The Crystal Structure of Glycyl-L-threonine Dihydrate

By V. S. Yadava and V. M. Padmanabhan

Nuclear Physics Division, Bhabha Atomic Research Centre, Trombay, Bombay-85, India

The crystal structure of glycyl-L-threonine dihydrate, $C_6N_2O_6H_{16}$, has been determined from three-dimensional X-ray diffraction data. The peptide glycyl-L-threonine crystallizes in the orthorhombic space group $P2_12_12_1$ with two molecules of water per asymmetric unit. Cell dimensions are $a = 9.72$ (2), $b = 10.03$ (2), $c = 10.73$ (2) Å. The structure was solved using the symbolic addition method and refined by full-matrix least-squares calculations to an R index of 0.094 for 830 observed reflexions. Both the carboxyl and the peptide groups are planar; the dihedral angle between the planes is $56.2°$. The dimensions of the peptide group are in agreement with the respective weighted mean values proposed by Marsh & Donohue [*Advanc. Protein Chem.* (1968). **22**, 235–256]. The peptide chain is in the extended conformation. The conformation of the threonine side group in the dipeptide is different from that in free threonine.

Introduction

Detailed investigations of the crystal structures of amino acids and simple peptides are carried out mainly because of their importance in arriving at the conformation of the polypeptide chain in proteins. In the crystal structure analysis of L-threonine (Shoemaker, Donohue Schomaker & Corey, 1950) the C^α–C^β distance in the side chain has been found to be abnormally short, 1.50 (1) Å. Also the bond angles at the β-carbon atom differ significantly from the expected tetrahedral angle of $109.5°$. The crystal structure analysis of glycyl-L-threonine dihydrate has been carried out to study further the geometry of the threonine molecule, and forms part of a programme of studies on crystal structure of peptides.

Experimental

The crystals were obtained by slow evaporation of an aqueous solution of glycyl-L-threonine at room temperature. The crystal data determined from Weissenberg photographs together with physical data are summarized in Table 1. The presence of two molecules of water of crystallization per asymmetric unit was inferred from the measured density and subsequently confirmed by chemical analysis of the crystals.

Three-dimensional intensity data were collected by the multiple-film equi-inclination Weissenberg technique with Cu $K\alpha$ radiation for six layers each about the a and c axes from approximately cylindrical crystals having diameters 0.28 and 0.42 mm respectively. The intensities of 830 independent reflexions were estimated visually with a standard scale. The unobserved reflexions were estimated as $I = \frac{1}{2}I_{min}$. Corrections for Lorentz, polarization and absorption factors were made in the usual way. Interlayer scaling was achieved using cross-layer data and the corrected intensities were put on an absolute scale by Wilson's method.

Structure determination

The structure was determined by the symbolic addition method (Karle & Karle, 1966). The three origin defining reflexions together with an enantiomorph defining reflexion and a reflexion with a symbol are listed in Table 2. During the symbolic addition procedure there were several indications that the symbol a had a value close to π ($184.7°$ in the final structure). An E synthesis with 350 reflexions ($E \geq 1.0$) revealed 12 atoms of the structure. The remaining two non-hydrogen atoms were located from a difference Fourier synthesis.

Table 1. *Crystal data*

Composition of the asymmetric unit	$C_6H_{12}N_2O_4 . 2H_2O$
Crystal system	Orthorhombic
Space group	$P2_12_12_1$
Lattice constants	$a = 9.72$ (2) Å
	$b = 10.03$ (2)
	$c = 10.73$ (2)
Molecules per unit cell	$Z = 4$
Density (measured)	$d_m = 1.38$ g.cm^{-3}
Density (calculated)	$d_c = 1.39$ g.cm^{-3}
Linear absorption coefficient (for Cu $K\alpha$ radiation)	$\mu = 10.9$ cm^{-1}
	$\lambda = 1.5418$ Å

