INTERPRETATION OF THE SMALL-ANGLE X-RAY DIFFRACTION OF COLLAGEN IN VIEW OF THE PRIMARY STRUCTURE OF THE α1 CHAIN

W. CLAFFEY, *Department of Macromolecular Science, Case Western Reserve University, Cleveland, Ohio 44106 U.S.A.*

ABSTRACT The small-angle X-ray diffraction pattern of collagen has been calculated using the sequence of the α1 chain and a Hodge-Petruska scheme for the packing of the collagen molecules. The molecular stagger giving the best fit of calculated-to-observed structure factors has been found to be 236 or 237 amino acid residues for three tendon collagens. But this result depends on the approximation that the molecular conformation is uniform throughout the molecule. A comparison of the observed and calculated electron density profiles in axial projection leads to a corrected model, in which the COOH-terminal telopeptide is contracted in a way suggesting a saddle-shaped electron density distribution near the collagenase site.

INTRODUCTION

Meridional small-angle X-ray diffraction patterns from collagen fibrils have often been obtained and analyzed (Kaesberg and Shurman, 1953; Tomlin and Worthington, 1956; Ericson and Tomlin, 1959; Ellis and McGavin, 1970; Chandross and Bear, 1973). The purpose of these studies was to determine the axial projection of the electron density distribution within the fiber period. It was hoped that the analysis of this density profile could be related to the sequence and conformation of amino acid residues along the fiber axis. The reason for taking up the problem again is that recent advances in the understanding of collagen structure appear to allow a more complete interpretation of the small-angle X-ray data.

The amino acid sequence of the α1 chain has been completed (it has been listed by Hulmes et al., 1973, and by Walton, 1974). Although the sequence is that of rat and calf skin collagen, it will be argued below that species differences do not affect the analysis of the small-angle X-ray data. The axial projection of the electron density can be calculated from the sequence information by assuming a Hodge-Petruska (1963) scheme for the packing of collagen molecules. Phases derived from this calculation can be used to determine the electron density profile according to experimental data, a meaningful procedure if the calculated and real profiles are not too dissimilar. Differences between the observed and calculated profiles may then give information about aspects of collagen structure not included in the calculation.

ELECTRON DENSITY CALCULATION

The tropocollagen molecule consists of two $\alpha 1$ chains and one $\alpha 2$ chain in a parallel supercoiled polyglycine II conformation. We assume that this conformation is uniform throughout the molecule, although this is unlikely to be correct in the telopeptide regions. The question of telopeptide conformation will be further discussed below. We must also assume that the molecule can be represented as consisting of three $\alpha 1$ chains, since the $\alpha 2$ sequence is not fully known, although at least some parts of it are similar and homologous to the $\alpha 1$ (Piez et al., 1972; Fietzek et al., 1972). Since the resolution of the X-ray data is about 15 Å, several consecutive $\alpha 2$ residues would have to differ from the corresponding $\alpha 1$ residues to introduce a serious error into this approximation.

The tropocollagen molecule is about 4.5 D long where D is the period of the fibril in projection on the fibril axis. If five molecules contribute to a period, neighboring molecules must be staggered by mD where m is an integer. There is a total of 4! structurally distinct ways of arranging the five molecules (Doyle et al., 1974), but in axial projection all have the same electron density. The arrangement used in the calculation has $m = 1$ and is shown in Fig. 1. Each chain within the molecule is shifted past the next by one residue (Ramachandran and Kartha, 1955; Rich and Crick, 1961), a scheme which can be represented as:

—Gly—X—Y—Gly—X—Y—
—Y—Gly—X—Y—Gly—X—
—X—Y—Gly—X—Y—Gly—

In subsequent references to the molecular stagger, the central of the three chains will be used as a marker.

The projected electron density at a fractional coordinate u within the period D is given by $\rho(u) = \sum_{i=1}^{15} \rho_i(u)$ where $\rho_i(u)$ is the density contributed by the ith chain. The u-coordinate of each atom in each amino acid residue is unknown. The simplest approximation is to let $\rho_i(u) = Z_j$, where Z_j is the total number of electrons in the jth residue. This means that the electron density distribution of each residue is represented as a point, which seems a good approximation for the resolution of the data. A way of taking into account the fact that each residue is actually of finite size will be considered below. The gap regions as represented in Fig. 1 are assumed to have the same density as the water and polysaccharide ground substance surrounding the fibril, so that they contribute nothing to the collagen diffraction.

