Solid-State Conformation of Monodisperse Homo-L-oligopeptides: An X-ray Diffraction Study[1]

Antonio Del Pra and Claudio Toniolo*

Biopolymer Research Centre, C.N.R., Institute of Organic Chemistry, University of Padova, 35100 Padova, Italy. *Received February 14, 1978*

ABSTRACT: Observations of extended peptide chains, whose direction is perpendicular to the fiber axis (cross-β structures), have been hitherto confined to fibrous proteins and to some synthetic polydisperse polypeptides of rather low molecular weight. This structure has now been found in some monodisperse linear homooligopeptides with aliphatic hydrocarbon side chains. X-ray diffraction photographs of t-Boc-(L-X)$_n$-OMe (X = Ala, Val, Leu, and n = 5, 6, 7) show the characteristic reflections of this form. In addition, the good orientation of suitably prepared specimens has enabled a fairly complete determination of the unit cell of the pentapeptides to be made. t-Boc-(L-Ala)$_5$-OMe molecules are packed in a monoclinic lattice with a = 4.80 Å, b = 28.0 Å, c = 5.82 Å, and β = 65.6°. On the contrary, orthorhombic unit cells are proposed for t-Boc-(L-Val)$_5$-OMe and t-Boc-(L-Leu)$_5$-OMe with dimensions a = 4.80 Å, b = 43.80 Å, c = 18.60 Å and a = 9.60 Å, b = 37.52 Å, c = 11.90 Å, respectively. As regards chain orientation, an appreciable proportion of the structure of the Leu pentapeptide is in the antiparallel arrangement, while from the available experimental data of the two remaining pentapeptides it has not been possible to establish whether the arrangement of the chains within the sheets is parallel or antiparallel.

In recent years we have performed a solid-state conformational analysis of a number of monodisperse all-L linear homooligopeptide series (to the heptapeptide) derived from N-unsubstituted α-amino acids using infrared absorption and vacuum–ultraviolet circular dichroism.[2-5] From the combined results strong evidence was found that the ordered secondary structure adopted by these peptides is the pleated sheet β structure, the content of which in the molecule increases with increasing chain length. This finding is not surprising, since it is well established that in the case of the α-amino acid residues, the homopolypeptides of which assume the α-helical conformation, even the heptapeptides have a chain length below the critical one for the formation of this structure in the solid state.[2,3] Also, it was shown that the type of ordered structure formed does not depend upon the conditions of preparation of the solid samples. It was additionally suggested that in the β structure the relative orientation of peptide chains that is favored, either parallel or antiparallel, is determined by two structural characteristics of the side chain, its overall bulkiness and the presence of branching at the β-carbon atom. Both features were indicated to favor the parallel-chain arrangement.

The present paper describes the results of our X-ray diffraction study of t-Boc-(L-X)$_n$-OMe, where X = Ala, Val, Leu and n = 5, 6, 7.[6] In addition, a detailed structural investigation of the three homopentapeptides is reported.

Experimental Section

Materials. The synthesis of the monodisperse homooligopeptides derived from L-alanine, L-leucine, and L-valine has been carried out according to procedures already published.[7-9] All the peptides are chemically and optically pure.

Methods. Films of t-Boc-(L-X)$_n$-OMe, where X = Ala, Val, Leu and n = 5, 6, 7, were obtained by stroking out a drying film until it was solid; fibers were also drawn by means of claws from a viscous solution of the sample. The solvent used was 2,2,2-trifluoroethanol. Films and fibers were then dried, in vacuo, for several hours to remove solvent molecules. The density was determined by the flotation method using benzene–carbon tetrachloride mixtures allowing the sample to reach the equilibrium position. Films and fibers were examined in a flat camera with sample to photographic film distance, d, variable from 4 to 10 cm, with Ni filtered Cu K$\bar{\alpha}$ radiation, from a fine and normal focus X-ray tube and a 0.6-mm pinhole. Tilted specimen and equi-inclination techniques were also used. Low-angle reflections were collected on a Rigaku Denky camera. Molecular models were built from CPK space-filling components (1 Å = 1.25 cm).

