Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

A list of organic kryptoracemates

László Fábián[a] and Carolyn Pratt Brock[b]*

[a]Pfizer Institute for Pharmaceutical Materials Science, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, and [b]Department of Chemistry, University of Kentucky, Lexington, KY 40506-0055, USA

Correspondence e-mail: cpbrock@uky.edu

Received 1 November 2009
Accepted 12 December 2009

A list of 181 organic kryptoracemates has been compiled. This class of crystallographic oddities is made up of racemic compounds (*i.e.* pairs of resolvable enantiomers) that happen to crystallize in Sohnke space groups (*i.e.* groups that include only proper symmetry operations). Most (151) of the 181 structures could have crystallized as ordered structures in non-Sohnke groups. The remaining 30 structures do not fully meet this criterion but would have been classified as kryptoracemates by previous authors. Examples were found and checked with the aid of available software for searching the Cambridge Structural Database, for generating and comparing InChI strings, and for validating crystal structures. The pairs of enantiomers in the true kryptoracemates usually have very similar conformations; often the match is near-perfect. There is a pseudosymmetric relationship of the enantiomers in about 60% of the kryptoracemate structures, but the deviations from inversion or glide symmetry are usually quite easy to spot. Kryptoracemates were found to account for 0.1% of all organic structures containing either a racemic compound, a *meso* molecule, or some other achiral molecule. The centroid of a pair of enantiomers is more likely (99.9% *versus* 99% probability) to be located on an inversion center than is the centroid of a potentially centrosymmetric molecule.

1. Introduction

Kryptoracemates, which are thought to be very rare, are racemic crystals in which the enantiomers are crystallographically independent, *i.e.* are not related by any space-group symmetry. The Greek prefix *krypto* refers to the compound's racemic nature being in some sense hidden. Kryptoracemates must crystallize in Sohnke[1] space groups, *i.e.* in the 65 space groups that do not include any improper symmetry element (*i.e.* inversion center, mirror plane, glide plane or rotoinversion axis). The value of Z', which is the number of molecules in the asymmetric unit, must be greater than 1 unless the molecules are located on rotation axes.[2]

[1] The spellings 'Sohnke' and 'Sohncke' both seem to be in common use.
[2] When determining Z' it is conventional to count individual crystallographically independent molecules rather than to count the pairs of enantiomers that constitute the racemic compound. The result is that Z' for a structure containing a single pair of enantiomers (*i.e.* for one formula unit of a racemic compound) is usually given as 1 if the space group includes improper symmetry elements and as 2 if it does not. This convention has been used so widely and for so long that it must be accepted. The spreadsheet associated with this paper, however, lists Z' for the racemic compound (*i.e.* for the number of independent pairs of enantiomers. This usage is consistent with the usual description of other solid-state compounds, such as co-crystals of isomers.

Kryptoracemates are considered oddities because very few have been identified, and because molecules that can be arranged around crystallographic inversion centers almost always are. The observation that inversion symmetry is usually retained has been made by many crystallographers over many decades. It was recently quantified by Pidcock *et al.* (2003), who showed that 99% of molecules[3] that can adopt inversion symmetry only are located on crystallographic inversion centers. In racemic compounds it is the 1:1 DL compound, rather than a single molecule, that occupies an inversion center, but the principle is the same. The pairs of molecules of a racemic compound can together conform to inversion symmetry but only rarely to any other symmetry.

Another indication that enantiomers should be related by improper symmetry elements is the observation that in $Z' > 1$ structures of enantiomerically pure materials the two molecules are often related by an approximate inversion center. Marsh (1999) reported that in the hundreds of $P1$, $Z' > 1$ structures he examined he found the two homochiral molecules to be related by approximate inversion in about a third of the structures. Inversion symmetry has long been thought to be very favorable for crystal packing.

Finally we note that it is not unusual to discover that a crystal grown from a sample of quite high enantiomeric purity actually contains the racemic compound (see, *e.g.*, Kwiatkowski *et al.*, 1989). In such cases the racemic crystal must have a significantly more favorable free energy than the crystal of one of the pure enantiomers (Brock *et al.*, 1991).

An early list of racemic compounds that crystallize in Sohnke groups was given by Brock *et al.* (1991), but only one of the seven examples was identified correctly. We have been unable to uncover the errors made in compiling that list.

Morales & Fronczek (1996), who published the structure of a kryptoracemate (refcode TABLUD in the CSD; Allen, 2002), cited Ivan Bernal (1995*a*,*b*) as having found 'less than three dozen examples', many of which contained transition metals. The molecules in many of the examples were said to be related by a pseudo-inversion center. The Morales & Fronczek paper also seems to be the first in which the term kryptoracemate, which was attributed to Bernal (1995*a*), appeared.

In the same year Bernal *et al.* (1996) listed 11 additional organic compounds that crystallize as kryptoracemates, but only four of the structures are reliable (BINGOU, DTYROS, CARVOX02 and RCPICM10). The others either appear in the CSD without coordinates, have $Z' = 1$ with no indication of disorder, or cannot be found in the literature. The latter two of the reliable examples are mixed crystals in which the enantiomeric molecules occupy the same site and in which the composition can vary.

In 1999 Lynch and co-workers (Lynch *et al.*, 1999) published the structure of another kryptoracemate (HIKGEN01), and added another reliable example (DLMSUC01) to the list.

Dalhus & Görbitz (2000) seem to have been the first to look systematically for kryptoracemates.[4] They list 17 organic kryptoracemates of which three had been mentioned previously and one has an incomplete entry in the CSD. One of the new examples is a mixed crystal with $Z' = 1$, and three more are partially mixed crystals composed of an enantiomerically pure host and a disordered racemic solvent. With the publication of this paper the number of reliable structures of organic kryptoracemates reached 20, of which six are mixed or partially mixed crystals.

A few years later Flack (2003) wrote:

> It is thought that *ca* 50 of these chiral crystal structures of racemates are known at present from 250 000 structures in the CSD, although it is not possible to give an authoritative and citable reference for this number. For various technical reasons associated with nomenclature and self-consistency, a direct search on the CSD does not produce reliable results, and one must critically consult the primary literature to identify those racemates with chiral crystal structures.

Very recently Bishop & Scudder (2009) described three more kryptoracemates (MUSDOT, XETBED and WASWAO), which they termed 'false conglomerates'. They also characterized the problem of locating all such structures in the CSD as 'a daunting task'.

We realised that the search problem had been largely solved by the availability of software tools that identify the absolute chirality of asymmetric atoms present in molecular models (Stein *et al.*, 2003, 2006). Furthermore, it is now possible to investigate in a routine, and at least semi-automatic, way the possibility that a higher-symmetry space group should have been used to describe the structure. Computational tools for that task are available in *PLATON* (Spek, 2009) and *Mercury* (Macrae *et al.*, 2008).

