Proc. Nat. Acad. Sci. USA
Vol. 72, No. 8, pp. 3009–3013, August 1975
Biochemistry

Interchain disulfide bonds in procollagen are located in a large nontriple-helical COOH-terminal domain

(collagen precursors/tadpole collagenase/pulse-label experiments/chick bone)

PETER H. BYERS*, EVA MARIE CLICK*, ELVIN HARPER†, AND PAUL BORNSTEIN*

* Departments of Biochemistry and Medicine, University of Washington, Seattle, Wash. 98195; and the † Department of Chemistry, University of California at San Diego, La Jolla, Calif. 92037

Communicated by Earl P. Benditt, May 22, 1975

ABSTRACT **Tadpole collagenase (EC 3.4.24.3) cleaved chick cranial bone procollagen into two triple-stranded fragments, PC^A and PC^B. Only PC^B, with an estimated molecular weight of about 60,000 for each component chain after reduction, was found to contain interchain disulfide bonds. The analogous cleavage of collagen is known to produce a large NH_2-terminal fragment with a molecular weight of 70,000 for each chain and a small COOH-terminal fragment containing chains of about 25,000 molecular weight. Since PC^B was too small to represent the product NH_2-terminal to the site of collagenase cleavage, localization of interchain disulfide bonds to a COOH-terminal domain in procollagen was indicated. This assignment was confirmed by Dintzis-type short-term labeling experiments. Procollagen obtained by acid extraction of bone lacked the COOH-terminal disulfide-bonded domain. The findings support a model for procollagen consisting of three proα chains each containing nonhelical NH_2-terminal extensions of 20,000 molecular weight and COOH-terminal extensions of about 35,000 molecular weight, the latter linked by interchain disulfide bonds.**

Since the initial identification of a biosynthetic precursor to collagen, it has become apparent that the nontriple-helical portion of the molecule is important for several functions, which may include molecular assembly, maintenance of intracellular solubility, transmembrane transport, and regulation of fibril formation (see ref. 1 for a review). Early structural studies of collagen precursors suggested, on the basis of electron micrographs of segment-long-spacing crystallites, that there was a single "noncollagenous" domain at the NH_2-terminus of each precursor chain (2, 3). Subsequent chemical studies of these precursors and of those procollagen-derived proteins which accumulated in the tissues of animals affected with dermatosparaxis tended to confirm the structural studies (4, 5).

More recent studies of (*a*) peptides derived from procollagen secreted into culture medium by cells, (*b*) precursor chains protected from proteolysis during purification, and (*c*) peptides secreted into culture medium by cranial bone have suggested that the nonhelical, noncollagenous domain is considerably larger than was previously thought (6–9). Tanzer and his colleagues (10) have recently indicated that there is a noncollagenous peptide extension at the COOH-terminus of the procollagen molecule, in addition to that at the NH_2-terminus. In this report, we present data to show that the COOH-terminal domain of procollagen is larger than the NH_2-terminal extensions, that the composition is different, and that interchain disulfide bonds occur exclusively in the COOH-terminal region.

Abbreviations: EDTA, ethylenediaminetetraacetic acid; DATD, *N,N'*-diallyltartardiamide; Bis, methylene bisacrylamide.

MATERIALS AND METHODS

Procollagen, pulse labeled with L-[2,3-^{3}H]proline (43.1 Ci/mmol, New England Nuclear) for 18 min or with L-[*G*-^{3}H]tryptophan (7.9 Ci/mmol, New England Nuclear) for 2 hr (7, 36) was extracted from 17-day-old chick embryo cranial bones with 1 M NaCl, 50 mM Tris·HCl, pH 7.5, to which 25 mM EDTA, 10 mM *N*-ethylmaleimide, 1 mM diisopropylfluorophosphate or phenylmethylsulfonylfluoride, and 1 mM benzamidine·HCl had been added to inhibit proteolysis. The bones were homogenized in that solvent, centrifuged at 39,000 × *g* for 15 min, and the labeled procollagen together with coextracted collagen was precipitated from the supernatant with 20% NaCl. The precipitate was collected by centrifugation, redissolved in extraction buffer, precipitated with 5% cold trichloroacetic acid and then dissolved in extraction buffer containing 0.1 M Tris·HCl, pH 7.5. Acid-extracted procollagen was obtained as previously described (4), except that extraction was limited to 1 hr, and was further purified by chromatography on DEAE-cellulose by minor modifications of the method of Smith *et al.* (8, 11). Isotopically labeled collagen was obtained as described by Vuust and Piez (12) and was used to generate standard fragments with tadpole collagenase.

