# Structural studies of collagen-like sequential polypeptides*

**E. Heidemann, H. G. Neiss, K. Khodadadeh, G. Heymer, E. M. Sheikh and Ö. Saygin**
*Macromolecular Chemistry, Technische Hochschule Darmstadt, W. Germany*
*(Received 19 October 1976)*

Sequential polyhexapeptides, synthesised by combination of sequences from collagen type Gly–X–Y (X = Ala, Pro, Ser; Y = Ala, Gly, Lys, Pro), were characterized by the temperature dependence of circular dichroism spectra. Under comparable conditions these studies revealed that alternating triplets of Gly–Pro–Pro or Gly–Pro–Ala combined with Gly–Pro–Lys or Gly–Pro–Glu exhibit collagen-like structures in aqueous solutions. In case of unstructured chains of (Gly–Pro–Ala) ≈ 12 it can be shown that N-terminal crosslinking of three chains produces a similar ordered structure.

## INTRODUCTION

The sequence of the helical portion of the collagen molecule is composed of different types of tripeptides. The well-known strong triple helix forming tripeptides, Gly–Pro–Hyp and Gly–Pro–Pro, have been studied in their oligo- or poly-tripeptides in solution; in the solid state and at varying chain lengths. The question arises as to how the other types of triplets with low or zero imino-acid content contribute to the stability of the triple helix because their total amount in the molecule is about 90%.

Up till now a number of collagen-like polytripeptides of the type Gly–X–Y from selected amino-acids were synthesised and their conformation studied in solution and in the solid state. (For reviews see Piez and Traub[1] and Goren[2].)

The exact determination of the structure of a collagen model in solution and in the solid state is generally difficult. However, the collagen structure is the most probable one, if one experimentally excludes the α-helix and the β-structure in such polypeptides. This could be supported by the assumption that tripeptides which are typical in the collagen sequence should predominantly form triple helix type structures.

Certainly various types of tripeptides have a different tendency to form helical structures. In order to study that, the following tripeptides were synthesised: Gly–Ala–Ala, Gly–Pro–Ser, Gly–Ser–Pro, Gly–Pro–Ala, Gly–Pro–Glu, Gly–Pro–Lys, Gly–Glu–Lys.

In addition, we combined these triplets to heterohexapeptides by attaching Gly–Pro–Pro sequences to them, assuming adjacent triplet mutual influence.

Collagen model peptides in solvents of different polarity were, also, examined. However, only the results obtained in water are relevant to the understanding of the forces in native structure. Furthermore, we have taken into account that the evaluation of the stability of the helix formation demands comparable methods and conditions, e.g. solvent composition, chain length etc. We analysed the thermal transition curves obtained by circular dichroism spectra.

Previous studies of collagen have shown that the triple helix folding is a cooperative process. The triple helix formation is slower than the α-helix formation. Studying the thermodynamics[3,4] and the kinetics[5,6] of the folding of collagen models of different sequences and lengths renders an insight into the underlying mechanism.

In interpreting folding studies, we have kept in mind the possibility of non-specific alignment of the 3 chains. This property, caused by backfolding[3] and non-specified aggregation[7] of different types of triplets (also called hybridization), is typical for collagen single chains. Therefore, we have chosen polypeptides with chain lengths allowing for reversible thermal transition. In order to overcome slow folding reactions, which are clearly a matter of statistically probable collisions of the individual chain, three chains were crosslinked at their N-terminals. In this paper we present results showing that those specific chains produce a fast folding reaction.

## EXPERIMENTAL

The synthesis of the monomers and the polycondensation reaction has been described elsewhere[8–11]. Purification and molecular weight determinations of the polypeptides were carried out by gel chromatography on standardized columns using Sephadex G 50s and Biogel P4. The columns were calibrated with oligomers obtained from stepwise peptide synthesis and with cyanogenbromide peptides, $\alpha_1$ CB3, $\alpha_1$ CB4, $\alpha_1$ CB5, of the $\alpha_1$-chain of calf skin collagen.