Table 2. *Initial phase assignments*

h	k	l	E	Phase	
7	6	0	2.7	$\pi/2$	Origin
2	0	5	1.9	$\pi/2$	
9	3	0	2.3	$\pi/2$	
0	11	6	1.9	$\pi/2$	Enantiomorph
4	7	4	2.3	a	

Refinement

The structure was refined by the full-matrix least-squares method using the computer program *ORFLS* (Busing, Martin & Levy, 1962) which minimizes the

function $\sum w(|F_o| - K|F_c|)^2$. A few cycles of isotropic refinement with unit weights for the observed reflexions (zero weight for the unobserved ones) gave an R index of 0·15, but the thermal parameters of some of the atoms had assumed rather low values. This could be due to extinction. Therefore, an approximate correction for isotropic, secondary extinction was applied using Zachariasen's (1967) general theory. $\bar{T}$, the absorption

weighted mean path length in the crystal, was calculated for each reflexion from the following relation (Coppens & Hamilton, 1970):

$$\bar{T} \simeq -\log A/\mu \,,$$

A being the absorption correction factor ($\mu r < 0.5$). Six cycles of anisotropic refinement, with the extinction constant g' (Coppens & Hamilton, 1970) as a variable parameter and with a Cruickshank (1961) type of weighting scheme based on an error analysis plot, gave a final R of 0·094 for the 830 observed reflexions (unobserved reflexions not included in the final cycle of refinement).

Scattering factors for C, N and O were taken from *International Tables for X-ray Crystallography* (1962). The extinction constant, g', had the value 4·49, corresponding to 1·3 sec for the equivalent mosaic spread parameter, η. For the worst affected reflexion, 111, F_0^2/F_c^2 was 0·37. The observed and calculated structure amplitudes as obtained in the final cycle of refinement are given in Table 3. The fractional positional coordinates and the anisotropic thermal parameters are listed in Table 4. The intramolecular bond lengths and angles calculated with the computer program *ORFFE* (Busing, Martin & Levy, 1964) are shown in Fig. 1.

Table 3. *Observed and calculated structure factors*
The successive columns give h, k, l, 10 F_o and 10 F_c.

Discussion

Molecular packing and hydrogen bonds

The crystal structure projected down the *b* axis is

Fig. 1. Bond distances and angles. O(2)–C(4)–C(3) = 105·8°. C(5)–C(4)–C(3) = 113·8°.

shown in Fig. 2. The glycyl-L-threonine molecules run approximately parallel to the c axis and are linked by a three-dimensional network of hydrogen bonds through water molecules and by those between neighbouring peptide molecules. Table 5 lists various short contacts found in this structure which are acceptable as hydrogen bonds from geometrical considerations (Hamilton & Ibers, 1968). Except for the peptide group nitrogen N(2), all other proton donor atoms N(1), O(2), O(4) are also proton acceptors. These possible hydrogen bonds have been shown by dashed lines in Fig. 2.

Bond distances and bond angles

The intramolecular bond distances and angles are shown in Fig. 1. The average estimated standard deviations in bond lengths and angles are 0·013 Å and 0·9°, respectively. The C^{α}–C'[C(3)–C(6)] bond length of 1·511 Å and the C(1)–N(1) length of 1·480 Å are in good agreement with the corresponding weighted mean val-

ues of 1·526 (4) and 1·487 (4) Å respectively (Marsh & Donohue, 1967). In the side chain, the C^{α}–C^{β}[C(3)–C(4)] distance of 1·585 Å and C^{β}–O^{γ} [C(4)–O(2)] distance of 1·462 Å are significantly longer than the normal values 1·54 and 1·42 Å for these bonds. In the structure of L-threonine (Shoemaker *et al.*, 1950) C^{α}–C^{β} is abnormally short, 1·50 (1) Å. In both these structure it is observed that the bond angles at the β-carbon atom show considerable deviations from the expected tetrahedral angle of 109·5° (104·1, 113·4, 110·5° in L-threonine; 105·8, 113·8, 111·7° in this work). A similar situation exists in L-threonyl-L-phenylalanine-*p*-nitrobenzyl ester hydrobromide (Mallikarjunan, Rao, Venkatesan & Sarma, 1969) in which as in L-threonine (Shoemaker *et al.*, 1950), the C^{α}–C^{β} distance in the threonyl residue is abnormally short (1·46 Å) and the bond angles at the β-carbon atom show deviations from the tetrahedral value. These deviations are possibly due to steric hindrance in the intermolecular environment.

