

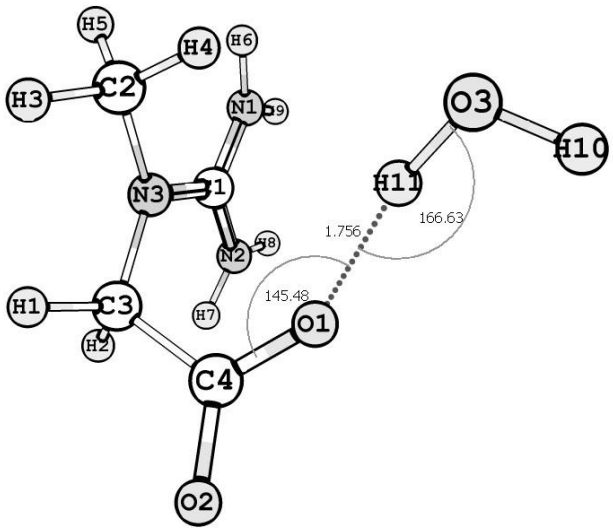

Figure 5.1. The bond distance, angle and torsion parameters of the H-Bond in the experimental neutron structure of creatine monohydrate that were optimised.

Table 5.1. List of methods and basis sets used to optimise the H-bond length in creatine monohydrate. The differences between the O1•••H11 distances of all methods were compared to the distance in Figure 5.1.

Method	Basis set	O1•••H11 (Å)	ΔO1•••H11 (Å)
Experimental		1.756	0.0000
HF	6-311++G(d,p)	1.738	-0.0018
B3LYP	6-311++G(d,p)	1.736	-0.0200
B97D	6-311++G(d,p)	1.728	-0.0280
M06-HF	6-311++G(d,p)	1.738	-0.0018
M06-2X	6-311++G(d,p)	1.740	-0.0016
M06-L	6-311++G(d,p)	1.739	-0.0017
PBEPBE	6-311++G(d,p)	1.722	-0.0340
B2PLYPD	6-311++G(d,p)	1.737	-0.0019
M06-2X	cc-pVTZ	1.740	-0.0016
M06-2X	aug-cc-pVTZ	1.740	-0.0016

The various methods and basis sets used to perform the geometry optimisation are listed in Table 5.1. The accuracy of each method was measured by comparing the interaction distance between atom O1 and H11 to the value of the experimental neutron structure in Figure 5.1. As mentioned previously, the HF method is a very basic method that does not take electron correlation effects into account[5]. The B3LYP method is defined as an exchange-correlation hybrid functional due to the fact that the method has an exchange functional consisting of a pure DFT exchange and an exact Hartree-Fock exchange[6]. Furthermore, the B3LYP method has been parameterised to account for intermolecular interactions and is known to provide accurate results for geometry optimisation calculations[7], including H-bond geometries[8]. However it does not describe dispersion well, therefore the remainder of the methods tested were chosen for their ability to model dispersion effects.

The PBEPBE method is an exchange–correlation method that makes use of generalized gradient approximations (GGAs), which generally improve upon the accuracy of the local-spin-density (LSD) approximation by applying various corrections[9,10]. These corrections will not be discussed here; for further details please see the paper by Perdew *et al.*[11]. The B97D method is also based on the GGA and has been fitted using the dispersion correction developed by Grimme[12]. The various M06 methods in Table 5.1 are meta-GGA methods, which potentially improve on the GGA methods by including the second derivative of the electron density (the Laplacian). The M06 methods have also been described by several authors to model H-bonded systems extremely well when performing geometry optimisations and calculating interaction energies, with results comparable to high level of theory WFT methods[13-15]. The B2PLYPD method is an exchange-correlation method that is based on GGA with an additional perturbative second-order correlation part. This makes the method a mixture of a standard DFT with an MP2 method[16,17].

The choice of basis set can also have a significant effect on the description of the orbitals in a molecule, hence three basis sets were tested. 6-311++G(d,p) is an all-electron basis set with additional p-functions added to the hydrogen atoms and d-functions to heavier atoms. The cc-pVTZ basis set (correlation consistent type) has three basis functions (triple-zeta) to represent each orbital and has p-functions and d-functions added to the hydrogen atoms and heavier atoms respectively. The cc-pVTZ basis set applies a pseudopotential functional to help recover the correlation energy of the valence electrons. The aug-cc-pVTZ basis set has additional diffuse functions (aug prefix) that help describe loosely bound electrons, e.g. the study of anions.

The results in Table 4.1 show that the M06 methods provide the most accurate representation of the H-bond distance when compared to the experimental neutron structure. The PBEPBE method proved to be the least accurate method to reproduce the H-bond distance, which is possibly due to the corrections of this particular method resulting in an inaccurate representation of system used in this study. A study by Karr *et al.* found that the PBEPBE method is capable of overestimating the dispersion forces of H-bonded systems, which results in the shortening of H-bond interaction distances[18]. Table 4.1 shows that the bond distance obtained with the PBEPBE method is considerably shorter, suggesting that the PBEPBE method does perhaps overestimate the dispersion forces in the system used in this study. The HF method also provided very good results, but owing to its inability to take electron correlation effects into account, it was not considered as a viable method for calculations that will be performed in this study.

The M06-2X method provided the most accurate results in Table 4.1. To test the basis set dependence of the M06-2X method, calculations were also performed at the M06-2X/cc-pVTZ and M06-2X/aug-cc-pVTZ levels of theory. The results are also included in Table 4.1 and show no significant increase in accuracy. Therefore, in the following study M06-2X/6-311++G(d,p) will be used to analyse the role of the functional group on the H-bond strength of a range of compounds forming H-bonds with water.

5.2. CSD searches for H-bonds in different functional groups

The CSD was systematically searched for carboxylates, nitroxides, nitrogen dioxide, nitrates, sulfonates, phosphates and phosphonates that are in the hydrated form and contain an H-bond. As mentioned in previous chapters, the nitrogen dioxide functional group cannot be defined as a pharmaceutically acceptable anion, but it was still included in this study as it is a neutral radical group that very often forms H-bonds in pharmaceutical compounds. In order to obtain more hydrates, zwitterions were not excluded from these searches. The structures obtained from the searches were only selected if they were of an appropriate size (< 50 atoms), as the computational cost of a system will increase exponentially with the number of atoms. Furthermore, the structures were only chosen if they adhere to the following requirements:

- X-ray or Neutron data
- Only organic structures
- 3-D coordinates must be determined
- No disorder

Table 5.2. Refcodes of hydrated salt and zwitterionic structures for each functional group.

Carboxylates	Codes	Nitroxide	Codes	Sulfonates	Codes	Phosphate	Codes
ARGIND11*	C1	JODCAH	N1	CBZSUL01*	S1	SPINPH01*	P1
BINFAF02*	C2	ACOHAC	N2	CYSTAC01*	S2	WEFBIU	P2
HISTCM12*	C3	CACGUI01	N3	PICSUL01*	S3	YAJPAB	P3
LECFIJ*	C4	AMHPYR	N4	BBZSUL01*	S4	**Phosphonate**	
LSERMH10*	C5	**Nitrogen dioxide**		SLBZAC01*	S5	SATHIE01*	P4
LYSCLH11*	C6	NINDOD01*	N5	TFMSUL02*	S6	BACDOY	P5
ISIFUM	C7	LIPWEM02	N6	TOLSAM12*	S7	FEBZER	P6
IWIRAH	C8	COWYAP	N7	ABERAB	S8	KIBLOW	P7
WOBXIU	C9	IVUKUG	N8	NOFXEL	S9	ZIVBIP	P8
LAPFAL	C10	XECBEO	N9	ROFNUW	S10	OKUWIB	P9
		Nitrate		TIXGEM	S11		
		BONKOE01	N10	EVOJUV	S12		
		GOLFAP*	N11				

*These structures have neutron radiation data available

Table 5.2 lists the CSD refcodes of the 42 structures investigated in this study, listed by functional group. The codes assigned to each structure in Table 5.2 will be used in subsequent sections. The number of structures for each functional group varies according to the trends described in Chapter 3. It should also be noted that only the components involved in the H-bond between the water molecule and the functional group were used for geometry optimisations, similar to the calculations performed for creatine (see Figure 5.1). This reduces the computational cost and time, whilst still allowing us to investigate the effect of the functional group on the strength of the H-bond. Furthermore, the geometries of structures that have neutron data available was not optimised.

5.3. Geometry optimisation

Geometry optimisations were performed using a combination of the M06-2X and MP2 methods. It was observed that the two components involved in the H-bond in some hydrates moved apart during the geometry optimisation. This effect is illustrated in Figure 5.2 and shows the H-bond disappearing in structure C9. The structures would continue to oscillate and a minimum energy conformation would never be found.

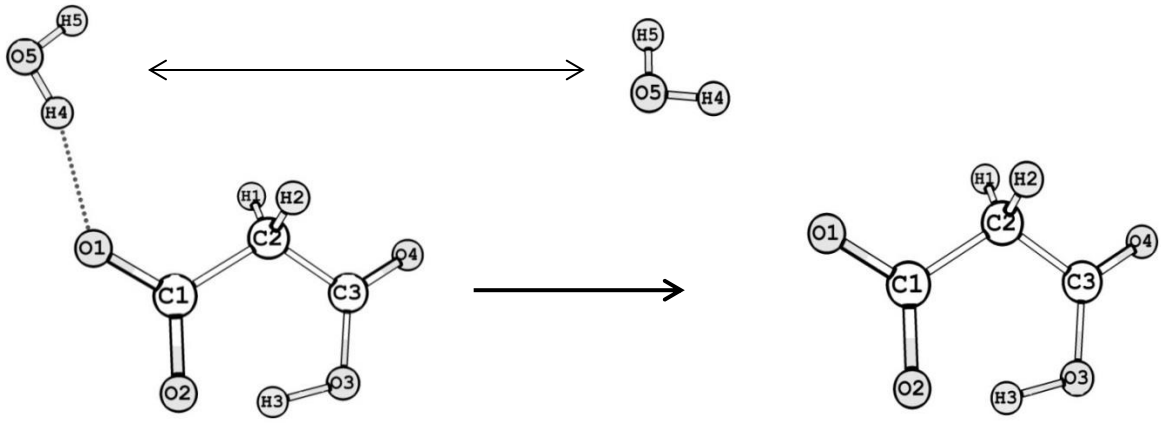

Figure 5.2. Illustration of the water molecule moving away from structure C9 during geometry optimisation.

A solution to this problem was to separate the geometry optimisation procedure into three parts, such that a specific set of the internal coordinates were optimised and the rest of the system was kept rigid. The system in Figure 5.2 will be used to describe the procedure. All three steps involved optimisation at the M06-2X/6-311++G(d,p) level of theory:

- Part 1: In the first part we only optimised the distance, angle and torsion angle parameters of the two atoms involved in the H-bond, e.g. O1 and H4 were optimised.
- Part 2: The coordinates from Part 1 were used to optimise all atoms involved in the H-bond. This was done by allowing the water molecule to optimise along with all atoms in the functional group, e.g. H5, O5, H4, O1 and C1.
- Part 3: The coordinates from Part 2 were used to fully optimise the system.

By using this procedure, it was possible to find optimised geometries for all of the structures in Table 5.2 with an H-bond between the functional group and the water molecule. After a structure was fully optimised at the M06-2X/6-311++G(d,p) level of theory, it was then subsequently optimised using the MP2 method in combination with the cc-pVTZ and aug-cc-pVTZ basis sets.

5.4. The H-bond strength with different functional groups

The H-bond strength was determined for each structure by performing counterpoise corrected single point calculations[19]. Interaction energies were calculated with various DFT and WFT methods. The success of the GGA and meta-GGA methods in optimising the H-bond led to the addition of four additional methods, namely B971, B97-D3, M06-2X-D3 and TPSS. B971 is a GGA exchange-correlation based method chosen known to yield accurate energetics[20,21]. Both the B97-D3 and the M06-2X-D3 method have been fitted to include the 3rd generation dispersion correction developed by Grimme *et al.* in 2011[22]. The B971 and M06-2X methods and their dispersion corrected counterparts were included to investigate the effect of the dispersion correction on the accuracy of the methods in our study. The TPSS method is a meta–GGA based method that has been parameterised to yield correlation energies, where LSD and GGA fail badly. Furthermore, studies have shown that the TPSS provides exceptionally accurate energies for H-bonded systems[23]. All of the DFT methods were used in combinations with the 6-311++G(d,p) basis set. Benchmark values were calculated at the MP2/aug-cc-pVTZ level of theory.

Table 5.3. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the carboxylate functional group.

Carboxylates						
Structure	**M062X***	**M062X-D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
C1	-9.82	-10.10	-8.14	-9.16	-9.65	-9.68
C2	-6.74	-6.85	-5.64	-6.40	-6.23	-6.44
C3	-10.74	-10.94	-8.58	-9.61	-9.59	-9.73
C4	-6.16	-6.35	-4.91	-5.58	-5.73	-6.20
C5	-11.30	-11.45	-9.46	-10.42	-10.01	-10.12
C6	-3.21	-3.35	-2.04	-2.72	-2.56	-3.48
C7	-18.49	-16.92	-14.07	-15.11	-14.86	-16.05
C8	-18.92	-18.20	-15.40	-16.32	-16.84	-17.44
C9	-8.33	-8.03	-6.59	-7.23	-6.64	-8.33
C10	-15.30	-15.49	-12.95	-13.93	-13.57	-13.97
Average	-11.81	-11.65	-9.63	-10.51	-10.32	-10.76

Table 5.4. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the nitroxide functional group.

Nitroxides						
Structure	**M062X***	**M062X-D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
N1	-9.34	-9.49	-6.21	-7.17	-7.60	-6.79
N2	-9.06	-9.26	-5.99	7.25	-7.45	-7.77
N3	-8.15	-9.54	-6.78	-7.95	-7.61	-8.14
N4	-10.97	-11.13	-6.68	-8.11	-8.23	-9.28
Average	-9.38	-9.86	-6.42	-7.74	-7.72	-8.00

Table 5.5. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the nitrogen dioxide functional group.

Nitrogen dioxide						
Structure	**M062X***	**M062X-D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
N5	-8.16	-8.27	-7.19	-7.42	-7.35	-7.83
N6	-5.85	-5.99	-5.59	-5.89	-5.97	-5.46
N7	-9.44	-9.60	-7.01	-8.00	-7.82	-7.61
N8	-10.37	-10.51	-8.72	-9.37	-9.46	-8.89
N9	-4.57	-5.29	-4.16	-4.76	-4.82	-4.32
Average	-7.68	-7.93	-6.53	-7.09	-7.08	-6.82

Table 5.6. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the nitrate functional group.

Nitrates						
Structure	**M062X***	**M062X -D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
N10	-17.94	-17.99	-16.24	-16.73	-16.47	-16.19
N11	-17.95	-17.99	-16.24	-16.73	-16.47	-16.19
Average	-17.95	-17.99	-16.24	-16.73	-16.47	-14.72

Table 5.7. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the sulfonate functional group.

Sulfonates						
Structure	**M062X***	**M062X-D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
S1	-14.69	-14.84	-13.61	-14.29	-14.46	-13.44
S2	-10.24	-10.43	-8.39	-9.29	-9.32	-9.43
S3	-6.97	-7.10	-5.71	-6.42	-6.00	-6.37
S4	-9.63	-9.83	-7.53	-8.57	-8.47	-9.16
S5	-9.85	-10.06	-8.42	-9.17	-9.45	-9.37
S6	-11.76	-11.96	-10.48	-11.20	-11.21	-11.02
S7	-9.68	-9.79	-8.95	-9.31	-9.27	-9.04
S8	-12.41	-12.55	-10.47	-11.15	-11.60	-11.39
S9	-18.06	-18.21	-14.29	-15.52	-16.00	-15.46
S10	-18.17	-18.32	-14.35	-15.58	-16.04	-15.54
S11	-17.26	-15.61	-13.32	-13.32	-13.59	-13.60
S12	-34.05	-32.61	-31.47	-31.99	-31.90	-30.90
Average	-14.40	-14.28	-12.25	-12.98	-13.11	-12.90

Table 5.8. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the phosphate functional group.

Phosphates						
Structure	**M062X***	**M062X -D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
P1	-25.65	-36.27	-24.29	-25.08	-25.38	-24.18
P2	-19.02	-19.12	-15.69	-16.70	-16.90	-16.37
P3	-17.18	-17.24	-14.51	-15.43	-15.65	-15.23
Average	-20.62	-24.21	-18.16	-19.07	-19.31	-18.59

Table 5.9. List of intermolecular interaction energies (kcal/mol) for each structure calculated with different DFT methods for the phosphonate functional group.

Phosphonates						
Structure	**M062X***	**M062X-D3***	**TPSS***	**B971***	**B97-D3***	**MP2****
P4	-10.90	-18.04	-15.28	-16.43	-16.63	-14.59
P5	-29.51	-29.69	-25.28	-26.56	-27.40	-27.34
P6	-21.00	-21.26	-18.26	-19.12	-20.16	-13.41
P7	-17.27	-17.06	-14.26	-14.91	-15.50	-16.01
P8	-14.92	-14.68	-12.29	-11.90	-11.48	-13.58
P9	-32.16	-32.30	-25.83	-27.88	-27.10	-26.98
Average	-20.05	-22.17	-18.53	-19.47	-19.71	-18.70

The interaction energies calculated at different levels of theory for each functional group are summarised in Tables 5.3-5.9. In these tables all DFT and WFT methods utilised the 6-311++G(d,p) and aug-cc-pVTZ basis set, represented by one asterisk (*) and two asterisks (**) respectively.

Table 5.10. Average E_{INT} values for each functional group from the benchmark MP2 level of theory.

Functional group	**E_{INT}**
Carboxylates	-10.76
Nitroxides	-8.00
Nitrogen dioxide	-6.82
Nitrates	-14.72
Sulfonates	-12.90
Phosphates	-18.59
Phosphonates	-18.70

Table 5.10 shows the average E_{INT} values obtained for each functional group. As discussed in previous chapters, we expect the strength of the H-bond to vary according to the polarity and electronegativity of each functional group.

By using this simple approach we can hypothesise that the H-bond strength will vary according to the following trend:

Phosphate > Phosphonate > Sulfonate > Nitrate > Carboxylate > Nitroxide > Nitrogen Dioxide

We however observe the following trend from the values in Table 10:

Phosphonate > Phosphate > Nitrate > Sulfonate > Carboxylate > Nitroxide > Nitrogen dioxide

Firstly, the phosphate group is slightly weaker than the phosphonate group owing to only three structures contributing to the average E_{INT} value for the phosphate group, resulting in a slightly weaker bond. The two functional groups that do not follow the expected trend are the nitrate and sulfonate group. This can be explained by the five contributions to the total energy of the H-bond, which are electrostatics (E_{el}), polarization (E_{pol}), charge transfer (E_{ct}), dispersion (E_{disp}), and exchange repulsion (E_{er}). Studies have shown E_{el} and the E_{ct} have the largest contribution to the strength of the H-bond. The compounds used to obtain the data in Table 5.7 consist of substituted sulfonates, which allow the negative charge to be distributed over the entire structure. On the other hand, the compounds in Table 5.6 consist of unsubstituted nitrates so that the negative charge of the nitrate functional group is distributed over the four atoms in the functional group, leading to H-bonds with a larger charge transfer contribution to the total energy[24,25]. We have also previously discussed the directionality of the H-bond and how small changes in the bond length, angle and torsion angle of the H-bond can have a significant effect on the H-bond strength. The absence of any substituents on the nitrate group will correspondingly also allow an H-bond to form with fewer geometrical parameter constraints than for the sulfonates.

There is also a great variation in H-bond strength between the structures within each functional group. Considering the benchmark E_{INT} values in Table 5.3, the difference between the weakest (E_{INT} = -3.48 kcal/mol) and the strongest H-bond (E_{INT} = -17.44 kcal/mol) is 13.96 kcal/mol. These two H-bonds could have completely different properties, as they are classified as weak and strong H-bonds according to Jeffrey[26]. Weak H-bonds have an electrostatic and dispersive interaction type, while the strong H-bonds are strongly covalent. The large difference in the nature of the H-bonds strongly suggests that the different substituents on the functional groups can play a crucial role in the strength and properties of the H-bond.

5.5. Accuracy of DFT methods reproducing WFT results

We proceeded to test the accuracy of each DFT method by calculating root mean squared deviations (RMSD) from the results obtained by calculations performed at the benchmark MP2/aug-cc-pVTZ level of theory. The RMSD values are determined for each functional group by performing the calculation in Equation 1:

$$\text{RMSD} = \sqrt{\frac{\sum(E_{DFT} - E_{MP2})^{2}}{N}} \qquad \text{(eq.1)}$$

Where the two symbols, E_{MP2} and E_{DFT}, represent the energy calculated with the DFT method that is to be compared to the MP2 benchmark level of theory.

Table 5.11. The RMSD values (kcal/mol) of each DFT method for each functional group.

	C1-C11	N1-N4	N5-N9	N10-N11	S1-S12	P1-P3	P4-P10	Total
M06-2X*	1.07	2.78	1.08	1.76	1.87	1.94	3.41	2.08
M06-2X-D3*	0.82	1.58	1.26	1.80	1.57	7.25	4.87	2.96
TPSS*	1.43	3.49	0.41	0.05	0.89	0.58	2.87	1.79
B971*	0.67	3.69	0.43	0.54	0.46	0.57	3.02	1.71
B97-D3*	0.96	0.96	0.47	0.28	0.53	0.80	3.76	1.65

The RMSDs of each method for every functional group are summarised in Table 5.11. The total RMSD for all 44 structures was also calculated for each method by calculating the average of the RMSD values calculated for all functional group for each method. The results in Table 5.11 show that all three methods that do not include dispersion corrections provide good results in comparison to the benchmark theory. Following the success of the M06-2X method in the previous chapter and in describing the H-bond in Figure 5.1, it was decided that the small difference in accuracy between M06-2X and other methods in Table 5.11 is not enough for us to select one of the other methods in Table 5.11 for single point calculations. Furthermore, other studies have found that the M06-2X method can reproduce interaction energies that have been obtained at the CCSD(T)/aug-cc-pVTZ level of theory[14,27]. Calculations at this level of theory were however beyond the scope of this study. Therefore, the M06-2X method was used for all subsequent single point calculations.

