The Structure of o-Bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline Ethyl Acetate Monohydrate: a Substrate of the Enzyme, Collagenase

By T. Ueki,* S. Bando, T. Ashida and M. Kakudo

Institute for Protein Research, Osaka University, Osaka, Japan

The crystal and molecular structure of a synthetic oligopeptide, o-bromocarbobenzoxy-glycyl-prolyl-leucyl-glycyl-proline ethyl acetate monohydrate, has been established by the X-ray diffraction method. The monoclinic crystal belongs to the space group $P2_1$ with $Z=2$. Cell dimensions are $a=13\cdot710$, $b=14\cdot007$, $c=10\cdot861$ Å and $\beta=114\cdot60°$. As in the case of p-bromocarbobenzoxy-glycyl-prolyl-leucyl-glycine(tetrapeptide), the peptide chain is folded back at the Pro(1) and Leu residues to form two intra-molecular hydrogen bonds between two glycine residues. This folding conformation (called 'α-U-folding' in this paper) is almost identical to that in the tetrapeptide. This conformation is found in other oligopeptides and also in some protein structures; hence, it is understood to be one of the specific and stable conformations, in contrast with that proposed for the collagen molecule. In relation to this conformation, the 'β-U-folding' conformation is also discussed. In the crystal structure, intermolecular hydrogen bonds are observed between the peptide bonds of Pro(1)-Leu residues *via* the water molecule. The ethyl acetate molecule is in the hydrophobic environments. Generally speaking, intermolecular interaction is weak in this crystal, which is consistent with the lower density of the crystal compared with that of tetrapeptide.

Introduction

During a long-range research project in our laboratory to establish the crystal and molecular structures of amino acids, oligopeptides and protein crystals, a series of oligopeptides such as carbobenzoxy (Z)-Gly, Z-Gly-Pro, Z-Gly-Pro-Leu, Z-Gly-Pro-Leu-Gly, Z-Gly-Pro-Leu-Gly-Pro, has been studied by X-ray diffraction methods (Sasada, Tanaka, Ogawa & Kakudo, 1961; Sasada & Kakudo, 1961; Kakudo, Sasada, Katsube, Sakakibara & Akabori, 1963). These substances were well known especially in connexion with the substrate specificity of the enzymatic reaction of collagenase(Nagai & Noda, 1959; Nagai, Sakakibara, Noda & Akabori, 1960; Sakakibara & Nagai, 1960). Among these peptides, it has been shown that only pentapeptide is subjected to a high degree of specific degradation by collagenase and that the sequence of amino acid residues, -Pro-X-Gly-Pro-, is essentially important for the substrate specificity.

The ORD studies of these peptides showed that a critical difference was observed between two groups: one containing Z-Gly, Z-Gly-Pro, Z-Gly-Pro-Leu and the other Z-Gly-Pro-Leu-Gly, Z-Gly-Pro-Leu-Gly-Pro (Hamaguchi, 1970). This indicates that there must be a substantial difference in the conformation of the molecules in the two groups, while a close similarity may be expected in their conformation of tetra- and penta-peptide.

The X-ray analyses of these peptides afford some structural basis for problems of biochemical aspects, and also contribute to the interpretation of optical rotatory dispersion (ORD) and spectroscopic measurements. Furthermore, the establishment of the molecular structures of tetra- and penta-peptide may contribute some basic knowledge to the partial structures of the crystalline region of the collagen molecule. Apart from these interests, comparison of conformation with other peptides such as Ferrichrome *A* (Zalkin, Forrester & Templeton, 1966), cyclohexaglycyl (Karle & Karle, 1963), CysH-Pro-Leu-GlyNH₂ (segment of oxytocin) (Low, Lovell & Rudko, 1969), and ⌐(Gly)₄-(*D*-Ala)₂⌐ (Karle, Gibson & Karle, 1970) is important for the structural chemistry of oligopeptides.

Recently, the crystal and molecular structure of Z-Gly-Pro-Leu-Gly was established (Ueki, Ashida, Kakudo, Sasada & Katsube, 1969), and its conformation was found to be the 'folded' rather than the extended structure, contrary to the expectation from preliminary study (Kakudo *et al.*, 1963). The present paper deals with the crystal and molecular structures of the pentapeptide, o-bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline which is crystallized with ethyl acetate and a water molecule. The serial numbering of the atoms in a peptide molecule is shown in Fig. 1.*

---

* Present address: Faculty of Engineering Science, Osaka University, Toyonaka, Japan.

* In the present paper the numbering of the amino-acid residues is, from the amino terminal, Z-Gly(1)-Pro(1)-Leu-Gly(2)-Pro(2). The numbering of the atoms in serial manner, as listed in the first column in Table 1, is used in the sections of structure analysis and crystal structure. In the rest of the paper, the usual nomenclature of amino acid residues, as defined by Edsall, Flory, Kendrew, Liquori, Nemethy, Ramachandran & Scheraga (1966), is adopted as listed in the last column of Table 1.

Experimental

The crystals of o-bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline [Z(o-Br)-GPLGP] and p-bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline [Z(p-Br)-GPLGP], obtained from the ethyl acetate and water solution, are opaque crystals of ill-defined shape and are both equally poor for diffraction work. Unit-cell dimensions of the crystals are:

	Z(o-Br)-GPLGP	Z(p-Br)-GPLGP
	$C_{28}H_{38}N_5O_8Br \cdot C_4H_8O_2 \cdot H_2O$	
a	$=13\cdot710+0\cdot007$ Å	15·47 Å
b	$=14\cdot007+0\cdot004$	14·88
c	$=10\cdot861+0\cdot012$	21·41
β	$=114\cdot60+0\cdot02°$	120°
Z	$=2$	4
d_{obs}	$=1\cdot33$ g.cm^{-3} .	

Both crystals are monoclinic, with space group $P2_1$. In Z(p-Br)-GPLGP the dimension of the c axis is nearly twice that in Z(o-Br)-GPLGP, and the former contains two molecules in an asymmetric unit. The characteristic diffraction pattern in Z(p-Br)-GPLGP, however, is that the reflexions with l odd indices are very weak. This fact suggests that the locations of the two crystallographically independent molecules along the c axis are very closely related as z and $z+\frac{1}{2}$. This situation was also observed in gramicidin S (Schmidt, Hodgkin & Oughton, 1957), cytochrome c from bonito heart (Ashida, Ueki, Tsukihara, Sugihara, Takano & Kakudo, 1971) and in cyclohexaglycyl; the structure analyses of this type of crystal usually suffer in detecting the small differences of nearly identical molecules responsible for the weak l-odd reflexions. In these cases, structure analyses that ignore these weak reflexions result in the averaged structure of nearly identical molecules. In the present case, it is also suggested that both crystals may be isostructural with respect to the molecular conformation of the peptide. Accordingly, we carried out the structure analysis of Z(o-Br)-GPLGP and avoided the difficulties in Z(p-Br)-GPLGP due to the weak l odd reflexions.

Only one piece of the crystal was available. It was large in size ($0\cdot29 \times 0\cdot77 \times 0\cdot48$ mm) and, though poor for diffraction work, was used throughout the experiments. Unit-cell dimensions were determined on the diffractometer by using Cu Kα radiation ($\lambda = 1\cdot5418$ Å). All measurements were done on the automatic four-circle diffractometer controlled by an Aicom C-2 computer with 8K core memory (this system is called Rigaku AFC-II).

The molecular weight in an asymmetric unit calculated from the crystal data is 759, which is less than the weight of the chemical formula cited above. This suggests that some part of the crystalline solvents (ethyl acetate or water) was lost in this crystal, as is often observed in cases of opaque crystals.