Some standards were checked by analytical ultracentrifugation and compared to those of gel filtration. All peptides were stored in the appropriate solvent for 5 days or more at low temperature (2°–10°C) before carrying out studies. The concentration was approximately 2.5 mg/ml. C.d. spectra were recorded after stepwise enhancing of the temperature and after attainment of equilibrium at each temperature. The Cary 60 spectropolarimeter with c.d. equipment 6002 was used. The $|\theta|$ values were calculated for the mean molecular weights of the respective amino-acids.

## RESULTS

In spite of the above-mentioned restrictions, an insight into

* Presented at the First Cleveland Symposium on Macromolecules, Structure and Properties of Biopolymers, Case Western Reserve University, Cleveland, Ohio, USA, October 1976.

the tendency of triple helix folding can be derived from the folding formation or from the unfolding reaction. This was followed by the temperature dependence of the c.d. signals

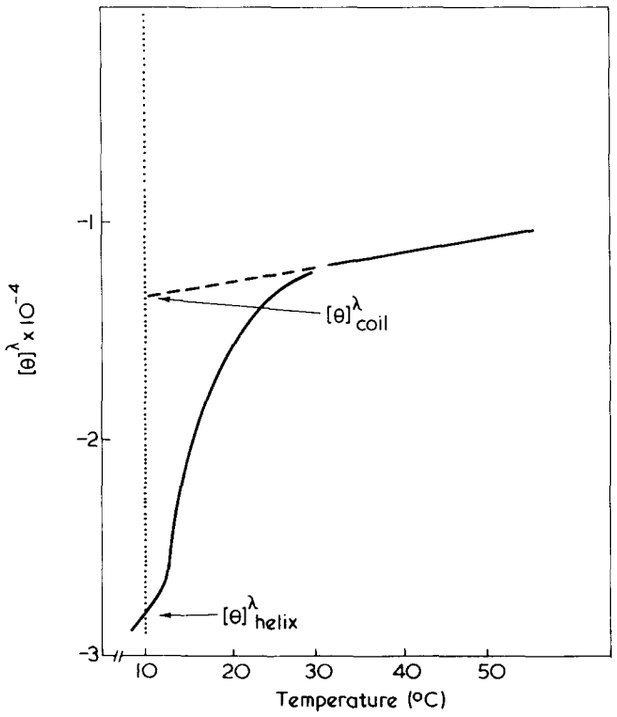


*Figure 1* Diagram illustrating the determination of the structure index *Si*

in solution after incubation at low temperatures. The bands at 198 nm and, in the case of a more stable structure, at 225 nm, show a sigmoidal shape dependent upon temperature. In the range between 10° and 50°C we define a structure index (*Si*) as the ratio of the $\theta$ values at 10°C divided by those at 50°C (*Figure 1*).

$$Si = \frac{|\theta|^{\lambda}\ \text{helix}}{|\theta|^{\lambda}\ \text{coil}}$$

Between 1.0 for non-sigmoidal shape and none structure and 5.0 for collagen all gradiations can be found. It can be seen from *Table 1* that Gly–Pro–Pro sequences in the polypeptide chain stimulates helix formation of the tripeptides Gly–Pro–Glu or of Gly–Pro–Lys if they coupled to polyhexapeptides. However, in the case of a combination with Gly–Glu–Lys the stimulation of the folding is very poor.

The sequence Gly–Pro–Ala has been extensively investigated[9] in its oligomer forms. Helix formation starts at $n = 6$, but even at $n$ = 12–13 the c.d. spectrum shows no maximum at 225 nm. However, the maximum is reached at $n \approx 31$. The sigmoidal shape is significantly expressed with increasing chain length at 198 nm. Likewise, increasing chain lengths at 225 nm approaches stepwise the zero line, shown in *Table 2*. Substances with $|\theta|_{198nm}$ values lower than 15 000 render no sigmoidal melting curves. Substances with maxima at 225 nm have a $|\theta|_{198nm}$ from 20 000–25 000 and more.

