Sung, S., and Laskowski, M., Sr. (1962), *J. Biol. Chem. 237,* 506-511.
Vogt, V. M. (1973), *Eur. J. Biochem. 33,* 192-200.
von Hippel, P. H., and Felsenfeld, G. (1964), *Biochemistry 3,* 27-39.
Wang, J. C. (1974), *J. Mol. Biol. 87,* 797-816.
Weigand, R. C., Godson, G. N., and Radding, C. M. (1975), *J. Biol. Chem. 250,* 8848-8855.
Weiss, B., Live, T. R., and Richardson, C. C. (1968), *J. Biol. Chem. 243,* 4530-4542.
Wells, R. D., Blakesley, R. W., Hardies, S. C., Horn, G. R., Larson, J. E., Selsing, E., Burd, J. F., Chan, H. W., Dodgson, J. B., Jensen, K. F., Nes, I. F., and Wartell, R. M. (1977), *CRC Crit. Rev. Biochem. 4,* 305-340.
Wingert, L., and von Hippel, P. H. (1968), *Biochim. Biophys. Acta 157,* 114-126.

Procollagen Complementary DNA, a Probe for Messenger RNA Purification and the Number of Type I Collagen Genes†

Anna Maria Frischauf, Hans Lehrach, Cynthia Rosner, and Helga Boedtker*

ABSTRACT: Type I procollagen mRNAs were separated from contaminating low-abundance messenger and nuclear RNAs by chromatography over Sepharose 4B in 0.65 M NaCl at room temperature. All of 27S rRNA and four-fifths of procollagen mRNAs bind to Sepharose under these conditions, while 18S rRNA and about three-fourths of other poly(A)-containing RNAs do not bind. AMV reverse transcriptase was used to prepare complementary DNA to procollagen mRNA at each purification step. Hybridization studies, in RNA excess, were carried out to establish the enrichment at each step both with respect to total RNA and to poly(A)-containing RNA. While "purified" procollagen mRNA preparations still consist of about 50% 27S rRNA, over 80% of cDNA prepared from it back hybridizes to its template at a log $cr_0t_{1/2}$ of -1.9. This type I procollagen cDNA hybridizes in DNA excess to DNA isolated from chicken erythrocytes and from embryonic chick calvaria at a log $c_0t_{1/2}$ of 3.1, demonstrating that procollagen cDNA is complementary to unique gene sequences in both tissues and that procollagen genes are not reiterated.

At least four different types of collagen are expressed in different tissues of the same organism or in the same tissue at different developmental stages (Miller and Matukas, 1974; Gross, 1973; Miller, 1976). The expression of these collagen genes is closely coupled to major differentiational events starting in gastrula (Green et al., 1968; Golob et al., 1974) and thus offers an extremely interesting system for studying the regulation of gene expression and its role in development in eucaryotes.

The isolation and characterization of procollagen mRNAs constitute an essential prerequisite to the study of control of the expression of these genes. We report here the extension of the initial purification of the two mRNAs coding for type I pro-α1 and pro-α2 collagen (Boedtker et al., 1976) resulting in an mRNA preparation which is essentially free of contaminating poly(A)-containing RNA species. AMV reverse transcriptase was used to synthesize complementary DNA to both purified procollagen mRNAs and to partly purified preparations to provide a quantitative determination of both the yields and the purification at each step.

cDNA prepared from purified procollagen mRNA was used to investigate the possibility of collagen gene amplification in both collagen and noncollagen producing tissues.

Materials and Methods

Isolation of Procollagen mRNAs. Type I procollagen mRNAs were isolated as described previously (Boedtker et al., 1976), except the 0.1 M KCl wash was omitted in the oligo(dT)-cellulose chromatography of total calvaria RNA, and the poly(A)-containing RNA was chromatographed over Sepharose 4B as described below.

Isolation of Embryonic Chick Calvaria DNA. Total calvaria RNA was separated from calvaria DNA following phenol extraction by pelleting through a cushion of 6.1 M CsCl, a modification of the procedure described by Glisin et al. (1974). The CsCl containing the DNA was pooled, dialyzed against 3 mM Na_2EDTA (pH 7.0) at 4 °C for 3 days, and precipitated with 3 volumes of absolute ethanol.

Isolation of Chick Erythrocyte DNA. Reticulocytes from an adult anemic chicken were frozen in liquid nitrogen. DNA was isolated according to the procedure of Sullivan et al. (1973). After alcohol precipitation, the DNA was dissolved in 0.1 M NaCl, 0.05 M Tris-HCl,[1] 10 mM Na_2EDTA (pH 7.0) and incubated overnight at 37 °C in RNase (100 μg/mL). The RNase was then removed by phenol extraction followed by dialysis against 3 mM Na_2EDTA (pH 7.0).

Sepharose 4B Chromatography. Sepharose 4B (Pharmacia, Sigma) was washed with diethyl pyrocarbonate (100 ppm), using 50 mL for a 10-mL column. After thoroughly rinsing with sterile deionized water, the column was equilibrated with 0.65 M NaCl, 0.02 M Tris-HCl (pH 7.5), 2.5 mM Na_2EDTA (pH 7.5), and 0.1% recrystallized Sarkosyl (ICN, K & K Laboratories, Inc., Plainview, N.Y.) at room temperature.

† From the Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138. *Received December 13, 1977.* This investigation was supported by National Institutes of Health Grant HD-01229 and by a grant from the Muscular Dystrophy Association, Inc.

[1] Abbreviations used are: Tris-HCl, 2-amino-2-hydroxymethyl-1,3-propanediol hydrochloride; Na_2EDTA, disodium (ethylenedinitrilo)-tetraacetic acid; DTT, dithiothreitol; $NaDodSO_4$, sodium dodecyl sulfate.

