Amino Acid Sequence of Rat α-Lactalbumin: A Unique α-Lactalbumin†

Rajani V. Prasad, Ralph J. Butkowski, James W. Hamilton, and Kurt E. Ebner*

ABSTRACT: The amino acid sequence of rat α-lactalbumin has been determined. Unlike other α-lactalbumins which contain 122 or 123 amino acids, rat α-lactalbumin is unique in that it contains 140 amino acids. The extra amino acids are a 17 amino acid extension at the carboxyl terminus. The amino acid sequence of this extension is Gly[124]-Ala-Pro-Ala-Leu-Val-Val[130]-Pro-Ala-Leu-Asp-Gly[135]-Glu-Thr-Pro-Val-Pro[140]. The extension is proline rich, which may contribute to the anomalous structural properties of rat α-lactalbumin. The amino acid sequence from residues 1 to 123 is similar to that of other α-lactalbumins. One possible explanation for the 17 amino acid extension is a mutation at the termination codon.

α-Lactalbumin is found in the milk of all species which synthesize lactose. The biological role of α-lactalbumin was described by Ebner et al. (1966), who showed that it was required for significant rates of lactose synthesis.

Unlike α-lactalbumins from most species, α-lactalbumin from rat milk contains 13.4% w/w carbohydrate and is unique in that it exists as multiple forms, each of which contains carbohydrate and is active in the lactose synthesis reaction (Brown et al., 1977). The molecular basis for the existence of the three forms of rat α-lactalbumin is dependent upon the carbohydrate units (Prasad et al., 1979).

The primary structures of various α-lactalbumins, including bovine (Brew et al., 1970), human (Findlay & Brew, 1972), guinea pig (Brew, 1972), goat (McGillivray et al., 1979), rabbit (Hopp & Woods, 1979), and the partial sequence of the kangaroo (Brew et al., 1978), have been determined, and these proteins show a high degree of homology and similarity of physical properties.

The properties of rat α-lactalbumin with respect to molecular weight, behavior on polyacrylamide gels, and circular dichroism spectra were quite different from those of other α-lactalbumins, suggesting structural differences. At the time the primary structural studies on rat α-lactalbumin were started, it was the only α-lactalbumin which was known to exist as a glycoprotein. Recently, rabbit α-lactalbumin was shown to be a glycoprotein (Hopp & Woods, 1979), but it had properties similar to other α-lactalbumins. Studies directed to understanding the control and expression of genes during differentiation of the mammary gland are performed often in the rat system (Matusik & Rosen, 1978; Richards et al., 1981a,b), because α-lactalbumin is a protein synthesized only in the mammary gland. In order to study further the structure and control of the rat α-lactalbumin genome, it is important to know the primary structure of the protein.

This study presents the complete amino acid sequence of rat α-lactalbumin (see paragraph at end of paper regarding supplementary material), and it shows that rat α-lactalbumin is indeed unique in that it contains 140 amino acids, in contrast to other α-lactalbumins, which contain 122 or 123 amino acids. The additional 17 amino acids represent an extension at the carboxyl terminus, which is characterized by being proline rich.

Results

Rat α-lactalbumin was purified as described by Brown et al. (1977). The flow diagram (Figure 1) describes the source, cleavage, and mode of purification of the peptides used to determine the complete sequence. By use of tryptic peptides and aminoethylated rat α-lactalbumin, every residue from positions 1 to 58, except 27, 33, 34, and 41, was identified previously (Prasad et al., 1979). Assignments at these positions were accomplished by sequence analysis of cyanogen bromide peptides.

Cyanogen Bromide Peptides. Rat α-lactalbumin contains two methionines, and cleavage with CNBr[1] yielded three peptides which were easily purified on a Bio-Gel P-10 column. The amino acid sequence of CB_{1-15} and CB_{16-110} confirmed the previous results (Prasad et al., 1979) and assigned threonine at the previously unidentified residues at positions 27 and 33.

The $CB_{111-140}$ peptide was useful in identifying the carboxyl terminus of the protein. The sequence of residues 111–122 is homologous to the sequence of other α-lactalbumins which terminate at residue 122 or 123. However, rat α-lactalbumin is the only protein sequenced to date which is apparently longer than 123 amino acids. The above fact is confirmed by the sequence of another peptide ($AP_{120-140}$) obtained by cleavage of the aminoethylated rat α-lactalbumin with arginine-specific protease.