Intensity data were collected using Cu Kα radiation monochromatized by a nickel foil of proper thickness, with a scintillation counter connected to a pulse-height analyser. Measurements were controlled automatically under the following conditions:

Range of measurement: (sin θ/λ) less than $0\cdot562$ Å^{-1}
Method of scanning: continuous $\omega - 2\theta$ scan
Scanning range: $\Delta(2\theta) = 2\cdot00° + 0\cdot40°$ tan θ.
Attenuators of nickel foil were used to maintain measurements within the linear region of the scintillation counter. A total of 2910 independent reflexions was measured, within the quadrant of the measuring
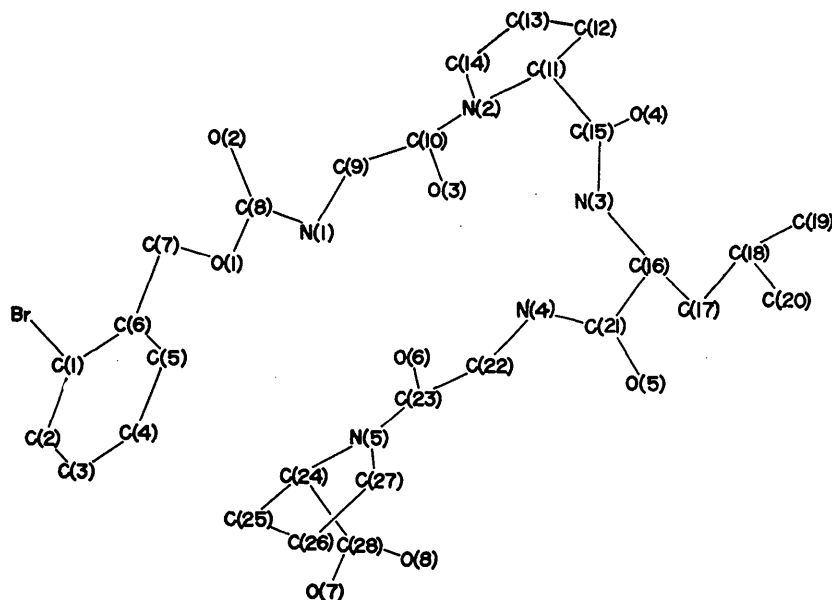

Fig. 1. Serial numbering of atoms in a peptide molecule.

sphere, in five days. During measurement the 080 reflexion was monitored to inspect the deterioration of the crystal by X-rays or the change of the setting parameters. The change in intensity of this reflexion was about 13% in total. This change occurred almost linearly with time, and the intensity data were corrected correspondingly. The absorption correction was not made, though it was needed for this large crystal.

Structure analysis

The parameters of the bromine atom were easily found in the Patterson maps sharpened with $B = 4.0$ Å^2.

Using these parameters, the heavy-atom Fourier synthesis and minimum function maps were obtained. The superposition of these maps revealed the phenyl group and C(7) atom. Subsequently, these atoms were included in the least-squares calculation followed by the Fourier summation. More than 10 steps were needed to yield the positions of the light atoms in the peptide chain; in each step, atoms whose temperature factors became unreasonably large after the least-squares calculation were discarded in calculating the phases of the Fourier summation, and a new set of atoms was picked up in the Fourier maps for the next step. After the atomic positions of the peptide chain

Table 1. *Positional parameters (fractional) with e.s.d.'s* (Å $\times 10^3$) *in parentheses and isotropic thermal parameters*

	x	y	z	B	
Br	0·0371 (3)	0·2500 (4)	0·1495 (3)	8·8*	
C(1)	0·1158 (21)	0·1631 (22)	0·0827 (21)	6·1	Z–C(1)
C(2)	0·1699 (22)	0·2067 (22)	0·0198 (22)	6·7	Z–C(2)
C(3)	0·2244 (26)	0·1433 (27)	−0·0313 (26)	7·9	Z–C(3)
C(4)	0·2300 (23)	0·0488 (25)	−0·0130 (23)	7·0	Z–C(4)
C(5)	0·1755 (25)	0·0073 (26)	0·0553 (25)	8·1	Z–C(5)
C(6)	0·1115 (20)	0·0719 (20)	0·1033 (20)	5·7	Z–C(6)
C(7)	0·0639 (23)	0·0234 (24)	0·1992 (23)	6·9	Z–C(7)
C(8)	−0·1132 (20)	0·0222 (20)	0·1907 (20)	5·5	Z–C(8)
C(9)	−0·2869 (16)	0·0146 (17)	0·1800 (16)	4·0	gly(1)C$_\alpha$
C(10)	−0·3389 (15)	−0·0775 (16)	0·1276 (16)	3·5	gly(1)C′
C(11)	−0·4692 (15)	−0·1997 (15)	0·1197 (15)	3·4	pro(1)C$_\alpha$
C(12)	−0·5734 (19)	−0·1942 (20)	0·1541 (20)	5·3	pro(1)C$_\beta$
C(13)	−0·5481 (22)	−0·1090 (22)	0·2516 (22)	6·3	pro(1)C$_\gamma$
C(14)	−0·4724 (17)	−0·0416 (17)	0·2210 (17)	4·3	pro(1)C$_\delta$
C(15)	−0·5128 (15)	−0·2280 (15)	−0·0339 (14)	3·3	pro(1)C′
C(16)	−0·5940 (17)	−0·1759 (17)	−0·2712 (17)	4·2	leuC$_\alpha$
C(17)	−0·6932 (18)	−0·1178 (18)	−0·3465 (18)	4·6	leuC$_\beta$
C(18)	−0·7839 (20)	−0·1435 (21)	−0·2982 (20)	5·7	leuC$_\gamma$
C(19)	−0·8189 (26)	−0·2395 (32)	−0·3250 (27)	9·5	leuC$_\delta$(1)
C(20)	−0·8785 (30)	−0·0798 (30)	−0·3795 (30)	9·5	leuC$_\delta$(2)
C(21)	−0·5067 (16)	−0·1483 (16)	−0·3239 (16)	3·7	leuC′
C(22)	−0·3296 (17)	−0·0954 (17)	−0·2870 (17)	4·2	gly(2)C$_\alpha$
C(23)	−0·2961 (14)	0·0078 (15)	−0·2464 (14)	3·0	gly(2)C′
C(24)	−0·1831 (19)	0·1443 (16)	−0·2489 (17)	4·1*	pro(2)C$_\alpha$
C(25)	−0·0836 (17)	0·1438 (18)	−0·2812 (18)	4·6*	pro(2)C$_\beta$
C(26)	−0·1087 (21)	0·0736 (19)	−0·3891 (24)	4·6*	pro(2)C$_\gamma$
C(27)	−0·1117 (20)	−0·0117 (18)	−0·3587 (21)	4·0*	pro(2)C$_\delta$
C(28)	−0·2673 (31)	0·2116 (20)	−0·3425 (34)	5·5*	pro(2)C′
C(29)	0·1326 (27)	0·2200 (41)	0·5673 (30)	8·7*	EA–C(1)†
C(30)	0·2350 (24)	0·2273 (30)	0·5440 (23)	7·2*	EA–C(2)
C(31)	0·3747 (36)	0·1549 (33)	0·5085 (36)	6·9*	EA–C(3)
C(32)	0·3754 (34)	0·1006 (34)	0·4238 (34)	7·2*	EA–C(4)
N(1)	−0·2119 (14)	0·0411 (14)	0·1208 (14)	4·4	gly(1)N
N(2)	−0·4256 (12)	−0·0998 (12)	0·1478 (12)	3·4	pro(1)N
N(3)	−0·5478 (12)	−0·1537 (12)	−0·1240 (12)	3·0	leuN
N(4)	−0·4102 (13)	−0·1242 (13)	−0·2402 (13)	3·7	gly(2)N
N(5)	−0·2208 (11)	0·0455 (12)	−0·2771 (11)	3·2*	pro(2)N
O(1)	−0·0524 (13)	0·0328 (14)	0·1167 (13)	5·9	Z–O(1)
O(2)	−0·0669 (16)	−0·0041 (17)	0·3113 (16)	7·7	Z–O(2)
O(3)	−0·3061 (11)	−0·1344 (11)	0·0630 (11)	4·1	gly(1)O
O(4)	−0·5166 (11)	−0·3119 (11)	−0·0642 (11)	4·1	pro(1)O
O(5)	−0·5358 (15)	−0·1561 (15)	−0·4472 (15)	7·0	leuO
O(6)	−0·3375 (12)	0·0577 (12)	−0·1827 (12)	4·8	gly(2)O
O(7)	−0·3379 (16)	0·1854 (17)	−0·4347 (19)	9·3*	pro(2)O(1)
O(8)	−0·2527 (25)	0·2962 (16)	−0·2938 (28)	8·3*	pro(2)O(2)
O(9)	0·2781 (20)	0·3041 (18)	0·5597 (18)	7·3*	EA–O(1)
O(10)	0·2709 (18)	0·1532 (19)	0·5198 (21)	6·4*	EA–O(2)
O(11)	−0·5193 (12)	0·0442 (12)	−0·1218 (12)	5·0	water

* Equivalent temperature factors are those defined by Hamilton (1959).
† Ethyl acetate.

were obtained, the position of O(11) of water and atoms of ethyl acetate were found in the Fourier maps; the peaks belonging to ethyl acetate were very low and diffused in the maps [even in the final Fourier maps, peak heights of carbon atoms in this molecule were 2·3, 3·4, 2·5 and 2·4 e.Å^{-3}, while those in the peptide chain were from 5·6 to 6·8 e.Å^{-3} except for atoms in Pro(2)].