Tadpole collagenase (EC 3.4.24.3) was purified as previously described (13), lyophilized, and stored at −20° until use. About 100 μg of collagenase were used to digest the procollagen and collagen extracted from three cranial bones; the digestion was carried out in 0.15 M NaCl, 50 mM Tris·HCl, 5 mM $CaCl_2$, pH 7.4, at 30° for 16 hr and terminated by chilling to 4°.

Analytical gel electrophoresis in sodium dodecyl sulfate was performed as described by Goldberg *et al.* (14) with dansylated collagen α1 chains, β_{11} components, and α1(I)-CB7 as internal standards. In some experiments (identified in the text), 7.5% acrylamide crosslinked with 0.3% *N,N'*-diallyltartardiamide (DATD) was used (15). The gels were sliced at 1 mm intervals and digested with 30% H_2O_2 [gels crosslinked with methylene bisacrylamide (Bis)] or 2% periodic acid (DATD crosslinked gels) and the radioactivity was counted by standard methods.

In experiments designed to demonstrate which of the fragments resulting from collagenase cleavage of procollagen was derived from the COOH-terminus of the molecule, one dozen bones were incubated for 20 min in the presence of either 100 μCi of [^{3}H]proline or L-[3,4-^{3}H]leucine (50.5 Ci/mmol, New England Nuclear). The ^{3}H-labeled bones were then combined with the same number of bones incubated separately with 100 μCi of the respective [^{14}C]amino

acid (L-[*U*-^{14}C]leucine, 0.29 Ci/mmol or L-[*U*-^{14}C]proline, 0.24 Ci/mmol, New England Nuclear) for shorter times (3, 4.5, 6, and 10 min with leucine or 3, 5.5, 7.5, and 10 min with proline) and the procollagen was extracted as described above. Following cleavage of procollagen with tadpole collagenase the resultant fragments were electrophoresed and the total radioactivity in ^{3}H and ^{14}C was determined for each gel slice.

RESULTS

The fragments that resulted from digestion of procollagen and collagen with tadpole collagenase were analyzed by electrophoresis in sodium dodecyl sulfate acrylamide gels. Cleavage of procollagen, prior to reduction of disulfide bonds, resulted in three major bands, two in the region of the dansylated α1 marker and one that migrated between the α and β markers (Fig. 1, upper panel). Following reduction, the bands in the α region were unchanged but the higher molecular weight band was replaced by two incompletely separated bands that migrated between the α1 and α1-CB7 markers (Fig. 1, lower panel). In contrast, cleavage of collagen with tadpole collagenase (Fig. 2) produced the fragments expected from the known specificity of this enzyme (16). The high-molecular-weight chains, migrating more rapidly than α1 represent the α1 and α2 segments in TC^{A}‡, and the lower molecular weight peptides near α1-CB7 are the fragments in TC^{B}.

Since no bands were observed which migrated in the region of α1-CB7 after cleavage of procollagen (Fig. 1), it is apparent that the biosynthetic precursor contains additional sequences at the COOH-terminal ends of the proα chains (as well as NH_2-terminal extensions). We have arrived at the conclusion, to be supported by the data which follow, that the band migrating between α and β markers (Fig. 1, upper panel) represents the COOH-terminal region of procollagen (PC^{B}), crosslinked by disulfide bonds. After reduction, this fragment yielded peptides migrating 40–50 mm from the top of the gel (Fig. 1, lower panel). The NH_2-terminal fragment (PC^{A}) after dissociation in sodium dodecyl sulfate buffer yielded chains which migrate more slowly than those derived from TC^{A} (compare Figs. 1 and 2), indicating the presence of NH_2-terminal extensions. Since migration on sodium dodecyl sulfate acrylamide gels was not affected by reduction (Fig. 1), we conclude that interchain disulfide bonds do not exist in this NH_2-terminal domain.