0006-2960/78/0417-3243$01.00/0 © 1978 American Chemical Society

Poly(A) (500 μg) containing RNA was dissolved in 500 μL of sterile deionized water, heated at 65 °C for 30 s, fast cooled to room temperature, an equal volume of 2× binding buffer was added, and the sample was applied to the column. One-milliliter fractions were collected at a rate of 0.5 column volume/h until the absorbance fell to one-half or less of its maximum value. The column was then eluted with 0.1 M NaCl, 0.02 M Tris-HCl (pH 7.5), 2.5 mM Na_2EDTA (pH 7.5), 0.1% Sarkosyl. Both the bound and unbound fractions were pooled and precipitated with 2 volumes of absolute ethanol.

Assay for in Vitro Synthesis of Procollagen. Wheat germ cell-free extracts were prepared from wheat germ obtained from the Pillsbury Co. according to the procedure of Roberts and Paterson (1973) with the minor modifications described previously (Boedtker et al., 1976), except that no preincubations were carried out, 1,4-dithiothreitol was used instead of mercaptoethanol, and the wheat germ extracts were predigested with *Micrococcal* nuclease, as described by Pelham and Jackson (1976). Protein synthesis assays, collagenase digestion of the wheat germ product, and $NaDodSO_4$-polyacrylamide slab gel electrophoresis were carried out as described previously (Boedtker et al., 1976).

Size Fractionation on 85% Me_2SO, 5–20% Sucrose Gradients. To obtain better resolution of RNA sedimenting between 18 and 30 S, the Me_2SO–sucrose gradients used before (Boedtker et al., 1976) were modified as follows. Me_2SO (85%) was prepared by adding 15 volumes of 0.33 M Trizma base (Sigma) and 0.033 M EDTA (not neutralized) with 85 volumes of Me_2SO (Matheson, Coleman and Bell, spectroquality) and titrated to pH 7.6 with 1 M HCl. Sepharose bound poly(A)-containing RNAs (60–90 μg) were dissolved in 30 μL of 0.05 M Tris-HCl, 0.005 M Na_2EDTA (pH 7.6), heated at 60 °C for 30 s, fast cooled to room temperature, and combined with 170 μL of 100% Me_2SO containing 0.05 M Tris base, 0.005 M EDTA, and titrated at pH 7.6 with 1 M HCl. The sample was immediately layered on a preformed 5–20% sucrose gradient containing 85% Me_2SO, 0.05 M Tris-HCl, 0.005 M Na_2EDTA (pH 7.6). The latter was centrifuged for 25 h at 50 000 rpm in a Beckman SW 56 Ti rotor at 23 °C. Fractions were collected after puncturing the bottom of the tube with a sterile syringe needle.

Preparation of cDNA. AMV reverse transcriptase, kindly supplied by Dr. J. Beard, was used to synthesize cDNA to procollagen mRNAs at various stages of purification. The 10-μL reaction mixture contained 10 mM Tris-HCl (pH 8.3); 8 mM $MgCl_2$; 10 mM DTT; 100 mM KCl; 500 μM dATP, dGTP, and dTTP; 200 μM (50 μCi) d[^{3}H]CTP (ICN), 20 μg/mL oligo$(dT)_{12-18}$ (Collaborative Research Inc.), 100 μg/mL actinomycin D, 150 μg/mL RNA, and 250 units/mL reverse transcriptase. The sample was incubated for 1 h at 42 °C, after which 2 μL of 0.1 M Na_2EDTA, 2 μL of 10% $NaDodSO_4$, and 7 μL of 1 N NaOH were added, and the mixture was incubated at 37 °C overnight. After neutralization, the mixture was put over Sephadex G-150 in 20 mM NH_4HCO_3. The excluded fractions were pooled, lyophilized, and redissolved in 5 mM Na_2EDTA (pH 7.0).

Hybridizations were carried out as described by Crouse et al. (1976).

Gene Copy Determination. cDNA and total cellular DNA were mixed in the ratio to be used for renaturation experiments [2.4 mg of DNA ($\sim 3 \times 10^{-15}$ mol of unique sequences) and 6000 to 10 000 cpm of procollagen cDNA ($\sim 1.5 \times 10^{-15}$ mol of procollagen sequences)] and degraded to an average length of approximately 300 nucleotides by heating the mixture in 0.3 M NaOH in a sealed ampule to 100 °C for 40 min. The DNA was then neutralized, was ethanol precipitated, and the pellet was washed with 70% ethanol to remove residual salt and redissolved at a concentration of approximately 30 mg/mL. An aliquot was end labeled with [γ-^{32}P]ATP and polynucleotide kinase as described by Maxam and Gilbert (1977) and separated from unincorporated label and small material on a 5-mL Sephadex G-75 column equilibrated in 30 mM NaOH, 2 mM EDTA. The first two-thirds of the acid-precipitable labeled material was pooled, neutralized, brought to 0.4 M sodium acetate, and ethanol precipitated. Total DNA, containing ^{3}H-labeled cDNA, and ^{32}P-labeled DNA were mixed so the amounts of ^{3}H and ^{32}P radioactivity were approximately equal.

DNA was denatured by boiling for 10 min in water. After fast cooling, the solution was made 0.6 M NaCl, 0.02 M Tris-HCl (pH 7.0) and then incubated at 68 °C. At various intervals, 3-μL aliquots were diluted into 1 mL of 0.12 M sodium phosphate (pH 6.9) and then chromatographed over hydroxylapatite (Bio-Rad HTP), equilibrated with the same buffer at 68 °C. The fractions which eluted at 68 °C (single-stranded denatured DNA) and those which eluted at 99 °C (double-stranded denatured DNA) were pooled and counted in Aquasol.

Since it is our experience that some nonhybridizable labeled material is formed during the alkali treatment described above, we normalized the renaturation values to correct for the different final extents obtained. The extrapolated initial and final values used were determined by linear extrapolation of the extent vs. c_0t and the extent vs. $1/c_0t$ plots to zero and infinite time.

