Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

β Turns, water cage formation and hydrogen bonding in the structures of L-valyl-L-phenylalanine

Carl Henrik Görbitz

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

Correspondence e-mail: c.h.gorbitz@kjemi.uio.no

Received 26 October 2001
Accepted 8 February 2002

L-Valyl-L-phenylalanine has been crystallized as an orthorhombic dihydrate (1) in the shape of needles and as a monoclinic trihydrate (2) with $Z = 16$ ($P2_1$, $Z' = 8$) in the shape of thin plates. Peptide molecules in these two structures occur in three basic conformations, termed c_1, c_{2A} and c_{2B}. c_{2B} has not been observed previously for dipeptides. Together with c_1 it forms a model pair for Type I and Type II β-turns in protein structures. The crystal packing of (2) is remarkable in that some of the L-Val side chains are exposed to the solvent region of the crystal rather than being located in a hydrophobic layer. The crystal packing thus offers a unique and detailed view of hydrogen-bond cage formation around the hydrophobic groups by the 24 cocrystallized water molecules. The eight $-NH_3^+\cdots{}^-OOC-$ contacts in the structure are unusually short and the minimum N$\cdots$O distance of 2.649 (5) Å represents a new extreme limit for this type of hydrogen bond in peptide structures.

1. Introduction

In the crystal structures of dipeptides with one or two completely or partly hydrophobic side chains (such as the $-CH_2-CH_2-$ group of L-Glu), peptide main chains are usually segregated into hydrophilic layers, while side chains (or the hydrophobic parts of them) form hydrophobic layers (Görbitz & Etter, 1992). Within a hydrophilic layer each amino group forms two hydrogen bonds to carboxylate groups, which by repetition gives rise to the so-called 'head-to-tail' chains (Suresh & Vijayan, 1985). Whenever there is an N- or C-terminal Gly residue the last amino H atom may be accepted by a third main chain carboxylate group, but for other dipeptides this H atom points straight into a hydrophobic layer where it is accepted by a functional group in a side chain, or, in the lack of such, a cocrystallized organic solvent molecule such as an alcohol or DMSO (Görbitz, 1999*a*). When both side chains are entirely hydrophobic and there are no organic solvent molecules, however, the following packing problem arises: How can a hydrogen-bond acceptor be positioned close to the third amino group? Research into the crystal structures of hydrophobic dipeptides has provided a surprisingly diverse set of solutions to this problem, including the formation of hydrophobic nanotubes (Görbitz & Gundersen, 1996*a*; Görbitz, 2002), hydrophilic nanotubes (Görbitz, 2001*a*) and hydrophobic layers in structures that may have additional hydrophobic columns (Stenkamp & Jensen, 1975; Görbitz, 1997, 1999*b*, 2000, 2001*b*; Görbitz & Gundersen, 1996*b*). The two crystal forms of L-Val-L-Phe

Table 1
Experimental details.

	(1)	(2)
Crystal data		
Chemical formula	$C_{14}H_{20}N_2O_3\cdot 2H_2O$	$C_{14}H_{20}N_2O_3\cdot 3H_2O$
Chemical formula weight	300.35	318.37
Cell setting, space group	Orthorhombic, $P2_12_12_1$	Monoclinic, $P2_1$
a, b, c (Å)	5.6595 (4), 8.3306 (6), 33.022 (2)	10.3137 (3), 45.2095 (12), 15.5891 (4)
β (°)	90	105.171 (1)
V (Å^3)	1556.90 (19)	7015.5 (3)
Z	4	16
D_x (Mg m^{-3})	1.281	1.206
Radiation type	Mo $K\alpha$	Mo $K\alpha$
No. of reflections for cell parameters	8448	49294
θ range (°)	2.47–27.47	1.43–29.57
μ (mm^{-1})	0.097	0.094
Temperature (K)	150 (2)	150 (2)
Crystal form, colour	Needle, colourless	Plate, colourless
Crystal size (mm)	1.50 × 0.15 × 0.05	1.10 × 1.0 × 0.01
Data collection		
Diffractometer	Siemens SMART CCD	Siemens SMART CCD
Data collection method	Sets of exposures each taken over 0.3° ω rotation scans	Sets of exposures each taken over 0.3° ω rotation scans
Absorption correction	Empirical	Empirical
T_{min}	0.865	0.902
T_{max}	0.995	0.999
No. of measured, independent and observed parameters	13 537, 3380, 3120	72 689, 19 842, 18 611
Criterion for observed reflections	$I > 2\sigma(I)$	$I > 2\sigma(I)$
R_{int}	0.0250	0.0346
θ_{max} (°)	27.47	29.57
Range of h, k, l	$-7 \rightarrow h \rightarrow 7$ $-10 \rightarrow k \rightarrow 10$ $-41 \rightarrow l \rightarrow 40$	$-14 \rightarrow h \rightarrow 14$ $-60 \rightarrow k \rightarrow 62$ $-19 \rightarrow l \rightarrow 21$
Refinement		
Refinement on	F^2	F^2
$R[F^2>2\sigma(F^2)]$, $wR(F^2)$, S	0.0302, 0.0715, 1.089	0.0751, 0.2081, 1.196
No. of reflections and parameters used in refinement	3380, 230	19 842, 1475
H-atom treatment	Mixed	Constrained
Weighting scheme	$w = 1/[\sigma^2(F_o^2) + (0.0350P)^2 + 0.2523P]$, where $P = (F_o^2 + 2F_c^2)/3$	$w = 1/[\sigma^2(F_o^2) + (0.0713P)^2 + 9.9786P]$, where $P = (F_o^2 + 2F_c^2)/3$
$(\Delta/\sigma)_{max}$	0.001	0.041
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})	0.179, −0.194	0.395, −0.417

