THE STRUCTURE AND EXPRESSION OF NORMAL AND ABNORMAL GLOBIN GENES

R.A.Flavell, G.C.Grosveld, F.G.Grosveld, R.Bernards,
J.M.Kooter and E.De Boer

Section for Medical Enzymology and Molecular
Biology, Laboratory of Biochemistry, University
of Amsterdam, Amsterdam, The Netherlands

P.F.R.Little

Department of Biochemistry, St. Mary's Hospital
Medical School, University of London, London,
Great Britain

I. Introduction

Most mammalian genes are present as a single copy per haploid genome and hence comprise only about one part in 10^6 of the total nuclear DNA. This fact impeded work on single-copy genes, but recently recombinant DNA technology (e.g. ref. 1) and sensitive gene mapping [2-4] has led to the elucidation of the structure of a number of eukaryotic genes [5-10].

II. The Structure of the Rabbit β-Globin Gene

Our initial approach was to construct a detailed physical map of the rabbit β-globin gene using the blotting-filter hybridization approach [4,5]. This showed that the β-globin gene consists of at least two non-contiguous blocks of coding sequences separated by an intervening sequence (IVS) of 600 base pairs (bp). The rabbit β-globin gene has now been cloned and we have analysed its
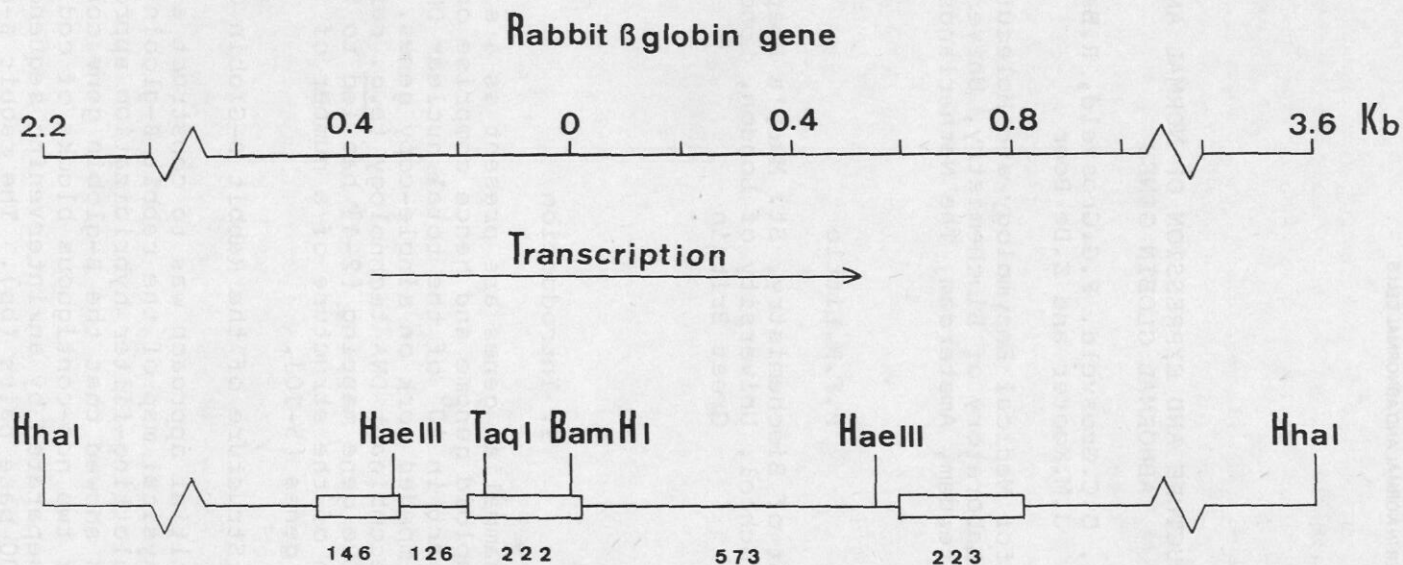
Fig. 1. A restriction map of the rabbit ß-globin gene. The map depicts RßG1 DNA (a 5 kb segment of rabbit chromosomal DNA inserted into plasmid pCRI). The blocks signify the coding segments of the ß-globin gene.

structure in more detail including the determination of the DNA sequence (in collaboration with Charles Weissmann and his colleagues [11]). These experiments show that two intervening sequences are present in the rabbit β-globin gene (Fig. 1), one of 126 bp in length present between the DNA sequences for amino acid 30 and 31 and the one already mentioned above of 573 bp long, present between the sequences coding for amino acids 104 and 105 [11]. For a discussion of the DNA sequence of the rabbit and mouse β-globin genes, see the article by Weissmann et al. [this volume].

