The location of the globin mRNA sequence within its 16S precursor

Kate Smith, Paul Rosteck, Jr., and Jerry B. Lingrel

Department of Biological Chemistry, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA

Received 23 September 1977

ABSTRACT

The coding sequence of globin mRNA has been located at or very close to the 3' end of its poly(A)-containing 16S precursor. The 16S RNA was annealed to globin cDNA and the hybrid digested with ribonuclease H. The undigested fragment did not bind to oligo(dT)-cellulose and its size was that expected for the intact 5' portion of the precursor.

INTRODUCTION

Comparison of the general structural features of mammalian cytoplasmic mRNA and its putative precursor, heterogeneous nuclear RNA (HnRNA), has given rise to several models relating to the location of the mRNA within its precursor. The mRNA sequence may be derived from the 3' end, 5' end or from an internal position of a larger RNA. For example, both cytoplasmic mRNA and HnRNA have poly(A) at their 3' termini (1-3) and at least some of the poly(A) of HnRNA is conserved (4). Thus, mRNA could be derived from the 3' end of its precursor. In support of this model, mRNA sequences have been found adjacent to poly(A) in HnRNA of HeLa cells (5). Both HnRNA and mRNA also contain a 5' methylated structure, suggesting that at least some mRNAs could be derived from the 5' end (6-9). It has further been proposed that mRNA sequences may be derived from internal positions, as short oligo(A) regions are found internally in some HnRNA molecules (10). These oligo(A) sequences could function as markers for processing and as primers for adding poly(A).

Studies dealing with mixed populations of precursors and mRNAs have not provided the evidence needed to distinguish between these hypotheses. However, several laboratories (11-20) including our own (21) have reported the existence of RNA species which are larger than mature globin mRNA and contain the globin mRNA sequence. The kinetics of labeling during both continuous exposure to radioactive precursor and in pulse chase experiments suggest that these RNAs are precursors to mature globin mRNA (18,20).

The predominant species migrates at 15S-16S and is 2-3 times larger than mature globin mRNA. In erythroleukemia cells this RNA has been shown to contain only β -mRNA sequences (22). Thus, the identification of a specific mRNA precursor allows the location of one mRNA to be determined. Our approach has been to hybridize the poly(A)-containing 16S RNA with full-length globin cDNA, treat the hybrid with ribonuclease H which hydrolyzes the RNA annealed to cDNA (23) and examine the undigested material both with respect to its size and for the presence of poly(A). Our results suggest that the globin mRNA sequence is located at or very near the 3' end.

MATERIALS AND METHODS

(a) Isolation of RNA

Nucleated erythroid cells were obtained from the spleens of anemic mice (24) and incubated as previously described (21). The cells were labeled with ^3H -uridine and the total RNA isolated by the proteinase K procedure as reported earlier (21). RNAs containing globin mRNA sequences were purified by chromatography on cDNA cellulose (25). The RNA binding to the column following the first application was reapplied to insure that non-specific sequences were absent from the preparation. Poly(A)-containing RNA was isolated by binding to oligo(dT)-cellulose in 0.5 M NaCl (26).

(b) Sucrose Density Gradient Centrifugation

RNA samples for both analytical and preparative procedures were heated at 80° for 5 min in 0.1% SDS solution and then rapidly cooled on ice. They were layered directly on 10-30% (w/w) linear sucrose gradients made in 10 mM Tris pH 7.5, 0.2 mM EDTA, 0.1% SDS and centrifuged in an SW 40 rotor at 40,000 rpm for 17 hr at 17°. Fractions were collected of 0.3-0.4 ml.

(c) Formamide Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis in formamide was carried out according to Orkin *et al.* (27) with the following modifications. Four percent polyacrylamide gels containing 3.64% acrylamide and 0.36% N,N-methylene-bisacrylamide were polymerized in formamide containing 20 mM phosphate and 0.1 M acetic acid. The electrode buffer was 20 mM phosphate (pH 6.5) and was recirculated during electrophoresis. Samples were applied to the gels in 20-30 μl of buffered formamide containing 20% sucrose and 0.05% bromophenol blue after heating at 75° for 10 min. Electrophoresis was for 30 min at 1 mA/gel, then for 3 hr at 2.5 mA/gel at 20-25°. Gels were divided into 1 mm slices and radioactivity in each was measured after overnight incubation at 37° in 0.2 ml Soluene 100 (Packard Instruments, Inc.).

