Journal of Supramolecular Structure 2:99–102 (1974)

SELF-ASSEMBLY OF COLLAGEN

John H. Fessler

Molecular Biology Institute and Biology Department, Univesity of California, Los Angeles, California 90024

The structure and self-assembly of collagen and procollagen molecules are reviewed. The registration peptides of procollagen have specific recognition properties which assure 1) selection of component polypeptide chains and 2) registration of their N-termini, facilitating orderly folding into a collagen helix. The stability of this helix relative to body temperature is critically altered by post-ribosomal hydroxylation of proline residues. The registration peptides of procollagen may have additional functions such as preventing intracellular fiber formation.

Collagen fibers are composed of collagen molecules (which have also been called tropocollagen molecules). Under appropriate conditions collagen molecules will self-assemble into fibers, and conversely fibers can be dissolved to collagen molecules. Fiber formation is an endothermic process and is accompanied by loss of solvated water (1–4). K. A. Piez's contribution to this issue deals in detail with the mutual arrangement of collagen molecules within a fiber. The following properties of collagen molecules are given as a brief background.

Collagen molecules are highly elongated cylinders: about 3000 Å long but only about 17 Å in diameter, with a mass of about 300,000 daltons. They consist of three polypeptide chains, each of about 100,000 daltons, which extend throughout the length of the molecule. Each of these chains is called an alpha chain and is folded in the special configuration of a collagen helix. The three helices of the three chains are closely fitted together in a superhelix, and the three chains mutually stabilize the configuration, partly by H bonds between the chains.

The amino acid composition of the collagen alpha chains varies with species and tissues. In the collagen of cartilage all three alpha chains of one molecule are identical [called $\alpha 1$ (type II)] and the molecule is denoted $[\alpha 1 \text{ (type II)}]_3$. The commonly investigated collagens of vertebrate bone and skin are of the form $[(\alpha 1)_2\ (\alpha 2)_1]$. Here two $\alpha 1$ (type I) chains are combined in each molecule with a slightly different chain called $\alpha 2$, and a problem of self-assembly is how these closely related chains are selected in the 2:1 ratio for each molecule. In both type I and type II collagens the amino termini of each of

a set of three chains are together at one end of the collagen molecule. The arrangement of collagen molecules within a fiber is also unidirectional with respect to the amino-carboxyl direction of the polypeptide chains.

The folding and mutually close arrangement of three collagen chain helices in each molecule only leave room for glycine, with its negligible sidechain of one H atom, in the inner regions near the axis of the superhelix. Structural considerations therefore require that glycine residues are placed regularly as every third residue in each chain, and this has been strikingly borne out in the amino acid sequence determined for $\alpha 1$ chains (see Ref. 28 for further detail).

From the point of view of self-assembly of the triple-stranded molecule this suggests that the regions of the helices which interact most closely are monotonously similar, and that the three chains could easily assemble without their ends being properly in register. Unmatched portions of such chains from one triplet might then associate with unmatched regions of other triplets. It is believed that this happens when denatured alpha chains are allowed to renature rapidly in relatively concentrated solutions, as in the setting of table jello (gelatin being a form of denatured and partly degraded collagen). Well-matched triple-stranded molecules can be obtained by careful, slow renaturation and annealing of alpha chains, but it was suspected that some special mechanism facilitated this process in vivo.

Speakman (5) postulated a biosynthetic precursor, procollagen, in which each alpha chain has an extension peptide at the N-terminus, with the specific function of assuring the initial, correct association of the ends of each triplet set. This material has been found (see Refs. 6, 7, and 8 for reviews). The extension peptides do not contain sufficient glycine to be folded in the collagen helix form and, unlike the native collagen helix polypeptide region, are susceptible to pepsin and chymotrypsin attack. Normal biological excision of the extension peptides is performed by one (or more) special procollagen peptidase(s) (9). Short regions at the ends of the final collagen molecules are also not in the collagen helix fold and play a special role in the association of collagen molecules and their eventual covalent cross-linking within fibers (10).

The stability of the collagen fold is also aided by the high proline content which accounts for about 20% of the residues of collagen. About half of these prolines are hydroxylated after ribosomal synthesis. Comparison with synthetic polyprolines shows that the pyrollidine ring stabilizes the collagen fold (for review see 10). A correlation exists between hydroxyproline content and denaturation temperature of various collagens (11) and studies with synthetic polypeptides showed that the hydroxyl groups substantially increase thermal stability (12). Such stabilization could occur by hydrogen bonding, through a water molecule, to an adjacent chain of a collagen triplet (13).

Hydroxylation modifies bioassembly in a novel way. It is known that hydroxylation follows ribosomal synthesis, and the enzyme proline hydroxylase has been located electron microscopically in the walls and lumen of the rough endoplasmic reticulum (14). Non-hydroxylated collagen chains have denaturation temperatures below body temperature (15–17) and proline hydroxylase acts on collagen chains in the "random" coil form. With hydroxylation the denaturation temperature of the collagen is increased to body temperature or just above and correspondingly folding occurs.

Our recent studies (18) show that procollagen chains are synthesized individually, and that subsequently the extension peptides at their amino ends associate specifically and become disulfide linked to each other. We postulate (19) that the disulfide bridges determine the correct mutual alignment of the chains. In other work (17) we showed that this association of the extension peptides proceeds similarly even when hydroxylation is prevented by addition of inhibitors, and that the collagen portions of the resulting disulfide-linked trimer are denatured at body temperature. As the trimer molecules were shown to have the stoichiometric composition $[(\text{pro } \alpha 1)_2 \ (\text{pro } \alpha 2)_1]$, we concluded that the extension peptides by themselves have the configurational properties to associate in this way, independently of the collagen portions of the chains.