We therefore decided to compile and study a list of kryptoracemates because it had finally become feasible to make such a list and because studies of classes of outliers often lead to new insights. By looking in detail at this group of exceptions to a strong and general rule we hoped to understand more about crystal packing and pseudosymmetry.

2. Methodology

Compounds containing metal complexes were not considered because of the difficulty of reliably identifying asymmetric metals and defining the boundaries of a metal-containing molecule or molecular ion. The list of 'organic' kryptoracemates therefore includes no transition metals. Some other heteroatoms were allowed.

[3] The term 'molecules' as used here is an abbreviated form of 'molecules and molecular ions' and so does not imply electrical neutrality. A 'molecule' is then a group of atoms connected by covalent bonds.

[4] Dalhus & Görbitz (2000) actually searched for all noncentrosymmetric structures having $Z' > 1$ and composed of separable enantiomers. The 17 kryptoracemates they identified form a subset of their final list (see their Table 7 and their supplementary material).

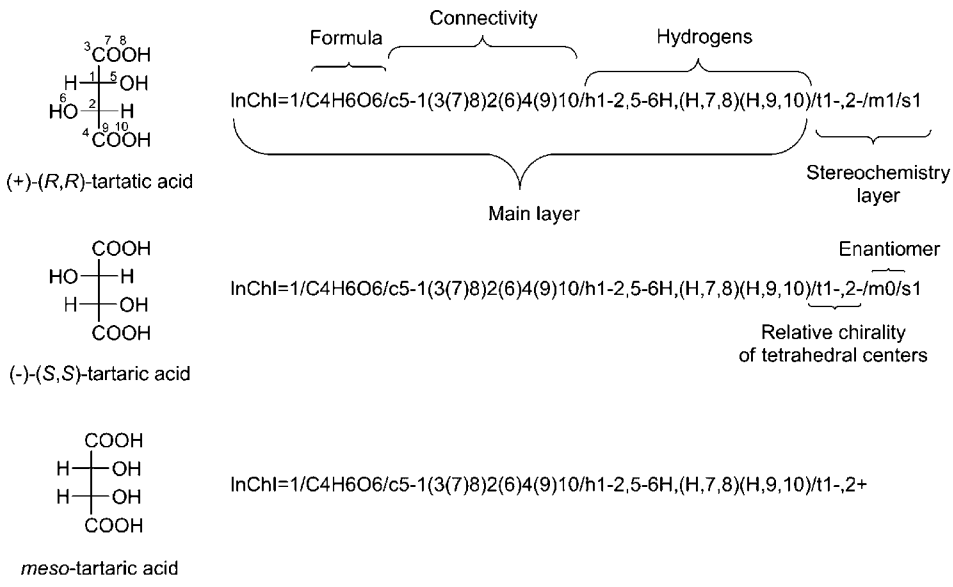

Figure 1
InChI string representation of tartaric acid stereoisomers. The main layer, which defines the connectivity, is the same for all three isomers, while their stereochemistry layers differ. The only difference between the InChI strings of (+)- and (−)-tartaric acid is the /*m*0 or /*m*1 enantiomer sublayer. Note that the string for *meso*-tartaric acid does not have an enantiomer sublayer, because the molecule is achiral.

2.1. Automated search

Searches were performed on Version 5.30 (November 2008) of the CSD (Allen, 2002) and all updates through May 2009. Search criteria were:

(i) coordinates archived in the CSD;

(ii) Sohnke (*i.e.* proper) space group;[5]

(iii) $Z' > 1$ (note that this test would exclude a kryptoracemate having both molecules located on twofold rotation axes);[6]

(iv) permitted elements restricted to C, H, N, O, S, Se, Te, P, As, B, Si, Ge, F, Cl, Br, I and a Group IA or IIA counterion;[7]

(v) asymmetric atom present.

Structures determined from powder data (*e.g.* HISRIL01 and YIXVAD) were allowed.

The refcodes located by this search were exported as mol2 files, which were split into individual model files (one covalently bonded unit per file) using a program written at the CCDC. Each file was then converted to an IUPAC International Chemical Identifier (InChI™) string (Stein *et al.*, 2003, 2006) using Version 2.2.1 of *Open Babel* (http://openbabel.org/). InChI strings provide a unique textual representation of chemical substances; these strings are composed of layers, which describe the substance in different levels of detail (see Fig. 1). The InChI strings for enantiomers are composed of identical main and stereochemistry layers, except for a '/*m*1' or '/*m*0' enantiomer sublayer, which indicates whether the structure is inverted or not. Thus enantiomeric units in a structure could be identified by comparing the generated InChI strings. If two strings for a structure differed only in their '/*m*' stereochemistry layer but were otherwise the same, the units were considered to be enantiomers and the refcode was retained as a hit. Structures rejected by this test included enantiomerically pure materials having $Z' > 1$ and compounds of both configurational and *E*/*Z* diastereomers.

The test on the InChI strings led to a list of structures in Sohnke space groups that are composed of a pair of enantiomers that are, except in three cases, resolvable. We thought that this list might be incomplete because of the failure of the InChI-string generation if the CSD entry included a 'delocalized' bond type.[8] This bond type is, however, separate from the 'aromatic' type and is seldom encountered. Checks of the 365 affected structures turned up no kryptoracemate.

The list of hits was then run through *PLATON* (Version 190509; Spek, 2009) in batch mode three times: once with default criteria for missed symmetry (angle deviation for any metrical symmetry $\leq 1^\circ$; atoms related by inversion or translation ≤ 0.45 Å apart; atoms related by other symmetry elements ≤ 0.25 Å apart), and then with those tolerances doubled and halved.

2.2. Initial screening for chemical composition

The hits were examined individually using the CSD software. If there was any question about a structure, such as about atoms marked in the CSD as disordered, the original literature was consulted. Structures that could crystallize in a Sohnke group without disorder were assigned to group 1 (strict definition) while those structures that could not were assigned to group 2 (relaxed definition). Minor conformational disorder was considered unimportant, but minor disorder associated with the presence of an enantiomeric or

[5] We have avoided the commonly used term 'chiral space group' because while some Sohnke groups (*e.g.* $P3_1$ and $P3_2$) cannot be superimposed on their mirror images and are therefore chiral, other groups (*e.g.* $P2_1$) are not chiral. For a discussion of this point see pp. 914–915 of Flack (2003).

[6] A search was also carried out with the number of chemical units >1 to find any structures that might have been encoded as cocrystals or salts but this precaution proved unnecessary.

[7] The only hits including IA, IIA metals that survived the full set of automatic tests were WONTIC (Li^+) and ZIKDEC (Cs^+), but they were deleted because the ligands were organized around the ions in a way typical of coordination complexes.