(d) RNA-DNA Hybridization

Complementary DNA was synthesized from purified globin mRNA as described by Wood and Lingrel (25) with the following modifications. The reaction mixture contained 20 units/ml of reverse transcriptase, 20 µg/ml of oligo(dT)₁₂₋₁₈, and 0.5 mM nucleotide triphosphates. Globin mRNA of high specific radioactivity (>100,000 cpm/µg) was prepared using ³H-uridine as described by Merkel *et al.* (28). Two sets of hybridization conditions were used. When comparing the hybridization kinetics of 9S and 16S RNA, incubations were performed in 0.12 M phosphate, 1 mM EDTA with 100 µg/ml of tRNA as carrier. Volumes ranged from 24 µl to 0.05 ml. Following denaturation at 95° for 5 min an excess of unlabeled cDNA (at least 10X) was annealed to 2,000 cpm of globin mRNA or 6,000 cpm of 16S RNA at 65° for 15 min. Reactions which were taken to a Dot (29) of 1 and 10 were carried out for 10 hr at 65°. The extent of duplex formation was determined by digestion with RNase A (20 µg/ml) and RNase T₁ (10 units/ml) for 1 hr at 25°. Trichloroacetic acid precipitable material was measured and expressed as percent of control incubations without ribonuclease digestion. Hybrids for RNase H digestion were formed for 2 hr at 65° in 0.2 M NaCl, 20 mM Tris-HCl, pH 7.5 with 200 µg/ml of tRNA. After 2 hours, 500 µg of poly(A) was added to the hybridization mixture in 0.2 M NaCl and the incubation continued at 65° for a further 60 min. Before addition of RNase H, aliquots were removed and the extent of duplex formation was tested as described above.

(e) Digestion of Hybrids with Ribonuclease H

DNA-RNA hybrids and controls were transferred from the 65° incubation to an ice bath. The incubation mixture was adjusted to 10 mM MgCl₂, and transfer RNA was added to a concentration of 250 µg/ml. The concentration of NaCl was thereby reduced from 0.2 M to 0.16 M. Ribonuclease H was added and the digestion allowed to proceed for 30 min at room temperature. The digestion was terminated by the addition of EDTA to 20 mM. Ten microliters of the ribonuclease H preparation, generously supplied by Dr. Stavrianopoulos, digested 3 µg of RNA in 15 min at 35° when diluted 1:1,000. For these experiments the original preparation was diluted 1:20 and enough enzyme was added to each digestion to give a final dilution of 1:200. The amount of RNA containing globin mRNA sequences was estimated for both controls and experimental incubations, to be no more than 0.03 µg per digestion reaction.

(f) Digestion of hybrids with Ribonuclease T₂

Hybrids between 5 µg of cDNA and 0.1 µg of ³H-poly(A) were formed in 200 µl of 0.2 M NaCl 0.02 M Tris-HCl pH 7.5 at 37° for 1½ hr. They were

then placed in a 65° water bath for 60 minutes with and without the addition of 500 µg of unlabelled poly(A). After cooling in an ice bath, ribonuclease T₂ was added to a concentration of 25 units/ml and digestion was carried out for 30 min at 37°. The amount of label remaining in hybrid form was determined by precipitability in 10% TCA. The digestion conditions were adequate to render 500 µg of poly(A) > 95% acid soluble in the absence of cDNA.

RESULTS

(a) Purification of the 16S Globin mRNA Precursor

When the total RNA isolated from nucleated erythroid cells labeled in the presence of ³H-uridine for 10 min is applied to a globin cDNA cellulose column, approximately 0.5% of the labeled RNA binds on the second application (21). This RNA is resolved into two major components when analyzed by polyacrylamide gels run in the presence of formamide (Figure 1). The faster moving component comigrates with α- and β-globin mRNAs. The slower moving RNA migrates with a molecular weight of 610,000 when compared to RNAs of known size (Figure 2).

