Journal of Supramolecular Structure 2:121–137 (1974)

THE STRUCTURE OF COLLAGEN FIBRILS

Karl A. Piez* and Andrew Miller†

*Laboratory of Biochemistry, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20014

†Laboratory of Molecular Biophysics, Department of Zoology, Oxford University, Oxford OX1 3PS, United Kingdom

INTRODUCTION

Collagen in most tissues of higher animals and in many tissues of lower animals takes the form of a rope with a high degree of order. Like a rope, which has several levels of coiling, the collagen fibril has four structural levels of which at least three are coils. The polypeptide chain, the molecule, and the microfibril are helical structures; the fibril may consist of parallel or perhaps coiled microfibrils.

The mechanical role of collagen is fulfilled much like a rope which may be the fibril itself with diameters of tens to hundreds of nanometers or may be bundles of fibrils forming a fifth level of structure. Collagen fibrils may be organized in various ways including the largely disordered arrangement in bone, the networks around blood vessels, the highly ordered orthogonal arrays in the cornea, and the parallel bundles in tendons.

Unlike a rope, the collagen fibril is composed of a single molecular species which is itself highly ordered and which packs in a highly ordered way. This is necessary because of the biological demands placed upon collagen. Its synthesis, the site and manner of fibril formation and crosslinking, its removal during tissue turnover, and its interactions with cells and other tissue constituents must be under careful biological regulation. For this to be achieved, the collagen monomer has the same degree of specificity and sophistication in its structure that has been widely demonstrated for globular proteins.

The techniques of protein chemistry, electron microscopy, and x-ray diffraction have in recent years made possible the understanding of the interrelationships between the several levels of structure in the collagen fibril. We would like here to review these findings with emphasis on molecular packing.

MOLECULAR STRUCTURE

The collagen molecule contains three similar or identical polypeptide chains which are parallel and extend the full length of the molecule. Each chain assumes a polyproline II-type helix and the three chains are coiled together in a superhelix with a pitch of about 9 nm to form a rod-like molecule about 300 nm long and 1.4 nm in diameter. The details of this coiled-coil structure have been reviewed elsewhere (1).

For the present discussion, it is important to note that the molecular structure is dependent upon the presence of glycine as every third residue and sufficient proline and hydroxyproline to direct the conformation. The remainder of the amino acids are not critical to molecular structure and in large part are available for the interactions that direct molecular packing.

Because the three chains are parallel and aligned and the residue spacing is constant, the molecule can be considered to be one-dimensional in the sense that the position of an amino acid residue in the sequence determines and is a measure of its position along the axis of the molecule. This feature is reinforced by the fact that the three chains are similar or identical. This means that, to a first approximation, the molecule "looks" much the same from whatever side it is viewed. This simplification is important in the preliminary analysis of molecular interactions reviewed here.

In actual fact the molecule has no true symmetry at the atomic level. This characteristic permits a unique three-dimensional packing and must be taken into account in more detailed analyses.

ONE-DIMENSIONAL STRUCTURE

Amino Acid Sequence

The amino acid sequence of the $\alpha 1$ chain of what is referred to as type I collagen has been determined through the combined efforts of several laboratories. Type I collagen is the major collagen of skin, tendon, and bone and contains two identical chains, the $\alpha 1$ chains, and one similar but genetically distinct chain, the $\alpha 2$ chain. The sequence has been reproduced elsewhere (2). It was determined partly on calf skin collagen and partly on rat skin collagen. Species differences are, however, few and do not affect the conclusions discussed here.

Of the 1052 residues in the $\alpha 1$ chain, 1011 consist of triplets of the form Gly-X-Y where X and Y can be any amino acid. This triplet repeat is required for triple-chain packing of the type characteristic of collagen. The N-terminal 16 residues and the C-terminal 25 residues have a different kind of sequence. The lysyl residues in these regions are precursors of crosslinks (1, 3). It is likely that these regions play other specific roles perhaps including a directing influence in molecular packing. However, since no definitive data are available, they will not be considered further here.

Charge Distribution and Ultrastructure

Because of the one-dimensional nature of the collagen molecule, it is possible to relate information contained in the sequence to primary structure information obtained by electron microscopy. Collagen forms an artificial aggregate, the SLS (segment-long-spacing) form, in which the molecules are parallel with their ends aligned. Positive staining with uranyl acetate or phosphotungstic acid (or double staining with both) of SLS aggregates produces a band pattern perpendicular to the molecular axes, which is characteristic of all collagens so far examined. The ubiquity of the pattern suggests that the distribution of charged groups along the molecule, which are responsible for the staining, is representative of structural features that are critical to the ability of the collagen monomer to form ordered fibrils.

It has been shown that a plot of the distribution of charged residues as determined from the amino acid sequence of peptides from collagen correlates very well with the SLS band pattern (4–6). This is shown in Fig. 1 for the entire length of the molecule. Although the charge distribution taken from the sequence refers only to the $\alpha 1$ chain and the molecule contains an $\alpha 2$ chain in addition to two $\alpha 1$ chains, the correlation is excellent. It is known that the $\alpha 2$ chain is closely homologous to the $\alpha 1$ chain (7, 8).

Sequence and Molecular Stagger in the Fibril

It has been known for some time that collagen molecules in the native fibril are axially displaced or staggered by multiples of a unit referred to as D, which has been variously estimated at 64–68 nm. The best value, determined by x-ray diffraction, is now known to be close to 66.8 nm. This stagger relationship is shown in Fig. 2 superimposed on a negatively stained electron micrograph of a native collagen fibril. The collagen molecule is 4.4D long as measured on electron micrographs and 4.47D long as measured by x-ray diffraction (9). The latter is probably the more accurate value. The nonintegral length gives rise to alternate light and dark regions on the electron micrograph corresponding to regions of high and low protein density.

Although the SLS band pattern and the native fibril band pattern (Fig. 2) must contain information about molecular interactions in the fibril, it has not been possible to utilize it to understand these interactions. The sequence data, however, provide similar information but at a much higher resolution. It has recently been shown by analysis of the sequence that interactions between large hydrophobic amino acid side chains and between side chains of opposite charge on adjacent molecules are maximal when the molecules are staggered by multiples of 234 ± 1 residues (2).

