Proc. Nat. Acad. Sci. USA
Vol. 71, No. 10, pp. 4208–4212, October 1974

The Identification of Collagen Messenger RNA

(chick embryo calvaria/formamide gel electrophoresis/cell-free protein synthesis)

HELGA BOEDTKER, RADOMIR B. CRKVENJAKOV, JEROLD A. LAST, AND PAUL DOTY

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Mass. 02138

Contributed by Paul Doty, July 15, 1974; revised September 23, 1974

ABSTRACT RNA isolated from calvaria of 16- to 18-day-old chick embryos, assayed in rabbit reticulocyte lysates, programs the synthesis of a collagenase-sensitive protein with the molecular weight of collagen pro-α-chains. When RNA labeled with [³H]uridine for 2 hr and chased for 1 or 2 hr was electrophoresed on aqueous polyacrylamide gels, most of the radioactivity not in 28S or 18S rRNA migrated with an apparent molecular weight of about 1,800,000. After oligo(dT)-cellulose chromotography and analysis in 99% formamide gels, this nonribosomal, rapidly labeled calvaria RNA species migrates at 28S–30S and thus has a molecular weight of at least 1,600,000.

Both the ability to program the synthesis of collagenase-sensitive protein in reticulocyte lysates and the presence of a single prominent rapidly labeled 30S peak in acrylamide gels strongly support the deduction that there is only one major mRNA species in calvaria and that this species is collagen messenger RNA.

Understanding the control of collagen biosynthesis is essential to understanding normal growth and development in vertebrates. Not only is collagen their major structural protein, but the onset of collagen synthesis is coincident with the onset of major differentiational events. An important step in the study of collagen biosynthesis must be the purification and characterization of the messenger RNAs which code for the five or more different collagen pro-α-chains.

Collagen messenger RNA (mRNA) has already been detected in polysomes prepared from the wings and legs of 8- or 9-day-old chick embryos by assaying for collagen synthesis in homologous cell-free extracts (1, 2) and in heterologous cell-free extracts prepared from Krebs II ascites cells (3). Collagen synthesizing polysomes have also been isolated from L-929 fibroblasts (4). However, collagen mRNA has not been successfully isolated from any of these polysomes.

More recently, Benveniste, Wilczek, and Stern (5) reported that total RNA isolated from the calvaria of 15-day-old chick embryos contained active collagen mRNA. They found that more than half of the [³H]proline-labeled protein made in ascites tumor extracts when calvaria RNA was added could be solubilized by collagenase. Also, some of this protein cochromatographed with carrier collagen on carboxymethyl cellulose. We have thus isolated RNA from calvaria. From a combination of pulse-chase studies of this RNA, under both aqueous and denaturing conditions, and assays of its ability to code for the synthesis of intact procollagen in rabbit reticulocyte lysates, we have been able to show that calvaria contain predominantly one class of mRNA, having a molecular weight of about 1,600,000–1,800,000. Several arguments indicate that this is collagen mRNA.

Abbreviations: NaDodSO$_4$, sodium dodecyl sulfate.

MATERIALS AND METHODS

Preparation and Fractionation of RNA from Calvaria. Calvaria RNA was prepared from 16- to 18-day-old chick embryos according to the procedure described by Benveniste *et al.* (5). The RNA was further purified by centrifugation through 5.7 M CsCl as described by Glisin *et al.* (6).

To prepare labeled RNA, calvaria from thirty 17-day-old chick embryos were incubated at 37° in modified Dulbecco's medium containing 0.5 or 1 mCi of [³H]uridine (0.04 μmol/ml). In pulse-chase experiments, the calvaria were further incubated in fresh medium containing 0.02 mM [¹H]uridine.

Calvaria RNA was fractionated (7) on oligo(dT)-cellulose. RNA (0.5 mg/ml) was applied to the column in 0.5 M KCl–0.01 M Tris·acetate, pH 7.5, and fractions were collected until no further RNA measured as A_{260} was eluted. The column was then eluted with 0.01 M Tris·acetate, pH 7.5, and the fractions containing material absorbing at 260 nm were pooled and precipitated with 2.5 volumes of 95% ethanol.

