The structure of the human β-globin gene in β-thalassaemia

R.A.Flavell, R.Bernards, J.M.Kooter and E.de Boer
Section for Medical Enzymology and Molecular Biology, Laboratory for Biochemistry, University of Amsterdam, PO Box 60.000, 1005 GA Amsterdam, Netherlands

P.F.R.Little, G.Annison and R.Williamson
Department of Biochemistry, St. Mary's Hospital Medical School, University of London, London, UK

ABSTRACT

Twenty-one cases of β^{0}- and β^{+}-thalassaemia have been analysed by restriction endonuclease mapping. In most cases no deletion in the regions surrounding the β- and δ-globin genes could be detected. However, in a single Asian case of β^{0}-thalassaemia, homozygous clinically, one of the homologous chromosomes contained a β-globin gene with a deletion of 600 base pairs of DNA and comprising most or all of the 3' end of the structural gene including the EcoRI restriction site within the β-globin coding sequence.

Thalassaemia is characterised by a complete absence or reduced amount of either the α- or β-globin chain. It shows a highly non-random geographical distribution, with hetero-zygote levels of up to 25% being found in some areas, notably parts of Italy, Greece and some of the Mediterranean islands[1]. In this paper we concern ourselves only with β-thalassaemia. β-Thalassaemia may be divided into a number of categories: β^{+}-thalassaemia, characterised by the presence of a low level (2 - 20%) of normal β-globin chain; β^{0}-thalassaemia, characterised by a complete absence of β-chains; the rare $\delta\beta^{0}$-thalassaemia, in which there is the complete absence of both δ- and β-globin chains, and finally Hb-Lepores, which are caused by a fusion of the δ- and β-globin genes[1 - 3].

At the molecular level several forms of β-thalassaemia have been demonstrated. $\delta\beta^{0}$-Thalassaemia results from (at least) partial deletion of the δ- and β-globin coding sequences[4, 5], whereas in β^{0}-thalassaemia the β-globin gene is usually intact as judged by cDNA/DNA hybridisation[6, 7]. In β^{+}-thalass-aemia normal globin mRNA is present, albeit at a reduced level[8, 9]; in some forms of β^{0}-thalassaemia mRNA is either

absent or present at variable levels[10 - 12]. In some
β°-thalassaemias globin mRNA sequences are present in the
nucleus but not in the cytoplasm of the erythroid cells[13].
In other cases the globin mRNA is missing a coding sequence
towards the 3' end of the molecule[11].

The development of techniques for the 'mapping' of
restriction endonuclease sites in eukaryotic DNA around
single copy genes[14 - 16] makes it possible to examine whether
deletion of DNA near to the δ- or β-globin genes has occurred
in the β-thalassaemias. Such deletions can be mapped even
when none of the structural gene sequence is affected. We
have recently derived such a 'map' of the human δ- and β-globin
gene locus (Fig. 1)[2] as have others[3] and this makes it
possible to select specific regions of DNA in, and adjacent to,
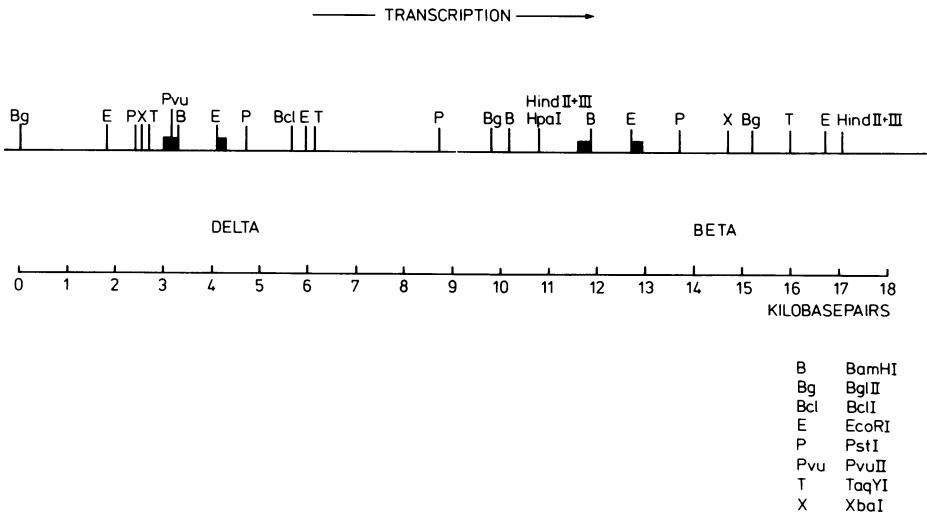the β-globin gene for study.

