CrystEngComm

Accepted Manuscript

Which intermolecular interactions have a significant influence on crystal packing?[†]

Robin Taylor

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK. Email: robin@justmagnolia.co.uk

Abstract: The tendency for an interaction to occur in crystal structures is not a simple function of its calculated energy *in vacuo*. This was shown by ranking intermolecular atom…atom interactions in organic crystal structures on the ratio (R_F) of their observed frequency of occurrence to the frequency expected at random, *i.e.* if determined solely by the exposed surface areas of atoms. The study was based on line-of-sight interactions in structures taken from the Cambridge Structural Database. Only one interaction per atom was included in the analysis, the one with the smallest value of d-V, where d is the interatomic distance and V the sum of the atoms' van der Waals radii. 95% confidence intervals were determined for each R_F value, enabling identification of interactions that occur significantly more often than expected by chance. Strong hydrogen bonds have the highest R_F values, followed by two halogen-bonding interactions, I…N and I...O. These strong interactions typically occur 3 to 10 times more often than would be expected by chance. Although comparatively weak in energetic terms, C-H...F and C-H…Cl have R_F values significantly in excess of the random expectation value of 1, and higher, for example, than those of Br…O and Cl…O. R_F values clearly reveal the effects of polarisation on the propensity for C-halogen groups to form halogen bonds and C-H groups to form hydrogen bonds to oxygen,

---

[†] Electronic supplementary information (ESI) available: Description of statistical method used to validate R_F confidence intervals, comma separated value files of full results obtained with coarse and fine base-atom types.

and highlight the dramatic differences between the interactions of phenyl and pentafluorophenyl.

Introduction

It did not take long for the pioneers of chemical crystallography to recognise the importance of hydrogen bonding. For example, Robertson observed that “the occurrence of the -OH group is generally characterized in the solid state by unusually small intermolecular distances”.[1] Crystallographers tacitly assumed that hydrogen bonding was confined to O-H and N-H donors until Sutor postulated the existence of C-H…O hydrogen bonds.[2,3] However, her suggestion was challenged trenchantly by Donohue,[4] whose scepticism dominated thinking for over a decade. In the 1980s and 1990s, several authors concluded that Sutor was correct after all.[5-13] The pendulum then swung so emphatically towards this point of view that Cotton and co-workers protested, saying “We strongly disagree with the newer and more relaxed definitions that do not distinguish between a ‘hydrogen bond’ and what is nothing more than a classical van der Waals interaction.”[14] But they were swiftly rebutted[15] and the more relaxed definition was subsequently enshrined by an IUPAC task force.[16]

Meanwhile, the importance of several other types of stabilizing intermolecular interaction was asserted. Notable among these were halogen bonds, about which there is now an extensive literature,[17-24] together with the related chalcogen and pnicogen bonds.[25] Halogen bonds are generally believed to be confined to chlorine and heavier halogens but arguments were made recently that fluorine can also donate them under appropriate conditions.[26,27] Other interactions of carbon-bound fluorine have also received attention. It has been argued that the C-H…F interaction is a weak hydrogen bond that can play a significant role in stabilizing certain crystal structures.[28,29] It was also suggested that interactions such as F…F and C-F…π stabilize molecular assemblies.[30-32] Largely on geometrical evidence, Aakeröy and co-workers concluded that C-H…Cl interactions are hydrogen bonds for both anionic and neutral chlorine.[33] It was also suggested that C-Cl…π and C-Br…π interactions can stabilize protein-ligand complexes and small molecule crystal structures.[34] The importance of C-H…π interactions and their value as supramolecular synthons was highlighted.[35,36] Dougherty and colleagues emphasised the significance of cation…π interactions.[37,38] Not to be outdone, anion…π interactions also had their advocates,[39] although their importance has been strongly disputed.[40]

In the face of so many "special" interactions, it is hard not to feel a trace of doubt. Can they all have a significant role in stabilizing crystal structures? Dunitz and Gavezzoti expressed reservations, writing "Only atoms on the outer surfaces of molecules can come into contact with atoms of other molecules. Since these are usually hydrogen atoms, C-H…X contacts in organic crystals are almost unavoidable. One must therefore be very careful about their significance as main contributors to intermolecular stabilization."[41] The same authors conceded that weak hydrogen bonds such as C-H…F, C-H…Cl, C-H…O, C-H…N and C-H…π may be categorised as existent rather than non-existent on the basis of geometric, spectroscopic and even energetic criteria.[42] But they added "The question is not whether weak hydrogen bonds 'exist' but rather to what extent are they relevant in distinguishing one possible crystal structure from another?" This goes to the heart of the matter. Whether we distinguish an interaction with a special name is less important than whether the interaction distinguishes itself by playing a special role in stabilizing crystal structures.

At least three properties of an interaction may indicate that it has an important stabilizing role: energy, geometry and frequency of occurrence. The attention they have received decreases in the order they were just listed. Energy is the most often studied, usually by quantum mechanical calculations. The implicit assumption is that the more attractive the calculated energy of an interaction, the more likely it is that the interaction is important in stabilizing crystal structures. Geometry has also received considerable attention, especially angular directionality: it is a natural topic of interest because interactions such as halogen bonds that show exquisite directional preferences[43] offer crystal engineers a powerful tool in designing crystal structures. It is argued that van der Waals (henceforth "vdw") forces are isotropic, so an intermolecular interaction that shows significant directional preferences must be more than a vdw contact.[15,33]

The final and least-studied property, frequency of occurrence, has the inestimable virtue of capturing what actually happens in practice. Whatever its calculated energy, if an interaction does not occur appreciably more often in crystal structures than would be expected by chance, the significance of its role in distinguishing one possible crystal structure from another must be questioned. Allen and co-workers ranked interactions on the ratio N_a/N_p, where N_a is the number of structures in the Cambridge Structural Database (CSD)[44] in which the interaction occurs, and N_p is the number in which it could possibly occur.[45,46] However, they did not determine confidence intervals for their ratios. The elegant logit-regression method of Galek and colleagues[47] is also worthy of note, but has been primarily focussed on the competition between the different hydrogen bonds that may occur in a given structure. In this paper, a new method for analysing and comparing frequencies of occurrence is presented and applied to a wide variety of intermolecular interactions.

Experimental

Data set

The study was based on 137,560 crystal structures taken from the CSD, version 5.35. All satisfied the following criteria: no elements other than H, C, N, O, F, P, S, Cl, Br or I; R-factor ≤ 7.5% (structures with no quoted R-factor were omitted); no missing atomic coordinates; no disorder (no disorder comment, no suppressed atoms); only one representative from each “refcode family” (a refcode family comprises independent determinations of the same chemical compound; the structure with the lowest R-factor was selected). Hydrogen atom positions were normalised[48] by moving the hydrogen atoms along the observed X-H bond vectors so that the XH distances were set to average neutron-diffraction values[49] (CH = 1.083Å, NH = 1.009Å, OH = 0.983Å).

Interactions included in analysis

The study was confined to intermolecular line-of-sight (LoS) interactions. A LoS interaction is one in which the interacting atoms “see” each other because no third atom intrudes between them. Fig. 1 shows two possible interactions between a pair of atoms A and B. The solid circles represent the vdw spheres of the atoms, P_A and P_B being the points at which these spheres intercept the AB line segment. The interaction shown at the top is longer than V_A+V_B, the sum of the vdw radii of A and B. This interaction is LoS if there is no other atom in the crystal structure whose vdw sphere intercepts the line between P_A and P_B. The bottom interaction is shorter than V_A+V_B. It is LoS if there is no atom whose vdw sphere encloses the line segment P_AP_B. The broken circles in Fig. 1 show example positions of a third atom that would render the A…B interaction not LoS. The purpose of using only LoS interactions is to focus as far as possible on interactions that are not artefacts of other, stronger interactions. For example, the N…O interaction in an N-H…O hydrogen bond is artefactual (and invariably not LoS) because the primary stabilizing interaction is the H…O contact (which is LoS).

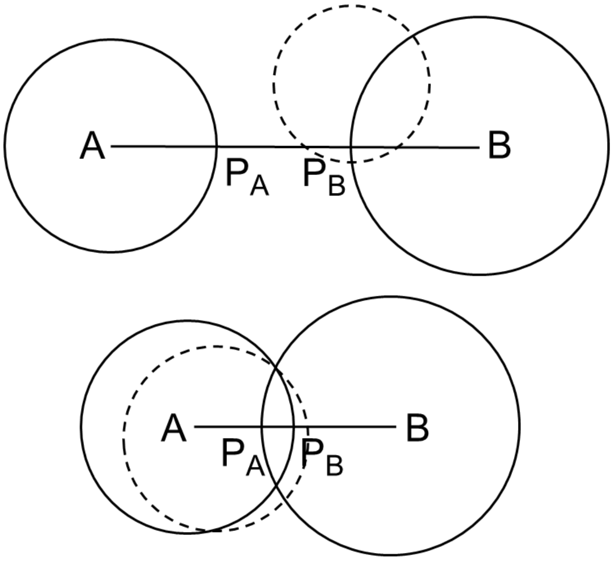

Fig. 1 Two possible interactions between atoms A and B. Broken circles represent example positions of a third atom that would prevent A…B from being line-of-sight.

