Type I Collagen Segment Long Spacing Banding Patterns

EVIDENCE THAT THE $\alpha_2$ CHAIN IS IN THE REFERENCE OR A POSITION*

Edward Bender, Frederick H. Silverz, Kimiko Hayashi, and Robert L. Trelstad

From the Department of Pathology, University of Medicine and Dentistry of New Jersey, Rutgers Medical School, Piscataway, New Jersey 08854

(Received for publication, February 19, 1982)

Densitometric scans of electron micrographs of type I collagen segment long spacing crystallites stained with uranyl acetate or phosphotungstic acid and uranyl acetate have been correlated with computer-synthesized scans derived from the sequence of the $\alpha_1(1)$ and $\alpha_2$ chains. Three models that differ in the location of the $\alpha_2$ chain were used in the computer synthesis; Models A, B, and C have the $\alpha_2$ chain in the A, B, and C chain positions, respectively. For all 13 experimental scans, the order of decreasing correlation with the models was found to be A, B, C. The probability of getting the same order of decreasing correlation all 13 times is $6/6^{13}$. It was also determined at the 0.99 confidence level that the mean of the differences in the correlation coefficients among the models is greater than 0, supporting the conclusion that sequence-derived models best fit the experimental data when the $\alpha_2$ chain is in the A position. Our results also agree with recent studies that show that uranyl ions bind to both positively and negatively charged residues on collagen type I.

The triple helical structure of type I collagen that has been proposed based on compositional and x-ray diffraction data is well known (1). Each molecule is composed of 1 $\alpha_2$ and 2 $\alpha_1(1)$ chains wound around each other to form the rigid collagen helix. Each $\alpha$ chain consists of 338 triplets, Gly-X-Y, in the triple helical region, where X and Y can be any amino acid except glycine and tryptophan. At each end of the triple helical region are short nonhelical telopeptides. Each $\alpha$ chain is staggered by 1 or 2 residues in the triple helical region with respect to the other 2 chains so that there is a glycine at every axial position along the molecule. The A chain is the $\alpha$ chain that has the first glycine in the triple helical region starting from the NH$_2$ terminus; the B chain is shifted by 1 residue with respect to the A chain so that its first triple helical glycine is the second of the molecule. The C chain is shifted by 2 residues with respect to the A chain so it has the third triple helical glycine in the molecule. The relative shift between chains allows for residues separated by 1 or 2 residues in neighboring chains to have the same axial position along the molecule (2). In type I and III collagens, this becomes important since many of the charged residues are found in attractive pairs separated by 1 and 2 residues.

In the presence of adenosine triphosphate at acid pH, collagen molecules align without stagger to form a lateral aggregate termed SLS' crystallite. These aggregates have a characteristic banding pattern when stained with heavy metals and viewed in an electron microscope, consisting of 58 bands according to the nomenclature introduced by Bruns and Gross (3). The positive staining pattern generated is the result of interactions between the stain and the charged amino acid residues. Studies by von der Mark et al. (4) suggested that a direct correlation could be made between the position of charged residues and the staining pattern of a 112-residue peptide in the COOH terminus. The staining pattern of the SLS is therefore directly related to the charge distribution of the collagen molecule. Since the positions of the charged residues in $\alpha_1(1)$ and $\alpha_2$ are not identical (5) and the chains are displaced by 1 or 2 residues relative to each other in the molecule, the SLS banding pattern potentially contains information concerning the position of the $\alpha_2$ chain in the triple helix.

The exact nature of inter- and intrachain interactions that stabilize the molecule as well as intermolecular interactions that stabilize fibrils depends on the location of $\alpha_1(1)$ and $\alpha_2$ chains, i.e., whether they are in the A, B, or C position. Several models of intrachain, interchain, and intermolecular interactions have been proposed which demonstrate the role of charge-charge and hydrophobic interactions in the genesis and stability of type I fibrils (2, 5–14). Experimental studies (14) suggest that many of the charged residues on collagen type I are not free but are involved in interactions. Interactions between attractive charged pairs separated by 1 and 2 residues may be characterized by both electrostatic and hydrophobic components (2). In this report, we present indirect evidence that the $\alpha_2$ chain is in the A position.