Calvaria RNA was fractionated on 5–20% sucrose gradients in 0.1 M NaCl–0.01 M Na acetate, pH 5, by centrifugation for 17 hr at 23,000 rpm at 4° in a Beckman SW 27.1 rotor.

Polyacrylamide Gel Electrophoresis of RNA. RNA preparations were routinely monitored by polyacrylamide gel electrophoresis (8, 9). RNA preparations were analyzed under denaturing conditions on 99% formamide gels using a modification (9) of the procedure described by Staynov *et al.* (10).

Assay for In Vitro Synthesis of Collagen. Rabbit reticulocyte lysates were prepared as described by McDowell *et al.* (11). The assay was performed using the conditions of Palmiter (12). After being incubated for 2 hr at 26°, the reaction mixture was incubated in 0.5 M Tris base for 30 min at 37° and then dialyzed against 0.01 M Tris·HCl, pH 7.5. Eighty-five micrograms of denatured calf skin collagen (Sigma) was added and the solution was made 16% (w/v) in ammonium sulfate. After 17 hr at 4°, the precipitate was centrifuged out and the pellet was dissolved in 1% sodium dodecyl sulfate (NaDodSO$_4$), 1% mercaptoethanol, 8 M urea and heated at 100° for 2 min. NaDodSO$_4$-polyacrylamide gel electrophoresis was carried out on 5% gels according to the procedure developed by Weber and Osborn (13).

To determine if the proteins synthesized in reticulocyte lysates were sensitive to collagenase, the ammonium sulfate precipitate was dissolved in 0.1 M Tris·HCl, pH 7.5, and dialyzed against 0.1 M NaCl–0.01 M Tes [*N*-tris(hydroxymethyl)methyl-2-aminomethane sulfonic acid], pH 7.5. After clarification by low-speed centrifugation, 1 ml of solution was made 1 mM in CaCl$_2$ and about 2 μg of a highly specific col-

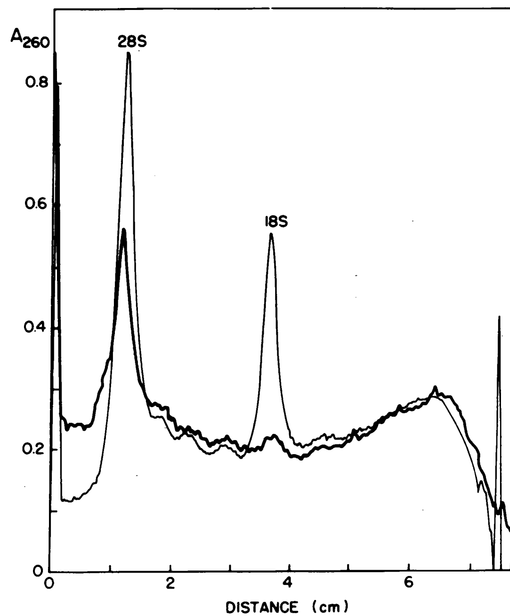

FIG. 1. Polyacrylamide gel electrophoresis of calvaria RNA. *Thin line*, unfractionated total RNA; *thick line*, RNA bound to oligo(dT)-cellulose.

lagenase from *Clostridium histolyticum* (14) was added. The solution was incubated at 37° for 30 min, precipitated with 5% Cl_3CCOOH, and then processed for electrophoresis as described above.

Assays in wheat germ S-30 extracts were performed as described by Roberts and Paterson (15).

RESULTS

Messenger activity of calvaria RNA

Calvaria RNA was applied to an oligo(dT)-cellulose column and both the bound fraction containing poly(A) sections linked to mRNA and that which did not bind were collected and analyzed on aqueous polyacrylamide gels as shown in Fig. 1. The fraction not bound is indistinguishable from unfractionated RNA. Both contain 28S and 18S rRNA, as well as three small UV-absorbing peaks of intermediate mobility that have previously been identified as fragments of 28S rRNA (9). The bound fraction, which contained 2.5% of the total RNA applied to the column, is predominantly 28S rRNA. However, a distinct slower migrating shoulder could be observed on the 28S peak in each of four preparations analyzed.

