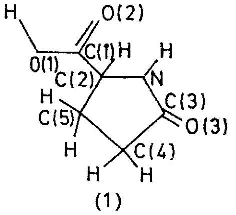
Crystal and Molecular Structure of Pyroglutamic Acid (5-Oxoproline)

By **Vasantha Pattabhi** * and **K. Venkatesan,** Centre of Advanced Study in Physics, University of Madras, Madras 600025, India

The crystal structure of the title compound has been determined by the symbolic addition method from 934 observed three-dimensional photographic data. Crystals are monoclinic with $a = 8{\cdot}14$, $b = 8{\cdot}86$, $c = 9{\cdot}32$ Å (all ±0·02 Å), and $\beta = 116{\cdot}5°(2)$, $Z = 4$, space group $P2_1/c$. The structure was refined by full-matrix least-squares method to R 0·091. The structure is stabilised by N–H · · · O and O–H · · · O type hydrogen bonds. The dimensions of the amide group, which is significantly non-planar, are compared with those found in other peptides.

PYROGLUTAMIC acid (PGA) is found in the naturally occurring peptides eisenine (L-pyroglutamyl-L-glutaminyl-L-alanine) and L-pyroglutamyl-L-glutaminyl-L-glutamine,[1] in peptide hormones such as gastrin, caerulein, and eledoisn, in venom peptide, and in rat collagen alpha-2 (fragment). In all these cases the amino-terminal end is a pyroglutamyl residue.[2] It is also a part of the porcine thyrotropin releasing hormone [TRH, L-pyroglutamyl-L-histidyl-L-proline(NH_2)],[3] isolated from the hypothalamus. In view of its occurrence in these biologically important peptides we decided to investigate its molecular structure. We also wished to determine accurately the dimensions of the amide group in pyroglutamic acid. In recent years compounds such as dioxopiperazine,[4] dimethyldioxopiperazine,[5] L-alanyl-L-alanyl dioxopiperazine,[6] and cyclotetrasarcosyl,[7] containing *cis*-peptide units have been studied by *X*-ray crystallographic methods. We hoped to compare the dimensions of the amide group in pyroglutamic acid (1) with its dimensions in these other peptides. A preliminary account of this work has appeared.[8]



(1)

EXPERIMENTAL

Crystal Data.—$C_5NO_3H_7$, $M = 129$. Monoclinic, $a = 8{\cdot}14 \pm 0{\cdot}02$, $b = 8{\cdot}86 \pm 0{\cdot}02$, $c = 9{\cdot}32 \pm 0{\cdot}02$ Å, $\beta = 116{\cdot}5°(2)$, $U = 601{\cdot}3$ Å^3, $D_m = 1{\cdot}509$ (by flotation), $Z = 4$, $D_c = 1{\cdot}424$. Space group $P2_1/c$ (C^5_{2h} No. 14) from systematic absences: $h0l$, for l odd, $0k0$ for k odd. Cu-K_α radiation, $\lambda = 1{\cdot}5413$ Å; μ(Cu-K_α) = 22·42 cm^{-1}.

Rotation and Weissenberg photographs were taken with Cu-K_α radiation. Three-dimensional intensity data were collected with Cu-K_α radiation for the layers $hk0$—8 and 0—$1kl$ by mounting the crystal about the c and a axes. Intensities were estimated visually and were corrected for Lorentz, polarization, and spot-shape factors.[9]

Of 1323 reflections theoretically available, it was possible to record only 1138 photographically, of which 204 were too low to have their intensities measured. The two data sets (a and c axes) were correlated by the method of Rollett and Sparks.[10]

Structure Determination and Refinement.—The structure was solved by a straightforward application of the symbolic addition procedure.[11] The program used was a modified version of that written by Dr. Thyagaraja Rao for an IBM 1130 computer. Atomic scattering factors used for carbon, nitrogen, oxygen, and hydrogen were taken from ref. 12.

