University of Strathclyde Department of Pure and Applied Chemistry

Salt Selection for Pharmaceutical Use

By Catriona A. Morrison

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy in the Faculty of Science of the University of Strathclyde

February 2012

Abstract

This work details the creation of a library of 255 single crystal structures of systematically related salt forms of phenylethylamines, collated together with phase specific physicochemical property data on melting point and aqueous solubility. This library has been used to investigate a number of structure-property relationships relevant to salt selection in the pharmaceutical industry.

Sixty-four salt forms of enantiopure and racemic methylephedrine were subject to detailed structural analysis, including identification of common graph-sets and isostructural groups with respect to cation packing. This showed the reoccurring presence of a common $C_2^2(9)$ graph-set which encompassed both OCO and OSO anion functionalities. PIXEL energy calculations were utilised to determine the strongest cation-cation and cation-anion interactions present within the salts. Close inspection of the densities showed no evidence to support Wallach's rule.

The library of 255 salt structures was analysed for trends in hydrate formation. It was found that hydrate formation was most likely where there was increased presence of polar groups and especially when there was an excess of hydrogen bond acceptor atoms over donor atoms.

For anhydrous salts, correlation studies showed linear trends between salt melting point and log aqueous solubility and between melting point of the salt and melting point of the parent free acid. The exact nature of the correlations varied and each was specific to certain chemical groups. Random forest analysis was used to build both regression and classification training models using 37 well characterised anhydrous methylephedrinium salts. These models successfully encompassed a wide variety of anion types. Prediction of aqueous solubility using the methylephedrinium salt regression training model was successful for a series of benzoate derived salts and for a series of hydrated methylephedrinium salts, the latter with the application of a simple correction constant.

Acknowledgements

Table of Contents

Table of Figures

xi

Table of Tables

Chapter 1
Introduction

1.1 Desirable properties of pharmaceutical drugs

Pharmaceutical drugs are designed to impart a desired effect within the human body. However, it is only on very rare occasions that any individual drug substance can be used in its chemically pure form to obtain the necessary therapeutic effect. Instead, the active pharmaceutical ingredient, API, is formulated into an appropriate form for effective pharmaceutical use.[1] Drugs are formulated as many different forms, such as injectable solutions, creams, inhalants, sprays, powders, capsules and tablets. Which form is chosen depends on the required application and the API's own characteristics.

The majority of modern drugs are dispensed in the solid form, namely pills and capsules, which allows for oral administration.[2] The transformation from the original drug molecule through to the final formulated pill or capsule that is delivered to market must be both technically feasible and economically viable for full scale production. There are currently only about 10000 drug-like compounds available to the pharmaceutical industry. A drug-like compound is defined as one which has sufficient absorption, distribution, metabolism and excretion (ADME) properties, along with low chemical reactivity-related toxicity properties within the human body, to pass Phase 1 clinical trials.[3] (Phase 1 clinical trials are intended to estimate the maximum tolerable dose in healthy humans.[4]) As an aid to designing suitable drug-like compounds, the 'rule of 5' was established by Lipinski et al.[5] This helps to predict when a drug molecule that is orally bioavailable will have poor solubility and membrane permeability. The rule predicts that a compound will have poor absorption or permeation when;

- The compound's molecular weight is over 500
- The compound contains more than 5 hydrogen bond donors
- The compound contains more than 10 hydrogen bond acceptors
- The compounds LogP is over 5. LogP is a measure of the partition ratio of a compound between two immiscible solvents, normally water and octanol.

The solid-state structure of the drug affects many of its physical properties. During large scale production important physicochemical properties to be considered include melting point, solubility, dissolution rate, stability, particle size and shape, and powder consistency.[6] During the manufacturing process not only must the API itself be chosen to have the desired ADME properties, but the correct formulation and 'packaging' of the drug must also be obtained in order to acquire the desired physicochemical properties.[7] Figure 1.1, shows the pharmacokinetic process for an oral formulated drug within the body.

```
                  ┌─────────────────────────┐
                  │  Administration of drug │
                  └─────────────────────────┘
                              ⇓
                  ┌─────────────────────────┐
                  │ Release of drug substance│
                  └─────────────────────────┘
                              ⇓
                      ┌──────────────┐
                      │  Absorption  │
                      └──────────────┘
                              ⇓
                   ┌──────────────┐        ┌──────────────┐
                   │ Distribution │⇔⇒      │ Site of action│
                   └──────────────┘        └──────────────┘
              ⇓          ⇓
    ┌──────────────┐ ┌──────────────────┐
    │  First-pass  │ │ Biotransformation │
    │  metabolism  │ └──────────────────┘
    └──────────────┘          ⇓
              ⇒       ┌──────────────┐
                      │  Elimination │
                      └──────────────┘
```
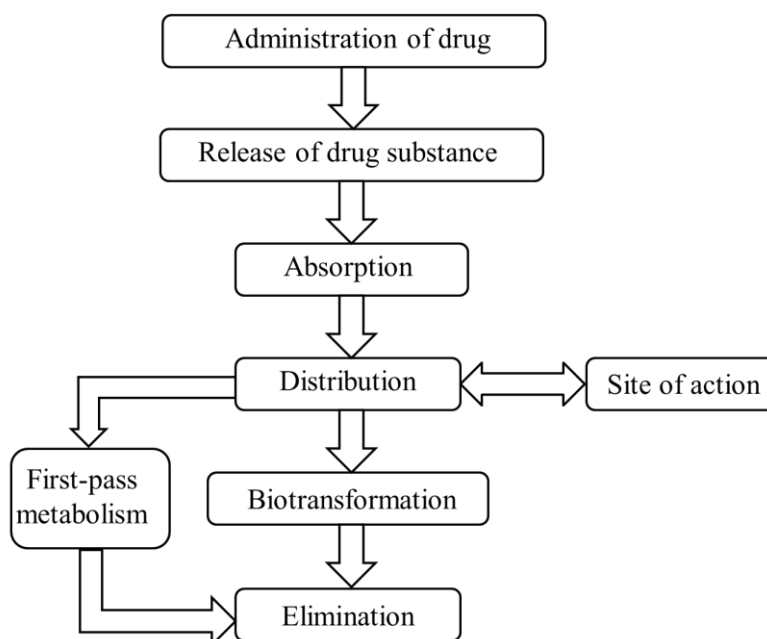
Figure 1.1 Principle steps of pharmacokinetics for a given pharmaceutical drug for oral administration, adapted from reference 1[1]

The chosen drug candidate must have a high enough melting point to withstand the rigorous manufacturing process. It may be stressed under high pressure and raised temperature throughout this process, but especially in the pressing stage to form the final pellet or pill. The thermal stability, solution stability (at various pH values), and the light sensitivity of the drug are also all important factors that are to be considered. Light can cause oxidation, hydrolysis and loss of potency to some sensitive medications. The chemical and physical stability of any new drug must be

known, as this affects the manufacturing and packaging process as well as the shelf life of the retail product.[8,9]

A drug's particle size must be uniform as this influences the mixing of the powder during manufacturing. A homogenously mixed powder is necessary, as uneven distribution of particle size can influence the bioavailability of the API, the strength of the resulting tablet and the flowability of the powder. The desired powder must flow smoothly and be easily transferred through the machinery, and also have the correct constituency so that it can be pressed and remain as tablets without crumbling. A classic example of where this problem exists is with the common drug paracetamol. Paracetamol is a hard material and hence difficult to tablet. It also has a very low overdose limit and so ensuring an accurate dose per tablet is crucial. This has led to extensive work on investigating alternative solid-state phases to try and combat this problem.[10] The presence of water in the solid drug can also be detrimental to bulk manufacturing. The stability of the drug, either during manufacturing or after tableting can be critically affected by the absorption or desorption of water. This could cause the product to decompose physically or chemically and thus severely reduce its shelf life. [7]

The solubility of a drug must be easily manipulated to acquire the desired pharmaceutical effects. In recent years this area has become of particular interest in the medicinal chemistry sectors. Due to the introduction of combinatorial chemistry and high-throughput screening to discovery new chemical entities, NCE, there has been a sharp decrease in the solubility of new drug molecules. Until the nineteen eighties the solubility of NCEs would rarely be below 20 µg/mL, whereas in comparison nowadays solubility values of less than 1 µg/mL are common.[11] This is an extremely important issue that needs to be addressed as a drug with poor solubility is likely to have a poor ADME profile.

Another important factor to be considered is the dissolution rate. Dissolution is the process by which solid compounds dissolve in a liquid medium. The kinetic descriptor used to enumerate the rate at which this happens, is known as the

dissolution rate.[1] During the formulation process, this property is typically of utmost importance. High solubility is directly linked to high dissolution rate and both are seen to be linked to the bioavailability of the compound. [8]

1.2   Drugs as salts

For new drugs one of the simplest ways to optimise and manipulate the material properties, hopefully without fundamentally altering the pharmacology, is to change a free acid or base into an available salt form of that drug. [8]

According to the Bronsted Lowry definition, acids are compounds that can donate protons and bases are compounds that can accept protons.[12] A salt is an ionic compound whose cation comes from a base and whose anion comes from an acid. As many chemical entities contain either an acidic or basic functionality, or indeed both, somewhere within their structure, this should in theory allow a salt to be formed.

Nowadays, it is estimated that around half of all drug molecules that are used for medicinal therapy are administrated as salts;[13] therefore the salification process of the drug substance has become a vital step in drug development. The drug's solid-state properties as well as its properties in solution can be altered and manipulated by salt formation. For pharmaceutical companies salt selection can be the most cost effective way of improving the potential drug properties. The characteristics of the salt can be modified without altering the drug's chemical structure by changing the counter ion and converting the drug to a new salt form. Each counter ion imparts unique properties to the parent compound. Therefore, the search for the most suitable salt form is of high importance, and salt selection may open up many new opportunities and have significant consequences.[1] Table 1.1, illustrates some of the common pharmaceutical anions that are used to synthesis salts of basic drugs.

Table 1.1 Classification of pharmaceutical salts of basic drugs, table adapted from reference 14[14]

Salt class	Examples
Inorganic acids	Hydrochloride, hydrobromide, sulphate, nitrate, phosphate
Sulfonic acids	Mesylate, esylate, isethionate, tosylate, napsylate, besylate
Carboxylic acids	Acetate, propionate, maleate, benzoate, salicylate, fumarate
Anionic amino acids	Glutamate, aspartate
Hydroxyacids	Citrate, lactate, succinate, tartrate, glycollate
Fatty acids	Hexonate, octanoate, decanoate, oleate, stearate
Insoluble salts	Pamoate (embonate), polystyrene sulfonate (resinate)

1.3 Solubility and dissolution rate of salts

Salt formation is generally the first approach to be considered as a way of increasing a drug's aqueous solubility.[15] The aqueous solubility must be measured, as this simulates the environment the drug will encounter when administrated orally into the gastro intestinal tract. Solubility affects the overall drug absorption process, the stability of the chemical and ultimately whether the drug will be effective or not. It is also an essential and defining factor in controlling the dissolution rate of the drug.[16]

Information on the aqueous solubility of an acidic or basic drug as a function of pH is an indication as to whether or not the compound will form suitable salts, and when salts are formed what some of the physiochemical properties might be.[1] Variable pH solubility measurements of the drug can dictate what counter ions are necessary to form a desirable salt. They can predict how easily the salt will dissociate back into its free acid or free base form and what the dissolution behaviour will be under different gastro-intestinal conditions.[11]

When considering the formation of a salt from a basic drug molecule, the pH-solubility interrelationship profile of the free base and its salt indicates how the solubility varies greatly with pH when the salt form and the free base are in equilibrium with each other in a saturated solution. When a soluble acid salt of a

weakly basic drug is added to solution this will cause the pH to drop. This drop in pH will then cause more of the drug to dissolve until a point when the pH of maximum solubility is reached.[17] When a basic compound or its salt is dissolved in water, the following equilibrium exists, where BH^+ and B are the protonated and free base forms of the compound respectively.

$$BH^+ + H_2O \xrightleftharpoons{K_a} B + H_3O^+ \qquad \text{(Eqn 1)}$$
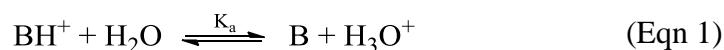
$$K_a = \frac{[B][H_3O^+]}{[BH^+]} \qquad \text{(Eqn 2)}$$
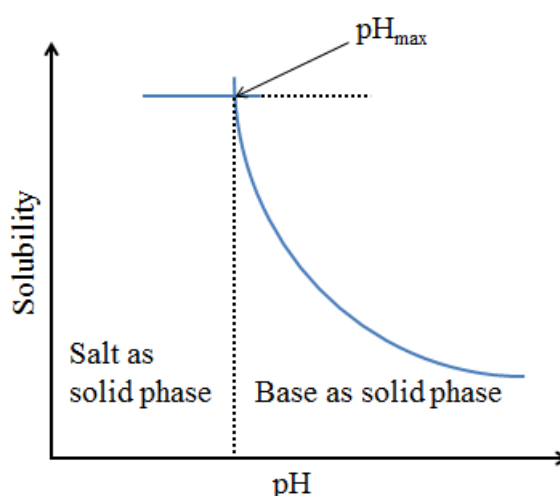
Figure 1.2 pH solubility profile of a basic drug, adapted from reference 18[18]

Figure 1.2, indicates how the solubility profile can by represented by two curves, one where the base is in the excess solid phase and the other where the excess solid phase is comprised of the salt. The point at which these two curves intersect is pH_{max}, the pH of maximum solubility. At any pH value above pH_{max} the free base is the equilibrium species, which will only convert to the salt when it is equilibrated with a solution at a pH below pH_{max}. The equilibrium can be shifted towards and beyond pH_{max} by the addition of an acid or a counter ion, therefore producing the salt in the solid phase.

The two equations that represent the conditions on either side of pH_{max} are given below, where S_T is the total solubility at that pH, BH^+ and B represent the protonated

(salt) and free base form of the compound respectively and the subscript 's' represents the saturation species

$$S_{T, pH<pHmax} = [BH^+]_S + [B] = [BH^+]_S\left(1 + \frac{K_a}{[H_3O]^+}\right) \qquad \text{(Eqn 3)}$$

$$S_{T, pH>pHmax} = [BH^+] + [B]_s = [B]_s\left(1 + \frac{[H_3O]^+}{K_a}\right) \qquad \text{(Eqn 4)}$$

The two independent curves can be obtained by varying the hydrogen ion concentration or the pH. The free base and salt can only both coexist as solids at $pH_{max.}$ This is of special relevance to this project as if the solid phase in a slurry experiment can be shown to be pure salt form, then the measured concentration at that point will be that of the salt - and this concentration will be relatively invariant with respect to changing pH.

The above Figure 1.2 is only a general profile, but many literature studies have proposed that it is a good description for the solubilities of free bases and their salts.[19-22] In all these cases, the salt forms had a higher aqueous solubility than the corresponding free base, but different salt forms of a specific base varied in their solubility, therefore choosing the correct counter ion is of great importance.

Increasing the solubility of a drug increases the dissolution rate but although the two properties are linked there is a distinct difference between them. Dissolution is the process by which the solid dissolves in a liquid, and the rate at which this occurs is called the dissolution rate. It has been proposed that this property best shows the bioavailability of a drug.[11] Dissolution and solubility are linked by the diffusion layer model. The model assumes that when introduced to solvent the outermost layer of the drug dissolves instantly into a thin film to from a saturated solution around the drug, known as the static layer. The transfer of the drug throughout the solvent occurs by diffusion of the drug molecule through this static layer. The thickness of the static layer is dependent on the stirring rate of the solution. Figure 1.3, illustrates the dissolution process across the static layer, which can be thought of three distinct

steps; 1. The molecule leaves the solid, 2. The molecule diffuses across the static layer and 3. The molecule leaves the static layer and enters the well stirred bulk solution.[23]
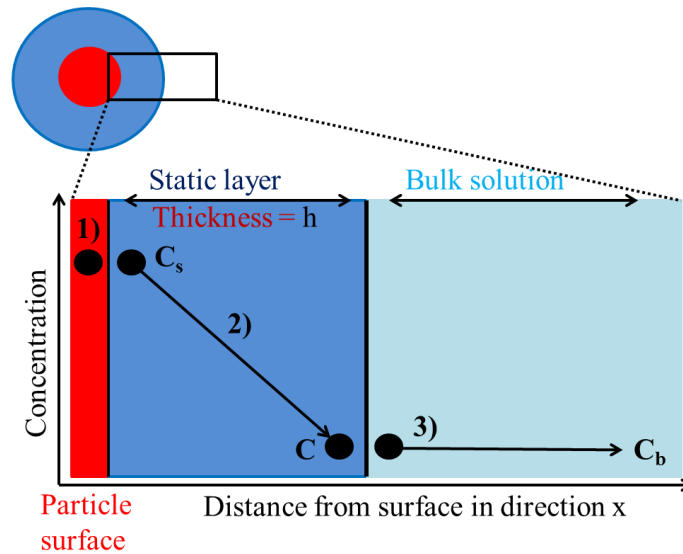
Figure 1.3 Dissolution process across the static layer

The dissolution rate (J) of solid per unit surface area is given by the equation;

$$J = \frac{D}{h}\,(C_s - C_b) \qquad \text{(Eqn 5)}$$

where D is the diffusion coefficient of the solute, h is the diffusion layer thickness during dissolution at the surface of the solid, C_s is the saturated solubility of the solid in the dissolution medium and C_b is its concentration in the bulk medium. Note that as the diffusion process is the rate determining step here, the concentration at the particle surface is the saturated concentration. The equation shows that increasing $(C_s\text{-}C_b)$ increases the rate of dissolution, and this is why substances with high maximum concentrations tend to give higher dissolution rates.

Salts have the ability to speed up their dissolution rate by effectively acting as their own buffer and in doing so they can alter and control the pH of the diffusion layer. Buffering the pH of the diffusion layer controls and increases the solubility through the layer and in doing so increases the dissolution rate. In the human body, the

dissolution rate of a compound is used to determine the rate of build-up of drug levels with time and also as a guide to the maximum level that is obtained.[8]

1.4 Alternative ways to alter physicochemical properties

Changing the free acid or free base into the salt form of the drug is normally the preferred way of altering the physicochemical properties. However, other options do exist. You can change the drug molecule itself, or attempt to use a specific polymorph, co-crystal, amorphous solid or hydrated or solvated form of the drug. Each form has the ability to influence the bioavailability, manufacturability, stability and other properties of the drug.[2] Figure 1.4, illustrates the various types of solid forms available to the pharmaceutical company. Table 1.2, outlines the potential range of the different solid forms and the relevance of their composition. Additionally to add further complication to the study as well as experiencing different polymorphic forms of the API by itself, polymorphs can also exist of the salt, solvate and co-crystal forms.[24]
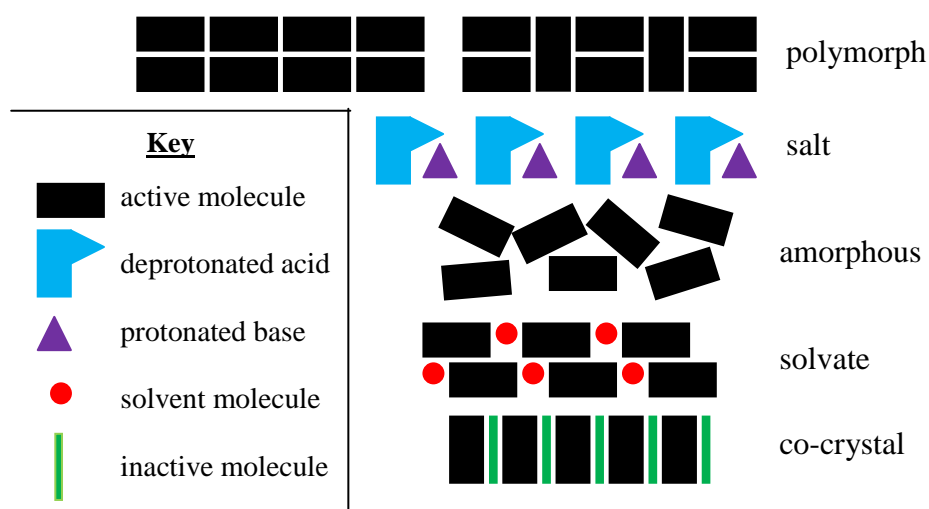
Figure 1.4 Schematic depictions of various types of solid forms, adapted from reference 25 [25]

Table 1.2 Outline of potential solid forms and their compositions, adapted from reference 26[26]

Type	Description	Composition
Crystalline solid	Solid consisting of highly ordered three dimensional molecular packing	API
Polymorph	Chemically identical crystalline form where the molecules adopt a varied packing arrangement	API
Crystalline solvate	Crystalline solid that contains solvent molecules within the crystal lattice	API + solvent
Crystalline salt	Ionic crystalline solid which contains an ionised API and a counterion	Ionised API + counterion
Co-crystal	Crystalline solid comprising of API and additional chemical entity.	API + co-crystal former
Amorphous	Non crystalline solid that contains no long range order (no crystal lattice)	API

1.4.1 Changing the drug molecule

When a new drug is required, the drug discovery process involves the screening of a vast number of compounds whose synthesis may have taken place over many years by the company's research department. Also, libraries of thousands of compounds that have been synthesised and organised into structurally related series by combinational chemistry techniques might be screened. From these databases a short list of possible drug candidates is constructed. From this list the required drug molecule is then chosen and developed.[14] If however, somewhere further down the development track it is discovered that the drug does not exhibit desirable physicochemical properties it may be possible to retreat back and chose an alternative starting drug molecule from the original short list. This might not be practical as it could require the full redevelopment of a new compound, which would not be economically favourable for the company, and would require a lot of pharmaceutical retesting. There is also no guarantee that the newly developed drug molecule will possess the desired physicochemical properties.[27]

1.4.2 Polymorphism

A polymorphic material is one in which a given chemical can crystallise with two or more distinct crystal structures.[19] Within the different crystal structures different

intermolecular interactions, such as hydrogen bonds and van der Waals interactions, will be present. The different polymorphs will therefore have different free energies and different physical properties such as solubility and chemical stability. Therefore, by choosing a different polymorph of the drug molecule different characteristics can be achieved.[28] Experimentally, the number of polymorphs generally seen for any given drug molecule rarely exceed five, but as the following statement made by McCrone in 1963 suggests we can never be sure that all the possible polymorphs have been accounted for.

> "virtually all compounds are polymorphic and the number of polymorphs of a material depends on the amount of time and money [spent] in research of that compound"[29]

This statement implies that if you research a compound for years using various crystallisation conditions you are very likely to find several more polymorphs, than if you perform a standard polymorph screen that will take a lot less time and money. This is a very worrying statement especially in industry where incomplete understanding of a material's polymorphism can cost money. Many polymorphic systems undergo concomitant crystallisation, where two or more different polymorphs crystallise from the same solution. This is a very undesirable and costly outcome for pharmaceutical companies, as the resulting properties of the drug will then be compromised and may cause the material to fail to meet specifications.[30]

One of the most famous cases of the commercially detrimental effect of polymorphism is that of the Abbott HIV treatment Ritonavir. First discovered in 1992, Ritonavir went into commercial production in 1996 as an ethanol/water solution or as a semi-solid, as the solid form was not bioavailable. Over the next two years, 240 batches were produced without any complications. What followed was highly unexpected and caused a large amount of financial inconvenience. Batch 241 failed quality control due to a solubility problem, as did over 50% of batches synthesised after this date. After analysis of the batch by microscopy and X-ray powder diffraction a second polymorphic form was discovered. This polymorph,

form II, was found to be four times less soluble than the original form I and was established to be the thermodynamic product.[31,32] This four-fold solubility difference between the two polymorphs of Ritonavir is higher than commonly seen for other polymorphs. A study by Pudipeddi et al.[33] looked at solubility ratios of 55 polymorphic compounds. They found the average polymorph solubility ratio to be 2.0 although this covered a range from 1.0, where the two polymorphs had identical solubility to an incredible 23.1 in the case of Premafloxacin.[34]
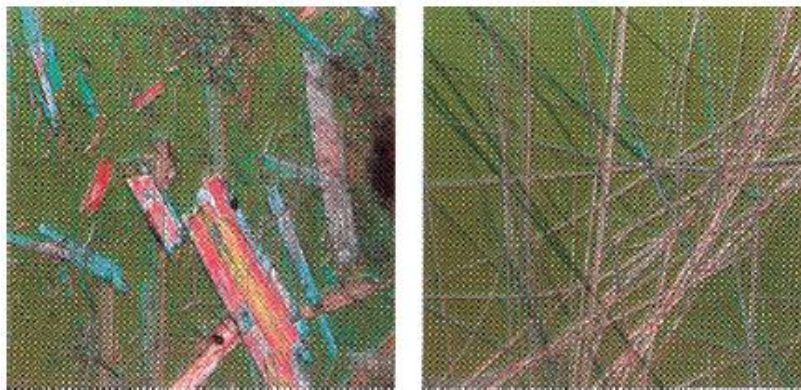
Figure 1.5 Video micrograph of Ritonavir crystal for I (left) and form II (right), reprinted with permission from S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, K. McFarland, *Organic Process Research & Development* **2000**, *4*, 413-417. Copyright 2011 American Chemical Society[32]

Figure 1.5, above shows the two different polymorphic forms of Ritonavir as video micrographs. After the first discovery of the new Form II, the synthesis of form I became disfavoured due to seeding problems during crystal growth. The only solution to this was to either reformulate utilising Form II or (as was eventually done) to get rid of all Form II seeds, both of which are highly expensive options. Generally pharmaceutical companies prefer to commercialise the most thermodynamically stable form of a drug, as this gives greater stability and avoids spontaneous phase transformation. A less thermodynamically stable form (a kinetic form) may be chosen to impart a particular property such as increased solubility – but here stability issues must be thoroughly investigated to avoid future problems. Thus polymorphism is often a problem in drug design, but it may also provide useful solutions if care is first taken to understand potential solid-state transformations.

1.4.3 Co-crystals and solvates

Co-crystals and solvates are another alternative to pharmaceutical salts. Both types of compounds are closely linked and the definition of both is often debated. They differ from salts as they contain neutral molecules that are chemically distinct.[35] A solvate, or pseudopolymorph as it is also sometimes rather improperly known, is a multi-component crystalline material formed from one or more solids and a material which exists in the liquid phase at room temperature. When the liquid phase is water the compound is known as a hydrate. In comparison a co-crystal is a multi-component crystalline material formed from two or more materials that are solids at room temperature. Figure 1.6, shows the breakdown of the different multi-component crystals currently available to pharmaceutical companies.
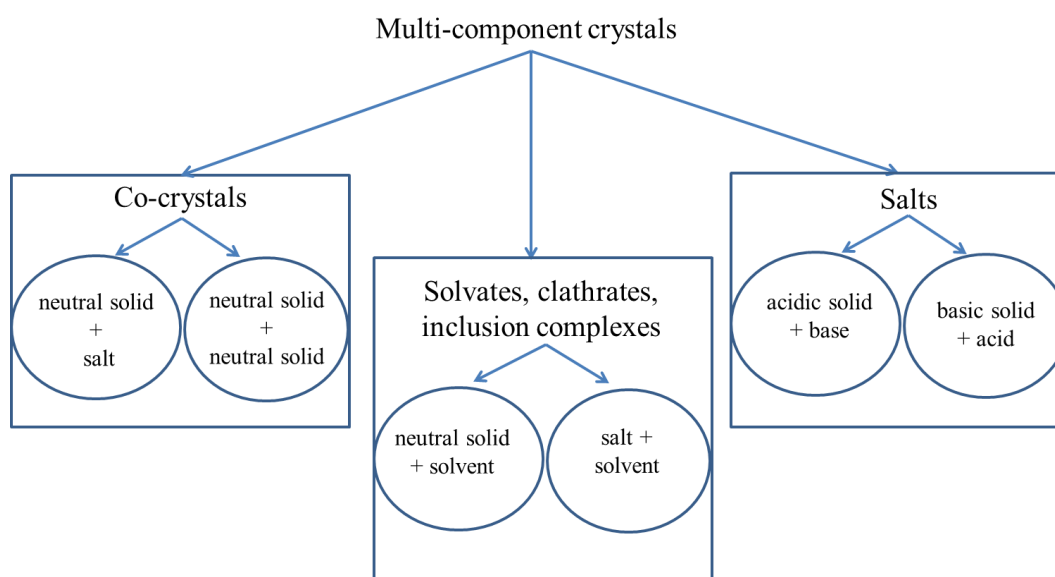
Figure 1.6 Types of multi-component crystals, adapted from reference 36[36]

Whether one of the components is a solid or liquid at room temperature greatly affects the physical properties and stability of the material. The physical properties and stability of solvates vary under different temperature and pressure, which influences the storage conditions and their reliability as pharmaceutical compounds.[37] In comparison co-crystals are normally more stable. Co-crystals, like salts, allow for the chemical and physical properties of the API to be altered without making or breaking covalent bonds.[38] Their advantage over salts is that this can be

achieved regardless of whether the API has suitable acidic or basic groups present.[39] The investigation into the development of pharmaceutical co-crystals is relatively new and, although in theory they can be thought of as a replacement to pharmaceutical salts, in practise they are not currently considered due to a number of underlying issues. The common counterions used in salt synthesis (e.g. Na^+, Cl^-) are known not to have any detrimental effects on the human body. They are generally ions that are already present in the body, they are also cheap and readily available. In contrast, the organic molecules that are used in the synthesis of co-crystals may be relatively expensive and have yet to undergo vigorous testing to demonstrate that they are not toxic and are safe to be present within the human body. Another factor is that salts are often used as they increase the water solubility of many APIs, something that may not be achieved with co-crystals where many organics components may only be sparingly soluble in water. It may also be difficult to attain the correct stoichiometry in co-crystals as there is no charge restraint such as that found for salts.[35]

1.5 Crystal Engineering

Crystal engineering is a concept that has grown rapidly in scope over the last decade, but it was first introduced in 1971 by Schmidt relating to his work on photodimers.[40] It is now widely used in dealing with the supramolecular chemistry of crystals packing, crystal nucleation and growth[41], polymorphism and co-crystal design.[42] In 1989, Desirajui described crystal engineering as follows and his concept is still widely accepted today.

> "The understanding of intermolecular interactions in the context of
> crystal packing and in the utilisation of such understanding in the
> design of new solids with desired physical and chemical
> properties."[43]

This concept is used in industry to attempt to make new tailor-made pharmaceutical drugs that have the desired physicochemical properties, whether that be in the

manipulations of pharmaceutical salts or co-crystals, or indeed to investigate different polymorphs of a given drug compound.[44] New polymorph prediction programs greatly assist in crystal engineering by predicting the intermolecular interactions that will be present in possible polymorphs which may lead to finding the thermodynamic product of pharmaceutical drugs.[45]

1.6   Choosing the right salt

Choosing the correct counterion to be used in salt synthesis for any given drug is still a complicated procedure. During the manufacturing process, issues such as yield, rate and quality of crystallisation, as well as cost and availability of the counterion must be considered. This along with the stability, solubility, and processability of the resulting salt has led to large scale investigations being carried out on varying salts from different counterions. Some counterions can be instantaneously ruled out for potential salt formation due to safety aspects; for example toxicological effects such as chronic and acute dosing of the salt.[17]

According to the 'Orange Book Database' thirty-eight different anions and fifteen different cations have been approved by the Food and Drug Administration (FDA) for the formation of pharmaceutical salts.[13] The current project centres on basic drug molecules, therefore only anionic counterions will be discussed in detail. Table 1.3, shows a small selection of the current anions used in salt synthesis and the percentage of current salts that are in that form. Other counterions that may be used are each present in less than one percentage of salts. Even though salt formation of pharmaceutical drugs is extremely common, very little is currently known about how the choice of a counter ion may change the structure of the resulting solid or what physicochemical properties the resulting salt will display. Despite this fact, there are a few general rules that are followed. These are described below.

Table 1.3 Selection of anionic counterions used for salt synthesis, adapted from reference 17[17]

Counterion	Percentage usage for salt formation
Hydrochloride	42.98
Sulphate	7.46
Bromide	4.68
Chloride	4.17
Tartrate	3.54
Phosphate/diphosphate	3.16
Maleate	3.03
Iodide	2.02
Mesylate	2.02
Hydrobromide	1.90
Acetate	1.26
Pamoate	1.01

Although theoretically any compound that exhibits acidic or basic properties can participate in salt formation, in practise the occurrence and stability of the formed salt is reliant on the relative strengths of the acid and base involved in the chemical reaction. The best salt formation is between a strong acid and strong base, where the difference in pKa is substantial. When a new drug is first synthesised knowledge of the pKa value for each of the ionisable groups in the molecule is essential in determining whether the free acid or base will form a salt.[46] This information is also used in determining a short list of suitable counter ions for the salt formation. If a stable salt is to be formed, it is generally thought that a minimum difference of 3 pKa units is required between the free acid and base counterions.[14]

As can be seen from Table 1.3 hydrochloride counterions are the most commonly used in the formation of pharmaceutical salts. They are often the first choice for weakly basic drug molecules due to cost implications and the required difference in pKa for salt formation. Hydrochloride salts are easy to form and can normally be recrystallised with relative ease from organic solvents. This is however not always a great advantage as this stops the screening of other possible counterions. Hydrochloride salts often have unacceptable high acidity, and may cause corrosion. They may also have poor biological solubility due to risk of salting out, a process where the salt precipitates out of solution due to the common ion effect.[17] Table 1.4,

illustrates the significance of choosing the correct counterion for a drug, so that the desired solubility can be achieved. This example is the drug avitripan (Figure 1.7). It can be seen that the hydrochloride salt has solubility five times less than that of the salts of some organic acids.
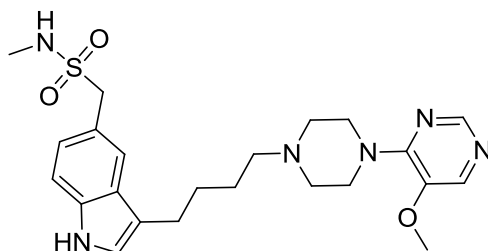
Figure 1.7 Chemical structure of avitripan

Table 1.4 Aqueous solubilities of mono-salts of avitripan containing various counterions, adapted from reference 11[11]

Acid used	pK$_a$ of acid	Solubility (mg/mL at 25°C)
HCl	-6.1	3.4
Methanesulfonic acid	-1.2	16.3
Tartaric acid	3.0, 4.4	14.7
Lactic acid	3.9	15.2
Succinic acid	4.2, 5,6	16.1
Acetic acid	4.8	16.5

When considering a new pharmaceutical salt, whether the chosen counterion gives an anhydrous or a hydrated form of the solid is important as this can alter the observed properties.[47] As described previously, hydrates are solvates where the solvent is water. Often the water occupies definite positions in the crystal lattice, which is achieved by the water molecules forming hydrogen bonds or coordination bonds with the anhydrous drug. Incorporation of the water molecules into the crystal lattice thus alters the physicochemical properties of the drug molecule. Solid hydrates are less water-soluble than their anhydrate solid form if analysed at the same temperature. This is due to the fact that the hydrate already has strong interactions with water within the solid structure; therefore the free energy that is released during further water interaction during crystal dissolution is less for the hydrate than it is for the anhydrate, making it less water soluble.[48] A related problem is hygroscopicity, the tendency to absorb water from air. Hygroscopic drugs are unstable and have limited shelf life, or require special storage instructions.

1.7 Enantiopure and Racemic salts

Recently more attention has been paid to the characterisation of enantiopure drugs in comparison to the racemic adduct in salt formation due to the different pharmacological effects between the two drug forms. An enantiopure drug is one that contains one or more chiral centres, all of which have the same chirality on each molecule, whereas a racemic drug contains equal amounts of both mirror-image enantiomer species. More than half of all marketed drugs contain chiral centres,[49,50] however these drugs are often marketed as racemates, as the preparation of these molecules in achiral environments often results in the racemic species being formed.[51]

In the solid-state chiral compounds can crystallise in different ways. Enantiopure compounds must crystallise to be enantiomerically pure throughout the lattice (Figure 1.8 (a)). Racemic compounds have the ability to crystallise in three different ways; they can spontaneously resolve into an enantiomerically pure arrangement (Figure 1.8 (a)), they can form a heterochiral arrangement with equal numbers of the alternative enantiomers throughout (Figure 1.8 (b)), or the enantiomers can randomly arrange themselves to form a solid solution (Figure 1.8 (c)).[52]

(a)

R R R R R R R R
R R R R R R R R
R R R R R R R R
R R R R R R R R

(b)

R S R S R S R S R
S R S R S R S R S
R S R S R S R S R
S R S R S R S R S

(c)

R S R R R S R S R
S R S R S S S R S
R S R R R S R S R
S R S R S S S R S
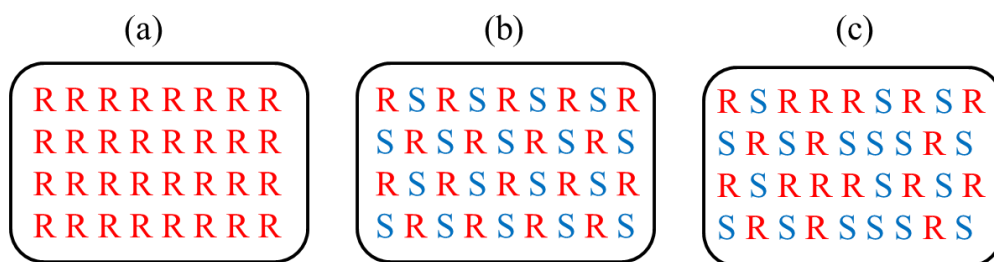
Figure 1.8 Packing arrangements in homochiral and racemic crystals, (a) homochiral packing (same chirality throughout lattice), (b) heterochiral or racemic crystal (regularly arranged alternative enantiomers throughout lattice), (c) solid solution (randomly arranged enantiomers throughout lattice), adapted from reference 52[52]

These three possible outcomes for solid-state arrangement of racemic crystals all have different thermodynamic connotations, which are illustrated in the phase diagrams of Figure 1.9. The racemic conglomerate, where the crystals spontaneously

resolve into an enantiomerically pure species occurs in 5-10% of crystalline racemates, whereas the racemic compound occurs in 90-95% of crystalline racemates. The formation of a solid solution is rare and therefore will not be further discussed. The preference to form a racemate over a conglomerate is normally linked to the theory that during nucleation only half of the molecules arriving at the surface of an enantiopure nuclei are of the right enantiomer for nucleation and then growth, whereas for racemic crystallisation there is no 'wrong' enantiomer for nucleation and during crystal growth the 'wrong' form has only to slide into place to facilitate crystal growth. The ratio of conglomerates vs. racemic compounds has been suggested by Jacques et al. to be altered with the presence of racemic salts, with the formation of racemic conglomerates two to three times more readily than with neutral molecules.[51,53]
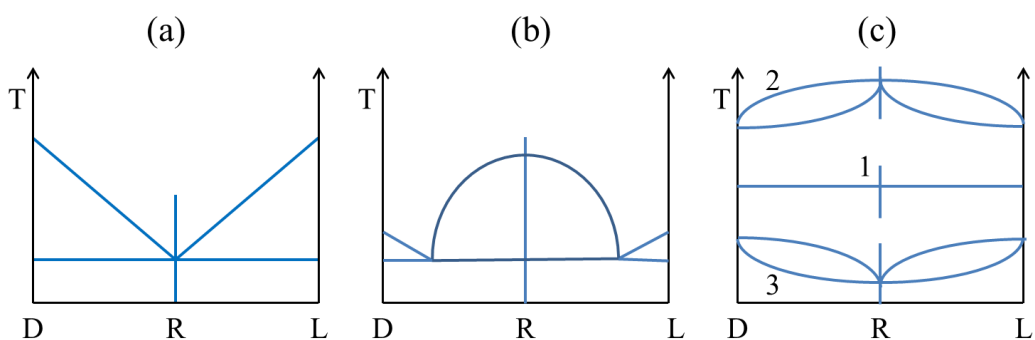
Figure 1.9 Phase diagrams of possible racemic species, (a) racemic conglomerate, (b) racemic compound, (c) pseudoracemate/solid solution (1: ideal melting temperature, 2: maximum melting temperature, 3: minimum melting temperature), adapted from reference 51, page 32[51]

Wallach's rule,[54] states that racemic crystals are inclined to be denser than their enantiopure counterparts. This rule is backed by the study by Brock et al. which demonstrates that on average racemic crystals are more dense than enantiopure crystals (including conglomerates).[55] The molecular arrangement in solid crystal structures is different for enantiopure and racemic compounds due to variation in their packing ability. Racemic compounds can crystallise in any of the 230 space group options, whereas enantiopure compounds are limited to crystallising in only 65 chiral space groups.[55] In theory, as enantiopure crystals are only allowed to use proper symmetry operators (i.e. rotations and translations) they are unable to be more tightly packed than the racemic crystals, that can also use improper symmetry

operators (i.e. centres of inversion, mirror planes, glide planes and alternating rotation-inversion), resulting in a less dense compound. Meanwhile Kitaigorodskii's "principle of close packing" states that close packing ultimately equates to thermodynamic stability and that therefore the densest material will be the most thermodynamically stable.[56] Together with optimising hydrogen-bonding and other directional intermolecular interactions, the need for packing efficiency is often taken as the driving force behind crystal formation. Taking Wallach's rule together with Kitaigoroskii's principle gives the notion that enantiopure solids are less thermodynamically stable than racemic forms.

There are many investigations and opinions on the comparison of melting point of enantiopure and racemic molecules. In the study by Jacques et al.[51](see page 94) a study of 36 pairs of enantiopure and racemic compounds, the racemic compounds appear to be more stable and have a higher average melting point than the enantiopure counterparts, with an average melting point of 405 K compared to 395 K respectively. However, in comparing the melting points for each individual pair, there are 21 pairs where the racemic compound has a higher melting point than the enantiopure equivalent and 14 where the enantiopure melting point is higher than that of the racemic compound which indicates no real trend. The remaining pair has identical melting points.

Li et al.[57] studied the melting points of 25 pairs of enantiopure and racemic drugs and salts. Excluding the 5 pairs where a conglomerate was formed in place of a racemic compound, there are 10 pairs where the racemic compound has a higher melting point and 10 pairs where the enantiopure compound has a higher melting point. Looking at the five conglomerate compounds in comparison to the enantiopure counterpart there is a more obvious trend. For four out of the five pairs the melting point of the enantiopure species is at least 20 $^{\circ}$C higher than that of the conglomerate, with the one case where the conglomerate has the higher melting point having a difference of 3 $^{\circ}$C between the pair. This trend is supported by the Schoeder-van Laar equation.[57] While it is accepted that the melting point of a compound is linked to its stability it is concluded that there is no clear method for predicting which compound

of a pair of enantiopure and racemic counterparts will have the higher melting point based on theoretical considerations alone, the exception perhaps being when the racemic species crystallises as a conglomerate. One investigative strand of this project will be to see if the answer to this problem lies in the analysis of the solid-state interactions within the crystal structure.

A similar debate is present regarding the solubility of enantiopure and racemic salts of the same molecules. Many believe that salts synthesised from the pure enantiomer form of the free acid or base are generally more soluble than those synthesised from the racemic acid or base.[58] However, it is also stated in literature that the solubility of a racemic compound is not related to that of the enantiopure form; therefore like the melting point the solubility can be either greater or smaller, see page 194 of reference 51. For the case of conglomerates, there is a rule linking the solubility of enantiopure compounds and their corresponding racemate. The Meyerhoffer double solubility rule states a conglomerate has the solubility equal to the sum of the solubilities of the corresponding enantiomers.[59] This can be explained by looking at the enantiopure, racemic and conglomerate crystals in terms of their solution-solid mixture.

Racemic and enantiopure compounds display their own solubility independently. At the solution-solid equilibrium a racemic compound will have an equal number of 'R' and 'S' molecules as its solubility characteristic whereas an enantiopure compound will have solely 'R' or 'S' molecules as its solubility characteristic. A conglomerate contains both 'R' and 'S' molecules in enantiopure crystals, where both enantiopure crystals have identical solubility characteristics. A saturated solution of 'R' crystals will thus have the same solution concentration as a saturated of 'S' crystals. In a conglomerate when the 'R' and 'S' crystals equilibrate together the two enantiopure species will display their own solubility independently, and thus the total solubility will be double that seen for the enantiopure compound.[60]

1.8   Intermolecular interactions

As mentioned previously, different salt forms such as polymorphs, solvates and co-crystals all have different material properties. The main reason for this is the differing intermolecular interactions within their structures. These interactions can be split into several types of supramolecular interactions. One of the main purposes of this work is to correlate solid-state structure with material properties. Thus, each type of interaction is described in detail below.

1.8.1   Ionic bonding

Ionic bonding is one the first supramolecular interactions that should be considered. Ion-ion interactions are of comparable strength to covalent bonding, with typical bond energies of 100-350 kJ/mol. The interaction comprises the attraction between oppositely charged anions and cations. A simple example of this is the salt sodium chloride. This ionic solid has an infinite cubic lattice where each sodium anion is surrounded by six chloride cations and vice versa. The structural arrangement of the ions is illustrated in Figure 1.10.[61] The lattice structure is stable in the solid-state but breaks down in solution giving species such as $[Na(H_2O)_6]^+$, which are held together by ion-dipole interactions.
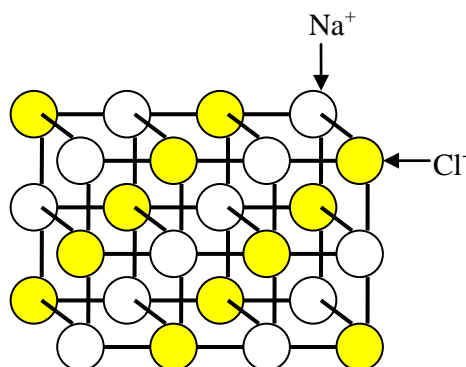
Figure 1.10 The sodium chloride ionic lattice

1.8.2   Ion-dipole interactions

Ion-dipole interactions are the next strongest type of non-covalent bonding interactions with typical strengths of 50-200 kJ/mol. These interactions are seen between ions and polar molecules, with such bonding being observed in both the

solid-state and in solution. As mentioned previously, an example is an ion such as Na^+ interacting with a polar molecule such as water. In the example with sodium and water, the lone pairs on the oxygen atoms are attracted to the cationic positive charge of the sodium, creating a hydration sphere around the metal, see Figure 1.11.[61]
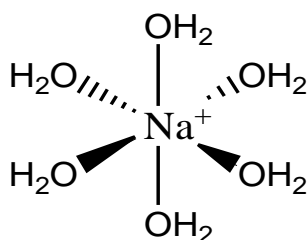
Figure 1.11 Ion-dipole interactions in the hydration sphere of sodium

1.8.3 Dipole-dipole interactions

Dipole-dipole interactions occur when one dipole aligns itself with the opposite dipole on another molecule. The interactions can be split up into two classes: type I where the interactions result from the single alignment of two opposing poles on adjacent molecules and type II, where the interaction is a result of alignment of a pair of dipoles as shown in Figure 1.12. This type of interaction has a bonding energy ranging from 5-50 kJ/mol. Type II dipole-dipole interactions are important in the solid-state for carbonyl compounds, where the interaction strength is similar to that of hydrogen bonds. However, dipole-dipole interactions are relatively weak when in solution.[61]
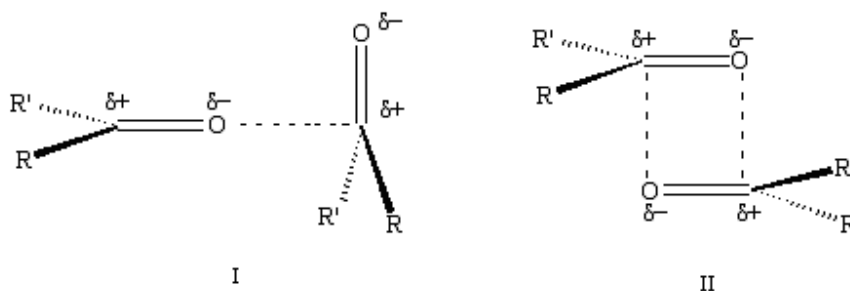
Figure 1.12 Type I and type II dipole-dipole interactions

1.8.4   Hydrogen bonding

Hydrogen bonding is a special kind of dipole-dipole interaction, which occurs when the hydrogen atom attached to an electronegative atom or other electron withdrawing group (the donor), is attracted to the dipole on a neighbouring adjacent molecule or functional group (the acceptor). Hydrogen bonding is a relatively strong interaction and has a high influence on the overall observed behaviour of molecular arrays. The typical length for an O⋯O separation in a hydrogen bond may vary from 2.5 to over 3.0 Å. Longer bond lengths are observed from the interaction of hydrogen with larger atoms such as chlorine. The variation of lengths, strengths and geometries of hydrogen bonds are able to determine the solid-state structure, as well as having an influence on the structure in solution and in the gas phase.

When a hydrogen bond exists between neutral species it is thought that there is a direct relationship between the strength of the hydrogen bond and the crystallographically determined distance between the donor and acceptor atoms of that bond. That is, the shorter the bond length the greater the strength and therefore influence on the structure that bond entails, where the shortest bonds are consistent with large differences in electro-negativity between the donor and acceptor atoms.

The concept of hydrogen bonds has changed quite considerably over the years. The original definition was the interactions between NH⋯O, OH⋯O, or FH⋯F only, but now extends to weaker interactions such as those between CH⋯O and OH⋯π. These are interactions which have very little covalent character present and are only slightly electrostatic. As a result of this, the range of energies for hydrogen bonds is significant, from 4 kJ/mol for weak distant interactions up to 120 kJ/mol for strongly covalent, electronegative (and often intramolecular) interactions. Within a structure it is possible for the donor atom to interact with two distinct hydrogen acceptors simultaneously, where the electro-negativity is split between the two bonds. This makes each individual interaction relatively weak in comparison to an interaction involving solely two components and is known as a bifurcated bond. Table 1.5, demonstrates given examples of strong, moderate and weak hydrogen bonds and examples of how they are utilised within structures.[61]

Table 1.5 Properties of hydrogen bond interactions, adapted from reference 61[61]

Type	Strong	Moderate	Weak
A-H···B interactions	Mainly covalent	Mainly electrostatic	Electrostatic
Bond energy (kJ/mol)	60-120	16-60	<12
Bond length (Å)			
H···B	1.2-1.5	1.5-2.2	2.2-3.2
A···B	2.2-2.5	2.5-3.2	3.2-4.0
Examples	Gas phase dimers with strong acid/bases, HF complexes	Acids, alcohols, DNA base pairs	Minor components of bifurcated bonds, C-H and O-H···π hydrogen bonds

1.8.5   Cation- π interactions

Two possible interactions that are less well known and that have smaller energies than those mentioned above are cation - π interactions and π - π stacking. Cation - π interactions are a non-covalent binding force and are described as the attraction between a cation and the face of a simple π system, such as benzene or ethylene. The interaction can be explained by recognising that sp^2 carbon is more electronegative than hydrogen, thus introducing a $C^{\delta-}$-- $H^{\delta+}$ bond dipole into the system. In π systems such as benzene, the bond dipoles all combine to produce a focused negative charge over the centre of the π system. The cation is then able to bind to this system with average bond energy of 5-80 kJ/mol. This is illustrated in Figure 1.13.
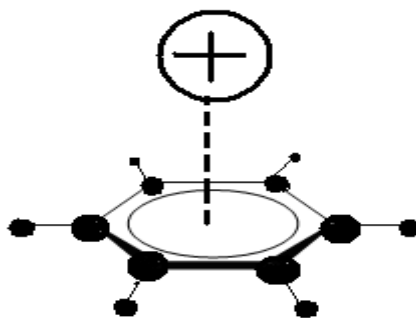
Figure 1.13 Cation- π interaction, adapted from reference 62[62]

The variation in binding energy is dependent on the size of the cation. K^+ binds well to benzene with a binding energy greater than that of K^+ to water. Smaller ions have the ability to bind more tightly due to the more focused intensity of the charge enhancing the electrostatic interactions. For example K^+ and NH_4^+ are of similar size

and therefore have similar cation- π binging energies, whereas $N(CH_3)_4^+$ has a much larger radii and consequently a lower binding energy to benzene. The binding energy is also affected by substituents on the benzene ring. Halides present on the benzene ring or other electron withdrawing groups lessen the strength of the cation- π interactions.[62]

1.8.6 π - π stacking interactions

When a weak electrostatic interaction occurs between two adjacent aromatic rings, where one is electron rich and the other electron poor, this results in π- π stacking. Figure 1.14, shows the two main types of interaction: face to face and edge to face, although there are many intermediate geometries known. π – π stacking is a relatively weak interaction with a bonding energy of 0-50 kJ/mol. A simple model can be used to explain the variety of observed geometries. The model is constructed from the competing electrostatic and van der Waals influences on the molecules. The overall attractive van der Waals interaction is proportional to the contact surface area of the two π-systems, where the negatively charged π-electron cloud of one molecule is attracted to the positively charged σ-framework on an adjacent molecule and vice versa. In contrast, the electrostatic repulsions determine the relative orientation of the two interacting molecules and are caused by repulsions between the two negative π-systems.[61]
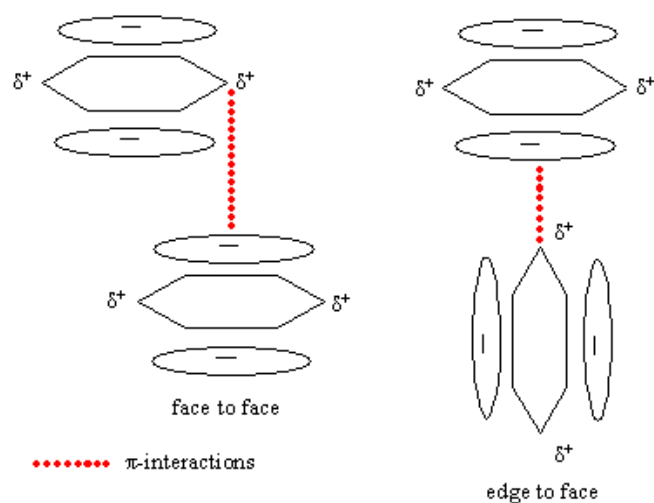
Figure 1.14 Face to face and edge to face π - stacking

1.8.7   van der Waals interactions

The weakest available interactions in solid-state structures are van der Waals forces, each typically with energy of less than 5 kJ/mol. Such van der Waals forces are electrostatic interactions that arise from the polarisation of an electron cloud by an adjacent nucleus. They are divided into dispersion (London) and exchange-repulsion terms. The London, or dispersion, element is resultant of fluxuating multipoles in adjacent molecules, which decreases with a distance of r^{-6}. This is an additive interaction where every bond in the molecule contributes to the overall observed interaction. The exchange-repulsion element defines the molecular shape and is a shorter range interaction, which is related to the overlap of electron clouds.[61]

Individually the majority of the above interactions are weak and do not have a major effect on the solid-state packing of molecules. However, when many of these individual interactions are combined, the resulting overall energy has a significant influence on the molecules packing abilities in the solid-state.

1.9   Etter's rules

Etter established a set of rules that attempt to predict when and where hydrogen bonds will occur within organic molecules.[63,64] These rules were based on analysis of the vast amounts of structures present in the Cambridge Structural Database.[65] The three rules are as follows;

1. All good proton donors and acceptors are used in hydrogen bonding.
2. Five, six or seven membered intramolecular hydrogen bonds will form in preference to intermolecular hydrogen bonds.
3. After the formation of intramolecular hydrogen bonds the best remaining proton donors and accepters will form intermolecular hydrogen bonds to one another.

In order to determine whether a hydrogen bond donor or acceptor is termed 'good' or not two categories of acceptors and donors have been determined, namely reliable and occasional hydrogen bond donor and acceptors. The functional groups within these two categories are tabulated in Table 1.6.[66] The formation of intramolecular hydrogen bonds instead of intermolecular hydrogen bonds affects the water solubility of drug compounds, with closed (intramolecular bonded) systems having a lower water-solubility than open (intermolecular bonded) systems.[67]

Table 1.6 Reliable and occasional hydrogen bond donor and acceptors, adapted from reference 66[66]

Type	Functional group involved
reliable donor	-OH, -NH$_3$, -NHR, -CONH$_2$, -CONHR, COOH
occasional donor	-COH, XH, SH, CH
reliable acceptors	-COOH, -COHNCO-, -NHCONH-, -CON< (1-3°), >P=O, > S=O, -OH
occasional acceptors	>O, -NO$_2$, -CN, -CO, -COOR, -N<, -Cl

1.10 Graph-set notation and analysis

Graph-set analysis is used to describe complicated hydrogen bond patterns in a systematic and consistent way.[64,68] The hydrogen bond motif is described as the set of molecules that are hydrogen bonded to one another. The motif can be infinite or finite and cyclic or non-cyclic. Intermolecular hydrogen bonds can assemble to form one of three motifs namely discrete, chain and ring, which have the designators D, C and R respectively. Discrete and ring hydrogen bonds are finite whereas chain graph-sets form an infinite propagation throughout a structure. Intramolecular motifs are named as the "self" graph-set and have the designator S.
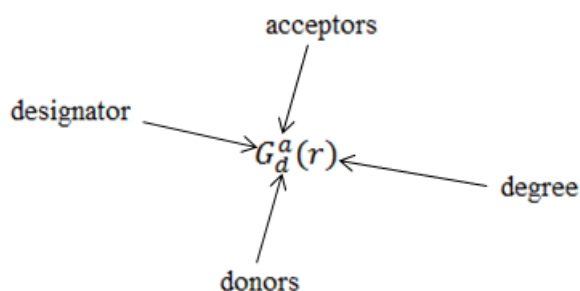
Figure 1.15 Specified graph-set (G)is the designator, (r)is the degree of the graph-set, (d)is the number of donor atoms and (a) is the number of acceptor atoms

Figure 1.15, shows how a graph-set is specified where 'G' is the designator and 'r' is the degree which represents the number of atoms involved in the graph-set. 'a' and 'd' are the number of hydrogen bond acceptor atoms and hydrogen bond donor atoms respectively. The building blocks that make up the graph-sets are best illustrated by the following examples.

1.10.1 Graph-set analysis for (1R,2S)(-)methylephedrinium chloride

This compound can be described with three graph-sets, illustrated in Figure 1.16. The two discrete graph-sets X and Y both have notation $D_1^1(2)$, as both contain one acceptor atom and one donor atom and there are a total of two atoms involved in the bond. Combination of these two graph-sets produces the chain motif $C_2^1(7)$. In Figure 1.16 the two donor atoms are coloured yellow, the acceptor is mauve and the 7 atoms that form the total motif described are numbered 1 to 7. The graph-sets are combined with X in the forward direction and Y in the reverse direction as they both share a common hydrogen bond acceptor. The notation for this is >X<Y.
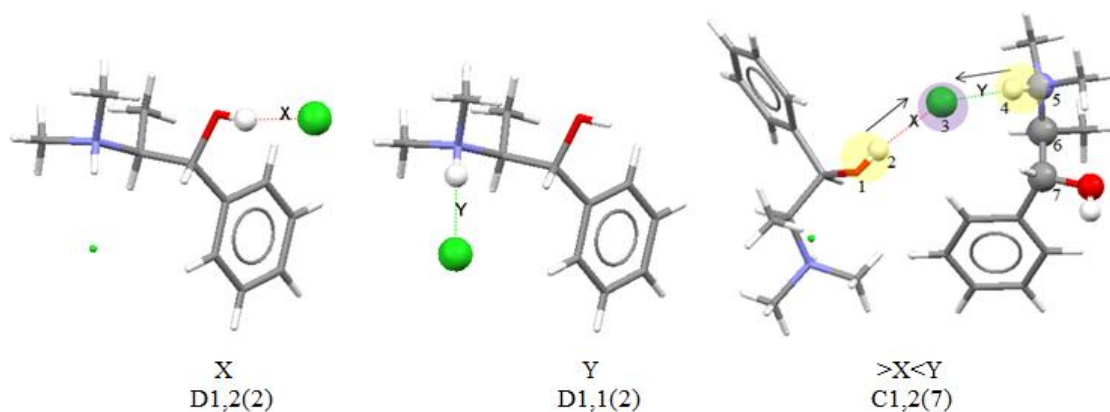
Figure 1.16 Graph-set of (1R,2S)(-)methylephedrinium chloride.

1.10.2 Graph-set analysis for (+/-)methylephedrinium hydrogen-sulfate monohydrate

A sub-section of the compound's graph-set has been used to illustrate the ring motif present between the anion and the water molecule, see Figure 1.17. The ring motif is made up of the two discrete graph-sets X and I. For the formation of the ring the graph-sets X and I are operating in the same direction, >X>I>X>I. The motif has four hydrogen bond acceptor and four hydrogen bond donor atoms as highlighted, and there are twelve atoms in total, hence the final graph-set is $R_4^4(12)$.
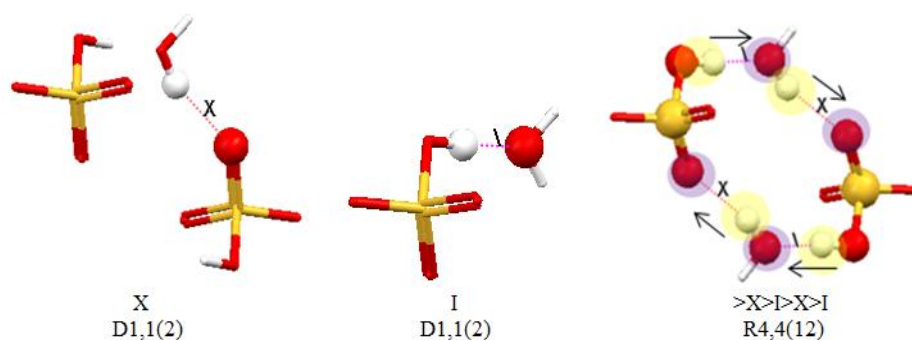
Figure 1.17 Graph-set of (+/-)methylephedrinium hydrogen-sulfate monohydrate

1.10.3 Graph-set analysis for (+/-)methylephedrinium malonate

The final motif to be illustrated is the self graph-set. This is shown by the intramolecular hydrogen bond of the hydrogen-malonate anion. The graph-set contains one hydrogen bond donor and one hydrogen bond acceptor and six atoms involved in the closed loop, see Figure 1.18. This gives the overall graph-set $S_1^1(6)$.
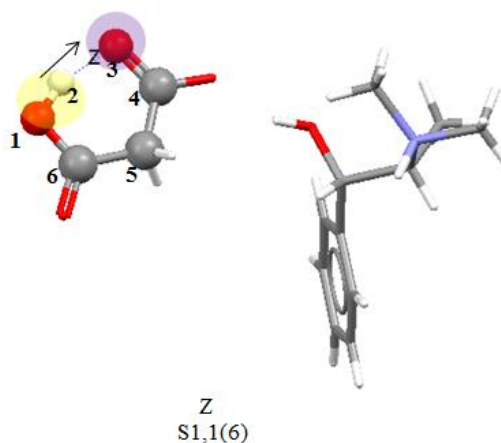
Figure 1.18 Graph-set of (+/-)methylephedrinium hydrogen-malonate

1.11 The PIXEL method

When looking at organic crystals the interaction energies that hold adjacent molecules together are much smaller than those occurring between atoms within the organic molecules.[69] Semi Classical Density Sums or the PIXEL method use a semi-empirical computational procedure to calculate the coulombic, polarization, dispersion and repulsions energies between separate rigid molecules.[70,71] This is achieved by using a fully delocalised electron density cloud for each molecule. The electron density is transposed into a grid of about 10^6 points (or pixels) which are then compressed into superpixels of size n x n x n (where n = 3, 4 or 5). By then eliminating any pixels with a negligible electron density the overall molecule can be described by approximately 10 000 pixels. Symmetry and space group operations are then used to determine the electron density of the entire molecular cluster of the crystal array. Figure 1.19, shows how the entire molecule is represented by electron density instead of just the individual atoms.
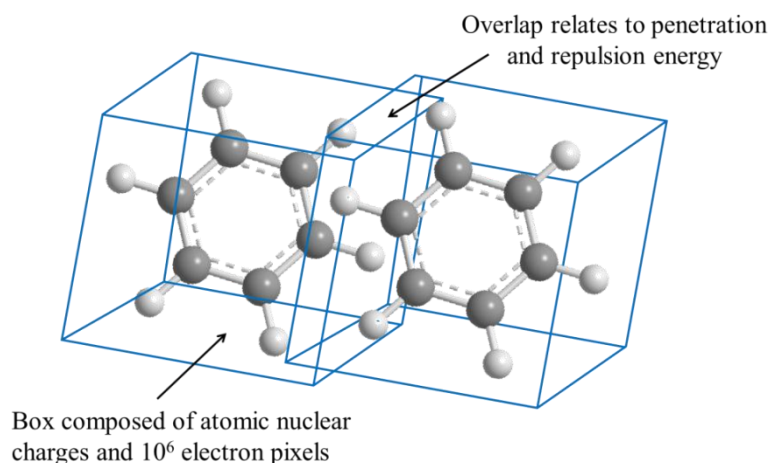
Figure 1.19 PIXEL electron density depiction of benzene, adapted from reference 69[69]

The four different types of forces that are calculated using the PIXEL method are shown in Table 1.7, along with their variation with intermolecular distance. As well as calculating the overall terms in the crystal individual energies can be extracted for molecules separated by a given symmetry operator. This allows comparison of individual molecular interactions as well as understanding which interactions have a weak overall energy possibly caused by a strong repulsion term.

Table 1.7 Forces involved in PIXEL calculations

Force	Arises from	Varies as
coulombic	charge–charge interaction	r^{-1}
polarisation	permanent dipole-dipole interaction	r^{-3}
dispersion	inductive dipole-dipole using London type formula	r^{-6}
repulsion	repulsion between spins	overlap of electron clouds

Using the PIXEL method has many advantages over more traditional force field methods. Firstly it takes into account the electron density over the entire molecule, and it is able to use this property to determine the separate coulombic, polarisation, dispersion and repulsion energies involved in intermolecular interactions. The electron densities were generated using Gaussian03[72] with the MP2 level of theory and 6-31G** basis set. A crystal calculation can be run on a standard computer and will take no longer than an hour for small molecules of the type under consideration here. The main downside to the program is its rigid closed-shell approach, this is a disadvantage when hydrogen bonding occurs between two molecules as this may rearrange the electron arrangements.[73] Despite this limitation the PIXEL method has been used successfully for many different molecular applications[74-78] as well as examining molecules under extreme pressure conditions.[79]

1.12 Current practice of drug companies

For pharmaceutical companies the current drug development process can be a trial and error system. Firstly they must choose an appropriate drug molecule which possesses desirable pharmaceutical effects. This is then reacted with a counterion to produce a salt that exhibits the sought-after physicochemical properties. As previously illustrated the perfect candidate drug molecule linked with an inappropriate counterion could produce an overall API that is deemed useless, as it does not have the correct physicochemical properties. For example if a salt molecule has too low an aqueous solubility, then whether the drug contains the desirable pharmaceutical effects is irrelevant as it will be unable to dissolve into the blood-stream to a large enough extent for the drug to be able to work efficiently.

