

N-(*tert*-Butoxycarbonyl)-*L*-valyl-*L*-valine methyl ester: a twisted parallel β -sheet in the crystal structure of a protected dipeptide

Øyvind Jacobsen,^a Hadgu Girmay Gebreslasie,^b Jo Klaveness,^a Pål Rongved^a and Carl Henrik Görbitz^{b*}

^aSchool of Pharmacy, University of Oslo, PO Box 1068 Blindern, N-0316 Oslo, Norway, and ^bDepartment of Chemistry, University of Oslo, PO Box 1033 Blindern, N-0315 Oslo, Norway

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

Received 18 May 2011

Accepted 8 June 2011

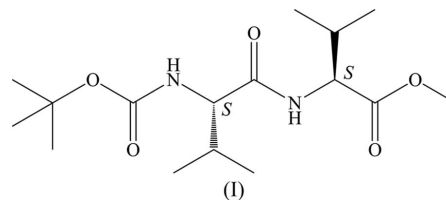
Online 5 July 2011

The title compound, C₁₆H₃₀N₂O₅, crystallizes with three molecules in the asymmetric unit, each adopting a β -strand/polyproline II backbone conformation. The main-chain functional groups are hydrogen bonded into tapes having the characteristics of parallel β -sheets. Each tape has a left-handed twist and thus forms a helix, with six peptide molecules needed to complete a full 360° rotation. A comparison of hydrogen-bond lengths and twisting modes is made with other related structures of protected dipeptides and with a hexapeptide derived from amyloid- β containing the Val–Val segment. Additionally, a comparison of the backbone conformation is made with that of the Val141–Val142 segment of the water channel aquaporin-4 (AQP4).

Comment

The water channel aquaporin-4 (AQP4) is a tetrameric integral membrane protein selectively transporting water across the cell membrane, and constitutes an important clinical target for the treatment of brain oedema (Manley *et al.*, 2000; Amiry-Moghaddam & Ottersen, 2003). In addition to its primary function as a passive water transporter, it has also been postulated that AQP4 may play a role in cell–cell adhesion (Hiroaki *et al.*, 2006). The structure of the M23 isoform of rat AQP4 (rAQP4) was recently determined to 3.2 Å resolution by electron diffraction (ED) on two-dimensional crystals [Protein Data Bank (PDB) code 2D57; Hiroaki *et al.*, 2006]. This has subsequently been improved to 2.8 Å resolution for an ED structure of an S180D mutant of rAQP4M23 (PDB code 2ZZ9; Tani *et al.*, 2009) and 1.8 Å for an X-ray structure of a truncated version of human AQP4 (PDB code 3GD8; Ho *et al.*, 2009). The ED structures revealed that adhesive interactions between rAQP4 tetramers in contiguous membranes are mediated by Pro139 and Val142 in the extracellular loop C.

The Ser140–Gly144 segment of loop C forms a short 3₁₀-helix. We believe that compounds mimicking the primary structure and conformation of a loop C segment containing at least one of the residues mediating adhesion can potentially have affinity for AQP4 and serve as lead compounds for the development of selective AQP4 ligands and, eventually, AQP4 inhibitors. As part of our ongoing synthetic and structural studies of mimics of loop C segments (Jacobsen *et al.*, 2009), the title compound, (I), corresponding to the Val141–Val142 segment of loop C, was synthesized and its X-ray crystal structure determined.



