Acta Crystallographica Section B
**Structural Science**
ISSN 0108-7681

# Non-centrosymmetric racemates: space-group frequencies and conformational similarities between crystallographically independent molecules

**Bjørn Dalhus* and Carl Henrik Görbitz**

Department of Chemistry, University of Oslo, PO Box 1033, Blindern, N-0315 Oslo, Norway

Correspondence e-mail: bjornda@kjemi.uio.no

Received 8 November 1999
Accepted 7 February 2000

DL-Allylglycine (DL-2-amino-4-pentenoic acid, $C_5H_9NO_2$) yields crystals with $Pca2_1$ symmetry and two crystallographically independent yet pseudo-inversion-related enantiomers. The distribution among the common space groups of other crystalline racemates with more than one molecule in the asymmetric unit has been established. The conformational similarities between crystallographically independent enantiomers in 114 non-centrosymmetric racemates were quantified using the r.m.s. deviation for a molecular superposition. The analysis shows that in the majority of crystals the conformations of the crystallographically independent molecules are very similar with mean r.m.s. deviation = 0.190 Å. In almost 80% of the structures the mean r.m.s. deviations is in the interval 0–0.2 Å. It is estimated that racemates constitute 23% of the centrosymmetric organic structures in the Cambridge Structural Database.

## 1. Introduction

The structure of DL-allylglycine (DL-2-amino-4-pentenoic acid) was determined as part of an ongoing programme to provide accurate H-atom positions in crystal structures of hydrophobic amino acids (Dalhus & Görbitz, 1999, and references therein). Intuitively, we had expected this racemate to crystallize in a space group with a center of symmetry, but surprisingly the space group proved to be non-centrosymmetric ($Pca2_1$). To our knowledge, no systematic investigation dealing with this phenomenon has been presented previously. We thus decided to retrieve the relevant data from the Cambridge Structural Database (CSD; Allen & Kennard, 1993) to establish the frequency with which racemates crystallize in non-centrosymmetric space groups, and also the degree of similarity between the two enantiomeric molecules in the asymmetric unit.

## 2. Experimental

Racemic DL-allylglycine was obtained from Sigma, and an aqueous solution was mixed with tetramethoxysilane in the ratio 10:1 (volume/volume) and left for some minutes to polymerize. Relatively large crystals appeared as ethanol diffused into the gel at room temperature.

The structure was solved and refined with *SHELXTL* (Sheldrick, 1994). All non-H atoms were refined anisotropically. Amino H atoms were refined isotropically, while other H atoms were placed geometrically and refined with constraints to keep all C—H distances and all C—C—H angles on one C atom the same. Further, isotropic displacement parameters for the H atoms were fixed at $1.2U_{eq}$ of the bonded