Except where otherwise stated, the interactions included in the analysis were further restricted as follows. Each crystallographically independent atom in the data set (the *base* *atoms*) was considered in turn. For any such atom, B, all intermolecular A…B LoS interactions shorter than $V_A+V_B+1Å$ were found (the use of the 1Å tolerance meant that at least one LoS interaction was found for the large majority of atoms). Only the interaction with the smallest value of Δ_{AB}:

$$\Delta_{AB} = d_{AB} - V_A - V_B \qquad (1)$$

(where d_{AB} is the AB distance) was accepted for inclusion in the study. The chosen interaction is termed the *primary interaction* of B. The purpose was to focus the analysis on the interactions that are most likely to reflect the inherent preferences of the base atoms, while allowing these interactions to be substantially longer than the sum of vdw radii if necessary. It has been established that interactions may be longer than the sum of vdw radii yet still have a significant stabilizing role.[50]

As an example, Fig. 2 (left) shows all A…B intermolecular interactions shorter than $V_A+V_B+1Å$ around the crystallographically-independent molecule in CSD structure AABHTZ; Fig.2 (centre) shows the subset of these interactions that are LoS; and Fig. 2 (right) shows the subset of these interactions chosen for inclusion in the analysis, one per base atom. Over the complete data set, about 4.5% of base atoms had no LoS interactions shorter than $V_A+V_B+1Å$, and were omitted from the study. Some of these atoms were inspected visually and invariably appeared sterically hindered, *e.g.* tertiary carbon atoms, and atoms in the interiors of crowded molecules such as CSD entry ADABOX.

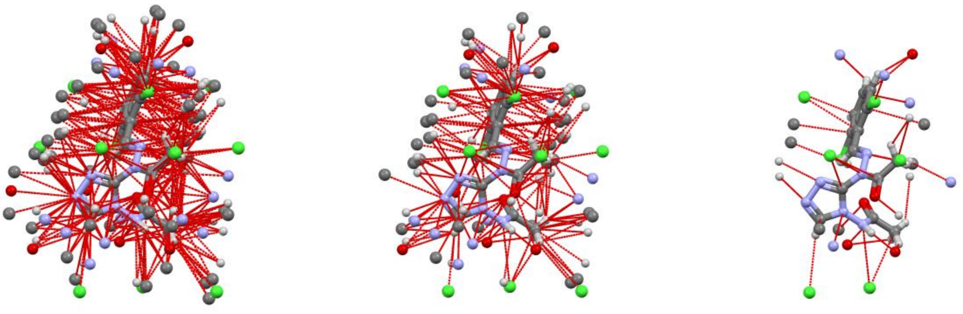

Fig. 2 Intermolecular interactions formed by the crystallographically independent molecule in CSD entry AABHTZ. *Left:* all interactions shorter than sum of vdw radii + 1Å; *Centre:*

line-of-sight interactions; *Right:* shortest line-of-sight interactions relative to vdw radii, one for each atom in molecule.

Atom typing

Interactions were classified by assigning atom types to the interacting atoms. Two main atom-typing schemes were used, *coarse* and *fine*. In both, atoms were assigned types depending on their substructural environments. An ordered list of substructures was defined using SMARTS strings[51] and each atom assigned the type corresponding to the first substructure it matched (Tables 1, 2). The coarse atom types were used in preliminary investigations and the fine atom types in follow-up studies. Further atom types, described later, were used for more detailed analyses of specific interaction types.

Surface area calculations

The exposed surface area of each crystallographically independent atom in the data set was computed by placing points randomly on its vdw surface and counting how many did not fall within the vdw envelope of any other atom in the same molecule or ion.[52] 15,000 points per atom were used, which is sufficient to produce results with an average standard deviation varying from 0.07Å^2 for hydrogen to 0.18Å^2 for iodine (an uncertainty of about 1% in the calculated exposed surface area).

Statistical considerations

Suppose we wish to test whether base atoms of atom type T_B prefer to form their primary interactions to atoms of type T_A. For a given structure containing both types of atoms, the null hypothesis is stated as:

$$H_0: p = S(T_A) / S(\text{total}) \qquad (2)$$

and the alternative hypothesis as:

$$H_1: p > S(T_A) / S(\text{total}) \qquad (3)$$

where p is the probability that a base atom of type T_B will form its primary interaction to an atom of type T_A, S(total) is the total surface area of all molecules and ions in the asymmetric unit, and $S(T_A)$ is the exposed surface area of atoms of type T_A. The null hypothesis therefore states that the probability of a primary interaction to an atom of a given type depends solely on the proportion of molecular surface area contributed by atoms of that type.

Let N be the number of base atoms of type T_B in a given structure, excluding any that do not form a LoS interaction shorter than the sum of vdw radii plus 1Å. Under the null hypothesis, the expected number of primary interactions to atoms of type T_A is binomially distributed. Each of the N base atoms constitutes a “trial” (remembering that there is one primary interaction per base atom) and the probability of success in a trial (an interaction to an atom of type T_A) is $S(T_A)/S(\text{total})$. The mean of a binomial distribution is (number of trials) x (probability of success). Therefore, the expected number of $T_B \ldots T_A$ primary interactions under the null hypothesis is:

$$E(T_B,T_A) = NS(T_A)/S(\text{total}) \qquad (4)$$

(This quantity will be non-integral, of course.) The following statistic (R_F, <u>r</u>atio of <u>f</u>requencies) can then be defined:

$$R_F(T_B,T_A) = \Sigma O(T_B,T_A)_i \,/\, \Sigma E(T_B,T_A)_i \qquad (5)$$

where the summations are over all crystal structures that contain both types of atoms, $O(T_B,T_A)_i$ is the observed number of $T_B \ldots T_A$ primary interactions in the ith structure, and $E(T_B,T_A)_i$ is the expected number. An interaction with R_F=x occurs x times more often than expected by chance.

The uncertainty of $R_F(T_B,T_A)$ was estimated by bootstrapping. Suppose the structures containing atoms of types T_B and T_A are (S_1, S_2, … $S_{NSTRUCT}$). 100,000 estimates of the statistic were made, each based on NSTRUCT structures chosen randomly from the set (S_1, S_2 … $S_{NSTRUCT}$) by sampling with replacement (*i.e.* so that any given structure might appear more than once, or not at all). The 2.5th and 97.5th percentiles of the resulting distribution were used as the 95% confidence interval of the true $R_F(T_B,T_A)$. If the lower limit of the confidence interval exceeds 1, the null hypothesis can be rejected at the 95% confidence level, *i.e.* it can be concluded that the interaction occurs more often than expected by chance. The confidence intervals were validated by using a second statistical method, based on binomial probability analysis, to determine whether the null hypothesis should be rejected. When tested on a large sample of interactions, the two methods gave closely similar results, indicating that the confidence intervals derived for R_F by bootstrapping are reliable. Full details of the second statistical method have been deposited as supplementary material.

vdw radii

The values of $S(T_A)$ and S(total), and thus the null-hypothesis probability that a base atom will form its primary interaction to an atom of type T_A, will depend on the choice of vdw radii. By far the most widely used radii are those from the seminal work of Bondi.[53] His estimate for element X was based on X…X contact distances in crystal structures carefully chosen from the relatively small number available in 1964. In addition, he ensured that his values were compatible with crystal densities at 0K. Bondi's radii have been shown to have a strong relationship with the shortest intramolecular nonbonded contact distances in organic molecules.[54]

Rowland and Taylor (henceforth "R&T") derived vdw radii for common non-metallic elements from intermolecular distance distributions (excluding distances strongly influenced by hydrogen bonding) determined from about 28,000 CSD crystal structures.[55] For as many element pairs X…Y as possible (depending on the availability of sufficient data), they estimated the distance, d_{XY}, at which the distribution reached half its maximum height. Radii were then chosen to achieve optimum least-squares agreement between the d_{XY} and the radii sums ($V_X + V_Y$). Recently, Alvarez produced radii for a large number of non-metallic and metallic elements.[56] He used the "half-height" methodology of R&T, but had access to many more crystal structures. His radii for oxygen and hydrogen were based on O…O and neutron-diffraction D…D distributions, respectively; most of the other radii, V_X, were based on X…O distributions (the oxygen radius having already been determined from the O…O distribution).

Alvarez's procedure therefore differs significantly from R&T's, whose radius for a given element X was a compromise based on several different X…Y distributions.

Table 3 gives the three sets of radii for the elements of interest in this work. While the overall consensus is good, there are some noticeable differences. In particular, the smaller V_H value of R&T almost certainly arises because it was a compromise radius based on several different types of H…Y distributions, whereas the Bondi and Alvarez values were derived solely from H…H (or D…D) contact distances. Hydrogen atoms in organic molecules usually carry a small net positive charge, so H…H contacts are likely to be slightly lengthened by electrostatic repulsion. Conversely, the opposite will tend to happen in some other types of interaction, such as H…C. It is a moot point which hydrogen radius is more appropriate for the current work. In any case, as Bondi pointed out, there is an inherent arbitrariness in how a vdw radius is defined. In view of the differences between the three sets of radii, the analysis was performed in triplicate, using each set in turn. Alvarez noted that the anionic radii of electronegative elements are all very similar to the vdw ones, so radii for halide anions were assumed identical to the corresponding uncharged halogen radii.