III. Transcription of the Rabbit β-Globin Gene

Of the many models to explain the way in which mature mRNA is obtained from split genes (see e.g. refs 12,13,5), RNA splicing [12-15] is by far the most probable, although formal proof for many of the steps is lacking. Splicing was proposed to explain the primary observation that adenovirus messenger RNAs (mRNAs) were transcripts of non-contiguous segments of the viral genome. Two subsequent observations on cellular gene transcripts have supported this model.

First, it has been known for some time that in mouse a β-globin pre-mRNA of 15S (about 1200-1500 nucleotides (NT) long) exists. The size of this precursor is approximately that expected of a transcript of the coding segments plus intervening sequences. R-loop mapping of the pre-mRNA on the cloned mouse β-major globin gene showed a continuous R-loop in the same region as the R-loops formed with β-globin mRNA which establishes that the pre-mRNA is a transcript of the coding regions plus at least the large intervening sequence [16].

Second, Rutter's and Abelson's groups have shown that an enzyme activity is present in yeast which can remove the intervening sequence present in yeast pre-tRNA transcripts to generate mature tRNA (see e.g. ref. 17).

IV. Transcription Mapping of Rabbit β-Globin pre-mRNA

Does the 15S RNA contain RNA complementary to the extragenic regions 5' and 3' of the coding regions or is it simply a transcript of the coding sequences and intervening sequences? The electron microscopical data cannot answer this definitively

and we have, therefore, taken another approach.

Berk and Sharp [18] have shown that it is possible to map transcripts on a DNA accurately by forming DNA-RNA hybrids, degrading the unhybridized DNA with S_1 nuclease and analysing the size of the protected DNA on alkaline agarose gels. Spliced mRNAs form hybrids with single-stranded DNA loops which are destroyed on digestion with S_1 nuclease to give DNA fragments of the same length as the coding segments. A pre-mRNA containing all the intervening sequences forms a single S_1 nuclease resistant hybrid of the same length as the pre-mRNA.

We have analysed the transcripts of the rabbit β-globin gene in this way. Bone-marrow RNA was prepared from an anaemic rabbit, size-fractionated on a sucrose gradient and then hybridized to various fragments of the cloned rabbit β-globin gene. The single-stranded nucleic acids were degraded with S_1 nuclease, the resistant DNA fractionated by alkaline agarose gel electrophoresis and detected by Southern blotting and hybridization to probes for the β-globin structural gene (pβG1 DNA [19]) or the cloned chromosomal DNA fragment (RβG1 DNA [11]). Fig. 1 gives a schematic map of the RβG1 DNA in which the HhaI, BamHI and TaqY1 sites discussed here are shown.