The large RNA is also resolved by centrifugation in sucrose density gradients (Figure 3). It can be purified from such gradients by successive pooling of the RNA designated by the bars in Figure 3a and 3b. After two such recentrifugations the pooled material sediments as a single component at

Figure 1. Formamide-polyacrylamide gel electrophoresis of RNA which hybridized to cDNA cellulose. RNA was isolated from nucleated erythroid cells exposed for 10 min to ³H-uridine (1). The RNA was dissolved in 20 µl of buffered formamide and electrophoresis was performed as described in MATERIALS AND METHODS. The gels contained 6,000 cpm of globin cDNA bound RNA and 4S, 9S, 18S, and 28S RNAs from mouse reticulocytes as internal markers. The positions of the markers were determined by staining in 0.02% methylene blue for 1 hr followed by destaining in water. Radioactivity was determined as described above.

16S (Figure 3c). When gradient purified 16S RNA is applied to oligo(dT)-cellulose, approximately 50% is bound in 0.5 M salt at room temperature.

(b) The 16S RNA Contains One Globin mRNA Sequence

As the 16S RNA is isolated by virtue of its binding to globin cDNA cellulose, it should contain globin mRNA sequences. In order to prove this, solution hybridization was used to analyze and compare the hybridization

Figure 2. The relationship between the logarithm of the molecular weight and the electrophoretic migration of RNAs in denaturing polyacrylamide gels. The conditions of electrophoresis were as described in MATERIALS AND METHODS and legend of Figure 1. The arrow indicates the position of the 16S RNA.

Figure 3. Sucrose density gradient purification of 16S RNA. Packed nucleated erythroid cells (4 ml) were incubated for 10 min in medium containing 500 μ Ci/ml of 3 H-uridine and the isolation and centrifugation of the RNA containing globin mRNA sequences was carried out as described under MATERIALS AND METHODS. 1 μ l aliquots were sampled from each gradient fraction for counting. a) Total globin cDNA bound RNA; b) Recentrifugation of the material under the bar of a); c) Recentrifugation of the material under the bar of b).

▲-----▲ cpm

— A_{260} of 18S, 9S and 4S markers in a parallel gradient

kinetics of this RNA with those of authentic globin mRNA. Figure 4 shows the result of such an analysis. Globin cDNA hybridizes to purified globin mRNA with a $\text{Dot}_{1/2}$ of 0.8×10^{-3} . The reaction plateaus at about 90%, demonstrating that the cDNA preparation is an almost full-length copy. The 16S RNA hybridizes to globin cDNA with a $\text{Dot}_{1/2}$ of 1×10^{-3} and the plateau value is about 1/3 that of the globin mRNA. The close approximation of the $\text{Dot}_{1/2}$ values proves that the 16S RNA does indeed contain globin mRNA sequences rather than being complementary to a minor contaminating cDNA species. Furthermore, although the larger RNA is more than twice the size of mature globin mRNA, cDNA protects only 33% of it from RNase digestion. Therefore, no more than one copy of the globin mRNA is present in each 16S RNA molecule.

(c) Location of the Globin mRNA Sequence in the 16S Globin mRNA Precursor

The experimental design for locating the globin mRNA sequence within the 16S RNA precursor was to hybridize the poly(A)-containing precursor to globin cDNA, digest the RNA annealed to cDNA with ribonuclease H and determine the size and poly(A) content of the undigested fragment(s). The poly(A) sequence provides a marker for the 3' end of the molecule. Ribonuclease H selectively degrades RNA annealed to DNA but does not attack single-stranded RNA (23). Figure 5b shows that after treatment with ribonuclease H, 16S RNA migrates in the same position as untreated 16S RNA (Figure 5a). Globin mRNA however is totally digested by ribonuclease H after hybridization with cDNA (Figure 5d).