5.6. The effect of steric density on H-bond interaction strength

The steric bulk of neighbouring groups can have a significant effect on the strength of H-bonds. A study by Filarowski *et al.* investigated steric repulsion in the structure of 2-(N-methyl-iminomethyl)-4-Cl-phenol[28]. Their results showed that the alkyl substituent in this structure pushes the phenol group towards the nitrogen atom of the imino group, which decreases the OH…N distance and increases the strength of the H-bond. The opposite effect can also be anticipated if the alkyl substituent was placed in a different position that would result in the phenol group being pushed away from the nitrogen atom, which increases the OH…N distance and decreases the strength of the H-bond.

The effect of sterics on H-bonded compounds can be assessed by calculating hydrogen bond propensities, which was developed by Galek *et al.* and is incorporated into the Mercury software package[29]. Hydrogen propensity values are determined from a statistical model that calculates the propensity and steric density values of structures by identifying all of the possible bond acceptors and donors with a particular structure. Steric density is determined by the number of bulky substituents that surround an H-bond donor or acceptor functional group, where larger groups will results in an increase of the steric density value[29]. As this model calculates propensity and steric density values for each possible donor-acceptor pair, we will only be discussing the values of the donor–acceptor pairs that correspond to the observed H-bonds in the optimised geometries of the compounds in Table 5.2.

The steric density, hydrogen bond propensity values and H-bond energies calculated at the MP2/aug-cc-pVTZ (indicated by MP2**) level of theory are summarised in Table 5.12 for each compound in values of kcal/mol (E_{INT}). All hydrogen propensity values were above 0.5, indicating that there is more is a 50% chance that the H-bond will be found in the compound. This also confirms the ability of the hydrogen bond propensity model to predict the formation of H-bonds. We have excluded the H-bond donor steric density values from the table as they have a constant value of 2.33, due to the hydrogen atom in the water molecule always acting as the donor group.

Table 5.12. Steric density, hydrogen bond propensity and E_{INT} (kcal/mol) values for the H-bonds of compounds in Table 5.2.

Structure	H-bond donor	H-bond acceptor	E_{INT} (MP2**) (kcal/mol)	Acceptor steric density	Propensity
C1	O2	O3	-9.7	31.1	0.97
C2	O6	O5	-6.4	20.8	0.99
C3	O3	O2	-9.7	32.15	0.97
C4	O5	O4	-6.2	32.54	0.95
C5	O4	O1	-10.1	28.39	0.90
C6	O2	O3	-3.5	32.93	0.97
C7	O4	O3	-16.1	29.17	0.94
C8	O5	O4	-17.4	26.81	0.95
C9	O5	O1	-14.5	20.88	0.91
C10	O5	O1	-14.0	34.22	0.51
N1	O6	O1	-6.8	31.79	0.74
N2	O2	O1	-7.8	39.97	0.77
N3	O2	O1	-8.1	32.24	0.76
N4	O3	O1	-9.3	34.78	0.52
N5	O6	O4	-7.8	28.53	0.89
N6	O5	O3	-5.5	31.74	0.89
N7	O9	O6	-7.6	32.3	0.81
N8	O3	O5	-8.9	30.44	0.89
N9	O7	O2	-4.3	27.35	0.85
N10	O5	O3	-16.2	7.22	1.00
N11	O10	O9	-16.2	7.22	1.00

Table 5.12 (continued).

Structure	H-bond donor	H-bond acceptor	E_{INT} (MP2**) (kcal/mol)	Acceptor steric density	Propensity
S1	O6	O4	-13.4	31.65	0.99
S2	O6	O4	-9.4	32.08	0.99
S3	O11	O7	-6.4	36.87	0.98
S4	O5	O2	-9.2	31.74	0.99
S5	O7	O3	-9.4	34.74	0.98
S6	O7	O3	-11.0	20.22	0.91
S7	O7	O3	-9.0	32.34	0.74
S8	O5	O3	-11.4	31.18	0.99
S9	O6	O5	-15.5	31.18	0.99
S10	O13	O6	-15.5	32.37	0.99
S11	O10	O5	-13.6	34.02	0.99
S12	O1	O3	-30.9	11.05	1.00
P1	O3	O5	-24.2	13.3	1.00
P2	O3	O1	-16.4	15.56	0.91
P3	O19	O5	-15.2	11.05	0.96
P4	O9	O1	-14.6	38.4	0.96
P5	O10	O3	-27.3	31.18	0.99
P6	O8	O3	-13.4	44.28	1.00
P7	O9	O3	-16.0	38.4	0.98
P8	O11	O2	-13.6	32.58	1.00
P9	O3	O14	-27.0	31.18	0.94

It is clear from the table that the compounds with the lowest steric density, i.e. below 15, are those that are the smallest, containing the functional group of interest with only a methyl or ethyl substituent. Furthermore, some of these compounds have some of the strongest H-bonds in Table 5.12. A noticeable example is compound S12, which consists of an unsubstituted sulfonate anion and has a steric density value of 11.05 and interaction energy of -30.9 kcal/mol. On the other hand, steric densities higher than 30 correspond to compounds with bulky substituents, which could result in weak H-bonds. Here, a good example is the -6.2 kcal/mol interaction in compound C4, which has a large steric density value of 32.54.

There are however some exceptions where the steric repulsion in structures results in stronger H-bonds, as suggested by Filarowski *et al.*[28]. A few noticeable examples are C7, C8, S9, S10, P5, P7 and P9. We have mentioned previously that a strong H-bond can be described as having the following parameters: a distance in the range of 1.2 – 1.5 Å; a bond angle of 170-180° and bond energy in the range of 15 – 40 kcal/mol. After investigating the H-bonding parameters of each of these exceptions, several observations can be made. Firstly, the two carboxylate structures have strong H-bond energies, but not long bond distances and angles well below 170°. They are, however, compounds with aromatic rings and substituents on different positions of the ring. These substituents could be involved in other intermolecular interactions or result in various inductive and resonance effects, which could increase or decrease the strength of the H-bond. It was found that of all the remaining compounds, only compound P8 does not contain an aromatic ring, which shows an example of where steric repulsion can lead to a stronger H-bond.

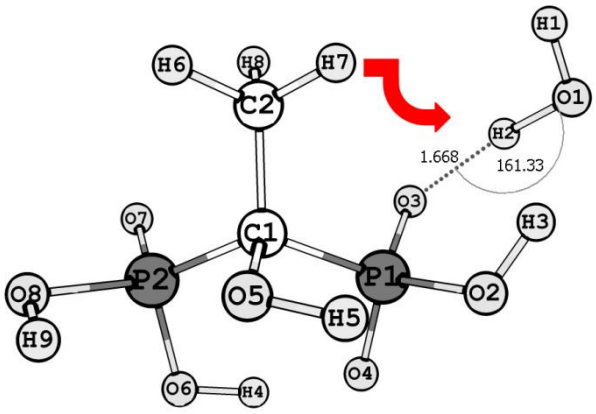

Figure 5.3. Steric repulsion in structure P8.

Figure 5.3 illustrates the steric repulsion from the methyl group forcing the H-bond to be more linear and have a shorter bond distance. It is clear that sterics play a crucial role in the strength of the H-bond, but the role can vary for different systems. The results given in Table 5.12 also show that the effect of substituents on aromatic rings can have a significant effect on the strength of the H-bond and could serve as a possible method for improving or changing the properties of pharmaceutical compounds.

5.7. The effect of the crystalline environment on the H-bond interaction energy

It has been reported that the structural packing in the crystalline environment is greatly influenced by intermolecular interactions[30]. The effect of the crystalline environment on the strength of the H-bond was investigated by initially only optimising the parameters involved in the H-bond and keeping the rest of the structure rigid. This optimisation was required as structures that have been determined by X-ray diffraction have an inaccurate description of the hydrogen atom positions as has been mentioned in previous chapters. Therefore, we can only accurately investigate how crystalline environment affects the H-bond interaction energy once the H-bond parameters of each structure have been optimised. The crystal structures listed in Table 5.2 determined by neutron diffraction analysis were excluded from this study, as these structures were not optimised in the previous study and thus the effect of the crystalline environment could not be investigated for these structures. Structures were optimised with the M06-2X and MP2 method, utilising the 6-311++G(d,p) basis set. Single point calculations were performed for optimised structures at the M06-2X/6-311++G(d,p) and MP2/6-311++G(d,p) levels of theory. All structures that have optimised H-bond parameters will henceforth be referred to as crystalline effect structures for clarity.

Table 5.13. Comparison of E_{INT} values from Table 5.3-5.9 and E_{INT} values that demonstrate the lattice effect.

	E_{INT} (kcal/mol)		
Structure	E_{INT} from Table 3-9 M06-2X	Crystalline effect M06-2X	Difference ΔM06-2X
C7	-18.49	-14.90	-3.59
C8	-17.40	-16.56	-0.84
C9	-17.38	-11.73	-5.65
C10	-15.30	-15.46	0.16
N1	-9.34	-8.90	-0.44
N2	-9.06	-9.19	0.13
N3	-8.15	-7.46	-0.69
N4	-10.97	-11.29	0.32
N5	-8.16	-8.09	-0.07
N6	-5.85	-2.01	-3.84
N7	-9.44	-3.44	-6.00
N8	-10.37	-9.77	-0.60
N9	-4.57	-4.57	0.00
N10	-17.94	-1.63	-16.31
S8	-12.41	-13.50	1.09
S9	-18.06	-15.70	-2.36
S10	-18.17	-10.08	-8.09
S11	-17.26	-11.50	-5.76
S12	-34.05	-28.50	-5.55
P2	-19.02	-20.40	1.38
P3	-17.18	-11.02	-6.16

Table 5.13. Continued.

P5	-29.51	-28.42	-1.09
P6	-20.58	-14.60	-5.98
P7	-17.27	-5.60	-11.67
P8	-14.92	-9.50	-5.42
P9	-32.16	-26.30	-5.86

ΔM06-2X and ΔMP2 represents the difference between the values in Tables 3-9 and the values calculated to measure the effect of the lattice.

The E_{INT} values for crystalline effect structures are summarised in Table 5.13, along with the E_{INT} values listed in Table 5.3-5.9 in order to compare the E_{INT} values of optimised structures to crystalline effect structures. From Table 5.13 we see a great variation in the effect of the lattice on the strength of the H-bond. The majority of the H-bond interactions are stronger in the optimised structures. This is expected as the structure is allowed to optimise towards a minimum energy conformation without the restriction from neighbouring fragments in the lattice.

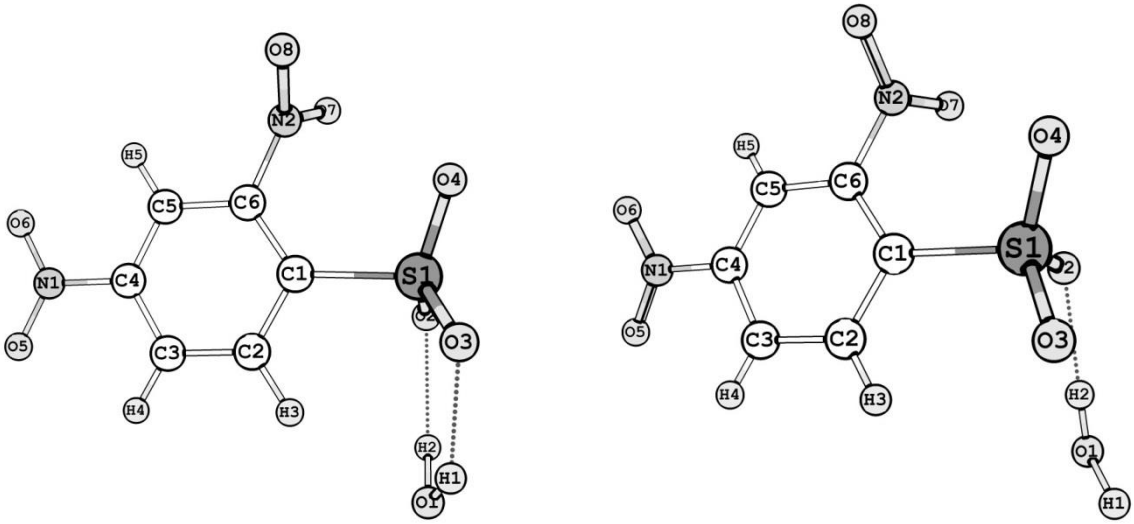

Figure 5.4. The geometries of the fully optimised (left) and the crystalline effect (right) structure S11.

Structure S11 is one of the structures with a weaker H-bond in the crystalline effect structure. The illustration of two structures in Figure 5.4 shows two H-bonds for the fully optimised structure on the left and only one H-bond for the crystalline effect structure on the right. The presence of more than one H-bond has been shown to stabilize structures in a nearly additive manner[31]. Furthermore, multiple intermolecular interactions between two fragments have been described as leading cooperative stabilisation of the non-covalent complex. The magnitude of cooperative stabilisation will depend on the number of interactions as well as the properties of the functional groups[32]. The stronger H-bond for the fully optimised structure in Figure 5.4 can thus be attributed to a cooperative stabilisation from the two H-bonds that are present.

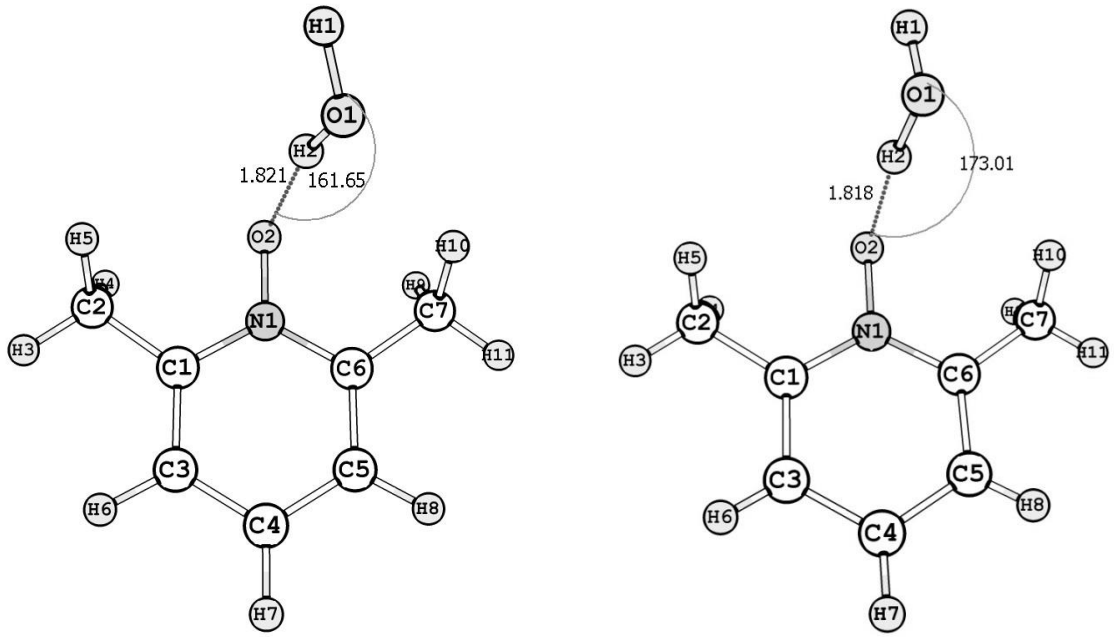

Figure 5.5. The geometries of the fully optimised (left) and the crystalline effect (right) structure for N2.

Structure N2 is one of the structures with a slightly stronger H-bond in the crystalline effect structure. From the illustration of the two structures in Figure 5.5, we observe that the crystalline effect structure has a shorter, more linear H-bond when compared to the fully optimised structure. We have mentioned earlier that more linear bonds will result in stronger H-bonds. The presence of the short distance and more linear bond in the lattice effect structure can be explained by the packing arrangement of structure N2, containing two 2,6-dimethylpyridine N-oxides and one water fragment as illustrated in Figure 5.6.

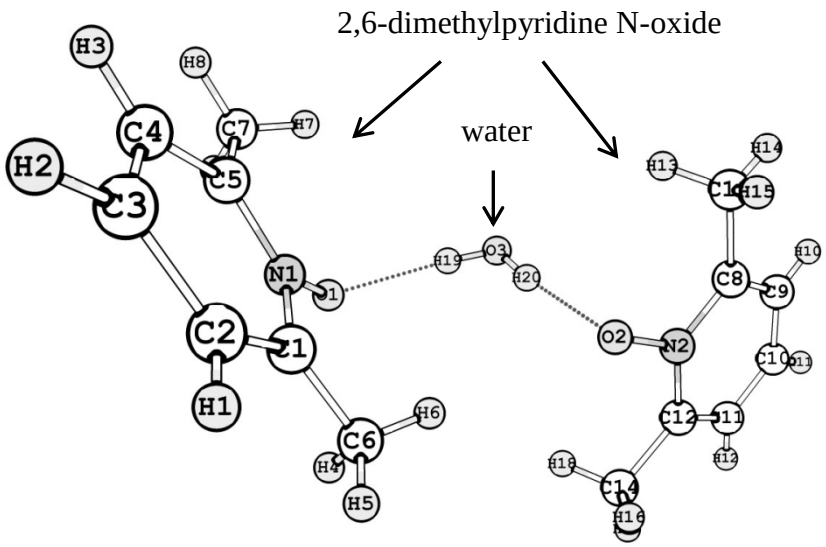

Figure 5.6. Illustration of the packing in structure N2.

Both of the 2,6-dimethylpyridine N-oxide molecules are seen forming a H-bond to the same water molecule in Figure 5.6. Therefore, the specific packing of this particular structure results in the more linear H-bond that is observed in the crystalline effect structure. On the

other hand, when the structure is allowed to optimise, the second H-bond no longer plays a role in the alignment of the water molecule relative to 2,6-dimethylpyridine N-oxide and it loses some of its linear character, resulting in a slightly weaker H-bond.

From the results in this study, we can conclude that effect of the crystalline environment can have a significant effect on the strength of some H-bonds, as it can restrict the interaction of functional groups with the water molecule and thus not allow cooperative stabilisation to strengthen the H-bond interaction. Furthermore, the H-bond is often used as a supramolecular synthon in crystal engineering when designing new crystalline materials[33,34]. The results in this study suggest that H-bond interaction energies are weaker in the crystalline environment and this should be considered when we investigate factors that influence the strength of the H-bond interaction in the gas phase.

5.8. Interaction energy calculations in implicit polarisable continuum

After performing computational studies of the H-bond in compounds in the gas phase described above, we went on to investigate the influence of a changing electrostatic environment on the interaction energy of the H-bond. This was done by optimising five of the H-bonded structures in an implicit polarisable continuum model (PCM). The nature of the H-bond is different for each of the structures, which will allow us to investigate how the solvent will affect each one individually. The five structures consist of one representative structure from five different functional groups. These five structures are P5, C4, N9, N4 and S10 and represent phosphonate, carboxylates, nitrogen dioxide, nitroxide and sulfonates respectively. Five solvent environments were then chosen with increasing relative permittivity values, namely heptane (ε = 1.911), n-octanol (ε = 9.863), 1-propanol (ε = 20.524), dimethyl sulfoxide (DMSO, ε = 46.826) and water (ε = 78.3553). The relative permittivity can be seen as a measure of the polarity of the solvent, with a high relative permittivity value corresponding to a highly polar solvent.

All five structures were fully optimised in all solvents using a combination of the M06-2X and MP2 methods with the 6-311++G(d,p) basis set. All optimised geometries were then subsequently used to calculate H-bond energies at the M06-2X/6-311++G(d,p) level of theory. This particular level of theory was specifically chosen as it provided accurate results in previous studies performed for the same structures. The basis set superposition error (BSSE) correction can be applied in Gaussian by implementing the counterpoise = 2 keyword. The BSSE could however not be corrected by applying the conventional counterpoise correction as this particular calculation is not possible with the Gaussian software used in this study, due to the complexity of the counterpoise correction calculation in a PCM[35,36].

The BSSE was however reduced by using a large basis set, which was suggested in a study performed by Bernstein *et al.*[37]. This procedure required three separate single point calculations, which can be explained by the following equation for calculating interaction energies:

$$E_{INT\,(SOLV)} = E\,(\text{Fragment } 1{+}2) - (E\,(\text{Fragment } 1) + E\,(\text{Fragment } 2))$$

1 2 3

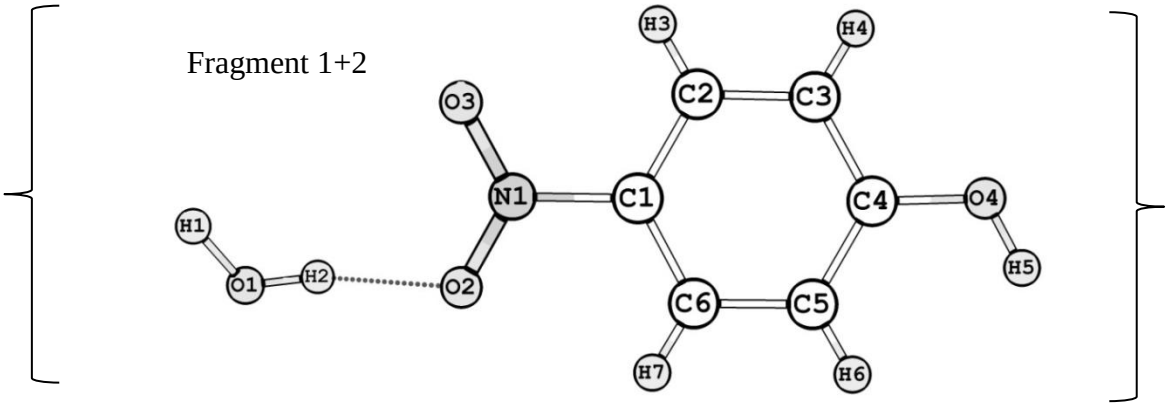

Figure 5.7. Illustration of the H-bond between Fragments 1 and 2 of structure N8.