Hybridization of RNA of approximately 15S (12-18S) to RβG1 DNA cleaved with HhaI (which cuts in the plasmid regions of RβG1 to give an approximately 6 kb fragment containing all of the rabbit DNA), yields a series of fragments in the alkaline gel (Fig. 2) . The largest fragment is 1250 NT long, a size consistent with a transcript of the coding sequences plus intervening sequences (1295 NT). In addition, fragments of 223 and 146 NT are present which are exactly the size expected for the DNA pieces generated by S_1 nuclease treatment of hybrids formed between β-globin mRNA and the cloned DNA. The IVS loops will be cut in such hybrids to give segments of 146, 222 and 223 NT in length. Finally, there is evidence for a pre-mRNA in which only the small IVS has been removed by processing. A prominent 1000 NT fragment (Fig. 2) is found in hybrids formed with RβG1 x HhaI (from this pre-mRNA we should expect a 1018 NT fragment) and a 735 NT fragment is found with RβG1 x HaeIII hybrids (expected 753 NT). Other faint bands are present which may be derived from still other splicing intermediates. These have not yet been
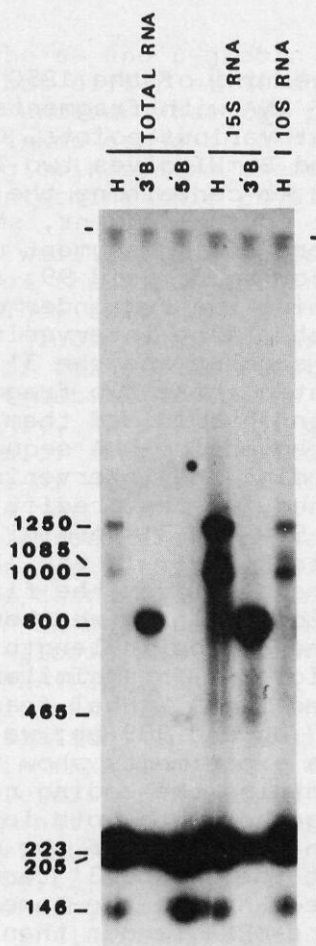
Fig. 2. Detection of the β-globin pre-mRNA from rabbit bone-marrow DNA by S_1 nuclease mapping. Total bone-marrow DNA, a 12-18S RNA fraction and an 8-12S fraction were hybridized with RβG1 DNA cleaved with HhaI (H), or the two separated globin gene-containing fragments of RβG1 cleaved with HhaI + BamHI (3' B and 5' B: see Fig. 1). Hybridization was performed under conditions where only DNA-RNA hybrids form. The sample was digested with S_1 nuclease and electrophoresed on a 2% alkaline agarose gel [18]. Phage ØX DNA cleaved with TaqYI, HaeIII and HapII (5' end-labelled with ^{32}P) was used as a molecular weight marker. The DNA was transferred to a nitrocellulose filter and hybridized with ^{32}P-labelled RβG1 DNA as described [4,5,23]. The figure shows an autoradiogram of the nitrocellulose filter.

mapped.

To map the termini of the 1250 NT transcript we hybridized 15S RNA with fragments of RβG1 where the gene is cut at various points. Cleavage of RβG1 with HhaI and BamHI gives two large fragments; a fragment of 2.2 kb containing the 5' extragenic region, the first coding segment, small intervening sequence and second coding segment up to the sequence coding for amino acid 99, and a fragment of 3.6 kb containing the remainder of the second coding segment, the large intervening sequence, the final coding segment and the 3' extragenic region. We separated these two fragments by gel electrophoresis and hybridized them to the '15S' RNA pool. Since the entire DNA sequence of the β-globin gene (coding and intervening sequences) has been determined, we can predict precisely the positions of the 5'- and 3'-termini of the β-globin mRNA on the map in Fig. 1. If the 1250 NT pre-mRNA is a transcript reading from the first coding segment through to the last, we should obtain hybrids of 477 and 815 bp in length. We find 465 bp and 800 bp (Fig. 2). In a similar experiment with the separated TaqY1 + HhaI fragments we should expect 307 bp and 989 bp; we find 320 bp and 960 bp. These experiments show that the 1250 NT transcript contains the coding regions of the β-globin gene together with both intervening sequences. Within the error of the measurements we can conclude that the 5' and 3' termini of the mRNA and this pre-mRNA are the same. We have no evidence for a pre-mRNA larger than the 1250 NT component. The question whether the termini of this transcript represent the initiation and termination points for transcription, however, still remains open. This is likely to remain unanswered until in vitro transcription systems have been set up or until cloned DNA segments have been brought to expression in cell systems [20].

V. The Human Globin Genes

Two aspects of the human globin gene complex prompted our initial attention:
1. Genetic evidence suggested 'close' linkage of the non-α-globin genes in one of two arrangements, either Gγ-Aγ-δ-β or Aγ-δ-β-Gγ (the two non-allelic foetal γ-globins differ by a single amino acid at position 136, Gγ has glycine and Aγ alanine at

this position; the δ- and β-globin genes are also closely related but differ at 10 amino acid sites). It was, however, unclear what the physical distance is between two linked genes on a mammalian chromosome. The human globin system seemed, therefore, appropriate to determine this physical linkage directly.