The least-squares refinement including all atoms with isotropic thermal parameters resulted in $R = 14·6\%$, where the occupancies of atoms in the ethyl acetate molecule were tentatively assigned as 0·75 since some molecules in the crystal were estimated to be randomly lost. A trial to refine the occupancies was not made. The difference Fourier synthesis at this stage gave a number of high peaks in the maps, indicating a high degree of anisotropic thermal vibrations or disorder. Among these peaks two had heights more than 1·0 e.Å^{-3} at the positions very close to the carboxyl group of Pro(2).

Finally, a refinement with anisotropic thermal parameters for bromine, and atoms of the Pro(2) residue and ethyl acetate resulted in a fairly good convergence of parameters. Even at this stage, many peaks existed around the carboxyl atoms of Pro(2) residue and ethyl acetate in the difference Fourier maps.

Positions of the hydrogen atoms were found for the peptide chain (except for those at the carboxyl group of Pro(2) residue and the methyl groups of Leu residue) and the water molecule in the difference Fourier maps by referring to the calculated positions. Peak heights in the maps of these hydrogen atoms ranged from 0·19 to 0·52 e.Å^{-3}. The final R value was 11·6% (hydrogen atoms were not included in the refinements).

Final parameters of all atoms are listed in Tables 1, 2 & 3. Atomic parameters of the C_γ, C_δ, C', O(7) and O(8) atoms are especially affected by the disordering, as can be seen from the anisotropic thermal param-
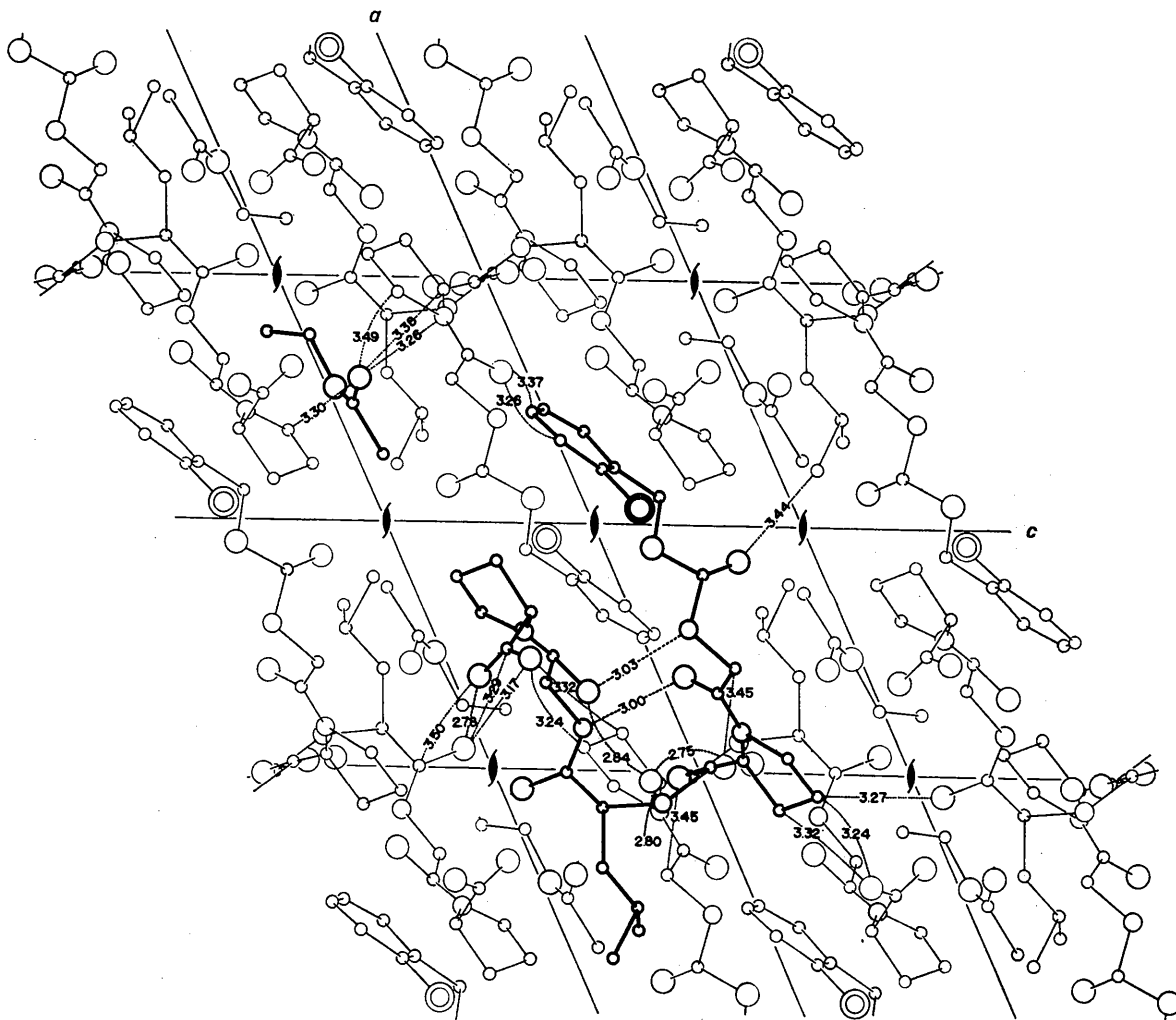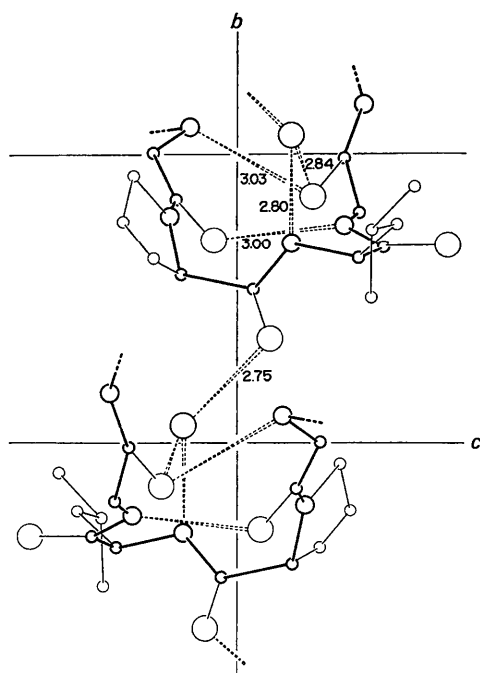

Fig. 2. Crystal structure viewed down the b axis. Hydrogen bonds are drawn in heavy dotted lines, some intermolecular short contacts in thin broken lines.

Fig. 3. Part of the crystal structure, as viewed along the a axis.