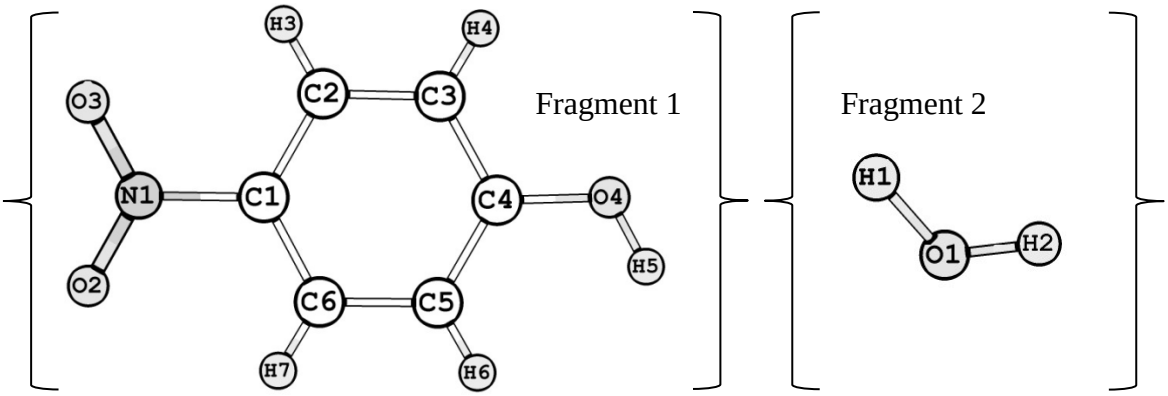

Figure 5.8. Illustration of Fragment 1 (left) and Fragment 2 (right) of structure N8.

An example of the fragments used for each of the three single point calculations is illustrated in Figures 5.7 and 5.8.

Table 5.14. Interaction energies (E_{INT}) of the phosphonate, carboxylate, nitroxide, nitrogen dioxide and sulfonate functional groups in 5 chosen solvents obtained at the M06-2X/6-311++G(d,p) level of theory.

		Solvent relative permittivity (ε)				
		1.911	9.863	20.524	46.826	78.3553
Structure P5	Solvent	Heptane	n-Octanol	1-Propanol	DMSO	Water
	E_{INT} (kcal/mol)	-23.3	-15.5	-14.4	-13.8	-13.6
Structure C4	Solvent	Heptane	n-Octanol	1-Propanol	DMSO	Water
	E_{INT} (kcal/mol)	-15.8	-11.4	-10.8	-10.9	-10.4
Structure N9	Solvent	Heptane	n-Octanol	1-Propanol	DMSO	Water
	E_{INT} (kcal/mol)	-5.1	-4.1	-3.9	-3.8	-3.6
Structure N4	Solvent	Heptane	n-Octanol	1-Propanol	DMSO	Water
	E_{INT} (kcal/mol)	-12.8	-10.9	-10.5	-10.1	-10.1
Structure S10	Solvent	Heptane	n-Octanol	1-Propanol	DMSO	Water
	E_{INT} (kcal/mol)	-14.9	-9.8	-9.1	-8.7	-8.5

The H-bond interaction energies for each of the five representative structures in the five solvent environments are listed in Table 5.14. The results show a decrease in intermolecular interaction as the relative permittivity increases. We can explain this occurrence by looking at the free energy of solvation (ΔG_s) that is provided by the PCM solvent model. The value of ΔG_s is determined by the solute–solvent interactions, which can be defined as the energy needed to transfer a molecule from the gas phase into solution. These solute-solvent interactions have been reported to have a significant effect on the total interaction energy (E_{tot}) of intermolecular H-bond interactions[38]. As mentioned in Chapter 1, the E_{tot} consists of contributions from electrostatics (E_{el}), polarization (E_{pol}), charge transfer (E_{ct}), dispersion (E_{disp}), and exchange repulsion (E_{er}). Gora *et al.* described the influence of solvent on intermolecular interactions as a decrease in electrostatic contributions (E_{el}) as the relative permittivity increases[39]. The E_{INT} values in Table 5.14 suggest that the same trend can be observed for the H-bonds in this study.

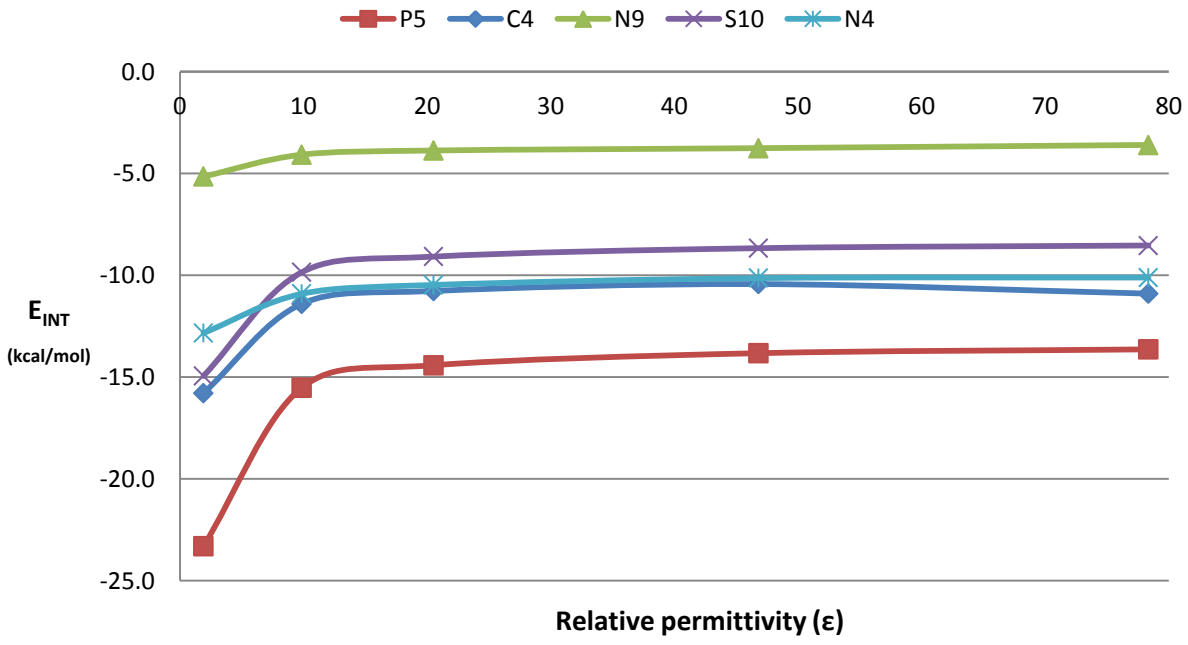

Figure 5.9. Evolution of the E_{INT} value for each representative structure as the relative permittivity is increased.

A graphical representation of the change in interaction energy as the relative permittivity increases is illustrated in Figure 5.9 for each representative structure. Each plot shows the relationship between E_{INT} and the relative permittivity for each representative structure. Furthermore, we see that the charged groups, such as phosphonate, carboxylates, nitroxide and sulfonates show the strongest dependence on the relative permittivity. This is explained in a study that was performed by Aquino *et al.*[40], which found that functional groups with a large dipole moment will interact strongly with polar solvents, leading to relatively high solvation energies. The structures used in this study have polar functional groups, which results in high solvation energies. The high solvation energy will correspond to stronger solute-solvent interactions that lower the electrostatic contribution of the H-bonds. The structures in Figure 5.9 are arranged according to the polarity of each functional group and difference in interaction energies obtained for the solvent with the lowest (heptane) and highest (water) relative permittivity.

Table 5.15. Structures arranged in decreasing polarity with the ΔE_{INT} values.

Structure	ΔE_{INT} (kcal/mol)
P5	9.7
S10	6.4
C4	5.4
N4	2.7
N9	1.5

$\Delta E_{INT} = E_{INT(water)} - E_{INT(heptane)}$

The values in Table 5.15 show that most polar functional groups will have stronger solute-solvent interactions in solvents with high relative permittivity's and correspondingly experience the largest lowering of H-bond interaction energy.

The decrease in strength of an H-bond in a compound that is placed in a polar solvent system such as water will result in many difficulties in the design of pharmaceutical compounds. The lack of strong H-bonds will not only complicate the formation of stable pharmaceutical hydrates, but also the ability of compounds that can form hydrates to remain in the hydrated form during product storage[41]. The crucial role of the H-bond during processing as well as storage of pharmaceutical hydrates should thus not be overlooked when designing novel pharmaceutical hydrates.

Chapter Summary

To summarise, we have investigated the various factors that can influence the strength of H-bonds in pharmaceutical hydrate structures. It was found that geometry optimisation as well as single point calculations performed at the M062X/6-311++G(d,p) level of theory provide accurate results when compared to the MP2/aug-cc-pVTZ benchmark level of theory. The results show that the strongest H-bonds are found for compounds that have charged polar functional groups. The ability of functional groups to form strong H-bonds is also greatly influenced by the neighbouring substituents. The ability of substituents to influence the

strength of H-bonds was additionally measured by determining steric density values for H-bond donor and acceptor atoms. Further investigation of these substituents is however required, as it was shown that some bulky substituents can results in stronger H-bonds, while the same substituent can result in a weaker H-bond in another compound.

Structures that contain aromatic rings show cooperative stabilisation of the H-bond interaction by neighbouring substituents on the aromatic ring, which results in stronger H-bond interaction energies. Utilising substituents on the aromatic ring could be a powerful method to fine-tune the strength of H-bonds and correspondingly the properties, including stability, of the hydrate. The stabilising effect of the surrounding crystalline environment was investigated by optimising the H-bond geometrical parameters and comparing the H-bond interaction energies to those obtained for fully optimised structures. The results showed that cooperative stabilisation can result in stronger H-bonds in fully optimised structures, while there are some examples where the surrounding crystalline environment could result in a slightly stronger H-bond. Lastly, the effect of the electrostatic environment on the strength of the H-bond was investigated by placing several representative structures in different electrostatic environments. The results showed a significant weakening of the H-bond in solvent systems with a high relative permittivity, which was mainly to an increase in solute-solvent interactions as the relative permittivity increases, which will in turn reduce the electrostatic contribution to the total interaction energy. This effect was found to be even more pronounced for structures that have polar functional groups.

References

(1) L. Pauling: The Nature of the Chemical Bond and the Structure of Molecules and Crystals - An Introduction to Modern Structural Chemistry. Oxford, *Oxford University Press* **1940.**

(2) Desiraju, G. R.: Supramolecular Synthons in Crystal Engineering—A New Organic Synthesis. *Angewandte Chemie International Edition in English* **1995**, *34*, 2311-2327.

(3) R. Desiraju, G.: Designer crystals: intermolecular interactions, network structures and supramolecular synthons. *Chemical Communications* **1997**, 1475-1482.

(4) G.Schaftenaar, J. H. N.: Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of computer-aided molecular design* **2000**, *14*, 123-134.

(5) Frank, N. H.: Note on the Hartree and Hartree-Fock Methods. *Physical Review* **1937**, *51*, 577-583.

(6) Becke, A. D.: Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics* **1993**, *98*, 5648.

(7) Zandler, M. E., D'Souza, F.: The remarkable ability of B3LYP/3-21G(*) calculations to describe geometry, spectral and electrochemical properties of molecular and supramolecular porphyrin–fullerene conjugates. *Comptes Rendus Chimie* **2006**, *9*, 960-981.

(8) Lusi, M., de Villiers, D., Esterhuysen, C.: Hydrogen-Bond Analysis: Statistical and Computational versus Experimental Position Refinement. *Crystal Growth & Design* **2014**, *14*, 3480-3484.

(9) Perdew, J. P., Burke, K., Ernzerhof, M.: Generalized Gradient Approximation Made Simple *Physical Review Letters* **1997**, *78*, 1396-1397.

(10) Perdew, J. P., Burke, K., Ernzerhof, M.: Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865-3868.

(11) Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C.: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **1992**, *46*, 6671-6687.

(12) Grimme, S.: Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27*, 1787-1799.

(13) Y. Zhao, Truhlar, D. G.: The M06 Suite of Density Functionals for Main Group Thermochemistry, Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06 Functionals and Twelve Other Functionals, *Theoretical Chemistry Accounts* **2008**, *120*, 215.

(14) Hohenstein, E. G., Chill, S. T., Sherrill, C. D.: Assessment of the Performance of the M05−2X and M06−2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules. *Journal of Chemical Theory and Computation* **2008**, *4*, 1996-2000.

(15) Walker, M., Harvey, A. J. A., Sen, A., Dessent, C. E. H.: Performance of M06, M06-2X, and M06-HF Density Functionals for Conformationally Flexible Anionic Clusters: M06 Functionals Perform Better than B3LYP for a Model System with Dispersion and Ionic Hydrogen-Bonding Interactions. *The Journal of Physical Chemistry A* **2013**, *117*, 12590-12600.

(16) Grimme, S.: Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of Chemical Physics* **2006**, *124*.

(17) Schwabe, T., Grimme, S.: Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Physical Chemistry Chemical Physics* **2007**, *9*, 3397-3406.

(18) Karr, P. A., Zandler, M. E., Beck, M., Jaeger, J. D., McCarty, A. L., Smith, P. M., D'Souza, F.: Predicting the site of electron transfer using DFT frontier orbitals: Studies on porphyrin attached either to quinone or hydroquinone, and quinhydrone self-assembled supramolecular complexes. *Journal of Molecular Structure: Theochem* **2006**, *765*, 91-103.

(19) Simon, S., M. Duran, and J.J. Dannenberg: How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *The Journal of Chemical Physics* **1996**, *105*, 11024-11031.

(20) Becke, A. D.: Density-functional thermochemistry. V. Systemic optimization of exchange-correlation functionals. *Journal of Chemical Physics* **1997**, *107*, 8554.

(21) Hamprecht, F. A., Cohen, A. J.: Development and assessment of new exchange-correlation functionals. *Journal of Chemical Physics* **1998**, *109*, 6264.

(22) Grimme, S., Ehrlich, S., Goerigk, L.: Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456-1465.

(23) Tao, J., Perdew, J. P., Staroverov, V. N., Scuseria, G. E.: Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Physical Review Letters* **2003**, *91*, 146401.

(24) Steiner, T.: The Hydrogen Bond in the Solid State. *Angewandte Chemie International Edition* **2002**, *41*, 48-76.

(25) Coombes, D. S., Price, S. L., Willock, D. J., Leslie, M.: Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *The Journal of Physical Chemistry* **1996**, *100*, 7352-7360.

(26) Jeffrey, G. A.: An Introduction to Hydrogen Bonding. Oxford, *Oxford University Press* **1997**.

(27) Kang, Y. K., Park, H. S.: Assessment of CCSD(T), MP2, DFT-D, CBS-QB3, and G4(MP2) methods for conformational study of alanine and proline dipeptides. *Chemical Physics Letters* **2014**, *600*, 112-117.

(28) Filarowski, A., Głowiaka, T., Koll, A.: Strengthening of the intramolecular O···H···N hydrogen bonds in Schiff bases as a result of steric repulsion. *Journal of Molecular Structure* **1999**, *484*, 75-89.

(29) Galek, P. T. A., Fábián, L., Motherwell, W. D. S., Allen, F. H., Feeder, N.: Knowledge-based model of hydrogen-bonding propensity in organic crystals. *Acta Crystallographica Section B* **2007**, *63*, 768-782.

(30) Cheng, B., Tehrani, A. A., Hu, M.-L., Morsali, A.: Supramolecular assemblies of Ru(II) organometallic half-sandwich complexes. *CrystEngComm* **2014**, *16*, 9125-9134.

(31) Shan, S.-o., Herschlag, D.: Energetic Effects of Multiple Hydrogen Bonds. Implications for Enzymatic Catalysis. *Journal of the American Chemical Society* **1996**, *118*, 5515-5518.

(32) Hunter, C. A., Ihekwaba, N., Misuraca, M. C., Segarra-Maset, M. D., Turega, S. M.: Cooperativity in multiply H-bonded complexes. *Chemical Communications* **2009**, 3964-3966.

(33) Perumalla, S. R., Sun, C. C.: Synthon preference in O-protonated amide crystals - dominance of short strong hydrogen bonds. *CrystEngComm* **2013**, *15*, 8941-8946.

(34) Chakraborty, S., Rajput, L., Desiraju, G. R.: Designing Ternary Co-crystals with Stacking Interactions and Weak Hydrogen Bonds. 4,4′-Bis-hydroxyazobenzene. *Crystal Growth & Design* **2014**, *14*, 2571-2577.

(35) Boys, S. F., Bernardi, F.: The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970**, *19*, 553-566.

(36) Simon, S., Duran, M., Dannenberg, J.: How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *The Journal of Chemical Physics* **1996**, *105*, 11024-11031.

(37) Rappé, A. K., Bernstein, E. R.: Ab Initio Calculation of Nonbonded Interactions: Are We There Yet? *The Journal of Physical Chemistry A* **2000**, *104*, 6117-6128.

(38) Buemi, G.: Ab initio study of 2,4-dihalosubstituted malonaldehyde and 2-halo-phenols in gas phase and solution. *Chemical Physics* **2002**, *277*, 241-256.

(39) Gora, R. W., Bartkowiak, W., Roszak, S., Leszczynski, J.: Intermolecular interactions in solution: Elucidating the influence of the solvent. *The Journal of Chemical Physics* **2004**, *120*, 2802-2813.

(40) Aquino, A. J. A., Tunega, D., Haberhauer, G., Gerzabek, M. H., Lischka, H.: Solvent Effects on Hydrogen BondsA Theoretical Study. *The Journal of Physical Chemistry A* **2002**, *106*, 1862-1871.

(41) Arora, K. K., Thakral, S.; Suryanarayanan, R.: Instability in Theophylline and Carbamazepine Hydrate Tablets: Cocrystal Formation Due to Release of Lattice Water. *Pharmaceutical Research* **2013**, *30*, 1779-89.

CHAPTER 6

The effect of aromatic ring substituents on hydrogen bond strength.

Introduction

The calculations described in Chapter 5 yielded some interesting results for aromatic structures that have been substituted at one of the positions on the aromatic ring. These substituents have the ability to strengthen or weaken the H-bond with water by cooperative stabilisation or by electron-donating and electron-withdrawing inductive effects. Since studies have suggested that the presence of strong H-bond acceptor and donor groups will result in a strong H-bond[1,2], we have investigated if this is indeed the case by measuring how cooperative stabilisation and the degree of electron-donation or withdrawal effects will affect the H-bond when various substituents are placed at the *ortho, meta* and *para* position of four compounds containing aromatic rings.

6.1. Geometry optimisation calculations of four aromatic structures

In the study described in Chapter 6, we identified four aromatic compounds with different functional groups to investigate. These are C4, N8, S5 and P5 with carboxylate, nitrogen dioxide, sulfonate and phosphonate functional groups. In order to generate the models used in this study the substituents other than the group forming the H-bond were removed from each of the aromatic rings and replaced by H-atoms, resulting in four new compounds that will henceforth be referred to as C_{Ph}, N_{Ph}, S_{Ph} and P_{Ph}. These four compounds were then optimised using a combination of the MP2 and M062X methods with the 6-311++G(d,p) basis set.

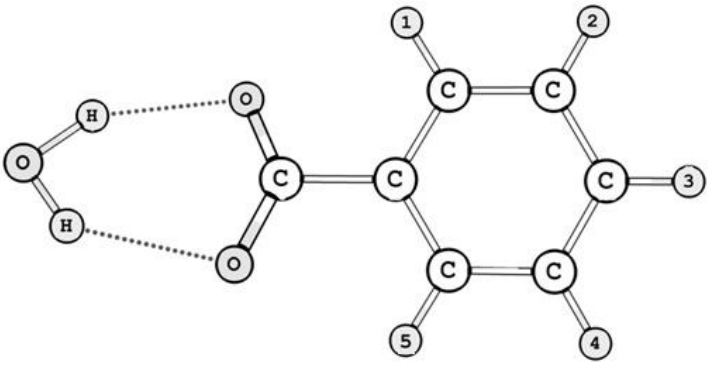

Figure 6.1. The five positions on the aromatic ring of optimised structure C_{Ph}.

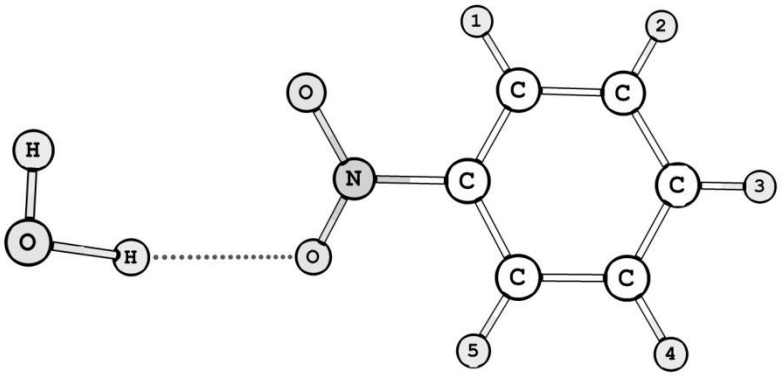

Figure 6.2. The five positions on the aromatic ring of optimised structure N_{Ph}.

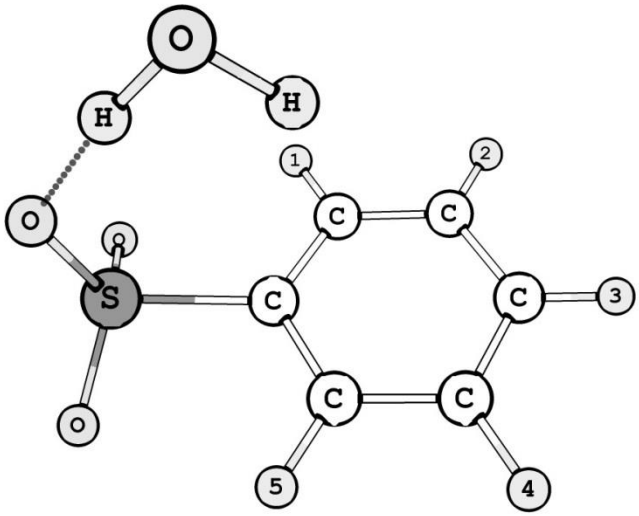

Figure 6.3. The five positions on the aromatic ring of optimised structure S_{Ph}.

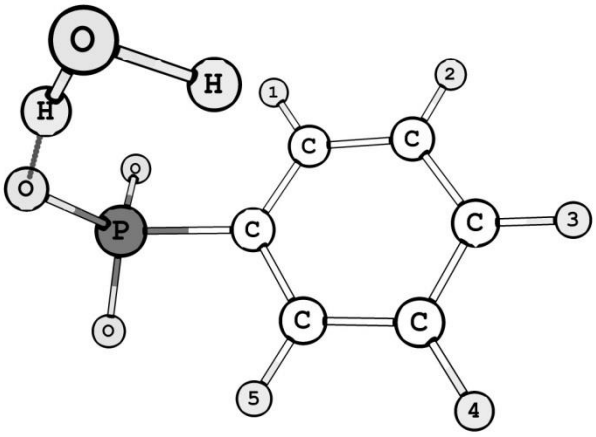

Figure 6.4. The five positions on the aromatic ring of optimised structure P_{Ph}.