Table 4. *Fractional coordinates and thermal parameters with standard deviations in parentheses*

The thermal parameters are defined by $T = \exp\left[-(\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}l^2 + 2\beta_{12}hk + 2\beta_{13}hl + 2\beta_{23}kl)\right]$.
All values are multiplied by 10^4.

	x	y	z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
N(1)	7083 (9)	0466 (9)	6251 (9)	53 (8)	60 (8)	66 (9)	−7 (7)	−3 (9)	25 (8)
C(1)	6138 (8)	1182 (10)	5390 (11)	66 (8)	26 (8)	65 (9)	−8 (8)	−1 (9)	9 (8)
C(2)	6783 (9)	2515 (10)	4987(11)	45 (9)	48 (7)	31 (7)	2 (8)	−5 (9)	0 (6)
O(1)	8021 (6)	2765 (7)	5208 (9)	32 (6)	93 (9)	93 (11)	−7 (6)	−4 (7)	28 (8)
N(2)	5924 (8)	3323 (7)	4401 (9)	49 (8)	31 (6)	62 (9)	2 (6)	−17 (8)	11 (6)
C(3)	6304 (9)	4594 (10)	3869 (10)	27 (8)	58 (10)	44 (10)	−11 (7)	−1 (9)	−7 (8)
C(4)	6179 (10)	4539 (11)	2401 (10)	44 (9)	79 (10)	39 (9)	−4 (9)	5 (10)	13 (9)
O(2)	4734 (7)	4252 (8)	2130 (8)	58 (8)	89 (8)	67 (9)	4 (7)	−19 (8)	0 (7)
C(5)	7144 (13)	3528 (15)	1786 (15)	62 (12)	156 (19)	97 (16)	−8 (12)	14 (13)	−24 (16)
C(6)	5483 (10)	5757 (8)	4363 (10)	50 (9)	39 (8)	57 (10)	−4 (7)	−30 (10)	7 (7)
O(3)	5852 (8)	6885 (6)	4110 (8)	58 (8)	51 (7)	98 (10)	3 (6)	−4 (8)	11 (6)
O(4)	4433 (7)	5482 (7)	5071 (7)	45 (6)	69 (7)	68 (8)	2 (6)	10 (7)	5 (6)
OW(1)	5350 (10)	8967 (11)	7661 (10)	139 (15)	176 (14)	78 (11)	−74 (12)	34 (12)	−4 (10)
OW(2)	6643 (7)	2281 (7)	8275 (8)	72 (8)	68 (7)	71 (9)	−6 (7)	6 (8)	2 (7)