Computer programs used: *SMART* (Bruker, 1998*b*), *SAINT* (Bruker, 1998*a*), *SHELXTL* (Sheldrick, 1997), *SADABS* (Sheldrick, 1996).

presented here provide two more ways of packing hydrophobic dipeptides in the solid phase.

2. Experimental

2.1. Crystal preparation

Crystals prepared by diffusion of simple alcohols into aqueous solutions of the peptide contained cocrystallized organic solvent molecules (Görbitz, 1999*c*). Slow evaporation from an aqueous solution produced very thin plate-shaped crystals. In an attempt to improve crystal quality various additives were tested and some of them changed the crystal habit to needles. The specimen used for data collection, form (1), was grown at room temperature by slow evaporation of an aqueous solution to which had been added a small amount of PEG-4000. After several new crystallizations by evaporation of pure aqueous solutions at 275 K plate-shaped crystals, form (2), with increased thickness were produced. The unstable crystals were harvested before evaporation had proceeded to dryness.

2.2. Data collection

The selected specimen of (2) was flash-frozen in N_2(l) before being transferred to the diffractometer. No such procedure was required for (1). More than a hemisphere of reciprocal space was collected by a combination of three sets of exposures for (1) and five sets of exposures for (2). Exposure times were 30 s. The crystal-to-detector distance was 5.0 cm for (1) and 6.0 cm for (2). Large crystals with dimensions up to 1.1–1.5 mm were used in order to maximize the diffraction intensities (Görbitz, 1999*d*).

2.3. Structure determination and refinement

Structure (1) was solved routinely by *SHELXTL* (Sheldrick, 1997). Positional parameters were refined for H atoms bonded to N and O atoms, other H atoms were included in theoretical positions with refinement of the distance to the bonded atom only. U_{iso} values for H atoms were $1.5U_{eq}$ (water) or $1.2U_{eq}$ (other) of the bonded atom, except that U_{iso} values were refined (with three free variables) for the freely rotating amino and methyl groups.

Attempts to solve structure (2) with *SHELXTL* (Sheldrick, 1997) were unsuccessful, but an almost complete structure including eight independent peptide molecules (*A*–*H*) was output by the program *XM* (Bruker, 2000), kindly provided by Professor G. Sheldrick. Subsequent refinement revealed an alternative conformation for the L-Val side chain of peptide molecule *C*, as well as an alternative position for the whole of molecule *E*, once again with a different orientation for the L-Val side chain. The minor components were termed *M* and *N*, respectively. The two disorder phenomena are linked and the distribution between the major [0.894 (3)] and minor

[0.106 (3)] components was refined with a single parameter. In order to reduce the number of refinement parameters, the same set of atomic displacement parameters (ADP's) was used for molecules *B* and *D*, which are related by close pseudotranslational symmetry. This procedure was also employed for molecules *F* and *H*. The *R*-factor increase associated with the substantial reduction in the number of refinement parameters from 1703 to 1475 was small, 0.0727 to 0.0751. Common ADP's as well as atomic coordinates were used for atoms in *C* and its minor component *M* outside the L-Val side chain. Independent O, N and C atom positions with 0.106 occupancy were refined isotropically. Peptide bond lengths and bond angles between C, N and O atoms were subject to loose restraints by *SHELXTL* SAME 0.01 0.02 commands.

In (2) there are 15 positions for water molecules with full occupancy and nine with occupancy 0.894. Seven water positions with occupancy 0.106 were also identified and an additional position with slightly (probably significantly) higher occupancy, 0.143 (15). Approximately 40 of the 48 major positions of water H atoms were located in the electron density maps; approximate positions for the remaining atoms could easily be elucidated from the hydrogen-bond pattern. The O—H bond lengths were tightly restrained to 0.85 Å, while some variation in the H—O—H angles was allowed by applying slightly looser 1.35 Å restraints to the internal water H···H distances.

Experimental data and refinement results are summarized in Table 1.[1]

2.4. Database searches

Structures were retrieved from the Cambridge Structural Database (CSD; April 2001 release; Allen & Kennard, 1993) by means of the program *ConQuest*1.2 without the use of any filters other than the presence of three-dimensional coordinates.

3. Results and discussion

The asymmetric units of (1) and (2) are shown in Figs. 1 and 2, respectively. The observation of eight molecules in the asymmetric unit, as in (2), is rare. A recent survey covering all structures in the CSD found a total of 28 structures (20 organic, eight organometallic) with $Z' = 8$ (Steiner, 2000). Bond lengths and bond angles are normal. Torsion angles are listed in Table 2.