Figure 1. A physical map of the δ- and β-globin genes. 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 δ- and β-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. These data are taken from Ref. 2.

Briefly, the technique we have used includes the isolation of high molecular weight DNA from whole, washed peripheral blood. This includes white cells and, since many thalassaemics are splenectomised, a large number of nucleated red cells. The DNA is cleaved to give defined fragments using a variety of suitable restriction endonucleases. The fragments are de-natured and then are separated by electrophoresis according to size on 1.2% agarose gels, transferred to cellulose nitrate filter sheets using the Southern "blotting" technique[14], and the δ- and β-globin gene-containing fragments visualised by hybridisation with ^{32}P-labelled pHβG1 DNA. pHβG1 is a recombinant pCR1 plasmid containing 540 base pairs of DNA complementary to β-globin mRNA[17].

Individuals, of known nationality, were diagnosed as suffering from homozygous β-thalassaemia by clinical present-ation, haematological analysis, familial studies and globin chain synthesis ratios[1]. This also allowed, in some patients, a further categorisation into β^{+}- or β^{o}-thalassaemia.

Figure 2 shows the results obtained when Pst I digests of DNA from patients with β^{+}- or β^{o}-thalassaemia are analysed in this way. Pst I normally gives a 5.0 kilobase (kb) fragment containing the β-globin gene and a 2.3 kb fragment containing the δ-globin gene (Fig. 1). For the majority of cases analysed here no difference can be seen between the DNA of β-thalassaemics and control placental DNA from a haematologi-cally normal subject. This suggests that deletions greater than 100 - 200 base pairs in the fragments containing the δ- or β-globin genes are not present in these patients. We have also obtained similar results with EcoRI, Xba I and Hind III digests. This 'normal' pattern has been observed with the following thalassaemic patients: Southern Italian: β^{+}- (3 patients) and β^{o}- (4 patients), Asian: β^{o}- (4 patients), Greek Cypriot: β^{+}- (1 patient), Turkish Cypriot: β^{+}- (6 patients) and unclassified β-thalassaemia major (1 patient). In addition 'normal' patterns were also obtained for two cases of Italian β^{o}-thalassaemia with detectable levels of globin RNA in the nucleus[13] (Fig. 2 and unpublished data).
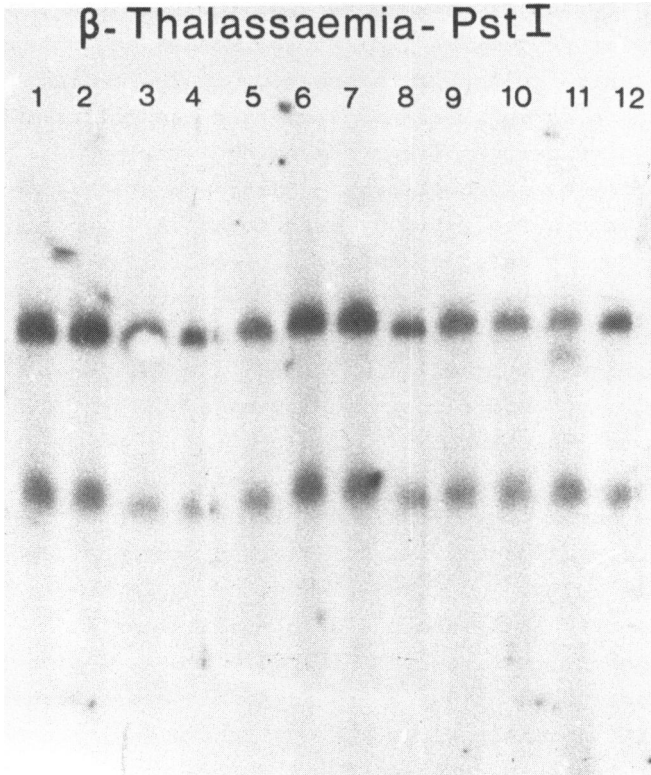
One patient (track 11 in Fig. 2) of Asian origin pre-