**Table 1**
Experimental details.

| | DL–Allylglycine |
|---|---|
| Crystal data | |
| Chemical formula | $C_5H_9NO_2$ |
| Chemical formula weight | 115.13 |
| Cell setting | Orthorhombic |
| Space group | $Pca2_1$ |
| $a$ (Å) | 9.8588 (5) |
| $b$ (Å) | 4.8125 (3) |
| $c$ (Å) | 25.0546 (13) |
| $V$ (Å$^3$) | 1188.73 (11) |
| $Z$ | 8 |
| $D_x$ (Mg m$^{-3}$) | 1.287 |
| Radiation type | Mo $K\alpha$ |
| Wavelength (Å) | 0.71073 |
| No. of reflections for cell parameters | 5506 |
| $\theta$ range (°) | 2.44–39.66 |
| $\mu$ (mm$^{-1}$) | 0.099 |
| Temperature (K) | 150 (2) |
| Crystal form | Plate |
| Crystal size (mm) | 0.40 × 0.40 × 0.10 |
| Crystal color | Colorless |
| | |
| Data collection | |
| Diffractometer | Siemens Smart CCD |
| Data collection method | $\omega$ scans |
| $\omega$-scan width (°) | 0.6 |
| No. of sets of exposures | 6 |
| Exposure time per frame (s) | 60 |
| Crystal-to-detector distance (cm) | 4.98 |
| Absorption correction | Multi-scan empirical (*SADABS*; Sheldrick, 1996) |
| $T_{min}$ | 0.961 |
| $T_{max}$ | 0.990 |
| No. of measured reflections | 26 563 |
| No. of independent reflections | 7312 |
| No. of observed reflections | 6251 |
| Criterion for observed reflections | $F > 4\sigma(F)$ |
| $R_{int}$ | 0.0369 |
| $\theta_{max}$ (°) | 40.10 |
| Range of $h$, $k$, $l$ | $-17 \rightarrow h \rightarrow 17$<br>$-8 \rightarrow k \rightarrow 8$<br>$-45 \rightarrow l \rightarrow 43$ |
| | |
| Refinement | |
| Refinement on | $F^2$ |
| $R[F^2>2\sigma(F^2)]$ | 0.0440 |
| $wR(F^2)$ | 0.1107 |
| $S$ | 1.054 |
| No. of reflections used in refinement | 7312 |
| No. of parameters used | 177 |
| H-atom treatment | Mixed |
| Weighting scheme | $w = 1/[\sigma^2(F_o^2) + (0.07P)^2]$, where $P = (F_o^2 + 2F_c^2)/3$ |
| $(\Delta/\sigma)_{max}$ | 0.000 |
| $\Delta\rho_{max}$ (e Å$^{-3}$) | 0.457 |
| $\Delta\rho_{min}$ (e Å$^{-3}$) | −0.286 |
| Extinction method | None |
| Source of atomic scattering factors | *International Tables for Crystallography* (1992, Vol. C, Tables 4.2.6.8 and 6.1.1.4) |
| | |
| Computer programs | |
| Data collection | *SMART* (Siemens, 1995) |
| Cell refinement | *SAINT* (Siemens, 1995) |
| Data reduction | *SAINT* (Siemens, 1995) |
| Structure solution | *SHELXTL* (Sheldrick, 1994) |
| Structure refinement | *SHELXTL* (Sheldrick, 1994) |

C atoms. The absolute structure cannot be unambiguously determined from the present refinement. However, a Flack (1983) parameter of 0.1 (5) suggests that the chosen polarity is correct, hence the Friedel pairs were not merged. Full experimental details are given in Table 1 and fractional coordinates in Table 2.

In this study only CSD (Allen & Kennard, 1993, October 1998 release) database entries including atomic coordinates have been accepted and the 'only organics' screen was also enabled where appropriate. Throughout the years various compounds have been subject to several X-ray structure investigations, reflected by the occurrence of more than one entry in the CSD, all within the same refcode family. By using an in-house program that keeps only one structure for each such family, care has been taken to eliminate duplicate structures in the crystallographic material.

## 3. Molecular and crystal structure of DL-allylglycine

The Bravais lattice is primitive orthorhombic with systematic absences corresponding to either $Pca2_1$ (non-centrosymmetric) or *Pcam* (centrosymmetric). It was not possible to conclude on the presence of a crystallographic inversion center from statistics for the normalized structure factors ($|E^2 - 1| = 0.831$ compared with the theoretical values 0.968 and 0.736 for centrosymmetric and non-centrosymmetric space groups, respectively). Attempts to solve the structure in the centrosymmetric space group *Pcam* were uniformly unsuccessful, but the structure was solved immediately in $Pca2_1$ and refined to a final $R = 0.044$. There are two enantiomers ($A$ and $B$, Fig. 1) in the asymmetric unit, related by a pseudo-inversion center located at $x = 0.3988$, $y = 0.2516$, $z = 0.1458$. Refinement in the centrosymmetric space group *Pcam* (one molecule in the asymmetric unit) was carried out after

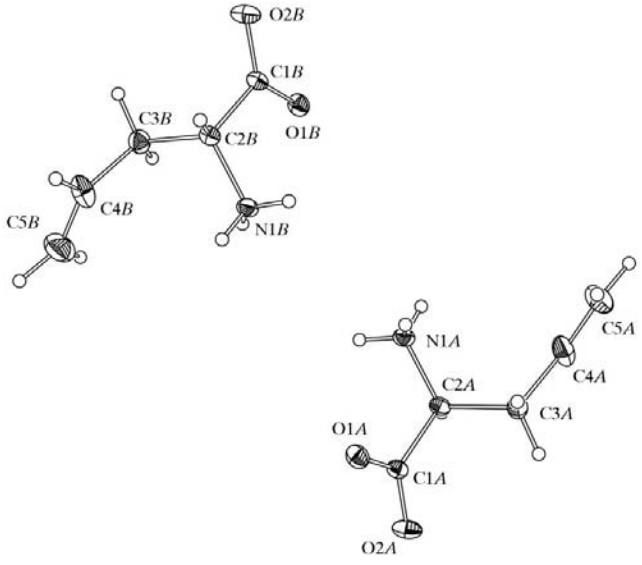


**Figure 1**
Molecular diagram with atomic numbering of DL-allylglycine. Displacement ellipsoids are drawn at the 50% probability level. H atoms are arbitrarily scaled.