Total calvaria RNA, as well as fractions which bound and which were not bound to oligo(dT)-cellulose, were assayed for mRNA activity by measuring proline incorporation as a function of RNA added, in an *in vitro* protein synthesizing system prepared from wheat germ. As shown in Fig. 2, the bound fraction was about six times more active than total RNA and about 20 times more active than the unbound fraction. This differential activity demonstrates that oligo(dT)-cellulose-bound RNA has been considerably enriched for messenger RNA.

Characterization of protein synthesized by rabbit reticulocyte lysates

To determine the nature of the product programmed by calvaria mRNA, we used reticulocyte lysates to translate total

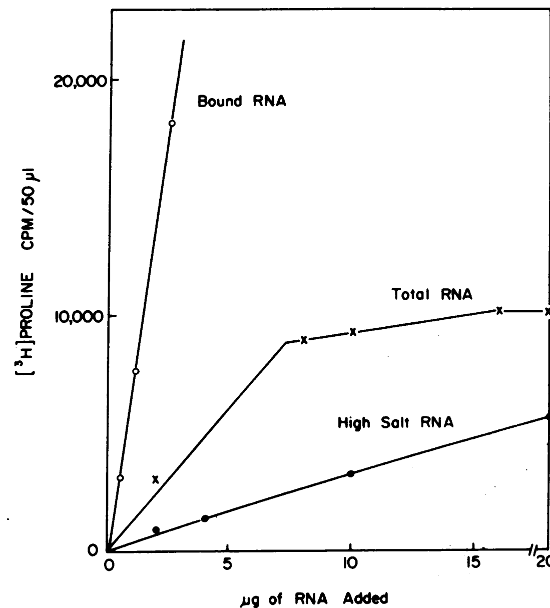

FIG. 2. Messenger activity of calvaria RNA assayed in wheat germ extract. ×, total RNA; ○, RNA bound to oligo(dT)-cellulose; ●, RNA not bound to oligo(dT)-cellulose.

RNA. These lysates rather than wheat germ extracts were used to characterize the product because we were able to obtain reproducible results with the lysates. After purification of the *in vitro* product as described in *Methods*, the product was analyzed on 5% polyacrylamide gels in $NaDodSO_4$; the result is shown in Fig. 3. Both the endogenous product and that made when calvaria RNA is added display a large globin peak and a second peak, of molecular weight about 65,000, which is probably not hemoglobin (16). When calvaria RNA is added, however, a distinctive third peak appears that has the same

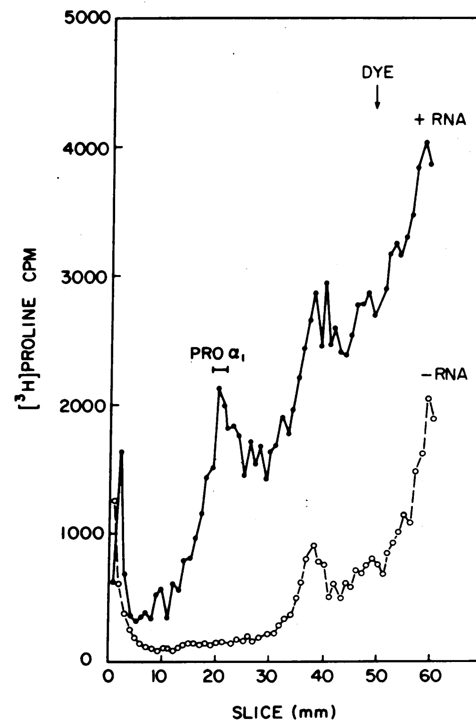

FIG. 3. $NaDodSO_4$-polyacrylamide gel electrophoresis of [^{3}H]proline-labeled proteins synthesized in reticulocyte lysates. ●, In the presence of 180 µg of calvaria RNA; ○, no RNA added.