Table 3. *Positional parameters of hydrogen atoms*

Parameters are taken from the difference Fourier maps.

	x	y	z	Bonded to
H(1)	0·14	0·24	0·00	C(2)
H(2)	0·20	0·18	−0·04	C(3)
H(3)	0·26	0·00	−0·04	C(4)
H(4)	0·18	−0·06	0·06	C(5)
H(5)	0·08	0·04	0·30	C(7)
H(6)	0·06	−0·04	0·18	C(7)
H(7)	−0·24	0·06	0·02	N(1)
H(8)	−0·34	0·06	0·14	C(9)
H(9)	−0·24	0·00	0·24	C(9)
H(10)	−0·52	0·02	0·20	C(14)
H(11)	−0·40	−0·02	0·26	C(14)
H(12)	−0·62	−0·08	0·26	C(13)
H(13)	−0·50	−0·12	0·30	C(13)
H(14)	−0·62	−0·16	0·10	C(12)
H(15)	−0·54	−0·26	0·20	C(12)
H(16)	−0·42	−0·26	0·18	C(11)
H(17)	−0·54	−0·10	−0·10	N(3)
H(18)	−0·60	−0·24	−0·28	C(16)
H(19)	−0·72	−0·14	−0·42	C(17)
H(20)	−0·70	−0·08	−0·32	C(17)
H(21)	−0·76	−0·12	−0·22	C(18)
H(22)	−0·38	−0·16	−0·20	N(4)
H(23)	−0·28	−0·14	−0·24	C(22)
H(24)	−0·38	−0·08	−0·36	C(22)
H(25)	−0·18	0·14	−0·18	C(24)
H(26)	0·00	0·14	−0·20	C(25)
H(27)	−0·08	0·20	−0·30	C(25)
H(28)	−0·02	0·06	−0·40	C(26)
H(29)	−0·14	0·06	−0·42	C(26)
H(30)	−0·24	−0·02	−0·52	C(27)
H(31)	−0·10	−0·06	−0·30	C(27)
H(32)	−0·42	0·06	−0·16	O(11) water
H(33)	−0·56	0·12	−0·12	O(11) water

eters listed in Table 2. Observed and calculated structure factors are tabulated in Table 4. Most calculations were carried out on the HITAC 5020E computer, University of Tokyo, by using the programs written by one of the authors (Ashida, 1967). The block-diagonal approximation was used in the least-squares calculation, in which all intensity data were used with unit weight. Atomic scattering factors were taken from *International Tables for X-Ray Crystallography* (1962). Anomalous dispersion terms were not included in the calculations.

Table 2. *Anisotropic thermal parameters* ($\times 10^4$)*

	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
Br	181	143	299	−31	256	−147
C(24)	119	51	143	−51	152	−17
C(25)	73	75	176	−52	117	5
C(26)	129	76	309	1	334	24
C(27)	135	67	234	18	290	−6
C(28)	308	66	600	83	807	153
C(29)	100	224	205	58	178	129
C(30)	87	130	116	−55	12	131
C(31)	175	105	316	−116	314	−120
C(32)	162	128	250	−75	249	−120
N(5)	58	64	94	6	84	−8
O(7)	131	151	353	59	214	298
O(8)	376	68	736	77	877	163
O(9)	145	95	150	−8	73	40
O(10)	105	81	255	−17	144	25

* These parameters were used in the form of:

$$\exp\left[-(\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}l^2 + \beta_{12}hk + \beta_{13}hl + \beta_{23}kl)\right].$$

Crystal structure and intermolecular hydrogen bonds

The peptide molecule is schematically approximated as a thick sheet, roughly perpendicular to the b axis, on which a 'U'-shaped peptide chain lies. Sheets are stacked with each other by the twofold screw axis. In this stacking structure, hydrophobic parts of the molecules (phenyl and pyrrolidine groups) contact each other between neighbouring molecules; hydrophilic parts (peptide groups between Pro(1) and Leu residues) interact with the upper and lower ones *via* water molecules located between these molecules. The arrangement of the molecules in the crystal is shown in Fig. 2 (viewed down the b axis) and in Fig. 3 (viewed down the a axis. Fig. 4 shows an enlarged and modified part of Fig. 2 looking along the twofold screw axis.

Van der Waals contacts

As seen in Fig. 2, the carbobenzoxy group seems fairly free; it has the closest contacts with O(3) of the Gly(1) residue, within 3·5 Å and close van der Waals contacts with the pyrrolidine rings of Pro(2) of the adjacent molecule. These facts are in accordance with the high-temperature factors of the atoms in the phenyl group. The side chain of the Leu residue is also free; atoms in this side chain have quite high temperature factors (9·5 Å^2 for the two C_δ atoms).

Note that the pyrrolidine ring of Pro(1) residue has contacts mainly with the oxygen atoms, O(8), O(5) and the O(9) of ethyl acetate within 3·5 Å. The pyrrolidine ring of the Pro(2) residue has a few short contacts with the neighbouring molecules. The only one strong interaction of Pro(2) is the hydrogen bond. As a result, the carboxyl group of Pro(2) residue retains some freedom of rotation around this hydrogen bond, causing severe disorder on these atoms.

The crystalline solvent (ethyl acetate molecules) is surrounded by the hydrophobic group, such as the side chain of the Leu residue and the pyrrolidine rings of proline residues. Of the atoms in ethyl acetate, only the oxygen atom of the carbonyl group has shorter contacts with other atoms.

Hydrogen bonds

The water molecule is situated in the middle of the two peptide bonds of Pro(1)-Leu, which are related by the twofold screw axis; it is located very closely to the screw axis. Two hydrogen atoms of the water molecule

Table 4. *Observe and calculated structure factors*
(on absolute scale)

point toward O(4′) of Pro(1′) of the upper molecule and O(6) of Gly(2) of the lower molecule; angle O(4)---H–O(11)–H---O(6), 103°, is quite satisfactory. The O(11) atom is also an acceptor of the hydrogen bond from N(3) of the Leu residue in the lower molecule. As a result of these three hydrogen bonds, the two peptide bonds, which are on the corner of the U folding, are strongly interconnected. Interaction between the molecules is strong in this part of the crystal.

In the hydrogen bonds between the carboxyl group of Pro(2) and the carbonyl group of the Leu residue, the O(7) atom may be the donor, though the bond distance, C(28)–O(7), is 1·12 Å and the other, C(28)–O(8), is 1·28 Å. The reason is that the O(7)---O(5′) distance is 2·78 Å while O(8)---O(5′) is 3·17 Å, and also the angles, C(28)–O(7)---O(5′) and C(28)–O(8)---O(5′), are 107 and 84°, respectively. The conflict in bond distances is due to the disorder observed in this carboxyl group. Angles of intra- and inter-molecular hydrogen bonds are given in Fig. 5.

Description of the molecular structure

The ethyl acetate molecule suffers from disorder, as indicated by the anomalous anisotropic thermal parameters. Hence, dimensions in the molecule have large deviations from those so far reported. They are shown in Fig. 6(c).

Geometry of amino acid residues

Bond lengths and angles in the peptide molecule are shown in Fig. 6. Because of the disorder or highly anisotropic thermal vibrations even for the atoms treated as isotropic to save the machine time of the computer, the estimated standard deviations for the non-heavy atoms range from 0·015 Å to about 0·05 Å for the atoms apparently having disorder. Those for the peptide bonds in this molecule are as follows.
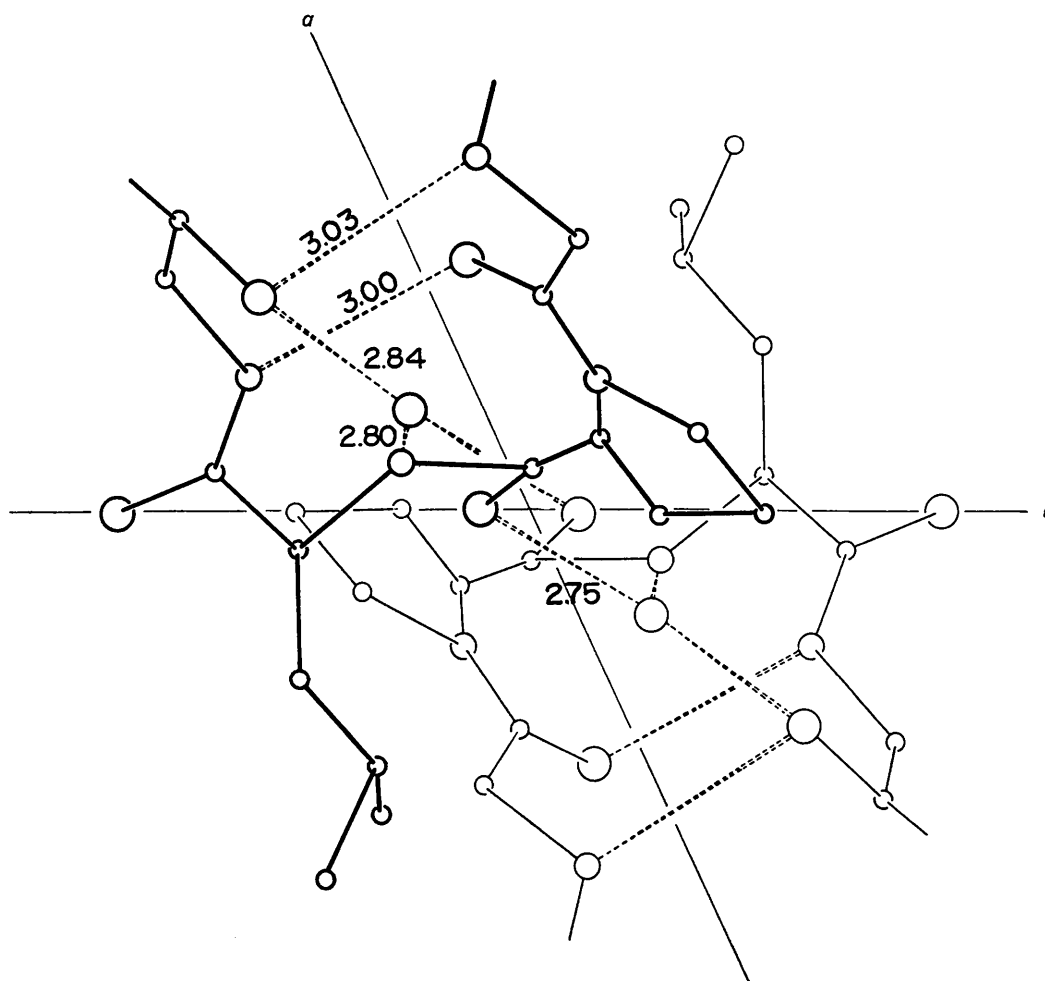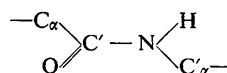