[8] Examples include complexes (*e.g.* of Ge) containing the acetylacetonato group or a closely related analogue, salts in which the cation contains a $-C(NR_2)_2^+$ group or the anion contains a metal-coordinated $-CO_2^-$ group, and molecules containing a phosphate group. Most such structures found contain atoms not usually found in organic molecules.

diastereomeric impurity meant the structure was assigned to group 2. The group 2 structures are not quite classical kryptoracemates, but they are unexpected crystallization products in which the two enantiomers are crystallographically independent. Previous authors counted structures like those in group 2 as kryptoracemates.

We excluded metal salts in which the cation (*e.g.* Ca^{2+}, Li^{+} or K^{+}) was clearly coordinated by an organic ligand. Salts in which the cation was in a less highly structured environment would have been retained but there were none.

The three structures (BSADAZ, CHATRZ and YEPLAH) in which the rate of racemization seemed likely to be rapid relative to the rate of crystal growth (the only asymmetric atom is an N, S or Se atom) were excluded. Structures with a P or an S atom (the latter as a sulfoxide) as the only asymmetric center(s) remained in the list.

PDTOMS11 was deleted because a careful study (Wong-Ng *et al.*, 1984) showed that the material, which was extracted from *Podopetalum ormondii*, contains two molecules that differ in composition as well as in chirality. The double bond in one molecule is a single bond in the other. PDTOMS11 is a quasiracemate (Lineberry *et al.*, 2008, and references therein) rather than a kryptoracemate.

Structures with $R > 0.10$ were excluded from the final count because an R factor that high often indicates an unresolved problem. That criterion, however, may have eliminated some true kryptoracemates from the count.

Solvates and salts were then considered. We eventually decided to include all structures of crystals that seemed to have been grown under conditions that would have been expected to produce a racemic compound. Salts were retained in group 1 if the composition is 1:1 and if the two counterions are either enantiomeric or achiral. Some other salts were assigned to group 2. We discarded structures corresponding to failed diastereomeric resolutions [*e.g.* KITLUV, which contains two (homochiral) ephedrinium cations and a pair of enantiomeric anions]. There are no salts of composition 2:1 (one anion per enantiomeric pair) in group 1, probably because few chemically reasonable dianions (oxalate is an exception) can be located, without disorder, on a site of inversion symmetry. ACUBAC, a 2:1 salt with a sulfate ion, was retained, but in group 2.[9] We found no 1:2 salts (one cation per enantiomeric pair of anions).

Solvates were treated similarly. A structure was retained in group 1 if there are two solvent molecules per enantiomer pair and if the two solvent molecules are either enantiomeric or achiral. Structures containing only one achiral solvent molecule per enantiomer pair were retained but in group 2.

Another class of structures included in group 2 are the 11 structures having an enantiomeric ratio other than 1:1. These structures, which are the products of so-called unbalanced crystallization (Albano *et al.*, 1969; Cai *et al.*, 2001), should probably be called co-crystals of a racemic compound and a pure enantiomer, but are included because they are so closely related to kryptoracemates. Nine have enantiomeric ratios of 2:1, but one (SOQQOE) has a ratio of 3:1 and the last (WIYSAZ) has a ratio of at least 5:1 (there is some disorder). Also included in group 2 is ABADUD (Basak *et al.*, 2004), which is an example of pseudo-unbalanced crystallization because the 'extra' molecule is not quite the same as either member of the enantiomeric pair.

The last set of structures in group 2 contains five in which an enantiomeric or diastereomeric impurity is present. The impurity is located at the site of one enantiomer but not at the other so that the structure is partially disordered. These crystals can be considered to be mixed crystals of the racemic compound and one of the pure enantiomers or of the racemic compound and a compound of diastereomers. Again, these structures were included because of their similarity to kryptoracemates.

Note that structures of mixed crystals (*e.g.* of crystals that are solid solutions) do not appear on the list if the two enantiomers are disordered at a single site so that they are listed in the CSD with $Z' = 1$. Such structures have been included in some previous lists.

2.3. Examination of conformational similarity and of pseudosymmetry

The structures remaining on the list were again examined using *Mercury* (Macrae *et al.*, 2008). We used a special overlay feature developed at the CCDC to assess the similarity of the independent molecules.[10] This superposition feature automatically calculates overlap both with and without inversion of one of the molecules.

The crystal packing was also examined using *Mercury*; deviations from improper symmetry were identified. If pseudo- or local inversion centers were found the coordinates of the centroid of the 'inversion-related' non-H atoms were calculated so that the deviations of those 'centers' from special coordinate values could be quantified. The results were compared with the results from the *ADDSYM* routine of *PLATON* and with space-group reinterpretations noted in the CSD. A comparison with space-group corrections tabulated in the literature (Marsh & Herbstein, 1988; Kapon & Reisner, 1990; Marsh, 1994*a*, *b*, 1995, 1997, 1999, 2002, 2004, 2005, 2009; Marsh & Bernal, 1995; Herbstein & Marsh, 1998; Marsh & Spek, 2001; Marsh *et al.*, 2002; Clemente, 2003; Marsh & Clemente, 2007) showed that the overlap between our list and the published lists is very small, because most of the structures corrected in those papers either included metal atoms, were enantiopure, or were originally in non-Sohnke space groups like *Cc*.

In only a small number of cases (*ca* 5%) was it difficult to decide whether the space group reported was the best possible description. If there was real doubt about deviations from improper symmetry the structure was excluded from the final count but was retained in a separate section of the final

[9] If the cations were related by crystallographic symmetry the sulfate anion would have to be located on a special position of symmetry *m*, 2 or $\bar{4}$. None of these possibilities seems very likely.

[10] This overlay option became generally available with the November 2009 release of the CSD.

Table 1
Counts of the CSD entries containing the atom types considered, having archived coordinates, and giving usable InChI strings.