The analysis was done by sliding the amino acid sequence past itself and scoring for complementarity at all possible staggers. This procedure is analogous to a search for internal homology except that complementarity rather than similarity is sought. A score of +1 was given whenever charges of opposite sign occurred on opposite chains within three residues to either side or whenever opposing large hydrophobic amino acids (valine, leucine, isoleucine, methionine, and phenylalanine) were found within two residues. The results, shown in Fig. 3, demonstrate that both hydrophobic and charge interactions are

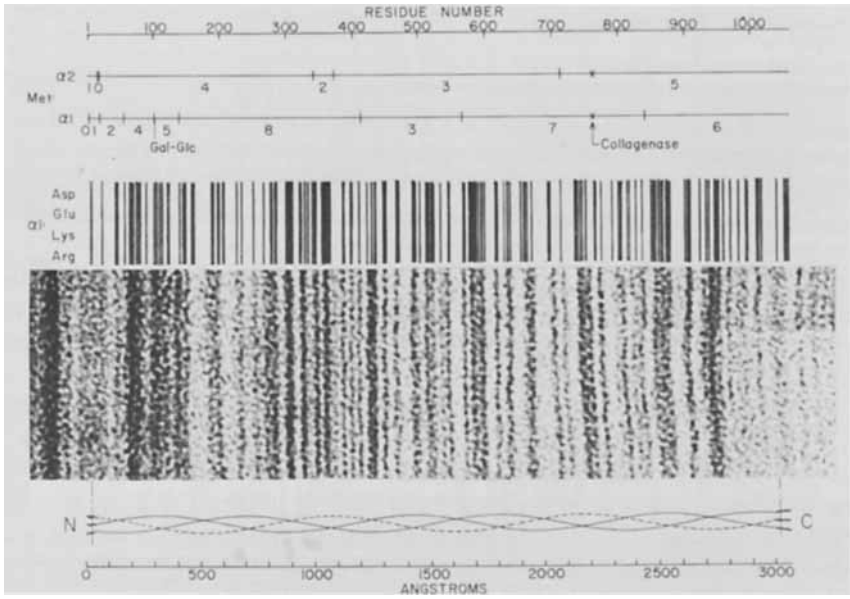


Fig. 1. A comparison of the distribution of charged amino acid residues derived from the sequence of the α1 chain of collagen (dark bars representing Asp, Glu, Lys, and Arg) and distribution of charge by electron microscopy of an SLS aggregate (center) positively stained with uranyl acetate and phosphotungstic acid. The electron micrograph was provided by K. Kühn. The figure also shows a representation of a collagen molecule (bottom) oriented to indicate the direction and position of molecules in the SLS aggregate. The α1 and α2 chains of collagen are represented schematically (top) to show the positions of methionine residues in the chains (short vertical lines), the point where carbohydrate is attached in the α1 chain, and the place where many animal collagenases cleave the α chains. Cleavage at the methionine residues yields peptides numbered as indicated which are the starting material for sequence and other chemical studies.

significant in determining molecular stagger. The peaks at multiples of 234 residues are significantly above the background and are equally spaced.

This result of course gives only a crude picture of molecular interactions and defines only the axial displacement between nearby molecules. A more detailed analysis in three dimensions may, however, be possible using similar principles. That the necessary information is present in the sequence is suggested by the observation that the large hydrophobic amino acids show approximate 2D/11 and D/11 spacings (2). Such regularities are probably related to super coiling of the chains (or molecules) and may help to define these parameters (10).

The value of D = 234 ± 1 residues from the sequence analysis and the value of D = 66.8 ± 0.2 nm from x-ray diffraction gives 0.286 ± 0.001 nm for the residue spacing. This is close to and may be better than the generally accepted value of 0.291 nm obtained by x-ray diffraction because of inherent difficulties in measuring this reflection.

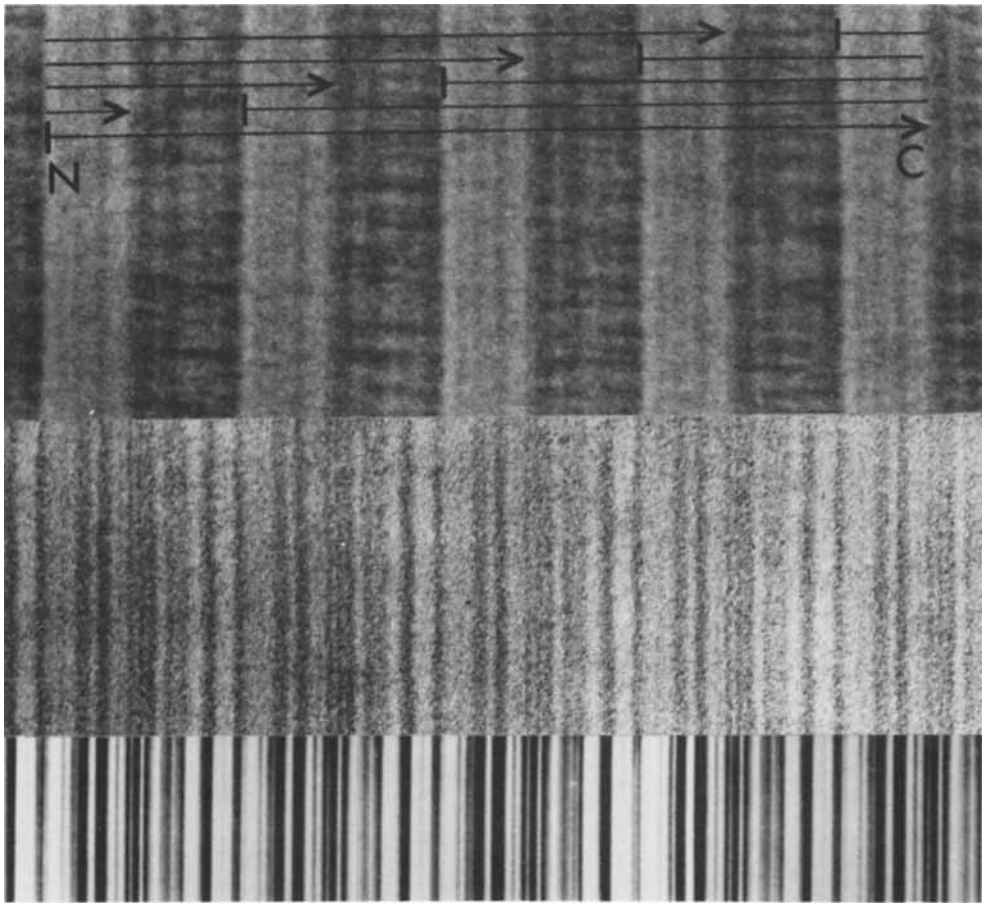


Fig. 2. Negatively stained (top) and positively stained (middle) native collagen fibrils. The Petrushka-Hodge two-dimensional packing structure is superimposed on the micrograph to show the position and direction of collagen molecules in the fibril. The period is 66.8 nm and each collagen molecule extends through 4.5 periods. The N-terminal ends of the α chains in the molecule are to the left. The two patterns can be aligned by the partial positive staining in the negatively stained fibril. The pattern at the bottom shows the native band pattern reconstructed from a plot of the distribution of charged residues in the amino acid sequence. The electron micrographs were provided by J. Chapman. The reconstruction was done by D. Hulmes.