Figure 4. A comparison of the kinetics of hybridization of globin mRNA and 16S RNA to excess nonlabeled globin cDNA. Hybridizations were carried out as described under MATERIALS AND METHODS.

The 16S globin mRNA precursor containing poly(A) sequences [about 50% of the 16S RNA contains poly(A)] was isolated by oligo(dT)-cellulose chromatography and annealed to an excess of cDNA. Since the cDNA contains the oligo(dT) sequence used to prime its synthesis, hybrid formation could occur between this oligo(dT) and the poly(A) tail of the RNA. Thus, regardless of where the globin portion of the cDNA anneals to the precursor, the poly(A) sequence could be hydrolyzed when the hybrid is digested with ribonuclease H. After formation of the cDNA-16S RNA hybrid at 65°, 500 µg of poly(A) was therefore added to the reaction mixture and the incubation was continued for a further 60 min at 65°. It was shown in a separate experiment that this poly(A) effectively competes for the excess oligo(dT) in the mixture. A hybrid was formed between cDNA and ³H-poly(A) at 37°. This hybrid was then incubated with and without the addition of 500 µg of cold poly(A) for 60 min at 65°. The conditions of salt and temperature were identical to those used for the cDNA-16S RNA hybrid. In the absence of 500 µg of unlabelled poly(A) the ³H-poly(A) was in hybrid form and was completely protected from digestion by ribonuclease T₂. However, if 500 µg of unlabelled poly(A) was added to the reaction, greater than 95% of the labelled poly(A) was digested by ribonuclease T₂. This proves that the excess poly(A) effectively competes for the oligo(dT) sequence of the

Figure 5. Effect of ribonuclease H on 16S RNA and on cDNA-mRNA hybrids. Nuclease digestion and sucrose gradient centrifugation were as described in MATERIALS AND METHODS. a) 16S RNA; b) 16S RNA incubated in the presence of a 1:200 dilution of ribonuclease H for 30 min at room temperature; c) 9S globin mRNA; d) globin 9S mRNA-cDNA hybrid incubated in the presence of a 1:200 dilution of ribonuclease H for 30 min at room temperature. The reaction volumes were 50 µl and contained 3,000 cpm of 16S RNA or 4,000 cpm of 9S in RNA. cDNA was used at 25 µg/ml.

cDNA and thus protects the poly(A) tail of the 16S RNA from cDNA dependent hydrolysis by ribonuclease H. Furthermore, when labelled poly(A)-containing 9S globin mRNA (0.1 μ g) was digested with RNase H in the presence of oligo(dT) (0.5–2.5 μ g) and 500 μ g of poly(A), more than 90% of the counts could be rehybridized to oligo(dT)-cellulose. Thus the poly(A) tail of the RNA is also protected from partial cleavage by ribonuclease H which might prevent its rebinding to oligo(dT) cellulose.

Following the treatment of the 16S RNA-globin cDNA hybrid with RNase H the material was analyzed by sucrose density gradient centrifugation (Figure 6). A major peak of radioactivity was seen migrating at about 9S and a smaller peak at about 5S. The size of the undigested fragments was more accurately determined by electrophoresis in polyacrylamide gels run in the presence of 98% formamide (Figure 7). Comparison of the mobility of the major peak to that of RNA standards gave a molecular weight of 316,000 or about 990 nucleotides. The total length of the 16S precursor has been calculated to be 1,900 nucleotides of which approximately 150 are in the poly(A) region (our unpublished observations). Since we know that our cDNA protects 90% of the 720 nucleotides (30)

Figure 6. Sucrose gradient analysis of the products of a 16S RNA-cDNA hybrid digested by ribonuclease H. 40,000 counts of poly(A)-containing 16S RNA were hybridized with 5 μ g of unlabelled cDNA in 0.2 M NaCl, 20 mM Tris-HCl pH 7.5, 200 μ g/ml tRNA for 2 hr at 65°. 500 μ g of poly(A) was added in 0.2 M NaCl and the incubation was continued for 60 min at 65°. The hybridization mixture (200 μ l) was adjusted to 20 mM MgCl₂ 250 μ g/ml tRNA and ribonuclease H added to a 1:200 dilution. The final concentration of NaCl was 0.16 M. The digestion was allowed to proceed for 30 min at room temperature and terminated with EDTA to a final concentration of 20 mM. The mixture was layered directly onto a 10–30% sucrose gradient and centrifuged as described above.

