STRUCTURAL SCIENCE
CRYSTAL ENGINEERING
MATERIALS

ISSN: 2052-5206
journals.iucr.org/b

Hydrophobic dipeptides: the final piece in the puzzle

Carl Henrik Görbitz

Acta Cryst. (2018). B**74**, 311–318

STRUCTURAL SCIENCE CRYSTAL ENGINEERING MATERIALS

ISSN 2052-5206

Hydrophobic dipeptides: the final piece in the puzzle

Carl Henrik Görbitz*

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

Received 6 March 2018
Accepted 13 May 2018

Edited by A. Nangia, CSIR–National Chemical Laboratory, India

Keywords: dipeptides; hydrogen bonding; crystal packing; conformations.

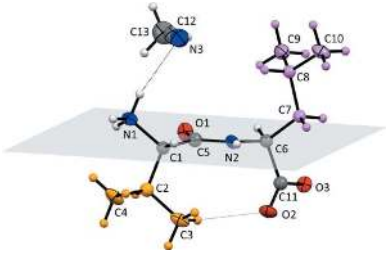
CCDC reference: 1825080

Supporting information: this article has supporting information at journals.iucr.org/b

The crystal structure of L-valyl-L-leucine acetonitrile solvate presented here adds to 24 previously reported structures of dipeptides constructed from the five nonpolar amino acids L-alanine, L-valine, L-isoleucine, L-leucine and L-phenylalanine. It thus constitutes the final piece in the 5 × 5 puzzle of hydrophobic dipeptide structures. This opportunity is taken to review the crystal packing arrangements and hydrogen-bonding preferences of a rather unique group of substances, with updated information on the various hydrogen-bonding patterns and the associated peptide conformations.

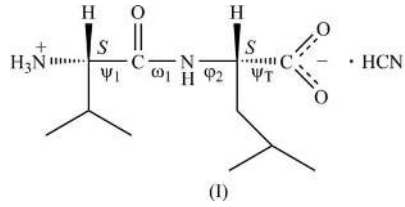
1. Introduction

In a previous review of the crystal structures of dipeptides (Görbitz, 2010), it was demonstrated that a vast majority of the 160 entries then available in the Cambridge Structural Database (CSD) (Groom *et al.*, 2016) had two or three *C*(8) head-to-tail hydrogen-bonded chains [for graph-set notation see Etter *et al.* (1990), Bernstein *et al.* (1995)]. Out of these, 97 fitted into a new classification scheme with four basic two-dimensional patterns called T4, T5, S4 and S5 (see Fig. 1). The letter in each code gives the symmetry relation between consecutive dipeptide molecules along the *C*(8) hydrogen-bonded chains, T for translation and S for screw axis, while the number indicates the chain type of the amide N—H donor as *C*(4) or *C*(5). For consistency, the T4 or S4 codes are used even when the H···O distances of the declared *C*(4) chains are well above 3.0 Å, which may happen when combining sterically bulky side chains with large co-crystallized molecules (see supporting information Fig. S1 for an example). The general molecular arrangements may be more or less retained even if one or even both *C*(8) chains are broken, in which case one or two asterisks are added to the original pattern codes. The two most common types of layers derived in this manner are S4* and T5** (Fig. 1). Finally, akin to how a graphene sheet may fold into a carbon nanotube, T4 and T5 sheets (but not S4 and S5 sheets with screw symmetry operations) have tubular versions. Of these T4-t occurs just for a single structure, Thr-Ala (Görbitz, 2005), and only T5-t will be discussed in further detail. The few structures that do not incorporate the four basic patterns or derivatives thereof can be divided into three group: nanotubular structures of the Val-Ala class (Görbitz, 2003), layered structures with antiparallel chains (one pattern shown in Fig. 1, four more in Fig. S2 in the supporting information) and a group of unique and quite complex structures that often have $Z' > 1$.



Loosely defining a 'hydrophobic amino acid residue' as one devoid of N and O atoms in its side chains, 'hydrophobic dipeptides', with two such residues, share the same set of strong hydrogen bond donors (amino + amide groups) and acceptors (amide + carboxylate groups) in their main chains, but vary in terms of the shape and size of the hydrophobic moieties in their side chains. For this reason they can serve as model compounds for investigations of hydrogen bond formation and yield valuable knowledge about the connection between peptide sequence and recurring crystal packing patterns.

From the five standard amino acids Ala, Val, Ile, Leu and Phe, here denoted by the common code Xaa, it is possible to construct 25 different L-Xaa-L-Xaa dipeptides. The first to have its solid-state structure determined was Ala-Ala as early as 1971 (Fletterick *et al.*, 1971), but 23 more years passed before the second one, Leu-Leu DMSO solvate, appeared (Mitra & Subramanian, 1994). In the following years, a plethora of substances were studied, some in several different forms (anhydrates, hydrates or other solvates). After the report of Phe-Ile in 2004 (Görbitz, 2004*c*), Val-Leu was the only Xaa-Xaa dipeptide for which there was no available crystal structure. The acetonitrile solvate (I) presented here thus completes a systematic investigation of a family of compounds that has no equivalent in the CSD.



