Five nucleotide changes in the large intervening sequence of a β globin gene in a β^+ thalassemia patient

Sally E.Spence, Robert G.Pergolizzi, Maryann Donovan-Peluso, Katherine A.Kosche, Carl S.Dobkin and Arthur Bank

Columbia University, Department of Medicine, and Department of Human Genetics and Development, Hammer Health Sciences Center, Room 1602, 701 West 168th Street, New York, NY 10032, USA

Received 17 November 1981; Revised and Accepted 22 January 1982

ABSTRACT

A β globin gene from a patient with homozygous β^+ thalassemia has been cloned and completely sequenced. No changes from normal are found in the 200 nucleotides 5' to the cap site, in the 3' untranslated region up to the poly A addition site, in the small intervening sequence (IVS 1), or in the coding sequence except for a third base change in codon 2. The only other differences are in the large intervening sequence (IVS 2). One of these, at a position 16 nucleotides from the 5' end of IVS 2, has been reported previously in normal individuals, and is probably a polymorphism. Four other changes, at positions 74, 81, 666, and 705 are also seen in IVS 2. Abnormal β globin mRNA precursors detected in the bone marrow cells of this patient, and abnormal β globin RNA splicing observed when this gene is transcribed in a tissue culture system taken together with these IVS 2 changes, suggest that the β^+ thalassemia phenotype is produced by a decrease in normal β globin mRNA processing.

INTRODUCTION

The β thalassemias are a heterogeneous group of inherited disorders characterized by decreased or absent synthesis of the β globin chains of hemoglobin A ($\alpha_2\beta_2$) (1, 2). β^0 thalassemia is associated with the absence of β globin, while β^+ thalassemia results in production of decreased amounts of structurally normal β globin. Both β^0 and β^+ thalassemia are associated with severe anemia in the homozygous state. Restriction mapping and direct nucleotide sequencing of cloned β globin genes has shown that β^0 thalassemia encompasses a heterogeneous group of defects. Deletion of the 3' portion of the β globin gene and its flanking region is found in Indian and Pakistani patients (3, 4). However, the vast majority of β^0 thalassemia patients and all β^+ thalassemia patients have no deletions

detectable by restriction enzyme analysis. An analysis of both cDNA and cellular DNA clones from several patients (5-10) has shown that nonsense mutations at codons 17 and 39 can generate a β^0 thalassemia phenotype with premature termination of globin chains. In addition, the β^0 thalassemia phenotype has been associated with a single nucleotide change at the 5' splice junction of the β globin IVS 2 (11, 12). This defect presumably prevents normal removal of IVS 2 and splicing between regions coding for amino acids 104 and 105.

To date, the only reported defect in β^+ thalassemia is a single nucleotide change in IVS 1 in two Cypriot patients (13, 14). In both cases, a G to an A substitution 22 nucleotides from the 3' junction of IVS 1 is found. This change in sequence creates a new splice site which may interfere with normal mRNA processing. We report here five nucleotide changes in the IVS 2 of a patient with a homozygous β^+ thalassemia phenotype.

METHODS AND MATERIALS

Large molecular weight DNA obtained from a bone marrow aspirate of a nine year old female of Italian-American descent was the gift of Drs. Jeffrey Ross and Lynne Maquat. Hematological data are consistent with this patient being homozygous for β^+ thalassemia (15). β globin chains are synthesized at a low level, precluding the possibility that the patient is homozygous for β^0 thalassemia.

DNA was partially digested with Eco RI by incubating 880 μ g of DNA with 220 units of Eco RI (Boehringer Mannheim) for 20 minutes at 37°C in the buffer recommended by the manufacturer. Twenty kilobase (kb) fragments were isolated by sucrose gradient centrifugation and cloned into charon 4A as previously described (16, 17). A recombinant phage containing a cellular DNA insert encoding the entire δ and β globin gene complex was isolated. Various fragments of the β globin gene were isolated from this clone and subcloned into plasmid pBR 322 (Figure 1). Certain subclones were resubcloned into the single stranded bacteriophage M13mp7 (18-20). These subclones were used for DNA sequence analysis by modifications of the dideoxy terminator technique of Sanger and Coulson (21).