Sources of uncertainty

The major sources of uncertainty in any R_F estimate will now be summarised. First, there will be random sampling error due to the finite number of crystal structures on which the estimate is based. Its likely magnitude is assessed by the bootstrapped confidence intervals of R_F. Second, there is the uncertainty due to the choice of vdw radii. Comparison of the results from the three radii sets gives insight into its importance. Third, there may be systematic error due to a biased set of crystal structures, *e.g.* a set containing a significant proportion of closely related molecules. This can be guarded against by confining the study, as far as possible, to interactions that have the potential to occur - because both of the relevant atom types are present - in a large number of structures. This should make the likelihood of bias small. Finally, there is a philosophical problem in assuming that the number of $T_B…T_A$ interactions is binomially distributed with a probability of success equal to $S(T_A)/S(total)$. Specifically, if one base atom forms an interaction to an atom of type T_A, it is conceivable that the probability of a second base atom also doing so is reduced. Whether the problem is of practical importance is imponderable and, in any case, it is difficult to see how it could be resolved without introducing further, and worse, assumptions or uncertainties. Nevertheless, it is as well to bear the reservation in mind.

Results and Discussion

Comparison of results from different vdw radii

R_F values were computed using all three sets of vdw radii and the coarse atom-typing scheme (Table 1) for both base atoms and the atoms to which they form their primary interactions (henceforth, *partner atoms*). Each combination of atom types was considered, giving a total of $17^2 = 289$ different types of interaction. Figs. 3 and 4 shows the R_F values from the Bondi radii plotted against those from R&T and Alvarez, respectively, based on interactions that have the possibility to occur in at least 500 structures. Interactions with $R_F>3$ are omitted so that the region of the plot where most interactions fall can be better visualised. The Spearman correlation coefficients of the R_F values from different radii are excellent, *viz.* 0.96, 0.98 and 0.98 for Bondi versus R&T, Bondi versus Alvarez, and R&T versus Alvarez, respectively. However, substantial differences are observed for a few interactions. For example, the R_F values for F…H[C] are 1.87, 1.80 and 1.55 from the Bondi, Alvarez and R&T radii, respectively; those for F…C[sat] are 0.59, 0.84 and 1.04; and those for F…C[unsat] are 0.55, 0.73 and 0.90. (Here and throughout, interactions are written with the base atom first and the partner atom second.)

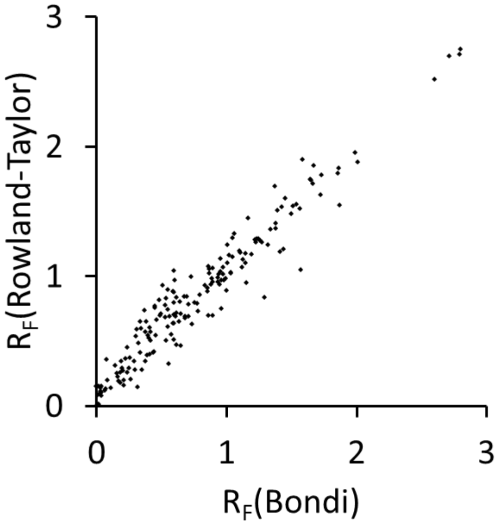

Fig. 3 Plot of R_F values from analyses based on vdw radii taken from Bondi[53] and Rowland and Taylor.[55]

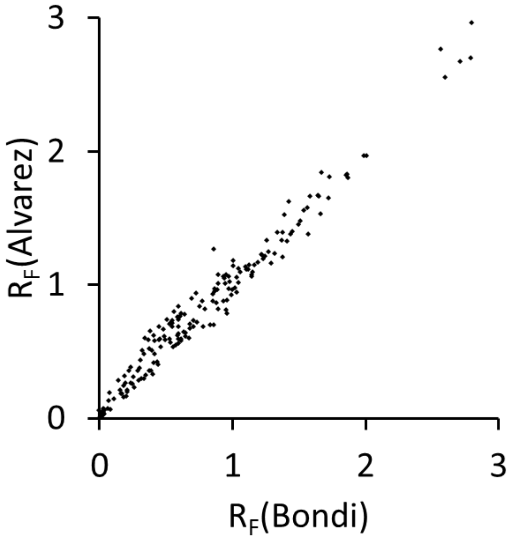

Fig. 4 Plot of R_F values from analyses based on vdw radii taken from Bondi[53] and Alvarez.[56]

Switching from one set of vdw radii to another can alter R_F values by changing atomic surface areas, by altering which interactions are LoS, and by changing which of the LoS interactions are considered primary interactions (since Δ_{AB} in equation 1 depends on the vdw radii). Manual examination of several structures suggested that the latter is the most important factor. Consider, for example, a fluorine atom that forms short contacts to both hydrogen and carbon atoms. Bondi assigned a relatively small radius to carbon and a relatively large one to hydrogen, whereas R&T did the opposite, with Alvarez taking an intermediate position. Therefore, it is possible that switching from Bondi to R&T radii could alter the primary interaction of the fluorine atom from F…H to F…C. This explains the R_F trends given at the end of the previous paragraph. Fig. 5 shows an example. When Bondi radii are used, the primary interaction of the labelled fluorine atom is to the labelled hydrogen; with R&T radii, it is to the labelled carbon. In this case, the Bondi result seems better, because the stabilizing interaction is surely F…H rather than F…C. But it is not always so clear. Often, switching from one set of radii to another alters primary interactions in such a way that both alternatives look equally reasonable.

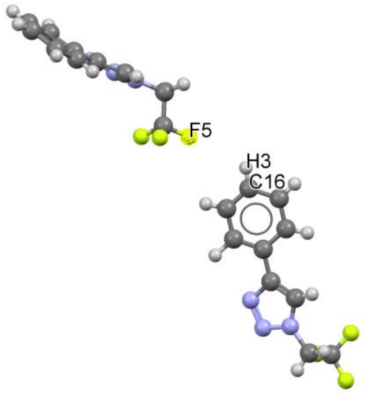

Fig. 5 F5 in CSD structure ACEJIC is deemed to form its primary interaction to H3 or C16, depending on which set of vdw radii is used.

To put things into perspective, Figs. 3 and 4 and the correlation coefficients quoted earlier show that the sensitivity of the results to the vdw radii is not a major problem. Nevertheless, it is the biggest source of uncertainty and cannot be ignored. Therefore, R_F values quoted from here on are averages from parallel analyses based on the Bondi, R&T and Alvarez radii. The confidence intervals are based on the lowest 2.5th percentile and the highest 97.5th percentile from the three analyses. These limits therefore take account of both sampling errors and uncertainties in the vdw radii.

Results based on coarse atom types

Table 4 lists interactions which have the potential to occur in at least 500 structures, ranked in descending order of R_F. The 50 top-ranked interactions are included, together with some of lower rank. Results for all interactions are deposited as supplementary material. Unsurprisingly, most of the highest-ranked interactions (9 of the top 10) are strong hydrogen bonds, the top two involving ionised acceptors. The highest-ranked halogen bond is I…N[acc] at rank 5. (N[acc] represents a nitrogen atom capable of accepting a hydrogen bond, *i.e.* with a localised lone pair.) I…O is at rank 11. It will be noted that R_F(X…Y) ≠ R_F(Y…X), and sometimes the difference is sizeable. For example, the R_F values of I…N[acc] and N[acc]…I are 5.4 and 1.8, respectively. This is because R_F(X…Y) expresses the preference that X has for interactions with Y *relative to other types of interactions that X might form*. N[acc] is the preferred partner for I, but I is only the second-highest ranked partner atom for N[acc], which much prefers H[polar] (R_F=5.7). I…O and O…I show similar behaviour.

Just as the top positions in Table 4 are mainly occupied by interactions that are obviously strongly stabilizing, so most of the lowest-ranked interactions, with $R_F << 1$, involve atoms that would be expected to repel each other strongly, such as halide…halide, halide...O, halide…N[acc], *etc*. For example, the R_F values for Br[-]…Br[-], Cl[-]…O and O…O are 0.0, 0.1 and 0.2, respectively.

Hydrogen bonds with C-H donors have high ranks, the three highest involving halide acceptors (ranks 12-14), with H[C]…O and H[C]…N[acc] slightly lower (ranks 15 and 20, respectively). The reversed interactions, where the acceptor is the base atom and the H[C] donor the partner atom, have smaller R_F values, some not significantly greater than 1 or significantly smaller than this value. For example, the R_F values for Br[-]…H[C] and Cl[-]…H[C] are 1.2 and 0.6, respectively. This is because a strong acceptor is a very favourable option for carbon-bound hydrogen, but a C-H donor is less favourable than O-H or N-H from the acceptor’s point of view. Perhaps the most striking result in Table 4 is the high ranking of H[C]…F and F…H[C] (ranks 18, 21), with H[C]…Cl and Cl…H[C] not too far behind (27, 38). These interactions are known to be weak, especially the latter.[57] In contrast, many atom pairs associated with halogen bonding fall relatively low in the table, *e.g.* R_F(Cl…O) is 1.0.

C[unsat]…H[C] is the only interaction of unsaturated carbon with R_F>1 (R_F=1.3). The reversed interaction, H[C]…C[unsat], has an R_F value of only 1.0; C-H groups prefer to hydrogen bond to halides, oxygen or nitrogen rather than interact with π systems. The R_F values of C[unsat]…halogen are 1.0, 1.0, 0.9 and 0.9 for F, Cl, Br and I, respectively, *i.e.* the C-halogen…π interactions do not occur more often than expected by chance. This is somewhat at variance with the reported utility of C-F…π interactions as synthons, or of C-Cl…π and C-Br…π as stabilizing interactions in protein-ligand complexes, although these interactions may perhaps be favoured in special circumstances.