2. Numerous well-defined cases of defects in the functioning of the human globin genes have been described. They fall into two classes:

a) Abnormal globin proteins are produced usually because of amino acid substitutions which result from point mutations in the structural gene; less commonly, deletions or gene fusions cause other types of abnormal haemoglobins.

b) The level of - otherwise normal - haemoglobin can be reduced or even be zero in the diseases known as the thalassaemias. Here we shall restrict the discussion to the β-thalassaemias and related diseases. In some cases of β°-thalassaemia no β-globin mRNA can be detected in the cell [21]; in others nuclear β-globin RNA is found but no cytoplasmic globin mRNA [22]; in still other cases cytoplasmic globin mRNA is found which is apparently inactive in translation. In β$^+$-thalassaemia reduced levels of β-globin mRNA are found with a parallel decrease in the levels of β-globin in the cell.

Our approach has been to construct physical maps of the normal human globin gene using the blotting procedure [2,4] and then to compare these with the corresponding maps of the DNA of patients with the various thalassaemias. In this way, large deletions in DNA regions within or around the globin genes can be detected directly in total genomic DNA.

VI. The Structure of the Human δβ-Globin Linkage Group

We have constructed a map of the two linked δ- and β-globin genes [23] (see Fig. 3). This map shows:

a) That the δ-globin gene is located about 7000 bp to the 5'-side of the β-globin gene. Both genes are transcribed from the same DNA strand.

b) The structures of the human δ- and β-globin genes strongly resemble the structures of the rabbit and mouse globin genes described earlier [5,6]. The blotting experiments show the presence
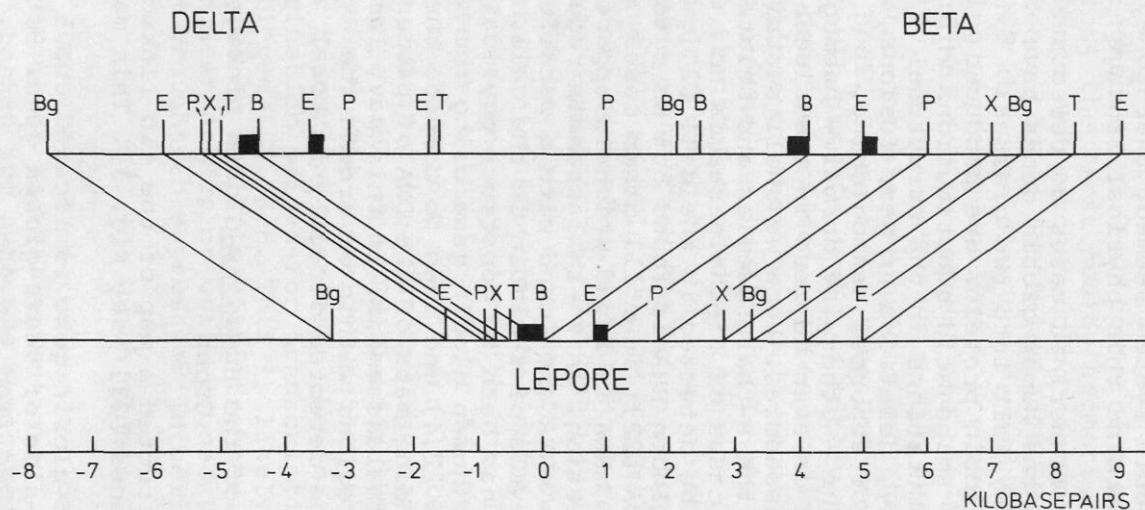
Fig. 3. A physical map of the ß- and δ-globin genes in normal and Hb Lepore DNA. The probable positions of the coding regions of the two globin genes are shown as filled boxes. It should be stressed that the only extragenic cleavage sites which can be detected for a given enzyme by this analysis are those closest to the gene examined. Although the ß- and δ-genes are presented as being composed of two coding segments in each case, the possibility that these segments are further split cannot be excluded from these data. Bg, BglII; E, EcoRI; P, PstI; X, XbaI; T, TaqYI; B, BamHI. Reproduced from ref. 23 with permission.

of the large intervening sequence in both human genes within the same region as in the rabbit and mouse genes. The human intervening sequences are 800-1000 bp long instead of 600 bp in the case of rabbit and mouse.

