organic compounds

L-Isoleucyl-L-asparagine 1.094-hydrate: a hybrid hydrogen-bonding pattern

Carl Henrik Görbitz[a]* and Paul Hoff Backe[b]

[a]Department of Chemistry, University of Oslo, Oslo, Norway, and [b]Institute of Clinical Biochemistry, University of Oslo, Oslo, Norway, and Institute of Medical Microbiology and Centre for Molecular Biology and Neuroscience (CMBN), Rikshospitalet University Hospital, Oslo, Norway
Correspondence e-mail: c.h.gorbitz@kjemi.uio.no

The title compound, $C_{10}H_{20}N_3O_4 \cdot 1.094H_2O$, crystallizes with two dipeptide molecules in the asymmetric unit, each participating in two head-to-tail chains with hydrogen bonds between the terminal amino and carboxylate groups. As with many other dipeptides, the resulting structure is divided into distinct layers, but as the amide groups of the two peptide molecules participate in different types of interaction, the observed hydrogen bonds within a peptide main-chain layer (as distinct from the side-chain/solvent regions) cannot adapt to any of the four basic patterns observed previously for dipeptides. Instead, a rare hybrid pattern is formed.

Comment

In a recent survey of the crystal structures of dipeptides (Görbitz, 2010) it was found that two or even three head-to-tail hydrogen-bonded chains, involving the N-terminal amino groups and C-terminal carboxylate groups, co-exist in more than two thirds of all structures. In most of them, two such chains define hydrogen-bonded layers that can be classified into four basic patterns called **S4**, **T4**, **S5** and **T5**, where the initial capital letter indicates the type of symmetry involved in moving from one molecule in the chain to the next (**T** = translation and **S** = screw axis), and the number indicates the type of hydrogen-bonded chain for the amide N—H group [**4** = C(4) chain, **5** = C(5) chain; for graph-set theory, see Etter *et al.* (1990)]. Furthermore, it was pointed out that the nature of the two residues of the dipeptide has a profound impact on both the type of pattern and the peptide conformation.

The prevalence of dipeptides with one polar and one hydrophobic residue in the Cambridge Structural Database (CSD, Version 5.31 of November 2009; Allen, 2002) is largely limited to structures where the polar part is Ser, but also to some extent Tyr and Trp. Dipeptides with Asn, Gln and Cys residues, on the other hand, are very scarce. As part of a programme to provide more information on dipeptides in this

group, the structure of the title compound, (I), has been reinvestigated (our first experiments with L-Ile-L-Asn were unsuccessful; see below). From previous experience (Görbitz, 2010), such a nonpolar–polar dipeptide was expected to form either a hydrated structure with a **T4** hydrogen-bond pattern or a nonhydrated **S5** structure.

(I)
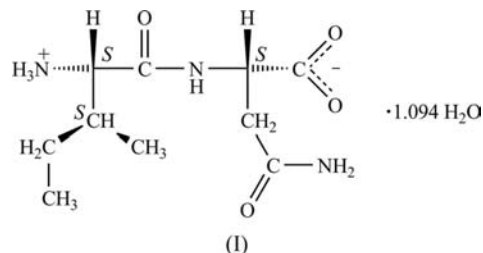
The asymmetric unit of (I) contains two peptide molecules, *A* and *B*, and also two [major disorder component, occupancy = 0.812 (2)] or three [minor disorder component, occupancy = 0.188 (2)] water molecules (Fig. 1). The minor arrangement retains water molecule 2 (with O2*W*), but molecule 1 is replaced by two different water molecules, *viz.* 3 and 4. Neighbouring unit cells cannot both have the minor arrangement, as this would bring O3*W* and O4*W* too close together. An illustration of this conflict is available in the *Supplementary Material*, which also includes additional information on the disorder of the molecule *A* carboxylate group indicated in Fig. 1.
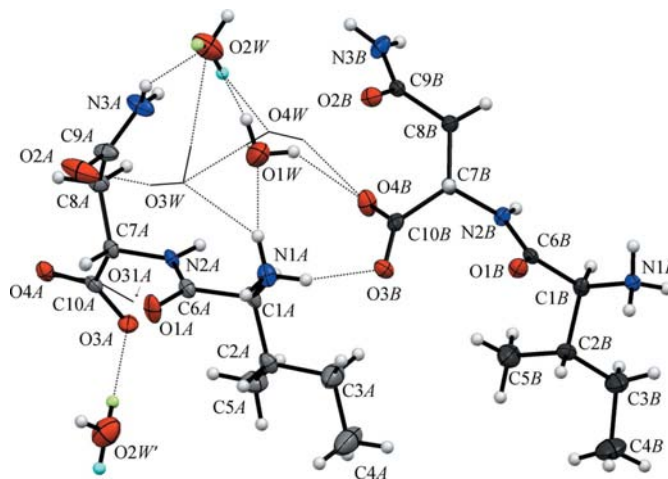

