Polypeptide Models of Collagen. II. Solution Properties of (Pro-Gly-Phe)$_n$

SAMIR K. BRAHMACHARI and V. S. ANANTHANARAYANAN,*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India; and RAO S. RAPAKA and RAJENDRA S. BHATNAGAR, *School of Dentistry, University of California, San Francisco, California 94143*

Synopsis

The conformation of (Pro-Gly-Phe)$_n$ in trifluoroethanol was investigated using CD, nmr and ir techniques. After making appropriate correction for the contribution of the phenylalanine chromophore to the observed CD spectra of the polytripeptide at several temperatures, it is found that (Pro-Gly-Phe)$_n$ can exist in a partially triple-helical conformation in this solvent at low temperatures. The nmr and ir data support this conclusion. In conjunction with recent theoretical sutdies, our data offer an explanation for the preferential occurrence of the Phe residue in position 2 of the tripeptide sequence Gly-R_2-R_3, in collagen.

INTRODUCTION

Phenylalanine is the major aromatic amino acid residue occurring in collagen. It should therefore be of interest to inquire about the conformational role of this amino acid in the collagen structure. An examination of the primary sequence of the α-1 chain of collagen[1,2] reveals that out of the total of 12 phenylalanine residues, 7 occur in the sequence of the type -Gly-Phe-Hyp-, while 5 appear as -Gly-Phe-R_3 (where R_3 is an amino acid at position 3 with Gly in position 1). Thus, Phe exhibits a unique preference for the 2 position in the tripeptide sequence as contrasted by several other amino acid residues which occur[1,2] in both positions 2 and 3. A similar preference by Phe for position 2 is also seen in the available primary sequence of the α-2 chain of collagen.[1,2] On the belief that the basis for this specific preference could lie in the conformational features of the tripeptide unit, we have studied the conformation of (Pro-Gly-Phe)$_n$ in solution. As in our earlier study,[3] an additional motivation of the selection of this model is to understand, in general, the role of the second residue in the sequence -Gly-R_2-Pro- in the hydroxylation of the third-position proline residue in collagen by peptidyl proline hydroxylase.[4]

The results obtained by using techniques of CD, nmr, and ir, show that (Pro-Gly-Phe)$_n$ can take up collagenlike conformation under favorable conditions. Thus, in agreement with recent theoretical studies,[5,6] the presence of the bulky phenyl ring in the second position does not appear

* To whom correspondence should be addressed.

Biopolymers, Vol. 17, 2097–2105 (1978)

0006-3525/78/0017-2097$01.00

to pose any stereochemical problem in the development of the triple-helical structure. A preliminary report of this work has appeared elsewhere.[7]

MATERIALS AND METHODS

Samples of (Pro-Gly-Phe)$_n$ were synthesized from the monomer *N*-Ac-Pro-Gly-Phe-OH. The details of the synthesis have been given elsewhere.[8] These samples were fractionated and their molecular weights determined by column chromatography using Sephadex. Two samples with molecular weights of 4100 and 11,000 have been used in these studies. *N*-Ac-Phe-ethyl ester was obtained from Sigma Chemical Co. The (Pro-Sar-Gly)$_n$ sample of molecular weight 3000 was the same as that used in our earlier study.[3] Spectroscopic grade trifluoroethanol (TFE) was obtained from Aldrich Chemical Co. Circular dichroism spectra were recorded with a Jasco J-20 spectropolarimeter on solutions containing 0.01–0.04% (w/v) of the polymer or the model compounds. The wavelength region was 300–205 nm, below which measurements were not feasible due to the high absorption of the polypeptide samples. The molar ellipticity values, $[\theta]$ (in deg cm^2 dmole^{-1}), for the polymer were expressed in terms of the mean residue weight. Nuclear magnetic resonance measurements on solutions of about 1% (w/v) polymer concentration were made using a Varian HA-100 spectrometer. Infrared spectra were obtained with a Carl-Zeiss UR-10 spectrophotometer using samples in nujol mull.