Results

Procollagen mRNA Purification Determined from cDNA Hybridization

Chromatography over Sepharose 4B. Procollagen mRNAs were isolated as described previously with minor modifications as detailed under Materials and Methods. The only major change was the use of Sepharose 4B chromatography in high salt. It has previously been shown that rat liver 28S rRNA is quantitatively retained on Sepharose 4B when applied in 0.5 M NaCl while 18S rRNA does not bind (Petrovic et al., 1971), probably because of the former's high guanine content (Petrovic et al., 1975). We confirm these results, but at slightly higher salt concentrations. Chick calvaria rRNA (400 μg) was chromatographed over Sepharose 4B, and aliquots of both the bound and unbound fractions were analyzed on polyacrylamide gels run in 99% formamide. A photograph of the resultant gels is shown in Figure 1. The bound fraction (left gel) contains only 27S rRNA and fragments of 27S, while the unbound RNA (right gel) contains only 18S rRNA.

Since every third amino acid in collagen is glycine, coded by GGX, it seemed likely that procollagen mRNAs would also bind to Sepharose 4B in high salt. To test this, poly(A)-containing calvaria RNA was chromatographed over Sepharose 4B. The Sepharose-bound and -unbound fractions were translated in a wheat germ S-30, and both the total mRNA activity and the collagenase sensitivity and size distribution of the product obtained with each were measured. Table I gives the total micrograms of RNA in each fraction, the messenger activity per microgram, the collagenase sensitivity of the proline-labeled product, and the total picomoles of proline in procollagen peptides. The fraction of RNA bound can vary from 0.45 to 0.55 depending on the amount of 27S rRNA in the poly(A)-containing RNA applied to the column. The presence of 27S rRNA in the Sepharose-bound fraction is reflected in the lower mRNA activity of this fraction, but the

TABLE I: Distribution of Type I Procollagen mRNA Activity after Chromatography Over Sepharose 4B.

RNA fraction	μg of RNA/fract	pmol of Pro/μg of RNA	% solubilized by collagenase	Total pmol of Pro in procoll/frac[a]
Sepharose bound	143	0.18	41	10.3
Sepharose unbound	117	0.26	8	2.5

[a] Total pmol of proline in procollagen per fraction equals μg of RNA/fraction × pmol of proline/μg RNA × fraction solubilized by collagenase.

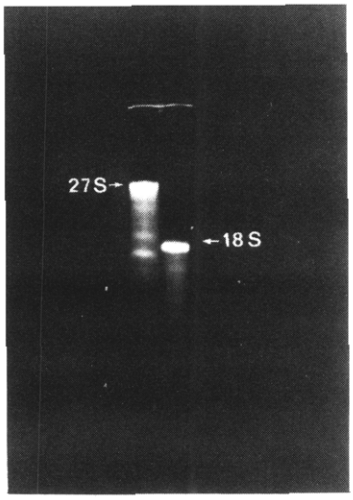

FIGURE 1: Fractionation of chick rRNA by Sepharose 4B chromatography. 3.1% polyacrylamide gel electrophoretic analysis of RNA bound to Sepharose (left gel) and RNA not bound to Sepharose (right gel). Electrophoresis carried out as described by Lehrach et al. (1977).

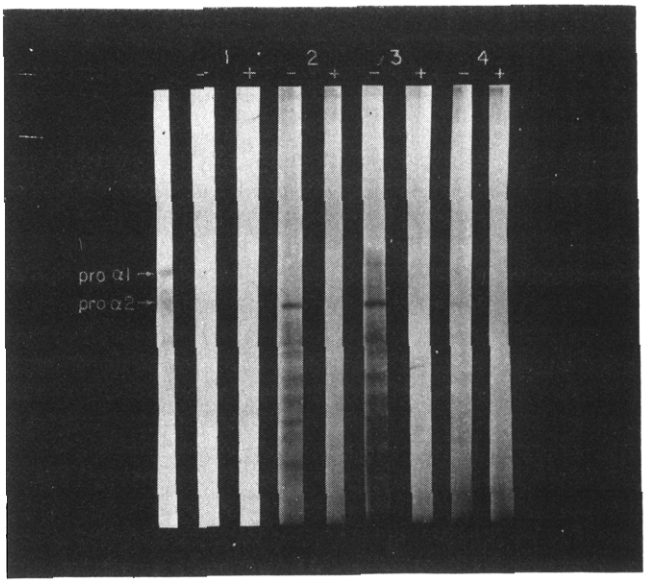

FIGURE 2: Fluorograph of wheat germ product obtained with calvaria RNA before and after Sepharose 4B chromatography. From left to right: Calvaria procollagen standards; minus RNA, minus collagenase; minus RNA, plus collagenase; 1 μg of oligo(dT)-cellulose bound RNA minus collagenase; 1 μg of oligo(dT)-cellulose bound RNA plus collagenase; 1 μg of Sepharose 4B bound RNA minus collagenase, 1 μg of Sepharose 4B bound RNA plus collagenase; 1 μg of Sepharose 4B unbound RNA minus collagenase, 1 μg of Sepharose 4B unbound RNA plus collagenase.

total mRNA activity (μg of RNA/fraction × pmol of proline/μg of RNA) is about the same for both fractions. Therefore roughly half of calvaria mRNA activity is bound to Sepharose and half is not. The collagenase sensitivity of the products obtained with the two fractions is very different however. Eighty percent of the collagenase-sensitive polypeptides synthesized with both fractions is made in response to the Sepharose-bound fraction. While this percent does vary from preparation to preparation, it has never been lower than 65%.

Definitive evidence for the location of most of the procollagen mRNA activity in the Sepharose-bound fraction is shown in the fluorogram of $NaDodSO_4$ gel analysis of the product obtained with both the Sepharose 4B bound and unbound fractions (Figure 2). Calvaria poly(A)-containing RNA, before Sepharose chromatography, is included for comparison. The Sepharose-bound fraction has clearly been enriched for procollagen mRNAs and stimulates the synthesis of high-molecular-weight polypeptides that comigrate with the procollagen standards as well as many smaller polypeptides resulting from incomplete translation of large mRNAs characteristic of wheat germ protein synthesis. All of the high-molecular-weight polypeptides and many of the smaller ones are degraded by collagenase. The Sepharose-unbound fraction stimulates the synthesis of some pro-α2 collagen as well as lower molecular weight polypeptides which are not solubilized by collagenase. Since almost twice as much radioactivity was applied to this gel, most of the labeled product must be in small polypeptides which run with the solvent front. We conclude that most, but not all, intact procollagen mRNA sequences bind to Sepharose 4B, while the majority of the heterogeneous low-abundance poly(A)-containing RNA does not bind.