Table 5. *Hydrogen-bond distances and angles*

N(1)$\cdots$OW(1)$^{y-1}$	2·72 Å	C(1)–N(1)–OW(1)$^{y-1}$	103·3°
N(1)$\cdots$O(4^{iii})$^{y-1}$	2·85	C(1)–N(1)–O(4^{iii})$^{y-1}$	110·2
N(1)$\cdots$OW(2)	2·86	C(1)–N(1)–OW(2)	94·1
N(1)$\cdots$O(2^{iii})	3·12	C(1)–N(1)–O(2^{iii})	114·3
N(2)$\cdots$O(1^{iii})	3·05	C(3)–N(2)–O(1^{iii})	127·3
		C(2)–N(2)–O(1^{iii})	107·4
O(2)$\cdots$O(3^{iv})	2·78	C(4)–O(2)–O(3^{iv})	117·4
O(2)$\cdots$N(1^{iii})	3·12	C(4)–O(2)–N(1^{iii})	131·6
O(4)$\cdots$OW(2^{iv})	2·74	C(6)–O(4)–OW(2^{iv})	122·5
O(4)$\cdots$N(1^{iii})	2·85	C(6)–O(4)–N(1^{iii})	114·1
O(4)$\cdots$OW(1^{iv})	2·87	C(6)–O(4)–OW(1^{iv})	123·2
OW(1)$\cdots$N(1)	2·72		
OW(1)$\cdots$OW(2^{iv})	2·76		
OW(1)$\cdots$O(4^{iv})	2·87		
OW(2)$\cdots$O(3^{ii})	2·72		
OW(2)$\cdots$O(4^{iv})	2·74		
OW(2)$\cdots$OW(1^{iv})	2·76		
OW(2)$\cdots$N(1)	2·86		

Symmetry code

Superscript	Atom at	Superscript	Atom at
None	x, y, z;	iii	$\frac{1}{2}+x$, $\frac{1}{2}-y$, $-z$
ii	$\frac{1}{2}-x$, $-y$, $\frac{1}{2}+z$;	iv	$-x$, $\frac{1}{2}+y$, $\frac{1}{2}-z$

The atoms with superscript $y-1$ are in the adjacent cell

Peptide group

The atoms C(1), C(2), O(1), N(2) and C(3) constitute the peptide group. The dimensions of the peptide group are not significantly different from the respective weighted mean values proposed by Marsh & Donohue (1967). However, the C^α-C' [C(1)-C(2)] distance of 1·539 Å and the angle C(2)-N(2)-C(3) as 124·6° are slightly larger (about 2σ) than the respective weighted mean values (Marsh & Donohue, 1967). The details of the least-squares planes fitted to the peptide group and the amide group atoms C(1), C(2), O(1), N(2) and deviations from these planes are listed in Table 6. As invariably found in peptides (Marsh & Donohue, 1967), the amide group is planar (maximum out-of-plane deviation 0·002 Å).
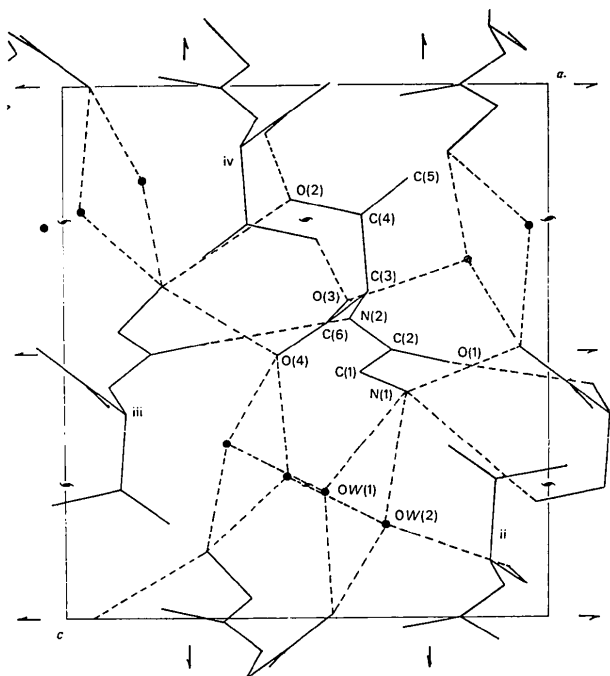

