The EMBO Journal vol.4 no.5 pp.1245-1250, 1985

α-Thalassaemia associated with the deletion of two nucleotides at position −2 and −3 preceding the AUG codon

F.Morlé, B.Lopez, T.Henni[1] and J.Godet

Laboratoire de Biologie Cellulaire U.A. 92 CNRS, Département de Biologie Générale et Appliquée, Université Claude Bernard Lyon-I, 43, boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France, and [1]Centre Pierre et Marie Curie, Hôpital Mustapha, Alger, Algeria

Communicated by J.L.Mandel

The nucleotide sequence of three single α-globin genes resulting from a rightward 3.7-kb deletion is described. The α genes were isolated from the DNA of three subjects homozygous for this deletion, the first being in addition homozygous for the structural mutation α^{G} Philadelphia (genotype $-\alpha^{G}/-\alpha^{G}$), the second, heterozygous for this structural mutation (genotype $-\alpha^{A}/-\alpha^{G}$) and the third homozygous for an α^{+}-thalassaemic gene (genotype $-\alpha^{+thal}/-\alpha^{+thal}$). The latter subject produced HbH in contrast to the two others. Whereas the two α^{A} and α^{G} genes are identical to the normal α1-globin gene (except for the α^{G} point mutation), the α^{+thal} gene has (i) a deletion of the two nucleotides at position −2 and −3 preceding the ATG codon, and (ii) a fusion between the 5′ part of the normal α2 gene and the 3′ part of the normal α1 gene. Using a dot-blot assay, we show that reticulocytes from the HbH subject contain at least as much αmRNA as reticulocytes from the two other subjects. In a transient expression system, the α^{+thal} gene leads to normally spliced transcripts. We conclude from these data that the defective output of α chains by the α^{+thal} gene, as evidenced by HbH production, results from a decreased efficiency of α-mRNA translation due to the two nucleotides deletion preceding the AUG codon.

Key words: α-thalassaemia/3.7-kb rightward deletion/translation efficiency

Introduction

Human α-globin is encoded by two non-allelic genes, α2 and α1, embedded within two adjacent homologous segments Zα1 and Zα2, respectively (Lauer *et al.*, 1980; Michelson and Orkin, 1983), and located on chromosome 16 (Deisseroth *et al.*, 1977). Unequal crossing-over between the α1 and α2 homologous segments leads to a rightward 3.7-kb deletion and generates chromosomes carrying a single α gene (Embury *et al.*, 1980; Lauer *et al.*, 1980; Lie-Injo *et al.*, 1981; Michelson and Orkin, 1983; Higgs *et al.*, 1984). Homozygosity for such a deletion usually has no clinical implications and is characterised by a moderate imbalance in globin chain synthesis without production of HbH (Milner and Huisman, 1976; Dozy *et al.*, 1979; Mathew *et al.*, 1983; Morlé *et al.*, 1984). HbH (i.e., β-chain tetramers) results from a marked imbalance in globin chain synthesis and is usually produced by subjects carrying only one functional α-globin gene (Weatherall and Clegg, 1981). We have, however, previously reported an Algerian patient who was homozygous for a rightward 3.7-kb deletion and who displayed HbH disease (Whitelaw *et al.*, 1980; Tabone *et al.*, 1981; Morlé *et al.*, 1984). This led us to conclude that the single α gene present on each of his chromosomes was carrying an α^{+}-thalassaemic determinant.

To identify this α^{+}-thalassaemic determinant, we compared the nucleotide sequence of the α^{+}-thalassaemic gene isolated from the subject with HbH (genotype $-\alpha^{+thal}/-\alpha^{+thal}$) with that of two other single α genes isolated from two Algerian subjects who were also homozygous for a 3.7-kb deletion but who produced no HbH. One of these subjects was in addition homozygous for the structural mutation α^{G} Philadelphia (genotype $-\alpha^{G}/-\alpha^{G}$), the other was heterozygous for this structural mutation (genotype $-\alpha^{A}/-\alpha^{G}$). We also compared the steady-state level of α-mRNA in reticulocytes from these three subjects and analysed splicing of transcripts from the α^{+thal} gene in a transient expression assay. We conclude from these analyses that the defective output of α chains from the α^{+thal} gene results from a decrease in α-mRNA translation initiation due to deletion of two nucleotides at position −2 and −3 preceding the AUG codon.