Refinement was initially carried out by the use of successive three-dimensional difference Fourier maps. However, in the final stages a full-matrix least-squares method was used. Three cycles of least-squares refinement on the CDC 3600, with unit weights, isotropic thermal parameters, and an overall scale-factor, reduced R from 0·31 to 0·18.[13] The distribution of $\Sigma(\Delta F)^2$ *vs.* $\langle|F_o|\rangle$ for the common reflections used for data correlation suggested a weighting scheme of the form:[14] $w = (1{\cdot}8 + |F_o| + 0{\cdot}033\ |F_o|^2)^{-1}$. The difference map computed at this stage indicated hydrogens at the positions expected at heights varying from 0·3 to 1·00 e A^{-3}.

Three cycles of isotropic refinement with Cruickshank's weighting scheme, for all the atoms including the hydrogens gave R 0·15. Hydrogens were included only in the structure-factor calculation and were assigned the isotropic temperature factors of the atoms to which they were bonded. A further three cycles of refinement, with anisotropic thermal parameters for all the non-hydrogen atoms and an overall scale factor, reduced R to 0·091. In the last cycle the maximum shifts in the parameters were $<\frac{1}{7}\sigma$.

Final positional and thermal parameters for non-hydrogen atoms are recorded in Table 1, and hydrogen atoms in Table 2. Observed and calculated structure factors are listed in Supplementary Publication No. SUP 20918 (2 pp.).† Of the 208 unobserved reflections 67 were calculated to be

---

† See Notice to Authors No. 7 in *J.C.S. Perkin II*, 1973, Index issue.

[1] J. P. Greenstein and M. Winitz, 'Chemistry of Amino-acids,' Wiley, New York, 1961, vol. II, p. 771.

[2] M. O. Dayhoff and R. V. Eck, 'Atlas of Protein Sequence and Structure,' National Biomedical Res. Foundation, Silver Spring, Maryland, 1968.

[3] C. Y. Bowers, A. V. Schally, F. Enzmann, J. Boler, and K. Folkers, *Endocrinology*, 1970, **86**, 1143.

[4] R. Degeilh and R. E. Marsh, *Acta Cryst.*, 1959, **12**, 1007.

[5] P. Groth, *Acta Chem. Scand.*, 1969, **23**, 3155.

[6] R. Benedetti, P. Corradini, M. Goodman, and C. Pedone, *Proc. Nat. Acad. Sci. U.S.A.*, 1969, **62**, 650.

[7] P. Groth, *Acta Chem. Scand.*, 1970, **24**, 780.

[8] V. Pattabhi and K. Venkatesan, *Cryst. Struct. Comm.*, 1972, **1**, 87.

[9] D. C. Phillips, *Acta Cryst.*, 1954, **7**, 746.

[10] J. S. Rollett and R. A. Sparks, *Acta Cryst.*, 1960, **13**, 273.

[11] J. Karle and I. L. Karle, *Acta Cryst.*, 1966, **21**, 849.

[12] 'International Tables for *X*-Ray Crystallography,' vol. III, Kynoch Press, Birmingham, 1962.

[13] P. K. Gantzel, R. A. Sparks, and K. N. Trueblood, University of California Program, UCLALS1, 1961.

[14] D. W. J. Cruickshank, A. Bujosa, F. M. Lovell, and M. R. Truter, 'Computing Methods and the Phase Problems in *X*-Ray Analysis,' eds. R. Pepinsky and J. M. Robertson, Pergamon, Oxford, 1961, p. 45.

ca. 1·5 × $|F_{\text{min.}}|$. The intensities of reflections too weak to be observed were taken [15] to be $I_{\text{min.}}/3$.

