Acta Crystallographica Section C
Crystal Structure Communications
ISSN 0108-2701

L-Histidyl-L-serine 3.7-hydrate: water channels in the crystal structure of a polar dipeptide

Carl Henrik Görbitz

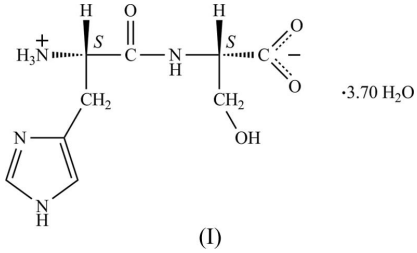
Department of Chemistry, University of Oslo, Oslo, Norway
Correspondence e-mail: c.h.gorbitz@kjemi.uio.no

Received 20 August 2010
Accepted 24 September 2010
Online 8 October 2010

Dipeptides may form nanotubular structures with pore diameters in the range 3.2–10 Å. These compounds normally contain at least one and usually two hydrophobic residues, but L-His-L-Ser hydrate, $C_9H_{14}N_4O_4\cdot3.7H_2O$, with two hydrophilic residues, forms large polar channels filled with ordered as well as disordered water molecules.

Comment

The structure of L-His-L-Ser was investigated by Padiyar (1998) [Cambridge Structural Database (CSD; Version 5.31 of November 2009; Allen, 2002) refcode WADXIJ] as a zwitterionic hydrate, (I), after having first been studied in a 1:1 complex with Gly-L-Glu by Suresh & Vijayan (1985), (II). In the process of elucidating the ability of small peptides to form nanotubular crystal structures (Görbitz, 2007), the former structure appeared to be of considerable interest because of its high water content (presented as a trihydrate), but it was difficult to get a clear picture of the actual water structure as no atomic coordinates were available in the CSD. A low-temperature reinvestigation of the uncomplexed dipeptide, (I), has thus been carried out.



(I)

The crystal structure of (I) is shown in Fig. 1(*a*). Bond lengths and bond angles are normal. The main chain is fairly extended, as reflected by the torsion angles listed in Table 1, and quite similar to the conformation of L-His-L-Ser in (II) (Fig. 1*b*), as also reported by Padiyar (1998). The L-His side-chain conformation, with N1—C1—C2—C3 = *gauche*− and C1—C2—C3—N2 = *gauche*−, coincides with the conformation taken by three out of six N-terminal His residues in dipeptide structures in the CSD (Allen, 2002) and effectively separates the imidazole moiety from the rest of the molecule, thus avoiding potential steric conflict between hydrogen-bond donors and acceptors interacting with the L-His side chain and functional groups in the backbone, respectively. In (II), C1—C2—C3—N2 is *gauche*+ (77.8°), giving a slightly different appearance. The L-Ser conformation, with N4—C7—C8—O2 = *gauche*+ and C7—C8—O2—H2 = *gauche*−, recurs in (II) and has also been observed for six out of 12 C-terminal Ser residues in other dipeptides, often with a cocrystallized water molecule as acceptor for the H atom of the hydroxy group as seen for (I) in Fig. 1(*a*).

In addition to the ordered water molecules 1 and 2 (Fig. 1*a*), which were also identified by Padiyar (1998), coordinates were refined for seven low-occupancy water sites (see *Experimental*). All are arranged along conspicuous water channels running parallel to the *a* axis, highlighted by ellipse 1 in Fig. 2(*a*), where they provide acceptors for the side-chain >N—H donor of L-His as well as for one of the H atoms of water molecule 1. These disordered water molecules define the solvent channel highlighted in ellipse 2 with a volume of 77.1 $Å^3$ per unit cell (calculated by *PLATON*; Spek, 2009) and an average cross section of 16.0 $Å^2$. The electron count within the channel, 27.3, corresponds to 3.41 water molecules per channel or 1.71 water molecules per peptide molecule, which fits nicely with the sum of refined occupancies for the seven disordered water molecules, 1.70. Together with water mol-