Fig. 4. Enlarged and modified part of Fig. 2, showing the hydrogen-bond system along the b axis.

	Present	Corey & Pauling (1953)	Average*
C_α–C′ distance	1·53 Å	1·53 Å	1·528 Å
C′–O	1·25	1·24	1·236
C′–N	1·32	1·32	1·321
N–C_α'	1·47	1·47	1·459
C_α–C′–O angle	119°	114°	114·7°
C_α–C′–N	120	121	119·8
O–C′–N	123	125	125·3
C′–N–C_α'	119	123	121·1

These values are very reasonable except angle C_α–C′–O. The best planes of the peptide bonds listed in Table 5 indicated that these peptide bonds have good planarity except for the first peptide bond between the carbo-benzoxy group and Gly(1) residue.

The geometry of the pyrrolidine rings in Pro(1) and Pro(2) residues is very similar to those so far

---

* Averages are estimated from the dimensions of the peptide bonds so far determined by the three-dimensional structure analyses (14 crystals).

reported. In the rings, the four atoms, N, C_α, $C_{,\beta}$ C_δ, are nearly coplanar, and the displacements of the C_γ atoms from these planes are 0·41 Å for Pro(1) and 0·49 Å for Pro(2) residue, respectively. The carbon atom of the amide group, Pro(1)C′, is located in the opposite direction to the C_γ atom with respect to the plane, while in the Pro(2) residue the C_γ and C′ atoms are in the same direction. The disorder in C_γ atoms is not observed, as in the Z-GPLG peptide (Ueki et al., 1969), and this is in contrast to the crystal of L-Leu-L-Pro-Gly (Leung & Marsh, 1958) in which the disorder in the C_γ atom was detected.

An interesting feature observed in Z-GPLG peptide, i.e. the widening of angle C_α–C_β–C_γ in the Leu residue, is not observed in the present case. It has the usual value of 111°. The reason for it is not yet clear. This result might be due to the experimental error.

As pointed out for the structure of the Z-GPLG molecule, the present molecule also has its C_α–H bond cis to the C′=O bond with respect to the C_α–C′ bond in the Pro(1) residue (Fig. 7). This conformation was found in myoglobin (Kendrew, 1963) and lysozyme structures (Blake, Mair, North, Phillips & Sarma,
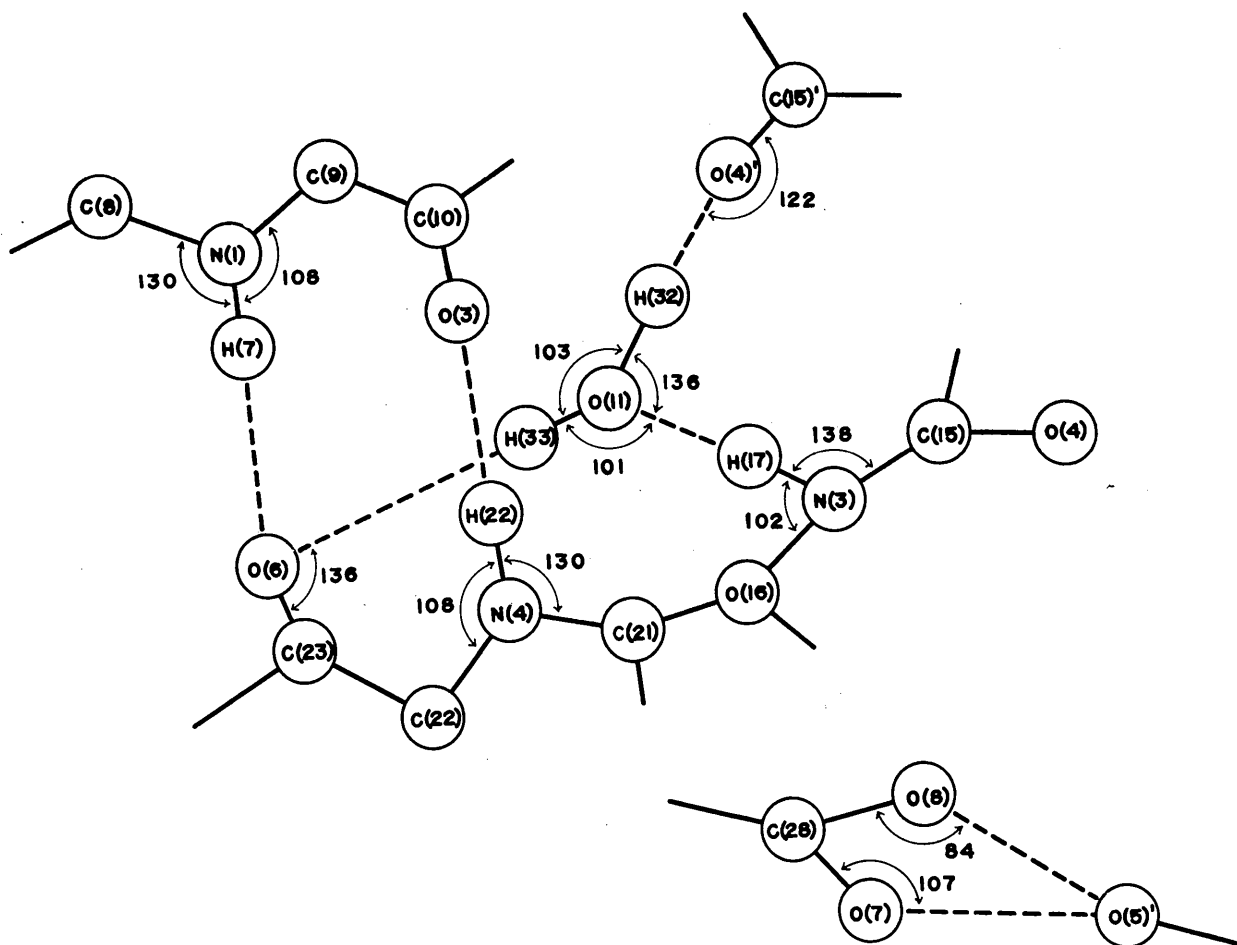