Results and Discussion

Diffraction pictures have shown that high crystallinity is present in all examined compounds, which also proved to be sufficiently oriented for a detailed analysis. As representative examples, the X-ray photographs of t-Boc-(L-Ala)$_5$-OMe, t-Boc-(L-Val)$_5$-OMe, and t-Boc-(L-Leu)$_5$-OMe are presented in Figures 1, 2, and 3, respectively. Very similar patterns are exhibited by their higher homologues. From the experimental data clear evidence was obtained that their structure is of the cross-β type.[10] Specimens prepared as described in the Experimental Section, mounted with the stroking direction horizontal, give diffraction photographs whose main features are a strong arc on the meridian and a strong reflection on the equator. The former reflection, the spacing of which is in the region 4.4–4.8 Å, is related to the distance between hydrogen-bonded peptide chains in a sheet (a axis of the unit cell). The latter reflection is related to the distance between consecutive sheets (c axis): its values are 5.30, 9.30, and 11.90 Å for the oligopeptides derived from L-alanine, L-valine, and L-leucine, respectively. The b axis, which depends on the nearly extended chain repeat, is perpendicular to the direction of stroking. The peptide chains are relatively so short that their side-to-side aggregation leads to the formation of a crystallite which is longest in the hydrogen-bond direction. Furthermore, the very diffuse nature of the reflections in which k is not zero supports the view that the crystallites are poorly developed along b.

The oriented fibers contain ribbon-like micelles with their longest dimension parallel to the fiber axis. The ribbons are packed face to face with a variable interfacial separation, and in this space some kind of amorphous matrix or solvent molecules are located. There are no sharp equatorial maxima in the small-angle region of the photographs; however, for each compound an intense, broad, and unresolved maximum is observed at about 80–100 Å, and this could be the first order of the mean separation of the micelles. Now we report a detailed structure analysis of the three homo-L-pentapeptides derived from an accurate inspection of their X-ray pictures.

***t*-Boc-(L-Ala)$_5$-OMe.** The spacings, d_0, and eye-estimated intensities, I_0, of the observed reflections are reported in Table I. The 5.30-Å equatorial reflection, which is related to the intersheet distance, is present also on the meridian. This suggest that the sample would consist at least of: (i) domains where the b axis of the unit cell is preferentially oriented perpendicular to the plane of the specimen, and (ii) domains where the c axis is perpendicular to the same plane. In both cases the a axis is preferentially oriented parallel to the stroking direction. We found that all observed reflections can be satisfactorily indexed according to a monoclinic unit cell with a = 4.80 Å, b = 28.0 Å, c = 5.82 Å, and β = 65.6°. We note

0024-9297/78/2211-0793$01.00/0

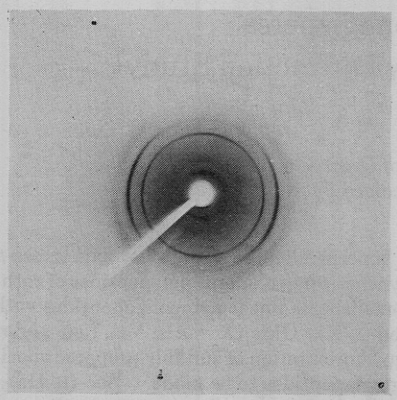

Figure 1. Normal beam X-ray diffraction pattern of *t*-Boc-(L-Ala)$_5$-OMe oriented film: flat camera, d = 6 cm, Cu K$\bar{\alpha}$ radiation; *t*-Boc-(L-Ala)$_5$-OMe film and stroking direction horizontal.

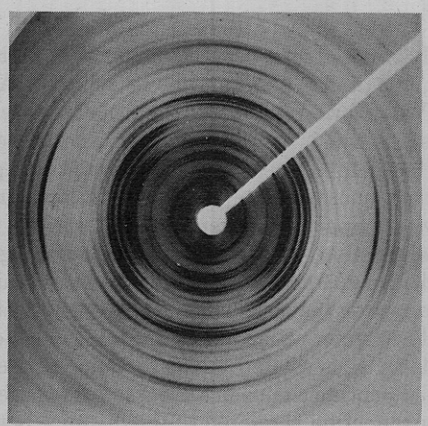

Figure 2. Normal beam X-ray diffraction pattern of *t*-Boc-(L-Val)$_5$-OMe oriented film: flat camera, d = 6 cm, Cu K$\bar{\alpha}$ radiation; *t*-Boc-(L-Val)$_5$-OMe film and stroking direction horizontal.

