LEARNING FROM OUTLIERS

Carolyn Pratt Brock

Department of Chemistry, University of Kentucky, Lexington, KY 40506-0055, USA

ABSTRACT

Progress in understanding packing in molecular crystals can be made by knowing what is normal and investigating the exceptions. Some studies that illustrate this approach are discussed. (1) In crystals most molecules adopt a conformation near that of the global energy minimum but biphenyls with H atoms in all four *ortho* positions are quite often found in a higher energy, near-planar conformation. This systematic effect occurs because the twisted molecules that would be found in the gas phase or in solution fill space inefficiently. (2) The size of the asymmetric unit is usually minimized but monoalcohols C_nH_mOH and *vic*-diols $C_nH_m(OH)_2$ are likely to crystallize with $Z' > 1$ (more than one molecule in the asymmetric unit) so that good H bonds can be formed. (Another possibility for these alcohols is crystallization in a high-symmetry space group). (3) Inversion symmetry is so favorable for crystal packing that the two enantiomers of a racemic compound are nearly always related by it. Even approximate inversion symmetry is favorable. Diastereomers that would be enantiomers but for the exchange of H and Me substituents often form pseudosymmetric co-crystals even though co-crystals are otherwise nearly always donor/acceptor complexes. Enantiomerically pure materials quite often have $Z' > 1$ structures that mimic inversion symmetry. (4) Short-period translational symmetry is the norm but $Z' > 1$ phases may occur when there is a packing conflict that can be resolved by a modulation, when a crystal is cooled, or when the best packing unit is a $Z' > 1$ aggregate. In structures having $Z' > 5$ at least two of these conditions are usually fulfilled.

1. INTRODUCTION

Crystallographers determine many structures both to further their own research and to help out colleagues. Some of those structures have surprising features that do not fit conventional expectations about packing in molecular crystals [1]. The following is a personal account of how serendipitous observations made while determining, reporting, and reading about individual crystal structures eventually led to studies of wider scope and to insights into the systematics of crystal packing. The resulting papers were all based on searches of the molecular crystal structures archived in the Cambridge Structural Database [2] (hereafter, the CSD).
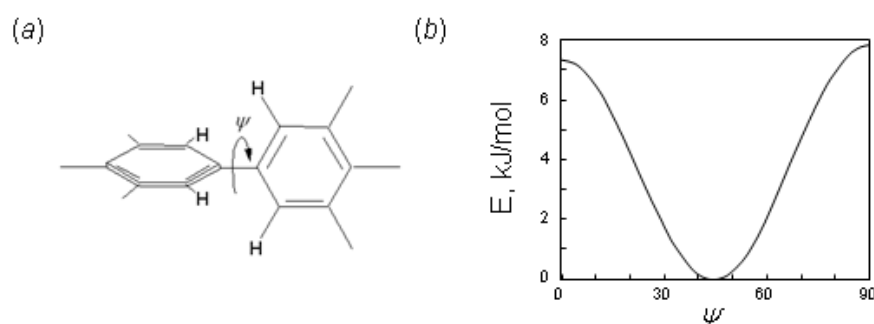
2. BIPHENYLS

When I was a graduate student my advisor, J. A. Ibers, suggested I look at "crystal packing forces", which were (and are) often invoked to explain unexpected results even though such forces were ill-defined. The use of X-ray diffraction for structure determination depends on the belief that the bonding and conformation in the solid state are not very different from what would be found in less condensed phases. Exceptions to this assumption have, however, been known for many years. Ibers pointed me to the specific example of pentaphenylantimony, Ph_5Sb, which

was known to crystallize with square-pyramidal, rather than the expected trigonal-bipyramidal geometry. The expected geometry had been found for the closely related compound Ph_5P.

Investigation of this problem led to several experimental and computational papers (*e.g.* [3] and [4]). At the end of a talk on the project at the 1974 ACA meeting a senior crystallographer asked how the work was related to the biphenyl problem. Biphenyl was known to be twisted in the gas phase but to be planar in the crystal, at least on average, because the molecule is located on an inversion center.

After the CSD became available to us and easy to use we compiled a list of all known molecular biphenyl structures in which there were H atoms at all four *ortho* ring positions and which were not otherwise sterically constrained [5]. We were lucky that quantum mechanical calculations and a gas-phase electron-diffraction study had just been published [6]; a reliable curve showing the variation of the molecular energy with rotation around the central C-C bond (Figure 1) was available. The energy curve has maxima at $\psi = 0°$ (planar conformation; +7.4 kJ/mol) and 90° (perpendicular rings), and a single minimum at 44°.

Figure 1. Drawing showing (*a*) the biphenyl fragment considered and defining the angle ψ, which is 0° for a planar molecule, and (*b*) a graph of the molecular energy as a function of ψ as determined for several biphenyl derivatives by quantum-mechanical calculations and gas-phase electron diffraction [6].

We determined the distribution of the twist angles ψ for structures found in the CSD. The observed distribution of 101 ψ values (Figure 2) showed a maximum at 38°, a possible subsidiary maximum near 20°, and a spike at 0°. The observed distribution was very asymmetric around the energy minimum; there are many more structures at low ψ values than at values above 44°.
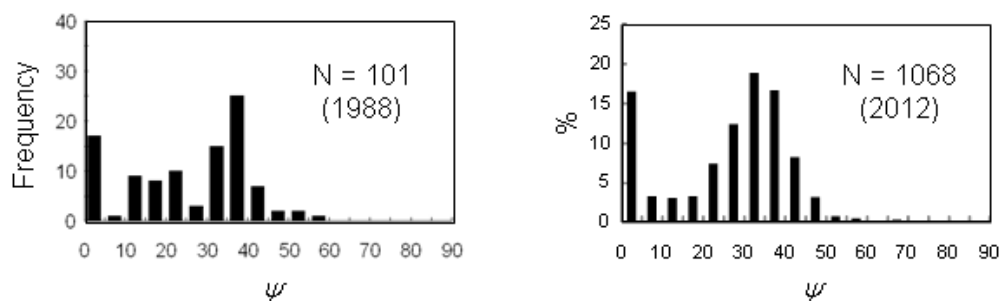
Figure 2. Histograms (5° bins) showing the frequencies of ψ values in structures archived in the CSD as of 1988 and as of 2012. The asymmetry of the distribution and the second peak at $0 - 5°$ indicate that crystallization systematically favors more planar conformations than would be observed in the gas phase.

Recently we constructed this histogram again because so many more structures are now available in the CSD. The new distribution is smoother, but displays the same basic features; the fraction of structures with $10 < \psi < 30°$ is, however, much reduced. The distribution can now be seen as a superposition of a wide, and somewhat asymmetric maximum with ψ *ca.* $30 - 35°$ and a much narrower maximum at $\psi = 0°$. A more thorough investigation that reached the same conclusions has just been published [7].

Our 1989 study showed that crystallization can systematically favor higher-energy molecular conformations if those conformations lead to better packing in the solid state. In the case of biphenyls it is easy to imagine that more planar molecules might be able to fill space more densely than very twisted molecules, and denser packing is usually associated with lower crystal energy. There is probably also (see below) an energy advantage to imposed inversion symmetry. The unusual histogram is therefore easy to explain. Such systematic stabilization of a higher-energy conformation, however, does not occur often; a systematic packing effect can only be expected if the overall shape of the molecule is very sensitive to a low-energy conformational change and if that change can improve the crystal packing.