Figure 1
The asymmetric unit of (I), showing the atom-numbering scheme. Displacement ellipsoids are drawn at the 50% probability level and H atoms are shown as small spheres of arbitrary radii. The C atoms of molecule *B* are coloured dark grey in all figures. Water molecule 1 (atom O1*W*) has an occupancy of 0.812 (2), and may be replaced by the low-occupancy [0.188 (2)] water molecules 3 and 4 (atoms O3*W* and O4*W*), shown in wireframe style. This disorder also affects the terminal carboxylate group of peptide molecule *A*. In the major orientation there is a hydrogen bond to O2*W* [included as O2*W'* in the figure; symmetry code (2 − *x*, ½ + *y*, ½ − *z*)] through atom H21*W* (green in the electronic version of the paper), but when water molecule 1 is not present, water molecule 2 instead donates atom H23*W* (cyan) to O4*W* of water molecule 4, and the carboxylate group relaxes to an alternative orientation, shown here in wireframe representation with only O31*A* labelled. Atom H22*W* (on O2*W*, white) is hydrogen bonded to O3*A* regardless of hydration pattern, and accordingly has an occupancy of 1.0.
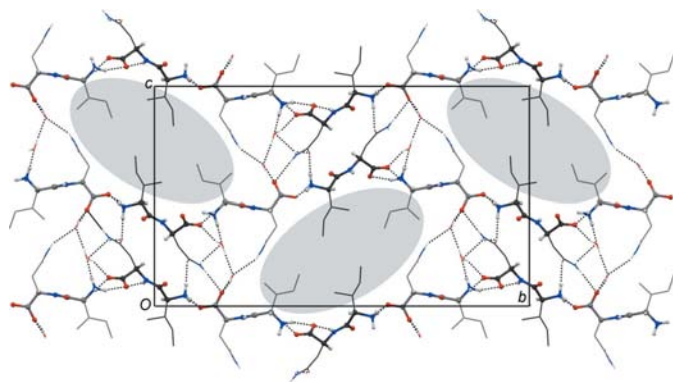
organic compounds

Figure 2
The unit cell and molecular packing of (I), viewed along the *a* axis. Side-chain atoms are depicted in wireframe style, and H atoms bonded to C atoms have been omitted for clarity. Wave-like two-dimensional layers of peptide main chains, in ball-and-stick representation, are seen edge-on. Grey-shaded ellipses show the aggregation of L-Ile side chains into hydrophobic columns.

The N-terminal L-Ile residue has the same conformation in both peptide molecules, with N1—C1—C2—C3 = *gauche*−, N1—C1—C2—C5 = *trans* and C1—C2—C3—C4 = *trans*. In contrast, as seen from the torsion angles in Table 1, the L-Asn side chain adopts two different conformations with either a *gauche*− [N2A—C7A—C8A—C9A = −63.7 (2)°] or a *trans* [N2B—C7B—C8B—C9B = 179.88 (15)°] orientation at the C^α–C^β bond. The r.m.s. value for the overlap between all non-H atoms in the two molecules is thus as high as 1.155 Å, but is only 0.547 Å when atoms O2, N3 and C9 are excluded (see *Supplementary Material*).

The crystal packing (Fig. 2) shows some typical features of mixed hydrophobic–polar dipeptides, such as the formation of distinct hydrogen-bonded layers by the main chains, which are in turn linked by additional interactions involving the side chains, and in this case also water molecules, to generate a three-dimensional hydrogen-bonding pattern.

