Mutation in an intervening sequence splice junction in man
(α-thalassemia/gene cloning/DNA sequence analysis/RNA processing/stable RNAs)

STUART H. ORKIN, SABRA C. GOFF, AND RACHEL L. HECHTMAN

Division of Hematology and Oncology, Children’s Hospital Medical Center and the Sidney Farber Cancer Institute, and the Department of Pediatrics, Harvard Medical School, Boston Massachusetts 02115

Communicated by David Baltimore, April 24, 1981

ABSTRACT The α2-globin gene of an individual with α-thalassemia associated with the absence of α2 mRNA was cloned in bacteriophage. This mutant globin gene was normally active in transcription in vitro. The DNA sequence of the gene, however, revealed a pentanucleotide deletion within the 5' splice junction of the first intervening sequence. Following the G of the invariant G-T dinucleotide normally located within such junctions, a deletion of T-G-A-G-G was found. No other sequence abnormalities within the mutant gene were present. We speculate therefore that this deletion within the splice junction is the primary genetic defect in this individual with thalassemia and that loss of a functional splice junction results in failure of stable mRNA formation.

Most eukaryotic genes are mosaic structures assembled from coding and intervening sequences (1, 2). In the case of mammalian globin genes two intervening sequences, situated in homologous positions in α-globin-like and β-globin-like loci, are normally present (3, 4). Within the nucleus of developing erythroid cells these sequences are transcribed as part of globin mRNA precursors that are subsequently processed into mature mRNA by a series of cleavage–ligation reactions (splicing events) (5–7). The nucleotide sequences of numerous cloned eukaryotic genes reveal striking similarities at the boundaries of coding and intervening sequences (splice junctions). These may be accommodated by consensus sequences of the general type A-G-G-T-A-A-G for the 5’-donor sites and T-T-N-C-A-G-G for the 3’-acceptor sites (8, 9). Almost without exception, intervening sequences have the dinucleotide G-T at their 5’t ends and A-G at their 3’t ends [the Chambron rule (8)]. Participation of ubiquitous nuclear U1 RNAs in appropriate recognition of these sites is suggested by their complementarity with sequences found at these splice junctions (10, 11). Although the effects of specific mutations in these junctional sequences have not been studied to date, analyses of specially constructed viral DNA mutants and recombinants, often bearing globin gene sequences and their splice junctions, have suggested a requirement for at least one splicing event in the formation of stable cytoplasmic mRNAs (12–15).

Genetic disorders of human hemoglobin synthesis, the thalassemias (16), provide a reservoir of naturally occurring mutant genes that may facilitate study of these complex cellular events. For example, in β-thalassemias, conditions in which mature β-globin mRNA and β-globin polypeptide are quantitatively reduced, the processing of β-globin RNA precursors in several patients appears inefficient (17, 18). In selected cases, abnormal β-globin mRNAs that may resemble processing intermediates are detectable (18). Furthermore, in one instance, nucleotide sequencing of a β-globin gene isolated from an individual with β-thalassemia identified a single base substitution within the body of the first intervening sequence (IVS-1) (19). This change could generate a new splicing signal and thereby alter the usual pathway of β-globin RNA processing.

Genetically defective globin genes that direct the production of no stable mRNA species may be especially useful in exploration of mRNA biogenesis. Our recent progress in defining the expression of the human α-globin genes (20) suggested for study a mutant locus with the desired phenotype. Whereas normal humans have stable RNA directed by both the 5’(α2) and 3’(α1) structural genes of the duplicated α-gene complex (20–22), some individuals with α-thalassemia have a profound (if not total) deficiency of α2 mRNA in the face of a grossly normal α-gene region as defined by restriction enzyme mapping (23). We have cloned the putative mutant α2-globin gene from one such individual, and we demonstrate here that it is specifically altered at a coding region–intervening sequence boundary. This lesion, a naturally occurring mutation of a splice junction, raises questions regarding the processing of precursor mRNAs and the generation of stable mRNAs.

