Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 262, No. 13, Issue of May 5, pp. 5947-5950, 1987

Basement Membrane Procollagen IV and Its Specialized Carboxyl Domain Are Conserved in *Drosophila*, Mouse, and Human*

(Received for publication, December 31, 1986)

Bruce Blumberg‡§, Albert J. MacKrell‡¶, Pamela F. Olson‡, Markku Kurkinen‖, Janet M. Monson, Jeanette E. Natzle‡‡, and John H. Fessler‡§§**

*From the ‡Department of Biology and Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90024, the ‖Department of Medicine, University of Medicine and Dentistry of New Jersey Robert Wood Johnson Medical School, Piscataway, New Jersey, 08854, the **Genetics Unit, Shriners Hospital, Montreal, H3G 1A6 Canada, and the ‡‡Department of Zoology, University of California at Davis, Davis, California 95616*

Interaction with the extracellular matrix is important for the proliferation and differentiation of cells during development. A specialized extracellular matrix, basement membrane, is built around a scaffold of procollagen IV molecules. We report the sequence of a 2.5-kilobase cDNA which contains the carboxyl end of a *Drosophila melanogaster* procollagen IV. The amino acid sequence of the carboxyl-terminal domain, which forms an essential intermolecular linkage between procollagen IV molecules, is 59% identical in *Drosophila* and vertebrate procollagens IV, and an additional 17% of residues are conservatively substituted. This implies that the nature of the linkage is also conserved. We suggest that intermolecular junctions through procollagen IV carboxyl domains are fundamental elements of the molecular architecture of Metazoan basement membranes and have been conserved during evolution. The isolation and identification of this basement membrane collagen gene of *Drosophila* will help in deducing the function of procollagen IV in basement membranes.

The immediate environment of cells influences their differentiated functions. Particularly important are the basement membranes, or basal laminae, which are ubiquitous condensations of specialized extracellular matrix at the basal surface of epithelia. They also surround muscle cells and nerves and are thought to play an important role in establishing and maintaining tissue organization (1). For example, culture of epithelial cells is greatly facilitated by basement membranes or gels made from their components (2).

To obtain insight into the functions of basement membranes during development we chose *Drosophila melanogaster*, which allows both genetic and recombinant DNA manipulation. From cultures of *Drosophila* cells we purified a collagen, laminin, and several other glycoproteins which are biochemically and ultrastructurally similar to their vertebrate counterparts and locate by immunofluorescence to the sites of basement membranes (3). The collagen resembles a major component of vertebrate basement membranes, procollagen IV (3, 4).

A vertebrate procollagen IV molecule consists of three polypeptides formed from two genetically distinct pro-α chains. Both homotrimeric (5) and heterotrimeric (6) procollagen IV molecules exist. The molecule has a central collagen triple helix flanked by amino and carboxyl domains which form junctions between molecules to produce a molecular network (7, 8). The noncollagenous carboxyl region of each polypeptide, also called NCl, has about 230 amino acid residues that are folded into a globular structure which is internally linked by six disulfide bridges (9). The three carboxyl peptide regions of each molecule together form a structure that is seen electron microscopically as a knob at one end of each thread-like molecule (10). The junctional complex of two fused knobs, *i.e.* six carboxyl peptides, is an important element of vertebrate basement membranes (7, 8, 11).

Whereas structural requirements of the collagen triple helix necessitate that every third residue is glycine, there are numerous nonhelical interruptions in the $(Gly\text{-}X\text{-}Y)_n$ sequence of vertebrate pro-α1(IV) and pro-α2(IV) (12, 13). These interruptions probably cause the flex or bend locations which are seen electron microscopically along the thread-like molecules (14). Similar interruptions were also found in chick types IX and X collagen which do not form fibers and are not basement membrane collagens (15, 16). Such interruptions have also been demonstrated in nucleotide sequences coding for collagens of *Caenorhabditis elegans* (17) and sea urchin (18). A sequence encoding interrupted collagen triple helix was isolated from a *Drosophila melanogaster* genomic DNA library, and it was suggested that, based on the interruptions and on the size and temporal expression of the mRNA, this could represent a portion of a *Drosophila* basement membrane collagen gene (19, 20). With the help of this cloned *Drosophila* sequence, DCG1, we demonstrate that procollagen IV chains exist in *Drosophila* and that they contain a highly conserved form of the specialized carboxyl peptide region of vertebrate procollagen IV, and we have identified the gene of which DCG1 is a part.