Both Mears et al. [24] and Lawn et al. [25] have also analysed the structure of the δ-β gene region and the results of the three groups do not show significant differences. In the latter case [25], the linkage of the δ- and β-globin genes has also been demonstrated.

VII. The Structure of the Human γ-Globin Genes

More recently we have also mapped the human γ-globin genes (Fig. 4) [26]. The map shows two linked γ-genes and the general organization of the genes resembles the β-δ locus.
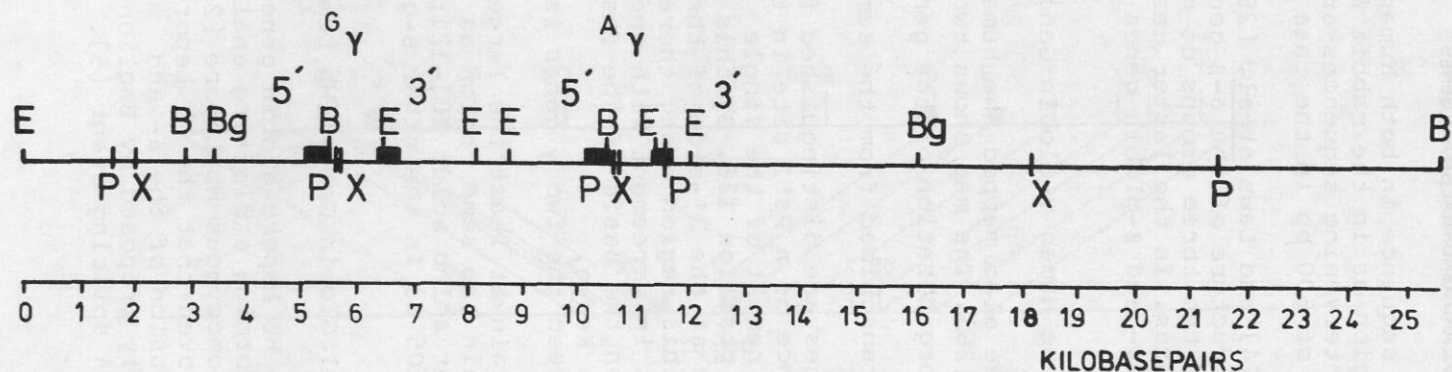
a) Both genes are transcribed from the same DNA strand.

b) The Gγ and Aγ genes are distinguished from one another by the presence of a PstI site in the latter gene. This is the result of the single amino acid difference at position 136. By this criterium, the Aγ gene is at the 3'-side of the Gγ gene; the gene order in this region must, therefore, be 5'-GγAγδβ-3', which is in agreement with one of the two models proposed on the basis of the fusion proteins Hb Lepore and Hb Kenya.

c) The distance between the two γ genes is 3500 bp.

d) Both γ genes contain at least the large intervening sequence within the same region as in β-globin genes (coding for amino acids 101-121; but presumably also 104-105 as in the other β-globin genes).

VIII. Analysis of Abnormal Globin Genes: Hb Lepore

The structure of the Hb Lepore globin gene has been analysed by performing a blotting analysis on DNA of patients with homozygous Hb Lepore [23]. This investigation has proven that the Hb Lepore gene is the result of a fusion of the δ- and β-globin gene as originally proposed by Baglioni [27] and not a δ-RNA-β-RNA splicing event [5].