To further purify procollagen mRNA, the Sepharose-bound RNA fraction was sedimented on Me_2SO-sucrose gradients and four fractions, >30, 27–30, <27–18, and <18 S were pooled, as shown in Figure 3. These fractions were translated in a wheat germ S-30, analyzed on gels, and a fluorogram of the gel is shown in Figure 4. The >30S fraction did not result in any significant stimulation, and the fluorograms of the gels were blank. The 27–30S gradient fraction stimulated the synthesis of proline-labeled polypeptides that comigrate with the calvaria procollagen standard as well as faster migrating polypeptides, all of which were solubilized by collagenase. This finding is not surprising, since 77–80% of the Cl_3AcOH-precipitable product produced by this fraction was solubilized by collagenase, while less than 30% of the product produced by the <27S–18S fraction and the <18S fraction were solubilized by collagenase. This is consistent with the fluorographs of the gel showing the products produced by the fractions smaller than 27S RNA. Some of the slower migrating polypeptides are resistant to collagenase digestion.

Properties of Procollagen cDNA. To provide the basis of making a quantitative estimate of the amount of procollagen mRNAs in the 27–30S fraction, it was transcribed into cDNA with AMV reverse transcriptase. Procollagen mRNA (1–2 μg) was used to prepare about 50 ng of cDNA transcript. While this appears to be a low yield, we will show later that it is actually a high yield considering the amount of 27S rRNA in our mRNA preparation and the length distribution of the transcripts. The size distribution of the [^{3}H]thymidine-labeled

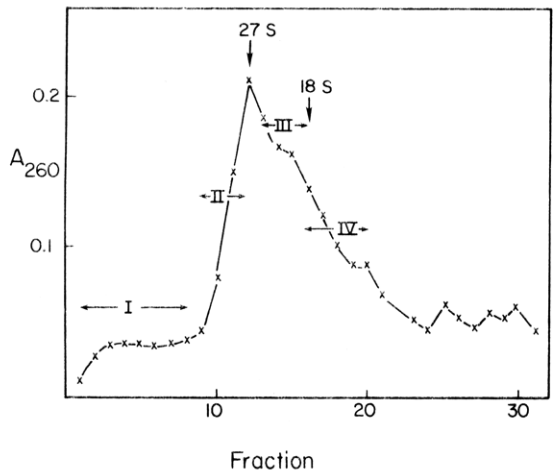

FIGURE 3: Absorbance profile of Sepharose 4B bound calvaria poly(A)-containing RNA fractionated on a Me_2SO–sucrose gradient. Details are described under Materials and Methods.

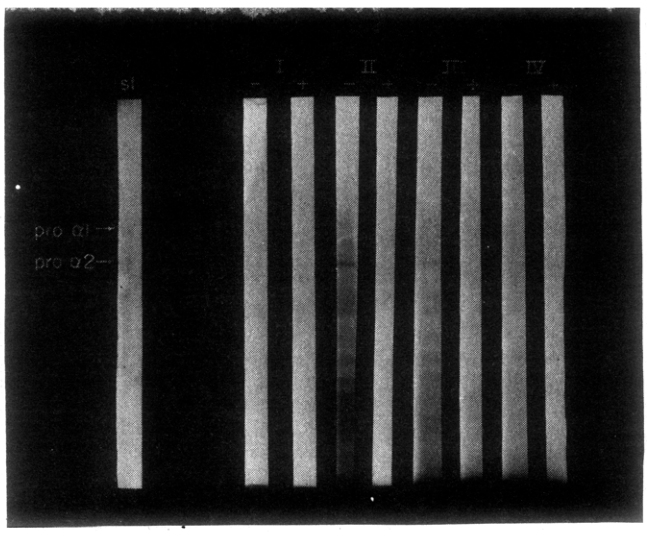

FIGURE 4: Fluorograph of wheat germ product obtained with Sepharose-bound RNA after Me_2SO–sucrose gradient fractionation. From left to right: Calvaria procollagen standard; >30 S minus collagenase, 0.25 μg, >30 S plus collagenase, 0.25 μg; 27–30 S minus collagenase, 0.9 μg, 27–30 S plus collagenase, 0.9 μg, <27–18 S minus collagenase, 1.0 μg, <27–18 S plus collagenase, 1.0 μg, <18 S minus collagenase, 0.8 μg, <18 S plus collagenase, 0.8 μg. The micrograms of each fraction added correspond to equal aliquots (~10%) of each fraction.

cDNA, analyzed on alkaline agarose gels (McDonell et al., 1977) using HpaI restriction fragments of T7 DNA as size markers, is shown in Figure 5. The linear relation between the square root of the molecular weight and log mobility (Lehrach et al., 1977) was used to extrapolate molecular weights of cDNA molecules shorter than the smallest restriction fragment. The size of the DNA ranges from 100 to 1000 nucleotides. Therefore, the longest transcripts correspond to at most one-fifth of procollagen mRNA (5000 NT).

To determine the fraction of cDNA that was double stranded, cDNA was self-hybridized in the absence of added RNA and then incubated with S-1 nuclease. Less than 5% of the radioactivity was resistant to S-1 nuclease. Therefore, 95% of the cDNA was not self complementary.