OH
|
(Gly–Pro–Glu)$_n$ gives no indication of folding but in com-

*Table 1* Sequential collagen models characterized by structure index obtained from circular dichroism spectra

| Polypeptide | $n$ | Solvent | $\|\theta\|_{198nm}$ | $\|\theta\|_{225nm}$ | Ratio of amplitudes 198nm : 225nm |
|---|---|---|---|---|---|
| (Pro–Ala–Gly–Pro–Ala–Gly)$_n$ | 12/13 | Water | −20 000 | − 300 | 100 : <0 |
| | 31 | Water | −45 000 | +2430 | 100 : 5.4 |
| | 31 | HFiP–EG | −52 000 | +5700 | 100 : 11 |
| tBu | 12/13 | Propandiol-1.3 | −35 000 | +2950 | 100 : 8.5 |
| (Pro–Glu–Gly–Pro–Ala–Gly)$_n$ | 10 | Water | −20 000 | + 360 | 100 : 1.8 |
| OH | 10 | HFiP–EG | −37 000 | +1100 | 100 : ~3.0 |
| (Pro–Glu–Gly–Pro–Ala–Gly)$_n$ | >10 | HFiP–EG | −47 000 | +4700 | 100 : 10 |
| (Gly–Pro–Glu–Gly–Pro–Pro)$_n$ | > 5 | HFiP–EG | −27 000 | +2700 | 100 : 10 |
| tBu tBu | | | | | |
| (Pro–Glu–Gly–Pro–Glu–Gly)$_n$ | =10 | HFiP–EG | −16 250 | − 300 | 100 : <0 |
| (Lys–Gly–Pro–Lys–Gly–Pro)$_n$ | >10 | Water | −15 800 | − 100 | 100 : near 0 |
| | >10 | HFiP–EG | −12 800 | | 100 : <0 |
| | >10 | Propandiol–1.3 | − 5 200 | | 100 : <0 |
| | >10 | MeOH–Water | −17 900 | | 100 : <0 |
| (Lys–Gly–Pro–Ala–Gly–Pro)$_n$ | >10 | Water | −15 000 | − 200 | 100 : <0 |
| | >10 | HFiP–EG | −12 700 | | 100 : <0 |
| | >10 | Propandiol 1.3 | −10 000 | | 100 : <0 |
| (Gly–Pro–Lys–Gly–Pro–Pro)$_n$ | >10 | Water | −35 000 | +2200 | 100 : 6.3 |
| | >10 | HFiP–EG | −12 000 | − 500 | 100 : 0 |
| OH | >10 | Propandiol 1.3 | −19 000 | +1000 | 100 : 5.3 |
| (Gly–Glu–Lys–Gly–Pro–Pro)$_n$ | >10 | Water pH 5.5 | −20 000 | + 400 | 100 : 2 |
| Collagen native | | Water | −54 000 | $+8000^{221nm}$ | 100 : 6.7 |
| Collagen random chain | | Water | −13 000 | $-2000^{221nm}$ | |

HFiP = 1,1,1,3,3,3-hexafluoro-2-propanol; EG = ethylene glycol

*Table 2* Circular dichroism parameters for polyhexapeptides and collagen

| Polyhexapeptides | | *n* | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 4 | 5 | 6 | 7 | 8 | 10 | 12 | 31 |
| Pro–Ala–Gly–Pro–Ala–Gly | Water | 1.0 | 1.0 | 1.03 | 1.05 | 1.09 | 1.1 | 1.35 | – |
| | HFiP–EG | – | 1.0 | 1.06 | 1.18 | 1.35 | – | – | 3.14 |
| tBu | Propandiol–1.3 | 1.0 | 1.04 | 1.19 | 1.42 | 1.82 | – | 2.0 | – |
| Pro–Glu–Gly–Pro–Ala–Gly | Water | 1.0 | 10 | 1.15 | 4.6 | – | – | 1.65 | – |
| OH | HFiP–EG | – | – | – | – | – | 10 | – | – |
| Pro–Glu–Gly–Pro–Ala–Gly | Water | 1.0 | 1.0 | 1.03 | | | 1.61 | 1.9 | |
| OH | | | | | | | | | |
| Gly–Pro–Glu–Gly–Pro–Pro | HFiP–EG | – | – | $n > 5$ | | | | | |
| tBu tBu | | | | 2.1 | | | | | |
| Pro–Glu–Gly–Pro–Glu–Gly | HFiP–EG | | | (insoluble in water) | | | 1.0 | | |