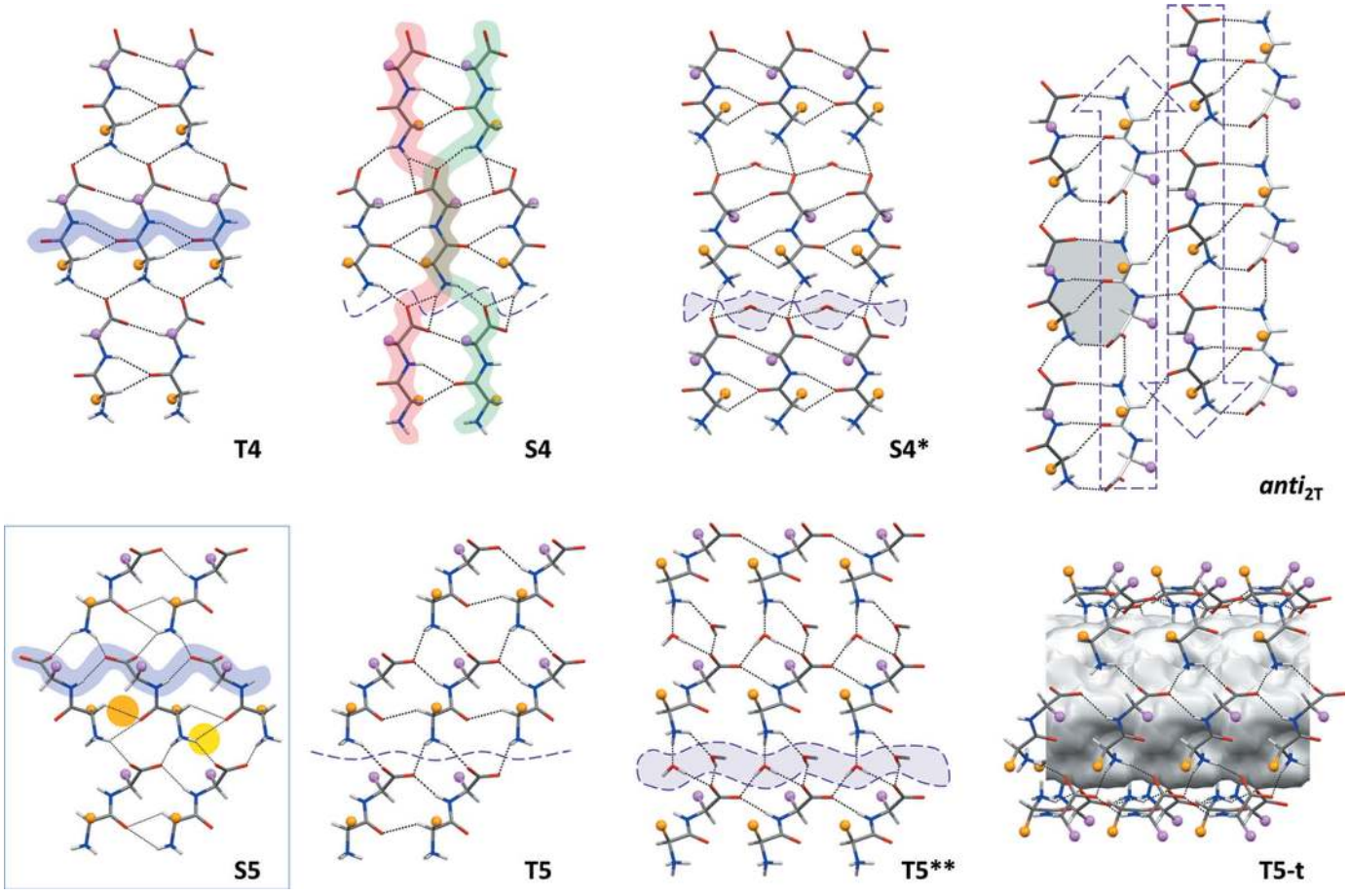


Figure 1
The four basic dipeptide aggregation patterns T4, S4, T5 and S5 compatible with the concomitant existence of two *C*(8) chains within a two-dimensional sheet, as shaded in red and green for S4 (Görbitz, 2010). The structure of (I) was used for the S5 drawing (inside box). Side chains (*N*-terminal orange, *C*-terminal violet) are curtailed beyond the C^{β}-atoms, shown as spheres. A *C*(4) (amide acceptor) and a *C*(5) chain (carboxylate acceptor) are shaded in blue for T4 and S5, respectively. Non-mandatory C—H···O contacts are quite abundant [C1—H11···O1 for (I) in orange], and S5 may involve a three-centre interaction (yellow). The S4* pattern can be derived from S4 by breaking one of the *C*(8) chains, here a result of inserting a water molecule at the interface between horizontal hydrogen-bonded tapes (purple dashes and shades). In an equivalent manner T5** mimics T5, but both *C*(8) chains are broken by water molecules, other solvent molecules, functional groups of the side chains or combinations thereof. Head-to tail chains in the $anti_{2T}$ pattern run in opposite directions (big, open arrows) with formation of conspicuous dimers (grey shade). T5-t is the tubular version of the flat T5 sheet, the shade is the void volume calculated by *Mercury* (Macrae *et al.*, 2008) for Phe-Phe (Görbitz, 2001*b*) after removal of disordered water molecules inside the channel. There is rotation (but not screw operations) along *C*(8) chains, so the use of the letter T (for translation) is not formally correct, but the notation has nevertheless been retained to stress the relationship with the regular T5 pattern.

In previous contributions we have focused on what we have dubbed 'the packing problem' of hydrophobic dipeptides (Görbitz & Gundersen, 1996*b*, Görbitz, 2007), originating from the fact that in any layered structure the number of $C(8)$ chains is limited to two. The third amino H atom may be donated to a functional group in the side chain of a polar residue (Görbitz & Etter, 1992), but if these are absent, the peptide carries an inherent challenge in finding the third acceptor. It is the different solutions to this 'packing problem' that give rise to a series of rather well-defined structural families, each sharing a particular type of molecular stacking and hydrogen-bonding network.

2. Methodology

2.1. Cambridge Structural Database searches

The CSD (Version 5.39 of November 2017; Groom *et al.*, 2016) was searched for unprotected, zwitterionic dipeptide structures (not redeterminations) of standard and nonstandard amino acids published in 2010 and later. All unique entries found (*i.e.* not pseudopolymorphs essentially identical to a previous structure except for incorporating a different co-crystallized molecule) were analysed in terms of hydrogen-bonding pattern and added to the previous overview (Görbitz, 2010). A complete list of all *hydrophobic* dipeptides was then prepared, including both redeterminations and pseudopolymorphs. A final shortlist, which served as the primary reference material for the present investigation, was limited to unique structure determinations for the 25 hydrophobic dipeptides discussed above, in addition to five other dipeptides with the same hydrogen-bonding pattern as (I), *i.e.* S5. All three lists are available in Tables S1–S3 of the supporting information.

Selected structures were visualized in *Mercury* (Macrae *et al.*, 2008), which was also used to prepare most illustrations. Unless otherwise noticed, all dipeptides discussed have L-chirality. Standard three-letter codes are used throughout, fGly is phenylglycine and Nva is norvaline.

