The nucleotide sequence of the 5' untranslated region of human γ-globin mRNA

Judy C.Chang*†, Raymond Poon*+, Kurt H.Neumann* and Yuet Wai Kan*+

*Hematology Service, San Francisco General Hospital, University of California, San Francisco and +Howard Hughes Medical Institute Laboratory, San Francisco General Hospital, San Francisco, CA, USA

Received 1 August 1978

ABSTRACT

The nucleotide sequence of the entire 5' untranslated region of human γ-globin mRNA has been determined. This was accomplished by analyzing complementary DNA (cDNA) synthesized from the mRNA with reverse transcriptase. The cDNA was labeled at its 3' end with ^{32}P using terminal deoxynucleotidyl transferase, digested with the restriction endonuclease Hae III and the end-labeled fragment isolated and sequenced by the method of Maxam and Gilbert. Including the initiation codon AUG, the 5' untranslated region of human γ-globin mRNA contains 57 nucleotides, compared to 41 in α- and 54 in β-globin mRNA. There is very little homology between α and γ sequences in the 5' region. There is considerable homology between β- and γ-globin mRNAs in the regions proximal and distal to the initiation codon, but the entire sequence shows less homology than the human and rabbit β-globin mRNAs. The hexanucleotide sequence CUUCUG is found near the 5' ends of all three human globin mRNAs, suggesting a possible role of this sequence in ribosomal binding. Both guanosine and cytidine were found at the 19th nucleotide position from the 5' end of the γ mRNA. We believe this heterogeneity arises from the difference in nucleotide sequence between the $^{A}\gamma$ and $^{G}\gamma$ loci.

INTRODUCTION

Mature eukaryotic mRNAs contain 5' and 3' noncoding sequences, which respectively preceed and follow the sequences normally translated into proteins. The sequences of the noncoding regions are of interest because they may play a role in the control of translation. Certain parts of the 5' region may be important for ribosomal binding during the initiation of protein synthesis.

Recent developments in rapid sequencing methodology and molecular cloning techniques have greatly facilitated the determination of the nucleotide sequences of many eukaryotic mRNAs[1,2]. The procedure involves synthesizing a complementary DNA (cDNA) from the mRNA with reverse transcriptase. By cloning the double stranded cDNA into the bacterial plasmid, a large amount of material can be obtained for sequence analysis. However, a drawback of the cloning technique is that variable lengths of 5' sequence are usually lost by S1 nuclease treatment during the preparation of the double stranded cDNA for cloning. One method of determining the 5' untranslated sequence is to analyze the ^{32}P-labeled cDNA of this region using a specific deoxyoligonucleotide primer complementary to the sequence around the initiation region[3-5]. We recently employed a

different approach in which we labeled the 3' end of the single stranded cDNA with ^{32}P and derived the sequence from the labeled end which corresponds to the 5' end of the mRNA. Using this method we obtained the entire 5' untranslated sequence of human α and β-globin mRNAs[6]. We have now applied this technique to determine the 5' untranslated region of human γ-globin mRNA.

MATERIALS AND METHODS

mRNA preparation and cDNA synthesis. RNA was extracted from cord blood reticulocytes and the poly (A)-rich RNA purified by oligo dT cellulose column chromatography as previously described[7]. cDNAs were synthesized from the mRNA with reverse transcriptase, according to the method of Verma et al.[8] and Efstratiadis et al[9]. The 100 µl reaction mixture contained 50 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$, 60 mM NaCl, 10 mM DTT, 400 µM each of dATP, dGTP, and TTP, 100 µM of ^{3}H-dCTP at the specific activity of 2.3 Ci/mmol (diluted from 23 Ci/mmol, New England Nuclear), 100 µg of actinomycin D, 0.06 µg of pT_{10}, and 4 µg of mRNA. Reverse transcriptase was added at 400 units/ml and the reaction mixture was incubated at 37°C for one hour. The cDNAs were extracted with phenol and recovered by ethanol precipitation. The RNA was hydrolyzed by incubation in 0.3 N NaOH for 16 hours at 37°C and the cDNAs purified on a G-75 Sephadex column.