| | | *n* | | |
|---|---|---|---|---|
| | | 7 | 8 | 10 |
| Lys–Gly–Pro–Lys–Gly–Pro | Water | 1.0 | 1.06 | 1.17 |
| | HFiP–EG | 1.0 | 1.04 | 1.04 |
| | Propandiol–1.3 | 1.0 | 1.04 | 1.15 |
| | Water: methanol, 6:4 | – | – | 2.35 |
| Lys–Gly–Pro–Ala–Gly–Pro | Water | – | 1.04 | 1.1 |
| | HFiP–EG | – | 1.0 | 1.05 |
| | Propandiol–1,3 | – | 1.0 | 1.02 |
| Glys–Pro–Lys–Gly–Pro–Pro | Water | – | – | 2.0 |
| | HFiP–EG | – | – | 1.0 |
| OH | Propandiol | – | – | 1.1 |
| Gly–Glu–Lys–Gly–Pro–Pro | Water | pH 5.5 | 1.04 | 1.1 |
| | Water | pH 0.9 | – | 1.0 |
| | Water | pH 12 | – | insoluble 1.0 |

| | | *n* | | | | |
|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | >10 |
| Ala–Ala–Gly–Ala–Ala–Gly | Water | 10 | | | 10 | 10 |
| Ala–Ala–Gly–Pro–Ala–Gly | Water | 10 | 10 | – | – | 11 |
| Pro–Ser–Gly–Pro–Ser–Gly | Water | 10 | | | | 10 |

bination with the sequence Gly–Pro–Ala the polyhexapeptide

OH
|
(Pro–Glu–Gly–Pro–Ala–Gly)$_n$

folds more significantly than (Pro–Ala–Gly–Pro–Ala–Gly)$_n$ of the same length. It seems likely that the sequence Pro–Ala–Gly may also stimulate the folding. But this is not the case, when e.g. (Lys–Gly–Pro–Ala–Gly–Pro)$_n$ was compared with (Lys–Gly–Pro–Lys–Gly–Pro)$_n$. In both cases the structure index is smaller than in (Pro–Ala–Gly)$_n$ at respective chain lengths and (Lys–Gly–Pro–Ala–Gly–Pro)$_n$ has a somewhat smaller index than (Lys–Gly–Pro)$_n$.

(Ala–Ala–Gly)$_n$ and (Pro–Ser–Gly)$_n$ form no structure in aqueous solution[12,13]. The stimulating effect in the copolypeptide (Ala–Ala–Gly–Pro–Ala–Gly)$_n$ is very small and lower than in the case of (Ala–Ala–Gly–Pro–Pro–Gly)$_n$ investigated by Segal[1]. We changed the proportion of the sequences of Ala–Ala–Gly and of Pro–Ala–Gly having synthesised (Ala–Ala–Gly–Ala–Ala–Gly–Pro–Ala–Gly)$_n$ and found no variation of structure even when molecular weights of 5000 Daltons were used at 198 nm.

(Pro–Ser–Gly)$_n$ forms a triple helix in organic solvents as reported by Braun *et al.*[12]. This does not resolve the problem whether these sequences promote the triple helix or not. An earlier extensive investigation on structure formation of the polymers of (Pro–Ser–Gly)$_n$, (Ser–Pro–Gly)$_n$, (Pro–Gly–Ser)$_n$, (Gly–Pro–Ser)$_n$, and (Gly–Ser–Pro)$_n$, (Ser–Gly–Pro)$_n$ in aqueous solution, even at molecular weights in some cases of more than 20 000 Daltons, gave no indication of structure[13]. This confirms serine's role in the α-helix formation. Serine seems to have a disarranging effect on helix formation in general.