Results

Nucleotide sequences of the three single α genes resulting from a rightward 3.7-kb deletion

The three single α genes α^{A}, α^{G} and α^{+thal} were isolated by cloning 16-kb *Bgl*II-cleaved α-gene fragments in λEMBL 4 vector (Frischauf *et al.*, 1983). 1.5-kb *Pst*I-cleaved α-gene fragments carrying the entire α-gene sequence were further subcloned in pEMBL 8 plasmid (Dente *et al.*, 1983). Nucleotide sequences of the three single α genes were determined from nucleotide −90, 5′ to the cap site, to the polyadenylation site. Results, in Figure 1, show that the nucleotide sequence of the α^{A} gene is identical to that of the normal α1 gene (Michelson and Orkin, 1983). The nucleotide sequence of the α^{G} gene is also identical to that of the normal α1 gene except for the nucleotide at position 361 which is G instead of C, reflecting the α^{G} Philadelphia mutation (Asn→Lys in codon 68). In contrast, the nucleotide sequence of the α^{+thal} gene differs in two ways from that of the normal α1 gene: (i) nucleotides at position −2 and −3 preceding the ATG translation initiation codon are deleted, and (ii) the 5′ end of the gene, from −90 to the end of exon 3, is identical (excepting the two-nucleotide deletion) to the normal α2 gene whereas the 3′ end of the gene, from exon 3 to the polyadenylation site, is identical to the normal α1 gene.

The two-nucleotide deletion preceding the ATG codon was confirmed by primer extension analysis performed on reticulocyte α-mRNA. Using a single-stranded end-labeled primer, complementary to a 38-nucleotide sequence in exon 1, we synthesised α-cDNA with reticulocyte RNA from a normal subject and from the HbH subject (Figure 2). As expected, the α-cDNA synthesised from the HbH subject RNA is two nucleotides shorter than that synthesised from the normal subject RNA, thus excluding a cloning artefact in the 5′ end of the α^{+thal} gene sequence. The sequence in the IVS 2 of the α^{+thal} gene was confirmed by *Dde*I and *Apa*I restriction mapping (data not shown).

The sequence data thus show that the α^{+thal} gene is a fusion-

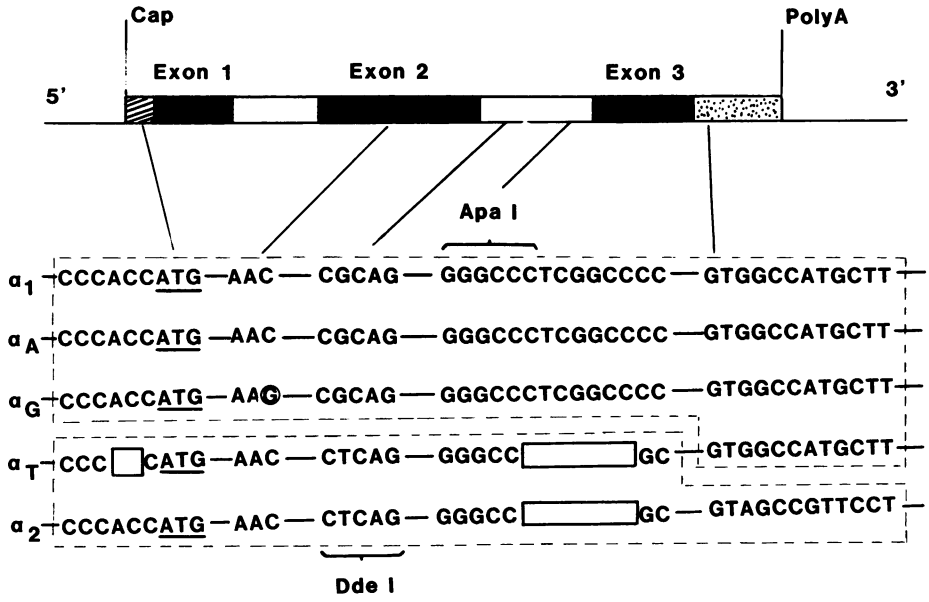


Fig. 1. Nucleotide sequence comparison between the α^A, α^G, α^{+thal} single genes and the $\alpha 1$, $\alpha 2$ normal genes. Normal $\alpha 1$ and $\alpha 2$ globin gene sequences are those recently revised by Michelson and Orkin (1983). α^T refer to the α^{+thal} gene. Only differences between the five α genes are reported. Open boxes indicate deletions whereas the α^G mutation is indicated by an open circle. *Apa*I and *Dde*I restriction sites allowing distinction of $\alpha 1$ and $\alpha 2$ ISV 2 are also positioned.