It was possible to assign the disulfide-bonded fragment (PC^{B}) to a COOH-terminal location in procollagen by a Dintzis-type experiment (17). In these short-term label experiments, procollagen was extracted from pooled tissues (see *Materials and Methods*) and then digested with tadpole collagenase. The resultant peptides were electrophoresed after reduction of disulfide bonds, and the relative specific activities (^{14}C/^{3}H) in PC^{A} and PC^{B} were calculated. In these experiments incorporation of the ^{3}H isotope, to produce a uniformly labeled protein, was limited to 20 min in order to minimize the synthesis of a labeled procollagen which might then be converted in the tissue to collagen. The presence of large amounts of labeled collagen might have complicated

‡ Cleavage of collagen with tadpole collagenase results in two fragments, TC^{A} (tropocollagen A) representing three quarters of the molecule NH_2-terminal to the cleavage site and TC^{B} (tropocollagen B) the COOH-terminal quarter (17). By analogy, the cleavage products of procollagen are designated PC^{A} (procollagen A) and PC^{B} (procollagen B). The chains resulting from dissociation of PC^{A} and PC^{B} fragments are designated proα^{A} and proα^{B}, respectively

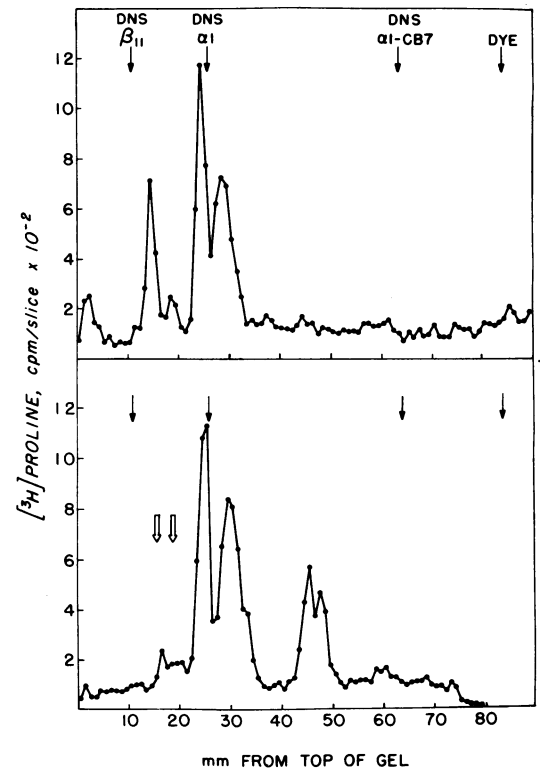

FIG. 1. Sodium dodecyl sulfate acrylamide gel electrophoresis of procollagen after digestion with tadpole collagenase; upper panel, unreduced; lower panel, reduced with 1% dithiothreitol for 30 min at 70°. The positions of dansylated collagen components used as internal markers are denoted by the thin arrows. These positions are relative since dansylation increases the mobility of collagen components in this system. The large open arrows denote the positions of proα1 and proα2 chains obtained after reduction of intact procollagen. Bis crosslinked gels were used.

the analysis of the results, since the PC^{A} and TC^{A} fragments were not completely separated by gel electrophoresis (compare Figs. 1 and 2). However, after 20 min the distribution of proline and of leucine between the PC^{A} and PC^{B} fragments was consistent with the calculated compositions (36) and so labeling closely approximated uniformity.

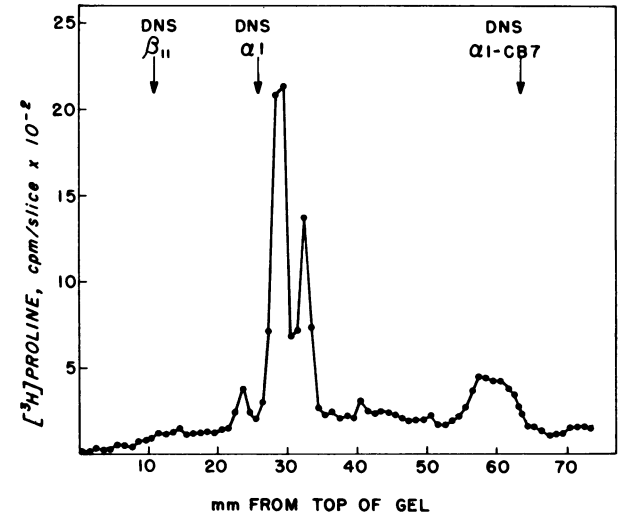

FIG. 2. Sodium dodecyl sulfate acrylamide gel electrophoresis of collagen after digestion with tadpole collagenase. Identical patterns were obtained with or without reduction. See legend to Fig. 1 for additional details.