## 3. MATCHED PAIRS OF ENANTIOMERICALLY PURE AND RACEMIC CRYSTALS

In early 1988 we did a structure as a favor for a colleague who needed to know the absolute configuration of a small molecule containing an S atom (Figure 3). The plan was to determine the structure locally with Mo Kα radiation and then to have data collected elsewhere with Cu Kα radiation so that the absolute structure could be determined reliably. It was a complete surprise when the space group of the "enantiomerically pure" crystals was found to be $P2_1/c$. Eventually we discovered [8] that crystals of the racemic compound are so much more stable than those of the enantiomerically pure material that the good crystals first obtained were racemic, even though the enantiomeric excess of the material from which the crystals were grown was >95%. Crystals of the enantiomerically pure material (space group $P2_12_12_1$) were eventually obtained;
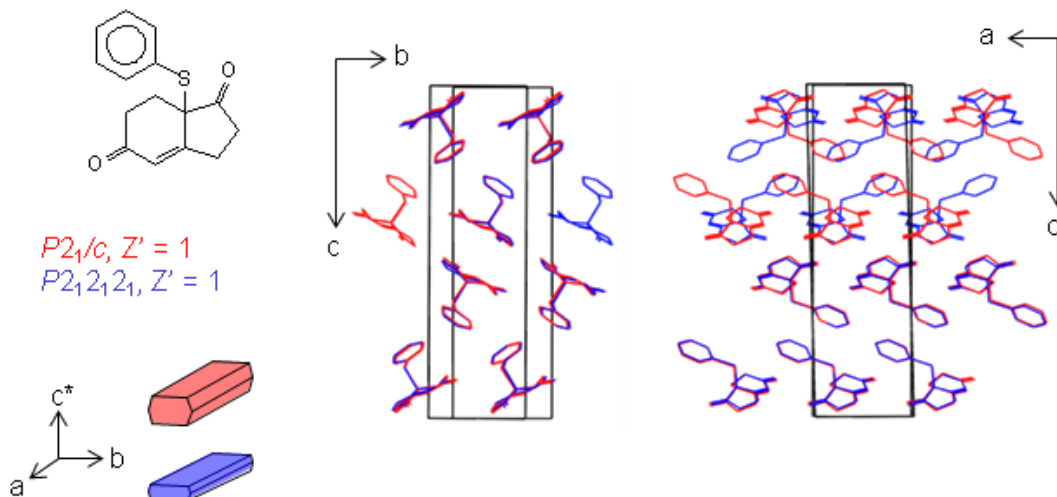
Figure 3. Chemical line drawing and packing diagrams for structures with refcodes KUHLAA (racemic compound) and KUHLEE (pure enantiomer) [8]. The two structures are built from identical bilayers that extend along a and b; adjacent bilayers are related by inversion centers and c cglides in the racemic compound and by 2_1 axes in the enantiomerically pure material.

they are very similar in appearance to the racemic crystals but somewhat thinner. The two structures are also very similar (Figure 3) but the $P2_12_12_1$ crystals are 1.5% less dense than the $P2_1/c$ crystals and have a melting point more than 45 K lower.

This experience led to a consideration of phase diagrams (Figure 4). We eventually concluded that in the case of this material the first crystals grown are very likely to be racemic unless the solution contains less than 1% of the "other" enantiomer.
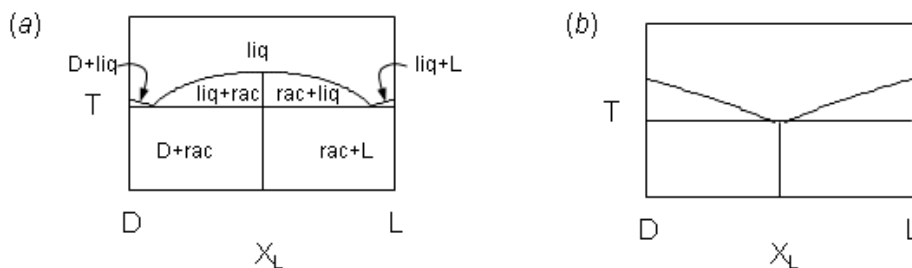
Figure 4. Phase diagrams for a mixture of separable enantiomers that form a racemic compound. (*a*) The racemic compound is considerably more stable than a conglomerate of enantiomerically pure crystals. As the melting point of the racemic compound rises the composition range in which it is possible to grow enantiomerically pure crystals shrinks but never completely disappears. (*b*) The racemic compound is just very slightly more stable than the conglomerate. If the melting point of the racemic crystals is lower than the eutectic temperature of the conglomerate then crystals of the racemic compound cannot be obtained unless kinetic effects favoring the racemic crystals are more important than the thermodynamic stabilities.

Later that year I was on sabbatical at the ETH with Jack Dunitz. We had talked previously about Wallach's Rule, which says that the density of a racemic crystal is normally greater than the density of the corresponding enantiomerically pure crystal. At the time of our first discussions we could think of no way of finding the necessary matched pairs of structures. In late 1988,

however, Jack's colleague Bernd Schweizer devised a way to do the search. We found and investigated 129 temperature-matched pairs of enantiomerically pure and racemic crystal structures of organic molecules [9].

The results were confusing until we realized it was necessary to divide the 129 pairs into two groups: (1) enantiomers that are separable (*i.e.*, that racemize slowly relative to the time necessary for crystal growth), and (2) enantiomers that interconvert faster than crystals grow. The members of first group are two-component systems (three-component if a solvent is included), while in the second group there is only one solute component. The phase diagrams for the two groups are different. (Phase diagrams for resolvable enantiomers are illustrated in Figure 4).

In the case of a single component the two crystal forms are polymorphs, which means that if two (or more) crystal forms are found at about the same temperature they can be assumed to have similar energies. Since the correlation of higher density with lower energy is strong the two (polymorphic) forms should then have similar densities. For the 64 pairs in this group there was no significant difference between the densities; the racemic crystals were, on average, 0.20(34)% denser (Figure 5). In the case of the 65 pairs of resolvable enantiomers, however, the racemic crystals were on average 0.92(29)% denser. Wallach, who had formulated his rule [10] from only eight examples, one of which was an exception, had been correct.
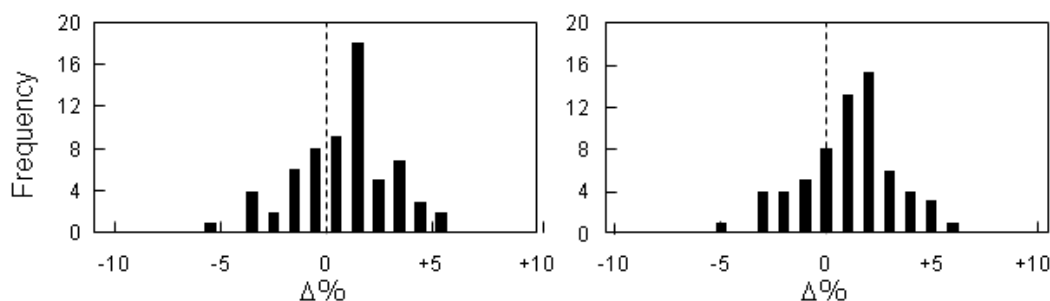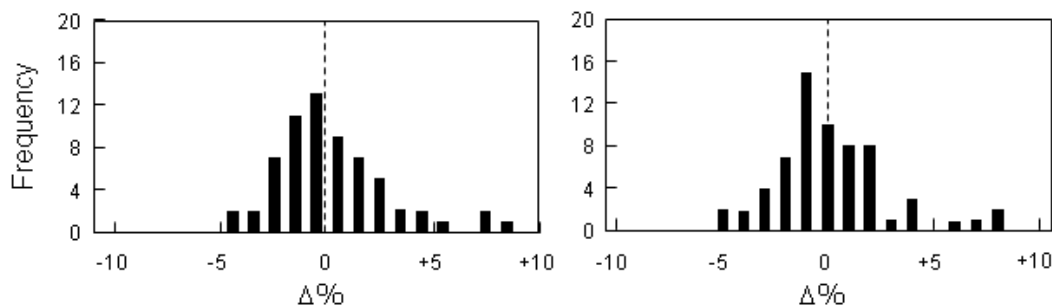
Figure 5. Histograms (1% bins with two origin choices) of the percentage difference between the densities of the racemic and enantiomerically pure members of the pairs of matched structures; $\Delta(\%) = 100(\rho_{rac} - \rho_{enant})/\rho_{avg}$. (*a*) Resolvable enantiomers, for which the difference in densities is significant. (*b*) Inseparable enantiomers, for which there is no measurable density difference.

Consideration of the phase diagrams led us, however, to the understanding that while Wallach's Rule may be correct for resolvable enantiomers, it cannot be proved by any comparison of matched pairs of enantiomerically pure and racemic crystals. If the racemic crystal is more stable, and thus probably denser, it is almost always possible to resolve the material carefully before growing crystals and to then grow crystals of the resolved material. If, however, the racemic crystal is much less stable than the enantiomerically pure crystals then growth of crystals from a racemic solution (at least under near-equilibrium conditions) produces a conglomerate (*i.e.*, a eutectic) of the two kinds of enantiomerically pure crystals. Racemic/homochiral pairs in which the racemic crystal is substantially less dense than its homochiral counterpart cannot exist unless (1) the density-energy correlation breaks down (as it might in the case of hydrogen bonding), (2) the conformational energy in the enantiomerically pure crystal is significantly higher, or (3) kinetic effects during the crystal-nucleation step are determining. Since these conditions are unusual, the sample of matched pairs is necessarily biased towards greater stability of the racemic crystals.

