Chapter 1
Chemical Insights from Systematic Structural Studies. The 'Stamp Collecting' Approach to Understanding the Solid State

Simon J. Coles

Abstract Many industries, most notably pharmaceutical, have for some time been using compound libraries as a systematic approach to comprehensively understanding a chemical landscape. Screening in medicinal chemistry is an obvious example and Quantitative Structure Activity Relationship (QSAR) takes a similar, in silico, approach. Understanding packing and structure-determining factors in the solid state is key in many areas, e.g. polymorphism or crystallisation, and of course we need to understand this behaviour if we are to control solid-state formation in any way. It is therefore rather surprising that there are relatively few systematic studies being conducted on the solid state in the way we now routinely work in these other areas. This paper presents numerous systematic studies of homologous series of compounds that have been studied as 'libraries' in the solid state—sometimes these families of related compounds can have as many as 200–300 crystal structures. Results that enable us to derive rules and begin to predict solid-state behaviour will be presented. Taking this concept further, some series of ultra-high resolution structural families are presented—from these studies a comparison of electron density distributions leads to very detailed correlations between bonding and reactivity. Finally, in order not only to rationalise the large amounts of data generated, but also to begin to analyse for prediction purposes, a concept for a statistical approach to describe and build models of crystal structures is outlined.

Keywords Structural systematics · X-ray crystallography · Charge density determination · Structural similarity · Crystal engineering

S.J. Coles (✉)
Department of Chemistry, University of Southampton, Southampton SO17 1BJ, UK
e-mail: s.j.coles@soton.ac.uk

P. Ramasami et al. (eds.), *Crystallizing Ideas – The Role of Chemistry*,
DOI 10.1007/978-3-319-31759-5_1

1.1 Background

2014 was the UNESCO International Year of Crystallography and it is at this time that it is worth reflecting on how far the subject has come and the directions in which it is heading. In just over a century it has evolved from the first diffraction experiments and structure determinations, through periods of increasing complexity, i.e. macromolecules, and increasing volume. Over the last 50 years a significant collection of small molecule structures has been amassed as we have gone from a time when a PhD student would have determined a handful of structures during the course of study to one where this number can be achieved in a day. This has a profound impact on the way we can now conduct the science and the new insights it can provide. We are at a turning point, where the structural systematics approach is able to make a significant impact.

Classical systematics studies tend to use crystal structures as a definitive source of information to probe molecular geometry and, in some cases, link this to physical properties. A prime example of this approach is a series of eight papers originating from our laboratory spread over about a decade up until 2006. Under the theme of "Structural Investigations of Phosphorus–Nitrogen Compounds", very accurate analysis of bond lengths and angles across a series (typically between 5 and 8 structures) of very closely related homologues was correlated with properties such as basicity and electron density and hence linked to reactivity [1].

However, during this time and increasingly of late, a much greater importance has been put on designing, analysing and understanding the whole crystal structure-that is the assembly of molecules, and relationship between them, in the solid state. In fact, in the UK the Directed Assembly [2] initiative has been identified as one of the most significant challenges that researchers should address in the coming decades: "The Directed Assembly Grand Challenge Network has been sponsored by EPSRC [3] to promote research into how assembly processes occur at the molecular and supramolecular levels, and how they can be controlled to develop materials with particular properties and function".

Crystal engineering is predicated on the concept that a solid-state system can be created where the structure, and often associated properties, are known a priori and intentionally designed into the system [4]. This can be achieved through utilising intermolecular interactions either from self-recognition where a molecular entity associates with an identical second unit or between two different entities that possess complementary interacting moieties. Crystal engineering, therefore, relies on being able to control the directing effects of molecular structure on crystal packing and has given rise to the notion of supramolecular synthons [5]. This approach was born out of the community concerned with engineering small organic compounds, however, this has very successfully spread to supramolecular chemistry [6], metal organic framework design [7] and the formation of coordination polymers [8].

In the search for new pharmaceuticals, companies attempt to engineer drug molecules that have specific interactions with active sites in biological entities. To

assist in this process it is quite common to build libraries of drug-like compounds [9] that can be used in screening experiments to find a candidate. Whilst in the early days of developing this approach these compounds were physically synthesised, and indeed some still are, as targets become more diverse and complex this rapidly becomes infeasible. Therefore, to move the approach forward, these studies are now generally performed in silico—not only does this avoid expensive synthesis, but also means that large and diverse virtual libraries can be built to probe significant areas of chemical space [10].

This approach can be very readily applied to crystallographic investigations and hence structural systematics. The Cambridge Structural Database [11] (CSD) is a library of crystal structures containing around 700,000 records and has been the basis for discovering trends and developing crystal engineering rules [12] in many studies, e.g. a recent paper by Taylor entitled "Which intermolecular interactions have a significant influence on crystal packing?" [13] to mention just one from tens or even hundreds. However, the records in the CSD are harvested from the published literature, the content of which is governed by characterisation of newly synthesised compounds. As such the structural ground which is covered can be sparsely populated and generally it is not of scientific interest to go to the extent of full crystallographic characterisation of every member of a homologous series—invariably only a representative study is performed.

Structural systematics can be defined as the comparison of sets of chemically related crystal structures with the aim to establish and describe relevant similarities and relationships. This approach is used to increase understanding of the assembly of organic molecules into crystal structures. Such investigations are invariably carried out on polymorphs, solvates, salts and molecular complexes, in which a particular molecule can occur in different crystal structure environments, but also with families of compounds, whose molecular structures are very closely related, through small but systematic modifications to a parent molecule [14]. The work described herein presents results arising from studies on homologous series of compounds, particularly organic drug-like systems—this begins with normal resolution studies and then progresses on to provide an insight into the potential of employing very high resolution, i.e. electronic distribution, approaches in structural systematics.

1.2 Similarity and Relationships in Molecular Systems

Systematic studies across a series of crystal structures are a powerful methodology for linking changes in structure to behaviour and function and was perhaps first realised as a viable ab initio approach in the form of crystal engineering about two decades ago [15]. Crystal engineering is the process of developing intentional properties and structural motifs in a crystal structure [4] and much work has been done engaging the supramolecular synthon and Aufbau principles put forward by Desiraju et al. [5, 16].

Clearly, it is important to be able to develop understanding and rules as to the behaviour of these synthons when in particular chemical environments and 'structural systematics' is our approach to addressing this problem.

Understanding relationships between members in families of molecular crystal structures requires a careful consideration of three notions—those of polymorphism, isostructurality and similarity.

1.2.1 Polymorphism

McCrone's 1965 definition of a polymorph [17] as "a solid crystalline phase of a given compound resulting from the possibility of at least two crystalline arrangements of the molecules of that compound in the solid state" leads one to conclude that polymorphism is a phenomenon where a chemical compound can occur in at least two crystalline forms. Polymorphs in the solid state can arise not only from the interplay between different intermolecular interactions and lattice packing, but also subtle differences in molecular conformation. Polymorphs exhibit different physical properties, such as melting point and solubility, yet the differences in total energy holding the lattice together are generally minute.