MATERIALS AND METHODS

Patient Material. Lymphocyte DNA of an Italian patient with α-thalassemia of the Hb H disease type was prepared as before (25). By restriction mapping with EcoRI digestion, this individual has the genotype 22.5 kb/2.6 kb (kb, kilobase) (23). On one chromosome, two α loci (one of the α2 and one of the α1 type) are present in the normal arrangement. On the other chromosome, a partially deleted α1 gene is found (24).

Cloning of the α-Globin Gene Complex. The National Institutes of Health guidelines for recombinant DNA research were followed. DNA fragments greater than 9 kb in length after digestion with EcoRI and BamHI were prepared by sucrose gradient centrifugation and cloned in phage arms isolated from λ bacteriophage Charon 30 (25). Recombinant phages were screened by the Benton–Davis procedure (26), using probe prepared from a 1.6 kb fragment isolated from normal α gene subclones (22) after Pst I digestion. From a positive phage containing the desired 14-kb EcoRI/BamHI fragment (21, 22), the α2 gene was removed from the phage DNA in a 4.3 kb fragment after Sac I digestion and then subcloned after Pst I digestion in plasmid pBR322.

In Vitro Transcription. Assays were performed precisely as described by Manley et al. (27). DNA Sequence Analysis. DNA nucleotide sequence was determined by the method of Maxam and Gilbert (28, 29). The strategy was that previously used by us in the analysis of an α1 gene (30). Critical regions were examined on complementary strands.

Gene Mapping. Restriction enzyme analyses of cellular DNA were performed as before (31) except dextran sulfate was added to the hybridization mixtures (32). DNA fragments greater than

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: IVS, intervening sequence; kb, kilobase(s).
10 kb after BamHI digestion were prepared prior to digestion with Hph I. This procedure eliminates the other allele [the 2.6-kb EcoRI fragment (24)] from the thalassemic sample and thereby permits direct examination of the chromosome region of interest.

S1 Nuclease Mapping of RNA. RNA mapping using the Weaver–Weissmann modification (33) of the Berk–Sharp procedure (34) was performed as described (20). Two probes spanning the 5’ portion of the normal α2 gene were utilized. A 5’-labeled fragment extended from the coding region HindIII site (codons 90/91) to the HindI site in the 5’-untranslated region. A 3’-labeled probe extended from this HindI site to the HindIII site.

RESULTS

The duplicated α-globin gene complex was cloned intact as a 14-kb EcoRI/BamHI fragment from the DNA of an Italian individual with “nondeletion” α-thalassemia (23, 35). In this entity a normal-appearing α-globin gene complex is present, as determined by restriction enzyme mapping, yet hematologic findings indicate that α mRNA sufficient for only a single α gene is found in erythroid cells. In the patient studied here S1 nuclease mapping previously demonstrated that the abundance of α2 mRNA was less than 3% that of α1-mRNA in peripheral blood erythroid cells (20). Because this assay revealed a specific deficiency of α2 mRNA, we focused on the function and structure of the cloned α2 gene.

In Vitro Transcription. We considered defects in gene transcription or in processing of precursor mRNA as most likely to account for the virtual absence of α2 mRNA. Although assays capable of testing all aspects of tissue-specific gene expression have yet to be developed, functional promoters can now be assessed in vitro. Cell-free systems permit the synthesis of RNA molecules initiated at their normal positions and terminated at sites determined by restriction enzyme cleavage of cloned DNA (27, 36). Transcription is dependent on DNA sequences 5’ to the gene, especially within the T-A-T-A-A-G box (4, 37, 38), and is markedly reduced from nonfunctional pseudogenes containing altered sequences in that region (4, 39). Cloned normal human α-globin genes are active in vitro in the transcription system of Manley et al. (4, 27). The cloned α2 gene isolated from the α-thalassemic individual was indistinguishable from normal in its capacity to direct synthesis of appropriate RNA species in vitro (Fig. 1). We conclude that this gene contains a promoter active in the heterologous system. In view of this result, we directed attention to the body of the gene itself.