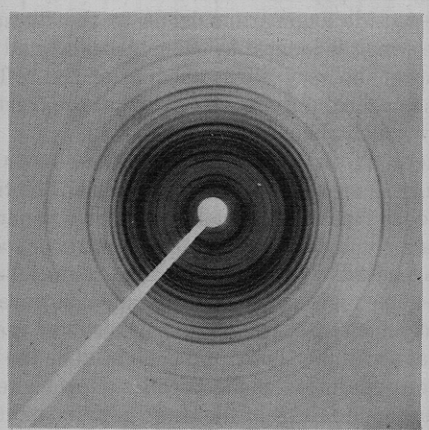

Figure 3. Normal beam X-ray diffraction pattern of *t*-Boc-(L-Leu)$_5$-OMe oriented film: flat camera, d = 6 cm, Cu K$\bar{\alpha}$ radiation; *t*-Boc-(L-Leu)$_5$-OMe film and stroking direction horizontal.

that the a and c axes have values similar to those found in β-poly-L-alanine.[11]

As regards the type of pleated sheet present (parallel or antiparallel chain), there are no reflections requiring a to be doubled, but it does not necessarily follow that the parallel-chain type is appropriate. For instance, the molecular model indicates that the length of a single antiparallel chain of a protected pentapeptide with fully extended end groups is about 26.5 Å. Nevertheless, taking into account two neighboring chains of the same sheet, this distance increases to about 28 Å, as a consequence of the steric hindrance of the bulky *t*-Boc groups. This latter value, which is higher than that obtained from a parallel-chain molecular model, agrees well with the experimental data. We believe, ultimately, that both these chain arrangements can be fitted into the available space. In addition, a random sheet arrangement, which has been already suggested for poly-L-alanine in the β form,[11] would also be possible. Another possibility is that the structure would be heterogeneous, either consisting of sheets with mixed chain systems or of mixed parallel and antiparallel sheets.

As far as the mode packing of the pleated sheets is concerned, the necessity of indexing the 4.37 and 2.40 Å reflections as 101 and 201, respectively, eliminates the possibility of an orthogonal or near-orthogonal chain arrangement of the type suggested by Warwicker for silk fibroin.[12] In the packing arrangement proposed by Marsh et al.[13] for the crystalline regions of silk fibroin the chain axes in adjacent sheets are displaced by $a/2$ as in Figure 4, and a displacement of this order must also be present in *t*-Boc-(L-Ala)$_5$-OMe to account for the intensity of the 101 and 201 reflections.

The sharpness of the $h0l$ reflections compared to those in which k is not zero suggests some disorder (or a shortening of the micelles) along the b axis; this can be due to the presence of the bulky N-blocking group which can hinder a good packing between neighboring sheets. Consequently, a decrease of crystalline coherence along the b axis must be expected,

Table I
X-ray Data for *t*-Boc-(L-Ala)$_5$-OMe (Monoclinic Unit Cell with a = 4.80, b = 28.0, c = 5.82 Å, and β = 65.6°)

hkl	d_c, Å	d_0, Å	I_0
	Equator		
010	28.00	27.2–29.3	vs
020	14.00	13.8–14.2	s
030	9.33	9.0–9.3	vw
001	5.30	5.30	s
090 (171)	3.11 (2.95)	2.93–3.11	m
002	2.65	2.67	m
	First-Layer Line		
101	4.37	4.37	vs
141, 131	3.71, 3.96	3.7–3.9	ms
	Second-Layer Line		
201	2.40	2.40	mw
241	2.27	2.28	w
220, 222, 230	2.16, 2.16, 2.13	2.15	w
232, 261	2.13, 2.13		
	Third-Layer Line		
301, 302, 311 312	1.58	1.59	w

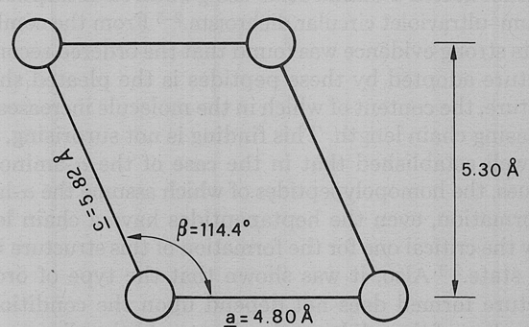

Figure 4. Diagrammatic representation of the b axis projection of the unit cell in the cross-β form of *t*-Boc-(L-Ala)$_5$-OMe. The direction of stroking is parallel to a.