The low structure index of many polytripeptides and their respective hetero-polytripeptides may provide a possible approach. (Pro–Ala–Gly)$_{n \approx 12}$ was isolated from polycondensation mixtures and was coupled with 1,2,3 carboxypropanic acid-tri-2.4-dinitrophenylester. We obtained a

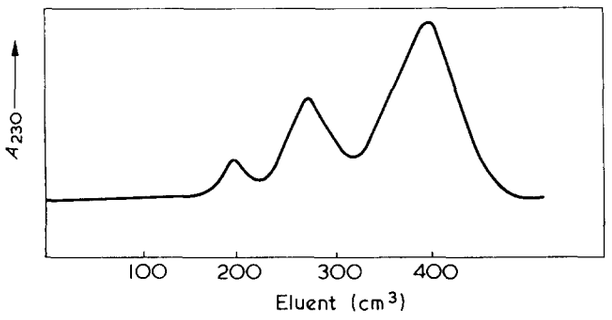


*Figure 2* Chromatography of crosslinked (Pro—Ala—Gly)$_{n\approx12}$ on Sephadex G50S (column: 150 X 3 cm), eluent: 0.05 M acetic acid

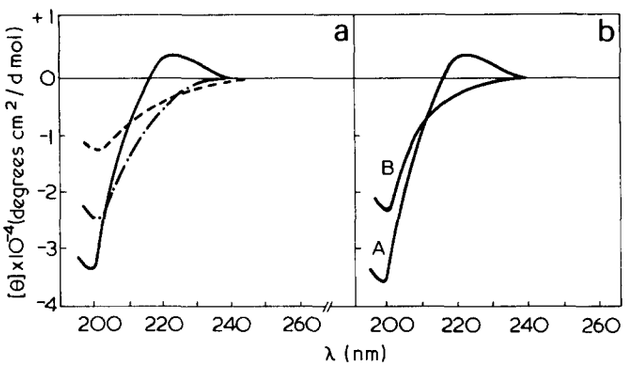


*Figure 3* Circular dichroism spectra of; (a) monomer I (- - -); dimer II (- · - ·); trimer III (——) crosslinked (Pro—Ala—Gly)$_{\approx12}$ at 8° C, (b) III at 8° C (A), at 40° C (B)
Solvent: water; concentration: 2 mg/ml

mixture of the monomer-, dimer- and trimer-crosslinked products determined by gel permeation chromatography in the presence of unfolding agent (urea) or at higher temperatures. Molecular weights of 9100, 6100 and 3000 were found as shown in *Figure 2.*

The crosslinked trimer shows a significant difference in the amplitude of the c.d. bands at 198 mm between the helical and the coiled state. The structure index was found to be 1.6 (*Figure 3b*). The large maximum at 225 nm with a ratio $|\theta|_{198}/|\theta|_{225}$ of 10:1 was surprising. The absolute value of the amplitude at 198 nm with $|\theta|$ 36 000 corresponds roughly to the length of the sum of the three chains.

The comparison of the c.d. spectra of the monomer- with the dimer- and the trimer-crosslinked substances is shown in *Figure 3a.* The amplitudes at 198 nm increase stepwise from the monomer to the trimer. The transition curves measured at 225 nm are shown in *Figure 4.* Both, the monomer and the dimer, give an indication of only a very small structure effect at low temperatures. The trimer shows a higher melting point in the range about 20°C.

*Figure 5* shows the relaxation curves after a temperature jump from 35° to 37°C for the three polymers. The folding of trimer is relatively fast and completed within a few minutes. In later stages no further enhancement of the ellipticity was detected (see the value after 2 weeks). For the dimer and monomer the folding is nearly completed after 45 min, but after two weeks – refer to the points – increased ellipticity was observed. It is remarkable, that this gain in folding is greatest in the case of the monomer. Clearly, the final product of the folding of the trimer crosslinked polymer is the most definitive one.