	Refcodes	Refcode families
All structures	186 459	174 465
All structures with $Z' = 2$	19 123	18 365
(% of all structures)	(10.3%)	(10.5%)
All structures in non-Sohnke groups	137 718	128 221
(% of all structures)	(73.9%)	(73.5%)
All structures with $Z' = 2$ in non-Sohnke groups	12 493	11 944
(% of non-Sohnke-group structures)	(9.1%)	(9.3%)
All structures in Sohnke groups	48 741	46 244
(% of all structures)	(26.1%)	(26.5%)
All structures with $Z' = 2$ in Sohnke groups	6630	6434
(% of Sohnke-group structures)	(13.6%)	(13.9%)
Meso molecules in Sohnke groups	725	699
Other achiral molecules in Sohnke groups	8590	7905
Kryptoracemates	181	181
(% of non-Sohnke-group structures plus Sohnke-group structures of unresolvable substances)	(0.1%)	(0.1%)
Unresolvable compounds (achiral, including *meso*, molecules; racemic compounds) in Sohnke groups	9496	8785
(% of Sohnke-group structures)	(19.5%)	(19.0%)
(% of all structures)	(5.1%)	(5.0%)
(% of non-Sohnke-group structures plus Sohnke-group structures of unresolvable substances)	(6.5%)	(6.4%)

spreadsheet. It is still possible, however, that a very few structures (*e.g.* BOGYUS, CUTGAZ and DAQSUJ) that we accepted as kryptoracemates should have been refined in space groups of higher symmetry. It is also possible that there are a few true kryptoracemates among the structures we classified as having been refined in a space group of too-low symmetry.

2.4. Spreadsheet

A spreadsheet available with the supplementary material[11] includes the following:

(i) refcode, year of publication, conventional R factor, space group, Z' for the racemic compound (usually 1, but note that this is not the conventional way of defining Z' for kryptoracemates; see footnote 2);

(ii) r.m.s. deviation of atoms for overlay after inversion of one of the molecules;

(iii) *PLATON* ADDSYM ALERT levels for default and half-default criteria;

(iv) centroid coordinates for possible inversion centers;

(v) comments about molecular overlays and packing pseudosymmetry;

(vi) qualitative (and admittedly subjective) indicators of conformational differences and deviations from improper symmetry (scale of 0–3);

(vii) classification of structure (*e.g.* kryptoracemate, pseudo-kryptoracemate, excluded).

One group of 'excluded' structures includes those 47 in which improper symmetry seems to have been overlooked; 29 of these 47 seem not to have been questioned previously. All 47 are included in a separate section of the spreadsheet.

2.5. Omissions

Kryptoracemates that have no asymmetric atom but have the same kind of chirality as occurs in hexahelicene (HEXHEL) and resolvable 1,1′-binaphthyls (*e.g.* 2,2′-dihydroxy-1,1′-binaphthyl, WANNII) could not be found because we could think of no way to search systematically for such molecules.[12]

At least one probable kryptoracemate was not found because of problems with the CSD entry: the coordinates of one of the two independent molecules in SIDXIM are missing in the original literature. It is likely that there are other similar cases.

Mixed crystals of a racemic compound and either a pure enantiomer or a compound of diastereomers were sometimes found (see above) and sometimes not (*e.g.* VEYBEH, which was not located by the original search but is included in group 2). This inconsistency is a consequence of the way disorder is handled in the CSD entries.[13]

We note the possibility that the CSD includes somewhat disordered, non-Sohnke group structures that should have been described as ordered kryptoracemates.

2.6. Other counts

We also made counts of refcodes and refcode families[14] for various classes of molecules containing the permitted elements. These values are given in Table 1 and were used to calculate actual and estimated frequencies of occurrence of kryptoracemates.

3. Final count

The final list includes 151 crystals (group 1) that are kryptoracemates according to a strict definition. Another 30 crystals (group 2) are kryptoracemates by a slightly relaxed definition (see above) that is at least as restrictive as the definitions used in previous studies. The total number of kryptoracemates according to a conventional definition is then 181. In the discussion that follows groups 1 and 2 will usually be consid-

[11] Supplementary data for this paper are available from the IUCr electronic archives (Reference: GP5035). Services for accessing these data are described at the back of the journal.

[12] AQABID, which is like hexahelicene, is in the list because it also contains an asymmetric C atom.

[13] If the crystal has *e.g.* 50:50 enantiomeric disorder at one site but not the other then the structure was identified as a kryptoracemate if the ordered molecule and the major component (an arbitrary choice for 50:50 disorder) are enantiomers, but was not found if the ordered molecule and the component defined in the CSD as major happen to be homochiral.

[14] The number of refcodes is an overestimate because multiple determinations of the same structure are included. The number of refcode families may be an underestimate because polymorphs belong to the same refcode family.

ered together. The average and median years of publication are 2000 and 2003. The average and median R factors are 0.055 and 0.053.

The total count of 181 is substantially larger than any previous estimate even though the list is more restricted than were previous lists, especially because our list includes neither metal complexes nor salts in which the counterion is enantiomerically pure. It must be noted, however, that the number of kryptoracemates has increased substantially since previous lists were made. Only 39 of the 181 structures on our list had been published by the end of 1995, when Bernal started making lists (Bernal *et al.*, 1996), and only 88 had been published by the end of 2002, when Flack (2003) was making an estimate.

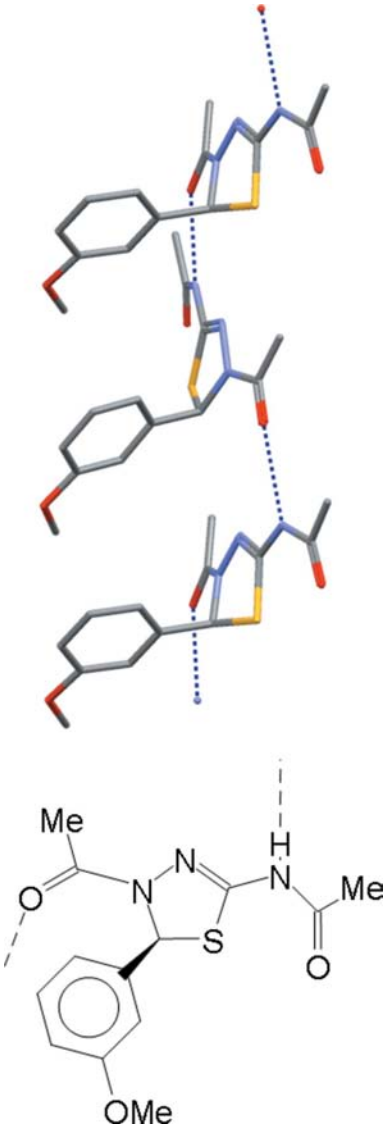

Figure 2
Drawing showing three molecules of IGAREO that form a hydrogen-bonded chain. The first and third molecules are related by translation; the middle molecule is the other enantiomer. The conformational difference between the enantiomers, a *ca* 75° rotation of the −PhOMe substituent, allows an approximate translation to relate that region of the two enantiomers. The chemical line drawing corresponds to the first and third molecules drawn.