Table II
X-ray Data for *t*-Boc-(L-Val)$_5$-OMe (Orthorhombic Unit Cell with a = 4.80, b = 43.80, and c = 18.60 Å)

hkl	d_c, Å	d_0, Å	I_0
	Equator		
020	21.90	21.9	s
040	10.95	10.9	m
002	9.30	9.30	vvs
050	8.76	8.75	ms
051	7.93	8.02	w
060	7.30	7.38	w
023	5.96	5.97	s
033	5.71	5.69	m
080	5.48	5.48	s
090	4.86	4.85	m
004	4.65	4.67	m
024	4.55	4.56	w
0,10,1	4.26	4.23	vw
0,10,2, 064	3.96, 3.92	3.94	vw
0,13,0, 094, 0,11,3	3.37, 3.36, 3.35	3.35	m
075, 0,13,1, 065	3.20, 3.31, 3.31	3.26	mw
{006, 0,14,0 / 0,12,3}	{3.10, 3.13 / 3.14}	3.13	vs
026, 036, 0,11,4	3.07, 3.03, 3.02	3.04	w
066, 0,12,4, 0,10,5	2.85, 2.87, 2.84	2.85	w
0,15,3	2.64	2.65	vvw
096, 0,16,2	2.61, 2.62	2.62	m
0,14,5, 0,18,1	2.39, 2.41	2.39	m
0,12,6, 0,16,4, 0,18,2	2.36, 2.36, 2.35	2.34	vvw
0,15,6, 0,20,2	2.13, 2.13	2.13	w
0,19,4, 0,21,0, 0,21,1	2.07, 2.08, 2.07	2.08	w
	First-Layer Line		
101, 120, 111	4.65, 4.69, 4.62	4.65	vvs
131	4.43	4.41	ms
140	4.39	4.38	w
141	4.28	4.28	w
151	4.11	4.12	vvw
142, 160	3.97, 4.01	3.98	vs
103, 152, 170	3.80, 3.83, 3.81	3.80	vvw
123, 171	3.74, 3.73	3.74	s
180	3.61	3.60	s
190	3.42	3.42	vw
104, 114	3.34, 3.33	3.33	w
1,10,1, 154, 183	3.19, 3.12, 3.12	3.16	w
105, 174, 193	2.94, 2.95, 2.99	2.96	mw
175, 1,13,1	2.66, 2.73	2.69	w
{186, 1,11,5, 1,13,4 / 1,16,0, 1,16,1}	{2.35, 2.36, 2.37 / 2.38, 2.36}	2.36	w
{196, 1,12,5, 1,14,4 / 1,16,2}	{2.30, 2.29, 2.28 / 2.30}	2.29	mw
{1,12,6, 1,17,3, 1,16,4 / 1,14,5}	{2.12, 2.13, 2.12 / 2.14}	2.13	m
1,19,0, 1,19,1, 1,15,5	2.08, 2.07, 2.07	2.08	m
1,21,0, 1,21,1	1.91, 1.90	1.91	mw
1,21,2, 1,18,5	1.87,1.87	1.86	w
1,23,1, 1,22,3, 1,20,5	1.76, 1.76, 1.75	1.75	w
	Second-Layer Line		
200	2.40	2.40	s
{202, 251, 232 / 242, 260}	{2.32, 2.30, 2.29 / 2.27, 2.28}	2.29	ms
{262, 233, 280 / 271}	{2.21, 2.21, 2.20 / 2.22}	2.21	ms
{282, 204, 291 / 290, 263, 214}	{2.14, 2.13, 2.14 / 2.15, 2.14, 2.13}	2.14	m
{205, 215, 274 / 293}	{2.02, 2.01, 2.02 / 2.03}	2.02	m
{2,12,2, 294, 255 / 2,13,0}	{1.96, 1.95, 1.96 / 1.96}	1.96	vw
{2,14,3, 2,12,4, 2,10,5 / 266}	{1.82, 1.84, 1.83 / 1.84}	1.83	vvw
	Third-Layer Line		
{320, 330, 301 / 311, 321, 300}	{1.60, 1.59, 1.59 / 1.59, 1.59, 160}	1.59	m
{303, 313, 352 / 370, 342}	{1.55, 1.55, 1.55 / 1.55, 1.56}	1.55	mw
{305, 325, 374 / 3,11,2, 3,12,0}	{1.47, 1.47, 1.47 / 1.47, 1.46}	1.47	vw

which is also in agreement with the presence of few broad $0k0$ reflections.

