Nucleotide Sequences of Human Globin Messenger RNA

Charles A. Marotta, Bernard G. Forget, Sherman M. Weissman, Inder M. Verma, Ronald P. McCaffrey, and David Baltimore

PNAS 1974;71;2300-2304
doi:10.1073/pnas.71.6.2300

Proc. Nat. Acad. Sci. USA
Vol. 71, No. 6, pp. 2300–2304, June 1974

Nucleotide Sequences of Human Globin Messenger RNA

(RNA-dependent DNA polymerase/RNA polymerase/polynucleotide kinase/thalassemia/hemoglobin Constant Spring)

CHARLES A. MAROTTA*§, BERNARD G. FORGET†, SHERMAN M. WEISSMAN*, INDER M. VERMA‡¶, RONALD P. McCAFFREY‡, AND DAVID BALTIMORE‡

* Department of Medicine, Yale University School of Medicine, New Haven, Connecticut 06510; † the Division of Hematology of the Department of Medicine, Children's Hospital Medical Center, and the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115; and ‡ the Department of Biology and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Communicated by Alexander Rich, March 18, 1974

ABSTRACT **Globin messenger RNA, isolated from human peripheral blood reticulocytes, was transcribed into complementary DNA by use of the RNA-dependent DNA polymerase of avian myeloblastosis virus. The complementary DNA was then transcribed into ^{32}P-labeled complementary RNA by *E. coli* RNA polymerase in the presence of α-^{32}P-labeled ribonucleoside triphosphates. The fingerprint pattern obtained from ribonuclease T1 digests of human globin complementary RNA was specific and reproducible. Different patterns were obtained from digests of duck, mouse, and rabbit globin complementary RNA. The fingerprint patterns obtained from digests of purified natural human 10S globin messenger RNA, labeled *in vitro* with ^{125}I or with [γ-^{32}P]ATP and polynucleotide kinase, were similar to that of the complementary RNA but contained some additional oligonucleotides. Sufficient nucleotide sequence information has been obtained from about 50% of the intermediate sized oligonucleotides (8–14 base residues long), to make possible examination of correspondence between these nucleotide sequences and globin amino-acid sequences. Approximately 70% of these oligonucleotide sequences can be matched to unique amino-acid sequences in the α- or β-globin chains. The other 30% do not match known amino-acid sequences and presumably correspond to untranslated portions of the mRNA; some of these sequences, however, can be matched to amino-acid sequences in the abnormally long segment of the α chain of hemoglobin Constant Spring, which is thought to result from a chain-termination mutation.**

The study of the nucleotide sequence of human globin messenger RNA (mRNA) provides a means to examine the molecular genetics of normal and abnormal human hemoglobin synthesis. Such studies are particularly relevant because there exist in man a large number of hereditary disorders that affect the structure and/or rate of synthesis of one or the other human globin chain, and that are presumed to be the consequence of single base substitutions in the mRNA or due to quantitative deficiency of chain-specific mRNA.

It is possible to isolate functional human globin mRNA from peripheral blood reticulocytes, and have it translated in various cell-free protein-synthesizing systems (1–3). However, most techniques used in RNA sequencing rely on autoradiography of fractionated digests of ^{32}P-labeled RNA of high specific radioactivity (4). Because it is not possible to obtain ^{32}P-labeled human globin mRNA by *in vivo* or cell culture labeling techniques, we have used a different approach to obtain ^{32}P-labeled RNA. Globin mRNA is an efficient substrate for the RNA-dependent DNA polymerase (or reverse transcriptase) of avian myeloblastosis virus (5–7): in the presence of oligo(dT), which serves as a primer presumably by binding to a poly(A) sequence situated at the 3′-terminus of the RNA, a DNA copy (cDNA) of the mRNA is synthesized (5–7). In the presence of actinomycin D, the cDNA synthesized is single-stranded, and can be shown to be nearly 100% complementary to the globin mRNA by hybridization assays (5). We have used the globin cDNA as a substrate for the DNA-dependent RNA polymerase of *Escherichia coli:* in the presence of [α-^{32}P]ribonucleoside triphosphates, we obtained the synthesis of a ^{32}P-labeled RNA (cRNA) of high specific radioactivity, which was shown, after analysis by standard RNA sequencing techniques, to contain most of the nucleotide sequence information of the natural 10S human globin mRNA. The fingerprint pattern of the cRNA is similar to that of digests of purified natural 10S RNA labeled *in vitro* by different techniques, and most of the nucleotide sequences identified can be matched to amino-acid sequences of the α- or β-globin chains of human hemoglobin (Hb) A [$\alpha_2\beta_2$]; other sequences cannot be matched, and presumably correspond to untranslated sequences of the mRNA.