**Table 2**
Fractional atomic coordinates and equivalent isotropic displacement parameters (Å$^2$).

$U_{eq} = (1/3)\Sigma_i\Sigma_j U^{ij} a^i a^j \mathbf{a}_i.\mathbf{a}_j$.

| | $x$ | $y$ | $z$ | $U_{eq}$ |
|---|---|---|---|---|
| O1*A* | 0.09520 (6) | 0.56822 (12) | 0.18784 (3) | 0.01745 (10) |
| O2*A* | −0.06823 (6) | 0.31216 (14) | 0.22600 (3) | 0.02104 (12) |
| N1*A* | 0.28759 (7) | 0.18606 (15) | 0.20740 (3) | 0.01549 (11) |
| C1*A* | 0.05368 (7) | 0.36671 (15) | 0.21479 (3) | 0.01362 (11) |
| C2*A* | 0.16103 (7) | 0.17524 (15) | 0.23957 (3) | 0.01352 (11) |
| C3*A* | 0.18821 (8) | 0.27327 (18) | 0.29687 (3) | 0.01810 (13) |
| C4*A* | 0.28927 (12) | 0.1015 (2) | 0.32668 (4) | 0.02662 (18) |
| C5*A* | 0.40348 (12) | 0.2001 (3) | 0.34706 (5) | 0.0387 (3) |
| O1*B* | 0.70017 (6) | −0.07034 (13) | 0.10500 (3) | 0.01816 (11) |
| O2*B* | 0.86465 (6) | 0.18270 (14) | 0.06682 (3) | 0.02159 (13) |
| N1*B* | 0.50974 (6) | 0.31731 (15) | 0.08508 (3) | 0.01513 (11) |
| C1*B* | 0.74251 (7) | 0.13005 (15) | 0.07796 (3) | 0.01383 (11) |
| C2*B* | 0.63630 (7) | 0.32251 (15) | 0.05285 (3) | 0.01361 (11) |
| C3*B* | 0.60759 (9) | 0.22035 (19) | −0.00413 (3) | 0.01938 (14) |
| C4*B* | 0.51141 (12) | 0.3988 (2) | −0.03492 (4) | 0.02719 (18) |
| C5*B* | 0.39443 (13) | 0.3126 (4) | −0.05470 (5) | 0.0405 (3) |

the proper origin shift. The refinement gave a final $R$ factor > 0.50. Furthermore, the molecular packing arrangement shows that the pseudo-inversion center is only a local pseudo-symmetry element (Fig. 2). We conclude that the space group is unambigously $Pca2_1$.

Corresponding bond lengths and angles are very similar in the two molecules. There are some small, but still significant differences between torsion angles (Table 3), the largest being 2.3° for N1—C2—C3—C4 ($\chi^1$) after inversion of *B* (or *A*).

Molecules *A* and *B* are stacked in consecutive layers parallel to the *xy* plane, and the crystal structure contains distinct hydrophilic and hydrophobic layers, Fig. 2. Our investigations (Dalhus & Görbitz, 1999) have revealed three major classes of molecular packing arrangements, each with a distinct hydrogen-bond pattern. The present racemate belongs to class I, as do the racemates of other unbranched amino acids such as DL-norleucine and DL-methionine (Dalhus & Görbitz, 2000). In this class two of the three amino H atoms in each independent molecule (H1*A*/H2*A* in *A*, H1*B*/H2*B* in *B*) form chains of hydrogen-bonded glide-plane-related enantiomers along **a**, Fig. 2. H3*A* and H3*B*, on the other hand, form hydrogen bonds across the hydrophilic layer joining pseudo-twofold related molecules with the same chirality at $C^{\alpha}$. Geometric parameters for all hydrogen bonds are given in Table 4.