1.2.2 Isostructurality

The quest for definitions of isostructurality has largely been driven by considerations of how different two structures can be and yet still be considered isostructural. Fabian and Kalman [18] proposed the following definition: isostructurality refers to the similarity of the spatial arrangements of the molecules of different compounds in their crystals. It is traditionally interpreted in three dimensions (i.e. isostructurality involves whole structures), which are infinite in three dimensions. However, it is possible to extend the interpretation of the phenomenon to one- and two-dimensional (1D and 2D, respectively) isostructurality. If two crystal structures contain similar infinite 2D molecular arrangements (layers or sheets) then they are termed two-dimensionally isostructural. Accordingly, structures with similar rows of molecules (columns, tapes or threads) are one-dimensionally isostructural.

1.2.3 Similarity

The general concept of similarity is largely overlooked as it is not obviously quantifiable and does not have a clear-cut definition. It is, however, very important and actually highly quantifiable in the solid state. For many years crystallographers have been identifying motifs in their structures and comparing them to other

structures already known—not just molecular substructure, but also intermolecular interactions. The XPac program [19], developed in the Southampton laboratory, allows comparison of different crystal structures through creation of sets of vectors between the component atoms of an arbitrary "seed" molecule within each crystal structure to be compared and the equivalent atoms of the neighbouring molecules of this "seed" within the crystal structures. This has the advantage that it is in principle possible to compare the arrangement of molecules in crystal structures where those molecules are really quite different from each other.

The angles and torsions generated by equivalent vectors in a pair of structures can then be compared with a match being generated when the differences are close to 0°. The number of matches of bond angles and torsion angles generated from these vector sets determines the level of similarity between the structures, so that if all match, they are considered 3D isostructural and if none match there is no similarity. Values between these two extremes indicate matching planes (2D similarity), tapes (1D similarity) or discrete assemblies [e.g. dimers (0D similarity)] between structures. The advantage of this approach is its flexibility, allowing comparison of multicomponent systems, $Z' > 1$ structures and families of related compounds as well as polymorphs. The XPac program can compute a quantitative assessment of each of these types of similarities [20] by means of the angle and torsion differences (Δa and Δp).

Whilst these topics have been recognised for quite some time, much work has been performed over the last 15 years or so into the solid-state aspects of these phenomena. Most work has been performed on identifying crystalline polymorphism and this has generally been inspired by the application to the interests of the pharmaceutical industry. In summary polymorphs relate to different crystal structures of the same molecular compound, whilst isostructurality is where different compounds exhibit essentially the same crystal structure arrangement. The notion of similarity enables comparison of anything occurring between these definitions and can be applied to collections of structures, be they polymorphs or homologous series. These topics are very much complicated when one begins to consider phase transitions, disorder, co-crystals, solvates and hydrates to name but a few, but it is not the purpose of this article to go into detail on these matters. Research over the last decade in the group at Southampton has concentrated on developing the XPac approach and illustrating its utility in defining similarity in large collections of solid-state structures. The following section summarises some of the highlights of this work in order to provide the reader with a context for the more detailed studies that follow.

The concept of supramolecular constructs (or "seeds"), as defined in the XPac software [20], can be used to assess the degree of similarity between members within large structural libraries. This was first demonstrated in 2006 [21] when comparing 25 related crystal structures based on the carbamazepine (CBZ) molecule. Two fragments of closely packed CBZ molecules, a stack and dimer, are identified as the dominating motifs in 24 of the 25 structures. The results of this work highlight the effect that molecular shape plays in the assembly of molecules in the solid state, even when hydrogen bonds are present.

In another exceptionally large study [22], over one hundred 4,4′-disubstituted benzenesulfonamidobenzenes were synthesised and their crystal structures determined. 74 % of the structures exhibit one of two motifs (dimer or chain) based on N–H···O=S interactions. The most common type is a series of 22 isostructures containing the simple dimer motif. A hierarchy for the classification of the 56 distinct structure types of this set was presented. Continuing on from this work, fourteen 4,5′-substituted benzenesulfonamido-2-pyridines, with tautomeric forms R^1–C_6H_4–SO_2–N=C_5NH_4–R^2 or R^1–C_6H_4–SO_2–NH–C_5NH_3–R^2, and with R^1=CF_3, I, Br, Cl, Me, F, H and R^2=CF_3 or I were compared [20]. All structures display a common 3D arrangement of N–H···N bonded centrosymmetric dimers. This isostructural series is exceptional in its completeness and in the diversity of the substituents involved. XPac plots of individual dissimilarity parameters illustrate geometrical similarities and differences. This study showed that the ability of two compounds to crystallise in fundamentally the same crystal structure depends on how much their molecules differ in shape and also on the flexibility of the crystal packing arrangement concerned.

Several more similar studies on different systems have been conducted, e.g. co-crystals of some 3, 5 and 6 monosubstituted salicylic acids with 4-aminopyridine [23], however, it is not the purpose of this article to comprehensively review the field, but merely provide an introduction and context.

1.2.4 Theory or Folklore—Testing Halogen Bonding and π···π Stacking

Halogen bonding, [24–26] and the role of fluorine–fluorine and hydrogen–fluorine interactions in supramolecular synthons, [27] is a topic of much current interest. This is a topic that can readily be probed via a systematic approach. The hypothesis was that F···F interactions are unfavourable and H···F interactions will be preferred in the arrangement of molecules with respect to each other in the crystal lattice. Recent work in our laboratory has involved the synthesis and structural characterisation of a range of systematically, fluorine-substituted benzylideneanilines. This system has been chosen for a number of reasons:

(a) a range of compounds with differing substitution patterns can readily be synthesised by a simple addition reaction between two components,
(b) they are relatively uncomplicated molecules, which limits the number of competing factors for crystal engineering,
(c) they are relatively planar, which increases the likelihood of interactions due to the probable stacking nature of the packing in the crystal lattice,
(d) substitution patterns can be deliberately chosen to test the hypothesis.

Therefore, by observing these design principles it is possible, provided the molecules stack on top of each other, that systems with ideal complementary,

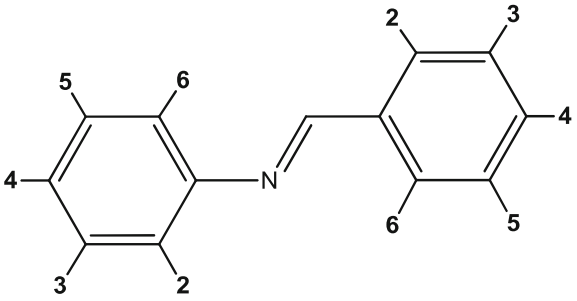

Fig. 1.1 The substitution pattern, and hence notation, of the benzylideneaniline system

perfect clashing and a hybrid of the two, will be produced, i.e. δ^+ regions will want to overlap with δ^- and δ^- regions will not want to overlap with each other. A schematic is presented in Fig. 1.1, which illustrates the numbering scheme—from this a nomenclature for the substitution patterns and combinations can be derived.