The hydrogen bonds within a peptide main-chain layer are shown in Fig. 3. The vast majority of known layered dipeptide structures belong to the four basic patterns described above, but in four previous structures with $Z' \geq 2$ the connectivities were different for the independent molecules in the asymmetric unit. As is clear from Fig. 3, the structure of (I) displays such a hybrid pattern, with interactions typical of both the **S4** pattern (the two lower rectangles in Fig. 3) and the **T5** pattern (two upper rectangles). Such an **S4**/**T5** hybrid, unexpected for a nonpolar–polar dipeptide, was also observed for L-Ala-L-Met hemihydrate with $Z' = 2$ (Görbitz, 2003). For regular **T4** and **T5** patterns the lack of screw operations means that similar chains along a head-to-tail sequence are always positioned on the same side of the specific main-chain layer, the difference between the two being that in a **T5** structure the two side chains of the dipeptide form independent layers in the crystal structure, while in **T4** structures there is only one type of side-chain/solvent region (Fig. 4). In contrast, the screw symmetry of **S4** and **S5** patterns means that side chains appear on alternating sides of the main-chain layer. This pertains also to the present hybrid structure, but in a two-up-
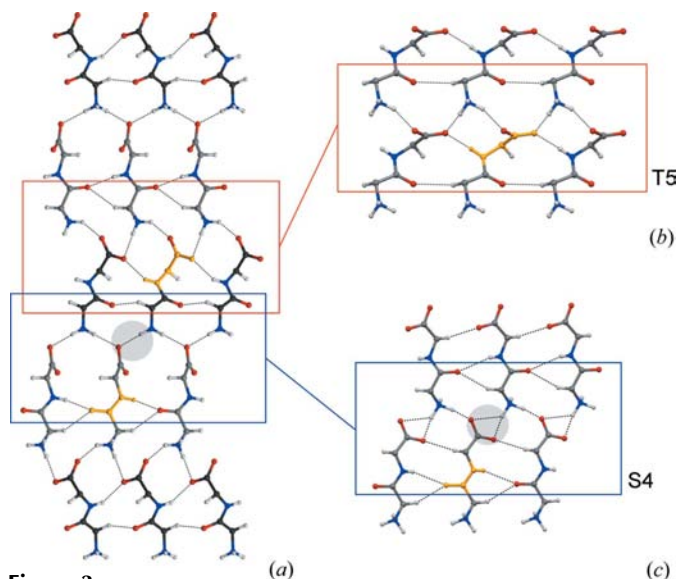

Figure 3
(*a*) A hydrogen-bonded layer in (I), compared with the idealized (*b*) **T5** and (*c*) **S4** patterns. In the electronic version of the paper, the repeating units of the amide $C(5)$ chain of the **T5** pattern and the amide $C(4)$ chain of the **S4** pattern are highlighted in orange. See *Comment* for further details.
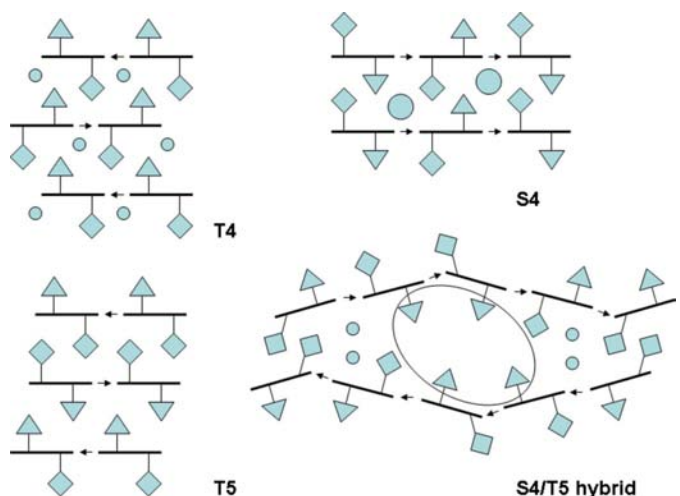

Figure 4
A schematic illustration of the side-chain arrangements in the four basic dipeptide aggregation patterns, **T4**, **T5** and **S4** (**S5** is similar), compared with the **S4**/**T5** hybrid structure of (I) (see also Fig. 2). Arrows indicate the directions of the head-to-tail chains, triangles and squares indicate the side chains of residues 1 and 2, respectively, small circles indicate water molecules, likely to be present in **T4** structures, and large circles indicate organic guest molecules (usually solvent), likely to be present in **S4** and **S5** structures. **T5** structures do not usually contain cocrystallized solvent molecules (Görbitz, 2010). The ellipse encircles a hydrophobic column, as in Fig. 2.

two-down fashion, as can be seen from Figs. 2 and 4. This gives rise to hydrophobic columns along the *a* axis, which are constructed from four different L-Ile side chains and are unusually large for this type of structure.