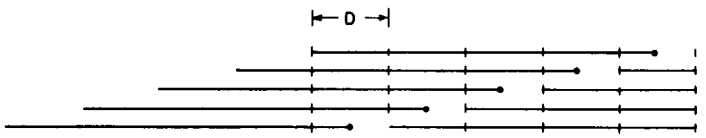

FIGURE 1 Arrangement of collagen molecules used in the electron density calculations. Closed circles indicate the COOH-terminal molecular ends.

The function $\rho(u)$ can thus be calculated at intervals of u corresponding to $1/D$ where D is the period expressed in residues, and $\rho(u)$ and its Fourier transform are sensitive to the stagger chosen.

COMPARISON WITH EXPERIMENTAL DATA

After the stagger giving the best fit to the small-angle X-ray data is determined, a value for the unit height can be derived as $H = D(\text{Å})/D(\text{residues})$. This value of the unit height should agree with the value determined by wide-angle X-ray diffraction if the assumption of uniform molecular conformation is correct. Yet there is a small but finite range of values of H to be considered, because of the breadth of the wide-angle meridional reflection. A value of $H = 2.91$ Å is given by Ramachandran (1967) for several collagens, while Rich and Crick give $H = 2.86$ Å. Taking $D(\text{Å}) = 670$ Å, these two values would lead to staggers of 230 or 234 residues, respectively. Recent work on the determination of the stagger by considering interactions between adjacent molecules tends to support the smaller value of H. Hulmes et al. found that the distribution of polar and large hydrophobic side groups favors a stagger of 234 ± 1 residues, while Walton found a value of 233 ± 1.

The stagger giving the best fit to the small-angle X-ray intensity data was determined as follows. The structure factors for the electron density distribution as derived in the preceding section can be written as

$$F_c'(h) = \sum_{u=0}^{1} \rho(u) \exp(2\pi ihu)$$

where the sum is taken at intervals of $1/D$. Calculations were made for density profiles resulting from staggers of 230 to 239 residues. Calculations were also made incorporating the effects of finite residue size. The residues were represented as spheres of radii R_j so that

$$F_c'(h) = \sum_{u=0}^{1} \rho(u)\, 3[\sin(kR_j) - kR_j \cos(kR_j)]/k^3 R_j^3 \exp(2\pi ihu)$$

where $k = 4\pi \sin\theta/\lambda$. Here the residue radii were obtained from a tabulation of residue volumes by Chothia (1975).

The intensity data of Tomlin and Worthington were used to compare the observed and calculated structure factors for three different wet tendon collagens, assuming that $F_o(h) = [I(h)]^{1/2}$ (Blaurock and Worthington, 1966). 25 orders were calculated for each of the 10 staggers, and the R-factor,

$$\Sigma \, | \, |F_o(h)| - |F_c(h)| \, | / \Sigma \, |F_o(h)|,$$

was computed. Here $F_c(h) = K \, |F_c'(h)| \exp(-B/d(h))$, where K is a scale factor that provides the best fit of the calculated to observed data and B is a pseudotemperature factor that describes the attenuation of the diffraction spectra due to disorder;

$d(h)$ is the interplanar spacing for a reflection of order h. The parameters K and B were found by least-squares refinement for each stagger.

RESULTS, ELECTRON DENSITY PROFILES, AND DISCUSSION

Fig. 2 shows a plot of the R-factor *vs.* D for rat tail tendon, beef achilles tendon, and kangaroo tail tendon as obtained from the point approximation to the residue electron densities. Also shown is R *vs.* D for rat tail tendon by the sphere approximation. For each value of D, Fig. 2 lists the fraction of the period occupied by all five collagen molecules. This fraction corresponds to the overlap zone or "bandwidth" traditionally postulated to account for the main features of the small-angle X-ray patterns from collagen. The minimum R-factor is found for a stagger of 236 or 237 residues; no choice can be made between these two values of D. Use of the sphere approximation produces a slightly lower R-factor at all staggers, but does not shift the position of the minimum. A stagger of 236 residues would give a fractional overlap of 0.457, in essential agreement with the width of the step function used as a first-order approximation to the electron density profile in most of the X-ray work cited in the Introduction. In what follows, the calculated amplitudes and phases derived from the model with $D = 236$ will be used as the best approximation to the Fourier series of the electron density profile in tendon collagen, although the results for $D = 237$ fit the data equally well. The amplitudes and phases for $D = 236$ are listed in Table I.