Figure 1:

Sequencing scheme of the β globin gene:

The structural β gene is shown by the rectangle on the top line. The solid areas are the coding sequences, and the clear areas IVS 1 and IVS 2. The Pst I (P), Bam HI (B), and Eco RI (E) sites in this portion of the cloned β T2-1 gene are indicated. The 1.8 kb B/B fragment containing the 5' flanking region and 5' end of the gene including IVS 1 was subcloned, and cleaved with Hae III (H) to yield fragments 130, 208, 272, and 73 bp in length which were sequenced. The 0.92 kb B/E fragment containing IVS 2 was subcloned and cleaved with Rsa I (R) and H to yield the fragments indicated. The 0.7 kb E/P fragment shown was isolated from a subclone, and cleaved with H to obtain the nucleotide sequence at the 3' end of the gene and its flanking region.

The following strategy was used to sequence the entire β globin coding and 5' and 3' flanking regions (Figure 1). A 1.8 kb Bam HI fragment containing the 5' flanking sequences, the first and most of the second coding regions, and IVS 1 of β globin was subcloned into pBR 322. After separation from plasmid DNA, the cellular Bam HI fragment was digested with Hae III generating fragments 130, 208, 272, and 73 base pairs (bp) in size representing the region between the 200 bp 5' to the cap site of mRNA and the second coding region (Figure 1). These fragments are suitable for cloning into M13mp7 and are small enough in size to obtain overlapping sequence from clones in both orientations. The 0.92 kb Bam HI/Eco RI fragment containing IVS 2, and small portions of the second and third coding regions of β globin was also subcloned into pBR 322 (Figure 1). After separation of the fragment from plasmid sequences, digestion by Hae III and Rsa I yielded fragments 102, 242, 132, 346, 75, and 18 bp in size (Figure 1). The 4.4 kb Pst I fragment was isolated directly from the charon 4A clone of cellular DNA, digested with Eco RI, and a fragment approximately 700 bp in length purified. This piece was further digested with

Hae III to yield three fragments 63, 101, and >390 bp in size, representing the majority of the third coding region and the 3' flanking region (Figure 1). All fragments with Hae III and/or Rsa I ends were subcloned into M13mp7 double stranded replicative form (RF) digested with Hinc II. Staggered ended fragments were incubated with *E. coli* DNA polymerase I large fragment (Klenow enzyme, New England Biolabs) to generate blunt ends, and subsequently ligated into M13mp7 RF digested with Hinc II.

Single stranded template DNA was isolated from 2 ml liquid culture, annealed to a 26 bp specific primer fragment (Bethesda Research Laboratories), and sequenced. Sequencing reaction products were fractionated by electrophoresis on thin polyacrylamide urea gels, and autoradiographed. All enzymes were obtained from Bethesda Research Laboratories, Incorporated, New England Biolabs, and Boehringer-Mannheim. Deoxy- and dideoxy-nucleoside triphosphates were obtained from PL Biochemicals and deoxyadenosine (α - ^{32}P) triphosphate (2000-3000 Ci/mmol) was purchased from Amersham.

RESULTS

Direct nucleotide sequence analysis of the β globin allele isolated from the genomic DNA of this patient permitted a comparison of this gene to the nucleotide sequence of a normal β globin gene (22). The 200 nucleotides 5' to the mRNA capping site, and 59 nucleotides 3' to the poly A addition site, as well as the 5' and 3' untranslated regions, and the IVS 1 are all identical to normal. In regions of the gene encoding protein (coding sequences) only one change was found. The triplet CAC, which encodes a histidine residue at codon 2, has been replaced by CAT.

Several changes in nucleotide sequence, however, are found in IVS 2 (Figures 2 and 3). One of these changes has been reported previously and is most likely a polymorphism. This alteration occurs 16 nucleotides from the 5' boundary of IVS 2 with a G replacing the C normally found at this position. This change eliminates the recognition site of the restriction endonucleases Asu I and Ava II. This altered Asu I site is

Figure 2:

Nucleotide sequence in the region of IVS 2 at positions including 74 and 81. The Rsa I site is at position 86, therefore, the sequence following this position is that of M13.

- M13 clone used as in text to obtain sequence of coding strand. The G, guanine, A, adenine, T, thymine, and C, cytosine, lanes are indicated at the top and the sequence at the left.
- M13 clone containing the same region of IVS 2, but with the subcloned insert in the opposite orientation, and the sequence of the non-coding strand.

detected by blot hybridization analysis of cellular DNA: A 0.8 kb Asu I fragment hybridizing to an IVS 2 probe is replaced by a 1.0 kb fragment when the polymorphism is present (23). This alteration has been previously reported in both normal and β thalassemia patients (23). Four other nucleotide changes occur at positions 74, 81, 666, and 705 of IVS 2. The change at position 74 is a G to a T, at 81, a C to a T, at 666, a T to a C, and at 705, a T to a G (Figure 3).