2.2. Crystallization and data collection

It is not trivial to crystallize Val-Leu, which in neutral aqueous solution is rapidly hydrolyzed and also converted to the corresponding cyclic dipeptide (diketopiperazine). For this reason, Sigma-Aldrich ships commercial samples as the hydrochloride. To grow crystals of the zwitterion, about 1 mg of Val-Leu·HCl was dissolved in 80 µL of water in a small test tube, followed by addition of $Ag_2O(s)$ (molar ratio 2:1). After precipitation of AgCl(*s*), the tube was left for 15 min; 30 µL of the clear solution was then transferred to a second test tube in which two types of crystals, large and small, appeared within 24 h by vapour diffusion of acetonitrile (∼1 ml) inside a bigger, capped tube. The former specimens proved to be a dipeptide-silver complex (to be discussed in detail elsewhere), while abundant smaller crystals of lower quality were the anticipated peptide solvate (I).

Table 1
Experimental details.

	(I)
Crystal data	
Chemical formula	$C_{11}H_{22}N_2O_3·C_2H_3N$
M_r	271.36
Crystal system, space group	Orthorhombic, $P2_12_12_1$
a (Å)	5.341 (3)
b (Å)	13.536 (6)
c (Å)	21.406 (8)
V (Å^3)	1547.6 (6)
Z	4
μ (mm^{-1})	0.083
Temperature (K)	100 (2)
Crystal size (mm)	0.42 × 0.15 × 0.01
Data collection	
T_{min}, T_{max}	0.612, 1.000
No. of measured, independent and observed [$I > 2\sigma(I)$] reflections	6287, 1572, 903
R_{int}	0.190
θ_{max} (°)	20.8
Refinement	
$R[F^2 > 2\sigma(F^2)]$, $wR(F^2)$, S	0.082, 0.184, 1.01
No. of reflections	1572
No. of parameters	140
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})	0.26, −0.28
CCDC No.	1825080

Single-crystal X-ray data collections were carried out with *APEX3* software (Bruker, 2016) on a D8 Venture single crystal CCD diffractometer equipped with an Oxford Cryosystems Cryostream *Plus* cooling unit and Mo $K\alpha$ radiation (λ = 0.71069 Å). Data integration and cell refinement were performed with *SAINTPlus* (Bruker, 2016) with subsequent absorption correction by *SADABS* (Bruker, 2016) and structure solution with *SHELXLT* (Sheldrick, 2015*a*). Despite using long exposure times, only a limited number of Bragg peaks were observed up to about 2θ = 40°. In order not to compromise the reflection-to-parameter ratio, non-terminal C atoms without excessive thermal motion were consequently kept isotropic during refinement in *SHELXL* (Sheldrick, 2015*b*).[1] H atoms were included in calculated positions and treated as riding: N—H = 0.88–0.91 Å, C—H = 0.98–1.00 Å with U_{iso}(H) = 1.5U_{eq} for the amino and methyl groups and 1.2U_{eq}(C) for other H atoms. Refinement results are summarized in Table 1.[1]

3. Results and discussion

The 'packing problem' of hydrophobic dipeptides has a number of solutions:

(1) Formation of a third $C(8)$ chain. This has been observed for (*S*)-fGly-(*R*)-fGly (Akazome *et al.*, 2007), but with L-L chirality only for the non-layered structures of Ala-Ala (Fletterick *et al.*, 1971) and two other dipeptides with small side chains (see Table S3 in the supporting information).

[1] A fully anisotropic refinement (175 rather than 140 parameters for 903 unique, observed reflections) decreased the *R*-value to 0.0791, but moderately *increased* standard uncertainties for geometric parameters. There were no significant changes to the molecular geometry.

(2) Utilization of the main-chain amide carbonyl group with formation of hydrophobic nanotubes. This was first seen for Val-Ala (Görbitz & Gundersen, 1996*b*) and has subsequently recurred for numerous other members of the Val-Ala class (Görbitz, 2003; Yadav *et al.*, 2015).

(3) Using an acceptor in a co-crystallized organic molecule. This is usually a solvent, but could also be a solid substance at room temperature, in which case the term 'co-crystal' may be more appropriate (Aitipamula *et al.*, 2012). The hydrophobic parts of such co-crystallized molecules are typically fully embedded in hydrophobic layers in the structure.

(4) Formation of $-NH_3^+\cdots$water hydrogen bonds in hydrates. These include the Phe-Phe class (Görbitz, 2001*b*), where water molecules are located within hydrophilic channels in structures with one-dimensional hydrogen-bonding patterns.

(5) Use of weaker acceptors in the side chains, such as the aromatic ring of Phe or the S atom of Met.

The molecular structure of (I) displayed in Fig. 2 is a typical example of alternative (1): two amino H atoms are engaged in head-to-tail chains in the standard S5 pattern shown in Fig. 1, while the third is accepted by a co-crystallized acetonitrile molecule. It follows that (I) has a layered structure, and the packing diagram in Fig. 3 shows that in the S5 class there is just one type of hydrophobic region, not two as for structures with T5 hydrogen-bonding (Görbitz, 2010). Table 2, listing hydrogen bond geometry parameters, has four entries with C—H donors, including C1—H11···O1 (Fig. 1) and the intramolecular contact in Fig. 2. The remaining two C—H···O interactions involve the acetonitrile molecules, which form two parallel head-to-tail chains inside solvent channels along the *a*-axis, Fig. 4.

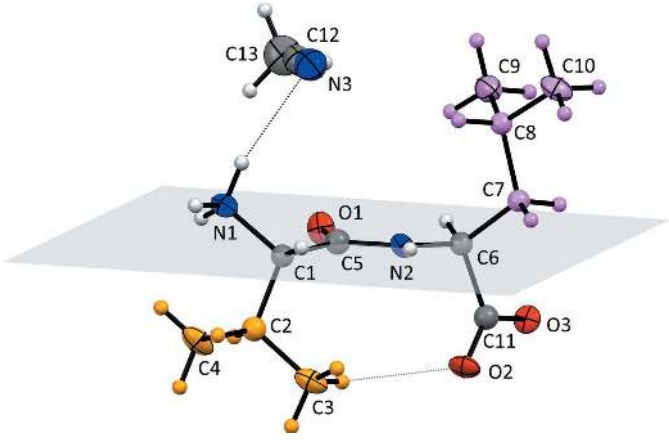


Figure 2
The molecular structure of (I) at 100 K. Displacement ellipsoids are shown at the 50% probability level; atoms depicted as spheres were refined isotropically. Bond length are normal, except C11—O3, which at 1.317 (13) Å is 0.079 Å longer than C11—O2. In other S5 structures the bond length difference between these two bonds is in the range 0.1–0.4 Å. The molecular conformation puts the Val side chain (orange) and the Leu side chain (violet) on opposite sides of the least-squares plane through the peptide bond (transparent grey). There is one intermolecular hydrogen bond to acetonitrile and one weaker intramolecular bond with a methyl donor (dashed lines).