A stagger of 236 residues leads to a unit height of 2.84 Å. This is slightly smaller

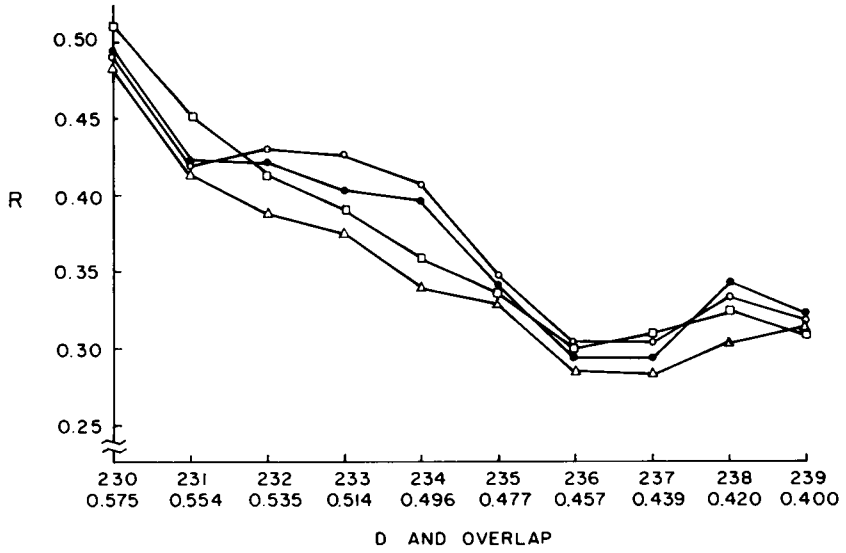

FIGURE 2 R-factor *vs.* D and fractional overlap for three tendon collagens. The closed circles are for rat tail tendon, the open circles are for beef achilles tendon, and the squares are for kangaroo tail tendon; the R-values were obtained by the point approximation. The triangles indicate R-values obtained by the sphere approximation for rat tail tendon.

TABLE I

ORDER h, OBSERVED STRUCTURE FACTOR F_o FOR KANGAROO TAIL TENDON COLLAGEN (DATA OF TOMLIN AND WORTHINGTON), CALCULATED STRUCTURE FACTOR MODULUS F_c, AND PHASE Φ FOR A STAGGER OF 236 RESIDUES

Calculations were based on the point approximation to the residue electron densities.

h	F_o	F_c	Φ
			rad
1	316	295	1.490
2	36	25	−0.182
3	119	112	0.969
4	28	15	1.081
5	62	122	0.323
6	27	36	1.157
7	35	48	0.127
8	30	27	−1.446
9	47	30	−0.004
10	27	24	0.805
11	17	18	0.146
12	30	20	1.420
13	7	20	0.081
14	14	5	−0.027
15	12	8	−0.115
16	10	13	0.596
17	10	5	0.036
18	10	5	−1.106
19	10	7	−0.425
20	20	3	−0.428
21	20	4	0.199
22	10	3	−0.153
23	0	2	−1.203
24	0	1	−0.911
25	10	1	1.270

than any previously reported unit height for collagen, suggesting that the assumption of uniform conformation is only approximately correct.

The question then arises: are certain parts of the molecule more deformed than others? Some insight into this can be given by calculating the electron density profile within a period as

$$\rho(u) = \sum_{h=1}^{25} F(h) \cos [2 \pi hu - \Phi(h)].$$

Although the calculated electron density $\rho_c(u)$ is directly available at high resolution, it is desirable to compare it to the observed electron density $\rho_o(u)$ at the reso-