MATERIALS AND METHODS

Materials. α-^{32}P-labeled ATP, GTP, CTP, and UTP (specific activity 50–150 Ci/mmole), [γ-^{32}P]ATP (20 Ci/mmole), and ^{125}I (carrier free) were obtained from the New England Nuclear Corp. [^{3}H]dGTP (10–15 Ci/mmole) was obtained from Schwarz-Mann or New England Nuclear. Oligo(dT)$_{12-18}$ and oligo(dT)–cellulose (type T2) were obtained from Collaborative Research, Inc., Waltham, Mass. The RNA-dependent DNA polymerase of avian myeloblastosis virus was purified from virions as described (8–10). The DNA-dependent RNA polymerase of *E. coli* was the gift of Dr. P. Lebowitz and was isolated and purified as described (11). DNase, electrophoretically purified, was obtained from Schwarz-Mann. Polynucleotide kinase (EC 2.7.1.78; ATP: 5′-dephosphopolynucleotide 5′-phosphotransferase) was purified from phage T4-infected *E. coli* cells (12). Purified *E. coli* alkaline phosphatase (EC 3.1.3.1) (RNase free) was the gift of Drs. J. Coleman and J. Taylor, and was prepared as described (13). Materials used in nucleotide sequence techniques (4, 14–17) have been described.

Preparation of Human Globin mRNA. Total cellular RNA was prepared from membrane-free lysates of reticulocyte-rich peripheral blood, by detergent and phenol–cresol extrac-

Abbreviations: cDNA, DNA synthesized from mRNA by DNA polymerase of avian myeloblastosis virus; cRNA, RNA synthesized from cDNA by *Escherichia coli* RNA polymerase; $(Me)_2SO$; dimethylsulfoxide; Hb, hemoglobin.

Present addresses: §Department of Psychiatry, Massachusetts General Hospital, Boston, Mass 02114; ¶Salk Institute, P.O.B. 1809, San Diego, Calif. 92112.

Please address reprint requests to Dr. Bernard G. Forget, Children's Hospital Medical Center, Boston, Mass. 02115.

tion (2, 18); the RNA was then fractionated by sucrose gradient centrifugation. The RNA sedimenting between 4 S and 18 S was precipitated with ethanol (18) and served as a partially purified mRNA fraction for synthesis of cDNA. In some cases, the RNA was further fractionated by acrylamide gel electrophoresis (19); then the 10S RNA was electrophoretically eluted (20), and used for cDNA synthesis.

The pure 10S globin mRNA used in the labeling experiments *in vitro*, was obtained by oligo(dT)–cellulose column chromatography (21) of the partially purified sucrose gradient mRNA fraction. The RNA eluted from the column by 10 mM Tris·HCl (pH 7.5), consisted of a single 10S RNA band when analyzed by electrophoresis on 5% acrylamide gels.

Synthesis of DNA Complementary to Globin mRNA. Single-stranded DNA complementary to globin mRNA was prepared by incubating the mRNA with the DNA polymerase of avian myeloblastosis virus (5); a typical reaction contained in 0.2 ml: 10 μg of 10S RNA; 5 μg/ml of oligo $(dT)_{12-18}$; 1 mM each dATP, dCTP, and dTTP; 0.2 mM dGTP; [^{3}H]dGTP (25 μCi/ml; final specific activity 40–60 cpm/pmole); 100 μg/ml of actinomycin D; 30 μl [15–20 units (9)] of purified DNA polymerase; 10 mM dithiothreitol; 6 mM magnesium acetate; 60 mM NaCl; and 50 mM Tris·HCl, pH 8.3. After incubation for 90–180 min at 37°, the solution was heated for 5 min in the presence of 0.3 N NaOH, at 100°, then neutralized with HCl and fractionated by gel filtration on Sephadex G-50 columns equilibrated with 0.05 M triethylamine bicarbonate (pH 7.0). The fractions containing the excluded material were pooled and lyophilized; the residue was resuspended in H_2O and used for synthesis of complementary RNA.