Isolation of 3' end labeled restriction fragment. cDNAs were labeled at their 3' ends with ^{32}P using terminal deoxynucleotidyl transferase (obtained from W. Salser) and (α-^{32}P) -GTP (New England Nuclear, 300 Ci/mmol) as previously described[6]. The labeled cDNAs were digested with the restriction endonuclease Hae III and the DNA restriction fragments were separated by electrophoresis on a polyacrylamide slab gel. The gel containing the 3' end γ-globin cDNA fragment was excised and the cDNA eluted according to the method of Maxam and Gilbert[2].

DNA sequencing. The DNA fragment labeled with ^{32}P at its 3' end was sequenced according to the method of Maxam and Gilbert[2]. Four chemical reactions were used in which cleavage occurs specifically or preferentially at G, A, T or C. The partially degraded products were analyzed on 20% polyacrylamide gel in 7M urea.

RESULTS

Because of their similarity in size, the single stranded α-, β- and γ-globin cDNAs synthesized from cord blood mRNAs were not resolved by electrophoresis on 4% polyacrylamide gel (Fig. 1, lane a). The cDNAs were then digested with the restriction endonuclease Hae III which cleaves single as well as double stranded DNAs at the sequence GGCC[10-12]. Because the distribution of the GGCC sequence differs for each of the cDNAs, three 3' labeled fragments were generated, and these were readily separated by

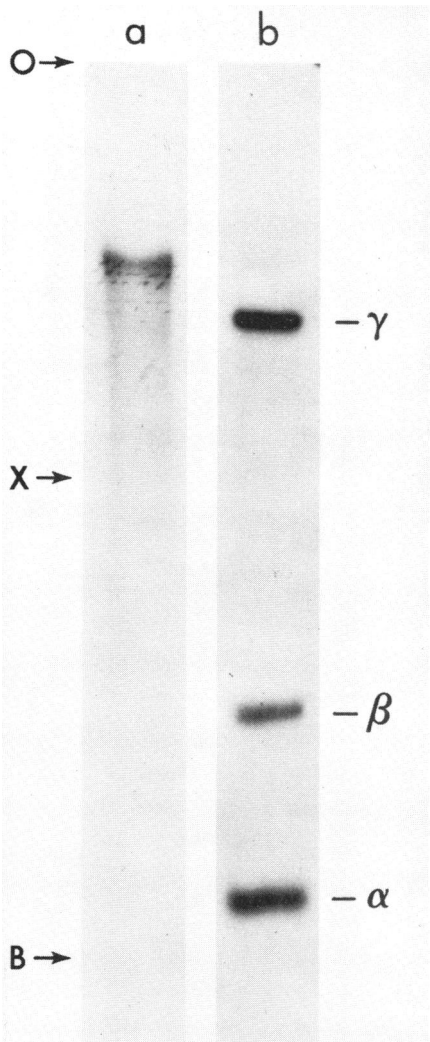

Figure 1. Autoradiograph of 3' end ^{32}P-labeled globin cDNAs on 4% polyacrylamide slab gel. (a) undigested cDNAs prepared from cord blood mRNAs; (b) Hae III digestion of the cDNAs. Arrows indicate positions of origin (O), xylene cyanol blue (X) and bromphenol blue (B) markers.

electrophoresis (Fig. 1, lane b). The two fast migrating fragments, 75 and 132 nucleotides in length, have previously been shown to correspond to the 5' region of human α- and β-globin mRNAs respectively and these sequences have been reported[6]. The slowest migrating fragment of approximately 450 nucleotides, which was present only when cord blood was used for cDNA synthesis, was derived from the γ-mRNA.