The calculated crystal density is $D_c = n \cdot 1.13$ g cm^{-3}, where n is the number of peptide chains crossing the monoclinic cell. Comparison with the measured density (D_o = 1.23 g cm^{-3}) indicates that such a cell is crossed by one β chain only. The difference between the observed and the calculated density is undoubtedly outside the range of the experimental error and can be only explained by the presence of solvent molecules in the borderline surface of the lamellar crystallites since stringent drying conditions have not been used.

***t*-Boc-(L-Val)$_5$-OMe.** The spacings, d_0, and eye-estimated intensities, I_0, of the observed reflections are listed in Table II. With these data we tried to define the unit cell dimensions. The equatorial reflection at about 21.9 Å can be directly related to the molecular chain length (b axis). Three well-developed layer lines, which appear mainly on the equi-inclination X-ray photographs, indicate that the crystallographic repeat along the meridian is 4.8 Å, i.e., twice the spacing of the 200 strong meridional reflection. Finally, the intersheet distance is 9.30 Å as in poly-L-valine.[14]

All observed reflections can be satisfactorily indexed according to an orthorhombic unit cell with a = 4.80 Å, b = 43.80 Å, c = 18.60 Å. The length of the b axis is approximately twice the normal molecular chain length indicating that the two chains of the asymmetric unit are different from a crystallographic point of view. The reason for giving the c axis the value of 18.60 Å instead of 9.30 Å (intersheet spacing) comes from the observation of several reflections with l odd, which can appear only if the repeat of pattern of electron density is 18.60 Å (or a multiple of this length). This implies that consecutive hydrogen-bonded sheets appear different, as in poly(β-*n*-propyl L-aspartate)[15] and in poly-*S*-carbobenzoxy-L-cysteine,[16] where it has been suggested that the side chains are highly ordered and alternate sheets have different side-chain conformation. There are no reflections requiring a to be doubled, but again it does not necessarily follow that the parallel chain is appropriate. It is our contention that the conclusions described above for the alanine pentapeptide can be applied also to this case.

By means of CPK space-filling components it is possible to show that the best packing is obtained when: (i) the *t*-Boc and OMe groups are bent with respect to the backbone direction (Figure 5a), and (ii) the chain axes of a sheet are displaced by $a/2$ with respect to the first neighboring sheets (Figure 5b). This lattice, which is in agreement with the experimental data and in particular with the very high intensity of the 101 reflection and with the absence of the 100 reflection, would be favored by the relatively bulky isopropyl side-chain groups too.

The calculated crystal density is $D_c = n \cdot 0.266$ g cm^{-3}, where n is the number of peptide chains in the orthorhombic cell. The comparison with the observed density, D_o = 1.08 g cm^{-3}, indicates that such a cell is crossed by four β chains.

***t*-Boc-(L-Leu)$_5$-OMe.** The spacings, d_0, and eye-estimated intensities, I_0, of the observed reflections are reported in Table III. The equatorial reflection at 18.76 Å can be directly related to the peptide chain length. Using CPK space-filling compo-