Using the value of D determined by the sequence analysis, it is possible to reconstruct the native band pattern as seen by electron microscopy of positively stained collagen fibrils from the sequence (6). If a plot of the charged amino acids along the molecule such as is shown in Fig. 1 is photographed and printed repeatedly on the same sheet of paper, moving the paper a distance equivalent to 234 residues each time, the native band pattern should be produced. That this is the case is shown in Fig. 2. Chapman and Hardcastle (11) find that the best fit is obtained if the stagger is 232 or 233 residues.

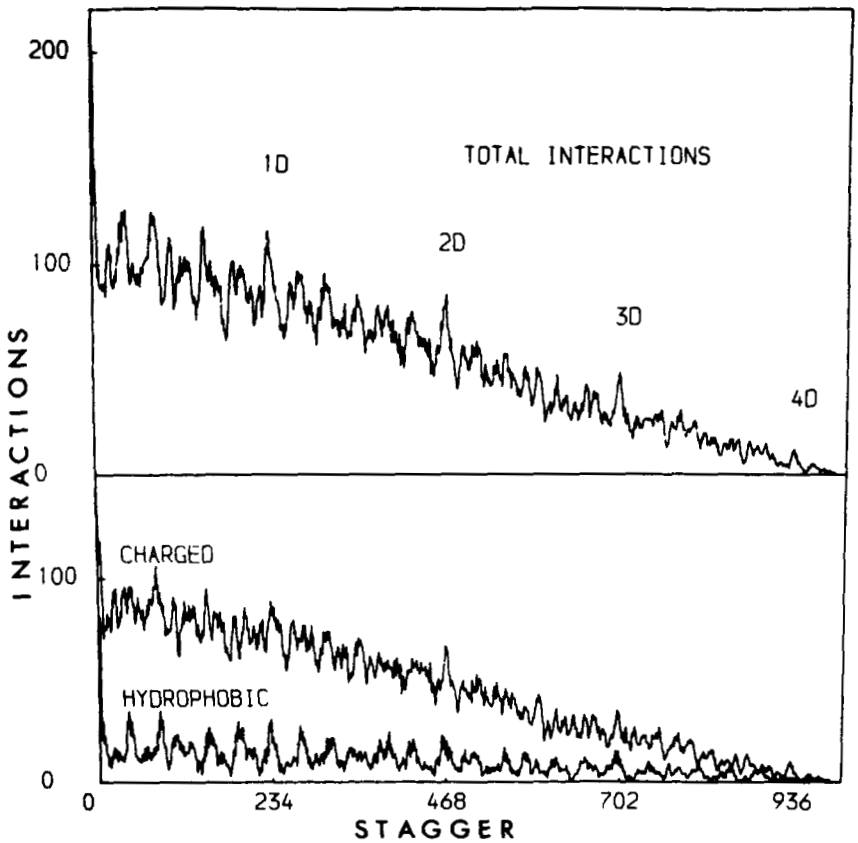


Fig. 3. The interaction pattern obtained by moving the amino acid sequence past itself and scoring for hydrophobic and charged interactions at each stagger. The maxima at multiples of 234 residues correspond to the 66.8 nm period observed in electron micrographs of native collagen fibrils. Reproduced from Hulmes et al. (2) with permission of Academic Press, Inc. (London).

A similar reconstruction was done some years ago by Hodge and Schmitt (12) using the SLS band pattern. The reconstruction from the sequence is more precise and allows the position on the micrograph of individual amino acids to be calculated (13).

MOLECULAR PACKING

Packing Models

Although the axial displacement between collagen molecules in the native fibril has been understood for some time, it has not been clear until recently how this structure could be extended to three dimensions. Models consisting of rolled sheets, pseudo-

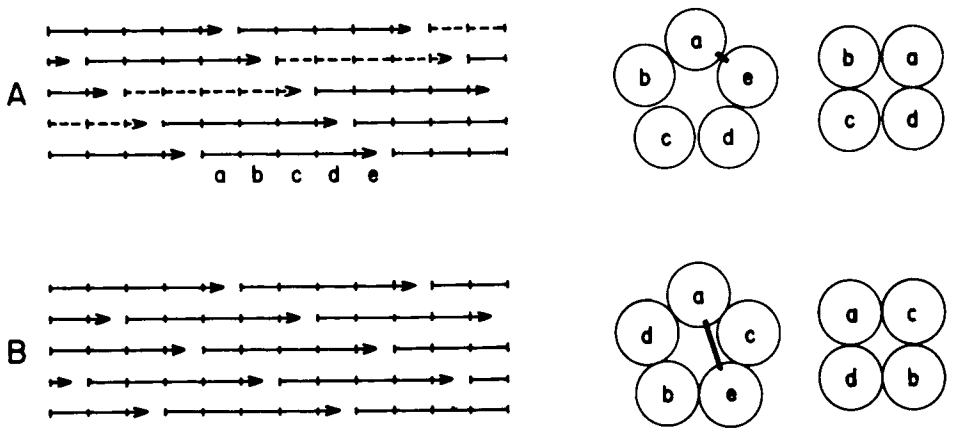


Fig. 4. Possible collagen microfibril structures in radial projection (left) and cross section (right) containing a true D period of 66.8 nm. Collagen molecules are represented by arrows in the projections and circles in the cross sections. The bars between segments a and e in the cross sections represent covalent crosslink positions that are favored by chemical data. (A) The Smith microfibril described by axial translations of 1D and 4D and rotations of $2\pi/5$ which may reduce to $\pi/2$ in gap regions as shown in the cross sections. The structure may be viewed as five-stranded with 0.5D gaps in the strands or as four-stranded with 0.5D overlaps in the strands. The latter situation is represented by the dashed molecules. (B) An alternative two-start helix described by axial translations of 2D and 3D and the same rotations.

hexagonally packed molecules and randomly accreted molecules have been proposed (1). However, none of these models incorporate the specificity possible and expected for interactions between protein molecules with the degree of order shown for collagen.