The scheme denotes first the number and position of fluoride substituents on the aniline ring and then on the benzyl ring, for example (E)-4,5-difluoro-N-(4-fluorobenzylidene)aniline becomes (4,5-4). When considering all possibilities it is possible to generate a theoretical 'matrix' where each ring is successively substituted from zero to five fluorides. The diagonals of this table represent structures that are capable of stacking in a perfectly complementary, or perfectly clashing, orientation and any structures that are off-diagonal will be 'frustrated', in that there will be a mixture of complementary and clashing groups. There are 20 possible substitution patterns for each ring, which in principle gives 400 compounds which can be synthesised and used to test the overall hypothesis and also several other questions within that scope.

From Fig. 1.2 it can be seen that the hypothesis is upheld, in that the cases of 0-1,2,3,4,5 and 3,4,5-2,6 which should overlap in a complementary fashion do in fact stack in this way. However, Fig. 1.3 illustrates examples of 2,5-2,5 and 3,5-3,5 which has fluorine groups directly stacked on top of each other—the hypothesis was that this was unfavourable for crystal structure formation, but on the face of this evidence it would appear this is not the case. Whilst we do not claim that F···F interactions are structure determining, it can clearly be seen that they can be present without adversely affecting the crystal packing. In fact it could also be concluded

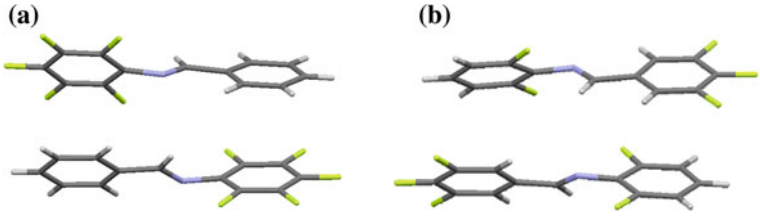

Fig. 1.2 Complementary overlap in (**a**) 0-1,2,3,4,5 and (**b**) 3,4,5-2,6

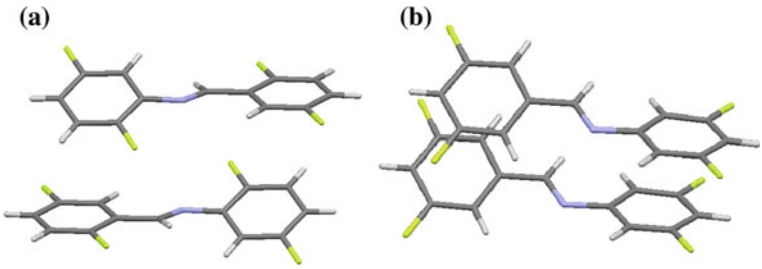

Fig. 1.3 Clashing overlap in (**a**) 2,5-2,5 and (**b**) 3,5-3,5

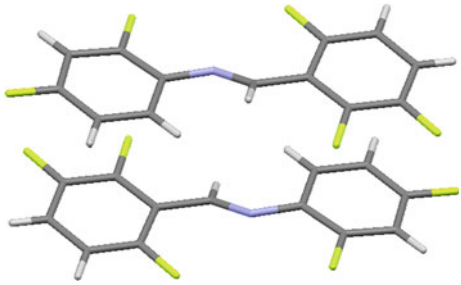

Fig. 1.4 The hybrid complementary—clashing system, 2,3,6-2,4

that the shape packing requirements in this case outweigh those of weak intermolecular interactions.

There are also cases, as exhibited by 2,3,6-2,4 in Fig. 1.4, where a combination of H···F and F···F interactions exist. However, the hypothesis is predicated on the fact that these molecules will naturally want to stack on top of each other as a result of π···π interactions. Hunter and Saunders [28] proposed arguments for this effect based on the fact that electron withdrawing groups reduce the negative quadrupole of the aromatic ring and therefore favour overlapping arrangements (whereas electron-donating groups increase the negative quadrupole and thereby favour offset arrangements). This evidence clearly argues that electrostatic effects are predominant. However, Rashkin and Waters [29] provide evidence to the contrary where meta- and para-substituted N-benzyl-2-(2-fluorophenyl)-pyridinium bromides stack in a parallel displaced conformation as a result of direct interaction of the edge of hydrogen atoms of one ring with the electronegative substituents on the other ring.

Our studies provide examples of both of these models. Moreover, the 'frustrated' systems can often produce 'head-to-tail' threads and side-to-side tapes, thereby avoiding these π–π interactions, in addition to stacks of the nature described above. With a total of 400 structures that could potentially be synthesised and then examined to test our hypothesis this study will be ongoing for a while, however, the first publication is currently in preparation.

1.2.5 Mandelic Acids

Quasiracemate formation frequency and structure, diastereoisomer resolvability and structure and the relationship between racemate and enantiomer structure are all important topics in organic solid-state structure research [30]. As a first stage into researching these phenomena, we have obtained numerous crystal structures of monosubstituted racemic mandelic acids and analysed their structural relationships [31]. Our current work is exploring polymorphism and enantiomeric behaviour of these substituted mandelic acids with a view to understanding diastereoisomer resolution in the solid state.

The substituents chosen were fluoro, chloro, bromo, iodo, trifluoromethyl, methoxyl and methyl and these have been located in the ortho, meta and para positions as indicated by Fig. 1.5. Of the 21 possible monosubstituted racemic mandelic acids outlined here, two methoxyl structures have proved elusive, otherwise it has been possible to compare 19 structures, with the additional inclusion of some polymorphs, using the XPac methodology [19].

These substituents have been used in previous crystal structure comparisons of large sets of related molecules and are chosen to probe structural similarity for a number of reasons. First, they lack strong hydrogen bond donating features, so avoiding interference with the patterns dominated by the hydroxyl and carboxyl groups of mandelic acids, which would complicate or obscure comparison and interpretation. Second, they are sterically undemanding, which further minimises complications in analysing resulting packing arrangements. For this reason also the attention of the study has been restricted to monosubstitution in order to avoid expanding the number and complexity of comparisons unduly.

All structural relationships discovered in this family can be considered as having the lowest common dimensionality (0D as defined by the XPac program) of either 8- or 10-membered hydrogen bonded dimer rings and these have been denoted A- and B-type, i.e. with graph set descriptors of $R_2^2(8)$ and $R_2^2(10)$ respectively (Fig. 1.6).