## 4. Statistical analysis of racemates in non-centrosymmetric space groups

Unfortunately, there is no efficient way of searching the Cambridge Structural Database for all racemic compounds. The only search field containing information about the (possible) stereochemistry of the registered compounds is the *NAME* field.

An initial search of CSD using various combinations of the stereochemical designators *D*, *L*, *R*, *S*, + and − as well as the strings '*racemate*', '*racemic*' and '*rac*' returned 1926 structures of racemic compounds, less than 2.5% of the organic crystal structures in the CSD. As this number seemed unreasonably low, a new strategy was employed.

The 77 986 unique organic structures in the CSD were separated into centrosymmetric (50 291 hits, 64.5%) and non-centrosymmetric (27 695 hits, 35.5%) space groups using appropriate bit screens in the CSD search program *QUEST* (Allen & Kennard, 1993). 6000 selected centrosymmetric structures (3 series, each with 2000 structures, from different parts of the database) yielded 5560 unique structures (93%). They were all manually inspected for chiral atoms with rejection of achiral as well as *meso* structures. In this way a total of 1263 racemates were identified. Assuming that both the racemates and the centrosymmetric organic structures in general are evenly distributed over the entire database, the estimated relative number of centrosymmetric organic racemates is around 23%, that is, approximately 11 500 structures. The majority (98.0%) of the 1263 retrieved racemates belong to the four space groups $P\bar{1}$ (26.6%), $P2_1/c$ (57.1%), $C2/c$ (7.6%) and $Pbac$ (6.7%), Table 5. $C2/c$ (7.6%) and $Pbac$ (6.7%), Table 5. For these space groups, there is no obvious

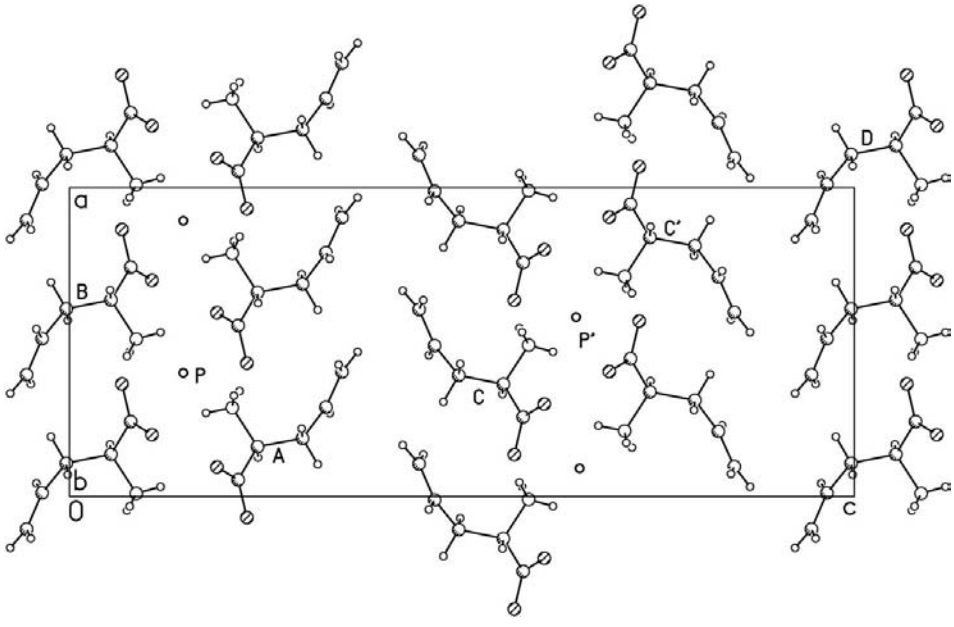


**Figure 2**
Packing diagram viewed down the *b* axis. Pseudo-inversion centers within the unit cell are marked with open circles. The pseudo-inversion center at $x = 0.3988$, $y = 0.2516$, $z = 0.1458$ is labelled *P*. Molecules *C* and *C*′ are related by the pseudo-inversion center at *P*′, but molecules *A* and *D* have the same chirality. Hence, the pseudo-inversion symmetry is local and confined to separate molecular layers.