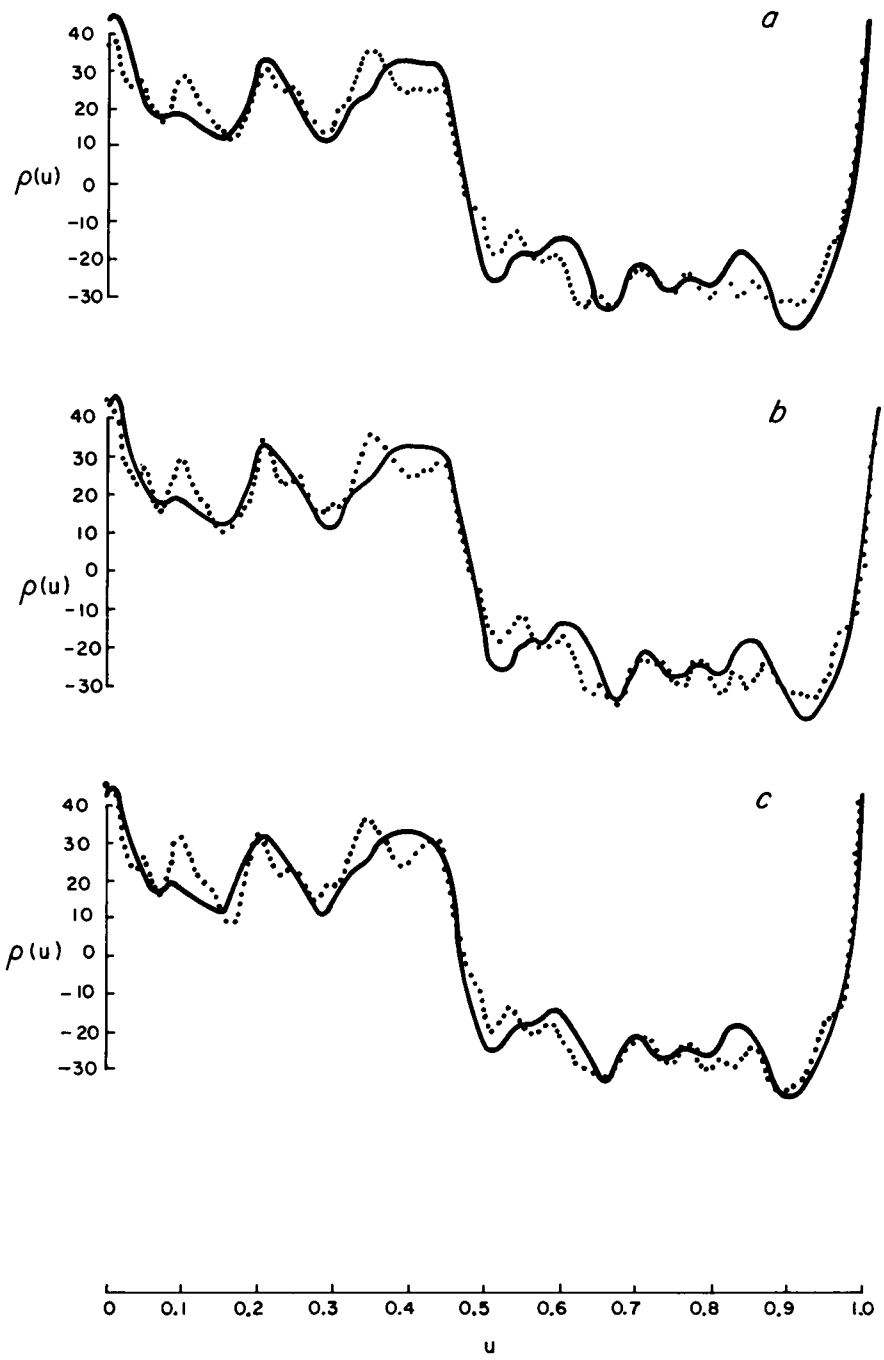

FIGURE 3 The calculated electron density distribution in axial projection is shown by a solid line, the observed by a dotted line for (*a*) rat tail tendon, (*b*) beef achilles tendon, and (*c*) kangaroo tail tendon.

lution imposed by experiment. Therefore, $\rho_c(u)$ was redetermined from its structure factors, and $\rho_o(u)$ was determined from the calculated phases and the observed structure factor moduli. Comparative graphs of these functions are shown for rat tail tendon, beef achilles tendon, and kangaroo tail tendon in Fig. 3 a, b, and c.

The three observed electron density profiles resemble each other, and deviate from the calculated profile at the same points, although not always by the same amounts. This suggests that species differences do not affect the general shapes of the profiles, and that we may try to interpret the deviations of the observed from the calculated profiles as arising from conformational features common to all three collagens.

Regions where $\rho_c(u) \sim \rho_o(u)$ should represent intervals where the conformation in each of the four or five superimposed molecules is relatively uniform. Regions where $\rho_c(u)$ and $\rho_o(u)$ are dissimilar should represent intervals where helix deformation occurs in one or more of the molecules. These statements are true only if the best normalization between the curves has been found; but this should indeed be the case because the best fit between the calculated and observed structure factors has already been found inreciprocal space.