Synthesis of Globin cRNA. The preparation and purification of RNA transcripts from globin cDNA using *E. coli* RNA polymerase and [α-^{32}P]ribonucleoside triphosphates is described (14). In summary, 1–3 μg of cDNA was incubated with 5 μg of RNA polymerase in the presence of 0.18 M KCl, 33 mM Tris·HCl (pH 7.9), 3.3 mM $MgCl_2$, 6.0 mM mercaptoethanol, three nonradioactive ribonucleoside triphosphates, (0.165 or 0.33 mM each), and 1–2 mCi (0.165 mM) of the fourth ribonucleoside triphosphate labeled with ^{32}P in the α position. After incubation at 37° for 50 min, an additional 5 μg of RNA polymerase were added and the mixture was incubated for 10 more minutes at 37°. Finally, the reaction was terminated by adding 2 μg of DNase, incubating for 5 min at 37°, then adding sodium dodecyl sulfate to 0.25%, and extracting the mixture with phenol. The ^{32}P-labeled RNA was separated from the nonincorporated [^{32}P]triphosphate by gel filtration over a column of Sephadex G-100, equilibrated with 10 mM Tris·HCl (pH 7.9), 10 mM $MgCl_2$, and 0.1 mM EDTA, then precipitated with ethanol in the presence of carrier RNA (14). The properties of rabbit globin cRNA have been described (39).

In Vitro Labeling of Natural mRNA. Pure 10S mRNA [purified by oligo(dT)–cellulose column chromatography] was digested simultaneously with T1 RNase and *E. coli* alkaline phosphatase. Then the 5′-terminus of each resulting oligonucleotide was labeled by incubating the digest in the presence of [γ-^{32}P]ATP and polynucleotide kinase (23, 24). Intact pure 10S RNA was labeled with ^{125}I by the method of Commerford (25). These techniques have been used by others to label globin mRNA for fingerprint analysis (24, 37).

RNA Sequence Analysis Procedures. General RNA sequence analysis methodology has been described (4, 14–17). The ^{32}P-labeled cRNA was completely digested with T1 RNase either directly after Sephadex G-100 gel filtration or after additional purification by $(Me)_2SO$–sucrose gradient centrifugation. Fingerprints were then prepared by electrophoresis of the digests on urea–Cellogel® strips, followed by homochromatography (14, 17).

RESULTS

Size of the Globin cRNA. The ^{32}P-labeled RNA isolated after incubation of human globin cDNA with *E. coli* RNA polymerase was fractionated by $(Me)_2SO$–sucrose gradient centrifugation (22) and by acrylamide gel electrophoresis in formamide (36) in order to determine its size. After $(Me)_2SO$–sucrose gradient centrifugation, the peak of radioactive cRNA sedimented slightly slower than 6 S, using as a standard the 6S RNA transcribed from the DNA of bacteriophage λ B_2B_5 (a gift of Dr. P. Lebowitz) (11), which was analyzed in a parallel gradient (data now shown). Analysis of the cRNA by acrylamide gel electrophoresis in formamide demonstrated that the bulk of the cRNA migrated with the faster moving components of reticulocyte transfer RNA (4S RNA), added as a marker (data not shown).

Fingerprint Pattern of the Globin cRNA. Fig. 1 shows the autoradiograph of a two-dimensional fractionation ("fingerprint") of a RNase T1 digest of human (sickle cell anemia) globin cRNA labeled with [α-^{32}P]GTP. The pattern is quite reproducible and specific for cRNA synthesized from human globin cDNA. The same pattern was obtained with cRNA derived from different mRNA preparations of the same patient and from mRNA of six unrelated patients with sickle cell anemia and that of another patient with Hb A and hemolytic anemia. Neither the use of different batches of DNA polymerase for the synthesis of cDNA nor the addition of manganese (0.8 mM) to the RNA polymerase reaction resulted in a qualitatively different fingerprint pattern of the cRNA. The same pattern was also obtained from the cRNA when, instead of partially purified (sucrose gradient-fractionated) mRNA, 10S RNA purified by acrylamide gel or oligo(dT)–cellulose chromatography was used for the synthesis of cDNA. On the other hand, different and distinctive fingerprint patterns were obtained with RNase T1 digests of mouse, rabbit, and duck globin cRNA (data not shown).

