Volume 16 Number 11 1988 Nucleic Acids Research

# A novel β thalassemia gene with a single base mutation in the conserved polypyrimidine sequence at the 3′ end of IVS 2

Cherif Beldjord[2,3], Claudine Lapoumeroulie[2], Josee Pagnier[2], M.Benabadji[3], Rajagopal Krishnamoorthy[3], Dominique Labie[2] and Arthur Bank[1]*

[1]Columbia University, College of Physicians and Surgeons, Departments of Medicine, and Genetics and Development, New York, NY, USA

Received February 2, 1988; Revised and Accepted May 5, 1988

ABSTRACT:

An adult Algerian patient with homozygous β thalassemia was found to have a unique β thalassemia gene. Cloning and sequencing revealed that the only abnormality present in this β gene is a transversion in the polypyrimidine stretch at the 3' end of the large intervening sequence (IVS 2) six bases 5' to the consensus AG dinucleotide sequence (CCGCCCACAG instead of CCTCCCACAG). In addition, digestion of the cloned fragment by the enzyme Mnl I demonstrates the disappearance of a restriction site as expected. This is the first example of a defect in the consensus sequence at the 3' end of an IVS leading to a thalassemia phenotype presumably due to decreased splicing.

## INTRODUCTION:

DNA haplotype analysis of the β globin gene cluster has provided a systematic strategy for the identification of novel β thalassemia alleles (1-3). Several new haplotypes have been found during the screening of Algerian β thalassemia patients (4) and a novel mutation associated with one of these haplotypes has been previously reported (5). In this paper, we identify a previously undescribed molecular defect in the DNA of an individual with another unique haplotype designated D (4). This patient is homozygous for β thalassemia and carries a haplotype 1 pattern on the other chromosome. The haplotype 1 was shown to be associated with a $\beta^0$ thalassemia gene due to a previously described one base pair (bp) deletion at codon 6 resulting in a frameshift mutation. The haplotype D associated β gene was cloned and sequenced and found to have a single nucleotide change, a novel T→G substitution, within the conserved polypyrimidine stretch at the 3' end of the second IVS of the β globin gene. Presumably, this mutation diminishes the splicing

© IRL Press Limited, Oxford, England.

of IVS 2 and results in decreased β globin synthesis. This is the first natural mutation to be reported due to a change in the polypyrimidine stretch at the 3' end of an IVS.

## METHODS:

Routine hematologic studies were performed as described (6). DNA preparation and haplotype analysis were performed as previously described (4) using blood drawn in EDTA. The β thalassemia gene carrying haplotype D was cloned in charon 21A as a 7.8 kilobase (kb) Hind III fragment and further subcloned in pBR 322 using standard procedures (7). The cloned insert was cleaved by Pst I, Bam HI, Eco RI, and Sph I (Boehringer Mannheim) and the three resulting fragments subcloned in M13 derivatives - MP 10, 11, 18, and 19 (8). Nucleotide sequence of these fragments covering the entire gene was performed using the dideoxy chain termination procedure of Sanger et al (9).

## RESULTS:

### Clinical Data:

The pedigree of the family is shown in Figure 1. The patient is a 40 year old female with classical features of thalassemia intermedia including anemia without significant requirement for blood transfusions. The hemoglobin level is 9g/dl; HbF 32%; $HbA_2$ 3%; peripheral smear shows abnormal cells with 350 erythroblasts/100 white blood cells; a splenectomy was performed when this patient was 15 years old. The patient has only needed transfusions during her pregnancies.

### Haplotype Analysis:

The patient's DNA was digested with a variety of restriction enzymes and the haplotype determined (Figure 2). The patient was found to be doubly heterozygous for haplotype 1 and a novel haplotype designated D. These two haplotypes differ in the 3' region; haplotype D has the characteristics of framework 3 as determined by restriction fragment length polymorphisms (RFLPs) surrounding the β gene (Figure 2). The cloning of the β thalassemia gene linked to haplotype D was verified by the absence of the Ava II polymorphic site within IVS 2 in this gene as compared to its presence in the gene

