Conformational Investigations on the Polypeptide and Oligopeptides with the Repeating Sequence L-Alanyl-L-prolylglycine*

B. B. Doyle,† W. Traub,‡ G. P. Lorenzi,§ and E. R. Blout‖

ABSTRACT: In order to examine the role of specific tripeptide sequences in collagen, the conformational properties of the polytripeptide $(Ala\text{-}Pro\text{-}Gly)_n$ and the homogeneous oligopeptides Boc-$(Ala\text{-}Pro\text{-}Gly)_n$-OMe, where n = 1–6, were investigated. X-Ray diffraction indicated that the polymer and the oligomers longer than the hexapeptide could adopt three different forms in the solid state, depending on the solvent to which they are exposed: (I) hydrogen-bonded sheets of polyproline II-like helices, when dried from aqueous solution; (II) a more compact hydrogen-bonded polyproline II type sheet structure, when precipitated from organic solvents or aqueous solution or when dried from ethylene glycol–hexafluoroisopropyl alcohol; and (III) a triple-helical structure, when dried from trifluoroethanol. The solid-state infrared spectra of the polymer and higher oligomers in these three forms were determined, and clear correlations were observed between the amide A, I, and II band positions and the structure for the higher oligomers. Solution studies, using infrared spectroscopy and circular dichroism spectroscopy, indicated that $(Ala\text{-}Pro\text{-}Gly)_n$ and the oligomers were unordered in aqueous solution, but the polymer was highly structured in ethylene glycol–hexafluoroisopropyl alcohol solution and had some order in trifluoroethanol solution. These results showed a striking difference between the properties of $(Ala\text{-}Pro\text{-}Gly)_n$, as described here, and those of $(Pro\text{-}Ala\text{-}Gly)_n$, which was found previously to be triple helical in the solid state and in aqueous solution.

The difference in properties suggests that X-imino acid-glycine tripeptide sequences may have a different role in collagen than imino acid-X-glycine sequences.

Collagen is a major extracellular structural protein in animals higher than the protozoa. From stereochemical considerations and X-ray diffraction data, the collagen molecule was found to adopt a supercoiled triple-helical conformation (Ramachandran and Kartha, 1955). These triple-helical molecules are organized extracellularly into fibers which are the functional unit. Collagens from all organisms and tissues have several distinctive amino acid features, including the presence of glycine as every third residue, a high proportion of imino acid residues, and the presence of the unusual residues, hydroxyproline and hydroxylysine. Chemical and enzymatic digestion of collagen and recent sequencing of some cyanogen bromide peptides indicate that collagen's amino acid sequence contains a large number of tripeptides of the form Gly-Pro-R[1] and a significant number of tripeptides of the form Gly-X-Hypro (Hannig and Nordwig, 1967; Piez *et al.*, 1968).

This repeating tripeptide character appears to be a feature of all collagens and must be important for the protein's function, which is primarily structural. Thus, one can try to relate the characteristic tripeptide sequences of collagen to its structural role by examining the conformation of synthetic polypeptides with repeating tripeptide sequences. Synthetic polytripeptides of the form $(Pro\text{-}R\text{-}Gly)_n$,[2] where R is hydroxyproline, proline, alanine, or lysine, have been found to be triple helical in the solid state (Andreeva *et al.*, 1963; Traub and Yonath, 1966; Traub *et al.*, 1969), and it appears that in some cases this structure is retained in solution as well (Engel *et al.*, 1966; Andreeva *et al.*, 1963; Brown *et al.*, 1969; Brown, 1970). $(Gly\text{-}Pro\text{-}Gly)_n$, on the other hand, forms hydrogen-bonded sheets of polyproline II type helices in the solid state (Traub, 1969), and adopts an aggregated structure in solution (Oriel and Blout, 1966).

To examine the role of tripeptides of the type X-imino acid-glycine in collagen, the polytripeptide $(Ala\text{-}Pro\text{-}Gly)_n$ was synthesized both in our laboratory (Lorenzi *et al.*, 1971) and independently at the Weizmann Institute of Sceince (Segal and Traub, 1969), and its conformational properties have been examined under a variety of conditions. A series of homogeneous oligopeptides of the form Boc-$(Ala\text{-}Pro\text{-}Gly)_n$-OMe, where n = 1–6, were also synthesized in our laboratory (Lorenzi *et al.*, 1971), and the physical properties of these oligomers were investigated and compared to those of the polymer.

In this paper physical–chemical investigations in the solid

* From the Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115. *Received March 22, 1971.* We are pleased to acknowledge the support, in part, of this work by U. S. Public Health Grants AM-07300 and AM-10794. W. T. was a National Science Foundation Senior Foreign Scientist at Harvard University, 1968–1969, and B. B. D. was a Public Health Predoctoral Fellow, No. 5-F1-GM-33,992-03.

† Present address: Laboratory of Biochemistry, National Institute of Dental Research, National Institutes of Health, Bethesda, Md.

‡ Present address: Department of Chemistry, The Weizmann Institute of Science, Rehovot, Israel.

§ Present address: Department of Industrial and Engineering Chemistry, Swiss Federal Institute of Technology, Zurich, Switzerland.

‖ To whom to address correspondence.

[1] The following abbreviations will be used in this paper: Gly for glycyl; Pro for L-prolyl; Hypro for L-hydroxyprolyl; Ala for L-alanyl; Ser for L-seryl; R for any amino acid residue; X for any amino acid residue except proline or hydroxyproline; Boc for *tert*-butyloxycarbonyl; OMe for methyl ester; TFE for trifluoroethanol.

[2] In order to be consistent and to facilitate comparisons between polymers, all collagen-like polymers, with glycine as every third residue, will be written with glycine in the third position, that is $(R\text{-}R'\text{-}Gly)_n$ for any polytripeptide. The references cited herein should be consulted for the procedures for the preparation of the various polypeptides.

state and in solution on the polypeptide (Ala-Pro-Gly)$_n$ and oligopeptides with this sequence are reported, and the structures indicated by these results are discussed.

Experimental Section

The samples of the polypeptide (Ala-Pro-Gly)$_n$ and the set of oligopeptides with this sequence were synthesized as described in the accompanying paper (Lorenzi *et al.*, 1971). Some samples prepared at the Weizmann Institute of Science (Segal and Traub, 1969) were also used for the solid state investigations.

X-Ray diffraction patterns of powders were photographed with a 114.6-mm diameter powder camera, with the sample in thin-walled glass capillaries. Oriented films were stroked out from the solution as it dried and were photographed with a Norelco microcamera. Nickel-filtered Cu Kα radiation from a Norelco fine-focus X-ray tube was used.

Infrared spectra were determined on a Perkin-Elmer 521 spectrophotometer. Solid samples were prepared in KBr pellets, and solutions were examined in 0.05-mm CaF_2 cells at concentrations near 2%.

A Cary 60 spectropolarimeter was used to determine the circular dichroism spectra. Error limits of 10% below 220 mμ and 5% above 220 mμ have been estimated for the instrument used (Carver, 1966). The concentrations of the solutions, by sample weight, were all near 0.02%, and cells of path lengths 2.00, 0.501, and 0.200 mm were used.

Elemental analyses were determined by Spang Microanalytical Laboratories, Ann Arbor, Mich. The densities were determined by the flotation method in a mixture of carbon tetrachloride and toluene.