The relatively high rankings of N[nonacc]…I[-] and N[nonacc]…Br[-] are partly artefactual. N[nonacc] is non-acceptor nitrogen and the SMARTS strings used to define it include groups such as $–NH_3^+$. While this is correct, since ammonio nitrogen is not a hydrogen-bond acceptor, it does result in primary interactions which are by-products of hydrogen-bonding, *e.g.* Fig. 6. Similarly, the high ranking of C[sat]…F is at least partly due to interactions in which the fluorine atom is “seen” by the carbon atom but also makes close (and presumably more stabilizing) interactions to hydrogen atoms, e.g. Fig. 7.

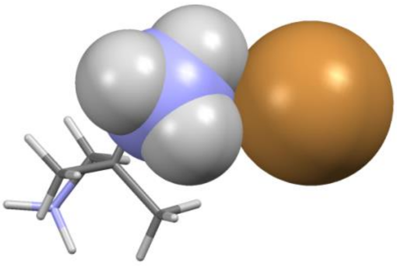

Fig. 6 CSD entry POPMOW, with atoms involved in short $–NH_3^+$…bromide contact shown in space-filling style. This is an N[nonacc]…Br primary interaction but is stabilized by N-H…Br hydrogen bonding.

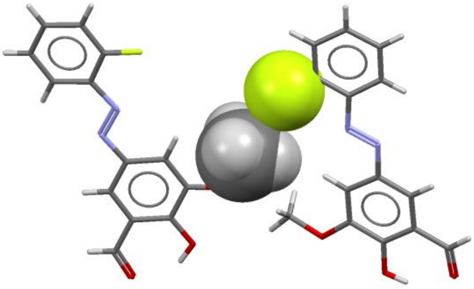

Fig. 7 CSD entry YAYVAW, with atoms involved in short methyl…F contact shown in space-filling style.

Effect of including secondary interactions

It was felt worthwhile to investigate the effect of including more than one interaction per base atom. Therefore, the analysis described in the preceding section was repeated, but allowing up to two interactions for each base atom, *viz*. the primary interaction and the second shortest interaction relative to the sum of vdw radii (*i.e.* the interaction with the second smallest value of Δ_{AB} in equation 1). The latter, which will be termed the *secondary interaction*, was still required to be shorter than the sum of vdw radii plus 1Å. The quantity N in equation (4) now becomes the total number of primary and secondary interactions formed by base atoms of type T_B, and $O(T_B,T_A)$ in equation (5) becomes the observed number of $T_B…T_A$ primary and secondary interactions in the ith structure.

In the ensuing discussion, R_F refers to a value from the original analysis, R_{F2} to a value from the new analysis, and results are based on interactions that have the potential to occur in at least 500 structures. Overall, R_F and R_{F2} correlate well (Spearman $r_S = 0.96$) and 76% of R_F,R_{F2} pairs are within 0.2 of each other. However, there is a strong tendency (statistically significant at the 99.9% confidence level) for R_{F2} to be closer to 1 than the corresponding R_F. Thus, of 70 interactions with $R_F \geq 1$, there are 63 with $R_F > R_{F2}$. The tendency was less obvious for the 124 interactions with $R_F < 1$, but still highly significant, some 84 having $R_F < R_{F2}$. The interactions for which the effect is largest are almost all strong hydrogen and halogen bonds. For example, R_F and R_{F2} for Br[-]…H[polar] are 10.9 and 8.3, respectively. Corresponding values for I…N[acc] are 5.4 and 3.0. The few interactions that go against the trend mostly involve H[C], especially when halogens or halides are involved. For example, Br…H[C] has R_F=1.1, R_{F2}=1.3, and Cl…H[C] has R_F=1.3, R_{F2}=1.4. Visual inspection of example structures suggested that this is because non-hydrogen atoms often make contacts to more than one of the hydrogen atoms on a $–CH_3$ or $–CH_2$ moiety, or to adjacent hydrogen atoms on an aromatic ring.

The results strongly suggest that secondary interactions are more likely to be due to random packing effects (hence, have R_F nearer to 1) than primary interactions, which tend to express the inherent preferences of the base atom. In the case of the strong hydrogen bonds, for example, the acceptor contact to H[polar] is the dominant interaction, and the nature of the secondary interaction is often artefactual, depending, for example, on which atoms are nearby in the donor molecule. Basing the study solely on primary interactions therefore seems to offer the best chance of focussing on interactions that are significantly influencing crystal packing rather than being artefactual. Only primary interactions were used in the remainder of the study.

Results based on fine atom types

Although the coarse atom-typing scheme is sufficient to reveal many interesting trends, it has limitations. For example, it does not discriminate between carbon-bound halogen and groups such as PF_6^-, Br_3^-, *etc*. It assigns all sulfur atoms to a single category, although it is known that the nonbonded interactions of this element change depending on whether it is in an electron-donating or -withdrawing environment.[58,59] All phosphorus atoms are classified together, but the behaviour of 3-coordinate phosphorus must surely be very different from that of the 4-coordinate atom. The investigation was therefore continued using the more discriminating atom-typing scheme of Table 2 for base atoms. The coarse scheme (Table 1) was retained for partner atoms, except that phosphorus atoms were subdivided into P[3-coord] and P[other] for their interactions with phosphorus base atoms. Statistics were computed for selected combinations of base-atom and partner-atom types. Full details of the results are deposited as supplementary material. Table 5 summarises R_F values and their 95% confidence intervals for selected interactions. All results in the table are based on at least 500 structures except where stated. The interactions for each base atom type are sorted into descending order of R_F.

The top third of the table presents data on various types of hydrogen bond. The first six lines relate to different types of C-H…O interactions. It is well known that the ability of a C-H group to donate a hydrogen bond correlates with its acidity,[9] and the R_F values are in line with this. Thus, the relatively acidic acetylenic CH has the highest R_F value, a remarkable 5.2. Even the weakest of the CH donors studied, methyl hydrogen, has R_F significantly greater than 1 (*viz.* 1.5). The next few lines of the table show the R_F values of interactions between four different types of oxygen and two types of hydrogen. As expected, the strongest hydrogen-bond acceptor, carbonyl oxygen, has the greatest preference for H[polar], followed by O[hydroxyl], O[ether] and O[conj]. The latter type corresponds to two-coordinate oxygen atoms conjugated to double or aromatic bonds (*e.g.* in furans or esters). It forms primary interactions to H[polar] less often than expected at random, and much prefers interacting with H[C]. This is consistent with previous studies.[60,61]

The middle section of the table lists all carbon-bound halogen interactions that have R_F>0.5 (but remembering that some interactions are missing because they failed to meet the 500 structures criterion). Of the organic fluorine interactions, F[C]…H[C] has by far the highest R_F (1.7). In contrast, R_F(F[C]…H[polar]) is only 0.6, confirming that organic fluorine rarely accepts hydrogen bonds from O-H and N-H donors.[62] Cl[C]…H[C] occurs significantly more often than expected by chance, though the R_F confidence interval is wide (1.0-1.6). Br[C]…H[C] interactions cannot be shown to occur more often than expected at random, while the R_F value of I[C]…H[C] is significantly less than 1. It should be remembered, however, that inclusion of secondary interactions would undoubtedly raise halogen…H[C] R_F values (see preceding section). The homonuclear halogen…halogen interactions all occur about as often as expected by chance. Three of the heteronuclear interactions, *viz*. Cl[C]…F, Cl[C]…Br and Br[C]…Cl, have R_F significantly above 1, probably indicating the effects of halogen bonding. Of the other potentially halogen-bonding interactions, only I[C]…N[acc] and I[C]…O have R_F values above 2. The corresponding bromine interactions have R_F of 1.8 and 1.5, respectively, both values being significantly greater than 1. Cl[C]…O and Cl[C]…N[acc] interactions occur about as often as expected by chance.

The bottom part of Table 5 lists interactions of two types of sulfur and of 3-coordinate phosphorus. The dramatic difference in the interaction preferences of electron-rich and electron-deficient sulfur is apparent. Thiocarbonyl sulfur strongly prefers the hydrogen bond, S[=C]…H[polar] having the very high R_F value of 9.7. In contrast, electron deficient sulfur abhors close contacts to polar hydrogen but has a strong preference for oxygen (R_F=1.7), indicative of chalcogen bonding. Its propensity to interact with N[acc] is much lower (R_F=1.2). The highest-ranked interaction for P[3-coord] is the hydrogen bond to H[polar], the only other interaction with R_F significantly greater than 1 being P[3-coord]…H[C].

Polarisation and halogen bonding

While the results in Table 5 clearly show the importance of halogen bonds involving iodine, the R_F values of other potential halogen-bonding interactions are only moderate and sometimes no higher than would be expected by chance. Hitherto, however, all carbon-bound halogen atoms of a given atomic number have been assigned to the same atom type. This is likely to obscure the important influence of polarisation on the ability of a chlorine or bromine atom to form a halogen bond. A further analysis was therefore performed, in which the atom types of these elements were subdivided as shown in Table 6 for base atoms. The coarse-atom typing scheme was used for partner atoms. Selected results are presented in Table 7. In view of the shortage of data for some of the new atom types, interactions are included in this table provided there were at least 200 structures in which they could theoretically occur, rather than the minimum of 500 used earlier. Additionally, results for four interactions of particular interest are included even though they were based on ≤50 structures. The possibility of systematic error in the results is, of course, correspondingly higher.