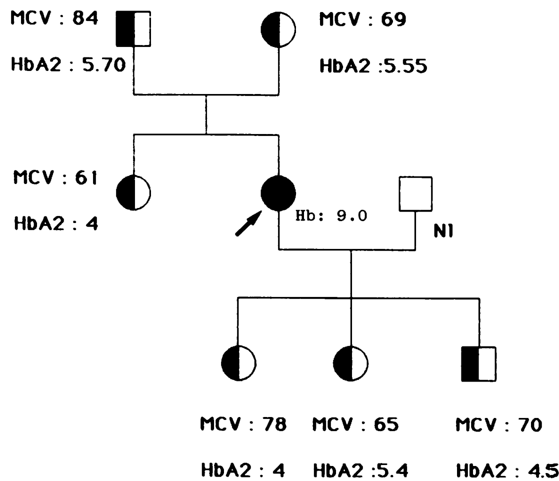



**Figure 1:**
**Haplotype Analysis of the Family Including the Propositus:**
The β thalassemia gene is indicated by filled squares or circles, and the normal gene by clear squares or circles. MCV: mean corpuscular volume; $HbA_2$: Hemoglobin $A_2$ level; Hb: Level of hemoglobin. NL: Indicates that the husband of the propositus had normal hematologic parameters.

present on haplotype 1 (Figure 2). The defect in the gene previously reported associated with framework 1 and haplotype 1 was shown by Mst II digestion to modify the sequence around

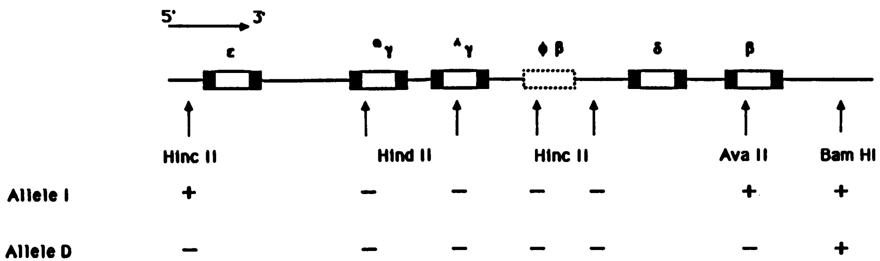



**Figure 2:**
**Haplotype Determination of the Propositus:**
The globin gene cluster is shown from the 5' to the 3' direction at the top. Individual restriction enzymes are indicated and their sites of digestion shown by the vertical arrows. Absence of cleavage: minus (-); presence of cleavage: plus (+). The two alleles of the patients are shown: allele 1, in the text, has been described previously; and allele D, a novel allele, is further described in the text.

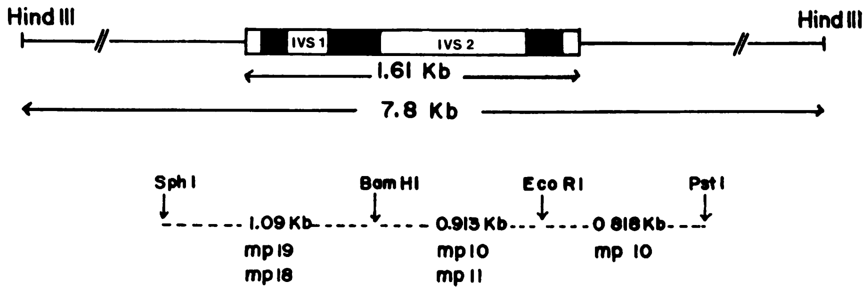



Figure 3:
Strategy for Sequencing the β Thalassemia Gene on Allele D:
A 7.8 kb Hind III fragment subcloned in pBR 322 was further digested with Sph I, Bam HI, Eco RI, and Pst I, as shown, and the three resulting fragments cloned into M13 vectors Mp derivatives as shown.

codon 6 consistent with the presence of a previously described gene (2) with a one bp deletion leading to a frameshift mutation (data not shown).

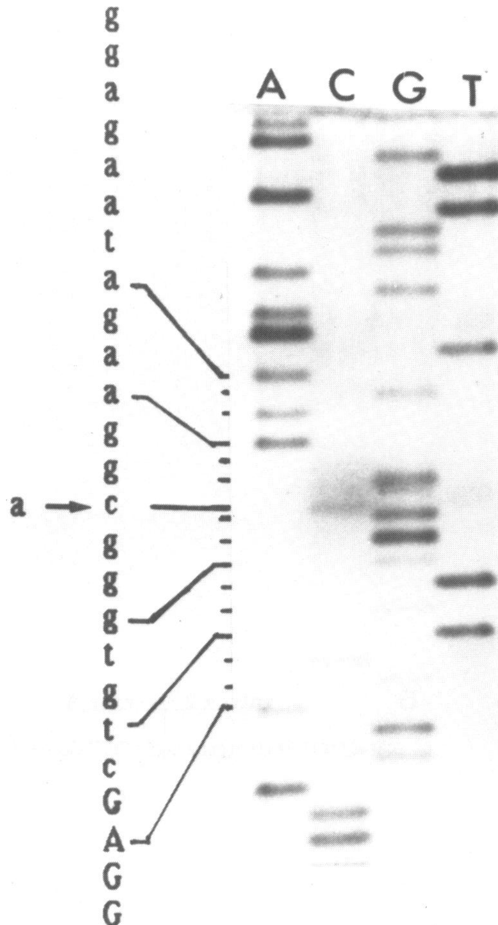
The father with less microcytosis (MCV 84) was found to have haplotype D, while the mother (MCV 69) had haplotype 1 (Figure 1).