Results

I. Solid State

The structures of (Ala-Pro-Gly)$_n$ and the Ala-Pro-Gly oligopeptides in the solid state were examined using X-ray diffraction and infrared spectroscopy.

A. X-Ray Diffraction. Three different X-ray patterns were observed for (Ala-Pro-Gly)$_n$, depending on the solvents to which it was exposed. Drying from aqueous solution gave one crystalline pattern (form I); precipitation from organic solvents or aqueous solution, or drying from bifunctional, somewhat hydrophobic solvents, including glycerin and ethylene glycol–hexafluoroisopropyl alcohol (2:1, v/v), gave a different set of crystalline reflections (form II); and drying from trifluoroethanol solution led to a third, less sharp X-ray pattern (form III).

X-Ray studies on the oligomers Boc-(Ala-Pro-Gly)$_n$-OMe, for n = 3–6, revealed, in each case, one of the three patterns mentioned above. For example, the nonapeptide (n = 3) gave the pattern characteristic of form II, the pentadecapeptide (n = 5) that of form III, and the octadecapeptide (n = 6) that of form I. In fact the different procedures used to purify the various oligopeptides were found to be responsible for the form that was observed. Any of these oligomers, with at least three tripeptide units, was found to be capable of taking up all three structures depending on the solvent to which it was exposed, and it was possible to convert any one form into any other by appropriate solvent treatment. The oligomer Boc-(Ala-Pro-Gly)$_4$-OMe can be used to illustrate this. The main fraction, after purification, was in form II. It was converted into form I by drying from aqueous solution, and to form III by drying from a solution in trifluoroethanol. All other possible interconversions were made, as shown in Figure 1. Plate I shows the X-ray diffraction powder patterns of Boc-(Ala-Pro-Gly)$_4$-OMe in each of the three forms.

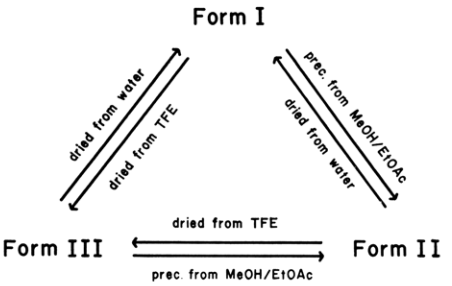

FIGURE 1: All possible interconversions of Boc-(Ala-Pro-Gly)$_4$-OMe between forms I, II, and III by appropriate solvent treatment.

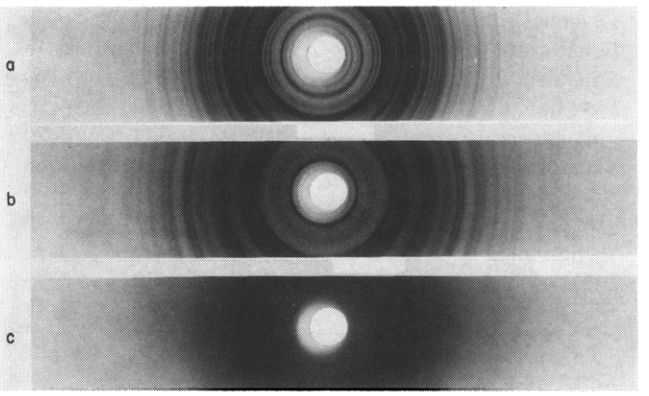

PLATE I: The X-ray diffraction powder photographs of Boc-(Ala-Pro-Gly)$_4$-OMe in (a) form I, as dried from water; (b) form II, as precipitated from methanol with ethyl acetate; (c) form III, as dried from trifluoroethanol.

The tripeptide was found to give a very sharp crystalline X-ray pattern quite unlike any of the three discussed above, and was unaffected by the solvent treatments which caused interconversions between the various forms of the higher oligomers. The hexapeptide also showed a unique crystalline pattern, but drying from water gave rise to some reflections characteristic of form I, and drying from trifluoroethanol gave a very diffuse pattern similar to that of form III observed for all higher oligomers. No pattern similar to that of form II was observed, but, with the solvents tried, the hexapeptide could not be precipitated from solution. The nonapeptide and dodecapeptide also showed distinctive sets of reflections in addition to the patterns of forms I, II, and III, but these were not found in the patterns of the pentadecapeptide or octadecapeptide.

i. FORM I. The X-ray pattern of the polymer dried from aqueous solution, form I, was described and interpreted by Segal and Traub (1969). They reported a structure of polyproline II type helices aggregated to form zigzag sheets, with neighboring chains anti-parallel and joined by one hydrogen bond per tripeptide ($N_1H_1 \cdots O_2$, according to the notation shown in Figure 2).

ii. FORM II. The X-ray pattern of form II was also described by Segal and Traub (1969) (their "organic" form), but they did not offer any interpretation. More recently, Schwartz *et al.* (1970) have suggested that this pattern can be indexed in terms of a hexagonal unit cell with a = 12.5 Å and c = 9.4 Å.

We have not succeeded in obtaining oriented films or fibers of this form of the polymer from the various precipitates and

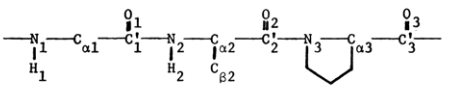

FIGURE 2: The structural formula of one tripeptide unit of (Ala-Pro-Gly)$_n$, indicating notation used in the text.

TABLE I: Observed and Calculated X-Ray Spacings for (Ala-Pro-Gly)$_n$ and the Higher Ala-Pro-Gly Oligomers Form II.[a]

Orientation	I_0	d_0 (Å)	hkl	d_c (Å)
Equatorial	ms	10.35	100	10.30
Diagonal	mw	7.03	101	6.95
Equatorial	ms	5.16	200	5.15
Equatorial	vs	4.69	110	4.69
Meridional and equatorial	m	4.25	102	4.28
			111	4.20
Equatorial	m	3.68	210	3.68
		3.45	300	3.43
Broad, unoriented	ms	3.30	211	3.43
			112	3.32
	mw	2.87	212	2.90
			310	2.88

[a] The orientations were obtained from a film of the oligopeptide Boc-(Ala-Pro-Gly)$_5$-OMe in form II (see text). Observed intensities (I_0) were estimated as very strong (vs), moderately strong (ms), medium (m), moderately weak (mw), and very weak (vw). Indices (hkl) were assigned to the reflections and their spacings calculated on the basis of an orthogonal cell with a = 10.30 Å, b = 4.27 Å, and c = 9.40 Å. Weak reflections were also observed at 2.67, 2.54, 2.43, and 2.30 Å.

solutions with which we have been working. However, we have been able to obtain a partially oriented form II pattern from a stroked film of the oligopeptide Boc-(Ala-Pro-Gly)$_5$-OMe precipitated from a methanol solution with ethyl acetate (Plate II). This has led us to index the reflections, in accordance with the observed spacings and orientations, in terms of an orthogonal unit cell with a = 10.30 Å, b = 5.27 Å, and c = 9.40 Å (Table I). In spite of the different cell dimensions, our indexing scheme and that of Schwartz *et al.* (1970) are very similar, as nearly all of the observed spacings correspond to $h0l$ reflections and the interplanar spacings along the a and c axes are nearly the same for the two unit cells.[3] However, considerations of the density and the packing of the molecules in the structure have led us to the conclusion that the orthogonal cell is, in fact, the correct one.