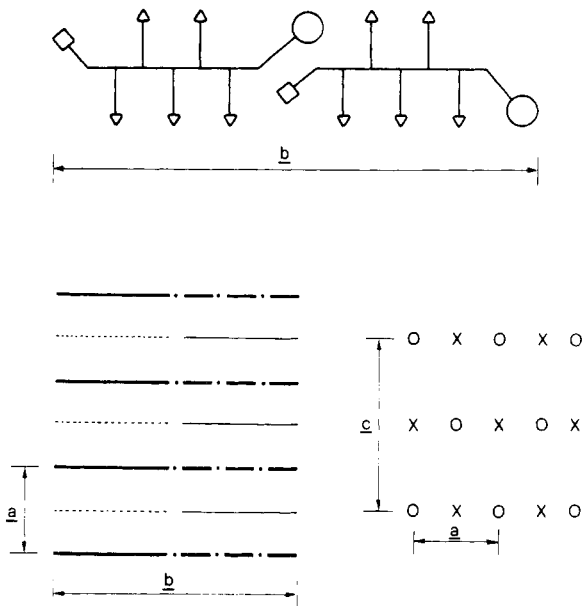

Figure 5. Schematic representation of the main features of the packing arrangement in the cross-β form of t-Boc-(L-Val)$_5$-OMe. The direction of stroking is parallel to a. (a) (top) This diagram brings into evidence the bending of the end groups of two pentapeptide molecules belonging to adjacent sheets. The chain axes of one sheet are displaced by $a/2$ with respect to those of the other one. Symbols used: O, t-Boc; □, OMe; △, isopropyl side chain. (b) (bottom) (1) Each line represents a pentapeptide chain, while a set of lines of the same kind represents a sheet. The nearest neighboring sheets represented by the thick and thin full lines, respectively, lie on the ab plane. On the contrary, the sheets represented by the dotted and broken lines lie in a plane parallel to ab, displaced from it by $c/2$. (2) Crosses and circles represent pentapeptide chains projected down b. In the crystal the chains represented by crosses are displaced by $b/2$ with respect to the chains represented by circles.

nents the antiparallel molecular model of t-Boc-(L-Leu)$_5$-OMe has been built up. Its length, which is about 22 Å, assuming a partial bending of the end groups, is noticeably greater than the observed spacing at 18.76 Å. Therefore, adjoining pleated sheets must interpenetrate themselves along the peptide chain direction in order to explain the experimental data; this interaction could be favored by the bulky side-chain isobutyl groups. An important feature of the equi-inclination X-ray photographs is a sharp meridional reflection at 3.20 Å, which must be indexed as 300.

In conclusion, we found that all observed reflections can be satisfactorily indexed according to an orthorhombic unit cell with a = 9.60 Å, b = 37.52 Å, and c = 11.90 Å. The length of the a axis is approximately twice the normal distance between adjacent chains in a β structure and indicates that they are antiparallel or that an appreciable proportion of the structure has an antiparallel orientation, while a relatively small number of parallel chains may, of course, occur in this material as faults in otherwise antiparallel sheets. The b value is about twice the expected length of a chain and can be justified by the above discussed mutual interpenetration. Finally, the intersheet distance, which is represented by the c value, is in agreement with that found in poly-L-leucine.[14]

As far as the mode of packing of the pleated sheets is concerned, the great intensity of 200 and 001 reflections shows that all chains lie on or very near the lattice points formed by the simple cell of sides $a/2$ and c. This type of packing is similar to that suggested by Warwicker for silk fibroin.[12]

The calculated density for n chains in the unit cell is $D_c = n \cdot 0.270$ g cm^{-3}. A comparison with the observed density, D_o = 1.07 g cm^{-3}, indicates that the orthorhombic cell is crossed by four peptide molecules.

Table III
X-ray Data for t-Boc-(L-Leu)$_5$-OMe (Orthorhombic Unit Cell with a = 9.60, b = 37.52, c = 11.90 Å)