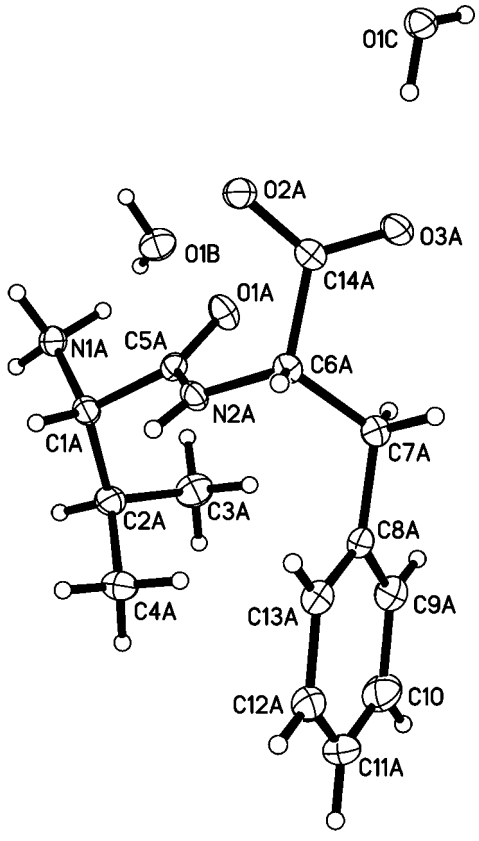


Figure 1
The asymmetric unit of (1) with atomic numbering scheme. Displacement ellipsoids are shown at the 50% probability level. H atoms are shown as spheres of arbitrary size.

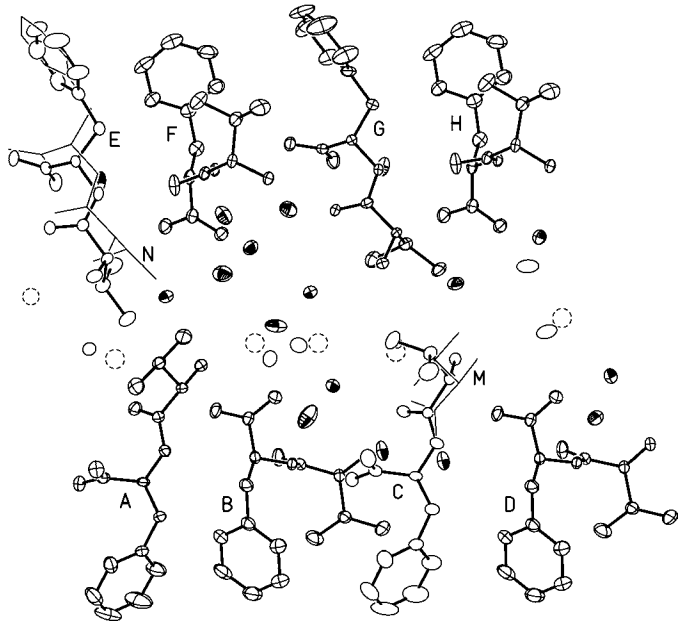


Figure 2
The asymmetric unit of (2). H atoms have been omitted for clarity. Atomic numbering in the peptide molecules follows the same general scheme as in Fig. 1 and is not shown. Atoms with full occupancy are shown as ellipsoids with principal ellipses, water molecules have additional shaded segments. Atoms with occupancy 0.894 are shown with boundary ellipses only. Water molecules with occupancy 0.106–0.143 are shown as dotted spheres, while minor positions for peptides appear as stick drawings.

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: OS0088). Services for accessing these data are described at the back of the journal.

Table 2
Torsion angles (°).

Torsion		(1)				(2)					
		A	*C*	*M*	*E*	*N*	*G*	*B*	*D*	*F*	*H*
N1—C1—C5—N2 (ψ_1)	151.35 (11)	125.9 (4)	140.8 (4)	†	136.4 (4)	163 (3)	160.5 (4)	−49.6 (5)	−62.7 (5)	−52.1 (5)	−49.1 (5)
C1—C5—N2—C6 (ω_1)	172.31 (11)	174.4 (3)	179.8 (4)	–	−179.6 (4)	175 (2)	176.2 (3)	−172.8 (3)	−174.3 (3)	−173.5 (3)	−170.3 (3)
C5—N2—C6—C14 (φ_2)	48.55 (16)	−60.6 (5)	−61.9 (5)	–	−59.5 (5)	−57 (3)	−59.8 (5)	−101.0 (4)	−92.5 (4)	−94.5 (4)	−102.6 (4)
N2—C6—C14—O2 (ψ_T)	48.45 (16)	−48.9 (5)	−49.5 (5)	–	−47.3 (5)	−45 (4)	−47.8 (5)	−10.4 (5)	−11.2 (5)	−22.7 (5)	−23.8 (5)
N1—C1—C2—C3 ($\chi_1^{1,1}$)	70.85 (15)	−64.2 (4)	59.8 (5)	−61 (3)	−63.4 (5)	67 (3)	59.7 (4)	−63.3 (5)	−62.7 (5)	−63.7 (5)	−60.0 (5)
N1—C1—C2—C4 ($\chi_1^{1,2}$)	−164.06 (11)	173.0 (4)	−174.2 (4)	176 (3)	173.7 (5)	−59 (3)	−68.6 (4)	172.8 (4)	174.1 (4)	173.0 (4)	176.3 (4)
N2—C6—C7—C8 (χ_2^1)	−50.08 (16)	−174.2 (3)	−168.5 (4)	–	−169.9 (4)	−165 (2)	−168.0 (4)	−61.5 (4)	−65.7 (4)	−62.6 (4)	−63.0 (4)
C6—C7—C8—C9 (χ_2^2)	138.07 (13)	91.6 (6)	88.2 (6)	–	92.4 (6)	87 (4)	85.0 (7)	102.8 (5)	106.7 (5)	102.5 (5)	106.5 (5)
C2—C1···C7—C8 (θ)	19.97 (12)	−107.1 (4)	−97.0 (4)	−103 (2)	−98.9 (5)	−69 (2)	−72.9 (4)	22.9 (4)	23.3 (4)	26.3 (4)	22.2 (4)