DNA Sequence of the Mutant α2-Globin Gene. The nucleotide sequence of the presumed mutant α2 gene is displayed in Fig. 2. For comparison we employed the sequence of a normal α2-globin gene reported by Liebhaber et al. (40). In addition, specific discrepancies between their sequence and our mutant gene were further analyzed in our work by direct sequence analysis of the normal α2 gene cloned by Lauer et al. (22). The only sequence abnormality of the mutant gene resided at the beginning of the first intervening sequence (IVS-1). Following the G of the invariant G-T contained within the donor splice site, a pentanucleotide deletion (T-G-A-G-C) was found (Figs. 2–4). In the normal sequence this region is particularly rich in short direct repeats (Fig. 4) that appear to contribute to the production of limited deletions (41). The pentanucleotide deletion in the mutant gene radically altered the junction of the coding region and IVS-1. The resulting junctional sequences do not conform to the Chambon rule (8) or to consensus sequences of known donor splice sites (8, 9).

We carefully examined the entire mutant gene for additional sequence abnormalities. None were found. Several apparent differences between the mutant gene sequence and the α2 sequence reported by Liebhaber et al. (40) near the end of IVS-2 (see Fig. 2 legend) were explored by reevaluation of the normal sequence. The sequences of this region of the mutant gene and the normal α2 gene of Lauer et al. (22) were identical. The differences between our sequence and that of Liebhaber et al. (40) in this region most likely reflect low frequency sequence errors in their analysis or perhaps polymorphism in this region of the α2 gene. Two third position codon differences (codons 54 and 123) between our sequences are also evident. Because we previously identified the codons found in the mutant gene in the normal α1 gene, we have not investigated this potential heterogeneity further (30).

Gene Mapping of Uncloned Thalassemic DNA. Rearrangements within cloned DNAs have been observed in phage propagated in Escherichia coli. For example, propagation of the entire α gene complex in bacteria leads to deletion events similar to those seen in deletion forms of α-thalassemia in which a struc-
molecular heterogeneity. In an individual of Asian background with nondeletion α-thalassemia, also associated with deficiency of α mRNA, the hybridization pattern after digestion with Hph I was entirely normal (not shown). We conclude that a different molecular defect, either within the α2 gene or affecting its expression, exists in the DNA of this unrelated individual. The molecular basis of this condition is heterogeneous.

Absence of Abnormal α2 mRNA Species. Previously a probe directed at the 3'-untranslated region of the α gene was used to distinguish and quantitate the RNA products of the
thalassemia. In an individual of Asian background with nondeletion α-thalassemia, also associated with deficiency of α mRNA, the hybridization pattern after digestion with Hph I was entirely normal (not shown). We conclude that a different molecular defect, either within the α2 gene or affecting its expression, exists in the DNA of this unrelated individual. The molecular basis of this condition is heterogeneous.

Absence of Abnormal α2 mRNA Species. Previously a probe directed at the 3'-untranslated region of the α gene was used to distinguish and quantitate the RNA products of the
talassemia. In an individual of Asian background with nondeletion α-thalassemia, also associated with deficiency of α mRNA, the hybridization pattern after digestion with Hph I was entirely normal (not shown). We conclude that a different molecular defect, either within the α2 gene or affecting its expression, exists in the DNA of this unrelated individual. The molecular basis of this condition is heterogeneous.

Absence of Abnormal α2 mRNA Species. Previously a probe directed at the 3'-untranslated region of the α gene was used to distinguish and quantitate the RNA products of the

periments in which the 3' untranslated portion of the mRNA was used as a marker for α2 mRNA, we conclude that no α2 mRNA, either unspliced or abnormally spliced at IVS-1, was present in the stable erythroid cell RNA of the patient.

**DISCUSSION**

We have observed a pentanucleotide deletion at a coding-intervening sequence splice junction of an α-globin gene isolated from an individual with α-thalassemia. Because this gene was selected for study on the basis of a profound deficiency of α2 mRNA in the patient’s erythroid cells, it is reasonable to speculate that we have localized the primary genetic defect in this instance and that it leads to the observed phenotype.