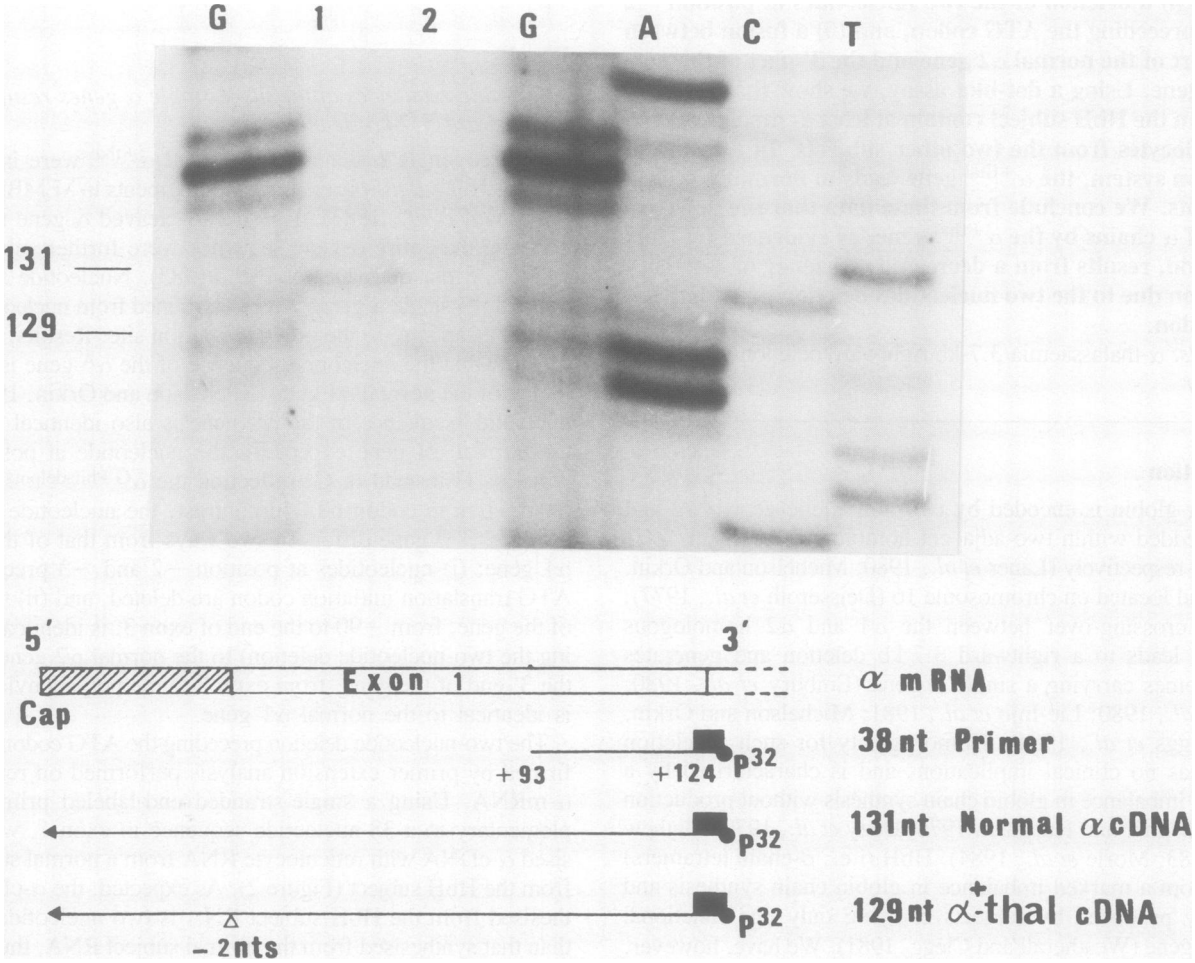
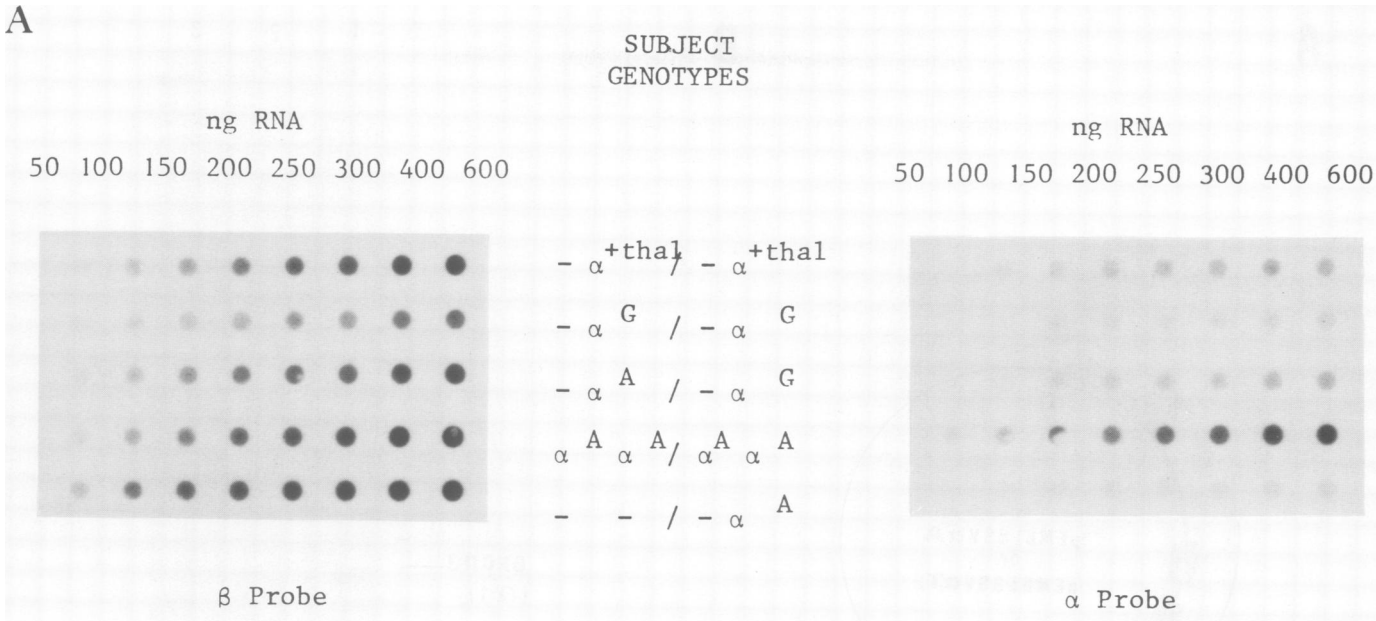