MATERIALS AND METHODS

UA-stained SLS of rat tail tendon collagen were prepared by dialysis of rat tail tendon collagen solutions of 4 °C versus 0.4% ATP, pH 2.7, at acid pH. A drop of the solution was placed on carbon/collodion-coated copper grids and positively stained with saturated aqueous UA. The pH of the unbuffered staining solution was 3.8. Stained crystallites were viewed with a Philips 300 transmission electron microscope and photographed at a magnification of 90,000. PTA-UA-stained calf skin type I SLS were obtained from Dr. Romaine Bruns, Developmental Biology Laboratory, Massachusetts General Hospital, Boston (3). One SLS was obtained from Ref. 16, and was similar to PTA-UA-stained SLS based on visual comparison. Negatives of UA and PTA-UA-stained SLS were scanned with a Gilford spectrophotometer in the visible range using a gel scanning attachment. The scans were manually digitized and entered into a PDP-11/03 minicomputer for data analysis.

* This work was supported by National Science Foundation Grant PCM 8118288 and National Institutes of Health Grant GM 30425. Parts of this work are based on the Biomedical Engineering Senior Project of E. B. at Boston University's College of Engineering. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed.

1 The abbreviations used are: SLS, segment long spacing; UA, uranyl acetate; PTA, phosphotungstic acid.
Theoretical charge density profiles of type I collagen SLS were synthesized using the sequences of the chick α1(I) (17) and α2 chain by varying the position of the α2 chain among the A, B, and C chain positions of the triple helix. The axial charge distribution of each α chain for the synthesis of the models based on the positions of both negatively and positively charged amino acid residues was obtained by assigning a value of 1 to the positions of charged residues (Glu, Asp, Arg, Lys, Hyl, His) and a value of 0 to all other positions. For the synthesis of models based on the positions of the negatively charged residues, a value of 1 was assigned to locations of Glu and Asp only. The charge distribution of the triple helix was obtained for different positions of the α2 chain by axially shifting and summing the charge distribution of the 3 chains. Using this approach, the sequence of each α chain becomes a series of ones and zeroes which were shifted (B and C chains) and summed to generate the profile of the three models. For example, in Model A, the α2 chain is in the reference or A position and the charge distribution for the molecule was obtained by summing the charge distribution of the α2 chain with the charge distribution of an α1(I) chain axially shifted by 1 residue (B position) and with another α1(I) chain shifted axially by 2 residues with respect to the A chain (C position). Models B and C were obtained in a similar manner with the α2 chain in the B and C chain positions, respectively. The 3 model scans were smoothed using “top hat,” triangular, and 1 + 2 cosine as the weighting functions in a weighted moving average. The triangular weighting function consistently yielded the best results (i.e., highest correlation coefficients).

The experimental scans were digitally stretched and shifted until bands 15 and 36 were visually aligned with peaks 15 and 36 of the synthesized scans. The average value of the experimental scan was then set at 0 and the minimum value was set at 0. An average scan was generated by summing and normalizing the experimental scans. The limiting factor in the resolution obtained from the micrographs was the sampling interval. Each experimental scan typically had approximately 250 points before the digital stretching was performed; with a molecular length of ~1000 residues, this corresponds to a resolution of about 1.2 nm, or 4 residues. AnHouston Instruments Digital Plotter (HIPLOT) was used in generating Figs. 1 and 2.

Correlation between experimental scans and models was performed by translating each experimental scan versus each model to get the highest correlation coefficient between each set, similar to the technique used by Meek et al. (18). The correlation coefficient \( r \) takes on values between +1 and −1; when \( r = 0 \), there is no correlation between the scans, when \( r = +1 \), the scans have the highest positive correlation, and when \( r = −1 \), the scans have the highest negative correlation. The formula used to calculate \( r \) is as follows

\[
r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{(n \sum X_i^2 - (\sum X_i)^2)(n \sum Y_i^2 - (\sum Y_i)^2))^{1/2}}
\]

The experimental and model data are the X and Y values, and \( n \) is the total number of points used in the calculation. The subscript \( i \) is the number of the residue with respect to the first triple helical glycine. For the UA-stained correlations, \( i = 235 \) to 705 (\( n = 451 \)), while with the PTA-UA-stained correlations, \( i = 165 \) to 945 (\( n = 781 \)); these are approximately SLS bands 13 through 38 and 9 through 52, respectively. Over these ranges, the experimental scans were most similar to each other and yielded the highest correlation coefficients with each other and with the models.