Cloning and Sequencing of the Haplotype D β Gene:

The β thalassemia gene carrying haplotype D was isolated after cloning Hind III cleaved cellular DNA from the patient into charon 21A digested with Hind III (7). The cellular DNA Hind III insert from phage containing the unique β gene was subcloned into pBR 322 and subsequently cleaved with Sph I, Bam HI, Eco RI, and Pst I (Figure 3); the three resulting fragments were subcloned into M13 derivatives - Mp 10, 11, 18, and 19 (Figure 3). The nucleotide sequence of these fragments was then performed on both strands. Normal β gene sequence at codon 6 of the cloned gene proved that the mutation associated with haplotype 1 was in the patient's other β gene. Intragenic polymorphisms characteristic of framework 3 and responsible for the absence of Hgia I and Ava II sites were determined. The only other sequence change was a T→G transversion 6 bp upstream of the AG acceptor site of IVS 2 (Figure 4).

Further Characterization of the Mutation:

Examination of the sequence surrounding the mutation




<u>Figure 4:</u>
<u>Radioautogram of Sequencing Gels of Haplotype D Gene:</u>
The nucleotide sequence obtained by dideoxy sequencing in the region of the mutation found is shown. At left is the actual sequence. The A→C substitution in the non-coding strand is demonstrated. In the coding strand, this substitution would be T→G. Nucleotides of the IVS or intron are shown in small letters, and of the exon in capital letters.

revealed the disappearance of an Mnl I recognition site, CCTC, due to the mutation (Figure 5). However, the small size of the fragments generated by Mnl I digestion prevented demonstration of the sequence change using this enzyme in genomic DNA. The subcloned Bam HI/Eco RI fragment of the cloned mutant $\beta$ gene was