Fraction of Procollagen mRNAs in Total RNA at Each Step in mRNA Purification. The purity and yield of the two procollagen mRNAs at each step in the purification were determined by hybridization of RNA in RNA excess to procollagen cDNA. The results obtained are shown in Figure 6 and

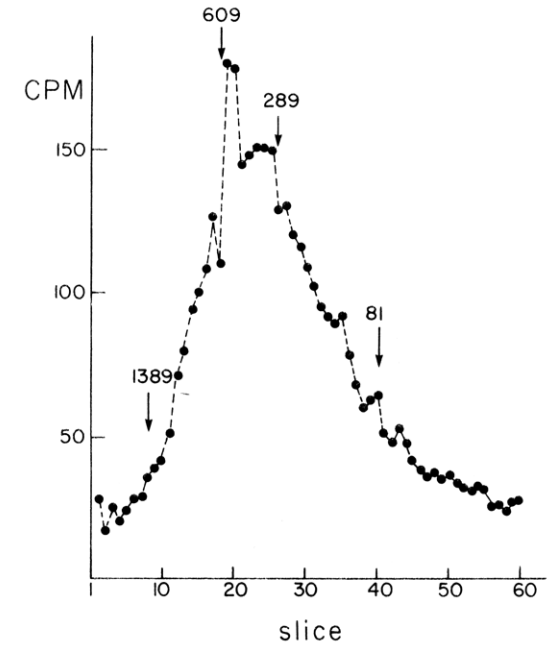

FIGURE 5: Alkaline–agarose gel electrophoretic analysis of procollagen cDNA. About 3 ng of [^{3}H]thymidine-labeled procollagen cDNA was electrophoresed on 2% agarose gels for 20 h at 0.5 mA/gel. The arrows locate *Hpa* I T7 DNA fragments.

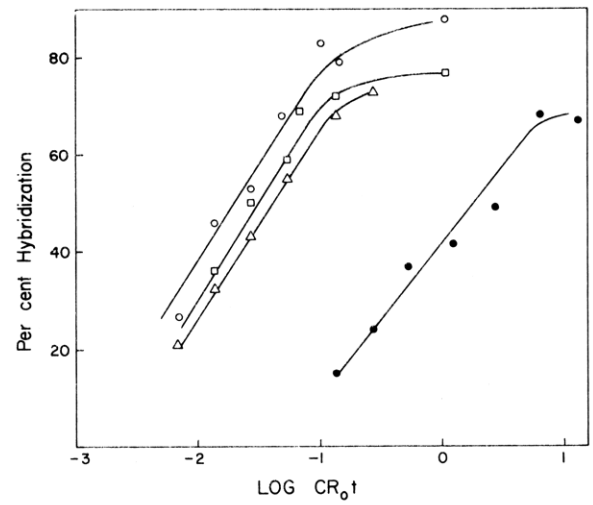

FIGURE 6: Hybridization of ^{3}H-labeled procollagen cDNA to procollagen mRNA at each step in purification. From right to left: (●) total RNA, (△) oligo(dT)-cellulose bound RNA, (□) Sepharose 4B bound RNA, and (○) 27–30S RNA from Me_2SO–sucrose gradient.

summarized in Table II. The amount of RNA in each fraction determined optically is given in column 1. Log $cr_0t_{1/2}$ measured from the data shown in Figure 6 is given in column 2. The percent procollagen mRNAs in each fraction given in column 3 was calculated, assuming that the two procollagen mRNAs, if 100% pure, would hybridize at a log $cr_0t_{1/2}$ of −2.24. From this and the total amount of RNA per fraction (column 1), we can calculate the actual micrograms of procollagen mRNAs in each fraction (column 4). These data clearly show that the major enrichment takes place during oligo(dT)-cellulose chromatography, as expected, since most of the rRNA is removed at this step. Rechromatography over oligo(dT)-cellulose would remove about half of the residual ribosomal RNA, but some ribosomal RNA remains even after further chromatography over this column (Longacre and Rutter, 1977). Chromatography over Sepharose 4B and size fraction on Me_2SO–

TABLE II: Purification of Type I Procollagen mRNAs Based on Hybridization of Procollagen cDNA to RNA Fractions.

RNA fraction	μg of RNA/fract	log $cr_0t_{1/2}$	% procoll mRNA[a]/fract	μg of procoll mRNA	fold purif
(1) total	34 600	−0.05	0.65	225	
(2) oligo(dT) bound	516	−1.64	25	130	39
(3) Sepharose bound	178	−1.76	33	59	51
(4) Sepharose unbound	147	−1.25	10	15	
Me_2SO-sucrose gradient on Sepharose-bound RNA					
(5) >30 S	10	−0.87	4.3	0.4	
(6) 27-30 S	30	−1.95	52	15.4	80
(7) <27-18 S	36	−1.72	30	10.9	
(8) <18 S	35	−1.88	44	15.3	

[a] Calculated assuming that 100% pure procollagen mRNAs would hybridize at a log $cr_0t_{1/2} = -2.24$. This value was calculated using the rate constant obtained for chick ovalbumin hybridization to cDNA (Hynes et al., 1977), correcting for the total complexity of the two procollagen mRNAs (10 100 nucleotides) being five times greater than that of ovalbumin mRNA (2020 nucleotides) and for the difference in salt concentration. In addition, we have assumed that the two procollagen mRNAs are present at identical concentrations. If the concentration of the pro-α1 mRNA were twice that of pro-α2 mRNA in keeping with the ratio of the α1 and α2 chains in type I collagen (Vuust, 1975), the values in column 3 should be multiplied by 0.62 and 0.31 to give the concentration of the higher and lower abundance species.

sucrose gradients provide an additional twofold enrichment.

Of the procollagen mRNA recovered after Sepharose 4B chromatography, 80% bound to Sepharose and 20% did not bind. This is in excellent agreement with the amount of collagen synthesis stimulated by the two fractions given in Table I. Since the recovery of both the RNA applied to the column and of procollagen mRNAs was only 60% and the percent procollagen mRNA per fraction only increased 8%, the efficiency of this chromatography could be questioned. As the protein synthesis results have already shown, and as the hybridization results we discuss below will confirm, the main function of this chromatography is to remove most of the contaminating nonprocollagen poly(A)-containing RNA.