Arginine Protease Peptides. In order to confirm the unique sequence at the carboxyl terminus and to provide overlap peptides, we used arginine protease to cleave aminoethylated rat α-lactalbumin. The sequence analysis of peptide $AP_{120-140}$ confirms the sequence of $CB_{111-140}$ which indicates that rat α-lactalbumin is longer than any other known α-lactalbumin. The sequence analysis of the peptide $AP_{106-119}$ indicates that the protease cleaved at residue 105 with low yields. Leucine was identified at that position, and hence the cleavage at leucine by the arginine protease is puzzling. Nevertheless, this unique cleavage provided an overlap into the $CB_{111-140}$ peptide. Sequence analysis of AP_{59-119} confirmed the sequence obtained by the staphylococcal protease peptides (discussed below) and provided overlaps in three regions, namely, residues 59–65, 67–71, and 81–88.

Cleavage with Staphylococcus aureus Protease. In order to obtain the sequence in the central region of the rat α-lactalbumin, we used CB_{16-110} as a starting material and cleaved

† From the Department of Biochemistry, University of Kansas Medical Center, Kansas City, Kansas 66103 (K.E.E., R.V.P., and R.J.B.), and the Calcium Research Laboratory, Veterans Administration Hospital, Kansas City, Missouri 64128 (J.W.H.). *Received September 21, 1981.* This research was supported by National Institutes of Health Grant AM 18257 and National Science Foundation Grant 20614.

[1] Abbreviations: PTH, phenylthiohydantoin; SAEC, *S*-(aminoethyl)cysteine; αLA, α-lactalbumin; CNBr, cyanogen bromide; CB, cyanogen bromide peptide; AP, arginine protease; SP, staphylococcal protease.

0006-2960/82/0421-1479$01.25/0

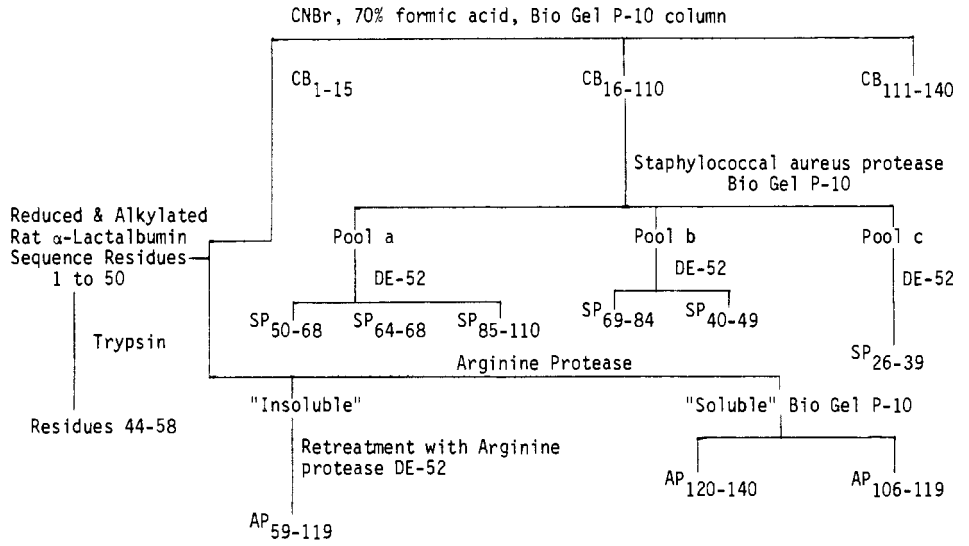

FIGURE 1: Flow diagram showing the isolation of the peptides used to determine the complete sequence of rat α-lactalbumin.

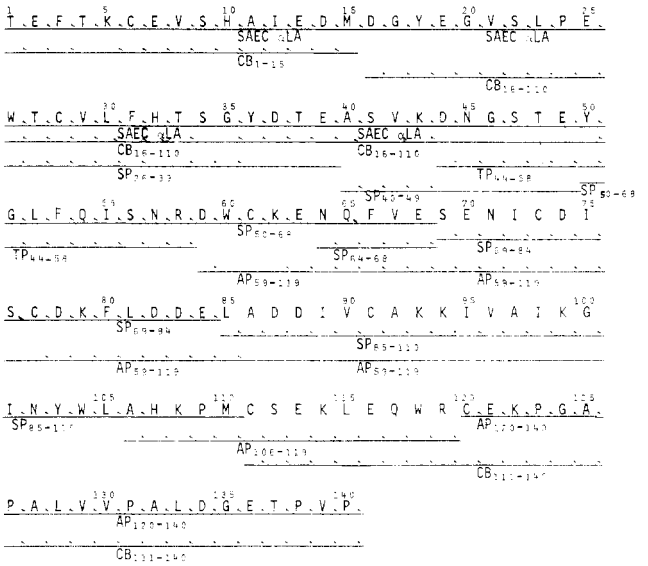

FIGURE 2: Proof of the entire sequence of rat α-lactalbumin. Identification of the amino acid (—); length of the peptide (—).