RESULTS AND DISCUSSION

Since the polytripeptide (Pro-Gly-Phe)$_n$ was insoluble in water, the CD spectra were obtained in solutions of TFE in which the samples dissolved freely. TFE is known to promote the collagen-type triple helix,[3,9] and hence one would expect the tendency, if any, for a polytripeptide to adopt this structure to be readily manifested in TFE. The CD spectra of (Pro-Gly-Phe)$_n$ of molecular weight 11,000 in TFE, at different temperatures, are shown in Fig. 1. Almost similar spectra (with a slight reduction in the magnitude of the ellipticity) were obtained with the sample of molecular weight 41,000. The CD of the polytripeptide in the 250–300-nm wavelength region showed a fine structure with a small ellipticity value whose magnitude was almost the same as that observed for simpler compounds containing the Phe residues,[10,11] including the monomer *N*-Ac-Pro-Gly-Phe-OH. This indicates the absence of any additional contribution by the polymer to the CD of the phenyl chromophore in this wavelength region.

The CD spectra of the polytripeptide in the region of 207–250 nm (Fig. 1) appear rather complex and do not immediately reveal any resemblance to the spectra of either the native triple-helical collagen molecule or the random coil form as in denatured collagen.[12] In fact, these spectra do not have the characteristics (either shape or magnitude) of any of the known

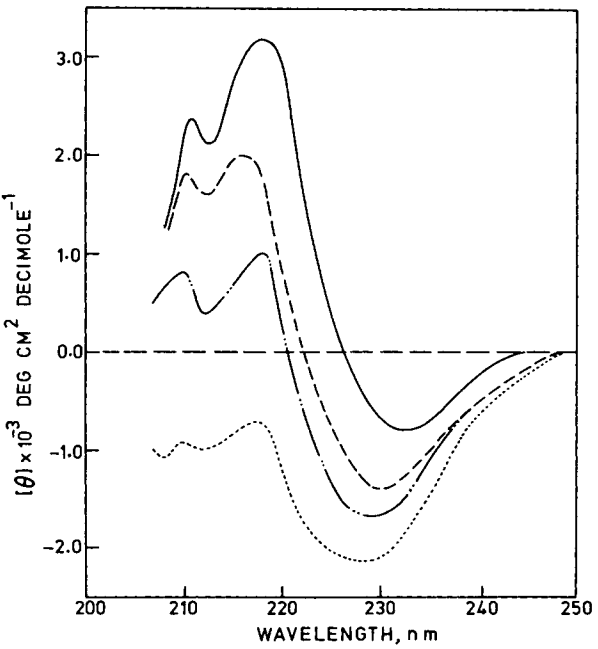

Fig. 1. CD spectra of (Pro-Gly-Phe)$_n$ (molecular weight: 11,000) in TFE: (—) at −20°C, (– –) at 3°C, (-· -·) at 25°C, (·· ··) at 70°C.

secondary structures in proteins, such as the α-helix or the β-structure. They have a resemblance to the theoretically calculated spectra for a special type of β-turn,[13] but the magnitudes do not agree. We realized, however, that the major source of complexity of the (Pro-Gly-Phe)$_n$ spectrum could arise from the Phe residue. It is well known that polypeptides and simpler compounds containing this residue exhibit CD spectral characteristics which include contribution from the aromatic ring.[10,11,14] It therefore became imperative to correct for the contribution from the phenyl chromophore, in order to arrive at the CD spectrum of the peptide backbone in this wavelength region. We have achieved this correction as described below.