On a detailed level, the involvement of carboxylate group(s) in hydrogen bonding within a basic pattern can vary (Görbitz, 2010). The two head-to-tail chains can thus involve both O atoms, called mode *A*, or just a single O atom, called

mode *B*. In Fig. 3, the **T5** pattern uses the *A* mode in both (I) and the model structure. The **S4** pattern of (I) (Fig. 3*a*) uses a plain *B* mode, highlighted by a grey circle, while the model structure in this case shows a more centred H-atom location between the two carboxylate O atoms; it is neither a regular *A* mode nor a regular *B* mode. Being closer to *A* than *B*, this is called an *Ac* mode, with 'c' for 'centred'. Both **S4** and **T5** patterns typically prefer *A* modes, so the *B* mode observed for the **S4**-like part of the structure of (I) is uncommon.

Hydrogen bonding along the short *a* axis is not limited to the interactions shown in Fig. 3, but in fact also includes additional contacts with side-chain donors and acceptors. It is well known that the H atoms on C$^{\alpha}$ next to a carbonyl group are weakly acidic, and the N-terminal residues of dipeptides (but not the C-terminal residue, for which C$^{\alpha}$—H is less acidic due to the negative charge of the carboxylate group) invariably form C$^{\alpha}$—H···O interactions that constitute important elements in molecular aggregation. The patterns shown in Fig. 3 all incorporate such interactions. The L-Asn side chain similarly has potential C—H donors at C$^{\beta}$, and indeed both molecules *A* and *B* are involved in C$^{\beta}$—H···O interactions. For *A*, a parallel β-sheet-like pattern results, similar to that found in the structure of L-Ser-L-Asn (Görbitz & Hartviksen, 2008). L-Met-L-Asn (Stievater & Srikrishnan, 2005) has almost the same interactions as *B*, although those involving the side chains are clearly weaker. An illustration of side-chain hydrogen bonding is available in the *Supplementary Material*. The third known dipeptide structure with C-terminal L-Asn in the CSD, Gly-L-Asn (Pasternak *et al.*, 1954), contains no such contacts. In the only dipeptide structure with an N-terminal Asn residue, L-Asn-L-Val (Bonge *et al.*, 2005), two out of three molecules in the asymmetric unit form C$^{\beta}$—H···O interactions. There are only two dipeptide structures available with Gln residues, Gly-L-Gln (Panneerselvam & Soriano-García, 1995) and L-Val-L-Gln (Görbitz & Backe, 1996). The former has only a 2.79 Å C$^{\gamma}$—H···O contact to a carboxylate group, the latter a 2.59 Å intramolecular contact. More generally, 11 out of 22 Asn residues in CSD peptides (all types, including cyclic) have C$^{\beta}$—**H**···**O** contacts shorter than 2.8 Å (after normalization of the C—H distance to 0.99 Å), while equivalent C$^{\gamma}$—**H**···**O** contacts are present for 15 out of 24 Gln residues.

In conclusion, this paper describes a well behaved hydrated structure in space group *P*2$_1$2$_1$2$_1$. When we first studied L-Ile-L-Asn 15 years ago (Backe, 1995), a crystal devoid of solvent water, with cell parameters *a* = 5.277 Å, *b* = 8.571 Å and *c* = 27.369 Å and angles close to 90°, was investigated. For unknown reasons the space group of this apparently orthorhombic system could not be established. This could be due to pseudo-merohedral or nonmerohedral twinning (scrutiny with modern programs is not possible as the original experimental data are no longer available). There appeared to be two molecules in the asymmetric unit, their conformations being rather similar to that of the *A* molecule of (I), but with a *gauche+/trans* orientation of the L-Asn side chain at C$^{\alpha}$–C$^{\beta}$. The molecular packing arrangement probably corresponds to an **S5** hydrogen-bonding pattern.

Experimental

Block-shaped crystals of (I) were grown by vapour diffusion of acetonitrile into an aqueous solution (30 µl) of the peptide (about 1 mg).