The specificity and fidelity of the cRNA fingerprint is further illustrated by the fact that the fingerprint pattern of natural 10S globin mRNA is similar to that of the cRNA. Pure human globin 10S mRNA, purified by oligo(dT)–cellulose column chromatography, was labeled with ^{125}I (25), then digested with RNase T1. The fingerprint pattern of this *in vitro* labeled natural 10S mRNA (derived from a patient with Hb A rather than HbS) is shown in Fig. 2. The fingerprint pattern is very similar to that of the cRNA (Fig. 1). Most of the large oligonucleotides present in the natural 10S mRNA fingerprint are also present in the cRNA, but the natural 10S RNA fingerprint does contain a few additional spots, indicated by the *arrows*. The mobilities of some of the spots differ somewhat in the two fingerprints, making exact matching occasionally equivocal. Differences in mobility in the first dimension may be caused by the modification of the cytosine residues by ^{125}I (37). Nevertheless, the overall patterns are unequivocally quite similar. A sample of the same natural 10S RNA, which was labeled *in vitro* (after simultaneous digestion with RNase T1 and alkaline phosphatase) by means of [γ-^{32}P]ATP and polynucleotide kinase, also gave a fingerprint pattern similar to that of the cRNA, but with a few additional spots (data not shown).

The absence of certain oligonucleotides in the cRNA

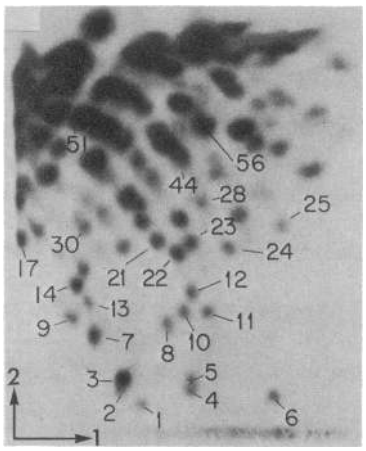

FIG. 1. Autoradiograph of the fingerprint of a T1 RNase digest of cRNA, synthesized from human (sickle cell anemia) globin cDNA, in the presence of [α-^{32}P]GTP. The RNA was partially purified by $(Me)_2SO$–sucrose gradient (22) before digestion. Fractionation, in the first dimension (*left* to *right*) was by electrophoresis on Cellogel®–urea strip at pH 3.5 (15–17), and in the second dimension (*bottom* to *top*) by homochromatography (17) on thin-layer plate of DEAE–cellulose.

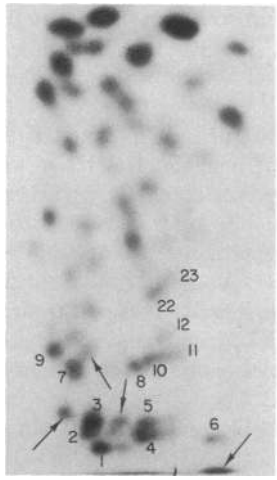

FIG. 2. Autoradiograph of the fingerprint of a T1 RNase digest of natural human 10S globin mRNA. 10S mRNA was purified from the reticulocyte RNA of a patient with Hb A and hemolytic anemia by sucrose gradient centrifugation, followed by oligo(dT)–cellulose column chromatography. The RNA was labeled with ^{125}I (25), then digested with T1 RNase, and fractionated as in Fig. 1. The spots are numbered as in Fig. 1; the *arrows* indicate spots present in the fingerprint of natural 10S RNA, which are absent in the fingerprint of the cRNA (Fig. 1).

fingerprint indicates that the cRNA (and presumably the cDNA) is not a complete copy of the natural mRNA. The globin cRNA, therefore, is a faithful although incomplete representation of the natural 10S globin mRNA and contains most, although not all, of its sequence information, probably in the form of a heterogeneous group of fragments shorter than the original mRNA and cDNA.