DISCUSSION

In this study, we have sequenced the β globin gene from

Figure 3:

Nucleotide changes in the β globin gene:

The gene is shown at the top with the densely shaded regions being the coding sequences, and IVS 1 and IVS 2 indicated.

The sequence changes are at the bottom: Normal and β^+ thalassemia sequences are compared. The change in codon 2 is shown. The position (POS) of the nucleotides from the 5' end of IVS 2 is indicated. The altered nucleotides at positions 16, 74, 81, 666, and 705 are circled.

an Italian-American patient presumed to be homozygous for β^+ thalassemia. Although this patient has a β^+ thalassemia phenotype, it is possible that she is doubly heterozygous for β^0 and β^+ thalassemia. We believe that the gene isolated is a β^+ thalassemia gene, but we cannot exclude the presence of a β^0 thalassemia gene. None of the nucleotide changes previously reported to be the cause of β^0 or β^+ thalassemia are seen in the sequence of this gene. Restriction mapping of both cellular DNA and cloned DNA from the patient reveals no detectable deletion, insertion, or sequence rearrangement in the β globin gene region. Mutations at codons 17 and 39 previously reported to generate premature termination signals, are not present (5-10). With the exception of a silent nucleotide change at codon 2, the coding sequence of the β globin gene is identical to that found normally. IVS 1 of this gene does not contain a false splice junction as is the case in two other β^+ thalassemia genes (13, 14). The normal 5' and 3' splice junctions of IVS 2 are intact in this gene. Sequences at least 200 bp 5' to the β globin gene, some of which have been implicated in initiation of transcription, are normal in this gene. In particular, the CCAAT box (72 bp to the mRNA cap site) and the ATAA box (27 bp 5' to the mRNA cap site) are identical to the normal gene (24). The possibility remains that the defect responsible for decreased β globin synthesis in this patient may be further than 200 bp 5' to the mRNA cap site. At the 3' end of the gene, the AATAAA signal

for poly A addition is present 108 bp 3' to the TAA termination codon, and is followed by the site of poly A addition after a GC pair 24 nucleotides further 3' (25).

Several considerations lead us to believe that the β thalassemia genotype of this gene is due to the changes in nucleotide sequence found in β IVS 2. First, the excessive Accumulation of large molecular weight precursors containing β globin mRNA sequences has been demonstrated in the bone marrow cells of this patient (15, patient 2), as well as others (15, 16). These *in vivo* findings are suggestive of abnormal β globin mRNA processing. However, the extent to which the gene isolated here contributes to this result as compared to the other β thalassemia gene in the patient is unknown. Second, preliminary experiments in this laboratory indicate that abnormal splicing of β globin RNA occurs when the gene described here is assayed in a monkey kidney cell culture system. Either a normal β globin gene or the gene described here has been placed in a plasmid containing pBR 322 sequences and an SV 40 origin of replication (27, 28). The RNA produced by cells transformed with each of these genes has been analyzed by S_1 nuclease mapping (29). An Eco RI-Bam HI IVS 2 DNA fragment 5' end labelled at the Eco RI end has been used as probe. With the normal globin gene, a 54 bp fragment representing spliced or processed β globin RNA, and a 920 bp fragment representing unprocessed β globin RNA precursor are seen. When the β^+ thalassemia gene described here is used, the normal 54 and 920 bp fragments are again seen, but, in addition, a unique 320 bp fragment is also seen. This fragment presumably reflects an alternative 3' splice site within IVS 2 approximately 580 bp from the 5' end of IVS 2.

In an attempt to define potential mechanisms by which the observed changes in IVS 2 in this gene could affect RNA processing, a computer program (30) calculating the most energetically favorable potential secondary structures in β globin mRNA precursors, was used to compare and analyze the normal and this β^+ thalassemia nucleotide sequence. The two conditions of the analysis were that secondary structure be entirely within IVS 2, and the free energy of the hairpin be at least -10 Kcal/mole (31, 32).

Only one such structure was found, involving nucleotides 22 and 82 of IVS 2 (Figure 4). In the normal globin mRNA precursor, the most stable predicted secondary structure within IVS 2 forms a hairpin between positions 22 and 42, and 62 to 82 β IVS 2 with a free energy of -14.0 Kcal/mole (Figure 4). A C to a T transition at position 74 of IVS 2 changes this potential structure slightly, changing the free energy to -13.9 Kcal/mole. The G to the T change at position 81 or IVS 2 alone, on the other hand, affected the free energy markedly, bringing it to -10.5 Kcal/mole. The reasons for this abrupt change can be seen in Figure 4, which shows that the C at position

Figure 4:

Computer predicted stable secondary structure in IVS 2: The hairpin stem begins at a position 22 nucleotides from the 5' end of IVS 2, and ends at position 82. The changes from normal at positions 74 and 81 are indicated. Hydrogen bonding is shown by horizontal lines.