† Same torsion angle as *C*.

3.1. Molecular conformations

A simplified description of the conformation of a dipeptide is provided by the torsion angle θ = C1$^{\beta}$—C1$^{\alpha}$···C2$^{\alpha}$—C2$^{\beta}$, which defines the relative positions of the two side chains. A previous investigation of zwitterionic L-Xaa-L-Xaa dipeptides (Xaa is not Gly or Pro; Görbitz, 2001*a*) revealed that side chains usually point in almost opposite directions, as reflected by 42 structures (out of 75) with $|\theta| > 135°$, 28 with $90 < |\theta| < 135°$, 3 with $45 < |\theta| < 90°$ and only two with $|\theta| < 45°$. Rare values for θ close to 0° were subsequently found for the four dipeptides L-Leu-L-Leu, L-Leu-L-Phe, L-Phe-L-Leu and L-Phe-L-Phe (Görbitz, 2001*a*), brought about primarily by C′—N—C$^{\alpha}$—C′ (φ_2) torsion angles around 50°. These conformations were shown to result from unique crystal packing arrangements with hydrophilic channels embedded in a hydrophobic matrix generated by the large non-polar side chains.

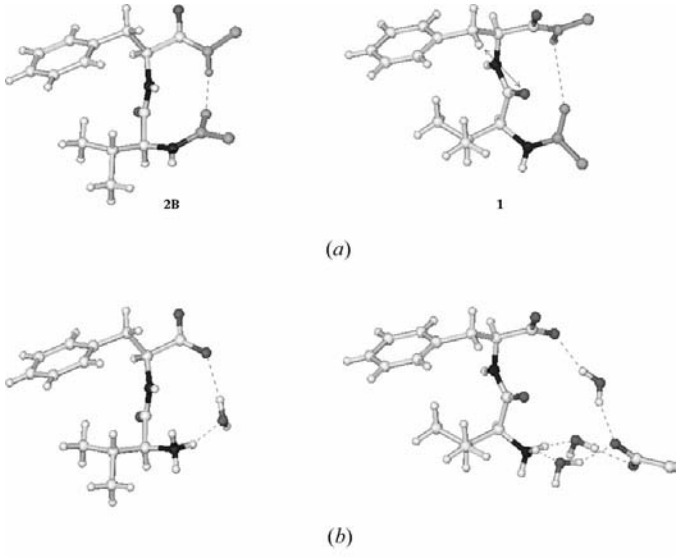


Figure 3
(*a*) Molecule *B* in (2) incorporated into a Type I β turn (left) and the peptide molecule in (1) incorporated into a Type II β turn (right). The arrow indicates the steric conflict for a non-Gly residue in position 3 of a Type II turn. Terminal modifications are colored in a gray half-tone. (*b*) The same molecules as in (*a*), but shown as they appear in the crystal structures.

The conformation of L-Val-L-Phe in (1), hereafter termed c_1, is actually very close to those of the abovementioned four dipeptides. The ten different molecules in (2) can, on the basis of the main chain conformations, be divided into two groups termed c_{2A} and c_{2B}. c_{2A} contains *A*, *C*, *M*, *E*, *N* and *G*, and c_{2B} contains *B*, *D*, *F* and *H*, Table 2. c_{2B} is very homogenous with r.m.s. values around 0.10 Å for the best all-atom overlap between individual molecules.

Both c_1 and c_{2B} have peptide side chains on the same side of the peptide plane, giving θ values in the range 20.0–26.4°, Table 2. The two conformations are distinguished by what can essentially be described as a flip of the peptide unit along the C$_1^{\alpha}$···C$_2^{\alpha}$ (= C1···C6) axis. c_{2B} thus corresponds closely to residue 2 and 3 in a Type I β turn in polypeptides, while c_1 corresponds to a Type II β turn (Fig. 3*a*). In Type II turns the carbonyl O atom of residue 2 becomes close to the C$^{\beta}$ atom of residue 3, which is therefore usually Gly. Indeed, all 22 molecules found in a CSD search for β turns with L-Phe as residue 3 had Type I turns. The incorporation of L-Phe at this position in c_1 is rendered possible by (i) a shift of the C5—N2—C6—C14 torsion angle (φ_2) from 90° in the idealized Type II turn to 48.6°, and (ii) a distinct deviation from planarity for the peptide bond giving C1—C5—N2—C6 (ω_1) = 172.3° (c_{2B} has ω_1 around −173°). The potential steric conflict is consequently alleviated to give a 2.91 Å H···O distance, as indicated in Fig. 3(*a*).