Interactions in Table 7 are sorted first by partner-atom type, then by the element type of the base atom, then into descending order of R_F. Some types of highly polarised chlorine and bromine have relatively high R_F values (>2) for interactions with O and N[acc], but atoms in less electron-withdrawing environments have $R_F \leq 1$, *e.g.* Cl[Car]…O and Br[CX_3Br]…O. The expected tendency for R_F to rise as the halogen becomes more polarised is exemplified by the Cl…O interactions, R_F decreasing in the order Cl[N] > Cl[CN_2] > Cl[ClC=CCl] > Cl[$CXCl_3$] > Cl[CX_2Cl_2] > Cl[Car] > Cl[CX_3Cl].

Differences can be significant when confidence limits overlap

It is safe to assume that two interactions have significantly different R_F values if their 95% confidence intervals do not overlap, but several pairs of interactions in the Cl…O series discussed immediately above have overlapping confidence intervals. However, this does not necessarily mean that their R_F values are not significantly different.[63] Additional calculations were performed to demonstrate this. Taking the pair Cl[$CXCl_3$]…O and Cl[CX_2Cl_2]…O as an example, 100,000 estimates of their R_F difference:

$$\Delta R_F = R_F(\text{Cl[CXCl}_3\text{]…O}) - R_F(\text{Cl[CX}_2\text{Cl}_2\text{]…O})$$

were made from each set of vdw radii in turn. Each estimate was based on random samples of the structures containing atoms of types Cl[$CXCl_3$] and O, and of those containing Cl[CX_2Cl_2] and O. These were selected by sampling with replacement, as in bootstrapping. If n_B, n_{RT} and n_A are the number of estimates of ΔR_F that were positive using the Bondi, R&T and Alvarez radii respectively, the probability that Cl[$CXCl_3$] …O has the higher R_F value can be estimated as $(n_B+n_{RT}+n_A)/300000$. The resulting probability exceeds 0.99, *i.e.* is highly significant. Results for other interaction pairs in the series whose confidence intervals overlap are summarised in Table 8. Not all differences have probability>0.95, so the order as a whole is not statistically significant, but it is more likely to be correct than not.

Comparison of halogen bonds with C-H…F interactions

Carbon-bound fluorine atoms were divided into two sub-types, those that belong to -CF_3 groups, F[CF_3], and those that do not, F[$C \neq CF_3$]. The following R_F values were then determined (95% confidence intervals in brackets): R_F(H[C]…F[CF_3])=1.6 (1.5-1.6); R_F(F[CF_3]…H[C])=1.5 (1.3-1.7); R_F(H[C]…F[C≠CF_3])=1.9 (1.9-2.0); R_F(F[C≠CF3]…H[C])=1.9 (1.6-2.0). All values are based on >500 structures. The results show that -CF_3 fluorine atoms are less likely to form interactions with H[C] than other C-F groupings, as might be expected on electrostatic grounds. The R_F values are comparable to, or higher than, many of the halogen-bonding interactions in Table 7. For example, R_F(H[C]…F[C≠CF_3]) is significantly higher than R_F(Br[Car]…O). At first sight, this is surprising. The calculated interaction energy of C-H…F is small;[64,65] for example, ΔE for H_3C-H…F-CH_3 has been calculated as -0.43 kcal mol^{-1}.[29] In contrast, ΔE for bromobenzene…acetone was calculated as -2.23 kcal mol^{-1}.[66] Even the weak halogen bonds H_2C=O…Cl-CH_3 and H_2C=O…Br-CH_3 have ΔE of -1.18 and -1.64 kcal mol^{-1}, respectively.[67]

Two explanations for the paradox may be suggested. Firstly, there is a limited range of favourable options available to a carbon-bound fluorine atom in a crystal structure. The atom must be close to something and, in many structures, a primary interaction to H[C] will be the best possible option. Interactions such as F[C]…O and F[C]…C[unsat] will usually be electrostatically unfavourable, and although the energy of F[C]…H[polar] is favourable (ΔE=-2.80 kcal mol^{-1}),[62] polar hydrogen atoms are likely to be tied up in hydrogen bonds to stronger oxygen or nitrogen acceptors. Another possible factor is that C-H…F interactions have only weak directional preferences, so, for a given molecule, there may be many packing arrangements that give rise to stable interactions of these types. In contrast, there may be far fewer ways to pack molecules so as to respect the highly directional requirements of halogen bonds.

Interactions of aromatic systems

Table 9 summarises R_F values for selected interactions in which the base atoms are the carbon atoms of phenyl groups (specifically, Csp^3-C_6H_5) and pentafluorophenyl groups (specifically, X-C_6F_5, X = H, C, N, O, F, P, S, Cl, Br, I). Coarse atom types were used for the partner atoms. Results for each base atom are given in descending order of R_F. The final column of the table gives the number of structures in which each interaction can theoretically occur; it will be seen that several results are based on rather few data. The table includes, *inter alia*, all interactions with R_F >1.

By far the highest-ranked interaction of phenyl carbon is $C[C_6H_5]…H[C]$, with an R_F value of 1.6. The R_F values of $C[C_6H_5]…C[unsat]$ and $C[C_6H_5]…H[polar]$ are 0.6 and 0.4, respectively, suggesting that π-stacking and O-H…π and N-H…π hydrogen bonding occur much less often than expected by chance. There is scant support for the hypothesis that halogens can interact favourably with the π systems of phenyl groups. The results for pentafluorophenyl are dramatically different from those of phenyl. The highest-ranked interactions are with electronegative atoms, including halide ions. Whether it is reasonable to refer to these as anion-π interactions is debatable,[40] but it is clear that $C[C_6F_5]…[Br-]$ and $C[C_6F_5]…[Cl-]$ occur more often than expected by chance.

Other interactions

Paulini and co-workers have discussed the potential importance of orthogonal dipolar interactions in stabilizing crystal lattices and biological complexes.[68] The R_F values of two example interactions, C-F…C=O and C-Cl…C=O, were therefore determined. The results (all based on >500 structures and with 95% confidence limits in brackets) were: R_F(Cl[C]…C[carbonyl])=1.3 (0.9-1.7); R_F(Cl[C]…O[carbonyl])=1.0 (0.8-1.2); R_F(F[C]…C[carbonyl])=1.0 (0.8-1.3); R_F(F[C]…O[carbonyl])=0.4 (0.3-0.6). None of the values is significantly greater than 1, although the first approaches significance. Thus, there is insufficient evidence to confirm the general importance of these interactions in stabilizing crystal structures.

In view of the difficulties of interpreting R_F values of interactions involving N[nonacc] in Table 4 (see Fig. 6), a more detailed analysis was performed of various types of nitrogen atoms that cannot act as hydrogen-bond acceptors. High (and statistically significant) R_F

values were found for the interactions between N[unsat] and the halide ions (2.2, 1.8 and 1.7 for iodide, bromide and chloride, respectively), where N[unsat] is defined by the SMARTS strings [NX3H0]=A and [nX3H0]. These corresponded to nitrogen in groups such as nitro and pyridinium. Fig. 8 shows two examples of these interesting and presumably favourable interactions.

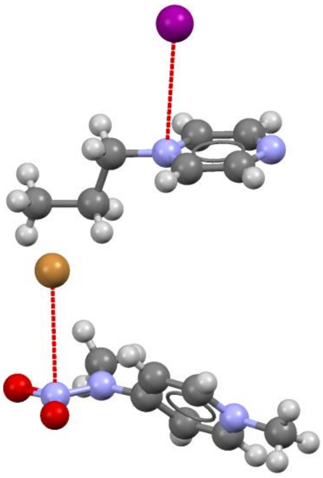

Fig. 8 Iodide…nitrogen and bromide…nitrogen interactions in CSD entries XENGOO and HIQYAH, respectively.

Conclusions

R_F has been defined as the ratio of the observed number of interactions of a given type in crystal structures to the number expected if determined solely by surface-area considerations. If the R_F value for an interaction, B…A, is significantly greater than 1, it indicates that atoms of type B form more interactions with atoms of type A than would be expected by chance. The relationship between R_F and the calculated energy of the B…A interaction is not necessarily simple. Interaction energies are usually calculated relative to the energy of the system with the interacting moieties separated at infinity. R_F(B…A), on the other hand, expresses the preference of B for B…A interactions relative to other types of interactions that B might form in crystal structures. A perfect correlation between R_F and interaction energies is therefore not to be expected. For example, the R_F value for C-H…F is higher than that of

Br…O, although the latter is undoubtedly the stronger interaction. R_F has the advantage of being a direct measure of what actually happens in practice.

The method for estimating R_F is objective and depends on no parameters other than vdw radii, its sensitivity to which has been well characterised. It exploits the huge number of well-resolved crystal structures in the CSD. For most of the interactions studied, the numbers of structures used to derive results were sufficiently large as to effectively preclude any danger of systematic errors due to sample bias. Random sampling uncertainties have been assessed by bootstrapping and the resulting confidence intervals validated against results from an alternative statistical approach. The confidence intervals have then been increased to take vdw radii uncertainties into account. The line-of-sight approach focusses on the interactions most likely to reveal the inherent preferences of the interacting atoms.