Figure 2. δ- and β-globin gene fragments in digest of DNA from patients homozygous for β-thalassaemia. DNA was isolated from either spleen or peripheral blood as described[2], digested to completion with Pst I and analysed for the δ- and β-globin gene fragments by blotting and hybridisation as described[2]. The figure shows an autoradiogram of the nitrocellulose filter after hybridisation to probes for the δ- and β-globin genes. Patient 5 is described in Ref. 13 and patient 11 is a β^{0} of Asian origin. Patients 2, 3 and 10 have been diagnosed as β^{0}-thalassaemics (Southern Italian origin, S. Ottolenghi pers. comm.) and the remainder as homozygous β^{+}-thalassaemics. Lanes 1, 4, 8 and 12 contain control DNAs (from a placenta from a normal Dutch baby).

senting clinically as homozygous β^{0}-thalassaemia, shows an abnormal pattern. Here, in addition to the normal fragments of 5.0 and 2.3 kb, a third fragment of 4.4 kb is visible. This suggests that this patient, while clinically homozygous, is 'molecularly' heterozygous for two different thalassaemia

genes. There are several possible explanations for the
presence of a new 4.4 kb fragment. First, it might arise by
a deletion of 600 base pairs from the 5.0 kb Pst I β-globin
gene fragment. Second, a new Pst I site(s) may have been
introduced into the 5.0 kb Pst I fragment, to generate a
fragment 600 base pairs shorter. Finally, it is possible that
deletion of Pst I sites around the δ-globin gene may have
given rise to a new fragment of 4.4 kb.

   To distinguish these alternatives we analysed DNA from
this patient with other restriction endonucleases. Fig. 3
displays some of these data and Table I contains details of
fragment sizes generated in all digests discussed here. Pst I,
BglII and Pst I plus BglII digests all show, in addition to
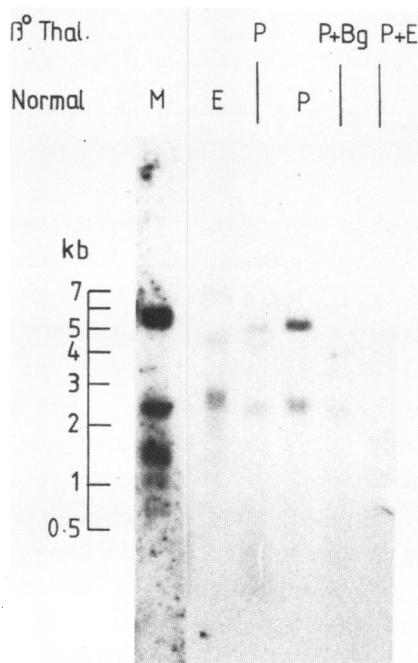the normal fragments, a novel fragment 600 base pairs smaller

Figure 3. Mapping the deletion in the DNA of a patient with
homozygous β°-thalassaemia. DNA from a patient, no. 11 in
Fig. 2, or from a control DNA (of Dutch origin) was digested
with Pst I, Pst I plus EcoRI, or Pst I plus BglII and
analysed for the δ- and β-globin gene fragments as des-
cribed[2].

TABLE I

The size of β-globin DNA fragments in a patient with
β°-thalassaemia compared with the normal β-globin gene frag-
ments.

	Pst I	Pst I + BglII	BglII	Pst I + EcoRI	EcoRI	BglII + EcoRI	BglII + Bam
Normal	5.0	4.0	5.4	4.0	6.5	3.0	1.8
					+	+	+
					4.0	2.4	3.3
	2.3	2.3	9.5	1.8	2.3	2.3	3.3
					+	+	+
					1.8	1.8	
β°-Thal	4.4	3.4	4.8	4.0	10.0	4.8	1.8
				(broad)			+
							3.3
							( )

Sizes are measured relative to the marker fragments described in
(2) and normal sizes are taken from Ref. 2.   The DNA of the
patient with β°-thalassaemia contains a second allele with sizes
as in normal DNA (see Fig. 2).