Fig. 4. A physical map of the γ-globin genes. The general comments in Fig. 3 apply here also. For abbreviations, see legend to Fig. 3.
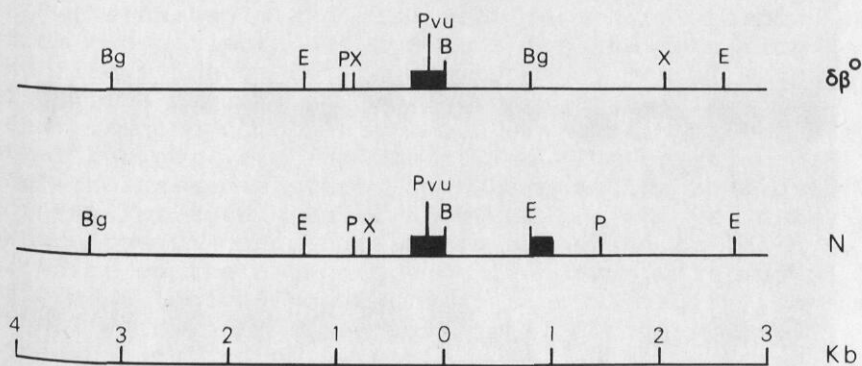
Fig. 5. A physical map of the δ+β-globin gene region in DNA from a patient homozygous for δβ°-thalassaemia. The DNA was isolated from a lymphocyte cell line, digested with the restriction enzymes indicated in single and in various double digests. The map is based upon the following fragments: BglII (Bg), 3.8 kb; BglII + BamHI (B), 3.1 kb; BamHI + PstI (P), 0.95 kb; BamHI + XbaI (X), 0.85 kb; BglII + EcoRI (E), 2.1 kb; BglII + PstI, 1.7 kb; PvuII (Pvu) + XbaI, 0.8 kb; BglII + XbaI, 1.65 kb; XbaI, 2.9 kb; EcoRI, 3.9 kb. N is the normal δ-globin gene.
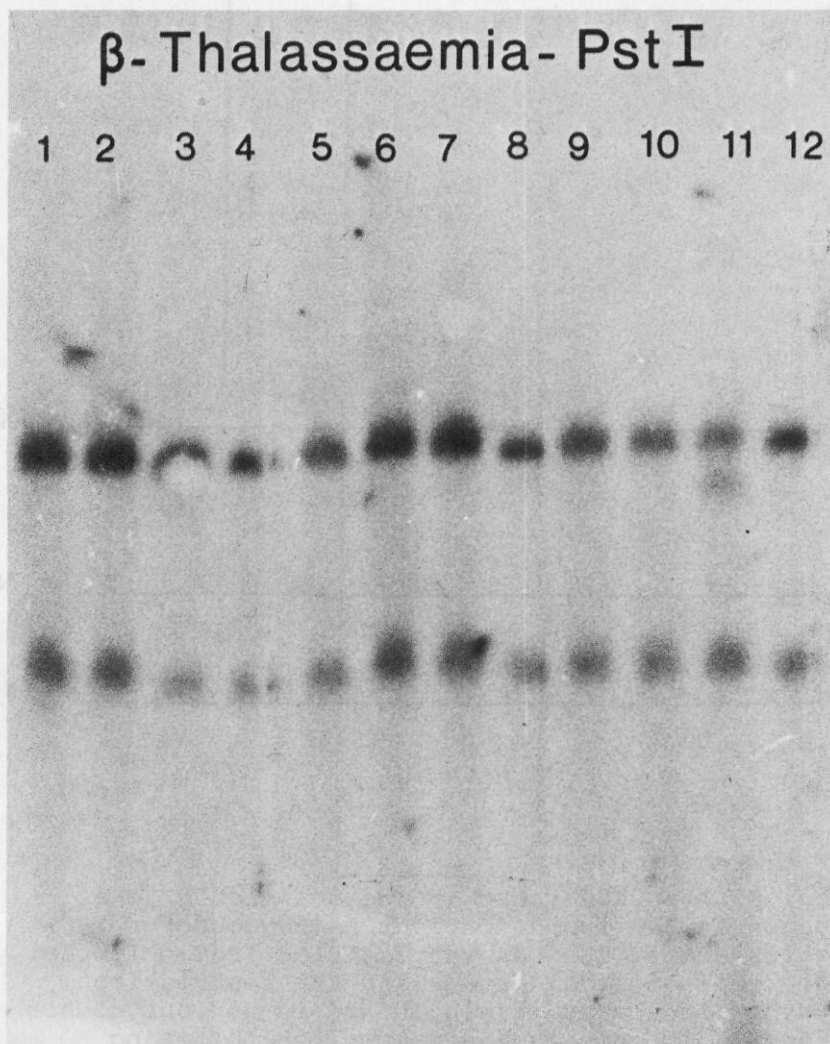
Fig. 6. β- and δ-globin gene fragments in digests
of DNA from patients homozygous for β-thalassaemia.
DNA was isolated from β-thalassaemics as described
[4] and digested to completion with PstI. The
samples were analysed for the β-δ-globin genes as
described [23]. Patient 5 is described by Comi et
al. [22]: Patient 11 in ref. 21. Patients 2, 3 and
10 have been diagnosed as β°-thalassaemia (Southern
Italian; Ottolenghi, S., personal communication)
and the remainder as homozygous β[+]-thalassaemics.
1, 4, 8 and 12 show control normal DNA (derived
from a placenta from a Dutch individual).