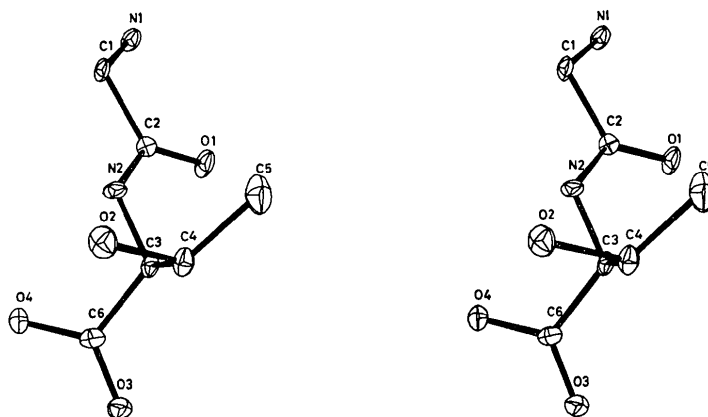Fig. 2. The crystal structure of glycyl-L-threonine projected down the *b* axis.

Table 6. *The deviations of the atoms from the least-squares planes* (Å)

The equation of least-squares plane is $ax + by + cz + d = 0$.

(i) Peptide group		
C(1)	0·018	a = 0·265
C(2)	−0·009	b = −0·424
O(1)	−0·002	c = −0·866
N(2)	−0·029	d = 3·948
C(3)	0·022	
(ii) Amide group		
C(1)	0·001	a = 0·260
C(2)	−0·002	b = −0·407
O(1)	0·001	c = −0·875
N(2)	0·001	d = 3·994
C(3)*	0·077	
(iii) Carboxyl group		
C(3)	0·003	a = 0·585
C(6)	0·009	b = 0·021
O(2)	0·004	c = 0·811
O(3)	0·003	d = −7·046
N(2)*	0·222	

* Not included in the calculation of the plane.

Carboxyl group

The carboxyl group in this structure is un-ionized [C-O(H) 1·304 Å, C-O 1·214 Å]. These dimensions are in close agreement with the mean values of 1·32 and 1·23 Å observed in monocarboxylic acids (Kanters, Kroon, Peerdeman & Schoone, 1967). The least-squares plane passing through the carboxyl group and deviations from this plane are listed in Table 6. The dihedral angle between the carboxyl and peptide planes is 56·2°. A stereo view (Johnson, 1965) of this dipeptide is shown in Fig. 3.

Molecular conformation

The notation followed for describing the conformation is the one suggested by the IUPAC-IUB Commission on Biochemical Nomenclature (1971). The rotations about N-C^α are denoted by φ, about N^α-C' by ψ, about C'-N by ω and about C^α-C^β by χ. The pertinent torsion angles are given in Table 7.

Fig. 3. Stereodiagram of the glycine-L-threonine molecule.

Table 7. *Torsion angles in glycyl-L-threonine*

Bond	atoms	designation	angle
C^α–C'	N(1)–C(1)–C(2)–N(2)	ψ_{11}	$-168\cdot7°$
	–O(1)	ψ_{12}	$11\cdot7°$
C'–N	C(1)–C(2)–N(2)–C(3)	ω	$-176\cdot2°$
N–C^α	C(2)–N(2)–C(3)–C(6)	φ_{21}	$-122\cdot9°$
	–C(4)	φ_{22}	$113\cdot3°$
C^α–C'	N(2)–C(3)–C(6)–O(3)	ψ_T^1	$169\cdot7°$
	–O(4)	ψ_T^2	$-8\cdot5°$
C^α–C^β	N(2)–C(3)–C(4)–C(5)	χ_{21}	$-62°$
	–O(2)	χ_{22}	$61°$

The backbone chain conformation around the α-carbon atom is given by the torsion angles, $\varphi_{21} = -122\cdot9°$, $\psi_T^1 = 169\cdot7°$. The pair of values lies within an allowed region of the (φ, ψ) conformational map of Ramachandran & Sasisekharan (1968) transposed into the standard convention. The torsion angle ω is $-176\cdot2°$. These conformational angles can be compared with the corresponding angles $(\varphi = \psi = \omega = 180°)$ for a fully extended conformation of the peptide chain.