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

D—H···*A*	*D*—H	H···*A*	*D*···*A*	*D*—H···*A*
N1—H1···N3	0.91	2.35	3.247 (12)	167
N1—H2···O2^{i}	0.91	2.05	2.851 (13)	146
N1—H2···O1^{ii}	0.91	2.50	3.071 (12)	122
N1—H3···O3^{iii}	0.91	1.91	2.778 (11)	158
N2—H4···O3^{ii}	0.88	2.07	2.922 (11)	162
C1—H11···O1^{ii}	1.00	2.51	3.173 (15)	124
C3—H32···O2	0.98	2.45	3.348 (14)	153
C13—H132···O2^{iii}	0.98	2.46	3.413 (14)	164
C13—H133···N3^{iv}	0.98	2.49	3.409 (15)	157

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

In Fig. 5 the crystal packing of (I) in (*a*) is compared with a number of other structures with corresponding S5 hydrogen bonding. When the space group is $P2_12_12_1$, head-to-tail chains in adjacent sheets along the vertical axis run in opposite directions. For Figs. 5(*a*), 5(*b*) and 5(*c*) this means that co-crystallized molecules, which may vary in size and number [one per asymmetric unit for 5(*a*) and 5(*b*), two for 5(*c*)] make contact across the hydrophobic interface to stacks of equivalent molecules in the other half of the hydrophobic region, Figs. 5(*a*) and 4(*b*). Such contacts are absent when the space group is $P2_1$, Figs. 5(*e*) and 5(*f*). The $P2_12_12_1$ structure of Phe-Val in Fig. 5(*d*) (CSD refcode XEGNAY; Görbitz, 2000) is a rare S5 structure devoid of co-crystallized solvent. In this case the third amino H atom is accepted by the aromatic ring of Phe, the enlarged view showing two N—H···C contacts, but also a C(ar)—H···O=C hydrogen bond. S5 patterns are compatible with Z'-values > 1, in which case independent molecules are always stacked in the direction of the $C(5)$ chain, *i.e.* along the viewing direction in Fig. 5. Figs. 5(*g*) and 5(*h*) show examples with $Z' = 2$ in space group $P2_1$ and $P2_12_12_1$, respectively.

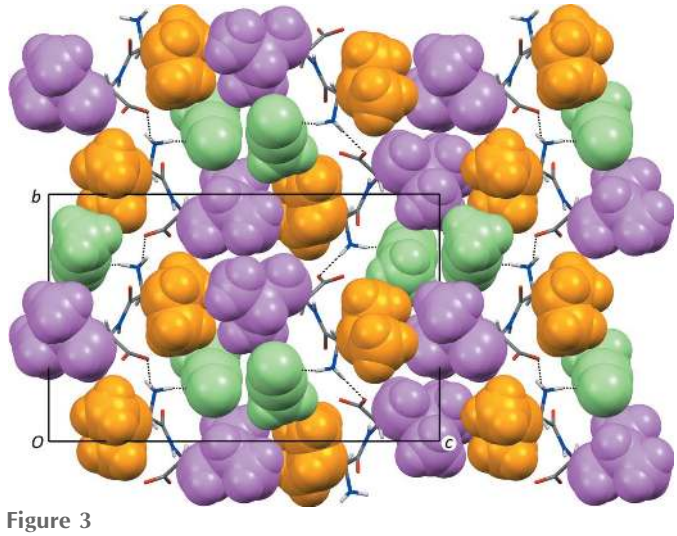


Figure 3
The molecular packing of (I) viewed along the *a*-axis with atoms constituting the hydrophobic regions of the crystal in spacefill representation. Colour coding as in Fig. 1, except that acetonitrile is depicted in light green. Hydrogen bonds occur only in narrow hydrophilic sheets.

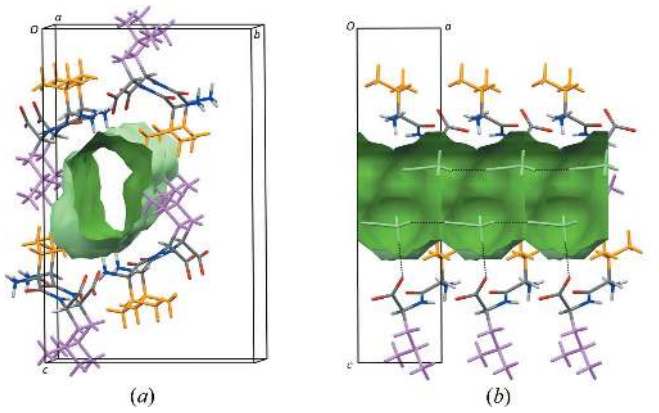


Figure 4
(*a*) Acetonitrile solvent channel (inner surface dark green, outer surface light green) running along the *a*-axis of (I). The illustration was prepared by removing acetonitrile and using the void calculator in *Mercury* with default probe radius 1.2 Å and grid spacing 0.7 Å. (*b*) Interior of channel viewed along the *b*-axis. Acetonitrile molecules are connected by weak hydrogen bonds, but also interact with the peptide carboxylate groups.