Figure 1.7 is a full structural relationship plot. As one moves up from the bottom of the structural relationship plot from the 'root' A- and B-type dimers, the degree of dimensionality increases at each level. Beginning with the 0D dimers, denoted A/B01, one moves up, through 1D and 2D to 3D, where the common supramolecular constructs are denoted A/B1*, A/B2* and A/B3* respectively. From the figure it can be seen that there are five 1D constructs, seven 2D constructs

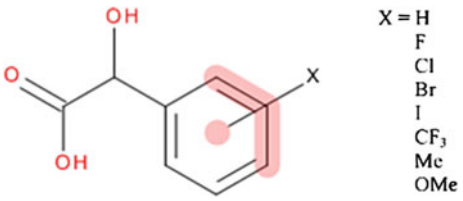

Fig. 1.5 Scheme for substitution patterns of mandelic acid

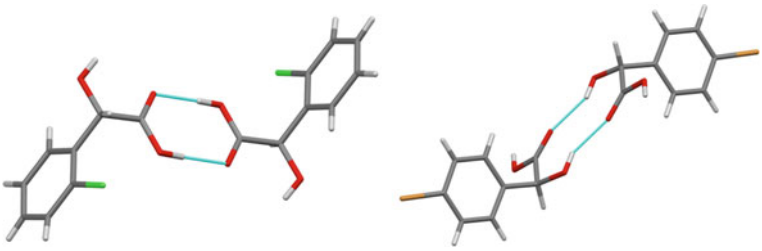

Fig. 1.6 A- and B-type dimers of the substituted mandelic acids

and four 3D constructs. 3D constructs are indicative of isostructurality and in this study the following isostructural groups are observed: AB31 = 2-bromo and 2-iodo; B31 = mandelic acid (polymorph 1), 4-methyl, 4-fluoro, 4-bromo and 4-trifluoromethyl; B32 = 2-fluoro (polymorph 2), 3-fluoro (polymorph 1), 3-chloro (polymorph 1); 3-chloro (polymorph 2), 3-methyl and 3-trifluoromethyl; B33 = 3-bromo and 3-chloro (polymorph 3). Additionally, of particular note, are three sets of relationships that are labelled AB constructs where a combination of A- and B-type assemblies are observed.

The structural relationship plot indicates that there are three relationships that are 1D constructs and a single 2D construct that are solely based on the A-type dimer. There are, however, two structures, 3-iodo and 4-methoxyl, that do not have any higher dimensional relationships and are based purely on the dimer arrangement. The B-type dimer construct is more prevalent than the A-type, which might be considered as contrary to the observation that the carboxylic acid dimer is probably the most common synthon in supramolecular chemistry [32]. This could be attributed to a greater degree of flexibility in the 10-membered ring, which enables the structure to make a strong structure directing contact and also simultaneously accommodate and/or optimise other packing requirements. The structural relationship plot also includes three AB labelled constructs coloured in purple. For a full discussion of the structural aspects of the relationships see reference [31].

The structural relationship plot demonstrates how the hierarchy of dimensionality is built up via a series of common 1D and 2D arrangements originating from just two predominant hydrogen bonding dimer (0D) motifs. Analysing these common motifs and the sets of structures that exhibit them has the potential to provide many interesting insights and highlight areas for exploration and further experimentation. Making a broad observation, it appears that 2-substituted structures are generally based on A-type dimers whilst 3- and 4-substituted tend to be B-type and accordingly the B01 group is more frequent than the A01 grouping.

It is not the purpose of this article to reproduce the work published elsewhere so for brevity and as an example, just one single construct is described in more detail here. Figure 1.8 illustrates the B12 construct (in red) as it is situated within the structure of polymorph 2 of mandelic acid. This construct is a superset of B23, B24 and AB22 (see Fig. 1.7). It is therefore also observed in the structures of

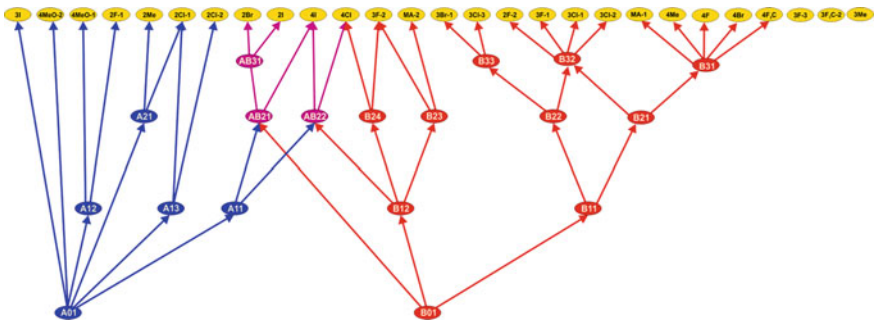

Fig. 1.7 Relationships between all the substituted mandelic acid structures

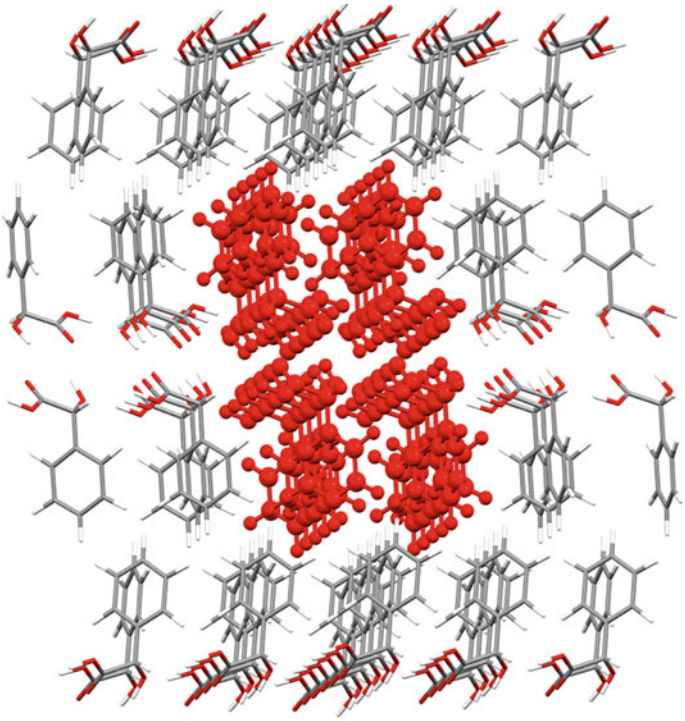

Fig. 1.8 The B12 construct highlighted in the structure of polymorph 2 of mandelic acid

3-fluoromandelic acid (polymorph 2), 4-iodomandelic acid and 4-chloromandelic acid. This relationship alone is remarkable—prior to performing the crystal structure similarity analysis there are no indicators that would lead one to predict that the structures of these compounds would have any similar features!

More specific observations on the structural relationship plot are that the unusual phenomenon of two isostructural polymorphs of 3-chloromandelic acid is observed [33]. The structural relationship plot also has the ability to clearly highlight the absence of expected structures, e.g. a missing 2-bromomandelic acid based on A21, a missing 3-methylmandelic acid based on B33 and a missing 4-bromomandelic acid based on AB22 are noted. An unexpected relationship worthy of note is that between 2-fluoro- and 3-fluoro-mandelic acids, two compounds that are isostructural within the B32 group, where the ortho and meta substituent positions lead to no overall difference in the crystal structure.

This work shows that there are extensive relationships of a 1, 2 and 3-dimensional nature between all the members of the set and indicates that building blocks comprising arrangements of common motifs can be the basis of varying degrees of similarity. Accordingly, the substituted mandelic acids appear to be polymorphically prolific. Two-dimensional relationships, that is sheets of molecules comprised of similar packing motifs, are the basis for a considerable amount of similarity.