There is still a lack of understanding as to how solid-state structure and physicochemical properties correlate. Therefore salt selection is still carried out by inefficient and time consuming trial and error methods involving the preparation and testing of numerous salt forms. A key reason for this lack of understanding is that solid-state structure-property correlation analysis is held back by the absence of suitable, large and systematically connected structural datasets and by the lack of physical property data to go with these datasets. Until recently single crystal diffraction measurements were relatively slow and tended to be reserved for compounds of some 'special interest'. However, due to increased instrument capacity within the last decade it is now possible to generate large numbers of crystal structures to examine one particular problem, within a reasonable timescale. [80]

1.13 References

(1)   Strahl, P. H.; Wermuth, C. G. *Handbook of Pharmaceutical Salts: Properties, Selection and Use*; VHCA: Zurich, 2008.
(2)   Bryn, S. R. *Solid State Chemistry of Drugs*; Academic Press: New York, 1982.
(3)   Lipinski, C. A. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235.
(4)   Storer, B. E. *Biometrics* **1989**, *45*, 925.
(5)   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Advanced Drug Delivery Reviews* **1997**, *23*, 3.
(6)   Stahly, G. P. *Crystal Growth & Design* **2007**, *7*, 1007.
(7)   Brittain, H. G. *Physical Characterization of Pharmaceutical Solids*; Marcel Dekker, Inc.: New York, 1995.
(8)   Berge, S. M.; Bighley, L. D.; Monkhouse, D. C. *J. Pharm. Sci.* **1977**, *66*, 1.
(9)   Huang, L. F.; Tong, W. Q. *Advanced Drug Delivery Reviews* **2004**, *56*, 321.
(10)   DiMartino, P.; Conflant, P.; Drache, M.; Huvenne, J. P.; GuyotHermann, A. M. *Journal of Thermal Analysis* **1997**, *48*, 447.
(11)   Serajuddin, A. T. M. *Advanced Drug Delivery Reviews* **2007**, *59*, 603.
(12)   Bronsted, J. N. *Angewandte Chemie* **1930**, *43*, 229.
(13)   Paulekuhn, G. S.; Dressman, J. B.; Saal, C. *J. Med. Chem.* **2007**, *50*, 6665.
(14)   Bastin, R. J.; Bowker, M. J.; Slater, B. J. *Organic Process Research & Development* **2000**, *4*, 427.
(15)   Lam, K. W.; Xu, J. J.; Ng, K. M. *Industrial & Engineering Chemistry Research* **2010**, *49*, 12503.
(16)   Chowhan, Z. T. *J. Pharm. Sci.* **1978**, *67*, 1257.
(17)   Gould, P. L. *Int. J. Pharm.* **1986**, *33*, 201.
(18)   Kramer, S. F.; Flynn, G. L. *J. Pharm. Sci.* **1972**, *61*, 1896.
(19)   Anderson, B. D.; Conradi, R. A. *J. Pharm. Sci.* **1985**, *74*, 815.
(20)   Ledwidge, M. T.; Corrigan, O. I. *Int. J. Pharm.* **1998**, *174*, 187.

(21) Li, S. F.; Wong, S. M.; Sethia, S.; Almoazen, H.; Joshi, Y. M.; Serajuddin, A. T. M. *Pharm. Res.* **2005**, *22*, 628.

(22) Bhattachar, S. N.; Deschenes, L. A.; Wesley, J. A. *Drug Discovery Today* **2006**, *11*, 1012.

(23) Carstensen, J. T. *Advanced pharmaceutical solids*; 1 ed.; Marcel Dekker: New York, 2001.

(24) Bernstein, J. *Chem. Commun.* **2005**, 5007.

(25) Hilfiker, R. *Polymorphism: In the Pharmaceutical Industry*; 1 ed.; Wiley, 2006.

(26) *Modern Pharmaceutics*; 5 ed.; Florence, A. T.; Siepmann, J., Eds.; Informa Healthcare USA: New York, 2009; Vol. 1.

(27) Tong, W. Q.; Whitesell, G. *Pharm Dev Technol* **1998**, *3*, 215.

(28) Davey, R. J. *Chem. Commun.* **2003**, 1463.

(29) McCrone, W. C. *Physics and Chemistry of the Organic Solid State*; Interscience: New York, 1963.

(30) Bernstein, J.; Davey, R. J.; Henck, J. O. *Angewandte Chemie-International Edition* **1999**, *38*, 3441.

(31) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharm. Res.* **2001**, *18*, 859.

(32) Chemburkar, S. R.; Bauer, J.; Deming, K.; Spiwek, H.; Patel, K.; Morris, J.; Henry, R.; Spanton, S.; Dziki, W.; Porter, W.; Quick, J.; Bauer, P.; Donaubauer, J.; Narayanan, B. A.; Soldani, M.; Riley, D.; McFarland, K. *Organic Process Research & Development* **2000**, *4*, 413.

(33) Pudipeddi, M.; Serajuddin, A. T. M. *J. Pharm. Sci.* **2005**, *94*, 929.

(34) Schinzer, W. C.; Bergren, M. S.; Aldrich, D. S.; Chao, R. S.; Dunn, M. J.; Jeganathan, A.; Madden, L. M. *J. Pharm. Sci.* **1997**, *86*, 1426.

(35) Mohamed, S.; Tocher, D. A.; Vickers, M.; Karamertzanis, P. G.; Price, S. L. *Crystal Growth & Design* **2009**, *9*, 2881.

(36) Monissette, S. L.; Almarsson, O.; Peterson, M. L.; Remenar, J. F.; Read, M. J.; Lemmo, A. V.; Ellis, S.; Cima, M. J.; Gardner, C. R. *Advanced Drug Delivery Reviews* **2004**, *56*, 275.

(37) Vishweshwar, P.; McMahon, J. A.; Bis, J. A.; Zaworotko, M. J. *J. Pharm. Sci.* **2006**, *95*, 499.

(38) Sarma, B.; Chen, J.; Hsi, H. Y.; Myerson, A. S. *Korean J. Chem. Eng.* **2011**, *28*, 315.

(39) Schultheiss, N.; Newman, A. *Crystal Growth & Design* **2009**, *9*, 2950.

(40) Schmidt, G. M. J. *Pure Appl. Chem.* **1971**, *27*, 647.

(41) Davey, R. J.; Allen, K.; Blagden, N.; Cross, W. I.; Lieberman, H. F.; Quayle, M. J.; Righini, S.; Seton, L.; Tiddy, G. J. T. *Crystengcomm* **2002**, 257.

(42) Almarsson, O.; Zaworotko, M. J. *Chem. Commun.* **2004**, 1889.

(43) Desiraju, G. R. *Crystal Engineering. The Design of Organic Solids*; Elsevier: Amsterdam, 1989.

(44) Chow, K.; Tong, H. H. Y.; Lum, S.; Chow, A. H. L. *J. Pharm. Sci.* **2008**, *97*, 2855.

(45) Chisholm, J.; Pidcock, E.; Van De Streek, J.; Infantes, L.; Motherwell, S.; Allen, F. H. *Crystengcomm* **2006**, *8*, 11.

(46) Gennaro, A. R. *Remington: The science and practice of pharmacy*; 19th ed.; Mack Pub. Co.: Easton, Pa, 1995.

(47) Reddy, J. S.; Ganesh, S. V.; Nagalapalli, R.; Dandela, R.; Solomon, K. A.; Kumar, K. A.; Goud, N. R.; Nangia, A. *J. Pharm. Sci.* **2011**, *100*, 3160.

(48) Khankari, R. K.; Grant, D. J. W. *Thermochim. Acta* **1995**, *248*, 61.

(49) Borman, S. *Chem. Eng. News* **1990**, *68*, 9.

(50) Millership, J. S.; Fitzpatrick, A. *Chirality* **1993**, *5*, 573.

(51) Jacques, J.; Collet, A.; Wilen, S. H. *Enantiomers, Racemates, and Resolutions*; Wiley: New York, 1981.

(52) Collet, A.; Ziminski, L.; Garcia, C.; VigneMaeder, F. *Supramolecular Stereochemistry* **1995**, *473*, 91.

(53) Duddu, S. P.; Grant, D. J. W. *Pharm. Res.* **1994**, *11*, 1549.

(54) Wallach, O. *Liebigs Ann Chem* **1895**, *286*, 90.

(55) Brock, C. P.; Schweizer, W. B.; Dunitz, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 9811.

(56) Kitaigorodskii, A. I. In *Organic Chemical Crystallography*; Consultants Bureau: New York, 1961.

(57) Li, Z. J.; Zell, M. T.; Munson, E. J.; Grant, D. J. W. *J. Pharm. Sci.* **1999**, *88*, 337.

(58) Yonemochi, E.; Yoshihashi, Y.; Terada, K. *Pharm. Res.* **2000**, *17*, 90.

(59) Meyerhoffer, W. *Ber. Dtsch. Chem. Ges.* **1904**, *37*, 2604.

(60) Izumi, T.; Blackmond, D. G. *Chemistry-a European Journal* **2009**, *15*, 3065.

(61) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; 2nd ed.; Wiley: Chicchester, 2000.

(62) Atwood, J. L.; Steed, J. W. *Encyclopedia of Supramolecular Chemistry*; Marcel Dekker: New York, 2004.

(63) Etter, M. C. *J. Am. Chem. Soc.* **1982**, *104*, 1095.

(64) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.

(65) Allen, F. H. *Acta Crystallographica Section B-Structural Science* **2002**, *58*, 380.

(66) Byrn, S. R.; Pfeiffer, R. R.; Stephenson, G.; Grant, D. J. W.; Gleason, W. B. *Chem. Mater.* **1994**, *6*, 1148.

(67) Kuhn, B.; Mohr, P.; Stahl, M. *J. Med. Chem.* **2010**, *53*, 2601.

(68) Etter, M. C.; Macdonald, J. C.; Bernstein, J. *Acta. Cryst.* **1990**, *B46*, 256.

(69) Dunitz, J. D.; Gavezzotti, A. *Angewandte Chemie-International Edition* **2005**, *44*, 1766.

(70) Gavezzotti, A. *J. Phys. Chem. B* **2002**, *106*, 4145.

(71) Gavezzotti, A. *J. Phys. Chem. B* **2003**, *107*, 2344.

(72) Frinch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyoto, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.;

Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Pittsburgh PA, 2003.

(73)   Gavezzotti, A. *Struct. Chem.* **2005**, *16*, 177.

(74)   Boese, R.; Clark, T.; Gavezzotti, A. *Helv. Chim. Acta* **2003**, *86*, 1085.

(75)   Cabeza, A. J. C.; Day, G. M.; Motherwell, W. D. S.; Jones, W. *Crystal Growth & Design* **2007**, *7*, 100.

(76)   Dunitz, J. D.; Gavezzotti, A. *Crystal Growth & Design* **2005**, *5*, 2180.

(77)   Eckhardt, C. J.; Gavezzotti, A. *J. Phys. Chem. B* **2007**, *111*, 3430.

(78)   Johnston, A.; Bardin, J.; Johnston, B. F.; Fernandes, P.; Kennedy, A. R.; Price, S. L.; Florence, A. J. *Crystal Growth & Design* **2011**, *11*, 405.

(79)   Wood, P. A.; Francis, D.; Marshall, W. G.; Moggach, S. A.; Parsons, S.; Pidcock, E.; Rohl, A. L. *Crystengcomm* **2008**, *10*, 1154.

(80)   Hursthouse, M. B. *Crystal. Rev.* **2004**, *10*, 85.

Chapter 2

Project aims and objectives

2.1 Aims

The aim of this project is to develop an improved understanding of the link between solid-state structure of organic salts and their physicochemical properties. A library of systematically related organic salts is presented along with corresponding physicochemical measured properties. The principal properties examined are solubility and melting point as these are of great importance with respect to the salt selection process employed in the pharmaceutical industry. The generated data is also analysed for trends in supramolecular structural features such as formation of hydrates. Correlation analysis is performed between measured physicochemical properties. Chemometric analysis of structural and measured properties is utilised to build regression and classification models that are developed to predict the aqueous solubility of salts, thus adding an element of rational design to the salt selection process.

2.2 Objectives

i. React a series of structurally related acid counterions with a systematic structural series of bases and characterise the products by single crystal X-ray diffraction to generate a library of related crystalline structures.

ii. Use structure analysis programs to compile information on supramolecular structural trends including hydrogen bonding motifs, molecular packing arrangements and isostructural compounds.

iii. Use structural and physicochemical properties collected to perform comparative analysis of enantiopure and racemic salts.

iv. Analyse structures that from hydrates using structural analysis program to examine common packing arrangements.

v. Measure solubility and melting point of the salts where suitable structural information has been determined.

vi. Perform correlation analysis on measured physicochemical properties to establish relationships between properties and structures.

vii.   Use chemometric programs to build and develop regression and classification models of solubility for database of salts, inputting measured and calculated physical and structural properties.

viii.   Test prediction capability of regression and classification models by utilising them to predict the solubility of salts.

Chapter 3
Materials and Methods

3.1   Materials
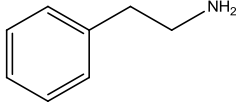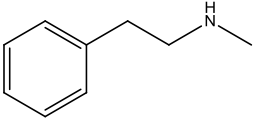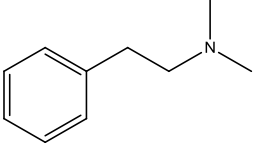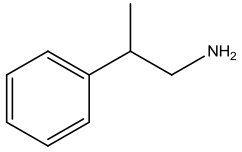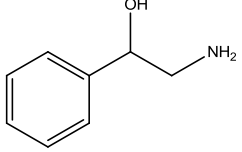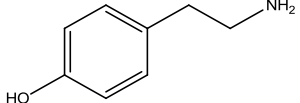
All materials were obtained commercially from Sigma-Aldrich, Fluka or Alfa Aesar and were used as supplied, with the exception of (+/-)methylephedrine which was formed from a 50:50 mixture of the commercially available (1R,2S)(-)methylephedrine and (1S,2R)(+)methylephedrine. To build the systematic database of associated families of organic salts eleven bases and forty-two acids were chosen. The eleven base molecules are all structurally related with the same phenethylamine (PEA) backbone. This similarity eases attempts to study the effect of small structural changes on solid-state packing and resultant physical properties. The phenethylamine family was chosen as this structural moiety lies at the core of asthma (bronchodilator) drugs, such as the Schering-Plough product salbutamol. An added advantage is that as well as being 'model' bases for more complex drugs, many of the simple PEAs are also of biological/pharmaceutical interest in their own right. Additionaly one of the few structural works present within the literature is on the salt family of the PEA compound (-)ephedrine. This was performed by Davey et al in 2006.[1] The eleven bases also allow for the comparison of enantiopure versus racemic structures of salts, with both (-)methylephedrine and (+/-)methylephedrine present. Other relevant work on PEAs can be found in references 2 through 6.[2-6]

The structures of the eleven PEA bases are shown below in Table 3.1 along with their abbreviated name, molecular weight and melting point. The stepwise changes between the different bases, with the addition of methyl or hydroxyl groups either on the benzyl ring or the aliphatic chain, can be seen. The addition of hydroxyl groups should provide an opportunity for more extensive hydrogen bonding which in turn may influence the packing of the molecules in the solid-state.

The forty-two acids that were chosen to build the database can be split into several groups of interest. Although several of the selected acids are not pharmaceutically acceptable counterions, they were chosen as they form structural series with those that are. A family of 18 benzoic acid derivatives and benzoic acid itself were used to investigate small changes in substituent on bonding and physical properties. The carboxylic acids were further investigated with a group of five aliphatic dicarboxylic

acids that have the ability to be either singly or doubly deprotonated. The remaining six carboxylic acids that were used consist of three enantiopure and racemic acid pairs. Four tetrahedral counterions were used which have a similar shape but which have differing overall charges and/or hydrogen bonding capabilities. In the halide series all three counterions (Cl^-, Br^- and I^-) have the same shape and charge and therefore the effects of stepwise change in size and hence charge density can be observed. The sulfonic acid group consists of a selection of sulfonic acids with varying substituent groups. Table 3.2 shows the structures of the acids, their abbreviated name, molecular weight, melting point and aqueous solubility information. The abbreviation of the base followed by that of the acid was used to name the salts, e.g. 2-phenylethylammonium 2-chlorobenzoate has the abbreviated name PEA2CB.

Table 3.1 Basic counterions selected for construction of systematic database including their molecular weight and melting point taken from MSDS *RMEpd melting point experimental not taken from MSDS

Base	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)
2-Phenylethylamine	PEA		121.18	n/a
2-Methylphenethylamine	MPEA		135.21	n/a
2-Dimethylphenethylamine	DMPEA		149.23	n/a
2-Phenylpropylamine	PPA		135.21	n/a
2-Hydroxyphenethylamine	HPEA		137.18	56-58
Tyramine	TYR		137.18	160-162
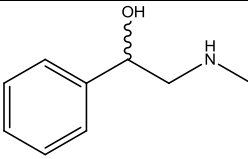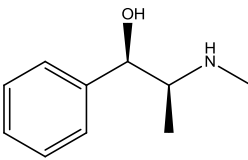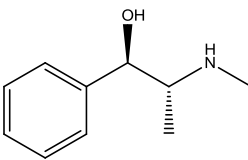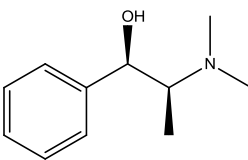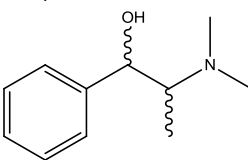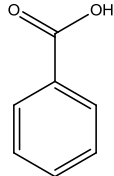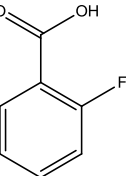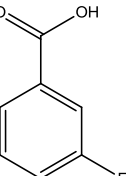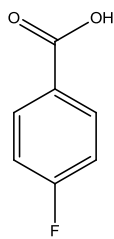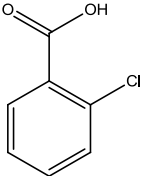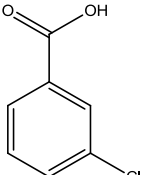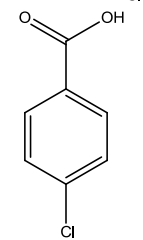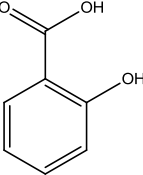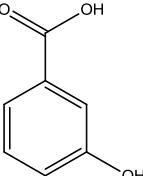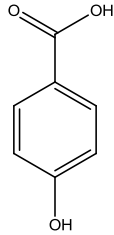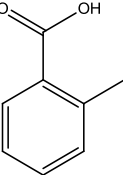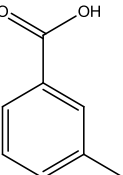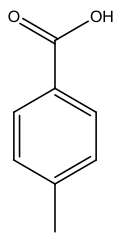
Base	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)
α(Methylaminomethyl) benzyl alcohol	MAMBA		151.23	74-76
(-)Ephedrine	Epd		165.23	36
(-)Pseudoephedrine	PEpd		165.23	118-120
(-)Methylephedrine	MEpd		179.26	85-88
(+/-)Methylephedrine	RMEpd		179.26	63

42
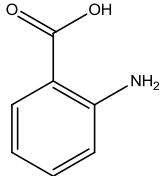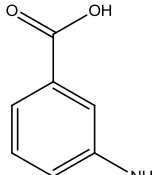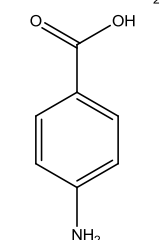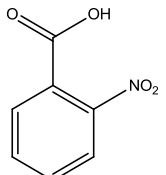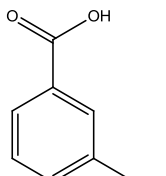
Table 3.2 Acidic counterions selected for construction of systematic database including their molecular weight and melting point (taken from MSDS) and aqueous solubility (extracted from "The handbook of aqueous solubility data"[7]).
*R-mandelic acid was purchased as l-mandelic acid, also known as (-)mandelic acid or D-mandelic acid
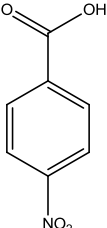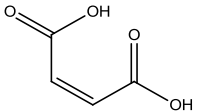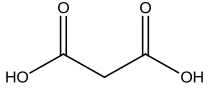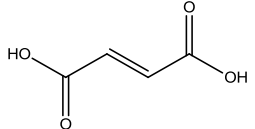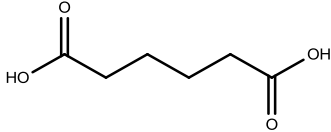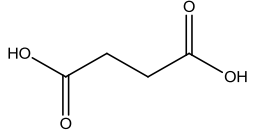
Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
Benzoic acid	Bz		122.12	122-125	0.0270
2-Florobenzoic acid	2FB		140.11	122-125	0.0514
3-Florobenzoic acid	3FB		140.11	122-124	0.0107
4-Florobenzoic acid	4FB		140.11	182-184	0.0086

43

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
2-Chlorobenzoic acid	2CB		156.57	138-140	0.0110
3-Chlorobenzoic acid	3CB		156.57	153-157	0.0255
4-Chlorobenzoic acid	4CB		156.47	238-241	0.0478
2-Hydroxybenzoic acid	2HB		138.12	158-161	0.0162
3-Hydroxybenzoic acid	3HB		138.12	200-203	0.0525

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
4-Hydroxybenzoic acid	4HB		138.12	213-217	0.0477
o-Toluic acid	oTol		136.15	102-104	0.0087
m-Toluic acid	mTol		136.15	107-113	0.0072
p-Toluic acid	pTol		136.15	177-180	0.0027

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
2-Aminobenzoic acid	2AB		137.14	144-148	0.0345
3-Aminobenzic acid	3AB		137.14	178-180	0.0507
4-Aminobenzoic acid	4AB		137.14	187-189	0.0368
2-Nitrobenzoic acid	2NB		167.12	146-148	0.0438
3-Nitrobenzoic acid	3NB		167.12	139-141	0.0177

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
4-Nitrobenzoic acid	4NB		167.12	237-240	0.0021
Maleic acid	Male		116.07	137-140	3.746
Malonic acid	Malon		104.06	132-135	5.724
Fumaric acid	Fum		116.07	298-300	0.0600
Adipic acid	Adp		146.14	151-154	0.1664
Succinic Acid	Suc		118.09	184-186	0.6389

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (°C)	Solubility (g/L at 25°C)
L-Malic acid	LMal		134.09	101-103	4.030
(+/-)-Malic acid	RMal		134.09	131-133	4.408
L-Tartartc acid	LTar		150.09	170-172	3.965
(+/-)-Tartaric acid	RTar		150.09	210-212	1.103
R-Mandelic acid*	LMD		152.14	130-133	0.9217

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
(+/-)-Mandelic acid	RMD		152.14	119-121	1.0200
Sulfuric acid	SO4		98.09	n/a	n/a
Tetrafloroboric acid	BF4		87.81	n/a	n/a
Perchloric acid	CLO4		100.46	n/a	n/a
Phorphoric acid	PO4		98.00	n/a	n/a
Hydrochloric acid	Cl	H—Cl	36.46	n/a	n/a
Hydrobromic acid	Br	H—Br	80.92	n/a	n/a
Hydroiodic acid	I	H—I	127.91	n/a	n/a
Methanesulfonic acid	MeSO3		96.11	n/a	n/a

Acid	Abbreviated name	Molecular structure	Molecular weight	Melting point (^{o}C)	Solubility (g/L at 25°C)
Ethanesulfonic acid	EtSO3		110.13	n/a	n/a
1,2-Ethanedisulfonic acid	EDS		190.19	111-112	n/a
Benzenesulfonic acid	BS		157.17	n/a	3.088
4-Hydroxybenzene-sulfonic acid	4HBS		174.17	n/a	n/a

All acids supplied as liquids or solutions were used neat and measured using an automatic pipette. Table 3.3 shows the concentrations of the liquids in percentage weight.

Table 3.3 Concentration of liquid acids used in synthesis in percentage weight

Acid	Percent weight (% wt.)
Sulfuric acid	95
Tetrafluoroboric acid	40
Phosphoric acid	85
Hydrochloric acid	37
Hydrobromic acid	48
Hydroiodic acid	55
Methanesulfonic acid	70
Ethanesulfonic acid	70
4-Hydroxybenzenesulfonic acid	65

3.2 Salt synthesis and crystallisation methods

3.2.1 Crystallisation method 1 – evaporation and cooling

The synthesis of all salts was initially attempted using the same method. This entailed the reaction of a partially dissolved aqueous solution of the free base with a 10 % excess of the selected acid. Additional water was added where appropriate. The resulting solution was stirred for 30 minutes with gentle heating to 50 $^{\circ}$C. The solution was then filtered and left to produce crystals by slow evaporation and cooling (Crystallisation method 1). Where unsatisfactory single crystals were obtained, recrystallisation or re-synthesis was attempted using one of the following methods.

3.2.2 Crystallisation method 2 - reduced evaporation from water

The unsatisfactory crystals obtained from crystallisation method 1 were re-dissolved in a minimal quantity of water and the resultant solution filtered into a narrow tube of approximate diameter 5 mm, to allow for slower evaporation and crystal growth.

3.2.3 Crystallisation method 3 – Diffusion experiments

The unsatisfactory crystals obtained from crystallisation method 1 were re-dissolved in a minimal quantity of an appropriate good solvent and filtered into a tube with approximate diameter 5 mm, which was then placed inside a vial that contained 1-2 cm^3 of poor solvent. The vial was sealed and the poor solvent allowed to diffuse into the good solvent to promote satisfactory crystal growth. Table 3.4 below lists the combinations of good and poor solvents used.

Table 3.4 List of good and poor solvents used in diffusion experiments

Good solvent	Poor solvent
Ethanol	Diethylether
Ethanol	Hexane
Water	Tetrahydrofuran
Water	Ethanol
Water	Methanol
Water	Acetone

3.2.4 Crystallisation method 4 – Synthesis with ethanol as solvent

The synthesis of MAMBARTar, MEpdRTar and MEpd4HB produced thick oils when attempted using water as the solvent. These salts samples were synthesised again in an ethanoic medium. Alternatively, the free base was first partially dissolved in ethanol, the free acid was then added in a quantity of 10 % excess. The solution was left to stir for 30 mins with gentle heating. The resultant solution was then filtered into two containers (a 2 dram vial and a small test tube with approximate diameter 5 mm) and left to crystallise.

3.2.5 Crystallisation method 5 - Recrystallisation with ethanol

The unsatisfactory crystals obtained from crystallisation method 1 were re-dissolved in a minimal quantity of ethanol and the resultant solution filtered into both a 2 dram sample vial and a narrow tube of approximate diameter 5 mm.

3.2.6   Crystallisation method 6 – Reaction with methanol:toluene as solvent

For some halide samples no reaction seemed to take place with water as the solvent. These experiments were repeated using a 50:50 methanol:toluene mixture as the solvent. The free base was first partially dissolved in a 50:50 methanol:toluene mixture, to the stirring mixture a equimolar quantity of the HX acid was added. The solution was then stirred for 20 minutes. The resultant solution was filtered and left to slowly evaporate and crystallise.

3.2.7   Crystallisation method 7 – Preparation of phosphate salts by alternative method

As previous attempts using crystallisation method 1 had resulted in MAMBAPO4 and MEpdPO4 forming oils, the method used for the formation of bis(synephrine) monohydrogen phosphate monohydrate was followed.[8] The free base was partially dissolved in 90:10 ethylacetate:water mixture. To the stirring solution phosphoric acid was added in a 10 % excess. This yielded the formation of the salt as a white precipitate. A further volume of 90:10 ethylacetate:water was then added to dissolve the salt. The resulting solution was filtered into two vials, one was left to slowly evaporate and crystallise at room temperature and the other was capped and placed in the fridge.

3.2.8   Crystallisation method 8 – Preparation of halide salts in a 50:50 ethanol:water solvent

A selection of halide salts failed to produce satisfactory single crystals from the crystallisation method 1. The synthesis was attempted again in a 50:50 ethanol:water medium. The free base was first partially dissolved in a 50:50 ethanol:water mixture, to the stirring mixture a equimolar quantity of the HX acid was added. The solution was then stirred for 20 minutes. The resultant solution was then filtered and left to slowly evaporate and crystallise.

3.2.9 Crystallisation method 9 – Preparation of iodide salts using Schlenk techniques

Preparation of MEpdI and RMEpdI by crystallisation method 1 produced crystals of the composition $[Mepd]_2[I^-][I_3^-]$ and $[RMEpd][I^-][I_2]$ respectively. The presence of I_2 and I_3^- may be due to the oxidation of the iodide ion by oxygen present in the air. To attempt to stop this oxidation the salts were prepared under a nitrogen atmosphere. The free base was placed in a Schlenk tube and the vessel was evacuated and filled with nitrogen. The base was then partially dissolved in degassed water. Hydroiodic acid was then slowly added in an equimolar quantity to the stirring suspension. The Schlenk tube was then sealed and the solution left to stir overnight. The next day ethanol was added to disperse the iodine droplets that had formed. The solution was then stirred under nitrogen for a further hour and the excess solvent removed under vacuum. The remaining solution was left to crystallise under nitrogen.

3.2.10 Crystallisation method 10 – Heating and cooling cycles

Several samples that had become viscous oils after synthesis by crystallisation method 1 were subjected to a series of heating and cooling cycles. This was an attempt to agitate the oil and induce nucleation and crystallisation. The samples were first heated to approximately 50 ^{o}C and then left to cool slowly. This process was then repeated up to five times. Where no crystallisation had occurred after these cycles a small amount of diethylether was added dropwise to try increase mobility and induce nucleation.

3.3 Preparation of salts and results achieved

Salt synthesis was attempted on 443 new acid-base combinations. From this 247 samples produced crystals suitable for analysis by single crystal X-ray diffraction. After analysis it was determined that 238 produced crystals of a salt form and nine gave an 'unwated' product of the free acid or free base. These 238 salt structures were added to 19 corresponding salt structures already present within the literature[1,9-]

[19] to give a database of 257 salt structures. The tables detailing the synthesis of all salts used to create the systematic structural database along with details of the crystallisation method used and the outcome of the synthesis can be found in Appendix A.

3.4   Single crystal X-ray diffraction

3.4.1   Data collection at the University of Strathclyde

Measurements were recorded at low temperature by Gemini or Xcalibur Oxford Diffraction diffractometers with monochromated Mo Kα radiation ($\lambda = 0.7107$ Å), or Cu Kα radiation ($\lambda = 1.5418$ Å). Data was processed, scaled and corrected by the Oxford Diffration CrysAlis software.[20] Full details of these and all other data collections and refinements are given in the cif files presented in the Appendix E.

3.4.2   Data collection by the National Crystallography Service, University of Southampton (NCS)

Measurements were recorded at low temperature by Bruker-Nonius CCD diffractometer with Mo Kα radiation ($\lambda = 0.7107$ Å). Data was processed using the APEX2, SAINT and SADABS software.[21-23]

3.4.3   Data collection at the synchrotron

Measurements on small and poor quality crystals were made by the NCS at station I19 at the synchrotron facility at Diamond Light Source, typically at 120 K. Data was collected using a Crystal Logics kappa-geometry diffractometer and a Rigaku Saturn 724+ CCD detector. Rigatu Crystal Clear was used to record images,[24] and the data was then transformed to Bruker format to allow processing via SAINT and SADABS.[22,23] The synchrotron X-ray wavelength was 0.6889 Å.

3.4.4 Structure solution and refinement

Structure solution used programs from the SHELXS or SIR families.[25-27] Refinement of non-hydrogen atomic co-ordinates with anisotropic thermal parameters was to convergence and by the full-matrix least-squares method on F^2 as implemented by SHELX-97[27] within the WinGX interface.[28] All H atoms attached to carbon were placed in geometric positions and refined in riding modes. Where possible, H atoms of potentially hydrogen-bonding groups (e.g. OH and NH) were placed as found by difference synthesis and refined freely. However, where this was not possible suitable restraints were used. Full details of this treatment and of the treatment of other effects such as disorder and twinning are presented in the cif files of Appendix E. All structures were checked with PLATON[29] and the cifcheck routine available from IUCr website. Molecular graphics were prepared using ORTEP-3 and Mercury.[30,31]

3.5 X-ray powder diffraction

3.5.1 Data collection

X-ray powder diffraction was used to check the phase and purity of the bulk materials before and after solubility analysis. The samples were first ground to achieve a fine powder which was analysed to check the purity of the bulk sample. After the saturated solution had reached equilibrium the wet powder recovered by filtration was analysed to check the phase of the material for which solubility was recorded. These measurements were made with a Bruker D5000 diffractometer operating in reflection mode and using copper Kα radiation ($\lambda = 1.5418$ Å). The data was collected at room temperature with a 2θ range of 4.000 ^{o} to 35.056 ^{o}. The data refinement was performed using Pawley fitting in DASH.[32]

3.5.2 X-ray powder diffraction results

Appendix B shows the unit cell parameters derived from the single crystals X-ray data and the X-ray powder data refinements for both the dry and wet powder

samples. The dry samples were examined prior to the solubility slurry experiment and the wet samples were examined after the solubility slurry experiment. Where the dry and wet unit cell parameters match those of the single crystal, the measured solubility was deemed to be that of the single crystalline phase and only then was it used in further analysis. Note that as the single crystal unit cells were obtained at low temperature and the powder diffraction cells at room temperature, some disparity between the two is expected.

3.6   Melting point determination

Melting points were collected in triplicate using a Buchi B-545 automatic melting point apparatus. The system was set to produce a melting point value instead of a range, producing three melting point values. The average measured melting point values obtained are presented in Table 3.5 below.

Table 3.5 Melting point of salts (stated is average of three measurements ^{o}C)

	MAMBA	Mepd	RMEpd	PEA	MPEA	DMPEA	PPA	HPEA	TYR	Epd	Pepd
2AB		106.3	106.1						148.6	124.8	
2CB		131.7	131.0		122.6				180.6	150.6	95.2
2FB	100.2	132.6	105.6						188.0		
2HB		130.0			64.5	87.2					126.8
2NB	103.8	80.1	103.2						135.9	176.9	103.1
3AB		142.1									
3CB	98.5	75.5	135.1	91.3				130.9	167.2	131.9	
3FB		138.2	116.0	103.0					161.2		
3HB			143.2	132.9					204.5		75.7
3NB			157.0	159.3					181.9		145.0
4AB		130.5	131.0	162.3				174.5	171.7	159.3	
4CB	126.2	177.5	163.5				120.7	154.6	134.7	145.6	92.8
4FB	74.0		133.9						163.8		76.0
4HB	181.5	160.3	142.3	118.0	176.0	100.2		194.9	176.5		
4HBS	149.2	147.6	120.7	84.7	102.7		121.3		105.5	122.3	179.8
4NB		161.7	168.0	167.3	128.4	58.4			177.3	152.3	168.0
Adp		95.0	110.3	146.5	86.6	78.8	153.4		115.9		
BF4		112.2		257.6							
Br	98.2	177.0	184.0	269.0					249.4	207.5	182.1
BS	116.3	111.1	103.9			86.9	102.5			172.7	120.2
Bz	47.6	83.2	115.1	91.0					167.8		
Cl		194.0	211.8	221.1					276.6	217.6	185.6

	MAMBA	Mepd	RMEpd	PEA	MPEA	DMPEA	PPA	HPEA	TYR	Epd	Pepd
ClO4											
EDS		102.2	98.4	373.9				359.4	295.3	120.2	77.0
EtSO3			208.2	118.5					9.0	147.0	
Fum				203.3			174.9		175.0		
I		85.5	140.8								
Lmal					133.0				129.8		
Male		125.7	131.3	143.5	79.0		78.9		158.7	94.0	125.4
Malon	108.3	113.6	114.9	96.6	84.7	97.9	101.6		128.7		
MeSO3				209.9			113.1		208.0		
MLD	37.2	123.2	88.9	127.9			118.3	152.6	166.6	165.0	
Mtar		83.6			127.2	51.6			164.0	77.5	171.3
mTol			110.3	103.1					151.3	119.4	
oTol		100.7	132.7	42.0				48.0	175.4		77.6
PO4											
pTol	121.0	160.0	142.9	109.0					138.2		
Rmal				152.3	107.4	88.5	69.0		68.1		124.8
RMD	60.0	102.9	120.5	139.0	101.8		137.5	167.2		110.3	
Rtar			128.3	165.8			156.9		172.3		94.1
SO4		82.4	90.6	44.9					88.1		95.8
Suc	108.0	102.8	85.0	166.9					202.6		

3.7 Solubility measurements

3.7.1 Preparation of samples for solubility measurements

Powder samples of the salts were checked for purity by powder diffraction, see Section 3.5 above. Approximately 0.5 g of each salt was used to produce saturated aqueous solutions (typically with 1 to 2 cm^3 of deionised water). These slurries were stirred in an incubator at 25 ^{o}C for three days to ensure equilibrium had been reached. The saturated solution was then extracted from the slurry by centrifugal methods (6000 rpm for 10 mins).

The solubility was determined by measuring the UV absorbance of the cation. A calibration curve was established for each free base dissolved in DMSO, within the linear domain of the Beer-Lambert law, which contained 5 points within the 0 to 1 absorbance domain to establish a linear relationship between the absorbance of the cation and the molarity, with a R^2 >99.9 %. (For each base an independent calibration curve using salt forms rather than the free base was also constructed and comparison of the two methods showed that they gave essentially identical results). The weighing and dilution of the calibration was checked with the preparation of quality control standards at 100 % and 10 % with an acceptability limit of 10 %. All salt solubility measurements were obtained in duplicate by diluting the saturated solution (obtained above) with DMSO until the absorbance fell within the linear response of the calibration curve. (DMSO was used to prevent potential problems with salting out). This allowed for the calculation of the cationic molarity of the saturated solution. Table 3.6, Table 3.7 and Table 3.8, below, show the average values achieved for the solubility measurements in mol/L of the cation and the associated standard deviations.

Table 3.6 Solubility results for salts of (-)methylephedrine, (+/-)methylephedrine, dimethylphenethylamine and (-)ephedrine

Cation →	MEpd		RMEpd		DMPEA		Epd	
Anion ↓	Solubility (mol/L base)	Standard Deviation	Solubility mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation
2AB	2.354	0.033	1.050	0.110			1.417	0.007
2CB	0.542	0.002	0.434	0.001			0.109	0.001
2FB	1.235	0.008	2.318	0.005				
2HB	0.190	0.000			3.125	0.081	2.579	0.196
2NB	0.313	0.019	0.149	0.001			0.096	0.001
3AB	1.621	0.037						
3CB	0.176	0.000	0.138	0.003			0.133	0.008
3FB	0.344	0.006	0.524	0.010				
3HB			0.496	0.014	2.164	0.129		
3NB			0.047	0.001			0.124	0.001
4AB	1.499	0.135	1.687	0.098			1.081	0.055
4CB	0.052	0.003	0.035	0.001			0.112	0.000
4FB			0.201	0.003				
4HB	0.440	0.006	0.885	0.009	2.575	0.096		
4HBS	1.173	0.134	1.185	0.035			0.055	0.001
4NB	0.128	0.002	0.066	0.005	2.753	0.053	0.137	0.014
Adp			1.704	0.018	2.570	0.014		
BF4	2.461	0.005						
Br	0.596	0.004	0.646	0.047			0.603	0.005
BS	4.523	0.258	2.712	0.116	0.348	0.021	0.246	0.022
Bz	0.327	0.006	1.629	0.108				
Cl	2.443	0.024	1.349	0.022	4.349	0.143	1.258	0.034
ClO4	0.396	0.001			4.206	0.035		
EDS	3.552	0.072	0.527	0.007			0.488	0.015

Cation →	MEpd		RMEpd		DMPEA		Epd	
Anion ↓	Solubility (mol/L base)	Standard Deviation	Solubility mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation
EtSO3	2.947	0.088	1.373	0.021			2.441	0.157
Base	0.048	0.001	0.102	0.001			0.396	0.001
Fum								
I	0.323	0.004	0.495	0.014			2.205	0.086
LMal								
LMD	1.821	0.063	2.558	0.008			0.222	0.002
LTar	2.521	0.044			1.428	0.043	1.679	0.031
Male	0.939	0.029	1.270	0.031	3.297	0.105	2.266	0.081
Malon	3.957	0.194	3.623	0.198	3.390	0.029		
MeSO3	1.531	0.030	3.982	0.060			2.239	0.073
mTol			1.387	0.035			0.121	0.001
oTol	1.663	0.102	0.286	0.002				
PO4								
pTol	0.128	0.002	0.109	0.002				
RMal					2.735	0.130		
RMD	3.415	0.083	4.132	0.072			1.670	0.072
RTar			3.774	0.308				
SO4	0.918	0.034	1.463	0.073			1.163	0.004
Suc	2.287	0.037	1.943		3.055	0.208		

Table 3.7 Solubility results for salts of hydrxyphenethylamine, α(methylaminomethyl)benzyl alcohol, methylphenethylamine and phenethylammine

Cation →	HPEA		MAMBA		MPEA		PEA	
Anion ↓	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation
2AB								
2CB					1.428	0.016		
2FB			2.767	0.109				
2HB					3.435	0.187		
2NB	0.542	0.013	1.387	0.174				
3AB								
3CB	0.265	0.001	1.784	0.062			3.132	0.140
3FB							3.081	0.001
3HB							2.573	0.014
3NB	0.356	0.009					0.070	0.001
4AB	0.373	0.001					0.607	0.002
4CB	0.306	0.001	0.189	0.013			0.139	0.010
4FB			0.370	0.036			2.875	0.149
4HB	0.458	0.022	0.281	0.003	0.135	0.006	0.462	0.003
4HBS			2.986	0.088	2.514	0.071	0.942	0.002
4NB					2.197	0.003	0.073	0.001
Adp					3.014	0.175	1.012	0.073
BF4							3.035	0.063
Br			6.084	0.202	2.726	1.126	3.076	0.007
BS			2.594	0.037				
Bz			1.888	0.176			3.953	0.242
Cl			2.743	0.261			3.602	0.251
ClO4							3.721	0.270

Cation →	HPEA		MAMBA		MPEA		PEA	
Anion ↓	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation
EDS	0.551	0.031	2.402	0.109	2.658	0.015	0.451	0.006
EtSO3							6.384	0.937
Base	9.558	0.254	0.594	0.000				
Fum							0.089	0.001
I			6.906	0.278	3.414	0.029	3.369	0.046
LMal					0.868	0.025		
LMD	1.063	0.004	1.713	0.089			0.978	0.005
LTar					1.330	0.048		
Male	0.443	0.013			2.954	0.011	0.207	0.001
Malon			4.371	0.153	3.693	0.156	5.512	0.402
MeSO3			1.978	0.064			1.340	0.011
mTol							0.281	0.002
oTol	0.885	0.043					0.603	0.070
PO4	1.010	0.020			1.970	0.024		
pTol	0.328	0.008	3.030	0.100			0.617	0.003
RMal	1.085	0.077			1.528	0.008	0.776	0.028
RMD	0.287	0.002	3.425	0.064	0.792	0.010	0.314	0.004
RTar							0.578	0.001
SO4							5.940	0.794
Suc			2.334	0.035			1.179	0.026

Table 3.8 Solubility results for salts of (-)pseudoephedrine, phenylpropylamine and tyramine

Cation →	PEpd		PPA		TYR	
Anion ↓	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation
2AB					0.932	0.058
2CB	0.159	0.016			0.223	0.004
2FB					0.252	0.005
2HB	0.173	0.000				
2NB	0.546	0.038			1.684	0.095
3AB						
3CB					0.142	0.004
3FB					0.220	0.005
3HB	0.286	0.002			0.125	0.001
3NB	0.160	0.005			0.127	0.002
4AB					0.152	0.001
4CB	0.159	0.002	0.147	0.018	0.612	0.002
4FB	0.848	0.067			0.400	0.004
4HB					1.407	0.004
4HBS	0.128	0.005	0.625	0.011	0.519	0.010
4NB	0.090	0.002			0.860	0.006
Adp	2.200	0.070	0.297	0.010	2.841	0.139
BF4					3.964	0.031
Br	1.351	0.004			2.161	0.072
BS	2.695	0.321	3.750	0.225		
Bz					0.347	0.001
Cl	2.923	0.105			1.116	0.008
ClO4	1.274	0.052			4.767	0.137
EDS	2.600	0.176			0.690	0.020

Cation →	PEpd		PPA		TYR	
Anion ↓	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation	Solubility (mol/L base)	Standard Deviation
EtSO3					2.045	0.019
Base	0.049	0.000			0.041	0.002
Fum			0.230	0.003	0.356	0.003
I	2.466	0.162			3.186	0.070
LMal					1.850	0.026
LMD			0.756	0.010	0.957	0.041
LTar	0.294	0.011			0.576	0.009
Male	1.882	0.090	1.807	0.058	1.041	0.061
Malon			2.889	0.116	1.980	0.140
MeSO3			3.413	0.200	1.854	0.034
mTol					0.329	0.002
oTol	0.324	0.023			0.319	0.010
PO4					1.074	0.003
pTol					0.439	0.001
RMal	1.947	0.178	2.889	0.116	1.753	0.053
RMD			0.296	0.009	0.978	0.022
RTar	0.839	0.009	0.331	0.006	0.455	0.021
SO4	0.697	0.083			5.641	0.120
Suc					0.088	0.000

3.7.2 UPLC conditions

Solubility analysis was performed using an Agilent 1200 UPLC, with a Waters X-Bridge column (C18, 5 μm, 2.1 x 50 nm). The system was set up with a column temperature of 60 ^{o}C and a gradient mobile phase at a flow rate of 1.00 ml/min, commencing with phase A (0.1 % NH_4 + H_2O, milli Q) and phase B (0.1 % NH_4 + ACN, HPLC grade) at 95 % and 5 % respectively for 0.3 minutes. The gradient was then introduced over a period of 2 minutes to reach a final ratio of 5 % A and 95 % B, which was held for a further 0.3 minutes to ensure all analytes had eluted form the column. The quantitative analysis of cation in the salt sample was determined using a UV-Vis detector at both 230 nm and 254 nm. The data was collected and processed using Chromeleon software.[33]

3.8   Structure analysis

3.8.1   Graph set notation and analysis

Graph set analysis[34,35] was used throughout this work to compare and contrast hydrogen bonding motifs. This was achieved using the graph set utility in Mercury CSD 2.3.[31] Parameters were chosen to find graph sets up to level two, with a maximum ring size of six hydrogen bonds and a maximum chain and discrete size of four hydrogen bonds.

3.8.2   Structure crystal packing similarity

The 'crystal packing similarity' module of Mercury CSD 2.3[31] was used to investigate similarities in the crystal packing throughout the dataset. This is achieved by taking a reference molecule and then examining the 3D geometry of the cluster of surrounding molecules. This cluster can then be compared with clusters from other structures to investigate their geometric similarity. A good example of the detailed use of this module is given by Childs et al.[36]

3.9   Calculations from structural information

3.9.1   Molecular Orbital PACkage

MOPAC (Molecular Orbital PACkage) is a semi-empirical quantum chemistry program that performs molecule based gas-phase calculations based on the Dewar and Thiel's NDDO approximation.[37] The calculations were set up using the AM1 Hamiltonian as this works with all atom types within the structures except boron.[38] The calculations were run using gas-phase molecular geometry optimisation and calculated the heat of formation and atomic partial charges, along with electronic and nuclear energies for all the structures. The obtained values were then fed into a structure-property correlation program to assist in predicting solubility measurements.

3.9.2   Pixel calculations

Pixel calculations are carried out as a submodule of the OPiX package and were used to calculate intermolecular interaction energies of crystals.[39-41] For this method all the charge density of the molecule is taken into account by using charge density distributions, as calculated by GAUSSIAN running in the CUBE mode with the MP2 level of theory and 6-31G** basis set.[42] Pixel calculations were only performed on salts that contain two molecular components (e.g. all hydrated salts and salts with Z'>1 were ignored) as salts with more than two molecular components encounter difficulties when processing. A positive charge was initially assigned to the proton attached to the nitrogen of the cation and the negative charge of the anion was shared equally between the two oxygen atoms of the carboxylate group or the three oxygen atoms of the sulfonate group or placed solely on the halide anion. (Several trial calculation using other initial charge distributions gave only small differences in the eventual outputs). The output from the Pixel calculations was used to help predict solubility measurements by inputting the obtained energies into the structure-property correlation programs.

3.10 Structure-property correlations

3.10.1 Random forest

Random forest[43] is a library package of the statistical computing program 'R'v2.11.1.[44] It is a classification and regression tool which is used herein to build models for chosen physical properties using a selection of calculated and measured molecular descriptors. The calculated molecular descriptiors were deliberated using Molecular Operating Enviroment (MOE)[45] from .mol files from ChemDraw. Except where otherwise noted, the corrected input value for each parameter was that of corresponding free base minus the corresponding free acid (Input = base value − acid value). The Randon forest optimal parameters were selected from the following; the default value of *mtry* was used, based upon the number of input parameters. *Ntree*, the number of trees grown, was increased steadily until no further improvement was seen in the model.

3.11 References

(1)   Collier, E. A.; Davey, R. J.; Black, S. N.; Roberts, R. J. *Acta. Cryst.* **2006**, *A62*, 498.
(2)   Li, Z. J.; Ojala, W. H.; Grant, D. J. W. *J. Pharm. Sci.* **2001**, *90*, 1523.
(3)   Li, Z. J.; Grant, D. J. W. *J. Pharm. Sci.* **1997**, *86*, 1073.
(4)   Zingg, S. P.; Arnett, E. M.; McPhail, A. T.; Bothnerby, A. A.; Gilkerson, W. R. *J. Am. Chem. Soc.* **1988**, *110*, 1565.
(5)   Ivanova, B.; Kolev, T.; Lamshoft, M.; Mayer-Figge, H.; Seidel, R.; Sheldrick, W. S.; Spiteller, M. *J. Mol. Struct.* **2010**, *971*, 8.
(6)   Cooke, C. L.; Davey, R. J.; Black, S.; Muryn, C.; Pritchard, R. G. *Crystal Growth & Design* **2010**, *10*, 5270.
(7)   Yalkowsky, S. H.; He, Y. *Handbook of aqueous solubilitry data*; CRC Press: Florida, 2003.
(8)   Dattagupta, J. K.; Meyer, E. F.; Mukhopadhyay, B. P. *Acta Crystallographica Section B-Structural Science* **1982**, *38*, 2830.
(9)   Barlow, R. B.; Johnson, O.; Howard, J. A. K.; Walton, D. C.; Koellner, G. *Acta Crystallographica Section B-Structural Science* **1989**, *45*, 396.
(10)   Bergin, R. *Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry* **1971**, *B 27*, 381.
(11)   Duddu, S. P.; Grant, D. J. W. *Pharm. Res.* **1994**, *11*, 1549.
(12)   Haynes, D. A.; Pietersen, L. K. *Crystengcomm* **2008**, *10*, 518.
(13)   Hearn, R. A.; Bugg, C. E. *Acta Crystallographica Section B-Structural Science* **1972**, *28*, 3662.
(14)   Hearn, R. A.; Freeman, G. R.; Bugg, C. E. *J. Am. Chem. Soc.* **1973**, *95*, 7150.

(15) Horn, E.; Tiekink, E. R. T.; Jones, G. P.; Naiola, B. P.; Paleg, L. G. *Acta Crystallographica Section C-Crystal Structure Communications* **1990**, *46*, 1575.

(16) Koleva, B. B.; Kolev, T.; Seidel, R. W.; Spiteller, M.; Mayer-Figge, H.; Sheldrick, W. S. *J. Mol. Struct.* **2008**, *888*, 138.

(17) Mathew, M.; Palenik, G. J. *Acta Crystallographica Section B-Structural Science* **1977**, *33*, 1016.

(18) Podder, A.; Dattagupta, J. K.; Saha, N. N.; Saenger, W. *Acta Crystallographica Section B-Structural Science* **1979**, *35*, 649.

(19) Rademeyer, M. *Acta Crystallographica Section E-Structure Reports Online* **2007**, *63*, O221.

(20) Oxford Diffraction; Oxford Diffraction Ltd: Abington, England, 2006.

(21) APEX2; Bruker AXS Inc.: Madison, Wisconsin, USA, 2007.

(22) SAINT; Bruker AXS Inc.: Madison, Wisconsin, USA, 2007.

(23) Sheldrick, G. M. University of Göttingen, Germany, 1996.

(24) Rigaku Corporation; The Woodlands: Texas, U.S.A, 2008.

(25) Altomare, A.; Cascarano, G.; Giacovazzo, C.; Guagliardi, A.; Burla, M. C.; Polidori, G.; Camalli, M. *J. Appl. Crystallogr.* **1994**, *27*, 435.

(26) Altomare, A.; Burla, M. C.; Camalli, M.; Cascarano, G. L.; Giacovazzo, C.; Guagliardi, A.; Moliterni, A. G. G.; Polidori, G.; Spagna, R. *J. Appl. Crystallogr.* **1999**, *32*, 115.

(27) Sheldrick, G. M. *Acta. Cryst.* **2008**, *A64*, 112.

(28) Farrugia, L. J. *J. Appl. Crystallogr.* **1999**, *32*, 837.

(29) Spek, A. L. *J. Appl. Crystallogr.* **2003**, 7.

(30) Farrugia, L. J. *J. Appl. Crystallogr.* **1997**, *30*, 565.

(31) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41*, 466.

(32) Friscic, T.; Fabian, L.; Burley, J. C.; Reid, D. G.; Duer, M. J.; Jones, W. *Chem. Commun.* **2008**, 1644.

(33) Harned, W. H. *Capturing and Reporting Electronic Data* **2002**, *824*, 66.

(34) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.

(35) Etter, M. C.; Macdonald, J. C.; Bernstein, J. *Acta. Cryst.* **1990**, *B46*, 256.

(36) Childs, S. L.; Wood, P. A.; Rodriguez-Hornedo, N.; Reddy, L. S.; Hardcastle, K. I. *Cryst. Growth Des.* **2009**, *9*, 1869.

(37) Guerrieri, P.; Rumondor, A. C. F.; Li, T.; Taylor, L. S. *AAPS PharmSciTech* **2010**, *11*, 1212.

(38) Kozma, D.; Bocskei, Z.; Simon, K.; Fogassy, E. *Journal of the Chemical Society-Perkin Transactions 2* **1994**, 1883.

(39) Gavezzotti, A. *J. Phys. Chem. B* **2002**, *106*, 4145.

(40) Gavezzotti, A. University of Milano, 2003.

(41) Gavezzotti, A. *J. Phys. Chem. B* **2003**, *107*, 2344.

(42) Frinch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyoto, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H.

P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Pittsburgh PA, 2003.

(43)   Liaw, A.; Wiener, M. *R News* **2002**, *2*, 18.

(44)   R Development Core Team; 2.11.1 ed.; R Foundation for Statistical Computing: Vienna, Austria, 2006.

(45)   MOE Chemical Computing Group; http://www.chemcomp.com: Quebec, Canada, 2002.

Chapter 4

Structural comparison of enantiopure and racemic methylephedrinium salts

4.1 Introduction

Salt synthesis was attempted with both the enantiopure and racemic forms of the base methylephedrine and 42 acids, (See Table 3.1 and Table 3.2). Of the possible 84 acid–base combinations 69 produced crystals suitable for analysis by single-crystal X-ray diffraction. It was determined that four of these combinations produced only crystals of the free acid, namely MEpd4FB, MEpdFum, RMEpd2HB and RMEpdFum despite a pK_a difference of more than three between the free acid and free base. The remaining 65 acid-base combinations all produced crystals of the desired salt forms.

Within these 65 salt forms there are 28 pairs of structures where both the enatiopure and racemic bases gave crystalline products with the same acid. Of importance are the 16 enantiopure/racemic pairs with chemically identical formulae. The remaining 12 pairs consist of six which are chemically different, mostly in their hydration state but in one case through the formation of tri-iodide rather than iodide counterions and, in the case of the succinate salts, through the formation of the 2:1 cation:anion salt for the racemic product but the 1:1 cation:anion salt for the enantiopure product. There are also four cases where the racemic salt spontaneously resolves to form a conglomerate. Finally, the mandelate and tartrate salts are omitted from the comparative discussion of enantiopure versus racemic structures because of the extra complication of the second stereocentre on the anion. For the record, the four acid-base combinations of the enantiopure and racemic methylephedrine bases with enantiopure and racemic mandelic acids attempted produced three different crystal structures, two of which were enantiopure and one of which was racemic. The tartrate equivalent produced two structures, one of which was enantiopure and another which was racemic. Interestingly the racemic tartrate product formed a co-crystal.

Four of the racemic salts each produced two different crystal phases that were analysed. The salts of 1,2-ethanedisulfonate and benzenesulfonate both gave an

anhydrous and monohydrated salt form. (+/-)Methylephedrinium iodide formed as a pure iodide salt and as a salt with both an iodide ion and a di-iodine molecule present, and the methylsulfonate salt produced two different polymorphs of the same salt. Finally, the dataset to be analysed consists of one final group - that of the nine salt forms where only one of the pair of enantiopure-racemic crystal structures has been obtained and analysed. Collectively, the above gives a total of 64 unique crystal structures that will be studied. We have previously published a partial analysis of this system consisting of a dataset of 37 crystal structures.[1]

There are several features general to all 64 independent crystal structures. All form contact cation-anion pairs through hydrogen bonds, but there are no direct cation-cation hydrogen bonded contact ion pairs. Anion-anion interactions are present in all structures with an anion that possesses a classic hydrogen bond donor except for the two hydrogen-sulfate salts. All salt structures conform to Etter's rule,[2] that predicts that intramolecular bonds will form in preference to intermolecular bonds; thus the hydrogen-maleate and hydrogen-malonate salts use their COOH groups as internal hydrogen bond donors to carboxylate, and the 2-aminobenzoate and 2-hydroxybenzoate salts similarly use their NH_2 and OH groups as donors. Of the 64 structures, 10 were isolated as monohydrates, one as a hemihydrate and one contains 0.3 water molecules per asymmetric unit. In all structures, except the three 1,2-ethanedisulfonate salts and (+/-)methylephedrinium succinate, the acid and base reacted to give a one to one, cation to anion, salt. In the four exceptions two to one, cation to anion, salts were produced. The only reported product to have a neutral free acid present is that of (+/-)methylephedrinium (+/-)tartrate tartaric acid. This compound contains a doubly deprotonated acid and a free acid molecule per asymmetric unit and can therefore be described as a co-crystal. The structure of (+/-)methylephedrinium iodide di-iodine can also be termed a co-crystal as its asymmetric unit contains neutral I_2 species.

To aid further discussion, the methylephedrinium structures will be separated into groups depending on their counterion type and initially discussed in terms of structure similarities through graph-set analysis, and hydrogen bonding network growth. The structures will also be analysed to determine different cation conformations and crystal packing similarities throughout the dataset. Finally the melting points and densities will be examined to see if the pairs of chemically identical salts conform to Wallach's rules.[3] Table 4.1 and Table 4.2 shows the codes for the compounds that will be used throughout this chapter along with their measured densities and melting points. Compounds with melting points recorded as n/a either have an unobtainable melting point or their powder diffraction data shows a phase change and therefore there is no guarantee that the measured melting point correlates to the obtained single crystal data.

Table 4.1 Codes, melting points and density measurements for investigated compounds

compound	code	Melting point (°C)	Density (g/cm⁻³)	compound	code	Melting point (°C)	Density (g/cm⁻³)
MEpdBz	1A	83.2	1.190	RMEpdBz2	1B	n/a	1.238
Mepd2AB	2A	106.3	1.259	RMEpd2AB	2B	n/a	1.224
MEpd2CB	3A	n/a	1.320	RMEpd2CB	3B	131.0	1.315
MEpd2FB	4A	132.6	1.297	RMEpd2FB	4B	n/a	1.297
MEpd2HB	5A	130.0	1.272	RMEpd2HB	5B	free acid	
MEpd2NB	6A	80.1	1.348	RMEpd2NB	6B	n/a	1.310
MepdoTol	7A	100.7	1.222	RMEpdoTol	7B	132.7	1.225
MEpd3AB	8A	142.1	1.279	RMEpd3AB	8B	no crystals	
MEpd3CB	9A	75.5	1.279	RMEpd3CB	9B	135.1	1.310
MEpd3FB	10A	138.2	1.295	RMEpd3FB	10B	116.0	1.292
MEpd3HB	11A	n/a	1.219	RMEpd3HB	11B	143.2	1.221
MEpd3NB	12A	no crystals		RMEpd3NB	12B	157.0	1.338
MEpdmTol	13A	no crystals		RMEpdmTol	13B	110.3	1.210
MEpd4AB	14A	130.5	1.227	MEpd4AB	14B	131.0	1.252
MEpd4CB	15A	177.5	1.333	RMEpd4CB	15B	163.5	1.333
MEpd4FB	16A	free acid		RMEpd4FB	16B	133.9	1.262
MEpd4HB	17A	160.3	1.217	RMEpd4HB	17B	142.3	1.217
MEpd4NB	18A	161.7	1.296	RMEpd4NB	18B	168.0	1.323
MEpdpTol	19A	160.0	1.247	RMEpdpTol	19B	142.9	1.247
MEpdEDS	20A	102.2	1.322	RMEpdEDS	20B	n/a	1.303
MEpdBS	21A	n/a	1.274	RMEpdBS	21B	n/a	1.311
MEpd4HBS	22A	147.6	1.402	RMEpd4HBS	22B	120.7	1.346

compound	code	Melting point (°C)	Density (g/cm⁻³)	compound	code	Melting point (°C)	Density (g/cm⁻³)
MEpdSO4	23A	n/a	1.360	RMEpdSO4	23B	n/a	1.402
MEpdMeSO3	24A	n/a	1.274	RMEpdMeSO3	24B	n/a	1.277
MEpdEtSO3	25A	n/a	1.301	RMEpdEtSO3	25B	no crystals	
MEpdMale	26A	125.7	1.259	RMEpdMale	26B	131.3	1.285
MEpdMalon	27A	113.6	1.271	RMEpdMalon	27B	114.9	1.315
MEpdAdp	28A	95.0	1.241	RMEpdAdp	28B	110.3	1.264
MEpdSuc	29A	102.8	1.289	RMEpdSuc	29B	85.0	1.242
MEpdLTar	30A	83.6	1.344	RMEpdLTar	30B	no crystals	
MEpdRTar	31A	no crystals		RMEpdRTar	31B	128.3	1.362
MEpdLMD	32A	n/a	1.261	RMEpdLMD	32B	n/a	1.271
MEpdRMD	33A	102.9	1.261	RMEpdRMD	33B	120.5	1.270
MEpdBr	34A	177.0	1.422	RMEpdBr	34B	184.0	1.474
MEpdCl	35A	194.0	1.241	RMEpdCl	35B	211.8	1.237
MEpdI	36A	85.5	1.937	RMepdI	36B	n/a	1.681
MEpdBF4	37A	112.2	1.354	RMEpdBF4	37B	no crystals	

Table 4.2 Codes, melting points and density measurements for investigated compounds

compound	code	Melting point (°C)	Density (g/cm⁻³)
RMEpdEDS2	20C	98.4	1.349
RMEpdBS2	21C	103.9	1.329
RMEpdMeSO32	24C	n/a	1.331
MEpdLMD2	32C	123.2	1.271
RMEpdI2	36C	140.8	2.032

4.2 Substituted benzoate salts

There are eight salt pairs with benzoic acid derived counterions which crystallise to produce chemically identical species for (1R,2S)(-)methylephedrinium and (+/-)methylephedrinium. The graph-set analysis for these eight pairs is shown in Table 4.3. The notation used for the naming of the methylephedrinium salts is explained in Table 4.1 and Table 4.2. A common feature in all but one (11A) of the structures is the $C_2^2(9)$ motif, involving the cation's OH and NH groups as donors and both O-atoms of the anion COO⁻ group as the acceptors, see Figure 4.1. These 15 structures all use the $C_2^2(9)$ chain to give one dimensional hydrogen-bonded networks, either along the crystallographic *a, b* or *c* directions, see Figure 4.2. The only structure of the 16 that does not contain the $C_2^2(9)$ motif is that of (1R,2S)methylephedrinium 3-hydroxybenzoate. This structure contains two cations and two anions per asymmetric unit and the network grows via a series of discrete motifs and a $C_2^2(14)$ motif to produce a three dimensional hydrogen bonded network. It can be seen that here the addition of a hydrogen bond donor group to the anion leads to the hydrogen bonding becoming more complicated. In all cases in Table 4.3 where formation of higher dimensional networks is recorded, this is through such anion-anion contact pairs.
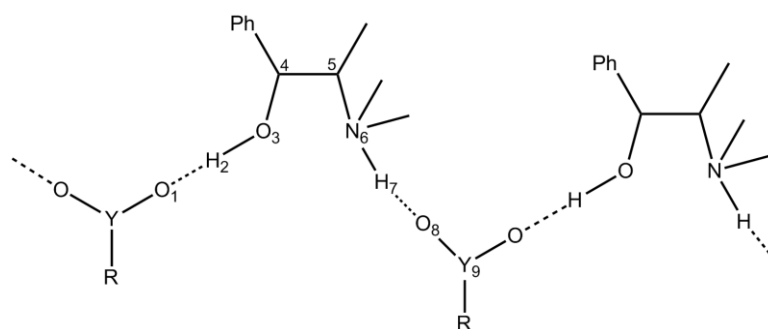
Figure 4.1 Common $C_2^2(9)$ graph-set motif. Y = C or SO

For the enantiopure-racemic pairs 2, 3, 6, 7 and 10 identical graph-sets are present throughout both structures. The enantiopure-racemic pair of 4-aminobenzoate, 14A and 14B, does not have identical graph-set, showing that the added presence of hydrogen bond donors, namely the NH$_2$ group of the anion, leads to more extensive and varied hydrogen bonding opportunities. This results in different ring and chain motifs being present and growth through a two dimensional sheet in the racemic

product compared to a three dimensional network in the enantiopure product. The enantiopure salts of 3-hydroxybenzoate and 4-nitrobenzoate (11A and 18A respectively) both contain two cations and two anions per asymmetric unit compared to their racemic salts which contain one of each. In the case of (1R,2S)methylephedrinium 4-nitrobenzoate, the carboxylic group of one anion forms hydrogen-bonds with the NH and OH of two different cations, whilst the crystallographically independent anion interacts with NH and OH functionalities of a single methylephedrinium. Thus the structure contains both ring, $R_2^2(9)$, and chain, $C_2^2(9)$, motifs, see Figure 4.3.
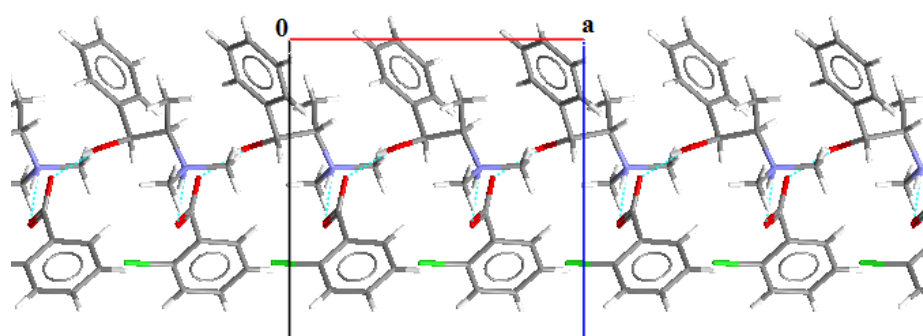
Figure 4.2 Propagation along the crystallographic *a* direction using a $C_2^2(9)$ motif in (1R,2S)methylephedrinium 2-chlorobenozte
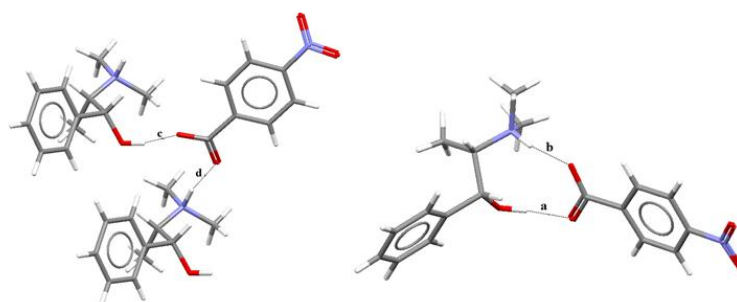
Figure 4.3 $C_2^2(9)$ and $R_2^2(9)$ motif observed in (1R,2S) methylephedrinium 4-nitrobenzoate

Kinbara *et al.* reported that they found significant differences in the hydrogen bonding motifs adopted between enantiopure and racemic primary amines and carboxylate anions.[4] We do not observe the same specific hydrogen bonding motif – nor significantly do we see any systematic difference between the enantiopure and racemic salts with respect to hydrogen bonding motifs. Except for the case of the 4-

aminobenzoate salts, the observed differences in hydrogen bonding motifs can be related to the number of cation-anion pairs per asymmetric unit and not to whether the salt is enantiopure or racemic.

Thirteen further "non-paired" benzoate based salt structures are available here for comparison. Four of these are from the species that produced conglomerates rather than racemic single crystals. Spontaneous resolution of the racemic product only occurred within the benzoate derived salts. Four spontaneous resolutions from 33 is slightly higher than the normally quoted 5 to 10% occurrence of the phenomena, but agrees with the suggestion that salts forms are more prone to spontaneously resolve than neutral molecules. Jacques et al.[5] states that conglomerates are more favourable for salts as the geometry imposed on structures by the relatively strong acid-base interaction, involved in the formation of a salt, may disfavour the formation of cetrosymmetric structure of the racemic compound.[5-7] Note that only five *para*-substituted benzoates were used here and that three of these five formed conglomerates. Only the *amino-* and *nitro-* derivative formed a racemic phase, whilst the *chloro-*, *hydroxyl-* and *methyl-* derivatives showed spontaneous resolution. The implication is that similarly shaped counterions all behaved in like manner with respect to spontaneous resolution.