There are two regions in Fig. 3 where the observed and calculated densities differ considerably. These are around $u = 0.1$ and around $u = 0.4$. The COOH-terminal telopeptide, 25 residues long, occupies the interval from $u = 0.356$ to $u = 0.457$. Helix breaking is to be expected in the telopeptide because glycine does not occur regularly at every third position. If the effect of unwinding the helix were to lengthen the peptide, the stagger derived from the present analysis would be smaller than the true stagger, because the peptide would extend further into the gap than would be the case for an undeformed helix. But assuming that the true stagger is 234 residues, the present analysis indicates that the COOH-terminal telopeptide appears contracted in projection, so that the fractional width of the overlap decreases. The $\rho_o(u)$ profiles of Fig. 3 near the telopeptide could be interpretated as follows: the telopeptide retracts, giving rise to a high-density region centered at $u = 0.34$. It then extends over a low-density interval but is again slightly compressed at the edge of the overlap. The net result is a saddle-shaped density distribution. This behavior of the telopeptide may be related to the fact that it is superimposed upon the point (at $u = 0.4$) within α1-CB7 where cleavage by many animal collegenases occurs (Piez and Miller, 1974). The site for enzymatic attack could thus be shaped by the telopeptide. However, the peak in $\rho_o(u)$ around $u = 0.1$ cannot be understood in terms of the residue distribution in this region. At present, no functional significance for this point in the period is known.

This work was supported by National Institutes of Health Grant HD-00669.

Received for publication 3 September 1976 and in revised form 28 February 1977.

REFERENCES

BLAUROCK, A. E., and C. R. WORTHINGTON. 1966. Treatment of low-angle X-ray data from planar and concentric multilayered structures. *Biophys. J.* **6**:305.

CHANDROSS, R. J., and R. S. BEAR. 1973. Improved profiles of electron density distribution along collagen fibrils. *Biophys. J.* **13**:1030.

CHOTHIA, C. 1975. Structural invariants in protein folding. *Nature (Lond.).* **254**:304.

DOYLE, B. B., D. J. S. HULMES, A. MILLER, D. A. D. PARRY, K. A. PIEZ, and J. WOODHEAD-GALLOWAY. 1974. A D-periodic narrow filament in collagen. *Proc. Roy. Soc. Lond. B. Biol. Sci.* **186**:67.

ELLIS, D. O., and S. MCGAVIN. 1970. The Structure of collagen. An X-ray study. *J. Ultrastruc. Res.* **32**: 191.

ERICSON, L. G., and S. G. TOMLIN. 1959. Further studies of low-angle X-ray diffraction patterns of collagen. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **252**:197.

FIETZEK, P. P., I. KELL, and K. KUEHN. 1972. The covalent structure of collagen. Amino acid sequence of the N-terminal region of α2-CB4 from calf and rat skin collagen. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **26**:66.

HODGE, A. J., and J. A. PETRUSKA. 1963. *In* Aspects of Protein Structure. G. N. Ramachandran, editor. Academic Press, Inc., New York. 289.

HULMES, D. J. S., A. MILLER, D. A. D. PARRY, K. A. PIEZ, and J. WOODHEAD-GALLOWAY. 1973. Analysis of the primary structure of collagen for the origins of molecular packing. *J. Mol. Biol.* **79**:137.

KAESBERG, P., and M. M. SHURMAN. 1953. Further evidence concerning the periodic structure in collagen. *Biochim. Biophys. Acta.* **11**:1.

PIEZ, K. A., G. BALIAN, E. CLICK, and P. BORNSTEIN. 1972. Homology between the α1 and α2 chains of collagen. *Biochem. Biophys. Res. Commun.* **48**:990.

PIEZ, K. A., and A. MILLER. 1974. The structure of collagen fibrils. *J. Supramol. Struc.* **2**:121.

RAMACHANDRAN, G. N. 1967. *In* Treatise on Collagen, G. N. Ramachandran, editor. Academic Press, Inc., New York. 103.

RAMACHANDRAN, G. N., and G. KARTHA. 1955. Structure of collagen. *Nature (Lond.).* **176**:593.

RICH, A., and F. H. C. CRICK. 1961. The molecular structure of collagen. *J. Mol. Biol.* 3:483.

TOMLIN, S. G., and C. R. WORTHINGTON. 1956. Low-angle X-ray diffraction patterns of collagen. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, **235**:189.

WALTON, A. G. 1974. Molecular aspects of calcified tissue. *J. Biomed. Mater. Res. Symp.* **5**:409.