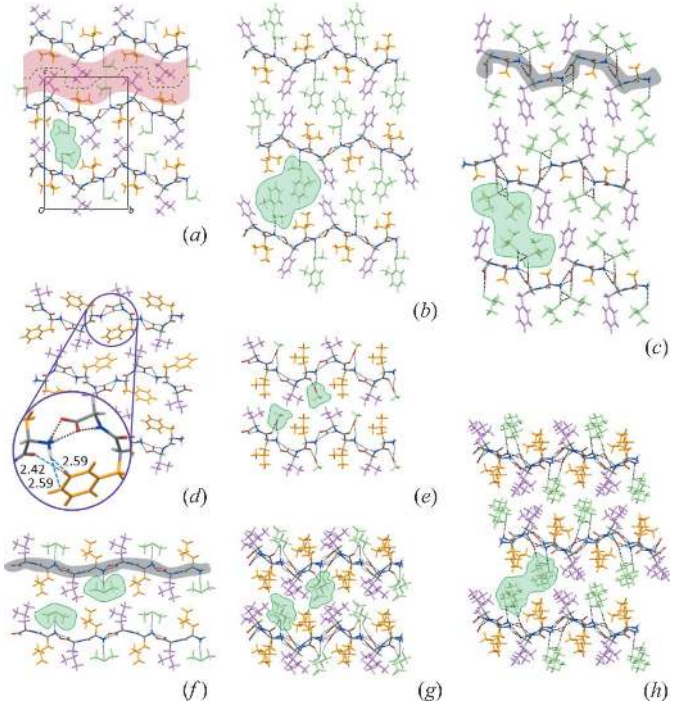


Figure 5
Crystal packing of a selection of structures with S5 hydrogen-bonding patterns. Side chains of *N*- and *C*-terminal residues are coloured in orange and violet, respectively, while co-crystallized molecules are light green. The structure of (I) is viewed along the *a*-axis in (*a*). Each hydrophobic region, one shaded in red, is divided into two parts by a central interface (red dashed line). Other structures are: (*b*) Ile-fGly methyl *o*-tolyl sulfoxide solvate (OKOZIY; Akazome *et al.*, 2010), (*c*) Ala-Phe 2-propanol solvate (1/2) (COCGEG; Görbitz, 1999*b*), (*d*) Phe-Val (XEGNAY; Görbitz, 2000) with a detail showing aromatic contacts in Å, (*e*) Leu-Val methanol solvate (SUWLIF; Görbitz & Torgersen, 1999), (*f*) Leu-Leu DMSO solvate (YORPEA; Mitra & Subramanian, 1994), (*g*) Leu-Ile trifluoroethanol solvate (IKOMOM; Görbitz, 2004*d*, and (*h*) Leu-Leu 2-methyl-1-propanol solvate (HIQWAF; Görbitz, 1999*a*). Green shades cover stacks of co-crystallized molecules, while grey shades in (*c*) and (*f*) emphasize layer undulation. See text for further details.

Anhydrates and hydrates

N-terminal residue Xaa-Xaa

C-terminal residue Xaa-Xaa	Ala	Val	Ile	Leu	Phe
Ala	ALAALA **T5-t**	WIRYEB NAYZET acetonitrile* hydrate*	AQARUF	RAVMOQ **S5*** tetrahydrate	QIMBUJ dihydrate
Val	XUDVOH XUDWAU 2-propanol* hydrate*	AQASIU hydrate*	AQASAM hydrate*	NAFZID 0.75 hydrate	XEGNAY **S5**
Ile	AQAROZ hydrate*	AQASEQ hydrate*	YAGZOW (**S5***) dihydrate	ETIWIN 0.75 hydrate HIZCOJ 2.5 hydrate	PAJPUM **S5/T5** hydrate*
Leu	DEZQOO **T5** hemihydrate		ETITUW **T5-t** hydrate*	IDUZOW **T5-t** hydrate*	IFABAS **T5-t** hydrate*
Phe		MOBYAD **T5**(t)** dihydrate MOBYEH **c** Trihydrate	ETONIK **T5**(t)** dihydrate	IDUZUC **T5-t** hydrate*	IFABEW **T5-t** hydrate*

(*a*)

Solvates

N-terminal residue Xaa-Xaa

C-terminal residue Xaa-Xaa	Ala	Val	Ile	Leu	Phe
Ala				TELVOV **S4** DMSO FEHPAK **S4** BMSO	
Val				JUCSEF01 **S5** 2-propanol SUWLIF **S5** methanol SUWLOL **S5** ethanol	
Ile				IKOMOM **S5** TFE	
Leu		***This work*** **S5** acetonitrile		HIQWAF **S5** 2-m-1-prop. JUQQIV ***anti*$_{2T}$** ethanol YORPEA **S5** DMSO	
Phe	COCGEG **S5** 2.0 2-propanol	COKGIK **S4** 2-propanol		COCGOQ **S4** 2-propanol hydrate*	JOQLIM **S4** 2.0 methanol

(*b*)

Val-Ala class, hydrophobic pores	Hydrophobic columns, no pores	2-D layered structure	Phe-Phe class, hydrophilic pores

Figure 6
Overview of the crystal packing of the 25 hydrophobic dipeptides, with anhydrates and hydrates in (*a*) and solvates in (*b*). Each structure is identified by its CSD refcode; codes for hydrogen-bonding patterns are taken from Fig. 1. Additionally, c for Val-Phe trihydrate (MOBYEH; Görbitz, 2002) denotes a complex pattern for a structure with $Z' = 8$, while S5/T5 for Phe-Ile hydrate (PAJPUM)(Görbitz, 2004*c*) represents a hybrid pattern for a structure with $Z' = 2$ (see Fig. S3 in the supporting information). The T5**(t) code is used for a subset of the T5** structures, see text for details. Box colour coding (legend at the bottom) identifies major structural classes. An asterisk after a solvent name means a non-stoichiometric composition. Solvent abbreviations: DMSO is dimethyl sulfoxide; BMSO is benzyl methyl sulfoxide; TFE is trifluoroethanol; 2-m-1-prop. is 2-methyl-1-propanol. Multiple entries within each boxes are different in at least one important aspect (hydrogen-bonding pattern, Z'-value or space group). Val-Ala class structures (yellow) and structures with hydrophobic columns (pink) have three-dimensional hydrogen-bonding patterns, as has Leu-Ala tetrahydrate (RAVMOQ; Görbitz, 1997). Leu-Val 0.75-hydrate (NAFZID; Görbitz & Gundersen, 1996*a*) and Leu-Ile 0.75-hydrate (ETIWIN; Görbitz, 2004*e*) are isostructural, as are Val-Phe dihydrate (MOBYAD; Görbitz, 2002) and Val-Ile dihydrate (ETONIK; Görbitz, 2004*b*). Packing diagrams for several of the structures in (*a*) are provided in Fig. S4 in the supporting information.