Table 4.3 Graph-set analysis of substituted benzoate salts with chemically identical pairs for the enantiopure and racemic species of methylephedrine

compound		hydrogen bond	graph-set		growth	compound		hydrogen bond	graph-set		growth
2A	a	$OH_c \cdots COO^-$	c	$C_1^1(6)$	*ab*	**2B**	a	$OH_c \cdots COO^-$	c	$S_1^1(6)$	*c*
	b	$NH_c \cdots COO^-$	d	$S_1^1(6)$			b	$NH_c \cdots COO^-$	d	$S_1^1(6)$	
	c	$NH_a \cdots COO^-$	>a<b	$C_2^2(9)$			d	$NH_a \cdots COO^-$	>a<b	$C_2^2(9)$	
	d	$NH_a \cdots COO^-$	>a>c<a	$D_3^2(9)$			e	$NH_a \cdots COO^-$	>a<b	$C_2^2(9)$	
			>b>c<b	$D_3^3(11)$							
3A	a	$OH_c \cdots COO^-$	>a<b	$C_2^2(9)$	*a*	**3B**	a	$OH_c \cdots COO^-$	>a<b	$C_2^2(9)$	*a*
	b	$NH \cdots COO^-$					b	$NH \cdots COO^-$			
6A	a	$OH_c \cdots COO^-$	>b<c	$D_2^1(3)$	*a*	**6B**	a	$OH_c \cdots COO^-$	>b<c	$D_2^1(3)$	*b*
	b	$NH \cdots COO^-$	>a<c	$D_2^2(5)$			b	$NH \cdots COO^-$	>a<c	$D_2^2(5)$	
	c	$OH_w \cdots COO^-$	>a<d	$D_2^2(8)$			c	$OH_w \cdots COO^-$	>a<d	$D_2^2(8)$	
	d	$OH_w \cdots NO_2$	>b<d	$D_2^2(8)$			d	$OH_w \cdots NO_2$	>b<d	$D_2^2(8)$	
			>a<b	$C_2^2(9)$					>a<b	$C_2^2(9)$	
			>c<d	$C_2^2(9)$					>c<d	$C_2^2(9)$	
7A	a	$OH_c \cdots COO^-$	>a<b	$C_2^2(9)$	*b*	**7B**	a	$OH_c \cdots COO^-$	>a<b	$C_2^2(9)$	*b*
	b	$NH \cdots COO^-$					b	$NH \cdots COO^-$			
10A	a	$OH_c \cdots COO^-$	>a<b	$C_2^2(9)$	*a*	**10B**	a	$OH_c \cdots COO^-$	>a<b	$C_2^2(9)$	*b*
	b	$NH \cdots COO^-$					b	$NH \cdots COO^-$			
11A	a	$OH_c \cdots COO^-$	>a<f	$D_2^1(3)$	*abc*	**11B**	a	$OH_c \cdots COO^-$	c	$C_1^1(7)$	*c*
	b	$NH \cdots COO^-$	>b<e	$D_2^1(3)$			b	$NH \cdots COO^-$	>a<b	$C_2^2(9)$	
	c	$NH \cdots COO^-$	>c<d	$D_2^1(3)$			c	$OH_a \cdots COO^-$	>b>c<bf	$D_3^2(10)$	
	d	$OH_a \cdots COO^-$	>b<c	$R_1^2(4)$					>a>c<a	$D_3^3(12)$	
	e	$OH_c \cdots COO^-$	>a<g	$D_2^2(5)$							
	f	$NH \cdots COO^-$	>b<d	$D_2^2(5)$							
	g	$OH_a \cdots COO^-$	>c<e	$D_2^2(5)$							
			>d<e	$D_2^2(5)$							
			>f<g	$D_2^2(5)$							
			<a>b	$D_2^2(8)$							
			<a>c	$D_2^2(8)$							

compound	hydrogen bond		graph-set		growth	compound	hydrogen bond		graph-set		growth
			<e>f	$D_2^2(8)$							
			>a>d	$D_2^2(9)$							
			>b>g	$D_2^2(9)$							
			>c>g	$D_2^2(9)$							
			>e>g	$D_2^2(9)$							
			>f>d	$D_2^2(9)$							
			>d>g	$C_2^2(14)$							
14A	a	OH_c···COO^-	c	$C_1^1(8)$	*abc*	**14B**	a	OH_c···COO^-	c	$C_1^1(8)$	*ab*
	b	NH_c···COO^-	d	$C_1^1(8)$			b	NH_c···COO^-	d	$C_1^1(8)$	
	c	NH_a···COO^-	>c<d	$C_2^2(6)$			c	NH_a···COO^-	>a<b	$C_2^2(9)$	
	d	NH_a···COO^-	>a<b	$C_2^2(9)$			d	NH_a···COO^-	>b>c<b	$D_3^2(11)$	
			>c>d	$C_2^2(16)$					>b>d<b	$D_3^2(11)$	
			>a>c<a	$D_3^2(11)$					>a>c<a	$D_3^3(13)$	
			>b>d<b	$D_3^2(11)$					>a>d<a	$D_3^3(13)$	
			>a>d<a	$D_3^3(13)$					>c<d>c<d	$R_4^2(8)$	
			>b>c<b	$D_3^3(13)$					>c>d<c<d	$C_4^3(20)$	
			>c>d<c<d	$C_4^4(22)$					>c>d>c>d	$R_4^4(32)$	
			>c>c>d>c>c<d	$R_6^6(38)$					>c>c>d>c>c<d	$R_6^5(36)$	
			>c>d>d<c>d>d	$R_6^6(38)$					>c>d>d<c>d>d	$R_6^5(36)$	
18A	a	OH_c···COO^-	>c<d	$C_2^2(9)$	*a*	**18B**	a	OH_c···COO^-	>a<b	$C_2^2(9)$	*a*
	b	NH···COO^-	>a<b	$R_2^2(9)$			b	NH···COO^-			
	c	OH_c···COO^-									
	d	NH···COO^-									

In the four conglomerate forming cases 4, 15, 17 and 19 where only the enantiopure structure was accessible, the common $C_2^2(9)$ chain was observed as above in all but the *p*-toluate salt, see Table 4.4. In the *p*-toluate salt a similar one dimensional network can be seen but the network's propagation occurs through only one oxygen atom of the COO⁻ group resulting in a $C_2^1(7)$, see Figure 4.4.
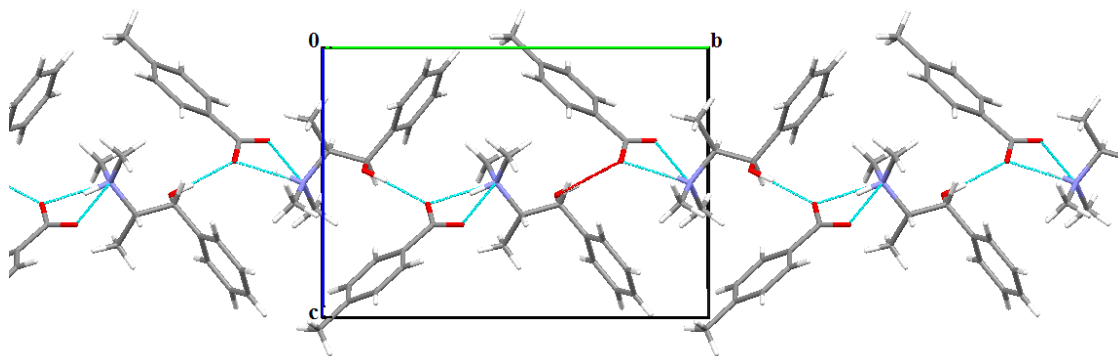
Figure 4.4 Propagation along the crystallographic *b* direction using a $C_2^1(7)$ motif in (1R,2S)methylephedrinium *p*-toluate

In a further four cases, enantiopure and racemic benzoate structures are available, but these salts did not form chemically identical pairs. The enantiopure species are monohydrates whilst the racemic compounds are anhydrous (see Table 4.4 compounds 1A, 1B, 9A and 9B) and as such identical hydrogen bonding motifs are not seen for the enantiopure-racemic pairs. Theory suggests that a possible function of water in hydrates is to 'fill' space.[8] If enantiopure species cannot pack as efficiently as racemic species, this would result in more 'space' and therefore more hydrates being present. The remaining three structures presented here do not consist of an enantiopure racemic pair as crystals of both forms were not obtained, (see Table 4.5 compounds 8A, 12B and 13B). However, within these structures there are still similarities to the previously discussed enantiopure-racemic pairs. All the above mentioned structures form one dimensional hydrogen bonded ribbons via the $C_2^2(9)$ chain as seen before, with the exception of (1R,2S)methylephedrinium 3-chlorobenzoate, 9A. This salt forms a two dimensional sheet with network growth through NH···water···COO⁻ in the crystallographic *b* direction forming a $D_2^2(4)$ graph-set and through solvent separated anions along the crystallographic *a* direction

forming a $C_2^2(6)$ chain, see Figure 4.5. This structure is thus different from all other benzoate salts studied herein, in that it does not form direct N-H···OOC hydrogen-bonds. Indeed 9A is even more of an outlier, in that all bar one of the other methylephedrinium salts reported here make direct hydrogen-bonds between NH and the formally charged group of the anion. The other exception is the benzenesulfonate 21B, see below, Sulfonate and hydrogen sulphate salts section.
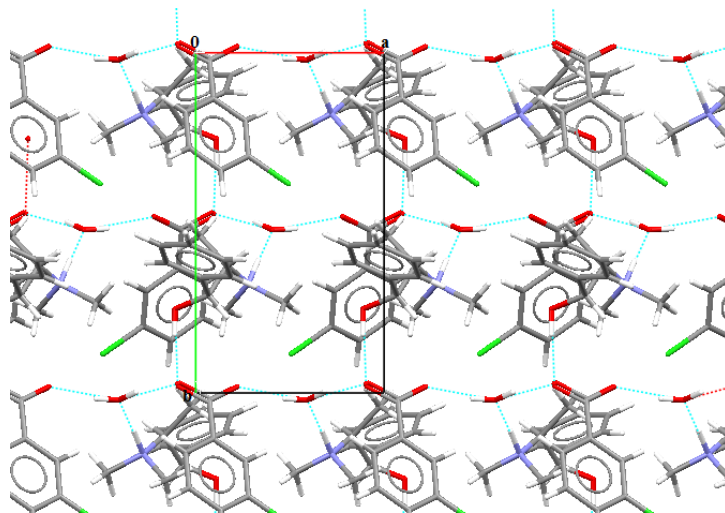
Figure 4.5 Formation of two dimensional sheet along the crystallographic *a* and *b* direction, (1R,2S)methylephedrinium 3-chlorobenzoate

Table 4.4 Graph-set analysis of substituted benzoate salts with non-chemically identical pairs for the enantiopure and racemic species and conglomerate forming salts of methylephedrine

compound		hydrogen bond	graph-set		growth	compound		hydrogen bond	graph-set		growth
1A	a	$OH_c\cdots COO^-$	>a<d	$D_2^1(3)$	*b*	**1B**	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*b*
	b	$NH\cdots COO^-$	>b<c	$D_2^1(3)$			b	$NH\cdots COO^-$			
	c	$OH_w\cdots COO^-$	>a<c	$D_2^2(5)$							
	d	$OH_w\cdots COO^-$	>b<d	$D_2^2(5)$							
			>c<d	$C_2^2(6)$							
			>a<b	$C_2^2(9)$							
4A	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*b*	**4B**		Crystallises as a conglomerate, same parameters as, (1R,2S)(-)-Methylephedrinium 2-fluorobenzoate			
	b	$NH\cdots COO^-$									
5A	a	$OH_c\cdots COO^-$	c	$S_1^1(6)$	*b*	**5B**					
	b	$NH\cdots COO^-$	>a<b	$C_2^2(9)$				Crystallises as the free acid 2-hydroxybenzoic acid			
	c	$OH_a\cdots COO^-$									
9A	a	$OH_c\cdots COO^-$	>a<c	$D_2^1(3)$	*a b*	**9B**	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*b*
	b	$NH\cdots O_w$	>b>c	$D_2^2(4)$			b	$NH\cdots COO^-$			
	c	$OH_w\cdots COO^-$	>b>d	$D_2^2(4)$							
	d	$OH_w\cdots COO^-$	>a<d	$D_2^2(5)$							
			>c<d	$C_2^2(6)$							
			<a>b	$D_2^2(8)$							
15A	a	$OH_c\cdots COO^-$	>b<c	$R_1^2(4)$	*b*	**15B**		Crystallises as a conglomerate, same parameters as, (1R,2S)(-)-Methylephedrinium 4-chlorobenzoate			
	b	$NH\cdots COO^-$	>a<b	$C_2^1(7)$							
	c	$NH\cdots COO^-$	>a<c	$C_2^2(9)$	*b*						
16A		Crystallises as the free acid 4-fluorobenzoic acid				**16B**	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*c*
							b	$NH\cdots COO^-$			
17A	a	$OH_c\cdots COO^-$	>b<f	$D_2^1(3)$	*a b c*	**17B**					
	b	$NH\cdots COO^-$	>c<e	$D_2^1(3)$							
	c	$OH_c\cdots COO^-$	>a<f	$D_2^2(5)$				Crystallises as a conglomerate, same parameters as (1R,2S)(-)-Methylephedrinium 4-hydroxybenzoate			
	d	$NH\cdots COO^-$	>d<e	$D_2^2(5)$							
	e	$OH_a\cdots COO^-$	>a<b	$C_2^2(9)$							
	f	$OH_a\cdots COO^-$	>c<d	$C_2^2(9)$							

compound		hydrogen bond	graph-set		growth	compound	hydrogen bond	graph-set	growth
			>a>e	$D_2^2(10)$					
			>b>e	$D_2^2(10)$					
			>c>f	$D_2^2(10)$					
			>d>f	$D_2^2(10)$					
			>e>f	$C_2^2(16)$					
19A	a	$OH_c\cdots COO^-$	>a<b	$C_2^1(7)$	*b*	**19B**	Crystallises as a conglomerate, same parameters as,		
	b	$NH\cdots COO^-$					(1R,2S)(-)-Methylephedrinium *p*-toluate		

Table 4.5 Graph-set analysis of substituted benzoate salts where only one of the enantiopure-racemic pairs formed a salt of methylephedrine

compound		hydrogen bond	graph-set		growth
8A	a	$OH_c\cdots COO^-$	>b<e	$D_2^1(3)$	*a*
	b	$NH_c\cdots COO^-$	>a<e	$D_2^2(5)$	
	c	$OH_c\cdots COO^-$	>c>e	$D_2^2(9)$	
	d	$NH_c\cdots COO^-$	>d>e	$D_2^2(9)$	
	e	$NH_a\cdots COO^-$	>a<b	$C_2^2(9)$	
			>c<d	$C_2^2(9)$	
12B	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*b*
	b	$NH\cdots COO^-$			
13B	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*b*
	b	$NH\cdots COO^-$			

4.3   Sulfonate and hydrogen sulphate salts

Of the five enantiopure-racemic pairs of YSO_3^- (Y = R or OH) salt structures obtained, four are of pairs of salts that are chemically identical (20, 21, 23 and 24). There are also two polymorphs of (+/-)methylephedrinium methylsulfonate, along with both an anhydrous and monohydrated salt of the racemic 1,2-ethanedisulfonate and benzenesulfonate salts. Graph-set and structural analysis was also carried out on (1R,2S)methylephedrinium ethanesulfonate where only the enantiopure product was obtained. The graph-set analysis of all sulfonate salts is shown in Table 4.6, Table 4.7 and Table 4.8. Looking at the graph-set analysis detailed in Table 4.6, only the salts of benzenesulfonate (21A and 21B) have the same hydrogen bonding present for both structures of the pair. This is in direct contrast to what was found for the benzoate salts. However, we do not find a racemic graph-set and an enantiopure graph-set as per reference 4. One similarity throughout all the nine anhydrous structures and the two hydrogen-sulfate monohydrate structures is that only two of the three sulfonate oxygen atoms (or three of the four sulfate oxygen atoms) are utilised as hydrogen-bond acceptors. This allows for a $C_2^2(9)$ graph-set to be present as the means of network propagation for the salts of anhydrous benzenesulfonate, anhydrous 1,2-ethanedisulfonate and hydrogen-sulfate monohydrate as well as the (+/-)methylephedrinium methylsulfonate (24B) and (1R,2S)methylephedrinium ethanesulfonate. Despite the change from COO to RSO_3 functionality, this motif is analogous to the $C_2^2(9)$ chain seen with the benzoate salts, where nine of the fourteen sulfonate and hydrogen sulfate salts have the $C_2^2(9)$ chain present. The methanesulfonate salts, 24A and 24C, do not have the $C_2^2(9)$ chain present and instead display discrete $D_2^2(8)$ motifs. All crystallographically independent sulfonate oxygen atoms are involved in hydrogen bonding for the structures of (+/-)-methylephedrinium 4-hydroxybenzenesulfonate hemihydrate (22B), (+/-)methylephedrinium 1,2-ethanedisulfonate monohydrate (20C) and (+/-)methylephedrinium benzenesulfonate monohydrate (21B). This implies that there is a need for the additional hydrogen bond donors to be present (here with the presence of a water molecule) in order for all the sulfonate oxygen atoms to be involved in hydrogen bonding.

Table 4.6 Graph-set analysis of sulfonate and hydrogen-sulfonate salts with chemically identical pairs for the enantiopure and racemic species of methylephedrine

compound	hydrogen bond		graph-set		growth	compound	hydrogen bond		graph-set		growth
20A	a	$OH_c\cdots SO_3$	>a<c	$D_2^2(8)$	*a c*	**20B**	a	$OH_c\cdots SO_3$	>a<a	$D_2^2(8)$	*c*
	b	$NH\cdots SO_3$	>a<d	$D_2^2(8)$			b	$NH\cdots SO_3$	>b<b	$D_2^2(8)$	
	c	$OH_c\cdots SO_3$	>b<c	$D_2^2(8)$					>a<b	$C_2^2(9)$	
	d	$NH\cdots SO_3$	>b<d	$D_2^2(8)$					>a<a>b<b	$C_4^4(24)$	
			>a<b	$C_2^2(9)$					>a<b>a<b	$R_4^4(24)$	
			>c<d	$C_2^2(9)$							
21A	a	$OH_c\cdots SO_3$	>a<b	$C_2^2(9)$	*a*	**21C**	a	$OH_c\cdots SO_3$	>a<b	$C_2^2(9)$	*c*
	b	$NH\cdots SO_3$						b	$NH\cdots SO_3$			
23A	a	$OH_a\cdots O_w$	>c<d	$D_2^1(3)$	*a b c*	**23B**	a	$OH_w\cdots SO_4$	>b<c	$D_2^1(3)$	*a b c*
	b	$OH_c\cdots SO_4$	>b<d	$D_2^2(5)$			b	$OH_w\cdots SO_4$	>a<c	$D_2^2(5)$	
	c	$NH\cdots SO_4$	>b<e	$D_2^2(5)$			c	$OH_c\cdots SO_4$	>a<d	$D_2^2(5)$	
	d	$OH_w\cdots SO_4$	>c<e	$D_2^2(5)$			d	$NH\cdots SO_4$	>b<d	$D_2^2(5)$	
	e	$OH_w\cdots SO_4$	>b>a	$D_2^2(6)$			e	$OH_a\cdots O_w$	>c>e	$D_2^2(6)$	
			>c>a	$D_2^2(6)$					>d>e	$D_2^2(6)$	
			>a>d	$C_2^2(6)$					>a<b	$C_2^2(6)$	
			>a>e	$C_2^2(6)$					>c<d	$C_2^2(9)$	
			>d>e	$C_2^2(9)$					>a>e>a>e	$R_4^4(12)$	
			>b>c	$C_2^2(9)$					>b>e>b>e	$R_4^4(12)$	
24A	a	$OH_c\cdots SO_3$	>a<d	$D_2^2(5)$	*b*	**24B**	a	$OH_c\cdots SO_3$	>a<b	$C_2^2(9)$	*c*
	b	$NH\cdots SO_3$	>b<e	$D_2^2(5)$			b	$NH\cdots SO_3$			
	c	$OH_c\cdots SO_3$	>c<f	$D_2^2(5)$		**24C**	a	$OH_c\cdots SO_3$	>a<d	$D_2^2(5)$	$\perp$ *b*
	d	$NH\cdots SO_3$	<a>b	$D_2^2(8)$			b	$NH\cdots SO_3$	>b<c	$D_2^2(5)$	
	e	$OH_c\cdots SO_3$	<c>d	$D_2^2(8)$			c	$OH_c\cdots SO_3$	<a>b	$D_2^2(8)$	
	f	$NH\cdots SO_3$	<e>f	$D_2^2(8)$			d	$NH\cdots SO_3$	<c>d	$D_2^2(8)$	

Table 4.7 Graph-set analysis of sulfonate salts with non-chemically identical pairs for the enantiopure and racemic species of methylephedrine

compound	\	hydrogen bond	\	graph-set	growth	compound	\	hydrogen bond	\	graph-set	growth
22A	a	$OH_a\cdots SO_3$	a	$C_1^1(8)$	*a b*	**22B**	a	$OH_a\cdots SO_3$	>e<h	$D_2^1(3)$	*a b*
	b	$OH_c\cdots SO_3$	>b<c	$C_2^1(7)$			b	$OH_a\cdots SO_3$	>c>g	$D_2^2(4)$	
	c	$NH\cdots SO_3$	>b>a<b	$D_3^3(13)$			c	$OH_c\cdots O_w$	>c>h	$D_2^2(4)$	
			>c>a<c	$D_3^3(13)$			d	$NH\cdots SO_3$	<g>h	$D_2^2(5)$	
							e	$OH_c\cdots SO_3$	>a<f	$D_2^2(5)$	
							f	$NH\cdots SO_3$	>a<g	$D_2^2(5)$	
							g	$OH_w\cdots SO_3$	>b<d	$D_2^2(5)$	
							h	$OH_w\cdots SO_3$	>b<e	$D_2^2(5)$	
									>b<h	$D_2^2(5)$	
									>d<e	$D_2^2(5)$	
									>d<h	$D_2^2(5)$	
									>f<g	$D_2^2(5)$	
									<c>d	$D_2^2(8)$	
									<e>f	$D_2^2(8)$	
									>d>a	$D_2^2(10)$	
									>e>a	$D_2^2(10)$	
									>f>b	$D_2^2(10)$	
									>g>b	$D_2^2(10)$	
									>h>a	$D_2^2(10)$	
									>a>b	$C_2^2(16)$	

88

Table 4.8 Graph-set analysis of sulfonate salts with additional hydrated structures and sulfonate structures where only one of the enantiopure-racemic pair formed a methylephedrine salt

compound	hydrogen bond		graph-set		growth
20C	a	$OH_c \cdots O_w$	>a>c	$D_2^2(4)$	*ab*
	b	$NH \cdots SO_3$	>a>d	$D_2^2(4)$	
	c	$OH_w \cdots SO_3$	>c<d	$C_2^2(6)$	
	d	$OH_w \cdots SO_3$	<a>b	$D_2^2(8)$	
			>b<c	$D_2^2(8)$	
			>b<d	$D_2^2(8)$	
21B	a	$OH_c \cdots SO_3$	>b>c	$D_2^2(4)$	*a*
	b	$NH \cdots O_w$	>b>d	$D_2^2(4)$	
	c	$OH_w \cdots SO_3$	>a<c	$D_2^2(5)$	
	d	$OH_w \cdots SO_3$	>a<d	$D_2^2(5)$	
			>c<d	$C_2^2(6)$	
			<a>b	$D_2^2(8)$	
25A	a	$OH_c \cdots SO_3$	>a<b	$C_2^2(9)$	*a*
	b	$NH \cdots SO_3$			

The crystal structure for compound (1R,2S)(-)-methylephedrinium 1,2-ethanedisulfonate, 20A, has two base molecules per asymmetric unit compared to one per asymmetric unit for the equivalent (+/-)-methylephedrinium salt, 20B and the monohydrated salt, 20C. The structure of 20A forms hydrogen-bonds with two cations and two anions in a ring formation, $R_4^4(24)$, see Figure 4.6. In the racemic structure, 20B, instead of the hydrogen bonding forming a ring, the ethanedisulfonate ion lies along the crystallographic *a* direction and forms a second $C_2^2(9)$ chain making the overall network a two dimensional sheet. In the racemic monohydrated salt, 20C, the presence of the water molecule prevents the formation of any $C_2^2(9)$ chains, and network propagation occurs instead through discrete graph-set motifs and a solvent separated $C_2^2(6)$ chain to form a two dimensional sheet. The formation of a ring is also seen in the structure of (+/-)-methylephedrinium hydrogen-sulfate monohydrate where the anions and water molecules hydrogen-bond to form two $R_4^4(12)$ rings, see Figure 4.6.
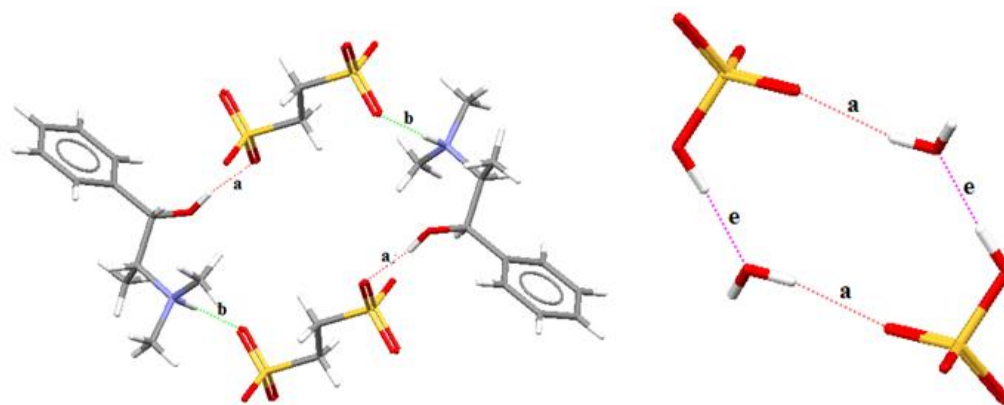
Figure 4.6 $R_4^4(24)$ ring formation of (1R,2S)methylephedrinium 1,2-ethanedisulfonate (left) and $R_4^4(12)$ ring of (+/-)methylephedrinium hydrogen-sulfate monohydrate (right)

4.4 Dicarboxylic acid derivatives salts

Of the four enantiopure-racemic pairs of dicarboxylic acid salt structures only two pairs are chemically identical, namely the hydrogen-maleate and hydrogen-malonate salts. The two hydrogen-maleate salts (26A and 26B in Table 4.9) have identical hydrogen bonding present, with the hydrogen-malonates' hydrogen bonding differing due to the (1R,2S)(-)-methylephedrinium malonate salt having two ion pairs per asymmetric unit, which results in the growth of two parallel units along the crystallographic *a* direction, see Figure 4.7. All four salts conform to Etter's rules,[2] with the presence of internal hydrogen bonds giving the graph-sets $S_1^1(7)$ and $S_1^1(6)$. The common $C_2^2(9)$ chain seen for both the benzoate derivatives and the sulfonate salts is absent here. This is because one of the oxygen atoms of the carboxylic is involved in intramolecular hydrogen bonding. In the case of the hydrogen-maleate salts, $C_2^2(12)$ chains propagate through oxygen atoms at either end of the maleate anion. The hydrogen-malonate pairs have different hydrogen bonding from each other, with the enantiopure salt showing a discrete graph-set and the racemic salt propagating through $C_2^1(7)$ chains with both the NH and OH of the cation hydrogen bonding to the same oxygen of the anion.

Table 4.9 Graph-set analysis of dicarboxylic acid derived salts of chemically identical pairs of enantiopure and racemic salts of methylephedrine

compound		hydrogen bond	graph-set		growth	compound		hydrogen bond	graph-set		growth
26A	a	$OH_c\cdots COO^-$	c	$S_1^1(7)$	*a*	**26B**	a	$OH_c\cdots COO^-$	c	$S_1^1(7)$	*a*
	b	$NH\cdots COO^-$	>a<b	$C_2^2(12)$			b	$NH\cdots COO^-$	>a<b	$C_2^2(12)$	
	c	$COOH\cdots COO^-$					c	$COOH\cdots COO^-$			
27A	a	$OH_c\cdots COOH$	e	$S_1^1(6)$	*a*	**27B**	a	$OH_c\cdots COO^-$	c	$S_1^1(6)$	*c*
	b	$NH\cdots COO^-$	f	$S_1^1(6)$			b	$NH\cdots COO^-$	>a<b	$C_2^1(7)$	
	c	$OH_c\cdots COOH$	>a<d	$D_2^2(7)$			c	$COOH\cdots COO^-$			
	d	$NH\cdots COO^-$	>b<c	$D_2^2(7)$								
	e	$COOH\cdots COO^-$	<a>b	$D_2^2(8)$								
	f	$COOH\cdots COO^-$	<c>d	$D_2^2(8)$								
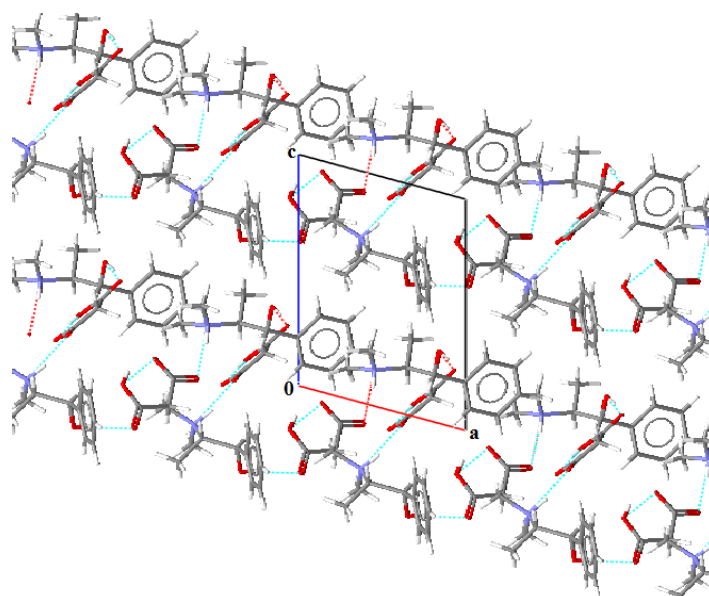
Figure 4.7 growth of two parallel chains along the crystallographic *a* direction for (1R,2S)methylephedrinium malonate

The enantiopure-racemic structural pairs of adipate and succinate do not form chemically identical salts. The enantiopure salt of adipate forms a monohydrate whereas the racemic salt is anhydrous. The salts of succinate differ from each other in that the enantiopure salt forms a one to one cation to anion ratio salt whereas the racemic salt has two cations for every one anion. The salt of (1R,2S)methylephedrinium (+/-)tartrate tartaric acid is unique in that it is a co-crystal. In each asymmetric unit there are two cations, one doubly deprotonated anion and a free acid molecule. Even though these five structures all differ significantly in their chemical make-up, four out of the five structures all possess the common $C_2^2(9)$ motif, see Table 4.10. The exception to this is (+/-)methylephedrinium adipate (28B) where the chain propagates through a $C_2^1(7)$ graph-set using only one oxygen atom of the COO$^-$ as an acceptor - as seen previously in (1R,2S)methylephedrinium *p*-toluate, 19A, above. The two adipate structures both form two dimensional sheets with growth in the second direction through anion-anion contact pairs propagating through a $C_1^1(9)$ graph-set, see Figure 4.8.

Table 4.10 Graph-set analysis of dicarboxylic acid derived salts of non-chemically identical pairs of enantiopure and racemic salts of methylephedrine

compound		hydrogen bond	graph-set		growth	compound		hydrogen bond	graph-set		growth
28A	a	$OH_c\cdots COO^-$	c	$C_1^1(9)$	*ab*	**28B**	a	$OH_c\cdots COO^-$	c	$C_1^1(9)$	*ac*
	b	$NH\cdots COO^-$	>a<b	$C_2^2(9)$			b	$NH\cdots COO^-$	>a<b	$C_2^2(7)$	
	c	$COOH\cdots COO^-$	>b>c<b	$D_3^2(12)$			c	$COOH\cdots COO^-$	>a>c<a	$D_3^3(14)$	
			>a>c<a	$D_3^3(14)$					>b>c<b	$D_3^3(14)$	
29A	a	$OH_c\cdots COO^-$	c	$C_1^1(7)$	*a*	**29B**	a	$OH_c\cdots COO^-$	>a<b	$C_2^2(9)$	*c*
	b	$NH\cdots COO^-$	>a<b	$C_2^2(9)$			b	$NH\cdots COO^-$			
	c	$COOH\cdots COO^-$	>b>c<b	$D_3^2(10)$							
30A	a	$OH_c\cdots COO^-$	c	$C_1^1(5)$	*ab*	**31B**	a	$OH_c\cdots COOH$	>b>h	$D_2^2(4)$	*ab*
	b	$NH\cdots COO^-$	e	$C_1^1(7)$			b	$NH\cdots OH_a$	>b<c	$R_1^2(5)$	
	c	$OH_a\cdots COO^-$	>d>g	$C_2^2(4)$			c	$NH\cdots COOH$	>e<f	$R_1^2(5)$	
	d	$OH_a\cdots O_w$	>a<g	$D_2^2(7)$			d	$OH_c\cdots COO^-$	>a>h	$D_2^2(7)$	
	e	$COOH\cdots COO^-$	>b<g	$D_2^2(7)$			e	$NH\cdots COO^-$	>b<g	$D_2^2(7)$	
	f	$OH_w\cdots COOH$	>d>f	$C_2^2(7)$			f	$NH\cdots OH_a$	>c>h	$D_2^2(7)$	
	g	$OH_w\cdots OH_a$	>f<g	$C_2^2(7)$			g	$OH_a\cdots COOH$	>f<h	$D_2^2(7)$	
			>a<f	$D_2^2(8)$			h	$OH_a\cdots COO^-$	>f>g	$D_2^2(7)$	
			>a>d	$D_2^2(8)$					>a<c	$C_2^1(7)$	
			>b<f	$D_2^2(8)$					>d<e	$C_2^1(7)$	
			>b>d	$D_2^2(8)$					>a<g	$D_2^2(8)$	
			>a<b	$C_2^2(9)$					>c<g	$D_2^2(8)$	
			>e<c	$C_2^2(10)$					>d<h	$D_2^2(8)$	
			>c>e	$C_2^2(12)$					>d>g	$D_2^2(8)$	
			>a>c<a	$D_3^2(8)$					>e<h	$D_2^2(8)$	
			>b>e<b	$D_3^2(10)$					>e>g	$D_2^2(8)$	
			>b>c<b	$D_3^3(10)$					>a<b	$C_2^2(10)$	
			>a>e<a	$D_3^3(12)$					>d<f	$C_2^2(10)$	
			>f>e<f	$D_3^3(12)$					>g>h	$R_2^2(11)$	
			>g>c<g	$D_3^3(12)$							
			>g>e<g	$D_3^3(12)$							

compound	hydrogen bond	graph-set		growth	compound	hydrogen bond	graph-set		growth
30A cont.		<d>c>d	$D_3^3(14)$						
		<d>e>d	$D_3^3(14)$						
		>f>c<f	$D_3^3(14)$						
		>c>c<e	$R_3^3(15)$						
		>c>e<c<e	$R_4^4(22)$						
		>c>e>c<e<e	$R_5^5(29)$						
		>c>e>e<c<e<e	$R_6^6(36)$						
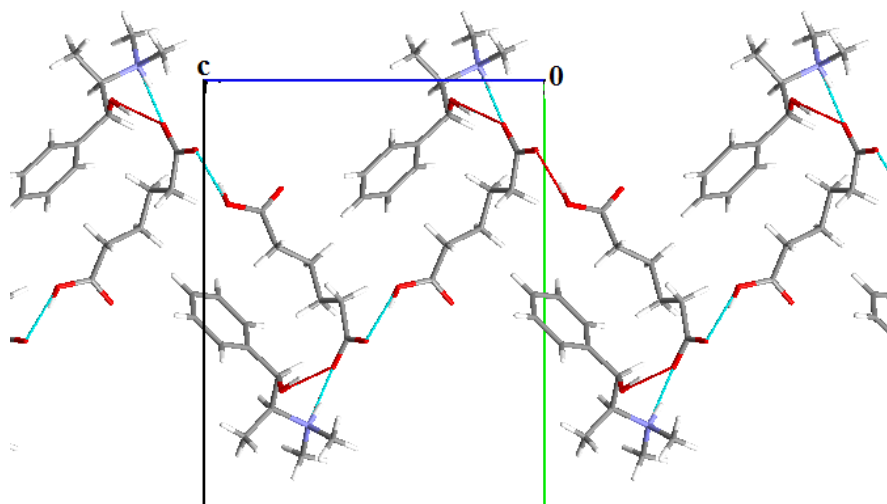
Figure 4.8 Structure of (+/-)methylephedrinium adipate showing crystal growth through $C_1^1(9)$ graph-set chains

4.5 Mandelic acid derived salts

The reaction of enantiopure and racemic methylephedrine with enantiopure and racemic mandelic acid gave three different salt structures, anhydrous and monohydrated enantiomerically pure salts and an anhydrous racemic salt, 32A, 32C and 33B respectively. The two anhydrous salts have identical graph-set present and form one dimensional ribbons using the $C_2^2(9)$ graph-set along the crystallographic b direction, see Table 4.11. Unlike previously seen cases the addition of the water molecule in (1R,2S)methylephedrinium L-mandelate monohydrate, 32C, does not contribute to an extra dimension in network growth, instead the structure propagates to form a one dimensional ribbon along the crystallographic b direction as with the anhydrous salts, see Figure 4.9.

Table 4.11 Graph-set analysis of mandelic acid derived salts of methylephedrine

compound		hydrogen bond	graph-set		growth	compound		hydrogen bond	graph-set		growth
32A	a	$OH_c \cdots OH_a$	d	$C_1^1(5)$	b	**33B**	a	$OH_c \cdots COO^-$	d	$C_1^1(5)$	b
	b	$OH_c \cdots COO^-$	>a<b	$R_1^2(5)$			b	$NH \cdots OH_a$	>b<c	$R_1^2(5)$	
	c	$NH \cdots COO^-$	>b<c	$C_2^2(9)$			c	$NH \cdots COO^-$	>a<c	$C_2^2(9)$	
	d	$OH_a \cdots COO^-$	>a<c	$C_2^2(10)$			d	$OH_a \cdots COO^-$	>a<b	$C_2^2(10)$	
			>b>d<b	$D_3^2(8)$					>a>d<a	$D_3^2(8)$	
			>a>d<a	$D_3^3(8)$					>b>b	$D_3^3(8)$	
			>c>d<c	$D_3^3(10)$					>c>d<c	$D_3^3(10)$	
32C	a	$OH_c \cdots COO^-$	d	$C_1^1(5)$	b						
	b	$NH \cdots OH_a$	e	$C_1^1(2)$							
	c	$NH \cdots COO^-$	>a<f	$D_2^2(5)$							
	d	$OH_a \cdots COO^-$	>c<f	$D_2^2(5)$							
	e	$OH_w \cdots O_w$	>b<c	$R_1^2(5)$							
	f	$OH_w \cdots COO^-$	>b<f	$D_2^2(6)$							
			>a<c	$C_2^1(7)$							
			>a<b	$C_2^2(10)$							
			<f>e>f	$D_3^3(7)$							
			>f>d<f	$D_3^2(8)$							
			>b>d<b	$D_3^3(8)$							
			>a>d<a	$D_3^3(10)$							
			>c>d<c	$D_3^3(10)$							
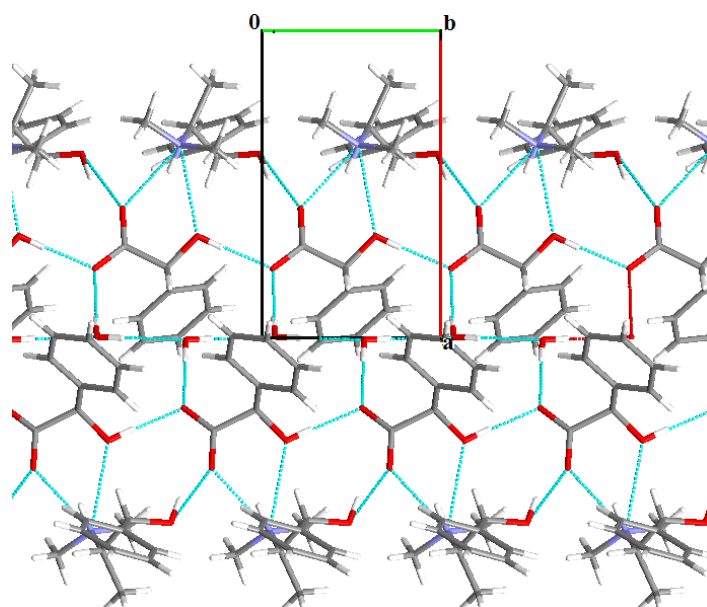
Figure 4.9 Crystal growth of (1R,2S)methylephedrine L-mandelate monohydrate along the crystallographic b direction

4.6 Halide salts

Three pairs of halide salt structures were obtained. For the chloride and bromide salts, the enantiopure and racemic salt pairs are chemically identical. The 'iodide' salts differ as the (1R,2S)(-)-methylephedrinium salt is a mixed iodide, tri-iodide structure, whilst the racemic salt was isolated as both a simple iodide and as an iodide salt with a neutral I_2 molecule also present. The bromide and chloride pairs form identical hydrogen bonding patters and all seven salts propagate via a $C_2^1(7)$ chain, see Table 4.12. This $C_2^1(7)$ motif is a monoatomic variation on the common $C_2^2(9)$ motif seen for the organic anions. All excluding the above mentioned iodide-tri-iodide structure form one dimensional ribbons. In the (1R,2S)(-)-methylephedrinium iodide tri-iodide structure, 36A, network propagation is in the *a* and *c* directions via $C_2^1(7)$ chains, giving the same hydrogen-bond motif as present in the other halides. The tri-iodide anions in 36A and the iodine molecules in 36C form channels but make no hydrogen bonding interactions, as illustrated in Figure 4.10. (-)Methylephedrinium tetrafluoroborate was the only other salt with a spherical counterion that was obtained. Like the two hydrogen sulfate structures it also contains the common $C_2^2(9)$ graph-set, see Table 4.13.

Table 4.12 Graph-set analysis of halide salts of enantiopure and racemic methylephedrine

compound	hydrogen bond		graph-set		growth	compound	hydrogen bond		graph-set		growth
34A	a	$OH_c{\cdots}X$	>a<b	$C_2^1(7)$	*b*	**34B**	a	$OH_c{\cdots}X$	>a<b	$C_2^1(7)$	*a*
	b	$NH{\cdots}X$					b	$NH{\cdots}X$			
35A	a	$OH_c{\cdots}X$	>a<b	$C_2^1(7)$	*b*	**35B**	a	$OH_c{\cdots}X$	>a<b	$C_2^1(7)$	*c*
	b	$NH{\cdots}X$					b	$NH{\cdots}X$			
36A	a	$OH_c{\cdots}X$	>a<c	$D_2^1(3)$	*a c*	**36B**	a	$OH_c{\cdots}X$	>a<b	$C_2^1(7)$	*b*
	b	$NH{\cdots}X$	>a<d	$D_2^1(3)$			b	$NH{\cdots}X$			
	c	$OH_c{\cdots}X$	>b<c	$D_2^1(3)$		**36C**	a	$OH_c{\cdots}X$	>a<b	$C_2^1(7)$	*a*
	d	$NH{\cdots}X$	>b<d	$D_2^1(3)$			b	$NH{\cdots}X$			
			>a<b	$C_2^1(7)$							
			>c<d	$C_2^1(7)$							

Table 4.13 Graph-set analysis of tetrafluoroborate salt of methylephedrine

compound	hydrogen bond		graph-set		growth
37A	a	$OH_c{\cdots}BF_4$	>a<b	$C_2^2(9)$	*c*
	b	$NH{\cdots}BF_4$			
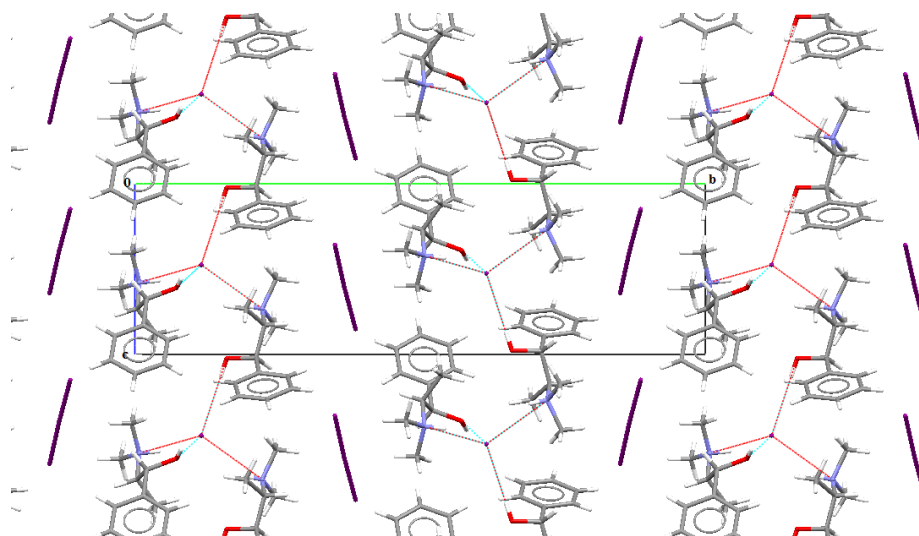
Figure 4.10 Structure of (1R,2S)methylephedrine iodide tri-iodide showing tri-iodide ions forming channels in between $C_2^1(7)$ graph-set chains

4.7 PIXEL calculation of energies of $C_2^2(9)$ graph-set

When analysing organic crystal structures the main and sometimes only bonding type commonly discussed is hydrogen bonding. The reason for this is twofold, firstly hydrogen bonds are very easy to identify as they are easily picked out by structural viewing programs. Secondly, they are known to be significant interactions as they can have an energy range from 4 kJ/mol for weak distant interactions up to 120 kJ/mol for strongly covalent, electronegative interactions.[9] However there are many other types of intermolecular interactions possible, including π-π stacking interactions, van der Waals interactions and halogen bonds[10] to name a few. These interactions are usually weaker than hydrogen bonds, for instance van der Waals interactions typically have energies of less than 5 kJ/mol. They can also often be hard to identify and are sometimes nearly impossible to visualise and therefore are often omitted from routine structural discussions. It is often unknown what effect collectively all the lesser interactions have on the overall energy of the crystal structure and therefore is it right to just discuss hydrogen bonds when examining a crystal structure? Previous studies suggest that studying all possible interactions and interaction types is necessary.[11-13]

In order to try and quantify the energy associated with hydrogen bonds in comparison to the overall energy of all the intermolecular interactions, the common reoccurring $C_2^2(9)$ chain was studied using the PIXEL program. All anhydrous compounds with a Z'=1 that contained the $C_2^2(9)$ graph-set as a means of hydrogen bond network propagation through cation-anion-cation chains were examined and the results shown in Table 4.14. The table shows the coulombic, polarisation, dispersion and repulsion energies associated with the motif, as well as the attractive energy for the $C_2^2(9)$ interaction where $E_{attractive\ =\ }E_{coul} + E_{pol} + E_{disp}$. The total energy for the individual NH···X and OH···X hydrogen bonds (where X is COO⁻, SO₃⁻ or BF₄⁻) and the energy for the $C_2^2(9)$ interaction have also been calculated, where $E_{total} = E_{coul} + E_{pol} + E_{disp\ +} E_{rep}$. Generally, the NH···X hydrogen bond has a much greater average energy associated with it compared to the OH···X hydrogen bond, average values -45.1 kJ/mol and -13.9 kJ/mol respectively. Structure 37A is calculated to have abnormally weak attraction energies (-6.2 and -7.4 kJ/mol) which is reassuring with respect to the integrity of the methodology used as here the hydrogen bond acceptor is the BF₄ anion – known to be a weak acceptor. The two benzenesulfonate salts 21A and 21 C are the other two structures to give appreciably low energies, with respect to the NH component of the motif (-27.5and -29.4 kJ/mol respectively). This low energy is reflected in the N···O separation distances which are 2.801 and 2.913 Å, comparably longer than the average N···O separation distance of 2.683 Å, see Table 4.15. The other compound to have a larger N···O separation distance is 33B, here the energy associated with the NH component is not overly low, with a value of -35.6 kJ/mol however it is still considerably below the observed average NH component energy.

That the NH···X interaction tends to be stronger than the OH···X equivalent may be expected as the first mentioned contains an interaction between a formally positively charged proton and the negatively charged carboxylate, sulfonate or tetrafluoroborate group. However for the compounds 5A, 21C and 33B this is not the case and the total energy associated with the OH···X hydrogen bond is comparable to or greater than the energy associated with the NH···X interaction. For 21C and 33B this large

energy can be associated with the O⋯O separation distance which is considerably smaller than the N⋯O distance for both these compounds, see Table 4.15. However this is not the case for 5A. From examining the structure of 5A it is not understood why the OH⋯X interaction has such a large energy. One possibility is that it could be to do with the internal hydrogen bond on the anion. Figure 4.11 shows that the O⋯O(F) distance does not vary with energy but that the N⋯O(F) distance does ('F' is present instead of 'O' for the BF_4 counterion). Somewhat counter-intuitively the larger energies are associated with long N⋯O(F) separations. The graph makes it more evident that the observed OH⋯X interaction for 5A is an outlier as its data point lies on the NH⋯O correlation line.

Table 4.14 shows the percentage of the compounds' attractive energy and total energy that is associated with the $C_2^2(9)$ graph-set. For the total attractive energy an average of 45.7 % of this can be assigned to the $C_2^2(9)$ interaction, and an average of 27.5 % of the total energy can be assigned to the $C_2^2(9)$ interaction. These values can be looked at in two ways; from one stand point 45.7 % of the compounds attractive energy and 27.5 % of the compounds total energy is significant, proving that it is correct to regard this interaction as being greatly important. However, looking at this result from a different perspective, this interaction does not account for 54.3 % of the total attractive energy and more importantly 72.5 % of the total energy associated with the compound. For the majority of these compounds the $C_2^2(9)$ graph-set is the only hydrogen bond interaction present and therefore the other 'lesser' intermolecular interactions are indeed significant in contributing to the crystals total energy.

Table 4.14 Coulombic, polarisation, dispersion and repulsion energies associated with the $C_2^2(9)$ chain graph-set (kJ/mol)

Compound	$C_2^2(9)$ E_{coul}	$C_2^2(9)$ E_{pol}	$C_2^2(9)$ E_{disp}	$C_2^2(9)$ $E_{attractive}$	$C_2^2(9)$ E_{rep}	$C_2^2(9)$ E_{total}	NH E_{total}	OH E_{total}	% $E_{attractive}$	% E_{total}
1B	-117.4	-74.7	-48.8	-240.9	199.5	-41.6	-33.8	-7.8	46.9	19.7
2A	-118.3	-63.6	-34.3	-216.2	166.7	-49.6	-37.8	-11.8	44.2	26.0
3A	-116.9	-64.9	-48.3	-230.1	170.0	-60.0	-46.8	-13.2	43.7	25.4
3B	-116.1	-74.7	-57.6	-248.4	195.8	-52.4	-41.3	-11.1	54.0	30.2
4A*(no disorder)	-128.1	-71.4	-41.0	-240.5	181.2	-59.3	-52.8	-6.5	52.9	35.0
5A	-166.3	-74.1	-54.9	-295.3	202.0	-93.2	-34.6	-58.6	56.9	41.7
7A	-132.0	-73.4	-43.3	-248.7	187.0	-61.7	-53.9	-8.1	52.1	34.1
7B	-130.3	-79.6	-58.5	-268.4	213.9	-54.5	-44.5	-10.0	57.3	33.1
9B	-143.5	-76.6	-39.3	-259.4	203.3	-56.1	-44.8	-11.3	48.5	26.6
10A	-139.4	-72.6	-35.6	-247.6	188.4	-59.3	-45.6	-13.7	47.7	26.0
10B	-147.7	-78.1	-38.2	-264	205.8	-58.1	-45.5	-12.6	50.8	28.4
11B	-103.8	-71.6	-51.4	-226.8	169.4	-57.4	-57.9	0.5	36.7	21.5
12B	-133.4	-71.1	-36.3	-240.8	180.0	-60.8	-50.0	-10.8	46.8	27.7
13B	-145.7	-78.6	-42.4	-266.7	203.4	-63.2	-52.6	-10.6	52.5	32.2
14A	-124.2	-67.5	-38.5	-230.2	157.0	-73.1	-63.3	-9.8	39.4	27.4
14B	-115.2	-67.5	-52.2	-234.9	160.0	-75.0	-65.4	-9.6	42.1	30.9
15A	-120.9	-68.3	-47.6	-236.8	164.8	-72.1	-56.8	-15.3	50.2	37.5
16B	-131.8	-80.8	-49.3	-261.9	208.1	-53.8	-45.0	-8.8	51.0	23.7
18B	-140.2	-75.7	-36.2	-252.1	191.9	-60.2	-46.3	-13.9	46.6	24.8
21A	-48.6	-43.2	-56.2	-148.0	103.1	-44.7	-27.5	-17.2	36.2	22.5
21C	-57.9	-45.0	-60.9	-163.8	109.9	-53.9	-29.4	-24.5	35.5	22.7
25A	-87.3	-52.8	-19.0	-159.1	153.3	-36.5	-30.6	-5.9	43.5	21.4
29A	-100.4	-73.0	-41.1	-214.5	172.7	-41.7	-47.1	5.4	36.1	18.6
32A	-136.8	-77.7	-65.6	-280.1	195.2	-84.9	-76.9	-8.0	47.2	30.9
33B	-117.3	-62.7	-62.6	-242.6	145.8	-96.9	-35.6	-61.3	41.5	36.8
37A	-24.4	-15.2	-27.9	-67.5	54.0	-13.6	-6.2	-7.4	27.3	10.9
Average	**-117.1**	**-67.5**	**-46.1**	**-230.2**	**172.4**	**-59.0**	**-45.1**	**-13.9**	**45.7**	**27.5**

Table 4.15 N...O(F*) and O...O(F*) distances and angles associated with the $C_2^2(9)$ chain graph-set of methylephedrine salts

Compound	N⋯O(F)* distance	O⋯O(F)* distance	NH⋯O angle	OH⋯O angle
1B	2.625	2.674	166.11	176.37
2A	2.618	2.733	165.83	175.62
3A	2.646	2.709	163.85	163.03
3B	2.642	2.686	169.29	157.46
4A*(no disorder)	2.656	2.678	173.08	170.39
5A	2.638	2.700	168.93	177.70
7A	2.677	2.656	171.20	172.64
7B	2.641	2.677	169.44	165.61
9B	2.625	2.625	166.32	174.27
10A	2.627	2.601	155.05	163.41
10B	2.630	2.620	167.97	177.20
11B	2.721	2.714	173.82	166.53
12B	2.654	2.639	163.18	171.53
13B	2.653	2.609	166.31	169.24
14A	2.692	2.713	173.04	175.52
14B	2.722	2.694	160.57	164.88
15A	2.722	2.707	133.12	177.85
16B	2.629	2.673	168.26	171.30
18B	2.626	2.624	165.36	171.43
21A	2.801	2.760	152.56	154.54
21C	2.913	2.707	144.82	170.07
25A	2.694	2.695	163.98	167.21
29A	2.675	2.655	167.05	175.84
32A	2.623	2.763	172.12	172.62
33B	2.833	2.698	134.41	157.85
37A*	2.762	2.852	147.80	172.40
Average	**2.683**	**2.687**	**162.44**	**169.71**
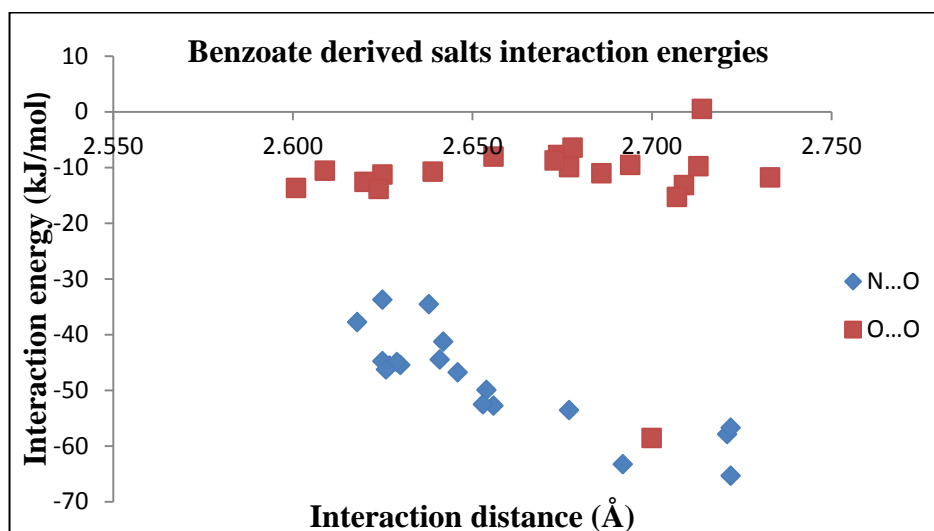
Figure 4.11 Plot of interaction distance versus interaction energy for the N⋯O(F) and O⋯O(F) of the $C_2^2(9)$ chain graph-set of methylephedrine salts

4.7.1 Are $C_2^2(9)$ interactions the strongest cation-anion interaction observed?

It is thought that the strongest interaction observed with any of the compounds should be between the positively charged cation and the negatively charged anion. It has already been established that $C_2^2(9)$ interactions are important with respect to the overall energy of the compound, but are they the strongest? Or is there another interaction between the cation and anion that has a larger contribution to the overall energy of the compound and if so what is this interaction? This was investigated for all the above compounds and the results are shown in Table 4.16.

For five compounds the largest interaction between the cation and anion does not take part in the $C_2^2(9)$ motif, namely compounds 3A, 10A, 18B, 21A and 37A. For all these interactions the coulombic energy is significantly smaller than observed for the $C_2^2(9)$ interaction, whereas the polarisation and dispersion contributions are comparable. The reason for the larger total energy in these five compounds is the very small repulsion energies observed, approximately 20 kJ/mol compared to over 100 kJ/mol for the repulsion contributions for the $C_2^2(9)$ interactions. For all five compounds this interaction is between a cation and an anion on adjacent $C_2^2(9)$ chains, as shown in Figure 4.12. This explains the low repulsion energy contribution for the interaction as when two adjacent chains pack together their optimum position is with maximum attraction and minimum repulsion energy.
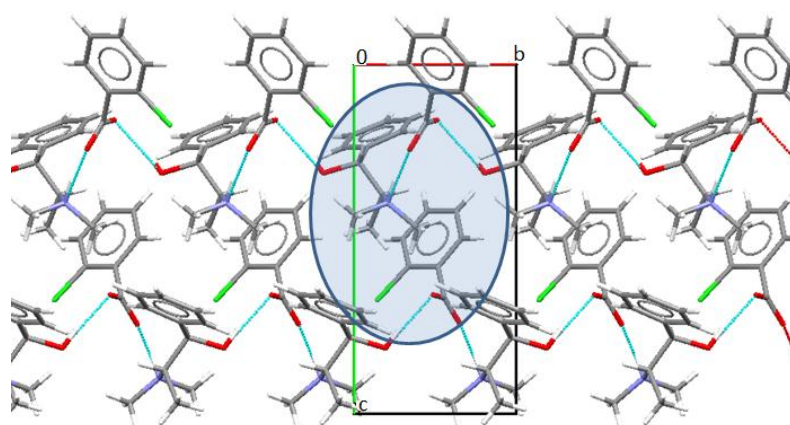
Figure 4.12 Interaction of cation and anion on adjacent $C_2^2(9)$ chains for (1R,2S)methylephedrinim 2-chlorobenzoate highlighted in blue

Table 4.16 Coulombic, polarisation, dispersion and repulsion energies for the strongest interaction between the cation and anion of methylephedrinium salts (kJ/mol)

Compound	Distance (Å)	E_{coul}	E_{pol}	E_{disp}	E_{rep}	E_{total}	C2,2(9) ?
1B	6.533	-82.3	-44.7	-29.0	122.3	-33.8	Yes
2A	7.295	-86.1	-41.4	-22.2	111.9	-37.8	Yes
3A	4.479	-33.9	-10.4	-27.0	20.3	-51.0	**No**
3B	6.303	-89.1	-47.4	-33.5	128.6	-41.3	Yes
4A	7.368	-95.4	-42.7	-22.8	108.0	-52.8	Yes
5A	5.525	-82.8	-34.1	-29.8	88.0	-58.6	Yes
7A	7.174	-93.9	-39.6	-23.5	103.1	-53.9	Yes
7B	6.081	-94.7	-48.2	-33.5	131.9	-44.5	Yes
9B	6.092	-89.8	-42.9	-26.0	113.9	-44.8	Yes
10A	5.131	-35.4	-10.9	-26.5	22.2	-50.6	**No**
10B	6.144	-90.5	-44.2	-25.6	114.8	-45.5	Yes
11B	6.230	-90.0	-41.4	-30.7	104.2	-57.9	Yes
12B	6.466	-82.3	-38.2	-22.7	93.2	-50.0	Yes
13B	6.524	-94.0	-46.5	-29.2	117.0	-52.6	Yes
14A	8.092	-94.4	-39.0	-20.2	90.3	-63.3	Yes
14B	7.772	-86.8	-36.4	-21.7	79.5	-65.4	Yes
15A	6.823	-85.5	-38.6	-26.8	94.1	-56.8	Yes
16B	6.479	-93.6	-48.1	-28.5	125.2	-45.0	Yes
18B	5.088	-45.4	-14.8	-32.3	28.6	-63.9	**No**
21A	5.750	-17.2	-7.0	-19.6	14.6	-29.2	**No**
21C	5.606	-17.4	-19.9	-38.1	46.0	-29.4	Yes
25A	5.942	-58.6	-30.2	-34.3	92.6	-30.6	Yes
29A	5.981	-79.8	-38.0	-20.3	90.9	-47.1	Yes
32A	5.117	-111.2	-51.5	-37.7	123.6	-76.9	Yes
33B	5.889	-76.4	-30.7	-24.6	70.5	-61.3	Yes
37A	4.224	-6.9	-3.8	-18.5	19.5	-9.6	**No**

4.8   Different conformations of the methylephedrinium cation

Within the salt structures the methylephedrinium cation adopts one of three different conformations. These three conformations are illustrated in Figure 4.13. Four torsion angles can be used to describe these conformations, and these encompass the two stereocentres at carbon atoms C7 and C8. Between each of the three classes, significant variation is based largely on torsion angle four. Overlay of the cations that represent the three different conformations was performed using the structural

overlay feature in Mercury CSD 2.3.[14] This is illustrated in Figure 4.14 and the torsion angle measurements for the three different conformations are shown in Table 4.17, Table 4.18 and Table 4.19
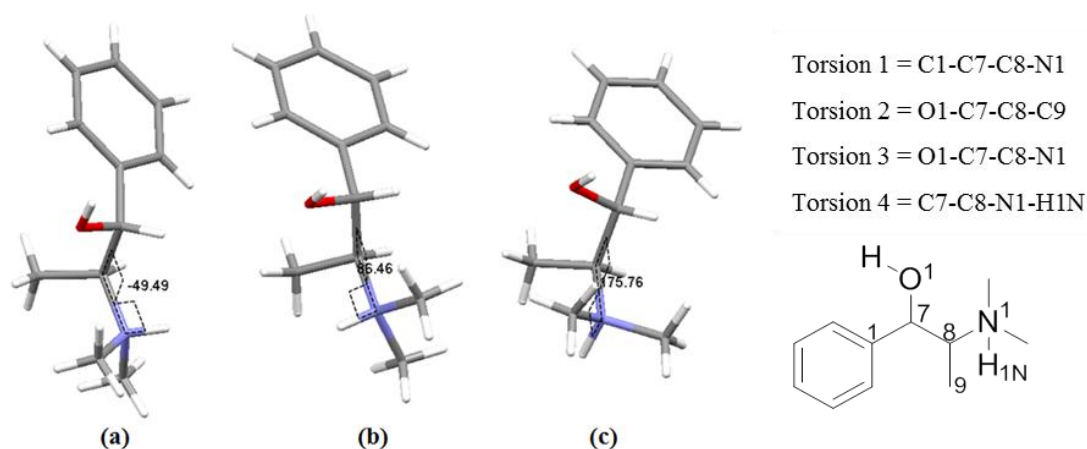
Figure 4.13 Difference conformations of cation with varying torsion four angles

Of those structures where more than one crystallographically independent cation exists in the asymmetric unit, only 17A and 20A contain two different conformers. Conformation (a) is the most common, being present in 36 cations, with a torsion angle 4 range of -26.27 ° (in 18A cation 1) to -66.67 ° (in 32A). Conformation (b) occurs in 18 of the cations and has a range of 53.88 ° (in 21B) to 91.29 ° (in 26B). Conformation (c) is the least reoccurring, being present in only 12 cations and has a torsion angle 4 range of 169.81 ° (in 5A) to -165.77 ° (in 14B). Interestingly, with the exception of 14B, only arrays of enantiopure cations exist in this third conformation, that with the amine proton *anti* to the phenylethylene chain. Related work on ephedrinium salts found two cation configurations.[7,15,16] The work by Collier et al.[16] showed that the difference in energy between the 'extended' conformation (comparable to conformation (c)) and the 'folded' conformation (comparable to conformation (a)) was insignificant as both structures represent overall energy minima. In our hands, DFT calculations using Gaussian 03[17] similarly suggests only small variations in energy, indeed in the gas phase conformation (b) and (c) were identical after geometry optimization.