1.3 Beyond Atomic Resolution—Systematically Probing the Effect of Weak Interactions

The systematic approach detailed above and illustrated in just a couple of examples of structural families is based entirely on knowledge of molecular structure and packing in the crystal lattice. In order to be able to rationalise this behaviour, and to probe chemistry based on weak intermolecular interactions, it is worthwhile to study structure at electronic (as opposed to atomic) resolution. Our research in this area brings together the concept of structural systematics with the advanced technique of charge density analysis (both experimental and theoretical approaches) in order to provide the necessary further insight. Experimental charge density analysis is a well-established technique [34]. Developments over the last 20 years in computing, software and CCD area detectors have allowed crystallographers to collect ultra-high resolution, exceptional quality diffraction data, needed to model the electron density distribution in the crystalline state. These advances now make it possible to adopt a structural systematics approach using the charge density technique [35].

Analysis of electron density distribution using Bader's Quantum Theory of Atoms In Molecules (QTAIM) [36] allows various bonding interactions to be investigated. In addition to characterising covalent interactions, it is an invaluable tool for quantifying intermolecular interactions. Analysis of the electron density distributions across series of related molecules offers the opportunity to explain in greater depth how alterations to a common molecular scaffold can influence

electronic and physical properties. Atomic resolution crystallography does not take into account the effect of chemical bonding, where some electron density is shared with other atoms. In charge density analysis the Hansen–Coppens multipole formalism discretely models the core and valence electron density in a crystal structure [37]. The (aspherical) valence density describes the deformation from a spherical electron density distribution, associated with both covalent bonding and the presence of lone pairs of electrons.

In QTAIM theory space is divided into disjoint regions known as atomic basins. Various properties relating to the charge distribution may be calculated for these basins and between different basins, e.g. trajectories of maximum electron density linking two atoms are known as bond paths and the combination of these bond paths represents the bonding between the atoms in the crystal structure. Partitioning the electron density into atomic volumes and integrating across this volume calculates the charges of the atoms in the structure. The Laplacian of the electron density ($\nabla^2\rho(r)$), depicts areas of local charge concentrations and depletions and therefore reveals the fine details of the electron density distribution in the crystal structure. This ability to calculate properties of the electron density distribution enables a direct quantitative comparison, which is invaluable for the systematic approach.

Although the geometric criteria used to identify and classify hydrogen bonding in standard resolution X-ray diffraction studies are valid for stronger interactions, they must be viewed with caution when applied to weaker intermolecular interactions. Only charge density analysis can definitively establish the presence of weaker interactions, with the added attraction of being able to compute numerous properties associated with them. These, and numerous other, properties of the charge density may be calculated and compared across families of compounds in a systematic fashion and a recent review [35] provides a complete background of this type of work, which is relatively unexploited. Systematic charge density studies can link observed behaviour to the electronic distribution across molecules in a crystal and pinpoint how alterations to the structure affect this distribution. A variety of chemical insights are available from classifying and comparing the nature and strength of bonds, both covalent and hydrogen bonding interactions, to understanding reactivity of compounds with varying substituent patterns.

Our adoption of charge density analysis in regard of the structural systematics approach is impacting several research areas. Examples include investigating reactivity between electrophile and nucleophile in close proximity, providing a solid-state model for bond formation; understanding unusual bonding in transition and main group coordination chemistry; probing the effect of polymorphism and providing insights to better understand and predict co-crystal formation. However, furthering understanding of anion binding in the field of supramolecular chemistry is the application described below that is used to exemplify the method.

1.3.1 *Systematic Charge Density as a Tool for the Supramolecular Chemist*

Supramolecular chemists design and synthesise molecules to associate with specific guests of interest via non-covalent intermolecular interactions. For example, the position and nature of hydrogen bond donor groups in neutral anion receptor compounds determines the selectivity of the receptor for anions [38], whilst appending electron withdrawing or electron-donating groups to the parent hydrogen bond donor scaffold can modulate the acidity of the hydrogen bond donor groups and hence the strength of the hydrogen bond interactions formed and so modulate affinity [39]. Systematic charge density studies provide information on the nature of the host–guest interaction (for example the strength of the hydrogen bonding interactions and hence the stability of the complex) and the ability to observe how these interactions change as functional groups on the periphery of the receptor are altered and as the receptor binds a variety of different guests.

To probe this phenomenon a systematic family (see Fig. 1.9) of urea-based anion receptor complexes were synthesised and crystallised [40]. By altering the bound anion across a series whilst maintaining a common receptor molecule, the

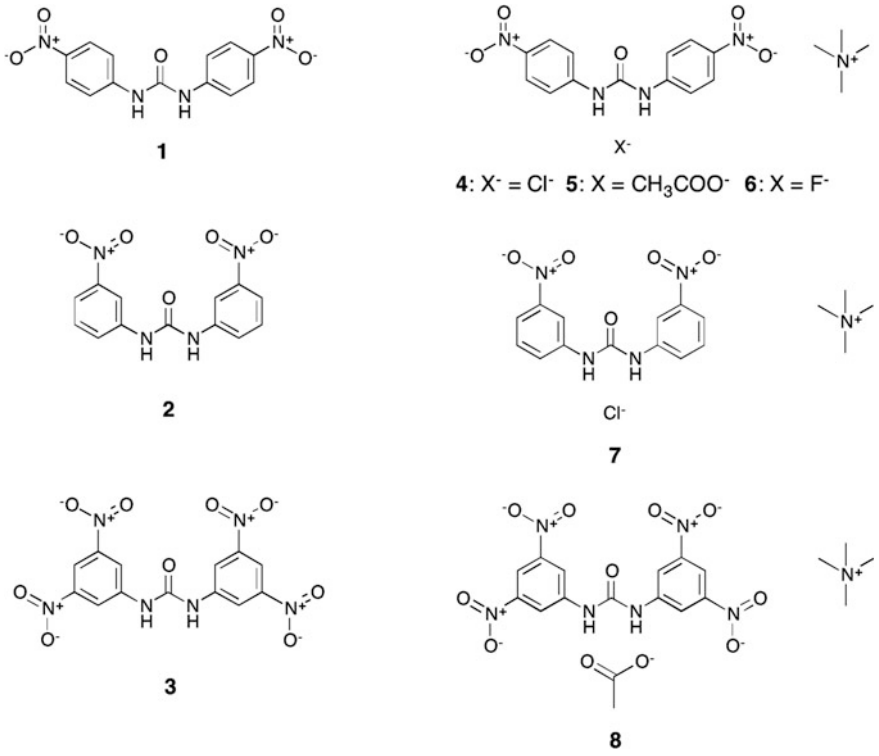

Fig. 1.9 The supramolecular family design

influence of anion basicity on the nature of the interaction between the two can be probed. Additionally, by further including receptors with functional groups in different substitution patterns into the family of complexes studied, it is also possible to assess the effect of this variation on the nature of the interaction with the anion.

The full details of the charge density analysis are provided in a separate publication [41] with just a summary outline and some example results below.