The *gauche* conformation observed for the L-Phe side chain in c_{2B} is in agreement with the 16:4:2 *gauche*−:*gauche*+:*trans* distribution for L-Phe in the 22 Type I turn CSD structures. For

Table 3
Hydrogen-bond parameters in (1) (Å, °).

Hydrogen bond	D—H	H···A	D···A	D—H···A
N1A—H1A···O1C^{i}	0.904 (19)	2.073 (19)	2.9201 (16)	155.5 (16)
N1A—H2A···O1C^{ii}	0.977 (17)	1.814 (17)	2.7855 (15)	172.4 (15)
N1A—H3A···O1B	0.970 (18)	1.930 (18)	2.8439 (16)	156.1 (15)
N2A—H4A···O3A^{iii}	0.811 (18)	2.173 (17)	2.8972 (15)	148.9 (15)
O1B—H1B···O2A^{ii}	0.89 (2)	1.93 (2)	2.7759 (14)	158.1 (17)
O1B—H2B···O2A^{iv}	0.87 (2)	2.05 (2)	2.8780 (15)	158.8 (17)
O1C—H1C···O3A	0.916 (19)	1.839 (19)	2.7514 (14)	175.5 (16)
O1C—H2C···O2A^{v}	0.91 (2)	0.91 (2)	2.6657 (14)	172.4 (17)

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

the L-Val side-chain model studies have shown that steric conflict with the peptide bond >N—H is least for the *gauche*,*trans* rotamer, as observed for c_{2B}.

In c_1 the L-Phe side chain is also *gauche*, just as in the closely related L-Leu-L-Phe and L-Phe-L-Phe structures (Görbitz, 2001*a*). The L-Val *gauche*+,*trans* orientation is distinct from the *gauche*−,*trans* conformation of L-Val in the Type II turn of *cyclo*(-Gly-L-Thr-L-Phe-L-Leu-L-Tyr-L-Val) (Morita *et al.*, 1996), the only structure in the CSD with L-Val as residue 2 in an isolated turn (although there are several examples of incorporation of L-Val into 3_{10} and α helices). There is hardly any steric conflict, and a *gauche*+,*gauche*− should theoretically also be possible.

Members of the c_{2A} family of structures share a slightly more extended, but still fairly uncommon, peptide backbone with θ values between −72.7 (*G*) and −107.4° (*A*). The L-Phe side chain is always *trans*, but there is a choice of all three different rotamers for the L-Val side chain. The most frequently observed *trans*, *gauche*− conformation (Benedetti *et al.*, 1983; Ashida *et al.*, 1987) is found for molecules *A*, *E* and *M*, while *trans*, *gauche*+ occurs for *C* and *gauche*+, *gauche*− for *G* and *N*.

In summary, L-Val-L-Phe has a completely new dipeptide backbone conformation in c_{2B} and rare conformations in c_1 and c_{2A} with a total of five different combinations of side-chain conformations.

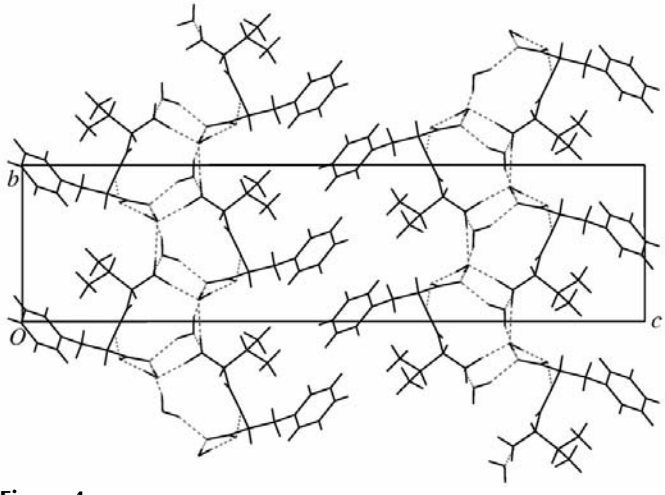


Figure 4
The unit cell and crystal packing of (1) viewed along the *a* axis with hydrogen bonds indicated in gray.

3.2. Crystal packing arrangement and hydrogen bonding

The crystal packing of (1) is shown in Fig. 4, while hydrogen-bonding parameters are listed in Table 3. Structures in which both dipeptide side chains are located in the same hydrophobic layer have previously been observed only for L-Tyr-L-Lys (Urpi *et al.*, 1988) and the racemates D,L-Ala-L,D-Met (Stenkamp & Jensen, 1974) and D,L-Ala-L,D-Val (Murali *et al.*, 1986). In the latter two structures as well as in the additional racemate Gly D,L-Phe (Marsh *et al.*, 1976) there are hydrophilic double layers composed of two hydrogen-bonded sheets connected by one or two hydrogen bonds, reminiscent also of the arrangement found in the crystal structures of hydrophobic amino acids (Dalhus & Görbitz, 1999).