(a)   (b)   (c)
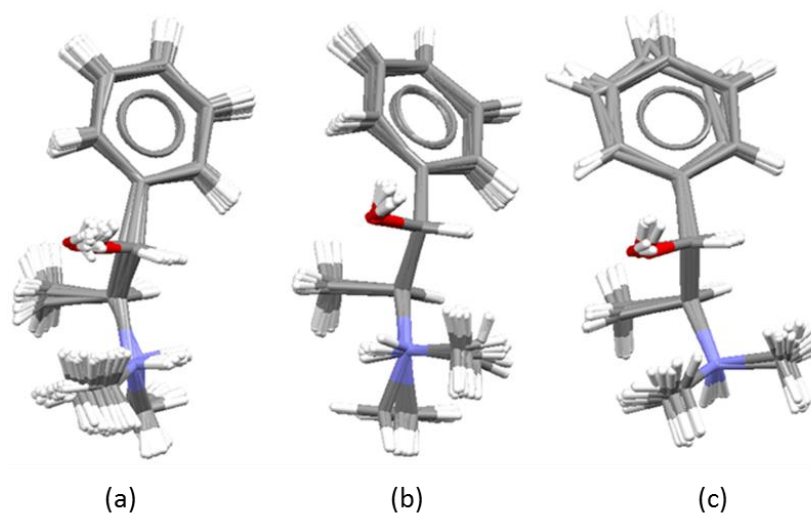
Figure 4.14 Overlay of cations from the three different conformations

Table 4.17 Torsion angle measurement form compounds containing cation conformation (a)

Structure	Torsion 1	Torsion 2	Torsion 3	Torsion 4
1A	159.548	50.552	-76.479	-40.363
1B	168.411	59.009	-67.802	-36.94
2B cation 1	157.458	49.407	-77.44	-37.69
2B cation 2	160.166	49.036	-78.117	-38.712
3A	156.419	45.656	-81.266	-49.492
3B	162.641	54.041	-73.492	-37.712
6A	150.918	41.497	-85.932	-53.18
6B	164.576	55.473	-71.322	-40.88
7B	162.547	53.703	-73.391	-41.051
8A cation 1	153.267	43.427	-83.917	-50.901
8A cation 2	153.329	42.232	-84.202	-58.724
9B	157.02	46.982	-79.789	-52.887
10A	149.013	39.084	-88.101	-57.001
10B	154.518	44.596	-82.202	-49.194
11A cation 1	161.677	45.124	-81.573	-51.435
11A cation 2	156.937	48.281	-78.958	-42.224
11B	166.867	57.974	-69.057	-36.442
12B	157.998	48.343	-78.795	-54.346
13B	153.633	42.967	-84.021	-46.754
16B	168.833	58.614	-67.827	-34.554
17A cation 1	162.837	53.898	-73.169	-46.931
18A cation 1	179.893	69.028	-56.956	-26.269
18A cation 2	149.505	39.476	-87.017	-52.086

Structure	Torsion 1	Torsion 2	Torsion 3	Torsion 4
18B	147.528	37.301	-88.941	-51.315
22A	140.838	30.205	-95.953	-56.404
22B cation 1	162.748	52.872	-73.438	-47.499
22B cation 2	165.391	55.429	-71.385	-47.509
27A cation 1	166.087	54.507	-72.2	-48.396
27A cation 2	161.433	50.957	-76.283	-56.349
27B	152.215	41.116	-85.946	-60.656
28A	154.9	44.395	-82.48	-61.188
28B	151.363	40.772	-85.786	-59.695
29A	156.623	46.672	-80.646	-53.326
29B	157.512	48.45	-77.604	-37.612
30A	160.464	49.254	-76.997	-47.572
31B cation 1	149.665	39.67	-85.958	-52.376
31B cation 2	152.424	42.131	-84.662	-52.667
32A	147.814	38.983	-89.001	-66.673
32C	156.614	44.289	-82.482	-51.443
33B	164.569	53.723	-72.512	-40.131
34A	150.079	41.89	-86.317	-54.145
34B	149.872	39.442	-87.019	-63.75
35A	150.346	41.689	-86.154	-58.647

Table 4.18 Torsion angle measurement form compounds containing cation conformation (b)

Structure	Torsion 1	Torsion 2	Torsion 3	Torsion 4
17A cation 2	174.568	61.165	-60.587	76.628
20A cation 2	171.366	56.074	-65.766	85.091
20B	172.442	57.059	-63.687	86.032
20C	174.157	60.099	-61.513	56.979
21A	-179.495	65.456	-55.561	69.316
21B	179.478	64.661	-57.418	53.882
21B	170.755	56.213	-65.086	73.513
21C	-175.094	69.728	-51.45	67.185
23B	174.732	59.635	-62.003	80.721
24A cation 1	179.805	65.625	-56.731	76.266
24A cation 1	175.106	56.222	-63.692	86.455
24A cation 1	170.688	55.613	-65.791	87.465
24C cation 1	175.915	61.208	-60.639	88.783
24C cation 2	178.032	61.852	-59.303	87.031
25A	174.504	59.656	-61.474	84.61

Structure	Torsion 1	Torsion 2	Torsion 3	Torsion 4
26B	171.604	57.658	-64.081	91.288
35B	175.065	60.724	-61.287	86.227
36A cation 1	179.855	66.667	-56.348	76.802
36A cation 2	178.583	64.428	-58.01	77.093
36B	169.688	55.81	-66.166	79.544
36C	172.838	59.639	-63.212	82.121
37A	175.975	62.347	-58.829	72.261

Table 4.19 Torsion angle measurement form compounds containing cation conformation (c)

Structure	Torsion 1	Torsion 2	Torsion 3	Torsion 4
2A	145.76	34.947	-90.755	172.629
4A	166.283	55.765	-70.106	-169.589
5A	159.843	46.625	-78.2	169.814
7A	171.35	59.074	-66.444	-178.257
9A	159.14	47.827	-77.711	175.759
14A	175.631	64.085	-62.705	-166.918
14B	169.339	59.072	-66.588	-165.773
15A	167.257	55.866	-69.32	179.575
19A	166.232	55.09	-70.734	175.646
20A cation 1	-178.751	67.964	-56.612	-170.616
23A	166.268	54.51	-70.9	-176.816
26A	172.775	61.062	-64.363	-173.643

4.9 Crystal packing similarities

Similarities in the crystal packing for the 64 methylephedrinium salt structures were investigated by using the 'crystal packing similarity' module of Mercury CSD 2.3.[14] This method is explained in experimental Section 3.8.2. For this work the analysis looked only at the largest molecular component of the salts, namely the cation, in order to prevent interference from differing anions and the presence of water molecules. By investigating similar packing in various cluster sizes from 2 to 15 cations, with a distance and torsion angle tolerance of 20 %, tree diagrams can be constructed to convey structure similarity over a large group, where structures that have 15 out of 15 cation molecules with the same geometry are said to be

isostructural, at least with respect to cation packing, see Figure 4.15. In the tree diagram less similar structures are named at the top of the diagram with the similarity increasing towards the bottom. The tree branches convey similar substructures. The colours of the small circles highlight the different cation molecular conformations, as discussed above, to show reoccurrence throughout the tree. The colours of the large ovals indicate distinct cation to cation pair geometries, or packing motifs. As the diagram indicates, all but nine structures (present at the top of the tree) have a common packing motif for at least two cations. Several of the racemic structures are able to adopt two different packing motifs shown with the corresponding two coloured half and half oval.
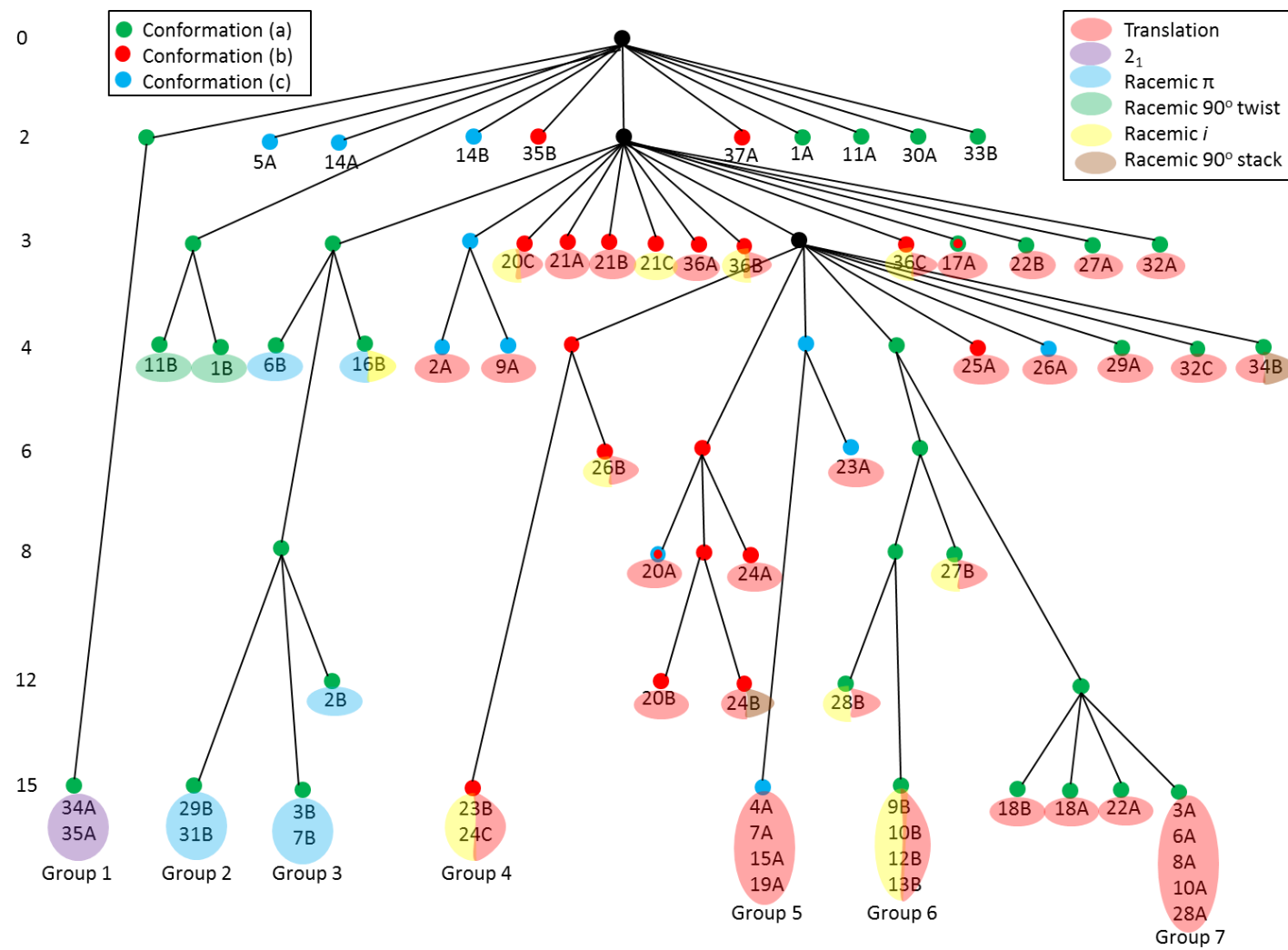
Figure 4.15 Tree diagram of similarities with respect to cation packing

The six different packing motifs are shown in Figure 4.16, In Figure 4.15 they are represented as follows; motif α (translation conformer, pink ovals), motif β (2_1 conformer, purple ovals), motif γ (racemic π conformer, blue ovals), motif δ (racemic 90° twist conformer, green ovals), motif ε (racemic *i* conformer, yellow ovals) and motif ζ (racemic 90° stack conformer, brown ovals).
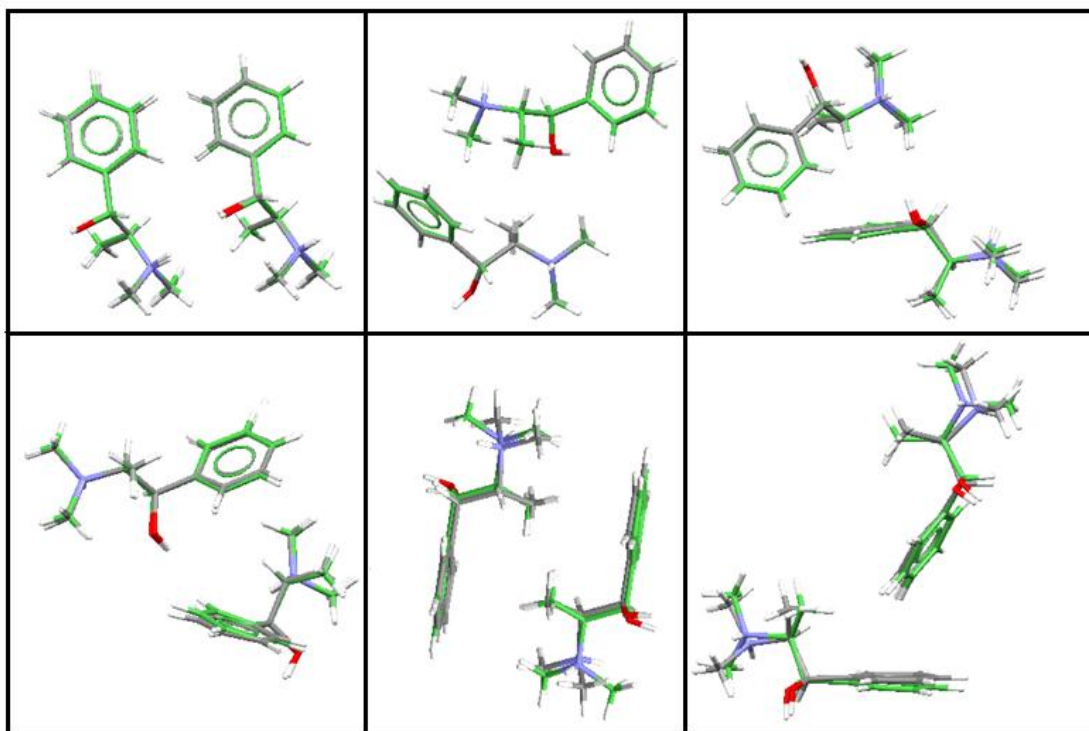
Figure 4.16 Six different packing conformers for two cation packing motif α (translation conformer, top left), motif β (2_1 conformer, top middle), motif γ (racemic π conformer, top right), motif δ (racemic 90° twist conformer, bottom left), motif ε (racemic *i* conformer, bottom middle) and motif ζ (racemic 90° stack conformer, bottom right).

Geometric constraints mean that all enantiopure salts are confined to packing motifs α and β, whereas racemic crystals are seen to use all the packing motifs except the rare motif β. Motif α, the 'translation conformer' is by far the most common occurring in 25 enantiopure salts and 18 racemic salts, with the motif ε, the 'racemic *i* conformer' being the second most common occurring in 13 racemic salts. Packing motif γ and motif ζ both feature π interactions, in motif γ the interaction is between a methine hydrogen atom and the aromatic ring, and in motif ζ there is an edge to face π interaction with the two aromatic rings, see Figure 4.17.
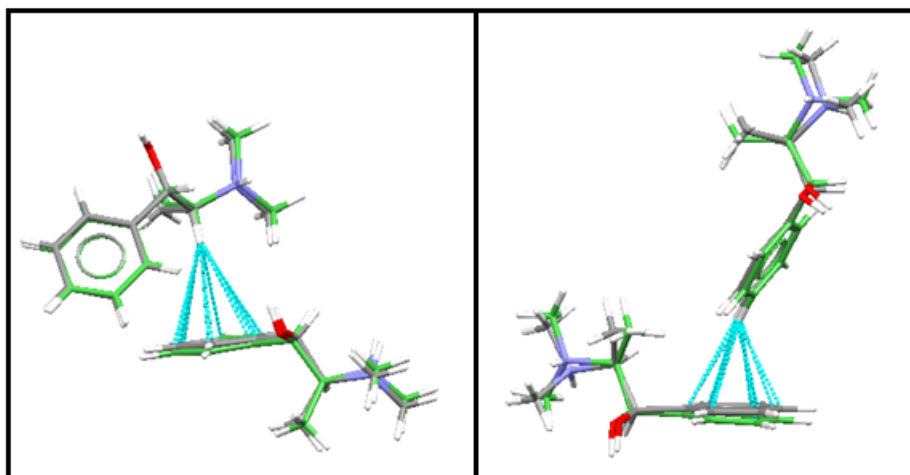
Figure 4.17 π-interactions of packing motif γ, the racemic π conformer and motif ζ, the racemic 90° stack conformer

The 'translation conformer' is the only packing motif that accommodates all three different cation molecular conformations, as conformation (c) is not present in any of the other packing motifs. This is not surprising as conformation (c) only contained one racemic salt, 14B, which is highlighted as even more of an outlier here as it does not have a common packing motif even at the two cation level. Figure 4.18 contains all the racemic salts except for 1B and 11B which are the only two salts to adopt the 'racemic 90° twist conformer'. The diagram shows which salts are able to adopt more than one packing motif. Interestingly only salts with cation molecular conformation (a) adopt the 'racemic π' packing motif; it is possible that the position of the two methyl groups and the proton attached to the nitrogen atom prevent cation conformation (b) adopting this packing motif. The cation molecular conformation (b) is able to pack using motif ε and motif ζ.
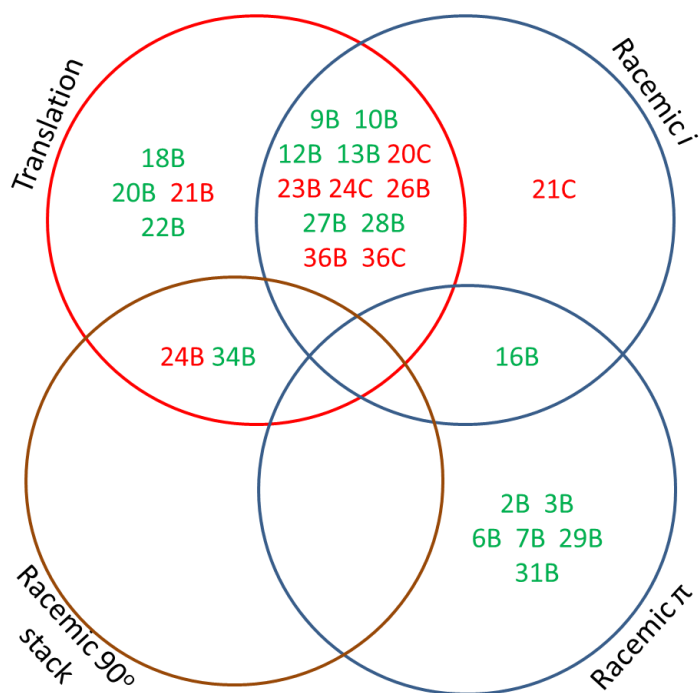
Figure 4.18 Racemic salts able to adopt various packing motifs. **Green** compounds are those that form cation conformation **(a)** and **red** compounds are those that form cation conformation **(b)**

4.10   PIXEL calculations for two cation packing conformers

To investigate the different energies of the two cation packing conformers PIXEL calculations were performed for all the anhydrous compounds with Z' = 1. The two chloride structures were also excluded due to repeated failures when attempting to run the PIXEL calculation. This left the analysis of 34 structures, of which five do not have any common packing motif for two cations. Where the compound can adopt two different common packing motifs, energies for both have been calculated. Table 4.20 shows the coulombic, polarization, dispersion, repulsion and total energies for the different two cation packing conformers.

Table 4.20 Coulombic, polarisation, dispersion and repulsion energies for the two cation packing conformers (kJ/mol)

Compound	Conformer	Distance	E_{coul}	E_{pol}	E_{disp}	E_{rep}	E_{total}
2A	translation	6.924	-5.7	-4.4	-14.4	7.5	-17.0
3A	translation	6.148	-4.5	-7.0	-24.6	16.7	-19.4
4A*(no disorder)	translation	5.729	-0.5	-8.3	-29.4	22.3	-15.7
7A	translation	5.940	-2.5	-6.6	-24.2	16.3	-17.0
9B	translation	5.849	-3.1	-9.8	-29.1	21.7	-20.4
10A	translation	5.832	6.5	-9.3	-29.4	17.9	-14.3
10B	translation	5.748	0.5	-10.7	-32.3	24.3	-18.2
12B	translation	5.767	-3.7	-10.5	-31.6	23.2	-22.6
13B	translation	6.022	5.9	-7.6	-23.9	14.3	-11.3
15A	translation	5.992	2.8	-5.8	-24.2	13.2	-14.0
18B	translation	5.828	8.9	-9.2	-28.3	17.0	-11.6
19A	translation	6.044	3.4	-5.2	-23.1	11.7	-13.2
21A	translation	7.423	-3.5	-6.1	-16.2	10.4	-15.4
22A	translation	6.140	3.0	-7.2	-23.2	12.1	-15.3
25A	translation	8.223	-11.4	-4.6	-13.6	6.8	-22.4
26A	translation	5.776	1.2	-5.9	-23.4	10.1	-18.1
26B	translation	5.921	8.2	-5.7	-20.8	7.1	-11.1
27B	translation	5.708	13.9	-9.3	-28.4	14.7	-9.0
28B	translation	5.854	4.1	-10.5	-26.6	19.0	-14.0
29A	translation	6.094	-1.3	-7.6	-21.6	14.1	-16.4
33A	translation	7.409	-7.2	-7.4	-19.9	13.3	-21.2
34B	translation	5.738	6.7	-9.6	-29.6	17.7	-14.7
34A	2_1	7.167	-9.6	-2.0	-11.4	4.0	-19.1
3B	rac π	6.024	0.3	-8.8	-29.0	14.7	-22.7
16B	rac π	6.101	-2.2	-8.6	-27.4	13.8	-24.4
7B	rac π	6.032	1.5	-8.8	-28.0	13.7	-21.6
11B	rac 90° twist	5.996	-11.5	-7.8	-25.8	14.5	-30.6
1B	rac 90° twist	6.042	-11.6	-7.5	-26.7	14.8	-30.9
9B	rac i	4.831	-23.9	-11.8	-42.3	29.8	-48.3
10B	rac i	4.732	-27.1	-12.8	-45.1	34.5	-50.5
12B	rac i	4.932	-21.4	-10.7	-38.6	24.7	-45.9
16B	rac i	6.156	-2.2	-2.1	-16.7	4.8	-16.2
28B	rac i	4.899	-25.6	-12.1	-41.4	30.9	-48.2
21C	rac i	6.036	-19.6	-4.7	-19.7	8.9	-35.1
26B	rac i	5.349	-19.7	-7.9	-28.9	11.2	-45.3
13B	rac i	4.771	-24.9	-12.7	-44.0	32.7	-49.0
34B	rac 90° stack	8.192	3.7	-6.4	-22.4	15.9	-9.2

The total energy for the common translational conformer has a range of -9.0 kJ/mol in 27B to -21.2 kJ/mol in 12B, with an average total energy of -16.0 kJ/mol. The total energy values seen for the single 2_1 and racemic 90° stack conformers have total energies of -19.1 and -9.2 kJ/mol (in compounds 34A and 34B respectively) which both lie in this range. The racemic 90° stack interaction does not appear to be a particularly strong interaction, this may be related to the 8.192 Å distance between the two cation centers. In comparison the other racemic interactions appear to be comparatively strong. The values for the rac π interactions indicate a slightly stronger interaction with an average total energy of -22.9 kJ/mol. However, it is the racemic 90° twist and racemic i conformers that clearly produce the strongest interactions between two cations. Compounds 1B and 11B are the only two to adopt the racemic 90° twist conformer and are both isostructural with respect to four cations packing, with an average total energy for two cation packing of -30.8 kJ/mol.

By far the strongest interaction with respect to two cation packing is the racemic i conformer. The average total energy for the racemic i interaction type is -42.3 kJ/mol. Most structures that adopt this conformer have an associated energy that lies within the range of -35.1 kJ/mol in compound 21C and -50.5 kJ/mol in compound 10B, though there is an apparent outlier in structure 16B with a much lower energy of -16.2 kJ/mol. This energy is comparable to the NH$^{\cdots}$X energy observed for the $C_2^2(9)$ motif which has an average energy of 45.1 kJ/mol. The Venn diagram in Figure 4.18 above shows that 16B is the only structure with a racemic i interaction that also makes a second type of racemic interaction, namely the racemic π interaction. The above values calculated for the two cation conformers are in line with those reported for other pair packing energies using PIXEL.[18-20]

Looking at the individual contributions to the overall energy it is observed that the polarization, dispersion and repulsion energy is comparable across all six packing conformers. It is the observed difference in coulombic energy that greatly influences the overall energy with the average coulombic energy of the translation conformer

significantly less than the average coulombic energy for the racemic *i* conformer, 1.159 kJ/mol and -20.550 kJ/mol respectively. These cation-cation packing pair interactions thus contribute considerably to the overall energy of the compounds. Especially important for the discussion below on the properties of enantiopure versus racemic crystal forms is the observation that racemic packing modes give stronger pair energies than other methods of relating pairs of cations.

4.11   Isostructural salts

From the tree diagram, Figure 4.15, you can see that there are several structures that are isostructural with respect to cation packing. These structures are located at the bottom of the tree with 15 matching neighbours in the similarity packing search. Three of the seven isostructural groups (groups one, three and six) have both identical orientation of the cations and identical hydrogen bonding present within the structures (either the $C_2^2(9)$ graph-set for groups three and six or the corresponding $C_2^1(7)$ graph-sets for the two halide salts of group one). The two structures of group three are both *ortho-* substituted benzoates and the four structures of group six are all *meta-* substituted benzoates. Because of this the anion molecules are able to orient themselves in identical ways between the cation molecules, see Figure 4.19.

Group five contains four structures; two o*rtho-* substituted and two *para-* substituted benzoates, all of which are anhydrous and contain only one cation-anion pair per asymmetric unit. The group however does not contain identical graph-set throughout. The two *ortho-* substituted benzoates both contain the $C_2^2(9)$ graph-set, Figure 4.20 shows that their corresponding anions can directly overlay. The *para-* substituted benzoate 19A only contains a $C_2^1(7)$ graph-set and therefore the slight change in hydrogen bonding means the anions do not directly overlay, however this slight alteration is not considered significant, see Figure 4.20.
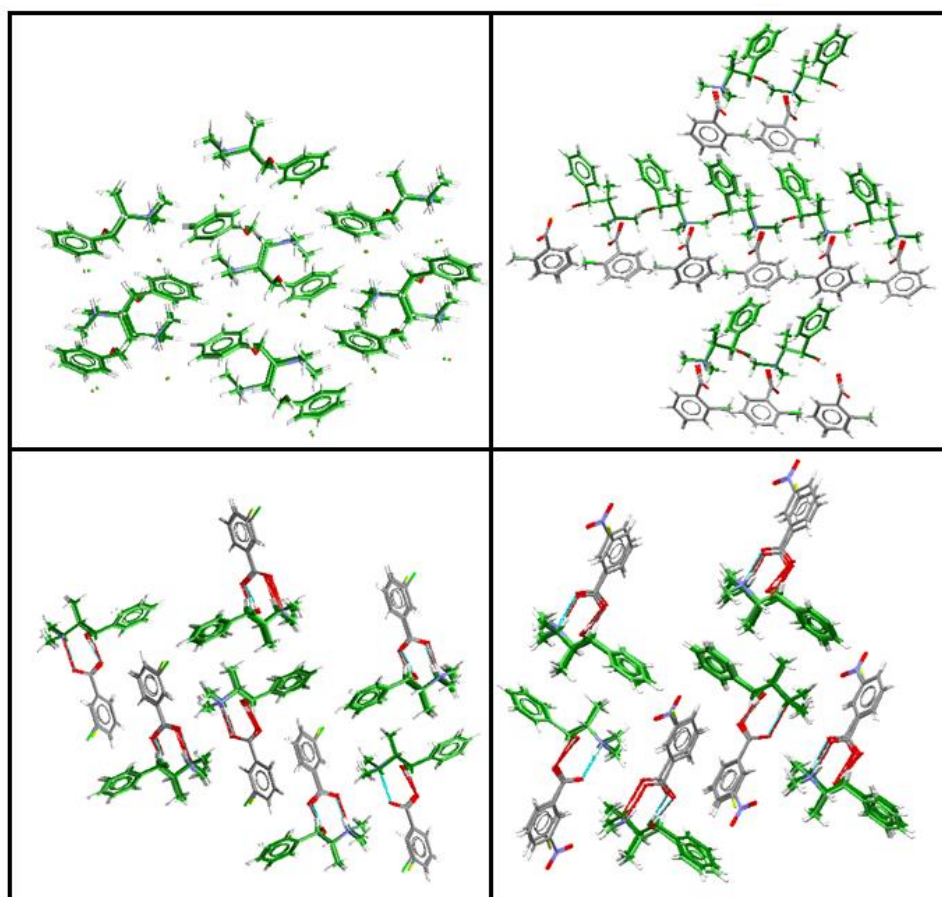
Figure 4.19 Overlay of structures from isostructural groups. Group 1 (top left), group 3 (top right), group 6 (bottom right and bottom left)
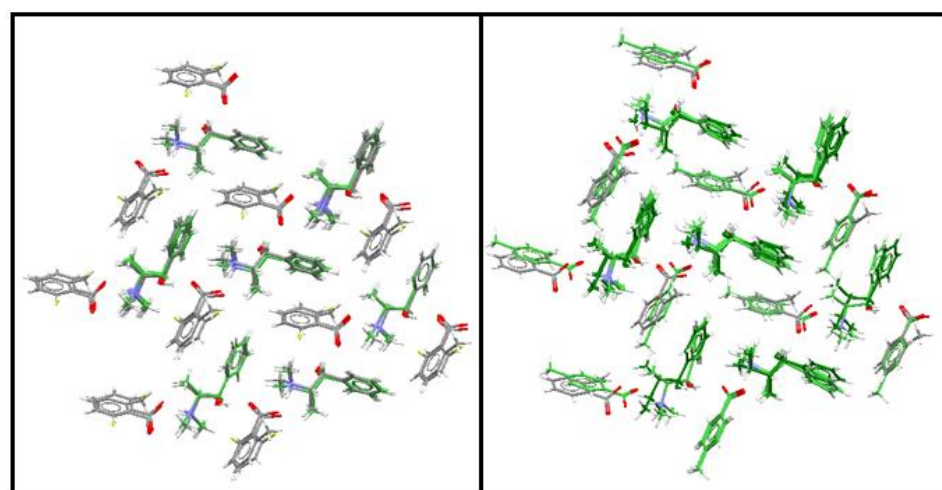
Figure 4.20 Overlay of structures from isostructural group 5. *Ortho-* substituted benzoate salts (left) and *ortho* and *para*-substituted salts (right)

Group two contains 29B, (+/-)methylephedrinium succinate, and 31B, (+/-)methylephedrinium tartrate tartaric acid, both of which have completely different chemical make-up. There is much more extensive hydrogen bonding present in 31B, due to the additional hydrogen bond donor and acceptor atoms available. The addition of the free acid allows the hydrogen bonding to form a two dimensional sheet instead of the one dimensional ribbon consisting of a $C_2^2(9)$ graph-set which is present in 29B. Group four, 23B and 24C, also has additional hydrogen bond potential present due to a water molecule in the hydrogen-sulfate structure. This results in increased dimensionality of the resulting motif, namely a three dimensional network for the hydrogen-sulfate monohydrate structure compared to the more common one dimensional ribbon of the methanesulfonate structure, see Figure 4.21.
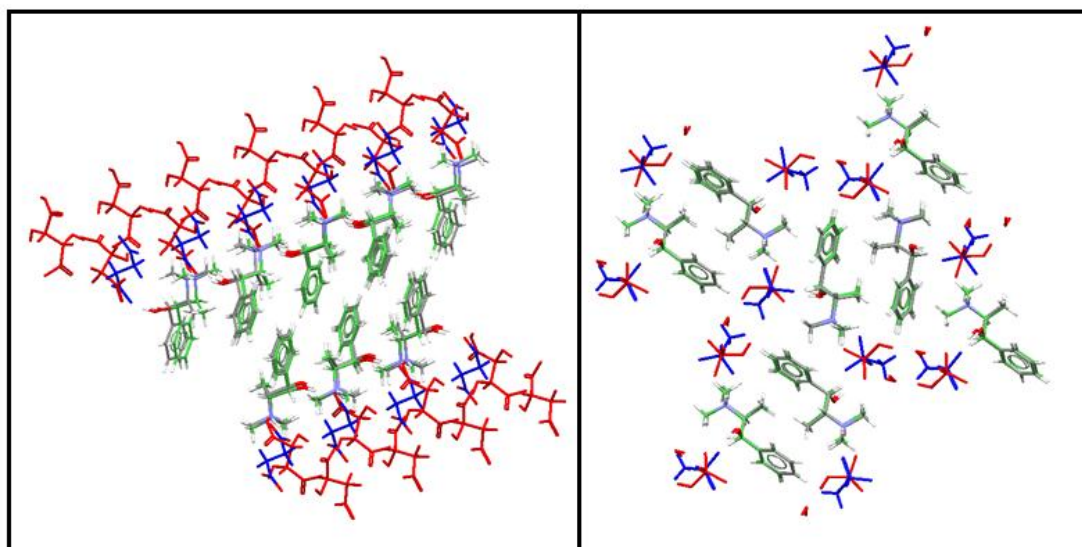
Figure 4.21 Overlay of structures from isostructural groups. Group 2 (left) and group 4 (right)

The analysis of structures 23B, 24C, 29B and 31B (group 4 and 2) suggests that here hydrogen bonding is a consequence of the cation crystal packing array and not the other way round. This conclusion was reached as the cation packing is isostructural and does not change despite the addition of water molecules or free acid molecules and major alterations to the hydrogen bonding motifs. This theory is further validated with analysis of group seven. This group contains a mixture of *ortho-* and *meta-*substituted benzoate salts with varying hydration states and number of cation-anion pairs per asymmetric unit as well as a hydrated adipate salt. The substituted benzoate

salts all form one dimensional chains however the hydrogen bonding patterns are different from structure to structure. In further contrast the adipate salt forms a two dimensional sheet. This point is further emphasised in Figure 4.22, which shows the comparison of the hydrated 2-nitrobenzoate salt with 3-aminobenzoate and the benzoate salt with the hydrated adipate salt.
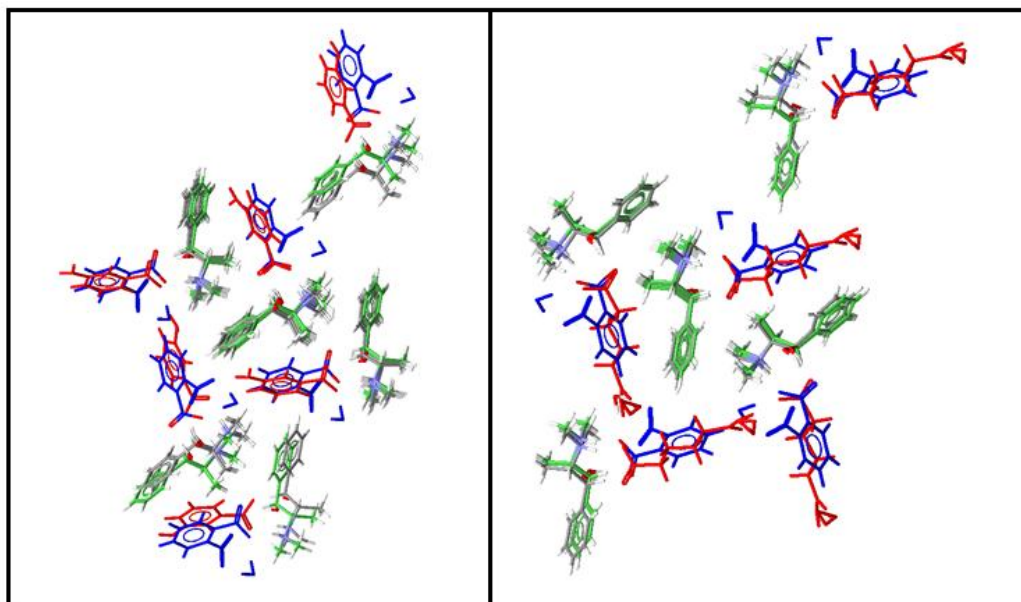
Figure 4.22 Overlay of structures from isostructural group 7. Comparison of hydrated 2-nitrobenzoate salt with 3-aminobenzoate (left) and the hydrated 2-nitrobenzoate salt with the hydrated adipate salt (right)

4.12 Density comparison

4.12.1 General trends

The densities of all the methylephedrinium structures were analysed to see if they conformed to Wallach's rule that predicts greater density for racemic forms of compounds.[3] This was first done using all the crystallographically measured density values and then further investigated using only the enantiopure-racemic pairs that are chemically identical. For comparison of the enantiopure versus racemic densities the values for all the tartrate and mandelate salts were excluded as these introduce complications with the possible formation of diastereomers. Also excluded were the racemic salts that spontaneously resolved to form a conglomerate. The values used are those in Table 4.1 and Table 4.2. The average density of the enantiopure salts is

1.308 g/cm^{-3}, whereas the average density for the racemic salts is approximately 2.0 % greater at 1.334 g/cm^{-3}. Excluding the three iodide salts, as these may distort the average density values, the average density of the enantiopure salts is 1.287 g/cm^{-3}, whereas the average racemic salt density value is only 0.7 % greater at 1.296 g/cm^{-3}. This observation of a small increase in density for racemic forms is in line with previous studies carried out by Brock et al.[21] The four conglomerate forming compounds, 4A, 15A, 17A and 19A do not appear to have enantiopure phases with significantly higher densities than those of comparable species. Indeed the density of 17A (1.217 g/cm^3) is the lowest recorded here. Thus the "extra" stability of these salts is not reflected in their densities.

4.12.2 Evaluation of chemically identical pairs of enantiopure-racemic salts

There are 16 pairs of enantiopure-racemic salts that have identical chemical make-up, for clarity these salts are listed in Table 4.21 below with their corresponding density and melting point values. The average density of the enantiopure salts is 1.288 g/cm^{-3}. The average density for the racemic salts is 1.301 g/cm^{-3}. The difference between the average densities of the 16 pairs is 1.0 %, This observation of a small increase in density for racemic forms is in line with previous studies carried out by Brock et al.[21] However, looking individually at the 16 salt pairs, only eight conform to Wallach's rules where the racemic compound is denser than the enantiopure compound. Of the remainder, there are five pairs that have effectively identical density values (differences of less than 0.005 g/cm^{-3}) and three salt pairs (2-aminobenzoate, 2-nitrobenzoate and 1,2-ethanedisulfonate, samples 2, 6 and 20) where the enantiopure compound is considerably denser than the racemic compound. (In addition, for 24 only one of the two polymorphic racemic structures is denser than the enantiopure form). Figure 4.23, shows the density differences for the 16 enantiopure-racemic salt pairs.

Table 4.21 Density and melting point measurements for chemically identical enantiopure-racemic salt pairs of methylephedrine

Salt	Density (g/cm^{-3})	Melting point (^{o}C)	Salt	Density (g/cm^{-3})	Melting point (^{o}C)
2A	1.259	106.3	2B	1.224	n/a
3A	1.320	n/a	3B	1.315	131.0
6A	1.348	80.1	6B	1.310	n/a
7A	1.222	100.7	7B	1.225	132.7
10A	1.295	138.2	10B	1.292	116.0
11A	1.219	n/a	11B	1.221	143.2
14A	1.227	130.5	14B	1.252	131.0
18A	1.296	161.7	18B	1.323	168.0
20A	1.322	102.2	20B	1.303	n/a
21A	1.274	n/a	21C	1.329	103.9
23A	1.360	n/a	23B	1.402	n/a
24A	1.274	n/a	24B	1.277	n/a
			24C	1.331	n/a
26A	1.259	125.7	26B	1.285	131.3
27A	1.271	113.6	27B	1.315	114.9
34A	1.422	177.0	34B	1.474	184.0
35A	1.241	194.0	35B	1.237	211.8
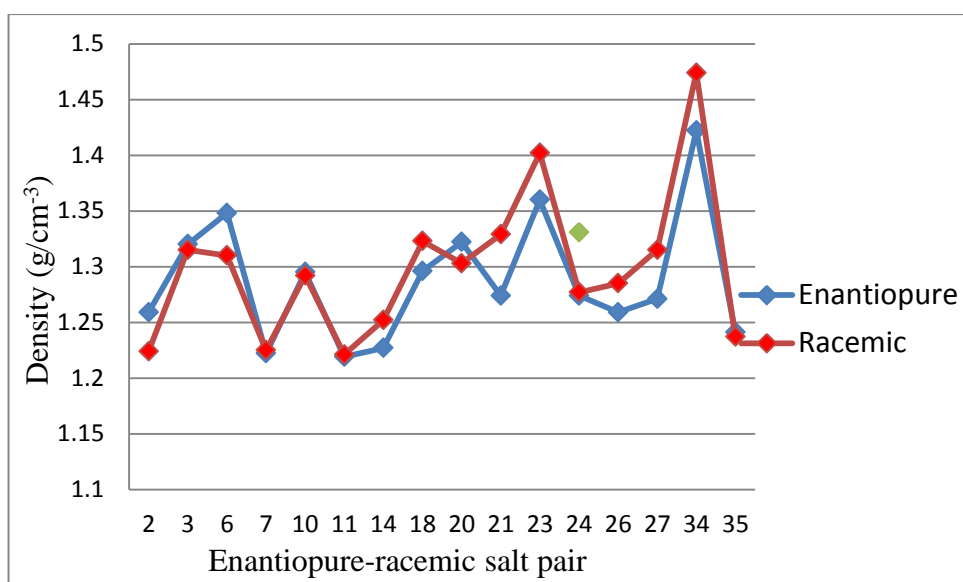
Figure 4.23 Density comparison for 16 chemically identical enantiopure-racemic salt pairs of methylephedrine

To investigate potential structural reasons for the exceptions to Wallach's rule, compounds 2, 6 and 20 were examined for differences between their structures and those of the other salts. The 1,2-ethanedisulfonate salt pair has significant differences in hydrogen bonding between its enantiopure and racemic forms, see graph-set analysis in Table 4.6. Note that the hydrogen bonding in enantiopure 20A propagates

in two dimensions rather than the one dimensional chain seen for racemic 20B. It is proposed that this difference in hydrogen bonding lies at the root of the observed difference in packing efficiency. This ties in with previous suggestions that significant differences in hydrogen bonding may lie at the root of observed differences in behaviour of racemic and enantiopure salts.[4] Packing efficiencies in compounds 20A and 20B are illustrated by void calculations shown in Figure 4.24. The void analysis was set with probe radius 0.5 Å and approximate grid spacing of 1.0 Å. The racemic structure clearly shows more inefficient packing than the enantiopure salt which leads to the structure being considerably less dense. A similar change from 2-dimensional to 1-dimensional hydrogen bonding motifs is also present for 2A and 2B. Of the 16 structural pairs in Table 4.21, only four show changes in the dimensionality of hydrogen bonding between enatiopure and racemic forms. All have higher dimensional hydrogen-bonding for the enantiopure forms. As shown above two of these four show a reversal of Wallach's rule. Another of these four (pair 11) also fails Wallach's rule by having essentially identical densities for its two forms. The only such pair to fit Wallach's rule is pair 14. These *para*-aminobenzoate salts differ from the other three in changing from a 2-dimensional to a 3-dimensional structure. A rule can thus be suggested.

"*Methylephedrinium salts with racemic forms that have 1D hydrogen bonding networks and enantiopure forms that have higher dimensional hydrogen bonding networks will fail Wallach's rule.*"

In contrast, structures 6A and 6B have identical one dimensional hydrogen bonding motifs (the common $C_2^2(9)$ graph-set) and so differences in hydrogen bonding is not the explanation here. Interestingly, 2B and 6B are two of only four of the 16 structures that adopt the 'racemic π' cation pair packing motif. The other two 'racemic π' motifs here are 3B and 7B, both of which are in the group with essentially identical densities for the enantiopure and racemic structures. This suggests that the 'racemic π' packing motif is less packing efficient than the homochiral 'translation' and '2_1' motifs. Looking at the PIXEL calculations in Section 4.10 it may be suggested that the less efficient packing of the 'racemic π'

motif is compensated by the greater energy stabilisation of approximately 6.8 kJ/mol compared to the 'translation' packing motif. Thus we have a second potential rule.

"*Methylephedrinium salts with racemic forms based on the racemic π cation-cation interaction will fail Wallach's rule.*"
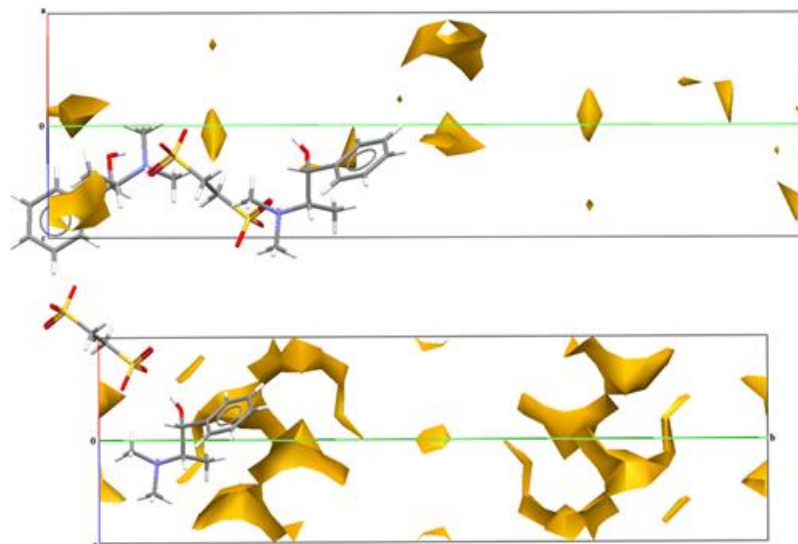
Figure 4.24 Void space analysis for (1R,2S)methylephedrinium 1,2-ethanedisulfonate (top) and (+/-)methylephedrinium 1,2-ethanedisulfonate (bottom)

4.13 Melting point comparison

4.13.1 General trends

Melting point is a simple parameter of interest both as a crucial material characteristic for working and tableting pharmaceutical compounds,[22] and as a feature correlated to other vital parameters such as lattice energy and solubility.[15,16,23-25] For comparison of the enantiopure versus racemic melting points the values for all the tartrate and mandelate salts were excluded as these introduce complications with the possible formation of diastereomers. Also excluded were the racemic salts that spontaneously resolved to form a conglomerate. The values compared are those in Table 4.1 and Table 4.2. The average melting point of the enantiopure salts is 126.4 $^{\circ}$C, whereas the average melting point for the racemic salts is approximately 10 $^{\circ}$C greater at 136.2 $^{\circ}$C. This is consistent with the racemic

phases being thermodynamically favoured over the enantiopure phases – as would be expected by combining Wallach's rule with Kitaigorodskii's "principle of close packing".[26]

4.13.2 Evaluation of chemically identical paris of enantiopure and racemic melting points
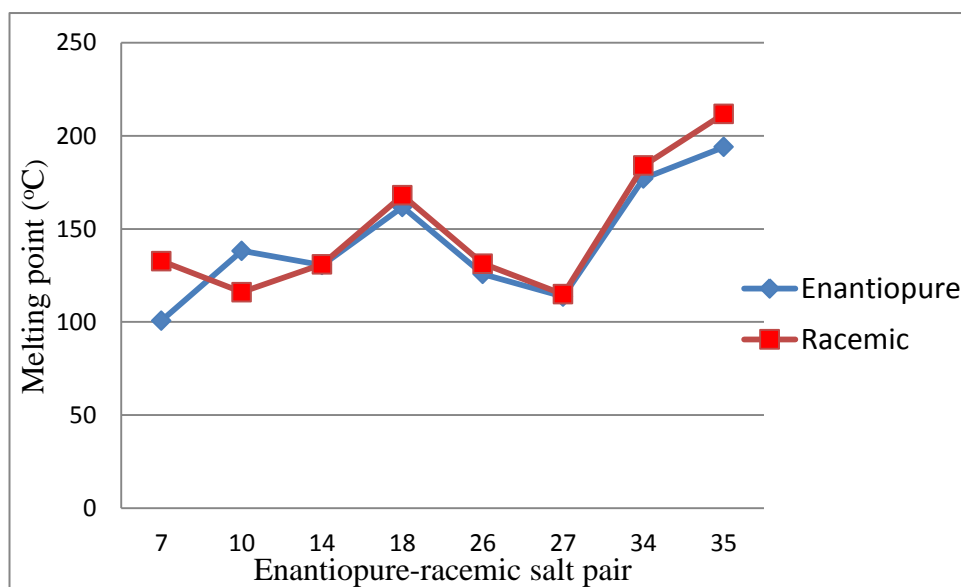
Figure 4.25 Melting point comparison for 8 chemically identical enantiopure-racemic salt pairs of methylephedrine

There are eight pairs of enantiopure-racemic salts that have identical chemical make-up and where their melting point was recorded (7, 10, 14, 18, 26, 27, 34 and 35). This number is greatly reduced from the 16 identical pairs compared with respect to density above as one if not both of the compounds of the excluded pairs are known to undergo phase changes (determined by X-ray powder analysis). Thus there is no guarantee that the measured melting point corresponds to the determined single crystal structure in these cases. The melting points are listed in the Table 4.21 above. The average melting point of the enantiopure salts is 142.7 ^{o}C, which is considerably higher than the average melting point of all the enantiopure salts; however this value does not contain any hydrate structures which generally have a lower melting point. The average melting point for the racemic salts is 148.7 ^{o}C again a value that is considerably higher than the collective value noted for all the racemic salts. The difference between the average melting points of the eight pairs is approximately 6.0

^{o}C. Looking individually at the eight salt pairs, Figure 4.25, only one does not conform to the expectation that the racemic compound should have a higher melting point than the enantiopure compound. The exception is the pair of 3-fluorobenzoate salts, where enantiopure salt (10A) has a melting point of 138.2 ^{o}C and the racemic salt (10B) has a melting point of 116.0 ^{o}C. Although there is no obvious structural reason for this, it should be noted that 10A and 10B have essentially the same density and thus do not strictly obey Wallach's rules.

4.14 Conclusions

This chapter investigated 65 salts which included 28 pairs of enantiopure and racemic methylephedrinium salts. Of the 28 pairs, 16 produced enantiopure/racemic crystal phases of identical chemical make-up, six pairs produced salts that crystallised as different chemical entities and four pairs had racemic species that spontaneously resolved to give conglomerates. Three of the four conglomerates had anions derived from *para-* substituted benzoic acids, but did not appear to be otherwise structurally different from the other salts or to have exceptionally high densities. Four of the racemic salts produced two different crystal phases that were analysed. The dataset also included 12 salts where only one of the enantiopure-racemic crystals has been obtained and three mandelate salts to give the total of 64 unique crystal structures that were studied. Exploration of hydrogen bonding, by graph-set analysis, and crystal packing was performed to establish similarities between structures. 52 of the salts were anhydrous and 12 hydrated. 60 of the salts reacted to give a one to one, cation to anion salts and four salts reacted to give a two to one cation to anion salt.

All the structures analysed formed contact cation-anion pairs, with no direct cation-cation hydrogen bonded contact ion pairs. Anion-anion interactions are present in all structures with an anion that possesses a classic hydrogen bond donor except for the two hydrogen-sulfate monohydrate salts. In all cases, except 9A and 21B, the cation-anion contact involved the cation's positively charged NH group as a hydrogen-bond

donor to the COO⁻ group of the anion. All salt structures conform to Etter's rule,[2] with intramolecular bonds forming in preference to intermolecular bonds where possible. Two reported products, 31B and 36C, produced a co-crystal with the presence of a free acid molecule or I_2 in the asymmetric unit. The commonest graph-set motif of $C_2^2(9)$ was found to be present in 38 of the 64 independent structures, and was flexible enough to include both OCO and OSO functionalities. Using PIXEL energy calculations it was determined that although significant the $C_2^2(9)$ motif did not account for the majority of the total energy of the compounds and other lesser intermolecular interactions need to be considered. The equivalent $C_2^1(7)$ motif was present in all seven halide structures. The network growth varied with 47 salts producing 1D hydrogen bonded chains, 12 salts producing 2D hydrogen bonded sheets and five salts producing 3D hydrogen bonding networks. The methylephedrinium cation is seen to adopt one of three different conformations and there are also six common cation-cation pair packing motifs, with 21 salt structures adopting seven different 3D isostructural groupings with respect to cation packing. Hydrogen bonding differences (and indeed differing hydration states and chemical phases) are seen within some of the otherwise isostructural cation arrays.

For the chemically identical enantiopure-racemic pairs, densities and where possible melting points were collated, in order to test the validity of Wallach's rules.[3] In general, the overall trend in densities agreed with Wallach's rules with the average racemic density being higher than the average enantiopure density. This though may be a somewhat counterfeit result – we can access crystals in cases where the stability of the enantiopure form is significantly less than that of the racemic form (as enantiopure starting material is available) but we cannot access structures of the racemic form if the enantiopure form is significantly more stable. As our comparison can only be made when both forms are available, there will always be a bias towards more stable (and hence presumably denser) average values for the set of racemic compounds. Close examination of the individual pairs finds that exceptions to Wallach's rule are common, with eight of 16 pairs failing to have a more dense racemic structure. Overall, there is thus little evidence to support the notion that

packing racemic ions with achiral ones must give a denser structure than packing enantiopure and achiral ions. One of the six observed cation packing motifs (the racemic π motif) is closely associated with failure of Wallach's rule. Also associated with failure of Wallach's rule is a significant change in hydrogen bonding dimensionality where racemic forms have 1D hydrogen bonded networks but enantiopure forms have higher dimensionality networks. Thus we have highlighted two different details of array structure that may explain inefficient packing.

4.15 References

(1)   Kennedy, A. R.; Morrison, C. A.; Briggs, N. E. B.; Arbuckle, W. *Crystal Growth & Design* **2011**, *11*, 1821.
(2)   Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.
(3)   Wallach, O. *Liebigs Ann Chem* **1895**, *286*, 90.
(4)   Kinbara, K.; Hashimoto, Y.; Sukegawa, M.; Nohira, H.; Saigo, K. *J. Am. Chem. Soc.* **1996**, *118*, 3441.
(5)   Jacques, J.; Collet, A.; Wilen, S. H. *Enantiomers, Racemates, and Resolutions*; Wiley: New York, 1981.
(6)   Jacques, J.; Leclercq, M.; Brienne, M. J. *Tetrahedron* **1981**, *37*, 1727.
(7)   Duddu, S. P.; Grant, D. J. W. *Pharm. Res.* **1994**, *11*, 1549.
(8)   Van der Sluis, P.; Kroon, J. *J. Cryst. Growth* **1989**, *97*, 645.
(9)   Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; 2nd ed.; Wiley: Chicchester, 2000.
(10)   Metrangolo, P.; Resnati, G. *Chemistry-a European Journal* **2001**, *7*, 2511.
(11)   Dunitz, J. D.; Gavezzotti, A. *Angewandte Chemie-International Edition* **2005**, *44*, 1766.
(12)   Hursthouse, M. B.; Montis, R.; Tizzard, G. J. *Crystengcomm* **2010**, *12*, 953.
(13)   Cabeza, A. J. C.; Day, G. M.; Motherwell, W. D. S.; Jones, W. *Crystal Growth & Design* **2007**, *7*, 100.
(14)   Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41*, 466.
(15)   Li, Z. J.; Ojala, W. H.; Grant, D. J. W. *J. Pharm. Sci.* **2001**, *90*, 1523.
(16)   Collier, E. A.; Davey, R. J.; Black, S. N.; Roberts, R. J. *Acta. Cryst.* **2006**, *A62*, 498.
(17)   Frinch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyoto, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.;

Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Pittsburgh PA, 2003.

(18)   Bacchi, S.; Benaglia, M.; Cozzi, F.; Demartin, F.; Filippini, G.; Gavezzotti, A. *Chemistry-a European Journal* **2006**, *12*, 3538.

(19)   Gibson, E. K.; Winfield, J. M.; Muir, K. W.; Carr, R. H.; Eaglesham, A.; Gavezzotti, A.; Parker, S. F.; Lennon, D. *PCCP* **2009**, *11*, 288.

(20)   Johnston, A.; Bardin, J.; Johnston, B. F.; Fernandes, P.; Kennedy, A. R.; Price, S. L.; Florence, A. J. *Crystal Growth & Design* **2011**, *11*, 405.

(21)   Brock, C. P.; Schweizer, W. B.; Dunitz, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 9811.

(22)   Strahl, P. H.; Wermuth, C. G. *Handbook of Pharmaceutical Salts: Properties, Selection and Use*; VHCA: Zurich, 2008.

(23)   Black, S. N.; Collier, E. A.; Davey, R. J.; Roberts, R. J. *J. Pharm. Sci.* **2007**, *96*, 1053.

(24)   Guerrieri, P.; Rumondor, A. C. F.; Li, T.; Taylor, L. S. *AAPS PharmSciTech* **2010**, *11*, 1212.

(25)   Anderson, B. D.; Conradi, R. A. *J. Pharm. Sci.* **1985**, *74*, 815.

(26)   Kitaigorodskii, A. I. In *Organic Chemical Crystallography*; Consultants Bureau: New York, 1961.

Chapter 5

Hydrates and their occurrence of formation

within the dataset

5.1 Introduction

It is thought that one third of pharmaceutical drugs are capable of forming hydrates.[1] As hydrates can have substantially different physicochemical properties than their equivalent anhydrates (see Section 6.4, Section 6.5, and Section 6.8) this makes studying the formation of pharmaceutical hydrates extremely important. Drug compounds may come into contact with water during many of the drug development stages, such as crystallisation, or indeed simply through the storage of the drug compound in humid atmospheres.[2] Two examples of unexpected hydrate formation with pharmaceutical consequences are; (a) A study by Kiang et al.[3] which investigated problems with a naphthyridinyl derivative intended for the treatment of osteoporosis. When the packaged drug was stored at raised humidity the tablets swelled to include up to five water molecules per asymmetric unit. The inclusion of the water molecules greatly reduced the stability of the drug, as was evident through cracking of the tablet, and; (b) A study on the non-benzodiazepine hypnotic drug zopiclone which showed that problems with stored tablets degrading was due to formation of an unexpected hygroscopic crystalline phase upon incomplete or inefficient drying of the original product.[4]

In pharmaceutical hydrates water molecules usually occupy definite positions within the crystal lattice by forming hydrogen bonds. Often the inclusion of water molecules allows for more ideal hydrogen bond geometries and therefore lower lattice energies.[5] Hydrates differ greatly in comparison to their anhydrous equivalent as the inclusion of water molecules leads to a new unit cell, which in turn will lead to different physical properties such as density, melting point, solubility and dissolution rate.[2] It is also possible to get different polymorphs of hydrates, which themselves will have different physical properties to one another, an example of this is succinyl sulfathiazole monohydrate[6] which has two polymorphic forms with different physical properties. It should be noted that two hydrated compounds are only deemed polymorphs when there is the same stoichiometry of water molecules present in both.

The solubility of an anhydrous compound in water has an equilibrium constant K_s as shown in equation 5.1. At the same pressure and temperature the equilibrium solubility of a hydrate has the equilibrium constant K_s', as shown in equation 5.2.

$$A(solid) \underset{}{\overset{K_s}{\rightleftharpoons}} A(aqueous) \qquad \text{(Eqn 5.1)}$$

$$A.mH_2O(solid) \underset{}{\overset{K_s'}{\rightleftharpoons}} A(aqueous) + mH_2O \qquad \text{(Eqn 5.2)}$$
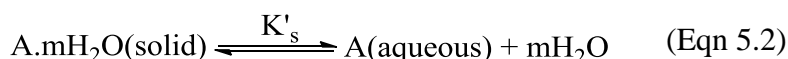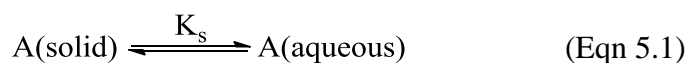
The formation of a hydrated crystal from an anhydrous crystal can be represented by the equilibrium constant K_h in equation 5.3, where K_h is dependent on the thermodynamic activities, a, of the anhydrous and hydrated crystals and free water, equation 5.4.

$$A(solid) + mH_2O \underset{}{\overset{K_h}{\rightleftharpoons}} A.mH_2O(solid) \qquad \text{(Eqn 5.3)}$$

$$K_h = \frac{a[mH_2O(solid)]}{a[A(solid)]a[H_2O]^m} \qquad \text{(Eqn 5.4)}$$

As the hydrated compound is already interacting with water in the solid state upon dissolution the free energy that is released is less than seen for the anhydrous compound.[7] Thus theory predicts that the anhydrous form of a compound should always have a greater aqueous solubility than the corresponding hydrate, when both are subject to the same conditions.

Hydrates can be classified in many different ways. Vippagunta et al.[8] split all hydrates into two distinct categories of stoichiometric and non-stoichiometric hydrates. The first describes structures with well-defined water content which results in a unique crystal structure that is different from the anhydrous form or indeed any other hydrate forms. The latter describes hydrates where water is present in variable compositions and where changing the composition leads to little change to the crystal structure, except for slight expansion of the crystalline arrangement to make room for the additional water molecules. Alternatively, Morris[9] distinguished hydrates by the following three classes,

- Class I: isolated site hydrates, where the water molecules are located on unique and isolated crystallographic sites.
- Class II: channel or planar hydrates, where the water molecules are located next to each other forming channel or planar networks.
- Class III: ion coordinated hydrates, where the water molecule coordinates to the cation or anionic centre.

Generally speaking class I hydrates are stoichiometric, class II hydrates are non-stoichiometric and class three hydrates fit into both categories.

Thus within a crystal water can have various different functions, it can participate in hydrogen bonding networks or occupy spaces within the lattice and have no strong interactions with the parent molecules or it can be used as a ligand to complete the coordination around a metal ion.[10] This can lead to the water having various different chemical environments. These environments can be looked at in terms of the hydrogen bonding present and Figure 5.1 shows the nine different hydrogen bonding environments that are possible.[11]
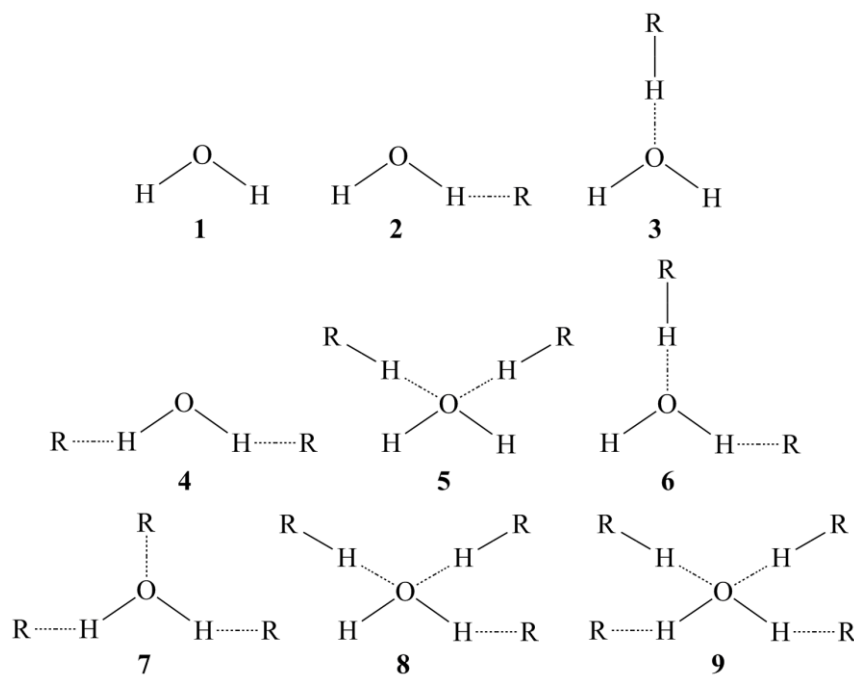
Figure 5.1 Nine different water environments possible with varying hydrogen bonding present, adapted from reference 11[11]

A study by Gillon et al.[11] found the most common solid state water environments were those where three or four hydrogen bonds were present between water and its neighbouring molecules.

Many studies have been carried out on organic hydrates present in the CSD to try and ascertain when a compound is most likely to form a hydrate.[5,11-15] One of the most recent of these studies performed by Infantes et al.[5] found that 6.6 % of the organic entries within the CSD were hydrates and 14 % of bioactive compounds were hydrates. This percentage is greatly increased when looking solely at salt compounds. A study by Haynes et al.[14] investigated hydrate formation in NH^+ containing salts of pharmaceutically acceptable counterions and found that hydrates were present in 22.5 % of CSD entries. It is known that those metal cations commonly used as salt formers in the pharmaceutical industry typically precipitate from solution as polyhydrates.[16-18] Indeed this is one reason why the otherwise easy to manufacture sodium salt may not be the chosen commercial form.[1] Similar prior knowledge about other counterions and their tendency to form hydrates would be helpful.

Studies have looked at functional groups present within compounds,[15] hydrogen bond donor-acceptor ratios,[19] and choice of counterion[11] in salts with a view to determining effect on hydrate formation. Following the guidelines of Etter's rules[20] which states that all good hydrogen bond donor and acceptor atoms will be involved in hydrogen bonding a study by Desiraju[21] concluded that the inclusion of water helps stabilise compounds where there is an imbalance of hydrogen bond donor and acceptor atoms present. Infantes et al.[19] contradict this statement and state that their studies showed no link between hydrate formation and donor acceptor ratios, instead they propose a link between hydrate formation and the sum of the average donor acceptor counts for functional groups along with increased polar surface area. This research also found no correlation between hydrate formation and molecular weight a theory based on higher molecular weight compounds having larger gaps available within their structure for water molecules to occupy.[19]

5.2 Analysis of hydrate formation looking at cation and anion contributions

Upon analysis of the 255 crystal structures that make up the present dataset it was discovered that 57 compounds were hydrates (containing either H_2O, H_3O^+ or $H_5O_2^+$). Hydrate formation in 22.4 % of the salts is directly comparable to the 22.5 % of hydrated NH^+ salts present in the CSD.[14] The types of hydrates present within the dataset are illustrated in Figure 5.2. The most common hydrate formed is monohydrate, occurring in 73.7 % of the hydrated compounds. Fifty-two of the hydrated compounds contain uncharged water molecules and four compounds contain charged H_3O^+ or $H_5O_2^+$ ions,[22] the remaining compound forms a hydrate species containing half a neutral water molecule and a hydronium ion per asymmetric unit. The three dihydronium compounds and the dihydronium hemihydrate all contain the same counterion, 1,2-ethanedisulfonate, with the hydronion compound containing a sulfate counterion. This is as expected as protonated water molecules are typically only found in the solid state of organic species when they are derived from strong acids, such as sulfonic acids.[22-24]
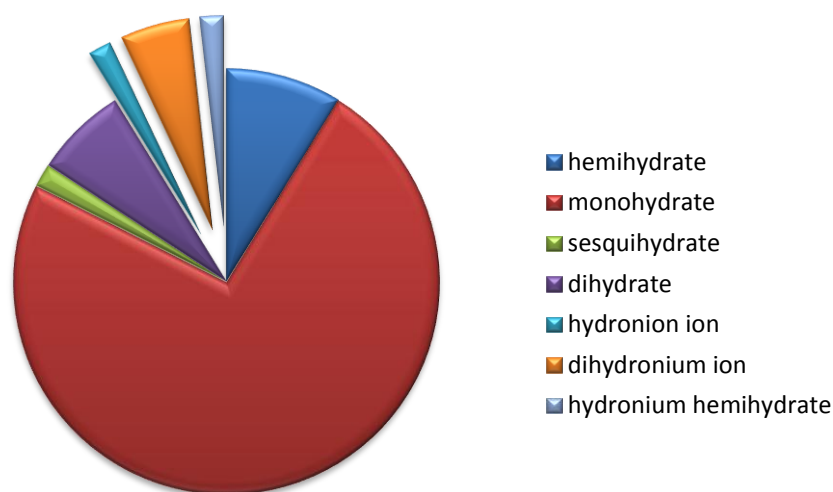
Figure 5.2 Chart of different hydrate structures present within the 57 structure dataset. The extended slice represents the charged H_3O^+ or $H_5O_2^+$ containing species

When looking at the hydrate formation in terms of what type of salt is more likely to form a hydrate, the data set was separated firstly by cation and secondly by anion

present within the salt. Figure 5.3 shows the number of anhydrous and hydrated salts derived from primary, secondary and tertiary amines. There is a slight decrease in hydrate formation with the trend of primary > secondary > tertiary, with 32.3, 27.7 and 25.8 % hydrate formation respectively. This observation is in line with literature expectations.[14] Looking individually at the different cations that make up the dataset, see Figure 5.4, there is a high occurrence of hydrate formation for the base (-)pseudoephedrine with 37.5 % forming hydrates, a figure well above the average for salts of secondary amines. It was found that the formation of hydrates was more common for enantiopure than for racemic methylephedrinium salts, see Section 4.2. The salts of (-)pseudoephedrine are also enantiopure therefore suggesting a pattern where enantiopure salts pack less efficiently allowing for more frequent formation of hydrates. Phenylethylammonium salts have a lower than average occurrence of hydrates for salts derived from primary amines with only 12.0 % of these salts forming hydrates. This low occurrence is also seen for the methylphenethylaminium salts with only 6.3 % of the salts forming a hydrated species. It is tempting to ascribe these latter results to the lack of additional polar groups on the amine. It has been suggested that polar groups favour hydrate formation.[5] The other two cations in this study to not contain any additional polar groups are phenylpropylammonium and dimethylphenethylaminium. These salts discredit the above theory as both have an average percentage of hydrate formation with regards to their cation class.
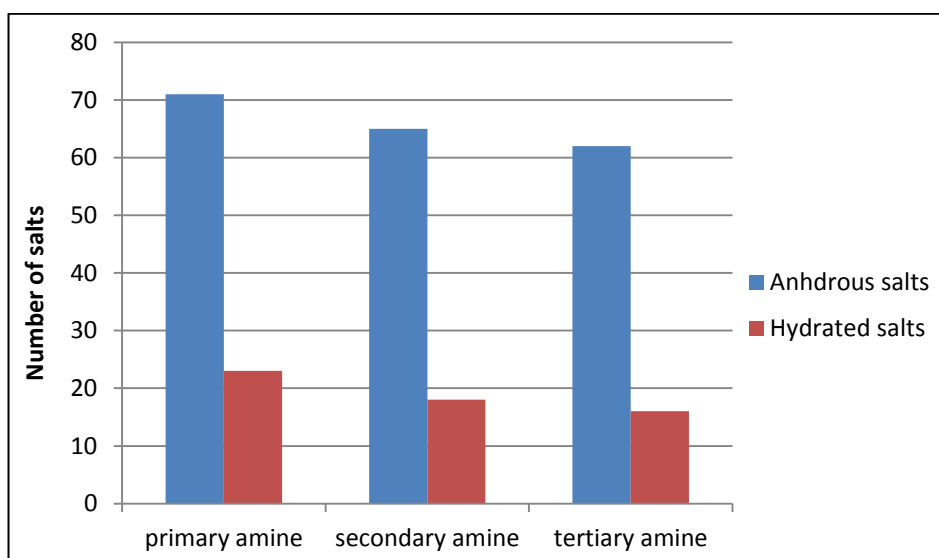
Figure 5.3 Number of primary secondary and tertiary amine anhydrous and hydrated structures present within the dataset
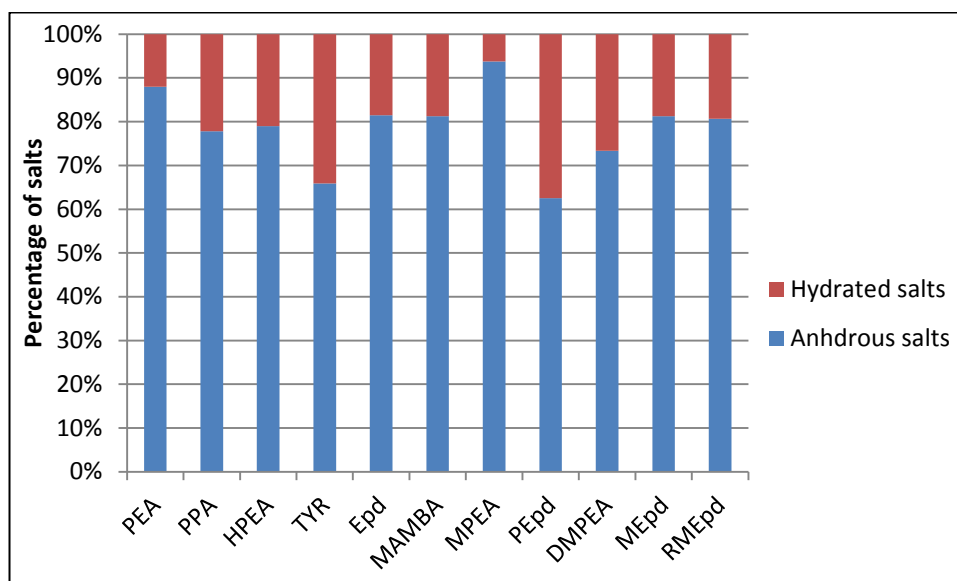
Figure 5.4 Percentage of anhydrous and hydrated structures for the cations present within the dataset

Figure 5.5 shows the number of anhydrous and hydrated salts formed from different counterion types. The category 'spherical' includes the spherical counterions sulfate, perchlorate, tetrafluoroborate and phosphate. There is a very low occurrence of hydrated compounds for benzoate and mandelate containing salts, and no hydrates present at all for halide containing salts. The presence of no hydrated halide salts contradicts the literature CSD study where a larger percentage of halide salts form hydrates than do "general" counterions.[14] The remaining three categories; sulfonate, spherical and dicarboxylate all contain a significant number of hydrated compounds with 28.6, 40.9 and 32.8 % of the compounds forming a hydrated species respectively. This observation and the low percentage of hydrate formation with benzoate and mandelate ions is possibly as expected, as it is believed that the occurrence of hydrate formation generally rises with increasing number of polar groups on a molecule and with increasing charge.[5]

Looking individually at the different anions that make up the dataset, Figure 5.6, it can be seen that the tartrate counterions have an extremely high occurrence of hydrated salts. This has been seen before in a study of salts of simple piperidine derivatives.[25] Within the dicarboxylate group there appears to be less chance of a hydrated compound forming with the presence of an intramolecular hydrogen bond.