When interactions are ranked on their R_F values, those at the top of the list are the strong hydrogen bonds, providing a degree of validation for the methodology. Somewhat surprisingly though, even these only have R_F values in the approximate range 3 to 10. A key question posed by Dunitz and Gavezzotti[41] has been answered. As they pointed out, C-H…X contacts are more or less inevitable in organic crystal structures, but it has been established that they occur significantly more often than would be expected by chance for, *inter alia*, X = O, F, Cl, and N when it has an available lone pair. In particular, C-H…F and C-H…Cl interactions occur surprisingly frequently, given their weak energies. The effects of polarisation on the frequencies of occurrence of various types of halogen bonds are clearly seen, as are the dramatically different interactions of phenyl and pentafluorophenyl groups. The method challenges some previously published suggestions. For example, halogen…C[unsat], F…F, C-F…carbonyl and C-Cl…carbonyl interactions have R_F values sufficiently low that their general importance in stabilising lattices of organic molecules must be doubted.

An interaction with a low R_F value is not necessarily energetically unfavourable (after all, the majority of intermolecular interactions observed in a crystal structure are likely to be stabilizing because crystals studied by X-ray diffraction are usually stable entities). But it does mean that it tends to be out-competed by other interactions. For that reason, if R_F(B…A) is less than or not much greater than 1, it should serve as a warning against the indiscriminate assumption that any short B…A interaction necessarily plays a significant role in stabilizing a crystal structure. Or, indeed, any stabilizing role at all: in organic structures, the most important contributor to lattice energy is close-packing,[69] and this will sometimes result in contacts that are present simply because close packing requires them to be.

Acknowledgements

The author thanks Peter Galek and Patrick McCabe of the Cambridge Crystallographic Data Centre for thought-provoking questions. Colin Groom, of the same organisation, is particularly thanked for a critical reading of the manuscript and the suggestion proposed in the last two sentences of the section on halogen bonds compared with C-H…F interactions. Anne Taylor is thanked for helpful comments on what the author erroneously thought to be the final draft. Spearman correlation coefficients were calculated with the web application http://scistatcalc.blogspot.co.uk/2013/10/spearman-rank-correlation-calculator.html.

References

1. J. M. Robertson, Proc. Roy. Soc. Lond. A, 1936, 157, 79-99.
2. D. J. Sutor, Nature, 1962, 195, 68-69.
3. D. J. Sutor, J. Chem. Soc., 1963, 1105-1110.
4. J. Donohue, in *Structural Chemistry and Molecular Biology*, ed. A. Rich and N. Davidson, W. H. Freeman, San Francisco, 1968, pp. 459-463.
5. G. A. Jeffrey and H. Maluszynska, *Int. J. Biol. Macromol.*, 1982, **4**, 173-185.
6. R. Taylor and O. Kennard, *J. Am. Chem. Soc.*, 1982, **104**, 5063-5070.
7. G. R. Desiraju, *J. Chem. Soc., Chem. Commun.*, 1989, 179-180.
8. G. R. Desiraju, *J. Chem. Soc., Chem. Commun.*, 1990, 454-455.
9. G. R. Desiraju, *Acc. Chem. Res.*, 1991, **24**, 290-296.

10. V. R. Pedireddi and G. R. Desiraju, *J. Chem. Soc., Chem. Commun.*, 1992, 988-990.
11. T. Steiner, *Cryst. Rev.*, 1996, **6**, 1-57.
12. T. Steiner, *Chem. Commun.*, 1997, 727-734.
13. C. H. Schwalbe, *Cryst. Rev.*, 2012, **18**, 191-206.
14. F. A. Cotton, L. M. Daniels, G. T. Jordan IV and C. A. Murillo, *Chem. Commun.*, 1997, 1673-1674.
15. T. Steiner and G. R. Desiraju, *Chem. Commun.*, 1998, 891-892.
16. E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.*, 2011, **83**, 1637-1641.
17. N. Ramasubbu, R. Parthasarathy and P. Murray-Rust, *J. Am.Chem. Soc.*, 1986, **108**, 4308-4314.
18. J. A. R. P. Sarma and G. R. Desiraju, *Acc. Chem. Res.*, 1986, **19**, 222-228.
19. P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, *Proc. Nat. Acad. Sci. USA*, 2004, **101**, 16789-16794.
20. P. Metrangolo, H. Neukirch, T. Pilati and G. Resnati, *Acc. Chem. Res.*, 2005, **38**, 386-395.
21. P. Metrangolo and G. Resnati, *Chem. Eur. J.*, 2001, **7**, 2511-2519.
22. P. Metrangolo and G. Resnati, *Cryst. Growth. Des.*, 2012, **12**, 5835-5838.
23. C. B. Aakeröy, M. Baldrighi, J. Desper, P. Metrangolo and G. Resnati, *Chem. Eur. J.*, 2013, **19**, 16240-16247.
24. S. Sirimulla, J. B. Bailey, R. Vegesna and M. Narayan, *J. Chem.Inf. Model.*, 2013, **53**, 2781-2791.
25. A. Bauzá, D. Quiñonero, P. M. Deyà and A. Frontera, *CrystEngComm*, 2013, **15**, 3137-3144.
26. P. Metrangolo, J. S. Murray, T. Pilati, P. Politzer, G. Resnati and G. Terraneo, *Cryst. Growth Des.*, 2011, **11**, 4238-4246.

27. P. Metrangolo, J. S. Murray, T. Pilati, P. Politzer, G. Resnati and G. Terraneo, *CrystEngComm.*, 2011, **13**, 6593-6596.

28. V. R. Thalladi, H. –C. Weiss, D. Bläser, R. Boese, A. Nangia and G. R. Desiraju, *J. Am. Chem. Soc.*, 1998, **120**, 8702-8710.

29. E. D’Oria and J. J. Novoa, *CrystEngComm*, 2008, **10**, 423-436.

30. A. R. Choudhury and T. N. Guru Row, *Cryst. Growth Des.*, 2004, **4**, 47-52.

31. K. Reichenbächer, H. I. Süss and J. Hulliger, *Chem. Soc. Rev.*, 2005, **34**, 22-30.

32. D. Chopra, *Cryst. Growth Des.*, 2012, **12**, 541-546.

33. C. B. Aakeröy, T. A. Evans, K. R. Seddon and I. Pálinkó, *New J. Chem.*, 1999, 145-152.

34. H. Matter, M. Nazeré, S. Güssregen, D. W. Will, H. Schreuder, A. Bauer, M. Urmann, K. Ritter, M. Wagner and V. Wehner, *Angew. Chem. Int. Ed.*, 2009, **48**, 2911-2916.

35. Y. Umezawa and M. Nishio, *Bioorg. Med. Chem.*, 1998, **6**, 493-504.

36. A. Matsumoto, T. Tanaka, T. Tsubouchi, K. Tashiro, S. Saragai and S. Nakamoto, *J. Am. Chem.Soc.*, 2002, **124**, 8891-8902.

37. D. A. Dougherty, *Science*, 1996, **271**, 163-168.

38. J. C. Ma and D. A. Dougherty, *Chem. Rev.*, 1997, **97**, 1303-1324.

39. P. de Hoog, P. Gamez, I. Mutikainen, U. Turpeinen and J. Reedijk, *Angew. Chem.*, 2004, **116**, 5939-5941.

40. B. P. Hay and R. Custelcean, *Cryst. Growth Des.*, 2009, **9**, 2539-2545.

41. J. D. Dunitz and A. Gavezzotti, *Chem. Soc. Rev.*, 2009, **38**, 2622-2633.

42. J. D. Dunitz and A. Gavezzotti, *Angew. Chem. Int. Ed.*, 2005, **44**, 1766-1787.

43. J. P. M. Lommerse, A. J. Stone, R. Taylor and F. H. Allen, *J. Am. Chem. Soc.*, 1996, **118**, 3108-3116.

44. F. H. Allen, *Acta Cryst.*, 2002, B**58**, 380-388.

45. C. Bilton, F. H. Allen, G. P. Shields and J. A. K. Howard, *Acta Cryst.*, 2000, B**56**, 849-856.

46. F. H. Allen, P. A. Wood and P. T. A. Galek, *Acta Cryst*., 2013, B**69**, 379-388.

47. P. T. A. Galek, L. Fábián, W. D. S. Motherwell, F. H. Allen and N. Feeder, *Acta Cryst*., 2007, B**63**, 768-782.

48. R. Taylor and O. Kennard, *Acta Cryst*., 1983, B**39**, 133-138.

49. F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor, *J. Chem. Soc. Perkin Trans. II*, 1987, S1-S19.

50. T. Steiner, *Angew. Chem. Int. Ed*., 2002, **41**, 48-76.

51. SMARTS – A Language for Describing Molecular Patterns; Daylight Chemical Information Systems, Inc.: Laguna Niguel, CA, USA; http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.