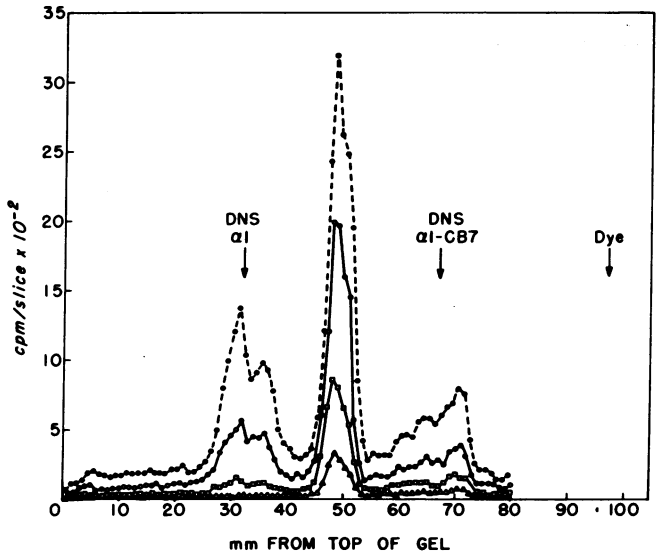

FIG. 3. Sodium dodecyl sulfate acrylamide gel electrophoresis of [^{3}H]leucine-labeled procollagen (●- - -●) and procollagen pulse-labeled with [^{14}C]leucine for 4.5 (△—△), 6 (□—□), or 10 (○—○) min. Procollagens were cleaved with tadpole collagenase and reduced with dithiothreitol. Gels were 7.5% in acrylamide and cross-linked with 0.3% DATD.

Radioactivity in leucine first appeared in the PC^B peptides and was seen only after a delay in the PC^A peptides (Fig. 3). At all time points from 3 to 10 min the relative specific activities of the PC^B peptides were higher than those for the PC^A peptides. With increasing time, the relative specific activities of PC^A and PC^B approached equality (Fig. 4) as incorporation of isotope extended into the NH_2-terminal region of the molecule. Identical results were obtained in experiments in which a proline label was used (data not shown). Thus, by the criteria established for short-term label

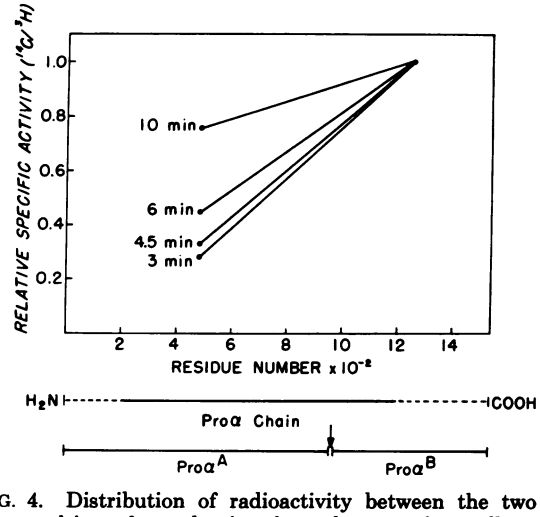

FIG. 4. Distribution of radioactivity between the two fragments resulting, after reduction, from cleavage of procollagen with tadpole collagenase. Procollagens were pulse-labeled with [^{14}C]leucine for 3, 4.5, 6, and 10 min and with [^{3}H]leucine for 20 min. The relative specific activities (^{14}C/^{3}H) were normalized to 1.0 for the fragment with the highest ratio of ^{14}C to ^{3}H at each time point. This peptide was located COOH-terminal to the site of cleavage by the protease. The points are plotted at the midpoint of the resulting fragments even though this may not represent the true distribution of leucine in the fragments. An intact proα chain (with solid and dashed lines representing triple-helical and nontriple-helical regions, respectively) and the site of cleavage with tadpole collagenase are presented diagrammatically below the plot.