Nucleotide Sequences of Human Globin cRNA. The nucleotide sequences of the individual oligonucleotides obtained after RNase T1 digestion of the cRNA were further studied by pancreatic RNase (and in some cases U2 RNase) digestion of the eluted spots. By combining the information, obtained from separate digests of the four types of labeled RNA, that were synthesized in the presence of only one radioactive ribonucleoside triphosphate, a great deal of sequence information can be obtained through "nearest-neighbor analysis"; the labeled phosphate, after digestion with T1 or pancreatic RNase, is associated with the nucleotide adjacent and to the 5′-side of the labeled precursor nucleotide. For instance, the labeled phosphate of [α-^{32}P]GTP, incorporated into the sequence —C*pGp— would be found with C*p after hydrolysis. In addition, the 5′-terminal nucleotide sequences of the various spots was confirmed by analysis of the corresponding oligonucleotides of a RNase T1 digest of natural 10S RNA, labeled by means of [γ-^{32}P]ATP and polynucleotide kinase. Additional sequence information was also obtained from analysis of the oligonucleotides of ^{125}I-labeled natural 10S mRNA. Sufficient sequence information is available, in the case of approximately half of the intermediate-sized oligonucleotides (8–14 base residues), to establish a single or limited number of linear sequence possibilities. This information was then analyzed by a computer, which considered all possible sequence permutations consistent with the sequence information of each spot and matched them to the possible codons for the amino-acid sequence of each globin chain; [a similar technique has been used by others (27, 28)]. Table 1 lists nine nucleotide sequences of the cRNA that can be matched to unique amino-acid sequences in the α- or β-globin chains. None of these nucleotide sequences (or their permutations) can be matched to other amino-acid sequences in either α- or β-globin chains. The probability of error in sequences of this length is relatively low in view of the uniqueness of the match between nucleotide and amino-acid sequence (27, 28).

The validity of the sequence assignments of Table 1 is reinforced by the fact that the oligonucleotides that match β-chain sequences are very prominent, whereas those that match α-chain sequences are very faint and barely visible in the fingerprint of cRNA of a patient with α thalassemia (Hb H disease) and marked deficiency of α-chain mRNA, previously demonstrated by RNA-DNA hybridization (26).

All but two of the amino-acid sequences represented in Table 1 are situated in the distal (COOH-terminal) two-thirds of the β chain. The two α-chain sequences are situated in the distal 15% of the α chain. These results suggest that the cDNA is primarily a representation of the central and 3′-terminal portions of the mRNA. The 5′-terminal portion of the mRNA may not be represented in the cDNA. The oligonucleotides present in the natural 10S mRNA (Fig. 2) but absent in the cRNA (Fig. 1) are therefore likely to correspond to the 5′-terminal portion of the mRNA.

Table 2*A* lists the sequences of two other oligonucleotides of intermediate length, which cannot be matched to any α- or β-globin amino-acid sequences, in spite of all the permutations allowed by the ambiguousness of their sequences. These sequences presumably represent untranslated portions of the mRNA. In addition, these sequences do not match amino-acid sequences in the abnormally long segment of the Hb Constant Spring chain, which presumably results from a chain-termination mutation, allowing translation of a normally untranslated portion of the α-chain mRNA (29, 30).

Three other sequences, however, have been identified that do match amino-acid sequences of the α Constant Spring-globin chain. They are listed in Table 2*B*. Spot no. 14 is the longest and it is a unique match. Spot no. 51 is a mixture of at least two sequences, (G)CCCUG(G) and (G)[C_2,U]CG(G),

TABLE 1. *Nucleotide sequences of human globin cRNA that match unique globin-chain amino-acid sequences*

Spot no.	Sequence	Chain
3†	ACC[U][C]CAAAUACCG	
	136 137 138 139 140 141	
	Leu Thr Ser Lys Tyr Arg	Alpha
	NN(G)–ACC–UCC–AAA–UAC–CG(U)	
7*	ACAACCUCAAG	
	79 80 81 82 83	
	Asp Asn Leu Lys Gly	Beta
	(G)AC–AAC–CUC–AAG-(G)NN	
11*	CCCAUCACUUUG	
	115 116 117 118 119	
	Ala His His Phe Gly	Beta
	(G)CC–CAU–CAC–UUU-G(G)N	
12*	U[AUC][ACU]AAG	
	144 145 146	
	Lys Tyr His Term	Beta
	NN(G)–UAU–CAC–UAA-G(C)N	
13*	GC[AACC][CU]AAG	
	56 57 58 59 60	
	Gly Asn Pro Lys Val	Beta
	(G)GC–AAC–CCU–AAG-(G)NN	
21*	AACUUCAG	
	101 102 103 104	
	Glu Asn Phe Arg	Beta
	NN(G)–AAC–UUC–AG(G)	
22†	CCUCA[C][U]UG	
	123 124 125 126	
	Ala Ser Leu Asp	Alpha
	(G)CC–UCA–CUU–G(A)N	
28*	CACCUUUG	
	83 84 85 86	
	Gly Thr Phe Ala	Beta
	N(G)C–ACC–UUU–GNN	
30*	CCACACUG	
	86 87 88 89	
	Ala Thr Leu Ser	Beta
	(G)CC–ACA–CUG-(A)NN	