IX. The β-Thalassaemias

δβ°-Thalassaemia is a rare condition that is characterized by the complete absence of β- and δ-globin chains. Complementary DNA (cDNA) titration hybridizations have shown that at least partial deletion of the β- or δ-globin genes has occurred although the extent of the deletion could not be defined. Recently we have constructed a map of cleavage sites in δβ°-thalassaemia. We have analysed DNA from two sibling patients homozygous for δβ°-thalassaemia and also from peripheral blood from two patients heterozygous for the condition. We find evidence for two components, one of which is the 5' regions of the δ-globin gene (Fig. 5) and a second component, which gives faint bands, that has not yet been identified. The deletion in δβ°-thalassaemia starts somewhere to the 3'-side of the intragenic BamHI site. We are not yet sure where the DNA sequences which are fused to the 3'-side of the δ-gene are derived from, although the most logical possibility is a DNA region to the 3'-side of the β-globin gene. Curiously, no transcripts of the δ-globin gene can be detected in δβ°-thalassaemia. This suggests either that the mere presence of the 5' regions of the δ-globin gene are not sufficient to ensure transcription of this region, or that transcription of this half-gene does occur but the transcripts are rapidly degraded.

In contrast to δβ°-thalassaemia, the structure of the β-globin gene in β°- and β⁺-thalassaemia is in most cases indistinguishable from the normal β-globin gene in our blotting hybridization experiments. Fig. 6 shows PstI digests of DNA from a normal placental DNA preparation and from patients with various forms of β-thalassaemia. In only one case can a difference be seen (no. 11). Here a second β-globin DNA fragment is found which is 600 bp shorter than the normal PstI β-gene fragment. Mapping experiments have shown that this results from a 600 bp deletion, which includes the 3' coding regions of the β-globin gene (coding for amino acids 104-146) and probably the 3' untranslated region (Fig. 7).

In the majority of β-thalassaemia cases, however, we can detect no alterations in the β-globin gene. To determine the molecular nature of the lesion, it will probably be necessary to
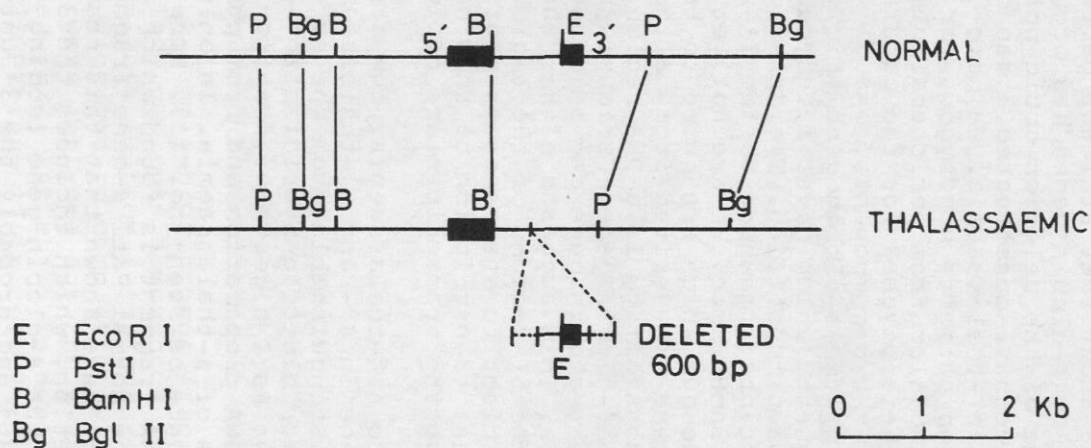
Fig. 7. A physical map of the ß-globin gene region in a normal individual and a patient with ß°-thalassaemia. The map shows a 600 bp deletion centered round the ß-globin intragenic EcoRI site which is missing in this ß°-thalssaemia gene. The 600 bp deletion is arbitrarily drawn centered on this EcoRI site. Also shown are the possible limits of the deletion which stretch 600 bp either side of this site.

clone the thalassaemic genes and to compare these genes with the normal β-globin gene. The problems involved with this type of approach have been discussed elsewhere [20].