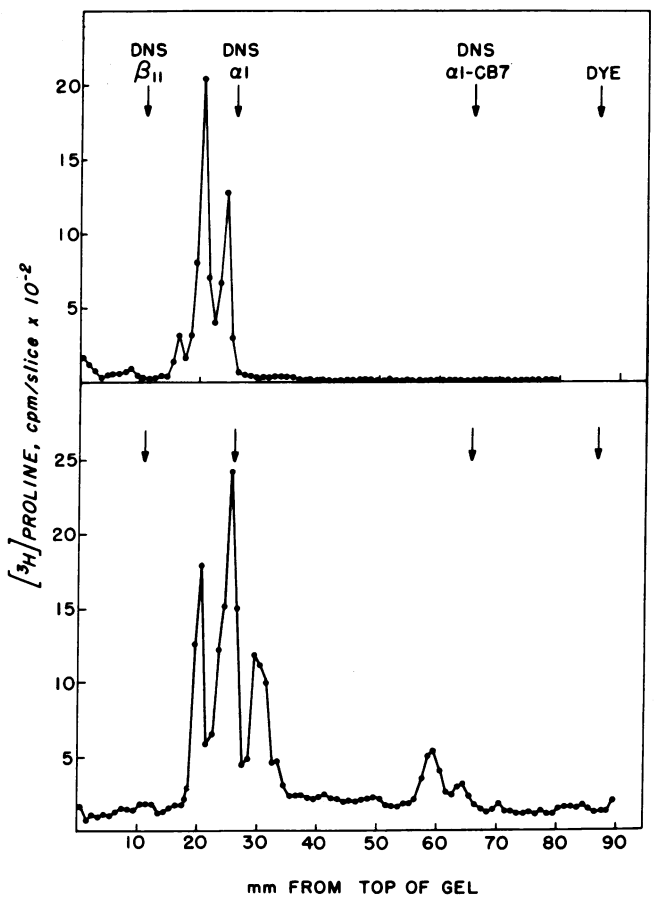

FIG. 5. Sodium dodecyl sulfate acrylamide gel electrophoresis of acid-extracted procollagen (upper panel), and a tadpole collagenase digest of this procollagen (lower panel). The patterns were identical whether the samples were reduced prior to electrophoresis or not. See legend to Fig. 1 for additional details.

experiments (12, 17), those peptides which migrate between the α chain and the α1-CB7 markers, and which are linked by disulfide bonds, are derived from the COOH-terminal region of the procollagen molecule.

Recent studies (7, 8) suggest that acid-extracted procollagen is a derivative of procollagen which lacks a fraction of the nonhelical region as well as interchain disulfide bonds. To determine more directly if acid-extracted procollagen lacks the COOH-terminal domain, consistent with its lack of interchain disulfide bonds, this procollagen was cleaved with tadpole collagenase and the products were examined by sodium dodecyl sulfate acrylamide gel electrophoresis (Fig. 5). In this experiment, cleavage was incomplete and the products include a fraction of undigested chains. However, it can be seen that the smaller peptides have the same mobility as those derived from collagen (Fig. 2), indicating an absence of COOH-terminal extensions, whereas the larger peptides have the mobility of those derived from intact procollagen.

When procollagen was labeled with tryptophan rather than proline, virtually all the radioactivity was found in the PC^B fraction (data not shown). This experiment suggests that the majority of the tryptophan in procollagen exists in the COOH-terminal region. A small amount of tryptophan has been identified in acid-extracted procollagen (18) and is therefore located in the NH_2-terminal domain (see below).

The hybrid nature of the PC^B peptides, consisting as they do of collagenous and noncollagenous sequences, makes an estimation of their molecular weight by sodium dodecyl sul-

fate acrylamide gel electrophoresis difficult (18). The mobility of these peptides (Fig. 2) suggests a molecular weight of about 60,000 for the proα^B peptides and therefore a molecular weight of about 35,000 for the COOH-terminal extension of each proα chain. This agrees closely with the molecular weight of 33,000 determined for the reduced product of the triple-chain disulfide-bonded fragment released into the medium of cultured bone during conversion of procollagen to collagen (8). Similarly, an estimated molecular weight of about 90,000 for the proα^A peptides (Fig. 2) is consistent with a molecular weight of 70,000 for each chain of the large collagenase-produced collagen fragment (19) augmented by an NH_2-terminal extension of about 20,000 molecular weight (18).

DISCUSSION

The experiments described in this paper establish a number of important structural features of the procollagen molecule and, in addition, reconcile data from different laboratories which had previously appeared contradictory. The presence of a large COOH-terminal domain in procollagen, first suggested by the studies of Tanzer *et al.* (10), is conclusively established. Although both the NH_2- and COOH-terminal extensions contain cysteine, interchain disulfide bonds are limited to the latter domain. The recent findings that disulfide bond formation occurs relatively late in molecular assembly (20, 21) and that disulfide bond formation and helix formation are coordinate (22) are consistent with a model in which molecular assembly occurs through interaction of specific sequences within the COOH-terminal domain after completion of chain synthesis. None of these studies, however, exclude initial association through NH_2-terminal extensions.