The spot numbers refer to Fig. 1. The sequences listed are based on the composite information obtained by pancreatic RNase (and in some cases, U2 RNase) digestion of the oligonucleotide from T1 RNase digests of the cRNA, labeled, in separate experiments, with a different one of the four [α-^{32}P]nucleoside triphosphates. The sequence of the 5′-terminus of the oligonucleotides was confirmed by analysis of the corresponding oligonucleotides of the natural 10S mRNA labeled by polynucleotide kinase. Many internal sequences were confirmed by study of the corresponding oligonucleotides of the ^{125}I-labeled natural 10S mRNA, analyzed by partial and complete pancreatic RNase digestion and U2 RNase digestion. Listed first is the sequence determined solely by analysis of nucleotide sequence data. The positions of the nucleotides in *brackets* can be interchanged on the basis of sequence data alone. Listed under the sequence is the unique amino-acid sequence matched to the nucleotide sequence by computer analysis, and the corresponding codon grouping of the nucleotide sequence. The nucleotides in *parentheses* at the 3′-terminus (*right*) and 5′-terminus (*left*) of the hyphenated nucleotide sequence are not contained in the numbered oligonucleotide but are known to be present in the indicated position by "nearest-neighbor" analysis (see *text*) and by the nature of the nuclease (T1 RNase) digest, respectively. *N* indicates an unknown nucleotide in a codon. The *numbers* over the sequences refer to the position of the indicated amino acid in either the α- or β-globin chain.

* The oligonucleotide is prominent in the fingerprint of alpha thalassemia cRNA.

† The oligonucleotide is faint or absent in the fingerprint of alpha thalassemia cRNA.

one of which, (G)CCUCG(G), matches the Hb Constant Spring α-chain sequence positions 145 to 147.

Spot no. 56 also contains a mixture of sequences: (G)UUAAG(C), (G)UAAUG, and (G)AUUAG(C), among them the sequence (G)UUAAG(C) which matches the α-chain COOH-terminal amino acid (no. 141), the chain-termination codon (position 142), and the start of the codon for amino-acid no. 143 of the α Constant Spring chain. The sequence of spot no. 56 is entirely consistent with the current hypothesis of the origin of Hb Constant Spring by chain-termination codon mutation: UAA to CAA (term. to Gln) (30).

The sequences of spots no. 51 and no. 56 listed in Table 2*B* were sequences predicted for α chain mRNA by Seid-Akhavan *et al.* by correlating the α-chain amino-acid sequence of Hb Constant Spring to that of Hb Wayne, a presumed frameshift mutation in the α-chain mRNA at amino-acid position no. 139 (31). If both of these mutant hemoglobins resulted from the type of mutation proposed, then a unique mRNA sequence of 26 base residues can be predicted, from the genetic code, for this region of the α-chain mRNA (31, 32). The sequence of spot no. 3 (Table 1) extends this derived sequence of the α-chain mRNA 10 more residues to the 5′-side.

Presence of Poly(A) Sequences in the cRNA. When cRNA is synthesized in the present of [α-^{32}P]ATP, the fingerprint of the RNase T1 digest of the RNA reveals a very prominent spot that has properties similar to that of the long poly(A) sequence identified by Brownlee *et al.* in immunoglobulin L-

TABLE 2. *Nucleotide sequences of human globin cRNA*

Spot no.	Sequence
(*A*) Sequences that do not match known globin chain amino-acid sequences	
6*	[G]U[AAU,AU,U_{3-5},C_{2-3}]AAAG(G)
25*	[G]U[AU_{1-2},U_{1-2}]AUG(A)
(*B*) Sequences that match or overlap amino-acid sequences in the α Constant Spring-globin chain	
14†	(G)[CU,C,C,C]CAACG(G)
	155 156 157 158
	Ala Ser Gln Arg
	(G)CC–UCC–CAA–CG(G)
51†	145 146 147
	Ala Ser Val
	(G)CC–UCG–(G)NN
56†	141 142 143
	Arg Term Ala
	N(G)U–UAA–G(C)N

The sequence information was obtained and is expressed as in legend of Table 1. The exact sequence of the nucleotides in *brackets* cannot yet be determined from the sequence information alone. The sequence listed for spot no. 51 also matches the amino-acid sequence of α-chain position 130–131. The sequence listed for spot no. 56 also matches [in a different codon grouping: NN(G)-UUA-AG(C) = y-Leu-Ser] the following amino-acid sequences: β 87–89; α 1–3; α 79–81; α 82–84 and α 100–102.