First, we obtained the CD spectra of the model compound, *N*-Ac-Phe-ethyl ester, between 200 and 240 nm in TFE at the four temperatures used for the polytripeptide sample (Fig. 2). Corresponding to the absorption spectrum of the phenyl chromophore in the far-uv region (which shows maxima at 210 and 190 nm involving the $\pi \rightarrow \pi^*$ transition), the CD spectra show a shoulder around 217 nm and indicate a maximum around 196 nm[10,11]; the latter is not shown in Fig. 2, owing to the excessive noise in the far-uv region at the extreme temperatures. (The origin of optical activity of the phenyl chromophore in Phe has been discussed by Goodman et al.[15]). There appears to be very little, if any, contribution from the N-terminal amide group of this compound in this wavelength region, since the spectra of *N*-Ac-Phe-methyl ester[10] and *N*-Ac-Phe-ethyl ester obtained by us show

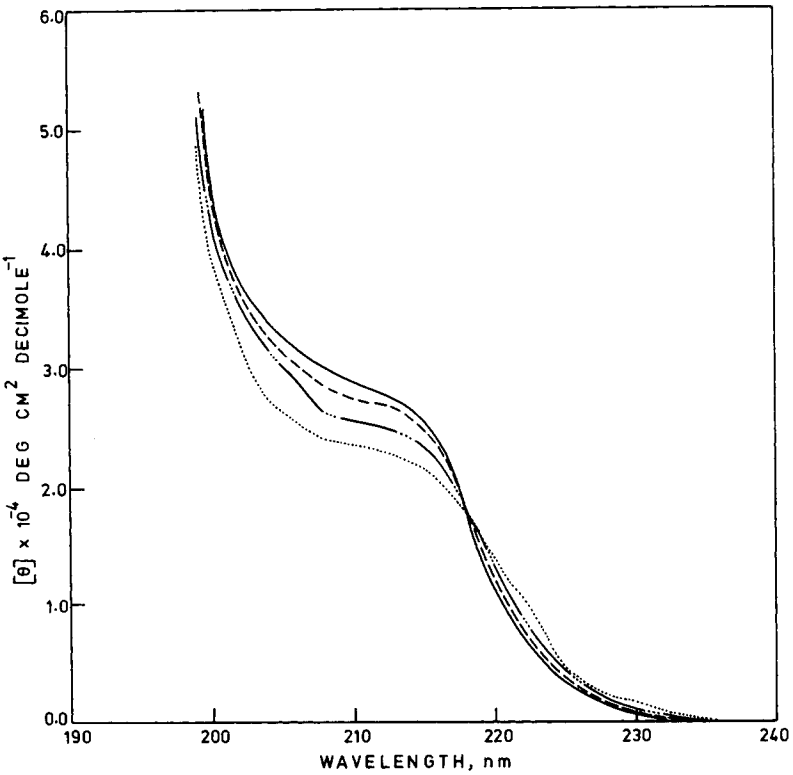

Fig. 2. CD spectra of *N*-Ac-Phe-ethyl ester in TFE. Temperatures and symbols are the same as in Fig. 1.

the same far-uv pattern as that of the free amino acid in the same solvent. Increasing the temperature causes small changes in the ellipticity values at different wavelengths. We have subtracted the CD spectrum of the model compound *N*-Ac-Phe-ethyl ester from that of $(Pro\text{-}Gly\text{-}Phe)_n$ at the respective temperatures. (In order to achieve this correction on a molar basis with respect to the Phe residue, the ellipticity values of the model compound were divided by a factor of three before subtraction from the polymer spectra.) The corrected spectra are shown in Fig. 3. It can be seen that the spectrum at the lowest temperature (−20°C) resembles that of $(Pro\text{-}Gly\text{-}Ala)_n$ in trifluoroethanol obtained by Doyle et al.[9] and that of $(Pro\text{-}Gly\text{-}Sar)_n$ in ethylene glycol-hexafluoroisopropanol mixture obtained by us.[3] In all these cases, the spectra show a band (appearing either as a maximum or as a shoulder) near 220 nm, which is characteristic of the triple-helical structure.[3,16] We can therefore conclude that, at low temperatures, $(Pro\text{-}Gly\text{-}Phe)_n$ exists as a mixture of triple-helical and random coil structures in TFE, as was the case with $(Pro\text{-}Gly\text{-}Ala)_n$ in TFE studied by Blout and coworkers.[9] With increasing temperatures, the maximum around 220 nm decreases and finally vanishes, due to the collapse of the triple-helical structure.