Molecular weights ranging from approximately 115,000 for the constituent chains of dermatosparactic and acid-extracted procollagen (18, 23) to 147,000 for the proα1 chain of chick bone procollagen (36) have been reported. It is now apparent that acid-extracted and dermatosparactic procollagens represent truncated precursors which have lost the nonhelical COOH-terminal domain. Very likely, some of the lower molecular weight forms identified in the medium of cultured fibroblasts (11, 24) and extracted from tissues of normal animals (25, 26) are similarly altered.

The relatively short COOH-terminal extension present in collagen is markedly acid-labile (27); presumably cleavage of similar bonds during extraction results in the fraction identified as acid-extracted procollagen. There is also variation in the length of the collagen molecule at the COOH-terminal end when genetically different collagens are compared (28). Some of these differences may result from variable cleavage during extraction but differences in the extent to which the COOH-terminal nonhelical extensions are retained in tissues could contribute to tissue-specific structural and functional properties of collagen. In type III collagen the interchain disulfide bonds present in the COOH-terminal region of this molecule (29) may reside in a nonhelical domain very close to the termination of the helix.

In dermatosparaxis and its human counterpart, Ehlers–Danlos syndrome, type VII (30), the lack of an enzymatic activity responsible for cleavage of the NH_2-terminal extensions (31) results in the accumulation in tissues of derivatives of procollagen lacking the COOH-terminal domain. Presumably, therefore, at least two different enzymes must be involved in the normal conversion of procollagen to collagen. Intermediates consistent with such a complex conversion scheme have been identified and partially characterized (32).

The structural differences between the NH_2- and COOH-terminal domains in procollagen are reflected in differences in amino-acid composition (6, 8, 18) and in antigenicity (33). The compositions presented by Sherr *et al.* (6) are similar to those reported by Murphy *et al.* (8), suggesting that the large, disulfide-linked trimer isolated after bacterial collagenase digestion of procollagen derives from the COOH-terminus rather than from the NH_2-terminus as the former authors proposed. It can be noted that the total incorporation of leucine into PC^B exceeded that into PC^A despite the lower molecular weight of the PC^B fragment (Fig. 3). This is consistent with the assignment of the procollagen-derived fragment released into the medium of cultured bone to the COOH-terminal region of the molecule, since it contains as much leucine as the remainder of the procollagen molecule (8). The findings of Nist *et al.* (33) that antibodies developed against acid-extracted procollagen do not react with a disulfide-bonded peptide isolated after bacterial collagenase digestion of procollagen, and vice versa, confirm the structural differences between the two domains.

We are not certain that the procollagen used as a substrate in these experiments contains all the additional sequences present in the native precursor. If an even higher molecular weight precursor exists, additional enzymes may be involved in its conversion to procollagen. We have not, however, detected evidence for the single chain precursor suggested by others (10, 34). Sodium dodecyl sulfate acrylamide gel electrophoresis of tadpole collagenase digests of procollagen partially resolved the polypeptides derived from proα1 and proα2 in both fragments (Fig. 1). Calculation of the relative specific activity ($^{14}C/^{3}H$) of individual gel slices across the PC^A and PC^B bands obtained after short term labeling (Fig. 3) failed to reveal the gradient of activity that would be expected if procollagen were synthesized as a single chain. These, and other pulse-label studies (35), argue for separate initiation of individual proα chains but do not exclude the existence of a polycistronic mRNA.

We thank Janet Monson and Jeffrey Davidson for their assistance with the preparation of isotopically labeled procollagens and for many helpful discussions. This work was supported by National Institutes of Health Grants AM 11248, DE 02600, and AM 15367. P.H.B. is a fellow of the Helen Hay Whitney Foundation.