TABLE 1

Positional and thermal parameters * of non-hydrogen atoms, with standard deviations in parentheses

(*a*) Thermal parameters

Atom	x	y	z
C(1)	0·4479(5)	0·2654(4)	0·3677(4)
C(2)	0·6392(5)	0·2272(4)	0·3804(4)
C(3)	0·9198(4)	0·1399(4)	0·5869(4)
C(4)	0·9663(5)	0·2895(4)	0·5315(5)
C(5)	0·7760(5)	0·3640(4)	0·4315(5)
N	0·7371(4)	0·1137(3)	0·5047(4)
O(1)	0·3771(4)	0·3890(3)	0·2826(4)
O(2)	0·3696(4)	0·1882(4)	0·4285(4)
O(3)	1·0323(3)	0·0523(3)	0·6904(3)

(*b*) Thermal parameters (× 10^4)

Atom	B_{11}	B_{22}	B_{33}	B_{12}	B_{13}	B_{23}
C(1)	112(7)	62(4)	118(6)	−14(8)	91(10)	20(7)
C(2)	112(7)	64(4)	110(5)	−13(8)	100(10)	8(7)
C(3)	106(7)	71(4)	112(6)	−9(7)	106(10)	−4(7)
C(4)	113(8)	85(5)	173(7)	−54(9)	128(12)	23(8)
C(5)	128(8)	71(4)	162(7)	−35(9)	129(12)	27(8)
N	94(6)	69(4)	135(5)	−21(7)	87(9)	22(6)
O(1)	119(5)	104(4)	172(6)	32(7)	127(9)	89(7)
O(2)	152(6)	138(5)	266(7)	41(9)	246(11)	172(9)
O(3)	101(6)	103(4)	141(5)	−5(7)	69(9)	40(6)

* The temperature factor is of the form $T = \exp[-(B_{11}h^2 + B_{22}k^2 + B_{33}l^2 + B_{12}hk + B_{13}hl + B_{23}kl)]$.

TABLE 2

Positional and thermal parameters of the hydrogen atoms

Atom	x	y	z	$B/\text{Å}^2$
H(1) [O(1)]	0·300	0·450	0·300	2·91
H(2) [C(2)]	0·617	0·192	0·267	2·15
H(3) [N]	0·683	0·017	0·500	2·29
H(41) [C(4)]	1·017	0·272	0·458	2·81
H(42) [C(4)]	1·033	0·300	0·633	2·81
H(51) [C(5)]	0·772	0·400	0·328	2·65
H(52) [C(5)]	0·766	0·364	0·533	2·65

DISCUSSION AND DESCRIPTION OF THE STRUCTURE

Bond Lengths and Angles.—Bond lengths and angles corresponding to the final co-ordinates are shown in Figure 1; the mean σ for bond lengths is 0·005 Å, and for angles 0·3°, errors in cell parameters having been taken into account in calculating these. The bond distances C(4)–C(5) and C(5)–C(2), which are C(sp^3)–C(sp^3) tend to be slightly longer than C(1)–C(2) and C(3)–C(4) as expected, since these are C(sp^2)–C(sp^3). C(1)–C(2) (1·545 Å) and C(3)–C(4) (1·532 Å) are significantly (*ca.* 6·5 σ) greater than is usual (1·5 Å) for a C(sp^2)–C(sp^3) single bond. The weighted average bond distance for this bond in the amino-acid is reported to be 1·527 ± 0·003 Å.[16,17] In the five-membered ring the angles at the tetrahedral carbon atoms are significantly less than the ideal (109·5°). Bond distances involving hydrogens range from 0·86 to 1·04 Å.

The carboxy-group in this structure is not ionized with C(1)–O(1) 1·324 and C(1)–O(2) 1·232 Å. As in most amino-acid and simple peptide structures, the C=O bond in the present structure is *cis* to nitrogen. The equation of the least-squares plane through atoms C(1), C(2), O(1), and O(2) and the deviations are recorded in Table 3.