The four fully optimised aromatic compounds are illustrated in Figures 6.1 – 6.4. All five positions were considered in this study, and not just the placing of one substituent at the *ortho*, *meta* and *para* position, as the H-bonds in Figure 6.1 - 6.4 are not symmetrical with respect to the ring. Therefore, we expect substituents that are placed at positions 1 and 5 will influence the H-bond differently.

We chose four substituents with different induction strengths: namely methyl ($-CH_3$), hydroxyl (-OH), formyl (-CHO) and nitro groups (-NOO). The methyl and hydroxyl groups are electron-donating groups, while the formyl and nitro groups are electron-withdrawing groups[3]. Furthermore, these substituents range from strongly activating to strongly deactivating groups, which is achieved by a combination of inductive and resonance effects. Every one of the four substituents was placed at each of the five positions available on the phenyl ring. This required a total of 80 geometry optimisation calculations, which were performed at the MP2/6-311++G(d,p) level of theory.

6.2. The inductive effect of substituents on H-bond energies.

The H-bond strength was determined for each structure by performing counterpoise corrected single point calculations at the M06-2X/6-311++G(d,p) level of theory. The M06-2X method was utilised due to its success with describing the systems discussed in previous chapters. In order to analyse the effect of the substituents, the H-bond interaction energies of the unsubstituted ring are included in Table 6.1, where an unsubstituted ring was defined as any structure that only contains the functional group attached to a phenyl ring.

Table 6.1. H-bond energies (kcal/mol) for each representative structure with phenyl group.

	Method (E_{INT} kcal/mol)
Structure	M06-2X/6-311++G(d,p)
C4	-20.74
N8	-4.98
S5	-13.46
P5	-29.39

The values of the H-bond strength of the unsubstituted aromatic structures of each functional group are summarised in Table 6.1. It should be noted that we will not consider the stabilisation of the structures by the substituents on the structure, but rather the effect of the substituents on the strength of the H-bond.

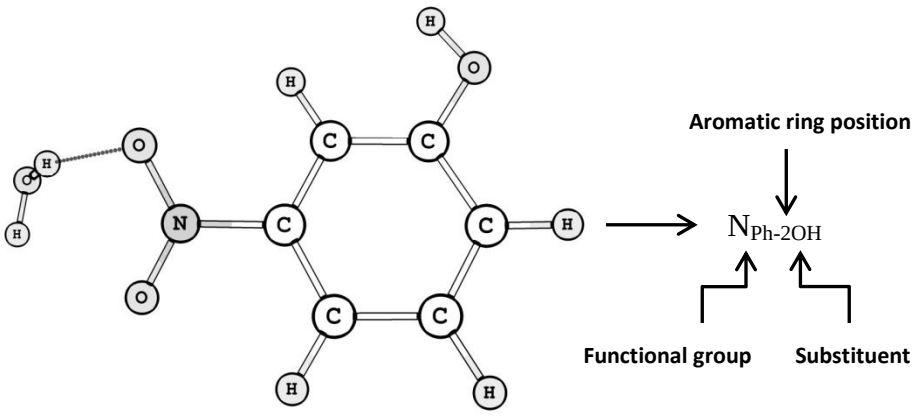

Figure 6.5. The nomenclature used to identify each of the 80 optimised compounds is illustrated for N_{Ph} with a hydroxyl substituent at position 2.

The nomenclature used to identify each of the 80 optimised compounds is described in Figure 6.5. An investigation of the H-bonds in all 80 compounds was performed, we will only be discussing the compounds with the weakest and strongest H-bonds for each substituent. This was done by comparing the relative interaction energy (E_{INT}) values of the strong and weak H-bonds and then studying the optimised geometry conformation in order to determine what factors result in the strength difference between different H-bonds.

6.2.1. The effect of aromatic substituents on the H-bond in compounds that contain the carboxylate functional group

Table 6.2. Relative E_{INT} (kcal/mol) of the H-bonds of the carboxylate compound with substituents placed at positions 1-5 obtained at the M06-2X/6-311++G(d,p) level of theory.

C_{Ph}		
Substituent	E_{INT} (kcal/mol)	ΔE_{INT} (kcal/mol)
C_{Ph}	-20.74	0.00
$C_{Ph\text{-}1OH}$	-18.24	2.50
$C_{Ph\text{-}2OH}$	-17.24	3.50
$C_{Ph\text{-}3OH}$	-16.26	4.48
$C_{Ph\text{-}4OH}$	-15.96	4.78
$C_{Ph\text{-}5OH}$	-14.12	6.63
$C_{Ph\text{-}1CH3}$	-17.64	3.10
$C_{Ph\text{-}2CH3}$	-16.65	4.09
$C_{Ph\text{-}3CH3}$	-16.51	4.23
$C_{Ph\text{-}4CH3}$	-16.30	4.44
$C_{Ph\text{-}5CH3}$	-16.03	4.71
$C_{Ph\text{-}1CHO}$	-19.28	1.47
$C_{Ph\text{-}2CHO}$	-15.83	4.91
$C_{Ph\text{-}3CHO}$	-15.60	5.14
$C_{Ph\text{-}4CHO}$	-15.92	4.82
$C_{Ph\text{-}5CHO}$	-15.34	5.41
$C_{Ph\text{-}1NO2}$	-18.82	1.92
$C_{Ph\text{-}2NO2}$	-14.02	6.72
$C_{Ph\text{-}3NO2}$	-15.12	5.62
$C_{Ph\text{-}4NO2}$	-15.52	5.22
$C_{Ph\text{-}5NO2}$	-15.51	5.23

The H-bond energies of all structures that have been substituted at one of the five positions in compound C_{Ph} are summarised in Table 6.2. Additionally, the difference between the E_{INT} values for each substituted and the unsubstituted compounds have also been included in the table and are represented by the ΔE_{INT} symbol. All of the results show that the H-bond is weaker for all substituted structures compared to the unsubstituted counterpart. A study by Rakatin *et al.* investigated the energy barrier that has to be overcome when rotating the aromatic carboxylate group out of a planar conformation[4]. Their results showed that rotation of the carboxylate group from 0 to 90° with respect to the plane of the aromatic ring will result in an energy penalty of 3.4-6.8 kcal/mol. From Figure 6.1 we observe a very planar orientation of the water molecule relative to the benzoate compounds, which allows the formation of two H-bonds. Investigation of the optimised structures in Table 6.2 show that the addition of substituents on the aromatic ring results in the rotation of the carboxylate

group, which will correspondingly result in a change in the orientation of the water molecule. It is therefore suggested that the degree of rotation will affect the strength of the H-bond as well as the energy penalty described by Rakatin *et al.*[4].

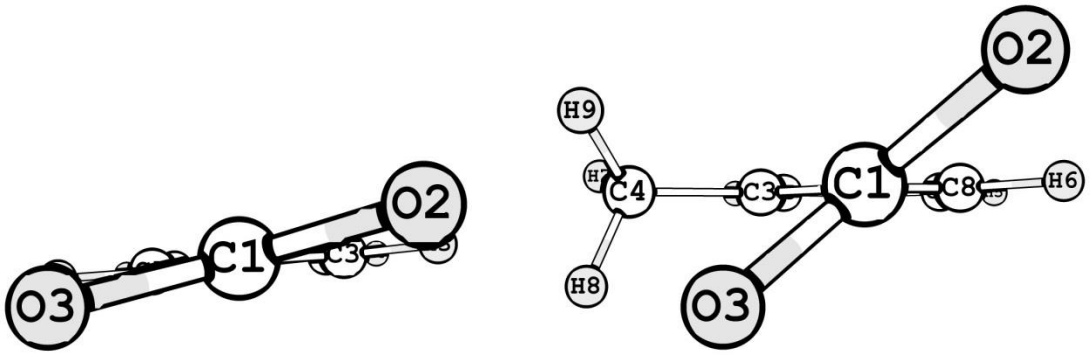

Figure 6.6. View down the carboxylate group of the unsubstituted (left) and substituted with a methyl at position 5 compound (right). The water molecules were omitted for clarity of the illustration.

The effect of the substituents on the rotation of the carboxylate group is illustrated in Figure 6.6, which shows that the addition of the methyl substituent at position 5 results in a rotation of the carboxylate group. Furthermore, the rotation of the carboxylate group results in a different orientation of the water molecule relative to the benzoate, which is illustrated in Figure 6.7.

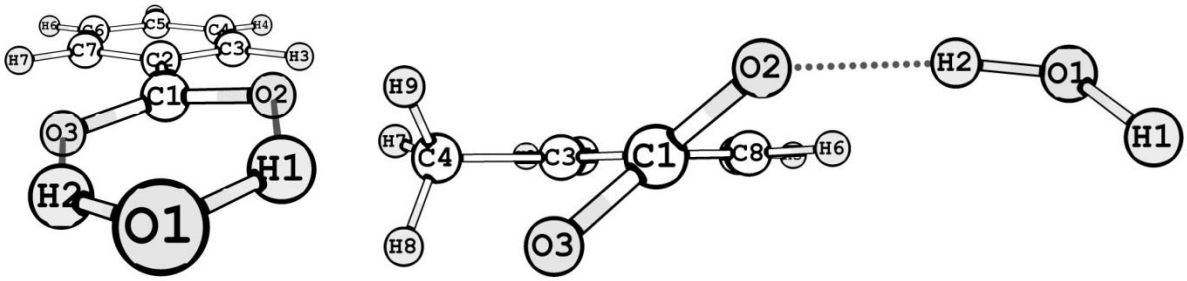

Figure 6.7. The orientation of benzoate relative to the water molecule for the unsubstituted (left) and the substituted (right) compounds.

Figure 6.7 shows that there are two H-bonds for the unsubstituted compound. This is accompanied by cooperative stabilisation, which can be identified as more than one H-bond interaction that stabilise one another and increase the H-bond interaction[5,6]. This is however not observed for the substituted compound, where the water molecule now has different geometrical parameters that do not allow cooperative stabilisation. We can thus conclude that the substituents result in the rotation of the carboxylate functional group and this yields lower E_{INT} values for all substituted compounds in Table 6.2. We are however interested in the effect of substituents on the H-bond and will thus still consider which substituents results in the strongest H-bond in Table 6.2 and how this is achieved.

Of all the positions where substituents can be added, the strongest H-bonds are found when substituents are added at position 1. After investigation of the optimised geometries for structures with an added substituent at position 1, it was found that there is a cooperative stabilisation of the H-bond when substituents are placed in this position (see below).

Furthermore, the strongest H-bond is found when the formyl substituent is placed at position 1, which is indicative of the substituent's ability to provide the largest cooperative stabilisation to the H-bond.

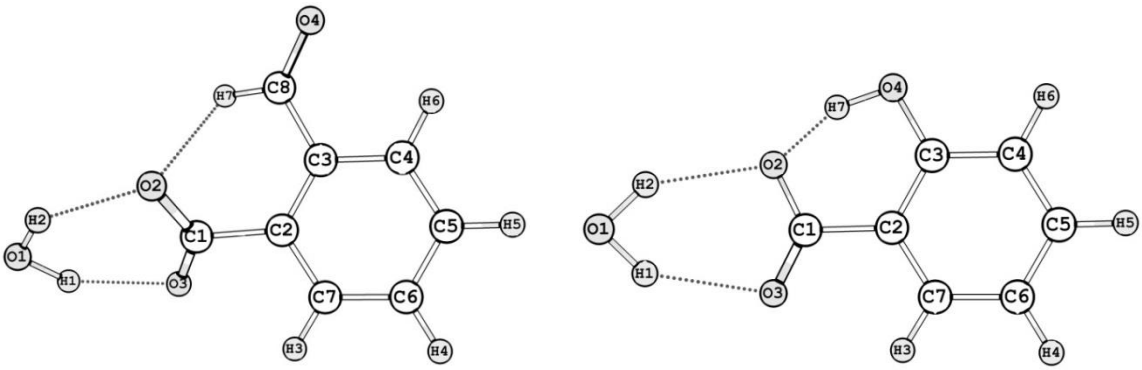

Figure 6.8. The cooperative stabilisation of the H-bond when a formyl (left) and an hydroxyl (right) substituent is placed at position 1 of the aromatic ring.

The cooperative stabilisation of the H-bond interaction in compound C_{Ph} when formyl and hydroxyl substituents are placed at position 1 is illustrated in Figure 6.8. It should also be mentioned that all of the hydrogen atoms numbering do not correspond to the numbering system in Figures 6.1 - 6.4. Both substituents form an additional H-bond to the carboxylate functional group, which results in a stabilisation and increase in H-bond interaction energy. From the results shown in this study, it is suggested that the strength of the H-bond that forms between benzoate and a water molecule will be increased if substituents that are capable of cooperative stabilisation are placed at a position adjacent to the carboxylate functional group. Furthermore, the effect of the substituent on the rotation of the carboxylate group should also be considered as substituents that result in a significant rotation of the carboxylate group will lower the strength of the H-bond.

6.2.2. The effect of aromatic substituents on the H-bond in compounds that contain the nitrogen dioxide functional group

Table 6.3. Relative E_{INT} (kcal/mol) of the H-bonds of the nitrogen dioxide neutral radical compound with substituents placed at positions 1-5 obtained at the M06-2X/6-311++G(d,p) level of theory.

	N_{Ph}	
Substituent	E_{INT} (kcal/mol)	ΔE_{INT} (kcal/mol)
N_{Ph}	-4.98	0.00
$N_{Ph\text{-}1OH}$	-8.94	-3.96
$N_{Ph\text{-}2OH}$	-4.91	0.07
$N_{Ph\text{-}3OH}$	-5.08	-0.10
$N_{Ph\text{-}4OH}$	-5.10	-0.12
$N_{Ph\text{-}5OH}$	-4.19	0.79
$N_{Ph\text{-}1CH3}$	-6.22	-1.24
$N_{Ph\text{-}2CH3}$	-5.14	-0.16
$N_{Ph\text{-}3CH3}$	-5.17	-0.19

Table 6.3. Continued.

$N_{Ph\text{-}4CH3}$	-5.12	-0.14
$N_{Ph\text{-}5CH3}$	-6.22	-1.24
$N_{Ph\text{-}1CHO}$	-9.48	-4.50
$N_{Ph\text{-}2CHO}$	-4.83	0.15
$N_{Ph\text{-}3CHO}$	-6.57	-1.60
$N_{Ph\text{-}4CHO}$	-4.87	0.11
$N_{Ph\text{-}5CHO}$	-5.96	-0.98
$N_{Ph\text{-}1NO2}$	-10.53	-5.55
$N_{Ph\text{-}2NO2}$	-5.53	-1.16
$N_{Ph\text{-}3NO2}$	-5.97	-0.99
$N_{Ph\text{-}4NO2}$	-5.53	-0.55
$N_{Ph\text{-}5NO2}$	-6.10	-1.12

The E_{INT} and ΔE_{INT} values are summarised in Table 6.2 for all structures that have been substituted at one of the five positions in compound N_{Ph}. The results show that some of the substituents result in a stronger H-bond. Compared to the strong H-bond that was observed for the unsubstituted carboxylate compound in Table 6.2, the H-bond for the unsubstituted compound in Table 6.3 is significantly weaker than the majority of H-bonds in the substituted compounds. We observe from Figure 6.2 that there is no cooperative stabilisation present in the optimised geometry of structure N_{Ph}. The strongest H-bond energies were once again found to occur with the addition of substituents at position 1. Similarly to the carboxylate group, cooperative stabilisation increases the strength of the H-bond when substituents are placed adjacent to the functional group. It must however be mentioned that the substituents are not involved in the cooperative stabilisation, but it is rather one of the neighbouring hydrogen atoms at position 5 that is involved in the cooperative effect. This suggests that the substituent at position 1 results in a rotation of the functional group to a conformation where the hydrogen atom can be involved in a second H-bond.

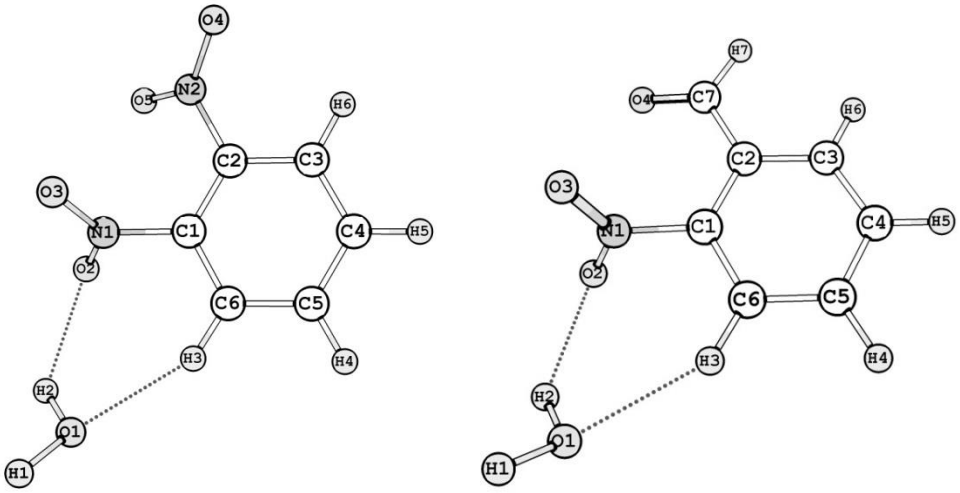

Figure 6.9. The cooperative stabilisation of the H-bond when a nitro (left) and a formyl (right) substituent is placed at position 1 of the aromatic ring.

The cooperative stabilisation of water by compound N_{Ph} when nitro and formyl substituents are placed at position 1 is illustrated in Figure 6.9.

6.2.3 The effect of aromatic substituents on the H-bond in compounds that contain the sulfonate functional group

Table 6.4. Relative E_{INT} (kcal/mol) of the H-bonds of the sulfonate compound with substituents placed at positions 1-5 obtained at the M06-2X/6-311++G(d,p) level of theory.

S_{Ph}		
Substituent	E_{INT} (kcal/mol)	ΔE_{INT} (kcal/mol)
S_{Ph}	-13.46	0.00
$S_{Ph\text{-}1OH}$	-14.05	-0.59
$S_{Ph\text{-}2OH}$	-14.38	-0.91
$S_{Ph\text{-}3OH}$	-14.48	-1.02
$S_{Ph\text{-}4OH}$	-14.38	-0.91
$S_{Ph\text{-}5OH}$	-13.41	0.05
$S_{Ph\text{-}1CH3}$	-16.57	3.11
$S_{Ph\text{-}2CH3}$	-18.03	-4.57
$S_{Ph\text{-}3CH3}$	-18.19	-4.73
$S_{Ph\text{-}4CH3}$	-17.95	-4.49
$S_{Ph\text{-}5CH3}$	-14.52	-1.06
$S_{Ph\text{-}1CHO}$	-12.86	0.60
$S_{Ph\text{-}2CHO}$	-13.23	0.24
$S_{Ph\text{-}3CHO}$	-13.05	0.41
$S_{Ph\text{-}4CHO}$	-13.32	0.14
$S_{Ph\text{-}5CHO}$	-17.95	-4.49
$S_{Ph\text{-}1NO2}$	-16.04	-2.58
$S_{Ph\text{-}2NO2}$	-15.96	-2.49
$S_{Ph\text{-}3NO2}$	-16.91	-3.45
$S_{Ph\text{-}4NO2}$	-15.45	-1.99
$S_{Ph\text{-}5NO2}$	-15.85	-2.39

The E_{INT} and ΔE_{INT} values are summarised in Table 6.4 for all structures that have been substituted at one of the five positions in compound S_{Ph}. For this particular compound, the strongest H-bonds are found when methyl, hydroxyl and nitro groups are placed at position 3. This trend can be explained comparing the nitro substituted compounds. From these compounds, it was observed that the two H-bonds in the $S_{Ph\text{-}3NO2}$ compound are very similar with respect to their geometrical parameters, which results in a planar arrangement of the water molecule with respect to the sulfonate functional group. Comparing now, for example, the two H-bonds in the $S_{Ph\text{-}1NO2}$ compounds, we observe that there is no symmetrical arrangement of the water molecule with respect to the sulfonate functional group on the

aromatic ring. This is further accompanied by one of the H-bonds being 0.18 Å shorter than its counterpart. These observations are shown in Figure 6.10, which shows the difference between the two compounds by looking down the plane that contains C2 and C3.

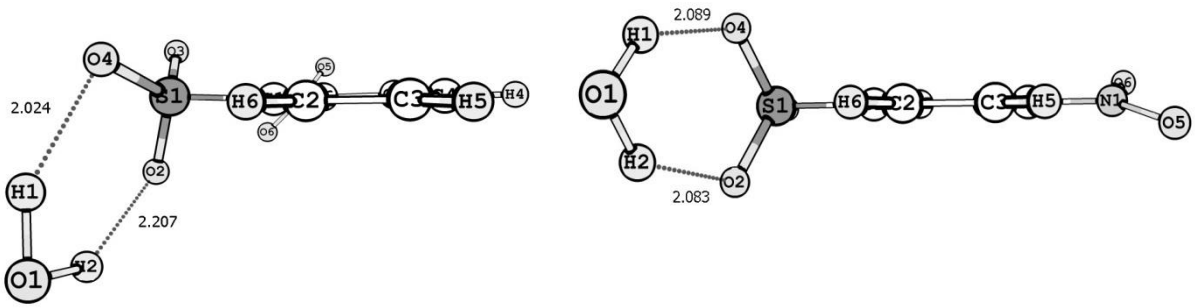

Figure 6.10. The difference in cooperative stabilisation of the H-bond when a nitro substituent is placed at position 1 (left) and position 3 (right). The values included represent the H-bond distance (Å).

Figure 6.10 shows the difference in geometrical parameters of the two H-bonds when the nitro substituent has been placed at position 1 and position 3. It is found that when the nitro group is placed at position 3, the sulfonate has the freedom to rotate towards a minimum energy conformation that results in a stronger H-bond than when the nitro group is placed at position 1.

The formyl substituent follows a similar trend to the previous functional groups. A stronger H-bond is found when the formyl group is placed at position 5, which results in the only formyl substituted compound that shows cooperative stabilisation.