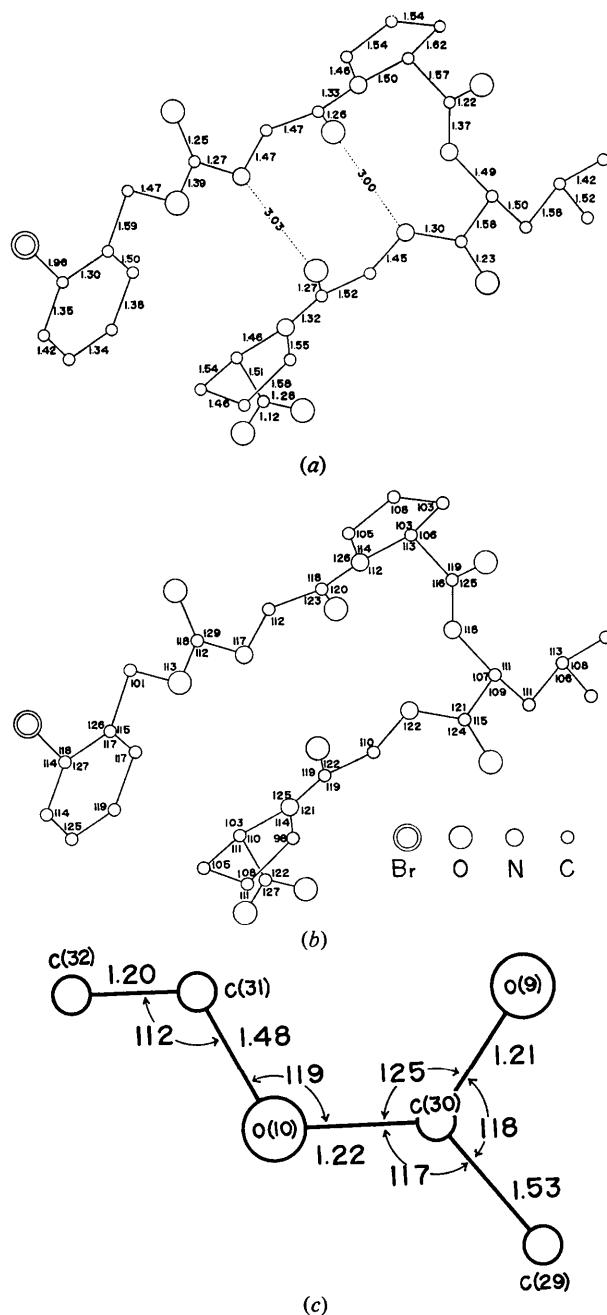Fig. 5. Angles in the hydrogen bond system.

(a)

(b)

(c)

Fig. 6. Peptide molecule showing (a) bond distances, (b) bond angles. Bond distances and angles in ethyl acetate molecule are shown in (c).

1967), but was not found in polyproline II (Cowan & McGavin, 1955), in the model of collagen (Ramachandran & Sasisekharan, 1965), or in Leu-Pro-Gly and tosyl-L-Pro-L-HyPro (Fridrichsons & Mathieson, 1962). Therefore, this conformation seems to occur in a certain oligopeptide having a particular conformation, as discussed later. In a previous paper, we suggested that this conformation may, somehow, have a correlation with the difference in the substrate specificity of

the tetra- and penta-peptide to the reaction of collagenase. The present analysis, however, revealed that this conformation occured in the pentapeptide as well, in spite of the difference in biochemical behavior of these peptides.

As far as the Pro(2) residue is concerned, the molecular dimensions are not usable for considering any precision, because of the disorder as stated previously.

Conformation of the peptide molecule

The most important feature of the molecule is the folding of the chain at the Pro(1) and Leu residues. This folding structure is constructed by rotating around the bonds of C_α–C' in the Pro(1) and of N–C_α in the Leu residue from the chain's fully extended conformation. From a qualitative consideration, rotation of either of these bonds seems to result in four different types of structures, keeping other bonds unchanged:

(1) Structure as the present molecule, which is stabilized by the intrachain hydrogen bonds [Fig. 8(c)].

(2) Structure with either of these bonds rotated by 180° (if the C_α–C' bond is rotated, the C_α–H bond becomes *trans* to the C'=O bond with respect to the C_α–C' bond, which is seen in the structures of collagen and polyproline II); in this structure, the bulky hydrophobic side chain of the Leu residue may produce the repulsive interaction with the Gly(1) residue; hence, this structure seems unstable.

(3) Twisted structure in which either of these bonds is rotated by an angle between 0 and 180°; this rotation results in an intermediate structure between (1) and (2).

(4) Structure in which both bonds are rotated by about 180°, which gives a structure similar to (1) with two intrachain hydrogen bonds. In this structure, the C_α–H bond is *trans* to the C'=O with respect to the C_α–C' bond.

For the peptide with a sequence such as (–Pro-amino acid with bulky side chain–) the folded conformation in (1) and (4) may be favoured. These conformations are discussed later in comparison with other peptide molecules.

Distances of the intrachain hydrogen bonds are: Gly(1)N-H---Gly(2)O, 3·03 Å, and Gly(1)O---H-Gly(2)N, 3·00 Å. These two hydrogen bonds greatly stabilize this folded conformation, which together with the hydrogen bonds seems to be the antiparallel pleated-sheet structure (β-structure). However, the location of the C_α atoms of Gly(1) and Gly(2) residues are not in the same direction with respect to the hydrogen bonds; one is above the hydrogen bond and the other below it. Interatomic short contacts in the chain (less than 3·5 Å) are shown in Fig. 9.

Between the conformation of Z-GPLG and Z-GPLGP, a remarkable difference is observed in the internal rotation angle of the C_α–C' bond of the Gly(1) residue which causes the hydrogen atom of Gly(1)N to stick inside the chain of Z-GPLGP (see internal rotation angles in Fig. 10). This makes the additional hydrogen bond possible between Gly(1)N-H---Gly(2)O.

Orientation of the carbobenzoxy group is different in both molecules; the φ value of Gly(1) is 83° in Z-GPLGP and 271° in Z-GPLG. This indicates about 180° difference in internal rotation angle in these molecules. As a result, the Z-O(2) atom is outside the chain in Z-GPLGP, but is inside the chain in Z-GPLG.

U-folding, special conformation in peptide chains

In the previous paper on Z-GPLG, comparisons of conformations of amino acid residues and peptide chains with those of cyclohexaglycyl, Ferrichrome A, lysozyme, carboxypeptidase and myoglobin were described, and we pointed out some common features of the conformation in these structures. The present structure analysis makes it possible to discuss this point further. In the present case, a comparison was made by drawings of the peptide chains projected on a special plane, and also by the internal rotation angles, as shown in Fig. 8(a), (b), (c), and (d) for Z-GPLGP, cyclohexaglycyl, Z-GPLGP and Ferrichrome A, respectively. It is surprising to note that the Z-GPLGP and Z-GPLG representations can be

Table 5. *Best planes, dihedral angles between the planes, and displacements from the planes*

(a) Equations of the best planes†

I	$0.0234X - 0.9533Y - 0.3011Z = -0.9018$	peptide Z–gly(1)
II	$-0.2531X + 0.4457Y - 0.8587Z = -0.2376$	peptide gly(1)–pro(1)
III	$0.9963X + 0.0835Y + 0.0215Z = -7.1424$	peptide pro(1)–leu
IV	$-0.2894X + 0.9520Y - 0.0998Z = -0.0828$	peptide leu–gly(2)
V	$0.3892X - 0.3122Y + 0.8666Z = -3.2894$	peptide gly(2)–pro(2)
VI	$-0.8302X - 0.1873Y + 0.5250Z = -0.5366$	carboxyl pro(2)
VII	$-0.5138X - 0.1229Y - 0.8491Z = -1.6199$	phenyl group
VIII	$-0.2195X + 0.3611Y - 0.9063Z = -0.4971$	pro(1) ring
IX	$0.2704X - 0.3730Y + 0.8875Z = -3.2299$	pro(2) ring

† Coordinates (X, Y, Z) in Å are referred to the orthogonal axes, a, b, and c^*.