The orthogonal cell described above contains two tripeptide units, each of mol wt 225, which leads to a calculated density of 1.46 g/cm^3, in reasonable agreement with the observed value of 1.52 g/cm^3 for the polymer in form II. On the other hand, the hexagonal cell should have densities of 0.29, 0.59, 0.88, 1.18, 1.47, or 1.76 g per cm^3, depending on whether

[3] Whereas for an orthogonal unit cell the 100 spacing equals the a axis, for an hexagonal unit cell it is 0.866 a, and therefore, 10.8 Å for a = 12.5 Å. A comparison of the two indexing schemes is further complicated by minor discrepancies in the two sets of observed spacings, as well as errors in the calculated spacings for the hexagonal cell.

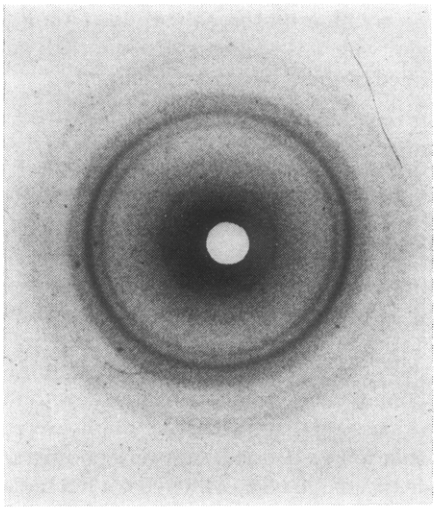

PLATE II: Partially oriented X-ray pattern of Boc-(Ala-Pro-Gly)$_5$-OMe in form II, stroked out after precipitation from methanol with ethyl acetate.

they are, respectively, 1, 2, 3, 4, 5, or 6 tripeptides in the unit cell. Only five tripeptides per cell will lead to reasonable agreement with the observed density, and this number is inconsistent with a hexagonal or even a pseudohexagonal cell. Elemental analysis of samples in form II suggests that some solvent may be present in the unit cell.[4]

The 9.4-Å c axis is suggestive of the poly-L-proline II conformation (Cowan and McGavin, 1955; Sasisekharan, 1959), which has been reported for form I of (Ala-Pro-Gly)$_n$ and several other polypeptides containing proline and/or glycine (Crick and Rich, 1955; Traub and Yonath, 1967; Traub, 1969). This is supported by the observation that drying of form I of Boc-(Ala-Pro-Gly)$_4$-OMe over P_2O_5 *in vacuo* for 1 week reproducibly yielded the form II of the oligopeptide. Such treatment could not cause large molecular rearrangements, and suggests that form II cannot be very different from form I, and the removal of water is important in its formation.

We have investigated the possible packing in the orthogonal unit cell of a polyproline II like conformation of the polymer, using space-filling and skeletal models, as well as graphical methods described by Traub (1969). In fact we have found that the only possible mode of packing allowing interchain hydrogen bonding is the type of anti-parallel zigzag arrangement reported for form I (Segal and Traub, 1969). For a small range of molecular orientations, $N_1H_1 \cdots O_2$ hydrogen bonds can be made between neighboring anti-parallel chains about 4.8 Å apart, while acceptable van der Waals contacts are maintained between neighboring parallel chains. These parallel chains are separated by 5.3 Å instead of 7.2 Å as in form I, giving a more compact structure, but the separation between hydrogen-bonded sheets of molecules is 1.7 Å greater in form II. The close relationship between the two structures is illustrated in Figure 3, and provides further strong support for the choice of an orthogonal unit cell. We should not exclude the possibility that the molecular conformation is somewhat different from that of polyproline II, and this might allow

[4] Elemental analysis of the pentadecapeptide in form II, after precipitation from methanol with ethyl acetate showed: *Anal.* Calcd. for $C_{56}H_{87}N_{15}O_{18}$: C, 53.44; H, 6.97; N, 16.69. Found: C, 49.43; H, 7.12; N, 13.8. But if ten molecules of methanol are bound to each pentadecapeptide molecule, the results are in satisfactory agreement. *Anal.* Calcd for $C_{56}H_{87}N_{15}O_{18}$ + 10MeOH: C, 50.19; H, 8.05; N, 13.31.

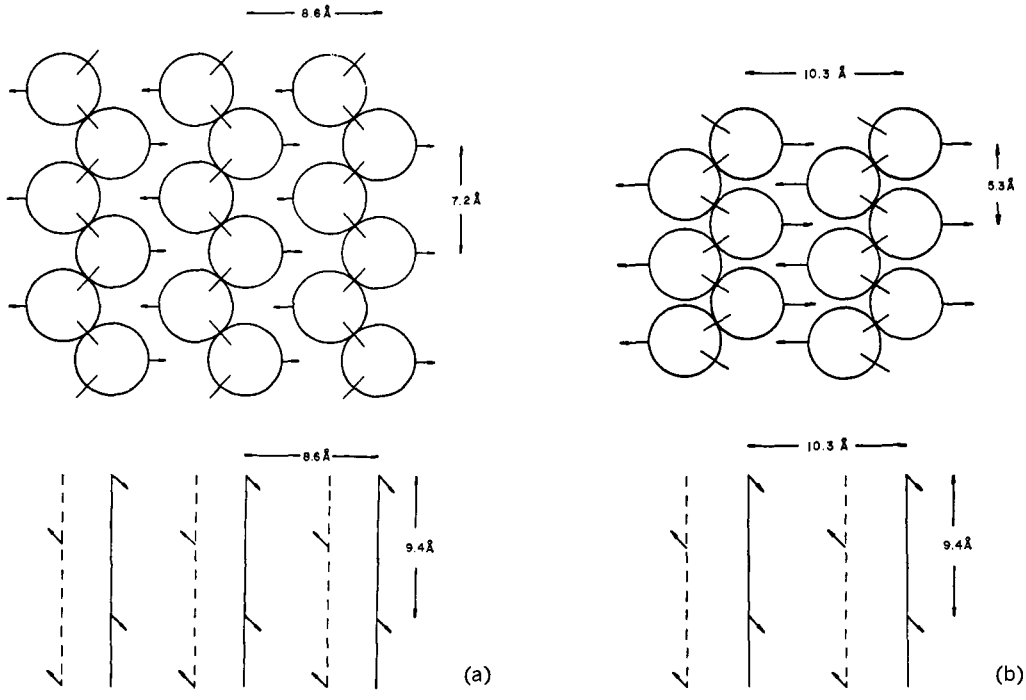

FIGURE 3: Schematic illustration of (a) form I and (b) form II of $(Ala\text{-}Pro\text{-}Gly)_n$ with polypeptide chains represented by circles, pyrrolidine rings by arrows, and the hydrogen bonds by straight lines.

formation of an additional hydrogen bond between parallel molecules in the packing arrangement of Figure 3b. Thus with the removal of water, it is possible that $N_2H_2\cdots O_1$ hydrogen bonds could be formed between adjacent parallel chains, in addition to the $N_1H_1\cdots O_2$ bonds already formed between anti-parallel chains. Alternatively, there could be $N_2H_2\cdots O_2$ bonding between anti-parallel molecules in forms I and II, with an additional $N_1H_1\cdots O_3$ bond between parallel molecules in form II. However, both schemes would require appreciable distortion of the polyproline II conformation and therefore seem unlikely in view of the close correspondence of the *c* axis with that of polyproline II.

iii. FORM III. A stroked film of the polymer from trifluoroethanol gave a well-oriented form III pattern (Plate III). Unlike the oriented patterns of forms I and II, Plate III shows a strong resemblance in spacing and intensity to the X-ray pattern of collagen. Its salient features include a meridional reflection at 2.85 Å and near-meridional reflections at 3.74 and 7.07 Å, corresponding to the seventh and third layer lines, respectively, which indicates similar helical parameters to collagen, as well as first and second orders of a 10.4-Å equatorial spacing, which indicates the lateral dimensions appropriate to a collagen-like triple helix. It seems clear that $(Ala\text{-}Pro\text{-}Gly)_n$ in form III, like a number of other polytripeptides and polyhexapeptides containing proline and glycine residues (Traub *et al.*, 1969), does, in fact, assume this type of molecular conformation.