A single sheet in (1) is shown in Fig. 5. The interactions between the terminal $-NH_3^+$ and $-COO^-$ groups that usually dominate the hydrogen-bonding patterns of dipeptides are missing in (1); all amino H atoms are instead accepted by solvent water molecules. Other dipeptide structures without head-to-tail chains also have cocrystallized water molecules in addition to either a charged multiple donor in the side chain [L-Arg-L-Glu·2H_2O (Pandit *et al.*, 1983) and L-Arg-L-Asp·2H_2O (Ramakrishnan & Viswamitra, 1988)] or a L-Tyr residue [L-Tyr-L-Phe·H_2O (Murali & Subramanian, 1987), L-Tyr-L-Val·H_2O (Ramakrishnan *et al.*, 1984), L-Tyr-L-Tyr·2H_2O (Cotrait *et al.*, 1984)]. The structure of L-Tyr-L-Lys (Urpi *et al.*, 1988) has no solvent water, but it is not directly comparable to the others as the positive charge is located in the side chain, while the main chain amino group is neutral. It is interesting to find that despite very different crystal packing arrangements and peptide conformations, hydrogen-bonding sheets in L-Tyr-L-Phe·H_2O (φ_2 = −70.6°, θ = −83.6°) and L-Tyr-L-Val·H_2O

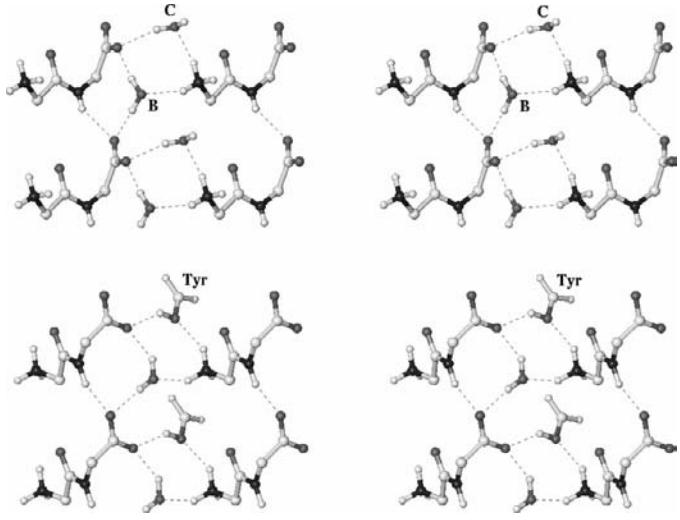


Figure 5
Stereoview of hydrogen bonding in a single hydrophilic sheet in (1) (top) and in L-Tyr-L-Phe·H_2O (Murali & Subramanian, 1987, bottom). The L-Tyr OH group is shown together with the three closest aromatic C atoms only.

($\varphi_2 = -79.4°, \theta = -92.8°$) are remarkably similar to the pattern in (1), Fig. 5, the only difference being the substitution of water molecule *C* in (1) with the phenolic —OH group of the L-Tyr side chain.

The crystal structure of (2) is best visualized along the *ac* diagonal as in Fig. 6. The aromatic rings together with four of the L-Val side chains form separate hydrophobic regions in the crystal. The most surprising aspect of the structure is the way the remaining four L-Val side chains belonging to molecules *A*, *C*, *E* and *G* are exposed to the solvent water region of the crystal. The side chain of molecule *G*, shown in Fig. 7, is completely surrounded by water molecules, amino groups and carboxylate groups. Even if static, this structure gives a unique, detailed picture of cage formation around hydrophobic groups in aqueous solution. The first hydration shell is constructed from ring systems including four to seven molecules. The intermolecular distances from the methyl C atom to O atoms are normally in the range 3.9 ± 0.2 Å, but reach as low as 3.53 and 3.44 Å for the methyl groups in molecule *G*. The associated H···O distances and C—H···O angles are 2.60 Å/158° and 2.78 Å/124°, respectively.

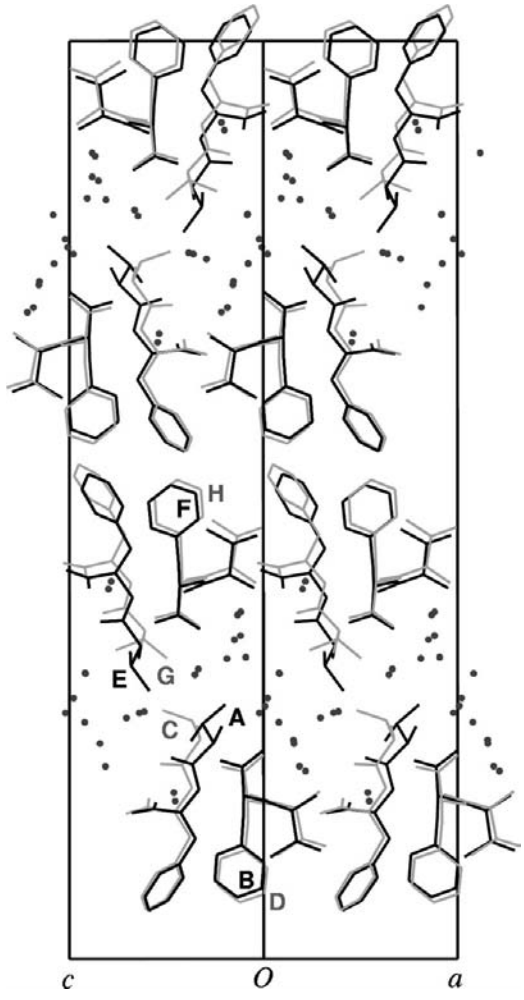


Figure 6
The unit cell and crystal packing of (2) viewed along the *ac* diagonal. H atoms have been omitted for clarity and water molecules appear as small spheres. Molecules *A*, *B*, *E* and *F* are drawn in black, while molecules *C*, *D*, *G* and *H* are drawn in a gray tone.