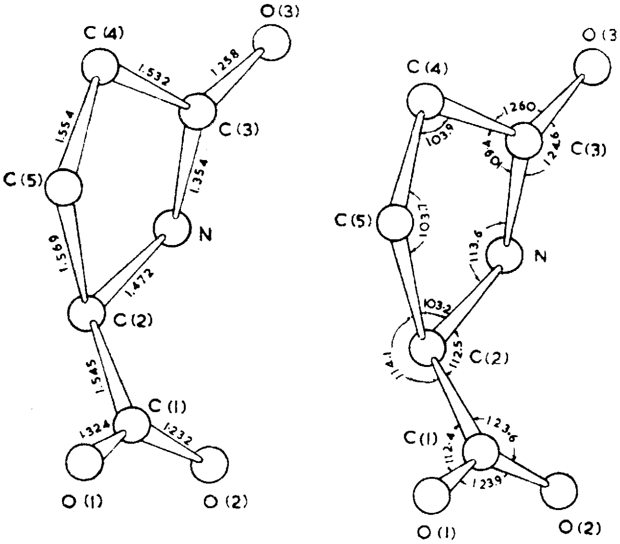


FIGURE 1 Bond lengths (Å) and angles (deg.) in the molecule

The carboxy-group is strictly planar with the nitrogen atom deviating from the plane by 0·33 Å; the torsion angle N–C(2)–C(1)–O(2) is −15°.

TABLE 3

Least-squares planes. Equations are in the form $AX + BY + CZ = D$, with respect to the crystallographic axes *a*, *b*, and *c** where *X*, *Y*, *Z* are in Å. Deviations (Å) of atoms from the planes are given in square brackets

Plane (1): O(1), C(1), C(2), O(2)
$0\cdot0609X + 0\cdot5477Y + 0\cdot8345Z = 3\cdot9712$
[O(1) −0·0014, C(1) 0·0043, C(2) −0·0012, O(2) −0·0017, N 0·3293]

Plane (2): C(2)—(4), O(3), N
$-0\cdot5323X + 0\cdot4883Y + 0\cdot6916Z = 1\cdot2913$
[C(4) 0·0172, C(3) 0·0132, O(3) −0·0272, N 0·0349, C(2) −0·0413, C(5) 0·3660]

Plane (3): C(4), C(3), O(3), N
$-0\cdot5157X + 0\cdot4785Y + 0\cdot7107Z = 1\cdot4632$
[C(4) −0·0017, C(3) 0·0060, O(3) −0·0023, N −0·0020, C(2) −0·1125, C(5) 0·3059]

Plane (4): C(2)—(5), N
$-0\cdot4835X + 0\cdot4124Y + 0\cdot7721Z = 1\cdot7756$
[C(4) −0·0291, C(3) 0·0751, C(5) 0·1471, C(2) −0·2464, N 0·0033]

Plane (5): C(2)—(4), N
$-0\cdot5405X + 0\cdot5068Y + 0\cdot6716Z = 1\cdot2088$
[C(2) −0·0145, N 0·0205, C(3) −0·0203, C(4) 0·0125, C(5) * 0·3994]

* Atom not included in defining the plane.

The *cis*-peptide unit C(4), C(3), O(3), N, and C(2) [plane (2) of Table 3] (C-1α, C′, O, N, and C-2α) in PGA

---

[51] W. C. Hamilton, *Acta Cryst.*, 1955, **8**, 185.

[16] M. Sundaralingam and E. F. Putkey, *Acta Cryst.*, 1970, **B26**, 790.