Fig. 6 gives an overview of the investigated structures of the 25 Xaa-Xaa dipeptides composed of Ala, Val, Ile, Leu and Phe, with nanoporous structures of the Val-Ala class and the Phe-Phe class (Görbitz, 2007) in yellow and blue, respectively. Selected structures are illustrated in Fig. 7. For anhydrates and hydrates in Fig. 6(*a*) the tremendous impact of the side chain on the crystal packing is immediately clear. Not only is size of importance, small side chains being more likely to form Val-Ala class structures (Fig. 7*d*) and *vice versa* for bulky side chains and the Phe-Phe class (Fig. 7*h*), but selection between these classes is also dictated by shape: Ile promotes the former while Leu promotes the latter (Görbitz, 2007). This means that for hydrophobic dipeptides any substitution of Leu for Ile (or opposite) leads to a completely different structure, the only exception is the isostructural pair Ile-Leu hydrate (ETITUW; Görbitz, 2004*d*) and Leu-Leu hydrate (IDUZOW; Görbitz, 2001*b*). In contrast, Val and Ile can always be interchanged without major structural changes, only Ile-Ile dihydrate (YAGZOW; Fig. 7*b*) (Görbitz, 2004*a*), for which the combined bulk of the two side chains is evidently incompatible with the Val-Ala class packing arrangement, is different from its Val-Ile (AQASEQ) and Ile-Val siblings (AQASAM; Görbitz, 2003). Furthermore, symmetry around the Ala-Ala – Phe-Phe diagonal in Fig. 6(*a*) is almost perfect, being broken only by the hydrates of Leu-Ile (HIZCOJ, ETIWIN; Görbitz & Rise, 2008; Görbitz, 2004*e*), which are both different from Ile-Leu hydrate (ETITUW; Görbitz, 2004*d*), and Leu-Val hydrate (NAFZID, Fig. 7*d*) (Görbitz & Gundersen, 1996*a*), which has no counterpart for Val-Leu. For co-crystals and solvates, such symmetry is absent, Fig. 6(*b*). Disregarding Phe-Phe methanol solvate (JOQLIM) (Mason *et al.*, 2014), there are three solvates of Xaa-Phe peptides, but none for Phe-Xaa peptides. Even more striking is the asymmetry for compounds with Leu residues. Other than Leu-Leu, there are seven entries for Leu-Xaa dipeptides in Fig. 6(*b*), but only one, the structure of (I), for *C*-terminal Leu. In total, there are 11 co-crystals incorporating Leu, and four additional pseudopolymorphs (Table S2 in the supporting information) bring the number up to 15. In contrast, Leu-Ile·trifluoroethanol (IKOMOM; Görbitz, 2016), with a very potent hydrogen bond acceptor in the alcohol, is the only solvate of a dipeptide with Ile, *i.e.* there are no known co-crystals of hydrophobic dipeptides with *N*-terminal Ile. This rather unexpected observation is not the result of lack of trying to crystallize such solvates.

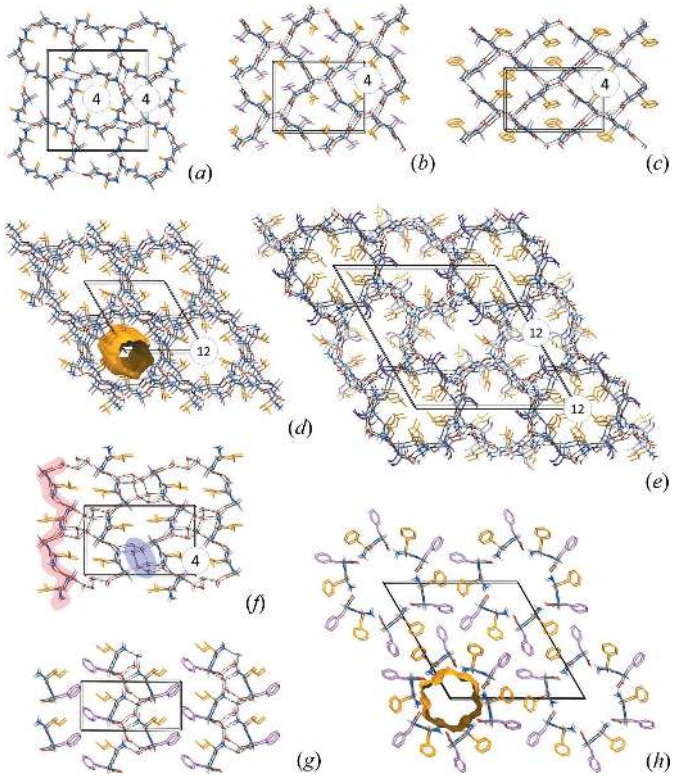


Figure 7
Crystal packing of hydrates and anhydrates of Xaa-Xaa dipeptides. Side-chain H atoms are omitted. The structures are: (*a*) Ala-Ala (ALAALA; Fletterick *et al.*, 1971), (*b*) Ile-Ile dihydrate (YAGZOW; Görbitz, 2004*a*), (*c*) Phe-Ala dihydrate (QIMBUJ; Görbitz, 2001*a*), (*d*) Val-Ala (WIRYEB; Görbitz & Gundersen, 1996*b*), (*e*) Leu-Val 0.75-hydrate (NAFZID; Görbitz & Gundersen, 1996*a*), (*f*) Leu-Ala tetrahydrate (RAVMOQ; Görbitz, 1997), (*g*) Ile-Phe dihydrate (Görbitz, 2004*b*) with layers and an unusual molecular conformation similar to (*h*) Phe-Phe (IFABEW; Görbitz, 2001*b*). Numbers inside circles indicate how many side chains contribute to each hydrophobic column [in (*a*) and (*e*) there are two different types]. Channels have been highlighted in (*d*) and (*h*); the latter are hydrophilic and no number in circle is given. The two Ile side chains in (*b*) have roughly the same volume (4 + 4 C atoms) as the Phe and Ala side chains in (*c*) (7 + 1 C atoms); these dipeptides can thus be isostructural. In (*f*) there are S5* dipeptide layers (red shade), but due to columns of water molecules (blue shade) the overall hydrogen-bonding pattern is three-dimensional.

In hydrates and anhydrates of dipeptides with one or two polar residues, the four basic hydrogen-bonding patterns are common (Table S3 in the supporting information), but for hydrophobic dipeptides they occur only twice, for Val-Phe (XEGNAY, S5; Görbitz, 2000) and Ala-Leu hemihydrate (DEZQOO, T5; Görbitz, 1999*c*), *i.e.* in two out of 23 structures (Fig. 6*a*). This clearly demonstrates that layered structures with embedded water molecules as receptors for the third amino H atom are not easily formed. Instead, hydrophobic groups aggregate into large columns with contributions from four to 12 side chains, and the hydrogen-bonding patterns become three-dimensional, see Fig. 6(*a*) and Fig. 7. The picture changes dramatically with the co-crystals, where 13 out of 14 structures have basic S4 or S5 patterns. It follows that formation of solvates/co-crystals automatically entails introduction of crystallographic screw axes along the *C*(8) hydrogen-bonded chains.