(I) crystallizes in the orthorhombic space group $P2_12_12_1$ with three molecules in the asymmetric unit (Fig. 1). The molecular conformations are roughly similar and, in particular, the r.m.s. value for the best overlay between molecules *A* and *C* is only 0.30 Å. Molecule *B* has a different orientation for the side chain of the C-terminal valine residue, Val2, as defined by the N2–C11–C12–C13 and N2–C11–C12–C14 torsion angles in Table 1, which together with a slightly larger main-chain C5–N1–C6–C10 torsion angle contribute to the slightly higher r.m.s. values of 0.79 and 0.69 Å for the best overlays with molecules *A* and *C*, respectively. The average values for the C5–N1–C6–C10 and N1–C6–C10–N2 (φ , ψ) torsion angles of Val1 are –94 and 130°, respectively, which are quite close to the ideal values for a parallel β -sheet (Loughlin *et al.*, 2010). The Val2 residues adopt polyproline II conformations, with the average dihedral angles being –59 and 145°, respectively. Polyproline II is one of the dominant local backbone conformations in unfolded peptides and proteins (Makowska *et al.*, 2006; Shi *et al.*, 2006). The overall backbone conformation seen for all molecules in Fig. 1 is dominant among 29 peptide molecules in 18 other structures of acyclic L,L-dipeptides in the Cambridge Structural Database (CSD, Version 5.32 of November 2010; Allen, 2002) that are *N*-terminally protected as a carbamate and *C*-terminally protected as an ester (see *Supplementary material*). In particular, ten out of 12 molecules devoid of aromatic residues in the protecting groups and the side chains take on roughly the same folding as observed for (I). In contrast to the dipeptide (I), the dihedral angles of Val141 and Val142 in loop C of rAQP4 are –63/–26 and –26/–39°, respectively, placing the Val141–Val142 segment in the 3₁₀- or mixed 3₁₀-/ α -helical region of the Ramachandran plot. However, the difference in backbone structure between (I) and the Val141–Val142 segment of rAQP4 loop C could be explained by the fact that (I) is too short to form the stabilizing intramolecular *i* → *i*+3 hydrogen bonds (to a peptide carbonyl acceptor three residues down the chain) which are characteristic of 3₁₀-helices (the subscript 10 referring to the size of the hydrogen-bonded ring).

The molecular packing of (I) is shown in Fig. 2. The molecules are hydrogen bonded along the *c* axis in a manner similar to a parallel β -sheet, thereby forming ribbons with a left-handed twist. A structural analogy in proteins is provided by the β -sheet of a Rossmann fold (Rao & Rossmann, 1973), which is frequently found in nucleotide-binding proteins, *e.g.* in the NADP(H)-binding proton pump transhydrogenase (Jeeves *et al.*, 2000). The Rossmann fold is a supersecondary structural motif which consists of a β -sheet comprising three or more parallel β -strands that are connected by α -helices, β -strands or unfolded loops. The topological structure of the Rossmann fold is β - α - β - α - β . The structures of NAD(P)-