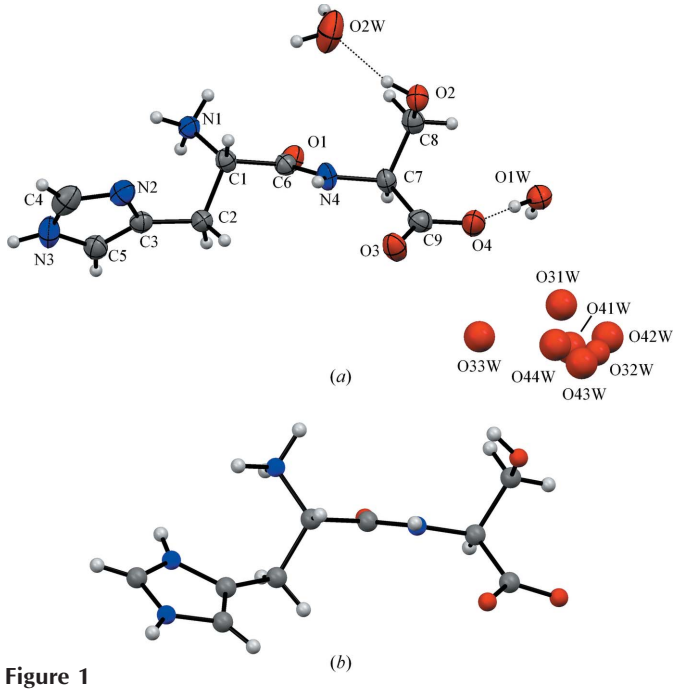


Figure 1
(*a*) The asymmetric unit of (I), showing the atom-numbering scheme. Displacement ellipsoids (or spheres for O atoms in water molecules with low occupancy) are drawn at the 50% probability level; H atoms are spheres of arbitrary size. (*b*) Ball-and-stick model of L-His-L-Ser as it appears in the complex with Gly-L-Glu (Suresh & Vijayan, 1985). The CSD entry (refcode DIYZOA; Allen, 2002) does not contain coordinates for H, so H atoms were introduced in idealized positions with the hydroxy group pointing in the direction of the closest acceptor atom.