1. Bornstein, P. (1974) *Annu. Rev. Biochem.* **43**, 567–603.
2. Stark, M., Lenaers, A., Lapière, C. M. & Kühn, K. (1971) *FEBS Lett.* **18**, 225–227.
3. Jimenez, S. A., Dehm, P., Olsen, B. R. & Prockop, D. J. (1973) *J. Biol. Chem.* **248**, 720–729.
4. Bornstein, P., von der Mark, K., Wyke, A. W., Ehrlich, H. P. & Monson, J. M. (1972) *J. Biol. Chem.* **247**, 2808–2813.
5. Furthmayr, H., Timpl, R., Stark, M., Lapière, C. M. & Kühn, K. (1972) *FEBS Lett.* **28**, 247–250.
6. Sherr, C. J., Taubman, M. B. & Goldberg, B. (1973) *J. Biol. Chem.* **248**, 7033–7038.
7. Monson, J. M. & Bornstein, P. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3521–3525.
8. Murphy, W. H., von der Mark, K., McEneany, L. S. G. & Bornstein, P. (1975) *Biochemistry*, in press.
9. Dehm, P., Olsen, B. R. & Prockop, D. J. (1974) *Eur. J. Biochem.* **46**, 107–116.
10. Tanzer, M. L., Church, R. L., Yaeger, J. A., Wampler, D. E. & Park, E. D. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 3009–3013.
11. Smith, B. D., Byers, P. H. and Martin, G. R. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3260–3262.
12. Vuust, J. & Piez, K. A. (1970) *J. Biol. Chem.* **245**, 6201–6207.
13. Harper, E., Bloch, K. J. & Gross, J. (1971) *Biochemistry* **10**, 3035–3041.

14. Goldberg, B., Epstein, E. H. & Sherr, C. J. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3655–3659.
15. Anker, H. S. (1970) *FEBS Lett.* **7**, 293.
16. Gross, J., Harper, E., Harris, E. D., Jr., McCroskery, P. A., Highberger, J. H., Corbett, C. & Kang, A. H. (1974) *Biochem. Biophys. Res. Commun.* **61**, 555–562.
17. Naughton, M. A. & Dintzis, H. M. (1962) *Proc. Nat. Acad. Sci. USA* **48**, 1822–1830.
18. von der Mark, K. & Bornstein, P. (1973) *J. Biol. Chem.* **248**, 2285–2289.
19. Kang, A. H., Nagai, Y., Piez, K. A. & Gross, J. (1966) *Biochemistry* **5**, 509–515.
20. Harwood, R., Grant, M. E. & Jackson, D. S. (1973) *Biochem. Biophys. Res. Commun.* **55**, 1188–1196.
21. Fessler, L. I., Burgeson, R. E., Morris, N. P. & Fessler, J. H. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2993–2996.
22. Schofield, J. D., Uitto, J. & Prockop, D. J. (1974) *Biochemistry* **13**, 1801–1806.
23. Lenaers, A., Ansay, M., Nusgens, B. V. & Lapière, C. M. (1971) *Eur. J. Biochem.* **23**, 533–543.
24. Lichtenstein, J. R., Byers, P. H., Smith, B. D. & Martin, G. R. (1975) *Biochemistry* **14**, 1589–1594.
25. Veis, A., Anesey, J., Yuan, L. & Levy, S. J. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1464–1467.
26. Byers, P. H., McKenney, K. H., Lichtenstein, J. R. & Martin, G. R. (1974) *Biochemistry* **13**, 5243–5248.
27. Rauterberg, J. (1973) *Clin. Orthop. Relat. Res.* **97**, 196–212.
28. Stark, M., Miller, E. J. & Kühn, K. (1972) *Eur. J. Biochem.* **27**, 192–196.
29. Chung, E., Keele, E. M. & Miller, E. J. (1974) *Biochemistry* **13**, 3459–3464.
30. Lichtenstein, J. R., Martin, G. R., Kohn, L. D., Byers, P. H. & McKusick, V. A. (1973) *Science* **182**, 298–300.
31. Kohn, L. D., Isersky, C., Zupnik, J., Lenaers, A., Lee, G. & Lapière, C. M. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 40–44.
32. Davidson, J. M. & Bornstein, P. (1975) *Fed. Proc.* **34**, 562.
33. Nist, C., von der Mark, K., Hay, E., Olsen, B. R., Bornstein, P., Ross, R. & Dehm, P. (1975) *J. Cell Biol.* **65**, 75–87.
34. Park, E., Tanzer, M. L. & Church, R. L. (1975) *Biochem. Biophys. Res. Commun.* **63**, 1–10.
35. Vuust, J. & Piez, K. A. (1972) *J. Biol. Chem.* **247**, 856–862.
36. Monson, J. M., Click, E. M. & Bornstein, P. (1975) *Biochemistry*, in press.