**Table 3**
Selected geometric parameters (Å, °).

| | | | |
|---|---|---|---|
| O1*A*—C1*A* | 1.2505 (9) | O1*B*—C1*B* | 1.2503 (10) |
| O2*A*—C1*A* | 1.2619 (10) | O2*B*—C1*B* | 1.2618 (10) |
| N1*A*—C2*A* | 1.4862 (10) | N1*B*—C2*B* | 1.4864 (10) |
| C1*A*—C2*A* | 1.5345 (10) | C1*B*—C2*B* | 1.5330 (10) |
| C2*A*—C3*A* | 1.5349 (11) | C2*B*—C3*B* | 1.5364 (11) |
| C3*A*—C4*A* | 1.4946 (12) | C3*B*—C4*B* | 1.4938 (13) |
| C4*A*—C5*A* | 1.3244 (17) | C4*B*—C5*B* | 1.3220 (17) |
| | | | |
| O1*A*—C1*A*—C2*A*—N1*A* | −26.27 (9) | O1*B*—C1*B*—C2*B*—N1*B* | 26.94 (9) |
| O2*A*—C1*A*—C2*A*—N1*A* | 157.17 (7) | O2*B*—C1*B*—C2*B*—N1*B* | −156.59 (7) |
| N1*A*—C2*A*—C3*A*—C4*A* | −60.95 (10) | N1*B*—C2*B*—C3*B*—C4*B* | 63.23 (10) |
| C2*A*—C3*A*—C4*A*—C5*A* | 120.91 (12) | C2*B*—C3*B*—C4*B*—C5*B* | −120.03 (12) |

**Table 4**
Hydrogen bonds in DL-allylglycine.

N—H, H···O[a] and N—H···O based on experimental H-atom positions, H···O[b] for N—H bonds normalized to 1.030 Å (Taylor & Kennard, 1983).

| N—H···O | N—H | H···O[a] | H···O[b] | N···O | N—H···O |
|---|---|---|---|---|---|
| N1*A*—H1*A*···O2*A*$^{i}$ | 0.86 (2) | 1.99 (2) | 1.824 | 2.841 (1) | 170 (2) |
| N1*A*—H2*A*···O2*A*$^{ii}$ | 0.80 (2) | 2.07 (2) | 1.852 | 2.826 (1) | 159 (2) |
| N1*A*—H3*A*···O1*B*$^{iii}$ | 0.86 (1) | 1.91 (1) | 1.749 | 2.763 (1) | 168 (1) |
| N1*B*—H1*B*···O2*B*$^{iii}$ | 0.93 (1) | 1.91 (1) | 1.844 | 2.837 (1) | 172 (1) |
| N1*B*—H2*B*···O2*B*$^{iv}$ | 0.92 (2) | 1.95 (2) | 1.813 | 2.836 (1) | 162 (1) |
| N1*B*—H3*B*···O1*A*$^{i}$ | 0.82 (2) | 1.95 (2) | 1.737 | 2.764 (1) | 175 (2) |

Symmetry codes: (i) $x+\frac{1}{2}, -y+1, z$; (ii) $x+\frac{1}{2}, -y, z$; (iii) $x-\frac{1}{2}-y, z$; (iv) $x-\frac{1}{2}, -y+1, z$.

**Table 5**
Distribution of 1263 selected racemates among 5560 organic structures in centrosymmetric space groups.

$Z_{eff}$ = reported $Z$ value, $Z_0$ = standard $Z$ value given in *International Tables for Crystallography* (1996, Vol. A).

| Space group | Racemates | Racemates with $Z_{eff} > Z_0$ | Overall in CSD with $Z_{eff} > Z_0$ |
|---|---|---|---|
| $P\bar{1}$ | 336 | 69 (20.5%) | 15.5% |
| $P2_1/c$ | 721 | 52 (7.2%) | 7.9% |
| $C2/c$ | 96 | 3 (3.1%) | 4.5% |
| *Pbca* | 85 | 2 (2.4%) | 4.2% |
| Others | 25† | – | - |

† *P*2/*c* (1), *Pcca* (1), *Pccn* (1), *Pbcn* (5), *Pnma* (1), $P4_2/n$ (4), $I4_1/a$ (4), $R\bar{3}$ (4), $R\bar{3}c$ (3) and unknown (1).

difference in the relative number of structures with two or more independent molecules in the asymmetric unit between racemic crystals and organic crystals in general, Table 5.