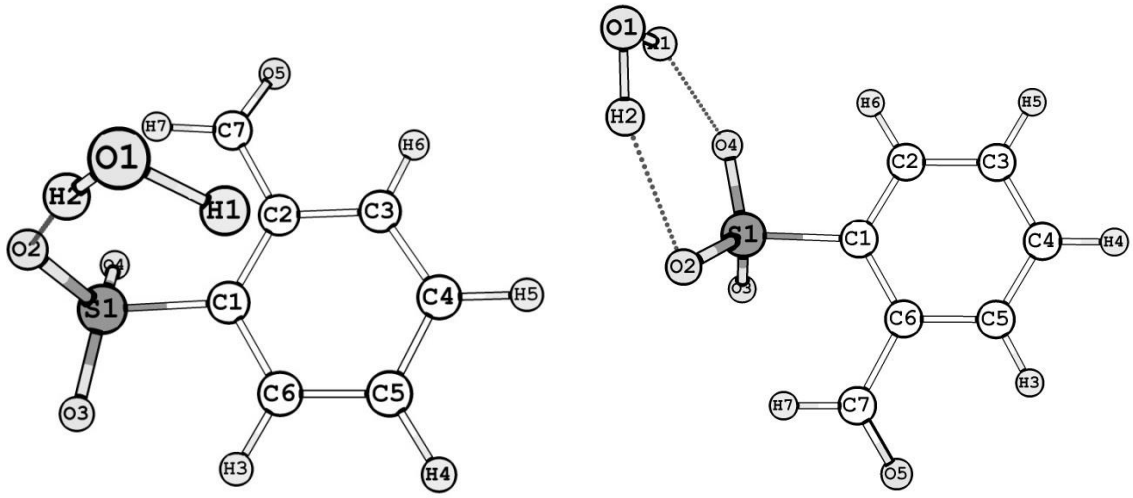

Figure 6.11. The H-bond for compounds with the formyl substituent at position 1 (left) and position 5 (right) on the aromatic ring.

Figure 6.11 shows the absence of any cooperative stabilisation when the formyl substituent is placed on position 1 and presence of cooperative stabilisation when the formyl group is placed in position 5.

6.2.4. The effect of aromatic substituents on the H-bond in compounds that contain the phosphonate functional group

Table 6.5. Relative E_{INT} (kcal/mol) of the H-bonds of the phosphonate compound with substituents placed at positions 1-5 obtained at the M06-2X/6-311++G(d,p) level of theory.

	P_{Ph}	
Substituent	E_{INT} (kcal/mol)	ΔE_{INT} (kcal/mol)
P_{Ph}	-29.39	0.00
$P_{Ph\text{-}1OH}$	-24.60	4.79
$P_{Ph\text{-}2OH}$	-34.00	-4.61
$P_{Ph\text{-}3OH}$	-34.65	-5.26
$P_{Ph\text{-}4OH}$	-34.34	-4.95
$P_{Ph\text{-}5OH}$	-28.25	1.14
$P_{Ph\text{-}1CH3}$	-28.84	0.55
$P_{Ph\text{-}2CH3}$	-29.08	0.31
$P_{Ph\text{-}3CH3}$	-29.15	0.24
$P_{Ph\text{-}4CH3}$	-33.60	-4.21
$P_{Ph\text{-}5CH3}$	-34.50	-5.11
$P_{Ph\text{-}1CHO}$	-26.47	2.93
$P_{Ph\text{-}2CHO}$	-27.57	1.82
$P_{Ph\text{-}3CHO}$	-26.50	2.89
$P_{Ph\text{-}4CHO}$	-27.27	2.12
$P_{Ph\text{-}5CHO}$	-31.61	-2.22
$P_{Ph\text{-}1NO2}$	-25.78	3.61
$P_{Ph\text{-}2NO2}$	-26.86	2.53
$P_{Ph\text{-}3NO2}$	-32.01	-2.62
$P_{Ph\text{-}4NO2}$	-27.15	2.24
$P_{Ph\text{-}5NO2}$	-28.05	1.34

The E_{INT} and ΔE_{INT} values are summarised in Table 6.5 for all structures that have been substituted at one of the five positions in compound P_{Ph}. The values in Table 6.5 show that the strongest H-bonds for the hydroxyl and nitro substituents are found when they are placed at positions 2, 3 and 4. All of the compounds contain cooperative stabilisation, but the compounds with substituents at positions 2, 3 or 4 show a more symmetrical arrangement, similar to the sulfonate compounds. The hydroxyl substituent at position 1 results in a completely different orientation of the water molecule relative to the phosphonate functional group, which results in a lower H-bond strength.

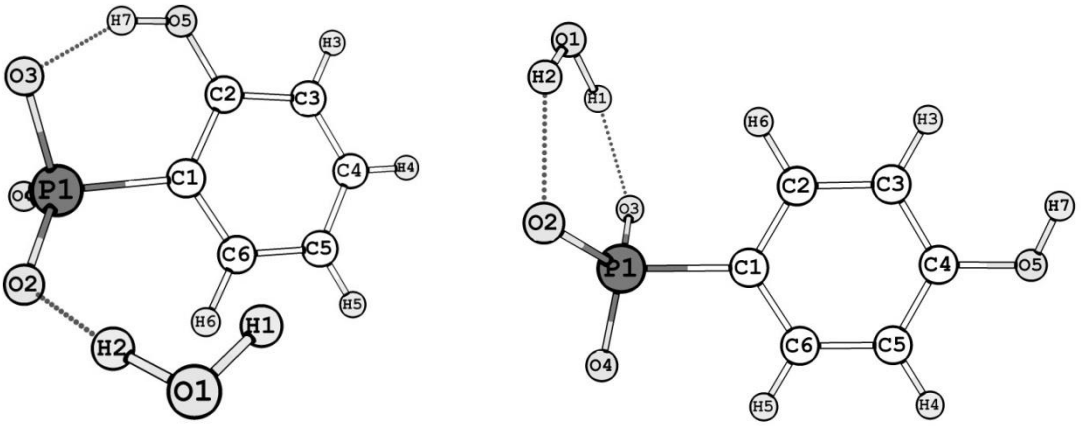

Figure 6.12. The H-bond for compounds with the hydroxyl substituent at position1 (left) and position 3(right) on the aromatic ring.

Figure 6.12 shows the difference between the H-bonds when the hydroxyl group is placed at positions 1 and 3.

The strongest H-bonds for the formyl and methyl substituents are found when they are placed at position 5, which are the only compounds where the H-bond is strengthened by cooperative stabilisation.

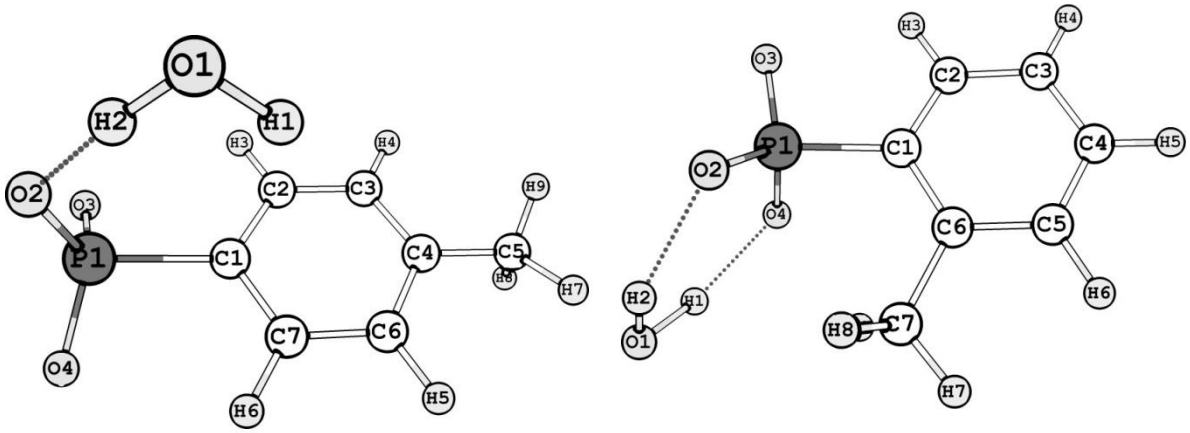

Figure 6.13. The H-bond for compounds with the methyl substituent at position 3 (left) and position 5 (right) on the aromatic ring.

Figure 6.13 shows the presence of cooperative stabilisation when the methyl substituent is placed at position 5, whereas no cooperative stabilisation is present when the methyl substituent is placed at position 3.

Several conclusions can be made from the substitution studies of carboxylate, nitrogen dioxide, sulfonate and phosphonate compounds. Firstly, substituents that are placed adjacent to the functional group can form an additional H-bond, which will result in cooperative stabilisation of the H-bond. Secondly, each functional group shows a different degree of rotation when a substituent is placed at one of the five positions of the aromatic ring. Each substituent will affect this rotation differently by steric or repulsive effects. Furthermore, the degree of rotation can influence the orientation of the water molecule relative to the aromatic compound, which will in turn influence the strength of the H-bond.

In this study we have investigated how substituents that are placed on different positions of the aromatic ring will affect the H-bond energies of different functional groups. Several other studies have also investigated the effect of substituent on interaction energies. A study by Mata *et al.*[7] investigated the substituent effect of R groups on the interaction energy in the FH···FR series of compounds. Their results suggested that the R group increases the strength of the electric field surrounding the F-atom, which shifts the position and depth of the minimum well in the H-bond interaction potential. The changes in the interaction potential and minimum well depth can be described by the Lennard-Jones potential, a mathematical model that gives the relationship between interaction energy and the distance between two atoms that form the interaction[3,8]. The stronger electric field will result in shorter distances between the H-bond acceptor/donor pair and stronger interaction energies.

The results in this study identified several other groups that are also capable of increasing the H-bond interaction energy, which could be due to a mixture of the factors that have been discussed and those mentioned above. The increase in H-bond strength by the addition of substituents to the aromatic ring does not just have the potential to increase the propensity of pharmaceutical compounds to form hydrates, but can also increase the strength of H-bonds in other pharmaceutical compounds that are not necessarily in the hydrated form. Many studies have suggested that the strength of the H-bond plays a crucial role in the activity of pharmaceutical compounds. For instance, Zhang *et al.* investigated the protein-binding properties of a steroidal lactam compound. They found that an increase in the strength of the H-bond led to a reduced toxicity as well as an enhanced pharmaceutical activity[9].

Another study by Reynisson *et al.* suggested that tuning the strength of the H-bond can result in stronger binding of ligands to proteins, but can also result in steric clash, which blocks the binding of a ligand to the protein target[10]. Changes to the free energy of the system as small as 1.4 kcal/mol can lead to a tenfold increase or decrease in the affinity of the ligand at physiological temperature. The electronic effects can also lower the free energy of system. Some ligands bind to specific low free energy conformations. A study by Vieth *et al.* found that the binding affinity of a ligand for a protein can only occur if there is a smooth energy surface with a clear global minimum[11]. Lastly, the hydrophobicity of pharmaceutical compounds can also be significantly affected by the addition of certain substituents. Hydrophobic interactions describe the relationship between water and molecules with low water-solubility. These molecules generally have long chains of carbon atoms that do not interact with water molecules. Hydrophobic interactions are due to van der Waals interactions and depend on changes in enthalpy and entropy, thus having a significant effect on the phase transformation of organic compounds[12]. It is important that all of these factors be taken into account when substituents are added to pharmaceutical compounds.

The strong H-bond interaction energies that were found in Tables 6.2 - 6.5 were dominated by cooperative stabilisation and rotation of the functional group, which does not provide us with any information regarding the inductive effects of these substituents that are placed on the aromatic ring. In order to further investigate the nature of the effect that the addition of substituents to the aromatic ring will have on the intermolecular H-bonds in the structures in

this study, we have studied changes in the electron density of the system by performing AIM analysis.

6.3. AIM analysis of compound C7

In order to simplify as well as minimise the computational resources required for this study, we only investigated the electron density changes by choosing one representative structure that contains one of the functional groups in this study, which is compound ISIFUM (C7). This compound was chosen instead of C_{Ph} as its optimised geometry contains two H-bonds with very similar geometrical parameters. Furthermore, when the substituents were removed and a geometry optimisation was once again performed, the symmetry was maintained. In this section we have investigated the electron density [$\rho(\boldsymbol{r})$], the Laplacian of the electron density [$L(\rho(\boldsymbol{r})$] and the electronic energy density (H_b) at the intermolecular bond critical points (BCPs) for each of the substituted structures by utilising the AIMALL (v.11.04.03) suite[13]. Furthermore, we chose to investigate the electron density changes by adding three electron-donating and three electron-withdrawing groups at the *ortho*, *meta* and *para* position of compound C7, which is illustrated in Figure 6.14.

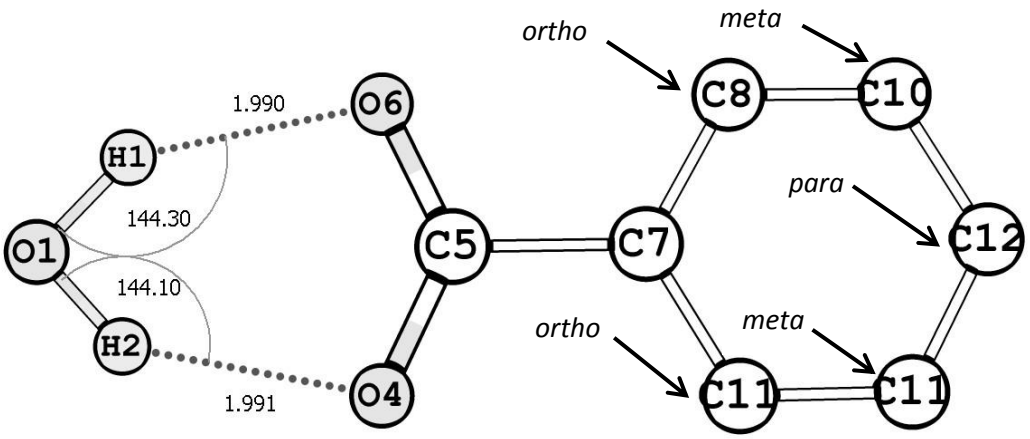

Figure 6.14. Geometrical parameters of the two H-bonds in compound C7 and the relative ortho, meta and para positions where substituents can be added.

The difference in induction strength of these substituents will also be investigated to determine if the induction effect of substituents can be manipulated in order to fine-tune the strength of H-bonds.

Table 6.6. The inductive strength of the electron-withdrawing and electron-donating groups used in this study.

Induction	Electron-withdrawing Group	Electron-donating Group
Strong	Nitro (-NO_2)	Hydroxyl (-OH)
Moderate	Carboxyl (-COOH)	Carboxylic ester (-$OOCCH_3$)
Weak	Halide (-Cl)	Methyl (-CH_3)

Table 6.6 summarises the inductive properties of the substituents used in this study. From the study of Bhavaraju *et al.* we can identify the electron donating groups in Table 6.6 as *ortho–para* directing activators, whereas the electron-withdrawing groups can be identified as *meta*

deactivators[3]. Following the addition of substituents at the *ortho*, *meta* or *para* positions, each structure was optimised the MP2/cc-pVTZ level of theory. Performing optimisation calculations at a high level of theory results in a very good representation of the electron density changes in the system, as wave function files (*wfx files) are generated from fully optimised structures. All *wfx files were generated at the MP2/aug-cc-pVTZ level of theory and single points calculations were performed for all optimised structures at the M06-2X/6-311++G(d,p) level of theory.

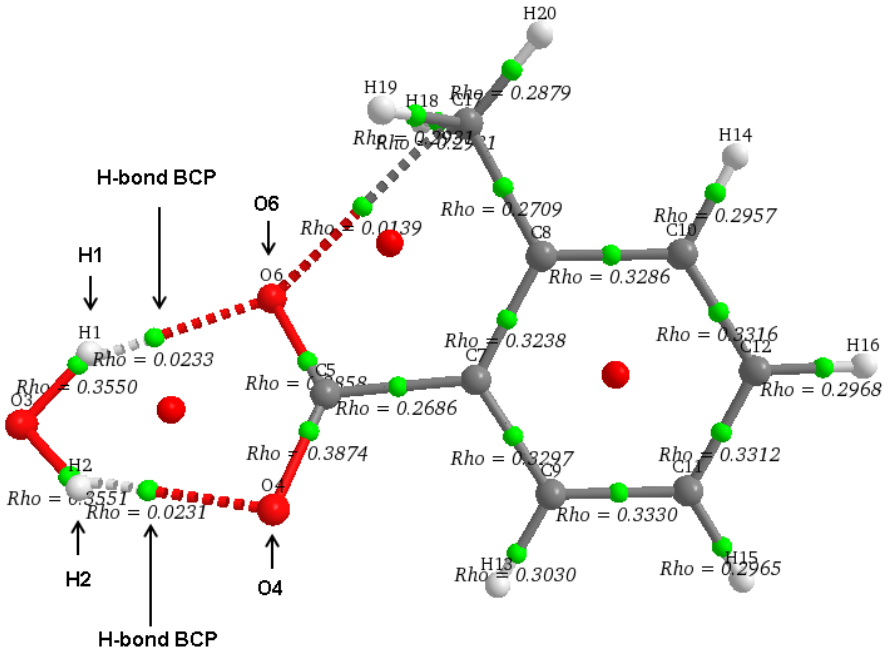

Figure 6.15. Illustration of the charge density plot of structure C7 that has been substituted by a methyl group at the ortho position. The two H-bond BCPs that will be investigated in this study are found between H1 and O6 and H2 and O4. Rho values are equivalent to electron density.

In order to study what effect the addition of substituents will have on the properties of the H-bond in compound C7, we have investigated the parameters at the two BCPs (shown in Figure 6.15). The ring critical points, as well as all other critical points between covalent bonds will not be discussed in this study. The Rho values represent electron density $\rho(r)$, the Laplacian of the $\rho(r)$, $L(\rho(\boldsymbol{r}))$, and the total electronic energy density (H_b).

Table 6.7. Values of E_{INT} (kcal/mol), ρ(*r*) at the BCPs in (ea_0^{-3}), the Laplacian of the electron density [Lρ(*r*)] in (ea_0^{-5}) and the total electronic density H_b in atomic units at the BCPs of the H-bonds for each respective substituent that has been placed on the aromatic ring at the *ortho*, *meta* or *para* position.

		BCP 1 (H1-O6)			BCP 2 (H2-O4)		
Substituent	E_{INT} (kcal/mol)	ρ(*r*) (ea_0^{-3})	L(ρ(r) (ea_0^{-5})	H_b (au)	ρ(*r*) (ea_0^{-3})	L(ρ(r) (ea_0^{-5})	H_b (au)
None	-20.31	0.023	0.088	0.001	0.023	0.088	0.0014
p-CH3	-21.26	0.023	0.088	0.001	0.023	0.088	0.0014
o-CH3	-20.98	0.023	0.088	0.001	0.023	0.088	0.0015
p-OOCCH3	-20.25	0.022	0.083	0.002	0.020	0.079	0.0020
o-OOCCH3	-19.69	0.022	0.085	0.002	0.022	0.086	0.0017
p-OH	-21.81	0.022	0.083	0.002	0.021	0.081	0.0019
o-OH	-18.41	0.015	0.064	0.003	0.027	0.094	-0.0003
m-NOO	-19.39	0.016	0.067	0.002	0.024	0.087	0.0010
m-COOH	-20.23	0.024	0.087	0.001	0.017	0.071	0.0024
m-Cl	-20.31	0.023	0.086	0.002	0.023	0.086	0.0016

The various parameters of the two BCPs in compound C7 are listed in Table 6.7, where we note that the E_{INT} (kcal/mol) values are larger when electron donating groups are placed at the *para* position. All EINT values are measure in kcal/mol. By comparing the relative inductive strengths of the electron-donating and electron-withdrawing groups to their relative ρ(*r*), L(ρ(***r***) and H_b values in the table, no definitive trend can be identified. If inductive effects were to play a role, we would expect that the ρ(*r*) values at the BCPs would be higher if electron-donating groups are placed at the *ortho* and *para* positions. This is however not observed as some of the ρ(*r*) values for the hydroxyl substituents are actually the lowest in the table. The opposite effect would be expected for the electron-withdrawing groups, with less ρ(*r*) at their BCPs when they are placed at the *meta* position. Furthermore, we expect stronger H-bonds when electron-donating groups have been placed on the ring and the opposite trend for the electron-withdrawing group. The strongest H-bond is actually observed for the *para* hydroxyl substituted compound, but we also observe the weakest H-bond for the *ortho* hydroxyl substituted compound. The slightly weaker H-bond interaction energies of the three electron-withdrawing groups would suggest that there are some inductive forces at play, but the ρ(*r*) values do not correspond to this observation. A study by Kim *et al.* found that both electron-donating and electron-withdrawing groups can increase energy, which is due to a subtle interplay between electronic terms contributing to the total interaction energy[14].

Investigation of the optimised geometries of these substituted compounds revealed that cooperative stabilisation and rotation of the carboxylate functional group are the dominant factors that affect the interactions energies in Table 6.7. The strongest H-bond in the minimum conformation of the p-OH substituted compound shows a planar arrangement with very little rotation of the functional group and cooperative stabilisation of the H-bond. All of the other substituted compounds show cooperative stabilisation with rotation of the functional group, which is accompanied by an energy penalty to the interaction energy as was described by Rakatin and cowokers[4]. The rotation of the functional group is illustrated below by comparing the planar arrangement in the *para* hydroxyl substituted compound to that of the

meta carboxyl substituted compound by looking down the plane of C8, C5 and C11 and defined in Figure 6.16.

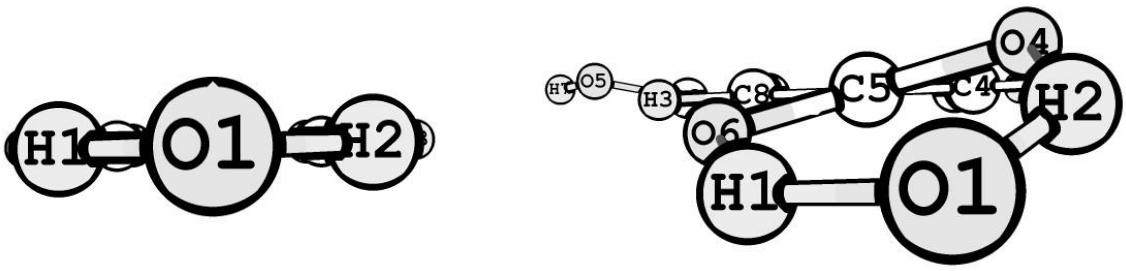

Figure 6.16. The rotation of the carboxylate functional group when compound C7 has been substituted with the hydroxyl substituent at the *para* position (left) and the carboxyl substituent at the *meta* position (right).