(b) Dihedral angles (°) between the planes

	I	II	III	IV	V	VI
I	—	100	94	152	87	90
II		—	104	54	169	109
III			—	102	68	146
IV				—	120	89
V					—	79

(c) Displacements (Å) of atoms from the planes

I		II		III	
Z–O(1)	0.087	Gly(1)C$_\alpha$	0.004	Pro(1)C$_\alpha$	−0.014
Z–C(8)	−0.019	Gly(1)C$_\beta$	−0.006	Pro(1)C′	0.018
Z–O(2)	−0.023	Gly(1)O	0.000	Pro(1)O	−0.002
Gly(1)N	−0.087	Pro(1)N	0.007	LeuN	0.012
Gly(1)C$_\alpha$	0.061	Pro(1)C$_\delta$	−0.004	LeuC$_\alpha$	−0.013
Z–C(7)*	−0.003	Pro(1)C$_\alpha$*	−0.259	Pro(1)N*	0.579
Gly(1)C$_\beta$*	1.434	Gly(1)N*	0.343	Pro(1)C$_\beta$*	−1.578
		Pro(1)C′*	0.842	LeuC$_\beta$*	−0.977
		Pro(1)C$_\beta$*	−0.116	LeuC$_\gamma$*	−2.453

IV		V		VI	
LeuC$_\alpha$	0.007	Gly(2)C$_\alpha$	−0.003	Pro(2)C$_\alpha$	0.018
LeuC′	0.012	Gly(2)C$_\beta$	0.001	Pro(2)C′	−0.036
LeuO	−0.017	Gly(2)O	−0.006	Pro(2)O(2)	0.012
Gly(2)N	−0.024	Pro(2)N	0.029	Pro(2)O(1)	0.010
Gly(2)C$_\alpha$	0.026	Pro(2)C$_\alpha$	−0.010	Pro(2)N*	0.455
LeuN*	0.167	Pro(2)C$_\delta$	−0.011		
LeuC$_\beta$*	1.151	Gly(2)N*	0.011		
Gly(2)C$_\beta$	1.282	Pro(2)C$_\beta$*	0.303		

VII		VIII		IX	
Br	0.022	Pro(1)N	0.096	Pro(2)N	0.084
Z–C(1)	0.022	Pro(1)C$_\alpha$	−0.053	Pro(2)C$_\alpha$	−0.079
Z–C(2)	−0.052	Pro(1)C$_\beta$	0.014	Pro(2)C$_\beta$	0.048
Z–C(3)	−0.018	Pro(1)C$_\delta$	−0.050	Pro(2)C$_\delta$	−0.049
Z–C(4)	−0.006	Pro(1)C$_\gamma$*	−0.407	Pro(2)C$_\gamma$*	−0.493
Z–C(5)	0.036	Gly(1)C′*	0.110	Pro(2)C′*	−1.449
Z–C(6)	0.084	Gly(1)O*	0.239	Gly(2)O*	0.299
Z–C(7)	−0.078	Pro(1)C′*	1.157	Gly(2)C′*	0.234
Z–O(1)*	1.225	Pro(1)O*	0.985	Pro(2)O(2)*	−2.266
		Gly(1)C$_\alpha$*	0.002	Gly(2)C$_\alpha$*	0.342
				Pro(2)O(1)*	−1.464

*Atoms not included in the calculation of the best planes.

throughly overlapped with each other, even with respect to the side chain of the Leu residue. To make the comparison clearer, the internal rotation angles of the peptide chains were accumulated (Table 6) together with the calculated values in the similar peptide system (Venkatachalam, 1968). In Table 6, the (φ_2, ψ_2) and (φ_3, ψ_3) values are the internal rotation angles of the
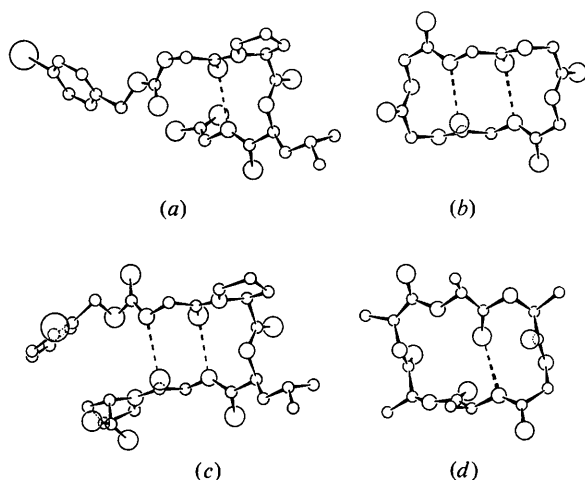
Fig. 9. Interatomic short contacts (less than 3·5 Å) in the chain.

second and third residues. C_α^1–C_α^4 is the distance between the C_α atoms of the first and fourth residue in the chain. In the cases of cyclic peptides, the equivalent estimations as mentioned before were made.

The structures of the peptide molecules so far established by the X-ray diffraction method can be classified in two groups as in Table 6. In the peptide molecules in Group I such as Z-GPLGP, Z-GPLG, cyclohexaglycyl and $\ulcorner(Gly)_4$-$(D$-$Ala)_2\urcorner$, the average values of (φ_2, ψ_2) and $\overline{(\varphi_3, \psi_3)}$ are (113, 152°) and (82, 188°). Although the latter two molecules consist mainly of a particular amino acid residue (glycine which does not have a C_β atom), it seems reasonable to include these molecules in group I. We call this conformation 'α-U-folding'. The conformation of CysH-Pro-Leu-GlyNH$_2$ (Low et al., 1969) seems to belong to this group, though details cannot yet be presented. Values (φ_2, ψ_2) in group I are those between the right-handed α-helix (3·6$_{13}$) and the δ-helix (3·0$_{10}$), while the (φ_3, ψ_3) values are not observed in any well-defined conformations. These observed values agree roughly with calculated values. However, the distance, C_α^1–C_α^4, is observed to be about 5·0 Å for Z-GPLGP and Z-GPLG and quite different from the calculated value, which is too small. This difference is probably due to the accumulation of small differences in the internal rotation angles. The same effect also occurs in cyclohexaglycyl.

We call the structure of Ferrichrome A a 'β-U-folding' conformation, in which (φ_2, ψ_2) and (φ_3, ψ_3) are about (120, 310°) and (260, 280°). In this group, the (φ_2, ψ_2) values agree well with those in collagen and polyglycine II, and the (φ_3, ψ_3) values agree with those in γ-helix (5·1$_{17}$). However, more experimental values are necessary to define the (φ_2, ψ_2) and (φ_3, ψ_3) values of this conformation. The C_α^1–C_α^4 distance in Ferrichrome A is too big for this type of conformation, and is probably due to the steric deformation by the rest of the residues in the ring.

The difference between these two folding conformations is seen in the ψ_2 and φ_3 values, both of which

Fig. 7. Internal rotation angles of the C_α–C' bond of proline residue. (a) Pro(1) and (b) Pro(2) residues in the present molecule; (c) that in tetrapeptide; (d) model used in the collagen structure.

Fig. 8. Drawings of the molecules projected on the special plane, defined by the atoms, C' of Gly(1), C_α of Pro(1) and N of the leu residue. (a) Z-GPLG molecule; (b) cyclohexaglycyl; (c) Z-GPLGP molecule; (d) Ferrichrome A (the backbone of the ring).

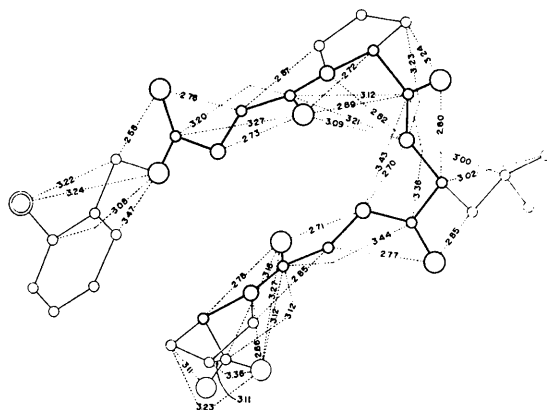Table 6. *Parameters of the folding conformation*

	φ_2	ψ_2	φ_3	ψ_3	N–H---O	C_α^1–C_α^4
Group I						
Z–GPLGP	115°	153°	75°	188°	3·00 Å	5·10 Å
Z–GPLG	122	147	76	188	2·97	4·97
Average	119	150	76	188	2·98	5·04
Calculated*	120	150	70	200	2·85	4·40
Cyclohexaglycyl	111	151	86	188	2·96	5·16
	111	150	88	184	3·02	
	111	150	85	187	3·03	5·16
	112	149	87	188	3·09	
Average	111	150	87	187	3·03	5·16
Calculated*	110	150	90	190	3·10	4·73
⌐(Gly)₄–(D–ala)₂⌐	110	165	74	196	3·04†	
	–114	–165	–49	–211	3·16‡	
Average value§	113	152	82	188	3·02	
Group II						
Ferrichrome A††	123°	312	262	279	2·98	5·61
Calculated*	120	310	260	180	2·84	5·04

* Calculated values are those given by Venkatachalam (1968).
† (–Gly–Gly–) part in ⌐(Gly)₄–(D–Ala)₂⌐ .
‡ (–D–Ala–D–Ala–) part in ⌐–(Gly)₄–(D–Ala)₂–⌐ .
§ In calculating these average values, the values of the (–D–Ala–D–Ala–) part in the ⌐(Gly)₄–(D–Ala)₂–⌐ molecule are excluded.
†† Sequence is ⌐Orn–Orn–Orn–Ser–Ser–Gly⌐ and the folding part is (–Ser–Gly–).

differ by 180° as pointed out in the description of structures (1) and (4) in the previous section. The conformation, α-U-folding, is observed in myoglobin and lysozyme (observed values in them are not accurate enough to discuss in detail), and the β-U-folding is observed in lysozyme.