As far as the 27 695 non-centrosymmetric organic crystals are concerned, the entries were further subdivided into crystal structures without (*e.g.* *P*1, $P2_1$ and $P2_12_12_1$) or with (*e.g.* *Pc*, *Cc* and $Pca2_1$) mirror symmetry. The latter category contains 3819 structures, all of which have been inspected manually. A total of 632 racemates (16.5%) were identified. A majority (93.3%) of these racemates belong to one of the five space groups *Pc* (6.2%), *Cc* (21.5%), $Pca2_1$ (23.3%), $Pna2_1$ (36.9%) and *Fdd*2 (5.5%), Table 6. The relative number of structures with two or more independent molecules in the asymmetric unit for the five space groups is approximately the same for both racemic crystals and organic crystals in general, Table 6.

No attempts have been made to estimate the number of organic racemates in the 23 874 unique structures in chiral space groups. In order for a racemate to crystallize in such a space group, the crystal must contain (at least) two crystallographically independent molecules of opposite chirality. Despite this restriction on possible structures, a search for racemates in these space groups would require a detailed inspection of all chiral centers in all structures with two or more molecules in the asymmetric unit. However, we believe this number to be low. A racemate in a chiral space group is rather extraordinary and this should indeed be reflected in the title of the CSD entry. The initial search for stereochemical designators in the *NAME* field, as discussed above, gave 139 hits for racemates belonging to a chiral space group. A thorough investigation of the 139 structures, using both molecular graphics (*Sybyl*6.4; Tripos Inc., 1997) and checking the original publication, revealed that only 17 (12.2%) were in fact racemates and had (at least) two independent enantiomers, Table 7. Out of the

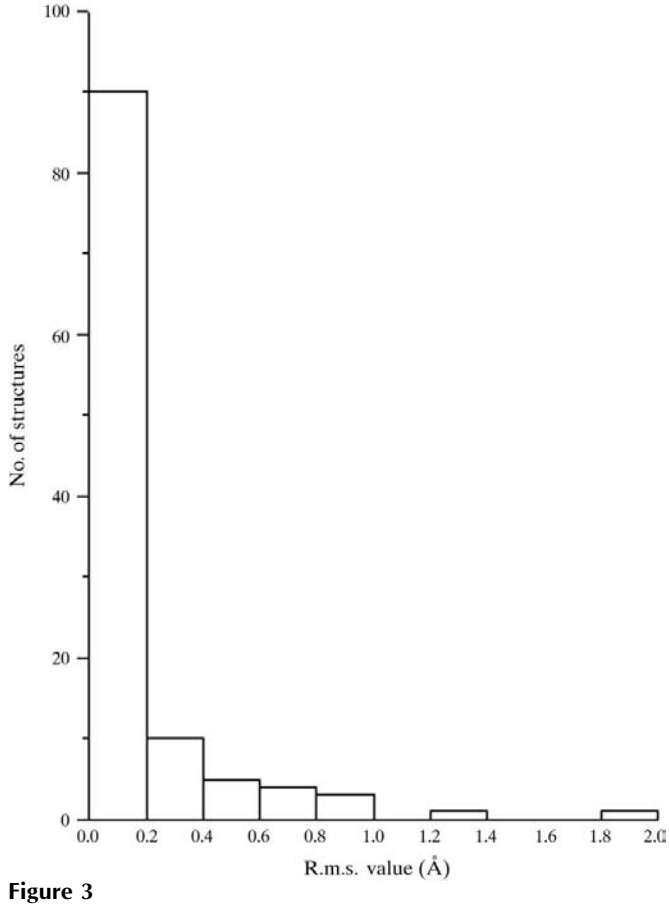


**Figure 3**
R.m.s deviations in atomic coordinates between pairs of crystallographically independent molecules in 114 non-centrosymmetric racemates.