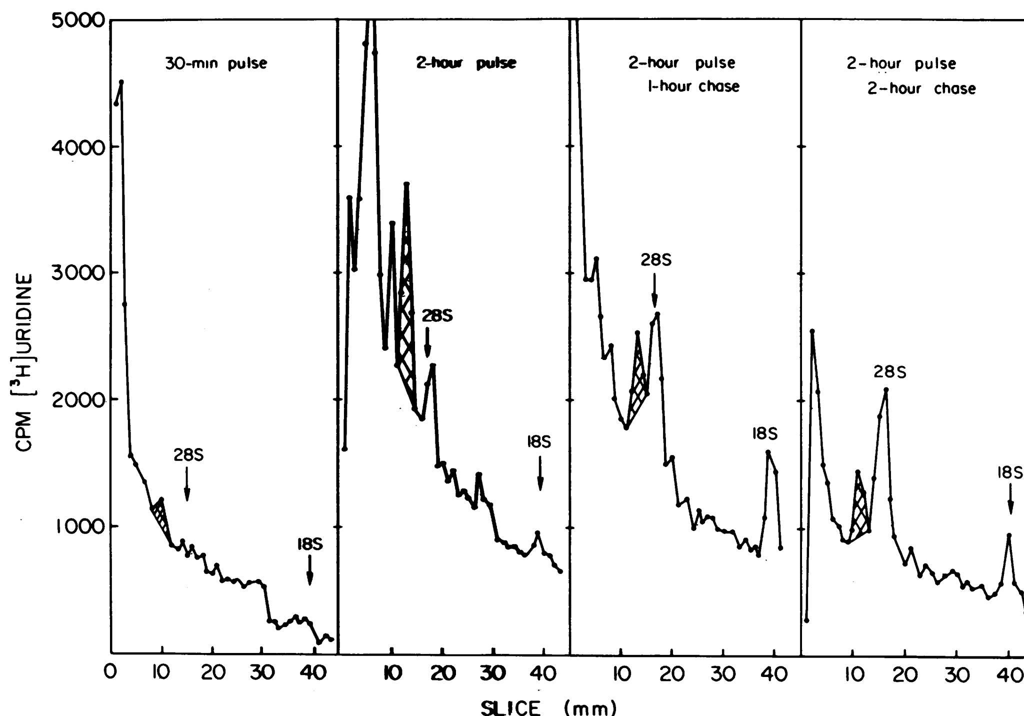

FIG. 4. Polyacrylamide gel electrophoresis of pulse- and pulse-chase-labeled calvaria RNA. From left to right: 30-min pulse; 2-hr pulse; 2-hr pulse, 1-hr chase; and 2-hr pulse, 2-hr chase. Crosshatched peaks are explained in *Results*.

mobility as procollagen. This peak contains about 8% of the counts applied to the gel; the remainder of the counts appear under the endogenous globin and the 65,000 dalton product peaks. The difference in the amount of globin seen with and without added calvaria RNA is a reflection of variability in the amount of globin which precipitates in 16% $(NH_4)_2SO_4$.

To see if the high-molecular-weight peak seen in Fig. 3 was indeed collagen, the *in vitro* product was treated with collagenase and electrophoresed. About 40% of the radioactivity electrophoresing in the molecular weight range of pro-α and α chains of collagen was found to be solubilized by collagenase, while neither globin nor the endogenous 65,000 dalton protein was affected by treatment with this enzyme.

To define the approximate size of the mRNA that programs this labeled collagen product in reticulocyte lysates, calvaria RNA was fractionated on sucrose gradients and five size fractions, >28S, 28S, <28S->18S, 18S, and <18S, were collected. In each case, the entire amount which was present in a given fraction was assayed in reticulocyte lysates, and the resultant product analyzed on $NaDodSO_4$ gels. The amount of [^{3}H]proline-labeled protein that comigrates with collagen pro-α chains was found to be 2700 cpm for the >28S fraction, 3600 cpm for 28S fraction, 1180 cpm for the <28S->18S fraction, and none for the 18S and the <18S fractions. The finding that collagen mRNA activity was most pronounced in fractions that sedimented at 28 S and greater than 28 S was completely consistent with our finding that the major ultraviolet-absorbing band found on gels after calvaria RNA is bound to oligo-(dT)-cellulose coelectrophoreses with 28S rRNA and contains a slower migrating shoulder. Indeed, each of the size fractions that stimulated the synthesis of procollagen *in vitro* contained appreciable amounts of the 28S component.