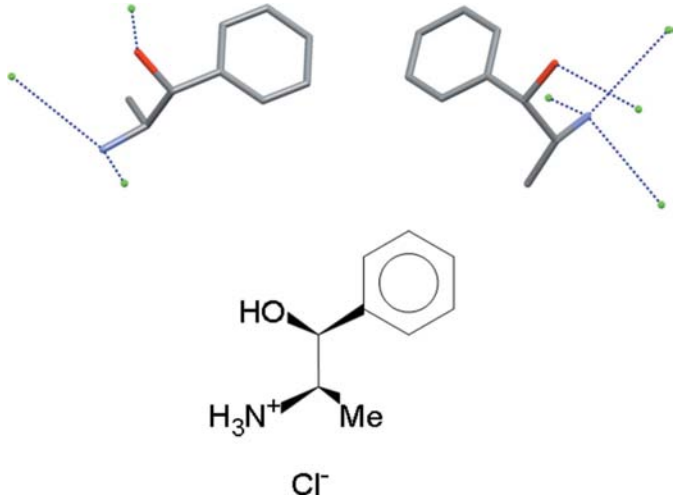

Figure 3
Drawing showing the two molecular conformations and two hydrogen-bonding patterns in NEPHCL. The $H_3N^+-C-C-OH$ torsion angle varies by *ca* 120° between the two cations. The chemical line drawing corresponds to the enantiomer shown on the right. Both −OH substituents are donors in hydrogen bonds to chloride ions. One $-NH_3^+$ subsituent is the donor in hydrogen bonds to three chloride ions; the other $-NH_3^+$ subsituent is the donor in two such hydrogen bonds.

Thirteen of the 151 structures in group 1 have $Z' = 2$ for the racemic compound, *i.e.* have two enantiomeric pairs in the asymmetric unit. The percentage (9%) is just a little lower than that for all structures considered (10%). The percentage is not calculated for group 2 because of the complication of the structures in which the ratio of enantiomers is not 1:1.

4. Discussion

We were surprised not only by the number of kryptoracemates found but by the ease with which the possibility of missed symmetry could usually be ruled out. The deviations from improper symmetry in true kryptoracemates are almost always easy to spot. Sometimes the two enantiomers have conformations that are visibly different. More often the enantiomers have different orientations with respect to the cell axes or are arranged around local or pseudo-inversion centers that have coordinates clearly different from special values like 0, $\frac{1}{4}$, and $\frac{1}{2}$.[15]

4.1. Differences in molecular conformations

The two enantiomers in a kryptoracemate usually have very similar conformations. If the two conformations are not the same the differences between them are almost always small. Examples of small differences include rotation of a phenyl ring by 10–90° (or a substituted phenyl ring by 10–180°; see Fig. 2)

[15] The only space groups in the triclinic, monoclinic or orthorhombic systems that have symmetry elements spaced by 1/8th are *Fdd*2 (No. 43) and *Fddd* (No. 70). A coordinate of a pseudo-inversion center that is close to an odd multiple of 0.125 is therefore seldom a cause for concern.

Table 2
Space groups and obvious pseudosymmetry for the kryptoracemates.

	Group 1			Group 2		
	Number	Pseudo $\bar{1}$	Pseudo *g*	Number	Pseudo $\bar{1}$	Pseudo *g*
*P*1	20	–	12	5	4	–
*C*2	3	–	1	1	–	–
$P2_1$	77	41	5	15	6	1
$P2_12_12_1$	47	30	1	7	3	–
Trigonal, tetragonal	4	3	–	2	–	–
Total	151	74	19	30	13	1

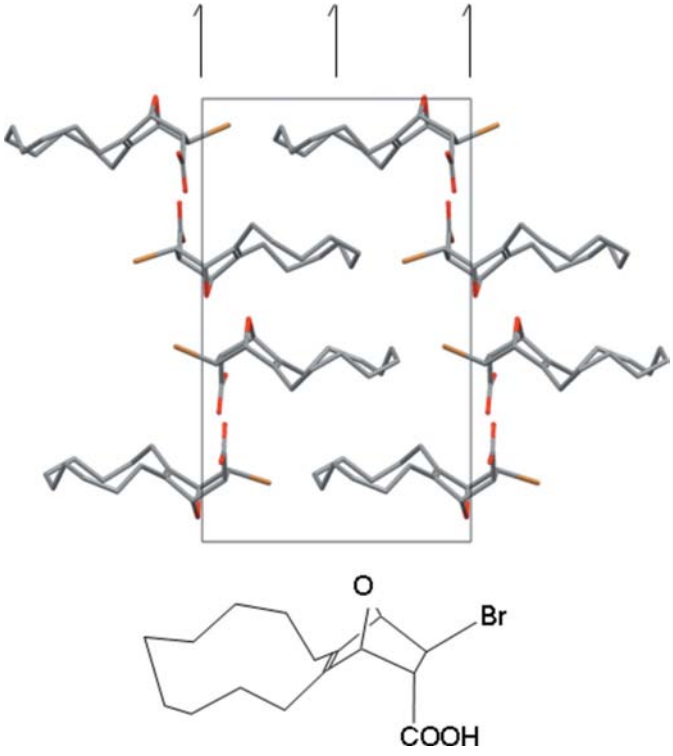

Figure 4
Projection down **a** of the packing in JAGQUD, for which a line drawing of one of the two enantiomers is also shown. The **b** axis of the $P2_1$ cell points upwards and the **c** axis points to the right. Enantiomers, which have nearly identical conformations and form hydrogen-bonded dimers, alternate along **b**. If the pseudo-inversion centers between the carboxylic acid groups were located much closer to $z = 0$ the bridgehead O atoms would interfere. If the centers were located near $z = \frac{1}{4}$ the bridgehead O atom would be too close to methylene groups on the adjacent dodecane ring.

and rotation of a $-CH_2CH_3$, $-$OMe or $-$C(=O)OMe group by *ca* 180° around the bond that attaches it to the rest of the molecule. Differences in rotations of groups including O and N atoms can sometimes be understood as necessary for the formation of a good set of hydrogen bonds (see Fig. 3).[16]

About two-thirds of the enantiomeric pairs on the list have conformations that are essentially indistinguishable. Superposition with inversion of the two molecules shows nearly perfect overlap of atoms. In over half of the structures the r.m.s. deviation of paired non-H atoms is less than 0.15 Å.[17] In about another 20% of the structures the conformational differences are small but easy to spot. We found a major conformational difference in fewer than 10% of the enantiomer pairs.

Dalhus & Görbitz (2000) found an average *r.m.s.* deviation of 0.19 Å for superposition of the non-H atoms of the independent molecules in 114 noncentrosymmetric (but not necessarily kryptoracemic) structures. We found a similar value (0.25 Å) for the 181 kryptoracemates in the final list. The median value (0.14 Å) is much lower.

Structures in which two crystallographically independent enantiomers have essentially the same conformation are sometimes suspected of having been refined in a space group of too-low symmetry. The list of kryptoracemates reported here indicates that the molecular conformations in kryptoracemates are considerably more likely to be indistinguishable than to be different.