The first model to do this in a convincing way was the one devised by Smith (14). He proposed a five-stranded microfibril in which the strands are lines of molecules with gaps of 0.5D between molecules. A radial projection of this structure is reproduced in Fig. 4 together with a related structure which is equally acceptable on *a priori* grounds (15). The Smith microfibril is a one-start helix while the other is a two-start helix. Both structures have the characteristic that all molecules are equivalent. A number of other five-stranded helices are possible but they require two sets of bonding rules for their construction and therefore seem less likely.

It should be noted that the Smith microfibril is an idealized structure. For example, it can also be viewed as a four-stranded microfibril in which the strands are linear aggregates of collagen molecules with an 0.5D overlap. The molecules which are dashed in Fig. 4 illustrate such a strand. Cross sections through the overlap region would intersect five molecules but through the gap region would intersect four molecules. The simplest model would show five-fold symmetry in cross sections with a space in the gap region. However, four-fold symmetry in the gap region changing to five-fold symmetry in the overlap region may be more realistic. Other less symmetric structures are also possible.

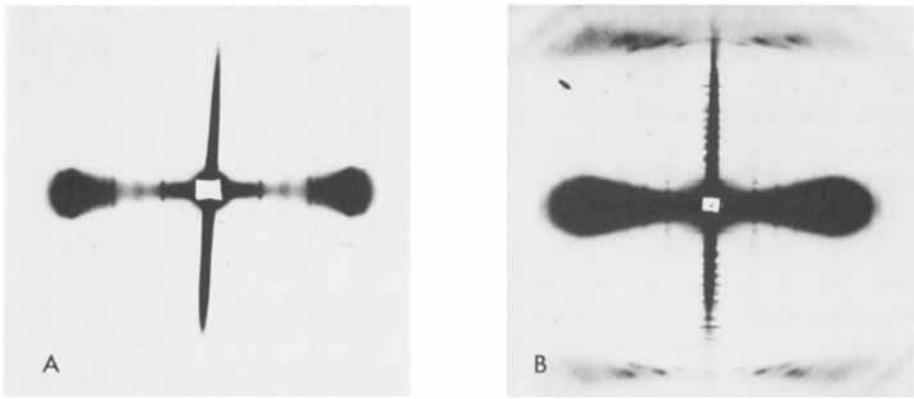


Fig. 5. Medium angle x-ray diffraction patterns of native rat tail tendon collagen (17). (A) Equatorial and near-equatorial reflections related to the microfibril structure. (B) A longer exposure showing intercepts of the row lines on the 0.95 nm layer line which arises from the collagen helix.

Similarly, the alternate two-start helix can be viewed in several ways.

The Smith microfibril is favored by chemical evidence which shows that covalent crosslinks occur between segments a and e (Fig. 4) of the collagen molecule. This would most likely occur between nearest neighbors which is possible only in the Smith microfibril. However, the evidence is not conclusive, since cross sections of the microfibril show that covalent crosslinks could be between nearest-but-one neighbors in the alternate two-start helix, particularly since crosslinks originate in nontriplet ends which may not have the rigid rod-like structure of the helix. This is illustrated in Fig. 4.

X-ray Diffraction

Collagen fibers have an x-ray diffraction pattern characterized by low-angle meridional reflections arising from the 67 nm axial repeat. The equator generally shows only a diffuse reflection at about 1.5 nm that is related to the lateral spacing of collagen molecules. Fine structure on or near the equator suggestive of lateral order was first observed by North et al. (16). Additional detail has recently been obtained which offers strong support for a highly ordered substructure (17–19). Medium-angle diffraction patterns of rat tail tendon are shown in Fig. 5. The reflections are consistent with a microfibril of the type proposed by Smith which has a native D period and a diameter of about 4.0 nm. The data also suggest that the molecules in the microfibril are supercoiled in such a way that the 5_1 Smith helix or the alternate 5_2 helix is converted to 4_1 symmetry. The microfibrils are packed on a square lattice in rat tail tendon fibers (19).

The pitch and handedness of the molecular supercoil is not known. However, a pitch near 70 nm is most likely. This is the third level of coiling alluded to earlier.

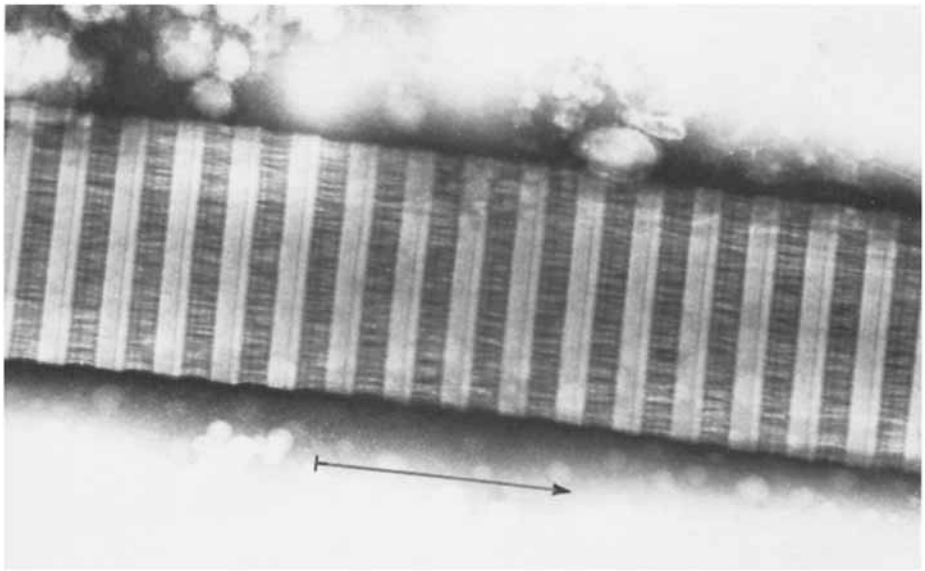


Fig. 6. An electron micrograph of a native collagen fibril from rat tail tendon negatively stained with 2% neutral phosphotungstic acid. The period is 66.8 nm. The filamentous structure is readily seen. The electron micrograph was taken by J. Woodhead-Galloway and B. Doyle.