A constant axial displacement per residue \( (h \text{ spacing}) \) was assumed for the SLS. As shown below, 2 different staining techniques (UA versus PTA-UA) yield the same order of correlation, which suggests that the \( h \) spacing in the SLS crystallizes is approximately constant. The high correlation between the staining pattern and location of charged residues indicates that ATP used in forming the SLS does not significantly change the charge distribution of the molecule.

**RESULTS AND DISCUSSION**

The staining pattern of collagen SLS crystallites is well known to represent the charge profile of the molecule. Until recently (22), it was believed that negatively charged residues stain with UA whereas both negatively and positively charged residues stain with a combination of PTA and UA. Since the α chains are staggered axially by either 1 or 2 residues in the collagen triple helix and the charge profile for the α2 chain is different from that of α1(I), the predicted SLS banding pattern based on the sequence depends on the location of the α2 chain. In our studies, we considered the 3 possible locations of the α2 chain: the A, B, and C positions (Models A, B, and C, respectively). The scans of the 6 UA-stained type I SLS and their average are shown in Fig. 1, and the scans of the 7 PTA-UA-stained type I SLS and their average are shown in Fig. 2.

Tables I and II show the correlation matrix for the scans derived from UA- and PTA-UA-stained SLS. As seen from these tables, the experimental scans show a high degree of correlation with each other as well as with the average scan. The experimental scans correlate with the average scan better than with the individual scans for both staining techniques. The average of the correlation coefficients shown in Row 7 of Table I (0.861) is less than the average of Row 8 in Table II.
(0.906). This result suggests that PTA-UA-staining may be more reproducible.

A smoothed version of the Model A, B, and C scans, derived from the positions of the negatively charged residues, is shown in Fig. 1; it is evident that Models A, B, and C are almost exactly alike and visually there is no way to decide which model fits the experimental data best. A smoothed version of the Model A scan derived from the positions of the negatively and positively charged residues is shown in Fig. 2; Models B and C are visually identical with Model A, as in Fig. 1, so they are not shown. Although the differences among the models appear too small to tell them apart, the correlation coefficients calculated by the computer reveal more reproducible. The maximum correlation coefficient between experimental and model scans generally occurs when a triangular weighted moving average of width, $N = 19$, is performed on the charge profile of the model. Best visual results are obtained with a triangular weighted moving average of width 15. The smoothing done by the weighted moving average is required since the sequence-generated charge profile has a higher resolution than the electron micrographs. Meek et al. (18) found that the correlation between the sequence and the native banding pattern was maximized using a top hat or square weighting function with a width of 11 residues.

Fig. 3 shows the results of correlations performed between the average of the UA-stained SLS scans and the 3 models which represent the positions of the negatively charged residues. Note that the order of decreasing correlation coefficients is A, B, C. The maximum correlation coefficient ($N = 19$) of 0.68 is below what was expected based on the high correlation of the sequence with the fibrillar banding pattern reported by Meek et al. (18). Later it was determined that the maximum correlation coefficient increased to about 0.88 when the average of the UA-stained SLS scans was correlated with the models obtained from the positions of both the negative and positive charges in the sequence (Fig. 4). This figure also indicates that the order of decreasing correlation is A, B, C.

Fig. 5 shows the results of correlations performed between the average of the PTA-UA-stained SLS scans and the 3 models that represent the positions of the negative and positive charges. For PTA-UA-staining, a maximum correlation coefficient of about 0.88 was also found ($N = 19$). As a check on the validity of the ordering, each of the 13 SLS scans (6 UA-stained and 7 PTA-UA-stained) was correlated with the models based on the positions of the negative and positive charges and the results are shown in Tables III and IV. In every case, the order of decreasing correlation is A, B, C.

If the order of correlation between the models and SLS scans were a random event, the probability of the same order occurring all 13 times would be $6/10^7$ or less than 1 in $10^9$. Based on probability, it is unlikely that the correlation order A, B, C is a random event. Although the difference among the correlation coefficients is very small (Tables III and IV), it was determined that at the 0.99 level of confidence the mean of the differences in the correlation coefficients among the models is greater than 0. Based on these experimental data, we conclude that the $\alpha_2$ chain is in the A or reference position. Previous theoretical interaction studies have suggested that the $\alpha_2$ chain is either in the A (19) or B position (20, 21).