Figure 7. Formamide-polyacrylamide gel electrophoresis of the unhydrolyzed material from the RNase H digestion of a 16S RNA-cDNA hybrid. The RNA from the gradient shown in Fig. 6 was pooled, ethanol precipitated, and applied to a polyacrylamide gel run in the presence of 98% formamide. 28S, 18S, 9S, and 4S RNAs were used as markers.

found in mature globin mRNA, then the unprotected fragment of 16S RNA should be about 1100 nucleotides in length. This is in reasonable agreement with the 990 nucleotide fragments found. It should be noted that the actual number of nucleotides found in globin mRNA, as determined by sequence analysis (30), is less than that obtained by polyacrylamide gel electrophoresis (31,32). However, because we have sized our fragments in gels the gel value for globin mRNA was used to be consistent.

Each fraction from the sucrose gradient (Figure 6) was applied to oligo (dT)-cellulose. None of the radioactivity was bound, indicating an absence of fragments containing Poly(A). This finding, coupled with the observed size of the major undigested peak of RNA, suggests that the globin mRNA sequence is located at the 3' end of the 16S precursor. However, a fragment containing as few as 5-10 nucleotides adjacent to the poly(A) sequence would probably represent too few counts to be detected following oligo(dT)-cellulose chromatography. It is possible therefore that the mRNA sequence is not contiguous with the poly(A) region of the precursor.

Figure 7 shows a small additional peak of radioactive RNA migrating with a molecular weight of about 50,000, or 155 nucleotides and lacking poly(A). The number of counts under this peak suggest a 1:1 stoichiometry between it and the larger 990 nucleotide fragment.

DISCUSSION

The purification and identification of the precursor to a specific mRNA has provided the opportunity to examine models of processing. The results reported here suggest that the globin mRNA derives from the 3' end of its 16S precursor. However, the presence of two size classes of undigested fragments after treatment of the 16S RNA-cDNA hybrid with ribonuclease H (Figure 7)

raises the possibility of an alternative model. The 155 nucleotide fragment might represent the undigested 5' ends of a smaller precursor RNA present in the 16S RNA preparation. However, our 16S preparations do not exhibit the level of contamination by smaller RNA compatible with such an explanation. Our results are also consistent with the existence of a discontinuous globin mRNA sequence in the 16S precursor, part of which is at the 3' end. Either the 155 nucleotide fragment, or the 990 nucleotide fragment, or both, could represent "silent" inserts which are absent in the mature 9S mRNA from which the cDNA was copied. Such a model of processing would require the excision of these inserts, and splicing of the two or three remaining sequences to form the mature messenger RNA.

ACKNOWLEDGEMENTS

This work was supported by grant GN-10999 from the U.S. Public Health Service, grant PCM-768022 from the National Science Foundation and grant NP-59H from the American Cancer Society. KS is a Fellow in Cancer Research supported by grant DRG-87F of the Damon Runyon-Walter Winchell Cancer Fund. Reverse transcriptase was supplied through the Office of Program Resources and Logistics, Viral Cancer Program, Viral Oncology Division of Cancer Cause and Prevention, National Cancer Institute, Bethesda, MD 20014. We are indebted to Dr. J.G. Stavrianopoulos for his generous gift of ribonuclease H and to Dr. Michael Mathews for his helpful discussion. The technical assistance of Julianna Sisson and Joe Raterman is gratefully acknowledged.