4.2. Identifying deviations from improper crystallographic symmetry

In many of the kryptoracemates it is obvious that improper crystallographic symmetry is absent. We found local or pseudosymmetry in just under 60% of the group 1 (strict definition) kryptoracemates, and in under 50% of the group 2 kryptoracemates

In many of the structures without pseudosymmetry it is easy to spot differences in the orientations of the independent enantiomers with respect to the crystallographic axes. The polarity of axes is also a good test. In a triclinic crystal either all axes are polar or none is polar, so a single polar axis guarantees the Sohnke group. It is usually easy to see if all the bonds of one type (*e.g.* a $-$C=O bond) have axial projections with the same sign. In the case of a monoclinic crystal a polar **b** axis (assuming the conventional setting) guarantees a Sohnke group.

In the cases of local or pseudosymmetry, especially inversion symmetry, the coordinates of the centroid of the related molecules are a good diagnostic tool because the coordinates are averages over so many independent values. If the original space group is of monoclinic or higher symmetry then at least two of the coordinates of the pseudocenter must have a special value like 0, $\frac{1}{2}$ or $\frac{1}{4}$ if higher symmetry has been missed. If a centroid coordinate deviates by 0.02 from a special value and the corresponding cell constant is ≥ 10 Å, then the centroid is displaced by at least 0.2 Å from a special position and there is

[16] See also Marsh (1999), which includes the statement 'Another method of attaining approximate centrosymmetry, particulary for amino acids, is for an N atom and an O atom to interchange places in an otherwise centrosymmetric coordination or hydrogen-bonding arrangement'.

[17] The r.m.s. deviation of a set of paired non-H atoms is an imperfect measure of conformational differences because its magnitude also varies with the size of the rest of the molecule and the precision of the determination. We found cases in which the r.m.s. deviation for two enantiomers having essentially the same conformation was larger than for two enantiomers in which the rotation of a phenyl ring varied by *ca* 10–15°.

little chance that the symmetry is crystallographic. In the case of a precisely determined structure finer distinctions can be made (see Marsh, 1999).[18]

4.3. Frequency of pseudo-inversion centers and glide planes

Local or pseudo-inversion centers are present in roughly half of the kryptoracemates, but in the group 1 racemates they are concentrated (see Table 2) in structures having 2_1 axes. Most of the pseudocenters in both groups 1 and 2 have at least one coordinate that is not special, *i.e.* that is clearly displaced from $x_i = 0, \frac{1}{4}$ and $\frac{1}{2}$. In some other structures the pseudocenters for two independent pairs of enantiomers, or for the cations and the anions, do not coincide. Sometimes it is possible to understand the reason for the displacement (*e.g.* JAGQUD; see Fig. 4), but sometimes it is not.

Pseudo-glide planes are much less frequent (see Table 2) than pseudo-inversion centers. The pseudo-glides are concentrated in the *P*1 structures of the group 1 kryptoracemates. In these structures there may be a very good pseudo-glide plane, but if so the glide is not perpendicular to a crystallographic translation (see, *e.g.*, Fig. 5). In such a case sets of molecules are related by pseudo-glide planes, but there is slippage between adjacent sets so that the glide operation is incompatible with the translational symmetry.

It is interesting that none of the 20 *P*1 structures of the group 1 kryptoracemates is even approximately centrosymmetric while four of the five *P*1 structures of the group 2 kryptoracemates are. Perhaps this difference is just an accident of the small numbers of structures found, but it may be a consequence of there being two different components present in each of the group 2 structures. Another factor may be hydrogen bonding; about half the group 1 *P*1 structures include chains of molecules linked by hydrogen bonds.

4.4. Analogy with $Z' = 2$ structures

Molecules are expected to crystallize in low-symmetry space groups (especially $P2_1/c$ and $P\overline{1}$) with asymmetric units that are as small as possible. This expectation is derived from space-group statistics (see Brock & Dunitz, 1994; Pidcock *et al.*, 2003, and references therein) and can be understood as indicating that the number of different intermolecular interactions is usually minimized. Kryptoracemates are then like the more general class of $Z' = 2$ structures in having an asymmetric unit that is twice as large as expected.

Sometimes the larger asymmetric unit can be related to the requirements for forming good hydrogen bonds (*e.g.* Brock & Duncan, 1994; Brock, 2002). Several of the kryptoracemates (see Figs. 2 and 3) can be understood in that way.

Not identified among the kryptoracemates is the large subgroup of $Z' = 2$ structures that includes pseudosymmetric (and usually twinned) low-temperature structures that result from cooling through a phase transition during which the asymmetric unit is doubled. Some of the kryptoracemate structures were done at low temperature and some are very pseudosymmetric, but we found no pair of closely related phases. Three of the kryptoracemates (HISRIL01, NOLFUP and YIXVAD) have a second polymorph that is centrosymmetric, but in each case the packing in the two polymorphs is very different.

The analogy with $Z' = 2$ structures suggests that it may not have been necessary to exclude the three structures