![Fig. 3. Correlation coefficient versus number of residues in the triangular smoothing function between an average densitometric scan of uranyl acetate-stained type I SLS and theoretical models. Models A, B, and C represent the theoretical scans obtained with the $\alpha_2$ chain in the A, B, and C positions of the triple helix. Models were constructed by assigning a value of 1 to Glu and Asp while all other residues were assigned a value of 0. Correlation was performed between residues 255 and 705 (bands 13 through 38).](image-url)
Type I Collagen SLS Banding Patterns

![Graph](image)

**Fig. 4.** Correlation coefficient versus number of residues in the triangular smoothing function between an average densitometric scan of uranyl acetate-stained type I SLS and theoretical models. Same as Fig. 3 except the models were constructed by assigning a value of 1 to all charged residues while a value of 0 was assigned to uncharged residues.

**Fig. 5.** Correlation coefficient versus number of residues in the triangular smoothing function between the average densitometric scan of PTA-UA-stained type I SLS and theoretical models. Models A, B, and C represent theoretical scans with the $\alpha_2$ chain in the A, B, and C positions, respectively, of the triple helix. Models were constructed by assigning a value of 1 to all charged residues and a value of 0 to all other residues. The correlation was performed from residues 165 to 945 (bands 9 through 52).

| Model | A   | B   | C   |
|-------|-----|-----|-----|
| A     | 0.7486 | 0.7457 | 0.7425 |
| B     | 0.8166 | 0.8139 | 0.8094 |
| C     | 0.7746 | 0.7703 | 0.7652 |

**Table III**

Maximum correlation coefficients between UA-stained type I collagen SLS scans and theoretical Models A, B, and C

Scan 7 is the average of experimental scans 1 through 6.

| Scan | Model | A   | B   | C   |
|------|-------|-----|-----|-----|
| 1    | A     | 0.8007 | 0.8000 | 0.7982 |
| 2    | B     | 0.7860 | 0.7823 | 0.7781 |
| 3    | C     | 0.7736 | 0.7692 | 0.7647 |
| 4    | A     | 0.7767 | 0.7739 | 0.7709 |
| 5    | B     | 0.7736 | 0.7692 | 0.7647 |
| 6    | C     | 0.7767 | 0.7739 | 0.7709 |
| 7    | A     | 0.8007 | 0.8000 | 0.7982 |
| 8    | B     | 0.7860 | 0.7823 | 0.7781 |
| 9    | C     | 0.7736 | 0.7692 | 0.7647 |

It was also possible to conclude based on correlation with the models (Figs. 3 and 4) that UA stains both negatively and positively charged residues in type I collagen. This conclusion is consistent with the recent findings of Tzaphlidou et al. (22) on the staining of reconstituted native fibrils with UA. The correlation coefficient with UA staining (0.88) was similar to that obtained with PTA-UA-staining; however, with the latter staining technique, the correlation could be performed over a larger portion of the molecule.

| Model | A   | B   | C   |
|-------|-----|-----|-----|
| A     | 0.7486 | 0.7457 | 0.7425 |
| B     | 0.8166 | 0.8139 | 0.8094 |
| C     | 0.7746 | 0.7703 | 0.7652 |

**Table IV**

Maximum correlation coefficients between PTA-UA-stained type I collagen SLS scans and theoretical Models A, B, and C

Scan 8 is the average of experimental scans 1 through 7.