REFERENCES

- 1 Edmonds, M., Vaughan, M.H., Jr., and Nakazato, H. (1971) *Proc. Natl. Acad. Sci. USA* 68, 1336-1340.
- 2 Lee, S.Y., Mendecki, J., and Brawerman, G. (1971) *Proc. Natl. Acad. Sci. USA* 68, 1331-1335.
- 3 Darnell, J.E., Wall, R., and Tushinski, R.K. (1971) *Proc. Natl. Acad. Sci. USA* 68, 1321-1325.
- 4 Perry, R.P., Kelley, D.E., and LaTorre, J. (1974) *J. Mol. Biol.* 82, 315-331.
- 5 Herman, R.C., Williams, J.G., and Penman, S. (1976) *Cell* 7, 429-437.
- 6 Perry, R.P., Kelley, D.E., Friderici, K.H., and Rottman, F. (1975) *Cell* 6, 13-19.
- 7 Salditt-Georgieff, M., Jelinek, W., Darnell, J.E., Furuichi, Y., Morgan, M., and Shatkin, A. (1976) *Cell* 7, 227-237.
- 8 Bajszar, C., Samarina, O.P., and Georgiev, G.P. (1976) *Cell* 9, 323-332.
- 9 Schihler, V., and Perry, R.P. (1976) *Cell* 9, 121-130.
- 10 Nabazato, H., Edmonds, M., and Kopp, D.W. (1974) *Proc. Natl. Acad. Sci. USA* 71, 200-204.
- 11 Williamson, R., Drewienkiewicz, C., and Paul, J. (1973) *Nature New Biol.* 241, 66-68.
- 12 Knöchel, W., and Tiedemann, H. (1975) *Biochem. Biophys. Acta* 378, 383-393.
- 13 Melli, M., and Pemberton, R.E. (1972) *Nature New Biol.* 236, 172-173.

- 14 Scherrer, K. (1973) in Karolinska Symposia on Research Methods in Reproductive Endocrinology, 6th Symposium, ed. Diczfalussy, D. (Karolinska Institute, Stockholm), pp. 95-129.
- 15 Ruiz-Carrilo, A., Beato, M., Schutz, G., Geigelson, P., and Allfrey, V.G. (1973) Proc. Natl. Acad. Sci. USA 70, 3641-3645.
- 16 Macnaughton, M., Freeman, K.B., and Bishop, J.O. (1974) Cell 1, 117-125.
- 17 Ostertag, W., Gaedicke, G., Kuge, N., Melderis, H., Weimann, B., and Dube, S.K. (1973) FEBS Lett. 32, 218-222.
- 18 Ross, J. (1976) J. Mol. Biol. 106, 403-420.
- 19 Curtis, P. J., and Weissmann, C. (1976) J. Mol. Biol. 106, 1061-1075
- 20 Bastos, R. N., and Aviv, H. (1977) Cell 11, 641-650
- 21 Kwan, S.-P., Wood, T. G., and Lingrel, J. B. (1977) Proc. Natl. Acad. Sci. USA 64, 178-182
- 22 Curtis, P., Mantei, N., VandenBerg, J., and Weissmann, C. (1977) Proc. Natl. Acad. Sci. USA 73, in press
- 23 Stavrianopoulos, J. G., and Chargaff, E. (1973) Proc. Natl. Acad. Sci. USA 70, 1959-1963
- 24 Merkel, C. G., Kwan, S.-P., and Lingrel, J. B. (1975) J. Biol. Chem. 250, 3725-3728
- 25 Wood, T. G., and Lingrel, J. B. (1977) J. Biol. Chem. 252, 457-463
- 26 Gorski, J., Morrison, M. R., Merkel, C. G., and Lingrel, J. B. (1974) J. Mol. Biol. 86, 363-371
- 27 Orkin, S. H., Swan, D., and Leder, P. (1975) J. Biol. Chem. 250, 8753-8760
- 28 Merkel, C. G., Wood, T. G., and Lingrel, J. B. (1976) J. Biol. Chem. 251, 5512-5515
- 29 Dot is defined as the concentration of DNA in moles of nucleotide/litre times time in seconds
- 30 Efstratiadis, D., Kafatos, F. C., and Maniatis, T. (1977) Cell 10, 571-586
- 31 Morrison, M. R., and Lingrel, J. B. (1976) Biochim. Biophys. Acta 447, 104-109
- 32 Hamlyn, P. H., and Gould, H. J. (1975) J. Mol. Biol. 94, 101-109