B. Infrared Spectroscopy. The solid-state infrared spectra of the three forms that the higher oligomers and $(Ala\text{-}Pro\text{-}Gly)_n$ can adopt were determined. Since the conformation of each form has been established by X-ray diffraction, it is possible to relate the infrared band positions to structural features. The amide bands of the polymer in the three different forms were very broad and difficult to correlate with conformation, possibly due to the molecular weight polydispersity of the preparation. On the other hand, the amide bands of the higher oligomers ($n \geqq 3$) were sharp and showed clear correlations between the band positions and structure. The amide band frequencies observed for the higher oligomers in the three forms are listed in Table II, and to illustrate this, the nfrared spectra of Boc-$(Ala\text{-}Pro\text{-}Gly)_4$-OMe in the three different forms are shown in Figure 4.

i. Form I (Figure 4a) has three bands in the 3500- to 3200-cm^{-1} region. The intermolecularly hydrogen-bonded amide N_1H_1 has the low amide A frequency at 3295 cm^{-1}, a position typical of amide hydrogens bonded between polypeptide chains (Miyazawa, 1962). The second amide N_2H_2 is probably not observed in this spectrum because of the low intensity of a nonbonded or weakly bonded fundamental stretching vibration (Huggins and Pimental, 1956), and because this band would coincide with the region of water absorption. The bands

TABLE II: The Amide A, Amide I, and Amide II Band Positions in the Solid-State Infrared Spectra of the Higher Oligopeptides, Boc-$(Ala\text{-}Pro\text{-}Gly)_n$-OMe ($n \geq 3$), in Forms I, II, and III.[a]

	Amide A	Amide I	Amide II
Form I	3470–3480 (ms)	1657–1665 (s)	1540–1549 (m)
	3400–3420 (ms)	1640–1641 (ms)	
	3295–3305 (ms)	1618–1625 (s)	
Form II	3330–3340 (s)	1670–1672 (s)	1535–1540 (m)
		1639–1640 (s)	
Form III	3305–3330 (s)	1638–1640 (s)	1530–1534 (ms)

[a] The relative intensities of the bands are indicated parenthetically: (s) strong; (ms) medium-strong; (m) medium; and (w) weak. The positions are all given in cm^{-1} units.

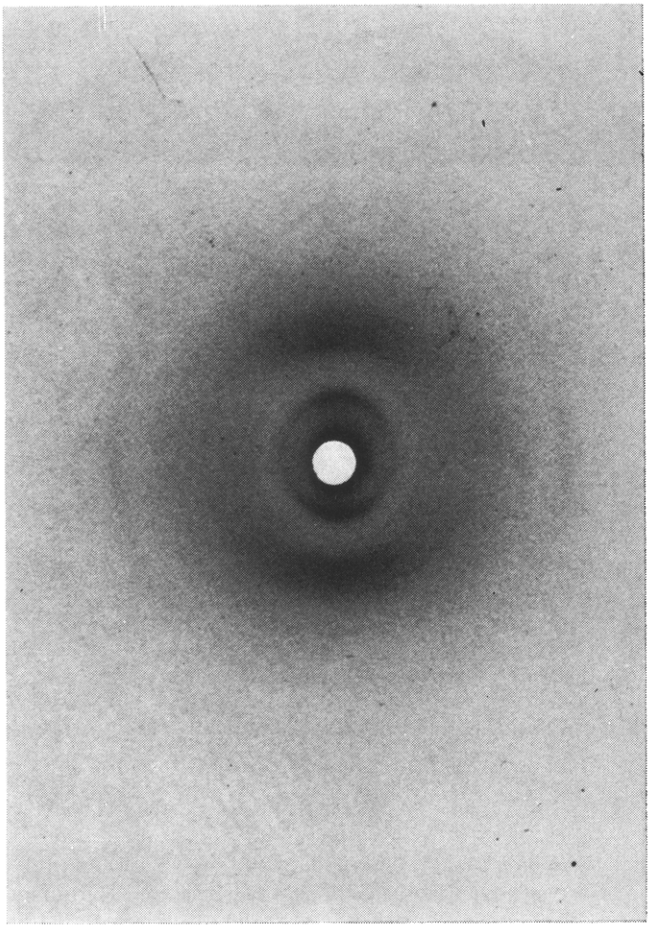

PLATE III: Oriented X-ray pattern of $(Ala\text{-}Pro\text{-}Gly)_n$ stroked out from trifluoroethanol (form III).

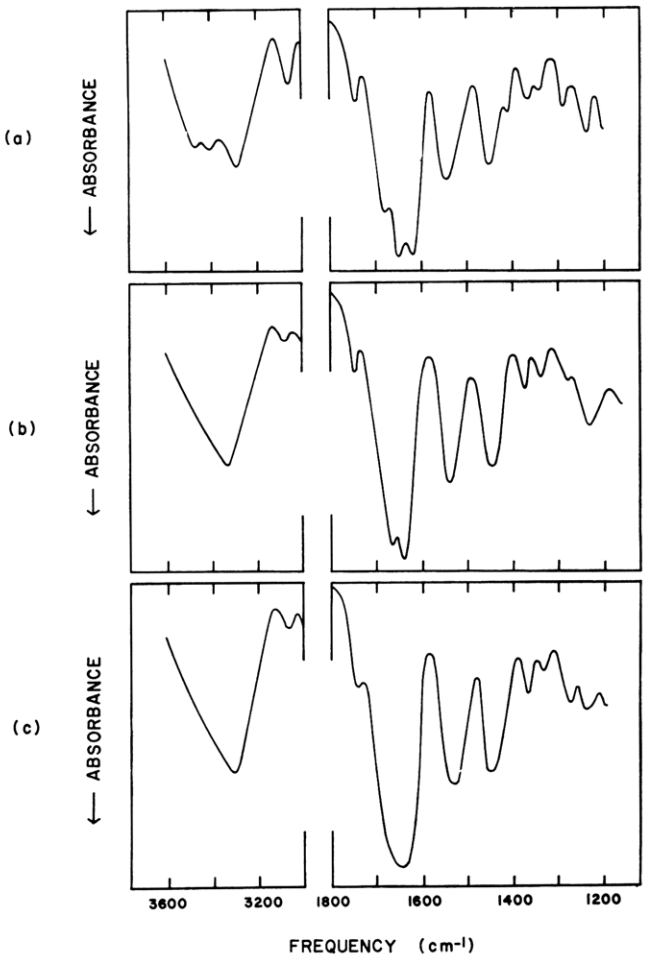

FIGURE 4: The infrared spectra of Boc-$(Ala\text{-}Pro\text{-}Gly)_4$-OMe in (a) form I, as dried from aqueous solution; (b) form II, as precipitated from methanol with ethyl acetate; and (c) form III, as dried from trifluoroethanol.