Crystal data

C$_{10}$H$_{19}$N$_3$O$_4$·1.094H$_2$O	V = 2734.8 (4) Å^3
M_r = 265.01	Z = 8
Orthorhombic, $P2_12_12_1$	Mo $K\alpha$ radiation
a = 4.7855 (4) Å	μ = 0.10 mm^{-1}
b = 18.2887 (15) Å	T = 105 K
c = 31.247 (3) Å	0.78 × 0.50 × 0.22 mm

Data collection

Bruker APEXII CCD area-detector diffractometer	18243 measured reflections
	3860 independent reflections
Absorption correction: multi-scan (*SADABS*; Bruker, 2007)	2902 reflections with $I > 2\sigma(I)$
T_{min} = 0.751, T_{max} = 0.978	R_{int} = 0.060

Refinement

$R[F^2 > 2\sigma(F^2)]$ = 0.043	H atoms treated by a mixture of independent and constrained refinement
$wR(F^2)$ = 0.104	
S = 1.04	
3860 reflections	$\Delta\rho_{max}$ = 0.22 e Å^{-3}
376 parameters	$\Delta\rho_{min}$ = −0.23 e Å^{-3}
25 restraints	

Table 1
Selected torsion angles (°) for molecules *A* and *B* of (I).

	Molecule *A*	Molecule *B*
N1—C1—C6—N2	126.30 (18)	130.70 (17)
C1—C6—N2—C7	176.95 (15)	174.21 (15)
C6—N2—C7—C10	−110.5 (2)	−93.1 (2)
N2—C7—C10—O3	12.1 (3)	75.8 (2)
C6—N2—C7—C11	−115.5 (3)	
N2—C7—C11—O31	−10.2 (6)	
N1—C1—C2—C3	−55.2 (2)	−64.2 (2)
N1—C1—C2—C5	−178.44 (17)	171.47 (19)
C1—C2—C3—C4	176.0 (2)	168.6 (2)
N2—C7—C8—C9	−63.7 (2)	179.88 (15)
C7—C8—C9—O2	−51.6 (3)	31.5 (3)
C7—C8—C9—N3	128.7 (2)	−149.65 (17)

All heavy atoms were refined anisotropically, except for O*W*3 and O*W*4 in water molecules with low occupancy, which were only refined isotropically. The carboxylate group of molecule *B* has two positions; atoms C10*B*, O3*B* and O4*B* define the major [0.812 (2)] orientation, while C11*B*, O31*B* and O41*B* define the minor [0.188 (2)] orientation. Atoms C10*B* and C11*B* (separation = 0.12 Å) were assigned the same set of anisotropic displacement parameters (command EADP in *SHELXL97*; Sheldrick, 2008), and likewise for O4*B* and O41*B* (separation = 0.20 Å), while atom O31*B*, 0.64 Å from O3*B*, was refined isotropically. A *SHELXL97* SAME command was used to restrain the geometries of the two carboxylate groups to be similar within an effective standard deviation of 0.03 Å for both bond lengths and 1–3 distances. An attempted splitting of atom O2*A*, which from the displacement ellipsoid in Fig. 1 is clearly affected by the disorder in the water structure, was abandoned, as no improvement with respect to the *R* factor, *etc.*, was achieved.

Peptide H atoms were positioned with idealized geometry and with fixed C/N—H distances of 0.88, 0.88, 0.91, 0.98, 0.99 and 1.00 Å for NH, NH$_2$, NH$_3$, CH$_3$, CH$_2$ and CH groups, respectively. Free

Table 2
Hydrogen-bond geometry (Å, °).