But does Wallach's Rule really have no predictive power? Spontaneous resolution is rare, and achiral molecules normally crystallize in space groups that include inversion centers or other improper symmetry operations. Improper symmetry operations are indeed favorable for crystal packing (see below), but that fact cannot be proved by any comparison of the properties of matched pairs of structures.

Much later we determined the structures of racemic (refcode GEJMEO; $C2/c$, $Z' = 1$) and enantiomerically pure (GEHHUX; $P2_1$, $Z' = 4$) 1,2-dicyclohexylethane-1,2-diol [11]. In this pair the racemic compound is 4.5% *less* dense at room temperature even though its melting point is 10 K *higher* than that of S,S or R,R crystals and an estimated ≥ 40 K higher than that of the conglomerate. Both types of crystals have the same H-bonding pattern (1-D ribbons), but half the molecules in the $P2_1$ crystals adopt a high-energy conformation in order to achieve that pattern; there are also indications of strain in the homochiral ribbons. Furthermore, the $P2_1$ structure has an unusually large number of molecules in the asymmetric unit ($Z' = 4$). This pair of structures led us to realize that a major advantage of improper symmetry elements is that they allow molecules to be related by rotations of arbitrary magnitude. In a racemic crystal any positive rotation relating molecules #1 and #2 can be offset by a corresponding negative (and improper) rotation relating molecules #2 and #3 so that molecules #1 and #3 are related by translation. In crystals in space groups having no improper symmetry operations (hereafter, Sohnke groups) the only operations (other than translations) that can relate molecules are rotations of magnitude $2\pi/n$, $n = 2, 3, 4$, or 6.

4. SPACE-GROUP FREQUENCIES

About 1981 a professor at a nearby college asked us to determine the structure of a crystal of bis(quinuclidine)bromine(I) tetrafluoroborate, which his students had synthesized. The first surprise was that the crystals were uniformly dark when viewed under crossed polarizers. Could the crystals really be cubic? Eventually the structure (Figure 6) was refined successfully in group $P2_13$, with each of the two ions lying on a threefold rotation axis [12]. A search of the CSD at that time showed that in essentially all of the $P2_13$ structures the molecules (or ions) lay on threefold axes so that $Z' = 1/3$.
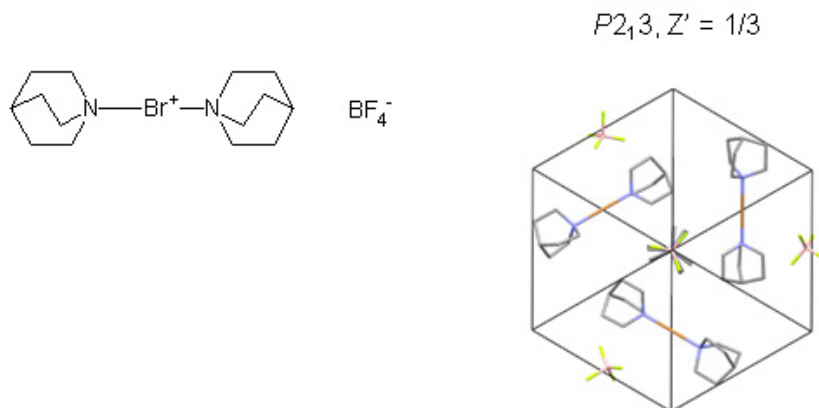
Figure 6. Chemical line drawing and packing diagram for the cubic structure of bis(quinuclidine)bromine(I) tetrafluoroborate (refcode BUWMOV) [12]. All ions lie on threefold axes.

Eventually we compiled a list from the CSD of the number of unique, well-determined structures in all space groups [13]. This table differed from previous work by considering the role of imposed symmetry; structures were classified by the value of Z' as well as by the space-group number.

We found that if the space group includes a mirror plane then at least one molecule nearly always has imposed mirror symmetry. Groups with 3-, 4-, and 6-fold axes do not occur unless a molecule (or aggregate of molecules, see below) lies on that axis. Space group $P1$ is especially likely to have $Z' = 2$. The overall value of Z is remarkably constant; in most crystals there are four orientations of the molecular inertia tensor.

We confirmed that most crystals have inversion symmetry. Occupation of these centers by centrosymmetric molecules is common. Some space groups, especially those with multiple mirror planes and rotation axes, are so rare that any molecular structure reported in them should be viewed with suspicion. In the trigonal, tetragonal, hexagonal, and cubic classes it is the space groups in the lower-symmetry Laue groups that are most frequent, which means that merohedral twinning is likely.

Later work by others [14] showed that if a molecule can conform to inversion symmetry that molecule is nearly always located on a crystallographic inversion center.

An important result of this work was the establishing of a more detailed baseline for expected packing in molecular crystals. Exceptions could then be recognized easily and investigated.

Among the $P2_13$ structures in the November 2011 version of the CSD, $Z' = 1/3$ is common but so is $Z' = 1$. In the latter group, however, essentially all structures have three independent formula units located on threefold axes so that $Z' = 3(1/3) = 1$. The symbol Z'' [15] has sometimes been used to describe the number of crystallographically independent units, which in this case would be 3, but Z'' (at least as originally proposed) can also include solvent molecules and counterions and so has no necessary relationship to Z'. It seems that a symbol that shows both the total number of crystallographically independent formula units and the imposed symmetry is needed. Zorkii's classification by "orbits" [16] is one possibility.

5. MONOALCOHOLS C$_n$H$_m$OH AND *VIC*-DIOLS C$_n$H$_m$(OH)$_2$

The structure of a sterol was determined as a favor for colleagues who needed to know the relative stereochemistry at three of the stereocenters. The structure (WEJDOE; $P2_1$) was found to have $Z' = 3$ [17], which we knew to be unusually large. There is no pseudosymmetry; rather, the three molecular orientations are present so that the hydroxyl groups can form threefold helices in which each O atom acts as both an H-bond donor and acceptor (Figure 7).
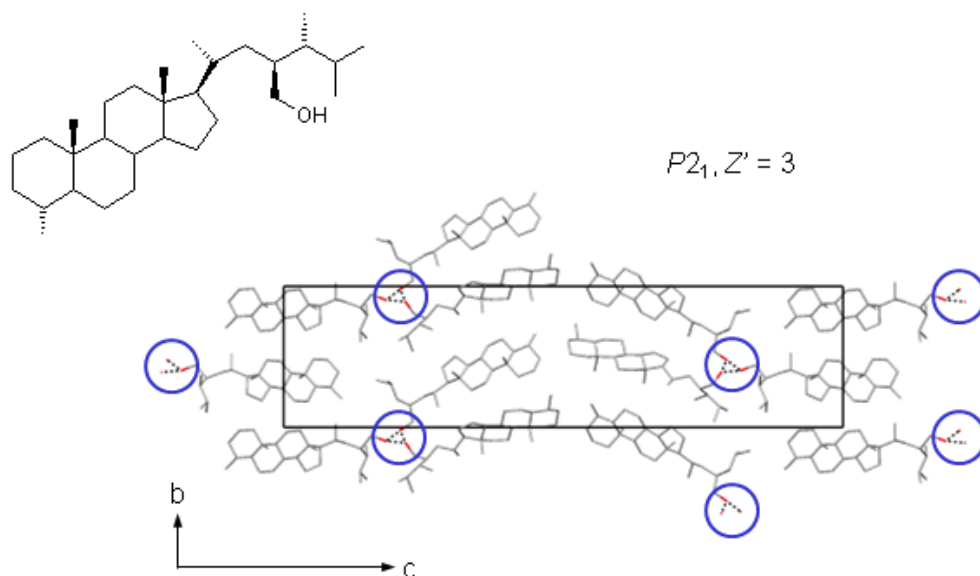
Figure 7. Chemical line drawing and packing diagram for structure with refcode WEJDOE [17]. The three independent monoalcohols form threefold, H-bonded helices parallel to a; these helices are outlined by circles.