EXPERIMENTAL PROCEDURES

RNA Isolation and cDNA Cloning—Total cellular RNA was isolated (21) from *Drosophila* Kc cells (22) which secrete collagen (3). cDNA was prepared from this RNA by the method of Gubler and Hoffman (23) except that the double-stranded cDNA was made blunt-ended with bacteriophage T4 polymerase (Pharmacia P-L Biochemicals), and *Eco*RI linkers were ligated onto the cDNA after treatment with *Eco*RI methylase (New England Biolabs). Chromatography on Sepharose CL-4B removed excess linkers and cDNA molecules smaller than 500 base pairs. The cDNA obtained was ligated into λgt10 (24) and packaged *in vitro* using Gigapack packaging extract (Stratagene Cloning Systems, San Diego, CA) and amplified. The

* This work was supported in part by United States Public Health Service Grant AG-02128 (to J. H. F.). 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J02727.

§ Recipient of National Research Service Award CA-09056.

¶ Recipient of National Research Service Award GM-07185.

§§ To whom correspondence should be addressed: Dept. of Biology and Molecular Biology Inst., UCLA, 405 Hilgard Ave., Los Angeles, CA 90024.

library contains 10^6 independent clones averaging greater than 1.9 kilobases.

Screening of the cDNA Library—The 1.6-kilobase *Eco*RI-*Bam*HI fragment of DCG1 sequenced by Monson *et al.* (19) was subcloned into the plasmid vector Bluescribe (Stratagene). Templates for *in vitro* transcription were prepared and a ^{32}P-labeled RNA probe for screening was synthesized using the procedure supplied by Stratagene. The RNA was separated from unincorporated nucleotides by Sephadex G-50 spun-column chromatography (25). Plaque hybridization (26) was performed with duplicate filters at 75 °C in 5 × SSPE (1 × SSPE is 180 mM NaCl, 50 mM NaH_2PO_4, 1 mM EDTA, pH 7.0), 0.1% sodium dodecyl sulfate overnight. The filters were washed at 75°C in 0.1 × SSC (1 × SSC is 150 mM NaCl, 15 mM sodium citrate, pH 7.0), 0.1% sodium dodecyl sulfate and exposed to Kodak XAR-5 film between intensifying screens. Positive clones were isolated, plaque-purified, and subcloned into Bluescribe for restriction mapping.

DNA Sequencing and Sequence Analysis—Restriction fragments were subcloned into M13mp18 and M13mp19 (27) and sequenced by the dideoxy method using [α-^{35}S]dATP and avian myeloblastosis virus reverse transcriptase (28).[1] The DNA was sequenced on both strands and across all restriction sites used in cloning. The DNA sequences were maintained and aligned using Staden's DB-system (29). DNA and amino acid sequences were analyzed using various programs of the University of Wisconsin Genetics Computer Group (30) and the National Biomedical Research Foundation (31). Hydrophobic correlation coefficients were calculated using the program HYDCOR and the optimal matching hydrophobicity scale (32) which was kindly made available by D. Eisenberg (UCLA).

RESULTS

We screened a cDNA library made from *Drosophila* Kc cells (22) with ^{32}P-labeled RNA prepared from DCG1. Several overlapping cDNA clones were isolated and restriction-mapped. Both strands of one of these clones, CDA3, were sequenced (28) (Fig. 1). Nucleotides 2347–2529 were determined from a separate clone, CDB5, on only one strand. The putative 3′ terminus of the mRNA is marked by two overlapping consensus polyadenylation signals (34) followed by 14 consecutive A residues. The (Gly-*X*-*Y*)$_n$ portion of the sequence in Fig. 1 has five interruptions. The fourth and fifth interruptions correspond to interruptions found in all three reported vertebrate pro-α(IV) chains (12, 13).