The hybridization results in Table II also confirm other results obtained in translation studies: a significant fraction of procollagen mRNA sequences are found sedimenting slower than 27 S when fractionated on Me_2SO-sucrose gradients. A total of 26 μg is located at <27 S, while only 15.4 μg is in the 27–30S fraction. It is not likely that intact procollagen mRNAs are being degraded during centrifugation, since there is little degradation of 27S rRNA. While some procollagen mRNAs may be degraded in vivo, most of the degradation probably takes place during isolation of total RNA.

The finding that almost 52% of the RNA in the 27–30S Me_2SO-sucrose fraction is procollagen mRNA is somewhat surprising in light of the large amount of 27S rRNA still present in this fraction (Boedtker et al., 1976). Because of the uncertainty of accuracy of the $cr_0t_{1/2}$ of 100% pure procollagen mRNA, the absolute percent of procollagen RNA per fraction could be only half that reported here. The relative purification of 80-fold is however a reliable measure of the extent of purification achieved.

Fraction of Procollagen mRNA in Poly(A)-Containing RNA. While hybridization of procollagen cDNA to various mRNA fractions in RNA excess gives the absolute amount of procollagen mRNA in each fraction, this value mainly indicates the amount of rRNA present. To obtain the fraction of procollagen mRNA in poly(A)-containing RNA, we prepared cDNA to each RNA fraction and hybridized these to the most highly purified procollagen mRNA fraction. As shown in Figure 6, hybridization of RNA preparations containing different amounts of a specific mRNA to the same cDNA will lead to hybridization transitions of identical extent centered at different $cr_0t_{1/2}$ values. Hybridization of the same mRNA to cDNAs made from RNAs containing different ratios of this mRNA to other poly(A)-containing RNAs will lead to hybridizations of different extents at identical $cr_0t_{1/2}$ (assuming differences in the size distribution of the different cDNAs will have negligible affect on the extent of hybridization). Although in principle the same information could be obtained by back hybridization of each cDNA to its template, hybridization to purified mRNA is less ambiguous, since the latter is uniquely identified by its constant $cr_0t_{1/2}$ value. This is best illustrated by the data shown in Figure 7 and summarized in Table III. The percent of each cDNA which hybridized to procollagen mRNA increases from 22% for the cDNA made to the oligo(dT)-cellulose bound fraction to 84% for the cDNA made to the purest procollagen mRNA, an almost fourfold enrichment of procollagen mRNAs relative to other poly(A)-containing RNAs, the maximum possible being 4.5-fold. Thus, while Sepharose 4B chromatography and Me_2SO-sucrose gradient fractionation only result in a twofold purification of procollagen mRNAs relative to total RNA, the result is an almost fourfold purification relative to poly(A)-containing RNA.

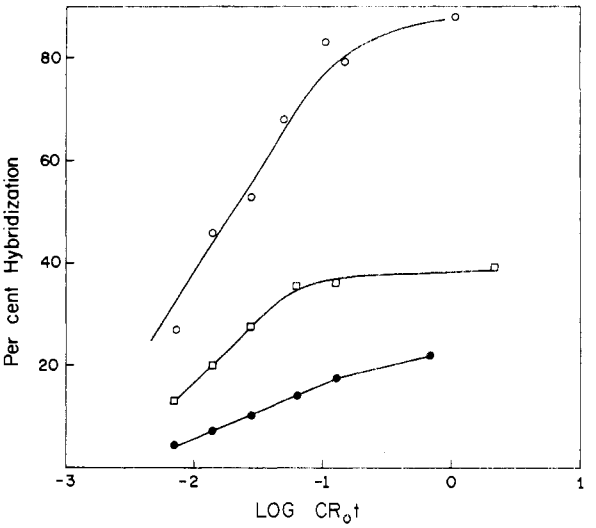

FIGURE 7: Hybridization of ^3H-labeled cDNAs prepared from calvaria poly(A)-containing RNA at each step during mRNA purification against procollagen mRNA in mRNA excess. Bottom curve: (●) cDNA to oligo(dT)-cellulose bound RNA; middle curve: (□) cDNA to Sepharose-bound RNA; upper curve: (○) cDNA to 27–30S RNA from Me_2SO-sucrose gradients.

The effectiveness of Sepharose 4B chromatography in separating procollagen mRNAs from other poly(A)-containing

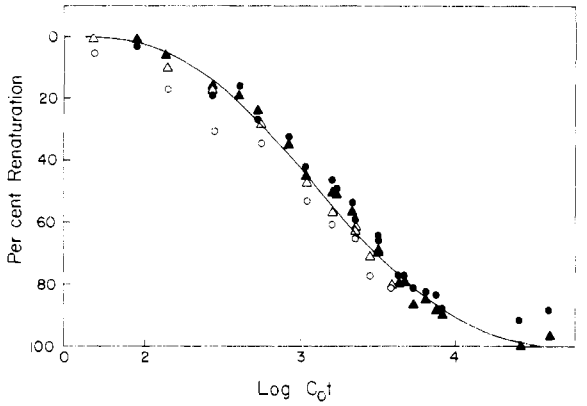

FIGURE 8: Renaturation of ^{3}H-labeled procollagen cDNA with ^{32}P labeled chick DNA from both collagen producing and noncollagen producing tissues: (○) ^{3}H-labeled cDNA renatured to ^{32}P-labeled embryonic chick calvaria DNA; (△) renatured embryonic chick calvaria DNA; (●) ^{3}H-labeled cDNA renatured to adult chicken reticulocyte DNA; (▲) renatured chick reticulocyte DNA. The data were normalized as described under Materials and Methods.

RNA can also be determined from the data in Table III. A total of 4.8 μg of cDNA was synthesized using the Sepharose 4B fraction as template, while 10.3 μg was synthesized with the Sepharose 4B unbound fraction. By subtracting the micrograms of procollagen cDNA in each fraction (obtained by multiplying the yield of cDNA by percent of cDNA hybridized to procollagen mRNA at low cr_0t), we can calculate that there were 2.9 μg of nonprocollagen cDNA in the bound fraction and 9.1 μg of nonprocollagen cDNA in the unbound fraction. Therefore, 76% of the nonprocollagen poly(A)-containing RNA is located in the Sepharose-unbound fraction. As a result, the subsequent Me_2SO–sucrose gradient fractionation results in a procollagen mRNA which is at least 84% pure relative to poly(A)-containing RNA.