X. Acknowledgements

XI. References

1   Tilghman, S.M., Tiemeier, D.C., Polsky, F., Edgell, M.H., Seidman, J.G., Leder, A., Enquist, L.W., Norman, B. and Leder, P. (1977) Proc. Natl.Acad.Sci.U.S. 74, 4406-4410.
2   Southern, E.M. (1975) J.Mol.Biol. 98, 503-517.
3   Botchan, M., Topp, W. and Sambrook, J. (1976) Cell 9, 269-287.
4   Jeffreys, A.J. and Flavell, R.A. (1977) Cell 12, 429-439.
5   Jeffreys, A.J. and Flavell, R.A. (1977) Cell 12, 1097-1108.
6   Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V. and Leder, P. (1978) Proc.Natl.Acad.Sci.U.S. 75, 725-729.
7   Breathnach, R., Mandel, J.L. and Chambon, P. (1977) Nature 270, 314-319.
8   Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977) Proc.Natl.Acad.Sci.U.S. 74, 3518-3522.
9   Goodman, H.M., Olson, M.V. and Hall, B.D. (1977) Proc.Natl.Acad.Sci.U.S. 74, 5453-5457.
10   Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. and Rutter, W.J. (1978) Proc.Natl. Acad.Sci.U.S. 75, 190-194.
11   Van den Berg, J., Van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R.A. and Weissmann, C. (1978) Nature 276, 37-44.

12   Berget, S., Moore, C. and Sharp, P.A. (1977)
      Proc.Natl.Acad.Sci.U.S. 74, 3171–3175.
13   Chow, L.T., Gelinas, R.E., Broker, T.R. and
      Roberts, R.J. (1977) Cell 12, 1–8.
14   Klessig, D.F. (1977) Cell 12, 9–21.
15   Dunn, A.R. and Hassell, J.A. (1977) Cell 12,
      23–36.
16   Tilghman, S.M., Curtis, P., Tiemeier, D.C.,
      Leder, P. and Weissmann, C. (1978) Proc.
      Natl.Acad.Sci.U.S. 75, 1309–1313.
17   Knapp, G., Beckmann, J.S., Johnson, P.F., Fuhr-
      man, S.A. and Abelson, J. (1978) Cell 14, 221–
      235.
18   Berk, A.J. and Sharp, P.A. (1977) Cell 12,
      721–732.
19   Maniatis, T., Kee, S.G., Efstratiadis, A. and
      Kafatos, F.C. (1976) Cell 8, 163–182.
20   Flavell, R.A., Little, P.F.R., Kooter, J.M.
      and De Boer, E. (1979) in Models for the Study
      of Inborn Errors of Metabolism (Hommes, F.A.,
      Ed.), North-Holland, Amsterdam, in press.
21   Tolstoshev, P., Mitchell, J., Lanyon, G.,
      Williamson, R., Ottolenghi, S., Comi, P.,
      Giglioni, B., Masera, G., Modell, B., Weather-
      all, D.J. and Clegg, J.B. (1976) Nature 259,
      95–98.
22   Comi, P., Giglioni, B., Barbarano, L., Otto-
      lenghi, S., Williamson, R., Novakova, M. and
      Masera, G. (1977) Europ.J.Biochem. 79, 617–622.
23   Flavell, R.A., Kooter, J.M., De Boer, E.,
      Little, P.F.R. and Williamson, R. (1978) Cell
      15, 25–41.
24   Mears, J.G., Ramirez, F., Leibowitz, D. and
      Bank, A. (1978) Cell 15, 15–23.
25   Lawn, R.W., Fritsch, E.F., Parker, R.C.,
      Blake, G. and Maniatis, T. (1978) Cell, in
      press.
26   Little, P.F.R., Flavell, R.A., Kooter, J.M.,
      Annison, G. and Williamson, R. (1979) Nature,
      in press.
27   Baglioni, C. (1962) Proc.Natl.Acad.Sci.U.S.
      48, 1880–1886 .