52. L. Infantes and S. Motherwell, *Struct. Chem.*, 2004, **15**, 173-184.

53. A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441-451.

54. R. Taylor, *J. Chem. Inf. Model*., 2011, **51**, 897-908.

55. R. S. Rowland and R. Taylor, *J. Phys. Chem*., 1996, **100**, 7384-7391.

56. S. Alvarez, *Dalton Trans*., 2013, **42**, 8617-8636.

57. A. Nangia, *CrystEngComm*, 2002, **4**, 93-101.

58. F. H. Allen, C. M. Bird, R. S. Rowland and P. R. Raithby, *Acta Cryst*., 1997, B**53**, 680-695.

59. F. T. Burling and B. M. Goldstein, *Acta Cryst*., 1993, B**49**, 738-744.

60. H. -J. Böhm, S. Brode, U. Hesse and G. Klebe, *Chem. Eur. J*., 1996, **2**, 1509-1513.

61. J. P. M. Lommerse, S. L. Price and R. Taylor, *J. Comput. Chem.*, 1997, **18**, 757-774.

62. J. D. Dunitz and R. Taylor, *Chem. Eur. J.*, 1997, **3**, 89-98.

63. M. J. Knol, W. R. Pestman and D. E. Grobbee, *Eur. J. Epidemiol*., 2011, **26**, 253-254.

64. J. A. K. Howard, V. J. Hoy, D. O'Hagan and G. T. Smith, *Tetrahedron*, 1996, **52**, 12613-12622.

65. J. D. Dunitz, A. Gavezzotti and W. B. Schweizer, *Helv. Chim. Acta*, 2003, **86**, 4073-4092.

66. P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7748-7757.
67. K. E. Riley and P. Hobza, *J. Chem. Theor. Comput.*, 2008, **4**, 232-242.
68. R. Paulini, K. Müller and F. Diederich, *Angew. Chem. Int. Ed.*, 2005, **44**, 1788-1805.
69. A. I. Kitaigorodsky, Molecular Crystals and Molecules, Academic Press, New York, 1973, p.6.

Table 1 Coarse atom-typing scheme

SMARTS[a]	Symbol	Description
[C,c]**H**	H[C]	Carbon-bound hydrogen
[N,n,O,o,F,P,p,S,s,Cl,Br,I]**H**	H[polar]	Polar hydrogen
[**C**X4]	C[sat]	Saturated (sp^3) carbon
[**C**!X4], **c**	C[unsat]	Unsaturated (sp^2 or sp) carbon
[**N**X3]-A=A, [**N**X3]-aa, [**N**X3]=A, [**n**X3], [**N**X4], [NX2]=[**N**X2]=[NX1], [NX2][**N**X2]#[NX1], [**N**X2]#[CX1]	N[nonacc]	Nitrogen not able to accept hydrogen bond (has delocalised or no lone pair)
[**N**X1], [**N**X2], [**n**X2], [**N**X3]	N[acc]	Nitrogen able to accept hydrogen bond (has localised lone pair)
[**O,o**]	O	Any oxygen
[**F**X0]	F[-]	Fluoride
[**F**!X0]	F	Unionised fluorine
[**P,p**]	P	Any phosphorus
[**S,s**]	S	Any sulfur
[**Cl**X0]	Cl[-]	Chloride
[**Cl**!X0]	Cl	Unionised chlorine
[**Br**X0]	Br[-]	Bromide
[**Br**!X0]	Br	Unionised bromine
[**I**X0]	I[-]	Iodide
[**I**!X0]	I	Unionised iodine

[a] Type is assigned to atom shown in bold. If type has >1 SMARTS, a match against any is sufficient. If atom matches SMARTS in >1 row, type assigned is that corresponding to highest row in table.

Table 2 Fine atom-typing scheme

SMARTS[a]	Symbol	Description
C[CX4H3]**H**	H[methyl]	Methyl hydrogen
[CX4H2](C)(C)**H**	H[methylene]	Methylene hydrogen
[CX4H1](C)(C)(C)**H**	H[methine]	Hydrogen on tertiary carbon
c**H**	H[Car]	Hydrogen on aromatic carbon
C#[CX2]**H**	H[Csp]	Hydrogen on acetylenic carbon
[NX4][CX4]**H**	H[CNX$_3^+$]	Hydrogen on sp^3 carbon alpha to cationic 4-coordinate nitrogen
O=C[**O**X2], C=C**O**C=C	O[conj]	2-coordinate oxygen in conjugated environment, *e.g.* in furan, esters.
[CX4][**O**X2H1]	O[hydroxyl]	Hydroxyl oxygen
[CX4][**O**X2][CX4]	O[ether]	Aliphatic ether oxygen
C=[**O**X1], C[**O**X1]	O[carbonyl]	Carbonyl oxygen (also includes carboxylate)
[c,C]-[**F**X1]	F[C]	Carbon-bound fluorine
[**P**X3]	P[3-coord]	3-coordinate phosphorus
P	P[other]	Other phosphorus
C=[**S**X1], C[**S**X1]	S[=C]	Thiocarbonyl sulfur (also includes thiolate, thiocyanate, *etc.*)
N=[C,N][**S**X2][C,N], [C,N]=N[**S**X2][C,N], [SX2]=N[**S**X2][C,N], [SX2][**S**X2]N=[C,N], [SX2][**S**X2]N=[SX2], [c,n][**s**X2]n, N[**S**X2][SX2], c[**s**X2][sX2]	S[δ+]	Sulfur in electron-withdrawing environment, *e.g.* in thiazole, isothiazole
[c,C]-[**Cl**X1]	Cl[C]	Carbon-bound chlorine

[c,C]-[**Br**X1]	Br[C]	Carbon-bound bromine
[c,C]-[**I**X1]	I[C]	Carbon-bound iodine

[a] See footnote to Table 1.

Table 3 vdw radii (Å)

Element	Bondi[53]	Rowland and Taylor[55]	Alvarez[56]
H	1.20	1.10	1.20
C	1.70	1.77	1.77
N	1.55	1.64	1.66
O	1.52	1.58	1.50
F	1.47	1.46	1.46
P	1.80	1.90[a]	1.90
S	1.80	1.81	1.89
Cl	1.75	1.76	1.82
Br	1.85	1.87	1.86
I	1.98	2.03	2.04

[a] Alvarez radius used as R&T did not determine a value.

Table 4 R_F values with confidence intervals and Δ_{mean} for interactions classified using the coarse atom-typing scheme for both base and partner atoms

Base type	Partner type	R_F	95% interval	Rank of R_F
Br[-]	H[polar]	10.9	9.0-13.8	1
Cl[-]	H[polar]	9.0	7.5-11.0	2
O	H[polar]	5.8	5.1-6.9	3
N[acc]	H[polar]	5.7	4.9-6.9	4
I	N[acc]	5.4	4.7-6.3	5
H[polar]	Cl[-]	4.9	4.7-5.2	6
H[polar]	Br[-]	4.8	4.5-5.0	7
H[polar]	N[acc]	4.1	3.8-4.4	8
H[polar]	O	3.8	3.4-4.1	9
S	H[polar]	2.9	2.5-3.4	10
I	O	2.8	2.6-3.1	11
H[C]	Br[-]	2.7	2.6-2.8	12
H[C]	I[-]	2.7	2.6-2.8	13
H[C]	Cl[-]	2.6	2.5-2.6	14
H[C]	O	2.0	1.9-2.0	15
N[nonacc]	I[-]	2.0	1.7-2.2	16
N[acc]	I	1.8	1.7-2.0	17
H[C]	F	1.8	1.8-1.9	18
Br	N[acc]	1.8	1.5-2.0	19
H[C]	N[acc]	1.8	1.7-1.8	20
F	H[C]	1.7	1.5-1.9	21
I[-]	H[C]	1.7	1.5-2.0	22

N[acc]	H[C]	1.7	1.6-1.8	23
O	H[C]	1.7	1.6-1.7	24
N[nonacc]	Br[-]	1.7	1.5-1.9	25
H[polar]	S	1.6	1.5-1.8	26
H[C]	Cl	1.6	1.5-1.6	27
P	H[C]	1.5	1.5-1.6	28
Br	Cl	1.5	1.2-1.8	29
C[sat]	F	1.5	1.5-1.6	30
Br	O	1.5	1.3-1.7	31
H[C]	Br	1.5	1.4-1.5	32
I	S	1.5	1.2-1.8	33
Cl	Br	1.5	1.2-1.9	34
H[C]	I	1.4	1.3-1.4	35
C[sat]	Cl	1.4	1.3-1.4	36
S	H[C]	1.3	1.2-1.5	37
Cl	H[C]	1.3	1.0-1.6	38
Cl	F	1.3	1.1-1.5	39
C[unsat]	H[C]	1.3	1.2-1.4	40
H[C]	S	1.3	1.2-1.3	41
N[nonacc]	Cl[-]	1.3	1.2-1.4	42
N[nonacc]	I	1.3	1.1-1.4	43
C[sat]	O	1.3	1.1-1.5	44
C[sat]	Br	1.3	1.2-1.3	45
O	I	1.2	1.1-1.4	46
N[nonacc]	O	1.2	1.2-1.3	47
I	H[polar]	1.2	0.9-1.6	48
N[nonacc]	F	1.2	1.1-1.3	49

C[sat]	Cl[-]	1.2	1.0-1.4	50
F	Cl	1.1	0.9-1.3	59
Br	H[C]	1.1	0.8-1.3	60
Br	Br	1.1	1.0-1.2	61
C[sat]	H[C]	1.1	0.9-1.2	63
Cl	Cl	1.0	1.0-1.1	67
F	F	1.0	0.9-1.1	70
Cl	O	1.0	0.8-1.2	75
C[unsat]	C[unsat]	1.0	0.9-1.1	78
Cl[-]	Cl[-][a]	0.0	0.0-0.0	193

[a] Equal bottom rank with Br[-]…Br[-] and I[-]…I[-].