Thus the effective removal of hydrogen bond donors and acceptors through prior intramolecular bond formation leads to lower than expected occurrence of hydrate formation for malonate and maleate salts. Alongside the halides, salts containing methanesulfonate and ethanesulfonate counterions as well as several benzoate derived salts are all anhydrous.
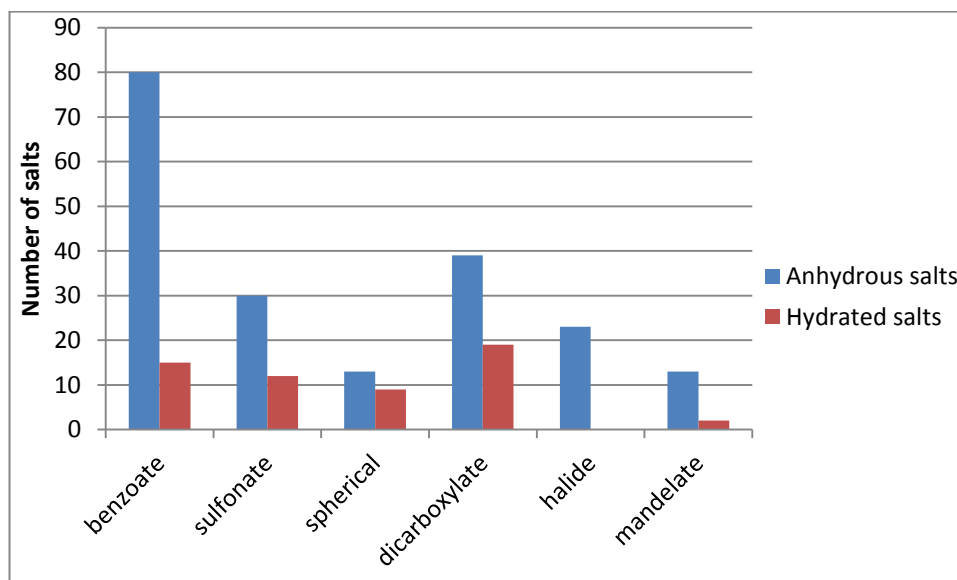
Figure 5.5 Number of anhydrous and hydrated structures from different anion classes
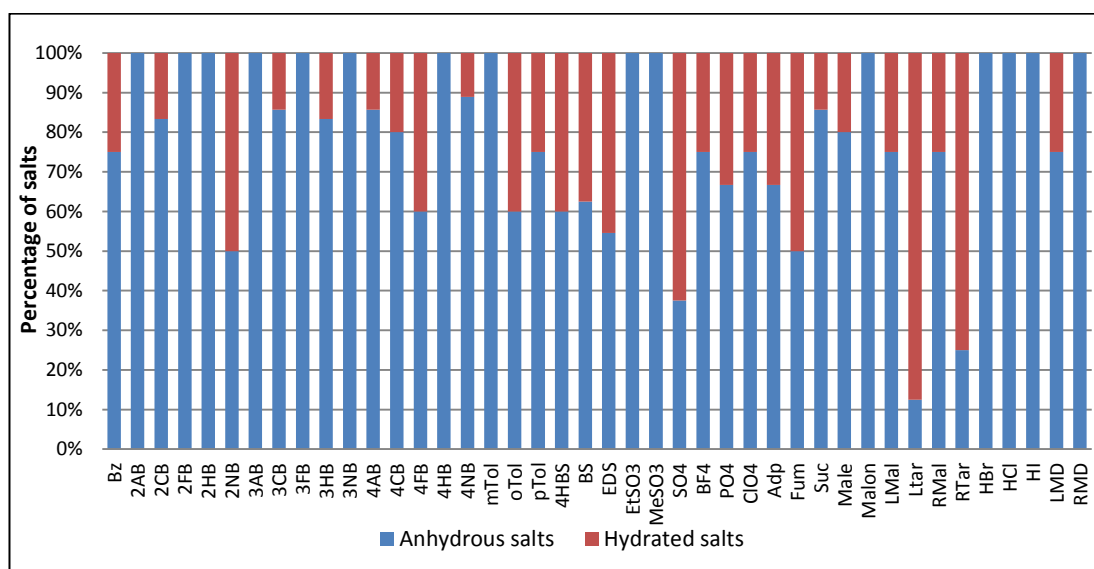
Figure 5.6 Percentage of anhydrous and hydrated structures for the anions present within the dataset

137

5.3   Water environment in hydrates

The bonding environment about the water molecules of the hydrated compounds was investigated to see if there were a common number of hydrogen donor and acceptor bonds present. It was found that the majority of the water molecules interacted with three other molecules by donating two hydrogen bonds and accepting one hydrogen bond. The percentage of structures that adopt each of the various water environments can be seen in Figure 5.7. The structure of (+/-)methylephedrinium hydrogen-sulfate monohydrate shows an example of the water acting as a typical two donor, one acceptor group which results in conjoining $R_4^4(12)$ graph-sets, see Figure 5.8. This result is in line with a literature study which found the most common environment of water molecules present within the CSD to be that involving two donor interactions and one hydrogen bond acceptor interaction.[11]
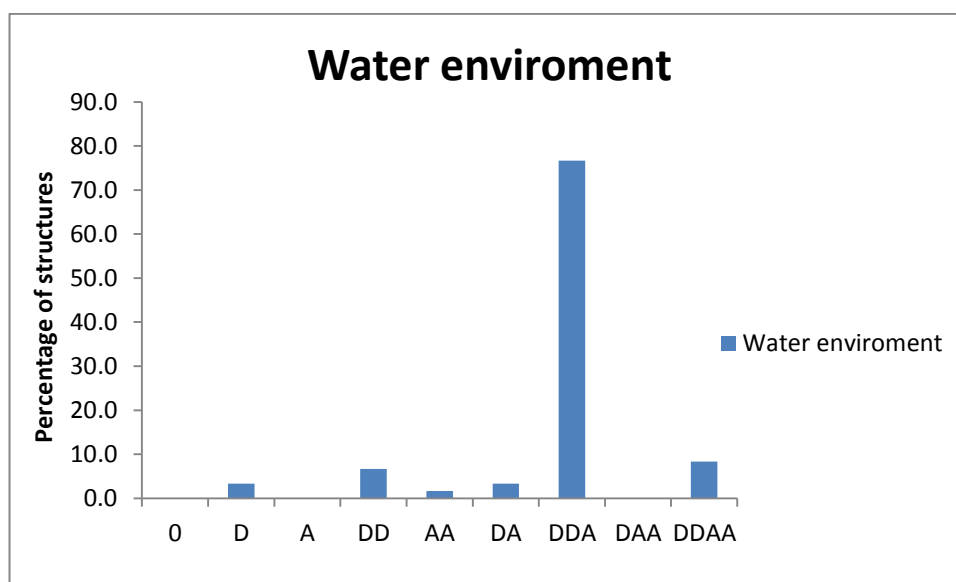
Figure 5.7 Bonding environments around water molecules. **D** is for a hydrogen bond donor and **A** is for water as a hydrogen bond acceptor.
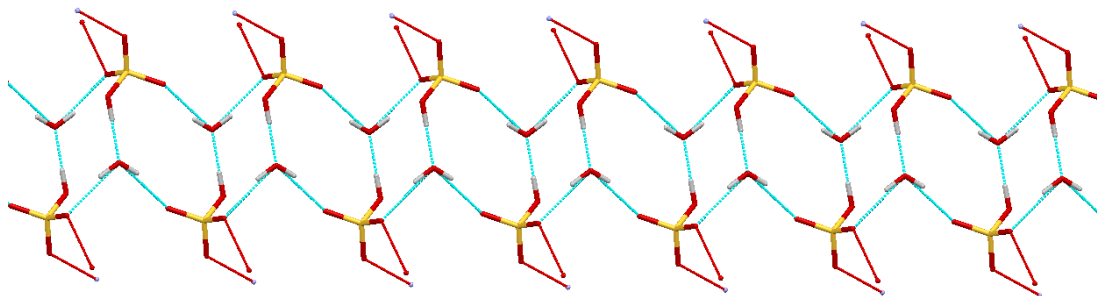
Figure 5.8 Example of water acting as a two donor one acceptor (DDA) in structure (+/-)methylephedrinium hydrogen-sulfate monohydrate

The complete dataset of structures, both anhydrous and hydrated were also analysed for their hydrogen bond donor/acceptor ratio in order to ascertain whether hydrate formation could be predicted solely by looking at the cation and anion present in the resulting salt. For the analysis the hydrogen bond donor and acceptor atoms were counted by assuming a one to one reaction between the cation and anion. In the spirit of prediction from zero prior knowledge, all structures that formed co-crystals were counted as if the free acid molecule was not present, all dicarboxylate and 'round' anions were assumed to be singly deprotonated and all 1,2-ethanedisulfonate anions were assumed to be doubly deprotonated. Only reliable hydrogen bond donor and acceptor atoms were counted, see Table 1.6, Sections 1.9.

Figure 5.9 shows the spread of the structures over all the donor/acceptor ratios from 1/3 (one donor and three acceptors) to 3.0 (three donors and one acceptor). The most common donor/accept ratio for both anhydrous and hydrated structures is one. However, where there is a stepwise decrease either side of one for the anhydrous salts, for the hydrated salts the stepwise decrease is only present with donor/acceptor ratios of less than one. This is made more apparent upon examining Figure 5.10, which shows the percentage of structures that fall into four distinct categories of donor/acceptor ratios. From these results it can be concluded that hydrate formation is most probable when there is an excess of hydrogen bond acceptor atoms present and unlikely if there is an excess of hydrogen bond donor atoms present. Of the five

hydrate structures with a donor acceptor ratio greater than one, four are benzoate derived monohydrate salts. This result is unusual in itself as benzoate derived salts have a very low occurrence of hydrate formation. (see Figure 5.5). As water primarily occurs where there is an excess of hydrogen bond acceptors, this could indicate that here the primary structural role of water molecules is to provide extra hydrogen bond donors. This observation agrees with the common DDA hydrogen bond environment observed surrounding water molecules (see Figure 5.7 above) and with the study by Desiraju.[21] The observation re acceptor to donor ratios also fits with the earlier statement (Figure 5.6) that "spherical", sulfonate, dicarboxylate and especially tartrate salts have high occurrences of hydrate formation. All these ion types have high numbers of acceptor atoms and few donor atoms.
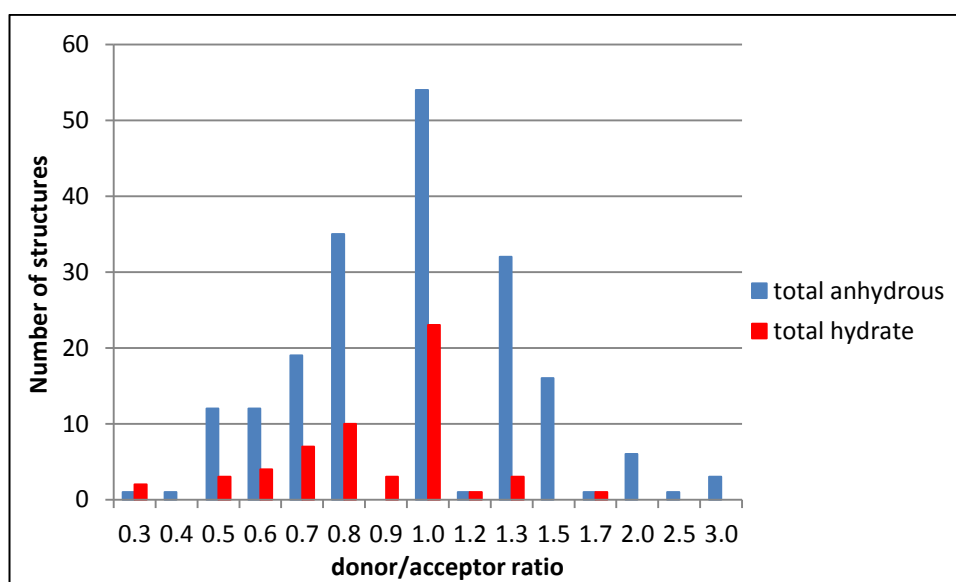
Figure 5.9 Calculated donor/acceptor hydrogen bond ratio for all anhydrous and hydrated structures from 1/3 (equivalent of one donor and three acceptors) to three (equivalent of three donors and one acceptor)
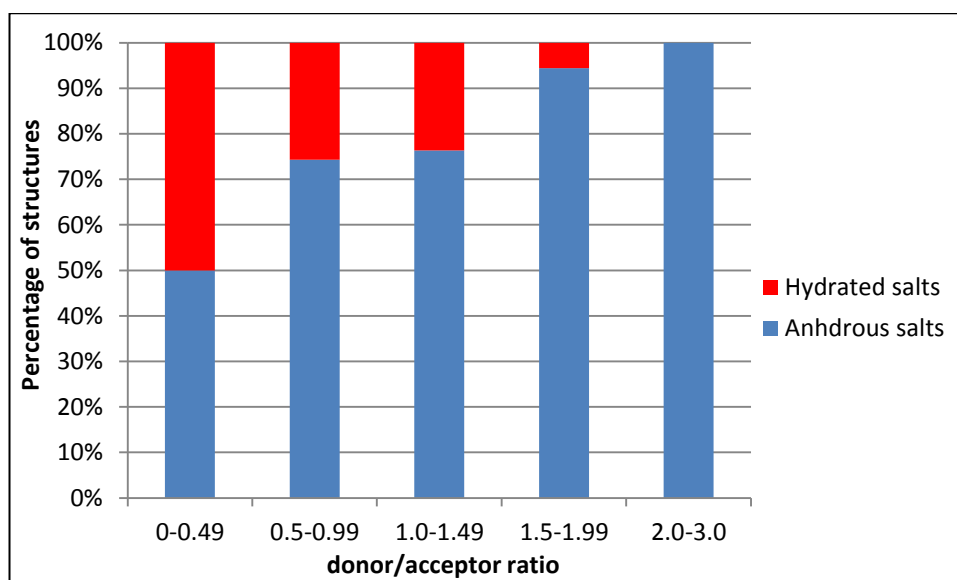
Figure 5.10 Percentage of anhydrous and hydrated structures in categories of hydrogen bond donor/acceptor ratios

5.4   Common graph-set encountered in hydrate structures

The discrete, chain and ring graph-sets present in the hydrate structures were analysed to see if any common graph-sets link the structures together. Only graph-sets involving water molecules were looked at as the other graph-set present between cation and anion molecules have already been discussed elsewhere. Table 5.1 shows the different hydrogen bond interactions that are observed involving water and how many of the 57 hydrated compounds have such interactions present. Hydrogen bonds where water acts as a donor to anion are present in all hydrate structures but water donating to cation type hydrogen bonds are only present in 11 of the 48 structures where this interaction is possible. When water acts as a hydrogen bond acceptor and the anion has the ability to act as a hydrogen bond donor there is very little difference between the number of compounds with cation to water and anion to water hydrogen bonds, 50.0 and 54.4 % respectively. Thus contrary ion charge appears to decrease the likelihood of cations acting as acceptors but not decrease the likelihood of anions acting as donors.

Table 5.1 Hydrogen bond interactions observed involving water

H-bond interaction (donor...acceptor)	No. of compounds	% of all hydrates	No. of compounds where possible	% of possible hydrates
anion...water	15	26.3	30	50.0
cation...water	31	54.4	57	54.4
water...water	3	5.3	57	5.3
water...anion	57	100.0	57	100.0
water...cation	11	19.3	48	22.9

The majority of hydrate compounds contain both discrete and chain graph-set that involve the water molecules as can be seen in Figure 5.11, however there are a significant number of compounds that only contain discrete graph-set. Upon visual analysis of these structures it was determined that these compounds tend to use water as a space-filler, with the water occupying sites where there would otherwise be a void space or in the case of PPAMale the water occupies the empty space between the layers of cation and anions, see Figure 5.12. It should be noted that TYRAdp is a co-crystal.

Graph-set that utilise water molecule

- discrete only
- discrete and ring
- discrete and chain
- discrete, chain and ring
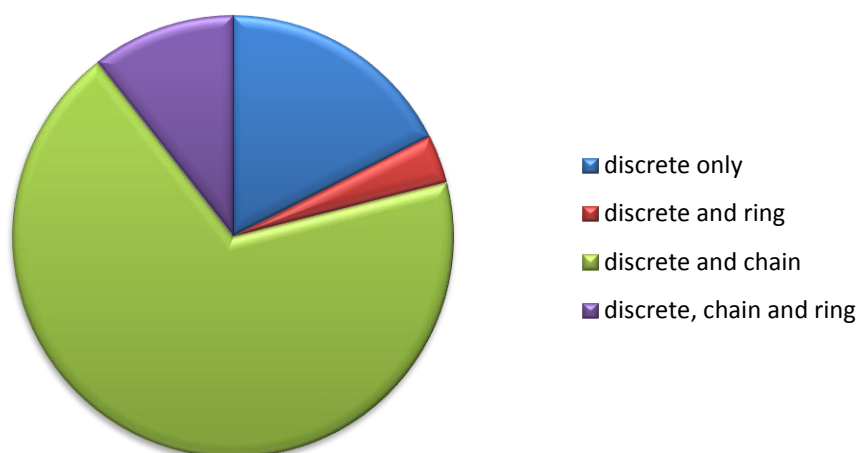
Figure 5.11 Types of graph-set present within hydrate structures
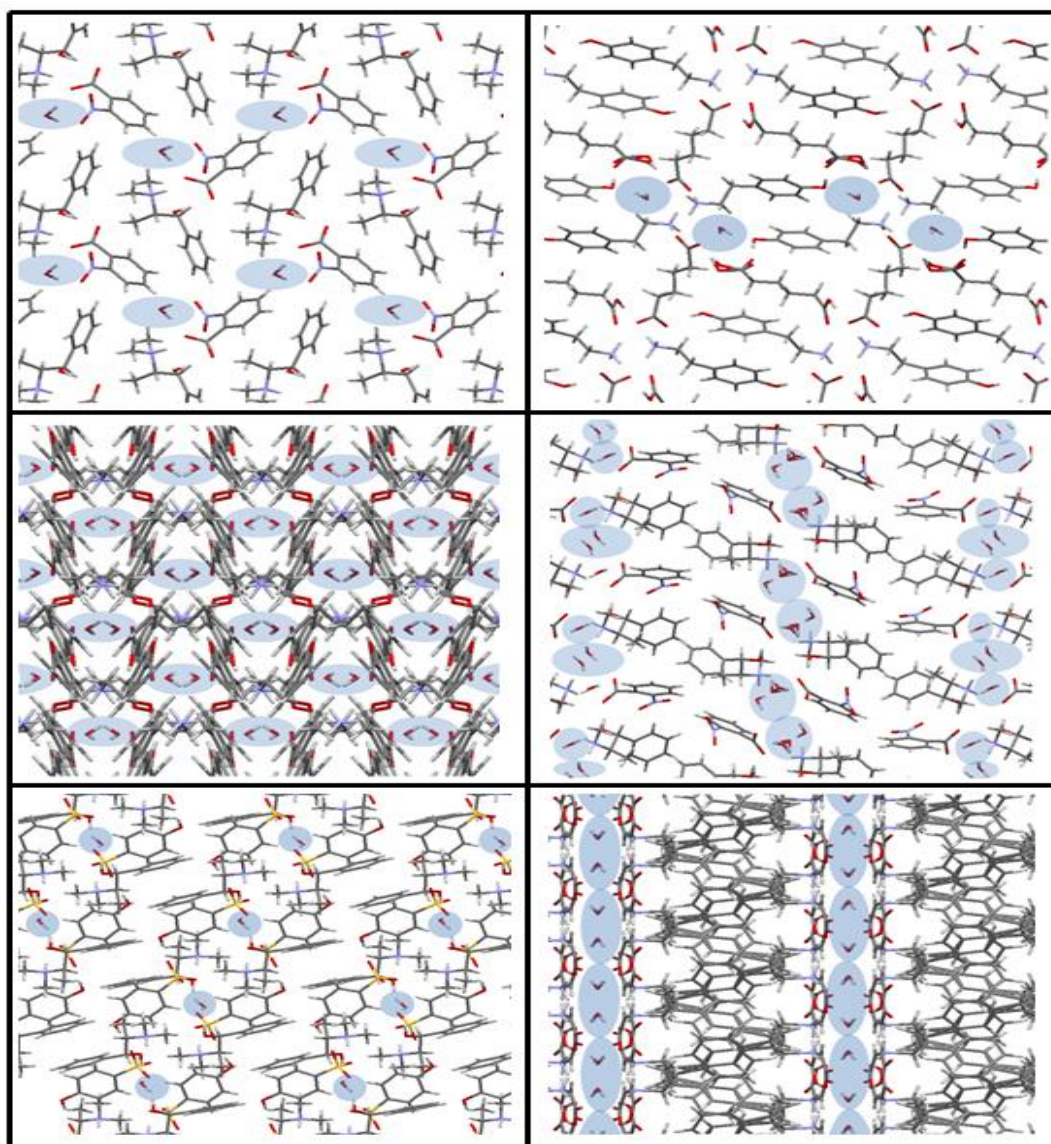
Figure 5.12 Examples of structures that only use discrete graph-set. Top right MEpd2NB, top left TYRAdp, middle right HPEApTol, middle left PEpd2NB, bottom right RMEpd4HBS, bottom left PPAMale.

Most of the hydrate structures contain water molecules that participate in the hydrogen bonding network through chain graph-sets, and the majority of these are anion-water-anion chains, see Figure 5.13. The most common graph-set is $C_2^2(6)$ which is present in 21 of the compounds. The compounds that contain solely discrete graph-set only form a maximum of two dimensional network growth, in contrast many of the structures containing a chain graph-set have network growth in three dimensions, the "extra" dimension most commonly through anion-water-anion chains, see Figure 5.14. The structure of dimethylphenthylaminium benzenesulfonate

monohydrate is a rare example of a structure that does not contain any direct cation-anion contact pairs, instead the structure uses a $C_2^2(6)$ graph-set with anion-water-anion interactions to form a one dimensional ribbon of solvent separated ion-pairs with the cations attached through hydrogen bonds to the water, see Figure 5.15. A similar situation is seen for (+/-)methylephedrinium benzenesulfonate monohydrate. Five other structures form one dimensional hydrogen bond ribbons. In these structures the anion-water-anion chains run parallel to the cation-anion chains and therefore do not add an extra dimensionality to the overall hydrogen bonding network, see Figure 5.16, (-)methylephedrinium benzoate monohydrate.
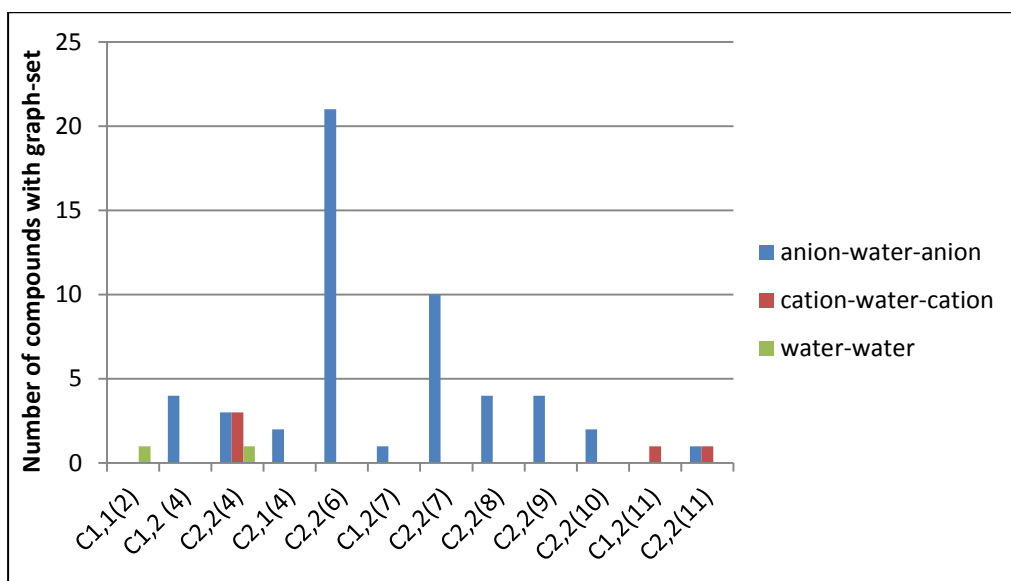
Figure 5.13 Chain graph-sets present within hydrated structures
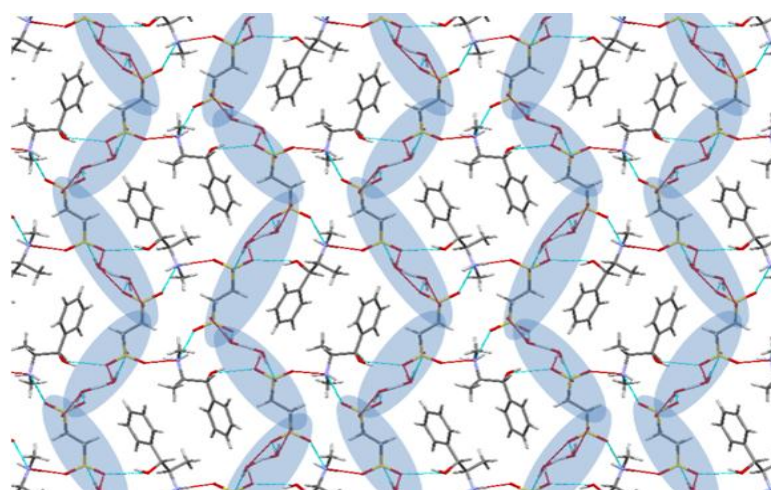
Figure 5.14 Anion-water-anion $C_2^2(6)$ chains present in (-)pseudoephedrinium 1,2-ethanedisulfonate hydronium
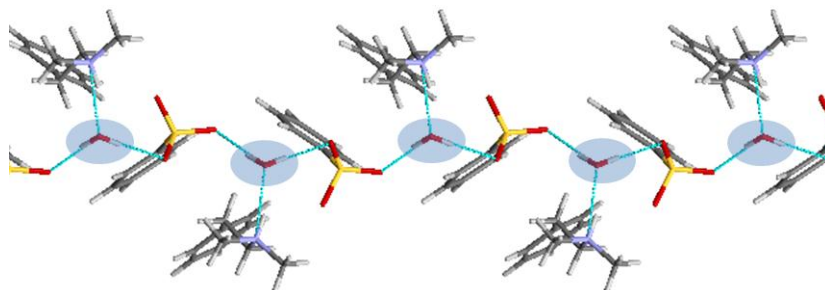
Figure 5.15 Formation of one dimensional ribbon through anion-water-anion $C_2^2(6)$ graph-set, structure of dimethylphenethylaminium benzenesulfonate monohydrate
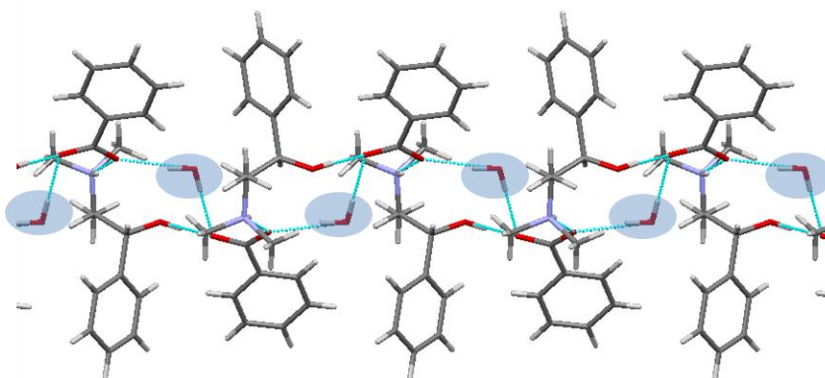
Figure 5.16 Anion-water anion $C_2^2(6)$ graph-set growing parallel to the cation-anion chains, structure of (-)methylephedrinium benzoate monohydrate

5.5 Crystal packing classifications

The crystal packing of the 57 hydrate structures were examined to see if they could be classified into groups of similar packing styles. Of the 57 structures three are co-crystals (DMPEA4NB, TYRAdp, TYRFum) and therefore will not be examined with the rest of the structures. The structures can be split into two main packing styles, described here as layered and paired. The layered structures can be further split into three sub-categories; hydrophilic-hydrophobic single layers, hydrophilic-hydrophobic double layers and alternative cation-anion layers. The paired structures can be split into two sub-categories; anion-cation pairs or anion-anion and cation-cation pairs. Table 5.2 shows which structures fit into which category of packing. Each of these categories will be discussed below with the help of examples.

Table 5.2 Different categories of packing styles

| hydrophilic-hydrophobic single layers | Layered Structures | | | | Paired structures | | Other |
| | hydrophilic-hydrophobic double layers | | alternate cation-anion layers | | anion-cation pairs | anion-anion and cation-cation pairs | |
	Alternate C-A in layers	C-C, A-A in layers	C-A-C layers	C-C-A or C-A-A layers			
Epd4HBS PEA4HBS TYRBS TYR4HBS	DMPEABS HPEAoTol MAMBA4FB MAMBALMD MEpdBz PEpd2CB PEpd2NB PEpd3NB PEpd4AB PEpdoTol RMEpdBS MEpdLMD2	HPEA4CB HPEApTol	DMPEAEDS DMPEALTar EpdLTar2 HPEALTar MAMBAEDS MEpdAdp MEpdLTar MPEAEDS PEARTar PEpd4AB RMEpd2NB RMEpd4HBS RMEpdEDS TYRBF4 TYRClO4 TYRLMal TYRLTar TYRPO4 TYRRMal TYRRTar TYRSO4 TYRSuc	EpdLTar EpdPO4 PEASO4 PEALTar PPARTar PPAMale	EpdMale MEpd2NB TYRAdp2	MEpd3CB MEpdSO4 PEpdSO4 RMEpdSO4	PEpdEDS

5.5.1 Hydrophilic-hydrophobic single layered structures

Hydrophilic-hydrophobic layered structures only occur in salts that possess an aryl group. There are four structures which form hydrophilic-hydrophobic single layered packing, all four of which contain an aryl sulfonate counterion. Here both anions and cations are present in all layers. The hydrophobic region contains the benzene rings and the hydrophilic region contains the functional groups and the water molecules. The layers are organised with two cations followed by two anions. Within these pairs the molecules face in opposite directions e.g. they are antiparallel. An example of this packing mode is the structure of phenylethylammonium 4-hydroxybenzenesulfonate monohydrate, see Figure 5.17, where blue is the hydrophobic layer and red is the hydrophilic layer.
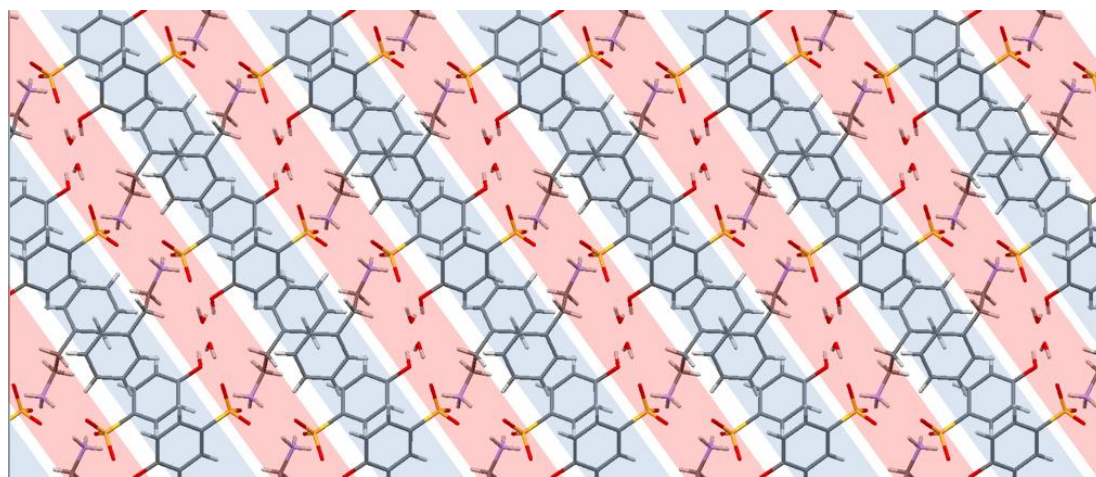
Figure 5.17 Hydrophilic-hydrophobic single layered packing, structure of phenethylammonium 4-hydroxybenzenesulfonate monohydrate. **Blue** is hydrophobic layer, **red** is hydrophilic layer

5.5.2 Hydrophilic-hydrophobic double layered structures

Again both anions and cations are present in all layers in these structures. This category can itself be spilt further into two sub-categories; one where the layers contain alternate cation-anion molecules and the other where the layers contain cation-cation followed by anion-anion molecules as seen in the single layer structures above. Water is of course contained only in the hydrophilic layers. There are thirteen structures which fit into the first sub-category and two structures that fit into the

second. In the majority of the double layer structures the water molecule is located in the hydrophilic layer in close proximity to the anion, as this is the molecule that always makes a hydrogen bond to the water. α(methylaminomethyl)benzyl alcohol 4-fluorobenzoate monohydrate is an example of a double layer hydrophilic-hydrophobic structure as can be seen in Figure 5.18, the hydrophobic region is in blue and the hydrophilic region is in red. Figure 5.19 is an example of the cation-cation, anion-anion double layered structure with yellow highlighting the anion and green highlighting the cation in hydroxyphenethylaminium 4-chlorobenzoate monohydrate.

Figure 5.18 Double layer hydrophilic-hydrophobic structure, α (methylaminomethyl) benzyl alcohol 4-fluorobenzoate monohydrate. **Blue** is hydrophobic layer, **red** is hydrophilic layer
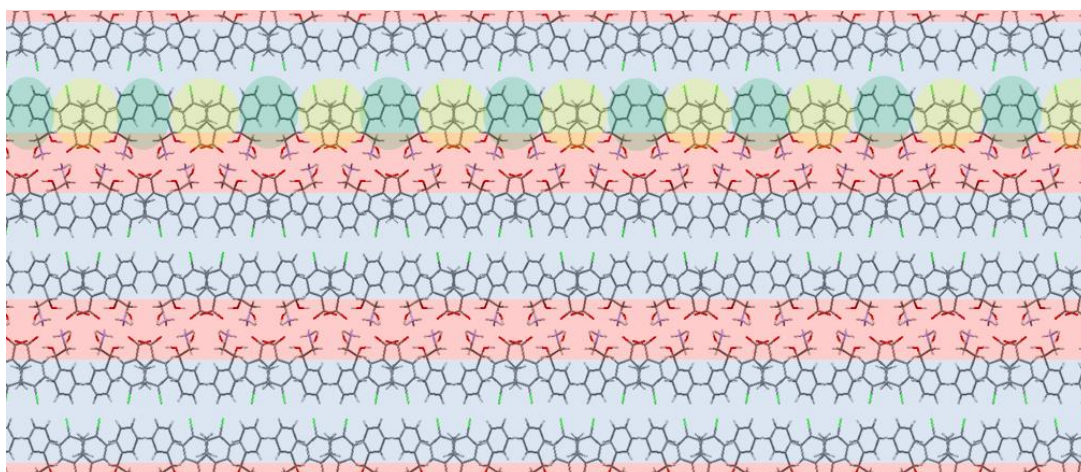
Figure 5.19 Cation-cation, anion-anion double layer hydrophilic-hydrophobic structure, hydroxyphenethylammonium 4-chlorobenzoate monohydrate. **Blue** is hydrophobic layer, **red** is hydrophilic layer, **green** is cations and **yellow** is anions

5.5.3 Alternate cation-anion layered structures

In contrast to the above structural types where both anions and cations are present in all layers, the alternate cation-anion layered structures are the most common occurring in 28 of the 54 structures. This group consists primarily of salts whose anion does not possess an aryl group. Twenty-two form simple cation-anion-cation-anion layers and six form either cation (or anion) bilayers of the form cation-cation-anion or anion-anion-cation layers (this latter is only in the case of phenylpropylammonium maleate monohydrate). The simplest form of cation-anion layering in illustrated with the structure of dimethylphenethylaminium L-tartrate dihydrate, with the cations forming one layer and the anions and water molecule forming another layer, see Figure 5.20. The green highlights the cations and the purple highlights the anions and water molecules.

An alternate packing which is seen with some structures is shown using tyrammonium dihydrogen-phosphate dihydrate as an example, see Figure 5.21. There are still distinct cation and anion/water layers but each layer contains two layers of cations or two layers of anions and water molecules. Again green highlights the cations and purple highlights the anion and water molecules. All of the structures containing hydronion and dihydronium ions adopt the cation-anion layered packing motifs. The structure of methylphenethylaminium ethanedisulfonate dihyronium shows how the anions line up separated by the dihydronium ions to form a layer with the cations lining up end to end to form a second layer, see Figure 5.22, again cations are highlighted in green and anion and dihydronium ions are highlighted in purple. In these structures water is always present in the anion layer. This indicates that anion choice has a big effect on occurrence of hydrate formation as implied in Section 5.2. As water commonly acts as a two-fold hydrogen bond donor, and anions are hydrogen bond acceptors this is in accordance with the water being present in this layer.
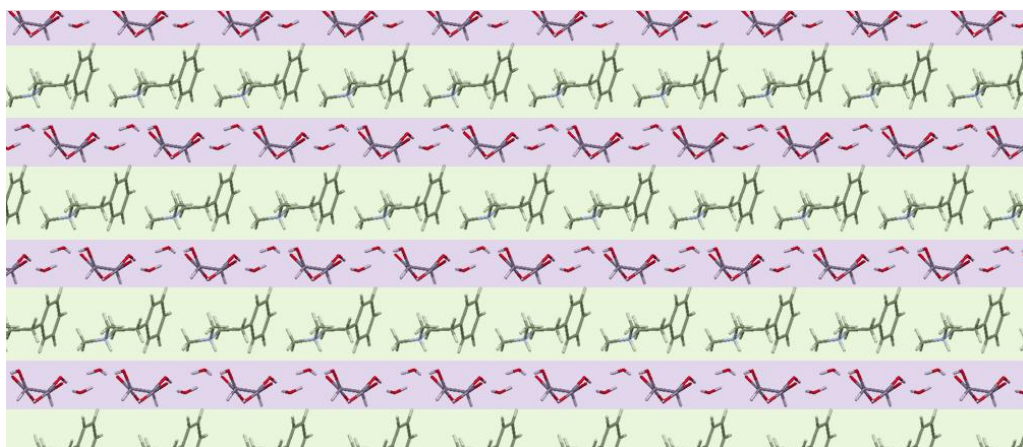
Figure 5.20 Cation anion layered structure, dimethylphenethylaminium L-tartrate dehydrate. **Green** layer is the cations and the **purple** layer is the anions and water molecules
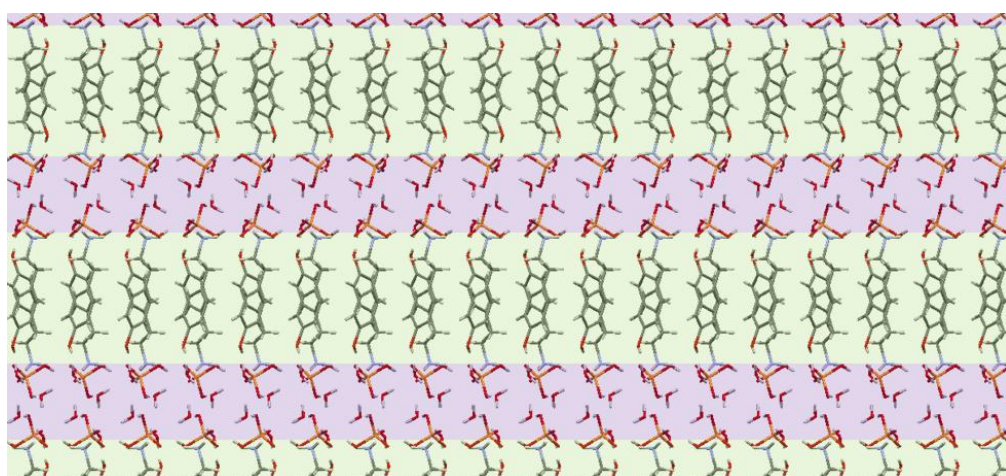
Figure 5.21 Cation anion bilayer structure, tyrammonium dihydrogen-phosphate dihydrate. **Green** layer is the cations and the **purple** layer is the anions and water molecules
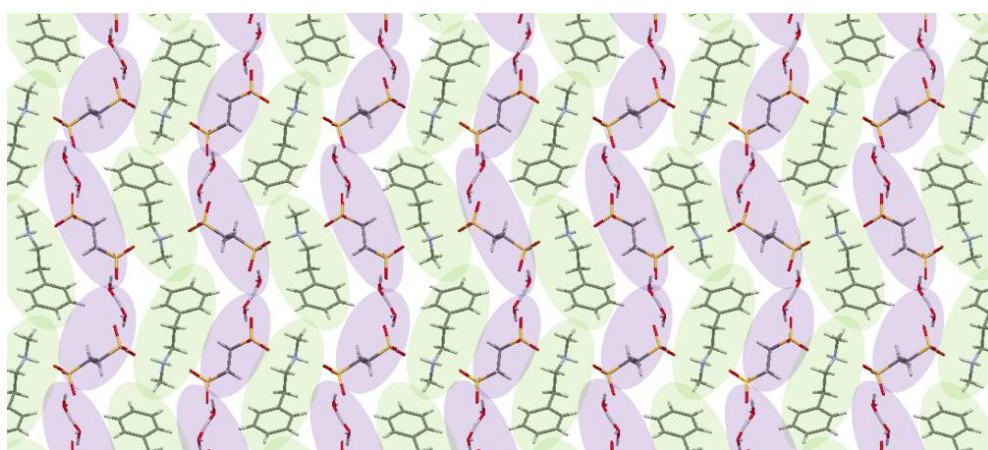
Figure 5.22 Cation anion layered structure, methylphenethylaminium ethanedisulfonate dihyronium. **Green** layer is the cations and the **purple** layer is the anions and dihydronium ions

5.5.4 Anion-cation paired and anion-anion, cation-cation paired structures

There are seven structures that adopt a pairing packing motif, which is packing that does not involve layer formation. Three of the seven pack with cation-anion pairs and four pack with cation-cation and anion-anion pairs next to each other. Systematic structural series of hydrates of simple organic salts are rare, but a recent study on benzoates also cites various layering modes as common packing features but with a distinct sub-group of non-layered channel structures.[16] An example of the cation anion pairs is shown above in Figure 5.12 (MEpd2NB) where the water fits in the gaps between the pairs. (+/-)Methylephedrinium hydrogen-sulfate monohydrate is an example of a compound that forms cation-cation and anion-anion pairs, see Figure 5.23. The water molecules also pair-up and sit in between the two anion molecules. Each purple block runs in a channel out of the page and therefore this can be described as a channel structure.
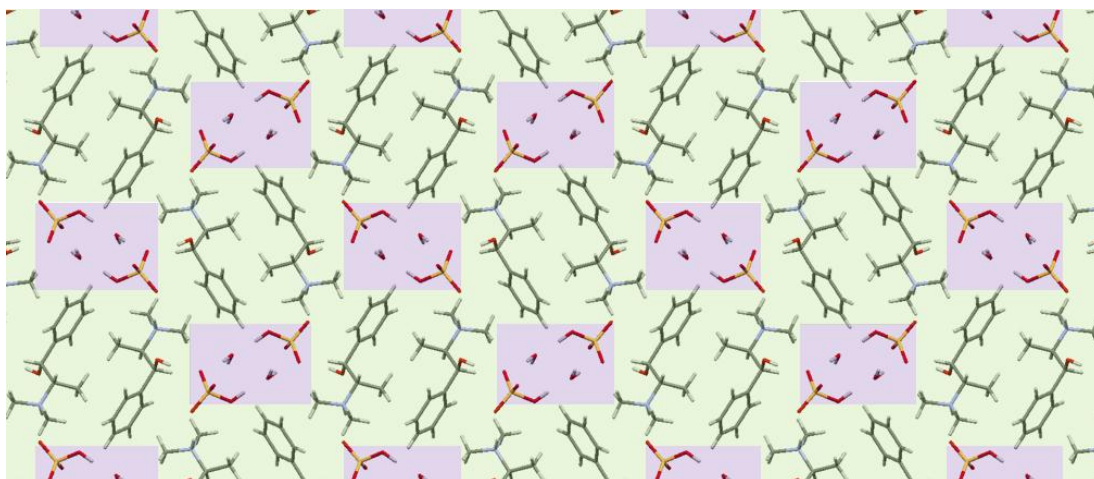
Figure 5.23 Cation-cation and anion-anion paired structure, (+/-)methylephedrinium hydrogen-sulfate monohydrate. **Green** highlights cations and **purple** highlights anions and water molecules

The structure of (-)pseudoephedrinium ethanedisulfonate hydronium is the only structure which does not fit exactly into one of the above categories. It is also odd as it forms a hydronium salt. It could be catergorised as an anion-cation layer structure, see Figure 5.14, the cations occur as pairs inside the pockets made by the anion and hydronium molecules.

5.6 General trends of anhydrous and hydrated salts

The 255 salt structures were examined to compare density, melting point and aqueous solubility measurements between the anhydrous and hydrated salts. For the density comparison the halide salts were excluded as none were seen to form hydrates and this may introduce bias to the results as the heavy halide anions result in compounds that are naturally denser than those with organic counterions. For the comparisons of melting point and solubility, only measurements that have a matching X-ray powder diffraction from the material recovered after the slurry experiment were used. This was to ensure that the reported measurements were phase specific to a known single crystal structure.

5.6.1 Density comparison of hydrated and anhydrous salts

With the exclusion of the halide structures, the density comparison looked at 175 anhydrous salts and 57 hydrated salts. Looking collectively at the two groups there is a marked difference in average density. The average density for the anhydrous salts is 1.310 g/cm^{-3}, 2.75 % lower than the average density for the hydrated salts which have an average value of 1.346 g/cm^{-3}. From this it may be concluded that the main role of water in the hydrated structures is to occupy otherwise vacant spaces within the crystal structures, acting as a space-filler and therefore making the resulting structure denser.[10]

Looking at the densities of the anhydrous and hydrated salts in terms of cation present it can be seen that the same trend is normally followed, with the hydrated salts being denser than the anhydrous salts in all but two cation classes, see Figure 5.24. For the (-)ephedrinium salts the density difference between the anhydrous and hydrated salts is negligible, with the anhydrous salts less than 0.5 % denser than the hydrated salts. For the hydroxyphenethylammonium salts the anhydrous salts are considerably denser than their hydrated counterparts with average densities of 1.400 and 1.322 g/cm^3 respectively. However, this is a counterfeit result as here the higher density is accounted for by the heavier counterions (such as perchlorate, sulfonate,

phosphate and ethanedisulfonate) all forming anhydrous salts with only benzoate and tartrates forming hydrated salts.
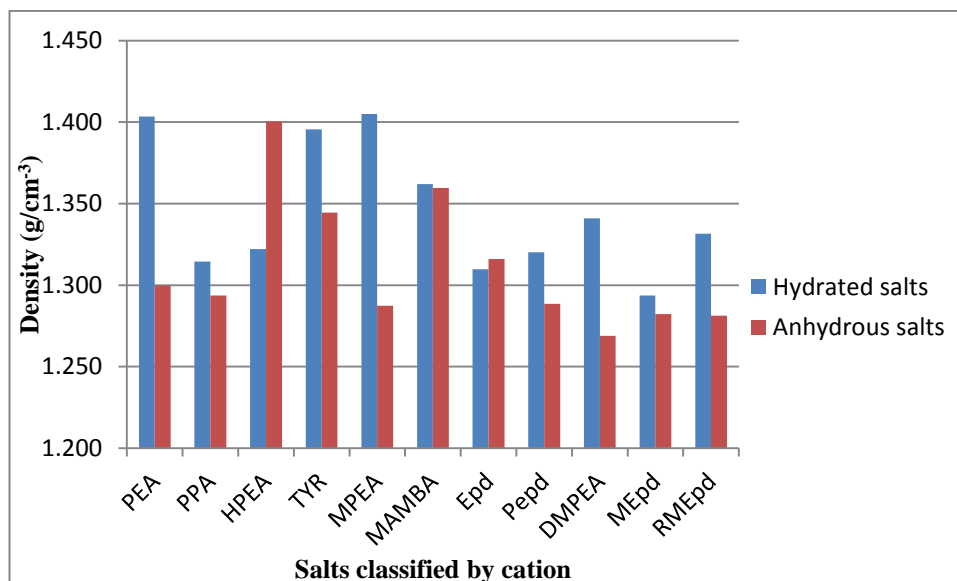
Figure 5.24 Density comparison for anhydrous and hydrated salts classified by cation present with salt

5.6.2 Melting point comparison of hydrated and anhydrous salts

With exclusion of the salts that did not have phase specific data, the melting point comparison looked at 133 anhydrous salts and 38 hydrated salts. The average melting point of the anhydrous salts is 149.5 ^{o}C, 48 % greater than the average melting point for the hydrated salts with a value of 101.0 ^{o}C. Looking at the average melting point in terms of cation present, the average melting point is considerably higher for the anhydrous salts in all cases, see Figure 5.25, with the average difference ranging from 18.1 % for the phenylpropylammonium salts to 53.1 % for the α-(methylaminomethyl) benzyl alcohol salts.
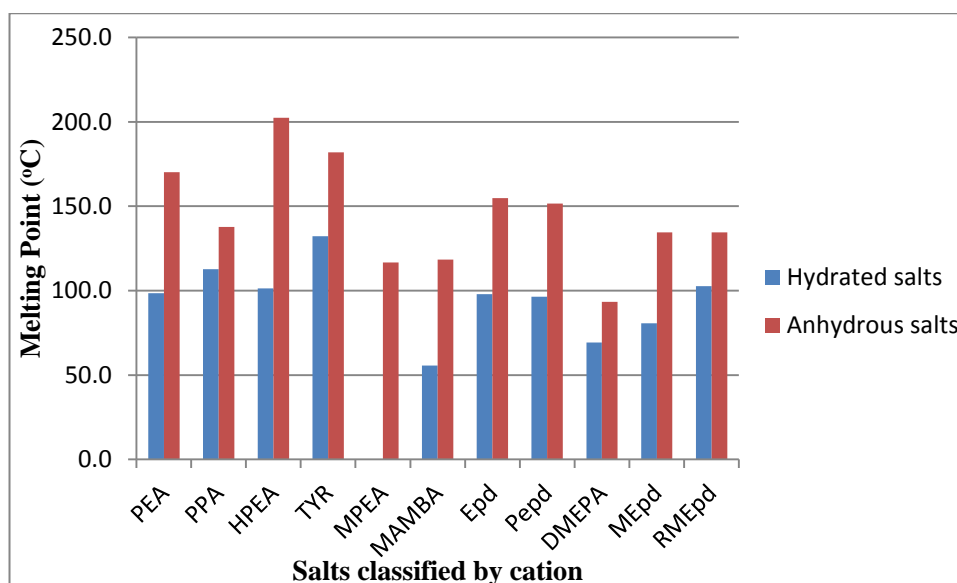
Figure 5.25 Melting point comparison for anhydrous and hydrated salts classified by cation present with salt

5.6.3 Solubility comparison of hydrated and anhydrous salts

With exclusion of the salts that did not have phase specific data, the aqueous solubility comparison looked at 138 anhydrous salts and 38 hydrated salts. The average solubility of the anhydrous salts is 1.463 mol/L and the average solubility of the hydrated salts is 1.284 mol/L. If there were the same types of anions present in the anhydrous and hydrated salt groups then this difference in solubility could be accounted to the lesser free energy observed upon dissolution for the hydrated compounds thus giving then a lower aqueous solubility.[7] However, looking at the salts when they are classified by cation there is no general trend to be seen, in fact in several cases the opposite trend is seen, see Figure 5.26. It is concluded that a solubility comparison between the hydrated and anhydrous salts available here is unwise and that the reading of great meaning into the results is tenuous as the results can be skewed very easily. This latter is true as the salt solubility is very much influenced by the counterion involved and we thus have no true like-for-like comparison available. For example there are several benzoate salts within the study which are unlikely to form hydrates (see Section 5.2) and from solubility analysis it has been shown that benzoates tend to have low solubility values (see Section 6.2) therefore the presence of anhydrous benzoates in salt classifications could result in

abnormally low average solubility measurements. On the contrary, tartrate salts tend to have a high occurrence of hydrate formation and are also associated with higher solubility measurement; therefore they could lead to a higher average solubility for the hydrated salts.
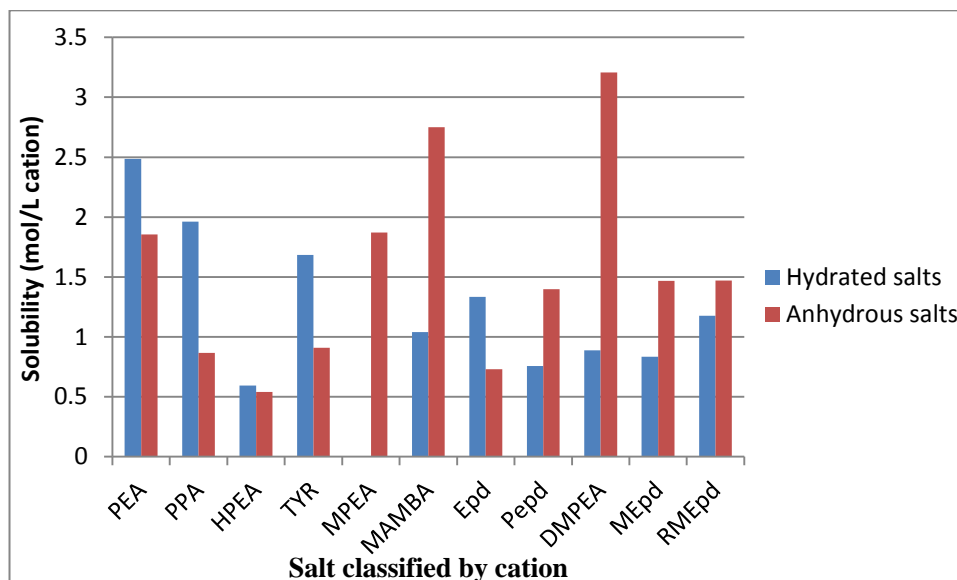
Figure 5.26 Solubility comparison for anhydrous and hydrated salts classified by cation present with salt

5.7 Comparison of anhydrous and hydrated salt forms

Both an anhydrous and a hydrated phase were characterised by single crystal diffraction for four of the otherwise chemically identical structures in the dataset. The four hydrated salts are all monohydrates and three of the pairs relate to the racemic methylephedrine base with the forth being a salt of enantiopure methylephedrine. Table 5.3 shows the molecular volume and density information for the structures as well as which one was the stable form after the slurry experiment. Only this stable form has a measured solubility and so solubility cannot be compared here. Similarly, melting point is known only for one form of each pair.

Table 5.3 Structures where both an anhydrous and hydrated phase where characterised by single crystal X-ray diffraction

Salt	Anhydrous salt		Hydrated salt		Thermodynamic species
	Mol Vol	Density	Mol vol	Density	
RMEpdBz	404.348	1.238	447.199	1.186	hydrate
RMEpdBS	421.726	1.329	450.305	1.311	anhydrous
RMEpdEDS	349.564	1.303	359.76	1.349	hydrate
MEpdLMD	431.762	1.275	456.328	1.271	hydrate*

*A slurry of pure MEpdLMD gave the hydrate, indicating that the hydrate is most stable under these conditions. However, a sample of MEpdLMD.H_2O obtained from racemic mandelic acid (rather than l-mandelic acid) and thus contaminated with some d-mandelic acid, gave a slurry where the final solid isolated was the pure anhydrate MEpdLMD.

In the case of RMEpdBS and RMEpdEDS the thermodynamic form appears to be the species which is the densest, the anhydrous for the benzenesulfonate salt and the hydrate for the ethanedisulfonate salt. For the case of RMEpdBz the hydrated salt, which is considerably less dense, is the thermodynamic species after the slurry experiment. This is generally regarded as unusual, though it should be noted that our previous analysis of Wallach's rule and melting point (Section 4.12 and 4.13) suggested that there was little correlation between density and stability. Moreover, here the hydrated species is a conglomerate whereas the anhydrate crystals are racemic. Therefore the driving force that causes the formation of a conglomerate from this slurry cannot be separated from whatever causes formation of the hydrate in this experiment. The densities of the two mandelate species are comparable, and indeed both species are seen as a thermodynamic product. The anhydrous salt was present after the slurry experiment of (-)methylephedrine with (+/-)mandelic acid and the hydrated salt was present after the slurry experiment of (-)methylephedrine with (-)mandelic acid.

The large difference in density between the two ethanedisulfonate salts can be illustrated by examining the way the cations pack together. In the anhydrous species the cations of neighbouring hydrogen bonded stacks interact via edge to face π-stacking, in contrast in the hydrated species the hydrogen bonded stacks are more closely packed via face to face π-stacking, see Figure 5.27. Our study shows that on average hydrates are denser than anhydrous salts, see Section 5.6.1, however here

this statement is untrue for three out of the four comparisons. Indeed here other contributing factors appear to be influencing the density whether it be formation of a conglomerate or that the anhydrous form is the thermodynamic product.
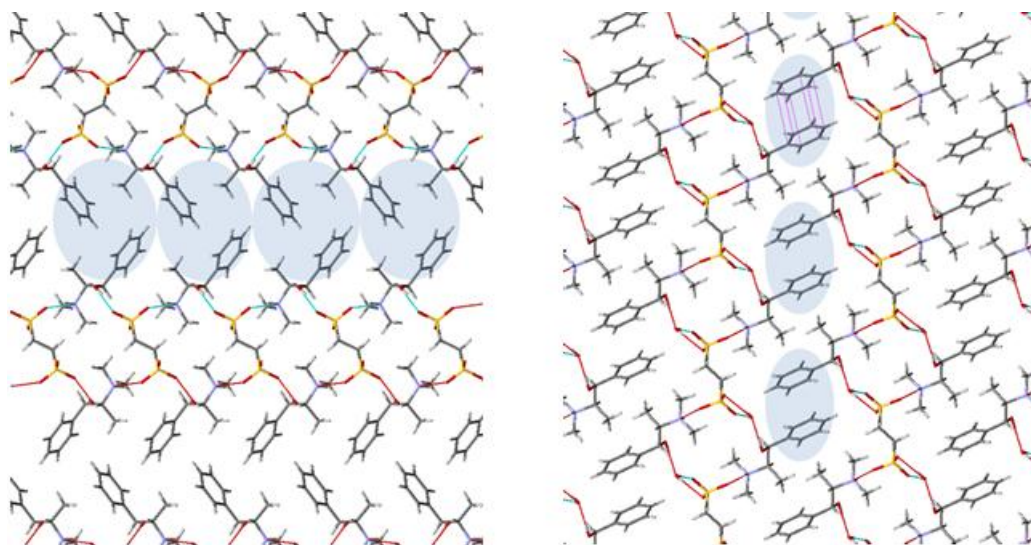
Figure 5.27 Crystal packing of the anhydrous (left) and hydrated (right) salt of (+/-)methylephedrinium 1,2-ethanedisulfonate. **Blue** highlights the edge to face π-stacking and face to face π-stacking in the anhydrous and hydrated structure respectively

5.8   Conclusions

Analysis of the database of salts found that 57 out of the 255 structures formed a hydrated species a ratio which is comparable to literature studies analysing NH^+ salts present within the CSD.[14] Within the dataset there is a decrease in hydrate formation with regards to the type of amine present of the order primary > secondary > tertiary. Upon analysis of the hydrate with regards to the cation present a high occurrence of hydrate formation is seen for (-)pseudoephedrinium salts and a low occurrence of hydrate formation is seen for salts containing phenethylammonium and methylphenethylaminium counterions, both of which do not contain any additional polar groups within their structure. Examining the dataset with regards to the anion present it is seen that there is a low occurrence of hydrate formation for the benzoate and mandelate salts and all halide salts are anhydrous. Hydrate formation is greatly increased within the sulfonate, 'spherical' and dicarboxylate salt categories again

supporting the theory that hydrate formation is linked to an increased presence of polar groups within the salt.[5]

The most common water environment present within the hydrated structures involved the water molecule providing two donor and one acceptor hydrogen bond interactions, a result that agrees with a previous literature study of the common water hydrogen bond patterns within the CSD.[11] Hydrogen bond donor/acceptor ratios were studied for the 255 salts present in the dataset, with the most common donor/acceptor ratio for the anhydrous and hydrated salts being one. However, the results show that hydrate formation is most probable when there is an excess of hydrogen bond acceptor atoms present and least likely when there is an excess of hydrogen bond donor atoms present. This indicates that the primary role of water is to provide extra hydrogen bond donors, a theory that is in agreement with the water molecule's most common environment acting as a two hydrogen bond donor and one hydrogen bond acceptor molecule.

Water acts as a hydrogen bond donor to the anion of the salt in all the hydrate structures. This commonly leads to the formation of an anion-water-anion chain graph-set, the most common of which is a $C_2^2(6)$ graph-set which occurs in 21 of the hydrate structures. The presence of such solvent-anion chains often leads to an extra dimensionality of crystal growth within the array. On the contrary to this the presence of a hydrogen bond interaction between a water donor and a cation hydrogen bond acceptor is much rarer, occurring in only 11 of the 48 structures where this interaction is possible. In circumstances where the water acts as a hydrogen bond acceptor, both the anion and cation are equally as likely to be the hydrogen bond donor.

With the exclusion of the co-crystal hydrate salts the 54 hydrated species used hydrogen bonding interactions to produce two main packing styles, namely layered and paired. The layered structures can be further split into three sub-categories; hydrophilic-hydrophobic single layers, hydrophilic-hydrophobic double layers and alternative cation-anion layers. The paired structures can be split into two sub-

categories; anion-cation pairs or anion-anion and cation-cation pairs. In all cases the water molecule is associated with the anion either being present in the hydrophilic or anion layer or encompassed next to the anion in the paired structures. The most common packing style observed is that of cation-anion layers which is present in 28 of the 54 structures.

5.9 References

(1)   *Towards Better Safety of Drugs and Pharmaceutical Products*; Stahl, H. P., Ed.; Elsevier/North-Holland Biomedical Press: Amsterdam, 1980.
(2)   Khankari, R. K.; Grant, D. J. W. *Thermochim. Acta* **1995**, *248*, 61.
(3)   Kiang, Y. H.; Xu, W.; Stephens, P. W.; Ball, R. G.; Yasuda, N. *Crystal Growth & Design* **2009**, *9*, 1833.
(4)   Shankland, N.; David, W. I. F.; Shankland, K.; Kennedy, A. R.; Frampton, C. S.; Florence, A. J. *Chem. Commun.* **2001**, 2204.
(5)   Infantes, L.; Chisholm, J.; Motherwell, S. *Crystengcomm* **2003**, *5*, 480.
(6)   Shefter, E.; Higuchi, T. *J. Pharm. Sci.* **1963**, *52*, 781.
(7)   Grant, D. J. W.; Higuchi, T. *Solubility Behaviour of Organic Compounds*; Wiley: New York, 1990.
(8)   Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. *Advanced Drug Delivery Reviews* **2001**, *48*, 3.
(9)   *Structural Aspect of Hydrates and Solvates*; 2nd ed.; Morris, K., Ed.; Informa Healthcare: New York, 2009.
(10)   Van der Sluis, P.; Kroon, J. *J. Cryst. Growth* **1989**, *97*, 645.
(11)   Gillon, A. L.; Feeder, N.; Davey, R. J.; Storey, R. *Crystal Growth & Design* **2003**, *3*, 663.
(12)   Gorbitz, C. H.; Hersleth, H. P. *Acta Crystallographica Section B-Structural Science* **2000**, *56*, 526.
(13)   Infantes, L.; Motherwell, S. *Crystengcomm* **2002**, 454.
(14)   Haynes, D. A.; Jones, W.; Motherwell, W. D. S. *Crystengcomm* **2005**, *7*, 342.
(15)   Clarke, H. D.; Arora, K. K.; Bass, H.; Kavuru, P.; Ong, T. T.; Pujari, T.; Wojtas, L.; Zaworotko, M. J. *Crystal Growth & Design* **2010**, *10*, 2152.
(16)   Arlin, J. B.; Florence, A. J.; Johnston, A.; Kennedy, A. R.; Miller, G. J.; Patterson, K. *Crystal Growth & Design* **2011**, *11*, 1318.
(17)   Kennedy, A. R.; Andrikopoulos, P. C.; Arlin, J. B.; Armstrong, D. R.; Duxbury, N.; Graham, D. V.; Kirkhouse, J. B. A. *Chemistry-a European Journal* **2009**, *15*, 9494.
(18)   Kennedy, A. R.; Kirkhouse, J. B. A.; McCarney, K. M.; Puissegur, O.; Smith, W. E.; Staunton, E.; Teat, S. J.; Cherryman, J. C.; James, R. *Chemistry-a European Journal* **2004**, *10*, 4606.
(19)   Infantes, L.; Fabian, L.; Motherwell, W. D. S. *Crystengcomm* **2007**, *9*, 65.
(20)   Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.
(21)   Desiraju, G. R. *Journal of the Chemical Society-Chemical Communications* **1991**, 426.

(22)   Wallace, S.; Huang, L.; Massa, L.; Mukhopadhyay, U.; Bernal, I.; Karle, J. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16798.