The rotation of the carboxylate group in Figure 6.116 results in an overall energy penalty to the carboxyl substituted group, which accounts for the change in interaction energy between these two compounds. The H_b values in Table 6.7 can however provide us with some characteristic properties of the H-bonds in each substituted compound. Nakanishi *et al.* investigated the properties of weak to strong interactions by utilising AIM[15]. In their study they constructed plots of electron energy densities (H_b) versus the Laplacians $L(\rho(r))$ at BCPs. From these plots they identified regions that correspond to different types of interactions or bonds, namely van der Waals interactions, H-bonds, molecular complexes and hypervalent adducts through charge transfer (CT) interactions, and covalent bonds. Values obtained from the work of Nakanishi *et al.* are listed in Table 6.8, which shows the AIM parameters ($\rho(r)$, $L\rho(r)$, H_b) that can be used to identify the specific interaction in a structure.

Table 6.8. Classification of different intermolecular interactions with AIM parameters according to reference 15 .

Interaction	$\rho(r)$	$L(\rho(r))$	H_b
Van der Waals	$0.00 < \rho(r) < 0.01$	$0.00 < L(\rho(r)) < 0.04$	$0.000 < H_b < 0.002$
H-bond	$0.01 < \rho(r) < 0.04$	$0.04 < L(\rho(r)) < 0.12$	$-0.004 < H_b < 0.002$
Molecular complex	$0.01 < \rho(r) < 0.03$	$0.02 < L(\rho(r)) < 0.06$	$-0.001 < H_b < 0.002$
Weak covalent bonds	$0.05 < \rho(r) < 0.17$	$-0.10 < L(\rho(r)) < 0.07$	$-0.130 < H_b < -0.030$

From Table 6.8 we see that the parameters of the H-bond BCPs in Table 6.7 fall within the classification of an H-bond as defined by Nakanishi *et al.*. Even though we observe slight changes in these parameters, the parameters remain within the ranges given for an H-bond.

Chapter summary

In this study we have investigated what effect various substituents will have on the H-bond strength when placed at five different positions of an aromatic ring. Furthermore, this was done for four aromatic rings with different functional groups, namely carboxylate, nitrogen dioxide, sulfonate and phosphonate. For all of these functional groups it was found that when substituents are placed at different positions on the ring, they will result in E_{INT} changes for each functional group. These changes were as a result of the rotation of the functional group, for which the degree of rotation will depend on the steric repulsion of substituents at different

positions on the ring. Furthermore, it was found that substituents that are placed on position 1 or 5 of the aromatic ring can result in cooperative stabilisation of the H-bond.

The strongest H-bonds for each functional group were observed when a specific substituent was placed at a position that allows the rotation functional group, which will result in a minimum energy conformation that can be cooperatively stabilised. The strongest H-bond in the carboxylate functional group was observed when the formyl substituent was placed at position 1 of the aromatic ring, The strongest H-bond for the carboxylate functional group was calculated to be -19.28 kcal/mol. This interaction energy is 1.47 kcal/mol weaker than the unsubstituted aromatic ring as the unsubstituted compounds shows very little rotation observed when no substituents are added to the ring. For the nitrogen dioxide functional group, it was found that the strongest H-bond has an interaction energy of -10.53 kcal/mol when the nitro group is placed at position 1, which is more than double in strength of the unsubstituted compound. The strongest H-bond for the sulfonate functional group was calculated to be -18.19 kcal/mol when the methyl group is placed at position 3, which is 5.61 kcal/mol stronger than the unsubstituted compound. The strongest H-bond for the phosphonate functional group was calculated to be -34.65 kcal/mol when the hydroxyl group is placed at position 3, which is 5.26 kcal/mol stronger than the unsubstituted compound.

Furthermore, AIM analysis showed that inductive effects do not affect the interaction energy, as there is no significant change in the charge density $\rho(r)$ values when electron-donating and electron-withdrawing groups are placed at the *ortho*, *meta* or *para* positions on the aromatic ring. The results in this study have rather suggested that cooperative stabilisation and the rotation of the functional group are the major factors that influence the strength of the H-bond. The results in this study have identified the substituent effect as being an effective method for increasing the strength of the H-bond. The methodology described in this chapter could serve as a viable fundamental basis when fine-tuning the strength of H-bonds in pharmaceutical compounds to increase the propensity of hydrate formation or modify and improve their properties.

References

(1) Lutz, H. D., Beckenkamp, K., Möller, H.: Weak hydrogen bonds in solid hydroxides and hydrates. *Journal of Molecular Structure* **1994**, *322*, 263-266.

(2) Briggs, N. E. B., Kennedy, A. R., Morrison, C. A.: 42 salt forms of tyramine: structural comparison and the occurrence of hydrate formation. *Acta Crystallographica Section B* **2012**, *68*, 453-464.

(3) Bhavaraju, M., Gwaltney, S. R.: A theoretical analysis of substituted aromatic compounds. *International Journal of Quantum Chemistry* **2013**, *113*, 1171-1179.

(4) Rakitin, A. R., Pack, G. R.: Necessity of Aromatic Carboxylate Anions To Be Planar To Induce Growth of Cationic Micelles. *Langmuir* **2004**, *21*, 837-840.

(5) Hunter, C. A., Ihekwaba, N., Misuraca, M. C., Segarra-Maset, M. D., Turega, S. M.: Cooperativity in multiply H-bonded complexes. *Chemical Communications* **2009**, 3964-3966.

(6) Ebrahimi, A., Habibi, M., Sadat Neyband, R., Reza Gholipour, A.: Cooperativity of small pi-stacking and hydrogen bonding interactions and substituent effects on X-ben pyridine H-F complexes. *Physical Chemistry Chemical Physics* **2009**, *11*, 11424-11431.

(7) Mata, I., Molins, E., Alkorta, I., Espinosa, E.: Tuning the Interaction Energy of Hydrogen Bonds: The Effect of the Substituent. *The Journal of Physical Chemistry A* **2011**, *115*, 12561-12571.

(8) Jones, J. E.: On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings of the Royal Society of London Series A* **1924**, *106*, 463-477.

(9) Zhang, H.-x., Liu, Y.: Protein-binding properties of a designed steroidal lactam compound. *Steroids* **2014**, *80*, 30-36.

(10) Reynisson, J., McDonald, E.: Tuning of hydrogen bond strength using substituents on phenol and aniline: A possible ligand design strategy. *Journal of Computer-Aided Molecular Design* **2004**, *18*, 421-431.

(11) Vieth, M., Hirst, J. D., Kolinski, A., Brooks, C. L.: Assessing energy functions for flexible docking. *Journal of Computational Chemistry* **1998**, *19*, 1612-1622.

(12) Chuajiw, W., Takatori, K., Igarashi, T., Hara, H., Fukushima, Y.: The influence of aliphatic amines, diamines, and amino acids on the polymorph of calcium carbonate precipitated by the introduction of carbon dioxide gas into calcium hydroxide aqueous suspensions. *Journal of Crystal Growth* **2014**, *386*, 119-127.

(13) Keith, T. A.: AIMAll *TK Gristmill Software* **2012**

(14) Lee, E. C., Kim, D., Jurečka, P., Tarakeshwar, P., Hobza, P., Kim, K. S.: Understanding of Assembly Phenomena by Aromatic−Aromatic Interactions: Benzene Dimer and the Substituted Systems. *The Journal of Physical Chemistry A* **2007**, *111*, 3446-3457.

(15) Nakanishi, W., Hayashi, S., Narahara, K.: Atoms-in-Molecules Dual Parameter Analysis of Weak to Strong Interactions: Behaviors of Electronic Energy Densities versus Laplacian of Electron Densities at Bond Critical Points. *The Journal of Physical Chemistry A* **2008**, *112*, 13593-13599.

CHAPTER 7

The effect of water activity on hydrate formation

Introduction

A common misconception about hydrate formation is that crystallisations with water as the solvent will lead to the incorporation of water in the crystal structure. Water is a very small and polar reactive solvent that can allow strong interactions between the host and guest molecules, as well as efficient packing[1]. Water as a crystallisation solvent can also result in a more stable hydrogen bond (H-bond) arrangement than when the organic molecule crystallises alone[2]. These properties are however influenced by the potential of water molecules within the crystalline environment[3], a thermodynamic property that was defined as water activity (a_w) in Chapter 2. It has been shown that whether the product of a crystallisation is the hydrate or the anhydrous phase of a pharmaceutical is determined by the activity of water in the crystallisation solvent. Furthermore, water activity has also been shown to be capable of influencing the catalytic activity of enzymes in a negative or positive manner[4]. The critical a_w is a specific value where the hydrated and anhydrous crystals are in equilibrium. A low water activity critical value corresponds to a compound that has a greater thermodynamic stability[5]. We aim to investigate the effect of water activity on hydrate formation in three binary systems. In each system a water-organic solvent mixture was added to a pharmaceutically acceptable compound.

7.1. Water activity

Water activity values can be calculated for simple binary systems, consisting of water and an organic solvent as the two components. Several models have been designed in order to calculate water activity values for these binary systems. We chose the non-random two-liquid (NRTL) model for our study as it has been specifically designed for experimentally-determined phase equilibrium data. A list of experimental parameters is required in order to calculate water activity values with the NRTL model. The Dortmund Data Bank (DDB) contains all of the experimental parameters that are required for the calculations in our study[6]. The NRTL model provides more flexibility than other activity models due to the additional alpha ($\propto_{12} = \propto_{21}$) parameter. The NRTL equation describes the relationship between the activity coefficients γ_i of a compound with its mole fraction x_i in a liquid phase[7]. Other models could also be used, but the NRTL model has specifically been parameterised to describe mixed solvent systems. The effect of water activity was investigated in three binary systems. These are water with ethanol, acetone or ethyl acetate. These three organic solvents are common solvents that have the necessary experimental data available to calculate water activity values. The water activity values for the three binary systems are summarised in Table 7.1 below. The calculations that were performed in order to obtain all water activity values were described in Chapter 2.

Table 7.1. Lists the mole fraction and water activity values for all three binary systems.

Ethanol			Acetone			Ethyl Acetate		
x_{Water}	x_{Ethanol}	a_{w}	x_{Water}	x_{Acetone}	a_{w}	x_{Water}	$x_{\text{Ethyl Acetate}}$	a_{w}
0	1	0.00	0	1	0.00	0.00	1.00	0.00
0.26	0.74	0.34	0.31	0.69	0.43	0.38	0.62	0.59
0.45	0.55	0.52	0.50	0.50	0.60	0.58	0.42	0.72
0.58	0.42	0.64	0.64	0.36	0.70	0.70	0.30	0.79
0.68	0.32	0.72	0.73	0.27	0.77	0.78	0.22	0.84
0.76	0.24	0.79	0.80	0.20	0.83	0.84	0.16	0.88
0.83	0.17	0.84	0.86	0.14	0.87	0.89	0.11	0.91
0.88	0.12	0.89	0.90	0.10	0.91	0.93	0.07	0.93
0.93	0.07	0.93	0.94	0.06	0.94	0.96	0.04	0.96
0.97	0.03	0.97	0.97	0.03	0.97	0.98	0.02	0.98
1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00

Crystal formation took place at different a_{w} values for the three binary systems. This effect can be explained by the direct relationship between the mole fraction of water and its activity. For binary mixtures, the water activity a_{w} will decrease as the mole fraction of water x_{Water} decreases.

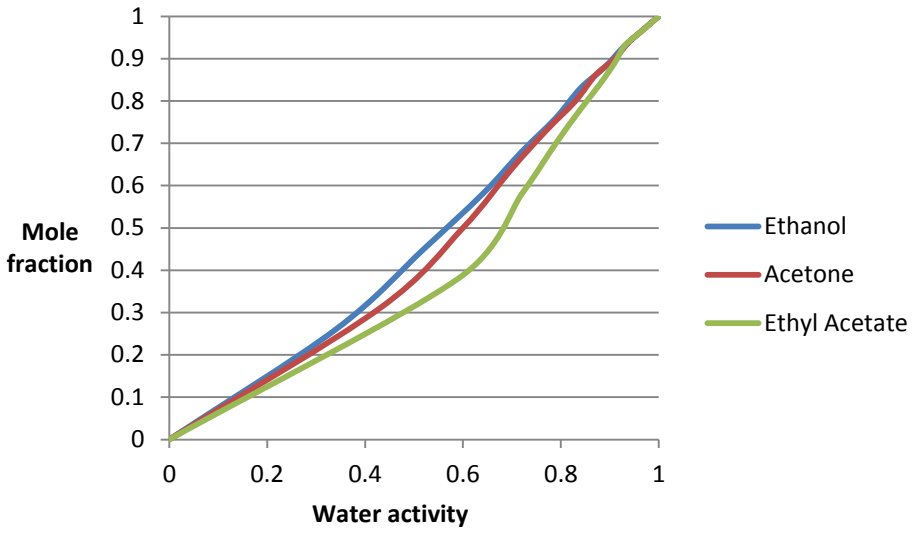

Figure 7.1. The relationship between water activity and mole fraction for the three binary systems

The difference in water activity values for each solvent system is graphically illustrated in Figure 7.1. It is clear that the potential of the water molecules in each solvent system is different. This can be explained by the difference in polarity of each organic solvent, with ethanol being the least polar and ethyl acetate the most polar solvent. More polar solvents will weaken the solute-solute interactions to a greater effect.

7.2. Crystal formation from organic solvent mixtures

Three organic compounds were investigated in this study, namely pyridine-4-carboxylic acid (isonicotinic acid), N-(aminoiminomethyl)-N-methylglycine (creatine) and benzoic acid. These three carboxylic acid compounds were chosen as they are pharmaceutically acceptable acids that should have a high propensity to form hydrates[8]. A series of crystallisations was performed by adding 1 mL of the water-organic solvent mixture (with a 1 mL disposable syringe) to approximately 0.02 g of one of the carboxylic acid compounds. The water component was added in increments of 0.1 mL with the corresponding organic solvent component making up the other part of the total 1 mL; for example, 0.4 mL of water and 0.6 mL of the organic solvent component. The a_w values were calculated for each solvent mixture in order to measure in what range crystal formation took place. Once the organic solvent was added to the pharmaceutically acceptable compound, crystal formation would generally take place after one to three days. All XRD analysis experiments were performed at 273 K.

7.2.1. Creatine monohydrate

Crystal formation of creatine in water-organic solvent mixtures took place after approximately two days for all a_w values above 0.52. The most uniform crystals were observed when a_w at 0.52 for the water and ethanol mixture. These crystals were used for further analysis with various experimental techniques. The thermogravimetric analysis (TGA) of creatine monohydrate is shown in Figure 7.2. We observe a loss of 12.66% weight (0.53 mg from the 3.86 mg sample) at approximately 99 °C for creatine monohydrate, which corresponds to the loss of a water molecule.

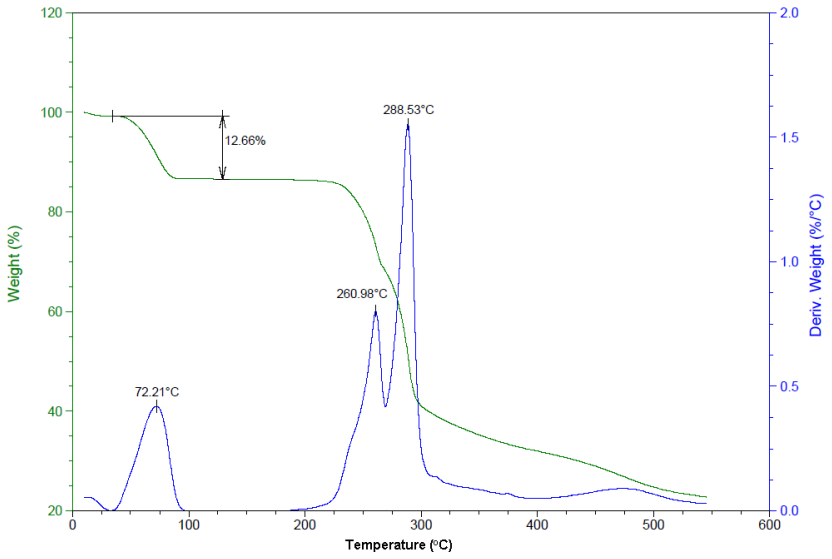

Figure 7.2. TGA analysis of the creatine monohydrate.

This was confirmed by the Differential Scanning Calorimetry (DSC) analysis of creatine monohydrate, which is shown in Figure 7.3. From the trace we observe one single thermal event at approximately 99°C that corresponds to the loss of water from the system.

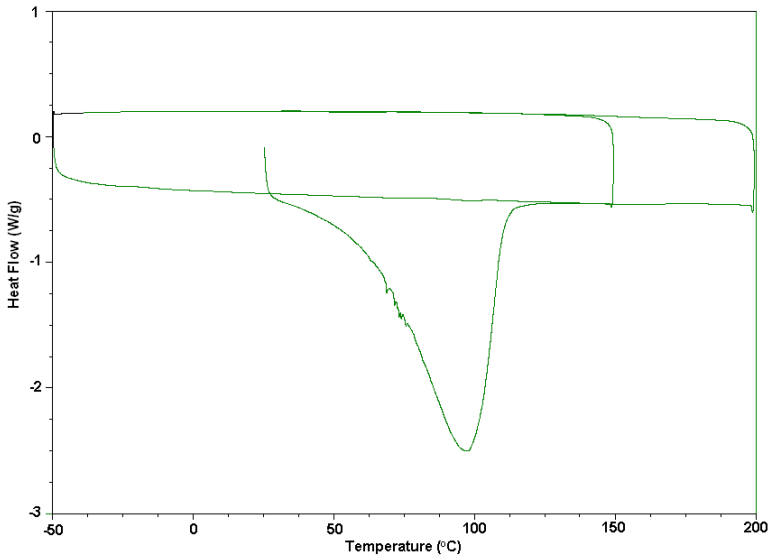

Figure 7.3. DSC trace of the creatine crystal from the water-ethanol mixture.

The structure of creatine monohydrate was determined by X-ray diffraction analysis of a suitable single crystal. The unit cell parameters were compared to those of the creatine monohydrate (refcode = CREATH04) given in the CSD; unit cell parameters are included in Table 7.2.

Table 7.2. Comparison of the reduced cell parameters of the experimental structure and CREATH04 from reference 9.

Reduced cell parameters		
Structure	Experimental structure	CREATH04
Space group	$P2_1/c$	$P2_1/c$
a (Å)	5.05	4.97
b (Å)	12.15	12.05
c (Å)	12.51	12.49
α (°)	108.90	109.10
β (°)	90.00	90.00
γ (°)	90.00	90.00
Volume (Å^3)	726.20	706.57

The values in Table 7.2 are in good agreement, which confirms that we have obtained the desired creatine monohydrate product.

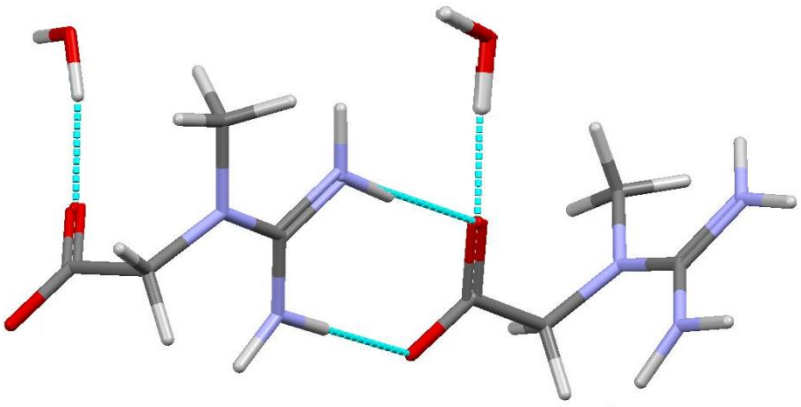

Figure 7.4. Illustrates the H-bond network for the structure of creatine monohydrate from reference 9. Dotted lines indicate the H-bonds.

Creatine forms a monohydrate crystal in the monoclinic space group $P2_1/c$. The structure has an H-bond between the carboxylate group and the water molecule as illustrated in hydrogen bonding network in Figure 7.4.

7.2.2. Anhydrous creatine

After obtaining the structure of creatine monohydrate, we wished to compare the stability of this structure with the anhydrous equivalent. Therefore we made use of sublimation, which is a common technique used for purification, as well as to induce phase transitions during crystallisation. The crystal structure of a resulting crystal was determined by X-ray diffraction analysis. The unit cell of the product corresponded to that of creatinine (refcode = CREATI) and is given in Table 7.3.

Table 7.3. Comparison of the reduced parameters of the experimental structure and creatinine from reference 10 .

Reduced cell parameters		
Structure	Experimental structure	CREATI
Space group	$P2_1/c$	$P2_1/c$
a (Å)	5.87	5.97
b (Å)	7.98	8.06
c (Å)	11.33	11.50
α (°)	93.59	95.94
β (°)	89.60	90.00
γ (°)	89.95	90.00
Volume (Å^3)	528.20	550.21

Table 7.3 shows that the creatinine structure was obtained and not the desired anhydrous product. This transformation has been previously reported by Dash *et al.*, who found that the structure of creatine monohydrate undergoes intramolecular cyclisation with the loss of an additional mole of water to form creatinine[11].

We then attempted to obtain the anhydrous product using supercritical CO_2, which is a modern technique that has been described as a new medium for the generation of novel crystalline forms of active pharmaceutical ingredients[12]. Unfortunately this too did not

provide the anhydrous product as the structure once again reverted to creatinine. The creatinine structure is illustrated in Figure 7.5.

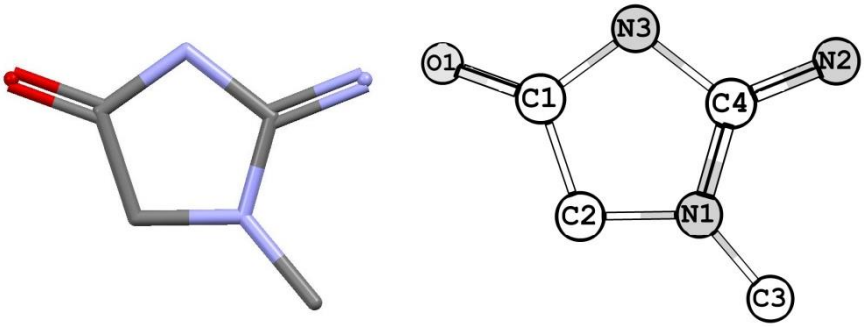

Figure 7.5. The structure of creatinine identified by X-ray diffraction analysis of crystals obtained from sublimation and supercritical CO_2.