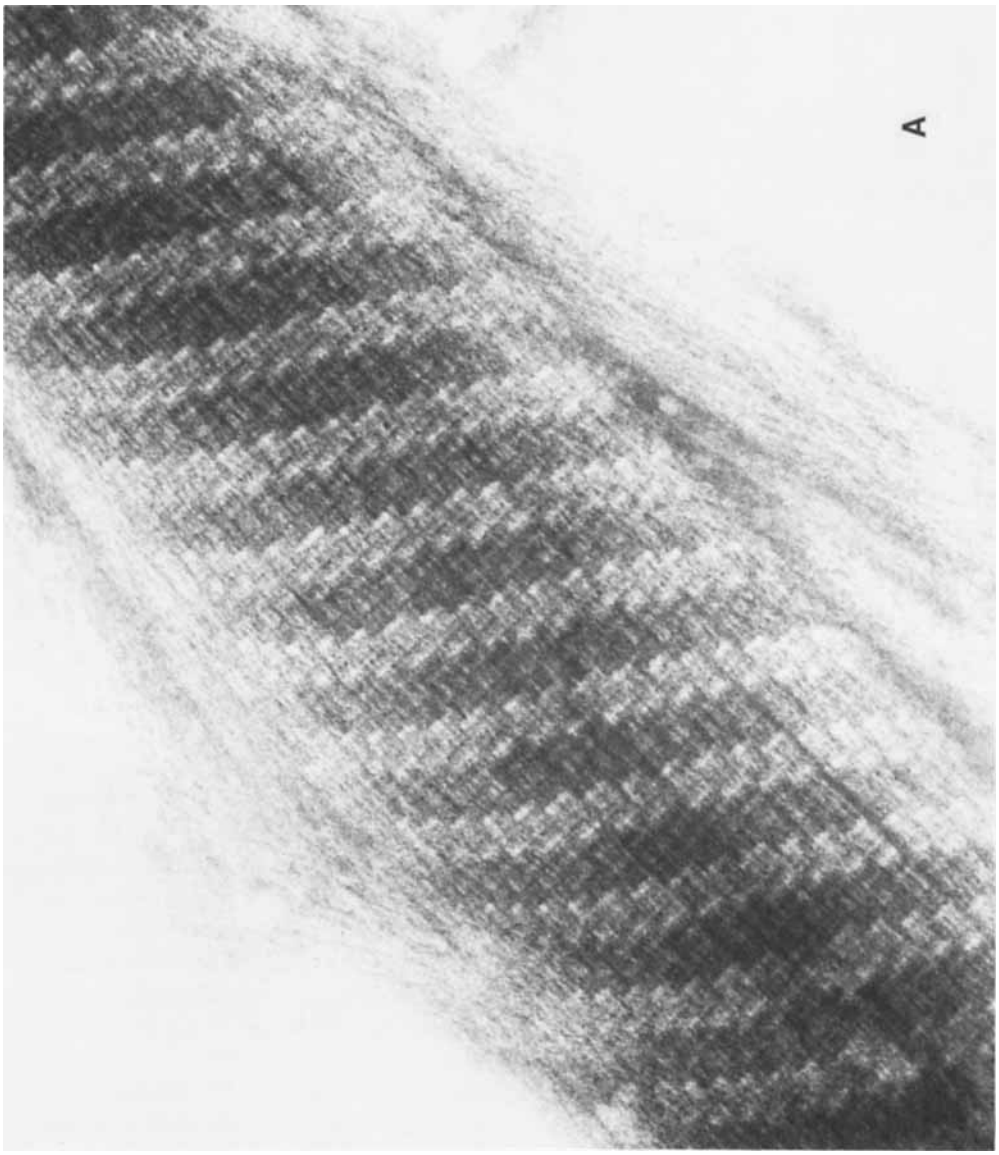
Electron Microscopy

That native collagen fibrils have a filamentous appearance has been noted by many investigators beginning more than twenty years ago (see Figs. 1 and 22 of Ref. 20) and has been described most recently by Boutielle and Pease (21). This characteristic can be seen in the negatively stained fibril in Fig. 2 but is more obvious in the fibril shown in Fig. 6. The filaments do not appear to be straight or regularly spaced. However, these features are difficult to interpret since staining and drying artifacts may exist.

That the filaments themselves are not artifacts of the preparation is evident from electron micrographs of fibrillar collagen reconstituted from solution. In this procedure soluble collagen in solution is precipitated by warming (or in other ways). The fibrils will sometimes be indistinguishable from native fibrils, but they often show defective packing. An example appears in Fig. 7. The native period can be seen but the packing appears to be "loose," showing the filaments more readily. Sometimes the native period cannot be seen at all. Filaments are evident in the background as well as in the fibrils.

We believe, as discussed later, that the filaments are the same in all cases but the degree of order in their packing varies; that is, the filaments have a regular 67 nm repeat and probably have the microfibril structure proposed by Smith.

A favorable case in which this could be demonstrated by electron microscopy presented itself in the form of the obliquely banded fibrils reported by Bruns et al. (22). An example of such a fibril is shown in Fig. 8. It was shown that this form consists of 14