The type of hydrogen-bonding pattern in a dipeptide structure has a profound impact on main-chain conformation. Table 3 summarizes torsion angle data for the patterns in Fig. 1 (except for S4* with five heterogeneous structures, Table S4 in the supporting information). According to Fig. 6(*a*), Ala-Ala can be classified as T5-t, just as the Phe-Phe-class, but the crystal packing is completely different, and Ala-Ala is not included in the statistics for T5-t. As is apparent also from Fig. 1, S4 and T4 share a common, extended conformation, represented in Fig. 8(*a*) by the blue S4 structure of Leu-Ala DMSO solvate (TELVOV; Mitra *et al.*, 1996). Structures of the S5 group are more variable in terms of molecular conformations, reflected by high sample standard deviations in Table 3,

Table 3
Peptide torsion angles (°)† for various dipeptide hydrogen-bonding patterns.

Class	N_s/N_m‡	ψ_1		ω_1§		φ_2		ψ_T	
S4	22/27	138.2	11.2	176.2	4.7	−153.7	8.0	−11.9	15.9
T4	7/7	133.7	11.4	174.0	8.2	−152.1	16.7	−11.3	11.3
S5	26/34	151.6	14.5	170.1	5.5	−96.2	19.2	−25.7	41.3
(I)		140.9	–	170.6	–	−95.9	–	−11.7	–
T5	15/15	162.1	6.0	168.1	6.0	−72.0	8.0	−23.7	6.0
T5-t	7/14	133.9	17.3	181.1	4.4	52.1	3.5	43.6	8.1
T5**	6/6	162.8	1.8	177.7	2.8	−83.3	13.9	−31.3	14.9
T5**(t)	4/4	153.8	8.6	170.1	2.8	51.2	2.9	45.9	5.9

† Torsion angle labels defined in Scheme 1, full statistics are given in Table S4 in the supporting information. For each angle: average value first column, sample standard deviation second column. ‡ N_s = number of CSD structures, N_m = number of independent molecules in these structures. § Calculated using only positive values; −175° was consequently entered as 185° *etc.*

but are usually more folded than S4/T4. The representative example in Fig. 8, Met-Asn (TARKUT; Stievater & Srikrishnan, 2005), still has side chains on opposite sides of the peptide plane, in a similar manner to (I) in Fig. 2. The comparatively extended geometry of Leu-Leu in the DMSO solvate (YORPEA; Mitra & Subramanian, 1994) with φ_2 = −149.5° leads to almost planar sheets in Fig. 5(*f*), while smaller φ_2-values give rise to more undulating sheets, as for Ala-Phe 2-propanol solvate (COCGEG, φ_2 = −77.6°) (Görbitz, 1999*b*) in Fig. 5(*c*). Both data in Table 3 and Fig. 5 demonstrate that (I) is a rather average S5 member.

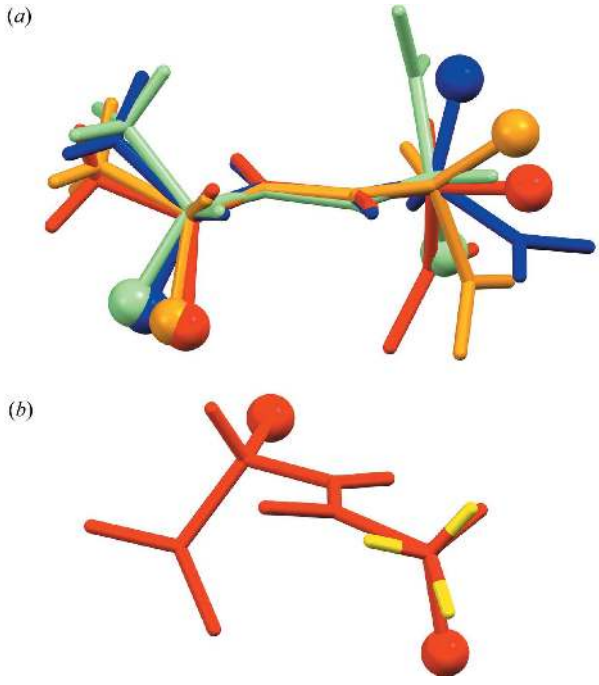


Figure 8
(*a*) Comparison of peptide main-chain conformations for main hydrogen-bonding classes. Side chains have been curtailed beyond the C^{β}-atoms shown as spheres. The overlap has been calculated from a least-squares fit for the five non-H atoms of the peptide bond. In blue colour is the S4 structure of Leu-Ala DMSO solvate (TELVOV; Mitra *et al.*, 1996), in orange the S5 structure of Met-Asn (TARKUT; Stievater & Srikrishnan, 2005), in red the T5 structure of Ale-Leu hemihydrate (DEZQOO; Görbitz, 1999*c*), and in light green the T5-t conformation of Ile-Leu 0.91-hydrate (ETITUW; Görbitz, 2004*d*). (*b*) Rotated view for Ala-Leu hemihydrate, amino H atoms in yellow.

For the very homogeneous T5 structures, represented in red in Fig. 8(*a*) by Ala-Leu hemihydrate (DEZQOO; Görbitz, 1999*c*), the side chain is more or less in the peptide plane. Not evident from Table 3 is the signature eclipsed conformation of the amino group, Fig. 8(*b*). The radically different T5-t conformation of Ile-Leu 0.91-hydrate (ETITUW; Görbitz, 2004*d*) in light-green colour in Fig. 8(*a*) puts side chains on the same side of the peptide plane, which is a prerequisite for formation of the hydrophilic channels of the Phe-Phe class.

During analysis of structures with T5** patterns, it became obvious that they had to be divided into two subgroups. The first (and larger), illustrated in Fig. 1, retains the T5** code and also the T5-conformation, except that the large deviation from planarity for ω_1 is absent (Table 3). Four closely related dihydrates in the second subgroup, including Val-Phe (MOBYAD; Görbitz, 2002) and Ile-Phe (ETONIK; Görbitz, 2004*b*) (Fig. 7*g*) as well as Nva-Phe (VIKWUJ; Görbitz & Yadav, 2013) and Ile-Trp (BEQJAJ; Sun & Oldfield, 2004) have, on the other hand conformations that largely coincide with the T5-t structures, even though they have layers rather than hydrophilic pores. The pattern code T5**(t) has been used for these structures.