Fraction of Procollagen mRNA Transcribed into cDNA. We can calculate the fraction of procollagen mRNA transcribed into cDNA by comparing the micrograms of procollagen cDNA synthesized with the micrograms of procollagen mRNA in the template. The micrograms of procollagen cDNA is obtained by multiplying the micrograms of RNA in the template (30 μg) by the total yield (4.9%) and then multiplying the product by the fraction of procollagen cDNA in the cDNA (84%). As a result, we find that 1.2 μg of procollagen cDNA was synthesized from 15.4 μg of procollagen mRNA (Table I, column 3), or 7.8% of procollagen mRNA was actually transcribed into procollagen cDNA. If each procollagen mRNA had served as a template but only 7.8% of it is transcribed, this would result in procollagen cDNAs about 400-nucleotides long. Since this value is not very much smaller than the average cDNA length of 500 nucleotides determined by gel electrophoresis, we can conclude that most of the procollagen mRNAs serve as a template in the reverse transcriptase reaction. The fact that the physical length of the cDNA synthesized is comparable to the percent of procollagen mRNA copied is good evidence against the possibility of the selective reverse transcription of a low-abundance contaminant into cDNA. Moreover, if only a small fraction of the poly(A)-containing RNA had been copied, this template would have to have been reused. This is contrary to everything known about the mechanism of action of the avian myeloblastosis viral reverse transcriptase enzyme. Finally, cDNA complementary to a small fraction of the poly(A)-containing RNA would have hybridized at a considerably higher $cr_0t_{1/2}$. There can be little doubt the poly(A)-containing RNA that stimulates the synthesis of procollagen polypeptides in wheat germ cell-free extracts and the poly(A)-containing RNA that serves as the major template for the reverse transcription are identical.

TABLE III: Purification of Type I Procollagen mRNA Based on the Percent of RNA Transcribed into Procollagen cDNA.

RNA fraction	μg of RNA/ fract	% RNA → cDNA	% cDNA hybrid. to procoll mRNA at low cr_0t
(1) oligo(dT)-cellulose bound	516	5.2	22
(2) Sepharose 4B bound	178	2.7	39
(3) Sepharose 4B unbound	147	7.0	11
Me_2SO–sucrose gradient on Sepharose 4B bound fraction			
(4) >30 S	10	5.4	62
(5) 27–30 S	30	4.9	84
(6) <27–18 S	36	3.6	65
(7) <18 S	35	3.2	35

Determination of the Number of Type I Collagen Genes per Genome

To determine the number of type I procollagen genes, ^{3}H-labeled procollagen cDNA was hybridized to an excess of both total adult chicken reticulocyte DNA and embryonic chick calvaria DNA both labeled at the 5′ end with ^{32}P. Since chromatography on hydroxylapatite columns was used to measure the extent of renaturation, both cDNA and total DNA were degraded to about 300 nucleotides (see Materials and Methods), since HAP chromatography is very sensitive to differences in the length of the driver and tracer DNA. The kinetics of renaturation of cDNA and total DNA shown in Figure 8 are essentially identical with a log $cr_0t_{1/2}$ of 3.1. The finding that type I procollagen cDNA hybridizes with the unique sequences of chicken DNA whether isolated from tissues which produce type I procollagen or from tissues which do not, signifies that there are single gene copies for procollagen α1 and α2 chains.

Discussion

The purification of type I procollagen mRNAs, determined by hybridization of procollagen cDNA, results in a mRNA that is over 84% pure in terms of poly(A)-containing RNA but only 50% pure in terms of total RNA. Since the major contaminant is 27S rRNA (Boedtker et al., 1976) which is neither translated into protein nor transcribed into cDNA (Verma and Baltimore, 1974), it does not interfere with either the identification of this mRNA as procollagen mRNA nor with the determination of the number of collagen genes using cDNA as the probe.

The extent of purification was determined at each step by two different hybridization procedures. The first and commonly used procedure is based on making cDNA to the purest mRNA fraction and hybridizing it, in RNA excess, to the RNA obtained at each step. In these hybridizations the extent of hybridization is always the same, but the rate of hybridization increases as the concentration of the mRNA per fraction increases. The second procedure involves making cDNA to the RNA obtained at each purification step and hybridizing it, again in RNA excess, to the purest mRNA fraction. In these hybridizations, the rate of hybridization or $cr_0t_{1/2}$ is always the same, but the extent of hybridization increases as the fraction of the specific mRNA in the poly(A)-containing RNA increases. In the case of large mRNAs, which are not com-

pletely transcribed into cDNA, the estimate of purity is only a minimum estimate (see below).

Using the first commonly used hybridization procedure, we have shown that we have achieved an 80-fold purification of procollagen mRNA relative to total RNA. The major purification occurs during oligo(dT)-cellulose chromatography in which the bulk of the ribosomal RNA is removed, resulting in a 39-fold purification with a loss of less than 40% of procollagen mRNA. While the remaining purification steps only double the overall purification with a loss of over 90% of the procollagen mRNA sequences present initially, a large fraction of the sequences lost were in degraded procollagen mRNAs, not in intact mRNA species. While our final yield of intact procollagen mRNA was only 7% of the amount of this RNA present initially in total RNA (corresponding to only 0.2% of that RNA), this relatively low yield is characteristic of results obtained in other eucaryotic mRNA isolations, resulting in highly purified mRNA preparations (Woo et al., 1975).