hkl	d_c, Å	d_o, Å	I_o
	Equator		
020	18.76	18.76	m
001	11.90	11.90	s
011	11.34	11.33	vs
021	10.05	10.08	mw
031	8.62	8.62	m
041	7.36	7.35	w
051	6.34	6.33	m
002	5.95	5.95	w
022	5.67	5.65	vw
061	5.53	5.52	s
003	3.97	3.97	ms
0,11,1	3.28	3.28	m
0,12,0	3.13	3.13	m
004, 0,14,0, 0,11,2	2.97, 2.96, 2.96	2.96	m
0,10,3	2.72	2.73	w
074, 0,13,2	2.60, 2.60	2.61	vw
0,15,1, 0,14,2, 0,12,3	2.45, 2.44, 2.45	2.46	m
005	2.38	2.38	vw
035, 0,10,4, 0,13,3 0,16,0	2.34, 2.33, 2.33 2.34	2.33	mw
065, 0,14,3, 0,17,0	2.22, 2.22, 2.21	2.22	vw
085, 0,12,4, 0,15,3	2.12, 2.15, 2.12	2.13	vw
0,16,3, 0,10,5	2.02, 2.01	2.02	w
006, 016, 026 0,18,2, 0,19,0	1.98, 1.98, 1.97 1.97, 1.97	1.97	vw
	First-Layer Line		
140	6.71	6.73	w
122	4.88	4.89	m
170, 132	4.68, 4.69	4.67	mw
142	4.45	4.45	ms
180, 171	4.22, 4.35	4.27	mw
	Second-Layer Line		
200	4.80	4.80	vvs
220	4.65	4.63	s
240	4.27	4.26	vw
231	4.19	4.17	vw
241, 250	4.02, 4.04	4.00	s
202	3.73	3.73	m
242	3.47	3.51	m
252, 280	3.34, 3.35	3.36	vw
	Third-Layer Line		
300	3.20	3.20	m
331, 350	3.00, 2.94	2.98	m
351, 360	2.86, 2.85	2.85	vw
	Fourth-Layer Line		
400	2.40	2.40	s
411, 421, 440	2.35, 2.33, 2.32	2.34	w
470, 432, 461	2.19, 2.19, 2.20	2.19	m
433, 482, 4,10,0	2.02, 2.01, 2.02	2.02	w
	Fifth-Layer Line		
500, 473, 4,10,2 4,12,0	1.92, 1.92, 1.91 1.90	1.91	vw
	Sixth-Layer Line		
600, 610, 620	1.60, 1.60, 1.59	1.60	mw
650, 660, 641	1.56, 1.55, 1.56	1.56	w

Other Oligopeptides. As regards the type of pleated sheet and crystal packing of the higher homologues (i.e., hexa- and heptapeptides) of these three series, basically the results obtained do not diverge from those described above for the corresponding pentapeptides. Finally, preliminary X-ray data on the N-deblocked derivatives of L-alanine, L-valine, and L-leucine in the ammonium form strongly indicate that they also assume the cross-β conformation in the solid state.

Conclusions

X-ray diffraction photographs of t-Boc-(L-X)$_n$OMe (X = Ala, Val, Leu, and n = 5, 6, 7) show the characteristic reflections of a cross-β structure. In addition, the good orientation of suitably prepared specimens has enabled a fairly complete determination of the unit cell of the pentapeptides to be made.

As regards chain orientation, an appreciable proportion of the Leu pentapeptide is in the antiparallel arrangement, while from the available data of the Ala and Val pentapeptides it has not been possible to establish whether the arrangement of the chains within the sheets is parallel or antiparallel.

References and Notes

(1) This work is part 50 of the series; for part 49 see C. Toniolo, *Macromolecules*, 11, 437 (1978)
(2) M. Palumbo, S. Da Rin, G. M. Bonora, and C. Toniolo, *Makromol. Chem.*, **177**, 1477 (1976).
(3) C. Toniolo and M. Palumbo, *Biopolymers*, **16**, 219 (1977).
(4) J. S. Balcerski, E. S. Pysh, G. M. Bonora, and C. Toniolo, *J. Am. Chem. Soc.*, **98**, 3470 (1976).
(5) M. M. Kelly, E. S. Pysh, G. M. Bonora, and C. Toniolo, *J. Am. Chem. Soc.*, **99**, 3264 (1977).
(6) The following abbreviations are used: t-Boc, *tert*-butyloxycarbonyl; OMe, methoxy; Ala, alanine; Leu, leucine; Val, valine.
(7) G. M. Bonora, D. Nisato, and C. Toniolo, *Makromol. Chem.*, **176**, 2535 (1975).
(8) G. M. Bonora, A. Maglione, and C. Toniolo, *Polymer*, **15**, 767 (1974).
(9) C. Toniolo, G. M. Bonora, and A. Fontana, *Int. J. Pept. Protein Res.*, **6**, 371 (1974).
(10) R. D. B. Fraser and T. P. MacRae in "Conformation in Fibrous Proteins and Related Synthetic Polypeptides", Academic Press, New York, N.Y., 1973.
(11) S. Arnott, S. D. Dover, and A. Elliott, *J. Mol. Biol.*, **30**, 201 (1967).
(12) J. O. Warwicker, *Acta Crystallogr.*, **7**, 565 (1954).
(13) R. E. Marsh, R. B. Corey, and L. Pauling, *Acta Crystallogr.*, **8**, 710 (1955).
(14) T. Komoto, K. Y. Kim, M. Oya, and T. Kawai, *Makromol. Chem.*, **175**, 283 (1974).
(15) E. M. Bradbury, L. Brown, A. R. Downie, A. Elliott, R. D. B. Fraser, W. B. Hanby, and T. R. R. McDonald, *J. Mol. Biol.*, **2**, 276 (1960).
(16) A. Elliott, R. D. B. Fraser, T. P. MacRae, I. W. Stepleton, and E. Suzuki, *J. Mol. Biol.*, **9**, 10 (1964).