[17] R. E. Marsh and J. Donohue, *Adv. Protein Chem.*, 1967, **22**, 234.

is not typical because of the constraints imposed by the five-membered ring. Non-planarity of the peptide linkage is highly significant since $\chi^1 = (\Sigma\Delta^2/\sigma^2) = 404$ and for $\nu = 2$, $P \ll 0{\cdot}001$. The best plane among the peptide group of atoms is that through C(4), C(3), O(3), and N [plane (3) of Table 3] from which C(2) deviates by 0·11 Å. Similar situations are observed in other structures containing *cis*-peptide linkages. From the calculation of the least-squares planes passing through different sets of atoms of the amide group in *NN*′-dimethyldioxopiperazine [5] and cyclotetrasarcosyl [7] we again found the best plane to be that through C-1α, C′, N, and O. In *NN*′-dimethyldioxopiperazine the C-2α atom is out-of-plane by *ca.* 0·13 Å, and in cyclotetrasarcosyl by *ca.* 0·10 Å. The situation is also similar to that in simple linear peptides containing a *trans*-peptide linkage, *e.g.* glycyl-L-leucine,[18] in which the best least-squares plane is through C-1α, C′, O, and N, with H and C-2α being significantly out-of-plane.

TABLE 4

Peptide dimensions in various structures

(*a*) Distances (Å)

Molecule	C′–N	C=O	C-α–N	C-α–C′	σ
PGA [a]	1·354	1·258	1·472	1·532	0·005
Dioxopiperazine (DOP) [b]	1·325	1·239	1·449	1·499	0·002
NN′-Dimethyl-DOP [c]	1·348	1·234	1·455	1·506	0·003
Cyclotetrasarcosyl [d]	1·352	1·235	1·454	1·530	0·004
L-Ala-D-Ala-DOP [e]	1·325	1·240	1·462	1·470	
L-Ala-L-Ala-DOP [e]	1·323	1·244	1·462	1·512	
	1·344	1·218	1·457	1·535	

(*b*) Angles and the torsion angle, (ω) (deg.)

	C-α–N–C′	N–C′=C-α	N–C′=O	C-α–C′=O	ω	σ
PGA	113·6	109·4	124·6	126·0	4·2	0·3
DOP	126·0	118·9	122·6	118·5	0·0	0·1
NN′-Dimethyl-DOP	124·6	118·1	123·6	118·3	7·1	0·2
Cyclotetrasarcosyl	123·9	119·1	121·7	119·2	8·1	0·25
L-Ala-D-Ala-DOP	127·9	119·2	122·9	117·8	0·0	
L-Ala-L-Ala-DOP	126·3	117·4	122·0	120·4	0·0	
	127·1	115·1	124·8	120·1	8·0	

[a] Present work. [b] Ref. 4. [c] Ref. 5. [d] Ref. 7. [e] Ref. 6.

The observations indicate that the nitrogen of the amide group has a slight pyramidal character. The torsion angle about the bond C(3)–N is 4·2° [C(4)–C(3)–N–C(2)]. Comparison of the *cis* peptide dimensions in the present structure with those in other molecules containing this unit is given in Table 4. The present bond distances of the peptide unit agree fairly well with those in the other structures, but some of the corresponding angles differ significantly. The internal angles C-α–N–C′ and N–C′–C-α in PGA are significantly smaller than in the other cases, a decrease to be expected in view of the constraints imposed by the five-membered ring. However the value of 124·6° for the angle N–C′ = 0, in PGA is very close to the mean (122·9 ± 1·5°). While the internal angle N–C′–C-α in PGA is less, by 8°, than the mean value (117·9°) for the other structures, the external angle C-α–C′=O has been correspondingly increased by about the same amount.

Table 4(*b*) shows that in some cases there is no rotation about the C–N bond (ω 0°), whereas in others ω has significant values. This is similar to the situation in L-alanyl-L-alanyl dioxopiperzine [6] in which there is no rotation about one amide bond, but significant rotation about the other bond (ω 8°). It would seem that crystal packing forces also influence the geometry of the peptide group.

Conformation of the Molecule.—For an ideally planar ring system the sum of the five internal angles should be 540°. The sum of these angles in the present structure is 523·8°. In fact, the deviations of atoms C(2), N, C(3), C(4), and C(5) from the least-squares plane passing through them (Table 3) show that the five-membered ring is non-planar. However, atoms C(2), N, C(3), and C(4) define a better plane. The deviation of C(5) from this plane is 0·40 Å suggesting that the five membered ring possesses an envelope conformation. The conformational features are brought out clearly from the calculated torsion angles, θ, the amount of rotation about the bonds in the molecule. These are shown in Figure 2.