The torsion angles ψ_T^1 and ψ_T^2 at the carboxyl terminus are in general found to be close to $180°$ and $0°$ ($360°$) with $\psi_T^1 = \psi_T^2 + 180°$. In the presence of a β-atom, however, ψ_T^2 is usually about $-10°$ (Lakshminarayanan, Sasisekharan & Ramachandran, 1967). The torsion angles at the C-terminus in this structure are $\psi_T^1 = 169\cdot7°$, $\psi_T^2 = -8\cdot5°$, close to the expected values.

The side chain conformation is described by the torsion angles around the bond C^α–C^β. These angles are close to 60, 180 and 300°, respectively, corresponding to the three staggered positions around the bond, and the γ-atom is said to go into positions I, II and III respectively (Lakshminarayanan *et al.* 1969). In this structure the C^γ atom [C(5)] is in the III position $(\chi_{21} = -62°)$ whereas in L-threonine (Shoemaker *et al.*, 1950) and L-threonyl-L-phenylalanine-*p*-nitrobenzyl ester hydrobromide (Mallikarjunan *et al.*, 1969), it is in the II position $(\chi_1 = 185$ and $176°$ respectively). The other γ-atom in the threonine side group is a hydroxyl oxygen atom. In this structure the O^γ [O(2)] atom goes into the I position $(\chi_{22} = 61°)$ whereas in L-threonine (Shoemaker *et al.*, 1950) and L-threonyl-L-phenylala-

nine-*p*-nitrobenzyl ester hydrobromide (Mallikarjunan *et al.*, 1969), the O^γ atom is in the III position $(\chi_2 = 305$ and $298°$ respectively).

References

BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. Report ORNL-TM-305. Oak Ridge National Laboratory, Oak Ridge, Tennessee. The CDC-3600 version incorporates modifications by C. K. JOHNSON & S. SRIKANTA. The extinction correction has been incorporated by S. K. SIKKA.

BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1964). *ORFFE*. Report ORNL-TM-306. Oak Ridge National Laboratory, Oak Ridge, Tennessee. The CDC-3600 version incorporates modifications by C. K. JOHNSON & S. SRIKANTA.

COPPENS, P. & HAMILTON, W. C. (1970). *Acta Cryst.* A26, 71–83.

CRUICKSHANK, D. W. J. (1961). *Computing Methods and the Phase Problem in X-ray Crystal Structure Analysis.* Oxford: Pergamon Press.

HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids.* New York: Benjamin.

International Tables for X-ray Crystallography (1962). Vol. III. Birmingham: Kynoch Press.

IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (1971). *Biochem. Biophys. Acta,* 229, 1–17.

JOHNSON, C. K. (1965). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Oak Ridge, Tennessee. The CDC-3600 version incorporates modifications by H. RAJAGOPAL & A. SEQUEIRA.

KANTERS, J. A., KROON, J., PEERDEMAN, A. F. & SCHOONE, J. C. (1967). *Tetrahedron,* 23, 4027–4033.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* 21, 849–859.

LAKSHMINARAYANAN, A. V., SASISEKHARAN, V. & RAMACHANDRAN, G. N. (1967). *Conformation of Biopolymers,* Vol. I. Edited by G. N. RAMACHANDRAN. London: Academic Press.

MALLIKARJUNAN, M., RAO, S. T., VENKATESAN, K. & SARMA, V. R. (1969). *Acta Cryst.* B25, 220–227.

MARSH, R. E. & DONOHUE, J. (1967). *Advanc. Protein Chem.* 22, 235–256.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Advanc. Protein Chem.* 23, 284–437.

SHOEMAKER, D. P., DONOHUE, J., SCHOMAKER, V. & COREY, R. B. (1950). *J. Amer. Chem. Soc.* 72, 2328–2349.

ZACHARIASEN, W. H. (1967). *Acta Cryst.* 23, 558–564.