All dipeptides discussed so far have L-L-chirality. D-D-Dipeptides would have equivalent mirror image structures, but L-L:D-D racemates would be different, as would enantiomeric L-D (and their mirror image D-L) structures and their L-D:D-L racemates. Very few such entries are available in the CSD; Table 3S lists four Gly LD-Xaa racemates, four L-Xaa-D-Xaa structures and two L-D:D-L racemates. Nevertheless, this small group suggests that heterochiral structures will also adopt the four basic patterns (four examples of S4 sheets) and patterns with *anti* chains (one example), but as illustrated in Fig. S1, heterochirality brings about certain new features, such as formation of amino acid-like double sheets for LD-Ala-DL-Met (Murali *et al.*, 1986).

4. Summary

Among organic substances, the collection of 25 dipeptides constructed from the five hydrophobic residues Ala, Val, Ile, Leu and Phe constitute a unique group where the balance between formation of favourable hydrogen-bonding networks and aggregation of hydrophobic moieties may be studied in detail and leads to very diverse families of structures. Within each group, the 'packing problem' of the peptide molecules, *i.e.* how to position three hydrogen-bond acceptors around the charged amino-group donor, has found a systematic and unique solution that depends not only on the size of the side chains, but also on their shape. Among the more unexpected observations for hydrophobic dipeptides is the remarkable tendency of compounds with a Leu-residue, particularly at the *N*-terminal, to incorporate co-crystallized (solvent) molecules. For related substances with Ile, with the same number of side-chain C atoms, this property is absent. Solvates and other co-crystals of these dipeptides have been found to crystallize in

space groups with screw axes and adopt S4 or S5 hydrogen-bonding patterns. The anhydrates and hydrates are dominated by the nanotubular structures of the Val-Ala and Phe-Phe classes, but may also have their side chains stacked in large hydrophobic columns (without pores), rather than the ubiquitous hydrophobic layers observed for other classes of dipeptides.

References

Aitipamula, S. *et al.* (2012). *Cryst. Growth Des.* **12**, 2147–2152.
Akazome, M., Doba, A., Matsumoto, S. & Ogura, K. (2010). *J. Org. Chem.* **75**, 660–665.
Akazome, M., Hirabayashi, A., Senda, K. & Ogura, K. (2007). *Tetrahedron*, **63**, 9933–9938.
Bernstein, J., Davis, R. E., Shimoni, L. & Chang, N.-L. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 1555–1573.
Bruker (2016). *APEX3, 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.
Fletterick, R. J., Tsai, C. C. & Hughes, R. E. (1971). *J. Phys. Chem.* **75**, 918–922.
Görbitz, C. H. (1997). *Acta Cryst.* C**53**, 736–739.
Görbitz, C. H. (1999*a*). *Acta Cryst.* C**55**, 670–672.
Görbitz, C. H. (1999*b*). *Acta Cryst.* C**55**, 2171–2177.
Görbitz, C. H. (1999*c*). *Acta Cryst.* C**55**, IUC9900149, cif-access.
Görbitz, C. H. (2000). *Acta Cryst.* C**56**, 1496–1498.
Görbitz, C. H. (2001*a*). *Acta Cryst.* C**57**, 575–576.
Görbitz, C. H. (2001*b*). *Chem. Eur. J.* **7**, 5153–5159.
Görbitz, C. H. (2002). *Acta Cryst.* B**58**, 512–518.
Görbitz, C. H. (2003). *New J. Chem.* **27**, 1789–1793.
Görbitz, C. H. (2004*a*). *Acta Cryst.* B**60**, 569–577.
Görbitz, C. H. (2004*b*). *Acta Cryst.* C**60**, o371–o373.
Görbitz, C. H. (2004*c*). *Acta Cryst.* C**60**, o810–o812.
Görbitz, C. H. (2004*d*). *Acta Cryst.* E**60**, o626–o628.
Görbitz, C. H. (2004*e*). *Acta Cryst.* E**60**, o647–o650.
Görbitz, C. H. (2005). *Acta Cryst.* E**61**, o2012–o2014.
Görbitz, C. H. (2007). *Chem. Eur. J.* **13**, 1022–1031.
Görbitz, C. H. (2010). *Acta Cryst.* B**66**, 84–93.
Görbitz, C. H. (2016). *Acta Cryst.* E**72**, 635–638.
Görbitz, C. H. & Etter, M. C. (1992). *Int. J. Pept. Protein Res.* **39**, 93–110.
Görbitz, C. H., Gundersen, E. (1996*a*). *Acta Chem. Scand.* **50**, 537–543.
Görbitz, C. H. & Gundersen, E. (1996*b*). *Acta Cryst.* C**52**, 1764–1767.
Görbitz, C. H. & Rise, F. (2008). *J. Pept. Sci.* **14**, 210–216.
Görbitz, C. H. & Torgersen, E. (1999). *Acta Cryst.* B**55**, 104–113.
Görbitz, C. H. & Yadav, V. N. (2013). *Acta Cryst.* C**69**, 1067–1069.
Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* B**72**, 171–179.
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.
Mason, T. O., Chirgadze, D. Y., Levin, A., Adler-Abramovich, L., Gazit, E., Knowles, T. P. J. & Buell, A. K. (2014). *ACS Nano*, **8**, 1243–1253.
Mitra, S. N., Govindasamy, L. & Subramanian, E. (1996). *Acta Cryst.* C**52**, 2871–2873.
Mitra, S. N. & Subramanian, E. (1994). *Biopolymers*, **34**, 1139–1143.
Murali, R., Lalitha, V., Subramanian, E. & Parthasarathy, R. (1986). *Int. J. Pept. Protein Res.* **27**, 160–164.
Sheldrick, G. M. (2015*a*). *Acta Cryst.* A**71**, 3–8.
Sheldrick, G. M. (2015*b*). *Acta Cryst.* C**71**, 3–8.
Stievater, T. & Srikrishnan, T. (2005). *J. Chem. Crystallogr.* **35**, 529–533.
Sun, H. & Oldfield, E. (2004). *J. Am. Chem. Soc.* **126**, 4726–4734.
Yadav, V. N., Comotti, A., Sozzani, P., Bracco, S., Bonge-Hansen, T., Hennum, M. & Görbitz, C. H. (2015). *Angew. Chem. Int. Ed.* **54**, 15684–15688.