Table 5 R_F values and their confidence intervals for selected interactions classified using the fine atom-typing scheme of for base atoms and the coarse scheme for partner atoms

Base type	Partner type	R_F	95% interval
H[Csp]	O	5.2	4.5-5.9
H[CNX_3^+]	O	2.7	2.5-2.9
H[Car]	O	2.3	2.2-2.3
H[methine]	O	2.0	2.0-2.1
H[methylene]	O	1.6	1.6-1.7
H[methyl]	O	1.5	1.5-1.6
O[carbonyl]	H[polar]	9.2	7.9-11.2
O[carbonyl]	H[C]	1.7	1.6-1.8
O[hydroxyl]	H[polar]	5.9	5.2-7.1
O[hydroxyl]	H[C]	1.2	1.1-1.3
O[ether]	H[polar]	4.0	3.6-4.7
O[ether]	H[C]	1.7	1.6-1.9
O[conj]	H[C]	1.7	1.6-1.7
O[conj]	H[polar]	0.9	0.8-1.0
F[C]	H[C]	1.7	1.5-1.9
F[C]	Cl	1.1	0.9-1.3
F[C]	F	1.1	1.0-1.1
F[C]	C[sat]	0.8	0.6-1.1
F[C]	C[unsat]	0.7	0.5-0.9
F[C]	H[polar]	0.6	0.4-0.7
F[C]	N[nonacc]	0.5	0.3-0.7
Cl[C]	Br	1.5	1.2-1.9
Cl[C]	F	1.4	1.2-1.7

Cl[C]	H[C]	1.3	1.0-1.6
Cl[C]	Cl	1.0	1.0-1.1
Cl[C]	O	1.0	0.8-1.2
Cl[C]	N[acc]	0.9	0.7-1.1
Cl[C]	C[unsat]	0.8	0.6-1.0
Cl[C]	C[sat]	0.7	0.5-1.0
Cl[C]	H[polar]	0.6	0.4-0.7
Br[C]	N[acc]	1.8	1.5-2.0
Br[C]	Cl	1.5	1.3-1.8
Br[C]	O	1.5	1.3-1.7
Br[C]	Br	1.1	1.0-1.2
Br[C]	H[C]	1.1	0.8-1.3
Br[C]	C[unsat]	0.7	0.6-0.9
Br[C]	C[sat]	0.6	0.3-0.8
I[C]	N[acc]	6.5	5.7-7.6
I[C]	O	2.7	2.5-3.1
I[C]	I	1.0	0.8-1.1
I[C]	N[nonacc]	0.6	0.3-1.0
I[C]	C[unsat]	0.6	0.5-0.7
I[C]	H[C]	0.6	0.4-0.7
S[=C]	H[polar]	9.7	8.0-11.7
S[=C]	H[C]	1.3	1.2-1.4
S[δ+]	O	1.7	1.5-2.0
S[δ+]	N[acc]	1.2	1.0-1.4
S[δ+]	H[C]	1.0	0.7-1.3
S[δ+]	C[unsat]	0.8	0.6-1.0
S[δ+]	N[nonacc]	0.6	0.2-0.9

S[δ+]	C[sat]	0.5	0.3-0.8
S[δ+]	H[polar]	0.5	0.2-0.7
P[3-coord]	H[polar]	2.6[a]	1.7-3.9
P[3-coord]	H[C]	1.7	1.6-1.8
P[3-coord]	O	1.1	0.8-1.4
P[3-coord]	P[3-coord]	0.7	0.5-0.9
P[3-coord]	N[acc]	0.4	0.1-0.7

[a] Based on 327 structures.

Table 6 Refined atom-typing scheme for chlorine atoms (analogous scheme used for bromine)

SMARTS[a]	Symbol	Description
[**Cl**X1][CX4](Cl)(Cl)	Cl[CXCl$_3$]	Chlorine in carbon tetrachloride, chloroform and other –CCl$_3$ groups.
[**Cl**X1][CX4](Cl)	Cl[CX$_2$Cl$_2$]	Methylene dichloride and other >CCl2 groupings.
[**Cl**X1][CX4]	Cl[CX$_3$Cl]	Methyl chloride and other Csp^3-Cl groupings.
[**Cl**X1][NX3]	Cl[N]	Chlorine bound to 3-coordinate nitrogen
[**Cl**X1][CX3]=[CX3]Cl	Cl[ClC=CCl]	Chlorine on 1,2-dichlorovinyl
[**Cl**X1][CX3](=N)N	Cl[CN$_2$]	Chlorine on sp^2 carbon bound to two nitrogen atoms, *e.g.* 2-position of imidazole
[**Cl**X1]c1[cH1][cH1][cH1][cH1]c1[CX4], [**Cl**X1]c1[cH1][cH1][cH1]c([CX4])[cH1]1, [**Cl**X1]c1[cH1][cH1]c([CX4])[cH1][cH1]1	Cl[Car]	Chlorine substituent on aromatic rings o-, m- and p-Csp^3-C$_6$H$_4$Cl

[a]See footnote to Table 1.

Table 7 R_F values and their confidence intervals for selected interactions classified using the atom-typing scheme of Table 6 for base atoms, and the coarse scheme for partner atoms

Base type	Partner type	R_F	95% interval
Cl[N]	O	3.2[a]	2.4-4.0
Cl[CN_2]	O	2.6[b]	1.3-3.8
Cl[ClC=CCl]	O	1.9	1.5-2.3
Cl[$CXCl_3$]	O	1.4	1.1-1.7
Cl[CX_2Cl_2]	O	0.9	0.6-1.2
Cl[Car]	O	0.8	0.5-1.2
Cl[CX_3Cl]	O	0.6	0.4-0.9
Br[CN_2]	O	2.5[a]	1.4-3.6
Br[Car]	O	1.5	1.2-1.9
Br[CX_3Br]	O	1.0	0.8-1.3
Cl[CN_2]	N[acc]	1.5[c]	0.7-2.1
Cl[$CXCl_3$]	N[acc]	1.2	0.8-1.6
Cl[Car]	N[acc]	0.9	0.4-1.3
Cl[CX_2Cl_2]	N[acc]	0.8	0.3-1.3
Cl[CX_3Cl]	N[acc]	0.5	0.2-0.9
Br[CN_2]	N[acc]	2.9[d]	1.9-4.2
Br[Car]	N[acc]	1.6	0.9-2.4
Br[CX_3Br]	N[acc]	1.5	0.8-2.2
Cl[$CXCl_3$]	Cl	1.0	0.9-1.1
Cl[Car]	Cl	1.0	0.8-1.2
Cl[ClC=CCl]	Cl	1.0	0.9-1.1
Cl[CX_3Cl]	Cl	0.9	0.8-1.0
Cl[CX_2Cl_2]	Cl	0.9	0.8-1.0

Based on [a]38, [b]31, [c]50, [d]34 structures.

Table 8 Statistical significance of differences in R_F values for different types of Cl…O interactions

Interaction A	Interaction B	Probability that $R_F(A) > R_F(B)$
Cl[N]…O	Cl[CN_2]…O	0.81
Cl[CN_2]…O	Cl[ClC=CCl]…O	0.92
Cl[CN_2]…O	Cl[$CXCl_3$]…O	0.99
Cl[ClC=CCl]…O	Cl[$CXCl_3$]…O	1.00
Cl[$CXCl_3$]…O	Cl[CX_2Cl_2]…O	1.00
Cl[$CXCl_3$]…O	Cl[Car]…O	1.00
Cl[CX_2Cl_2]…O	Cl[Car]…O	0.79
Cl[CX_2Cl_2]…O	Cl[CX_3Cl]…O	1.00
Cl[Car]…O	Cl[CX_3Cl]…O	0.98

Table 9 R_F values and their confidence intervals for interactions of phenyl and pentafluorophenyl carbon atoms

Base type	Partner type	R_F	95% interval	Number of structures
$C[C_6H_5]$	H[C]	1.6	1.5-1.7	18985
$C[C_6H_5]$	Br	1.1	1.0-1.4	898
$C[C_6H_5]$	Cl	1.1	0.9-1.2	1448
$C[C_6H_5]$	I	1.0	0.8-1.3	174
$C[C_6H_5]$	F	0.8	0.6-1.0	1061
$C[C_6H_5]$	C[unsat]	0.6	0.5-0.8	18985
$C[C_6H_5]$	H[polar]	0.4	0.4-0.5	9787
$C[C_6H_5]$	O	0.4	0.3-0.6	17241
$C[C_6H_5]$	N[acc]	0.4	0.2-0.6	6379
$C[C_6H_5]$	Cl[-]	0.3	0.1-0.4	378
$C[C_6H_5]$	Br[-]	0.2	0.1-0.4	190
$C[C_6F_5]$	Br[-]	2.5	1.6-3.5	18
$C[C_6F_5]$	Cl[-]	1.9	1.1-3.0	14
$C[C_6F_5]$	O	1.5	1.3-1.8	306
$C[C_6F_5]$	N[acc]	1.2	0.8-1.6	221
$C[C_6F_5]$	F	1.2	1.1-1.3	569
$C[C_6F_5]$	Cl	1.2	0.7-1.7	38
$C[C_6F_5]$	C[unsat]	1.0	0.9-1.1	569
$C[C_6F_5]$	H[C]	0.4	0.2-0.5	487
$C[C_6F_5]$	H[polar]	0.1	0.0-0.3	205