1.3.2 Complexing Different Anions

By studying complexes of 1,3-bis(4-nitrophenyl)urea and varying the anion from chloride to acetate to fluoride, following a trend of increasing basicity, we are able to study how variations in the electron density distributions across the crystal structures, **4** versus **5** versus **6** relate to the basicity of the anion. Figure 1.10 presents an example of such an electron density distribution (for **4**) along with some computed properties varying across the series (**4**–**6**).

When increasing the basicity of the anion from chloride to acetate to fluoride (**4**–**5**–**6**), the electron density and Laplacian values increase in magnitude, indicating a stronger interaction. This correlates well with the observed binding affinities in solution where chloride is shown to have a markedly weaker association with receptor **1** than acetate.

The suitability of using the D-H···A distance to evaluate hydrogen bond strength in atomic resolution crystal structures is verified in this series as an exponential relationship is shown to exist between the H···A distance and electron density (at the Bond Critical Point, BCP, of the hydrogen bond). Increased electron density is an indicator of increased hydrogen bond strength.

In this series two distinct types of hydrogen bond are observed. The first type is of stronger hydrogen bonding, contains the N–H···anion hydrogen bonds of the fluoride and acetate complexes (**5**, **6**, and **8**) and is characterised by electron density values at the BCPs > 0.19 e Å^{-3} and H···A distances <1.80 Å with the bond path between the D···A atoms shorter than the van der Waals radii of the individual atoms. The second type has weaker hydrogen bonding regions with the electron density at the BCPs < 0.15 e Å^{-3} and the H···A distance >2.15 Å and encompasses the N–H···Cl interactions of the chloride structures (**4** and **7**) and the C–H···O interactions in **5**.

A supramolecular chemist can therefore begin to tune the strength of interactions in their system required for their desired function based on the quantified strengths derived from electron density distributions. If strong affinity is required, hydrogen bond strength can be increased by moving to a more basic anion and using the first type of interactions. Additionally, in more involved systems, for example where a series of loading or unloading steps are required, e.g. in a transport based process, more desirable, weaker, second type of interactions can be selected.

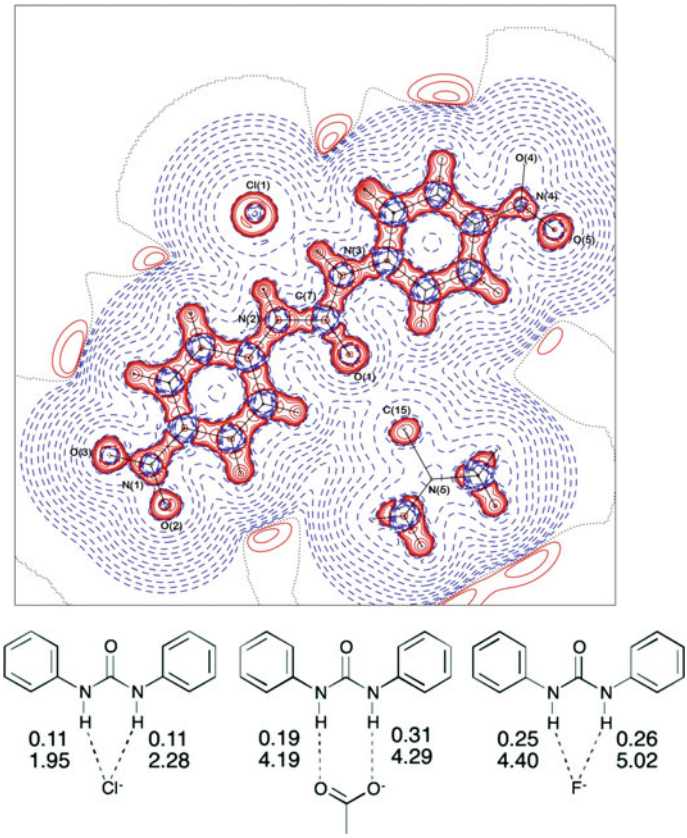

Fig. 1.10 An example of a Laplacian plot (Cl^-, structure **4**), from which trends in the electron density (*top value* in units of e Å^{-3}) and Laplacian values (lower value units of e Å^{-5}) can be computed—in this case at the BCPs between each urea N–H and anion in **4**, **5** and **6**

1.3.3 Charges

Charge density studies offer the ability to determine the individual charges on atoms in a crystal structure. Each atom is viewed as being contained within a surface whose boundaries are minima in the electron density and this is the surface over which integration of the electron density is approximated to the charge of the atom. This provides a probe of the charge transfer between individual units in a supramolecular system and is another handle on how changes to individual components affect not only particular areas but also the electron density distribution across the *entire* structure. This effect can be further correlated to changes in properties such as the electrostatic potential.

Electronegative atoms such as the oxygen atoms of the urea and nitro group, are negatively charged with hydrogen atoms positively charged, as are the nitrogen

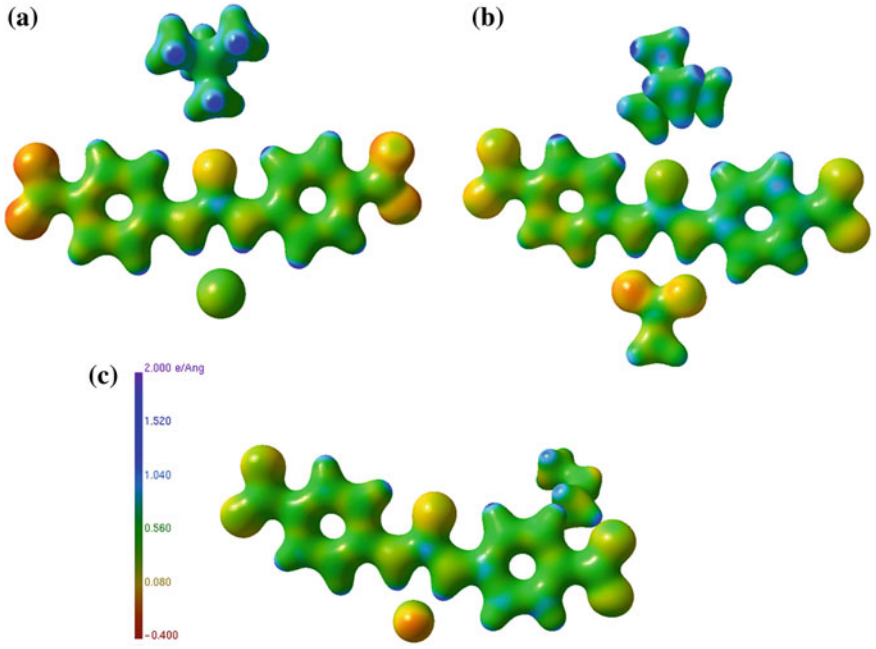

Fig. 1.11 Electrostatic potential plots (e Å^{-1}) of (**a**) **4**, (**b**) **5** and (**c**) **6**

atoms of the nitro groups. The acetate oxygen atoms in **5** are highly negative while the halide anions in **4** and **6** have less negative charge. From **4**–**5**–**6**, as the basicity of the anion is increased, the charges of the urea nitrogen atoms become less negative and the charge on the urea hydrogen atoms in **6** is significantly lower than in **4** and **5**. This perturbation of charge in the urea portion of the structures is shown to extend to the peripheral regions of the structure with the charge on the oxygen atoms of the nitro groups approaching closer to neutrality with increasing basicity. This is reflected in the electrostatic potential maps displayed in Fig. 1.11.