It was of interest to see whether the contribution from the phenyl chro-

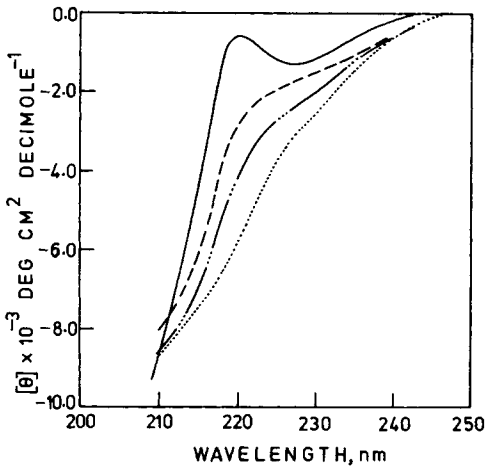

Fig. 3. CD spectra of $(Pro\text{-}Gly\text{-}Phe)_n$ in TFE after correction for the Phe contribution from the data in Fig. 2 (see text for details). Temperatures and symbols are the same as in Fig. 1.

mophore to the $(Pro\text{-}Gly\text{-}Phe)_n$ CD spectrum can be obtained by subtracting the CD spectrum of the monomeric unit, namely, *N*-acetyl-Pro-Gly-Phe-OH (which was the starting material for the polytripeptide synthesis; see Materials and Methods), from the spectrum of the polytripeptide. We found, however, that the monomer itself possessed an ordered structure in TFE, making it unsuitable for correction purposes. In fact, the subtraction of the spectrum of *N*-Ac-Phe-ethyl ester from that of the monomeric tripeptide yielded a β-turn CD spectrum for the latter; the presence of the β-turn was subsequently confirmed by nmr and ir techniques.[17] This observation thus lends support to the validity of the subtraction procedure described in the previous paragraph, using the model compound *N*-Ac-Phe-ethyl ester.

Another procedure that we have adopted to eliminate the contribution from the phenyl chromophore to the spectra of the polytripeptide shown in Fig. 1 is to subtract the spectrum at the highest temperature (70°C) from those at the lower temperatures and thus obtain a set of difference spectra. If the polytripeptide possessed a relatively more ordered structure at the lower temperatures than at the highest temperature, the difference spectra can be expected to represent the contribution from the excess ordered structure alone at these temperatures. An additional correction is, however, necessary to take into account the temperature dependence of the CD bands of the Phe chromophore which we obtained from the data in Fig. 2. The difference spectra thus obtained are shown in Fig. 4 and compared with the corresponding difference spectrum for collagen (obtained by subtraction of the spectrum of denatured collagen[12] from that of native collagen). Also shown in Figure 4, for comparison, is the difference spectrum between $(Pro\text{-}Gly\text{-}Sar)_n$ in ethylene glycol at 25°C (where it is known[3] to exist in a partially triple-helical conformation) and at 95°C (where it is essentially