at 3410 and 3470 cm^{-1} are most likely due to the hydrogen-bonded OH-stretching vibrations of water. The absorption maximum at 3410 cm^{-1} suggests the presence of water, which is strongly hydrogen bonded, perhaps forming hydrogen bonds between adjacent parallel chains in the structure. There are three distinct amide I bands, at 1660, 1640, and 1620 cm^{-1}. In the form I conformation one carbonyl ($C_2'O_2$) is strongly intermolecularly hydrogen bonded, one ($C_1'O_1$) is specifically bound to water, and one ($C_3'O_3$) is exposed to bulk solvent. The three different amide I frequencies could reflect the three nonequivalent carbonyls or could be due to more complex splitting phenomena in the vibrational spectrum.

ii. In the infrared spectrum of form II (Figure 4b) there is only one peak in the NH-stretching region, at 3325–3330 cm^{-1}. This high amide A frequency suggests hydrogen bonds which are weaker than those formed in the form I structure. The two strong amide I bands, which are present at 1670 and 1637 cm^{-1}, are reminiscent of the amide I region of $(Gly\text{-}Pro\text{-}Gly)_n$ (B. B. Doyle and E. R. Blout, to be published), which forms a compact sheet structure of polyproline II type helices with two hydrogen bonds per tripeptide (Traub, 1969).

iii. The collagen-like form, form III, has one amide A band that occurs between 3305 and 3320 cm^{-1} for the higher oligomers (Figure 4c) and at 3315 cm^{-1} in the polymer spectrum. The amide A band of collagen always appears at 3325–3330 cm^{-1}, at least 25 cm^{-1} higher than that observed in other protein spectra. The amide I band of the higher oligomers and the polymer in form III occurs at 1638 cm^{-1}. Although the major amide I band of collagen is found at a higher frequency (1660–1650 cm^{-1}), other polypeptides in a triple-helical form have amide I bands near 1640 cm^{-1} (B. B. Doyle and E. R. Blout, to be published). Since the amide band positions of form III samples are not typical of the collagen spectrum, one might suggest that the samples in this form are mostly unordered with only a small amount of triple-helical structure. Then the X-ray diffraction pattern would represent only the small amount of ordered collagen-like form, while the infrared bands would reflect the dominant random character of the sample. However, this hypothesis can be rejected because the polymer $(Ala\text{-}Pro\text{-}Gly)_n$ was oriented in form III, and showed perpendicular dichroism for the amide A and amide I bands, and parallel dichroism for the amide II band. This result indicates these amide bands are not caused by unordered material, but are representing to a significant extent the collagen-like structure observed in the X-ray pattern.

The infrared spectrum of the tripeptide has very sharp amide bands at positions different from those found in the higher oligomers in the three forms, and their positions do not change after the sample has been treated with solvents that cause structural change in the higher oligomers. The hexapeptide, as purified, has an infrared spectrum which does not resemble the spectra of the three forms the higher oligomers adopt. However, when the hexapeptide is dried from water and trifluoroethanol, the infrared spectra have amide bands similar to those found in the higher oligomers in form I and

TABLE III: The Amide I Band Positions for the Oligomers Boc-(Ala-Pro-Gly)$_n$-OMe, Where n Equals 1–6, in Deuterium Oxide (D_2O), Ethylene Glycol–Hexafluorisopropyl Alcohol (2:1, v/v) (EG–HFIP), and Trifluoroethanol (TFE) Solution Infrared Spectra.[a]

Sample	Solvents		
	D_2O	EG–HFIP	TFE
Boc-Ala-Pro-Gly-OMe	1653 (s)	1678 (s)	1680 (s)
	1623 (ms)	1639 (s)	1637 (ms)
Boc-(Ala-Pro-Gly)$_2$-OMe	1649 (s)	1675 (ms)	1680 (s)
	1622 (s)	1634 (s)	1639 (s)
Boc-(Ala-Pro-Gly)$_3$-OMe	1645 (s)	1671 (ms)	1680 (ms)
	1622 (s)	1633 (s)	1638 (s)
Boc-(Ala-Pro-Gly)$_4$-OMe	1648 (s)	1670 (sh)	1680 (ms)
	1620 (s)	1630 (s)	1640 (s)
Boc-(Ala-Pro-Gly)$_5$-OMe	1646 (s)	1672 (sh)	1680 (m)
	1623 (s)	1636 (s)	1638 (s)
Boc-(Ala-Pro-Gly)$_6$-OMe	1645 (s)	1672 (sh)	1680 (m)
	1623 (s)	1633 (s)	1637 (s)
(Ala-Pro-Gly)$_n$	1640 (s)	1657 (ms)	1637 (s)
	1622 (s)	1629 (s)	

[a] All the frequencies are in units of cm^{-1}, and the relative intensities are indicated parenthetically: (s) strong; (ms) medium-strong; (m) medium; (sh) shoulder. The values for the polymer in each solvent are given for comparison.

form III, respectively. These results are in good agreement with the X-ray diffraction patterns discussed above.

The above results show the advantage of examining a set of homogeneous oligomers as a complement to investigations of a higher molecular weight and heterogeneous polymer. This advantage was striking here since the conformationally sensitive infrared amide band positions corresponding to the different forms were easily identified for the higher oligomers, but not for the polymer.

II. Solution Investigations

Solution infrared spectroscopy and circular dichroism spectroscopy were used to investigate the conformations of the oligomers and (Ala-Pro-Gly)$_n$ in solution. For these studies, three solvents were chosen which, upon evaporation, produced the three different forms observed in the solid state: water (or deuterium oxide), form I; ethylene glycol–hexafluoroisopropyl alcohol (2:1, v/v), form II; and trifluoroethanol, form III. These three solvents are all reasonably transparent in the ultraviolet and infrared frequency ranges of interest.

A. Infrared Spectroscopy. The infrared spectra of (Ala-Pro-Gly)$_n$ and the Ala-Pro-Gly oligomers were determined in the region 1800–1500 cm^{-1} in the solvents indicated above. The positions of the amide I bands are listed in Table III. The infrared solution amide bands can be compared with the amide band positions for each of the three forms in the solid state (Table II), and the bands in the polymer spectrum can be compared to those in the spectra of the oligomers in the same solvent.

i. In D_2O two strong amide I bands near 1645 and 1620 cm^{-1} are found in the spectra of all the oligomers and the polymer. The similarity of this conformationally sensitive amide I region in the polymer and the oligomers suggests these samples are in the same conformation in this solvent. Previous investigations on (Ala-Pro-Gly)$_n$ indicated the polymer was unordered in aqueous solution (Segal and Traub, 1969), and it appears that all of the oligomers are unordered in aqueous solution too, as might be expected. Since it has been generally assumed that any unordered polypeptide or protein will have a single amide I band, the unperturbed vibration (Miyazawa, 1960), it was necessary to clarify the source of the two amide I bands observed for the unordered polymer and oligomers. To do this, the infrared spectra of several blocked dipeptides and tripeptides were examined in D_2O, and the amide band positions of these peptides are given in Table IV. The data in Table IV show that Ala-Pro imide-bond carbonyls in an unordered peptide chain have an absorption maximum near 1620 cm^{-1}, whereas Pro-Gly and Ala-Gly amide-bond carbonyls have an absorption maximum in the region of 1645 cm^{-1}. Thus the consistent observation of two amide I bands in the spectra of the Ala-Pro-Gly oligomers and (Ala-Pro-Gly)$_n$ is in accord with the conclusion that these molecules are unordered in aqueous solution. Two such amide I bands should be present in solution spectra in other solvents when samples with a significant number of both imino and amino acids are present in an unordered conformation.