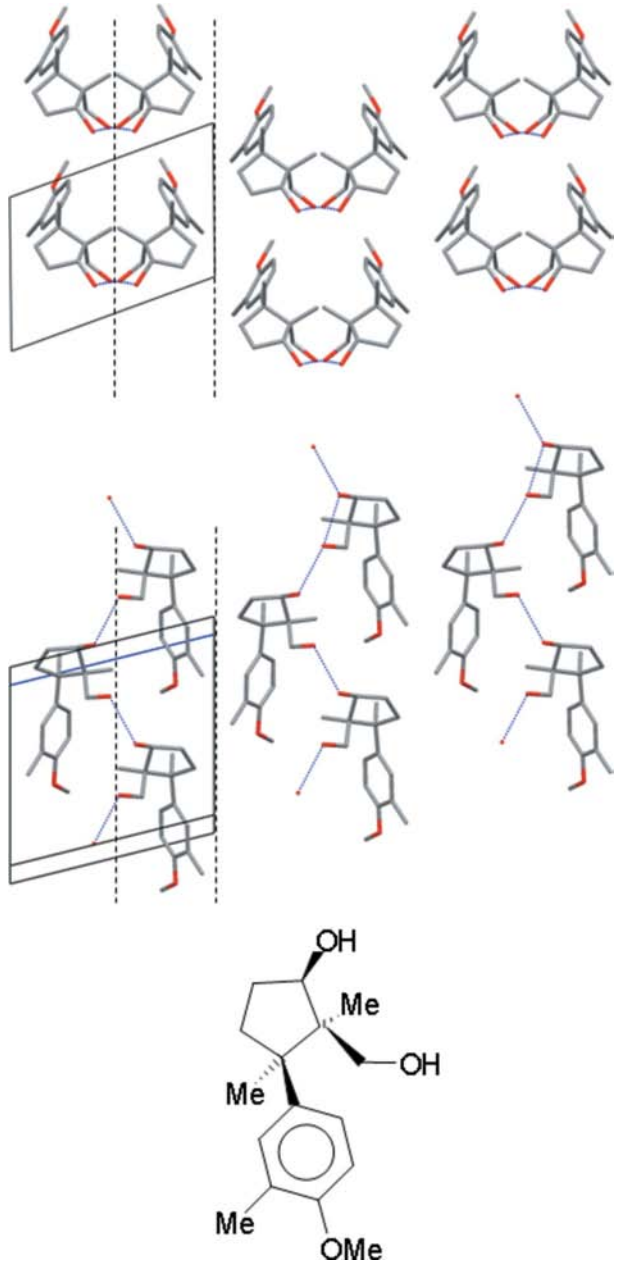

Figure 5
Two views and a chemical line drawing for the structure NAJCUX. The upper packing diagram shows a projection down **b**; the **a** axis points vertically downward. The lower diagram, which is related to the upper by a 90° rotation about the horizontal, shows that there is no crystallographic axis perpendicular to the pseudo-glide plane, which is indicated by dashed lines. This second diagram includes molecules having centroids with $0 < x < 1$, $-0.5 < y < 1$ and $0 < z < 3$. The line drawing corresponds to the enantiomer shown at the far left of the lower drawing.

[18] Note that the position of the centroid gives quite different information than does the r.m.s. deviation from centrosymmetry. If the local symmetry is strong the r.m.s. deviation can be very low even if the local center has coordinates that are far from special values like 0, $\frac{1}{4}$ and $\frac{1}{2}$.

(BSADAZ, CHATRZ and YEPLAH) containing conformational enantiomers (*i.e.* enantiomers that are expected to racemize rapidly).

4.5. Comments on the frequency of occurrence of kryptoracemates

Pidcock *et al.* (2003) found that 99% of molecules for which the only possible symmetry is inversion symmetry are found on $\overline{1}$ sites. A 1:1 racemic compound can also occupy a $\overline{1}$ site, although crystal packing of racemic compounds is not usually described in that way. The number of kryptoracemates found in this study, however, shows that their frequency is substantially lower than 1%; a better estimate would be *ca* 0.1% (see Table 1). An important difference between a 1:1 aggregate of two enantiomers and a possibly centrosymmetric molecule is that the size and shape of the two-molecule aggregate can be adjusted to a much greater extent to optimize the crystal packing. Intermolecular distances (such as those between the two enantiomers) are much more variable than bond lengths. This additional freedom means that the fraction of racemic compounds that can conform to inversion symmetry in a well packed crystal is greater than the fraction of possibly centrosymmetric molecules that can do so.

Another approach to an estimate is to multiply the fraction of *meso* molecules, other achiral molecules, and racemates that crystallize in Sohnke groups by the fraction of molecules that crystallize with $Z' = 2$. The first term gives the fraction of molecules (and enantiomeric pairs) that crystallize in a proper space group when they could crystallize in a more common non-Sohnke group. The second term gives the fraction of molecules that crystallize with an asymmetric unit larger than is required by the space-group symmetry. The resulting value is *ca* (0.06)(0.10) or 0.6%, which is somewhat higher than the estimate based on our count of kryptoracemates.

4.6. The real oddities

The most unusual group of structures found includes the 11 instances of 'unbalanced crystallization' in which the ratio of enantiomers is not 1:1. A two-component *T*–*X* phase diagram for these systems would have at least four maxima: one for each pure enantiomer and two for the D:L compounds, usually with compositions 1:2 and 2:1.[19] This group of structures demonstrates that if a type of crystal packing can be imagined it can very often be found.

The structure of ABADUD (Basak *et al.*, 2004) is noteworthy because the 'third' molecule present is very nearly the same as one of the enantiomers but not quite; in the third molecule a $-CMe_2-$ group is replaced by a $-CHMe-$ group. This structure is a striking example of a failure of fractional crystallization.

4.7. A comment about spontaneous resolution

Scientists have long wondered about the frequency of spontaneous resolution, first observed by Pasteur for sodium ammonium tartrate (see Flack, 2009). While it has been generally assumed that most substances that crystallize in Sohnke groups can be resolved, the split between Sohnke-group crystals grown from enantiomerically pure and from enantiomerically impure materials is unknown and probably unknowable. It has long been recognized, however, that a substantial number of achiral and *meso* molecules crystallize in Sohnke groups; a well known example is the $P3_1/P3_2$ polymorph of glycine, which is reported to be the most stable of the three known phases (Perlovich *et al.*, 2001). What was not known is the fraction of the Sohnke-group crystals that would have to have been grown from unresolved or unresolvable material.[20]

We found that 19–20% of the Sohnke-group structures that include the atom types considered here are *meso* or otherwise achiral (see Table 1). This rather surprising result may reduce any previous estimate of the likelihood of spontaneous resolution.

The best available estimate of the frequency of spontaneous resolution may come from knowing the partition of structures of racemic compounds, *meso* molecules, and achiral molecules between Sohnke and non-Sohnke groups. This estimate assumes that the split between the two kinds of groups is similar for resolvable and unresolvable materials. The percentage in Sohnke groups (6.4–6.5%; see Table 1) may be a little high because the count of achiral molecules includes resolvable hexahelicenes, 1,1′-binaphthyls *etc.*, but if apparently achiral molecules that might have separable conformational enantiomers are removed from the count the percentage drops to no lower than 6.1%. Our rough estimate of the probability of spontaneous resolution is then a little above 6%.

5. Summary

Kryptoracemates are materials that crystallize in Sohnke (*i.e.* proper) space groups when they could crystallize without disorder in non-Sohnke groups, of which centrosymmetric groups are a subset. The compiled list of 151 kryptoracemates that fit this definition plus 30 more that meet a slightly relaxed definition shows that they are rare, but not as rare as previously thought, even though the definition used for this list is more restrictive than the definitions used by previous authors.

The conformations of the two enantiomers in kryptoracemates are usually very similar. Strong similarity of enantiomeric conformations is not, by itself, a good reason to suspect a space-group error.

The deviations from improper symmetry in kryptoracemates are usually easy to spot. Many kryptoracemates are

[19] A maximum for a 1:1 racemate might also be present.