To obtain an accurate estimate of the purity of procollagen mRNA relative to other poly(A)-containing RNA at each purification step, the second hybridization procedure was carried out. Our finding that only 22% of the cDNA, made using oligo(dT)-cellulose bound RNA as template, hybridized to procollagen RNA appears surprising at first, since it is known that 62% of the protein synthesized in this embryonic tissue is collagen (Diegelman and Peterkofsky, 1972). The difference can be readily explained, however, by the size of our cDNA. Since the largest cDNA synthesized was 1000-nucleotides long, procollagen cDNA molecules correspond to less than a fifth of the RNA from which they were copied, while the cDNA copies of other poly(A)-containing RNAs correspond to a much larger fraction of their template. As a result, the estimate of the percent of procollagen mRNA in poly(A)-containing RNA at each purification step is only a minimum estimate. The 84% extent of hybridization at low $cr_0t_{1/2}$ we observed compares well with the 88% obtained by Longacre and Rutter (1977) for chick globin mRNA. Probably over 90% of the poly(A)-containing RNA in the final purification step is procollagen mRNA.

The relatively high purity achieved is the result of the separation of procollagen mRNAs from the bulk of other poly(A)-containing RNA by Sepharose 4B chromatography. If the Sepharose step is omitted, only 50% of the poly(A)-containing RNA in the 27–30S Me_2SO–sucrose gradient fraction is procollagen mRNAs. The separation of procollagen mRNAs from other RNAs is probably due to the high guanine content of procollagen mRNA. Under similar conditions, myosin heavy-chain mRNA does not bind to Sepharose (Morris et al., 1972). The separation, however, is extremely sensitive to the salt concentration. At slightly higher NaCl concentrations, other poly(A)-containing RNAs bind while procollagen mRNAs have been reported to elute from Sepharose with poly(A)-containing RNAs in 0.1 M NaCl, pH 5, 4 °C (Zeichner and Stern, 1977).

The finding that there is only one copy of each of the type I procollagen genes per genome in both chick reticulocyte and chick calvaria DNA is another example of the absence of gene amplification even in those differentiated tissues which are committed to the production of predominantly one or a few proteins.

Acknowledgments

We want to thank Tricia Bredbury for both her skill and patience in dissecting the many hundreds of chick embryos and preparing the total RNA without which these experiments could not have been carried out. We also want to thank Charles R. Swanson for developing the conditions for the effective separation of 27S and 18S rRNA by chromatography over Sepharose 4B. We are indebted to Dr. Richard Schwarz for his generous gift of purified collagenase and to Alan Hinnebusch for the *Hpa* II fragments of T7 phage DNA. Finally, we are most grateful to Dr. James Beard for the AMV reverse transcriptase and to the Pillsbury Co. for supplying us with fresh wheat germ. During part of this investigation, H. Lehrach was the recipient of a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research.

References

Boedtker, H., Frischauf, A. M., and Lehrach, H. (1976), *Biochemistry 15,* 4765–4770.

Crouse, G. F., Fodor, E. J. B., and Doty, P. (1976), *Proc. Natl. Acad. Sci., U.S.A. 73,* 1564–1567.

Diegelmann, R. F., and Peterkofsky, B. (1972), *Dev. Biol. 28,* 443–453.

Glisin, V., Crkvenjakov, R., and Byus, C. (1974), *Biochemistry 13,* 2633–2637.

Golob, R., Chetsanga, C. J., and Doty, P. (1974), *Biochim. Biophys. Acta 349,* 135–141.

Green, H., Goldberg, B., Schwartz, B., and Brown, D. D. (1968), *Dev. Biol. 18,* 391–400.

Gross, J. (1973), *Harvey Lect. 68,* 351.

Hynes, N. E., Groner, B., Sippel, A. E., Nguyen-Huu, M. C., and Schütz, G. (1977), *Cell 11,* 923–932.

Lehrach, H., Diamond, D., Wozney, J., and Boedtker, H. (1977), *Biochemistry 16,* 4743–4751.

Longacre, S. A., and Rutter, W. J. (1977), *J. Biol. Chem. 252,* 2742–2752.

Maxam, A. M., and Gilbert, W. (1977), *Proc. Natl. Acad. Sci., U.S.A. 74,* 560–564.

McDonnell, M. W., Simon, M. N., and Studier, F. W. (1977), *J. Mol. Biol. 110,* 119–146.

Miller, E. J. (1976), *Mol. Cell. Biochem. 13,* 165–192.

Miller, E. J., and Matukas, V. S. (1974), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 33,* 1197–1204.

Morris, G. E., Buzash, E. A., Rourke, A. W., Tepperman, K., Tompson, W. C., and Heywood, S. M. (1972), *Cold Spring Harbor Symp. Quant. Biol. 37,* 535–541.

Pelham, H. R. B., and Jackson, R. S. (1976), *Eur. J. Biochem. 67,* 247–256.

Petrovic, S. L., Novakovic, M. B., and Petrovic, J. S. (1975), *Biopolymers 14,* 1905–1913.

Petrovic, S., Novakovic, M., and Petrovic, J. S. (1971), *Biochim. Biophys. Acta. 254,* 492–495.

Roberts, B. E., and Paterson, B. M. (1973), *Proc. Natl. Acad. Sci., U.S.A. 70,* 2330–2334.

Sullivan, D., Palacios, R., Starnezer, J., Taylor, J. M., Faras, A. J., Kiely, M. L., Summers, N. M., Bishop, J. M., and Schimke, R. T. (1973), *J. Biol. Chem. 248,* 7530–7539.

Verma, I. M., and Baltimore, D. (1974), *Methods Enzymol. 29,* 125–130.

Von der Mark, H., Von der Mark, K., and Gay, S. (1976), *Dev. Biol. 48,* 237–249.

Vuust, J. (1975), *Eur. J. Biochem. 60,* 41–50.

Woo, S. L. C., Rosen, J. M., Liarakos, C. D., Choi, Y. C., Busch, H., Means, A. R., O'Malley, B. W., and Robberson, D. L. (1975), *J. Biol. Chem. 250,* 7027–7039.

Zeichner, M., and Stern, R. (1977), *Biochemistry 16,* 1378–1382.