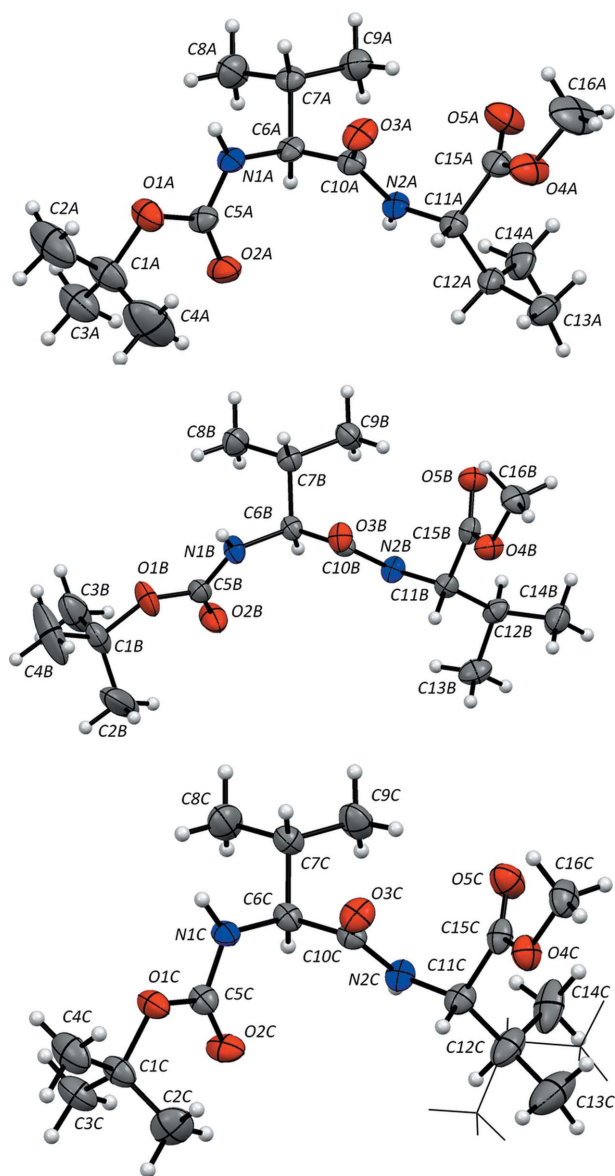


Figure 1

The structure of (I), showing the atom-numbering scheme for each molecule in the asymmetric unit (*A*, *B* and *C*). For comparison, the three molecules are shown in approximately the same orientation, which is not representative of their relative positions and orientations in the asymmetric unit. Displacement ellipsoids are shown at the 50% probability level and H atoms are shown as spheres of arbitrary size. The minor orientation of the second Val residue of molecule *C* [occupancy 0.157 (7)] is shown in wireframe representation.

dependent dehydrogenases feature two Rossmann folds with left-handed twisting of the β -sheets in their cofactor-binding domains (Kutzenko *et al.*, 1998).

Extracellular deposition of insoluble fibrous β -sheet protein aggregates (amyloid fibrils) is diagnostic of several debilitating and/or fatal diseases, *e.g.* Alzheimer's disease (AD). According to the amyloid hypothesis, formation of amyloid fibrils from amyloid- β ($A\beta$), a 39- to 42-residues long polypeptide formed by proteolytic cleavage of amyloid precursor protein (APP), also plays an important part in the pathogenesis of AD (Selkoe, 1991; Hardy & Allsop, 1991). Recent evidence, however, strongly suggests that the actual synaptotoxic species are soluble $A\beta$ dimers and oligomers (Jin *et al.*, 2011).

The hexapeptide Gly-Gly-Val-Val-Ile-Ala, (II), which constitutes a short segment of $A\beta$, forms both fibrils and microcrystals *in vitro*. The structures of the fibrillar and microcrystalline forms of (II) are believed to be closely related. X-ray crystallographic studies of microcrystals of (II) have revealed a cross- β spine structure, a quaternary structural feature characteristic of amyloid fibrils, consisting of two parallel β -sheets (PDB code 2ONV; Sawaya *et al.*, 2007). In a cross- β spine, side chains on the two β -sheets interdigitate to form a tight dry steric-zipper interface (Nelson *et al.*, 2005; Sawaya *et al.*, 2007). Importantly, residues 1 and 2 of (I) are identical to residues 3 and 4 of (II), suggesting that (I) could potentially serve as a minimal model compound for studies of the formation and properties of β -sheet structures associated with AD. Recent Monte Carlo simulations using a coarse-grained united-atom model of (II) have shown that (II) spontaneously forms three types of aggregation structures, including, like (I), a single-layer left-handed twisted β -sheet