Fig. 2. Primer extension analysis of α-mRNA in reticulocyte total RNA from a normal subject **(lane 1)** and from the HbH subject **(lane 2)**. **Lanes G, A, C** and **T**, by reference to a known sequence, allowed us to ascertain the expected lengths of α-cDNA. α-cDNA synthesis was initiated in exon 1 at position +93 (with respect to the mRNA cap site) with a 38-nucleotide 5′ end-labeled single-stranded DNA fragment which contains the antisense α-gene sequence from +93 to +124 and eight additional nucleotides from the M13 vector (dark block in **lower figure**, see Materials and methods).

A



B

Subjects	Normal	α-thalassaemic			
		without HbH		with HbH	
Genotypes ng RNA	$\alpha^A\ \alpha^A/\alpha^A\ \alpha^A$	$-\alpha^A/-\alpha^G$	$-\alpha^G/-\alpha^G$	$-\alpha^{+thal}/-\alpha^{+thal}$	$--/-\alpha^A$
50	0.94	0.35	0.40	0.45	0.18
100	0.85	0.36	0.33	0.44	0.19
150	0.70	0.33	0.34	0.44	0.19
200	0.80	0.37	0.31	0.42	0.19
250	0.82	0.37	0.32	0.41	0.20
300	0.81	0.39	0.33	0.41	0.21
400	0.80	0.40	0.33	0.40	0.22
600	0.78	0.41	0.30	0.36	0.22
Means ± SD	0.81 ±0.07	0.37 ±0.03	0.33 ±0.03	0.41 ±0.03	0.20 ±0.01

Fig. 3. Steady-state level of α-mRNA in reticulocyte total RNA. **(A)** Autoradiograph of the duplicated RNA filters after hybridisation with α and β probes, respectively (60-min exposure). **(B)** Ratios of α and β probes hybridised with reticulocyte total RNA. α/β mRNA ratios were determined using a dot-blot assay as reported in Materials and methods. Due to the higher specific activity of the β probe, the indicated ratios are only relative and do not refer to the molar ratio of α/β mRNA. Normal RNA were obtained from a non-α-thalassaemic patient with sickle-cell disease.