*Figure 6* shows the ellipticity of the crosslinked trimer after incubation at different concentrations at low temperatures. The respective values are independent of concentration indicating a first order reaction and, therefore, a multimer can be excluded. Surely, in the case of crosslinked

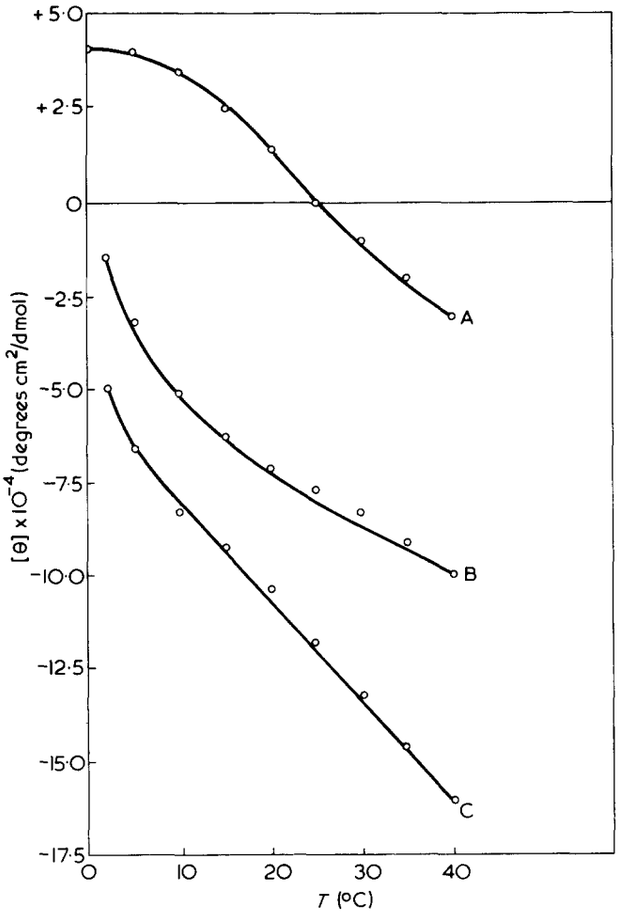


*Figure 4* Thermal transition curves of the monomer, dimer, trimer crosslinked (Pro—Ala—Gly)$_{\approx12}$. Solvent: water; concentration: 2 mg/ml

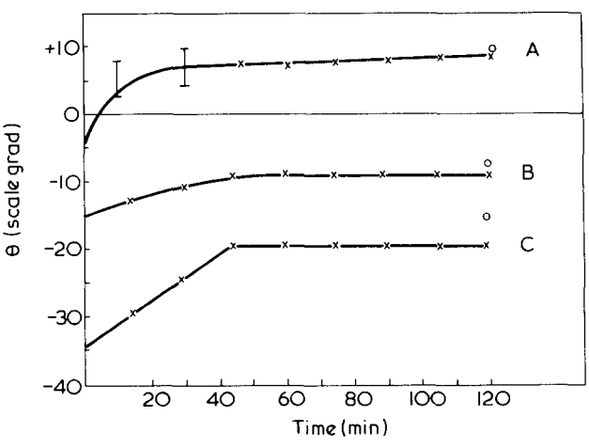


*Figure 5* Relaxation time curves after temperature jump from 35° – 37° C for trimer (A) dimer (B) and monomer (C) crosslinked (Pro—Ala—Gly)$_{\approx12}$ (scale grad). Points at 120 min represent the $\theta$ values after 2 weeks at 5° C. Solvent: water; concentration: 2 mg/ml

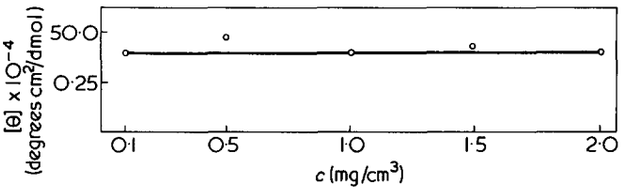


*Figure 6* The mean residue ellipticity of the trimer crosslinked $(Pro–Ala–Gly)_{12}$ after incubation at 5°C in different concentrations

dimer and the monomer, the folding products are composed irregularly and contain more than one polymer molecule.