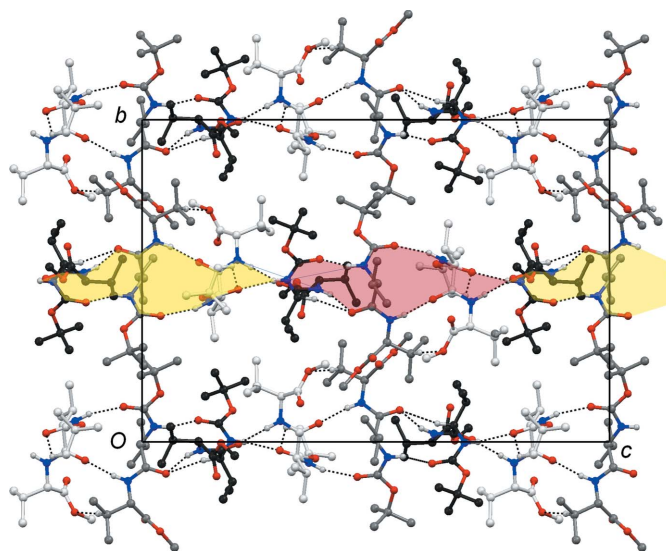
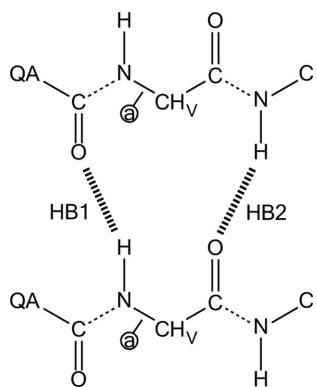


Figure 2

The molecular packing of (I) viewed along the *a* axis. Dashed lines indicate hydrogen bonds. H atoms not involved in hydrogen bonding have been omitted. In the electronic version of the paper, C atoms of molecules *A*, *B* and *C* have been coloured in black, grey and white, respectively. A hydrogen-bonded helix has been highlighted by the twisted tape (yellow and red).