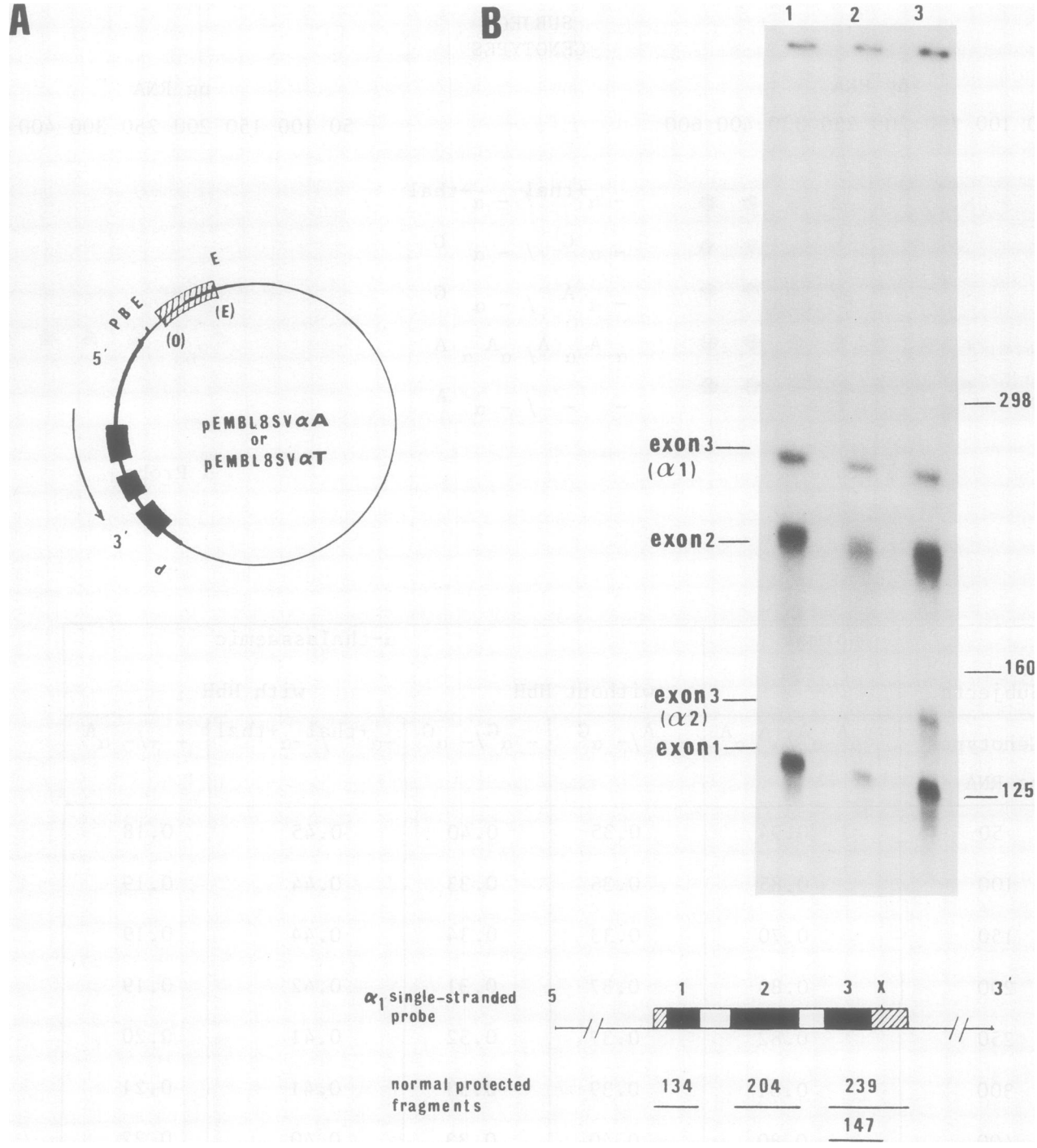


Fig. 4. Expression of the cloned α^{+thal} gene in HeLa cells. **(A)** Vectors used for HeLa cells transfection. **(B)** S1 mapping analysis of total RNA extracted from HeLa cells transfected by pEMBL 8 SV α^A **(lane 2)**, pEMBL 8 SV α^T **(lane 1)** and from normal reticulocytes **(lane 3)**. Lengths of the uniformly labeled α1-globin probe protected by normally processed α-mRNA are indicated at the bottom. X indicates the mismatch between α1 and α2 3′ untranslated sequences 15 nucleotides beyond the TAA stop codon leading to the additional protected α2 fragment of 147 nucleotides observed with normal reticulocyte RNA. Uniform labeling of the probe used in this experiment explains why the shorter α2-specific protected fragment appears no more intense than the longer α1-specific one, in contrast to previously published data using end-labeled probe.

gene carrying a two-nucleotide deletion at position −2 and −3 preceding the ATG codon.