## DISCUSSION

The c.d. spectra of the synthetic polypeptides are similar to collagen in respect to their shape, sign, and the wavelength of the bands. Therefore, structures from α-helix and β-type cannot be present. Furthermore, especially single chain polyproline structures are excluded, because such structures cannot be stable in triplets with only one or two proline residues and moreover the tendency of hydrogen bond formation cannot be suppressed.

However, backfolding and each form of hybridization must be considered. Therefore, molecular weight determination e.g. ultracentrifugation, light scattering, viscosity and gel permeation chromatography, to follow the folding formation do not indicate a collagen triple helix fold. Extensive folding studies with the greater CB-peptides from the α1-chain of calf skin show a strong hysteresis between folding and unfolding transition curves[4]. Similar experiments with longer chains of synthetic polypeptides from collagen type behave the same. This is a clear indication that non-specific folding, even at high concentrations most probably occur. Kobayashi *et al.*[14] carried out experiments with relatively short chains of $(Gly–Pro–Pro)_n$ of unique chain length. Compared to these findings the most of our polypeptides with short chain lengths were not stable enough to form helices. The trimer crosslinked molecule inevitably fits at the ends of the three chains. Therefore, the higher folding velocity is due to the very high local concentration of the three polymer chains crosslinked at the amino terminals. This is also an expression of the higher stability of the nucleus. The nucleus concentration should be enhanced by this crosslinking and consequently the cooperativity is lowered. The slope of the thermal transition curve in any case seems not to be greater than that from the non-crosslinked polymers. This confirms that no enhancement of cooperativity happens. Kinetic measurements and crosslink preparation of other polyhexapeptides are in progress. However, the analysis of the transition curve from the c.d. temperature dependent measurements has been shown to be an excellent tool in determining the folding formation, whether the chains are in triple helix formation fitted at the ends or not. The introduced structure index is only applicable under the above mentioned restrictions assuming the relative concentration of helical and randomly coiled chains.

## REFERENCES

1 Piez, K. and Traub, W. *Adv. Protein Chem.* 1971, **25**, 243
2 Goren, J. *CRC Crit. Rev. Biochemistry* 1947, p 197
3 Hauschka, P. V. and Harrington, W. F. *Biochemistry* 1970, **9**, 3734
4 Saygin, Ö. *Dissertation* TH Darmstadt (1976)
5 Weidner, H., Engel, J. and Fietzek, P. 'Peptides, Polypeptides and Proteins, (Eds. E. R. Blout, F. A. Bovey, M. Goodman and N. Lotan), Wiley, New York, 1974
6 Seba, J. *Diplomarbeit,* TH Darmstadt (1976)
7 Heidemann, E. R. Harrap, B. S. and Schiele, H. D. *Biochemistry* 1973, **12**, 2958
8 Neiss, H. G. and Heidemann, E. R. *Makromol. Chem.* 1976, **177**, 701
9 Heymer, G. and Heidemann, E. R. *Makromol. Chem.* 1976, **177**, 3299
10 Khodadadeh, K., Heppenheimer, K. and Heidemann, E. R. *Makromol. Chem.* in press
11 Heidemann, E. and Nill, H. W. *Z. Naturforsch.* 1969, **24**, 837, 843
12 Braun III, F. R., Hopfinger, A. H. and Blout, E. R. *J. Mol. Biol.* 1972, **63**, 101
13 Heidemann, E. R. unpublished results
14 Sakakibara, S., Iuoue, K., Shudo, K., Kishida, Y., Kobayashi, Y., Prockop, D. J. *Biochim. Biophys. Acta* 1973, **303**, 198; Kobayashi, Y., Sakai, R., Kakiuchi, K. and Isemura, T. *Biopolymers* 1970, **9**, 415