We then realized that it was likely that the tendency to form H bonds would lead to anomalous space-group frequencies in other monoalcohols C$_n$H$_m$OH, at least as long as a full set of hydrogen bonds is formed. Chains in which the molecules are related by translation, 2_1 axes, or glide planes are possible, but only if the molecule is relatively small. If the molecules are bulky aggregates are more likely to be formed by incorporating 3-, 4-, or 6-fold crystallographic symmetry or by having $Z' > 1$ (Figure 8).

A search of the CSD confirmed this expectation. Of 55 alcohols that met the criteria 37 (66%) have H bonding patterns in which each O atom acts as both a donor and an acceptor. Of those 37 only 6 (16%) crystallize in triclinic, monoclinic, or orthorhombic space groups with $Z' \leq 1$; the corresponding percentage for the all structures in the CSD was 91%. Just over half of the 37 structures (19) have $Z' > 1$ and another third (12) are in high-symmetry space groups.

Later we looked in detail at *vic*-diols C$_n$H$_m$(OH)$_2$ [18]. Only about 50% of those structures include complete, intermolecular H-bond patterns. In that half, however, the probabilities of structures being in high-symmetry space groups and having $Z' > 1$ are again substantially elevated, although not as much as in the monoalcohols.

We suggested that systematic deviations from expected space-group frequencies can be viewed as evidence of intermolecular interactions being structure-determining.
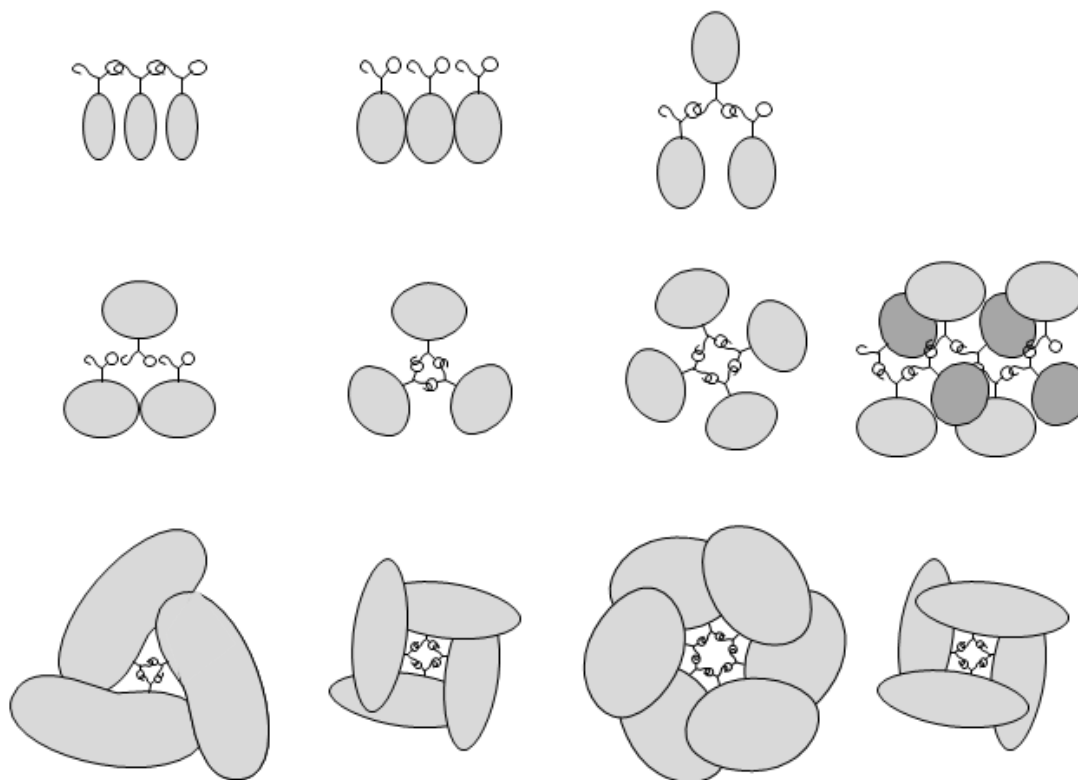
Figure 8. Schematic diagram of how monoalcohols can form aggregates. If the molecules are small then a chain can be formed by translation. If the molecules are somewhat larger the molecules can be related by a twofold screw or a glide operation. If the molecules are yet larger three- or fourfold rings may be formed or the asymmetric unit may contain two independent molecules. If the hydroxyl group is very small relative to the rest of the molecule aggregates can be formed around three-or four- (or six-) fold screw axes or around sites of three-bar or four-bar symmetry.

6. CO-CRYSTALS OF ISOMERS

Over the years we had become aware of several structures of co-crystals of separable isomers that are not enantiomers (Figure 9). Co-crystals of enantiomers (*i.e.*, racemic compounds) are expected (see above) but co-crystals of isomers are not, because in the latter case separation by fractional crystallization is usually successful. Co-crystals of isomers can, however, be very stable; the co-crystal of *cis*- and *trans*-2,3-tetralindiol (refcode RIHLUQ) has been known since the early 20th century [19] and dominates the phase diagram [20]. The stability of the compound is a consequence of the spatial complementarity of the two sets of H-bonding hydroxyl groups [21] (Figure 10).
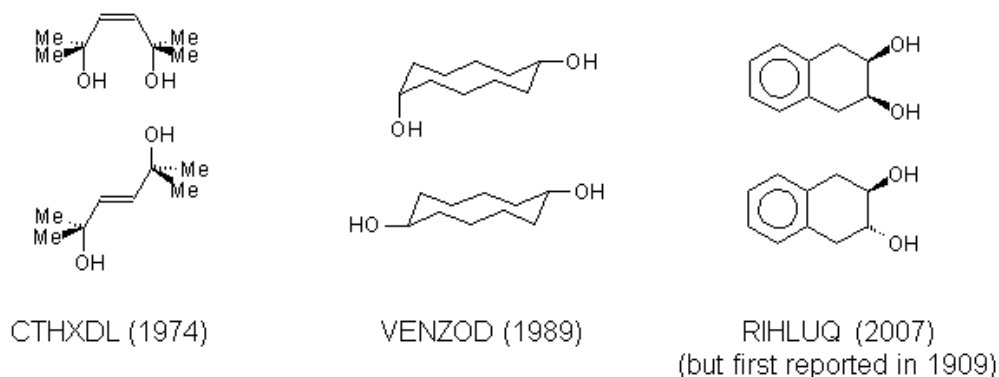
Figure 9. Some examples of ordered co-crystals of isomers.
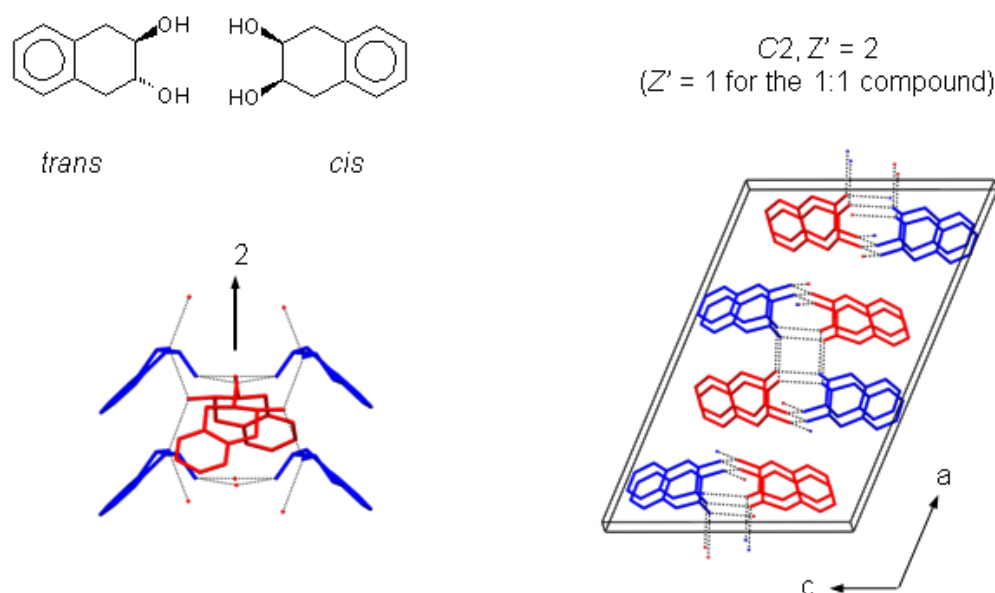
Figure 10 Chemical line drawing and packing diagram for the 1:1 compound with refcode RIHLUQ [21]. The compound forms (and dominates the phase diagram [20]) because equatorial hydroxyl groups on the two molecules make zigzag chains of H bonds while the up-down alternation of the two other hydroxyl groups allows formation of four-membered H-bonded rings. The overall arrangement also allows stacking of the aromatic rings. The ordered compound (shown) contains a *cis* molecule and one of the two *trans* enantiomers. The two *trans* enantiomers are disordered in crystals grown from racemic solutions.