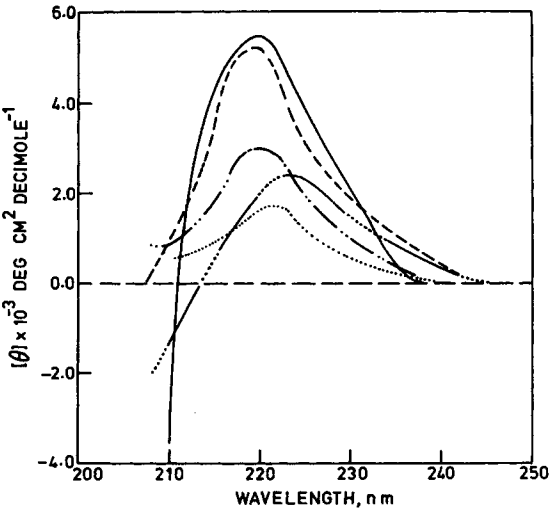

Fig. 4. CD difference spectra of: (—) collagen in water (native vs denatured); (—·····—) (Pro-Gly-Sar)$_n$ in ethylene glycol (25°C vs 95°C); and (Pro-Gly-Phe)$_n$ at (- - - -) −20°C, (-· -·) 3°C, and (· · · ·) 25°C, against spectrum at 70°C in TFE.

in the disordered state[3]). It can be seen that the difference spectra of (Pro-Gly-Phe)$_n$ at lower temperatures share the essential features, including the positive maximum around 220 nm, with those of collagen and of (Pro-Gly-Sar)$_n$ in ethylene glycol. We can thus conclude that (Pro-Gly-Phe)$_n$ at the lowest temperature (−20°C) has elements of the triple-helical structure, while at higher temperatures the magnitude of the positive band at 220 nm is reduced, indicating the progressive collapse of this structure. An assumption involved in this subtraction is that the contribution of the Phe residue to the observed CD is independent of the conformational state of the polytripeptide. This appears to be a valid assumption in view of the similarity of the difference spectra of collagen and the polytripeptide and in view of the absence of specific interaction involving the Phe residue in the triple-helical structure as suggested by model building and theoretical calculations[6] and by the near-uv CD and nmr data (see below) of the polytripeptide.

The procedures described above for correcting the observed CD spectrum of (Pro-Gly-Phe)$_n$ for the Phe contribution seem to be consistent in leading to the same conclusion about the conformation of the polytripeptide. Additional evidence for the presence of ordered structure in (Pro-Gly-Phe)$_n$ at low temperatures comes from nmr measurements on the polytripeptide at different temperatures. The peak due to the phenyl ring appearing at 7.23δ was found to be broad at the low temperatures, possibly due to the ordered conformation of the backbone; as the temperature increases, the peak sharpens progressively, showing the breakdown of the order. The position of the peak, however, remains the same at all temperatures, indicating the absence of specific interaction between the phenyl rings, as earlier concluded from CD data. The plot of the width at half-peak height against

temperature is shown in Fig. 5(a). In Fig. 5(b) the magnitude of $[\theta]_{220}$ (obtained from the CD data similar to those shown in Fig. 1) is plotted against temperature at 5° intervals from 3 to 70°C. From the arguments given earlier, the magnitude of the CD band at 220 nm can be assumed to be proportional to the degree of triple-helical order in the polytripeptide. The sigmoidal curves obtained from the nmr and CD data (Fig. 5) would thus indicate the collapse of the ordered triple-helical structure on heating, which is complete at about 50°C. This justifies the use of the high-temperature (70°C) data in Fig. 1 as representing the random coil structure in the subtraction procedure for CD spectral data (Fig. 4) described in the previous section.

The ir spectrum of the $(Pro\text{-}Gly\text{-}Phe)_n$ is shown in Fig. 6. The amide I bands are observed at 1674–1676 cm^{-1} and at 1648–1650 cm^{-1}, while the amide II band appears at 1545–1550 cm^{-1} and the amide A band is at 3320 cm^{-1}. These are close to those given by Doyle et al.[18] for collagen and a series of model polytripeptides. As pointed out by these authors, the amide A band of collagen and other triple-helical polytripeptides always appears about 25 cm^{-1} higher than that observed for other secondary structures