Steady-state level of α-mRNA in reticulocytes from the three subjects homozygous for a 3.7-kb rightward deletion

To establish whether the defective output of α chains in the HbH subject is due to a transcriptional or post-transcriptional defect, we determined the relative α/β mRNA steady-state levels in reticulocytes from the three subjects. As reported in Figure 3, reticulocytes from the HbH subject contain at least as much α-mRNA as reticulocytes from the other two subjects carrying the same number of α genes. The α/β mRNA steady-state level was also determined in reticulocyte RNA from a subject carrying only one functional α gene (genotype $--/-\alpha^A$) and thus producing HbH. As seen in Figure 3, α-mRNA in this subject was half that observed in the three subjects carrying two single α genes.

These results show that the defective output of α chains in the HbH subject, homozygous for the α^+-thalassaemic determinant and with the genotype $-\alpha^{+thal}/-\alpha^{+thal}$, cannot be explained by a reduced level of α-mRNA.

Transient expression of the α^{+thal} gene in HeLa cells

As the α^{+thal} gene is an α2/α1 fusion gene, the secondary structure of the primary transcripts could be altered, leading to abnormally spliced α-mRNA. To test this possibility, the α^A and the α^{+thal} genes were subcloned in a pEMBL SV expression vector (Figure 4A) and the recombinant plasmids were used to transfect HeLa cells. Transcripts obtained from the transfected cells were analysed by S1 mapping using a uniformly labeled single-stranded α-globin probe (Figure 4B). In both cases, only the protected fragments corresponding to the three normal exons, were observed. We concluded from this that the defective output of α chains from the α^{+thal} gene cannot be explained by abnormal splicing of α-transcripts.

Discussion

The aim of this study was to identify the α^+thalassaemic determinant responsible for the defective output of α chains in a previously reported Algerian subject homozygous for a rightward 3.7-kb deletion who produced HbH. The nucleotide sequence of the α^{+thal} gene from this subject, when compared with that of two other single α genes from subjects also homozygous for a rightward 3.7-kb deletion but who produced no HbH, revealed two unexpected characteristics: (i) it carries a deletion of two nucleotides at position −2 and −3 preceding the ATG codon, and (ii) it is an α2/α1 fusion gene, in contrast to the two other α-genes, which are identical to the normal α1 gene.

Independently of the two-nucleotide deletion, the structural differences observed between the three single α genes confirm the heterogeneity of the $-\alpha^{3.7}$/haplotype, found in other populations (Michelson and Orkin, 1983; Higgs *et al.*, 1984). The structure of the α^{+thal} gene results from a recombination event in segment II, as defined by Higgs *et al.* (1984), whereas the α^A and α^G genes result from a recombination event in segment I.

Which of these two characteristics was responsible for the defective output of α chains from the α^{+thal} gene? As reticulocytes from the HbH subject contain at least as much α-mRNA as those of the two subjects without HbH, the defect can only occur at a post-transcriptional level. Furthermore, S1 mapping studies of α-transcripts in a transient expression assay revealed that α-transcripts from the α^{+thal} gene are correctly spliced. Therefore, the only difference between α-mRNA from the HbH subject and the two subjects without HbH is the two-nucleotide deletion preceding the AUG codon. This leads us to conclude that the two-nucleotide deletion is responsible for the defective output of α chains, probably by reducing the efficiency of translation initiation.

Results published recently support our conclusion. A survey on 5′ untranslated sequences from 200 higher eukaryote mRNA has indeed shown that the sequence preceding the AUG initiation codon is highly conserved with the proposed CC$_G^A$CCAUG consensus sequence (Kozak, 1984a). In addition, it has been shown that point mutations, at position −3 preceding the ATG codon in the preproinsulin gene, affect the level of preproinsulin production in a transient expression system such as COS cells (Kozak, 1984b).

Interestingly, the two-nucleotide deletion observed in the α^{+thal} gene leads to a change A→C at position −3 preceding the ATG codon. Similarly, the same change A→C at position −3 in the preproinsulin gene leads to lower production of preproinsulin in COS cells. The naturally occurring mutant reported here is an *in vivo* confirmation of the effect of the nucleotide sequence preceding the AUG codon on translation efficiency.