The goal of making a list of ordered compounds of isomers (and other very closely related molecules) seemed impossible before a way of searching for such co-crystals could be found. A method based on the manipulation of InChI™ strings [22, 23] was suggested by Laszlo Fábián, who was then working at the Cambridge Crystallographic Data Centre. He realized that the strings could be generated automatically for all molecules present in crystal structures archived in the CSD and could then be compared. If at least two of the strings were the same except in the connectivity section of the main layer (in which case the molecules are skeletal isomers) or the stereochemical layer (in which case the molecules are diastereomers) then the compound is a co-crystal of interest (Figure 11).

Figure 11. InChI™ strings [22, 23] for the diastereomers L-isoleucine and D-*allo*-isoleucine (XADVED; [24]) and for the two enantiomers of isoleucine.

This systematic search could not, however, find co-crystals of closely related compounds having slightly different compositions. Some of these compounds are important as quasiracemates, which are composed of two quasienantiomers that would be inversion-related but for a minor switch (*e.g.,* H/Me) or substitution (*e.g.,* Br for Cl or Me for Cl). Quasiracemates formed from two diastereomers related by a minor switch; examples XADVED (*i.e.,* L-isoleucine and D-*allo*-isoleucine)[24 and LIPYUE ([25], Figure 12) were found by the automated search, but finding

Figure 12. Chemical line drawings and packing diagram for the quasiracemate (*P*1) with refcode LIPYUE [25]. Each independent molecules has two asymmetric C atoms (marked with asterisks); the configuration is the same at the asymmetric C atom that is part of a six-membered ring but is different at the other. The nearly perfect inversion symmetry of the structure is broken by the difference between the positions of the two methyl groups.

quasiracemates formed from two compounds of slightly different composition (*e.g.*, GOLVOS [26], a co-crystal of L-methionine and D-norleucine with a switch of –SMe to –CH$_2$Me) was considerably more labor intensive.

The search was limited to organic compounds to keep the project tractable, but some unusual hetereoatoms (Se, Te, As, B, Si, and Ge) were allowed, as were Na$^+$ and K$^+$ counterions.

The final list included structures of 270 CSD entries [27]. The two most important classes were co-crystals of configurational diastereomers (114 examples) and of quasienantiomers (114 examples). Together, however, these two classes account for only 51% of the total (157 entries) because 71 compounds are counted in both groups. Most of the remaining 113 examples could not have been predicted. Some (like RIHLUQ) are co-crystals of molecules that form better H bonds with one of their isomers than with themselves; others can be understood as H-bonded co-crystals of a desired molecule and an oxidized impurity. And there are other structures that were just surprises (Figure 13).

RIZJUG (2:1)   HULKOO (1:2)

XOLTOH (1:1 plus CH$_2$Cl$_2$)
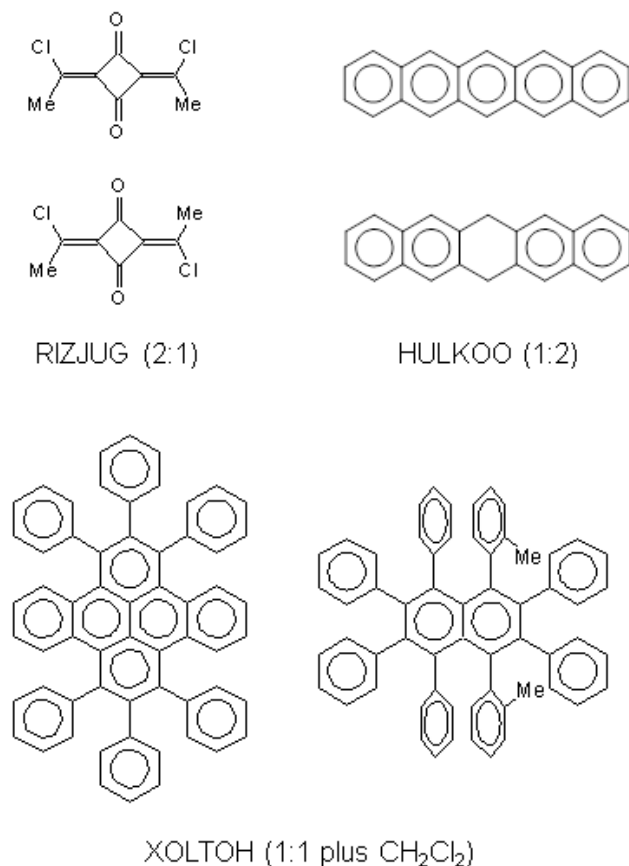
Figure 13. Some examples of unexpected ordered co-crystals found in the CSD.

This project did demonstrate, yet again, that inversion centers, even if only approximate, are very favorable for crystal packing. If pseudo inversion centers were not favorable then the number of quasiracemates found (42% of the total) would have been much lower. Diastereomers are especially likely to form co-crystals if the switch of an H atom and a Me group bound to the same C atom, or the inversion of a [2.2.1] or [2.2.2] cage, would make the molecules

enantiomers. This project also showed that reliable prediction of co-crystal formation by isomers and near isomers is likely to be difficult except in the case of nearly enantiomeric diastereomers.

7. KRYPTORACEMATES

Once the list of co-crystals of isomers had been generated we realized that comparisons of InChI strings could also be used to generate a list of kryptoracemates, *i.e.*, of ordered racemic compounds that crystallize in Sohnke groups so that the enantiomers are not related by any crystallographic symmetry. Structures of kryptoracemates have often been suspected of having higher symmetry than was recognized at the time of the structure determination.

In kryptoracemates the strings for the two molecules would differ in the enantiomer sublayer of the stereochemical layer (Figure 11). We again confined the search to "organic" compounds (with additional heteroatoms Se, Te, As, B, Si, and Ge) allowed. Salts with Group IA or IIA cations were also allowed. Each structure was examined carefully for missed symmetry using the display program *Mercury* [28] and the validation checks available within *PLATON* [29]. The original papers were all consulted.

A list of 181 structures was compiled; 151 of these (group 1) could have crystallized in non-Sohnke space groups without any change of composition. The other 30 structures (group 2) do not meet the strict definition but would have been classified as kryptoracemates by previous authors.

Group 2 examples include solvates (and one salt) in which there is one noncentrosymmetric solvent molecule (or counterion) per racemic pair. There are also 11 (ordered) structures in which the ratio of enantiomers is not 1:1. The process that leads to this second group has been called unbalanced crystallization, but the structures could also be viewed as co-crystals of the racemic compound and one of the pure enantiomers. Finally, there are five somewhat disordered structures in which an enantiomeric or diastereomeric impurity is present at a modest level.

The most important finding of this study was that the deviations from pseudosymmetry are usually easy to spot. While a pseudosymmetric relationship between the enantiomers was found in *ca.* 60% of 181 structures it was always clear that the space group had been assigned correctly. It proved very useful to calculate the centroid of the two enantiomers. If at least one of the coordinates was shifted away from a coordinate that should have been special (0, ½, or ¼) then the inversion center (or other improper operation) could only be approximate (Figure 14). Such an analysis may be the easiest way to satisfy skeptical referees.

Even though the conformations of the two enantiomers are essentially indistinguishable in *ca.* 65% of the structures some of the structures were not even approximately pseudosymmetric. When there are easily identifiable conformational differences between the two enantiomers they are usually minor (rotation of a phenyl ring; rotation of an ethyl or methoxy group around the bond that attaches it to the rest of the molecule).