1.3.4 The Substituent Positional Effect

A further comparison that this family of structures offers is that of different receptors complexing a given anion. Through the approach described herein, it is possible to analyse the effect of peripheral modification of the receptor on the electron density distribution by comparing two sets of structures where a common anion is complexed to a changing receptor scaffold. In the two acetate complexes (**5** and **8**) geometric analysis suggests the presence of C–H···O interactions between the phenyl ring hydrogen atoms and acetate oxygen atoms (with H···A distances and DHA angles in **5** of 2.510 Å 137.33° and 2.418 Å 136.78° and in 8 of 2.457 Å

132.98° and 2.623 Å 132.32°). However, bond paths between the respective H···A are only present in **5** implying that while these interactions are present in **5**, in **8** they are an artefact of the close proximity of the atoms to each other in the structure. This is a powerful example of how systematic charge density analysis provides additional information about the interactions in anion–receptor complexes. Additionally, this work demonstrates how modification of the receptor scaffold in this case has brought about changes in the intermolecular interactions observed in the crystal structure, as moving from **5** to **8** the change in receptor substitution pattern is accompanied by a movement of the acetate anion from co-planarity with the receptor in **5** to non co-planarity of the acetate and receptor in **8**.

The strength of interaction has been determined in a family of solid-state anion–receptor complexes. This has not been performed before, as the normal approach to measuring interaction strength is by performing NMR titration studies in solution. The N–H···anion hydrogen bonding interactions were classified as one of two types: strong (N–H···acetate and N–H···fluoride hydrogen bonds) and weak (N–H/C–H···chloride hydrogen bonding). Here we demonstrate new insights that are only possible from systematic studies using charge density analysis and hence provide the field of Supramolecular Chemistry, which is heavily reliant on the crystal engineering type of approach, with a new tool to aid the design of their systems.

1.4 Data Mining and Statistics as a Tool for the Future in Structural Science

The work presented above analyses large 'libraries' of crystal structures to find patterns and trends and thereby derive new data and knowledge. This approach is not entirely new, however, the volume of structures now being generated means we can no longer take this approach without developing new methods. Furthermore we live in a 'Big Data' age and now it is becoming possible to analyse this data in the context of other data—for example a trend in a family of crystal structures could be related to physical properties contained in other, completely unrelated, databases. In the UK, the EPSRC Chemistry Grand Challenge network 'Directed Assembly of Extended Structures' [2] heralds the medium–long term research future for some areas of the subject. The challenge can be described in a very simple sentence: "It is not yet possible to design a material with a particular property". As mentioned above, in the last decade the field of Crystal Engineering has become a vast discipline addressing materials design (although thus far very little attention is being paid to properties). However, this field focuses on complimentary interactions between molecules considered to be structure defining motifs. These motifs are then employed as building blocks in the design of extended structures. Success is variable, often the resulting structure is serendipitous and there are few real application areas—we are at a stage where we need to consider alternative approaches as we reach the limitations of the traditional approach and have a far

greater amount of data about crystal structures than we can hope to analyse 'by hand'.

The approach our research is now taking, inspired and as a result of the structural systematics work described above, combines structural chemistry and statistical analysis with the goal of developing an understanding of how structures form from an information-based route. In order to do this, structurally systematic libraries are constructed with a view to testing a particular question, e.g. (as above) are H···F interactions structure forming? Statistical approaches are then taken to look for correlations between molecular structure, crystal structure and properties. The first step in this process is to generate 'descriptions' for all of these that can be employed in statistical modelling. So called descriptors are produced—these are computable values that can be used to describe shape, geometry, connectivity, etc. Then, in much the way that Quantitative Structure Activity Relationship (QSAR) studies are performed, statistical tests are made to assess the descriptors and look for correlations. The resulting descriptors for the test set of structures are used to build statistical models that are capable of determining what the important descriptors are. Once we have this information it is then possible to make correlations with other data such as physical properties or the propensity to form a crystal.

1.5 Concluding Thoughts

In conclusion, the next 100 years of crystallography promises to deliver results that are equally as exciting as the first century of the subject. However, we are now shifting into a period where the technique itself is mature and we can focus more on the application. By assembling large *related* families of crystal structures we will be able to understand fundamental principles governing the formation of the solid state, which will in turn enable us to control design and ultimately with devices and properties in mind. Crystals are, and will continue to be, very important to many aspects of the world around us and will form the basis of many new innovations, so it is vital to be able to understand, control and predict their formation and behaviour—many challenges yet lie in wait.

Acknowledgements This article contains, and alludes to, a summary of a vast amount of work—it is most certainly not solely down to the author! Firstly, I would like to acknowledge the Engineering and Physical Sciences Research Council for its continued funding of the UK National Crystallography Service over a period spanning more than three decades (and has directly funded the author for two of these!). My mentor, Mike Hursthouse, introduced me to this entire subject area and did the majority of the founding work—and is continuing to do so with just as much insight from retirement. Terry Threlfall has a wealth of experience in this area—he began collaborating with the group around 20 years ago and is still driving us forward with as much enthusiasm and insight as ever. The researchers who conduct the work, past and present, are numerous and involve a wide range of people from summer students through to retired professors. The most notable, who directly contributed to the specific work described herein, are Graham Tizzard, Mateusz Pitak, Claire Wilson, Thomas Gelbrich, Philip Adler, Isabelle Kirby, Riccardo Montis, Amy Ellis, Jonathan Sarson, Ka Leung and Eleanor Dodd.