TABLE IV: The Amide I Band Positions of Several Small Peptides Related to (Ala-Pro-Gly)$_n$ in Deuterium Oxide (D_2O) Solution Infrared Spectra.[a]

Sample	Amide I Band Position
Boc-Ala-Pro-Gly-OMe	1653 (s)
	1623 (ms)
Boc-Ala-Ala-Gly-OMe	1650 (s)
Boc-Ala-Pro-OH	1622 (s)
Boc-Pro-Gly-OMe	1650 (s)

[a] All the frequencies are in units of cm^{-1}, and the relative intensities are indicated parenthetically: (s) strong; (ms) medium strong; (m) medium.

ii. In ethylene glycol–hexafluoroisopropyl alcohol the amide I region of the polymer has two sharp bands at 1657 and 1629 cm^{-1}. The sharpness of these bands and their positions, very different from those found in the tripeptide or other oligomers, suggest that (Ala-Pro-Gly)$_n$ is structured in ethylene glycol–hexafluoroisopropyl alcohol. It is possible that the structure the polymer is adopting in solution is the same that it adopts when dried from this solvent, namely form II. The solid-state infrared spectrum of the polymer and higher oligomers in form II has two strong amide I bands at 1670 and 1640 cm^{-1}, positions each approximately 12 cm^{-1} higher than the two bands observed in solution. This difference in frequency could be due to strong solvent–polypeptide interactions in solution which lower the stretching frequencies.

iii. In trifluoroethanol the polymer has only one amide I band, at 1637 cm^{-1}, with some broadness on the high-frequency side, whereas the oligomers have two amide I bands. The presence of only one strong amide I band in the polymer infrared spectrum suggests that (Ala-Pro-Gly)$_n$ is not completely unordered in trifluoroethanol solution, since in that case two strong amide bands should be observed as in D_2O. The positions of the amide I and amide II bands of the polymer in trifluoroethanol solution are the same as those observed

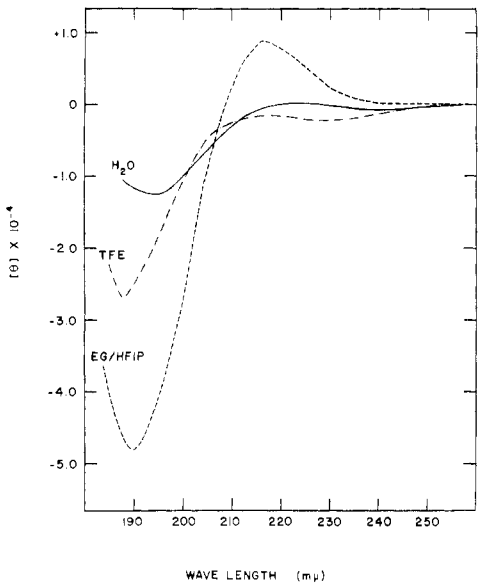

FIGURE 5: The circular dichroism spectra of $(Ala\text{-}Pro\text{-}Gly)_n$ in water, ethylene glycol–hexafluoroisopropyl alcohol (2:1, v/v), and trifluoroethanol. The values are given in mean residue ellipticities.

in the spectrum of the polymer dried from trifluoroethanol, where it is in the collagen-like conformation, form III. Thus the polymer may retain some of its triple-helical form in trifluoroethanol.

B. Circular Dichroism Spectroscopy. The circular dichroism spectra of $(Ala\text{-}Pro\text{-}Gly)_n$ in water, ethylene glycol–hexafluoroisopropyl alcohol (2:1, v/v), and trifluoroethanol are shown in Figure 5. The spectra of the water-soluble fraction of the polymer (molecular weight near 5000) had only one broad negative extremum in water, resembling the circular dichroism spectrum of denatured collagen. The circular dichroism spectrum of the water-insoluble polymer fraction (molecular weight near 14,000) in ethylene glycol–hexafluoroisopropyl alcohol showed a very large negative extremum at 190 mμ ($[\theta] = -48{,}300$; $\theta_{MT} = -145{,}000$)[5] and a positive extremum at 216–217 mμ ($[\theta] = +8700$; $\theta_{MT} = +26{,}000$), The magnitude and shape of the circular dichroism spectrum in ethylene glycol–hexafluoroisopropyl alcohol resemble the circular dichroism spectrum of native collagen ($[\theta]_{198\ m\mu} = -54{,}000$; $[\theta]_{220\ m\mu} = +8000$, Brown *et al.*, 1969), but the extrema are at considerable shorter wavelengths in the polymer spectrum compared with those in the collagen spectrum. In trifluoroethanol the water-insoluble fraction of the polymer has a negative extremum near 187 mμ ($[\theta] = -26{,}700$; $\theta_{MT} = -80{,}000$) and a very small maximum near 220 mμ, presenting a spectrum similar to that of partially denatured collagen.

There are several problems with basing even tentative conclusions about conformation on simple comparisons of the circular dichroism spectra of the polymer with the circular dichroism spectra of native and denatured collagen. The circular dichroism spectrum of native collagen is well established, but the spectrum of collagen in an unordered state is not well defined, so that it is not clear how much local structure may remain in a sample whose circular dichroism spectrum resembles that of "denatured" collagen. In addition, the solvents, ethylene glycol–hexafluoroisopropyl alcohol and trifluoroethanol, have not been commonly used for circular dichroism investigations, and the circular dichroism spectrum expected for a triple-helical conformation in these solvents is not known. Solvent effects could shift the positions or alter the magnitudes of the extrema from their usual values in aqueous solution.

To provide a different approach to interpreting the circular dichroism spectra of the polymer, in addition to direct comparisons, the circular dichroism spectra of the oligomers Boc-$(Ala\text{-}Pro\text{-}Gly)_n$-OMe, where $n = 1$–6, were determined in water, ethylene glycol–hexafluoroisopropyl alcohol, and trifluoroethanol. Comparisons between the circular dichroism spectra of the oligomers and that polymer in each solvent may indicate changes in conformation with increasing molecular size. The end groups of small peptides make a significant contribution to their circular dichroism spectra, the Boc and methyl ester groups in the case of these Ala-Pro-Gly oligomers. This contribution will be less important with increasing chain length and a difference in the circular dichroism spectra of different oligomers, and the polymer, could be due either to the diluting out of the end-group contributions or to conformation changes (or to both causes). In order to eliminate end-group effects so that only conformation changes will cause changes in the circular dichroism values, the following method, suggested by Dr. R. C. Thompson in this laboratory, was used. The circular dichroism values for each oligomer were calculated on a molar basis. Then the molar values of Boc-$(Ala\text{-}Pro\text{-}Gly)_{n-1}$-OMe were subtracted from those of Boc-$(Ala\text{-}Pro\text{-}Gly)_n$-OMe, for $n = 2$–6, giving the effect of one tripeptide addition to the conformation of each oligomer. If all of the oligopeptides have the same conformation or are unordered, these difference curves—the tripeptide from the hexapeptide, the hexapeptide from the nonapeptide, etc.—should all be very similar, with no misleading interference from end-group effects. In order to compare the circular dichroism spectrum of the polymer to that of the oligomers, its circular dichroism values must be calculated on a mean tripeptide basis, since these difference curves represent the contribution of a mean tripeptide unit. If some change in structure occurs with increasing chain length, due to peptide–peptide interactions or solvent–peptide interactions, the polymer circular dichroism spectrum should differ considerably from the difference curves.