Table 4
Average values and sample standard deviations for hydrogen bonds in (2) (Å, °).

Hydrogen-bond type	*N*	H···*A*	*D*···*A*	*D*—H···*A*
$-NH_3^+\cdots{}^-OOC-$	8	1.839, 0.045	2.713, 0.037	161.8, 8.3
$-NH_3^+\cdots OH_2$	17	2.019, 0.109	2.854, 0.060	155.1, 12.7†
>N—H···$^-$OOC—	4	2.005, 0.044	2.872, 0.045	168.8, 3.0
>N—H···O=C<	4	1.970, 0.053	2.841, 0.050	170.5, 2.4
H—O—H···$^-$OOC—	24	1.991, 0.105	2.798, 0.086	159.7, 9.4
H—O—H···O=C<	2	1.970, –	2.756, –	157.0, –
H—O—H···OH_2	22	2.005, 0.123	2.812, 0.098	158.6, 10.3

† 158.9, 6.7 for 15 two-center hydrogen bonds.

Statistics for 81 hydrogen bonds in (2), which can be classified into seven different groups, are given in Table 4. Average values for hydrogen-bond angles are around 160°, except those involving the peptide bond >N—H as the donor, which are more linear. The average N···O length of the $-NH_3^+\cdots{}^-OOC-$ interaction, 2.713 Å with estimated standard error 0.013 Å, is significantly shorter than the 2.840 Å previously reported in a survey of hydrogen bonds in peptide structures (Görbitz, 1989). A new CSD search (Allen & Kennard, 1993) confirmed that the 2.649 (5) Å N···O distance for the N1*E*···O3*H* contact is the shortest ever recorded for this type of hydrogen bond. The shortness of this particular hydrogen bond is not easy to explain, as it is not particularly linear (angle 165°, constrained refinement), and the carboxylate group of peptide molecule *H* accepts no less than four other H atoms. It is, however, interesting to notice that out of the other two amino H atoms of N1*E* one is involved in a three-center interaction (the only occurrence in the structure) with rather long H···O distances, while the second is accepted by a water molecule with low occupancy (0.143) and thus is not used in hydrogen bonding for most molecules. Both these observations should contribute to making the last amino H atom a very strong donor.

All peptide amino groups donate two H atoms to water molecules and one to a peptide carboxylate group. A special motif

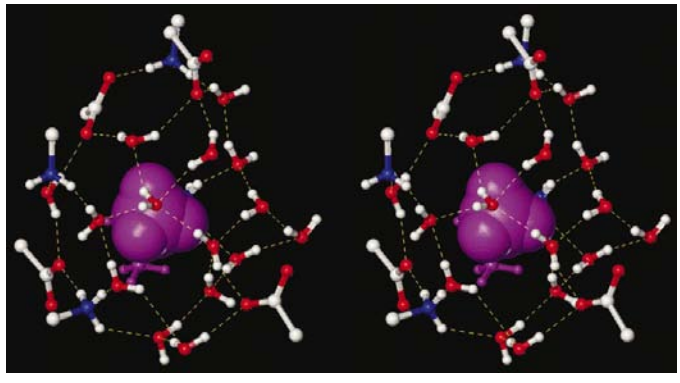
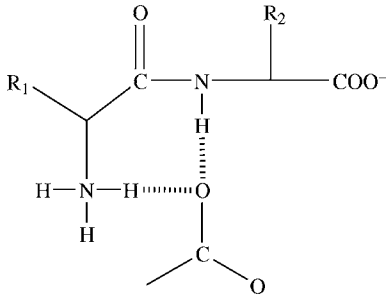


Figure 7
Stereo view of the cage formation around one of the L-Val methyl groups (depicted in violet space fill) of molecule *G* in (2).



is formed for peptides in the c_{2B} family. It may at first seem surprising that this simple pattern with second-level graph set $R_2^1(7)$ (Etter *et al.*, 1990; Bernstein *et al.*, 1995) has not previously been found in peptide structures, but this is a direct result of the molecular conformations, more specifically the values for N1—C1—C5—N2 (ψ_1) in the range −49.1 (5) to −62.7 (5)°. A search for structures of *N*-free peptides in the CSD showed that with one exception (ψ_1 = 8.7°; Flippen-Anderson *et al.*, 1994) all have $|\psi_1| > 101.6°$.

Fig. 3(*b*) shows how the typical hydrogen bond between residue 1 and residue 4 in the Type I β turn is replaced by a water molecule for peptide molecule *B* in (2). There is no such direct link between the *N* and *C* terminal groups in (1); the charged groups are connected by two water molecules and a carboxylate group.