Materials and methods

Cloning and sequencing of α-globin genes

DNA from white blood cells from the three subjects was digested to completion with restriction endonuclease *Bgl*II. 16-kb *Bgl*II-cleaved fragments were enriched by sucrose-density gradient, ligated to λEMBL 4 *Bam*HI-cut arms (Frischauf *et al.*, 1984) and packaged *in vitro* (Maniatis *et al.*, 1982). α Clones were screened until purity using an α1 1.5-kb *Pst*I nick-translated probe. α 1.5-kb *Pst*I α-globin fragments were further subcloned in pEMBL 8 plasmid (Dente *et al.*, 1983). Sequencing was performed by the method of Sanger (Sanger *et al.*, 1977), using subclones obtained in M13mp8 and M13mp9 single-stranded vectors and a *Bal*31 sequencing strategy initiated at the unique *Hind*III intragenic site.

Primer extension analysis

One of the *Bal*31 deleted subclones was used to prepare a fragment containing exon 1 sequence from nucleotide +93 to nucleotide +124 with respect to the mRNA cap site (+93 refers to a *Hha*I site and +124 to the end of the *Bal*31 deletion). This fragment was 5′ end-labeled with kinase and [γ-^{32}P]ATP and the 38 nucleotide single-stranded fragment complementary to mRNA sequence was purified on denaturing 12% polyacrylamide gel. 3000 c.p.m. of this 38 nucleotide primer were hybridised with 2 μg of reticulocyte total RNA in 10 μl of 0.4 M NaCl, 1 mM EDTA, 10 mM Tris pH 8.0 for 2 h at 65°C. The reaction was then adjusted in 50 μl to 500 μM dATP, dGTP, dTTP, dCTP, 5 mM DTT, 50 mM Tris pH 8.0, 8 mM $MgCl_2$, 20 units/ml RNAsine. 10 units of reverse transcriptase were added and an incubation was performed for 45 min at 37°C. The cDNA were ethanol-precipitated and analysed on 6% polyacrylamide sequencing gel.

HeLa cells transfection

The *Pvu*II/*Hind*III SV40 fragment, which spans over the two 72-bp repeats and the origin of replication, was subcloned using *Eco*RI linkers in the *Eco*RI site of the polylinker in pEMBL 8 plasmid (Dente *et al.*, 1983) leading to pEMBL 8 SV expression vector. The 1.5-kb *Pst*I α^A and α^{+thal} gene fragments were subcloned in the indicated orientation in pEMBL 8 SV leading to pEMBL 8 SV α^A and pEMBL 8 SV α^T plasmids. These plasmids were introduced independently into HeLa cells by calcium phosphate precipitation according to Mellon *et al.* (1981) with the following modifications: after 5 h of contact with DNA, the cells were incubated 4 min at room temperature in medium containing 15% glycerol. Cells were rinsed with phosphate buffered saline and fresh medium was added. After 48 h of additional incubation at 37°C, cells were frozen in liquid nitrogen until use.

S1 mapping analysis

Total RNA from transfected HeLa cells were extracted by the LiCl-urea method (Auffray and Rougeon, 1980). A uniformly labeled α1 globin probe was obtained using the method reported by Ley *et al.* (1982) as follows. 0.5 − 1 μg of single-stranded DNA from pEMBL 8 SV α^A was annealed with 4 ng of 15-bp universal primer (Bethesda Research Laboratory); the probe was synthesised 30 min at 37°C by Klenow DNA polymerase (Amersham) using [α-^{32}P]dCTP and dTTP (400 Ci/mmol, Amersham) and unlabeled dGTP and dATP. Following a 30-min chase with cold dCTP and dTTP, DNA was phenol-extracted, ethanol-precipitated, digested with *Bam*HI, denatured for 10 min at 100°C, cooled in ice and immediately loaded on 1.5% agarose gel. The 1500-nucleotide radioactive single-stranded probe was then recovered by an optimized squeeze-freeze method (Tautz and Renz, 1983). For S1 mapping, 10 000 c.p.m. of the probe were hybridised for 20 h at 42°C with 5 μg of total RNA from HeLa cells (or 1 μg of total reticulocyte RNA) in 20 μl of 50% formamide, 10 mM Tris pH 8.0, 1 mM EDTA, 0.4 M NaCl. Hybrids were digested for 1 h at 37°C with 100 units of S1 nuclease (Bethesda Research Laboratory) in 250 μl of S1 buffer (50 mM NaCl, 30 mM Na acetate pH 4.6, 1 mM $ZnSO_4$), 20 μg/ml of freshly denatured salmon sperm DNA), ethanol-precipitated and analysed on 6% polyacrylamide sequencing gel.