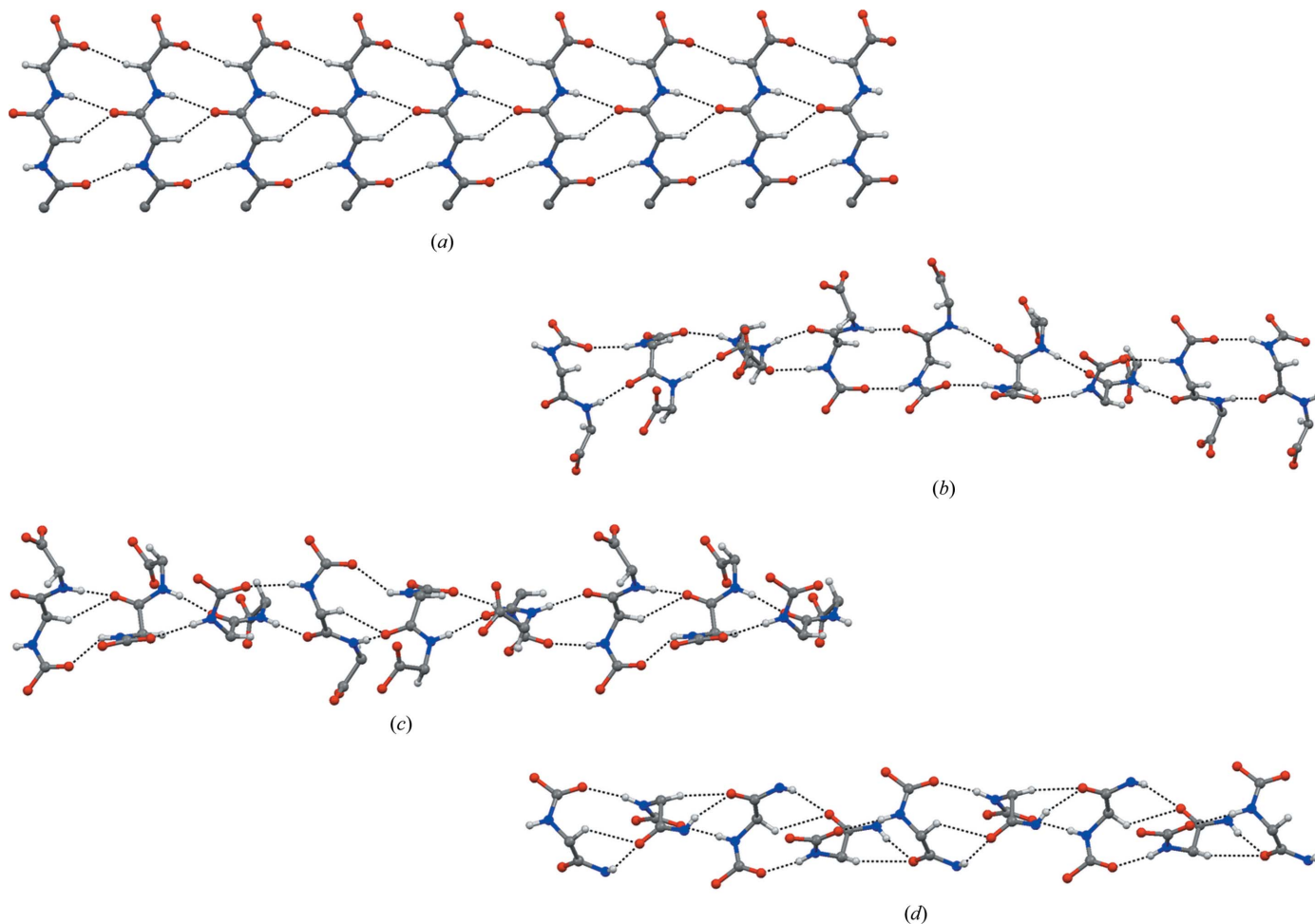

Figure 3

The search fragment used to find structures of acyclic peptides with parallel β -sheets in the CSD (Allen, 2002). The dashed C—N bonds have bond types ‘any’, the central N—C bonds have been defined as acyclic, the number of H atoms on the central C $^{\alpha}$ atom can be 1 or 2 (V = variable), and the atom type for QA atoms is either C or O. The two hydrogen bonds HB1 and HB2 were defined as intermolecular contacts with H \cdots O distances in the range 1.7–2.8 Å after normalization of the N—H bond lengths to 0.88 Å. The molecules connected by N—H \cdots O hydrogen bonds may additionally be connected by C—H \cdots O hydrogen bonds, but these were not specified in the search.

and a fibril-like cross- β spine structure comprising left-handed twisted parallel β -sheets with an average twisting angle of $12\pm 2^\circ$ (Mu & Gao, 2009). For comparison, the observed average twisting angles from one β -strand to the next in the crystal structures of (I) and (II) are 60 and 0° , respectively, the latter value corresponding to a flat sheet (Sawaya *et al.*, 2007).

To get an overview of small-molecule structures containing parallel β -sheets, the CSD was searched for peptides (not just dipeptides) containing the search fragment depicted in Fig. 3. The query returned a total of 92 CSD entries, and subsequent manual scrutiny revealed infinite tapes in 69 peptide structures including 48 regular dipeptides, two structures where two dipeptides are linked by a disulfide bridge, 11 tripeptides, six tetrapeptides, one pentapeptide and one folded decapeptide (see *Supplementary material*). Most tapes are planar and straight (67 observations in 49 structures), as seen in Fig. 4(a), but some structures have tapes that are twisted into helices, as seen for (I) in Fig. 2.

This second group of 20 peptides may be subdivided into smaller groups based on the number of molecules required to complete a full turn of the helix. The only example with eight


Figure 4

(a) A straight parallel β -sheet (CSD refcode CEPQOE; Gerhardt & Weck, 2006). Side chains are not shown. Hydrogen bonds with N—H and C—H donors are shown as dashed lines. (b) A twisted parallel β -sheet with eight peptide molecules for one complete turn (PIYSAS; Oku *et al.*, 2008). (c) The twisted parallel β -sheet of (I) with six peptide molecules for one complete turn. (d) A twisted parallel β -sheet with four peptide molecules for one complete turn (FABLUP10; Varughese *et al.*, 1986).