A

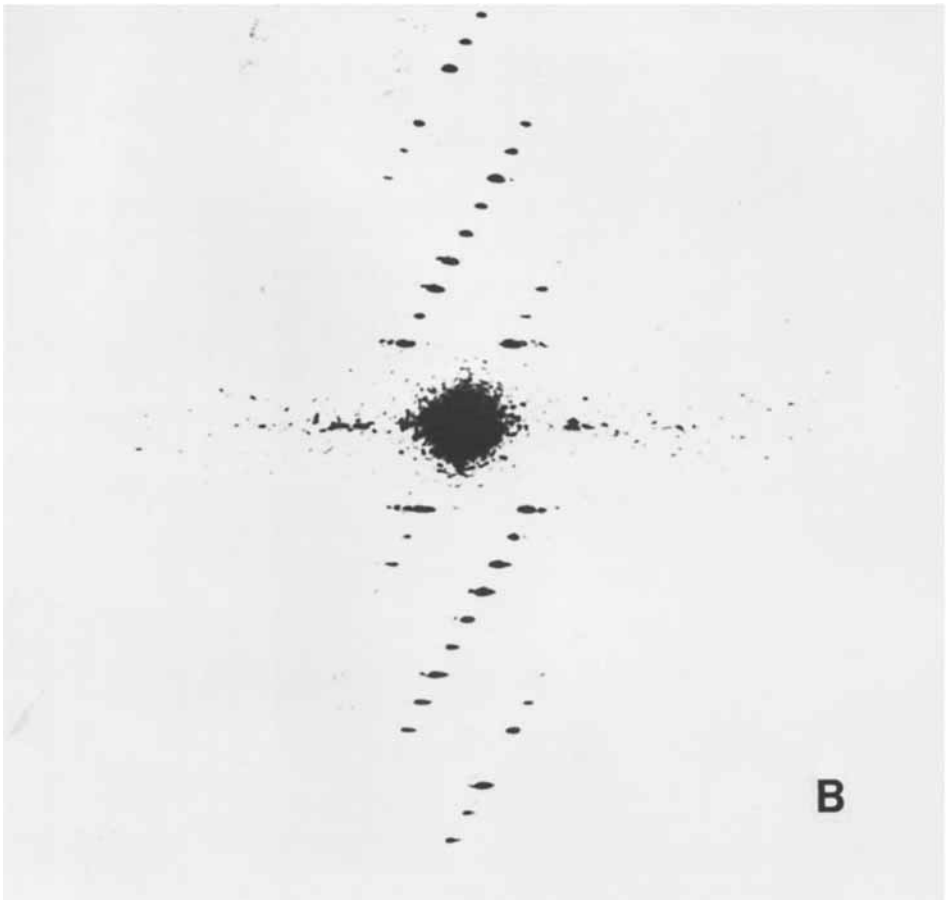


Fig. 8. An obliquely banded collagen fibril obtained by reconstitution of cartilage collagen and negatively stained. Reproduced from Bruns et al. (22) with permission of the American Association for the Advancement of Science (copyright 1973). (B) The optical diffraction pattern of (A). Reproduced from Doyle et al. (18) with permission of the Royal Society (London).

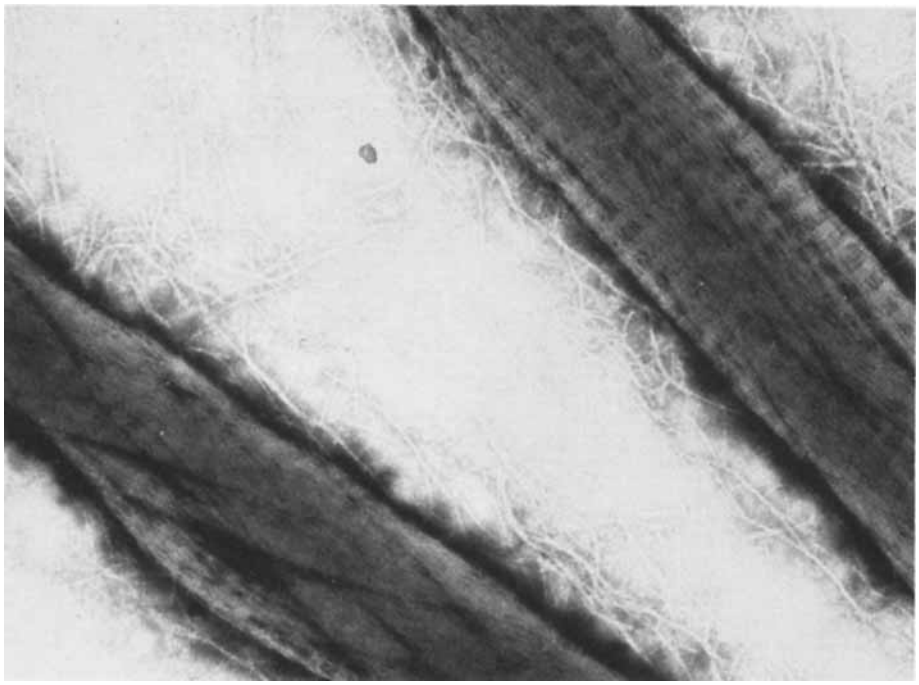
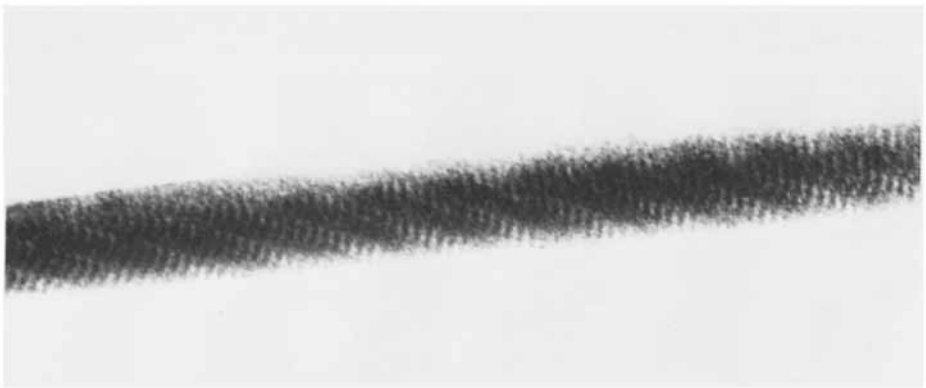


Fig. 7. Poorly ordered reconstituted collagen fibrils negatively stained with 1% neutral phosphotungstic acid. The 66.8 nm native period is barely visible. Loosely packed filaments in the structure and in the background can be readily seen. The electron micrograph was taken by J. Woodhead-Galloway and K. Piez.

nm subfibrils, each with the native repeat, axially displaced by about 10 nm. This interpretation is confirmed by the optical diffraction pattern of the electron micrograph, also shown in Fig. 8. The layer lines arising from the 67 nm repeat are sampled by diagonal row lines arising from the regular displacement of the subfibrils. The distance between diagonal row lines on a line parallel to the equator is the reciprocal of the subfibril width, 14 nm in the fibril in Fig. 8.

On repeating this work, it was found by Doyle et al. (15) that the subfibril width varied over a very wide range for unknown reasons, although the axial displacement of subfibrils was essentially constant. It was important to find the smallest subfibril width in which the native repeat was still present. An obliquely banded fibril in which the bands are quite steep, suggesting a very narrow subfibril, is shown in Fig. 9. The structure cannot be determined from the micrograph itself, but the optical diffraction pattern, also shown in Fig. 9, shows that it is closely related to the structure determined for the fibril in Fig. 8. The native 67 nm repeat is present and its reflections are sampled by diagonal row lines. Measurements of this and other patterns show the subfibril width to be about 4.0 nm. The structural relationships in the obliquely banded fibrils arising from subfibrils of several widths are shown schematically in Fig. 10.