3.3. Water structure

Four of the 24 water molecules in (2) with high occupancy have no hydrogen bonds to other water molecules and act as connectors between one amino and two carboxylate groups. Other water molecules are located in channels running through the structure in the direction of the *ac* diagonal, as seen in Fig. 6. They are involved in one (four molecules), two (seven), three (five) or four (two) hydrogen bonds to other water molecules. All water molecules are donors in two hydrogen bonds and they accept one (ten molecules) or two (14) H atoms. The most common acceptors in hydrogen bonds with water molecule donors are carboxylate groups (24 interactions) followed by other water molecules (22). The geometry differences for these two types of interactions are very small, Table 4. The peptide amino groups donate a total of 17 H atoms to water molecules. These hydrogen bonds are on average slightly longer than the two other types involving water molecules. There are no interactions between the peptide bond >N—H group and water molecules.

The water structure includes three antidromic ring systems (one double acceptor and one double donor; Saenger, 1979), one with four water molecules, one with five and one with seven. The latter two can be seen in Fig. 7.

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.

Ashida, T., Tsunogae, Y., Tanaka, I. & Yamane, T. (1987). *Acta Cryst.* B**43**, 212–218.

Benedetti, E., Morelli, G., Némethy, G. & Scheraga, H. A. (1983). *Int. J. Pept. Protein Res.* **22**, 1–15.

Bernstein, J., Davis, R. E., Shimoni, L. & Chang, N. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 1555–1573.

Bruker (1998*a*). *SAINT*. Version 6.01. Bruker AXS Inc., Madison, Wisconsin, USA.

Bruker (1998*b*). *SMART*, Version 5.054. Bruker AXS Inc., Madison, Wisconsin, USA.

Bruker (2000). *XM. Macromolecular Direct Methods*, Version 6.09. Bruker AXS Inc., Madison, Wisconsin, USA.

Cotrait, M., Bideau, J. P., Beurskens, G., Bosman, W. P. & Beurskens, P. T. (1984). *Acta Cryst.* C**40**, 1412–1416.

Dalhus, B. & Görbitz, C. H. (1999). *Acta Cryst.* B**55**, 424–431.

Etter, M. C., MacDonald, J. C. & Bernstein, J. (1990). *Acta Cryst.* B**46**, 256–262.

Flippen-Anderson, J. L., Hruby, V. J., Collins, N., George, C. & Cudney, B. (1994). *J. Am. Chem. Soc.* **116**, 7523–7531.

Görbitz, C. H. (1989). *Acta Cryst.* B**45**, 390–395.

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**, IUC9900149.

Görbitz, C. H. (1999*c*). *Acta Cryst.* C**55**, 2171–2177.

Görbitz, C. H. (1999*d*). *Acta Cryst.* B**55**, 1090–1098.

Görbitz, C. H. (2000). *Acta Cryst.* C**56**, 1496–1498.

Görbitz, C. H. (2001*a*). *Chem. Eur. J.* **7**, 5153–5159.

Görbitz, C. H. (2001*b*). *Acta Cryst.* C**57**, 575–576.

Görbitz, C. H. (2002). Submitted for publication.

Görbitz, C. H. & Etter, M. C. (1992). *Int. J. Pept. Protein Res.* **39**, 93–110.

Görbitz, C. H. & Gundersen, E. (1996*a*). *Acta Cryst.* C**52**, 1764–1767.

Görbitz, C. H. & Gundersen, E. (1996*b*). *Acta Chem. Scand.* **50**, 537–543.

Marsh, R. E., Ramakumar, S. & Venkatesan, K. (1976). *Acta Cryst.* B**32**, 66–70.

Morita, H., Kayashita, T., Shishido, A., Takeya, K., Itokawa, H. & Shiro, M. (1996). *Tetrahedron*, **52**, 1165–1177.

Murali, R., Lalitha, V., Subramanian, E. & Parthasarathy, R. (1986). *Int. J. Pept. Protein Res.* **27**, 160–164.

Murali, R. & Subramanian, E. (1987). *Int. J. Pept. Protein Res.* **29**, 187–192.

Pandit, J., Seshadri, T. P. & Viswamitra, M. A. (1983). *Acta Cryst.* C**39**, 1669–1672.

Ramakrishnan, B., Seshadri, T. P. & Viswamitra, M. A. (1984). *Acta Cryst.* C**40**, 1248–1250.

Ramakrishnan, B. & Viswamitra, M. A. (1988). *Acta Cryst.* C**44**, 1959–1961.

Saenger, W. (1979). *Nature*, **279**, 343–344.

Sheldrick, G. M. (1996). *SADABS*. University of Göttingen, Germany.

Sheldrick, G. M. (1997). *SHELXTL*, Version 5.10. Bruker AXS Inc., Madison, Wisconsin, USA.

Steiner, T. (2000) *Acta Cryst.* B**56**, 673–676.

Stenkamp, R. E. & Jensen, L. H. (1974). *Acta Cryst.* B**30**, 1541–1545.

Stenkamp, R. E. & Jensen, L. H. (1975). *Acta Cryst.* B**31**, 857–861.

Suresh, C. G. & Vijayan, M. (1985). *Int. J. Pept. Protein Res.* **26**, 311–328.

Urpi, L., Collect, M. & Subirana, J. A. (1988). *Int. J. Biol. Macromol.* **10**, 55–59.