(23)   Zhu, Z. B.; Gao, S.; Ng, S. W. *Acta Crystallographica Section E-Structure Reports Online* **2009**, *65*, O2687.

(24)   Yana, J.; Lee, V. S.; Nimmanpipug, P.; Dokmaisrijan, S.; Aukkaravittayapun, S.; Vilaithong, T. *Journal of Solid Mechanics and Materials Engineering* **2007**, *1*, 556.

(25)   Miller, G. J. Ph.D. Thesis - A Structural Database for Pharmaceutical Salt Selection, University of Strathclyde, 2010.

Chapter 6

Enantiopure and racemic salts; Comparison of solubilities and structure-property correlation

6.1   Introduction

This chapter investigates the physical properties of salts of a range of phenylethylamine derivatives. The properties investigated include those such as density and melting point already examined in detail for methylephedrine (see Section 4.12 and Section 4.13) but also the pharmaceutically vital property of aqueous solubility. Attempts are made to find correlations between measured values and other physical or calculated properties – including properties that can only be described with prior knowledge of solid state structure.

Work examining potential relationships between physicochemical properties has been performed for a wide variety of salts with varying and often apparently contradictory results. Neilsen et al.[1] attempted to correlate solubility to different crystal packing modes as determined from X-ray structure analysis. The results showed no correlation between the two; this is in contrast to the findings of Arlin et al.[2] who show that analysis of the crystal array can help predict the rank order of aqueous solubility for a series of alkali earth metal salts of benzoate derivatives. Another correlation found within this latter research is between the solubility of the free acid and the solubility of the salt, a rare example of a correlation that generally agrees with other works.[3,4] Indeed Parshad et al.,[3] in a study of *p*-benzoate derived salts of benzylamine, found a linear correlation between these two properties. Finally, the study by Arlin et al. shows little or no correlation between aqueous solubility and the hydrophilicity of the organic ion, a negative statement that is also found to be true by some other studies,[5,6] and false by others. Interestingly those studies which do suggest a connection between hydrophilicity and aqueous solubility split between examples where high hydrophilicity enhances solubility[7] and those where it reduces solubility.[3] Thus the correlation between hydrophilicity and aqueous solubility is variously positive, negative or non-existant, a confused and non-simple state of affairs that rather neatly sums up current understanding in the area.

A correlation that has been analysed repeatedly and is reasonably well established is that between melting point and solubility. Concentrating on salt species rather than neutral molecular solids, a study by Thomas et al.[8] found the expected inverse linear relationship between the log solubility and melting point and a study by Gould similarly found an linear relationship between solubility and the inverse melting point as well as a positive correlation between the melting point of the free acid and the melting point of the salt.[9] Other studies show correlation with increase in melting point correlating with a decrease in log solubility, however there is no linear or quantitative trend in any of these other studies.[5,6,10,11] It should be noted that the two studies where a linear relationship is claimed both have relatively small numbers of salts in their datasets (8 and 4 respectively). The study by Black et al.[10] is most relevant to this work as it is composed of a dataset of 17 (-)ephedrinium salts. Their non-linear trend between melting point and log solubility is echoed by a similar non-linear trend observed between log solubility and enthalpy of fusion. This study also finds no correlation between the calculated lattice energy and log solubility and the authors go as far as to claim that their data shows that solid state structure does not appreciably affect solubility and that solubility is thus determined by solution phase factors.[10]

In order to examine our large dataset it was first split down into sub groups depending on the cation present within the salt. The solubility measurements were performed in duplicate as described in Section 3.7. Only where PXRD indicated that the measured solubility was for a phase where SXD data was available, is a solubility value presented. The melting point measurements were performed in triplicate as described in Section 3.6.

6.2   General trends in solubility for methylephedrinium salts

Firstly, all the solubility values for both the enantiopure and racemic salts of methylephedrine were examined together and the solubility value described as sparingly soluble (<0.75 mol/L), soluble (0.75-1.75 mol/L) or very soluble (>1.75

mol/L), this information can be found in Section 3.7. The terms sparingly soluble, soluble and very soluble are arbitrary and thus related only to this work and not to outside studies. The varying solubility of the different salts is shown in Figure 6.1. The different subgroups of salts can be seen with the benzoate derived salts highlighted in red, the halides and spherical anions highlighted in green, the sulfonates highlighted in blue and finally the carboxylate salts highlighted in purple.
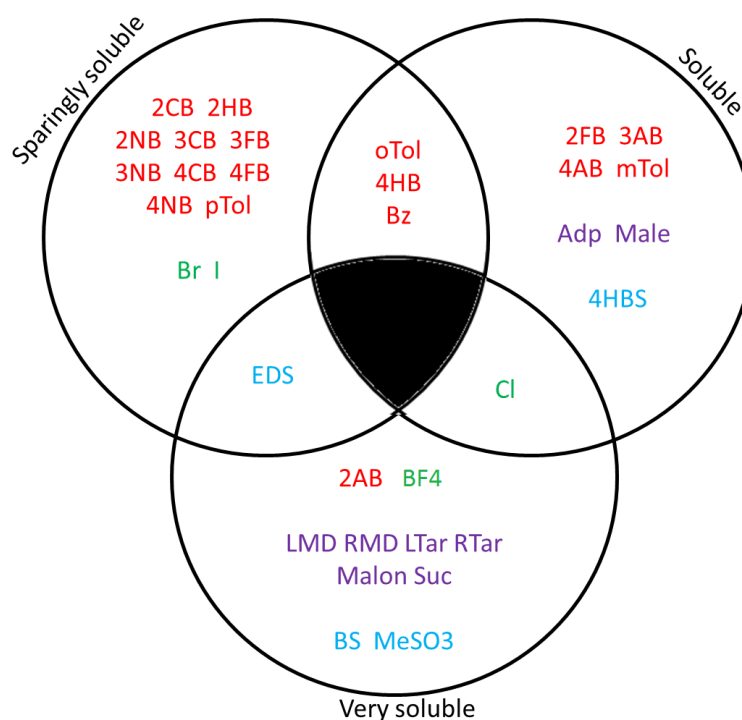
Figure 6.1 Solubility of different salt forms of methylepehdrine. Colours represent different anion groups; **red** = benzoate derivatives, **blue** = sulfonate, **purple** = dicarboxylate and mandelate and **green** = halides and tetrafluoroborate salts. Overlapping regions contain different solubility classification for the enantiopure-racemic salt pair

The majority of the enantiopure and racemic benzoate salts fit into the sparingly soluble category having a solubility value of less than 0.75 mol/L. All four of the aminobenzoate salts have much higher solubility ranging from 1.499 in MEpd4AB to 2.354 mol/L in MEpd2AB, this difference in behaviour for aminobenzoates is also seen elsewhere.[2] The amino acids behave differently as they are amphoteric (have the ability to react as an acid or a base).[12] Their solubility values are very pH dependant

because the predominant form present in solution changes with H^+ concentration. The different possible phases are shown below in equation 6.1.[13]

$$H_2A \longleftrightarrow HA^{\pm} + H^+ \longleftrightarrow A^- + 2H^+ \qquad \text{(Eqn 6.1)}$$
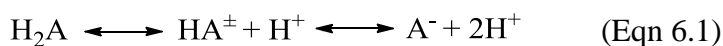
Generally solubility greatly increases on the formation of a salt; in this case the reaction with a base shifts the equilibrium to the right. However, the presence of the second basic group on the deprotonated aminobenzoate ensures that it will interact differently with the protoic aqueous environment compared to other deprotonated acids and therefore it is not strictly appropriate to directly compare the observed solubilities of these aminobenzote salts with the other benzoate salts.

The benzoate and 4-hydroxybenzoate salts are placed in the overlapping area of the sparingly soluble and soluble categories in Figure 6.1 as both have solubility values for their racemic forms that are much greater than their enantiopure forms' solubilities. The racemic forms of both give conglomerates in the solid phase and therefore the measured solubility value is not that for the true racemic phase. The other salts that form a conglomerate in their racemic phase, RMEpd4CB and RMEpdpTol, do not follow the same pattern and instead the racemic/conglomerate measured solubility is lower than that achieved for the enantiopure compound, all falling into the sparing soluble category. The Meyerhoffer double solubility rule states a conglomerate has the solubility equal to the sum of the solubility of the corresponding enantiomers.[14] Table 6.1, shows the solubility values collated for the observed enantiopure compounds and their corresponding conglomerates. Only two of the five sets of compounds agree with the Meyerhoffer double solubility rule.

Table 6.1 Enantiopure and conglomerate solubility to test Meyerhoffer double solubility rule

Compound	Enantiopure solubility (mol/L base)	Conglomerate solubility (mol/L base)	Agrees with Meyerhoffer rule
MEpd	0.048	0.102	yes
MEpdBz	0.327	1.629	no
MEpd4CB	0.052	0.035	no
MEpd4HB	0.440	0.885	yes
MEpdpTol	0.128	0.109	no

Other benzoates with relatively high aqueous solubility are the enantiopure salts of MEpdoTol and MEpd2FB with solubility values of 1.663 and 1.235 mol/L respectively. Interestingly, both of these salts are in the isostructural group 5 at the bottom of the similarity packing tree diagram, see Section 4.9, Figure 4.15. This group contains two *ortho-* substituted and two *para-* substituted benzoate salts, of which the two *para-* substituted salts have 'typical' benzoate solubility values of 0.052 and 0.128 mol/L. Is it possible that this packing arrangement has a much greater energy penalty for the *ortho-* anions providing a much less stable salt and resulting in a much greater solubility? The other benzoate derived salt of higher solubility is RMEpdmTol with a measured value of 1.387 mol/L, at present no argument presents itself as to why this has a solubility measurement of such magnitude.

For the halide salts, all bromide and iodide salts gave a very low solubility measurement. The chloride salts have a much higher solubility especially in the enantiopure salt where it fits into the very soluble category with a measured solubility of 2.443 mol/L. Whilst this is possibly to be "expected" as chloride anions are traditionally the first salt forms attempted in salt preparation partly due to their ability to greatly increase the solubility of basic drugs,[15] it is also "unexpected" as the chloride and bromide enantiopure salts are isostructural and theory indicates that the bromide should make weaker interactions and so be more soluble. A similar argument on this perhaps overly simple level explains the high solubility of the BF_4 salts. This could be attributed to the weak interactions made between cations and anions here − as illustrated by the PIXEL calculations shown in Section 4.7. High chloride salt solubility may thus be due not to solid-sate effects but to a higher hydration energy of chloride, as compared to heavier halides, in the solution phase.[16]

The carboxylate salts all have relatively high solubilities within a wide range from 0.939 mol/L for MEpdMale to 4.132 mol/L for RMEpdRMD. The enantiopure and racemic pairs of the carboxylate salts all have comparable values.

The last group, the sulfonates, make up an interesting set with RMEpdEDS being an outlier. All of the other five sulfonate salts have comparable and relatively high solubilities ranging from 1.173 mol/L for MEpd4HBS to 3.982 mol/L for MEpdMeSO3. In the case of the two ethanedisulfonate salts the enantiopure salt has a solubility of 3.552 mol/L whereas the racemic salt has a much lower solubility of 0.527 mol/L a measurement that coincides with many of the benzoate derived salts. Note that this is the largest solubility difference observed here however, these two salts are not chemically identical, the enantiopure salt is anhydrous whereas the racemic salt is a monohydrate. It is well known that hydrated forms of crystalline materials are less water soluble than anhydrous forms,[17] so this would appear to be the crucial difference here rather than any other structural difference between the two forms. Nonetheless this comparison is normally made between two species that are identical except for the presence of water, in the data examined here the two species are altogether different as one is an enantiopure salt and the other is the corresponding racemic salt. The other two salt pairs that contain an anhydrous and hydrated compound are the 3-chlorobenzoate and 4-hydroxybenzenesulfonate salts. These two salt pairs disagree with the above statement as both have hydrated salts with a higher solubility than the anhydrous pair, though the differences are not nearly as pronounced.

6.3 Comparison of enantiopure and racemic salt solubility measurements for methylephedrinium salts
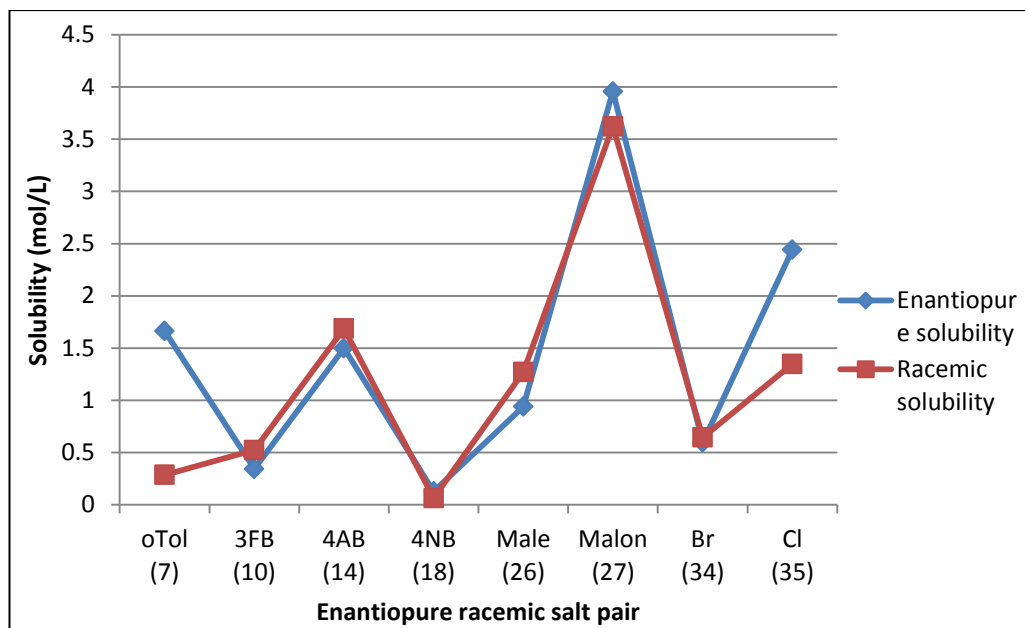
Figure 6.2 Solubility comparison for eight chemically identical enantiopure-racemic salt pairs of methylephedrine

There are only eight pairs of enantiopure and racemic salts that can be directly compared with respect to their solubility, see Figure 6.2. This is reduced from the 16 pairs that were compared with respect to density as phase changes occurred in one of the pair in all of the excluded samples except for the pair of 1,2-ethanedisulfonate salts which was excluded as the enantiopure species is anhydrous and the racemic species is a hydrate. These eight salt pairs all have identical chemical make-up and are composed of four benzoate derived salts, two carboxylate salts and two halide salts. All solubility data along with ESDs is shown in Table 6.2. The average enantiopure solubility is 1.466 mol/L, 22 per cent greater in comparison to the average racemic solubility which has a value of 1.181 mol/L.

Table 6.2 Solubility data for chemically identical pairs of enantiopure and racemic methylephedrinium salts

Compound	Sol (mol/L base)	ESD	Compound	Sol (mol/L base)	ESD
MEpdoTol	1.663	0.102	**RMEpdoTol**	0.286	0.002
MEpd3FB	0.344	0.006	**RMEpd3FB**	0.524	0.01
MEpd4AB	1.499	0.135	**RMEpd4AB**	1.687	0.098
MEpd4NB	0.128	0.002	**RMEpd4NB**	0.066	0.005
MEpdMale	0.939	0.029	**RMEpdMale**	1.27	0.031
MEpdMalon	3.957	0.194	**RMEpdMalon**	3.623	0.198
MEpdBr	0.596	0.004	**RMEpdBr**	0.646	0.047
MEpdCl	2.443	0.024	**RMEpdCl**	1.349	0.022

As discussed previously in Section 1.7 it is predicted that the more thermodynamically favoured racemic salts should have a lower solubility than their enantiopure counterparts as they should have greater stability caused by more efficient crystal packing. As previously seen with respect to density and melting point this statement does hold true for the group as a whole however, again as was seen in Section 4.12 and Section 4.13 many individual salt pairs do not adhere to this prediction. In the case of the solubility measurements this "rule" in fact holds true for only half of the salt pairs (4 from 8). The two salt pairs that have noticeably the largest difference in solubility between the enantiopure and racemic product are those of the *o*-toluate and chloride salts, both of which have an enantiopure compound with a much greater aqueous solubility. These large differences in solubility are what produce the observed difference in average values.

All of the eight enantiopure/racemic pairs for which solubility data is available have also been examined with respect to density. Three of the eight have been shown to fail Wallach's rule.[18] (Though all three are from the "equal density" category, unfortunately all three cases where the enantiopure form had a significantly greater density than the racemic form failed to give phase specific solubility data). In the case of 3-fluorobenzoate, the salt pair does not conform to Wallach's rules as the racemic phase has a lower density, this coincides with the racemic compound also

having a lower melting point and therefore the higher solubility observed for the racemic compound is arguably as would be expected. However, the other two salt pairs of the eight that do not conform to Wallach's rules are the chloride and *ortho*-toluate salt pairs, e.g. the two pairs that show massively increased solubility for the enantiopure forms. Thus failure to obey Wallach's rule is not a predictor of which enantiomorph will be the more soluble.

6.4 Correlation of physical properties for (1R,2S)(-) methylephedrinium benzoate derived salts

In an initial attempt to find correlations between measured physical properties, many data plots were constructed and the most interesting results are presented in this section. There are 14 (-)methylephedrinium benzoate derived salts where solubility measurements were obtained and where the solid-state phase of the slurry was confirmed by X-ray powder diffraction, see Table B.10, Appendix B. Of these 14 salts, three are monohydrates and the remainder are anhydrous. Prior work based on fundamental considerations of dissolution as an equilibrium between solid and solution phases shows that hydrated and anhydrous forms of materials should have different behaviour and thus these should be considered separately.[17,19]

It is well known that melting point often correlates with log solubility and so this was the first relationship to be investigated.[8,9] For the (-) methylephedrinium benzoate derived salts, a plot of log solubility versus melting point showed a linear association between the values for the anhydrous salts with a correlation factor 'R' of 0.799, see Figure 6.3. As expected, the salts with the lowest melting points have the highest solubilities. The values for the hydrated salts do not correlate to those of the anhydrous salts, or indeed to each other. Ignoring the hydrates, Figure 6.4 shows the variable line fit plot, which has an equation of log y = -0.0192x + 2.3617, with a 0.0048 error on the gradient and a 0.6824 error on the intercept. The residual plot (Figure 6.5) shows that the data is homoscedastic (there is constant variance

throughout all the data points) which is in agreement with the data being linearly correlated. This correlation, although containing a large associated error, is in agreement with other literature studies' correlations which contain only a handful of data points.[8,9]
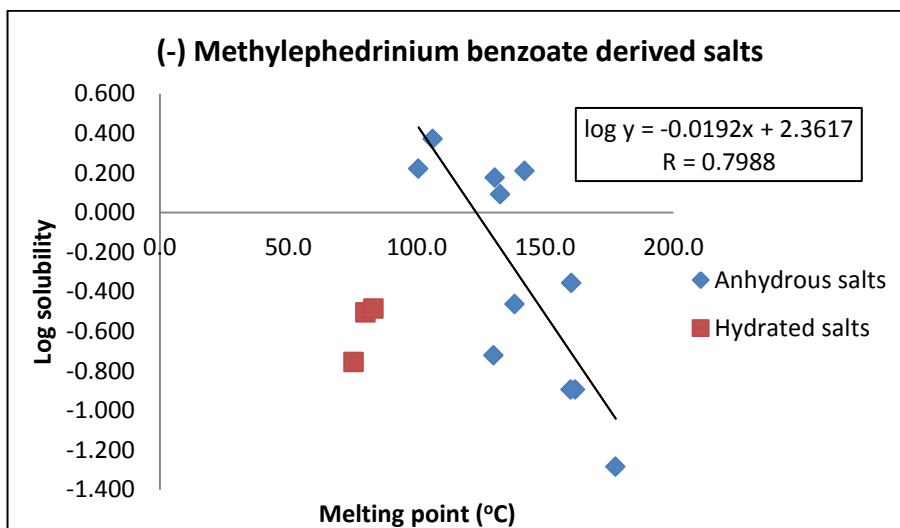

Figure 6.3 Plot of log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium benzoate derived salts
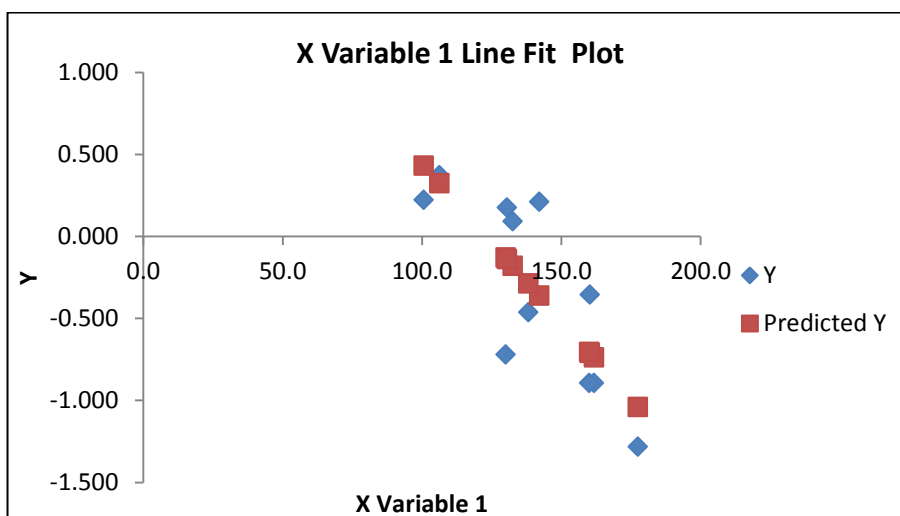

Figure 6.4 Variable line fit plot for log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium benzoate derived salts
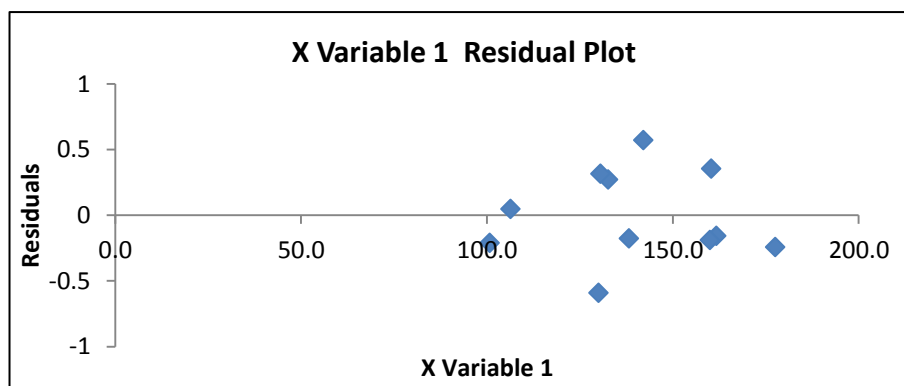
Figure 6.5 Residual plot for log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium benzoate derived salts

For this set of data a correlation was also found between the melting point of the free acid and the melting point of the salt, see Figure 6.6. A similar trend was seen in the study by Gould.[9] The data has a linear trend line with the equation $y = 0.4182x + 68.005$ with an error of 0.0980 and 17.4089 on the gradient and intercept respectively, with a correlation value of $R = 0.8182$. The range of the error can be seen in Figure 6.7. Again as before the residual plot, Figure 6.8, confirms the data is homoscedastic. As might be expected, this again was only true for the anhydrous species. We observe that the melting points of all three hydrates lie close to 80 °C and thus their relatively invariant melting points cannot help to predict the observed differences in solubility or be related to differences in free acid solubility.
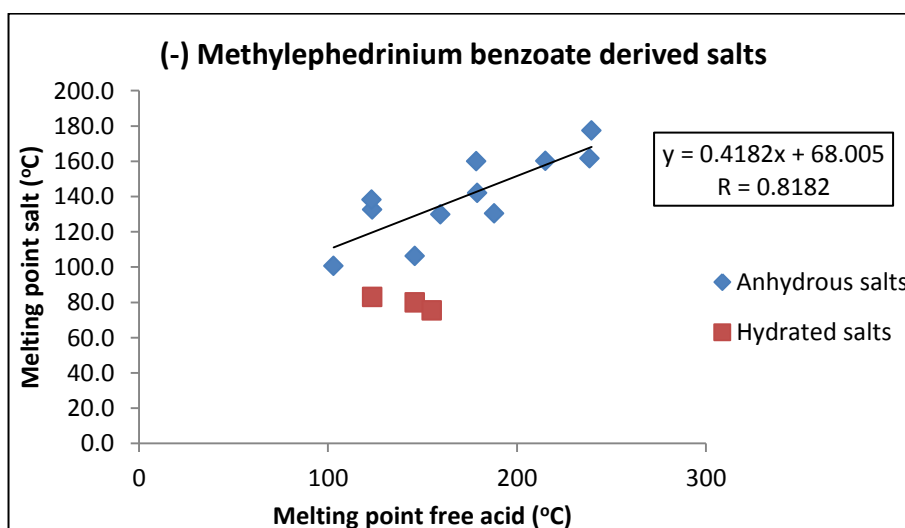
Figure 6.6 Plot of melting point of salt versus melting point of free acid for (1R,2S)(-) methylephedrinium benzoate derived salts

171
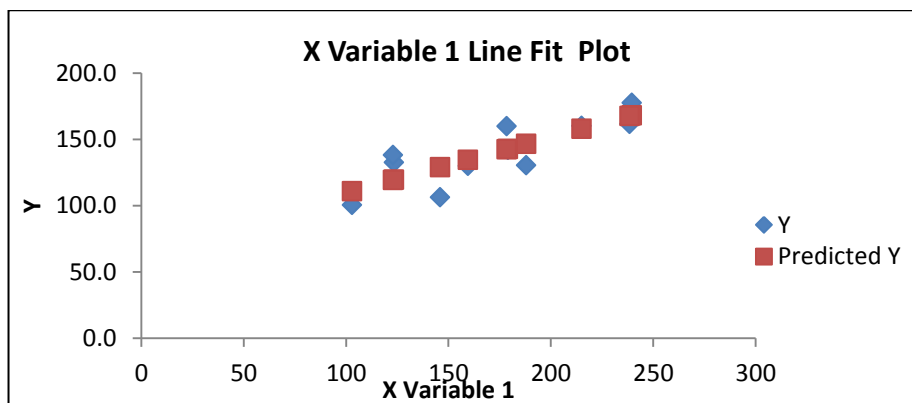
Figure 6.7 Variable line fit plot of melting point of salt versus melting point of free acid for (1R,2S)(-) methylephedrinium benzoate derived salts
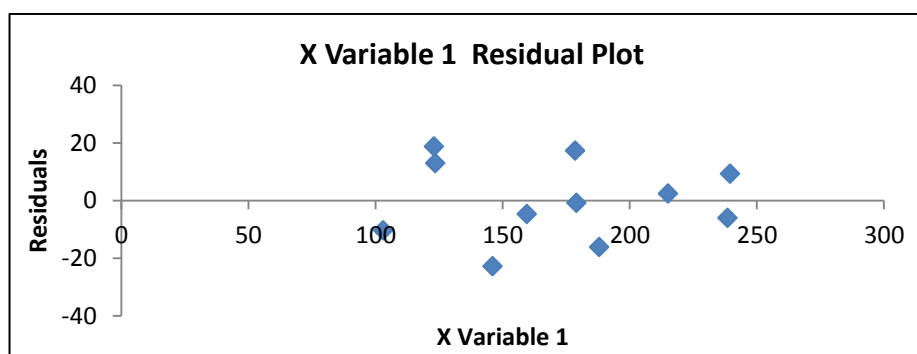
Figure 6.8 Residual plot for melting point of salt versus melting point of free acid for (1R,2S)(-) methylephedrinium benzoate derived salts

As there appears to be a correlation between melting point of salt and solubility and melting point of free acid and melting point of the salt, it was though that the melting point of the free acid may linearly correlate with the solubility of the salt. Curiously, this is not the case and although a negative correlation is apparent, there is no linear relationship seen between these properties, see Figure 6.9.

172
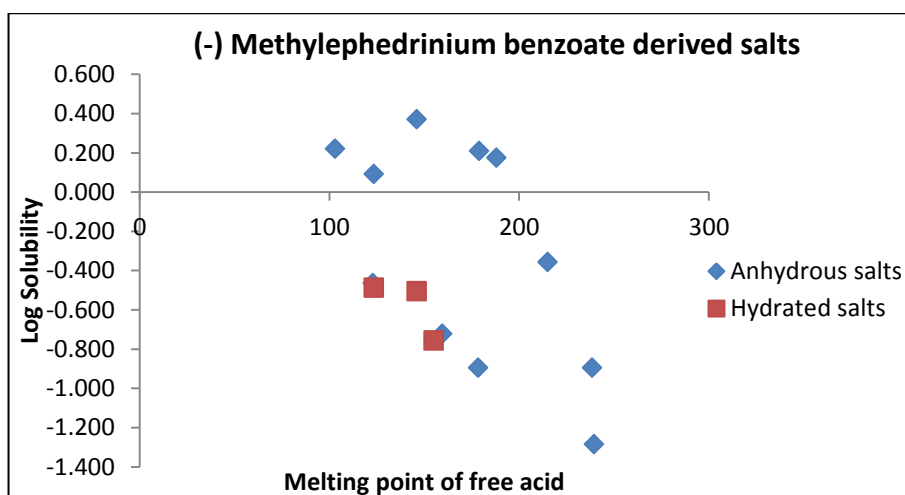
Figure 6.9 Plot of log solubility versus melting point of free acid for (1R,2S)(-) methylephedrinium benzoate derived salts

For these 14 benzoate derived salts no correlations were seen between density and either melting point or log solubility, as shown in Figure 6.10 and Figure 6.11. However, the plot of molecular weight versus log solubility is interesting, see Figure 6.12. There appears to be no correlation for the salts of lower molecular weight. These make up the bulk of the sample and have widely dispersed solubilities but fairly consistent molecular weights of between approximately 310 and 320. However, for the four heavier salts there appears to be some association with higher molecular weight correlating to higher solubility. Intriguingly this anti-intuitive trend encompasses both anhydrous and hydrated forms. In previous studies molecular weight has been found to be one of the important properties associated with predicting solubility.[20-22]
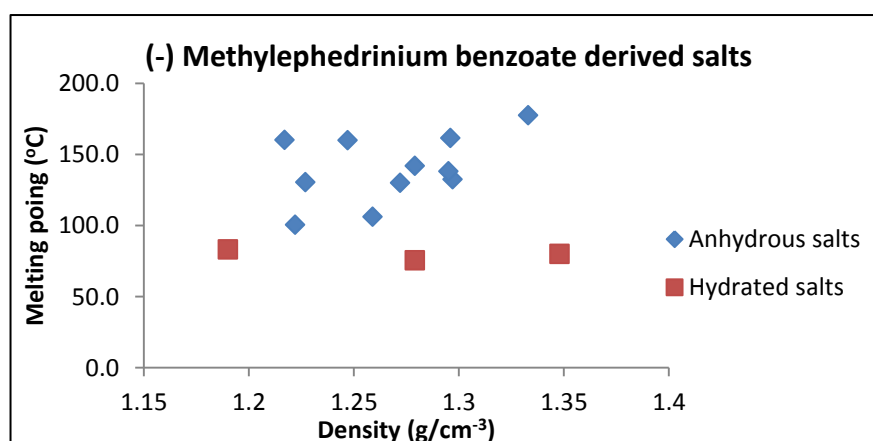
Figure 6.10 Plot of melting point of salt versus density for (1R,2S)(-) methylephedrinium benzoate derived salts
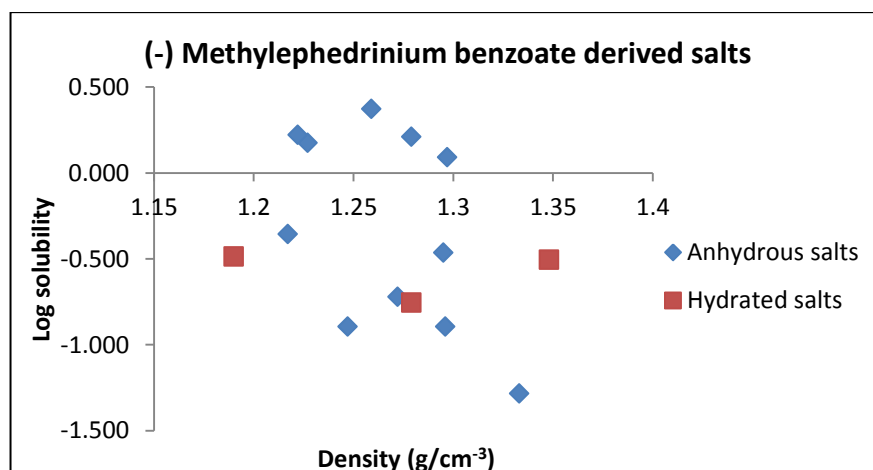
Figure 6.11 Plot of log solubility versus density for (1R,2S)(-) methylephedrinium benzoate derived salts
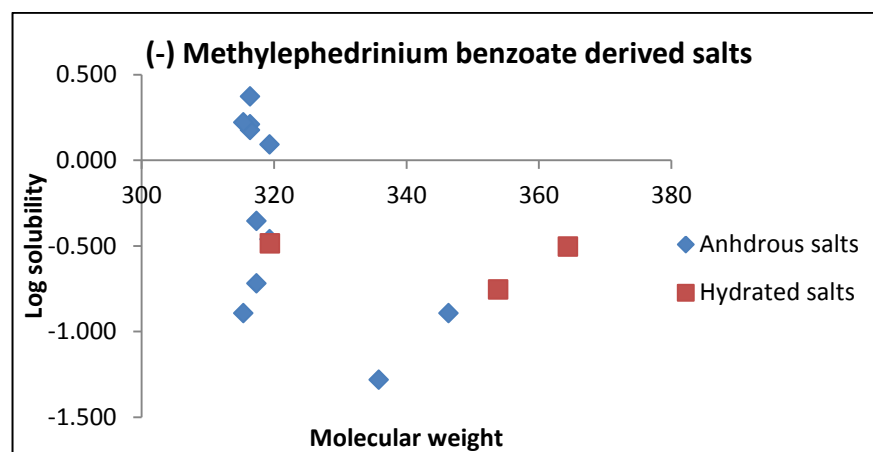
Figure 6.12 Plot of log solubility versus molecular weight for (1R,2S)(-) methylephedrinium benzoate derived salts

The final correlation that was attempted for the benzoate derived salts was that of solubility of free acid versus solubility of salt, as this has been shown to correlate in many literature studies.[2-4] For this analysis all free acid solubility values were taken from 'The handbook of aqueous solubility data'.[23] The extracted solubility data was measured at 25 $^{\circ}$C and where multiple solubility measurements were available an average value was used (excluding any obviously inaccurate values). The resulting plot of salt solubility versus free acid solubility is shown in Figure 6.13. This data shows no correlation between the two parameters.
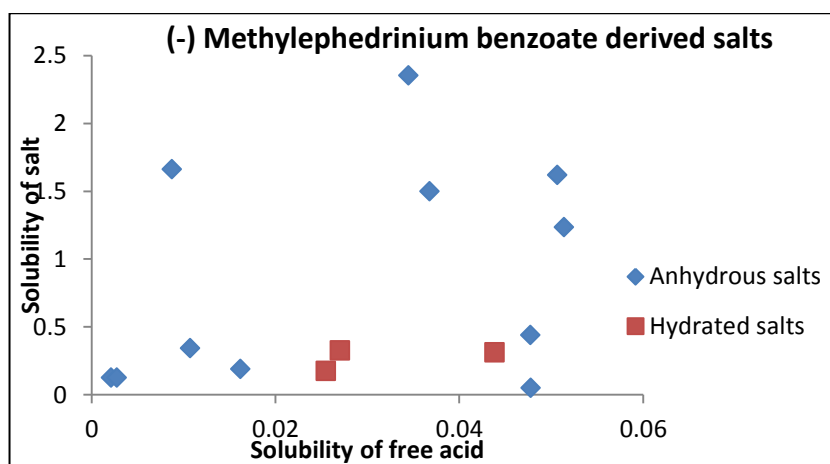
Figure 6.13 Plot of solubility of salt versus solubility of free acid for (1R,2S)(-) methylephedrinium benzoate derived salts

6.5 Correlation of physical properties for (+/-)methylephedrinium benzoate derived salts

The same analyses as above were performed for the racemic methylephedrinium salts of benzoate derivatives. There are 14 salts that are to be considered in this group. These consist of 10 anhydrous salts, and four salts that spontaneously resolve to form conglomerates, one of which is also a monohydrate. As seen for the enantiopure forms above the melting point of salt and log solubility data shows association, Figure 6.14. The anhydrous, conglomerate and hydrated conglomerate salt data all appears to correlate to each other in a linear relationship. It is suspicious here that the hydrate salt of RMEpdBz shows correlation, where the hydrate salts of (-) methylephedrine did not. It should be noted that the salt of RMEpdBz prior to the slurry stage of the solubility measurement was a racemic anhydrous salt and therefore as the hydration state appears to not affect the solubility measurement it is possible that the melting point of the wrong phase has been measured here. Alternatively, it may also simply be coincidence that this single value lies on the line. The latter is suspected to be true as a repeat melting point measurement of the material recovered from the slurry produced the same result. In either case, it was decided not to use this result as a data point when calculating the equation of the line.

A further curiosity is that fundamental thermodynamic treatments of solubility of conglomerates treat them as fundamentally different from racemic phases.[24] However, here there appears to be no fundamental difference. The solubility of conglomerates is discussed more fully in section 6.2 above. The equation of the line is log y = -0.0253x + 2.9294, with errors of 0.0062 and 0.8683 on the gradient and intercept respectively. As can be seen in Figure 6.15 and Figure 6.16 there is a reasonable linear relationship between the predicted and measured values with a correlating 'R' value of 0.8809.

Figure 6.14 Plot of log solubility versus melting point of salt for (+/-) methylephedrinium benzoate derived salts
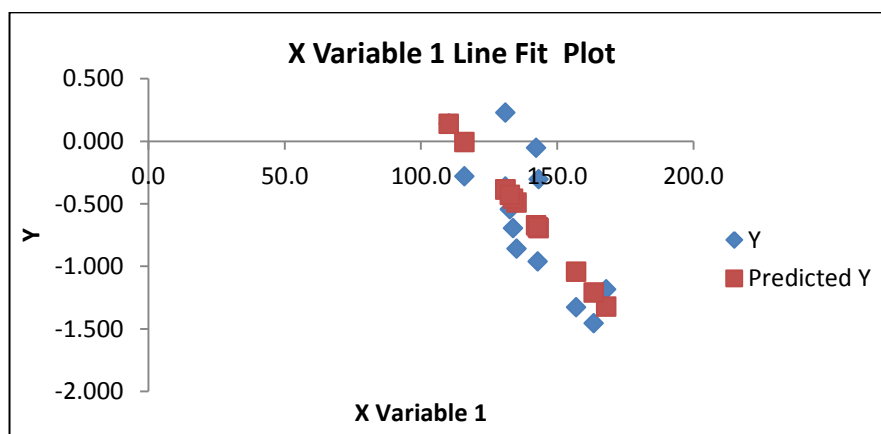
Figure 6.15 Variable line fit plot for log solubility versus melting point of salt for (+/-) methylephedrinium benzoate derived salts
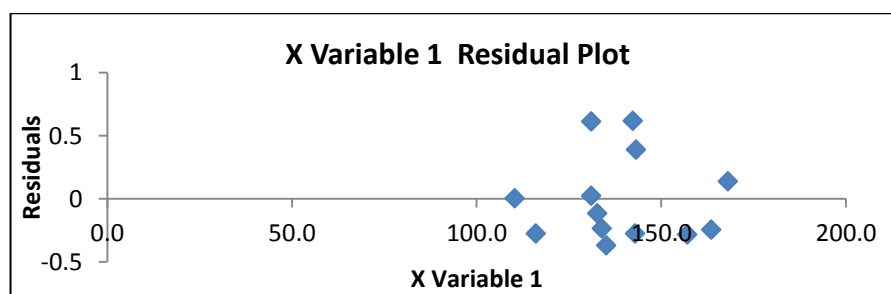
Figure 6.16 Residual plot for log solubility versus melting point of salt for (+/-) methylephedrinium benzoate derived salts

The graphs produced for the enantiopure and racemic methylephedrinium salts are equivalent within the errors of the data and therefore all the data can be collated together. This has been shown in Figure 6.17, with the new linear correlation of log y = -0.0217x + 2.5619, and regression coefficient of R = 0.7643. The error on the gradient and intercept are now 0.0038 and 0.5331 respectively.

Figure 6.17 Plot of log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium benzoate derived salts

The other data that shows a reasonable linear correlation for the (+/-)methylephedrinium salts is a comparison of the melting points of the free acids and the melting points of the salts, see Figure 6.18. Here again the monohydrated data is ignored and the conglomerate data is included. The association has a linear line of best fit, y = 0.2662x + 93.6585, with a regression coefficient R of 0.8556. The errors of the gradient and intercept are 0.0747 and 13.1456 respectively. Once again the

graph of the variable line fit plot and the residual plot confirm the linear correlation and the homoscedasticity of the dataset, see Figure 6.19 and Figure 6.20.
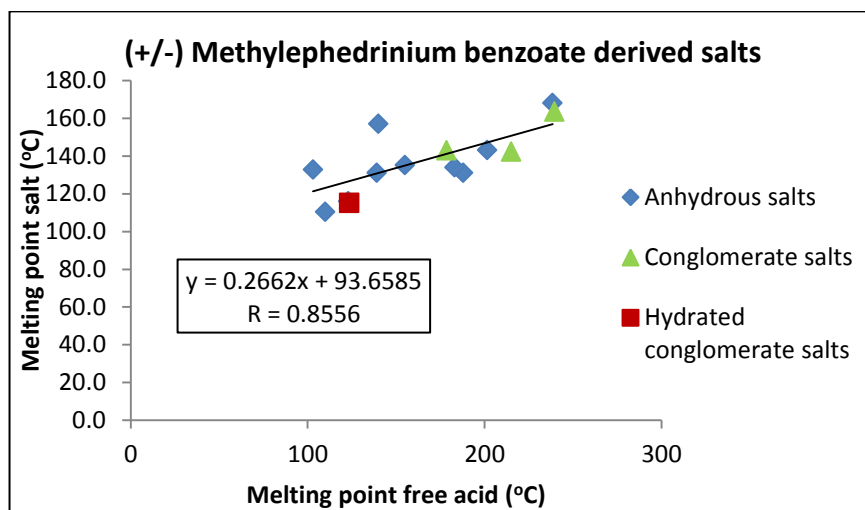
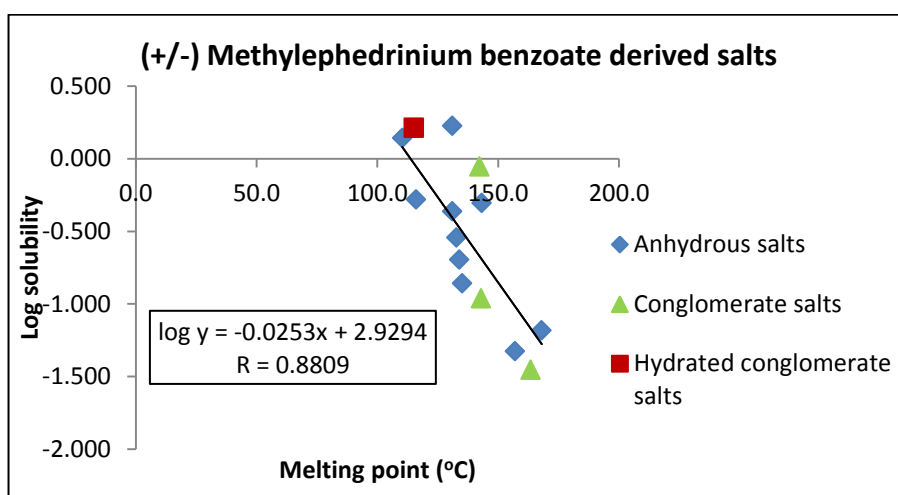**Figure 6.18** Plot of melting point of salt versus melting point of free acid for (+/-) methylephedrinium benzoate derived salts
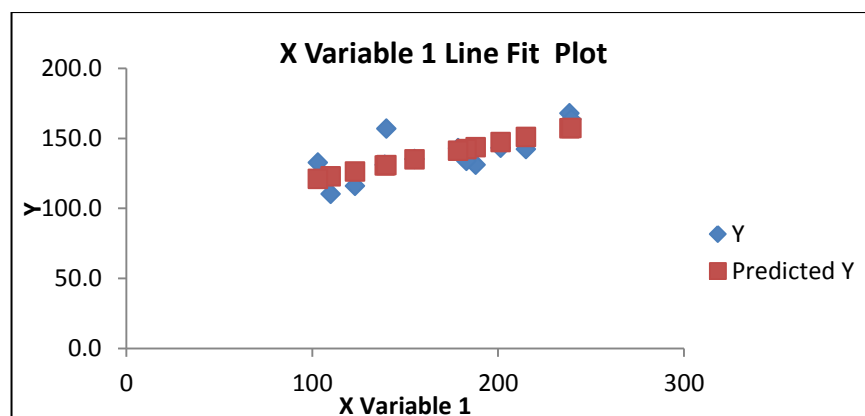
Figure 6.19 Variable line fit plot for melting point of salt versus melting point of free acid for (+/-) methylephedrinium benzoate derived salts
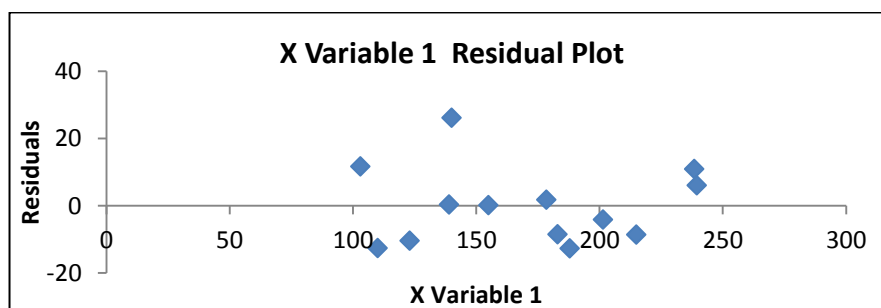
Figure 6.20 Residual plot for melting point of salt versus melting point of free acid for (+/-) methylephedrinium benzoate derived salts

178

As seen previously, with the melting point and log solubility data for the enantiopure and racemic salts, again the errors in the correlation lines of the enantiopure and racemic plots of melting point of free acid versus melting point of salt are within experimental error of each other and therefore the two data sets can be combined. This has been shown in Figure 6.21, with the new linear association of y = 0.3359x + 81.9680, and regression coefficient of R = 0.8763. The error on the gradient and intercept are now 0.0597 and 10.5622 respectively.
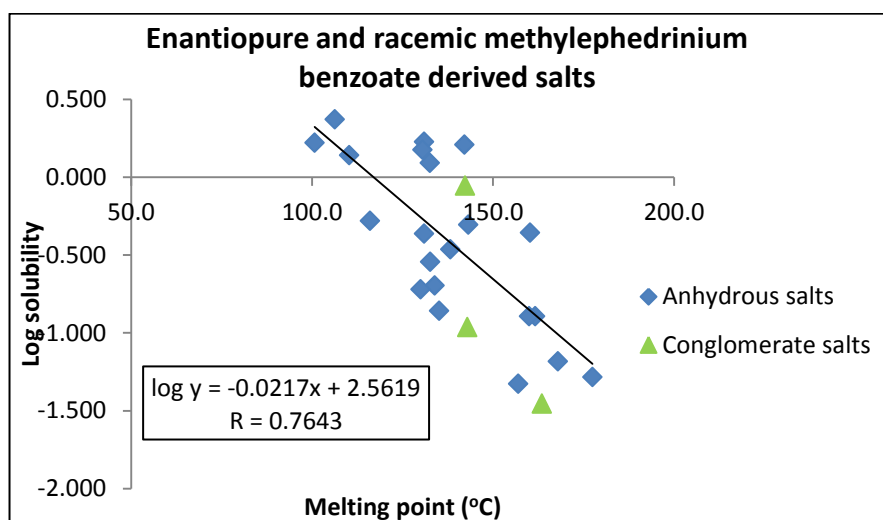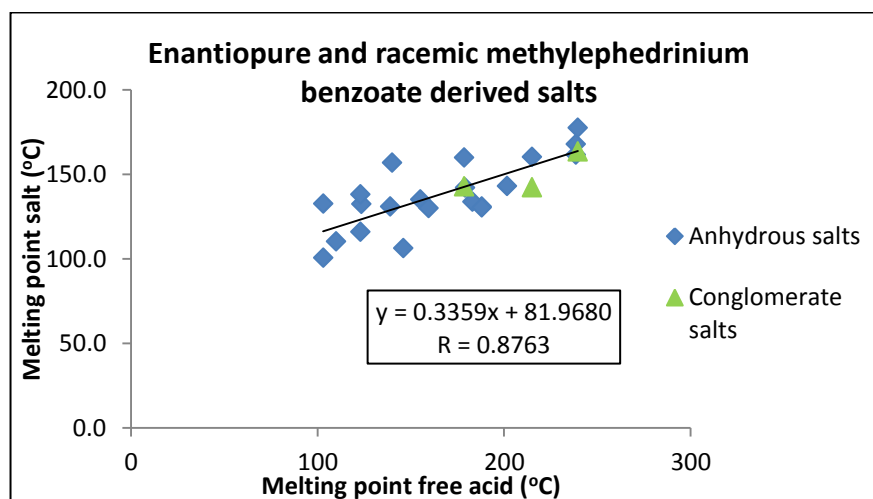

Figure 6.21 Plot of melting point of salt versus melting point of free acid for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium benzoate derived salts

For the racemic salts there is no trend seen at all for the plot of log solubility versus melting point of free acid, see Figure 6.22. As observed with the enantiopure methylephedrinium benzoate derived dataset there is no correlation between the density and melting point or density and solubility. The trend seen with the enantiopure salts for molecular weight of salt and solubility is absent here, see Figure 6.23.
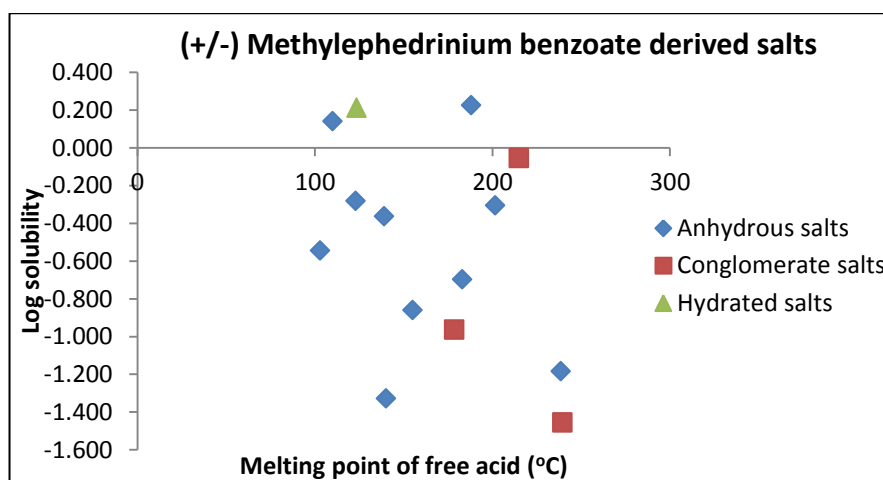
Figure 6.22 Plot of log solubility versus melting point of free acid for (+/-) methylephedrinium benzoate derived salts
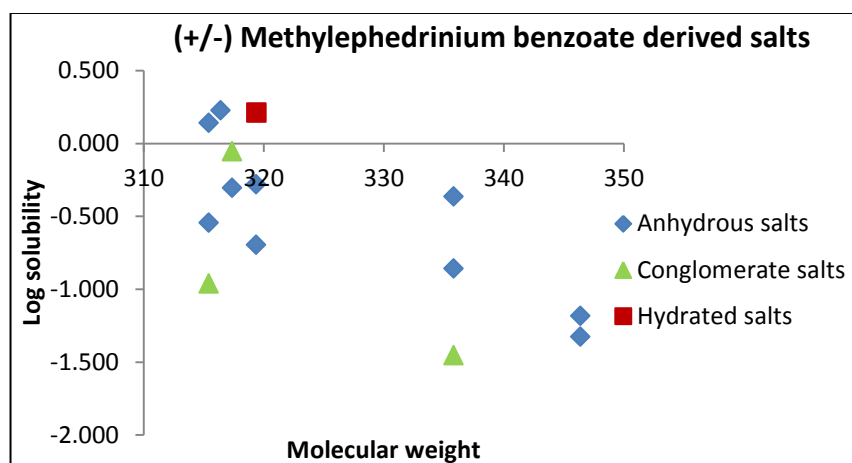

Figure 6.23 Plot of log solubility versus molecular weight for (+/-) methylephedrinium benzoate derived salts

6.6   Correlation of physical properties for the enantiopure and racemic methylephedrinium sulfonate, carboxylate and halide salts

For the comparison of the sulfonate, carboxylate and halide salts the enantiopure and racemic salts of each were analysed together as the benzoate comparison (above) did not show any difference in behaviour caused by the chirality of the cation. Unfortunately it is hard to decipher whether or not there is any correlation within sulfonate salt data sets as there are only six salts, two of which are hydrates. To narrow the dataset even further, melting point information is not available for RMEpdMeSO3 as that salt is hydroscopic. Figure 6.24 shows a plot of melting point

versus log solubility. Whilst it is hard to describe any concrete linear relationship, it can be noted that the hydrates and anhydrates seem to behave differently and that the higher melting point anhydrous forms have the lowest solubility.

Figure 6.24 Plot of log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium sulfonate derived salts

There are 12 carboxylate salts and due to their variety it was not expected that any association between the data points would be found. The group consists of three monocarboxylate and nine dicarboxylate salts. There is also a mix of one to one salts and two to one cation to anion salts as well as one hydrate and one co-crystal. It is conceivable that if more salts of the same group were measured it would be possible to correlate the data for melting point, density and solubility. However, with the available data it is not surprising that no linear correlation is seen, see Figure 6.25.

Figure 6.25 Plot of log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium carboxylate salts

181

For the halide salts both the iodide salts contain an I_2 within their structure. There appears to be a linear correlation between the melting point and the log solubility, see Figure 6.26. Interestingly, it has a positive association in comparison to the negative association seen with the benzoate salts. The correlation line has the equation log y = 0.0055x + 1.0319, with a regression value of R = 0.7935. This relationship suggests that the higher melting point correlates to higher solubility measurement and goes against the hypothesis that higher melting point relates to more stable compounds and therefore to lower solubility. For the melting point and the solubility the halide sequence is I < Br < Cl, so iodide has a low melting point and an unexpected low solubility. This relates to there being weak interactions with I^- in both the solid and solution forms. For the six halide salts there is also a linear association between density and log solubility with a correlation value of R = 0.8161, producing a linear equation of log y = -0.7528x + 1.0567, see Figure 6.27. Here the halide density sequence is I > Br > Cl. It is possibly a little odd that a linear association is observed here, as in the case of the halides the density does not correspond to packing efficiency but instead to the increasing size of the counterion. A negative correlation can also be seen between density and log melting point, see Figure 6.28. This linear correlation of the data has the equation log y = -0.3451x + 2.7386, with a correlation value of R = 0.8250.
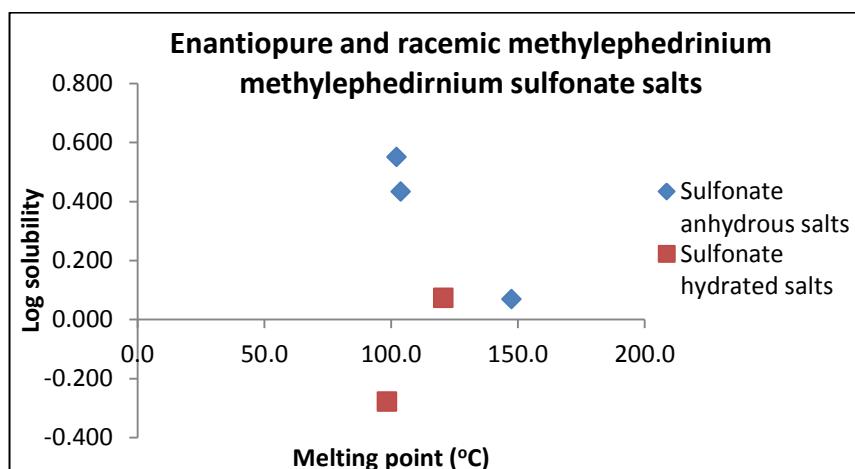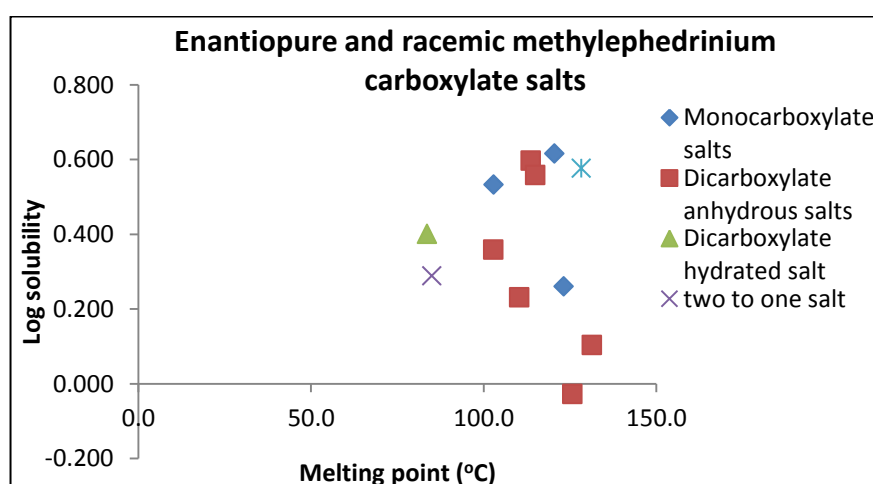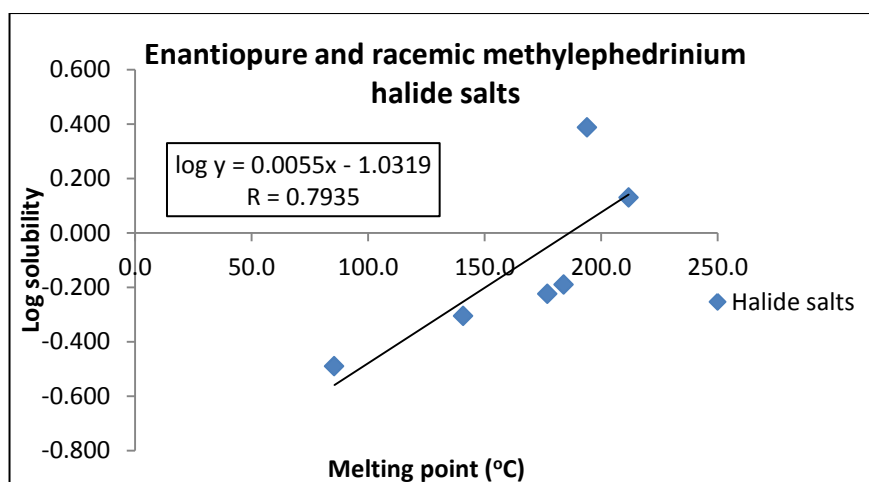

Figure 6.26 Plot of log solubility versus melting point of salt for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium halide salts
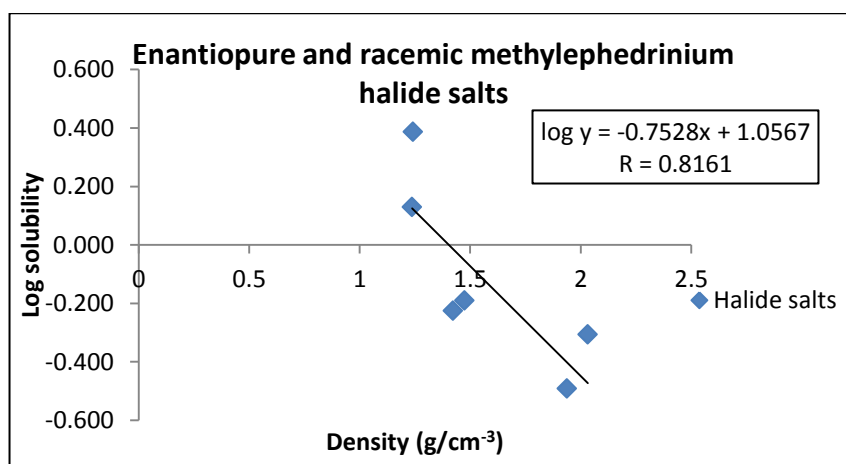
Figure 6.27 Plot of log solubility versus density for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium halide salts
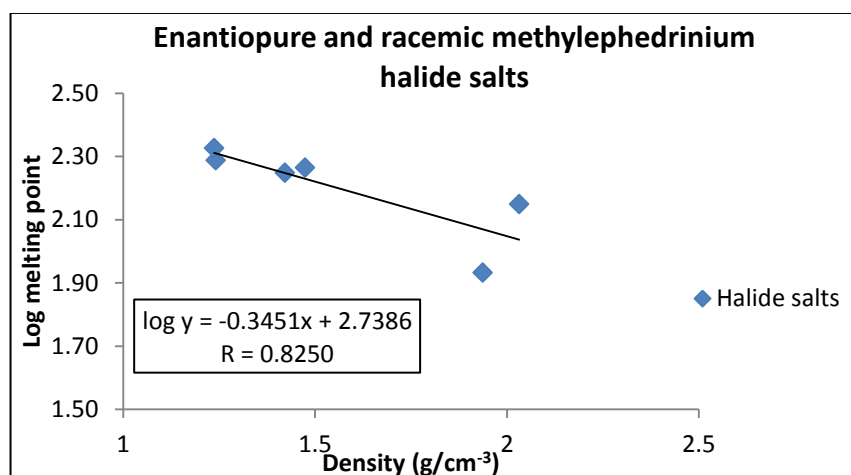
Figure 6.28 Plot of log melting point of salt versus density for (1R,2S)(-) methylephedrinium and (+/-) methylephedrinium halide salts

6.7 Investigation of different isostructural groups and effect on solubility

There are three groups at the bottom of the structure tree diagram which are isostructural with respect to their cation packing and contain more than two salts, see Section 4.9 Figure 4.15. These groups were examined for any trends or similarities in the data, as it may be expected that isostructural groups should have similar solubility behaviour as their identical packing may equate to similar stability. Group 5, which contains MEpd2FB, MEpd4CB, MEpdoTol and MEpdpTol has a linear correlation between melting point and log solubility, Figure 6.29, with a regression

coefficient of 0.9468. The linear equation is log y = -0.0208x + 2.5071 with errors of 0.0050 and 0.7291 on the gradient and intercept respectively. Group 6, which consists of RMEpd3CB, RMEpd3FB, RMEpd3NB and RMEpdmTol shows a similar linear correlation, Figure 6.30. The Equation is log y = -0.0299x + 3.2922 with a regression coefficient of 0.9793. For groups 5 and 6 the correlation line is the same as seen above for the collective benzoate salts however both lines are more accurate with a much better correlation. This indicates that similar structures lead to more similarity in solution behaviour. For group 7 the structures that were isostructural for ten cations was looked at and included MEpd2NB, MEpd3AB, MEpd3FB, MEpd4HBS, MEpd4NB and RMEpd4NB. Excluding the one hydrated structure (MEpd2NB) there is a linear correlation. This group is unique as it includes a sulfonate structure amongst the benzoates. This group also contains an aminobenzoate structure which has previously been shown not to correlate with the other benzoate derived salts. The linear correlation has the equation log y = -0.0376x + 5.2441 with an R value of 0.8027, see Figure 6.31. As with group 5 and 6 this shows that similarity in structure packing leads to similarity in solution behaviour regardless of the counterion be it a sulfonate or benzoate.
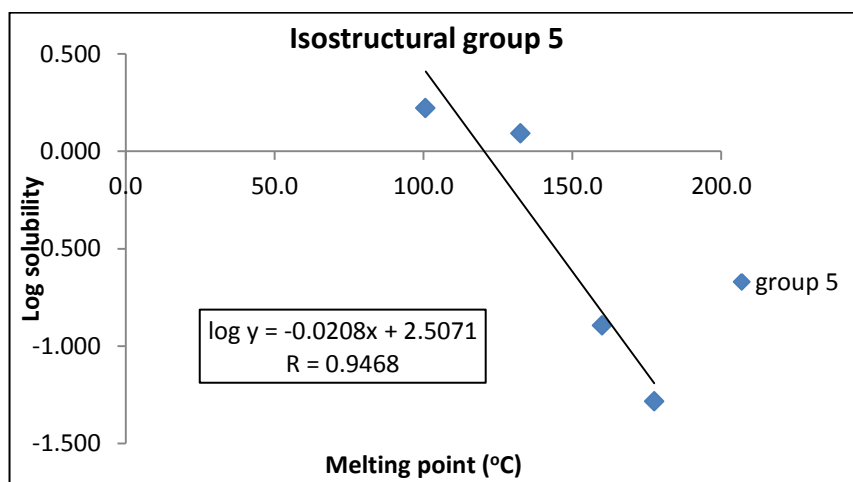
Figure 6.29 Plot of log solubility versus melting point of salt for methylephedrinium salts isostructural group 5
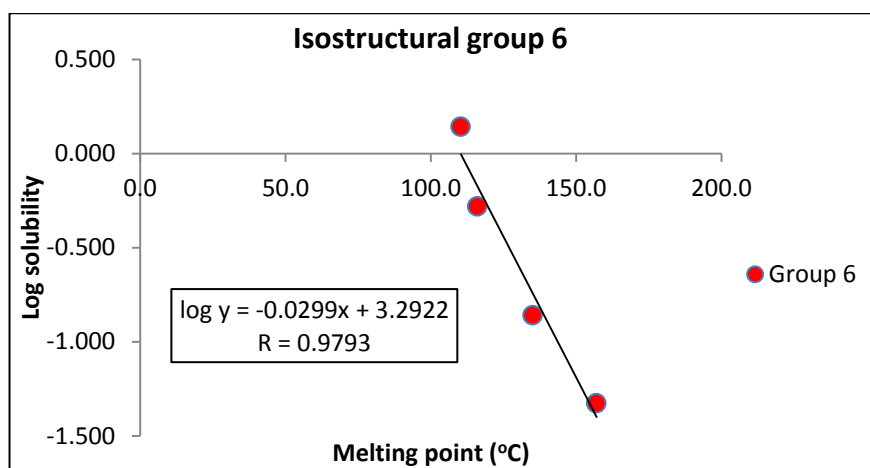
Figure 6.30 Plot of log solubility versus melting point of salt for methylephedrinium salts isostructural group 6
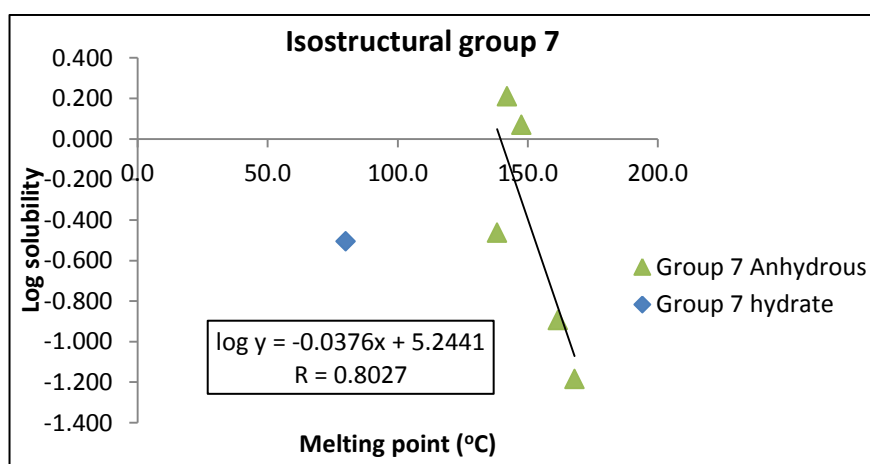
Figure 6.31 Plot of log solubility versus melting point of salt for methylephedrinium salts isostructural group 7

This section shows how it is possible to linearly correlate salt melting point and log solubility and melting point of salt and melting point of free acid for anhydrous salts. The enantiopure and racemic methylephedrinium benzoate derived salts melting point and solubility data correlates to produce the same negative linear relationship. When looking at isostructural groups of methylephedrinium salts the accuracy of this correlation greatly improves showing that similarity in structures relates to similar solubility behaviour. A positive linear correlation is seen between melting point of salt and log solubility for the enantiopure and racemic methylephedrinium halide salts. This correlation goes against the theory that higher melting point relates to more stable compounds and therefore lower solubility.