than the normal β-globin fragment. This demonstrates that
the novel fragment cannot be generated from the δ-globin
containing fragment, since it is found with two different
enzymes, both of which generate a fragment 600 base pairs
smaller than the normal β-fragment. Secondly, these digests
suggest that a deletion of 600 base pairs has generated the
novel fragments, since it is unlikely that new sites for both
BglII and Pst I would have been introduced at identical
positions. This is proved by use of EcoRI digests singly and
in double digests with Pst I and BglII. EcoRI digests of
this patient's DNA show a novel fragment of 10 kb (Fig. 4) and
in addition the novel fragments in BglII and Pst I digests are
not cut by EcoRI, whereas the normal β-globin fragments are
cut by this enzyme. Thus the intragenic EcoRI site in the
abnormal β-globin gene has been deleted, along with 600 bp of
DNA. The 10 kb fragment is explained by fusion around the
EcoRI site of the 6.5 and 4.0 kb fragments, and the loss of
0.6 kb from this fusion product. It is apparent from Fig. 2

that the novel Pst I fragment shows less intense hybridisation
than the normal β-globin fragment. This suggests that there
has been a loss of coding information from the affected
β-globin gene.

It is known that the intragenic EcoRI site is located
within the DNA at the position coding for amino acids 121 - 122
in the β-globin chain[18]. It would therefore be expected that
a 600 base pair deletion, including the EcoRI site, would
remove much of the coding sequence of the affected gene. How-
ever, it is also known[2, 3] that the human β-globin structual
sequences is interrupted by at least one 'intron' (a sequence
within a structural gene that does not code for mRNA) of about
900 bp within the DNA coding for amino acids 104 and 105[19].

We may therefore attempt to locate the deletion more
accurately by detecting the presence or absence of the BamHI
site located in the DNA coding for amino acids 99 - 100[18],
and by hybridisation with probes specific for the DNA to the 3'
or 5' side of the intragenic EcoRI site. A double digest of
DNA cut with BglII plus BamHI indicate that the novel BglII
fragment is cut by BamHI. In normal DNA the 5.4 kb BglII
fragment is cut by BamHI to give fragments of 3.3 and 1.8 kb
(Fig. 1). We would predict that a deletion of 600 base pairs
(including the intragenic EcoRI site) would generate a new
double digest fragment of 2.7 kb which contains the 3' exon of
the β-globin gene. This is not seen (Table 1).This is confirmed
by the pattern of hybridisation of 3' and 5' specific probes
shown in figure 4. With 5' probes, the 4.4Kb fragment is
readily seen whereas the 3' probe does not hybridise to this
fragment. It is not possible from these data to show that all
of the mRNA coding sequence to the 3' side of the normal EcoRI
site has been deleted, only that a substantial proportion has
been removed. Figure 5 shows the map of the deletion round the
β-globin gene in this patient. These data narrow down the site
of the 600 bp deletion to a region of 1,200 bp, 600 bp to either
side of the intragenic EcoRI site.

Since this form of β-thalassaemia seems exceptional we
should ask whether this patient really has homozygous
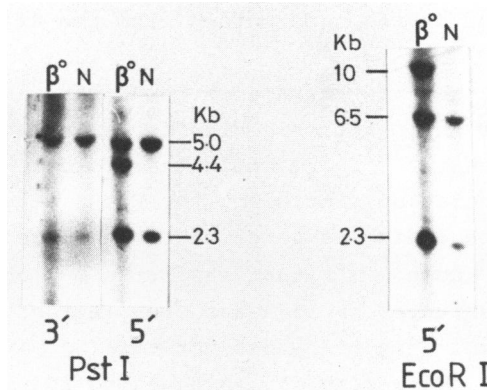β⁰-thalassaemia. Both parents of the patient showed elevated

Figure 4. Hybridisations with 3' and 5' specific probes.
pHβGI was fragmented with EcoRI plus BamHI plus Hind III. This
generates a 4 kb fragment containing the fragment of double
strand cDNA from the EcoRI site coding for amino acid 121 - 122
to the 3' end of the mRNA coding sequence, a 10 kb fragment
containing the DNA from the 5' end of the mRNA coding sequence
to the BamHI site coding for amino acids 99 - 100 and a small
67 base pair fragment containing the double stranded cDNA con-
tained between BamHI and EcoRI sites. The 4 and 10 kb frag-
ments are probes for the 5' and 3' regions of the gene respec-
tively.
        Conditions of hybridisation are as described in Figure 2
except a 3-fold excess of unlabelled DNA of opposite specificity
was included in each hybridisation.
        Note that the new EcoRI fragment of 10 kb in the β^{O} DNA
and that the novel 4.4 kb β^{O} fragment generated by Pst I only
hybridises with the 5' specific probe. The faint band seen in
EcoRI-digested β^{O} DNA between the 10 and 6.5 kb EcoRI fragment
is artifactual. (N = Normal, β^{O} = Thalassaemic)

Hb A$_2$ levels, characteristic of parents of β^{O}-thalassaemics,
and hence the patient is not a double heterozygote for
$\delta\beta^{O}$thalassaemia and β-thalassaemia defects.