The 3' end γ fragment was eluted from the gel and sequenced according to the

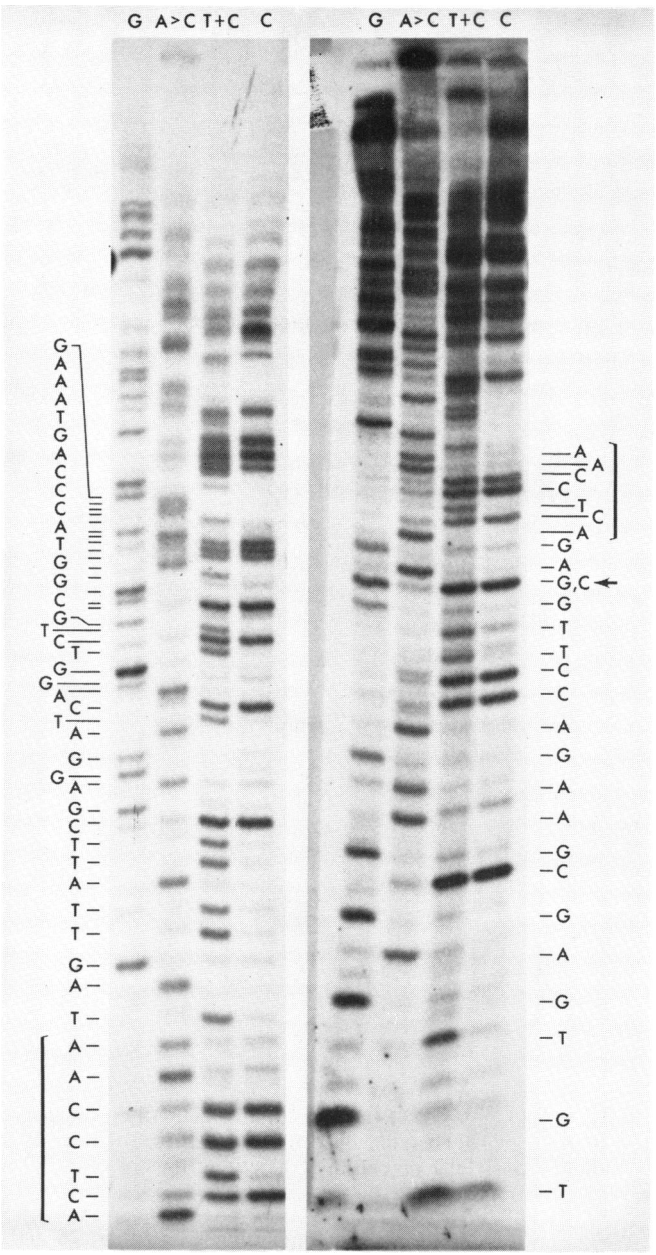

Figure 2. Autoradiograph of the sequencing gel of the 3' terminal Hae III fragment of γ-globin cDNA, labeled at its 3' end with (3',5'-^{32}P) GDP. Four chemical reactions were used in which cleavage occurs specifically at G (G), more frequently at A than C (A > C), at both T and C (T+C), and specifically at C (C). The left set was loaded on the gel 20 hours before the right. Brackets show the region of overlap. The arrow indicates the position where heterogeneity was observed.

method of Maxam and Gilbert (Fig. 2). The mRNA sequence that we derived from this cDNA sequence starts from the first nucleotide following the cap structure and extends to the nucleotides encoding the first 10 amino acids (Table 1). Including the cap and the initiation codon, the 5' noncoding region of the γ-globin mRNA contains 57 nucleotides. Of special note is position number 19 from the cap where both guanosine and cytidine were present (Fig. 2). These two nucleotides were also found at the same position in the cDNAs prepared from two other samples of cord blood. As in the case of the α-and β-globin mRNAs, a stable secondary structure with a hairpin loop with free energy (ΔG, 25^0) of -8.7 Kcal/mol can be constructed from the 5' untranslated region of the γ-globin mRNA (Fig. 3)[13].

TABLE 1 NUCLEOTIDE SEQUENCE OF THE 5' REGION OF HUMAN γ-GLOBIN mRNA

5' (m^7Gppp)-A-C-A-C-U-C-G-C-U-U-C-U-G-G-A-A-C-$\frac{G}{C}$-U-C-U-G-A-G-G-

U-U-A-U-C-A-A-U-A-A-G-C-U-C-C-C-U-A-G-U-C-C-A-G-A-C-G-C-C-A-U-G-G-G-U-
Gly

C-A-U-U-U-C-A-C-A-G-A-G-G-A-G-G-A-C-A-A-G-G-C-U-A-C-U - - - - - - - - - -3'
His Phe Thr Glu Glu Asp Lys Ala Thr