The comparison revealed another interesting feature: the chain of Z-GPLGP can be completely overlapped on the chain of cyclohexaglycyl to the extent of the second hydrogen bond. Agreement of the internal rotation angles in this region of the chain is very good. Considering that two hydrogen bonds in cyclohexaglycyl are equivalent bonds related by the centre of symmetry and only half of the molecule (three residues) is independent, the consistency of the chain with that in Z-GPLGP is of great interest. It is probably correct to consider that the conformation of Z-GPLGP is a stable and characteristic structure with two intrachain hydrogen bonds and with special values of internal rotation angles.

The structure of gramicidin S was proposed on the basis of the α-helix and β-structure by Schmidt et al. in 1957, but it seems interesting for us to re-examine its structure taking into account these folding structures. This specific conformation must have a relation to the folding part of the cross-β-structure. Furthermore, this type of conformation at the folding part of the peptide, observed in several small oligopeptides and even in proteins, may be reasonably classified as a term like α-helix or β-structure.
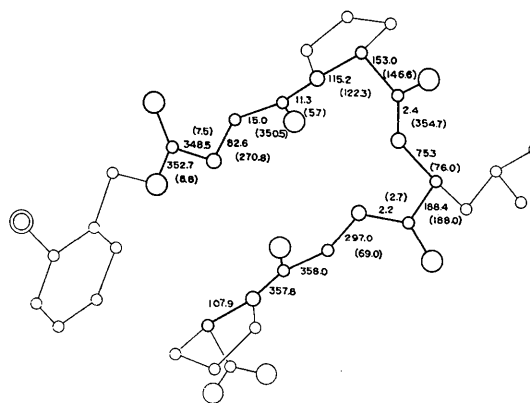
Fig. 10. Internal rotation angles of peptide chains. Definitions for these angles are those given by Edsall *et al.* (1966). The values in parentheses are those of Z–GPLG.

References

Ashida, T. (1967). *The Universal Crystallographic Computing System* I, p. 65. Edited by T. Sakurai. Tokyo: Japan Crystallographic Association.

Ashida, T., Ueki, T., Tsukihara, T., Sugihara, A., Takano, T. & Kakudo, M. (1971). To be published.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. Roy. Soc.* B**167**, 365.

COREY, R. B. & PAULING, L. (1953). *Proc. Roy. Soc.* B**141**, 10.

COWAN, P. M. & McGAVIN, S. (1955). *Nature, Lond,* **176**, 501.

EDSALL, J. T., FLORY, P. L., KENDREW, J. C., LIQUORI, A., NÉMETHY, J. T., RAMACHANDRAN, G. N. & SCHERAGA, H. A. (1966). *J. Mol. Biol.* **15**, 399.

FRIDRICHSONS, J. & MATHIESON, A. McL. (1962). *Acta Cryst.* **15**, 569.

HAMAGUCHI, K. (1970). Private communication.

HAMILTON, W. C. (1959). *Acta Cryst.* **12**, 609.

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

KAKUDO, M., SASADA, Y., KATSUBE, Y., SAKAKIBARA, S. & AKABORI, S. (1963). *Aspects of Protein Structure,* p. 69. Edited by G. N. RAMACHANDRAN. London and New York: Academic Press.

KARLE, I. L., GIBSON, J. W. & KARLE, J. (1970). *J. Amer. Chem. Soc.* **92**, 3755.

KARLE, I. L. & KARLE, J. (1963). *Acta Cryst.* **16**, 969.

KENDREW, J. C. (1963). *Science,* **139**, 1259.

LEUNG, Y. C. & MARSH, R. E. (1958). *Acta Cryst.* **11**, 17.

LOW, B. W., LOVELL, F. M. & RUDKO, A. D. (1969). *Acta Cryst.* A**25**, (S3), S188.

NAGAI, Y. & NODA, H. (1959). *Biochim. Biophys. Acta,* **34**, 298.

NAGAI, Y., SAKAKIBARA, S., NODA, H. & AKABORI, S. (1960). *Biochim. Biophys. Acta,* **37**, 567.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1965). *Biochim. Biophys. Acta,* **109**, 314.

SAKAKIBARA, S. & NAGAI, Y. (1960). *Bull. Chem. Soc. Japan,* **33**, 1537.

SASADA, Y. & KAKUDO, M. (1961). *Acta Cryst.* **14**, 1305.

SASADA, Y., TANAKA, K., OGAWA, Y. & KAKUDO, M. (1961). *Acta Cryst.* **14**, 326.

SCHMIDT, G. M. J., HODGKIN, D. C. & OUGHTON, B. M. (1957). *Biochem. J.* **65**, 744.

UEKI, T., ASHIDA, T., KAKUDO, M., SASADA, Y. & KATSUBE, Y. (1969). *Acta Cryst.* B**25**, 1840.

VENKATACHALAM, C. M. (1968). *Biopolymers,* **6**, 1425.

ZALKIN, A., FORRESTER, J. D. & TEMPLETON, D. H. (1966). *J. Amer. Chem. Soc.* **88**, 1810.

# The Structure of 11,11-Difluoro-1,6-methano[10]annulene

BY CARLO MARIA GRAMACCIOLI AND MASSIMO SIMONETTA

Istituto di Chimica Fisica, Università, Via Saldini 50, 20133 Milano, Italy

The crystal and molecular structure of 11,11-difluoro-1,6-methano[10]annulene has been determined and refined by least-squares methods. The crystals are orthorhombic, space group $Pna2_1$, with $a = 9\cdot233$, $b = 13\cdot235$, $c = 7\cdot055$ Å and four molecules per unit cell. Intensity data were collected visually from Weissenberg photographs. The solution of the phase problem was readily obtained by finding the orientation of a plausible molecular model with respect to the crystallographic axes through a systematic search of the Patterson synthesis around the origin; the position of the molecule in the unit cell was then found by molecular packing energy calculations. After refinement by least-squares, the final R index is $0\cdot080$. The presence of fluorine atoms does not seem to affect the conformation of 1,6-methano[10]annulene; the distance C(1)–C(6) is 2·25 Å, excluding any 'double norcaradiene' character for this compound, in agreement with chemical and spectroscopic evidence.

Introduction

In line with a recent interest in aromatic systems with $4n + 2$ π-electrons (*n* greater than 1) the chemistry of aromatic compounds with 10 π-electrons has been extensively developed by Vogel and co-workers (Vogel, 1967; 1968*a,b*). An interesting property of 1,6-methano[10]annulene is the influence of substituents in the bridge (position 11) on the stability of the aromatic system (I). Chemical and spectroscopic (n.m.r.) evidence are in favour of the existence of a direct C(1)–C(6) bond for the 11,11-dimethyl derivative which is consequently non-aromatic and has the so-called bisnorcaradiene structure (II).

The situation for the 11,11-difluoro derivative seems

(I)   (II)
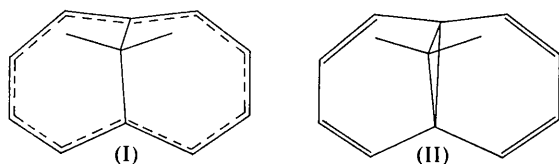
to be the opposite and the substance has the chemical character of an aromatic compound. An investigation of the structure of these two substances by X-ray diffraction has been suggested to us by Prof. Vogel.* As a first result, the structure of 11,11-difluoro-1,6-methano[10]annulene has been solved and refined. Although the extensive thermal vibration does not permit a very high precision in determining the molecular geometry at room temperature, it is felt that the method here adopted for solving the structure might be interesting and the results are anyway fully adequate to prove the aromaticity of this compound.

Experimental

Crystals of 11,11-difluoro-1,6-methano[10]annulene are in the form of greenish needles, elongated along **c**.