        Reticulocytes from this patient have been analysed for
globin mRNA sequences. Within the limits of detection, no
β-globin mRNA could be detected (P. Tolstoshev, personal
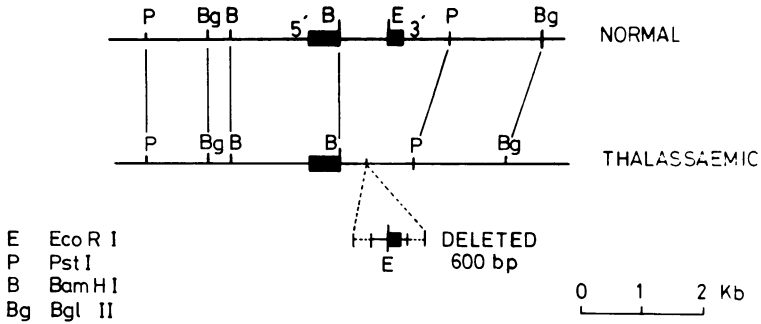communication). This implies that either the genes are not

Figure 5. A physical map of the β-globin gene regions in a normal and a patient with β°-thalassaemia. The map shows a 600 base pair deletion centred round the β-globin intragenic EcoRI site, which is missing in this β°-thalassaemia gene. Also shown are the possible limits of this deletion, including the intragenic EcoRI site and 600 bp to the 3' or 5' side.

transcribed or that the RNA transcripts are not stable within the cell. If the described deletion is the primary β-thalass-aemic lesion in this gene then it is possible that the loss of the 3'-end of the mRNA has resulted in lack of poly (A) addition, or in failure to terminate the primary transcript. Similarly, the deletion of the intron-exon junction would conceivably prevent processing of putative precursor RNAs. In each case we might expect an unstable RNA to be produced. It is also possible that the deletion of extragenic DNA to the 3'-side of the β-globin has prevented transcription of the abnormal globin gene.

It is likely that the deletion we have mapped is the primary lesion in this type of thalassaemia gene, but it is not proven. The deletion might be a secondary event which occurred in a β-thalassaemic gene of the type observed in all other cases we have studied. This is of course possible since the gene sequence is of no selective value once it has ceased to function. It would be useful, in this respect, to perform a population study in the regions from which this patient originates.

It is evident from these data that many cases of

thalassaemia presenting from Mediterranean countries, and some
Asian cases as well, do not result from large deletions of DNA
sequence within several thousand base pairs of the β-globin
structural gene. Since no major deletion can be detected, ante-
natal diagnosis by a change in band pattern, as described by
Orkin[20] for δβ°-thalassaemia, will not be possible for these
cases. However, Kan[21] has described an alternative approach,
which makes use of linked mutations or rearrangements of remote
restriction enzyme cleavage sites for the antenatal diagnosis
of sickle cell disease, and such an approach could be still
applicable for β-thalassaemias even where no deletion can be
demonstrated.

It is known that the thalassaemic lesion segregates
genetically with the β-globin gene, and therefore is unlikely
to be a remote mutation. The data presented here suggests that
the primary lesion in $\beta°$- and β^+-thalassaemias in general
affects a DNA sequence less than 100 - 200 base pairs in length,
and may possibly be a point mutation. Identifying such a
lesion may be possible when β-thalassaemic genomic DNA is
cloned in recombinants and the sequences compared with those
from normal individuals. These comparisons, in conjunction with
in vitro assays for gene expression (see, e.g., ref. 22) and
in vitro base-specific mutagenesis techniques (see, e.g., ref.
23) should reveal the nature of the primary lesion in the
β-thalassaemias.

We would like to stress that it is not formally possible
to prove a lesion is the primary thalassaemic event without
such in vitro methods to study expression. The experiments of
Kan[21] have shown that a tight linkage may exist between an
advantageous gene (in this case HbS) and an apparently
unrelated restriction site polymorphism. It is possible that
similar effects could occur for the advantageous β-thalassaemic
gene, Only by demonstration of lack of in vitro transcription by
correction of the lesion and restoration of transcriptional
activity, is it formally possible to discriminate between
tightly linked polymorphism and the primary thalassaemic lesion.