Pulse-chase labeling of calvaria RNA with [^{3}H]uridine

Since calvaria RNA can program the synthesis of collagen pro-α chains in reticulocyte lysates, it was obviously important to find out if rapidly labeled mRNA in calvaria was concentrated near 28 S as Fig. 1 and the location of mRNA activity on sucrose gradients suggest. To this end, calvaria were incubated in medium containing 0.5 or 1 mCi of [^{3}H]uridine for either 30 min or 2 hr. Three sets of calvaria, each incubated for 2 hr, were then chased with cold uridine for 0, 1, or 2 hr. The RNA isolated from these four labeled preparations was analyzed on aqueous polyacrylamide gels. The results are shown in Fig. 4. In contrast to the results obtained in Fig. 1, with higher concentrations of RNA, we find a marked occurrence of counts near the top of the gel that gradually tail into the gel with various peaks displayed above this continuously diminishing background. The first peak in each case occurs in the first 10 slices and was first presumed to be heterogeneous nuclear RNA, based on its appearance after only 30 min of labeling. While no well-defined peaks are seen after 30 min, a number of peaks and a heightened background appear after 2 hr of labeling. The most prominent of these peaks migrate more slowly than 28S rRNA. After a 1- or 2-hr chase, these slow-migrating peaks diminish in size, but some of the label at the top of the gel remains and the peak adjacent to the 28S rRNA peak appears to retain its identity and most of its magnitude. This peak is crosshatched in all four frames in Fig. 4. Its location at 30 S is consistent with a molecular weight of about 1,800,000, and it appears to be the same RNA species as the shoulder on the 28S peak shown in Fig. 1.

Although the location of this 30S peak was consistent with the location of the peak of mRNA activity in sucrose gradient fractions as measured by the appearance of a proline-labeled procollagen peak on $NaDodSO_4$ gels, the 30S RNA remained an unconvincing candidate for the putative collagen mRNA. The persistence of counts near the top of the gel, even after a 2-hr chase, made it likely that this polydisperse RNA distribution represents aggregates. Since collagen mRNA was expected to be the major component in calvaria mRNA, the posssibility remained that an RNA species other than 30S

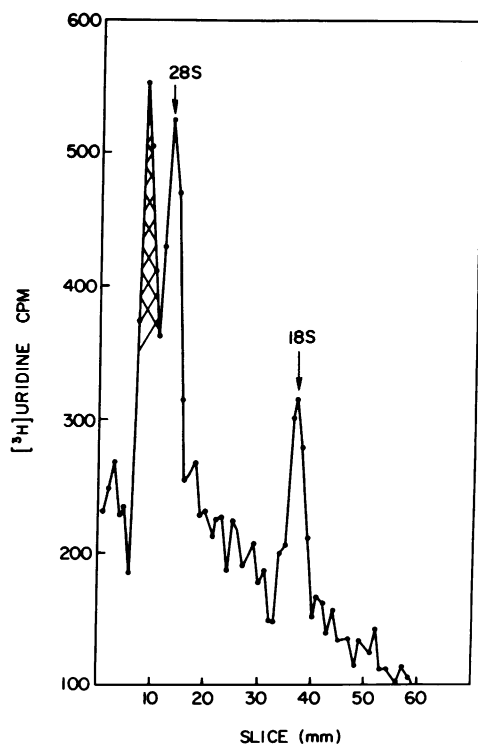

FIG. 5. Polyacrylamide gel electrophoresis of [^{3}H]uridine-labeled calvaria RNA on aqueous gels. RNA was incubated in 95% formamide for 2 min at 37° before it was applied to the gel.