A comparison of the amino acid sequence deduced for the carboxyl domain of this *Drosophila* collagen with those of mouse[1] (35) and human pro-α1(IV) (36, 37), and with mouse pro-α2(IV),[1] is shown in Fig. 2. The *Drosophila* sequence is 59.4% (137/231) identical with either mouse chain and 58.5% (135/231) identical with human pro-α1(IV). Many of the substitutions are conservative changes. The positions of the 12 cysteine residues are conserved; these residues are important in the three-dimensional structure of the molecules and in the intermolecular associations between carboxyl domains in the formation of networks of procollagen IV (9, 11). In addition to the consensus sequence of 112 residues shown in Fig. 2, an additional 40 positions have residues with similar chemical properties.

A duplicated domain, with 36% identity of amino acid residues, was identified within the sequence of the carboxyl domain of pro-α1(IV)[1] (35) and pro-α2(IV)[1] from mouse and human pro-α1(IV) (36, 37). The *Drosophila* sequence also shows this duplication (Fig. 3) and the two halves are 33% identical in amino acid sequence. Thus, this duplication appears to be a common feature of procollagens IV.

Sweet and Eisenberg (32) have defined a hydrophobic correlation coefficient which reliably predicts whether two aligned sequences share the same three-dimensional structure. Proteins with closely similar three-dimensional structures have a hydrophobic correlation coefficient greater than 0.4 (32). For example, the coefficient for horse α and β globins is 0.68. The correlation coefficient of *Drosophila* with mouse pro-α1(IV) is 0.78, with mouse pro-α2(IV) is 0.72, and between the two mouse chains is 0.82. The hydropathy plots (33) of these peptides are highly similar (not shown). We conclude that these carboxyl domains share a common major folding pattern. In contrast to the close similarity of this *Drosophila* collagen carboxyl peptide with those of procollagens IV there are only distant, although significant, homologies with the carboxyl propeptides of vertebrate procollagens I, II, III, and V.[2] We conclude that the *Drosophila* collagen sequence we describe here is a procollagen IV and designate it *Drosophila* pro-α1(IV).

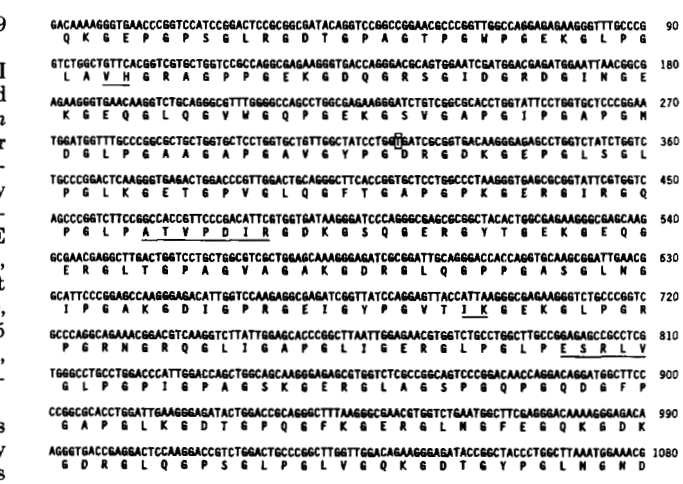

FIG. 1. **The DNA sequence and the deduced amino acid sequence of λCDA3 are shown.** Interruptions in the triple helical repeat are *underlined*. Cysteine residues are *circled*. The 3′ end of the overlap between CDA3 and DCG1 is at nucleotide 500. An *arrow* marks the beginning of the noncollagenous carboxyl domain. Two overlapping, consensus polyadenylation signals are *boxed*. CDA3 and DCG1 were obtained from DNA libraries derived from the Oregon R and Canton S strains, respectively, of *D. melanogaster*. Their sequences agree, except that at position 323 in the above sequence a T replaces a C of DCG1, which leaves the corresponding amino acid, glycine, unchanged.

[1] Kurkinen, M., Condon, M. R., Blumberg, B., Barlow, D. J., Quinones, S., Saus, J., and Pihlajaniemi, T. (1987) *J. Biol. Chem.*, in press.