peptide molecules for a full turn is shown in Fig. 4(b). There is also just a single example of a structure with a repeating unit of seven molecules (with $Z' = 7$). The structure of (I) (Fig. 4c) adds to a larger group of 11 other structures with a repeating unit of six peptide molecules. Eight of these belong to the hexagonal space group $P6_5$ with $Z' = 1$, one, with a rare 2-methylvalyl residue, belongs to $P6_1$, one to the trigonal space group $P3_2$ with $Z' = 2$, while the last structure, like (I), is orthorhombic, $P2_12_12_1$, with $Z' = 3$. The last group, with a repeating unit of four peptide molecules (Fig. 4d), comprises seven structures, of which five are tetragonal and two are monoclinic. It is interesting to find that when only C^α mono-substituted L-amino acids are involved (18 out of 20 structures), the twist is always left-handed. This is in accordance with the preferred sense of twist observed for β -sheets in proteins (Chothia, 1973; Chou *et al.*, 1983; note that our definition of handedness is the sense of rotation when going from one chain in the sheet to the next, while Chothia defines twisting as the observed rotation when viewing along a single polypeptide chain, which gives opposite results with respect to handedness). Another important observation is that the hydrogen bonds appear to get shorter when the twist increases. Thus, the average of the two hydrogen-bond lengths HB1 and HB2 in Fig. 3 is 2.140 Å for the flat tapes (or 2.095 Å when four clear outlier structures are removed from the statistics), but 2.087 and 2.057 Å for twisted tapes with a repeating unit of six and four molecules, respectively. Hydrogen-bond data for (I) are given in Table 2. After normalization of N—H bonds to 0.88 Å, the average length of the six N—H...O hydrogen bonds in the structure of (I) is 2.023 Å, which is short for this type of pattern.

Experimental

(I) was synthesized by standard solution-phase peptide coupling of L-valine methyl ester, which was generated *in situ* from L-valine methyl ester hydrochloride by deprotonation with *N,N*-diisopropylethylamine, with commercially available *N*-(*tert*-butoxycarbonyl)-L-valine. 3-[3-(Dimethylamino)propyl]-1-ethylcarbodiimide (EDC) was used as coupling reagent and 1-hydroxybenzotriazole (HOBt) as an additive to suppress epimerization (Jacobsen *et al.*, 2009). About 10 mg of (I) was dissolved in ethyl acetate (50 μ l). Needle-shaped crystals appeared as water diffused into the solution at room temperature.

Crystal data

$C_{16}H_{30}N_2O_5$	$V = 5897.3$ (10) Å ³
$M_r = 330.42$	$Z = 12$
Orthorhombic, $P2_12_12_1$	Mo $K\alpha$ radiation
$a = 11.6937$ (12) Å	$\mu = 0.08$ mm ⁻¹
$b = 18.6458$ (19) Å	$T = 150$ K
$c = 27.047$ (3) Å	$0.72 \times 0.15 \times 0.11$ mm

Data collection

Bruker APEXII CCD diffractometer	38750 measured reflections
Absorption correction: multi-scan (SADABS; Bruker, 2007)	5757 independent reflections
$T_{\min} = 0.900$, $T_{\max} = 0.980$	4918 reflections with $I > 2\sigma(I)$
	$R_{\text{int}} = 0.044$

Refinement

$R[F^2 > 2\sigma(F^2)] = 0.036$	H atoms treated by a mixture of independent and constrained refinement
$wR(F^2) = 0.089$	$\Delta\rho_{\text{max}} = 0.19$ e Å ⁻³
$S = 1.03$	$\Delta\rho_{\text{min}} = -0.18$ e Å ⁻³
5757 reflections	
651 parameters	
12 restraints	

Table 1

Selected torsion angles (°).