7.2.3. Isonicotinic acid

The hydrate formation of isonicotinic acid was specifically investigated due to the presence of the carboxylate group and aromatic nature of this compound. Similarly to the six-membered carbon aromatic ring, benzene, the heterocyclic pyridine ring also contains delocalised electrons and thus we could test the predictions made in Chapter 6, which suggest that substituted aromatic compounds with pharmaceutically acceptable anionic groups will form H-bonds if the rotation of the anionic functional group allows for the formation of the H-bond. Furthermore, substituents adjacent to the functional anionic group should cooperatively stabilise the H-bond. Crystal formation of isonicotinic acid in water-organic solvent mixtures took place after approximately two days for all a_w values above 0.60. The crystal obtained at a_w= 0.60 was analysed by X-ray diffraction.

Table 7.5. Comparison of the reduced parameters of the experimental structure and isonicotinic acid from reference 13.

Reduced cell parameters		
Structure	Experimental structure	ISNICA
Space group	$P\bar{1}$	$P\bar{1}$
a (Å)	6.41	6.39
b (Å)	7.11	7.23
c (Å)	7.46	7.47
α (°)	99.45	103.66
β (°)	117.94	114.88
γ (°)	106.53	106.19
Volume (Å^3)	270.10	274.90

The unit cell parameters in Table 7.5 showed that the anhydrous crystals of isonicotinic acid (refcode = ISNICA) were obtained. This suggests that the anhydrate will be the thermodynamically favoured product when placed in a water-organic solvent mixture. The structure of isonicotinic acid is illustrated in Figure 7.6.

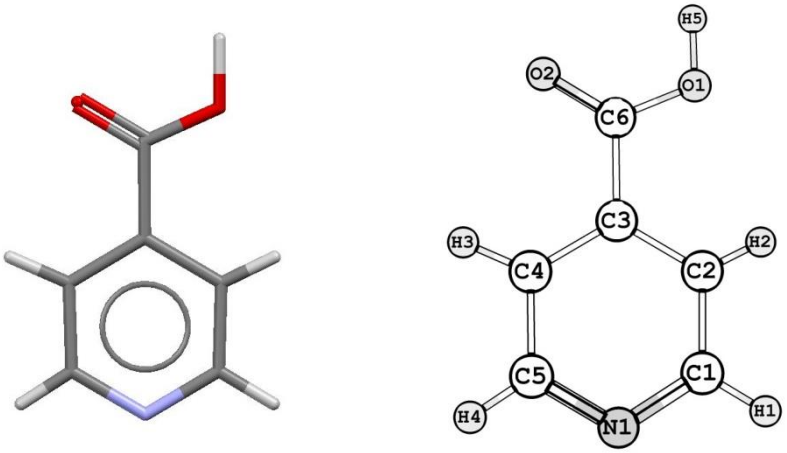

Figure 7.6. The structure of isonicotinic acid.

A CSD search for nicotinic acid derivatives produced the structure of 1-methylpyridinium-4-carboxylate monohydrate, which is illustrated in Figure 7.7.

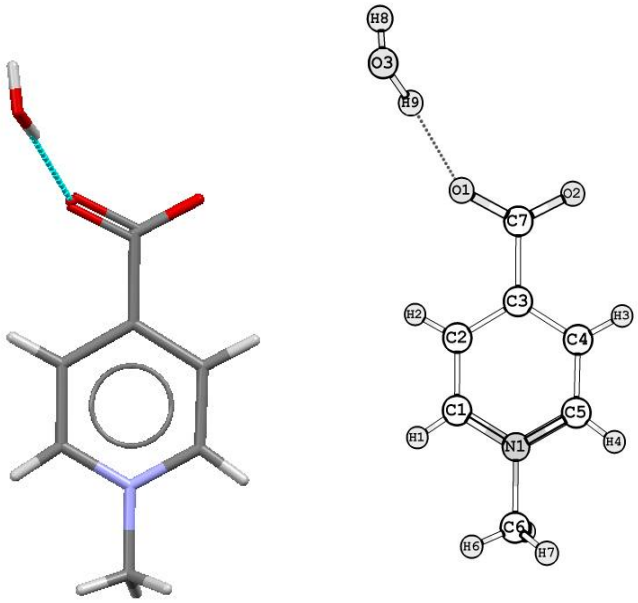

Figure 7.7. The structure of 1-methylpyridinium-4-carboxylate monohydrate from reference 14.

The presence of an H-bond in the crystal structure of 1-methylpyridinium-4-carboxylate monohydrate proves that the presence of substituents at specific positions on the aromatic ring can result in the formation of H-bonds. For this particular structure, the methyl substituent allowed for the rotation of the carboxylate functional group to form the H-bond to the water molecule. Furthermore, due to the fact that the thermodynamically favoured product is the hydrate, it suggests that the formation of the H-bond plays an important role in the propensity of compounds to form hydrates as was suggested in previous chapters.

7.2.4. Trimesic acid dihydrate

A series of crystallisations were performed for trimesic acid in various solvent mixtures as for creatine and isonicotinic acid. Trimesic acid is known to form a dihydrate, with the procedure to form this specific hydrate described by Fan *et al.*[15]. The process simply requires the addition of any amount of water to trimesic acid to form the trimesic acid dihydrate, since all solvent mixtures containing water produced trimesic acid dihydrate crystals. Furthermore, any solvent mixture that contains water as one of its components will still result in the trimesic acid dihydrate product. For this particular material we observed a significant change in the appearance of the crystals as the water activity values changed. Zobrist *et al.* performed a study to investigate the effect of water activity on crystal formation and proposed a new thermodynamical model to explain the nucleation of crystals, which can be described as the process by which new crystals are formed. The results showed that nucleation rate cannot be explained by dilution or solute concentration, but rather by solvent water activity[16].

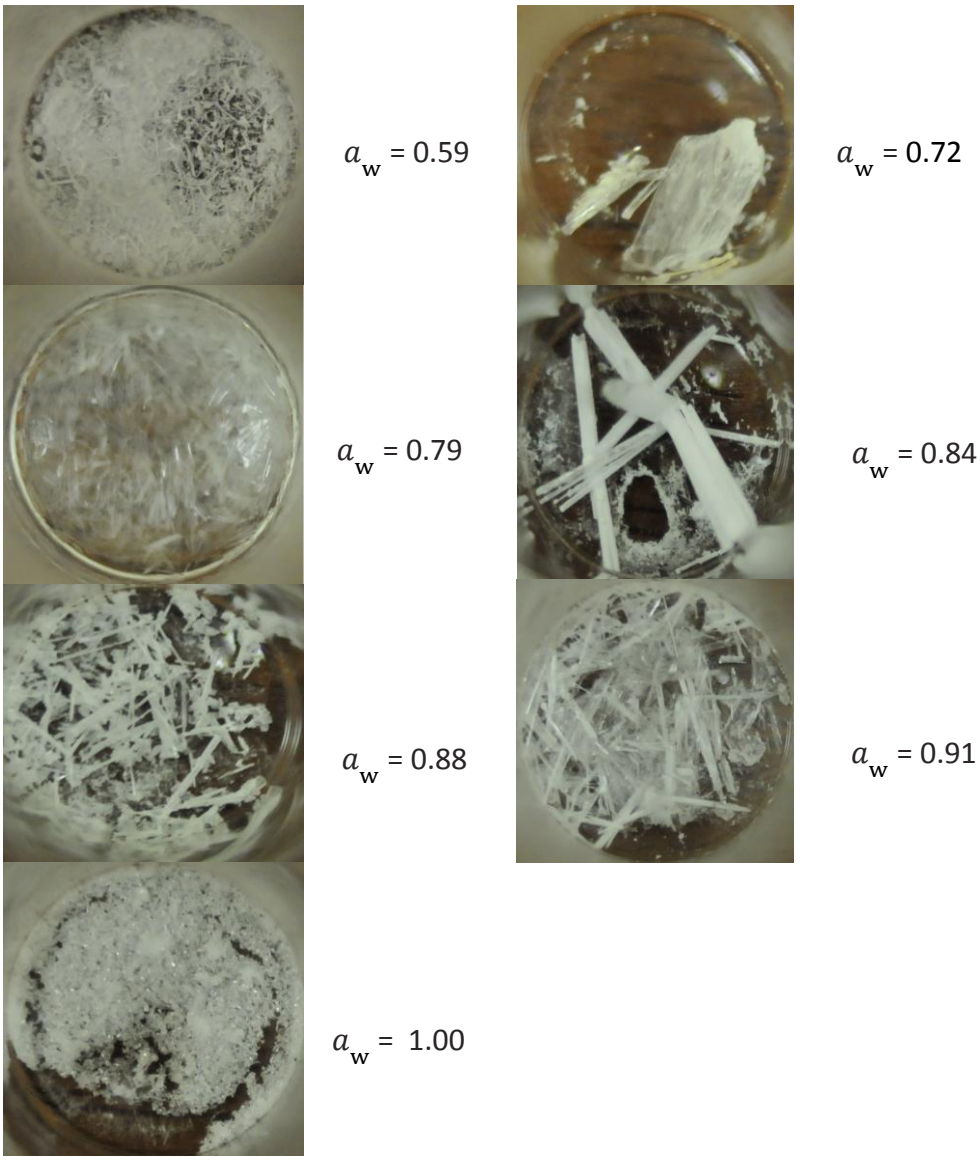

Figure 7.8. Change in trimesic acid hydrate crystal appearance at different a_w values. Crystals were obtained from a water-ethyl acetate solvent mixture.

The change in crystal shape and appearance of trimesic acid dihydrate crystals grown at different a_w values in the mixed solvent system consisting of water and ethyl acetate is illustrated in Figure 7.8. Only the a_w values where pronounced changes occur were included in Figure 7.8. It can be observed that the largest crystals were not obtained at a_w = 1.00. This then shows that the rate of nucleation at different a_w values can serve as a viable method for growing crystals of an appropriate size and shape to be used for SCD, TGA, DSC and other analytical techniques.

Chapter summary

The critical a_w values at which creatine monohydrate and isonicotinic acid crystals form were identified in this study. The crystals of creatine monohydrate, creatinine and isonicotinic acid were analysed by XRD by comparing the obtained unit cell parameters to those found in the CSD. The results in this study have shown that the hydrated or anhydrous product of different compounds can be obtained at different a_w values. Although it was not possible to obtain a hydrate of isonicotinic acid, a related compound with a methyl substituent on the aromatic ring that did form a H-bond was identified from the CSD. This suggests that the substitution affected the ability of the compounds to from an H-bond between the carboxylate group and the water molecule. We suggested therefore that the ability of this compound to form the H-bond is why the hydrated form is found experimentally. Lastly, the study shows that a_w can be utilised when attempting to grow crystals of an appropriate shape and size to be used for further analysis of the compound.

References

(1) Vippagunta, S. R., Brittain, H. G., Grant, D. J. W.: Crystalline solids. *Advanced Drug Delivery Reviews* **2001**, *48*, 3-26.

(2) Desiraju, G. R.: Hydration in organic crystals: prediction from molecular structure. *Journal of the Chemical Society, Chemical Communications* **1991**, 426-428.

(3) Cazier, J.-B., Gekas, V.: Water activity and its prediction: A review. *International Journal of Food Properties* **2001**, *4*, 35-43.

(4) Yang, Z., Zacherl, D., Russell, A. J.: pH Dependence of subtilisin dispersed in organic solvents. *Journal of the American Chemical Society* **1993**, *115*, 12251-12257.

(5) Jayasankar, A., Roy, L., Rodríguez-Hornedo, N.: Transformation pathways of cocrystal hydrates when coformer modulates water activity. *Journal of Pharmaceutical Sciences* **2010**, *99*, 3977-3985.

(6) Onken, U., Rarey-Nies, J., Gmehling, J.: The Dortmund Data Bank: A computerized system for retrieval, correlation, and prediction of thermodynamic properties of mixtures. *International Journal of Thermophysics* **1989**, *10*, 739-747.

(7) Renon, H., Prausnitz, J. M.: Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE Journal* **1968**, *14*, 135-144.

(8) Aitipamula, S., Wong, A. B. H., Chow, P. S., Tan, R. B. H.: Pharmaceutical Salts of Haloperidol with Some Carboxylic Acids and Artificial Sweeteners: Hydrate Formation, Polymorphism, and Physicochemical Properties. *Crystal Growth & Design* **2014**, *14*, 2542-2556.

(9) S. Frampton, C., C. Wilson, C., Shankland, N., J. Florence, A.: Single-crystal neutron refinement of creatine monohydrate at 20 K and 123 K. *Journal of the Chemical Society, Faraday Transactions* **1997**, *93*, 1875-1879.

(10) du Pre, S., Mendel, H.: The crystal structure of creatinine. *Acta Crystallographica* **1955**, *8*, 311-313.

(11) Dash, A. K., Mo, Y., Pyne, A.: Solid-state properties of creatine monohydrate. *Journal of Pharmaceutical Sciences* **2002**, *91*, 708-718.

(12) Padrela, L., Rodrigues, M. A., Velaga, S. P., Matos, H. A., de Azevedo, E. G.: Formation of indomethacin–saccharin cocrystals using supercritical fluid technology. *European Journal of Pharmaceutical Sciences* **2009**, *38*, 9-17.

(13) Takusagawa, F., Shimada, A.: Isonicotinic acid. *Acta Crystallographica Section B* **1976**, *32*, 1925-1927.

(14) Szafran, M., Bartoszak-Adamska, E., Koput, J., Dega-Szafran, Z.: Structure of 1-methylpyridinium-4-carboxylate monohydrate studied by X-ray, FT-IR, Raman, NMR and ab initio methods. *Journal of Molecular Structure* **2007**, *844–845*, 140-156.

(15) Fan, Z.-Z., Li, X.-H., Wang, G.-P.: Trimesic acid dihydrate. *Acta Crystallographica Section E* **2005**, *61*, o1607-o1608.

(16) Zobrist, B., Marcolli, C., Peter, T., Koop, T.: Heterogeneous Ice Nucleation in Aqueous Solutions: the Role of Water Activity. *The Journal of Physical Chemistry A* **2008**, *112*, 3965-3975.

CHAPTER 8

Conclusions and future work

8.1. Conclusions

The propensity of various pharmaceutically relevant anions to from hydrates was investigated in this study. A CSD study showed that increasing number of polar groups and increased charge density on anionic functional groups increases the likelihood of hydrate formation. Various other pharmaceutically acceptable anionic groups were also identified that were previously thought to be harmful, but have recently emerged as the components of novel pharmaceutical compounds. This led to a further investigation into the role that the H-bond plays in the propensity of these functional groups to form hydrates.

PES scans were utilised to investigate the chemical environment that surrounds the H-bond in the crystal structure of creatine monohydrate determined using neutron diffraction. A total of four scans was performed, namely horizontal, vertical, LTR diagonal and RTL diagonal scans, which resulted in the identification of six attractive H-bonds in the four curves that were generated utilising various DFT and WFT methods. It was found that the MP2/6-311++G(d,p) level of theory provided the most accurate results when compared to a CCSD(T)/6-311++G(d,p) benchmark level of theory. Comparing the interaction energies obtained for optimised geometries, it was found that they are up to 6.9 kcal/mol stronger than the H-bond interaction energy calculated for the experimental structure.

Various factors that influence the H-bond interaction energy were discussed in Chapter 5. In order to accelerate the geometry optimisations that were required for these calculations, various DFT and WFT methods were used to optimise the H-bond parameters in the neutron structure of creatine monohydrate. It was found that calculations that were performed at the M06-2X/6-311++G(d,p) level of theory provided the most accurate results when compared to the benchmark MP2/aug-cc-pVTZ level of theory. The M06-2X/6-311++G(d,p) level of theory was thus used in combination with high level WFT methods for this aspect of the study.

A total of 44 compounds were identified that contain various anionic functional groups. The structures that were not analysed by neutron diffraction were optimised with the MP2 method in combination with the cc-pVTZ and aug-cc-pVTZ basis sets. It was found that the average H-bond strength increases as the number of polar groups and charge density increases, which corresponds to the trend that was observed in the CSD survey.

The influence of the chemical environment around each functional group was further investigated by determining the steric density values of each structure by calculating hydrogen bond propensity values for H-bond donor and acceptor atoms. It was found that steric density has the most pronounced effect on structures that contain the sulfonate functional group. The sulfonate with the lowest steric density value of 11.05 yielded an H-bond interaction energy of -30.9 kcal/mol. Furthermore, the sulfonate structure with the highest steric density of 32.54 has the lowest interaction energy of -6.2 kcal/mol. It was however found that there is no direct relationship between interaction energies and steric density values as some structures with a high steric density still exhibit a high interaction energy. A different trend was observed for structures that contain aromatic rings, which are affected by ring substituents that have a significant effect on the strength of the H-bond even

in the presence of large steric density values. If all other non-aromatic structures are considered, we can conclude that steric effects from the surrounding environment can result in stronger interaction energies.

The effect of the crystalline environment on H-bond interaction energies was investigated by performing calculations that selectively optimise the H-bond geometrical parameters using a combination of the M06-2X and MP2 methods utilising the 6-311++G(d,p) basis set. This was only performed for 25 of the 44 structures that had not been determined by neutron diffraction. We saw that the majority of the structures have stronger H-bond interaction energies in the structures that have been allowed to undergo full geometry optimisation, which is expected as optimisation allows the structures to adopt minimum energy conformations that have higher H-bond interaction energies as a result of cooperative stabilisation and optimal geometrical conformations. It was found that the few noticeable examples of structures that have stronger H-bond interaction energies could be attributed to steric effects from the surrounding crystalline environment that reduce the interaction distance and make the interaction angle more linear.

The dependence of the H-bond interaction energy on the electrostatic environment was further investigated by performing optimisations for four anionic functional groups in various solvents using the implicit solvent model. These optimisations were performed at the MP2/ 6-311++G(d,p) level of theory, which was followed by interaction energy calculations at the M06-2X/6-311++G(d,p) level of theory. The results from the solvent dependency study showed that there is an overall decrease in interaction energy as the relative permittivity of the solvent increases, with the weakest H-bond interaction energy for each anionic functional group found when water is the chosen solvent. Furthermore, the decrease in interaction energy is more pronounced for some of the anionic functional groups. This is due to the increase in solute-solvent interactions for groups that are more polar, which in turn will result in the decrease of the electrostatic (E_{el}) contribution to the total interaction energy. To summarise, more polar groups will result in a smaller E_{el} contribution, which will then result in a weaker H-bond interaction energy.

It was noted in Chapter 5 that compounds that contain substituted aromatic rings showed interesting behaviour. This led to a study described in Chapter 6 that investigates the effect of the chemical environment neighbouring the H-bonded functional group on the interaction energy in aromatic compounds. Different substituents were placed at one of the five positions on the aromatic rings of compounds containing four different functional groups, namely carboxylate, nitrogen dioxide, sulfonate and phosphonate. A total of 80 structures were optimised using a combination of the M06-2X and MP2 methods utilising the 6-311++G(d,p) basis set, which was followed by counterpoise-corrected single point calculations at the M06-2X/6-311++G(d,p) level of theory. A total of four substituents with different induction strengths were used in the study, namely: methyl ($-CH_3$), hydroxyl (-OH), formyl (-CHO) and nitro groups ($-NO_2$).

It was found that two factors result in the increase or decrease of the H-bond interaction energy. Firstly, the strongest H-bonds for each functional group were observed when a specific substituent is placed at a position that allows the rotation of the anionic functional group. The steric density or repulsion from substituents that force the functional group to rotate will result in an energy penalty that weakens the H-bond interaction energy. The second factor involves the formation of more than one H-bond, which results in the cooperative stabilisation of the H-bond and a corresponding increase in interaction energy. This stabilising effect is found when substituents are placed at positions 1 and 5 of the aromatic ring, but will only occur if the substituents also allow the functional group to rotate towards a minimum energy conformation that allows the formation of cooperative stabilised H-bonds. It was however found that each of the functional groups are affected differently by the substituents on the aromatic ring, which is possibly due to the difference in polarity and charge distribution of each group.

The strongest H-bond interaction energy for compounds that contain the carboxylate functional group was found to be -19.28 kcal/mol when the formyl substituent is placed at position 1 on the aromatic ring. This is, nevertheless, still 1.47 kcal/mol lower than the unsubstituted aromatic ring, owing to the planar arrangement of the unsubstituted carboxylate compound. The highest interaction energy for compounds that contain the nitrogen dioxide functional group was found to be -10.53 kcal/mol, which is more than double the interaction energy of the unsubstituted compound and is found when a nitro group is placed at position 1 on the aromatic ring. The large difference in interaction energy is due to the cooperative stabilisation of the H-bond in the substituted compound. The highest interaction energy for compounds that contain the sulfonate functional group was found to be -18.19 kcal/mol when the methyl group is placed at position 3 on the aromatic ring, which is 4.73 kcal/mol stronger than the unsubstituted compound. The higher interaction energy of the substituted aromatic ring is due to the small amount of rotation that is observed when the nitro group is placed at position 3. The highest interaction energy for compounds that contain the phosphonate functional group was found to be -34.65 kcal/mol when the hydroxyl group is placed at position 3, which is 5.26 kcal/mol stronger than the unsubstituted compound. The large difference in interaction energy is once again due to the cooperative stabilisation of the H-bond in the substituted compound.

The electron-donating or electron-withdrawing properties of six substituents were also analysed by AIM analysis. These substituents are methyl, hydroxyl, formyl, nitro, carboxyl, halide and carboxylic ester. Compound C7 was chosen as its symmetric structure reduced the computational resources required for the study. Geometry optimisations were performed at the MP2/cc-pVTZ level of theory for compounds that have been substituted at the *ortho*, *meta* or *para* positions. Following the generation of wavefunction files at the MP2/aug-cc-pVTZ level of theory single point calculations were performed for all optimised structures at the M06-2X/6-311++G(d,p) level of theory. AIM analysis of the wavefunctions files revealed that electron-donating groups that are placed at the *para* position yield stronger H-bonds, once again accompanied by cooperative stabilisation. Furthermore, electron-withdrawing groups with sufficient inductive effects can result in a weaker H-bond when placed at the *meta* position. It must however be mentioned that the changes in interaction

energies are not reflected by the change in ρ(r) at the BCPs, which suggests that the rotation of the functional group and cooperative stabilisation are the dominant effects that play a role in the strength of the H-bond.