In water the circular dichroism difference curves of Boc-$(Ala\text{-}Pro\text{-}Gly)_{n-1}$-OMe subtracted from Boc-$(Ala\text{-}Pro\text{-}Gly)_n$-OMe, for $n = 2$–6, were all very similar, within experimental error, and resembled the circular dichroism spectrum of the polymer in water (Figure 6a). This strongly suggests that all of the oligomers and the polymer are unordered in water, in agreement with the infrared solution results and previous investigations (Segal and Traub, 1969). The similarity between the five circular dichroism difference curves and the polymer circular dichroism spectrum suggests this difference curve method can be used to eliminate end-group contributions, and thus to distinguish between conformation changes and group dilution effects.

The circular dichroism difference curves of the oligomers in ethylene glycol–hexafluoroisopropyl alcohol and the polymer circular dichroism spectrum in this solvent are presented in Figure 6b. The circular dichroism spectrum of the water-insoluble polymer in this solvent system is dramatically different from any of the difference spectra, suggesting the polymer is highly structured in this solvent system.

[5] The values for the polypeptide circular dichroism spectra are given both in mean residue ellipticities, $[\theta]$, and for comparison to the oligomer circular dichroism difference spectra, in mean tripeptide ellipticities θ_{MT}.

The circular dichroism difference curves in trifluoroethanol are shown in Figure 6c, and there appears to be significant variation between some difference curves. The spectrum of the water-insoluble polymer fraction has a much larger negative extremum than that found in any of the difference spectra, suggesting this fraction of the polymer has some structure in trifluoroethanol.

The agreement of these circular dichroism results with the infrared solution results presented in the last section increases the force of these conclusions, since they were arrived at by two independent properties of the molecule—its vibrational characteristics (the infrared spectrum) and its electronic transitions (the circular dichroism spectrum).

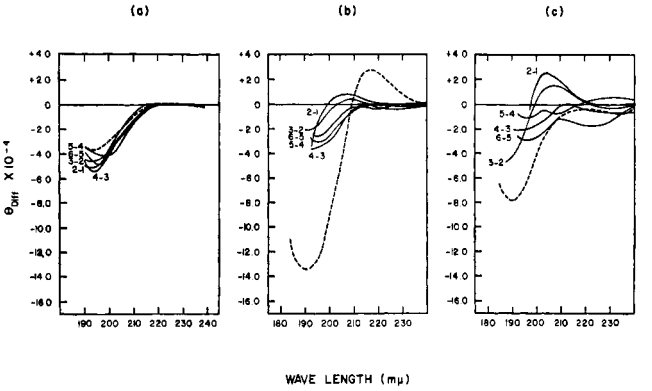

FIGURE 6: The circular dichroism difference curves of the oligomers, Boc-(Ala-Pro-Gly)$_{n-1}$-OMe subtracted from Boc(Ala-Pro-Gly)$_n$-OMe, where n = 2–6, in (a) water, (b) ethylene glycol–hexafluoroisopropyl alcohol (2:1, v/v), and (c) trifluoroethanol. The circular dichroism spectrum of the polymer (Ala-Pro-Gly)$_n$, as calculated on a mean tripeptide basis, is given in each solvent for reference (- - - - -).

Discussion and Conclusions

Collagen is known to contain a significant number of tripeptide sequences of the form X-imino acid-glycine, where X is any amino acid except proline and hydroxyproline (Greenberg *et al.*, 1964; Piez *et al.*, 1968; Butler, 1970), but the role of these sequences, of which Ala-Pro-Gly is an example, is not understood. Polymers with a repeating tripeptide sequence of the form Pro-R-Gly, such as (Pro-Ala-Gly)$_n$, (Pro-Pro-Gly)$_n$, and (Pro-Hypro-Gly)$_n$, were found to be triple helical in the solid state (Traub and Yonath, 1966; Andreeva *et al.*, 1963), and in aqueous solution as well for high enough molecular weight samples (Andreeva *et al.*, 1963; Brown *et al.*, 1969; Brown, 1970; Kobayashi *et al.*, 1970). These results suggest that a major role of Pro-R-Gly tripeptide sequences is to stabilize the triple-helical molecular conformation of collagen, both in the fibrous form and in aqueous solution. It is clear from the investigations described in this paper that Ala-Pro-Gly sequences, and perhaps therefore X-imino acid-Gly sequences in general, do not contribute this same degree of stability toward triple-helix formation as imino acid-R-Gly sequences, in either the solid state or in solution.

When dried from various solvents (Ala-Pro-Gly)$_n$ and the Ala-Pro-Gly oligomers longer than the hexapeptide always adopted some aggregate form of polyproline II type helices. Such sequences adopted the specific triple-helical aggregate when dried from trifluoroethanol or when in the polyhexapeptides (Pro-Pro-Gly-Ala-Pro-Gly)$_n$ and (Pro-Ala-Gly-Ala-Pro-Gly)$_n$ (Segal *et al.*, 1969; Segal, 1969). This suggests that X-imino acid-Gly sequences could easily fit into a triple-helical form in collagen fibers (approximately a solvated solid state). Since we have found (Ala-Pro-Gly)$_n$ and the Ala-Pro-Gly oligomers to be unordered in aqueous solution, it appears likely that similar sequences in collagen would not contribute to the stability of a triple-helical structure and might actually destabilize this conformation in physiological solution. The lower conformational stability of such sequences may play a role in solution by regulating the melting temperature of collagen, so that collagen can be effectively organized and degraded when necessary. Thus, it is possible that sequences of this sort, X-imino acid-Gly, contribute some structural stability for the triple-helical conformation in collagen fibers, while providing a control mechanism in solution.

It is interesting to consider why the conformational properties of (Ala-Pro-Gly)$_n$ differ from those of (Pro-Ala-Gly)$_n$ in both the solid state and in solution. Theoretical considerations (Hopfinger and Walton, 1970) have suggested that these two polymers have very similar energy values for a triple-helical structure, taking into account steric factors, electrostatic interactions, torsional factors, and interchain interactions. It is possible that the differences observed experimentally are due to solvent–polypeptide interactions, as suggested by Segal and Traub (1969). In (Pro-Ala-Gly)$_n$ the Ala-NH points outward from the triple helix and is totally exposed to solvent. For (Ala-Pro-Gly)$_n$ in a triple-helical form, the Ala-NH would point somewhat inward toward the helix center, so it is not totally exposed to solvent. Thus a polar solvent, such as water, might disrupt a triple-helical conformation of (Ala-Pro-Gly)$_n$ by bonding to the NH of alanine, but would not have the same disruptive effect on a (Pro-Ala-Gly)$_n$ triple helix.