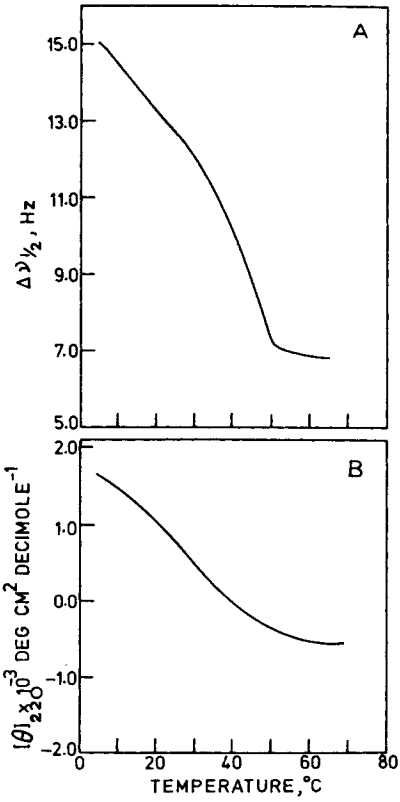

Fig. 5. (A) The nmr line-width of the phenyl proton resonance signal (at 7.23δ) and (B) mean residue molar ellipticity at 220 nm (from Fig. 1) for $(Pro\text{-}Gly\text{-}Phe)_n$ in TFE plotted against temperature.

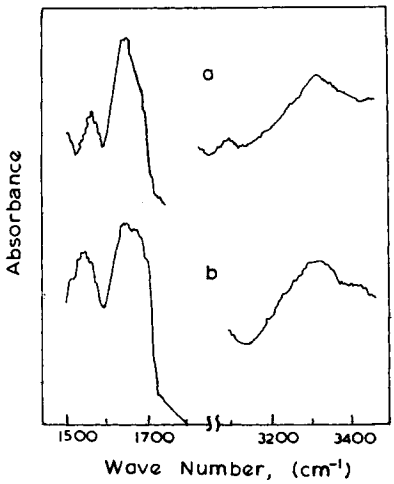

Fig. 6. The ir spectra of (a) (Pro-Sar-Gly)$_n$ and (b) (Pro-Gly-Phe)$_n$ in nujol mull.

and can thus be used specifically to identify the collagenlike structure. For example, while (Pro-Sar-Gly)$_n$ existing in the triple-helical conformation[3] shows the band around 3320 cm^{-1} (shown in Fig. 6 for comparison), (Pro-Gly-Sar)$_n$, which has a random-coil structure (in the absence of helix-promoting solvents),[3] does not exhibit this band.

Theoretical studies have recently been carried out by Traub[5] and Bansal[6] on the steric interactions of various amino residues in the collagenlike tripeptide sequence. With regard to the Phe residue, Traub[5] had concluded that it can easily be accommodated in position 2 in the Gly-R_2-R_3 sequence, but offers severe steric hindrance to the triple-helical conformation in the 3 position. The more recent studies of Bansal[6] also indicate that position 2 is stereochemically allowed for the Phe residue and is preferred over position 3 (which, while sterically restricted, is not totally disallowed). Thus, both the theoretical studies and the amino acid sequence data on collagen[1,2] support our experimental observations on (Pro-Gly-Phe)$_n$.

During the preparation of the manuscript, we learned about a recent paper by Tamburro et al.,[19] who have studied the conformation of (Pro-Gly-Phe)$_n$ and (Pro-Phe-Gly)$_n$. These authors have reported a β-bend conformation for (Pro-Gly-Phe)$_n$ in solution, mainly on the basis of the CD spectra. Their CD spectrum in TFE is identical to that reported here. However, no correction for the Phe contribution to the spectrum in the 200–250-nm region was made by these authors, either for this polytripeptide or (Pro-Phe-Gly)$_n$. The latter was also reported as existing in a β-bend conformation in the presence of ethylene glycol. As we have shown in this paper, the correction for Phe contribution is found to be necessary for meaningful interpretation of the CD data.