81 is a member of the only GC pair in the base of the hairpin.

When the changes at positions 74 and 81 occur together as in this gene, an additional effect on the potential secondary structure is also seen. Not only is the stability of the structure further reduced, but two alternative structures with similar free energies can form. Either positions 27 to 42 and 52 to 77 can pair with a free energy of -9.3 Kcal/mole, or positions 22 to 40 and 64 to 82 pair with a free energy of -9.5 Kcal/mole.

Several recent findings make it unlikely, however, that the nucleotide changes at position 81 alone, or together with those at positions 16, 74, and 666 of IVS 2 are responsible for the β^+ thalassemia phenotype of this gene. The change at position 16 eliminates a normal Asu I site, and has been found in clinically normal individuals (23). The change at position 74 has been reported in a cloned and sequenced β globin gene from a patient with β^0 thalassemia which also has a nucleotide change at the 5' end of IVS 2 that can explain the thalassemia phenotype (11, 12). A recently sequenced β^0 thalassemia gene contains the changes at positions 16, 74, and 81 identical to those described here, in addition to a two bp deletion and frameshift mutation presumably responsible for the β^0 thalassemia phenotype (9). Further analysis of this gene also shows the change at position 666 (33). In addition, a normal β gene from a Greek subject has been reported to contain the changes at 16, 74, 81, and 666 of IVS 2 (33). These data indicate that these changes are all polymorphisms, although their presence together in a single gene suggest a potential selective advantage for this haplotype.

The nucleotide sequence changes offer no simple explanation for the β^+ thalassemia phenotype or for the aberrant *in vitro* RNA processing of the gene described here. The only unique nucleotide change in this gene is at position 705. This change creates a new potential consensus 3' splice site sequence (AGG from ATG) as well as a sequence similar to the 5' IVS splice site of the β globin gene (AGGT from ATGT). However, it is unclear how the change at 705 affects processing. The abnormal splicing of this gene's transcript in monkey kidney cells does not use a 3' splice site at position 705. The change at 705

may exert its effect on the processing of β globin RNA indirectly, either acting alone or with the other changes in IVS 2 to promote the use of the cryptic 3' splice site at position 580 (there is a good consensus 3' splice sequence at this site). It will be of great interest to see if the change at 705 alone is sufficient to cause the β^+ thalassemia phenotype of this gene, or whether the other IVS 2 changes, particularly at position 81, are also required to produce the defect. Analysis of suitable recombinant genes created by site-directed mutagenesis should provide the necessary data to answer these questions.

The potential effect of nucleotide changes in the IVS on secondary structure and β globin mRNA processing in this β thalassemia gene suggests a general function of IVS. Nucleotide sequences of coding regions of genes most likely evolve to optimize amino acid sequence. At the same time, presumably, stability of mRNA precursors transcribed in a cell is also highly desirable, and in the eukaryotic nucleus, this may require considerable secondary structure. The IVS may provide stability to the secondary structure of transcribed mRNA precursors. In this view, the coding sequences of genes would evolve for optimal protein production without the constraint of also providing optimal secondary structure of mRNA precursors (34). This idea is consistent with the observation that the IVS in β globin genes are highly conserved. The data are also consistent with a potential role for IVS in promoting extensive gene diversity by single point mutations which either create or abolish specific 5' or 3' splice sites. A single base change of this type could result in new splicing patterns in the RNA produced such as that in the gene described here. The mutant genes resulting from these splice site changes might be selected either positively or negatively depending on the effect of the new RNAs produced.

ACKNOWLEDGMENTS

SES and KAK are predoctoral fellows on NIH-GM-07088 grant; RGP is a postdoctoral fellow on NIH-AM-06384 grant; CSD is a Fellow in Cancer Research of the Damon Runyon-Walter Winchell Cancer Fund, DRG-402-FT. This work was also supported by

NIH-AM-25274 grant, New York Community Trust 6-40664 grant, and NIH-CA-31768 grant. We thank Ms. Una T. Collins for preparing this manuscript.