Packing and Hydrogen Bonding Scheme.—The molecular packing down the *b* axis is depicted in Figure 3, which shows hydrogen bonding between centrosymmetrically related molecules and also between these pairs. Thus chains (or sheets, whichever is the case) of molecules run through the structure, and are held together by van der Waals forces.

Intermolecular contacts <3·6 Å are recorded in Table 5. There are two protons in the molecule which can take part in hydrogen bonding and both of them are involved in intermolecular hydrogen bonds. Hydrogen bond distances and angles are listed in Table 6. In view of the large positional errors associated with the hydrogen atoms, the H · · · A (acceptor) distances, and D(donor)–H · · · A and H–D · · · A angles were also estimated from calculated hydrogen positions and are given in parentheses in Table 6. The oxygen $O(2^{II})$ of

---

[18] V. Pattabhi, K. Venkatesan, and S. R. Hall, *Cryst. Struct. Comm.*, 1973, **2**, 223.

the centrosymmetrically related molecule is directed so as to be able to form an N–H · · · O(2^{II}) bond of length 2·96 Å, which is slightly long. However, it has been observed that in peptides the nitrogen atom of the amide group tends to form significantly longer hydrogen bonds than the protonated terminal amino-groups.[17] The oxygen O(3^{XI}) of the carbonyl group accepts the proton from the carboxy-oxygen O(1) to give a hydrogen bond of length 2·59 Å. This observation, together with the fact that C(1)–O(1) is 1·324 Å, establishes that the proton is attached to O(1). The fact that the angle C(2)–C(1)–O(1) (112·4°) is smaller than C(2)–C(1)–O(2) (123·6) is consistent with this. The angle at the acceptor atom O(3) is 116·1° and the deviation of O(3) from the plane of the atoms C(1), C(2), and O(1) is 0·176 Å. This would mean that the proton-donating atom is almost in the direction of a lone pair of the acceptor O(3). In the case of the N–H · · · O(2) bond, the proton-donating

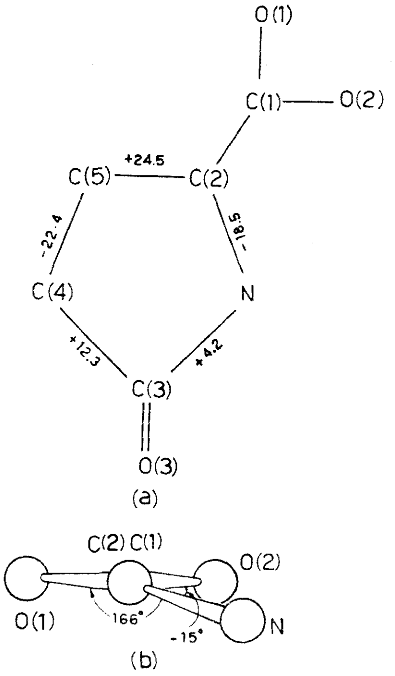


FIGURE 2 (a) Torsion angles (deg.) within the five-membered ring; (b) configuration about C(2)–C(1)

TABLE 5

Intermolecular contacts <3·6 Å

C(3) · · · O3^{I})	3·261	O(3^{X}) · · · C(5^{V})	3·567
O(3^{X}) · · · N^{II}	3·468	O(3^{X}) · · · O(1^{VI})	2·594
O(2) · · · N^{II}	2·963	O(3^{X}) · · · C(1^{VI})	3·434
C(3) · · · · C(3^{I})	3·516	O(3^{X}) · · · O(2^{VI})	3·508
O(3^{X}) · · · O(3^{IX})	3·473	C(4) · · · O(1^{VIII})	3·488
C(4) · · · O(3^{III})	3·481	C(3) · · · O(1^{VIII})	3·344
O(1) · · · C(2^{IV})	3·336	O(1) · · · O(2^{VII})	3·343
O(1) · · · N^{IV}	3·126		