* and † See Table 1.

chain mRNA (28). However, the size of this poly(A) sequence in globin cRNA has not yet been precisely determined. Whatever its size, it may not be a true representation of the poly(A) sequence of the natural mRNA, because a long poly(A) sequence could possibly be transcribed by "slipping" of the RNA polymerase on the oligo(dT)$_{12-18}$ primer that is used in the synthesis of the cDNA and is presumably linked to the 5'-terminus of the cDNA.

DISCUSSION

The ^{32}P-labeled cRNA transcribed by *E. coli* RNA polymerase from human globin cDNA can be used to study the nucleotide sequence of a mRNA that cannot be otherwise efficiently labeled with ^{32}P *in vivo* or in tissue culture. The method is specific and reproducible and has the added advantage that a great deal of sequence information can be obtained by the use of only one labeled nucleoside triphosphate precursor at a time, which results in the labeling of the nearest-neighbor nucleotide in the RNA. The process of transcription of mRNA into cDNA and of cDNA into cRNA appears to be faithful, as evidenced by the fact that, on the basis of sequence information from roughly 50% of the oligonucleotides 8–14 base residues long, approximately 70% of the oligonucleotides of the cRNA can be matched to unique amino-acid sequences of the α- or β-globin chains. This percentage is roughly what would be expected from the size of the natural mRNA, which is estimated to have a molecular weight of 200,000 to 220,000 (33–36) (approximately 650 base residues) and, therefore, is approximately 30% longer than necessary to simply code for globin chains of 141 to 146 amino acids. Although the cRNA is smaller than the natural mRNA (4 S compared to 10 S), it appears to contain the majority of the sequences of the natural mRNA: the fingerprint patterns of digests of natural 10S RNA, labeled by ^{125}I and by [γ-^{32}P]-ATP with polynucleotide kinase, are very similar to that of the cRNA and contain only a few additional oligonucleotides, not present in the fingerprint of the cRNA.

Salser *et al.* (38) have reported nucleotide sequence studies of rabbit globin cRNA; three sequences, 6–9 base residues long, were identified, which match amino-acid sequences in the rabbit α- or β-globin chain. Proudfoot and Brownlee have identified a unique nucleotide sequence (AUUGC) adjacent to the poly(A) sequence of rabbit globin mRNA, by sequence analysis of short cDNA transcripts obtained by incubating rabbit globin mRNA with *E. coli* DNA polymerase I (40).

We thank Dr. D. G. Nathan for helpful discussions; Miss D. Paci for skilled technical assistance, and Dr. P. Lebowitz for the gift of *E. coli* RNA polymerase. B.G.F. is the recipient of a Research Career Development Award AM-70234 of the U.S. Public Health Service. I.M.V. was fellow of the Jane Coffin Childs Memorial Fund for Medical Research. D.B. is American Cancer Society Research Professor. This work was supported in part by the following grants of the National Institutes of Health: CA-13472, AM-15929, AM-05581, CA-14051, AM-15035, and a contract from the Virus Cancer Program.