It is also possible that some unusual hydrogen-bonding properties are involved in the conformational differences between (Ala-Pro-Gly)$_n$ and (Pro-Ala-Gly)$_n$. Veis and Nawrot (1970) recently proposed that the carbonyl of a proline residue is more basic and thus capable of forming stronger hydrogen bonds than the carbonyl groups of serine, alanine, or glycine. The hydrogen bond formed in collagen-like polypeptides with the sequence (Pro-R-Gly)$_n$ and postulated for collagen (Traub *et al.*, 1969) is between the glycine amide hydrogen (N_1H_1) and the proline carbonyl (O_2C_2'), and because of the increased basicity of the proline carbonyl, this should be stronger than a normal NH $\cdots$ O bond. In a triple helix with the sequence (Ala-Pro-Gly)$_n$, if the same hydrogen-bonding pattern $N_1H_1 \cdots O_2$ is present, the hydrogen bond would be formed between the glycine amide hydrogen N_1H_1 and the alanine carbonyl ($C_2'O_2$), not the proline carbonyl, which is now $C_3'O_3$. This hydrogen bond would then be weaker than that found in (Pro-Ala-Gly)$_n$, and could explain the difference in stability of the triple-helical form for these two polymers. One problem with this explanation is that collagen and all triple-helical polypeptides have an unusually high-frequency infrared amide A band (due to the NH-stretching vibration) (Doyle, 1970), and this high frequency implies a weaker hydrogen bond than normal, not the strong hydrogen bond that the above hypothesis concerning hydrogen bonding would predict.

Our investigations on the conformation of (Ala-Pro-Gly)$_n$ and the Ala-Pro-Gly oligomers have suggested possible relations between types of tripeptide sequences in collagen and the structural function of collagen. These studies have also illustrated the usefulness of using complementary physical techniques in solution and in the solid state, and the value of examining a homogeneous set of oligomers to help interpret physical-chemical results on synthetic polypeptides.

References

Andreeva, N. S., Millionova, M. I., and Chirgadze, Y. N. (1963), *in* Aspects of Protein Structure, Ramachandran, G. N., Ed., London, Academic Press, p 137.
Brown, F. R., III (1970), Ph.D. Thesis, Harvard University.
Brown, F. R., III, Carver, J. P., and Blout, E. R. (1969), *J. Mol. Biol. 39*, 307.
Butler, W. (1970), *Biochemistry 9*, 44.
Carver, J. P. (1966), Ph.D. Thesis, Harvard University.
Cowan, P. M., and McGavin, S. (1955), *Nature (London) 176*, 501.
Crick, F. H., and Rich, A. (1955), *Nature (London) 176*, 780.
Doyle, B. B. (1970), Ph.D. Thesis, Harvard University.
Engel, J., Kurtz, J., Katchalski, E., and Berger, A. (1966), *J. Mol. Biol. 17*, 255.
Greenberg, J., Fishman, L., and Levy, M. (1964), *Biochemistry 3*, 1826.
Hannig, K., and Nordwig, A. (1967), *in* Treatise on Collagen, Vol. 1, Ramachandran, G. N., Ed., London, Academic Press, p 73.
Hopfinger, A. J., and Walton, A. G. (1970), *Biopolymers 9*, 29.
Huggins, C. M., and Pimental, G. C. (1956), *J. Phys. Chem. 60*, 1685.
Kobayashi, Y., Sakai, R., Kakiuchi, K., and Isemura, T. (1970), *Biopolymers 9*, 415.
Lorenzi, G. P., Doyle, B. B., and Blout, E. R. (1971), *Biochemistry 10*, 30460.
Miyazawa, T. (1960), *J. Chem. Phys. 32*, 1647.
Miyazawa, T. (1962), *in* Polyamino Acids, Polypeptides, and Proteins, Stahmann, M., Ed., Madison, Wis., University of Wisconsin Press, p 201.
Oriel, P. J., and Blout, E. R. (1966), *J. Amer. Chem. Soc. 88*, 2041.
Piez, K., Bladen, H., Lane, J., Miller, E., Bornstein, P., Butler, W., and Kang, A. (1968), *Brookhaven Symp. Biol. 21*, 345.
Ramachandran, G., and Kartha, G. (1955), *Nature (London) 176*, 593.
Sasisekharan, V. (1959), *Acta Crystallogr. 12*, 897.
Schwartz, A., Andries, J. C., and Walton, A. G. (1970), *Nature (London) 226*, 161.
Segal, D. (1969), *J. Mol. Biol. 43*, 497.
Segal, D., and Traub, W. (1969), *J. Mol. Biol. 43*, 487.
Segal, D., Traub, W., and Yonath, A. (1969), *J. Mol. Biol. 43*, 519.
Traub, W. (1969), *J. Mol. Biol. 43*, 479.
Traub, W., and Yonath, A. (1966), *J. Mol. Biol. 16*, 404.
Traub, W., and Yonath, A. (1967), *J. Mol. Biol. 25*, 351.
Traub, W., Yonath, A., and Segal, D. (1969), *Nature (London) 221*, 914.
Veis, A., and Nawrot, C. F. (1970), *J. Amer. Chem. Soc. 92*, 3910.

Studies on the Conformational Isomers of Deoxycytidylate Aminohydrolase*

Mosè Rossi, Ilka Dosseva,† Michele Pierro,‡ Marcello G. Cacace, and Eduardo Scarano§

ABSTRACT: Binding of regulatory ligands to deoxycytidylate aminohydrolase induces in the enzyme conformational changes that have been investigated by means of titration of thiol groups, reactivity of amino acid residues, difference spectra, quenching of fluorescence, and time rate of inactivation with proteolytic enzymes.

The data reported here indicate the existence of at least three conformational isomers of deoxycytidylate aminohydrolase, namely, the conformation of the enzyme with no ligands, the conformation of the enzyme–dCTP–Mg complex, and the conformation of the enzyme–dTTP–Mg complex.

Deoxycytidylate aminohydrolase is a typical allosteric enzyme. Its activity is regulated by highly specific allosteric effectors (Geraci *et al.*, 1967; Scarano *et al.*, 1967a). Deoxycytidine triphosphate is the allosteric activator and deoxythymidine triphosphate is the allosteric inhibitor. In the presence of Mg ions the enzyme binds four molecules of dCTP or four molecules of dTTP at saturating concentrations of the activator or of the inhibitor, respectively. With a homogeneous preparation of deoxycytidylate aminohydrolase we have described homotropic and heterotropic effects for the substrate and for the allosteric and isosteric effectors (Rossi *et al.*, 1967; Scarano *et al.*, 1967b), and have shown that no change in molecular weight of the enzyme occurs on binding of the regulatory ligands (Geraci *et al.*, 1967; Scarano *et al.*, 1967a). The enzyme with no ligands, enzyme–dCTP–Mg complex, and enzyme–dTTP–Mg complex all have a molecular weight of 1.2×10^5. Thus, monomer–polymer equilibrium plays no role in the regulation of the activity of deoxycytidylate aminohydrolase. In previous papers we have reported kinetic experiments which suggest the occurrence

* From the International Institute of Genetics and Biophysics, Naples, Italy, the Institute of Molecular Biology, University of Naples, Naples, Italy, and the Laboratory of Molecular Embryology, Naples, Italy. *Received October 20, 1970.* Preliminary reports of these findings were presented at the 4th, Oslo, 1967 (Scarano *et al.*, 1968), and at the 6th, Madrid 1969, Meetings of the Federation of the European Biochemical Societies. This is the sixth paper of a series on deoxycytidylate aminohydrolase. The preceding paper is published in *Biochemistry 9*, 2539 (1970).

† Fellowship from the International Atomic Energy Agency; present address: Bulgarian Academy of Sciences, Research Group on Epilepsy, Sofia 31, Bulgaria.

‡ Present address: Instituto di Puericultura, University of Naples, Italy.

§ Address correspondence to Professor E. Scarano, International Institute of Genetics and Biophysics.