| Scan | Model | A   | B   | C   |
|------|-------|-----|-----|-----|
| 1    | A     | 0.7486 | 0.7457 | 0.7425 |
| 2    | B     | 0.8166 | 0.8139 | 0.8094 |
| 3    | C     | 0.7746 | 0.7703 | 0.7652 |
| 4    | A     | 0.7613 | 0.7606 | 0.7601 |
| 5    | B     | 0.7858 | 0.7828 | 0.7795 |
| 6    | C     | 0.8405 | 0.8391 | 0.8373 |
| 7    | A     | 0.8344 | 0.8330 | 0.8301 |
| 8    | B     | 0.8779 | 0.8748 | 0.8712 |
| C     | 0.8779 | 0.8748 | 0.8712 |
As an extension of these findings with the SLS crystallites, we considered how the correlation of the native banding pattern of D-staggered fibrils with a sequence-generated pattern was affected with the \( \alpha_2 \) chain in different positions. If one assumes that the correlation order A,B,C of the triple helical models is correct, then it seems reasonable that the native banding pattern should correlate in the same order. Meek \textit{et al.}, however, did not find significant differences in the correlation coefficient with the \( \alpha_2 \) chain in different positions (18). The results of our initial studies indicate that the telopeptide model used in the synthesis of the native banding pattern affects the correlation more than the position of the \( \alpha_2 \) chain. We have found that by leaving out the telopeptides in the synthesis of the native banding pattern the order of correlation of the triple helical models (A,B,C) is preserved. A half-extended telopeptide model also maintains the order of correlation of the triple helical models; however, the differences among the models are very small. A fully extended telopeptide model gives us a correlation order of C,B,A. We feel these results indicate that a useful criterion for evaluating different telopeptide models is the effect they have on the correlation order of the triple helical models. Further work on different telopeptide models is now under way.

Acknowledgment—We would like to thank Dr. Michael Friedman of Rutgers University for helpful discussions on some of the statistical aspects.

REFERENCES
1. Ramachandran, G. N., and Ramakrishnan, C. (1976) in \textit{Biochemistry of Collagen} (Ramachandran, G. N., and Reddi, A. H., eds) Chap. 2, Plenum Press, New York
2. Silver, F. H. (1982) \textit{Collagen and Related Research}, in press
3. Bruns, R. R., and Gross, J. (1973) \textit{Biochemistry} 12, 908-815
4. von der Mark, K., Werdn, P., Rexrodt, F., and Kuhn, K. (1970) \textit{FEBS Lett.} 11, 105-108
5. Hofmann, H., Fietzek, P. P., and Kuhn, K. (1978) \textit{J. Mol. Biol.} 125, 137-165
6. Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A., and Woodhead-Galloway, J. (1973) \textit{J. Mol. Biol.} 79, 137-148
7. Doyle, B. B., Hukins, D. W. L., Hulmes, D. J. S., Miller, A., Rattew, C. J., and Woodhead-Galloway, J. (1974) \textit{Biochem. Biophys. Res. Commun.} 60, 858-864
8. Piez, K. A., and Torchia, D. A. (1975) \textit{Nature} 258, 87
9. Salem, G., and Traub, W. (1975) \textit{FEBS Lett.} 51, 94-99
10. Piez, K. A., and Trus, B. L. (1977) \textit{J. Mol. Biol.} 110, 701-704
11. Piez, K. A., and Trus, B. L. (1978) \textit{J. Mol. Biol.} 122, 419-432
12. Traub, W. (1978) \textit{FEBS Lett.} 92, 114-120
13. Kumar, N. G., and Cunningham, L. W. (1980) \textit{Biopolymers} 19, 1587-1595
14. Li, S. T., Golub, E., and Katz, E. P. (1975) \textit{J. Mol. Biol.} 98, 835-839
15. Silver, F. H., and Trelstad, R. L. (1980) \textit{J. Biol. Chem.} 255, 9427-9433
16. Piez, K. (1976) in \textit{Biochemistry of Collagen} (Ramachandran, G. N., and Reddi, A. H., eds) Chap. 1, Plenum Press, New York
17. Highberger, J. H., Corbett, C., Dixit, S. N., Yu, W., Seyers, J. M., Kang, A. H., and Gross, J. (1982) \textit{Biochemistry} 21, 2045-2055
18. Meek, K. M., Chapman, J. A., and Hardcastle, R. A. (1979) \textit{J. Biol. Chem.} 254, 10710-10714
19. Piez, K. A., and Trus, B. L. (1978) \textit{J. Mol. Biol.} 122, 419-432
20. Hofmann, H., Fietzek, P. P., and Kuhn, K. (1978) \textit{J. Mol. Biol.} 125, 137-165
21. Traub, W., and Fietzek, P. P. (1976) \textit{FEBS Lett.} 68, 245-249
22. Tzaphlidou, M., Chapman, J. A., and Al-Samman, M. H. (1982) in \textit{Micron}, in press