it at glutamic acid by using *S. aureus* protease. Six peptides were purified, and in addition to establishing the sequence within the region 50–108, they provided overlap segments and the identification of residues 34 and 41.

Carboxypeptidase Reactions. Both carboxypeptidases B and Y were separately used to digest aminoethylated rat α-lactalbumin. Carboxypeptidase B did not release any amino acids, which suggested that proline might be at the C terminus. No amino acids were released in the control digests lacking substrate. Carboxypeptidase Y showed a time-dependent release of 2.4 prolines/mol of protein and 0.8 valine/mol of protein, suggesting that 2 prolines are present at the carboxy terminus. However, the second proline is released only after the valine is released. Small amounts of threonine appear toward the end the digest, indicating that the sequence is Thr-Pro-Val-Pro_{COOH}. Proof of the complete sequence is shown in Figure 2.

Discussion

The complete amino acid sequence of rat α-lactalbumin, along with the peptides used to determine this sequence, is shown in Figure 2. Every amino acid has been directly identified, and the overlapping sequence is available for the entire molecule. Comparison of its amino acid sequence with that of other α-lactalbumins (Figure 3) shows that the rat α-lactalbumin is unique in that it is longer than any other

```
                1         5         10        15        20        25        30        35        40        45        50
R αLA           T E F T K C E V S H A I E D M D G Y E G V S L P E W T C V L F H T S G Y D T E A S V K D N G S T E Y
B αLA           E Q L T K C E V F R E L K D L K G Y G G V S L P E W V C T T F H T S G Y D T E A I V E N N Q S T D Y
GP αLA          K Q L T K C A L S H E L N D L A G Y R R I T L P E W L C I I F H I S G Y D T Q A I V K N S N H K E Y
H αLA           K Q F T K C E L S Q L L K D I D G Y G G I A L P E L I C T M F H T S G Y D T Q A I V E N N Q S T E Y
G αLA           E Q L T K C E V F Q K L K D L K D Y G G V S L P E W V C T A F H T S G Y N T Q A I V E N N Q S T D Y
Rb αLA          T Q L T R C E L T E K L K E L D G Y R D I S M S E W I C T L F H T S G L N T K I T V N N N G S T E Y

                          55        60        65        70        75        80        85        90        95        100
R αLA           G L F Q I S N R D W C K E N Q F V E S E N I C D I S C D K F L D D E L A D D I V C A K K I V A I K G
B αLA           G L F Q I N N K I W C K N D Q D P H S S N I C N I S C D K F L N N D L T N N I M C V K K I L D K V G
GP αLA          G L F Q I N N K D F C E S S T T V Q S R D I C D I S C D K L L N D N L T N N I M C V K K I L D I K G
H αLA           G L F Q I S N K L W C K S S Q V P Q S R N I C D I S C D K F L N D N I T N N I M C A K K I L D I K G
G αLA           G L F Q I N N K I W C K D D Q D P H S R N I C N I S C D K F L N N D L T D D I V C A K K I L D K V G
Rb αLA          G I F Q I N S K L W C V S K Q D P Q S K N I C D T P C E N F L D D N L T D D V K C A M K I L D K E G

                          105       110       115       120       125       130       135       140
R αLA           I N Y W L A H K P M C S E K L E Q W R C E K P G A P A L V V P A L D G E T P V P
B αLA           I N Y W L A H K A L C S E K L D Q W L L E K L
GP αLA          I N Y W L A H K P L C S D K L E Q W Y C E A Q
H αLA           I N Y W L A H K A L C T E K L E Q W L C E K L
G αLA           I N Y W L A H K A L C S E K L D Q W L C E K L
Rb αLA          I D H W L A H K P L C S E N L E Q W V C K K
```

FIGURE 3: Complete amino acid sequences of rat α-lactalbumin (R αLA), bovine α-lactalbumin (B αLA), guinea pig α-lactalbumin (GP αLA), human α-lactalbumin (H αLA), goat α-lactalbumin (G αLA), and rabbit α-lactalbumin (Rb αLA). The residues underlined are conserved in all the species.