[2] B. Blumberg, A. J. MacKrell, L. I. Fessler, and J. H. Fessler, unpublished observations.

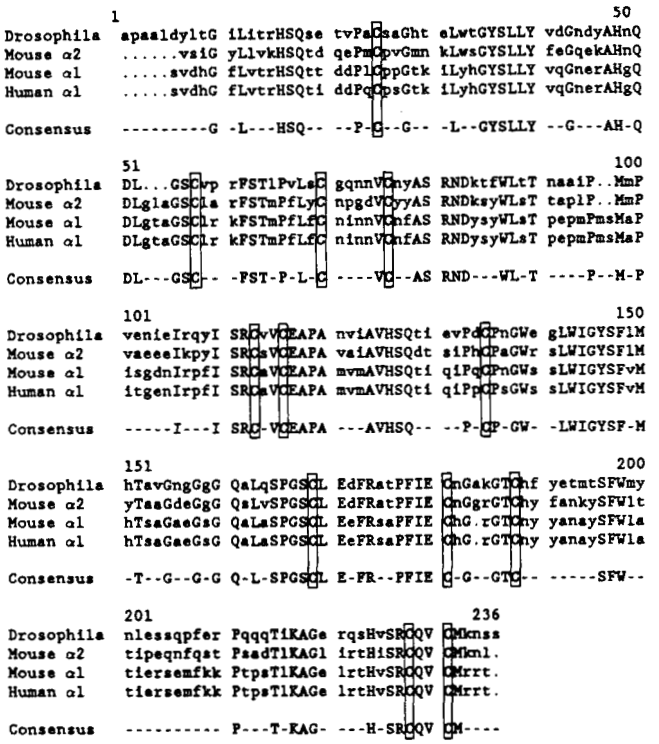

FIG. 2. The deduced amino acid sequences for *Drosophila* pro-α1(IV), mouse pro-α1(IV), mouse pro-α2(IV), and human pro-α1(IV) were aligned by the program GAP (30), which introduced gaps to maximize sequence similarity while minimizing gap length and number. *Capital letters* mark residues which are identical in all four sequences. Cysteine residues are *boxed*.

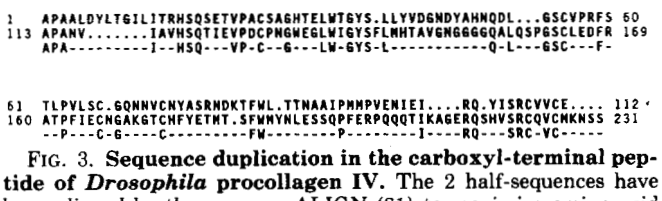

FIG. 3. **Sequence duplication in the carboxyl-terminal peptide of *Drosophila* procollagen IV.** The 2 half-sequences have been aligned by the program ALIGN (31) to maximize amino acid homology while introducing a minimum of gaps. The parameters used in running the program were the mutation data matrix + 6, a gap penalty of 6 and 1000 random runs. Residue *1* is the first residue of the carboxyl domain.

DISCUSSION

As insects have a radically different body plan from vertebrates, with an exoskeleton and hemocoel, the extracellular matrix of their basement membranes could differ substantially from that of vertebrates. Furthermore, in contrast to vertebrate basement membranes which are placed at the interfaces of cells with either extracellular matrix or another cell layer, some insect basement membranes are bathed in hemolymph. Yet the *Drosophila* collagen carboxyl domain has nearly as many identical amino acid residues with either mouse chain (137 out of 231) as the two murine chains have with each other (142 out of 229). As the two mouse chains together form a hexameric junctional complex, the residues which are common to all three chains may be responsible for key interactions within the complex. The high hydrophobic correlation coefficients support the concept of a common three-dimensional structure of these carboxyl domains. Presumably the internal tandem duplication, and the conserved cysteine residues which suggest conserved sets of disulfide links, are all particularly suited to fit into the complex of six peptides that is an important junctional complex in vertebrate basement membranes and which now appears to have been widely conserved during evolution.