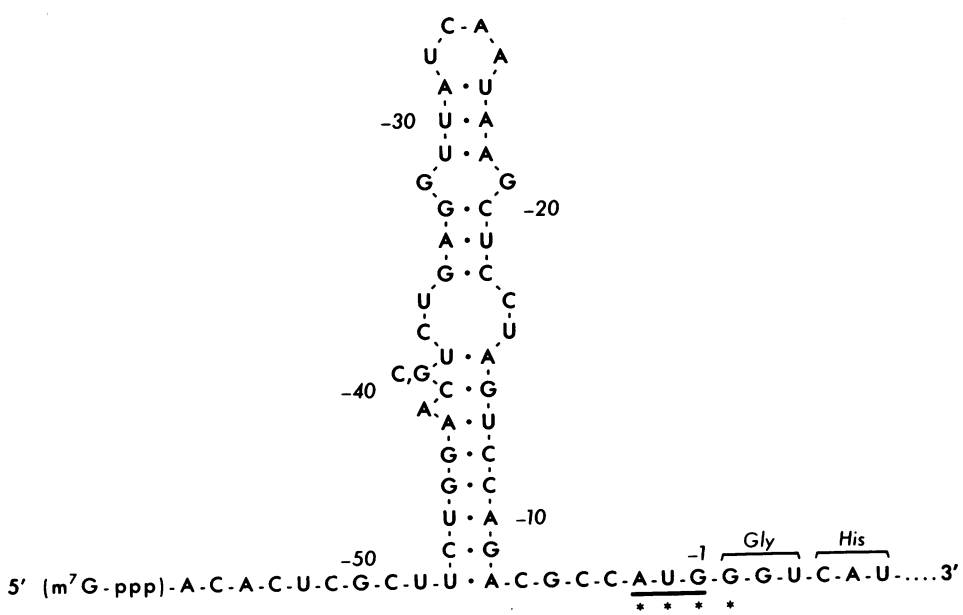

Figure 3. Possible secondary structure of the 5' untranslated region of human γ-globin mRNA.

DISCUSSION

The nucleotide sequence of the entire 5' untranslated region of human γ-globin mRNA has now been determined. This region contains 57 nucleotides as compared to 41 and 54 for human α- and β- mRNAs respectively[6]. In this region, the α sequences differ extensively from the non-α sequences. Comparison of human γ- with human β-and rabbit β-globin mRNA sequences in this same region reveals that the homology between human γ- and either of the two β-globin mRNAs is 60%, while that between the two β's is 80% (Fig. 4). These findings support the theory that γ-globin evolved from the β-globin gene before the divergence of human and rabbit (Fig. 4).

Comparison of the human β- and γ-globin mRNAs shows that the translated region is most highly conserved, with 75% homology between the two sequences. The 5' untranslated regions are less conserved but are still homologous in 60% of their sequences. The 3' untranslated regions show extensive divergence (Poon et al., unpublished). The high degree of homology in the translated regions of these globin genes is most likely due to the functional constraints of the hemoglobin macromolecule. The less divergence in the 5' untranslated regions as compared to the 3' untranslated ends suggests that certain sequences in the former are conserved because they may provide sites for ribosomal binding during the initiation of protein synthesis, as is seen in the prokaryotic system[14,15]. A hexanucleotide sequence of CUUCUG was found near the 5' end of all three human globin mRNAs. Recently, a complementary hexanucleotide sequence of GAAGAC was found 9-14 nucleotides from the 3' termini of several eukaryotic 18S rRNAs[16, 17]. An

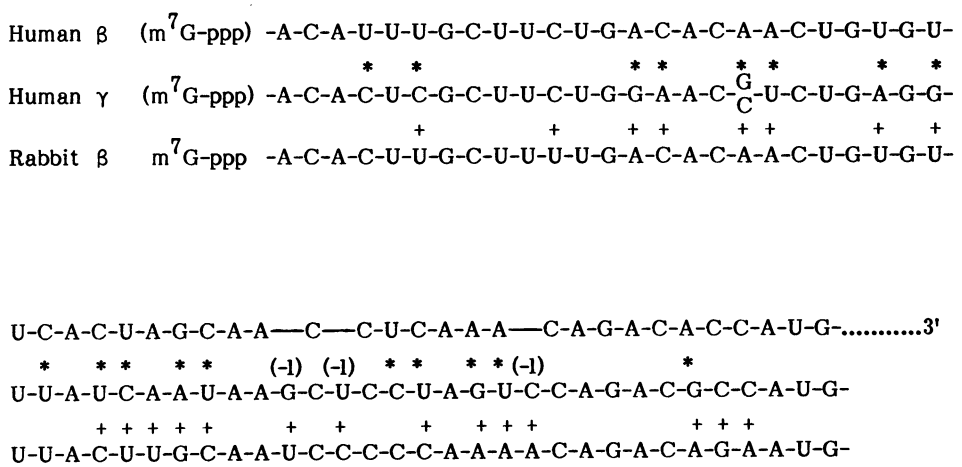

Figure 4. Sequence comparison of 5' untranslated regions of human β-, human γ- and rabbit β-globin mRNAs. (*) denotes differences in nucleotides between human β- and γ-globin mRNAs and (+) indicates the differences between human γ- and rabbit β-; (-1) represents deletion of one nucleotide in that sequence.