 doi:10.1107/S0108270110038254

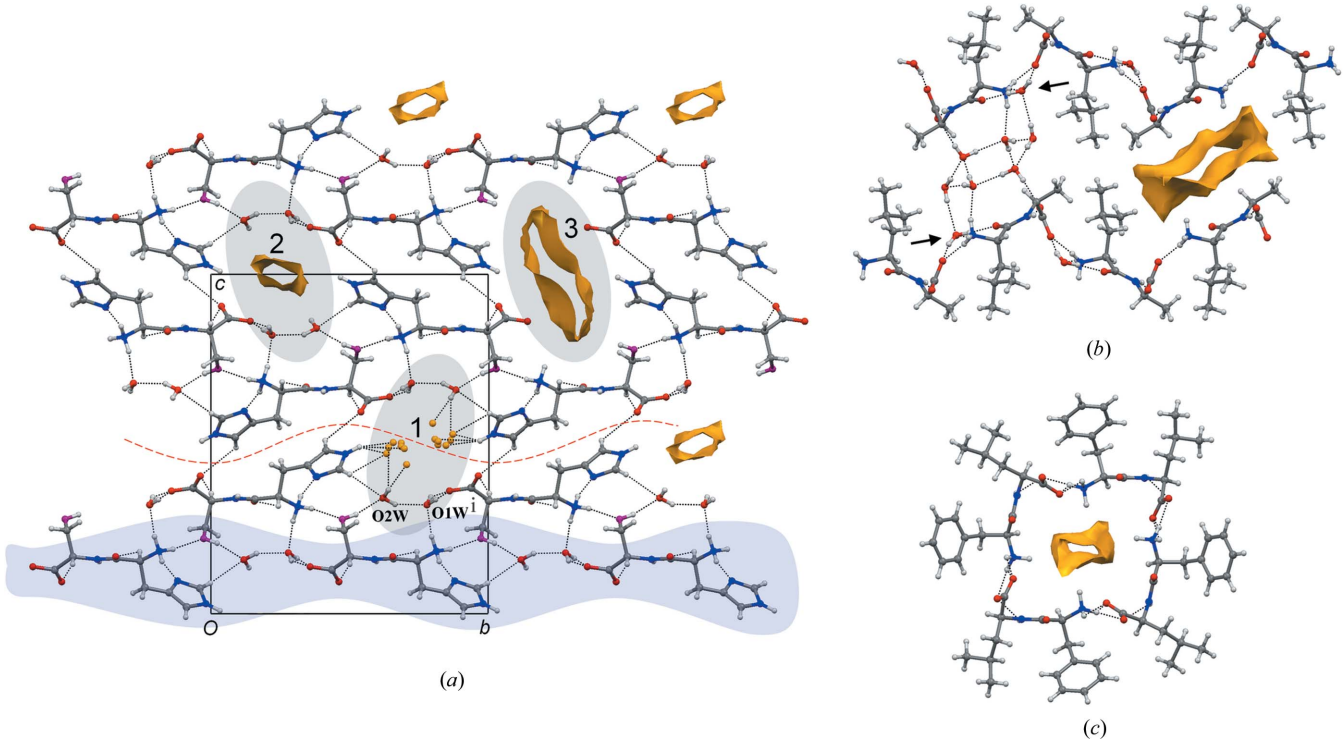


Figure 2
(*a*) Unit cell and molecular packing of (I), viewed along the *a* axis. A horizontal hydrogen-bonded sheet, encompassing ordered water molecules and peptide molecules in the *ab* plane, is highlighted in blue (colours relate to the electronic version of the paper). Notably, the L-Ser hydroxy group, with the O atom coloured in violet, is located outside the sheet and in fact constitutes an essential part of the neighbouring sheet. The two sheets connected in this way are said to form a layer. A wavelike interface between two such layers has been indicated by a dashed line, and it is crossed only by the indicated very weak C5—H51···O3($-x, y+\frac{3}{2}, -z+\frac{1}{2}$) contacts (H···O = 2.62 Å and C—H···O = 120°). The ellipses in grey show water-filled channels running through the crystal parallel to the *a* axis. In 1, the positions of all water molecules have been included, with low-occupancy water positions coloured in orange, while in 2 only the ordered water molecules (with O*W*1 and O*W*2) have been retained, leaving an empty column shown in orange (three more examples included). In 3, all water molecules have been removed, giving a much larger channel. (*b*) Water channels in the structure of L-Leu-L-Ala tetrahydrate (Görbitz, 1997) comparable in size to the channels of (I). Out of the four water molecules in the asymmetric unit, one (indicated by arrows) is tightly associated with the peptide molecules, while the other three form the bulk of the channel. To the right, a column is illustrated in the same way as column 3 in (*a*). (*c*) A hydrophilic channel without ordered water molecules running through the crystal structure of L-Phe-L-Leu (Görbitz, 2001). All illustrations are on approximately the same scale and were prepared by *Mercury* (Macrae *et al.*, 2008) with application of a 1.2 Å probe radius and 0.5 Å grid spacing for calculation of voids. [Symmetry code: (i) $-x+1, y+\frac{1}{2}, -z+\frac{1}{2}$.]

ecules 1 and 2, the complete solvent system, ellipse 3 in Fig. 2(*a*), has a total volume of 175.2 Å^3 with an average cross section of 36.4 Å^2.

Water channels have previously been found in the structures of highly hydrated dipeptide species like L-Val-L-Ser trihydrate (Johansen *et al.*, 2005), L-His-L-Asp trihydrate (Cheng *et al.*, 2005) and in particular L-Leu-L-Ile 2.5-hydrate (Görbitz & Rise, 2008) and L-Leu-L-Ala tetrahydrate (Fig. 2*b*) (Görbitz, 1997). All these channels have completely ordered water structures. The same applies to those members of the Phe-Phe class of nanotubular dipeptides which have channels small enough to be spanned by individual water molecules, *e.g.* L-Leu-L-Leu 0.87-hydrate and L-Leu-L-Phe 0.86-hydrate (Görbitz, 2001). In contrast, the only slightly larger channels of L-Phe-L-Leu 1.26-hydrate (Fig. 2*c*) (Görbitz, 2001), which are comparable in size (average cross section = 19.4 Å^2) to the central channel in ellipse 2 in Fig. 2(*a*) for (I), contain disordered water molecules. These two structures together with the hexagonal structure of L-Phe-L-Phe 2.47-hydrate (Görbitz, 2001) are unique among dipeptides in having channels with disordered and presumably mobile water molecules (Febles *et al.*, 2006).