D—H···A	D—H	H···A	D···A	D—H···A
N1A—H1A···O1W	0.91	1.87	2.755 (3)	164
N1A—H1A···O3W	0.91	2.23	2.959 (8)	136
N1A—H2A···O3B	0.91	1.91	2.792 (2)	163
N1A—H3A···O4B^{i}	0.91	2.03	2.695 (2)	128
N2A—H4A···O1A^{ii}	0.88	1.99	2.862 (2)	169
N3A—H5A···O2W	0.88	2.35	3.146 (2)	151
N3A—H6A···O2A^{ii}	0.88	1.98	2.846 (3)	167
C1A—H11A···O1A^{ii}	1.00	2.41	3.298 (2)	147
C8A—H81A···O2A^{ii}	0.99	2.52	3.383 (3)	146
N1B—H1B···O2B^{iii}	0.91	1.85	2.758 (2)	174
N1B—H2B···O4A^{iv}	0.91	1.93	2.835 (2)	172
N1B—H3B···O4A^{v}	0.91	2.06	2.968 (2)	179
N2B—H4B···O3B^{ii}	0.88	2.12	2.928 (2)	152
N3B—H5B···O3A^{vi}	0.88	2.09	2.946 (2)	163
N3B—H6B···O2B^{ii}	0.88	2.32	3.046 (2)	139
C1B—H11B···O1B^{ii}	1.00	2.27	3.098 (2)	139
C8B—H82B···O3B^{ii}	0.99	2.54	3.341 (2)	137
O1W—H11W···O2W	0.85 (1)	1.95 (1)	2.785 (3)	168 (3)
O1W—H12W···O4B	0.86 (1)	2.08 (1)	2.804 (3)	141 (2)
O2W—H21W···O3A^{vii}	0.85 (1)	2.17 (1)	3.019 (3)	179 (2)
O2W—H22W···O3A^{vi}	0.85 (1)	2.07 (1)	2.903 (3)	167 (2)
O2W—H22W···O31A^{vi}	0.85 (1)	1.93 (1)	2.749 (4)	161 (2)
O2W—H23W···O4W	0.85 (1)	2.31 (2)	3.153 (12)	170 (4)
O3W—H31W···O2A	0.86 (1)	2.09 (1)	2.917 (8)	164 (4)
O3W—H32W···O2W	0.85 (1)	2.31 (1)	3.157 (8)	172 (4)
O4W—H41W···O4B	0.85 (1)	2.10 (2)	2.867 (10)	150 (4)
O4W—H42W···O3W	0.86 (1)	2.42 (1)	3.281 (12)	179 (6)

Symmetry codes: (i) $x + 1, y, z$; (ii) $x - 1, y, z$; (iii) $x - \frac{1}{2}, -y + \frac{1}{2}, -z + 1$; (iv) $-x + \frac{1}{2}, -y + 1, z + \frac{1}{2}$; (v) $-x + \frac{3}{2}, -y + 1, z + \frac{1}{2}$; (vi) $-x + 1, y - \frac{1}{2}, -z + \frac{1}{2}$; (vii) $-x + 2, y - \frac{1}{2}, -z + \frac{1}{2}$.

rotation was permitted for amino and methyl groups. Restraints of 0.85 (1) and 1.35 (2) Å were imposed on the O—H and H···H distances, respectively, of the water molecules, giving O—H bond lengths in the range 0.85–0.86 Å and H—O—H angles in the range 103.8–105.9°. H atoms on OW3 and OW4, and also H23W on O2W, all with occupancy 0.188, could not be detected in the electron-density maps and so their positions were inferred from the nature of the surrounding donor and acceptor atoms, with appropriate restraints imposed on the pertinent intermolecular H···O distances during refinement. U_{iso}(H) values were set at $1.2U_{eq}$ of the carrier atom, or at $1.5U_{eq}$ for amino and methyl groups and water molecules. In the absence of significant anomalous scattering effects, 2638 Friedel pairs were merged. The absolute configuration was known for the purchased material.

Data collection: *APEX2* (Bruker, 2007); cell refinement: *SAINT-Plus* (Bruker, 2007); data reduction: *SAINT-Plus*; program(s) used to solve structure: *SHELXTL* (Sheldrick, 2008); program(s) used to refine structure: *SHELXTL*; molecular graphics: *SHELXTL*; software used to prepare material for publication: *SHELXTL*.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Backe, P. H. (1995). Masters thesis, University of Oslo, Norway.
Bonge, H. T., Rosenberg, M. L., Riktor, M. & Görbitz, C. H. (2005). *Acta Cryst.* E**61**, o524–o527.
Bruker (2007). *APEX2, SAINT-Plus* and *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
Etter, M. C., MacDonald, J. C. & Bernstein, J. (1990). *Acta Cryst.* B**46**, 256–262.
Görbitz, C. H. (2003). *Acta Cryst.* C**59**, o730–o732.
Görbitz, C. H. (2010). *Acta Cryst.* B**66**, 84–93.
Görbitz, C. H. & Backe, P. H. (1996). *Acta Cryst.* B**52**, 999–1006.
Görbitz, C. H. & Hartviksen, L. M. (2008). *Acta Cryst.* C**64**, o171–o176.
Panneerselvam, K. & Soriano-García, M. (1995). *Acta Cryst.* C**51**, 2718–2720.
Pasternak, R. A., Katz, L. & Corey, R. B. (1954). *Acta Cryst.* **7**, 225–236.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Stievater, T. & Srikrishnan, T. (2005). *J. Chem. Crystallogr.* **35**, 529–533.