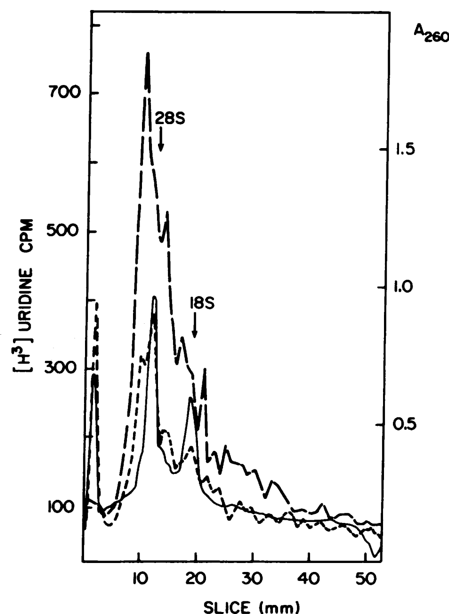

FIG. 6. Polyacrylamide gel electrophoresis of [^{3}H]uridine-labeled calvaria RNA after oligo(dT)-cellulose chromatography, on 99% formamide gels. — — —, cpm of RNA bound to oligo(dT)-cellulose; - - -, cpm and ——, A_{260} of RNA not bound to oligo(dT)-cellulose.

RNA was being masked by these aggregates at the top of the gel. Therefore, the pulse-chase RNA was reanalyzed on aqueous gels after first being denatured by heating at 37° for 2 min in 95% formamide. The distribution of radioactivity of RNA labeled for 2 hr and chased for 2 hr is shown in Fig. 5. The large amount of radioactivity previously found at the top of the gel has been significantly reduced, whereas the relative amount of label found in the 30S peak has more than doubled, while that in 28S and 18S rRNA has remained unchanged. Therefore, most of the labeled RNA near the top of the gel is not heterogeneous nuclear RNA but aggregates of the 30S species.

Having succeeded in disaggregating pulse-labeled RNA by formamide denaturation, we next analyzed this RNA on 99% formamide gels in order to obtain a molecular weight estimate of the 30S peak under conditions where RNA electrophoretic mobility is conformation-independent. Our initial results indicated that the only major pulse-labeled peak other than ribosomal RNA that migrated at 30S on aqueous gels was migrating at 25S on 99% formamide gels. However, after several different samples were examined, a different result became evident. Although we frequently found significant peaks of radioactivity with mobilities of RNA species having molecular weights of about 1,300,000 (25S) and 1,000,000 (22S), the major peak of radioactivity comigrated with 28S rRNA so that the ratio of the counts in 28 S to those in 18 S was much greater than two. If the 30S component seen on aqueous gels comigrates with 28S rRNA on formamide gels, their molecular weights must be the same, or 1,600,000.

In the expectation that the identity of the main pulse-labeled component could be more clearly resolved if the preparation were enriched for mRNA, RNA pulse-labeled for 2 hr and chased for 1 hr was chromatographed on oligo(dT)-cellulose and analyzed on 99% formamide gels. The results obtained for both the bound and nonbound fractions are shown in Fig. 6. Both fractions display a peak of radioactivity that migrates a little more slowly than 28S rRNA and apparently corresponds to the 30S peak found on aqueous gels, as well as the UV-absorbing shoulder found after oligo(dT)-cellulose chromatography (see Fig. 1). However, especially in the case of the bound fraction, the separation between the 30S and 28S components is so small that the possibility that there is only one component and that the main messenger peak comigrates with 28S rRNA cannot yet be ruled out. Hence, rather than assign a precise molecular weight to the 30S component seen clearly on aqueous gels but less well defined on formamide gels, we only conclude that it is at least as large as 28S rRNA, that is, at least 1,600,000 daltons. In addition to the main labeled peak, the oligo(dT)-cellulose-bound fraction is seen to contain two smaller labeled peaks mentioned above, which correspond to RNA species having molecular weights of about 1,300,000 and 1,000,000. Since there is little if any UV absorbance associated with these peaks, and since they are present to a much lesser extent in the RNA fraction that does not bind to oligo(dT)-cellulose, we identify these peaks as representing calvaria mRNA species other than intact collagen mRNA.