α-lactalbumins. This extension of 17 amino acids is located at the C-terminal end and is confirmed by sequence analysis of two different peptides ($CB_{111-140}$ and $AP_{120-140}$). The results from the carboxypeptidase digests indicate the sequence Thr-Pro-Val-Pro_{COOH}, which is in excellent agreement with the direct sequence analysis of $CB_{111-140}$ and $AP_{120-140}$.

Most of the α-lactalbumins sequenced to date share a great deal of homology in amino acid sequence. Relative to bovine α-lactalbumin, goat and human α-lactalbumins have the highest degree of identity, namely, 90% and 77%, respectively, while rat and rabbit α-lactalbumins have only 70% and 63% of their residues identical with those of bovine α-lactalbumin. Rabbit α-lactalbumin has 122 amino acids, whereas the other α-lactalbumins have 123 residues. Hopp & Woods (1979) have suggested that rabbit α-lactalbumin has a deletion/insertion mutation at the C terminus or a point mutation involving introduction of a termination codon.

The amino acids underlined in Figure 3 are conserved in all the species, and some of these probably represent crucial structural features. For example, the half-cystine residues are retained to maintain the three-dimensional structure necessary for the activity (Brew et al., 1970). Chemical modification of histidine by ethoxyformate (Prieels et al., 1979), nitration of tyrosine (Denton & Ebner, 1971), or sulfenylation of tryptophan (Shechter et al., 1974) results in significant loss of activity. Interestingly, His-32 and most of the tyrosine and tryptophan residues are conserved in the sequences of all the α-lactalbumins. A recent attempt to determine the region of bovine α-lactalbumin involved in binding to galactosyltransferase has involved labeling studies using [^{14}C]acetic anhydride (Richardson & Brew, 1980). The results indicate that lysine-5 and lysine-114 were protected against modification when the reaction was performed on the complex of galactosyltransferase and α-lactalbumin. With the exception of rabbit α-lactalbumin, the two lysines are conserved in all the other species.

Rat α-lactalbumin exhibits anomalous behavior upon polyacrylamide gel electrophoresis and gel chromatography (Brown et al., 1977). The possibility that the carbohydrate unit at Asn-45 might confer such properties is unlikely since rabbit α-lactalbumin is also glycosylated with a similar carbohydrate unit in the same position and does not exhibit these anomalous properties (Hopp & Woods, 1979). However, rat α-lactalbumin has a 17 amino acid extension at the C terminus which is proline rich, since 4 out of 7 prolines in rat α-lactalbumin are present within this 17 amino acid segment. The carboxyl terminus might, therefore, alter the structure of the protein, resulting in the observed anomalous properties.

Several possibilities exist to explain the carboxyl-terminal extension in rat α-lactalbumin. The most likely explanation is a mutation of the termination codon. Residue 124 is glycine and a point mutation in the termination codon UGA to GGA would code for glycine. The protein would, therefore, continue to be synthesized until the next termination codon is reached. Such a mutation has been proposed for a variant of human hemoglobin, hemoglobin Constant Spring (HbCS) (Clegg et al., 1971). This variant is a minor one, comprising only 2% of the total hemoglobin, and is characterized by a 31 amino acid extension at the carboxyl terminus of the α chain. The production of small amounts of hemoglobin Constant Spring could be explained either by an unstable HbCS or by a mutation at the termination codon in the gene which normally directs only 1% of the total α chain synthesized. In contrast, the mutation in rat α-lactalbumin must have occurred in the gene which directs the synthesis of all the α-lactalbumin, since within the limits of detection, no "normal" rat α-lactalbumin containing 123 amino acids was observed.

The β subunit of human chorionic gonadotropin, when compared to the β subunits of human leutenizing hormone, human follicle stimulating hormone, and human thyroid stimulating hormone, has a 31 amino acid extension at the C terminus (Fiddes & Goodman, 1980). The likelihood of a mutation at the termination codon has been proposed.

Point mutations occur in bovine α-lactalbumin (Aschaffenburg, 1963; Bhattacharya et al., 1963) as well as β-lactoglobulin and the caseins. Bovine α-lactalbumin has two genetic variants, A and B, which differ in a substitution of Gln for Arg at position 10 in the sequence. The B variant is dominant in Western breeds, whereas the African breeds contain both variants.