Northern blot hybridization with CDA3[2] and DCG1 detects an RNA of approximately 6.4 kilobases (19) which is of appropriate size to encode a pro-α(IV) chain. While this research was in progress, two *Drosophila* collagen-like genes encoding 6.3-kilobase mRNAs were identified and mapped to chromosome loci 9E and 25C (38). CDA3[2] and DCG1 hybridize exclusively to location 25C (20). Although the two genes at 9E and 25C could encode *Drosophila* pro-α1 and pro-α2 chains, the collagen which we isolated from *Drosophila* cell cultures consists predominantly of one polypeptide (3). The *Drosophila* carboxyl peptide is about equally similar to either of the vertebrate chains, which suggests that the gene duplication which gave rise to pro-α1(IV) and pro-α2(IV) occurred after the divergence of vertebrates and invertebrates.

Of the many invertebrate collagens (39), only one specifically identified basement membrane collagen has been purified, from the isolated intestinal basement membrane of the nematode *Ascaris suum* (40). Noelken *et al.* (40) isolated dimeric molecules with centrally located knobs which they suggested could correspond to the hexameric junctional complexes of vertebrate procollagen IV. Recently it was suggested, based on partial genomic sequences, that a basement membrane collagen exists in sea urchins (18), as judged by interrupted $(\text{Gly-}X\text{-}Y)_n$ sequences, mRNA size, and temporal expression, and by an apparent genomic organization of small exons separated by 400–1300-base pair introns. In contrast, the DCG1 portion of this *Drosophila* collagen gene has large coding regions separated by small introns (19). Thus, while there is conservation of amino acid sequence in one important region of the molecule, the genomic organization may not have been conserved in all evolutionary lines.

Interruptions of the $(\text{Gly-}X\text{-}Y)_n$ sequence must influence important properties of any collagen in which they occur. A full analysis of this problem for procollagens IV will require the complete sequences of the *Drosophila* and other pro-αIV chains. Here we note that two of the interruptions in the $(\text{Gly-}X\text{-}Y)_n$ sequence shown in Fig. 1 have approximate equivalents in the mouse and human chains (12, 13). Electron micrographs of the collagen produced by *Drosophila* cells (3, 10), from which our cDNA library was constructed, indicate that these thread-like molecules have bends and kinks which may correspond to some interruptions of $(\text{Gly-}X\text{-}Y)_n$ and could have a function in basement membranes.

We plan to modify the gene encoding *Drosophila* pro-α1(IV) and reintroduce it into embryos in order to investigate the formation of basement membranes and their role in development. In conclusion, the present findings strongly support the concept that key features of the molecular architecture of basement membranes have been conserved during the evolution of the Metazoa.

REFERENCES

1. Hay, E. D. (1981) in *Cell Biology of the Extracellular Matrix* (Hay, E. D., ed) pp. 379–409, Plenum Press, New York
2. Kleinman, H. K., McGarvey, M. C., Hassell, J. R., Star, V. C., Cannon, F. B., Laurie, G. W., and Martin, G. R. (1986) *Biochemistry* **25,** 312–318
3. Fessler, J. H., Lunstrum, G., Duncan, K. G., Campbell, A. G., Sterne, R., Bachinger, H. P., and Fessler, L. I. (1984) in *The Role of Extracellular Matrix in Development* (Trelstad, R., ed) pp. 207–219, Alan R. Liss, Inc., New York
4. Kefalides, N. A., Alper, R., and Clark, C. C. (1979) *Int. Rev. Cytol.* **61,** 167–228
5. Haralson, M. A., Federspiel, S. J., Martinez-Hernandez, A., Rhodes, R. K., and Miller, E. J. (1985) *Biochemistry* **24,** 5792–5797
6. Trueb, B., Grobli, B., Spiess, M., Odermatt, B. F., and Winter-