alternate ribosomal binding site is the region surrounding the initiation codon AUG, where base pairs can also be formed between the mRNA and the 18S rRNA. In globin mRNAs, the hexanucleotide AAUGGU in rabbit β-, the pentanucleotide AUGGU in rabbit α-, human α- and β-, and the tetranucleotide AUGG in human γ- are complementary to the invarient 3' terminal sequence of 18S rRNAs[18]. However, the exact role of these conserved complementary sequences in the translational process awaits functional analysis of modified or mutant mRNAs.

Polymorphism in nucleotide sequence has been found in different regions of human globin mRNAs. In the β-globin structural region it occurs at the third nucleotide of the codon for amino acid number 50[19]. In the α-globin, Wilson et al. described a possible polymorphism in the untranslated 3' region and suggested that the heterogeneity was due to the two α gene loci[20]. In the α- and β-globin mRNAs, we did not find polymorphism in the 5' untranslated region, although the sequencing technique we used may not detect certain heterogeneity such as the occurence of T with C in the same position. In the γ-globin, we found a polymorphism at the 19th position from the 5' end. Since it was present in three different cord blood samples, we believe that the heterogeneity represents the difference in sequence of the $^{G}\gamma$ and $^{A}\gamma$ loci. However, we have not yet been able to assign either nucleotide to a particular γ locus. This will require sequence analysis of the individual γ-cloned DNAs.

ACKNOWLEDGEMENTS

We thank Ms. Jennifer Gampell for editorial assistance and the Office of Program Resources and Logistics, Viral Cancer Program, Viral Oncology, Division of Cancer Cause & Prevention, National Cancer Institute, Bethesda, Maryland for the reverse transcriptase. This work was supported in part by grants from the NIH (AM 16666, HL 20985), The National Foundation-March of Dimes, and a contract from Maternal and Child Health, Department of Health, State of California. Y.W.K. is an Investigator of the Howard Hughes Medical Institute.

† To whom correspondence should be addressed at San Francisco General Hospital, San Francisco, California, 94110.

REFERENCES

1. Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. 94, 441-448.
2. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
3. Baralle, F. (1977) Cell 10, 549-558.
4. Baralle, F. (1977) Nature 267, 279-281.
5. Baralle, F. (1977) Cell 12, 1085-1095.
6. Chang, J.C., Temple, G.F., Poon, R., Neumann, K.H. and Kan, Y.W. (1977) Proc. Natl. Acad. Sci. USA 74, 5145-5149.

7. Temple, G.F., Chang, J.C. and Kan, Y.W. (1977). Proc. Natl. Acad. Sci. USA 74, 3047-3051.
8. Verma, I.M., Temple, G.F., Fan, H. and Baltimore, D. (1972) Nature New Biol. 235, 163-167.
9. Efstratiadis, A., Maniatis, T., Kafatos, F.C., Jeffrey, A. and Vournakis, J.N. (1975) Cell 4, 367-378.
10. Horiuchi, K. and Zinder, N.D. (1975) Proc. Natl. Acad. Sci. USA 72, 2555-2558.
11. Blakesley, R.N. and Wells, R.D. (1975) Nature 257, 421-422.
12. Godson, G.N. and Roberts, R.J. (1976) Virology 73, 561-567.
13. Tinoco, I., Borer, P.N. Bengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
14. Shine, J. and Dalgano, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342-1346.
15. Steitz, J.A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. USA 72, 4734-4738.
16. Alberty, H., Raba, M. and Grass, H.J. (1977) Nucleic Acids Res. 5, 425-434.
17. Hagenbüchle, O., Santer, M. and Steitz, J.A. (1978) Cell 13, 551-563.
18. Shine, J. and Dalgano, L. (1974) Biochem. J. 141, 609-615.
19. Wilson, J.T., Forget, B.G., Wilson, L.B. and Weissman, S.M. (1977) Science 196, 200-202.
20. Wilson, J.T., deRiel, J.K., Forget, B.G., Marotta, C.A. and Weissman, S.M. (1977) Nucleic Acids Res. 4, 2353-2368.