An unexpected bonus result of this project was an estimate of the frequency of spontaneous resolution. The InChI strings for chiral and achiral molecules are different, so we could determine what percentage of all structures found in Sohnke groups (*i.e.*, not just kryptoracemic structures) contained only achiral or *meso* molecules. That value was 19.5%. We then calculated the split between Sohnke and non-Sohnke groups for achiral molecules, *meso* mole-
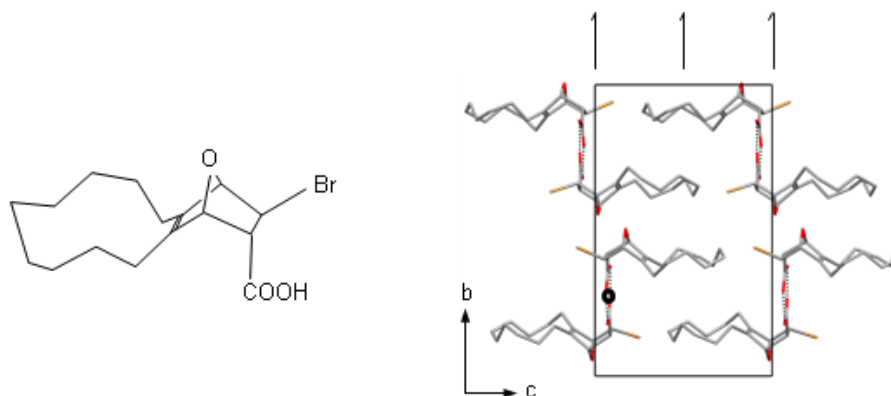
Figure 14. Chemical line drawing and packing diagram for the kryptoracemate ($P2_1$, $Z' = 2$) with refcode JAGQUD [30]. The approximate inversion center marked has coordinates $x = 0.330$ and $z = 0.075$ so that the deviations from space group $P2_1/c$ are obvious. If the inversion centers were moved to $z = 0$ the bridgehead O atoms on adjacent molecules would be too close together.

cules, and racemic compounds. The percentage in Sohnke groups (6.4%) gives an estimate of the likelihood of spontaneous resolution if it is assumed that the split between Sohnke and non-Sohnke groups is the same for resolvable and unresolvable materials.

8. HIGH-*Z'* STRUCTURES

Over the years we have been directly involved with solving and refining 25 structures containing at least three independent molecules (eleven Z'=3, two Z'=4, four Z'=5, one Z'=7, three Z'=8, one Z'=9, two Z'=10 structures, and one Z'=17 structure). We discovered that there is no one simple way to describe them all. Some are modulated, which is to say that small displacements of the molecular packing units would lead to a structure with higher symmetry and/or a smaller unit cell,* but other structures are not modulated.

Two of the modulated structures we studied are intermediate phases in sequences of transitions that occur with crystal cooling [31, 32]. They and the $Z' = 9$ structure [33] (which has not yet been archived in the CSD) are probably incommensurately modulated, although conventional refinements were satisfactory. It is certain that the Z'=17 structure (Figure 15) is modulated incommensurately [34].

The structures that are not modulated are easy to spot. Packing units composed of H-bonded aggregates are common (see Section 5). The $Z' = 8$ structure LAKMIU [35] is a good example. It has an H-bonded, octameric packing unit built from small (19 non-H atoms) molecules; if the aggregate is considered to be the basic packing unit then Z' is 1 rather than 8. The octamer (Figure 16) is not at all symmetric.

---

* Modulations occur often enough with crystal cooling that they are no surprise. As the amplitudes of the thermal motions decrease some intermolecular contacts become more unfavorable. Sometimes a small deformation of the structure occurs because it lowers the overall energy. While there might be a different packing arrangement with a still lower energy, transformation at low temperature to a very different structure is unlikely.
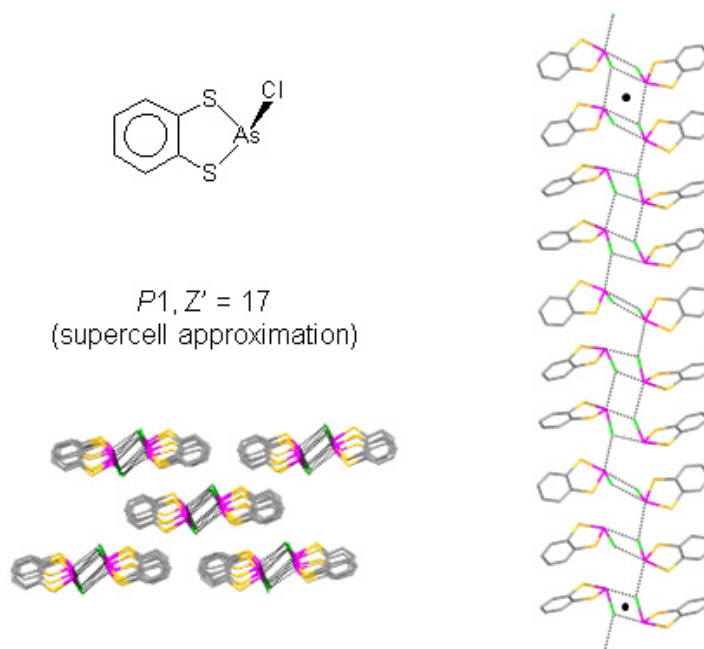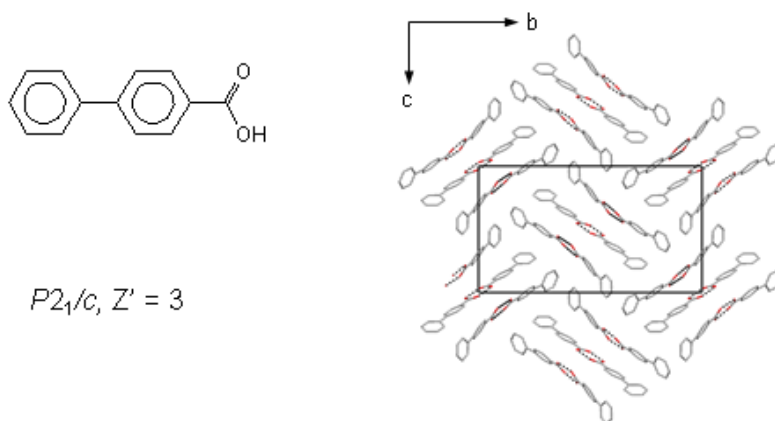
Figure 15. Chemical line drawing and packing diagrams for the commensurate approximation of an incommensurate structure that can be described well in a $Z' = 17$ supercell [34].
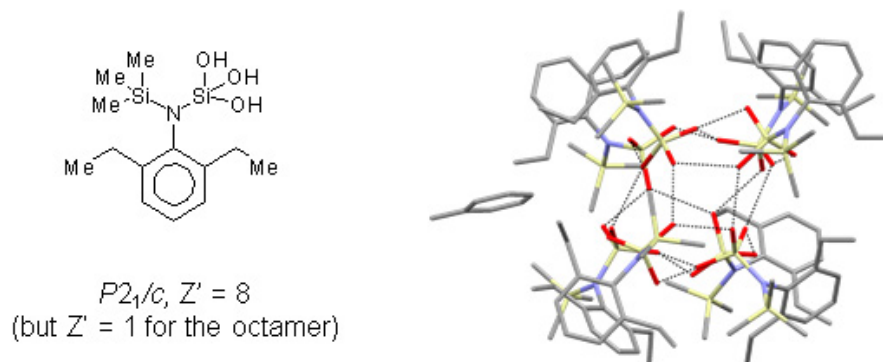
Figure 16. Chemical line drawing (included toluene not shown) and asymmetric unit for the structure with refcode LAKMIU [35]. The H-bonded octamer has no pseudosymmetry at all.

In 4-biphenylcarboxylic acid (BOPSEE; $P2_1/c$, $Z' = 3$) the packing unit is an aggregate of three RCOOH dimers [36]. In this aggregate (Figure 17) the COOH regions are closer together than are the biphenyl groups, which are splayed out. This aggregate must be energetically favorable but it cannot be extended indefinitely to fill space.