Considering the crystal packing arrangement of dipeptides, compounds with at least one hydrophobic residue frequently form distinct layers (Görbitz, 2010). When both residues are polar or charged, however, layers are usually less obvious or even absent. The crystal packing of (I) can nevertheless conveniently be regarded as being composed of layers, with each layer in turn being constructed from two individual sheets (Fig. 2*a*). Interactions between layers are very weak, essentially being confined to interactions involving the disordered water molecules at the centre of the solvent channel, as the long >C5—H51···O3($-x, y+\frac{3}{2}, -z+\frac{1}{2}$) contact included in Fig. 2(*a*) is not very significant (Wood *et al.*, 2009). This would explain the fragile nature of the crystals grown.

The hydrogen-bonding pattern within a sheet in the structure of (I) may be compared with patterns observed in other dipeptide structures. In a recent survey (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

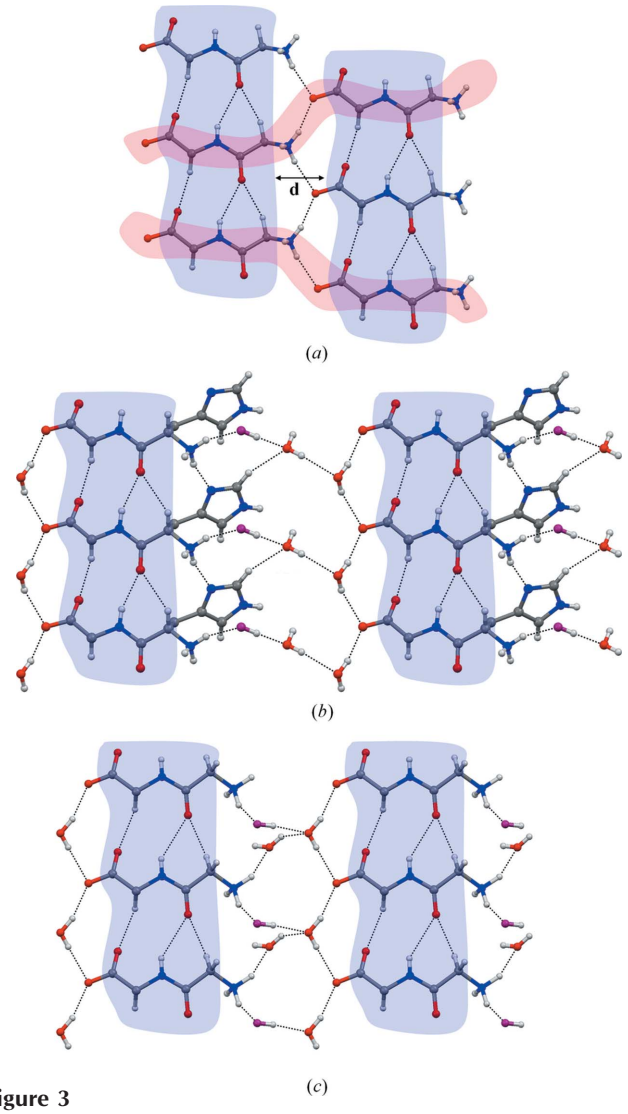


Figure 3
(*a*) A standard **T4** hydrogen-bonding pattern found in the crystal structures of dipeptides (Görbitz, 2010). The two horizontal head-to-tail chains are coloured in red (colours relate to the electronic version of the paper); the *C*(4) chain involving the amide N—H···O and two additional C^{α}—H···O interactions form vertical one-dimensional hydrogen-bonded tapes coloured in blue. The separation between tapes is called **d**. (*b*) Hydrogen bonding in an individual sheet of (I) (highlighted in blue in Fig. 2*a*), including interactions with the fully occupied water molecules, O1*W* and O2*W*, in the water channel. The O atoms of the L-Ser hydroxy groups coming from the adjacent layer have, as in Fig. 2(*a*), been coloured in violet. L-His H atoms on C^{β} have been omitted. (*c*) Hydrogen bonding in a sheet in the structure of Gly-L-Tyr dihydrate (Cotrait & Bideau, 1974). The L-Tyr side chains have been hidden, except for the hydroxy groups of dipeptide molecules in the adjacent layer. Hydrogen-bonded tapes have been coloured in blue in (*b*) and (*c*) in the same way as in (*a*).