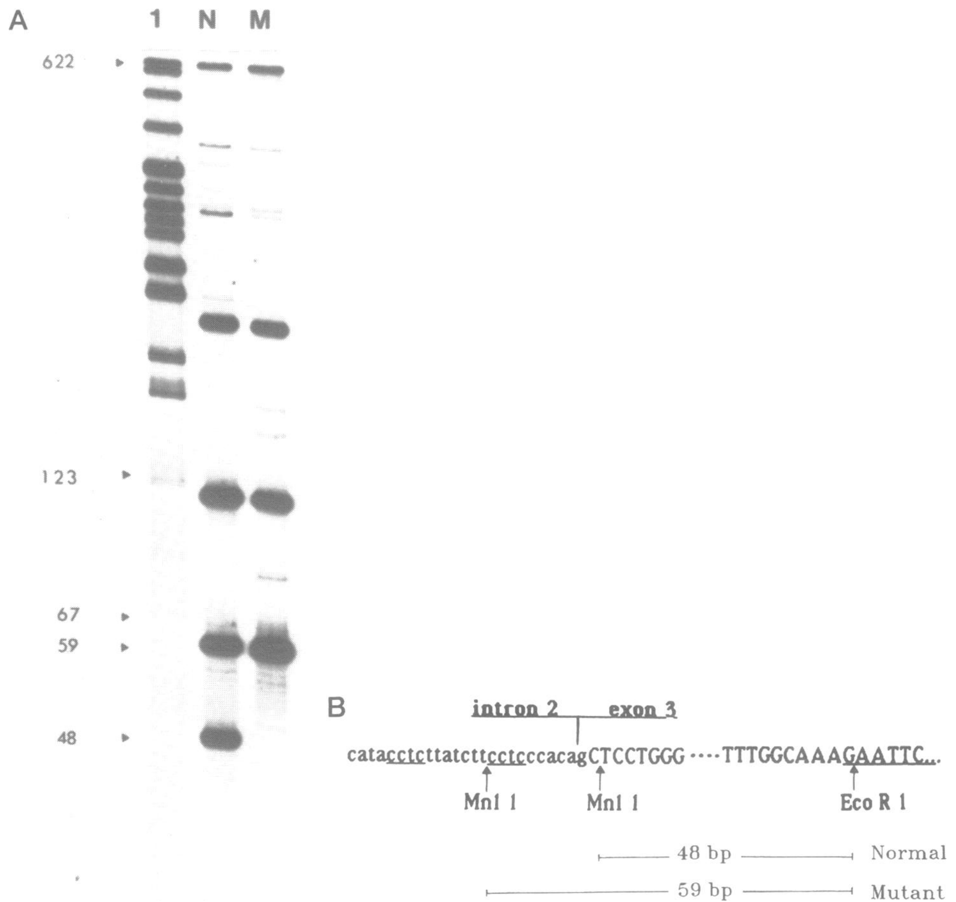



**Figure 5:**
**Mnl I Digestion of Cloned Normal (N) and Mutant (M) Globin Gene Fragments:**
**A. Radioautogram of End-Labelled Mnl I-Digested Cloned DNA:**
The cloned Bam/Eco fragment, containing part of exon 2, the entire IVS 2, and part of exon 3, was isolated and labelled with $P^{32}$ dATP at the Eco RI end by incubation with Klenow reagent. Fragments were then digested with Mnl I, electrophoresed on 8% polyacrylamide gels in 7M urea, and subsequently radioautographed. Lane 1: Represents size markers with fragment sizes indicated at left. The N gene demonstrates 59 and 48 bp fragments in the region of the mutation. The M gene shows only the 59 bp fragment. The extra bands between the 59 and 622 bp fragments were only seen when Mnl I digestion preceded labelling of the Eco RI site, and were absent when the Eco RI labelling was done prior to Mnl I digestion. These bands

are, therefore, presumed to be due to labelling of other DNA fragments, and not partial digestion.

B. The Nucleotide Sequence and Mnl I and Eco RI sites at the 3' End of IVS 2:

IVS 2 sequences are shown in lower case and exon 3 sequences as capitals. The vertical arrows with Mnl I indicate the cleavage sites of Mnl I while the underlined sequence, CCTC, indicates the recognition sequence for Mnl I. The 48 bp Mnl I fragment, generated by the normal gene, and the 59 bp fragment, expected from the mutant gene in which the more 3' CCTC sequence has been changed to CCGC, are shown.

end-labelled with $P^{32}$ at the 3' end and digested with Mnl I; the resulting fragments were separated by electrophoresis through denaturing polyacrylamide gels (acrylamide 8%; urea 7M) and the expected change in the Mnl I fragments generated was shown to result from the single nucleotide change in the mutated gene (Figure 5). Instead of the normal 48 bp Mnl I fragment, only the expected 59 bp Mnl I fragment is detected in mutant DNA. When a similar Bam HI/Eco RI fragment isolated from a normal β globin gene was similarly end-labelled and digested with Mnl I, the normal expected 48 bp Mnl I fragment was seen (Figure 5). However, surprisingly, the abnormal 59 bp fragment was also consistently found (Figure 5). This result is most likely due to the unique cleavage pattern of Mnl I and the close proximity of the two Mnl I sites in this region of the gene. Mnl I gives rise to a cleavage 7 bp downstream from its recognition sequence, and in the region of the mutation, the cleavage site of the 5' Mnl I sequence falls at the 5' end of the second more 3' Mnl I recognition sequence (Figure 5). It is probable, therefore, that if an initial Mnl I cleavage occurs at the more 3' Mnl I site the normal 48 bp fragment is generated. However, if the initial Mnl I cleavage on this fragment occurs at the more 5' Mnl I site, then this cleavage prevents subsequent cleavage at the more 3' Mnl I site and generates the mutant 59 bp fragment.

## DISCUSSION:

Analysis of a β thalassemia gene associated with a unique haplotype in an Algerian patient has identified a unique β gene in an individual with the clinical syndrome of thalassemia

intermedia. Both of the patient's β genes are affected since the patient has an anemia and is more severely affected than β thalassemia heterozygotes. Cloning and sequencing the entire novel β globin gene has shown that the only change present in this gene is in the conserved polypyrimidine stretch of IVS 2 at a position 6 bases upstream of the AG dinucleotide acceptor splice site. The other β thalassemia gene in this patient is a $\beta^0$ thalassemia gene due to a frameshift mutation at codon 6. Therefore, the β globin produced in this patient must be due to the activity of the gene with the defect in IVS 2.