Abolition of hydroxylation also interferes in the normal processing and transport of these materials, so that the nonhydroxylated procollagen trimers accumulate within the cells. If further synthesis is stopped with cycloheximide and *then* hydroxylation is restarted, the previously accumulated trimer is hydroxylated and concomitantly its denaturation temperature rises (17); processing and secretion from the cells also occur.

A problem in the self-assembly of collagen is that fibers are formed extracellularly, but not inside cells. Procollagens form fibers poorly, or not at all, and this may prevent intracellular fiber formation. Presumably conversion to collagen occurs at the time of secretion or extracellularly. Procollagen molecules can occur intracellularly at high concentrations, and have been found in vesicles associated with their transport (20–22). The secretion of collagen is interfered with by agents which destabilize microtubules, such as colchicine and elevated pressure (23, 24).

Little is known about the biological processes of extracellular collagen fiber formation. The in vitro reassembly of purified collagen obtained from existing fibers has been studied in some detail. An initial nucleation phase is followed by growth of fibers without substantial change in the distribution of fiber diameters (15). As might be expected these processes are quite sensitive to the ionic environment, and especially to the presence of polyelectrolytes. Whether such noncollagenous materials, which might well be present at the in vivo site of fibrillogenesis, participate in a catalytic manner is not known. We must distinguish between the clearly established ability of purified collagen, on its own, to self-assemble into typical collagen fibers, and the kinetic and biological problems of rapid fiber formation and eventual interaction with other materials to form a coherent matrix.

To obtain a measure of the ability of purified collagen to associate into fibers we have studied the temperature dependence of the equilibrium between these two states (3, 4, 26). A temperature of half-association, equivalent to Tm, was found to be 30° for rat skin collagen (26). This is considerably below 37°, the body temperature, at which most of the material exists as fibers. Empirically it is known that collagen fibers made in vitro at about 30° show much better order, as judged by their striated appearance in the electron microscope, than fibers made by warming solutions to 37°. Presumably there is a better chance in vitro at 30° than at 37° for initially mismatched molecules to find correct locations. This suggests that the in vivo mechanism of collagen fiber formation might involve some process which would facilitate good, rapid ordering of the associating molecules, though it would *not* determine the final arrangement. An intriguing possibility is that the final excision of the procollagen registration peptides might occur at the site of

the growing fiber, and somehow facilitate the ordered placing of collagen molecules on the fiber surface. This would be separate from its interference with further fiber formation, if not removed, as in the genetic disease of dermatosporaxis (27).

ACKNOWLEDGMENTS

Financial support is gratefully acknowledged from National Institutes of Health Grants AM 13748 and AM 11546, and American Heart Association Research Grant 71-1031.

REFERENCES

1. Jackson, D. S., and Fessler, J. H., Nature 176:69 (1955).
2. Gross, J., J. Exp. Med. 108:215 (1958).
3. Fessler, J. H., Biochem. J. 76:452, 463 (1960).
4. Fessler, J. H., In "Structure and Function of Connective and Skeletal Tissues," S. Fitton-Jackson (Ed.), p. 80, London, Butterworths (1965).
5. Speakman, P. T., Nature 229:241 (1971).
6. Schofield, J. D., and Prockop, D. J., Clin. Orthop. 97:175 (1973).
7. Bornstein, P., Ann. Rev. Biochem. 43 (in press).
8. Martin, G. R., Byers, P. H., and Piez, K. A., Adv. Prot. Chem. (in press).
9. Kohn, L. D., Isersky, C., Zupnik, J., Lenaers, A., Lee, G., and Lapiere, C. M., Proc. Nat. Acad. Sci. U.S. 71:40 (1974).
10. Traub, W., and Piez, K. A., Adv. Prot. Chem. 25:243 (1971).
11. Gustavson, K. H., Svensk. Kem. Tidskr. 65:70 (1953).
12. Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D. J., Biochim. Biophys. Acta 303:198 (1973).
13. Ramachandran, G. N., Bansal, M., and Bhatnagar, R. S., Biochim. Biophys. Acta 332:166 (1973).
14. Olson, B. R., Berg, R. A., Kishida, Y., and Prockop, D. J., Science 182:825 (1973).
15. Berg, R. A., and Prockop, D. J., Biochem. Biophys. Res. Comm. 52:115 (1973).
16. Rosenbloom, J., Harsch, M., and Jimenez, S., Arch. Biochem. Biophys. 158:478 (1973).
17. Fessler, L. I., and Fessler, J. H., J. Biol. Chem. (in press).
18. Morris, N. P., Fessler, L. I., Weinstock, A., and Fessler, J. H. F., (submitted for publication).
19. Fessler, L. I., Burgeson, R. E., Morris, N. P., and Fessler, J. H., Proc. Nat. Acad. Sci. U.S. 70:2993 (1973).
20. Weinstock, A., and Leblond, C. P., J. Cell. Biol. 51:26 (1971).
21. Trelstaed, R. L., J. Cell. Biol. 48:689 (1971).
22. Weinstock, M., and Leblond, C. P., J. Cell. Biol. 60:92 (1974).
23. Ehrlich, H. P., and Bornstein, P., Nature New Biol. 238:257 (1972).
24. Diegelman, R. F., and Peterkovsky, B., Proc. Nat. Acad. Sci. U.S. 69:892 (1972).
25. Bard, J. P. L., and Chapman, J. A., Nature New Biol. 246:83 (1973).
26. Fessler, J. H., and Tandberg, D., J. Supramol. Struct. (in preparation).
27. Bailey, A. J., and Lapiere, C. M., Eur. J. Biochem. 34:91 (1973).
28. Piez, K. A., and Miller, A., J. Supramol. Struct. 2:121 (1974).