carboxylate groups, co-exist in more than two thirds of all crystal structures. In most of them, two such chains generate hydrogen-bonded sheets that can be classified into four basic patterns called **S4**, **T4**, **S5** and **T5**, where the initial capital letter denotes 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 describes the hydrogen bonding of the amide N—H group [4 = *C*(4) chain, 5 = *C*(5) chain; for graph-set theory, see Etter *et al.*, 1990]. A model **T4** pattern is depicted in Fig. 3(*a*) showing hydrogen-bonded tapes incorporating the amide *C*(4) chains as well as C^{α}—H···O interactions. For each of the four basic patterns, the separation between tapes, called **d** for the **T4** pattern in Fig. 3(*a*), must be small enough for amino groups and carboxylate groups to form the direct hydrogen bonds that define the two *C*(8) chains. Occasionally, **d** is too large for this to be possible, and one or two *C*(8) chains may be lost compared to the parent pattern (Görbitz, 2010). In the pattern code, an asterisk is used to denote such a missing chain; **S4*** accordingly means a pattern that is derived from the regular **S4** pattern, but with only one remaining *C*(8) chain. Fig. 3(*b*) shows that the pattern of (I) can be classified as **T4****, with the charged termini being bridged by the solvent water molecules and the functional groups of the L-Ser and L-His side chains. The only other known example of a **T4**** structure, Gly-L-Tyr dihydrate (Cotrait & Bideau, 1974), is shown in Fig. 3(*c*). Despite its overall apparently quite different crystalline arrangement, the solvent water molecules and side-chain hydroxy groups play roles remarkably similar to the roles they play in the structure of (I).

Experimental

Fragile needle-shaped crystals were grown by vapour diffusion of acetonitrile into 30 µl of an aqueous solution containing about 1 mg of the peptide.

Crystal data

$C_9H_{14}N_4O_4\cdot3.7H_2O$
M_r = 308.89
Orthorhombic, $P2_12_12_1$
a = 4.812 (3) Å
b = 15.505 (8) Å
c = 18.958 (10) Å
V = 1414.6 (13) Å^3
Z = 4
Mo *Kα* radiation
μ = 0.13 mm^{-1}
T = 105 K
0.62 × 0.06 × 0.04 mm

Data collection

Bruker APEXII CCD diffractometer
Absorption correction: multi-scan (*SADABS*; Bruker, 2007)
T_{min} = 0.836, T_{max} = 0.996
7345 measured reflections
1544 independent reflections
1076 reflections with $I > 2\sigma(I)$
R_{int} = 0.083

Refinement

$R[F^2 > 2\sigma(F^2)]$ = 0.048
$wR(F^2)$ = 0.131
S = 1.08
1544 reflections
219 parameters
7 restraints
H atoms treated by a mixture of independent and constrained refinement
$\Delta\rho_{max}$ = 0.27 e Å^{-3}
$\Delta\rho_{min}$ = −0.26 e Å^{-3}

Peptide H atoms were positioned with idealized geometry and fixed C/N—H distances for NH_3, NH, CH_2, CH (methine) and CH (sp^2) groups of 0.91, 0.88, 0.99, 1.00 and 0.95 Å, respectively. Free rotation was permitted for the amino group. Restraints were imposed on the O—H distances of L-Ser and the ordered water molecules 1 and 2 (by use of the DFIX 0.85 0.01 instruction; Sheldrick, 2008) and on the H···H distances of water molecules (DFIX 1.35 0.02) to give O—H bond lengths in the range 0.84 (2)–0.85 (2) Å and H—O—H angles in the range 105–109°. The U_{iso}(H) values were constrained to

Table 1
Selected torsion angles (°).