Another group of $Z' > 1$ structures include what might be called alternations or perhaps ordered discontinuities. Consider $[Mg(H_2O)_2(15\text{-crown-5})](NO_3)_2$ (JAWQIH01; $P2_1/c$, $Z' = 3$) [37]. The three sets of ions are related by a pseudotranslation of $\mathbf{c}/3$, but the 15-crown-5 ring is inverted in the third cation. This general type of modulation would be described in the least-squares program JANA [38] with a sawtooth or crenel function.

Figure 17. Chemical line drawing and packing diagram for the structure with refcode BOPSEE [36]. The basic packing unit is a centrosymmetric set of three H-bonded dimers. The shorter inter-dimer distance at the center of the unit is favorable for the carboxylic acid groups but the biphenyl units require more space and so are splayed out.

Recently we have begun looking systematically at the high-*Z'* structures (*Z' > 4*) found in the CSD. It turns out that it is not too difficult to spot the modulations with the CCDC's visualization program *Mercury*, even though the modulations sometimes extend along directions other than the standard crystal axes. We have found that many high-*Z'* structures, perhaps even most, are modulated versions of structures built from packing units with $2 \leq Z' \leq 4$.

Most of the high-*Z'* structures fit one of the classes already mentioned. If there is a twofold modulation in a structure built from H-bonded tetramers then the resulting *Z' = 8* is exceptionally high. The same *Z'* value results from a twofold modulation of an AAAB alternation pattern (*e.g.*, BIVJIA [39, Figure 18). A fivefold modulation of a *Z' = 4* packing unit gives an
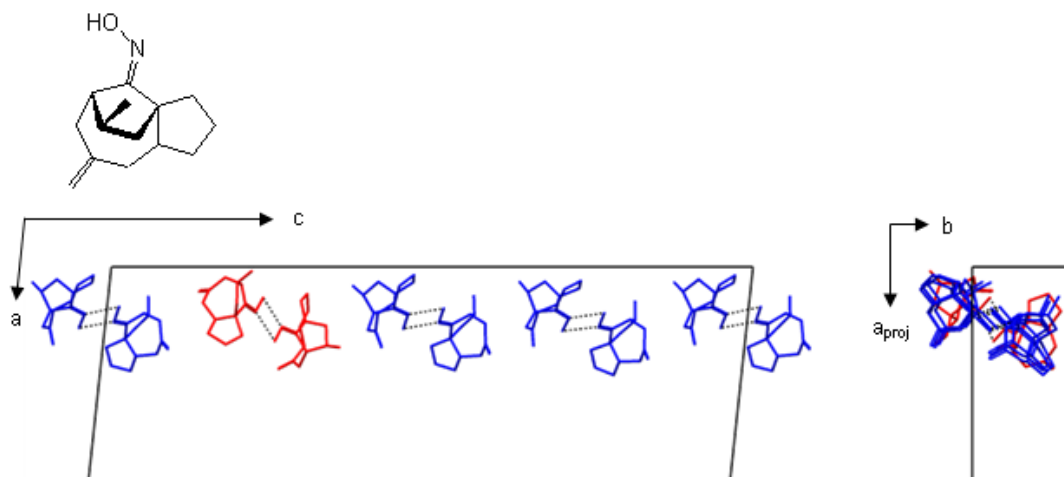
Figure 18. Chemical line drawing and partial packing diagrams for the structure with refcode BIVJIA [39]. The basic packing unit is a dimer; three of the dimers are related by pseudotranslations of c/4 but every fourth dimer has a different orientation.

astounding $Z' = 20$ (VUJBAE [40]; $P1$; Figure 19).

$P1, Z' = 20 = 4 \times 5$
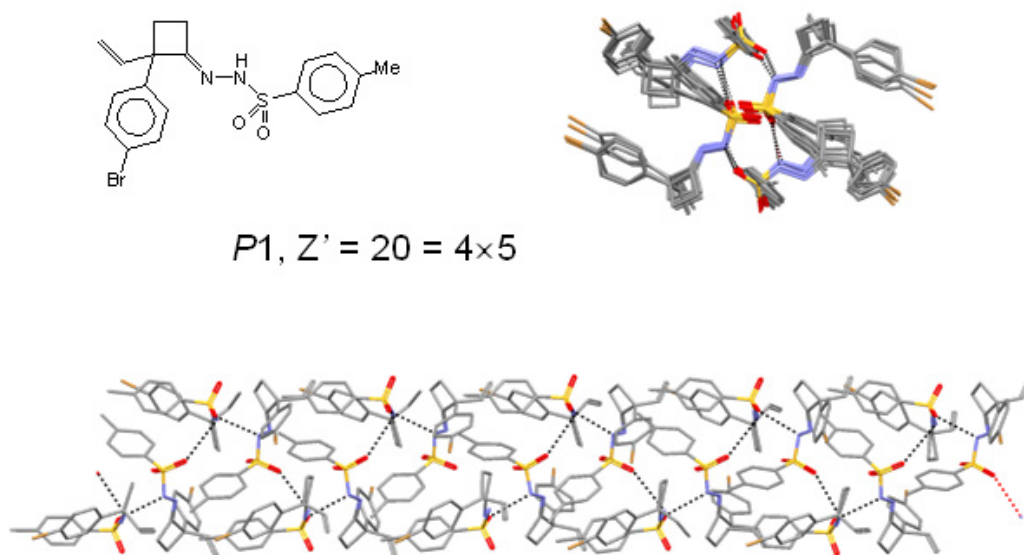
Figure 19. Chemical line drawing and packing diagrams for the structure with refcode VUJBAE [40]. The molecules form a fourfold, H-bonded helix; there is then a fivefold modulation along the helix axis, which is the crystallographic direction [2 1 -1].

There are also a few high-Z' structures in which there is no obvious pseudosymmetry relating the independent molecules. Notable is the structure of L-methionyl-L-alanine (OLOGEB [41]; $P6_1$, $Z' = 7$). In other cases like cholesterol (CHOEST20 [42], $P1$, $Z' = 8$ and CHOEST21 [43], $P1$, $Z' = 16$) there is local, but no long-range, pseudosymmetry.

9. CONCLUSIONS

9.1 Strong Intermolecular Interactions

Strong intermolecular interactions are usually associated with deviations from expected packing patterns. Hydrogen bonding is a good example, especially in simple molecules like monoalcohols C_nH_mOH, *vic*-diols $C_nH_m(OH)_2$, and amino acids. In the simple alcohols high-symmetry space groups and $Z' > 1$ structures are much more prevalent than they are in the CSD as a whole. The amino acids show systematic deviations from Wallach's Rule, with the enantiomerically pure structures having both higher energies *and* higher densities than their racemic counterparts [44]). The hydrogen bonding in the racemic crystals is usually more energetically favorable than in the enantiomerically pure crystals, but better H bonds often require more space than do less good H bonds.

9.2 Larger Asymmetric Units

There is an analogy between $Z' > 1$ structures, structures of kryptoracemates, and structures of co-crystals of isomers (and other very similar molecules); in all three types the asymmetric unit is larger than chemically necessary. It is usually energetically unfavorable, however, to enlarge the asymmetric unit   If two different molecules or molecular orientations are present then

optimizing the repeat distances for both is seldom possible. If the molecules are different there is always one direction, and often many, in which the optimum repeat distances for the different molecules are not the same. The best distance is then a compromise that is almost certainly too short for one component and too long for another. The more different the two molecules are, the greater the packing problem.* One reason that inversion symmetry is so favorable is that it reduces the number of distinct intermolecular vectors by a factor of two without imposing the constraints [13] associated with rotational and mirror symmetry.

It is usually possible to identify a reason for the existence of an asymmetric unit that is larger than chemically necessary. Scientists trying to make co-crystals usually rely on donor/acceptor interactions to favor a larger asymmetric unit. The possibility of satisfying H-atom (or electron) donors and acceptors explains the formation of many $Z' > 1$ structures.

Since $Z' > 1$ structures are more common for enantiomerically pure organic molecules than for the CSD as a whole it must be that improper symmetry operations, and especially inversion centers, are very favorable for crystal packing. It has been noted by many (e.g., Marsh [45]) that the independent molecules in $Z' = 2$, Sohnke-group structures are often arranged around pseudo inversion centers. There is then also an analogy between a $Z' = 1$ structure of a molecule that is potentially centrosymmetric and a $Z' = 2$ structure of one that is not because molecules that can conform to inversion symmetry so often do [14].