Recently, double-stranded cDNA libraries from poly(A) RNA isolated from lactating guinea pig (Craig et al., 1981) and human (Hall et al., 1981) have been constructed. Studies using cDNA clones of rat α-lactalbumin would shed more light on the uniqueness of the 17 amino extension of the protein.

Acknowledgments

We thank James Rouse for operating the sequenator and the amino acid analyzer and Dr. Jon Barr for assisting in the amino acid analyses of rat α-lactalbumin.

Supplementary Material Available

Details regarding the sequence data for rat α-lactalbumin (25 pages). Ordering information is given on any current masthead page.

References

Aschaffenburg, R. (1963) *J. Dairy Res. 36*, 259–262.

Bhattacharya, S. D., Roychaudhury, A. K., Sinha, N. K., & Sen, A. (1963) *Nature (London) 197*, 797–799.

Brew, K. (1972) *Eur. J. Biochem. 27*, 341–352.

Brew, K., Castellino, F. J., Vanaman, T. C., & Hill, R. L. (1970) *J. Biol. Chem. 245*, 4570–4575.

Brew, K., Steinman, H. L., & Hill, R. L. (1978) *J. Biol. Chem. 248*, 4739–4744.

Brown, R. C., Rish, W. W., Hudson, B. G., & Ebner, K. E. (1977) *Biochim. Biophys. Acta 441*, 82–91.

Clegg, J. B., Weatherall, D. J., & Milner, P. F. (1971) *Nature (London) 234*, 337–340.

Craig, R. K., Hall, L., Parker, D., & Campbell, P. N. (1981) *Biochem. J.* (in press).

Denton, W. L., & Ebner, K. E. (1971) *J. Biol. Chem. 246*, 4053–4059.

Ebner, K. E., Denton, W. L., & Brodbeck, U. (1966) *Biochem. Biophys. Res. Commun. 24*, 232–236.

Fiddes, J. C., & Goodman, H. M. (1980) *Nature (London) 286*, 684–687.

Findlay, J. B. C., & Brew, K. (1972) *Eur. J. Biochem. 27*, 65–71.

Hall, L., Davies, M. S., & Craig, R. K. (1981) *Nucleic Acids Res. 9*, 65–84.

Hopp, P. T., & Woods, K. R. (1979) *Biochemistry 18*, 5182–5191.

Matusik, R., & Rosen, J. (1978) *J. Biol. Chem. 253*, 2343–2347.

McGillivray, R., Brew, K., & Barnes, K. (1979) *Arch. Biochem. Biophys. 197*, 404–414.

Prasad, R., Hudson, B. G., Butkowski, R., Hamilton, J. W., & Ebner, K. E. (1979) *J. Biol. Chem. 254*, 10607–10614.

Prieels, J. P., Bell, J. E., Schindler, M., Castellino, R. J., & Hill, R. L. (1979) *Biochemistry 18*, 1771–1776.

Richards, D. A., Rodgers, J. R., Supowit, S. C., & Rosen, J. M. (1981a) J. Biol. Chem. 256, 526–532.
Richards, D. A., Blackburn, D. E., & Rosen, J. M. (1981b) J. Biol. Chem. 256, 533–538.
Richardson, R. H., & Brew, K. (1980) J. Biol. Chem. 255, 3377–3385.
Shechter, Y., Patchornick, A., & Burnstein, Y. (1974) J. Biol. Chem. 249, 413–419.

Isolation and Partial Characterization of the Amino-Terminal Propeptide of Type II Procollagen from Chick Embryos†

Samantha Curran and Darwin J. Prockop*

ABSTRACT: The amino-terminal propeptide from type II procollagen was isolated from organ cultures of sternal cartilages from 17-day-old chick embryos. The procedure provided the first isolation of the propeptide in amounts adequate for chemical characterization. The propeptide had an apparent molecular weight of 18 000 as estimated by gel electrophoresis in sodium dodecyl sulfate. It contained a collagen-like domain as demonstrated by its amino acid composition, circular dichroism spectrum, and susceptibility to bacterial collagenase. One residue of hydroxylysine was present, the first time this amino acid has been detected in a propeptide. The peptide contained no methionine and only two residues of half-cystine. Antibodies were prepared to the propeptide and were used to establish its identity. The antibodies precipitated type II procollagen but did not precipitate type II procollagen from which the amino and carboxy propeptides were removed with pepsin. Also, they did not precipitate the carboxy propeptide of type II procollagen. The data demonstrated that the type II amino propeptide was similar to the amino propeptides of type I and type III procollagens in that it contained a collagen-like domain. It differed, however, in that it lacked a globular domain as large as the globular domain of 77–86 residues found at the amino-terminal ends of the proα1 chains of type I and type III procollagens.