[20] The frequency (26%) of Sohnke-group structures for the atom types considered here is higher than for the CSD as a whole (18%), There are probably several reasons for this difference. First, organic crystals are almost certainly grown from resolved material more often than are crystals of metal-containing compounds. Second, organometallic complexes are more likely to have a low-energy, centrosymmetric conformation than are organic molecules. If such a conformation is accessible, crystallization in a space group that includes inversion symmetry is almost certain (Pidcock *et al.*, 2003).

not even especially pseudosymmetric. If there is pseudosymmetry the deviations from improper symmetry are often obvious: coordinates of a pseudo-inversion center have values that are not 0, $\frac{1}{4}$ and $\frac{1}{2}$, or the normal to a pseudo-glide plane is not parallel to any crystallographic axis.

The estimated frequency of organic kryptoracemates (*ca* 0.1%) is about ten times lower than an estimate based on the probability of retaining a molecular inversion center. A pair of enantiomers is more likely to conform to inversion symmetry than is a potentially centrosymmetric molecule because the van der Waals surface of a centrosymmetric pair can be more easily varied to optimize the packing than can the surface of a single molecule. If a molecule is located on an inversion center it has no translational freedom, but if an enantiomer pair is located on an inversion center there are still three adjustable molecular translations. Although the overall shape and size of the enantiomer pair may be limited by the inversion symmetry and by the requirement for close packing, the surface of the pair is still considerably more variable than is the surface of a single molecule. The low frequency of kryptoracemates is yet another indication of how favorable inversion symmetry is for crystal packing.

The fraction of crystals in Sohnke groups that contain *meso*, otherwise achiral, or racemic material (19–20%) is higher than expected.

Somewhat over 6% of molecules that could crystallize in a group that includes improper symmetry operations crystallize in a Sohnke group instead. This percentage may be as good an estimate as any of the frequency of spontaneous resolution.

We are grateful to Professor Ton Spek (University of Utrecht) for very useful advice. It is a pleasure to acknowledge the contributions of Dr Frank Allen of the CCDC, with whom we have had many interesting and stimulating discussions over many years. LF thanks Pfizer for funding *via* the Pfizer Institute for Pharmaceutical Materials Science.

References

Albano, V. G., Bellon, P. L. & Sansoni, M. (1969). *Chem. Commun.* pp. 899–901.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Basak, A., Bag, S. S., Mazumdar, P. A., Bertolasi, V. & Das, A. K. (2004). *J. Chem. Res.* pp. 318–321.
Bernal, I. (1995*a*). ACA Annual Meeting, Montreal, Quebec, Canada. Abstract 4a.1.e.
Bernal, I. (1995*b*). Personal communication.
Bernal, I., Cai, J., Massoud, S. S., Watkins, S. F. & Fronczek, F. R. (1996). *J. Coord. Chem.* **38**, 165–181.
Bishop, R. & Scudder, M. L. (2009). *Cryst. Growth Des.* **9**, 2890–2894.
Brock, C. P. (2002). *Acta Cryst.* B**58**, 1025–1031.
Brock, C. P. & Duncan, L. L. (1994). *Chem. Mater.* **6**, 1307–1312.
Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.
Brock, C. P., Schweizer, W. B. & Dunitz, J. D. (1991). *J. Am. Chem. Soc.* **113**, 9811–9820.
Cai, J.-W., Hu, X.-P., Chen, C.-H. & Ji, L.-N. (2001). *Acta Cryst.* C**57**, 394–396.
Clemente, D. A. (2003). *Tetrahedron*, **59**, 8445–8455.
Dalhus, B. & Görbitz, C. H. (2000). *Acta Cryst.* B**56**, 715–719.
Flack, H. D. (2003). *Helv. Chim. Acta*, **806**, 905–921.
Flack, H. D. (2009). *Acta Cryst.* A**65**, 371–389.
Herbstein, F. H. & Marsh, R. E. (1998). *Acta Cryst.* B**54**, 677–686.
Kapon, M. & Reisner, G. M. (1990). *Acta Cryst.* C**46**, 349–350.
Kwiatkowski, S., Syed, A., Brock, C. P. & Watt, D. S. (1989). *Synthesis*, pp. 818–820.
Lineberry, A. M., Benjamin, E. T., Davis, R. E., Kassel, W. S. & Wheeler, K. A. (2008). *Cryst. Growth Des.* **8**, 612–619.
Lynch, V. M., Bur, S. K. & Martin, S. F. (1999). *Acta Cryst.* C**55**, 622–624.
Macrae, C. F., Bruno, I. J., Chisholm, J. A., Edgington, P. R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. & Wood, P. A. (2008). *J. Appl. Cryst.* **41**, 466–470.
Marsh, R. E. (1994*a*). *Acta Cryst.* A**50**, 450–455.
Marsh, R. E. (1994*b*). *Acta Cryst.* B**50**, 112–116.
Marsh, R. E. (1995). *Acta Cryst.* B**51**, 897–907.
Marsh, R. E. (1997). *Acta Cryst.* B**53**, 317–322.
Marsh, R. E. (1999). *Acta Cryst.* B**55**, 931–936.
Marsh, R. E. (2002). *Acta Cryst.* B**58**, 893–899.
Marsh, R. E. (2004). *Acta Cryst.* B**60**, 252–253.
Marsh, R. E. (2005). *Acta Cryst.* B**61**, 359.
Marsh, R. E. (2009). *Acta Cryst.* B**65**, 782–783.
Marsh, R. E. & Bernal, I. (1995). *Acta Cryst.* B**51**, 300–307.
Marsh, R. E. & Clemente, D. A. (2007). *Inorg. Chim. Acta*, **360**, 4017–4024.
Marsh, R. E. & Herbstein, F. H. (1988). *Acta Cryst.* B**44**, 77–88.
Marsh, R. E., Kapon, M., Hu, S. & Herbstein, F. H. (2002). *Acta Cryst.* B**58**, 62–77.
Marsh, R. E. & Spek, A. L. (2001). *Acta Cryst.* B**57**, 800–805.
Morales, G. A. & Fronczek, F. R. (1996). *Acta Cryst.* C**52**, 1266–1268.
Perlovich, G. L., Hansen, L. K. & Bauer-Brandl, A. (2001). *J. Therm. Anal. Calorim.* **66**, 699–715.
Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2003). *Acta Cryst.* B**59**, 634–640.
Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.
Stein, S. E., Heller, S. R. & Tchekhovskoi, D. (2003). *An Open Standard for Chemical Structure Representation: The IUPAC Chemical Identifier*, in Proceedings of the 2003 International Chemical Information Conference (Nimes), Infonortics, pp. 131–143, http://www.iupac.org/inchi.
Stein, S. E., Heller, S. R. & Tchekhovskoi, D. V. (2006). *The IUPAC Chemical Identifier – Technical Manual*, National Institute of Standards and Technology, Gaithersburg, Maryland, US 20899-8380.
Wong-Ng, W., Cheng, P.-T. & Nyburg, S. C. (1984). *Acta Cryst.* B**40**, 151–158.