halter, K. H. (1982) *J. Biol. Chem.* **257,** 5239–5245
7. Timpl, R., Wiedemann, H., Van Delden, V., Furthmayr, H., and Kuhn, K. (1981) *Eur. J. Biochem.* **120,** 203–211
8. Yurchenko, P. D., and Furthmayr, H. (1984) *Biochemistry* **23,** 1839–1850
9. Weber, S., Engel, J., Wiedemann, H., Glanville, R. W., and Timpl, R. (1984) *Eur. J. Biochem.* **139,** 401–410
10. Fessler, J. H., Bachinger, H. P., Lunstrum, G., and Fessler, L. I. (1982) in *New Trends in Basement Membrane Research* (Kuhn, K., Schone, H., Timpl, R. ed) pp. 145–153, Raven Press, New York
11. Timpl, R., Oberbaumer, I., Von Der Mark, H., Bode, W., Wick, G., Weber, S., and Engel, J. (1985) *Ann. N. Y. Acad. Sci.* **460,** 58–72
12. Babel, W., and Glanville, R. W. (1984) *Eur. J. Biochem.* **147,** 217–224
13. Schwarz, U., Schuppan, D., Oberbaumer, I., Glanville, R. W., Deutzmann, R., Timpl, R., and Kuhn, K. (1986) *Eur. J. Biochem.* **157,** 49–56
14. Hofmann, H., Voss, T., Kuhn, K., and Engel, J. (1984) *J. Mol. Biol.* **172,** 325–343
15. Ninomiya, Y., and Olsen, B. R. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 3014–3018
16. Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T., and Olsen, B. R. (1986) *J. Biol. Chem.* **261,** 5041–5050
17. Kramer, J. M., Cox, G. N., and Hirsh, D. (1982) *Cell* **30,** 599–606
18. Venkatesan, M., De Pablo, F., Vogeli, G., and Simpson, R. T. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3351–3355
19. Monson, J. M., Natzle, J., Friedman, J., and McCarthy, B. J. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 1761–1765
20. Natzle, J. E., Monson, J. M., and McCarthy, B. J. (1983) *Nature* **296,** 368–371
21. Blumberg, B., Fessler, L. I., Kurkinen, M., and Fessler, J. H. (1986) *J. Cell Biol.* **103,** 1711–1719
22. Cherbas, P., Cherbas, L., and Williams, C. M. (1977) *Science* **197,** 275–277
23. Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25,** 263–269
24. Huynh, T. V., Young, R. A., and Davis, R. W. (1985) in *DNA Cloning: A Practical Approach* (Glover, D., ed) pp. 49–78, IRL Press, Oxford, United Kingdom
25. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Benton, W. D., and Davis, R. W. (1977) *Science* **196,** 180–182
27. Yanisch-Perron, C., Vierira, J., and Messing, J. (1985) *Gene (Amst.)* **33,** 103–119
28. Kurkinen, M., Bernard, M. P., Barlow, D. J., and Chow, L. T. (1985) *Nature* **317,** 177–179
29. Staden, R. (1980) *Nucleic Acids Res.* **8,** 3673–3694
30. Devereaux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
31. Dayhoff, M. O. (1976) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed) Vol. 5, Suppl. 2, pp. 4–6, National Biomedical Research Foundation, Washington, D.C.
32. Sweet, R. M., and Eisenberg, D. (1983) *J. Mol. Biol.* **171,** 479–488
33. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
34. Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263,** 211–214
35. Oberbaumer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, T., Siebold, B., Glanville, R. W., and Kuhn, K. (1985) *Eur. J. Biochem.* **147,** 217–224
36. Pihlajaniemi, T., Tryggvason, K., Myers, J. C., Kurkinen, M., Lebo, R., Cheung, M., Prockop, D. J., and Boyd, C. D. (1985) *J. Biol. Chem.* **260,** 7681–7687
37. Brinker, J. M., Gudas, L. J., Loidl, H. R., Wang, S., Rosenbloom, J., Kefalides, N. A., and Myers, J. C. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 3649–3653
38. Le Parco, Y., Cecchini, J., Knibiehler, B., and Mirre, C. (1986) *Biol. Cell* **56,** 217–226
39. Bairati, A., and Garrone, R. (eds) (1985) in *Biology of Invertebrate and Lower Vertebrate Collagens,* Plenum Publishing Corp., New York
40. Noelken, M. E., Wisdom, B. J., Dean, D. C., Hung, C.-H., and Hudson, B. G. (1986) *J. Biol. Chem.* **261,** 4706–4714