C1A—O1A—C5A—N1A	-178.4 (3)	N2B—C11B—C15B—O4B	141.6 (2)
O1A—C5A—N1A—C6A	-177.4 (2)	C11B—C15B—O4B—C16B	176.6 (2)
C5A—N1A—C6A—C10A	-87.6 (3)	N1B—C6B—C7B—C8B	-56.3 (3)
N1A—C6A—C10A—N2A	126.2 (2)	N1B—C6B—C7B—C9B	-179.5 (2)
C6A—C10A—N2A—C11A	-176.4 (2)	N2B—C11B—C12B—C13B	-67.2 (3)
C10A—N2A—C11A—C15A	-62.6 (3)	N2B—C11B—C12B—C14B	170.2 (2)
N2A—C11A—C15A—O4A	149.1 (2)	C1C—O1C—C5C—N1C	177.0 (2)
C11A—C15A—O4A—C16A	178.2 (3)	O1C—C5C—N1C—C6C	168.0 (2)
N1A—C6A—C7A—C8A	-64.3 (3)	C5C—N1C—C6C—C10C	-76.2 (3)
N1A—C6A—C7A—C9A	172.9 (2)	N1C—C6C—C10C—N2C	129.5 (2)
N2A—C11A—C12A—C13A	-168.9 (2)	C6C—C10C—N2C—C11C	-176.4 (2)
N2A—C11A—C12A—C14A	65.2 (3)	C10C—N2C—C11C—C15C	-58.3 (3)
C1B—O1B—C5B—N1B	174.5 (2)	N2C—C11C—C15C—O4C	145.5 (2)
O1B—C5B—N1B—C6B	-174.1 (2)	C11C—C15C—O4C—C16C	-178.0 (2)
C5B—N1B—C6B—C10B	-117.6 (2)	N1C—C6C—C7C—C8C	-60.2 (3)
N1B—C6B—C10B—N2B	134.3 (2)	N1C—C6C—C7C—C9C	176.3 (2)
C6B—C10B—N2B—C11B	177.5 (2)	N2C—C11C—C12C—C13C	-157.8 (4)
C10B—N2B—C11B—C15B	-57.0 (3)	N2C—C11C—C12C—C14C	73.9 (5)

Table 2

Hydrogen-bond geometry (Å, °).

$D-H\cdots A$	$D-H$	$H\cdots A$	$D\cdots A$	$D-H\cdots A$
N1A—H1A...O2C ⁱ	0.85 (2)	2.08 (2)	2.920 (3)	169 (3)
N2A—H2A...O3B	0.87 (2)	2.03 (2)	2.874 (3)	165 (3)
C6A—H61A...O3B	1.00	2.45	3.295 (3)	142
N1B—H1B...O2A	0.85 (2)	2.14 (2)	2.948 (3)	157 (3)
N2B—H2B...O3C	0.86 (2)	1.97 (2)	2.810 (3)	170 (3)
C6B—H61B...O3C	1.00	2.52	3.367 (3)	142
N1C—H1C...O2B	0.87 (2)	2.12 (2)	2.966 (3)	164 (3)
N2C—H2C...O3A ⁱⁱ	0.86 (2)	1.92 (2)	2.782 (3)	177 (3)
C6C—H61C...O3A ⁱⁱ	1.00	2.75	3.559 (3)	138

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

The side chain of residue 2 in molecule *C* is disordered over two positions, and atoms of the major orientation [occupancy 0.843 (7)] were refined in a normal manner. The covalent geometry of the minor orientation [occupancy 0.157 (7)] was loosely tied to the geometry of the major component by a *SHELX SAME* restraint (Sheldrick, 2008). The two positions defined for C^α (C11C and C11D) were constrained to occupy the same site. C^α (C11D) and C^β (C12D) received the same set of anisotropic displacement parameters as their major counterparts (C11C and C12C), while a common isotropic displacement parameter was refined for $C^{\gamma 1}$ (C13D) and $C^{\gamma 2}$ (C14D). Positional parameters were refined for H atoms bonded to N atoms, with a mild restraint of 0.88 (2) Å applied to the N—H distances. Other H atoms were positioned with idealized geometry and fixed C—H distances of 0.98 and 1.00 Å for CH₃ and CH groups, respectively. Free rotation was permitted for methyl groups. $U_{\text{iso}}(\text{H})$ values were set at $1.2U_{\text{eq}}$ of the carrier atom or at $1.5U_{\text{eq}}$ for methyl groups. In the absence of significant anomalous scattering effects, 4636 Friedel pairs were merged.