DISCUSSION

The isolation and identification of the eukaryotic mRNAs that have been described in the last few years have been achieved most readily when cells are available that synthesize predominantly one protein. The most outstanding example and thus far the best characterized is, of course, globin mRNA isolated from reticulocyte cells in which 90% of the protein made *in vivo* is globin.

More than 60% of the protein made in calvaria of 16-day-old chick embryos is reported to be collagen (17), and hence the collagen mRNA must be the major mRNA in this tissue. Total RNA prepared from calvaria of 16- to 18-day-old chick

embryos has been shown to program the *in vitro* synthesis of a proline-labeled polypeptide in reticulocyte lysates. This polypeptide is not seen in endogenous *in vitro* synthesis. It migrates with the mobility of calvaria procollagen on sodium dodecyl sulfate gels, and 40% of it is digested by collagenase under conditions where *none* of the smaller peaks programmed by endogenous mRNA are digested. It, therefore, seems reasonable to identify this peak as pro-α chains of collagen. Although the proportion of pro-α chains being made is indeed small relative to total protein synthesized, it is nevertheless comparable to the proportion of ovalbumin made in rabbit reticulocyte lysates (12).

Since pro-α chains of collagen contain about 1200 amino-acid residues (18–20), the coding sequence of the mRNA must consist of about 3600 nucleotides, corresponding to a molecular weight of 1,250,000. Any nontranslated regions would increase this expected molecular weight; for example, it would increase to 1,300,000 if 150 adenylic acid residues were added. With 1,300,000 as the minimum expected molecular weight of the mRNA, an electrophoretic mobility corresponding to at least 25 S would be expected. There have been recent reports of collagen mRNA sedimenting at 22 S (21) or 24 S (22) as well as both sedimenting and electrophoresing at 22 S (23). All of these values indicate molecular weights below the coding limit. However, we have found that those sucrose gradient size fractions that stimulated the most procollagen synthesis in rabbit reticulocyte lysates sedimented in the range of 28–31 S. When we fractionated a pulse-labeled calvaria RNA preparation in precisely the same fashion on a sucrose gradient and analyzed the size fractions on aqueous polyacrylamide gels, we found that each of the fractions (>28S, 28S, and <28S–>18S) that stimulated the *in vitro* synthesis of collagen pro-α chains contained the rapidly labeled 30S peak. Although this result indicates an unfortunately low level of resolution in this size range, it was important to find that the amount of collagen synthesis in each of the fractions was roughly proportional to the amount of the 30S component therein. Moreover, this 30S component binds to oligo(dT)-cellulose and migrates in 99% formamide at 28–30S. While the present data do not permit a precise molecular weight assignment, it is clear that the major nonribosomal, rapidly labeled RNA species found in calvaria is considerably larger than earlier suggestions indicated. Furthermore, it has also become evident that this large, 1,600,000 to 1,800,000 dalton RNA is exquisitely sensitive to specific scission, and in some preparations of calvaria RNA 1,300,000 and 1,000,000 fragments appear as prominent rapidly labeled peaks. Hence, it is not surprising that messenger activity as well as retention on oligo(dT)-cellulose has been reported for presumptive collagen mRNAs smaller than the molecular weight we report here.

If we compare rapidly labeled calvaria RNA with myosin mRNA, the only other very large eukaryotic mRNA that has been characterized, there are some interesting differences. Myosin mRNA has been assigned a molecular weight of 2,000,000 based on both aqueous and formamide gel electrophoresis, but sediments at 26 S in sucrose gradients (24, 25); sedimentation at 26 S implies a molecular weight of about 1,400,000, which is obviously too small to code for myosin. It seems likely that the anomolously low sedimentation constant is due to myosin mRNA having a more extended mean configuration than that of ribosomal RNAs. Collagen mRNA, however, has a lower molecular weight than myosin mRNA, based on its mobility on formamide gels, and yet both sediments and electrophoreses at 30 S, suggesting that the conformation of collagen mRNA is comparable to that of ribosomal RNAs.