Roman numeral superscripts refer to the following transformation of the atoms, relative to the reference molecule at x, y, z:

I $2 - x, -y - z$
II $1 - x, -y, 1 - z$
III $2 - x, \frac{1}{2} + y, \frac{3}{2} - z$
IV $1 - x, \frac{1}{2} + y, \frac{1}{2} - z$
V $1 - x, y - \frac{1}{2}, \frac{3}{2} - z$
VI $x, \frac{1}{2} - y, \frac{1}{2} + z$
VII $x, \frac{1}{2} - y, z - \frac{1}{2}$
VIII $1 + x, \frac{1}{2} - y, \frac{1}{2} + z$
IX $-x, -y, 1 - z$
X $x - 1, y, z$

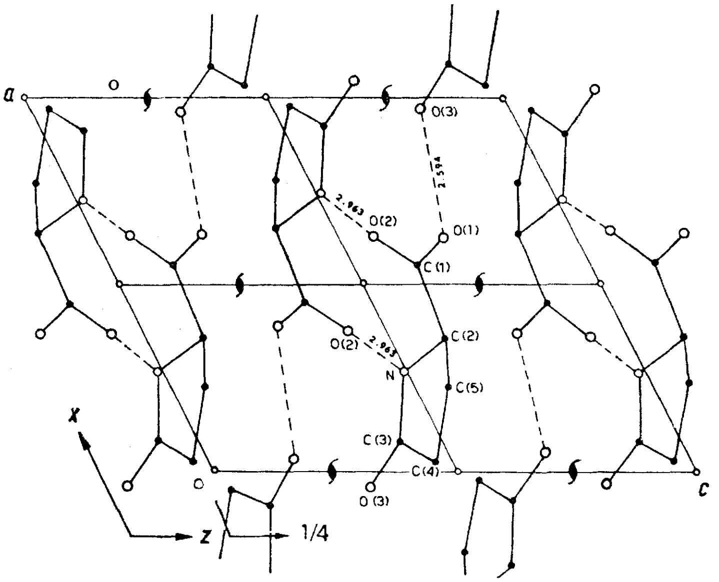


FIGURE 3 Packing of the molecule viewed down *b* axis

TABLE 6

Hydrogen bond lengths (Å) and angles (deg.)

Hydrogen bond	D · · · A	C–D · · · A	D–H · · · A	H–D · · · A
N–H · · · O(2^{II})	2·963	132·6	160·3 (159·7)	13·5 (13·5)
O(1)–H · · · O(3^{XI})	2·594	118·8	127·4 (159·6)	36·7 (12·6)

Roman numeral superscripts denote the following equivalent positions relative to the reference molecule at x, y, z:
II $1 - x, -y, 1 - z$ XI $x - 1, \frac{1}{2} - y, z - \frac{1}{2}$

atom is also approximately in the direction of a lone pair of O(2). The angle C(1)–O(2) · · · N is 149·1° and the deviation of O(2) from the plane defined by C(2), N, C(3) is 0·50 Å.

Calculation of the root-mean-square displacements along the principal axes of the thermal ellipsoids establish that all the atoms show the greatest vibration in a direction nearly parallel to the *c* axis. This appears to be related to the nature of the packing of the molecules in the unit cell. The molecules lie almost parallel to the *a* axis with the mean plane through all the atoms approximately parallel to (001). The N–H · · · O(3) bond is approximately parallel to the *b* axis whereas the O(1)–H · · · O(3) bond is in the *a* direction. It seems therefore reasonable that the direction of greatest motion of the atoms should be nearly parallel to the *c* axis.

We thank the University Grants Commission, New Delhi, for financial assistance (to V. P.).

[3/1806 *Received*, 30*th August*, 1973]