The consequences of obliteration of a single splice junction within a gene containing more than one intervening sequence have not been previously explored. Loss of a functional splice junction might lead to one of three possibilities. First, as prior experiments have suggested a minimal requirement for a single splicing event in the production of stable mRNAs (12–15), partially spliced yet stable RNA might be generated upon transcription of this mutant gene and subsequent processing of IVS-2. Second, an abnormal splicing event mediated by sequences 5' to the normal splice junction of IVS-1 might occur if upstream sequences resembled authentic donor sites. This could lead to the formation of abnormally spliced, but stable, RNA. Such RNA resulting from a "forbidden splice" has been observed for an immunoglobulin light chain gene missing an RNA splice signal due to abnormal DNA recombination (44). Third, the splice junction abnormality might entirely prevent generation of stable RNA products. The absence of α2 mRNA detectable by S1 nuclease mapping in our patient’s RNA makes this last possibility quite likely. This suggests either a preferred order of splicing events, such as processing observed prior to IVS-2 as reported for β mRNA precursors (7), or the existence of cellular mechanisms capable of recognition and destruction of aberrant RNAs. Instability of RNA transcripts from a mutant gene such as that described here could be manifest within the nucleus of developing erythroid cells, within their cytoplasm, or within circulating peripheral blood reticulocytes. Our findings suggest that alteration of the 5' boundary of IVS-1, at least in the case of the α-gene transcript, ultimately leads to a failure of stable RNA formation. Heterologous systems are now available that will permit a direct test of this conclusion (12–15). Formally, we have not excluded the possibility that the mutant gene is transcriptionally silent in vivo. This might result from the presence of additional mutations located upstream from the α2 gene, such as those found in pseudogenes (4, 39). Our in vitro transcription experiments demonstrate the presence of a functional promoter element in the cloned DNA but do not prove in vivo activity, because tissue-specific gene expression cannot as yet be reconstituted in vitro. Demonstration of processing abnormalities of RNA transcripts generated in heterologous systems will ultimately provide the most persuasive argument against the existence of mutations elsewhere in our patient’s DNA responsible for the phenotype observed here.

Studies of others involving cloned β-globin genes from β-thalassemic individuals lend further support to the concept that abnormalities within one intervening sequence may lead to a quantitative deficiency of RNA, such as that characteristic of thalassemia syndromes (45, 46). In the β-globin gene of a β'-thalassemic individual a single nucleotide substitution within the body of IVS-1 has been detected (19). In another gene isolated from a patient with β-thalassemia a substitution for the G of the G-T dinucleotide at the beginning of IVS-2 was observed (T. Maniatis, personal communication). In contrast with the le-
sion described here, these single base substitutions in intervening sequences may alter the efficiency of precursor mRNA processing more than its fidelity, as suggested by a mutation of an intervening sequence junction in a yeast tRNA gene (47). Taken together, the demonstration of different intervening sequence mutations among individuals with thalassemia syndromes provides encouragement that a spectrum of lesions may soon be identified. Correlation of specific gene defects with their RNA phenotypes in vivo, as well as analysis of RNA processing in heterologous systems, may define specific regions of critical importance in the normal processing of RNA transcripts and cellular pathways for metabolism of abnormal RNA products.

Note Added in Proof. Blot analysis using Hph I digests has provided additional evidence for molecular heterogeneity in “nondeletion” α-thalassemia. Of two Turkish individuals with the nondeletion complex (23), one had the abnormality reported here, and the other a normal mapping pattern.

We thank James Manley for patiently instructing us in his in vitro transcription system, and Tom Maniatis and Joyce Lauer for initially providing their cloned normal α2 gene. We appreciate helpful discussions with Alan Michelson and the continued support and encouragement of David G. Nathan. S.H.O. is supported by grants from the National Foundation—March of Dimes and the National Institutes of Health and a Research Career Development Award of the National Institutes of Health.