A

Fig. 9. (A) An obliquely striated collagen fibril similar to that seen in Fig. 8 but with steeper oblique bands. (B) The optical diffraction pattern of (A). Reproduced from Doyle et al. (15) with permission of the Royal Society (London).

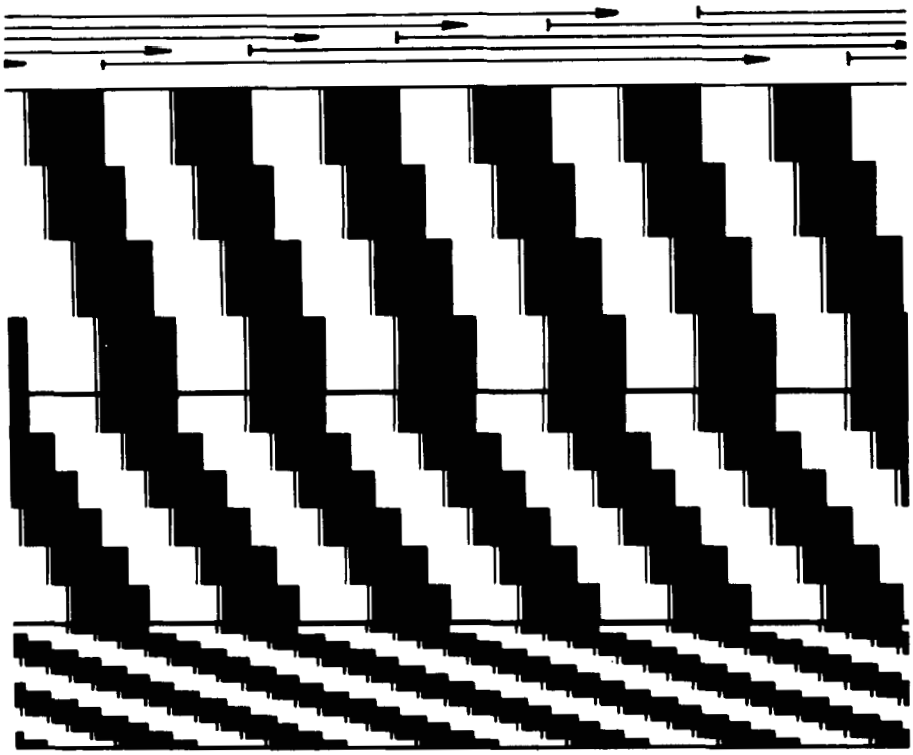


Fig. 10. A schematic illustration of obliquely banded fibrils showing the native subfibrils of varying widths (34, 14, and 4.0 nm) and constant stagger. The narrowest subfibril is the Smith microfibril. The relationship of the Petrushka-Hodge packing to these structures appears above. Compare with Figs. 8 and 9.

A diameter of 4.0 nm limits the number of possible structures that can have the native repeat to the five-stranded Smith microfibril or one of the closely related structures (15). As already noted, the Smith microfibril is preferred for independent reasons. The evidence provided by electron microscopy and x-ray diffraction therefore strongly supports a microfibril structure.

Symmetrically Banded Fibrils

The demonstration of a microfibril structure and the finding of a new polymorphic form of collagen has permitted the structure of symmetrically banded collagen fibrils to be interpreted (6). These fibrils are sometimes obtained by reconstitution from solutions of collagen for poorly understood reasons. They have a 67 nm period but have planes of symmetry.

The new polymorphic form, shown in Fig. 11, has a "checkerboard" pattern. It can be readily seen that it is composed of subfibrils, each with the native repeat like the

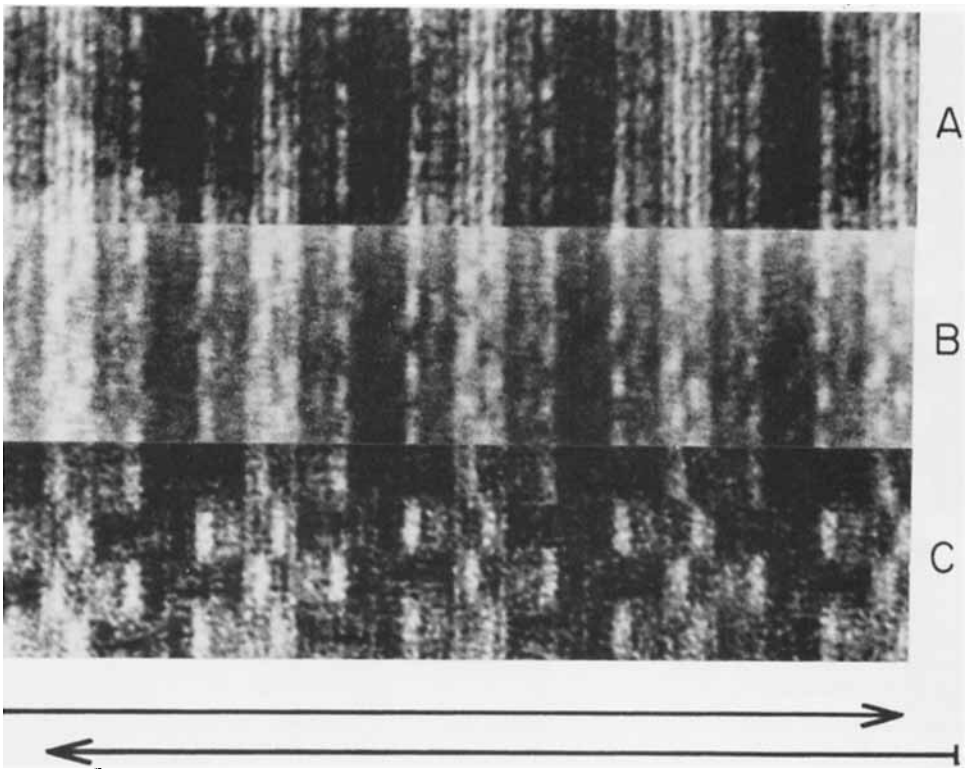


Fig. 11. A symmetrically banded fibril (A) and two "checkerboard" patterns (B and C) of different subfibril widths. The relationship can be readily seen by tilting the page. The arrows show the relationship between molecules in adjacent subfibrils. Reproduced from Doyle et al. (6) with permission. of the Royal Society (London).

obliquely banded fibrils, but the subfibrils are antiparallel. Molecules in adjacent subfibrils are staggered. Again, like the obliquely banded fibrils, the subfibril width varies. Examples about 25 nm and 15 nm wide appear in Fig. 11. A comparison of these to the symmetrically banded fibril, which has no apparent lateral substructure, shows them to be closely related. Apparently the symmetrically banded fibrils have a substructure not resolved in the micrographs which produces the native period, but the antiparallel arrangement of the substructure produces the planes of symmetry. The substructure is most likely the five-stranded Smith microfibril. The structural relationships are shown schematically in Fig. 12.