6.8 Correlation of melting point and log solubility for salts of other cations

The benzoate derived salts of the other nine phenylethylamine derived cations were investigated to see if there was a similar correlation between log solubility and melting point as seen with the methylephedrinium salts. With the exception of the tyrammonium salts, for each cation there are only a few salts available for data comparison. This limits the information available and hence the ability to detect true trends in the data. Thus for the α(methyaminomethyl)benzyl alcohol salts there appears to be no correlation between melting point and log solubility, as shown in Figure 6.32. However, as there are only four points this does not confirm that there is no correlation in the data. Assuming a similar trend line to that seen for methylephedrinium salts then the hydrate structure would be to outlier to the dataset – but the α(methyaminomethyl)benzyl alcohol data alone does not unambiguously show this.

Combining the data for the benzoate derived salts of phenethylamine, methylphenethylamine, dimethylphenethylamine and α(methylaminomethyl)benzyl alcohol salts shows a linear association between melting point and log solubility, with a correlation coefficient of 0.8176, see Figure 6.33. This data gives a line equation of log y = -0.0139x + 1.6884 with a gradient and intercept error of 0.0025 and 0.3109 respectively. These errors indicate that there is a correlation between this data and that of the methylephedrinium benzoate derived salts. To confirm and illustrate this correlation the data from all five cation benzoate salts is shown in Figure 6.34. This data associates linearly with a relationship of log y = -0.0172x + 2.0013 with a regression value R of 0.7864. As before this correlation is only consistent for the anhydrous salts. The data correlation of the enantiopure and racemic salts is understandable as they are otherwise chemically identical, however the further correlation with the α(methylaminomethyl)benzyl alcohol, methyl, dimethyl and phenethylammonium salts is unexpected as they are all distinctly different chemical species, both from each other and from the methylephedrinium

salts. Especially to be noted is their varying hydrogen bonding abilities which encompass cations derived from primary, secondary and tertiary amines as well as the presence or absence of the hydroxyl group on the phenethylamine backbone.
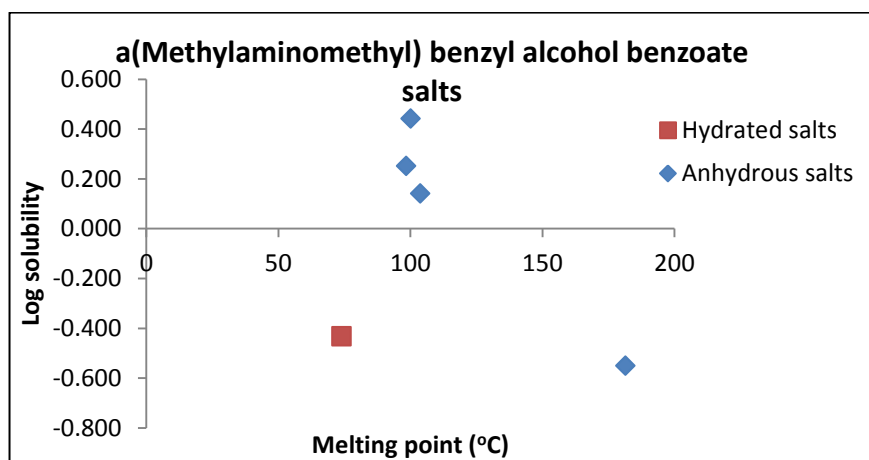
Figure 6.32 Plot of log solubility versus melting point of salt for α(methylaminomethyl) benzyl alcohol benzoate derived salts
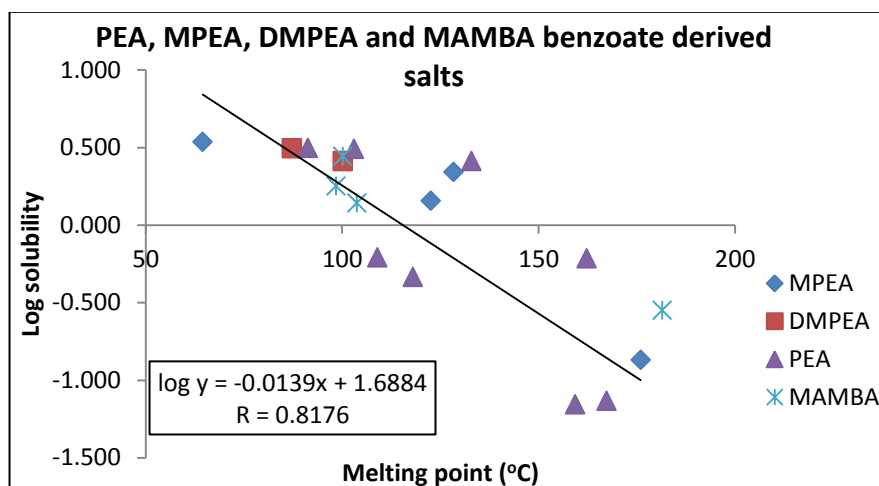
Figure 6.33 Plot of log solubility versus melting point of salt for phenethylammonium, methylphenethylaminium, dimethylphenethylaminium and α(methylaminomethyl) benzyl alcohol benzoate derived salts
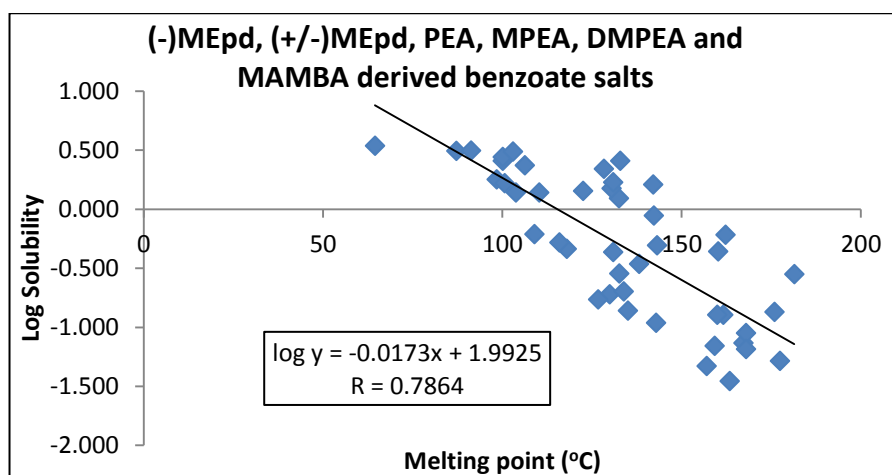
Figure 6.34 Plot of log solubility versus melting point of salt for (-) methylephedrinium, (+/-) methylephedrinium, phenethylammonium, methylphenethylaminium, dimethylphenethylaminium and α(methylaminomethyl) benzyl alcohol benzoate derived salts

The benzoate salts of the remaining four cations do not show similar trends to the above salts. Indeed there is limited correlation within the individual datasets. For the benzoate salts of (-)ephedrine and (-)pseudoephedrine, see Figure 6.35, several interesting points arise. Firstly the benzoate salts of (-)pseudoephedrine are mostly hydrates with five out of the seven known salts adopting a hydrated phase. This is extremely unusual as for all other salts there is only one or two hydrated salts present within the datasets. Secondly, with the exception of the (-)ephedrinium 4-aminobenzoate salt, there appears to be correlation between the two different cations. Also the hydrated salts seem roughly linear with the anhydrous salts, which has not been observed previously. Interestingly the hydrate salts are all more soluble which is unexpected from literature studies[17] but has previously been observed for the (-) methylephedrinium salts. Unlike the enantiomorphic methylephedrinium salts, (-) ephedrine and (-) pseudoephedrine are diastereomers and therefore possess different physical properties and chemical reactivity. For the salt of (-) ephedrinium 4-aminobenzoate there is no obvious reason why the data point does not correlate with the others, although as discussed in Section 6.2 and in the literatures anomalous solubility values are often measured for the aminobenzoate salts.[2]

The study by Black et al.[10] on the 17 structures of (-) ephedrine does not contain any benzoate salts. However, on comparison of the other salt solubility measurements to those measured during this study the solubility values are not always directly comparable (the previous study was performed using gravimetric analysis and this study uses HPLC analysis) but they are within a reasonable error of each other.
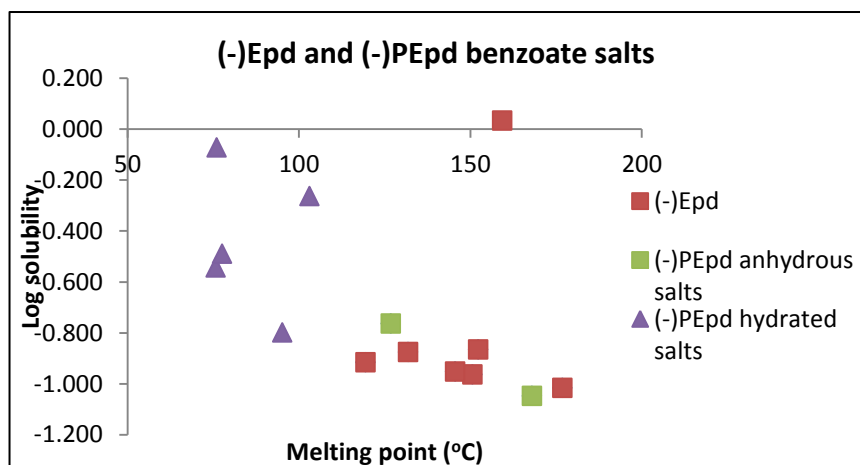
Figure 6.35 Plot of log solubility versus melting point of salt for (-) ephedrinium and (-) psuedoephedrinium benzoate derived salts

The tyrammonium benzoate salts again apparently show a linear association between melting point and log solubility, Figure 6.36, with a relationship of log y = -0.0098x + 1.0868. This line does not however encompass all the data points, tyrammonium 4-hydroxybenzoate (the top red spot) does not fit. There are two tyrammonium benzoate salts (TYR4CB and TYR4HB) that have a large number of cation anion pairs per asymmetric unit, here being eight and four respectively. For the case of TYR4HB it is possible that this large value of 'Z' equates to instability within the structure and therefore a greatly increased solubility. This would explain the outlying result of TYR4HB, however the measurement for TYR4CB does not appear to be an outlier and therefore weakens this hypothesis.
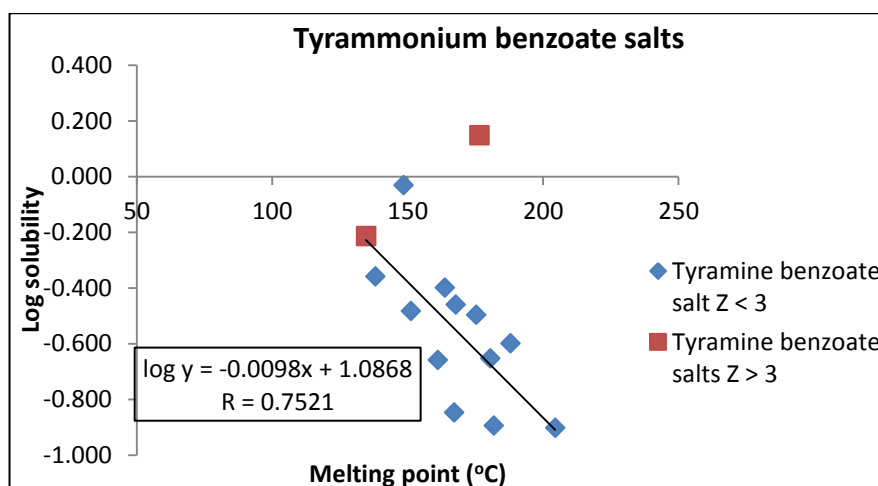
Figure 6.36 Plot of log solubility versus melting point of salt for tyrammonium benzoate derived salts

The final set of benzoate salts to consider is those of hydroxyphenethylamine. Unlike all of the above salts there appears to be a positive linear correlation between melting point and log solubility, Figure 6.37. This group contains three anhydrous salts and two hydrated salts, the hydrated salt of HPEA4CB has a higher melting point than expected so much so that it correlates with the anhydrous salts. This result may be due to the measured melting point actually being that of the anhydrous chlorobenzoate salt and not the monohydrate as expected. The melting point and log solubility are correlated by log y = 0.0037x + 1.0601 with an R value of 0.9931. This positive linear relationship is not seen for any of the other benzoate salts but is present in the halide salts of enantiopure and racemic methylephedrine.
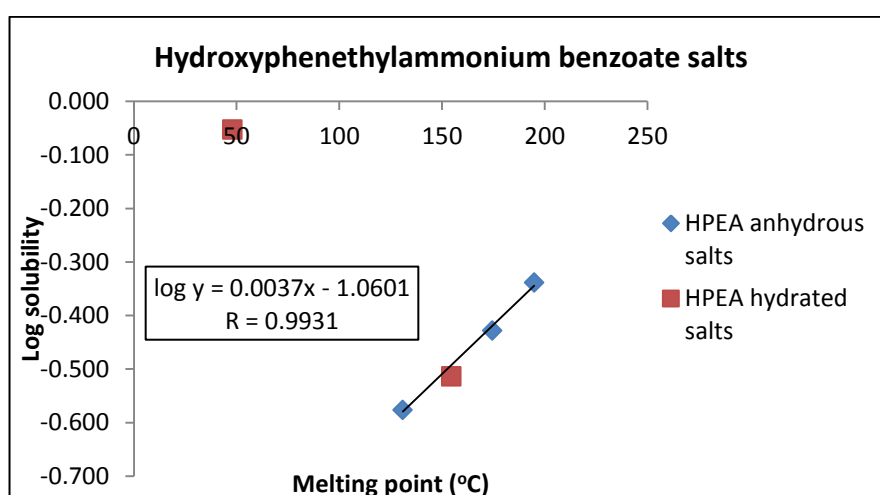
Figure 6.37 Plot of log solubility versus melting point of salt of salt for hydroxyphenethylammonium benzoate derived salts

The same negative correlation that is seen for the methylephedrinium benzoate derived salts is also seen with benzoate derived salts of α(methylaminomethyl)benzyl alcohol, phenethylammonium, methylphenethylaminium and dimethyl-phenethylaminium salts. This dataset correlation may be unexpected as it incorporates primary, secondary and tertiary amine salts. A negative correlation between salt melting point and log solubility is also apparent for the anhydrous benzoate salts of tyramine and (-)ephedrine and (-)pseudoephedrine. The hydrated salts of (-)pseudoephedrine also show a linear trend. The majority of the hydrated benzoate salts have a higher solubility than the anhydrous salts of the same cation. This is generally unexpected as anhydrous forms of crystalline materials are less water soluble than their anhydrous equivalent. A positive linear correlation is seen between melting point of salt and log solubility for the hydroxyphenethylammonium benzoate derived anhydrous salts, similar to that seen for the halide salts of methylephedrine.

6.9 Comparison of different anions present in benzoate derived salts to establish relationships across all cation domains

Throughout the analysis so far it has been established that aminobenzoate salts tend to not have the same associations between melting point and log solubility as the other benzoate salts studied. This is most notably seen for Epd4AB but is also the case for the methylephedrinium and tyrammonium aminobenzoate salts where the data points for these salts seem to lie on the edge of the correlations. The theory that the physical properties of aminobenzoate salts may trend when examining the anion instead of their cation was tested. Figure 6.38 below shows the plot of log solubility versus melting point for all isomers of aminobenzoate. It shows that there is a clear association with a relationship of log y = -0.0111x + 1.6416 and a regression factor, R = 0.9154. This confirms that the aminobenzoate salts correlate with respect to their anion and not their cation, as was seen for other benzoate salts. This is further illustrated in the other figures below which clearly show there is no or limited

association seen when data is grouped with respect to the other benzoate anions examined see Figure 6.39, Figure 6.40, Figure 6.41, Figure 6.42 and Figure 6.43 below. All hydrated salts were excluded from these comparisons.
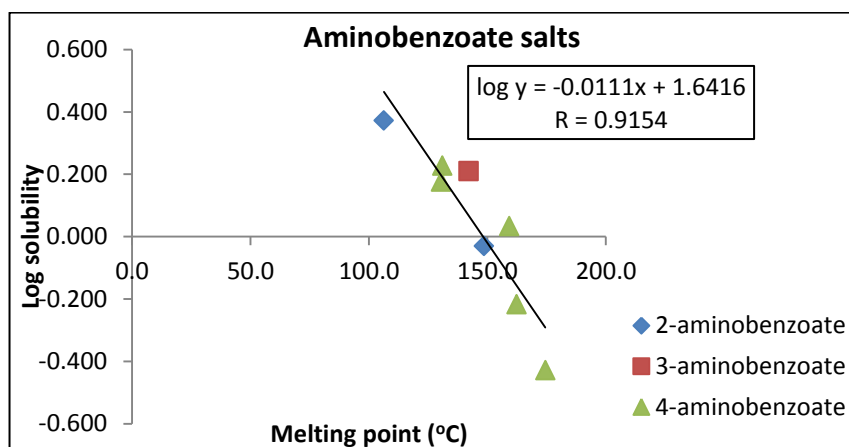
Figure 6.38 Plot of log solubility versus melting point of salt for aminobenzoate salts
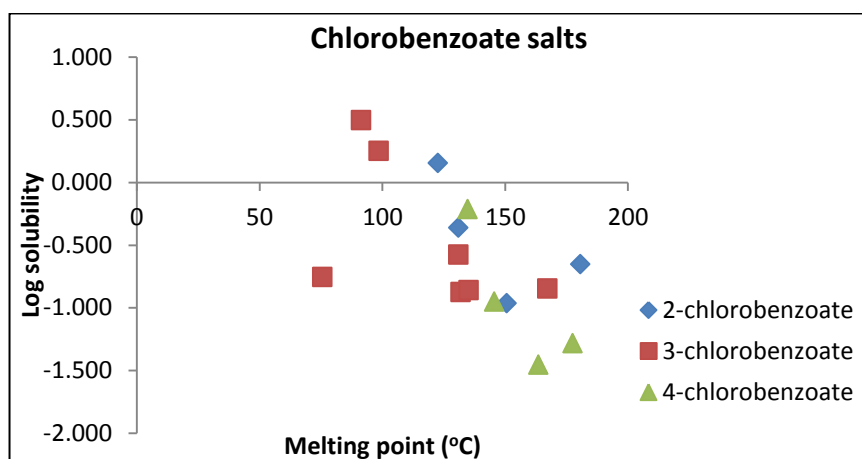
Figure 6.39 Plot of log solubility versus melting point of salt for chlorobenzoate salts
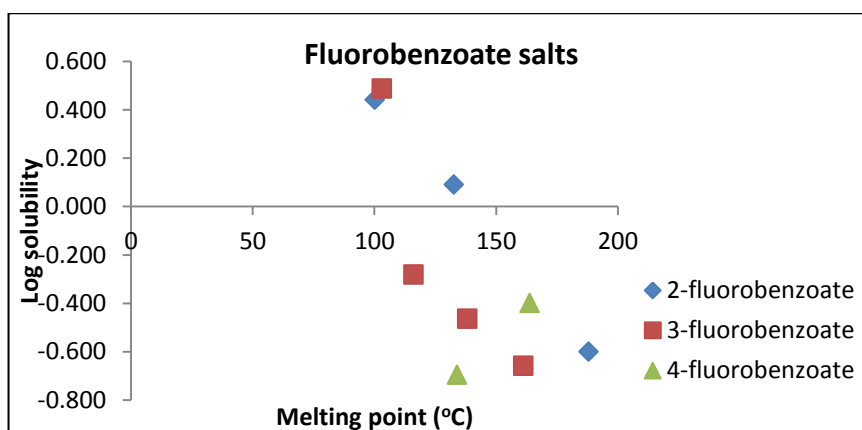
Figure 6.40 Plot of log solubility versus melting point of salt for fluorobenzoate salts
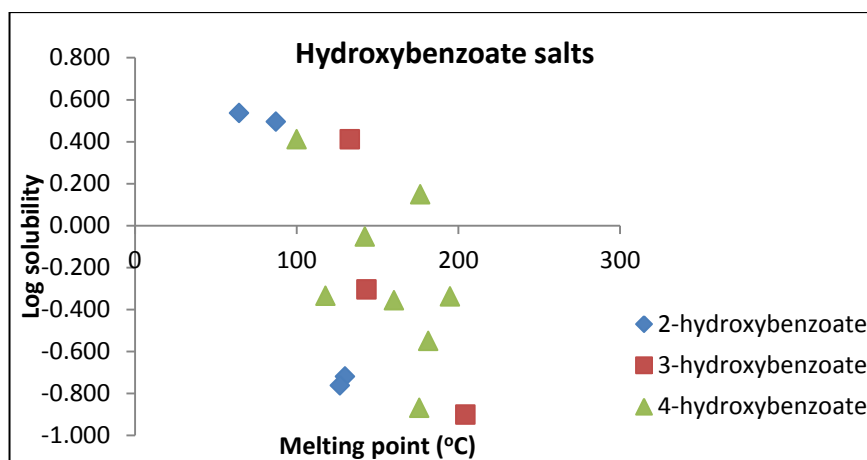
Figure 6.41 Plot of log solubility versus melting point of salt for hydroxybenzoate salts
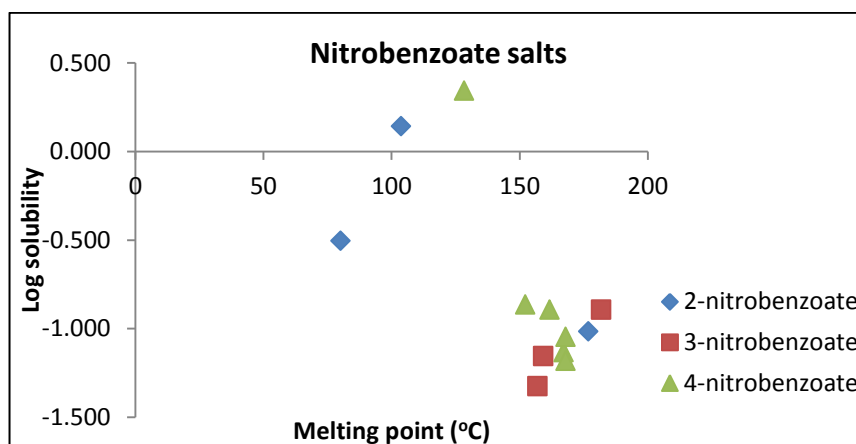
Figure 6.42 Plot of log solubility versus melting point of salt for nitrobenzoate salts
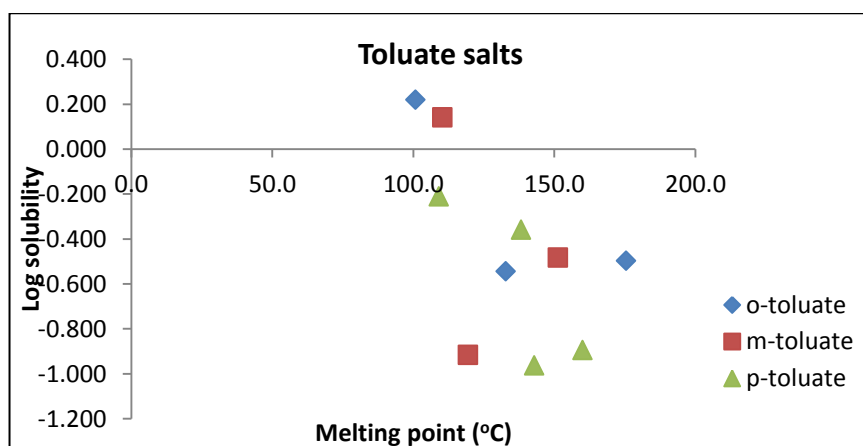
Figure 6.43 Plot of log solubility versus melting point of salt for toluate salts

There does not seem to be any association between the log solubility and melting point for all the other salts (sulfonates, halides, dicatboxylates).

6.10 Relationship between Hammett values and solubility

In a tangent to the above correlation analysis, Hammett values for the *meta-* and *para-* substituted benzoic acid derived salts were studied in order to ascertain any relationship between varying substituent type and/or position and solubility. The Hammett value is the classical method for describing both the electronic induction and resonance effects of a substituent with respect to its position on the aromatic ring. As it does not describe steric effects well, similar values for *ortho*-substituents are normally avoided. The constants used for this study are stated in Table 6.3 below and were taken from Exner, in Chapman and Shorter,[25] with the exception of the two hydroxyl constants which were extracted from Matsui et al.[26]

Table 6.3 Hammett values used for analysis taken from Exner, in chapman and Shorter[25] and Matsui et al.[26]

Group	σ_{para}	σ_{meta}
NH$_2$	-0.57	-0.09
OH	-0.38	0.13
CH$_3$	-0.14	-0.06
H	0	0
F	0.15	0.34
Cl	0.24	0.37
NO$_2$	0.81	0.71

For the methylephedrinium salts of *meta-* substituted benzoates, a plot of Hammett values against the corresponding log solubility produced a linear association, see Figure 6.44, with a linear equation of log y = -1.845x + 0.0465 and a correlation coefficient of 0.9504. This indicates that the electron donating substituents such as NH_2 and CH_3 at the *meta*-position result in a greater solubility than benzoic acid derived salts with electron withdrawing substituents such as F, Cl and NO_2. The only other salt group available with a substantial number of *meta-* substituted benzoic acid derived salts was the tyrammonium salt group. A plot of Hammett value versus log solubility again produces a linear correlation with the equation, log y = -0.5092x - 0.5814 and with a correlation coefficient of 0.7125, see Figure 6.45. This correlation value is not as high as that seen for the methylephedrinium salts, apparently due to the 3-hydroxybenzoate salt having a lower solubility value than might be expected.
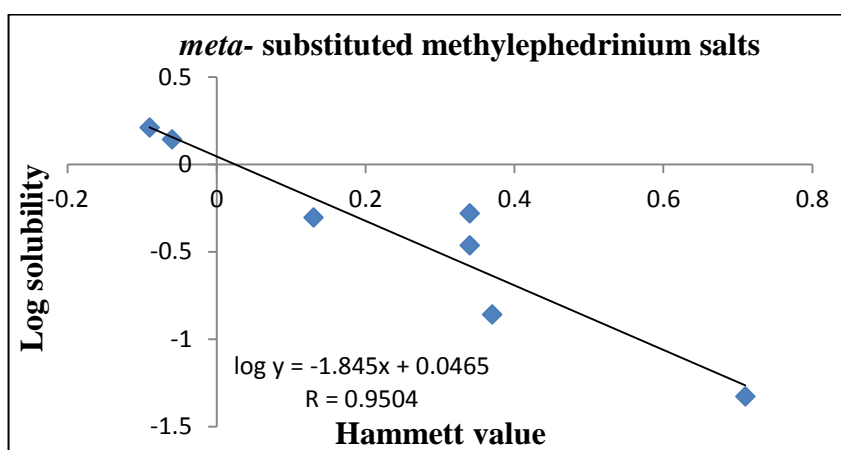
Figure 6.44 Plot of log solubility versus Hammett values for *meta-* substituted methylephedrinium salts
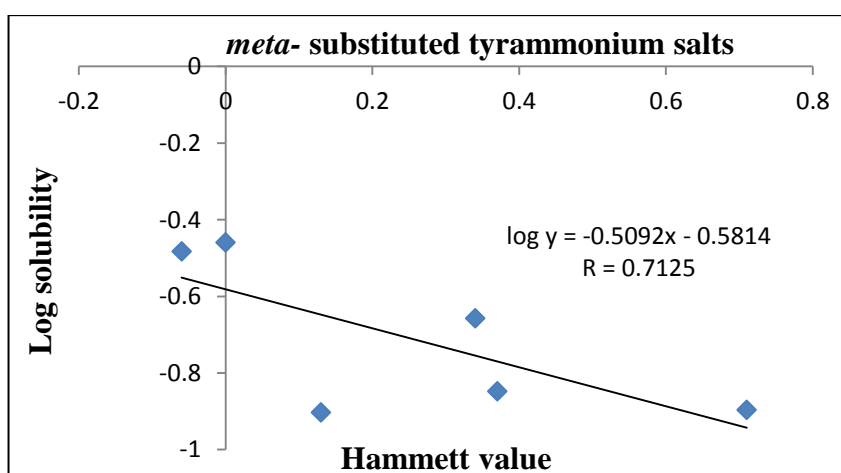
Figure 6.45 Plot of log solubility versus Hammett values for *meta-* substituted Tyrammonium salts

Similar plots to the above were attempted for the *para-* substituted benzoate derived salts for the methylephedrinium and tyrammonium salts. The results were interesting as no linear relationship was seen between the Hammett value and the log solubility of the corresponding salts. Instead the results appear to fit a parabola trend with salts with a high Hammett value having a higher solubility value than seen for the *meta-* substituted salts, see Figure 6.46 and Figure 6.47. The parabola contains electron releasing substituents on the negative slope and electron withdrawing substituents on the positive slope.

Traditionally non-linear Hammett plots suggest that the substituent affects the rate of reaction or a change of the rate determining step.[27] A study by Hancock et al.[28] shows a V-shaped Hammett plot for the solubility of *para-*substituted benzoic acids. It would be nice to assume that the non linear plot is due to a change in the solubility mechanism of the *para-* benzoate salts. However, unlike chemical reactions where there is a specific 'reaction' site it is important to consider that the entire structure of the compound, as it is all involved in the solution process.

para- substituted methylephedrium salts

$$\log y = 1.2411x^2 - 1.1577x - 0.9041$$
$$R = 0.9478$$
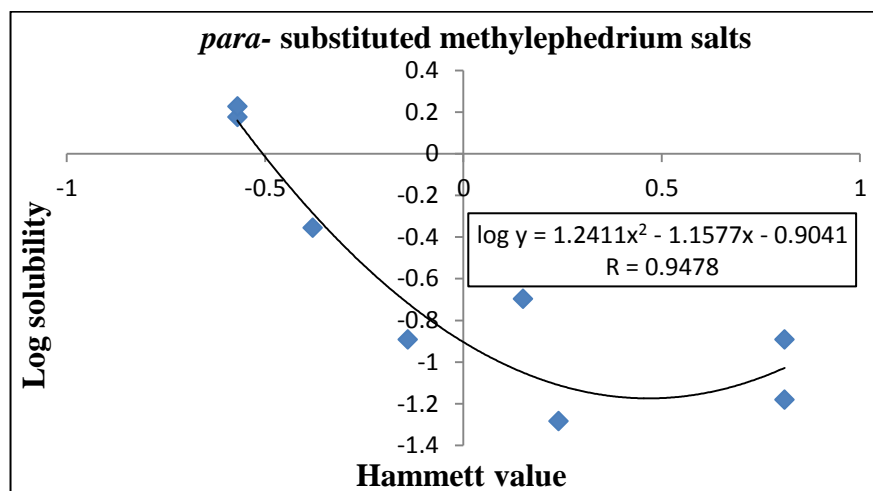
Figure 6.46 Plot of log solubility versus Hammett values for *para-* substituted methylephedrinium salts
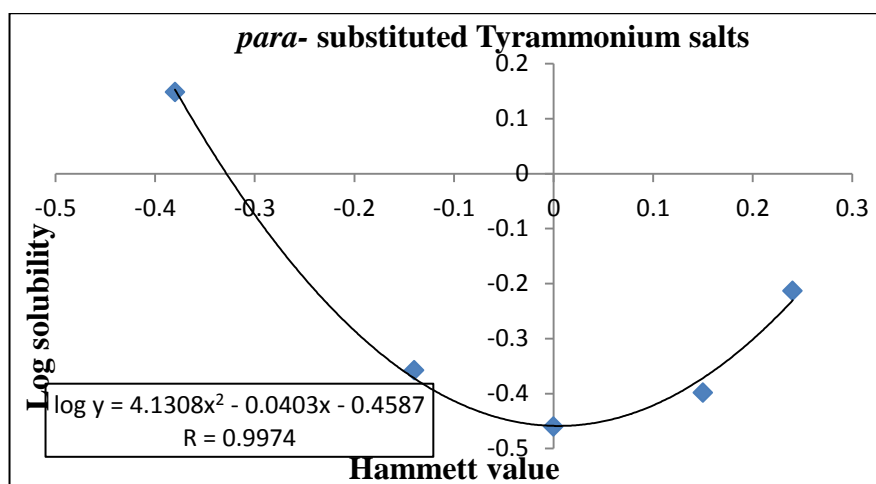
Figure 6.47 Plot of log solubility versus Hammett values for *para-* substituted Tyrammonium salts

When investigating this dataset of salts it is apparent how important the anion choice is on the observed solubility. In the case of the aminobenzoate salts the solubility correlates with the anion present instead of the cation, which is seen for all of the other benzoate salts. This is assumed to be caused by the fact that aminobenzoic acids are amphoteric. The solubility of the salt is affected by whether the benzoate counterion has an electron donating or electron withdrawing group on the *meta-* or *para-* position. There is a linear negative correlation between Hammett value and log solubility for *meta-* substituted benzoate salts and a parabolic correlation for *para-* substituted benzoate salts.

6.11 References

(1)   Bach Nielsen, A.; Frydenvang, K.; Liljefors, T.; Buur, A.; Larsen, C. *European Journal of Pharmaceutical Sciences* **2005**, *24*, 85.
(2)   Arlin, J. B.; Florence, A. J.; Johnston, A.; Kennedy, A. R.; Miller, G. J.; Patterson, K. *Crystal Growth & Design* **2011**, *11*, 1318.
(3)   Parshad, H.; Frydenvang, K.; Liljefors, T.; Larsen, C. S. *Int. J. Pharm.* **2002**, *237*, 193.
(4)   Galcera, J.; Molins, E. *Crystal Growth & Design* **2009**, *9*, 327.
(5)   Anderson, B. D.; Conradi, R. A. *J. Pharm. Sci.* **1985**, *74*, 815.
(6)   Rubino, J. T. *J. Pharm. Sci.* **1989**, *78*, 485.
(7)   Agharkar, S.; Lindenbaum, S.; Higuchi, T. *J. Pharm. Sci.* **1976**, *65*, 747.
(8)   Thomas, E.; Rubino, J. *Int. J. Pharm.* **1996**, *130*, 179.
(9)   Gould, P. L. *Int. J. Pharm.* **1986**, *33*, 201.

(10) Black, S. N.; Collier, E. A.; Davey, R. J.; Roberts, R. J. *J. Pharm. Sci.* **2007**, *96*, 1053.

(11) Guerrieri, P.; Rumondor, A. C. F.; Li, T.; Taylor, L. S. *AAPS PharmSciTech* **2010**, *11*, 1212.

(12) Mazzenga, G. C.; Berner, B. *J. Controlled Release* **1991**, *16*, 77.

(13) *Handbook of Industrial Crystallization*; Myerson, A. S., Ed.; Butterworth-Heinemann: Boston, 1993.

(14) Meyerhoffer, W. *Ber. Dtsch. Chem. Ges.* **1904**, *37*, 2604.

(15) Paulekuhn, G. S.; Dressman, J. B.; Saal, C. *J. Med. Chem.* **2007**, *50*, 6665.

(16) Arshadi, M.; Yamdagni, R.; Kebarle, P. *J. Phys. Chem.* **1970**, *74*, 1475.

(17) Shefter, E.; Higuchi, T. *J. Pharm. Sci.* **1963**, *52*, 781.

(18) Wallach, O. *Liebigs Ann Chem* **1895**, *286*, 90.

(19) Grant, D. J. W.; Higuchi, T. *Solubility Behaviour of Organic Compounds*; Wiley: New York, 1990.

(20) Gozalbes, R.; Pineda-Lucena, A. *Biorg. Med. Chem.* **2010**, *18*, 7078.

(21) Tantishaiyakul, V. *Int. J. Pharm.* **2004**, *275*, 133.

(22) Tantishaiyakul, V. *J. Pharm. Biomed. Anal.* **2005**, *37*, 411.

(23) Yalkowsky, S. H.; He, Y. *Handbook of aqueous solubilitry data*; CRC Press: Florida, 2003.

(24) Jacques, J.; Collet, A.; Wilen, S. H. *Enantiomers, Racemates, and Resolutions*; Wiley: New York, 1981.

(25) Exner, O. *Correlation Analysis in Chemistry: Recent Advances*; Plenum Press: New York, 1978.

(26) Matsui, T.; Ko, H. C.; Hepler, L. G. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **1974**, *52*, 2906.

(27) Schreck, J. O. *J. Chem. Educ.* **1971**, *48*, 103.

(28) Hancock, C. K.; Idoux, J. P. *J. Org. Chem.* **1967**, *32*, 1931.

Chapter 7

Random Forest Analysis

7.1   Introduction

Over the last decade there has been a large increase in the number of candidate drug molecules synthesised, this has been caused largely by extensive use of modern practices such as combinational chemistry techniques, high throughput screening and robotics.[1] These techniques have many advantages, though an unintended consequence is that they have caused the sharp decrease in the typical solubility of new candidate drug molecules, making the choice of the correct drug molecule at the early stage of development of paramount importance.[2] To assist in this decision the art of predicting physical properties from structure alone has been investigated. Quantitative structure-property relationships (QSPR) calculations have become easier to perform in recent years due to increased computer power, which allows for quick calculation of hundreds of 2D and 3D molecular descriptors and quick output from analysis programs.[3]

There are many different methods currently used in QSPR studies including; principle component analysis (PCA), multiple linear regression (MLR), partial least squares (PLS), artificial neutral network (ANN) and Random forest (RF). All these techniques have their associated advantages and disadvantages due to their varying complexities. Currently the most successful models use large and expansive training sets to predict aqueous solubility of organic molecules.[4-6] The study by Ghasemi et al.[4] had success using a MLR model with a 110 compound training set and 40 compound prediction set with low prediction errors. The study by Kovdienko et al.[6] contains a very large data set of 2800 organic compounds and found that the prediction was more accurate when using a Random forest model compared to a MLR model. Both these studies build a regression model based on numerical solubility data, the alternative is to build a classification model based on categorical solubility data, such as low, medium and high solubility.[5] These models were successful despite relying in literature solubility data of varying quality.

With respect to aqueous solubility prediction models for salts in the literature, there have been only a handful of studies performed. These have had changeable success. Tantishaiyakul et al.[7,8] performed two similar studies using PLS and ANN model for a group of diclofenac and benzylamine salts. These models contain far fewer salt compounds than observed for the above studies of organic compounds, namely 23 and 22 salts respectively. However, the presence of fewer salts in the training set still allows for an accurate model with R^2 values of 0.96 and 0.85 for the diclofenac and benzylamine model respectively. The important properties for both QSPR studies include Log P, binding energy between ions, molecular weight, surface area and hydrogen bonding ability. The study by Parshad et al.[9] which looked at 22 *p*-substituted benzoate salts of benzylamine also found molecular weight to be a contributing descriptor during the multivariate data analysis as well as other properties such as solubility of free acid and intrinsic dissolution rate. A total of five descriptors produced a model with a reasonable R^2 value of 0.86. It should be noted however that using dissolution rate to predict solubility does not help in the 'real world' as it is very uncommon for the dissolution rate of a compound to be known where the solubility is not. One of the main problems that appear to arise from QSPR studies of salts in comparison to those of organic molecules is the much smaller training sets that are constructed which often contain five times less input data fields. The study by Guerrieri et al.[10] is an example of this where the PLS analysis was performed on 11 salts of procaine. Here the model fails for salts of low solubility, possibly due to not enough input information being available for salts of differing chemical diversity.

Another property that has been looked at using QSPR studies is melting point. Although it is a much simpler property to measure than solubility, the prediction of melting point from structure properties appears to be much more difficult. The study by Hughes et al.[3] finds melting point a very hard thing to predict and attributes this to the lack of crystal packing information in the descriptors and to the complexity of melting points, achieving a R^2 correlation value for the training set of only 0.46. The studies by Modarresi et al.[11] and Deeb et al.[12] were slightly more successful with R^2

factors of 0.67 and 0.70 respectively, though these are still far from the standard of the solubility prediction models. With regards to predicting melting point the most important descriptors are found to be those linked to hydrogen bond strength.

Random forest[13] is a chemometric technique that has been used for a wide variety of regression and classification models in chemistry, process manufacturing and biology.[6,14-18] The program is a library package that is run using statistical computing environment 'R'.[19] Random forest was first developed by Breiman[20] and consists of a collaboration of single decision trees built up by a classification and regression tree algorithm (CART)[21] Some of the advantages of using RF are stated below;

- No need to pre-select descriptors as method allows for automated descriptor selection
- The dataset cannot be over trained
- Provides a measure of the descriptors importance through 'variable importance plots'
- Method has a built in procedure to estimate the model quality and the models ability at prediction
- Builds models that are tolerant to 'noise' in the input descriptors and outliers in the dataset
- Allows for measurement of correlation between input samples in the dataset

These advantages make the use of Random forest an obvious choice for QSPR studies. The studies by Palmer et al.[14] and Svetnik et al.[15] show that Random forest outperforms, or at a minimum gives comparable results to, other analysis techniques such as PLS, DT and ANN. The studies by Palmer et al.[14] and Kovdienko et al.[6] both use Random forest for regression QSPR studies of aqueous solubility of (non-salt) organic compounds. The training sets for both studies were extremely accurate with R^2 values of 0.99 and 0.98 respectively. This was coupled with excellent prediction results on the external datasets with R^2 values of 0.89 and 0.82 respectively. Although literature studies have shown Random forest to be used for many diverse

areas of research there are no published studies which use this method for analysis of aqueous solubility of salts.

This chapter uses a training set of 37 salt compounds to build both regression and classification models for salt solubility. The regression model will be built initially with 2D and 3D parameters that relate to the free base and free acid of the desired salt. After the training set has been constructed, it will then be manipulated in a stepwise manor with the addition of other parameters such as measured physical properties and calculated structural properties until the best model has been constructed. The classification model for salt solubility will contain three arbitrary categories of solubility; sparingly soluble (<0.75mol/L), soluble (0.75-1.75 mol/L) and very soluble (>1.75mol/L) as used in Section 6.2. As with the regression model a step wise approach will be used to build the best training model. After the training sets have been constructed their solubility prediction ability will be tested using salts containing different cations.

7.2   Training set using 2D and 3D parameters

The training set was first constructed using (-)methylephedrinium and (+/-)methylephedrinium salts. The dataset included all anhydrous salts that consisted of a one to one ratio of cation to anion and where the measured solubility is known to be that of the phase described by single crystal diffraction. Care was taken to have only strictly comparable species in this training set. Thus, the hydrated salts, and salts with a two cation to one anion ratio, and salts containing free acid molecules or other unexpected guests were all excluded from the dataset - as their chemical differences may reasonably be expected to give differences in solubility behaviour and therefore may interfere with construction of the model. The racemic salts that spontaneously resolved to form a conglomerate were also excluded. This was partly as the literature suggests that their solubility should be double that of the enatiopure forms[22] and partly as their input parameters would be largely identical to those of the enantiopure

salts. The initial dataset will therefore contain 37 salt compounds as listed in Table 7.1.

Table 7.1 List of 37 salt compounds used in the construction of the training data sets

37 salt compounds used in training sets			
Enantiopure compounds		**Racemic compounds**	
2A	MEpd2AB	**3B**	RMEpd2CB
4A	MEpd2FB	**7B**	RMEpdoTol
5A	MEpd2HB	**9B**	RMEpd3CB
7A	MEpdoTol	**10B**	RMEpd3FB
8A	MEpd3AB	**11B**	RMEpd3HB
10A	MEpd3FB	**12B**	RMEpd3NB
14A	MEpd4AB	**13B**	RMEpdmTol
15A	MEpd4CB	**14B**	RMEpd4AB
17A	MEpd4HB	**16B**	RMEpd4FB
18A	MEpd4NB	**18B**	RMEpd4NB
19A	MEpdpTol	**21C**	RMEpdBS2
22A	MEpd4HBS	**24C**	RMEpdMeSO32
26A	MEpdMale	**26B**	RMEpdMale
27A	MEpdMalon	**27B**	RMEpdMalon
29`A	MEpdSuc	**28B**	RMEpdAdp
32A	MEpdLMD	**33B**	RMEpdRMD
34A	MEpdBr	**34B**	RMEpdBr
35A	MEpdCl	**35B**	RMEpdCl
36A	MEpdBF4		

The 2D and 3D parameters were collated from the structure analysis programs MOE[23] and Pipeline Pilot.[24] The parameters were obtained from the input ChemDraw.mol files of all the free acid molecules and the free base molecule methylephedrine. For input of the dataset the free base parameters were subtracted from the corresponding free acid parameters to normalise the data. The initial attempt at training the methylephedrine dataset used all the 2D and 3D parameter output from MOE along with the physicochemical parameter output from Pipeline Pilot. This provided a data matrix of 37 compounds versus 337 parameters. Obviously in the final training set the number of parameters should not be more than the number of input compounds but for the initial trail run no parameters were excluded.

7.2.1 Initial Random forest regression training model containing all 2D and 3D parameters

The training set initially used 5000 trees and tried 112 variables at each split. For an initial attempt at training the data the results were promising. The model explains 65.9 % of the variance of the data. The plot of actual versus predicted solubility is shown in Figure 7.1. The equation of the line is y = 0.8215x + 0.2417 with errors of 0.0349 and 0.0649 on the gradient and intercept respectively. The variable importance plot, Figure 7.2, shows the parameters that were considered most important when performing the Random forest analysis. Here it is evident that the Log P parameters are highly regarded with the regression analysis.
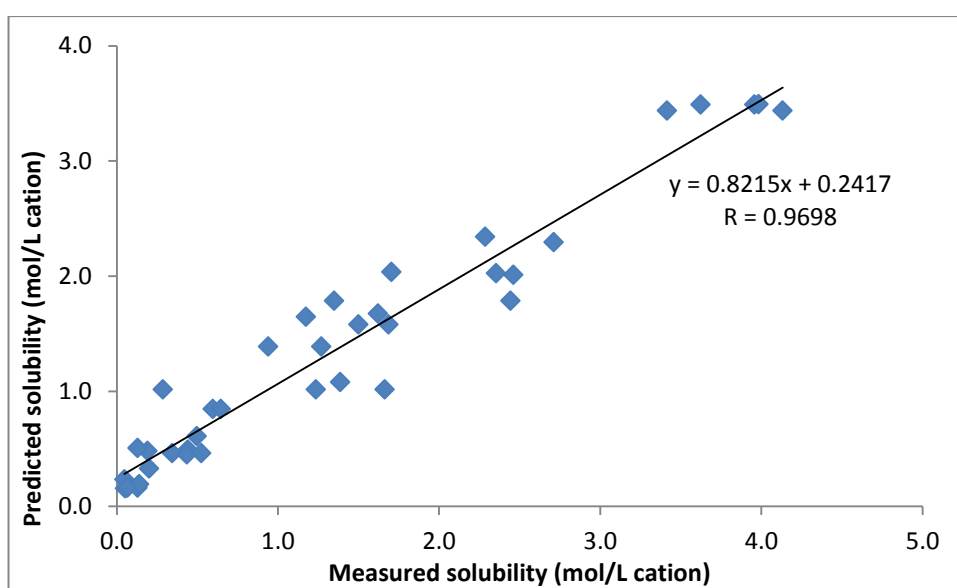
Figure 7.1 Plot of measured versus predicted solubility for regression training model containing all 2D and 3D parameters
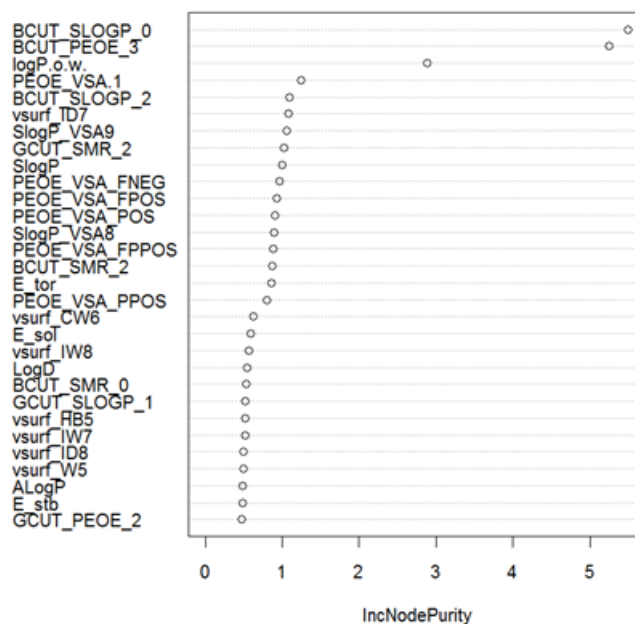
Figure 7.2 Variable importance plot for regression training model containing all 2D and 3D parameters

7.2.2 Initial Random forest classification training model containing all 2D and 3D parameters

Figure 7.3 shows the error achieved with the Random forest analysis of solubility class for the overall system, (black), the sparingly soluble category error (green), the soluble category error (red) and the very soluble category error (blue). It is clearly evident that the only category that has a reasonable error associated with it is that of sparingly soluble compounds. The variable importance plot again shows that the LogP parameters are important as they were in the regression analysis, see Figure 7.4. The proximity matrix (Figure 7.5) shows the clustering of the sparingly soluble compounds (green dots), as expected for such a poor fit there is little clustering of the soluble and very soluble red and blue dots. Finally, the confusion matrix shows how many compounds of each category were correctly predicted, see Table 7.2. This shows that the system correctly predicted fourteen from sixteen of the sparingly soluble compounds, but only five from eleven of the soluble compounds and five from ten of the very soluble compounds. As an initial trial run the results from the sparingly soluble category are reasonable, however it is clear that great improvement is required.

205
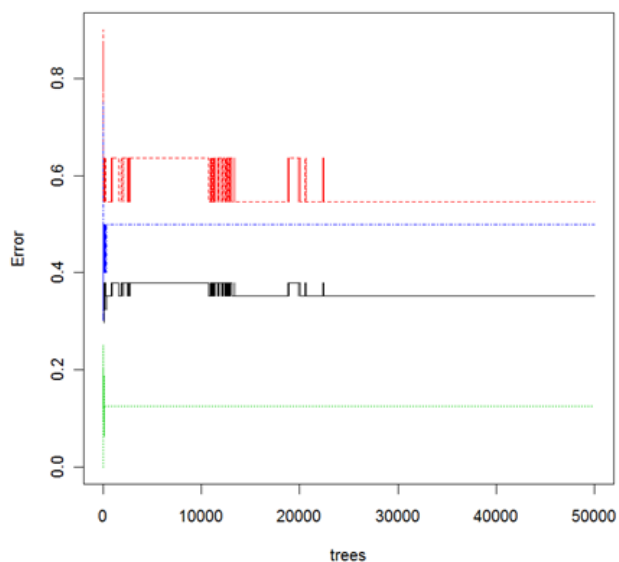
Figure 7.3 Error plot for classification model using all 2D and 3D parameters. **Black** represents the overall error, **green** represents sparingly soluble category error, **red** represents soluble category error and **blue** represents the very soluble category error.

Figure 7.4 Variable importance plot for classification model containing all 2D and 3D parameters

Figure 7.5 Proximity matrix for classification model using all 2D and 3D parameters. **Green** represents sparingly soluble compounds, **red** represents soluble compounds and **blue** represents very soluble compounds

Table 7.2 Observed and predicted solubility results for the classification training model containing all 2D and 3D parameters

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	Class error (%)
Sparingly soluble	14	2	0	12.5
Soluble	3	5	3	54.5
Very soluble	1	4	5	50.0

7.3   Training sets using selected 2D and 3D parameters

As the above preliminary results show some potential for both the regression and classification models, the datasets were then analysed and minimised in order to have less parameters than compounds in the training sets. This was done in two steps; firstly all parameters that did not appear to have any impact on the variable importance plot were excluded. There were many such parameters that were constant for all compounds. Secondly with the remaining parameters a Pearson correlation matrix was constructed and the parameters reduced by the removal of parameters that

correlated with a correlation coefficient of 0.7 or greater. The new dataset contained 37 samples and 23 variables.

7.3.1 Random forest regression training model with specific 2D and 3D parameters

The training set now uses 50000 trees and considers seven variables at each split. This new model is a slight improvement and explains 68.4 % of the variance of the data. The plot of actual versus predicted solubility is shown in Figure 7.6. The variable importance plot (Figure 7.7) again shows that logP (octanol:water) is the most influential of the tested parameters.
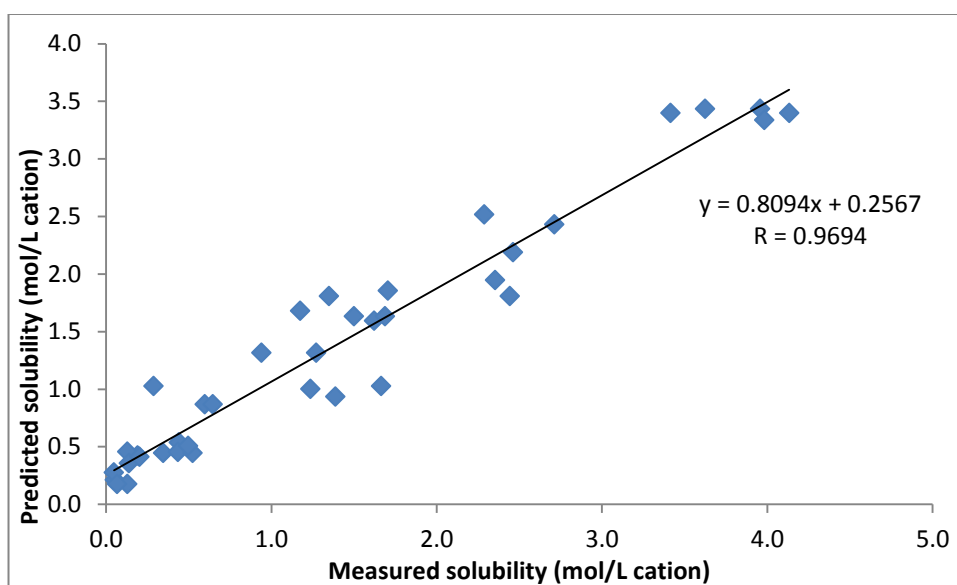
Figure 7.6 Plot of measured versus predicted solubility for regression training model containing specific 2D and 3D parameters
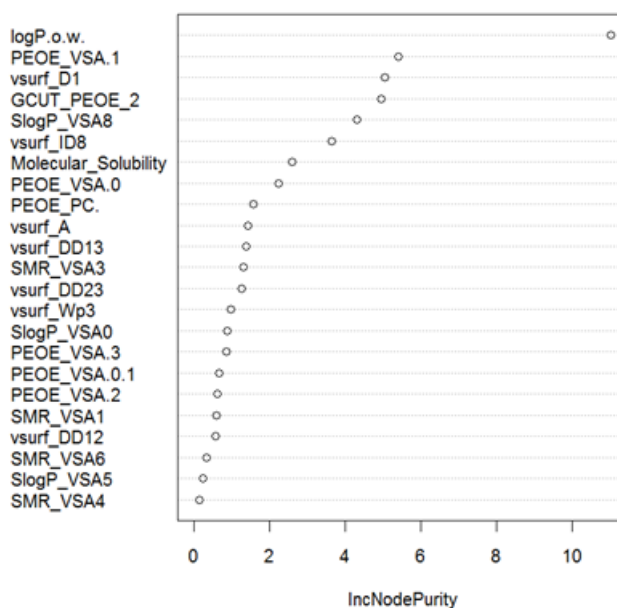
Figure 7.7 Variable importance plot for regression training model containing all 2D and 3D parameters

7.3.2 Random forest classification training model with selected 2D and 3D parameters

For the classification model the Random forest summary plot shows a similar error for the sparingly soluble category but a marked improvement for the soluble and very soluble categories, see Figure 7.8. The variable importance plot, is similar to that of the regression model, Figure 7.7, with the LogP (octananol:water) coefficient being the most important factor. The plot of the proximity matrix, Figure 7.9, shows a much tighter cluster for the sparingly soluble compounds (green dots) and the very soluble compounds are also showing some proximity in the right hand corner of the matrix (blue dots). There is still however much needed improvement in order to separate the three soluble (red dots) compounds from within the sparingly soluble cluster. The confusion matrix, Table 7.3, shows that although more tightly clustered than seen previously there is still a 12.5% error on the sparingly soluble category. There is, as expected, an improvement in the very soluble category with the error reduced from 50% to 30%. The predictive capability for the soluble category is still extremely poor with a prediction error of 45.5%.
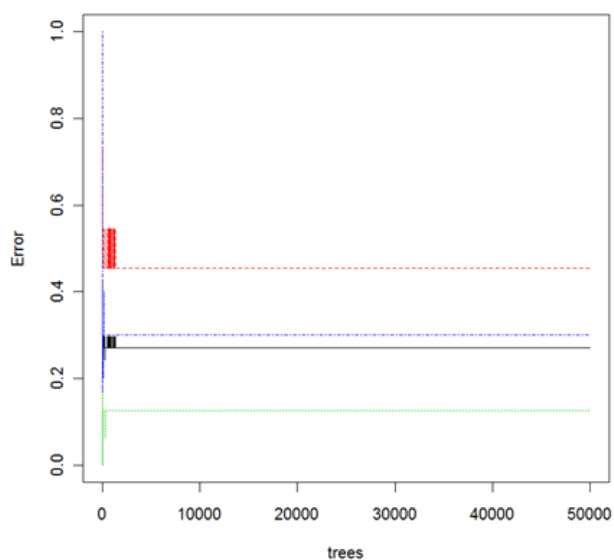
Figure 7.8 Error plot for classification model using all 2D and 3D parameters. **Black** represents the overall error, **green** represents sparingly soluble category error, **red** represents soluble category error and **blue** represents the very soluble category error.
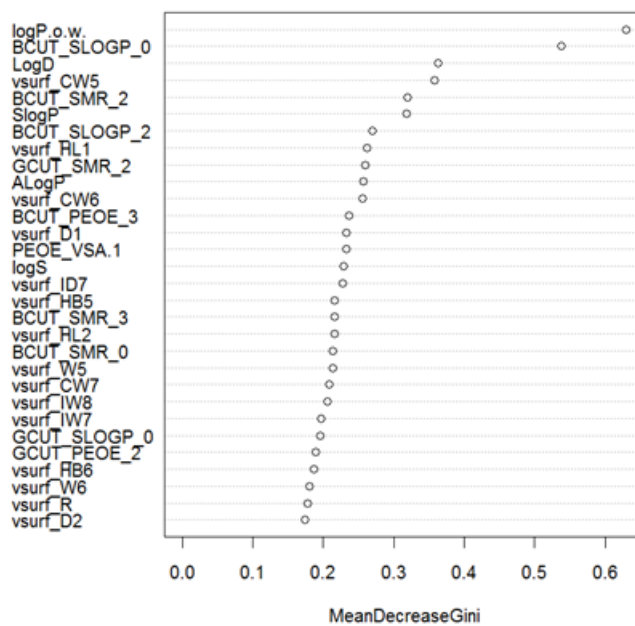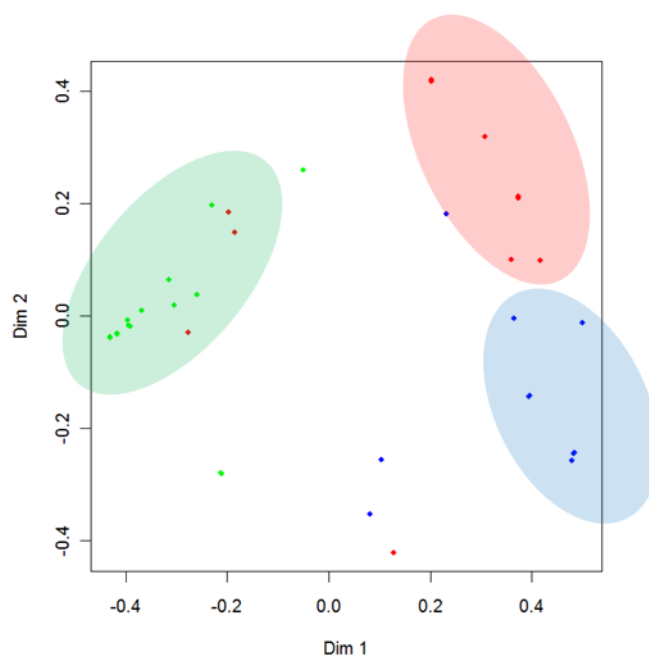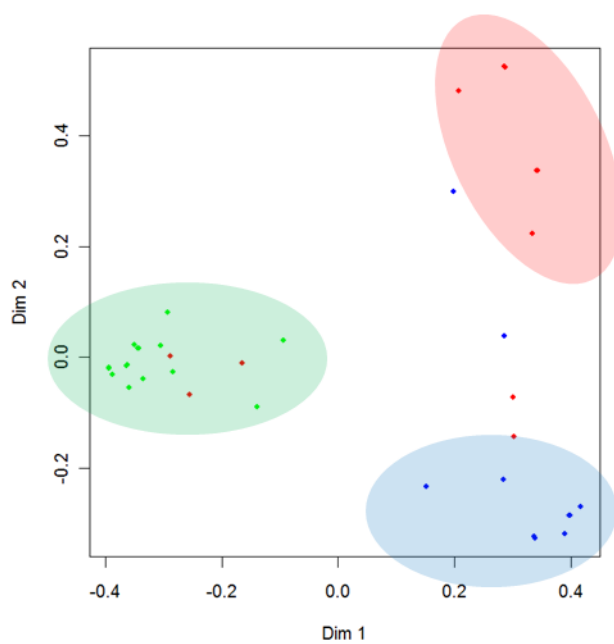

Figure 7.9 Proximity matrix for classification model using specific 2D and 3D parameters. **Green** represents sparingly soluble compounds, **red** represents soluble compounds and **blue** represents very soluble compounds

Table 7.3 Observed and predicted solubility results for the classification training model containing specific 2D and 3D parameters

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	Class error (%)
Sparingly soluble	14	2	0	12.5
Soluble	3	6	2	45.5
Very soluble	1	2	7	30.0

7.4 Training sets using 2D and 3D parameters and measured physical properties

The training sets based only on 23 calculated molecular descriptors from Section 7.3 were manipulated to try and achieve better models. To the correlation reduced 23 parameter variables, two physical property measurements were added namely density and melting point. It was thought that addition of parameters directly linked to the solid-state of the salts would greatly increase the accuracy of the training set. The density values were obtained from the crystallographic data and the melting point values were obtained experimentally (see Section 3.6). A melting point value could not be successfully obtained for (+/-)methylephedrinium methanesulfonate due to the salt being hygroscopic and therefore a zero was entered into the datasheet. These two added parameters did not correlate to any of the existing parameters and therefore the dataset was not further reduced. This dataset was then analysed using both regression and classification analysis to see if these physical properties have a positive effect on the training models.

7.4.1 Random forest regression training model built with 2D, 3D and physical parameters

The training set now uses 50000 trees and considers eight variables at each split. The training model containing the added physical parameters shows an improvement on the statistics. The model now includes 72.6 % of the data variance, with a line equation of y = 0.8240x + 0.2451 and a R value of 0.98, see Figure 7.10. The variable importance plot, Figure 7.11, shows how melting point and to a much lesser

extent density were used to build the model. It is not surprising that melting point has such a profound influence on the construction of the training set as it has been previously shown to correlate to solubility both within this work, see Section 6.4 to 6.9, and in literature studies.[25,26] Note that not having a melting point for the methanesulfonate salt makes little difference as the model can cope with holes in the dataset.
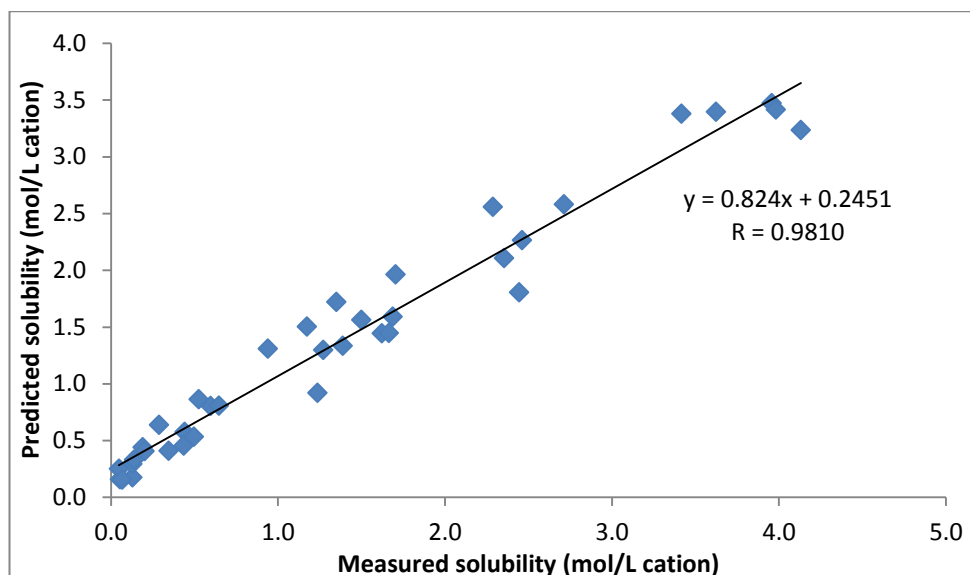
Figure 7.10 Plot of measured versus predicted solubility for regression training model containing specific 2D, 3D and physical parameters
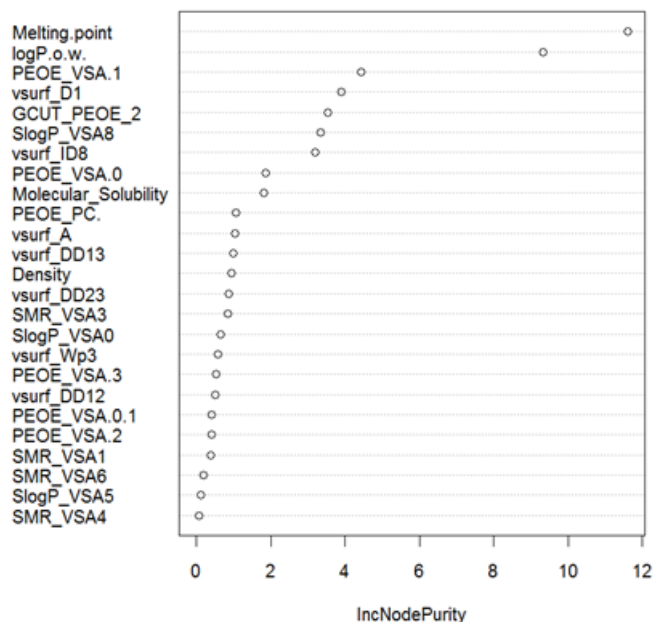
Figure 7.11 Variable importance plot for regression training model containing all 2D, 3D and physical parameters

7.4.2 Random forest classification training model built with 2D, 3D and physical parameters

As with the regression model, the addition of the physical parameters improves the classification model, with melting point having the greater influence. An improved error is seen on both the sparingly soluble and very soluble categories. The soluble category still has a large error present - therefore this model still does not correctly represent salts of this category. The proximity matrix from the Random forest analysis, Figure 7.12, shows a tighter cluster than seen previously for the sparingly soluble compounds (green dots) and for the very soluble compounds (blue dots) with the exception being the two compounds MEpd2AB and MEpdCl, which appear nearer the soluble data points. As expected there is no apparent clustering for the soluble compounds. The confusion matrix, Table 7.4, illustrates the improved predictability of the sparingly soluble and very soluble categories which now have class errors of 6.3 and 20.0 % respectively, and the increased error of the soluble class to 54.5 %. This new model only wrongly predicts one of the solubility values in the sparingly soluble category and two in the very soluble category.
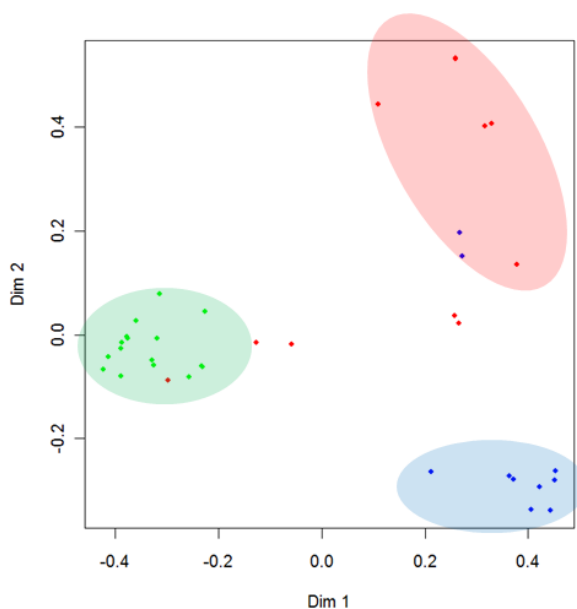
Figure 7.12 Proximity matrix for classification model using specific 2D, 3D and physical parameters. **Green** represents sparingly soluble compounds, **red** represents soluble compounds and **blue** represents very soluble compounds

Table 7.4 Observed and predicted solubility results for the classification training model containing specific 2D, 3D and physical parameters

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	class error (%)
Sparingly soluble	15	1	0	6.3
Soluble	3	5	3	54.5
Very soluble	0	2	8	20.0

7.5 Training sets using 2D and 3D parameters, measured physical properties and AM1 calculations

MOPAC[27] is a semi-empirical molecular orbital package that is used to study molecular structures in the solid-state. Within the package the AM1 Hamiltonian[28] was used to calculate the molecular orbitals heat of formation, electronic and nuclear energy and cosmo area and volume relating to molecular geometry in the gas phase, see Section 3.9.1. These parameters were added to the already constructed training sets containing the 2D, 3D and physical measured parameters. Training models were then built to ascertain if the added calculated parameters improved the regression and classification models. MOPAC AM1 calculations could not be performed for MEpdBF4 as there are no measurement parameters available for boron atoms. This compound was therefore removed from the dataset to leave thirty-six compounds. As previously, the new parameters were firstly added to the dataset followed by correlation analysis to preserve a dataset with the number of parameters being less than the number of samples analysed, (parameters with a correlation of above 0.7 were considered correlated and therefore only one of the two parameters was required and retained).