**Table 6**
Racemates in non-centrosymmetric space groups which do have symmetry operations of the second kind.

$Z_{eff}$ = reported $Z$ value, $Z_0$ = standard $Z$ value given in *International Tables for Crystallography* (1996, Vol. A).

| Space group | Racemates | Racemates with $Z_{eff} > Z_0$ | Overall in CSD with $Z_{eff} > Z_0$ |
|---|---|---|---|
| *Pc* | 39 | 13 (33.3%) | 28.5% |
| *Cc* | 136 | 14 (10.3%) | 15.0% |
| $Pca2_1$ | 147 | 45 (30.6%) | 27.9% |
| $Pna2_1$ | 233 | 34 (14.6%) | 12.8% |
| *Fdd*2 | 35 | 1 (2.9%) | <0.5% |
| Others | 42 † | 3 (6.8%) | – |

† *Pnc*2 (1), *Pba*2 (1), *Pnn*2 (2), *Aba*2 (4), *Iba*2 (14), $P\bar{4}$ (1), $I\bar{4}$ (8), $I4_1cd$ (2), $P\bar{4}2_1c$ (8) and *R*3*c* (1).

remaining 122 structures a wrong name had been assigned for 117, either in the CSD or in the original publication. Most errors occur as the name for a racemic product has been kept even though the enantiomers are spontaneously resolved upon crystallization. The space group is incorrect for five structures either as a result of erroneous CSD deposition (refcodes: FIBKIK, FIBKOQ, JAVKUM) or wrong assignment in the original paper (refcodes: SARSUG, SOXVOR). The latter racemates are probably correctly described in the centrosymmetric space group $P\bar{1}$, as pointed out by Marsh (1995). A complete list with refcode and error category for the 122 structures is available from the correspondence author upon request.[1]

A previous analysis of the extent of conformational similarity between crystallographically independent molecules in 399 crystal structures from 65 chiral space groups gave an average r.m.s. deviation of 0.326 Å (Sona & Gautham, 1992). They also observed, as did Padmaja *et al.* (1990), that the higher the symmetry of the space group, the lower the number of structures with more than one molecule in the asymmetric unit. From our data it looks as if this is not the case for racemates in non-centrosymmetric space groups: there is only one triclinic structure, 38 monoclinic, 85 orthorhombic, one tetragonal and finally two trigonal crystals (Tables 6 and 7).

[1] Supplementary data for this paper, including lists of CSD refcodes and r.m.s. values for the various structures included in Tables 6 and 7, are available from the IUCr electronic archives (Reference: OS0044). Services for accessing these data are described at the back of the journal.

**Table 7**
Racemates in space groups which do not have symmetry operations of the second kind.

| Space group | Racemates |
|---|---|
| *P*1 | 1 |
| $P2_1$ | 11 |
| $P2_12_12_1$ | 4 |
| $P3_121$ | 1 |

A least-squares superposition of the independent molecules in the various racemates in non-centrosymmetric space groups (Tables 6 and 7, and inclusive of DL-allylglycine) gives an average r.m.s. deviation of 0.190 Å (non-H atoms only) after inversion of one of the two molecules. Compared with the mean r.m.s. of 0.326 Å, as obtained by Sona & Gautham (1992), we may infer that when there are two independent molecules in the asymmetric unit the conformational difference (after a possible inversion) is smaller for molecules with opposite chirality than for molecules with the same chirality or achiral molecules. Almost 80% of the structures have r.m.s. values in the range 0–0.2 Å, Fig. 3. The corresponding frequency is approximately 55% for molecules in chiral space groups (Sona & Gautham, 1992).

The purchase of the Siemens SMART diffractometer was made possible through financial support from The Norwegian Council of Research (NFR).

## References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 1, 31–37.
Dalhus, B. & Görbitz, C. H. (1999). *Acta Cryst.* B**55**, 1547–1555.
Dalhus, B. & Görbitz, C. H. (2000). In preparation.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Marsh, R. E. (1995). *Acta Cryst.* B**51**, 897–907.
Padmaja, N., Ramakumar, S. & Viswamitra, M. A. (1990). *Acta Cryst.* A**46**, 725–730.
Sheldrick, G. M. (1994). *SHELXTL*. Version 5. Siemens Analytical X-ray Instruments Inc., Madison, Wisconsin, USA.
Sheldrick, G. M. (1996). *SADABS*. University of Göttingen, Germany.
Siemens (1995). *SMART and SAINT*. Siemens Analytical X-ray Instruments Inc., Madison, Wisconsin, USA.
Sona, V. & Gautham, N. (1992). *Acta Cryst.* B**48**, 111–113.
Taylor, R. & Kennard, O. (1983). *Acta Cryst.* B**39**, 133–138.
Tripos Inc. (1997). *Sybyl*6.4. Tripos Inc., St Louis, Missouri, USA.