Conformational Energy Calculations of the Effects of Sequence Variations on the Conformations of Two Tetrapeptides[1]

István Simon,[2] George Némethy, and Harold A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14853.

Received December 8, 1977

ABSTRACT: Conformational energy calculations were carried out on the two terminally blocked tetrapeptides *N*-acetyl-Thr-Asp-Gly-Lys-*N'*-methylamide and *N*-acetyl-Ala-Asp-Gly-Lys-*N'*-methylamide. The first peptide is a sequence variant of tetrapeptides studied earlier in this laboratory. The second peptide occurs in a bend at residues 94–97 in staphylococcal nuclease. A selection strategy is described which helps to accelerate the search of starting conformations used for energy minimization. The strategy involves exhaustive searches for conformations of fragments of the molecule which are stabilized by specific interactions and subsequent combination of fragments, prior to minimization. Several groups of low-energy conformations were found. They are compactly folded structures, but they differ from the "standard" chain reversals. One group, which is of low energy in both peptides, is stabilized by Asp···Asp and Asp···Lys backbone–side chain hydrogen bonds. Another group, of low energy in the Thr-containing peptides, is stabilized by a network of hydrogen bonds involving polar atoms of both backbone and side chains of the Thr, Asp, and Lys residues. The conformation corresponding to the sequence fragment in staphylococcal nuclease has relatively high energy, indicating that the bend observed in the protein is stabilized by interactions involving parts of the protein outside the tetrapeptide sequence.

I. Introduction

The balance between the roles of short-range and of medium- or long-range interactions in determining the existence and location of chain reversals is one of the important questions in the analysis of protein folding.[3,4] One can approach this problem by analyzing the tendency of oligopeptides to take up compactly folded conformations. This study is a continuation of earlier theoretical and experimental work from this laboratory[5] on specific tetrapeptide sequences.

In this paper, two tetrapeptides composed of L-amino acids are investigated. One of them, Ala-Asp-Gly-Lys, to be referred to as peptide A, appears as a bend conformation at residues 94–97 of staphylococcal nuclease.[6] The other, Thr-Asp-Gly-Lys, to be referred to as peptide T, is a permutation of four sequences (one of which appears as a bend at residues 35–38 of α-chymotrypsin) studied in earlier work[5] and should provide further information about the role of sequence variations when the amino acid composition is constant. In the two tetrapeptide sequences studied theoretically in this paper, only the first residue is different. Their comparison should, therefore, furnish some clues about the role of side-chain interactions in short-range conformational stability. The threonyl side chain in peptide T, for example, may participate in interactions which must be absent in peptide A, where alanine replaces threonine. An experimental study of these two tetrapeptides is in progress.[7]

Similar theoretical conformational studies on two biologically active tetrapeptides are reported in the following two papers.[8,9] The peptides are sequence variants of each other and show the large effect of variations in the sequence upon preferred conformations.

II. Computational Methods

Conformational Energy Calculations. The tetrapeptides were considered in the form of *N*-acetyl-*N'*-methylamides. The use of the CH_3CO and $NHCH_3$ blocking groups[10] eliminates charges at the terminal groups and in effect simulates an oligopeptide inside a protein sequence. The aspartyl and lysyl side chains were taken to be uncharged. The justification for this choice has been discussed elsewhere.[5,11]

0024-9297/78/2211-0797$01.00/0