We can conclude from these results that the substitution of aromatic compounds can have a significant effect on the H-bond interaction energy. Therefore, it could serve as a powerful method to fine-tune the strength of H-bond interactions and correspondingly increase the propensity of pharmaceutical compounds to form hydrates. Furthermore, the fine-tuning of the H-bond interaction could also improve or modify the properties of pharmaceutical compounds.

As described in Chapter 7, the effect of a_w on the crystal formation of creatine monohydrate, isonicotinic acid and trimesic acid was investigated in various water-organic solvent mixtures. Crystals obtained from a creatine-solvent mixture were analysed by XRD and found to be the creatine monohydrate. On attempting to obtain the anhydrous product by sublimation, XRD analysis of the resulting crystals showed that creatinine was obtained, which is the product of intramolecular cyclisation with the loss of an additional molecule of water. Crystals obtained from an isonicotinic acid-solvent mixture were analysed by XRD, which showed that the thermodynamically favoured product in the presence of water is still the isonicotinic acid compound. A CSD survey did however identify a substituted isonicotinic acid derivative that is hydrated, which confirms that the substituent effects described in Chapter 6 strengthen the H-bond in this compound and result in the thermodynamically favoured hydrate product. Lastly, it was shown that a_w affects the nucleation of trimesic acid dihydrate crystals, resulting in a variety of crystal shapes and sizes at different a_w values. This could serve as a possible technique for obtaining crystals of an appropriate size and shape as would be required for XRD, TGA, DSC and other analytical techniques.

8.2. Future work

Suggested future work entails performing the PES scans at a higher level of theory (increase of CCSD(T)/6-311++G(d,p) to the CCSD(T)/aug-cc-pVTZ level of theory). All rigid scans should also be repeated as relaxed scans to identify other minima and maxima. Furthermore, all of the geometry optimisations in Chapter 5, namely full geometry, crystalline environment and solvent models, should be repeated at the CCSD(T)/aug-cc-pVTZ level of theory, which has been described as the gold standard of computational methods. Calculations at the CCSD(T)/aug-cc-pVTZ level of theory will also serve as the benchmark for interaction energies.

The increase in H-bond interaction energy obtained by placing substituents on the aromatic ring can also be investigated by generating an electrostatic surface potential (ESPs), by using wave function and cube files (*.cube) that are generated from the corresponding optimised structures. The ESP could help explain the changes in interaction energies and also why electron-withdrawing and electron-donating groups have such a small effect on the interaction energy. Lastly, we wish to utilise the results in this study by applying them experimentally. This could be done by the addition of various substituents on the aromatic

ring by a synthetic procedure. We aim to use test two compounds, namely isonicotinic acid and benzoic acid. Both of these compounds are commonly available chemicals, and substituents may be added to the aromatic ring using various well-known synthetic procedures.

Due to the focus of Chapter 6 on aromatic compounds, we wish to extend the study of the effect of substituents to compounds that do not contain aromatic rings. This could be done by identifying a structure with a long carbon chain, upon which various substituents with increasing size and inductive strength could be added at different positions along the carbon chain. Geometry optimisations would be performed for the substituted compounds, whereafter H-bond interaction energies would be calculated. This study could show how the increasing steric density of large substituents affects the rotation of the functional anionic group and correspondingly the H-bond interaction energy.

Appendix A

Supplementary Information

Digital Supplementary Information

The attached CD contains all of the Gaussian output files that were generated in this study. The files were organised according to chapters. The naming of each file is explained below.

Chapter 4

Example 1: CREATH04_Scan_B3LYP_ 6-311++G(d,p)_Horizontal
Examlpe 2: CREATH04_OPT_MP2_ 6-311++G(d,p)_Horizontal

For the first example the first part of of each filename indicates the CSD refcode of the structure that was exported from Mercury as Cartesian coordinates to be used for computational calculations. This is then followed by the SCAN keyword, which shows that the PES scan was determined with this calculation. This was followed by the method and basis set. The last part of the filename describes which of the four scans was performed. It should be mentioned that all CCSD(T) scan files were compressed into one *zip file, which is due to the calculation of the full range was done by separating the calculations into different steps. The Scan keyword was replaced by the OPT keyword in the second example to indicated that these calculations were geometry optimisation calculations that were performed at the minimum conformation obtained from the PES curves.

Chapter 5

Example 1: BACDOY01_OPT_MP2_aug-cc-pVTZ.log
Example 2: BACDOY01_OPT_MP2_aug-cc-pVTZ_SCRF_Ethanol.log
Example 3: BACDOY01_OPT_MP2_aug-cc-pVTZ_CRYST.log

For the first example, the first part of the filename consists of the CSD refcode, which is followed by the OPT keyword. This suggests that a geometry optimisation calculation was performed. This was then followed by the method and basis set. For the second example, the last part of the filename consists of the SCRF keyword and the identity of the solvent, for example “SCRF_water” or “SCRF_ethanol”. The SCRF keyword implies that this calculation was performed in the solvent model. If this keyword is not included in the filename it is implied that the calculation was performed in the gas phase. The third example differs from the first example by one additional keyword, namely CRYST. This shows that crystalline environment geometry optimisations were performed.

Chapter 6

Example 1: LECFIJ_OPT_MP2_aug-cc-pVTZ_OH_Pos1.log
Example 2: ISIFUM_OPT_MP2_aug-cc-pVTZ_NH2_ortho.log
Example 3: ISIFUM_WFX_MP2_aug-cc-pVTZ_NH2_ortho.wfx

For the first example, the naming of these files are similar to what is described above for Chapter 5. The last two keywords in the example show that the hydroxyl substituent has been used and it was placed at position 1 of the aromatic ring. The Ph keyword indicates that the calculation was performed for the unsubstituted compound. For the second example the last keyword differs from the first example in that it describes ortho, meta or para position on

the aromatic ring. For the third example the OPT keyword has been replaced by the WFX keyword and it contains the *wfx suffix to indicate that the wave function is included in the file.

The following Gausian output files are included in on the USB flash drive in their respective chapter folders:

Chapter_4

PES Scans

Horizontal

CREATH04_SCAN_HF_6-311++G(d,p)_Horizontal.log
CREATH04_SCAN_M062X_6-311++G(d,p)_Horizontal.log
CREATH04_SCAN_B3LYP_6-311++G(d,p)_Horizontal.log
CREATH04_ SCAN_B97D_6-311++G(d,p)_Horizontal.log
CREATH04_SCAN_B3LYPD3 _6-311++G(d,p)_Horizontal.log
CREATH04_ SCAN_MP2_6-311++G(d,p)_Horizontal.log
CREATH04_ SCAN_CCSD(T)_6-311++G(d,p)_Horizontal.log

Vertical

CREATH04_SCAN _HF _6-311++G(d,p)_Vertical.log
CREATH04_SCAN _M062X _6-311++G(d,p)_Vertical.log
CREATH04_SCAN B3LYP _6-311++G(d,p)_Vertical.log
CREATH04_SCAN_B97D_6-311++G(d,p)_Vertical.log
CREATH04_ SCAN_B3LYPD3 _6-311++G(d,p)_Vertical.log
CREATH04_ SCAN_MP2 _6-311++G(d,p)_Vertical.log
CREATH04_ SCAN_CCSD(T)_6-311++G(d,p)_Vertical.log

LTR_Diagonal

CREATH04_SCAN_HF_6-311++G(d,p)_LTR_Diagonal.log
CREATH04_SCAN_M062X_6-311++G(d,p)_LTR_Diagonal.log
CREATH04_SCAN_B3LYP_6-311++G(d,p)_LTR_Diagonal.log
CREATH04_SCAN_B97D_6-311++G(d,p)_LTR_Diagonal.log
CREATH04_SCAN_B3LYPD3_6-311++G(d,p)_LTR_Diagonal.log
CREATH04_SCAN_MP2 _6-311++G(d,p)_LTR_Diagonal.log
CREATH04_SCAN_CCSD(T) _6-311++G(d,p)_LTR_Diagonal.log

RTL_Diagonal

CREATH04_SCAN_HF_6-311++G(d,p)_RTL_Diagonal.log
CREATH04_SCAN_M062X_6-311++G(d,p)_RTL_Diagonal.log
CREATH04_SCAN_B3LYP_6-311++G(d,p)_RTL_Diagonal.log
CREATH04_SCAN_B97D_6-311++G(d,p)_RTL_Diagonal.log
CREATH04_SCAN_B3LYPD3_6-311++G(d,p)_RTL_Diagonal.log
CREATH04_SCAN_MP2_6-311++G(d,p)_RTL_Diagonal.log
CREATH04_SCAN_CCSD(T)_6-311++G(d,p)_RTL_Diagonal.log

Scan_OPT

CREATH04_ OPT_MP2_ 6-311++G(d,p)_Horizontal.log
CREATH04_ OPT_MP2_ 6-311++G(d,p)_ Vertical.log
CREATH04_ OPT_MP2_ 6-311++G(d,p)_ LTR_Diagonal.log
CREATH04_ OPT_MP2_ 6-311++G(d,p)_ RTL_Diagonal.log

Chapter_5

OPT

ISIFUM_OPT_cc-pVTZ.log
IWIRAH_OPT_aug-cc-pVTZ.log
WOBXIU_OPT_cc-pVTZ.log
LAPFAL_OPT_aug-cc-pVTZ.log
JODCAH_OPT_cc-pVTZ.log
ACOHAC_OPT_aug-cc-pVTZ.log
CACGUI01_OPT_aug-cc-pVTZ.log
AMHPYR_OPT_cc-pVTZ.log
LIPWEM02_OPT_aug-cc-pVTZ.log
COWYAP_OPT_cc-pVTZ.log
IVUKUG_OPT_aug-cc-pVTZ.log
XECBEO_OPT_aug-cc-pVTZ.log
BONKOE0_OPT_aug-cc-pVTZ.log
ABERAB_OPT_aug-cc-pVTZ.log
NOFXEL_OPT_aug-cc-pVTZ.log
ROFNUW_OPT_aug-cc-pVTZ.log
TIXGEM_OPT_cc-pVTZ.log
EVOJUV_OPT_aug-cc-pVTZ.log
WEFBIU_OPT_aug-cc-pVTZ.log
YAJPAB_OPT_aug-cc-pVTZ.log
BACDOY_OPT_aug-cc-pVTZ.log
FEBZER_OPT_cc-pVTZ.log
KIBLOW_OPT_aug-cc-pVTZ.log
ZIVBIP_OPT_aug-cc-pVTZ.log
OKUWIB_OPT_cc-pVTZ.log

SCRF_1-Propanol

LECFIJ_OPT_MP2_cc-pVTZ_SCRF_1-Proponol.log
XECBEO_OPT_MP2_cc-pVTZ_SCRF_1-Proponol.log
AMHPYR_OPT_MP2_cc-pVTZ_SCRF_1-Proponol.log
BACDOY_OPT_MP2_cc-pVTZ_SCRF_1-Proponol.log
ROFNUW_OPT_MP2_cc-pVTZ_SCRF_1-Proponol.log

SCRF_Heptane

LECFIJ_OPT_MP2_cc-pVTZ_SCRF_Heptane.log
XECBEO_OPT_MP2_cc-pVTZ_SCRF_Heptane.log
AMHPYR_OPT_MP2_cc-pVTZ_SCRF_Heptane.log
BACDOY_OPT_MP2_cc-pVTZ_SCRF_Heptane.log
ROFNUW_OPT_MP2_cc-pVTZ_SCRF_Heptane.log

SCRF_ DMSO

LECFIJ_OPT_MP2_cc-pVTZ_SCRF_DMSO.log
XECBEO_OPT_MP2_cc-pVTZ_SCRF_DMSO.log
AMHPYR_OPT_MP2_cc-pVTZ_SCRF_DMSO.log
BACDOY_OPT_MP2_cc-pVTZ_SCRF_DMSO.log
ROFNUW_OPT_MP2_cc-pVTZ_SCRF_DMSO.log

SCRF_n-Octanol

LECFIJ_OPT_MP2_cc-pVTZ_SCRF_n-Octanol.log
XECBEO_OPT_MP2_cc-pVTZ_SCRF_n-Octanol.log
AMHPYR_OPT_MP2_cc-pVTZ_SCRF_n-Octanol.log
BACDOY_OPT_MP2_cc-pVTZ_SCRF_n-Octanol.log
ROFNUW_OPT_MP2_cc-pVTZ_SCRF_n-Octanol.log

SCRF_Water

LECFIJ_OPT_MP2_cc-pVTZ_SCRF_Water.log
XECBEO_OPT_MP2_cc-pVTZ_SCRF_Water.log
AMHPYR_OPT_MP2_cc-pVTZ_SCRF_Water.log
BACDOY_OPT_MP2_cc-pVTZ_SCRF_Water.log
ROFNUW_OPT_MP2_cc-pVTZ_SCRF_Water.log

CRYST

ISIFUM_OPT_6-311++G(d,p)_CRYST.log
IWIRAH_OPT_6-311++G(d,p)_CRYST.log
WOBXIU_OPT_6-311++G(d,p)_CRYST.log
LAPFAL_OPT_6-311++G(d,p)_CRYST.log
JODCAH_OPT_6-311++G(d,p)_CRYST.log
ACOHAC_OPT_6-311++G(d,p)_CRYST.log
CACGUI01_OPT_6-311++G(d,p)_CRYST.log
AMHPYR_OPT_6-311++G(d,p)_CRYST.log
LIPWEM02_OPT_6-311++G(d,p)_CRYST.log
COWYAP_OPT_6-311++G(d,p)_CRYST.log
IVUKUG_OPT_6-311++G(d,p)_CRYST.log
XECBEO_OPT_6-311++G(d,p)_CRYST.log
BONKOE0_OPT_6-311++G(d,p)_CRYST.log
ABERAB_OPT_6-311++G(d,p)_CRYST.log
NOFXEL_OPT_6-311++G(d,p)_CRYST.log
ROFNUW_OPT_6-311++G(d,p)_CRYST.log
TIXGEM_OPT_6-311++G(d,p)_CRYST.log
EVOJUV_OPT_6-311++G(d,p)_CRYST.log
WEFBIU_OPT_6-311++G(d,p)_CRYST.log
YAJPAB_OPT_6-311++G(d,p)_CRYST.log
BACDOY_OPT_6-311++G(d,p)_CRYST.log
FEBZER_OPT_6-311++G(d,p)_CRYST.log
KIBLOW_OPT_6-311++G(d,p)_CRYST.log
ZIVBIP_OPT_6-311++G(d,p)_CRYST.log
OKUWIB_OPT_6-311++G(d,p)_CRYST.log

Chapter_6

Subs

Ph

BACDOY_OPT_MP2-6-311++G(d,p)_Ph.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_Ph.log
LECFIJ_OPT_MP2_6-311++G(d,p)_Ph.log
XECBEO_OPT_MP2-6-311++G(d,p)_Ph.log

CH3

BACDOY_OPT_MP2-6-311++G(d,p)_CH3_Pos1.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CH3_Pos1.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CH3_Pos1.log
XECBEO_OPT_MP2-6-311++G(d,p)_CH3_Pos1.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CH3_Pos1.log
BACDOY_OPT_MP2-6-311++G(d,p)_CH3_Pos2.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CH3_Pos2.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CH3_Pos2.log
XECBEO_OPT_MP2-6-311++G(d,p)_CH3_Pos2.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CH3_Pos2.log
BACDOY_OPT_MP2-6-311++G(d,p)_CH3_Pos3.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CH3_Pos3.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CH3_Pos3.log
XECBEO_OPT_MP2-6-311++G(d,p)_CH3_Pos3.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CH3_Pos3.log
BACDOY_OPT_MP2-6-311++G(d,p)_CH3_Pos4.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CH3_Pos4.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CH3_Pos4.log
XECBEO_OPT_MP2-6-311++G(d,p)_CH3_Pos4.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CH3_Pos4.log
BACDOY_OPT_MP2-6-311++G(d,p)_CH3_Pos5.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CH3_Pos5.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CH3_Pos5.log
XECBEO_OPT_MP2-6-311++G(d,p)_CH3_Pos5.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CH3_Pos5.log

OH

BACDOY_OPT_MP2-6-311++G(d,p)_OH_Pos1.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_OH_Pos1.log
LECFIJ_OPT_MP2_6-311++G(d,p)_OH_Pos1.log
XECBEO_OPT_MP2-6-311++G(d,p)_OH_Pos1.log
AMHPYR_OPT_MP2-6-311++G(d,p)_OH_Pos1.log
BACDOY_OPT_MP2-6-311++G(d,p)_OH_Pos2.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_OH_Pos2.log
LECFIJ_OPT_MP2_6-311++G(d,p)_OH_Pos2.log
XECBEO_OPT_MP2-6-311++G(d,p)_OH_Pos2.log
AMHPYR_OPT_MP2-6-311++G(d,p)_OH_Pos2.log
BACDOY_OPT_MP2-6-311++G(d,p)_OH_Pos3.log

BBZSUL01_OPT_MP2-6-311++G(d,p)_OH_Pos3.log
LECFIJ_OPT_MP2_6-311++G(d,p)_OH_Pos3.log
XECBEO_OPT_MP2-6-311++G(d,p)_OH_Pos3.log
AMHPYR_OPT_MP2-6-311++G(d,p)_OH_Pos3.log
BACDOY_OPT_MP2-6-311++G(d,p)_OH_Pos4.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_OH_Pos4.log
LECFIJ_OPT_MP2_6-311++G(d,p)_OH_Pos4.log
XECBEO_OPT_MP2-6-311++G(d,p)_OH_Pos4.log
AMHPYR_OPT_MP2-6-311++G(d,p)_OH_Pos4.log
BACDOY_OPT_MP2-6-311++G(d,p)_OH_Pos5.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_OH_Pos5.log
LECFIJ_OPT_MP2_6-311++G(d,p)_OH_Pos5.log
XECBEO_OPT_MP2-6-311++G(d,p)_OH_Pos5.log
AMHPYR_OPT_MP2-6-311++G(d,p)_OH_Pos5.log

CHO

BACDOY_OPT_MP2-6-311++G(d,p)_CHO_Pos1.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CHO_Pos1.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CHO_Pos1.log
XECBEO_OPT_MP2-6-311++G(d,p)_CHO_Pos1.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CHO_Pos1.log
BACDOY_OPT_MP2-6-311++G(d,p)_CHO_Pos2.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CHO_Pos2.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CHO_Pos2.log
XECBEO_OPT_MP2-6-311++G(d,p)_CHO_Pos2.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CHO_Pos2.log
BACDOY_OPT_MP2-6-311++G(d,p)_CHO_Pos3.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CHO_Pos3.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CHO_Pos3.log
XECBEO_OPT_MP2-6-311++G(d,p)_CHO_Pos3.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CHO_Pos3.log
BACDOY_OPT_MP2-6-311++G(d,p)_CHO_Pos4.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CHO_Pos4.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CHO_Pos4.log
XECBEO_OPT_MP2-6-311++G(d,p)_CHO_Pos4.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CHO_Pos4.log
BACDOY_OPT_MP2-6-311++G(d,p)_CHO_Pos5.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_CHO_Pos5.log
LECFIJ_OPT_MP2_6-311++G(d,p)_CHO_Pos5.log
XECBEO_OPT_MP2-6-311++G(d,p)_CHO_Pos5.log
AMHPYR_OPT_MP2-6-311++G(d,p)_CHO_Pos5.log

NOO

BACDOY_OPT_MP2-6-311++G(d,p)_NOO_Pos1.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_NOO_Pos1.log
LECFIJ_OPT_MP2_6-311++G(d,p)_NOO_Pos1.log
XECBEO_OPT_MP2-6-311++G(d,p)_NOO_Pos1.log

AMHPYR_OPT_MP2-6-311++G(d,p)_NOO_Pos1.log
BACDOY_OPT_MP2-6-311++G(d,p)_NOO_Pos2.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_NOO_Pos2.log
LECFIJ_OPT_MP2_6-311++G(d,p)_NOO_Pos2.log
XECBEO_OPT_MP2-6-311++G(d,p)_NOO_Pos2.log
AMHPYR_OPT_MP2-6-311++G(d,p)_NOO_Pos2.log
BACDOY_OPT_MP2-6-311++G(d,p)_NOO_Pos3.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_NOO_Pos3.log
LECFIJ_OPT_MP2_6-311++G(d,p)_NOO_Pos3.log
XECBEO_OPT_MP2-6-311++G(d,p)_NOO_Pos3.log
AMHPYR_OPT_MP2-6-311++G(d,p)_NOO_Pos3.log
BACDOY_OPT_MP2-6-311++G(d,p)_NOO_Pos4.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_NOO_Pos4.log
LECFIJ_OPT_MP2_6-311++G(d,p)_NOO_Pos4.log
XECBEO_OPT_MP2-6-311++G(d,p)_NOO_Pos4.log
AMHPYR_OPT_MP2-6-311++G(d,p)_NOO_Pos4.log
BACDOY_OPT_MP2-6-311++G(d,p)_NOO_Pos5.log
BBZSUL01_OPT_MP2-6-311++G(d,p)_NOO_Pos5.log
LECFIJ_OPT_MP2_6-311++G(d,p)_NOO_Pos5.log
XECBEO_OPT_MP2-6-311++G(d,p)_NOO_Pos5.log
AMHPYR_OPT_MP2-6-311++G(d,p)_NOO_Pos5.log

AIM

ISIFUM_OPT_MP2- aug-cc-pVTZ _NH2_para.log
ISIFUM_OPT_MP2- aug-cc-pVTZ _NH2_ortho.log
ISIFUM_OPT_MP2-aug-cc-pVTZ _OH_para.log
ISIFUM_OPT_MP2-aug-cc-pVTZ _OH_ortho.log
ISIFUM_OPT_MP2-cc-pVTZ _OOCCH3_para.log
ISIFUM_OPT_MP2-aug-cc-pVTZ _OOCCH3_ortho.log
ISIFUM_OPT_MP2-cc-pVTZ _COOH_meta.log
ISIFUM_OPT_MP2-aug-cc-pVTZ _NOO_meta.log
ISIFUM_OPT_MP2-aug-cc-pVTZ _Cl_meta.log

WFX

ISIFUM_WFX_MP2-aug-cc-pVTZ _NH2_para.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _NH2_ortho.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _OH_para.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _OH_ortho.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _OOCCH3_para.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _OOCCH3_ortho.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _COOH_meta.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _NOO_meta.wfx
ISIFUM_WFX_MP2-aug-cc-pVTZ _Cl_meta.wfx