Applying the correction methods described above to the data of Tamburro et al.[19] on (Pro-Phe-Gly)$_n$, we find that this polytripeptide is essentially a random coil in TFE (at 22°C). It thus appears that (Pro-Gly-

Phe)$_n$ is more stable in the triple-helical conformation than (Pro-Phe-Gly)$_n$ in the same solvent system, namely, TFE. This is in contrast with observations on several other polytripeptides,[3,9,16] where the (Pro-Gly-R_2)$_n$ type are less stable in the triple-helical structure than the (Pro-R_3-Gly)$_n$ type in solution. The destabilizing effect of the Phe residue in the 3 position in the Gly-R_2-R_3 tripeptide sequence can be understood in terms of the severe steric hindrance[5,6] due to the bulky phenyl ring in position 3, but not in position 2. Thus, both the experimental evidence and theoretical calculations offer an explanation for the occurrence of Phe only in position 2 in the amino acid sequence of collagen.

The work reported here was supported at Bangalore by a grant from the Council of Scientific and Industrial Research and D.S.T. (India) and at San Francisco by a USPHS NIH Grant.

References

1. Piez, K. A. (1976) in *Biochemistry of Collagen,* Ramachandran, G. N. & Reddi, A. H., Eds., Plenum, New York, pp. 1–44.

2. Fietzek, P. P. & Kuhn, K. (1976) *Int. Rev. Connect. Tissue Res.* **7,** 1–60.

3. Ananthanarayanan, V. S., Rapaka, R. S., Brahmachari, S. K. & Bhatnagar, R. S. (1976) *Biopolymers* **15,** 707–716.

4. Bhatnagar, R. S. & Rapaka, R. S. (1976) in *Biochemistry of Collagen,* Ramachandran, G. N. & Reddi, A. H., Eds., Plenum, New York, pp. 479–523.

5. Traub, W. (1974) *Isr. J. Chem.* **12,** 435–439.

6. Bansal, M. (1977) *Int. J. Pept. Protein Res.* **9,** 224–234.

7. Brahmachari, S. K. & Ananthanarayanan, V. S. (1976) Paper presented at the 45th Annual Meeting of the Society of Biological Chemists (India), Hyderabad. [Abstracted in *Ind. J. Biochem. Biophys.* (Suppl. 5) **14,** 1977.]

8. Rapaka, R. S. & Bhatnagar, R. S. (1975) *Int. J. Peptide Protein Res.* **7,** 119–128.

9. Doyle, B. B., Traub, W., Lorenzi, G. P. & Blout, E. R. (1971) *Biochemistry* **10,** 3052–3060.

10. Adler, A. J., Greenfield, N. J. & Fasman, G. D. (1973) *Methods Enzymol.* **27,** 675–735.

11. Goodman, M. & Toniolo, C. (1968) *Biopolymers* **6,** 1673–1689.

12. Tiffany, M. L. & Krimm, S. (1972) *Biopolymers* **11,** 2309–2316.

13. Woody, R. W. (1974) in *Peptides, Polypeptides and Proteins,* Blout, E. R., Bovey, F. A., Goodman, M. & Lotan, N., Eds., Wiley, New York, pp. 338–350.

14. Peggion, E., Verdini, A. S., Cosani, A. & Scoffone, E. (1969) *Macromolecules* **2,** 170–174.

15. Goodman, M., Davis, W. G. & Benedetti, E. (1968) *Acc. Chem. Res.* **1,** 275.

16. Brown, F. R., Corato, A. D., Lorenzi, G. P. & Blout, E. R. (1972) *J. Mol. Biol.* **63,** 85–99.

17. Brahmachari, S. K., Ananthanarayanan, V. S., Rapaka, R. S. & Bhatnagar, R. S. (unpublished).

18. Doyle, B. B., Bendit, E. G. & Blout, E. R. (1975) *Biopolymers* **14,** 937–957.

19. Tamburro, A. M., Scatturin, A. & Del Pra, A. (1977) *Int. J. Pept. Protein Res.* **9,** 310–318.

Received October 28, 1977
Accepted January 6, 1978