Dot blot assay on reticulocyte RNA

Serial dilutions containing the indicated quantities of reticulocyte total RNA were denatured 15 min at 60°C in 50% formamide, 6% formaldehyde, 20 mM MOPS, 5 mM CH_3COO Na, 1 mM EDTA, diluted in ice-cold 20 x SSC and filtered in duplicates on nitrocellulose filters. After baking, duplicate filters were hybridised with a 1.5-kb α-*Pst*I or a 4.4-kb β-*Pst*I nick-translated globin probe (sp. act. 32 x 10^6 d.p.m./μg and 65 x 10^6 d.p.m./μg for α and β probes, respectively). After post-hybridisation washing, radioactive spots were cut, dried and

counted by liquid scintillation. For each dilution, the ratio α/β of radioactivity hybridised with α and β probes was determined.

Acknowledgements

λEMBL 4 vector was provided by J.C.Le Hegara, pEMBL vectors by L.Dente and reverse transcriptase by P.Beard. We thank M.Cohen-Solal, P.Couble, M. Ben-Chaïbi, M.Goossens, J.C.Le Hegara, R.Hen, E.May and J.Samarut for stimulating discussions, R.Grantham for reading the manuscript and M.Motta and P.M. Malet for their excellent technical contribution. This work was supported by grants from the Centre National de la Recherche Scientifique and the Institut National de la Santé et de la Recherche Médicale.

References

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303-314.
Deisseroth,A., Nienhuis,A., Turner,P., Velez,R., French-Anderson,W., Ruddel,F., Lawrence,J. Creagen,R. and Kucherlapati,R. (1977) *Cell*, **12**, 205-218.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645-1655.
Dozy,A.M., Kan,Y.W., Embury,S.H., Mentzer,W.C., Wang,W.C., Lubin,B., Davis,J.R. and Koenig,H.M. (1979) *Nature*, **280**, 605-607.
Embury,S.H., Miller,J.A., Dozy,A.M., Kan,Y.W., Chan,V. and Tood,D. (1980) *J. Clin. Invest.*, **66**, 1319-1325.
Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827-842.
Higgs,D.R., Hill,A.V.S., Bowden,D.K., Weatherall,D.J. and Clegg,J.B. (1984) *Nucleic Acids Res.*, **12**, 6965-6977.
Kozak,M. (1984a) *Nucleic Acids Res.*, **12**, 857-872.
Kozak,M. (1984b) *Nature*, **308**, 241-246.
Lauer,J., Shen,C.J., Maniatis,T. (1980) *Cell*, **20**, 119-130.
Ley,T.J., Anagnou,N.P., Guglielmina,P. and Nienhuis,A.N. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4775-4779.
Lie-Injo,L.E., Herrera,A.R. and Kan,Y.W. (1981) *Nucleic Acids Res.*, **9**, 3707-3717.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Mathew,C.G.P., Rousseau,J., Rees,J.S. and Harley,E.H. (1983) *Br. J. Haematol.*, **55**, 103-111.
Mellon,P., Parker,V., Gluzman,Y. and Maniatis,T. (1981) *Cell*, **27**, 279-288.
Michelson,A. and Orkin,S.H. (1983) *J. Biol. Chem.*, **258**, 15245-15250.
Milner,P.F. and Huisman,T.H.J. (1976) *Br. J. Haematol.*, **34**, 207-220.
Morle,F., Jaccoud,P., Dorleac,E., Motta,M., Delaunay,J. and Godet,J. (1984) *Human Genet.*, **65**, 303-307.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Tabone,P., Henni,T., Belhani,M., Colonna,P., Verdier,G. and Godet,J. (1981) *Acta Haematol.*, **65**, 26-31.
Tautz,D. and Renz,M. (1983) *Anal. Biochem.*, **132**, 14-19.
Weatherall,D.J. and Clegg,J.B. (1981) *The Thalassaemia Syndromes*, 3rd Edition, published by Blackwell Scientific Publications, Oxford.
Whitelaw,E., Pagnier,J., Verdier,G., Henni,T., Godet,J. and Williamson,R. (1980) *Blood*, **55**, 511-516.

Received on 31 December 1984; revised on 26 February 1985