Data collection: *APEX2* (Bruker, 2007); cell refinement: *SAINTE-Plus* (Bruker, 2007); data reduction: *SAINTE-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: JZ3205). Services for accessing these data are described at the back of the journal.

References

- Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
- Amiry-Moghaddam, M. & Ottersen, O. P. (2003). *Nat. Rev. Neurosci.* **4**, 991–1001.
- Bruker (2007). *APEX2, SAINT-Plus and SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
- Chothia, C. (1973). *J. Mol. Biol.* **75**, 295–302.
- Chou, K.-C., Nemethy, G. & Scheraga, H. A. (1983). *Biochemistry*, **22**, 6213–6221.
- Gerhardt, W. W. & Weck, M. (2006). *J. Org. Chem.* **71**, 6333–6341.
- Hardy, J. & Allsop, D. (1991). *Trends Pharmacol. Sci.* **12**, 383–388.
- Hiroaki, Y., Tani, K., Kamegawa, A., Gyobu, N., Nishikawa, K., Suzuki, H., Walz, T., Sasaki, S., Mitsuoka, K., Kimura, K., Mizoguchi, A. & Fujiyoshi, Y. (2006). *J. Mol. Biol.* **355**, 628–639.
- Ho, J. D., Yeh, R., Sandstrom, A., Chorny, I., Harries, W. E. C., Robbins, R. A., Miercke, L. J. W. & Stroud, R. M. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 7437–7442.
- Jacobsen, Ø., Klaveness, J., Ottersen, O., Amiry-Moghaddam, M. & Rongved, P. (2009). *Org. Biomol. Chem.* **7**, 1599–1611.
- Jeeves, M., Smith, K. J., Quirk, P. G., Cotton, N. P. J. & Jackson, J. B. (2000). *Biochim. Biophys. Acta*, **1459**, 248–257.
- Jin, M., Shepardson, N., Yang, T., Chen, G., Walsh, D. & Selkoe, D. J. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 5819–5824.
- Kutzenko, A. S., Lamzin, V. S. & Popov, V. O. (1998). *FEBS Lett.* **423**, 105–109.
- Loughlin, W. A., Tyndall, J. D. A., Glenn, M. P., Hill, T. A. & Fairlie, D. P. (2010). *Chem. Rev.* **110**, PR32–PR69.
- Manley, G. T., Fujimura, M., Ma, T., Noshita, N., Filiz, F., Bollen, A. W., Chan, P. & Verkman, A. S. (2000). *Nat. Med.* **6**, 159–163.
- Makowska, J., Rodziewicz-Motowidlo, S., Baginska, K., Vila, J. A., Liwo, A., Chmurzynski, L. & Scheraga, H. A. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 1744–1749.
- Mu, Y. & Gao, Y. I. (2009). *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **80**, 041927.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Riekel, C., Grothe, R. & Eisenberg, D. (2005). *Nature (London)*, **435**, 773–778.
- Oku, H., Yamada, K. & Katakai, R. (2008). *Biopolymers*, **89**, 270–283.
- Rao, S. & Rossmann, M. (1973). *J. Mol. Biol.* **76**, 241–256.
- Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, W., McFarlane, H. T., Madsen, A. Ø., Riekel, C. & Eisenberg, D. (2007). *Nature (London)*, **447**, 453–457.
- Selkoe, D. J. (1991). *Neuron*, **6**, 487–498.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Shi, Z., Chen, K., Liu, Z. & Kallenbach, N. R. (2006). *Chem. Rev.* **106**, 1877–1897.
- Tani, K., Mitsuma, T., Hiroaki, Y., Kamegawa, A., Nishikawa, K., Tanimura, Y. & Fujiyoshi, Y. (2009). *J. Mol. Biol.* **389**, 694–706.
- Varughese, K. A., Angus, R. H., Carey, P. R., Lee, H. & Storer, A. C. (1986). *Can. J. Chem.* **64**, 1668–1673.