Since we have established that rapidly labeled calvaria RNA contains only one major RNA species other than rRNA, with a mobility corresponding to 30 S on aqueous gels, and 28–30 S on formamide gels, and since we have been able to associate the ability to program the *in vitro* synthesis of what appears to be collagen pro-α chains in heterologous reticulocyte lysates with the presence of this peak in the larger sucrose size fractions, it is likely that the rapidly labeled peak with a molecular weight of at least 1,600,000 is indeed collagen mRNA. Obviously, final identification of this RNA as collagen mRNA awaits its isolation and purification as a single species able to direct the synthesis of collagen *in vitro*.

We are indebted to Richard Schwarz for providing us with *Clostridium histolyticum* collagenase of high specificity, to Tricia Bredbury for invaluable technical assistance, and to John Wozney for running formamide gel electrophoresis. This research was supported by NIH Grant HD-01229.

1. Lazarides, E. & Lukens, L. N. (1971) *Nature New Biol.* **232**, 37–40.
2. Kerwar, S. S., Kohn, L. D., Lapiere, C. M. & Weissbach, H. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2727–2731.
3. Crkvenjakov, R., Dewey, K. F., Schwarz, R. & Doty, H. B. (1973) *Fed. Proc.* **32**, 534 abstr.
4. Kerwar, S. S., Cardinale, G. J., Kohn, L. D., Spears, C. L. & Stassen, F. L. H. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1378–1382.
5. Benveniste, K., Wilczek, J. & Stern, R. (1973) *Nature* **246**, 303–305.
6. Glisin, V., Crkvenjakov, R. & Byus, C. (1974) *Biochemistry* **13**, 2633–2637.
7. Aviv, H. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1408–1412.
8. Loening, U. E. (1968) *J. Mol. Biol.* **38**, 355–365.
9. Boedtker, H., Crkvenjakov, R. B., Dewey, K. F. & Lanks, K. (1973) *Biochemistry* **12**, 4356–4360.
10. Staynov, D. Z., Pinder, J. C. & Gratzer, W. B. (1972) *Nature New Biol.* **235**, 108–110.
11. McDowell, M. J., Joklik, W. K., Villa-Komaroff, L. & Lodish, H. F. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2649–2653.
12. Palmiter, R. D. (1973) *J. Biol. Chem.* **248**, 2095–2106.
13. Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412.
14. Schwarz, R. (1974) Ph.D. Dissertation, Harvard University, Cambridge, Mass.
15. Roberts, B. E. & Paterson, B. M. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2330–2334.
16. Lodish, H. F. & Desalu, O. (1973) *J. Biol. Chem.* **248**, 3520–3527.
17. Diegelmann, R. F. & Peterkofsky, B. (1972) *Develop. Biol.* **28**, 443–453.
18. Jimenez, S. A., Dehm, P. & Prockop, D. J. (1971) *FEBS Lett.* **17**, 245–248.
19. Bornstein, P., von der Mark, K., Wyke, A. W., Ehrlich, M. P. & Monson, J. M. (1972) *J. Biol. Chem.* **247**, 2808–2813.
20. Harwood, R., Grant, M. E. & Jackson, D. S. (1973) *Biochem. Biophys. Res. Commun.* **55**, 1188–1196.
21. Ziechner, M. & Rojkind, M. (1973) *Fed. Proc.* **32**, 1305 abstr.
22. Benveniste, K., Wilczek, J. & Stern, R. (1974) *Fed. Proc.* **33**, 1794, abstr.
23. Harwood, R., Connolly, A. D., Grant, M. E. & Jackson, D. S. (1974) *FEBS Lett.* **41**, 85–88.
24. Morris, G. E., Buzash, E. A., Rourke, A. W., Tepperman, K., Thompson, W. C. & Heywood, S. M. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **37**, 535–541.
25. Mondal, H., Sutton, A., Chen, V. & Sarkar, S. (1974) *Biochem. Biophys. Res. Commun.* **56**, 988–996.