7.5.1 Regression training model built with 2D, 3D, physical parameters and AM1 input

The addition of the AM1 parameters has a negative effect on the regression model with the new model only explaining 62.75 % of the data. As a result it was determined not to use this data in the building of future training set.

7.5.2   Classification training model built with 2D, 3D, physical parameters and AM1 input

Again the AM1 parameters do not improve the overall error on any of the categories, in fact the error in both the soluble and very soluble categories is seen to increase slightly. This implies that the added parameters have a negative effect on the classification model. As with the regression model these parameters will be excluded from future training sets.

7.6   Training sets using 2D and 3D parameters, measured physical properties and PIXEL calculations

In another attempt to improve the model using calculated parameters, regression and classification models were built with input PIXEL parameters, see Section 1.11 and Section 3.9.2. The PIXEL calculations were performed only on salts within the training set with Z'=1, as for Z'>1 calculations cannot be performed. This reduced the training set to 32 salt compounds. Data for the two chloride salts also could not be calculated due to collision errors during the PIXEL calculations. This left 30 compounds in the training dataset. Several different datasets were constructed to try and ascertain which if any PIXEL parameters are best for modelling the solubility of the salts. The parameters used in the building of the various datasets include (in addition to the 2D and 3D parameters described above); the total coulombic, dispersion and repulsion energies for cation-cation, cation-anion and anion-anion interactions and the total polarisation energy. Also the coulombic, polarisation, dispersion, repulsion and pixel energies for the largest three cation-cation, cation-anion and anion-anion interactions and the overall total coulombic, polarisation, dispersion, repulsion and epixel energies. A final attempt used the energies between the ions of the $C_2^2(9)$, or equivalent, hydrogen bonded graph-set.

7.6.1 Random forest regression training model built with 2D, 3D, physical and PIXEL parameters

The best regression model using the PIXEL parameters contained the total cation-cation, anion-anion and cation-anion coulombic, dispersion and repulsive energies along with the total coulombic, dispersion, repulsion and polarisation energies for the salts. The model represents 57.53 % of the data variance, with a line equation of y = 0.7596x + 0.2920 and a R value of 0.99, see Figure 7.13. The variable importance plot, Figure 7.14, shows how the input PIXEL parameters were prominent in building the model especially the cation-cation dispersion energies. Although this model explains a smaller fraction of the data than previous models, it is interesting that the regression model now has a considerably different relationship between the experimental and predicted solubility values indicated by the line equation. However, although these added PIXEL parameters are influential in the construction of the regression model the decrease in the representative data variance indicates a negative influence.
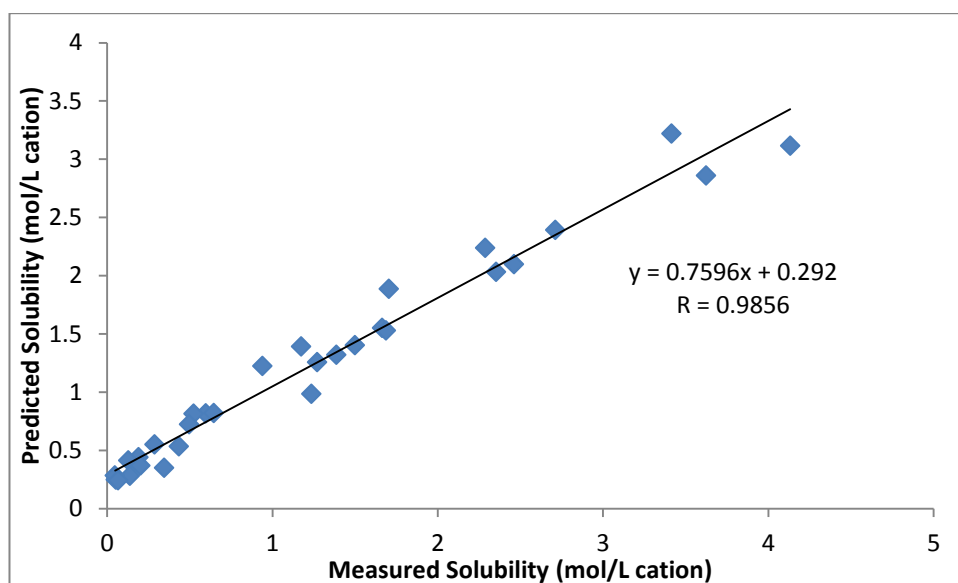
Figure 7.13 Plot of measured versus predicted solubility for regression training model containing specific 2D, 3D, physical and PIXEL parameters

Figure 7.14 Variable importance plot for regression training model containing all 2D, 3D, physical and PIXEL parameters

7.6.2 Random forest classification training model built with 2D, 3D, physical and PIXEL parameters

As with the AM1 parameters, the PIXEL parameters appear to have a negative effect on the classification model. Their presence does not improve the errors on any of the categories. This is best illustrated with the proximity matrix, Figure 7.15, which shows the salt compounds no longer forming neat clusters.
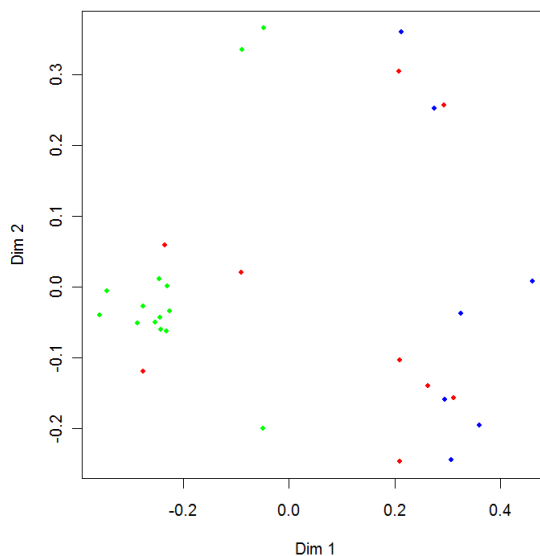
Figure 7.15 Proximity matrix for classification model using specific 2D, 3D, physical and PIXEL parameters. **Green** represents sparingly soluble compounds, **red** represents soluble compounds and **blue** represents very soluble compounds

It is concluded that although PIXEL calculations are good for comparing individual interactions, see Section 4.7, they did not help to build a stronger solubility classification model through Random forest analysis. For the regression model, although PIXEL parameters are prominent in constructing the model this is not in a positive manner as is testified to by the decrease in variance of the data explained by the model. PIXEL calculations are also very time consuming and they are limited as they can only be performed on salts with Z'=1 and two components per asymmetric unit. For the above reasons these parameters will not be further utilised within this study.

7.7 Training set containing 2D, 3D, physical measured and crystal parameters

It was thought that although the calculated AM1 and PIXEL parameters did not have a positive influence on the regression and classification model, other structural information might. To the existing dataset containing 2D, 3D and physical measured

parameters, data extracted from the single crystal structures were added. This included unit cell parameters, molecular volume (calculated from the unit cell volume and the corresponding value of Z per asymmetric unit), definitive torsion angles of the cation (see Section 4.8), number of hydrogen bond donor and acceptor atoms available, dimensionality of the hydrogen bonding network and descriptors of hydrogen bonding from graph-set analysis. Where two or more different angles were available for a given torsion an average of the angles was input into the dataset (for example when there were two cations present per asymmetric unit). As was done previously correlation analysis was performed in order to reduce the number of parameters. Hammett values were also added for the meta and para substituted benzoate salts as they have previously been shown to correlate with solubility (see Section 6.10).

7.7.1 Regression training model built with 2D, 3D, physical and crystal parameters

With the addition of the crystal parameters the accuracy of the correlation between the actual and predicted solubility did not appear to improve, see Figure 7.16, with the training set explaining 68.8 % of the data variance. This is an improved model from the AM1 containing model but not as good as with the exclusion of these parameters. However, some of the added parameters did appear to feature highly in the variable importance plot, see Figure 7.17, and therefore this model was further manipulated to see if a more accurate outcome could be reached. This was done by a stepwise trial and error process that encompassed correlation analysis and exclusion of parameters that had little influence on the model.
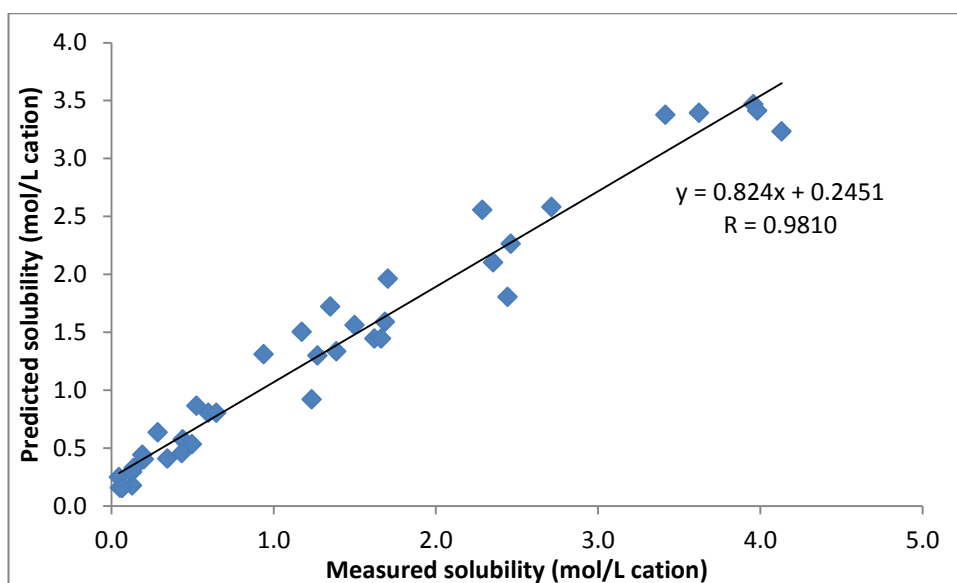
Figure 7.16 Plot of measured versus predicted solubility for regression training model containing specific 2D, 3D, physical and crystal parameters
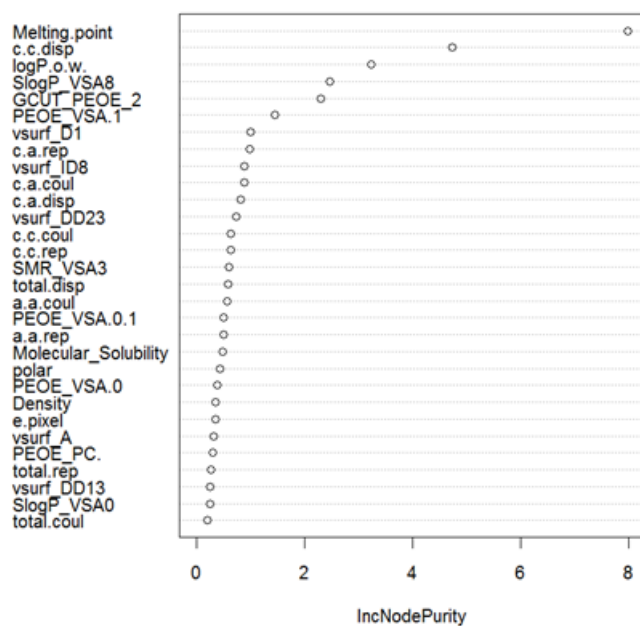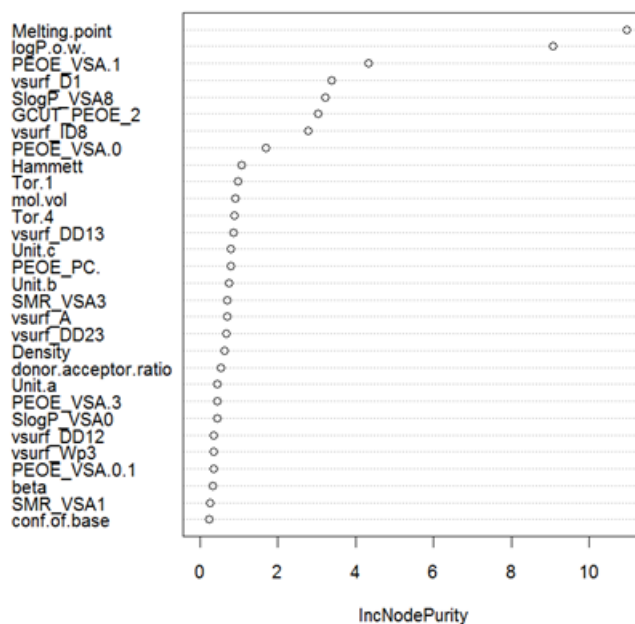

Figure 7.17 Variable importance plot for regression training model containing specific 2D, 3D, physical and crystal parameters

7.7.2 Improved regression training model built with 2D, 3D, physical and crystal parameters

After many stepwise correlation and exclusion analyses a final regression training set was achieved. It was concluded that the added 'crystal' parameters, like the AM1

parameters did not benefit the training model. The best model was seen to contain only four parameters. In order of importance they are melting point, logP (octanol:water), vsurf_D1 and GCUT_PEOE_2, see Figure 7.19. This model explains 74.1 % of the variance of the data. The predicted and actual solubility values are related by y = 0.829 x + 0.235, with errors of 0.033 and 0.062 on the gradient and intercept respectively and a regression coefficient of R = 0.973, see Figure 7.18. The partial dependence of the four contributing parameters can be seen in Figure 7.20. Melting point has been previously shown to have a correlation with solubility[25,26] as has hydrophilicity[9,29] thus the presence of the logP (octanol:water) and the Vsurf_D1 parameter which is a measure of hydrophobic volume make sense in the training set. Indeed log P itself has been found to be an essential parameter in other QSAR studies of solubility, for example see references 4 and 6. The final parameter, GCUT_PEOE_2, is related to the partial equalization of orbital electronegativity associated with the free acid and base and is calculated from a Mulliken type charge,[30] a parameter understandably related to solubility in a polar medium.
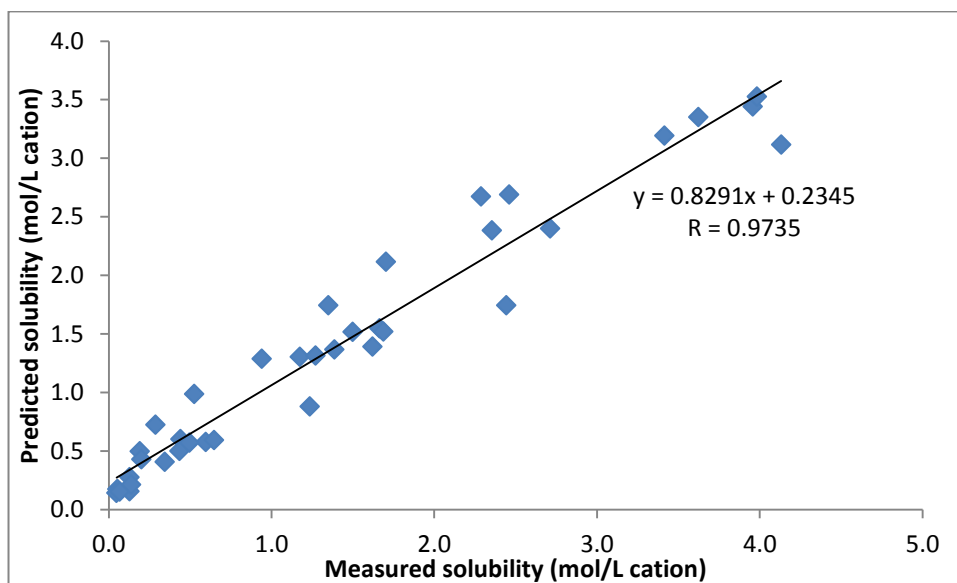
Figure 7.18 Plot of measured versus predicted solubility for finalised regression training model
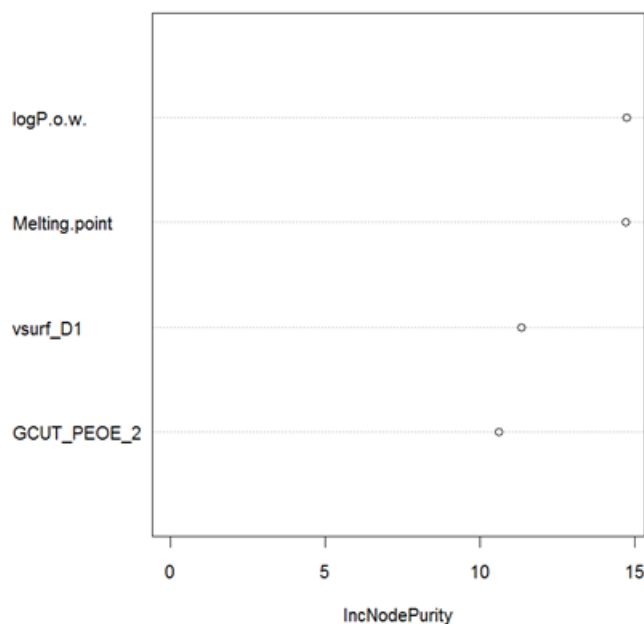
Figure 7.19 Variable importance plot for finalised regression training model
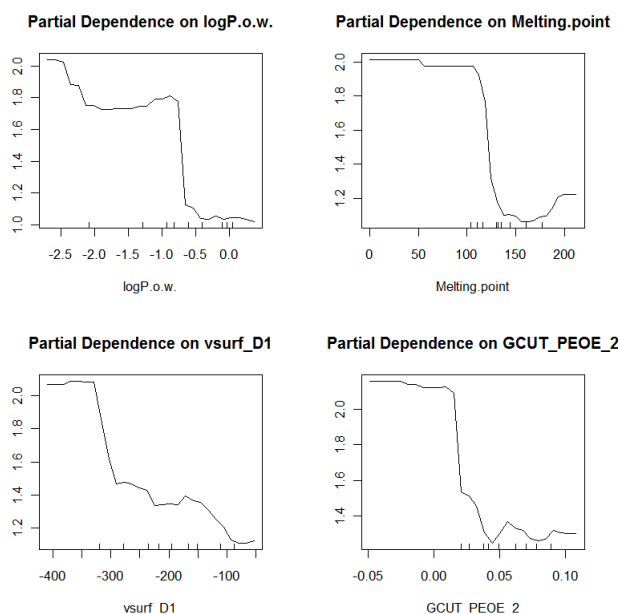
Figure 7.20 Partial dependence plots for four contributing parameters; logP(octanol:water), melting point, vsurf_D1 and GCUT_PEOE_2

In studies by Huuskonen et al.,[31] Hughes et al.[3] and Wang et al.[32] an association has been constructed to predict log solubility of organic compounds from the values of logP and melting point with the equations; logS = -1.01logP - 0.01mpt + 50, logS = 0.5 − 0.01(mpt-25) − logP and logS = 3.513- 0.001mpt − 1.112logP respectively. These three equations vary greatly in their composition and in analysis of this work

herein none of the above variations were able to predict solubility for the salt compounds.

7.7.3 Classification training model built with 2D, 3D, physical, and crystal parameters

With the addition of the crystal parameters to the classification training model there is an improvement in the error associated with the sparingly soluble category, now associated with no error. The variable importance plot, Figure 7.21, contains some of the added crystal structural parameters, notably the cation torsion angles and the Hammett constant, indicating their influence on the classification model. In the proximity matrix, Figure 7.22, the analysis shows that the sparingly soluble compounds (green dots) still cluster, however the two bromide compounds appear now to be separated from the main group. This may be understandable as the two bromide compounds are chemically different from the other sparingly soluble compounds which are all benzoate derived salts. The clustering from the very soluble (blue dots) compounds appears to be very similar to before and as expected the problematic soluble (red dots) compounds don't appear to cluster. The confusion matrix, Table 7.5, shows how the class error in the sparingly soluble category has fallen to zero. Manipulation of this model will be done by a stepwise trial and error process similar to that used for the regression model above in order to achieve the best classification model.

Figure 7.21 Variable importance plot for classification training model containing specific 2D, 3D, physical and crystal parameters
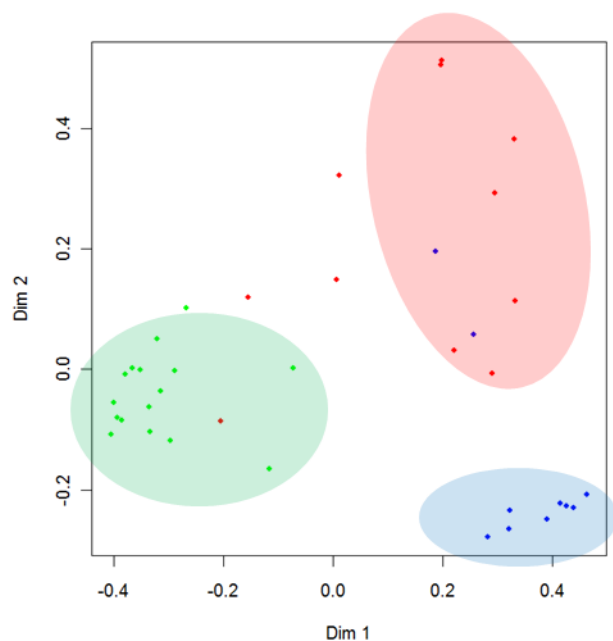
Figure 7.22 Proximity matrix for classification model using specific 2D, 3D, physical and crystal parameters. **Green** represents sparingly soluble compounds, **red** represents soluble compounds and **blue** represents very soluble compounds
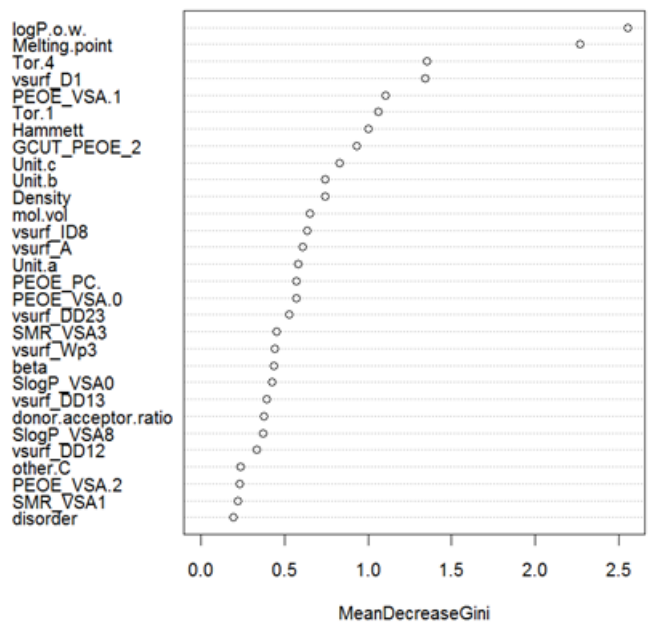
224

Table 7.5 Observed and predicted solubility results for the classification training model containing specific 2D, 3D, physical and crystal parameters

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	class error (%)
Sparingly soluble	16	0	0	0.0
Soluble	4	4	3	63.6
Very soluble	2	0	8	20.0

7.7.4 Improved classification training model built with 2D, 3D, physical and crystal parameters

After many stepwise correlation and exclusion analyses a final classification training set was achieved. This differs from the regression training set as there are crystal parameters as well as measured and calculated parameters present. Figure 7.23 shows the variable importance plot with the three calculated values (logP (octanol:water), vsurf_D1 and Hammett), two crystallographically obtained values (molecular volume and cell volume) and one measured value namely melting point. Of the parameters not also seen in the regression study, the Hammett values have been shown previously to correlate to solubility earlier in this work, see Section 6.10. The presence of the molecular volume and cell volume suggest that the molecular size and compactness of the salt has a contributing factor to the solubility. Molecular volume has previously been highlighted in QSPR predictions of aqueous solubility as an important factor.[4,32]

The proximity matrix, Figure 7.24, shows the close clustering of the sparing soluble compounds and also some clustering of the very soluble and to a lesser extent the soluble compounds. This classification model has an overall error of 18.92 % with zero error of the sparingly soluble category, 20 % error on the very soluble category and 45.5 % error on the intermediate soluble category, see Table 7.6. From this latter category, the three benzoate compounds (MEpd2FB, MEpdoTol and RMEpdmTol) are all wrongly classified as sparingly soluble, however note that the two toluate salts are located away from the main cluster with only the fluorobenzoate salt present

within the cluster. A similar classification model was attempted with only two solubility categories, but the results were no more successful.

Figure 7.23 Variable importance plot for finalised classification training

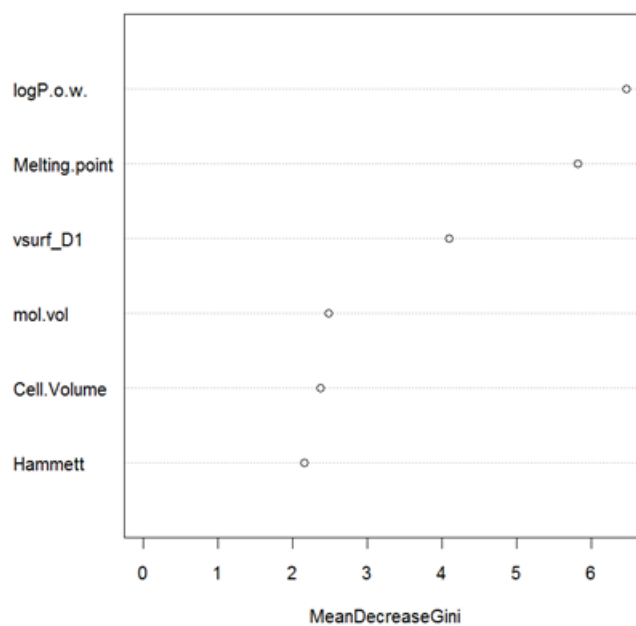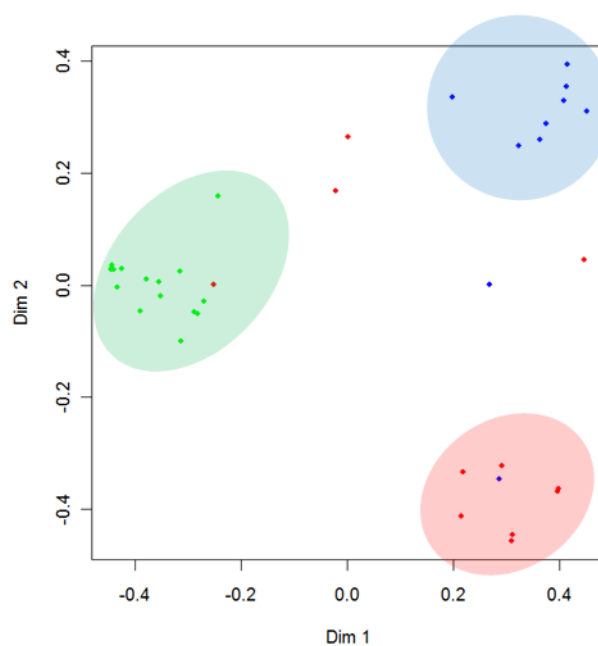**Figure 7.24** Proximity matrix for finalised classification model. **Green** represents sparingly soluble compounds, **red** represents soluble compounds and **blue** represents very soluble compounds

Table 7.6 Observed and predicted solubility results for the finalised classification training model

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	class error (%)
Sparingly soluble	16	0	0	0.00
Soluble	3	6	2	45.5
Very soluble	0	2	8	20.0

7.8 Testing the predictive power of the training sets

Both the regression and classification training sets were tested for their ability to predict solubility for salts omitted from the training set on the grounds of chemical or physical difference. Only salts where the x-ray powder pattern obtained from the solids isolated from the slurries used for solubility measurements matched the pattern predicted from the single crystal data were used. This was to ensure that chemical and phase identity was unambiguously assigned. The same limits for the categories of solubility were used as for the training datasets above.

7.8.1 Predicting solubility for (+/-)methylephedrinium conglomerates

This prediction dataset contains the three methylephedrinium conglomerate forming salts that were excluded from the initial training set, namely the 4-chloro, hydroxy and methyl benzoate salts. Figure 7.25 shows the actual versus predicted solubility for the three salts where the predicted values are based on the regression training set detailed in Section 7.7.2. It can be seen that the solubility of the three conglomerate salts can be accurately predicted using this model. The predictive accuracy of the classification model, Section 7.7.4, is also tabulated below, Table 7.7. Here you can see that all three salts have the same category for their predicted and observed solubility, namely sparingly soluble, giving no error in the predictive capability of the training set. These results show that the conglomerate forming salts do not behave differently than the enantiopure and racemic methylephedrinium salts and implies that the Meyerhoffer double solubility rule[22] has no validity here.
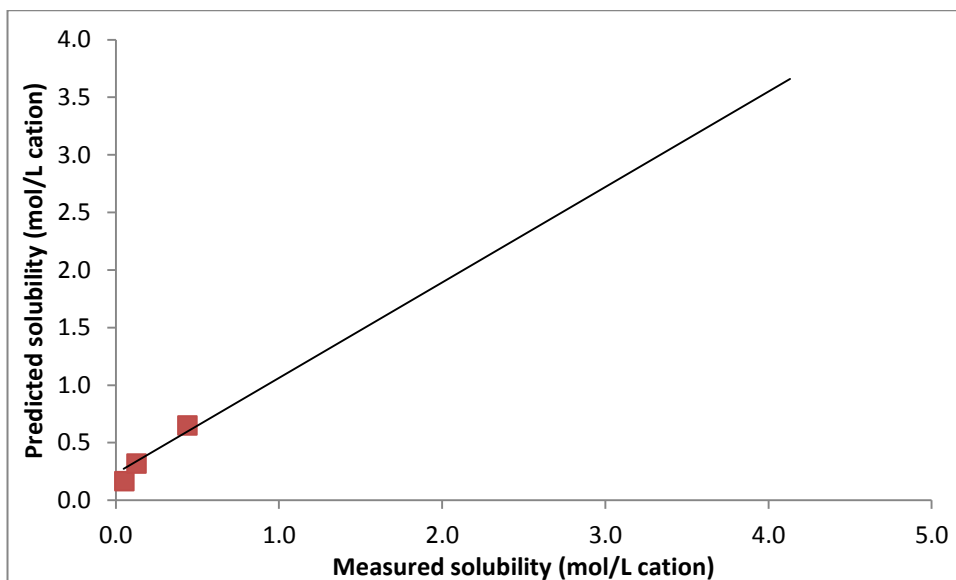
Figure 7.25 Plot of measured versus predicted solubility for the prediction test set of methylephedrinium conglomerate salts. Line obtained from training set, **red** represents methylephedrinium salts

Table 7.7 Observed and predicted solubility results for the prediction test set of methylephedrinium conglomerate salts

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	Prediction error (%)
Sparingly soluble	3	0	0	0.0

7.8.2 Predicting solubility of phenethylammonium salts

There are fifteen anhydrous phenethylammonium containing 1:1 salts that were analysed using the two training datasets. Unlike the conglomerate methylephedrinium salts, there appears to be no predictability of phenethylammonium salt solubility using either the regression or classification training models, see Figure 7.26 and Table 7.8.
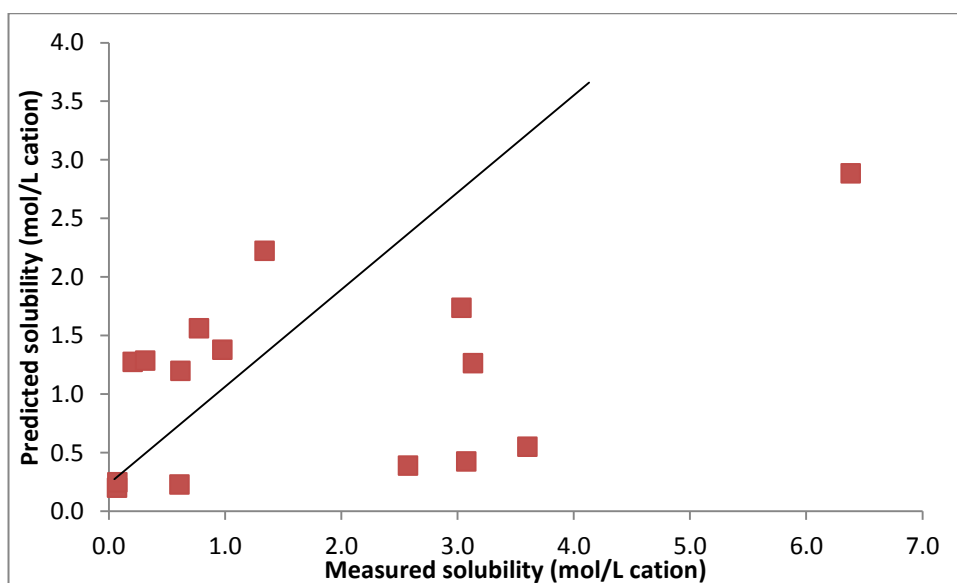
Figure 7.26 Plot of measured versus predicted solubility for the prediction test set of phenethylammonium salts. Line obtained from training set, **red** represents phenethylammonium salts

Table 7.8 Observed and predicted solubility results for the prediction test set of phenethylammonium salts

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	Prediction error (%)
Sparingly soluble	4	2	0	33.3
Soluble	1	2	0	33.3
Very soluble	5	1	0	100.0

Similar prediction studies were also performed for salts of other different cations, namely those of tyramine, phenylpropylamine, methylphenethylamine, α(methylaminomethyl)benzyl alcohol, (-)ephedrine, (-)pseudoephedrine and dimethylphenethylamine. As with the phenethylammonium salts, no accurate predictions were observed. For further details of these studies please refer to Appendix C.

7.8.3 Predicting solubility of benzoate derived salts

Predicting solubility of general non-methylephedrine based salts from the methylephedrine training set failed, see above. However, close examination of the

data showed that for the cations derived from α (methylaminomethyl) benzyl alcohol, (-)ephedrine, (-)pseudoephedrine, dimethylphenethylamine and methylephedrine there was some success in the prediction of solubility of the benzoate derived salts. Figure 7.27 shows the plot of measured versus predicted solubility for these salts. It is based on the regression analysis and has a correlating R value of 0.9282. In light of this success the classification model was used to attempt correct categorical prediction of this same set of benzoate salts. Table 7.9 shows the results. It is apparent that the classification model is not as accurate as the regression model in predicting solubility of benzoate derived salts.
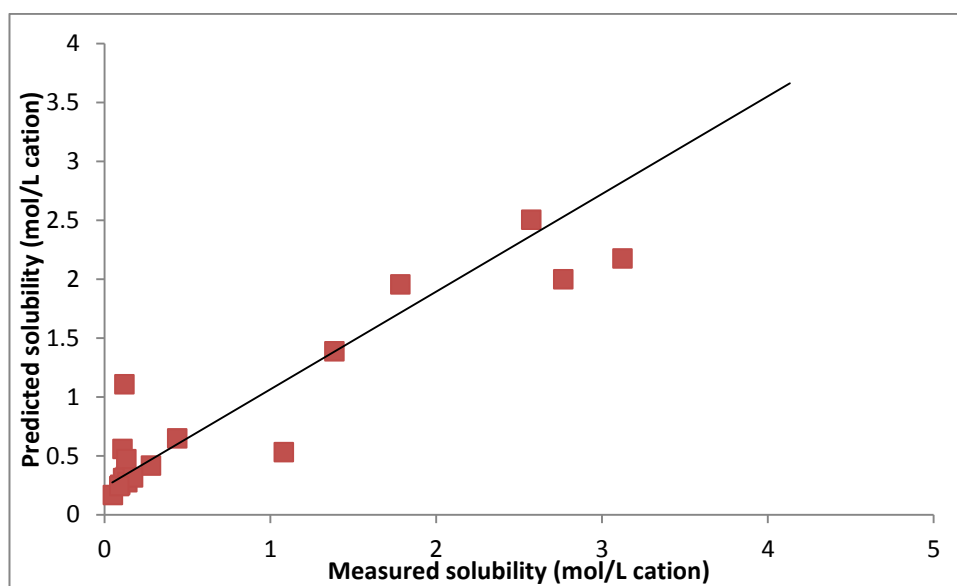
Figure 7.27 Plot of measured versus predicted solubility for the prediction test set of α (methylaminomethyl) benzyl alcohol, ephedrinium, pseudoephedrinium, dimethylphenethylaminium and methylephedrinium benzoate derived salts. Line obtained from training set, **red** represents benzoate derived salts

Table 7.9 Observed and predicted solubility results for the prediction test set of α (methylaminomethyl) benzyl alcohol, ephedrinium, pseudoephedrinium, dimethylphenethylaminium and methylephedrinium benzoate derived salts

Predicted → Observed ↓	Sparingly soluble	Soluble	Very soluble	Prediction error (%)
Sparingly soluble	11	2	0	15.4
Soluble	2	1	0	66.6
Very soluble	2	3	3	62.5

7.8.4 Predicting solubility using 2D and 3D parameters of the free acid only

The above training models were generated using the difference between the MOE-generated values of 2D and 3D parameters of the appropriate base and acid. As most individual comparisons were for salts of a given cation, the models were also calculated using input parameters of the free acid only. This had no effect on the training models, but the prediction results obtained using these training models with the salts of the non-methylephidrinium cations varied somewhat. However, in global terms the regression model was still unable to predict solubility of any salts consistently, other than those of the conglomerate methylephedrine. Interestingly, analysis of the group of benzoate derived salts where some predictability was previously observed did not correctly predicted solubility when using the uncorrected free acid only model. Meanwhile, the free acid only based classification model seemed to produce different predictions with greater errors appearing on the sparingly soluble category and slightly improved error on the soluble and very soluble category. Overall the use of 2D and 3D parameters based solely on free acid values was not a success. For reference to these results please see Appendix D.

7.9   Attempted solubility prediction for other methylephedrinium salts

The regression training set was able to be used successfully to predict the solubility of the methylephedrinium salts that had formed a conglomerate. With this in mind the training set was used to try and predict the solubility of other methylephedrinium salts that had previously been excluded from the training set. These salts included those that formed a salt with two cations and one anion per asymmetric unit, salts that contained neutral molecules (cocrystals), salts that formed hydrated species and salts that were excluded as a phase difference was observed between the single crystal experiment and powder diffraction data obtained during the solubility measurements. This set encompasses 18 methylephedrinium salts in total.

7.9.1 Attempted prediction of hydrated salts using regression model

There are seven methylephedrinium compounds that form a hydrated thermodynamic species, namely 1A, 6A, 9A, 20C, 22B, 30A and 32C, for which phase specific solubility measurements were obtained. All are monohydrates. Figure 7.28 shows the solubility prediction results for these compounds. The salt of (+/-)methylephedrinium 1,2-ethanedisulfonate monohydrate seems to be an outlier and does not correlate with any of the other hydrate structures. Chemically this salt is different from the other hydrates as it forms a salt with a two to one cation to anion ratio. The training set is only composed of salts with a one to one cation to anion ratio. Interestingly the hydrate salts of one to one salts (shown in red) appear to correlate to each other, they also have a simple relationship to the prediction line composed from the training set. From the equations of the two lines it can be seen that the gradient is the same and the difference in the lines is 0.794 on the intercept. This implies that the solubility of the hydrated salts can be predicted with the addition of a correction factor of -0.794. Figure 7.29 shows the hydrate data with the applied correction factor. All 1:1 salts now lie on the prediction line from the training set.

Figure 7.28 Plot of measured versus predicted solubility for the prediction test set of methylephedrinium hydrated salts. Line obtained from training set, **red** represents methylephedrinium hydrated salts, **green** represents (+/-)methylephedrinium 1,2-ethanedisulfonate monohydrate

Figure 7.29 Plot of measured versus predicted solubility for the prediction test set of methylephedrinium hydrated salts with associated correction. Line obtained from training set, **red** represents methylephedrinium hydrated salts

7.9.2   Attempted prediction of two to one and cocrystal salts using regression model

There are three anhydrous salts that have phase specific solubility data that do not form one to one cation to anion salts. (-)Methylephedrinium 1,2-ethanedisulfonate forms a two to one salt, (+/-)methylephedrinium iodide iodine contains a neutral I_2 species and (+/-)methylephedrinium tartrate tartaric acid contains two cations, one doubly deprotonated anion and a free acid molecule per asymmetric unit. Figure 7.30 shows the prediction results for these three salts. The ethanedisulfonate and the iodide salt appear to be correctly predicted, whereas the tartrate salt has a predicted solubility considerably less than the measured value. As there are only three salts to compare it is not possible to draw any conclusions from this data.
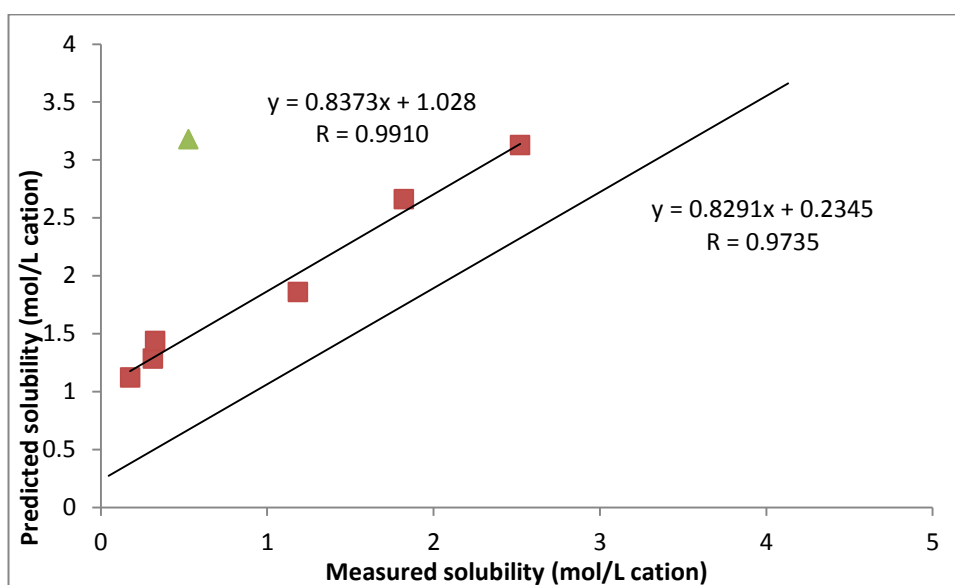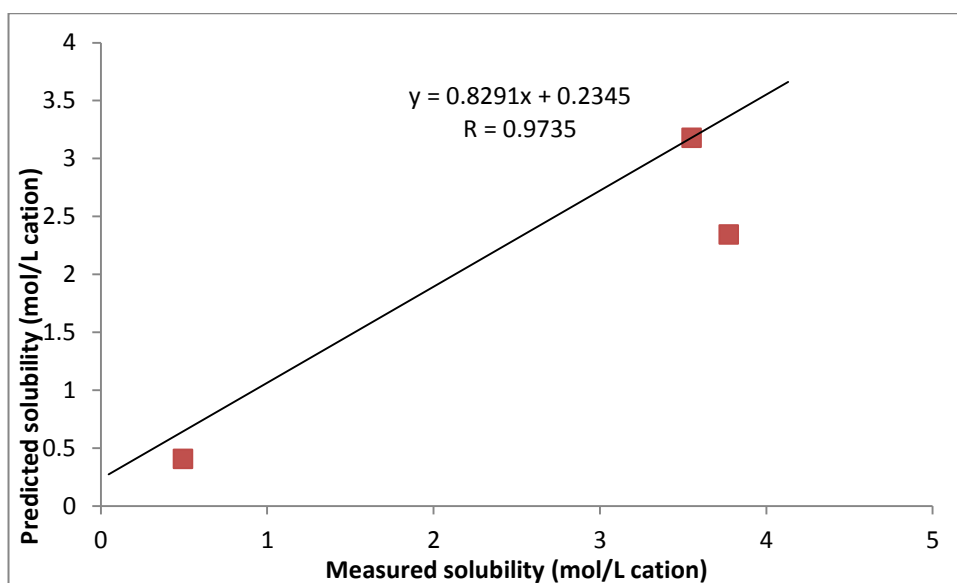
Figure 7.30 Plot of measured versus predicted solubility for the prediction test set of methylephedrinium two to one and cocrystal salt. Line obtained from training set, **red** represents methylephedrinium two to one and cocrystal salt

7.9.3 Attempted prediction of salt solubility with unknown crystal phase using regression model

There are eight methylephedrine salts for which solubility measurements have been acquired but which were excluded from the original training set as phase specific crystal data was not available. By using the regression model the solubility of these eight methylephedrine salts was predicted. The prediction plot below, Figure 7.31, with the two lines shown for typical anhydrate and hydrate behaviour implies that two of the salts have an anhydrous thermodynamic phase, four have a hydrated thermodynamic phase and two do not fit with either of these proposals. The two structures that do not fit either the anhydrous or the hydrate prediction line are both sulfonate structures. Although the single crystal information already known is that of the hydrogen-sulfate monohydrate it is possible that the thermodynamic phase may be a two to one cation to anion hydrated structure. This would explain the non-fit with the hydrate prediction line as it has been shown previously that this model cannot predict hydrated two to one salts. Figure 7.32 shows the four proposed hydrated salts with a 0.794 correction factor. The four salts now fit the line of the anhydrous prediction confirming the theory of their hydrated status.
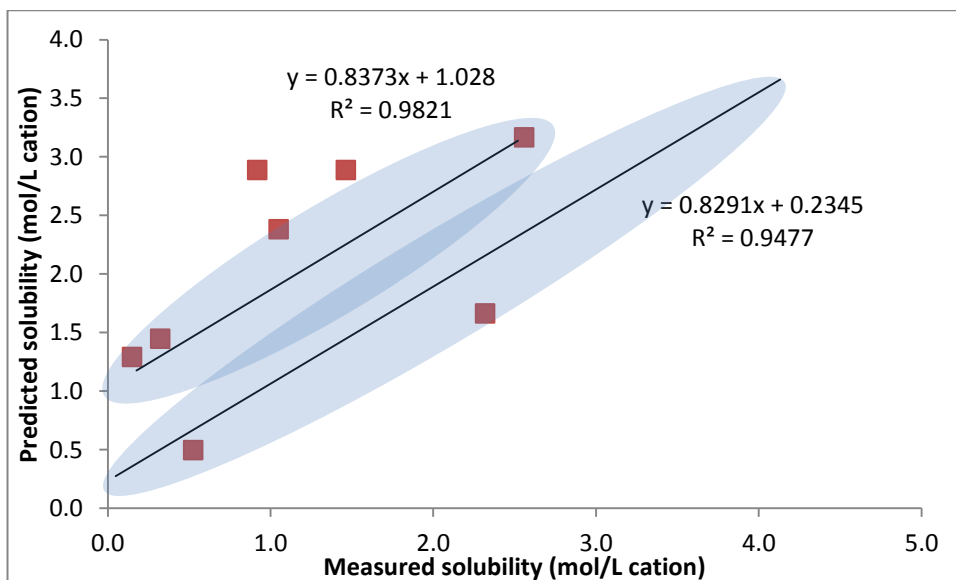
Figure 7.31 Plot of measured versus predicted solubility for the prediction test set of methylephedrinium salts of unknown phase information. Line obtained from training set, **red** represents methylephedrinium salts
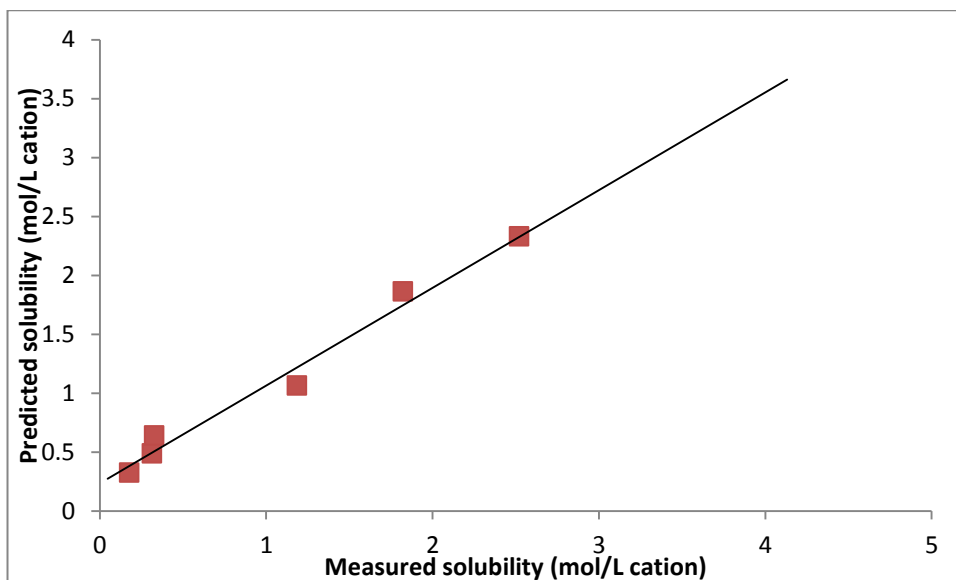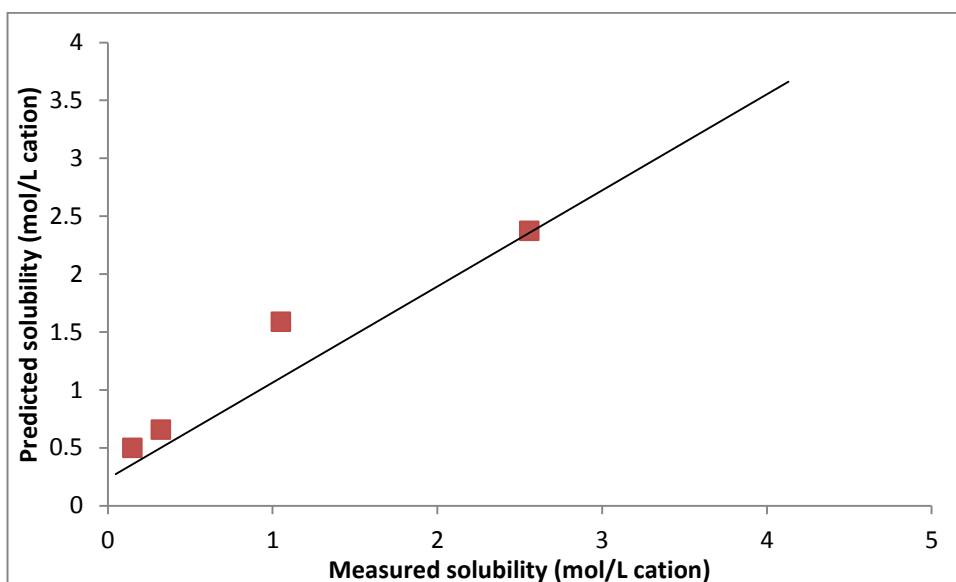

Figure 7.32 Plot of measured versus predicted solubility for the prediction test set of methylephedrinium salts of unknown phase with associated correction. Line obtained from training set, **red** represents methylephedrinium salts

7.10   Summary of prediction ability of methylephedrinium salts

The methylephedrine regression model has been shown to correctly predict the aqueous solubility of anhydrous and conglomerate methylephedrinium salts. It has also been shown to have the ability to predict the solubility of hydrated

methylephedrinium salts with the addition of a correction factor. This study only encompassed monohydrated salts and therefore it is unknown whether this model also holds for higher hydrates. The model cannot be used to predict the solubility of hydrated salts that are composed of two cations and one anion per asymmetric unit. Further studies need to be performed to confirm whether solubilities of two to one anhydrous salts and cocrystal salts can be successfully predicted using this model or a derivative of it.

7.11   References

(1)   Monissette, S. L.; Almarsson, O.; Peterson, M. L.; Remenar, J. F.; Read, M. J.; Lemmo, A. V.; Ellis, S.; Cima, M. J.; Gardner, C. R. *Advanced Drug Delivery Reviews* **2004**, *56*, 275.

(2)   Serajuddin, A. T. M. *Advanced Drug Delivery Reviews* **2007**, *59*, 603.

(3)   Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. *Journal of Chemical Information and Modeling* **2008**, *48*, 220.

(4)   Ghasemi, J.; Saaidpour, S. *Chemical & Pharmaceutical Bulletin* **2007**, *55*, 669.

(5)   Gozalbes, R.; Pineda-Lucena, A. *Biorg. Med. Chem.* **2010**, *18*, 7078.

(6)   Kovdienko, N. A.; Polishchuk, P. G.; Muratov, E. N.; Artemenko, A. G.; Kuz'min, V. E.; Gorb, L.; Hill, F.; Leszczynski, J. *Molecular Informatics* **2010**, *29*, 394.

(7)   Tantishaiyakul, V. *Int. J. Pharm.* **2004**, *275*, 133.

(8)   Tantishaiyakul, V. *J. Pharm. Biomed. Anal.* **2005**, *37*, 411.

(9)   Parshad, H.; Frydenvang, K.; Liljefors, T.; Larsen, C. S. *Int. J. Pharm.* **2002**, *237*, 193.

(10)   Guerrieri, P.; Rumondor, A. C. F.; Li, T.; Taylor, L. S. *AAPS PharmSciTech* **2010**, *11*, 1212.

(11)   Modarresi, H.; Dearden, J. C.; Modarress, H. *Journal of Chemical Information and Modeling* **2006**, *46*, 930.

(12)   Deeb, O.; Goodarzi, M.; Alfalah, S. *Mol. Phys.* **2011**, *109*, 507.

(13)   Liaw, A.; Wiener, M. *R News* **2002**, *2*, 18.

(14)   Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O. *Journal of Chemical Information and Modeling* **2007**, *47*, 150.

(15)   Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 1947.

(16)   Debeljak, Z.; Skrbo, A.; Jasprica, I.; Mornar, A.; Plecko, V.; Banjanac, M.; Medic-Saric, M. *Journal of Chemical Information and Modeling* **2007**, *47*, 918.

(17)   Zhang, Q.-Y.; Aires-de-Sousa, J. *Journal of Chemical Information and Modeling* **2007**, *47*, 1.

(18) Johnston, A.; Johnston, B. F.; Kennedy, A. R.; Florence, A. J. *Crystengcomm* **2008**, *10*, 23.

(19) R Development Core Team; 2.11.1 ed.; R Foundation for Statistical Computing: Vienna, Austria, 2006.

(20) Breiman, L. *Machine Learning* **2001**, *45*, 5.

(21) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth International Group Belmont, CA, 1984.

(22) Meyerhoffer, W. *Ber. Dtsch. Chem. Ges.* **1904**, *37*, 2604.

(23) MOE Chemical Computing Group; http://www.chemcomp.com: Quebec, Canada, 2002.

(24) Pipline Pilot In *Chemistry Collection: Basic Chemistry User Guide*; Accelrys Software Inc: San Diego, 2010.

(25) Thomas, E.; Rubino, J. *Int. J. Pharm.* **1996**, *130*, 179.

(26) Gould, P. L. *Int. J. Pharm.* **1986**, *33*, 201.

(27) Stewart, J. J. P.; Stewart Computational Chemisty: Colarado Springs, CO, USA, HTTP://OpenMOPAC.net (2008).

(28) Kozma, D.; Bocskei, Z.; Simon, K.; Fogassy, E. *Journal of the Chemical Society-Perkin Transactions 2* **1994**, 1883.

(29) Agharkar, S.; Lindenbaum, S.; Higuchi, T. *J. Pharm. Sci.* **1976**, *65*, 747.

(30) Mulliken, R. S. *J. Chem. Phys.* **1934**, *2*, 782.

(31) Huuskonen, J.; Rantanen, J.; Livingstone, D. *European Journal of Medicinal Chemistry* **2000**, *35*, 1081.

(32) Wang, J. M.; Hou, T. J.; Xu, X. J. *Journal of Chemical Information and Modeling* **2009**, *49*, 571.

Chapter 8
Conclusions

8.1 Conclusions

A series of crystallisation experiments gave a library of systematically related salt structures constructed using 11 bases from the phenethylamine family and 42 acid derived counterions. The 462 possible acid-base combinations produced 235 independent salt crystal structures; these were added to 20 salt structures already present within the literature[1-11] to give a library containing 255 crystal structures. The systematic variation in the chosen materials allowed investigation into small changes of substituent on inter-molecular bonding and physicochemical properties. These crystal structures along with the measured physicochemical properties of density, melting point, hydration state and aqueous solubility were investigated for trends and correlations in structural features and how these relate to variation in physicochemical properties.

In depth structural analysis was conducted on 64 crystal structures of methylephedrine which included 28 pairs of enantiopure and racemic methylephedrinium salts. All the structures analysed formed contact cation-anion pairs, with no direct cation-cation hydrogen bonded contact ion pairs. Anion-anion interactions were found to be present in all structures with an anion that possessed a classic hydrogen bond donor, except for the two hydrogen-sulfate monohydrate salts. The commonest graph-set motif of $C_2^2(9)$ was found to be present in 38 of the 64 independent structures, and was flexible enough to include both OCO and OSO functionalities. The equivalent $C_2^1(7)$ motif was present in all seven halide structures. No systematic difference was found between the hydrogen bonding of racemic amines and that of enantiopure amines. This is contrary to literature suggestions.[12] The methylephedrinium cation was seen to adopt one of three different conformations and there were six common cation-cation pair packing motifs. Analysis with the 'crystal packing similarity' module of Mercury CSD 2.3[13] showed that 21 salt structures adopted seven different 3D isostructural groupings with respect to cation packing. Hydrogen bonding differences (and indeed differing hydration states and chemical phases) were seen within some of the otherwise isostructural cation arrays. PIXEL energy calculations were utilised to establish the strongest

cation-cation and cation-anion interactions present within the structures. The common $C_2^2(9)$ chain was found to contain the strongest individual attractive interaction in 80.8 % of cases, but this was only 27.5 % of the total energy, showing that non-hydrogen bonding interactions are cumulatively more important, a point highlighted by the 19.2 % where the strongest individual interaction is between $C_2^2(9)$ chains. Within the $C_2^2(9)$ chains the N¨O distance was found to vary with energy, but the O¨O distance was invariant. PIXEL calculations also suggested that racemic i interactions were stronger than other cation-cation interactions.

Of the 28 pairs, 16 produced enantiopure/racemic crystal phases of identical chemical make-up. Note that this means that in 43 % of cases the racemic and enantiopure bases behave chemically differently, despite using an otherwise achiral environment. For these 16 structural pairs densities and where possible melting points were collated and compared, principally in order to test the validity of Wallach's rules.[14] Although general overall trends agreed with Wallach's rules individual assessment produced many cases where this statute was found to be false, with eight pairs where the racemic form fails to be denser than the enantiopure form. It is thus concluded that there is no experimental evidence for Wallach's rule. One of the six observed cation packing motifs (the racemic π motif) is closely associated with failure of Wallach's rule. Also associated with failure of Wallach's rule is a significant change in hydrogen bonding dimensionality where racemic forms have 1D hydrogen bonded networks but enantiopure forms have higher dimensional 2D networks. Thus we have highlighted two different details of structural array that may explain inefficient packing.

Hydration state analysis of the database found water to be present in 57 out of the 255 crystal structures. A decrease in hydrate formation was observed with regards to the type of amine present of the order primary > secondary > tertiary. Both the above statements are in general agreement with the findings of literature based surveys.[15] Specifically, a high occurrence of hydrate formation was found for (-

)pseudoephedrinium salts and a low occurrence found for phenethylammonium and methylphenethylaminium salts. Hydrate formation is seen to greatly increase with the presence of sulfonate, 'spherical' and dicarboxylate counterions and greatly decrease with the presence of benzoate and mandelate counterions. Hydrate formation appears to be non-existent for the halide salt species. These results support the theory that formation of a hydrate is linked to an increased presence of polar groups within the salt.[16]

The most common water environment present within the hydrated structures involved the water molecule providing two donor and one acceptor hydrogen bond interactions. A donor/acceptor ratio of one was most frequent for both the anhydrous and hydrated salts however, the results show that hydrate formation is most probable when there is an excess of hydrogen bond acceptor atoms present and least likely when there is an excess of hydrogen bond donor atoms present. This indicates that the primary role of water is to provide extra hydrogen bond donors a theory that is in agreement with the water's most common environment acting as a two hydrogen bond donor and one hydrogen bond acceptor molecule. This is in agreement with the literature study by Desiraju[17] but contrary to that by Infantes et al.[18]

With the exclusion of the co-crystal hydrate salts the 54 hydrated species used hydrogen bonding interactions to produce two main packing styles, namely layered and paired. In all cases the water molecule is associated with the anion either being present in the hydrophilic or anion layer or encompassed next to the anion in the paired structures. The most common packing style observed is that of cation-anion layers which was present in 26 of the 54 structures. Upon investigating trends in density, melting point and solubility for the anhydrous and hydrated salts, the hydrated salts were found to be on average 2.75 % denser than the anhydrous salts. They were also observed to have an average melting point 48 % lower than observed for the anhydrous species. The solubility comparison for the anhydrous and hydrated salts produced unpredictable outcomes with no steady trends observed.

Analyses of the measured physicochemical properties found linear correlations between salt melting point and log aqueous solubility and between melting point of salt and melting point of free acid. Both only held for anhydrous salts. Attempts to find correlations between a range of other parameters were unsuccessful. The enantiopure and racemic methylephedrinium benzoate derived salts melting point and solubility data correlated to produce the same negative linear relationship. When looking at isostructural groups of methylephedrinium salts the accuracy of this correlation greatly improves showing that similarity in structures relates to similar solubility behaviour. This is evidence that solid-state structure does materially affect solubility. Correlation was also found to be greatest between chemically "similar" groupings. As two extreme examples of this;

> Moving away from benzoates, a positive linear correlation was seen between melting point of salt and log solubility for the enantiopure and racemic methylephedrinium halide salts. This correlation goes against the theory that higher melting point relates to more stable compounds and therefore lower solubility.

> The amphoteric nature of the aminobenzoic acids results in solubility correlations being seen when aminobenzoate salts of a range of bases are grouped instead of the correlations for a range of salts of the same cation, as seen for all other benzoate salts.

Thus it is concluded that correlations are possible, but that these are only observed when chemically similar groupings are used. This may explain why some relatively large surveys did not report correlation – they were using too diffuse a group of species.[19]

The solubility of the salt is affected by whether the benzoate counterion has an electron donating or electron withdrawing group on the *meta-* or *para-* position. There was a linear negative correlation between Hammett value and log solubility for

meta- substituted benzoate salts and a parabolic correlation for *para-* substituted benzoate salts.

After collation of the structural and measured physicochemical properties for the database of salt structures, the chemometric program Random forest was used to build both regression and classification training models using 37 well characterised anhydrous methylephedrinium salts. This appears to be the largest scale QSPR study on salt solubility and to be the first such study to attempt to add solid-state-specific descriptors to the dataset. Following several correlation and variable importance treatments the regression model explained 74.1 % of the data variance and was observed to rely on four parameters, namely melting point, logP (octanol:water), vsurf_D1 and GCUT_PEOE_2. A linear correlation between measured and predicted solubility had a correlation 'R' value of 0.9735. Note that this model cuts across the chemical similarity barriers found for simple correlations of say melting point with solubility. For the simple correlations, different (indeed inverted) linear relationships were found for different chemical categories of salts. The Random forest model improves on this by including a wide variety of anion types. The final classification training model had an overall error of 18.9 % with zero error of the sparingly soluble category, 20 % error on the very soluble category and 45.5 % error on the intermediate soluble category. This model was built using three calculated parameters (logP (octanol:water), vsurf_D1 and Hammett), two crystallographically obtained parameters (molecular volume and cell volume) and one measured parameter namely melting point. Both the regression and the classification models include information on solid-state structure (i.e. melting point, cell volume) and information which is likely to be more relevant to the solution phase (logP (i.e. octanol:water), vsurf_D1). This reflects the essential nature of saturation solubility as an equilibrium between solid and solution phases.

Finally, prediction of aqueous solubility using the methylephedrinium salt regression training model was successful for a series of benzoate salts derived from the cations

of α (methylaminomethyl)benzyl alcohol, (-)ephedrine, (-)pseudoephedrine and dimethylphenethylamine. Success was also achieved with the correct prediction of aqueous solubility of "unknown" anhydrous and conglomerate methylephedrinium salts. It has also been shown that the model has the ability to predict the solubility of monohydrated methylephedrinium salts with the addition of a simple correction factor. Although it is widely known that hydrated solids are generically less soluble than anhydrates in aqueous medium,[20] we know of no previous suggestion that there is a simple relationship between the solubility values of the two types.

8.2 References

(1)   Barlow, R. B.; Johnson, O.; Howard, J. A. K.; Walton, D. C.; Koellner, G. *Acta. Cryst.* **1989**, *B45*, 396.
(2)   Bergin, R. *Acta. Cryst.* **1971**, *B27*, 381.
(3)   Collier, E. A.; Davey, R. J.; Black, S. N.; Roberts, R. J. *Acta. Cryst.* **2006**, *A62*, 498.
(4)   Duddu, S. P.; Grant, D. J. W. *Pharm. Res.* **1994**, *11*, 1549.
(5)   Haynes, D. A.; Pietersen, L. K. *Crystengcomm* **2008**, *10*, 518.
(6)   Hearn, R. A.; Bugg, C. E. *Acta. Cryst.* **1972**, *B28*, 3662.
(7)   Hearn, R. A.; Freeman, G. R.; Bugg, C. E. *J. Am. Chem. Soc.* **1973**, *95*, 7150.
(8)   Horn, E.; Tiekink, E. R. T.; Jones, G. P.; Naiola, B. P.; Paleg, L. G. *Acta. Cryst.* **1990**, *C46*, 1575.
(9)   Koleva, B. B.; Kolev, T.; Seidel, R. W.; Spiteller, M.; Mayer-Figge, H.; Sheldrick, W. S. *J. Mol. Struct.* **2008**, *888*, 138.
(10)   Mathew, M.; Palenik, G. J. *Acta. Cryst.* **1977**, *B33*, 1016.
(11)   Rademeyer, M. *Acta. Cryst.* **2007**, *E63*, O221.
(12)   Kinbara, K.; Hashimoto, Y.; Sukegawa, M.; Nohira, H.; Saigo, K. *J. Am. Chem. Soc.* **1996**, *118*, 3441.
(13)   Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41*, 466.
(14)   Wallach, O. *Liebigs Ann Chem* **1895**, *286*, 90.
(15)   Haynes, D. A.; Jones, W.; Motherwell, W. D. S. *Crystengcomm* **2005**, *7*, 342.
(16)   Infantes, L.; Chisholm, J.; Motherwell, S. *Crystengcomm* **2003**, *5*, 480.
(17)   Desiraju, G. R. *Journal of the Chemical Society-Chemical Communications* **1991**, 426.
(18)   Infantes, L.; Fabian, L.; Motherwell, W. D. S. *Crystengcomm* **2007**, *9*, 65.
(19)   Black, S. N.; Collier, E. A.; Davey, R. J.; Roberts, R. J. *J. Pharm. Sci.* **2007**, *96*, 1053.
(20)   Grant, D. J. W.; Higuchi, T. *Solubility Behaviour of Organic Compounds*; Wiley: New York, 1990.