The existence of a large number of quasiracemic crystals shows that approximate inversion symmetry is more favorable than the complete absence of inversion symmetry.

The other large group of $Z' > 1$ structures is composed of the modulated structures. Many of these structures are not really exceptional, especially if the crystal was studied at a temperature lower than that at which it was grown. Simple modulations (*e.g.*, a two- or threefold increase in some crystal direction) occur quite often with cooling; such a modulation is usually derived from a low-frequency lattice mode that is antisymmetric with respect to some symmetry element of the crystal. Longer modulations (including incommensurate modulations) occur more rarely but can often be understood in the same way. A classic example is biphenyl. At higher temperatures the amplitude of the intramolecular torsional mode, which is antisymmetric with respect to the crystallographic inversion center at the molecular center, is large enough that the repulsions between the *ortho* H atoms are tolerable. At low temperatures , where the amplitudes of low-frequency modes are reduced, the repulsions become important enough that there are transitions to two incommensurate phases. These phases contain averaged molecular conformations that are twisted rather than planar (see [46]).

9.3 Packing Conflicts

The percentage of $Z' > 1$ structures among small biphenyl molecules (234 structures having fewer than 21 C atoms) is high at 22%, suggesting a conflict between optimizing the intramolecular energy (which favors a twisted conformation) and the intermolecular energy (which almost certainly favors a more planar conformation that can fill space more densely). As the biphenyl fragment becomes a smaller part of the overall molecule that percentage drops to 13% (21-30 C atoms) and finally to 11% (more than 30 C atoms).

---

* If one of the molecules is much larger than the other, as in the case of included solvent molecules, the mismatch problem becomes much less important.

Systematic packing effects are possible if a low-energy conformational change can make a large difference in the way the molecules pack. Again, the classic example is the group of biphenyls with H atoms in all four *ortho* positions. These molecules can pack more densely if planar and the energy cost for making the molecules planar is low. Furthermore, many planar biphenyls can conform to inversion symmetry, which is a second packing advantage.

10. SUMMARY

Every crystal structure determined presents an opportunity to learn about crystal packing. Progress in understanding how molecules are arranged in crystals can be made by knowing what is normal and investigating the exceptions.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Kitaigorodskii, A.I. *Organic Chemical Crystallography*, (Consultant's Bureau: New York, 1961).
[2]   F.H. Allen, *Acta Cryst*. **B58** (2002), 380.
[3]   C.P. Brock and D.F. Webster, *Acta Cryst*. **B32** (1976), 2089.
[4]   C.P. Brock, *Acta Cryst*. **A33** (1977), 898.
[5]   C.P. Brock and R.P. Minton, *J. Am. Chem. Soc*. **111** (1989), 4586.
[6]   O. Bastiansen and S. Samdal, *J. Mol. Struct*. **128** (1985), 115.
[7]   A.J. Cruz-Cabeza, J.W. Liebeschuetz and F.H. Allen, *CrystEngComm* **14** (2012), DOI: 10.1039/c2ce25585e.
[8]   C.P. Brock, S. Kwiatkowski, D.S. Watt and A. Sayed, *Acta Cryst*. **B48** (1992), 719.
[9]   C.P. Brock, W.B. Schweizer and J.D. Dunitz, *J. Am. Chem. Soc*. **113** (1991), 9811.
[10]   O. Wallach, *Justus Liebigs Ann. Chem*. **286** (1895), 90.
[11]   B.O. Patrick and C.P. Brock, *Acta Cryst*. **B62** (2006), 488.
[12]   L.K. Blair, K.D. Parris, P.S. Hii and C.P. Brock, *J. Am. Chem. Soc*. **105** (1983), 3649.
[13]   C.P. Brock and J.D. Dunitz, *Chem Mater* **6** (1994), 1118.
[14]   E. Pidcock, W.D.S. Motherwell and J. Cole, *Acta Cryst*. **B59** (2003), 634.
[15]   B.P. van Eijck and J. Kroon, *Acta Cryst*. **B56** (2000), 535.
[16]   N.Yu. Chernikova, V.K. Bel'skii and P.M. Zorkii, *J. Struct. Chem*. **31** (1991), 661.
[17]   C.P. Brock, I. Stoilov and D.S. Watt, *Acta Cryst*. **C50** (1994), 434.
[18]   C.P. Brock, *Acta Cryst*. **B58** (2002), 1025.

[19]   H. Leroux, *Comptes Rendus* **148** (1909), 931.

[20]   H. Lettré and I. Lerch, *Chem. Ber*. **85** (1952), 394.

[21]   M.A. Lloyd, G.E. Patterson, G.H. Simpson, L.L. Duncan, D.P. King, Y.Fu, B.O. Patrick, S. Parkin and C.P. Brock, *Acta Cryst*. **B63** (2007), 433.

[22]   S.E. Stein, S.R. Heller and D. Tchekhovskoi *Proceedings of the 2003 International Chemical Information Conference (Nimes), Infonortics* (2003), 131.

[23]   S.E. Stein, S.R. Heller and D.V. Tchekhovskoi,). *The IUPAC Chemical Identifier – Technical Manual (National Institute of Standards and Technology, Gaithersburg*, 2006).

[24]   B. Dalhus and C.H. Görbitz, *Acta Cryst*. **B56** (2000), 720.

[25]   M. Walker, E. Pohl, R. Herbst-Irmer, M. Gerlitz, J. Rohr and G.M. Sheldrick, *Acta Cryst*. **B55** (1999), 607.

[26]   B. Dalhus and C.H. Görbitz, *Acta Cryst*. **C55** (1999), 1105.

[27]   S.P. Kelley, L. Fábián and C.P. Brock, *Acta Cryst*. **B67** (2011), 79.

[28]   C.F. Macrae, I.J. Bruno, J.A. Chisholm, P.R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. v. d. Streek and P.A. Wood, *J. Appl. Cryst*. **41** (2008), 466.

[29]   A.L. Spek, *J. Appl. Cryst*. **36** (2003), 7.

[30]   K. Peters, E.-M. Peters, A. Taugerbeck and W. Tochtermann, *Z .Kristallogr. New Cryst. Struct.* **213** (1998), 497.

[31]   M.A. Siegler, X. Hao, S. Parkin and C.P. Brock, *Acta Cryst*. **B67** (2011), 486.

[32]   M.A. Siegler, S. Parkin and C.P. Brock, *Acta Cryst*. **B68** (2012), 389.

[33]   P.A. Koutentis, R.C. Haddon, R.T. Oakley, A.W. Cordes and C.P. Brock, *Acta Cryst.* **B57** (2001), 680.

[34]   R.C. Bakus II, D.A. Atwood, S. Parkin, C.P. Brock and V. Petříček, in preparation.

[35]   G. Prabusankar, R. Murugavel and R.J. Butcher, *Organometallics* **24** (2005), 2124.

[36]   C.P. Brock, J.R. Blackburn and K.L. Haller, *Acta Cryst*. **B40** (1984), 493.

[37]   X. Hao, S. Parkin and C.P. Brock, *Acta Cryst*. **B61** (2005), 675.

[38]   V. Petricek, M. Dusek and L. Palatinus *JANA2006* (Institute of Physics, Praha, Czech Republic, 2006).

[39]   M. Harmata, P. Rashatasakhon and C.L. Barnes, *Can. J. Chem*. **84** (2006), 1456.

[40]   F. Kleinbeck and F.D. Toste, *J. Am. Chem. Soc.* **131** (2009), 9178.

[41]   C.H. Görbitz, *Acta Cryst*. **C59** (2003), o589.

[42]   H.-S. Shieh, L.G. Hoard and C.E. Nordman, *Nature (London)* **267** (1977), 287.

[43]   Y.-Y. Hsu, J.W. Kampf and C.E. Nordman, *Acta Cryst*. **B58** (2002), 260.

[44]   J.D. Dunitz and A. Gavezzotti, *J. Phys. Chem. B* **116** (2012), 6740.

[45]   R.E. Marsh, *Acta Cryst*. **B55** (1999), 931.

[46]   P. Launois, F. Moussa, M.H. Lemée-Cailleau and H. Cailleau, *Phys. Rev. B: Condens. Matter* **40** (1989), 5042.