References

1. Bartlett SW, Coles SJ, Davies DB, Hursthouse MB, Ibisoglu H, Kiliç A, Shaw RA, Ün I (2006) Structural investigations of phosphorus–nitrogen compounds. 7. Relationships between physical properties, electron densities, reaction mechanisms and hydrogen-bonding motifs of $N_3P_3Cl_{(6-n)}(NHBu^t)_n$ derivatives. Acta Crystallogr B 62:321–329 and references therein
2. The Directed Assembly Grand Challenge (2014) Accessed on 09 Dec 2014 http://beyondthemolecule.org.uk/
3. The Engineering and Physical Sciences Research Council (2014) Accessed on 09 Dec 2014 http://www.epsrc.ac.uk/
4. Desiraju GR (2002) Cryptic crystallography. Nat Mater 1:77–79
5. Desiraju GR (2003) Supramolecular synthons in crystal engineering—a new organic synthesis. Angew Chem Int Ed 34:2311–2327
6. Gale PA, Steed JW (2012) Supramolecular chemistry: from molecules to nanomaterials. Wiley. ISBN: 978-0-470-74640-0
7. Fujita M, Eddaoudi M, Eubank JF, Batten SR, Hubberstey P, Lin X, Champness NR, Schr öder M, Ghosh SK, Kitagawa S, Lin W, Wu S, Kim H, Chun H, Kim K, Collins DJ, Ma S, Zhou H-C, Friščíc T, Dutta S, Georgeiev IG, MacGillivray LR, Snur RQ, Özgür Yazaydin A, Dubbeldam D, Frost H (2010) Metal-organic frameworks: design and application, 1st edn. Wiley and Sons, NY
8. Larrea ES, Roberto Fernández de Luis JLM, Pizarro JL, urtiaga MK, Rojo T, Arriortua MI, Kremer C, Chiozzone R, Li Y, Son P, Li Y, Li X, Vinh-Thang H, Kaliaguine S, i Xamena FXL, Linares-Solano A, Cazorla-Amorós D, Marco-Lozar JP, Súarez-Garcia F, Jiang J, Fan Y-J, Zang S-Q, Wei Y-L, Du C-X, Hou H-W, Yamada T, Kitagawa H, Nuzhdin AL, Kovalenko KA, Fedin VP, Bukhtiyarova GA (2012) Coordination polymers and metal organic frameworks. Chemical engineering methods and technology. Nova Science Publishers, New York
9. Dolle RE, Bourdonnec BL, Worm K, Morales GA, Thomas CJ, Zhang W (2010) Comprehensive survey of chemical libraries for drug discovery and chemical biology: 2009. J Comb Chem 6:765–806 and references therein
10. Villar HO, Hansen MR (2009) Design of chemical libraries for screening. Expert Opin Drug Discov 12:1215–1220
11. Allen FH (2002) The Cambridge structural database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58:380–388
12. Nangia A (2002) Database research in crystal engineering. Cryst Eng Comm 4:93–101
13. Taylor R (2014) Which intermolecular interactions have a significant influence on crystal packing? Cryst Eng Comm 16:6852–6865
14. Hursthouse MB (2004) High-throughput chemical crystallography (htcc): meeting and greeting the combichem challenge. Cryst Rev 10:85–96
15. Orpen AG (1993) Structural systematics in molecular inorganic chemistry. Chem Soc Rev 22:191–197
16. Ganguly P, Desiraju GR (2010) Long-range synthon Aufbau modules (LSAM) in crystal structures: systematic changes in C_6H_6–nFn (0 <= n <= 6) fluorobenzenes. Cryst Eng Comm 12:81–833
17. McCrone WC (1965) In: Fox D, Labes MM, Weissberger A (eds) Physics and chemistry of the organic solid state vol 2. Interscience Publishers, pp 725–767
18. Fabian L, Kalman A (2004) Isostructurality in one and two dimensions: isostructurality of polymorphs. Acta Crystallogr B 60:547–558
19. Gelbrich T, Hursthouse MB (2005) A versatile procedure for the identification, description and quantification of structural similarity in molecular crystals. Cryst Eng Comm 7:324–336

20. Gelbrich T, Threlfall TL, Hursthouse MB (2012) XPac dissimilarity parameters as quantitative descriptors of isostructurality: the case of fourteen 4,5′-substituted benzenesulfonamido-2-pyridines obtained by substituent interchange involving CF_3/I/Br/Cl/F/Me/H. Cryst Eng Comm 17:5454–5464
21. Gelbrich T, Hursthouse MB (2006) Systematic investigation of the relationships between 25 crystal structures containing the carbamazepine molecule or a close analogue: a case study of the XPac method. Cryst Eng Comm 8:448–460
22. Gelbrich T, Hursthouse MB, Threlfall TL (2007) Structural systematics of 4,4'-disubstituted benzenesulfonamidobenzenes. 1. Overview and dimer-based isostructures. Acta Crystallogr B63:621–632
23. Montis R, Hursthouse MB (2012) Crystalline adducts of some substituted salicylic acids with 4-aminopyridine, including hydrates and solvates: contact and separated ionic complexes with diverse supramolecular synthons. Cryst Eng Comm 14:7466–7478
24. Tothadi S, Desiraju GR (2013) Designing ternary cocrystals with hydrogen bonds and halogen bonds. Chem Commun 49:7791–7793
25. Bui TTT, Dahaoui S, Lecomte C, Desiraju GR, Espinosa E (2009) The nature of halogen···halogen interactions: a model derived from experimental charge-density analysis. Angew Chem Int Ed 48:3838–3841
26. Metrangolo P, Neukirch H, Pilati T, Resnati G (2005) Halogen bonding based recognition processes: a world parallel to hydrogen bonding. Acc Chem Res 38:386–395
27. Panini P, Chopra D (2012) Role of intermolecular interactions involving organic fluorine in trifluoromethylated benzanilides. Cryst Eng Comm 14:1972–1989
28. Hunter CA, Sanders JKM (1990) The nature of pi–pi interactions. J Am Chem Soc 14:5525–5534
29. Rashkin MJ, Waters ML (2002) Unexpected substituent effects in offset pi–pi stacked interactions in water. J Am Chem Soc 9:1860–1861
30. See for example control and prediction of the solid state (2014). Accessed on 09 Dec 2014 http://www.cposs.org.uk/
31. Coles SJ, Ellis AL, Leung K, Sarson J, Threlfall TL, Tizzard GJ (2014) Relationships between the racemic structures of substituted mandelic acids containing 8- and 10-membered hydrogen bonded dimer rings. Cryst Eng Comm 16:10816–10823 and references therein
32. Desiraju G (1996) The crystal as a supramolecular entity. Wiley, p 43
33. Coles SJ, Threlfall TL, Tizzard GJ (2014) The same but different: isostructural polymorphs and the case of 3-chloromandelic acid. Cryst Growth Des 14:1623–1628
34. Macchi P (2013) Modern charge density studies: the entanglement of experiment and theory. Crystallogr Rev 19:58–101
35. Kirby IL, Pitak MB, Coles SJ, Gale PA (2014) Systematic experimental charge density: linking structural modifications to electron density distributions. Chem Lett 44:147–149
36. Bader RFW (1990) Atoms-in-molecules: a quantum theory. Oxford University Press, Oxford
37. Hansen NK, Coppens P (1978) Testing aspherical atom refinements on small-molecule data sets. Acta Crystallogr A 34:909–921
38. Gale PA (2006) Structural and molecular recognition studies with acyclic anion receptors. Acc Chem Res 39:465
39. Amendola V, Fabbrizzi L, Mosca L (2010) Anion recognition by hydrogen bonding: urea-based receptors. Chem Soc Rev 39:3889–3915
40. Kirby IL, Pitak MB, Wenzel M, Wilson C, Sparkes HA, Coles SJ, Gale PA (2013) Systematic structural analysis of a series of anion receptor complexes. Cryst Eng Comm 15:9003–9010
41. Kirby IL, Brightwell M, Pitak MB, Wilson C, Coles SJ, Gale PA (2014) Systematic experimental charge density analysis of anion receptor complexes. Phys Chem Chem Phys 16:10943–10958