REFERENCES

1. Bank, A., Mears, J. G., and Ramirez, F. (1980) *Science* 207, 486-493.
2. Maniatis, T., Fritsch, E. F., Lauer, J., and Lawn, R. M. (1980) *Ann. Rev. Genet.* 14, 145-178.
3. Orkin, S. H., Kolodner, R., Michaelson, A., and Husson, R. (1980) *Proc. Natl. Acad. Sci.* 77, 3558-3562.
4. Flavell, R. A., Bernard, R., Kooter, J. M., and deBoer, R. (1979) *Nucl. Acids Res.* 6, 2749-2760.
5. Temple, G. F., Chang, T. C., and Kan, Y. W. (1977) *Proc. Natl. Acad. Sci.* 74, 3047-3051.
6. Chang, T. C. and Kan, Y. W. (1979) *Proc. Natl. Acad. Sci.*, 76, 2886-2889.
7. Trecartin, R. F., Liebhaver, S. A., Chang, J. C., Lee, K. Y., Angus, A., Furbetta, M., and Cao, A. (1981) *J. Clin. Invest.* 68, 1012-1017.
8. Moschonas, N., deBoer, E., Grosveld, F. G., Dahl, H. H., Wright, S., Shewmaker, C., and Flavell, R. A. (1981) *Nucl. Acids Res.* 9, 4391-4401.
9. Orkin, S. H. and Goff, S. C. (1981) *J. Biol. Chem.* 256, 9782-9784.
10. Pergolizzi, R. G., Spritz, R. A., Spence, S. E., Goossens, M., Kan, Y. W., and Bank, A. *Nucl. Acids Res.*, in press.
11. Baird, M., Driscoll, M. C., Schreiner, H., Sciarratta, G. V., Sansone, G., Niazi, G., Ramirez, F., and Bank, A. (1981) *Proc. Natl. Acad. Sci.* 78, 4218-4221.
12. Shander, M., Van der Woude, S., Proudfoot, N., and Maniatis, T. (1981) *J. Supramol. Struct. Cell. Biochem. Suppl.* 5, 229 (Abstract).
13. Spritz, R. A., Jagadeeswaran, P., Choudary, P. V., Biro, P. A., Elder, J. T., deRiel, J. K., Manley, J. L., Gefter, M. L., Forget, B. G., and Weissman, S. M. (1981) *Proc. Natl. Acad. Sci.* 78, 2455-2459.
14. Westaway, D. and Williamson, R. (1981) *Nucl. Acids Res.* 9, 1777-1788.
15. Maquat, L. E., Kinniburgh, A., Beach, L. R., Honig, G. R., Lazerson, J., Ershler, W. B., and Ross, J. (1980) *Proc. Natl. Acad. Sci.* 77, 4287-4291.
16. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., and Maniatis, T. (1978) *Cell* 15, 1157-1174.
17. Burns, A. L., Spence, S. E., Kosche, K. A., Ramirez, F., Mears, J. G., Schreiner, H., Miller, C., Baird, M., Leibowitz, D., Giardina, P., Markenson, A., and Bank, A. (1981) *Blood* 57, 140-146.
18. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980) *J. Mol. Biol.* 143, 161-178.
19. Schreier, P. H. and Cortese, R. (1979) *J. Mol. Biol.* 129, 169-172.
20. Messing, J., Crea, R., and Seeburg, P. H. (1981)

- Nucl. Acids Res. 9, 309-315.
21. Sanger, F. and Coulson, A. R. (1978) FEBS Letter 82, 314-317.
 22. Lawn, R. M., Efstratiadis, A., O'Connell, C., and Maniatis, T. (1980) Cell 21, 647-651.
 23. Driscoll, M. C., Baird, M., Rachmilewitz, E. A., and Bank, A. (1981) J. Clin. Invest. 68, 915-919.
 24. Efstratiadis, A., Tosakany, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., deRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulder, C. C., and Proudfoot, N. J. (1980) Cell 21, 653-668.
 25. Proudfoot, N. J. and Brownlee, G. G. (1976) Nature 263, 211-214.
 26. Kantor, J. A., Turner, P. H., and Nienhuis, A. W. (1980) Cell 21, 149-157.
 27. Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T. (1981) Cell 27, 279-288.
 28. Gluzman, Y. (1981) Cell 23, 175-182.
 29. Dobkin, C. and Bank, A. Unpublished observations.
 30. Bush, B. Personal communication.
 31. Borer, P. N., Dengler, B., Tinoco, I., and Uhlenbeck, O. C. (1974) J. Mol. Biol. 86, 843-853.
 32. Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) Nature [New Biol.] 246, 40-41.
 33. Orkin, S. Personal communication.
 34. Leibowitz, D. Personal communication.