1. Nienhuis, A. W. & Anderson, W. F. (1971) *J. Clin. Invest.* **50**, 2458–2460.
2. Benz, E. J., Jr. & Forget, B. G. (1971) *J. Clin. Invest.* **50**, 2755–2760.
3. Metafora, S., Terada, M., Dow, L. W., Marks, P. A. & Bank, A. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1299–1303.
4. Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965) *J. Mol. Biol.* **13**, 373–398.
5. Verma, I. M., Temple, G. F., Fan, H. & Baltimore, D. (1972) *Nature New Biol.* **235**, 163–166.
6. Ross, J., Aviv, H., Scolnick E. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 264–268.
7. Kacian, D. L., Spiegelman, S., Bank, A., Terada, M., Metafora, S., Dow, L. & Marks, P. A. (1972) *Nature New Biol.* **235**, 167–169.
8. Verma, I. M., Meuth, N. L., Bromfeld, E., Manly, K. F. & Baltimore, D. (1971) *Nature New Biol.* **233**, 131–134.
9. Baltimore, D. & Smoler, D. F. (1972) *J. Biol. Chem.* **247**, 7282–7287.
10. Verma, I. M. & Baltimore, D. (1973) "Purification of the RNA-directed DNA polymerase from avian myeloblastosis virus and its assay with polynucleotide templates," in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 29, pp. 125–130.
11. Lebowitz, P., Weissman, S. M. & Radding, C. (1971) *J. Biol. Chem.* **246**, 5120–5139.
12. Richardson, C. C. (1965) *Proc. Nat. Acad. Sci. USA* **54**, 158–165.
13. Applebury, M. L., Johnson, B. P. & Coleman, J. E. (1970) *J. Biol. Chem.* **245**, 4968–4976.
14. Marotta, C. A., Lebowitz, P., Dhar, R. & Weissman, S. M. (1973) "Preparation of RNA transcripts of discrete segments of DNA," in *Methods in Enzymology*, eds., Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 29, pp. 254–272.
15. Brownlee, G. G., Sanger, F. & Barrell, B. G. (1968) *J. Mol. Biol.* **34**, 379–412.
16. Forget, B. G. & Weissman, S. M. (1968) *J. Biol. Chem.* **243**, 5709–5723.
17. Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11**, 395–399.
18. Benz, E. J., Jr., Swerdlow, P. S. & Forget, B. G. (1973) *Blood* **42**, 825–833.
19. Forget, B. G., Baltimore, D., Benz, E. J., Jr., Housman, D., Lebowitz, P., Marotta, C. A., McCaffrey, R. P., Skoultchi, A., Swerdlow, P. S., Verma, I. M. & Weissman, S. M. (1974) *Ann. N.Y. Acad. Sci.*, **232**, 76–87.
20. Forget, B. G., Marotta, C. A., Weissman, S. M., Verma, I. M., McCaffrey, R. P. & Baltimore, D. (1974) *Ann. N.Y. Acad. Sci.*, in press.
21. Aviv, H. & Leder, P. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 1408–1412.
22. Strauss, J. H., Jr., Kelley, R. B. & Sinsheimer, R. L. (1968) *Biopolymers* **6**, 793–807.
23. Szekely, M. & Sanger, F. (1969) *J. Mol. Biol.* **43**, 607–617.
24. Labrie, F. & Sanger, F. (1969) *Biochem. J.* **114**, 29 P (abstract).
25. Commerford, S. L. (1971) *Biochemistry* **10**, 1993–1999.
26. Housman, D., Forget, B. G., Skoultchi, A. & Benz, E. J., Jr. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 1809–1813.
27. Adams, J. M., Jeppesen, P. G. N., Sanger, F. & Barrell, B. G. (1969) *Nature* **223**, 1009–1014.
28. Brownlee, G. G., Cartwright, E. M., Cowan, N. J., Jarvis, J. M. & Milstein, C. (1973) *Nature New Biol.* **244**, 236–239.
29. Milner, P. F., Clegg, J. B. & Weatherall, D. J. (1971) *Lancet* **1**, 729–732.
30. Clegg, H. B., Weatherall, D. J. & Milner, P. F. (1971) *Nature* **234**, 337–340.
31. Seid-Akhavan, M., Winter, W. P., Abramson, R. K. & Rucknagel, D. L. (1972) *Blood* **40**, 927 (abstract).
32. Laux, B., Dennis, D. & White, H. B., III. (1973) *Biochem. Biophys. Res. Commun.* **54**, 894–898.
33. Labrie, F. (1969) *Nature* **221**, 1217–1222.
34. Williamson, R., Morrison, M., Lanyon, G., Eason, R. & Paul, J. (1971) *Biochemistry* **10**, 3014–3021.
35. Gaskill, P. & Kabat, D. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 72–75.
36. Gould, H. J. & Hamlyn, P. H. (1973) *FEBS Lett.* **30**, 301–304.
37. Robertson, H. D., Dickson, E., Model, P. & Prensky, W. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3260–3264.
38. Salser, W., Poon, R., Whitcome, P. & Fry, K. (1973) "New techniques for determining nucleotide sequences from eucaryotic cells," in *Virus Research*, eds., Fox, C. F. & Robinson, W. S. (Academic Press, New York), p. 545.
39. Baltimore, D., Verma, I. M., Fan, H. & Temple, G. F. (1973) "Synthesis by reverse transcriptase of DNA complementary to globin messenger RNA," in *Viral Replication and Cancer*, eds., Melnick, J. L., Ochoa, S., & Oro, J. (S. A. Calabria, Barcelona), pp. 155–169.
40. Proudfoot, N. J. & Brownlee, G. G. (1974) *FEBS Lett.* **38**, 179–183.