FIBRIL ASSEMBLY

The five-stranded microfibril appears to be a constant feature of several collagen aggregates, but the way the microfibrils are packed can vary widely. These structures in-

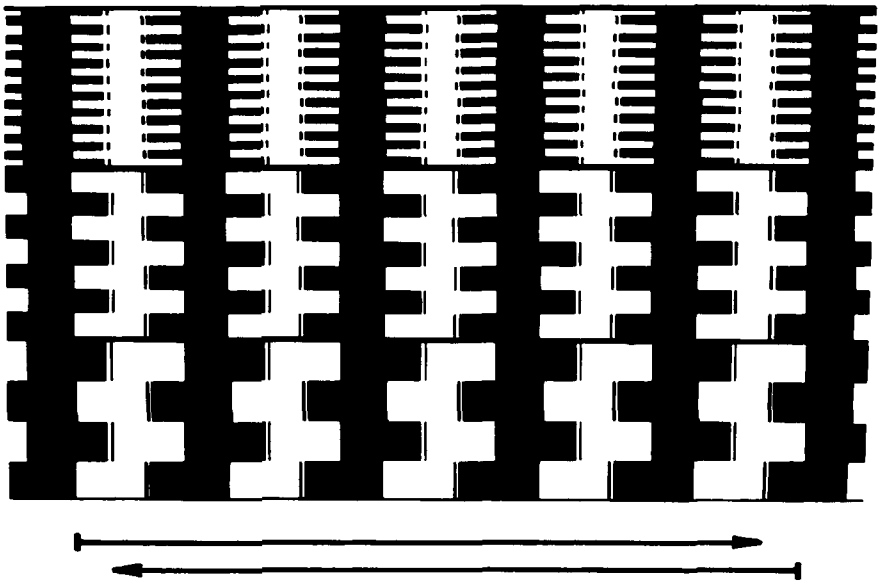


Fig. 12. A schematic illustration of antiparallel subfibrils of varying widths showing the structural relationship in the fibrils in Fig. 11. The widths are 25, 15, and 4.0 nm. The narrowest subfibril is the Smith microfibril.

clude parallel aligned microfibrils (native fibrils), parallel staggered microfibrils, antiparallel microfibrils, and loosely packed, poorly aligned microfibrils. This suggests that the intrafibrillar and interfibrillar interactions differ at least quantitatively.

This difference may be important in fibril assembly. The microfibril is a closed structure in its lateral dimensions, but can grow in length by the addition of single molecules. Packing of microfibrils, however, occurs by side-to-side association. Therefore, fibril length and width can be separately regulated, a useful biological situation.

Fibril assembly has not been studied in any detail. The recent findings related to fibril structure, reviewed here, and collagen chemistry should make it possible to design experiments that will tell us how this occurs and lead us to a better understanding of suprafibril structure.

SUMMARY

Chemical, x-ray diffraction, and electron optical studies have combined to show that collagen molecules are assembled into a substructure by a regular stagger and rotation of molecules to produce a five-stranded helical microfibril which has the native repeat of 66.8 nm. Side-to-side packing of aligned microfibrils produces the native fibril. Different arrangements of microfibrils produce other types of fibrils.

REFERENCES

1. Traub, W., and Piez, K. A., Adv. Prot. Chem. 25:243 (1971).
2. Hulmes, D., Miller, A., Parry, D., Piez, K. A., and Woodhead-Galloway, J., J. Mol. Biol. 79:137 (1973).
3. Gallop, P. M., Blumenfeld, P. O., and Seifter, S., Ann. Rev. Biochem. 41:617 (1972).
4. von der Mark, K., Wendt, P., Rexrodt, F., and Kühn, K., FEBS Lett. 11:105 (1970).
5. Balian, G., Click, E. M., Hermodsen, M. A., and Bornstein, P., Biochem. 11:3798 (1972).
6. Doyle, B., Hulmes, D., Miller, A., Parry, D., Piez, K. A., and Woodhead-Galloway, J., Proc. Roy. Soc. B. 186:67 (1974).
7. Piez, K. A., Balian, G., Click, E. M., and Bornstein, P., Biochem. Biophys. Res. Commun. 48:990 (1972).
8. Fietzek, P. P., Kell, I., and Kühn, K., FEBS Lett. 26:66 (1972).
9. Tomlin, S. G., and Worthington, C. R., Proc. Roy. Soc. A 235:189 (1956).
10. Segrest, J. P., and Cunningham, L. W., Nature New Biol. 234:26 (1971).
11. Chapman, J. A., and Hardcastle, R. A., Conn. Tissue Res. in press.
12. Hodge, A. J., and Schmitt, F. O., Proc. Acad. Sci. 46:186 (1960).
13. Chapman, J. A., Conn. Tissue Res. in press.
14. Smith, J. W., Nature 219:157 (1968).
15. Doyle, B., Humes, D., Miller, A., Parry, D., Piez, K. A., and Woodhead-Galloway, J., Proc. Roy. Soc. B, in press.
16. North, A. C. T., Cowan, P. M., and Randall, J. T., Nature 174:1142 (1954).
17. Miller, A., and Wray, J. S., Nature 230:437 (1971).
18. Miller, A., and Parry, D., J. Mol. Biol. 75:441 (1973).
19. Fraser, R., Miller, A., and Parry, D., J. Mol. Biol. 83:281 (1974).
20. Bear, R. S., Adv. Prot. Chem. 7:85 (1952).
21. Bouteille, M., and Pease, D. C., J. Ultrastructure Res. 35:314 (1971).
22. Bruns, R. R., Trelstad, R. L., and Gross, J., Science 181:269 (1973).