$\beta^0$ thalassemia genes have been described in which the invariant AG at the 3' end of IVS 1 and IVS 2 has been mutated (10-12). The presence of these genes leads to the complete absence of splicing of the IVS and the absence of mature β mRNA. Thus, the invariant AG at the 3' end of IVS in the β globin gene appears to be required for normal splicing. By contrast, the effect of modifying the consensus polypyrimidine sequence at position -6 3' to the AG on splicing efficiency is less well documented. Although the majority of genes examined have a pyrimidine in this position, there are several genes that have an adenosine here including some immunoglobulin genes that are expressed at high levels (13).

The most likely explanation for the decreased β globin expression due to the presence of the abnormal $\beta^+$ thalassemia gene described here is that the consensus sequence affected by this mutation is required for optimal splicing of IVS 2 (14, 15). This represents the first in vivo demonstration of a role for this consensus sequence in optimal splicing of the β globin gene and further indicates the complexity of the intragenic DNA sequences required for optimal β globin gene expression.

ACKNOWLEDGMENTS:
We wish to thank Oona Collins for editorial assistance. This work was supported by grants from INSERM to Unite 15, E. E. C., and TSD M-1313 (D. L.); the Fondation pour la Recherche Medicale Francaise and the Philipe Foundation (C. L.); Public Health Service Grants DK-25274 and HL-37069 from the National Institutes of Health and a grant (6-426) from the March of Dimes Birth Defects Foundation (A. B.).

[2]Institut National de la Santé et de la Recherche Médicale, Unité de Recherches en Pathologie Moleculaire, Institut de Pathologie Moleculaire, Faculte de Médecine Cochin Port-Royal, Paris, France and [3]Centre de Transfusion Sanguine Algiers, Hospital Mustapha, Algiers, Algeria

*To whom correspondence should be addressed

**REFERENCES:**

1. Antonarakis, S. E., Kazazian, H. H., and Orkin, S. H. (1985) Hum. Genet. 69, 1-14.
2. Orkin, S. H., Kazazian, H. H., Antonarakis, S. E., Goff, S. C., Boehm, C. D., Sexton, J. P., Waber, P. G., and Giardina, P. J. V. (1982) Nature 296, 627-631.
3. Kazazian, H. H., Orkin, S. H., Antonarakis, S. E., Sexton, J. P., Boehm, C. D., Goff, S. C., and Waber, P. G. (1984) EMBO J. 3, 593-595.
4. Beldjord, C., Lapoumeroulie, C., Baird, M., Girot, R., Adjrad, L., Lenoir, G., Benabadji, M., and Labie, D. (1983) Hum. Genet. 65, 204-206.
5. Lapoumeroulie, C., Pagnier, J., Bank, A., Labie, D., and Krishnamoorthy, R. (1986) Biochem. Biophys. Res. Commun. 139, 562-566.
6. Wintrobe, M. M. (1981) Clinical Hematology, eighth ed., Lea & Febiger Publishers, Philadelphia.
7. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A laboratory Manual, Cold Spring Harbor Laboratories, New York.
8. Messing, J. (1983) Methods in Enzymology, vol. 101, Academic Press, New York, pp. 20-78.
9. Sanger, F., Nicklen, S., and Coulson, R. (1977) Proc. Natl. Acad. Sci. 74, 5463-5467.
10. Antonarakis, S. E., Orkin, S. H., Chang, T., Scott, A. F., Sexton, J. P., Trusko, S. P., Charache, S., and Kazazian, H. H. (1984) Proc. Natl. Acad. Sci. 81, 1154-1158.
11. Atweh, G. F., Anagnou, N. P., Shearin, J., Forget, B. G., and Kaufman, R. E. (1985) Nucl. Acids Res. 13, 777-790.
12. Orkin, S. H. (1987) in The Molecular Basis of Blood Diseases, Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P., and Majerus, P. W., Eds., pp. 106-126, W. B. Saunders, Philadelphia.
13. Mount, S. M. (1982) Nucl. Acids Res. 10, 459-472.
14. Padgett, R. E. et al (1986) Annu. Rev. Biochem. 55, 1119-1150.
15. Green, M. R. (1986) Annu. Rev. Biochem. 20, 671-708.