Procollagens have been identified as precursors of three interstitial collagens, types I, II, and III. Several of the type-specific N and C propeptides[1] have been characterized. The primary structure has been determined for the N-terminal propeptide of the proα1(I) chain in calf (Hörlein et al., 1979), sheep (Rohde & Timpl, 1979), and chick embryos (Pesciotta et al., 1980) and for the N-terminal propeptide of the proα1(III) chain from calf [A. Brandt, D. Hörlein, P. Bruckner, R. Timpl, P. P. Fietzek, and R. W. Glanville, unpublished results; see Timpl & Glanville (1981)]. The N propeptides for both proα1(I) and proα1(III) chains have an N-terminal globular domain of 77–86 amino acid residues, followed by a collagen-like domain of about 40 amino acid residues [for review, see Timpl & Glanville (1981)]. The collagen-like domain is joined to the α chain by a short non-collagen sequence of 2–8 amino acids. The N propeptide of the proα2(I) chain has not been completely characterized, but in sheep (Becker et al., 1977) and chick embryos (Tuderman et al., 1978; Morris et al., 1979) it lacks the N-terminal globular domain and consists of about 60 amino acid residues that are collagen-like in structure. In rat, the N propeptide of the proα2(I) appears to be about the same size as the N propeptide of proα1(I) and proα1(III) and therefore may contain the globular domain (Smith et al., 1977). In the case of type II procollagen, the presence of an N-terminal propeptide has been established (Olsen et al., 1976; Merry et al., 1976; Uitto et al., 1977), and it has been shown to be about two-thirds the size of the N propeptides of proα1(I) or proα1(III) chains (Tuderman et al., 1978). Also, biosynthetic studies suggest that it contains mannose, which is not found in the N-terminal propeptides of either type I or type III procollagen (Guzman et al., 1978). We here report the first isolation of the N propeptide of type II procollagen in amounts adequate for chemical characterization.

Experimental Procedures

Materials

Dulbecco's modified Eagle's medium, Eagle's minimal essential medium, streptomycin, penicillin, trypsin, and Freund's complete and incomplete adjuvants were products of Gibco, Inc. The radioisotopes used were New England Nuclear's mixture of ^{14}C-labeled amino acids and New England Nuclear's [^{14}C]proline, 240 μCi/μmol. DEAE-cellulose (DE-52) and CM-cellulose (CM-52) were purchased from Whatman, Inc., and Bio-Gel P-2 (200–400 mesh) was purchased from Bio-Rad Laboratories. Advanced Biofactures Corp. provided the bacterial collagenase. Protein A–Sepharose 4B was purchased from Pharmacia Fine Chemicals, and the hemocyanin used was Calbiochem's A grade. Dr. Bjørn Olsen kindly provided the sheep anti-rabbit IgG (Olsen et al., 1977).

Methods

Purification of the Type II N Propeptide. Type II propeptides were isolated by a modification of the organ culture system developed by Olsen et al. (1977). Sterna were dissected from 40 dozen 17-day-old chick embryos and cleaned of adhering perichondrial tissue. The whole sterna were incubated in a Dulbecco's modified Eagle's medium supplemented with 40 μg/mL sodium ascorbate, 60 μg/mL β-aminopropionitrile, and 1–2 μCi/mL radioisotope. The incubations were carried

† From the Department of Biochemistry, University of Medicine and Dentistry of New Jersey—Rutgers Medical School, Piscataway, New Jersey 08854. *Received May 27, 1981.* This work was supported in part by National Institutes of Health Grant AM-16,516. A preliminary report was reported in abstract form (Curran et al., 1981).

[1] Abbreviations: N propeptide, amino-terminal propeptide of a procollagen; C propeptide, carboxy-terminal propeptide of a procollagen; CD, circular dichroism; $NaDodSO_4$, sodium dodecyl sulfate; DEAE, diethylaminoethyl; CM, carboxymethyl; IgG, immunoglobulin G; EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/82/0421-1482$01.25/0