N1—C1—C6—N4	159.6 (4)	C1—C2—C3—N2	−56.2 (6)
C1—C6—N4—C7	−179.0 (4)	C1—C2—C3—C5	121.1 (5)
C6—N4—C7—C9	−152.1 (4)	N4—C7—C8—O2	67.4 (5)
N4—C7—C9—O3	3.8 (6)	C7—C8—O2—H2	−85 (4)
N1—C1—C2—C3	−51.8 (5)		

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

D—H···*A*	*D*—H	H···*A*	*D*···*A*	*D*—H···*A*
O2—H2···O2*W*	0.85 (2)	1.85 (2)	2.684 (6)	167 (5)
N1—H1*A*···O2^{i}	0.91	1.88	2.755 (5)	160
N1—H1*B*···N2^{ii}	0.91	1.91	2.792 (6)	163
N1—H1*C*···O1*W*^{iii}	0.91	1.91	2.810 (5)	171
N3—H3···O32*W*^{iv}	0.88	1.98	2.724 (11)	142
N4—H4···O1^{v}	0.88	2.16	2.991 (5)	157
C1—H11···O1^{v}	1.00	2.44	3.197 (6)	132
C4—H41···O2*W*^{vi}	0.95	2.60	3.499 (8)	159
C7—H71···O3^{ii}	1.00	2.36	3.342 (6)	166
O1*W*—H11*W*···O4^{ii}	0.84 (2)	2.11 (3)	2.888 (5)	154 (6)
O1*W*—H12*W*···O4	0.85 (2)	1.90 (2)	2.725 (5)	166 (5)
O2*W*—H21*W*···O31*W*^{i}	0.85 (2)	1.79 (6)	2.504 (11)	140 (8)
O2*W*—H22*W*···O1*W*^{i}	0.85 (2)	2.09 (4)	2.889 (6)	156 (8)

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

kU_{iso}(C,N,O), where k = 1.5 for the –NH_3 and water H atoms and 1.2 otherwise.

In the standard refinement of the structure of (I), electron density from disordered solvent molecules within the channels was modelled by seven O atoms, three with occupancies in the range 0.42 (3)–0.457 (13) that were refined isotropically and four with occupancies in the range 0.042 (10)–0.159 (19) that were assigned a fixed isotropic displacement parameter of 0.06 Å^2. The H atoms of the low-occupancy water molecules were omitted from the model.

Structure refinement (not tabulated, available in archived CIF) has also been carried out with a modified *.hkl* file from which the contribution from disordered solvent had been eliminated by the SQUEEZE routine of the *PLATON* program (Spek, 2009). This procedure gave R = 0.045 and $wR(F^2)$ = 0.112, and reductions in the standard uncertainties of calculated geometric parameters, amounting to 0.001 Å for bond lengths and 0.1° for angles and torsion angles, but no significant changes to the geometric parameters themselves.

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*.

Supplementary data for this paper are available from the IUCr electronic archives (Reference: FG3190). Services for accessing these data are described at the back of the journal.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Bruker (2007). *APEX2*, *SAINT-Plus* and *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
Cheng, F., Sun, H., Zhang, Y., Mukkamala, D. & Oldfield, E. (2005). *J. Am. Chem. Soc.* **127**, 12544–12554.
Cotrait, M. & Bideau, J.-P. (1974). *Acta Cryst.* B**30**, 1024–1028.
Etter, M. C., MacDonald, J. C. & Bernstein, J. (1990). *Acta Cryst.* B**46**, 256–262.
Febles, M., Pérez-Hernández, N., Pérez, C., Rodriguez, M. L., Foces-Foces, C., Roux, M. V., Morales, E. Q., Buntkowsky, G., Limbach, H.-H. & Martin, J. D. (2006). *J. Am. Chem. Soc.* **128**, 10008–10009.
Görbitz, C. H. (1997). *Acta Cryst.* C**53**, 736–739.
Görbitz, C. H. (2001). *Chem. Eur. J.* **7**, 5153–5159.
Görbitz, C. H. (2007). *Chem. Eur. J.* **13**, 1022–1031.
Görbitz, C. H. (2010). *Acta Cryst.* B**66**, 84–93.
Görbitz, C. H. & Rise, F. (2008). *J. Pept. Sci.* **14**, 210–216.
Johansen, A., Midtkandal, R., Roggen, H. & Görbitz, C. H. (2005). *Acta Cryst.* C**61**, o198–o200.
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.
Padiyar, G. S. (1998). *J. Pept. Res*, **51**, 266–270.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.
Suresh, C. G. & Vijayan, M. (1985). *Int. J. Pept. Protein Res.* **26**, 329–336.
Wood, P. A., Allen, F. H. & Pidcock, E. (2009). *CrystEngComm*, **11**, 1563–1571.