Since the completion of this work, we have seen similar
data (Orkin, Old, Weatherall and Nathan, personal communication)

derived independently demonstrating the deletion for patient 11
and for two additional patients (submitted for publication).

REFERENCES

1.   Weatherall, D.J. and Clegg, J.B.   (1972).   The
      Thalassaemia Syndromes, Second Edition (Oxford: Blackwell
      Scientific Publications).
2.   Flavell, R.A., Kooter, J.M., de Boer, E., Little, P.F.R.
      and Williamson, R.   (1978).   Cell 15, 25 - 41.
3.   Mears, J.G., Ramirez, F., Leibowitcz, D. and Bank, A.
      (1978).   Cell, 15, 15 - 24.
4.   Ottolenghi, S., Comi, P., Giglioni, B., Tolstoshev, P.,
      Lanyon, W.G., Williamson, R., Russo, G., Musumeci, S.,
      Schiliro, G., Tsistrakis, G.A., Charache, S., Wood, W.G.,
      Clegg, J.B. and Weatherall, D.J.   (1976).   Cell 9, 71 - 80.
5.   Ramirez, R., O'Donnel, J.V., Marks, P.A., Bank, A.,
      Musumeci, S., Schiliro, G., Pizzarelli, G., Russo, G.,
      Russo, G., Luppis, G. and Gambino, R.   (1976).   Nature
      263, 471 - 475.
6.   Kan, Y.W., Holland, J.P., Dozy, A.M. and Varmus, H.E.
      (1975).   Proc. Natl. Acad. Sci. U.S.A. 72, 5140 - 5144.
7.   Tolstoshev, P., Mitchell, J., Lanyon, G., Williamson, R.,
      Ottolenghi, S., Comi, P., Giglioni, B., Masera, G.,
      Model, B., Weatherall, D.J. and Clegg, J.B. (1976).
      Nature 259, 95 - 98.
8.   Houseman, D., Forget, B.G., Skoultchi, A. and Benz, E.J.Jr.
      (1973).   Proc. Natl. ACad. Sci. U.S.A.   70, 1809 - 1813.
9.   Kacian, D.L., Gambino, R., Dow, L.W., Grossland, E.,
      Natta, C., Ramirez, R., Spiegelman, S., Marks, P.A. and
      Bank, A.   (1973).   Proc. Natl. Acad. Sci. U.S.A. 70,
      1886 - 1890.
10.   Forget, B.G., Benz, E.J.Jr., Skoultchi, A., Baglioni, C.
      and Houseman, D.   (1974).   Nature 247, 379 - 381.
11.   Old, J.M., Proudfoot, N.J., Wood, W.G., Longley, J.I.,
      Clegg, J.B. and Weatherall, D.J.   (1978). Cell 14,
      289 - 298.
12.   Benz, E.J.Jr., Forget, B.G., Hillman, D.G., Cohen-Solal, M.,
      Pritchard, J., Cavellesco, C., Prensky, W. and
      Houseman, D.   (1978).   Cell 14, 299 - 312.

13. Comi, P., Giglioni, B., Barbarano, L., Ottolenghi, S., Williamson, R., Novakova, M. and Masera, G. (1977). Eur. J. Biochem. 79, 617 - 622.
14. Southern, E.M. (1975). J. Mol. Biol. 98, 513 - 517.
15. Botchan, M., Topp, W. and Sambrook, J. (1976). Cell 9, 269 - 287.
16. Jeffreys, A.J. and Flavell, R.A. (1977). Cell 12, 429 - 439.
17. Little, P., Curtis, P., Coutelle, Ch., Van den Berg., J., Dalgleish, R., Malcolm, S., Courtney, M., Westaway, D. and Williamson, R. (1978). Nature 273, 640 - 643.
18. Marotta, C.A., Wilson, J.T., Forget, B.G. and Weissman, S.M. (1977). J. Biol. Chem. 252, 5040 - 5053.
19. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978). Cell 15, 1157.
20. Orkin, S.H., Alter, B.P., Altay, C., Mahoney, J., Lazarus, H., Hobbins, J.C. and Nathan, D.G. (1978). New England J. Med. 299, 166 - 172.
21. Kan, Y.W. and Dozy, A.M. (1978). Lancet ii, 910.
22. Trendelenburg, M.E. and Gurdon, J.B. (1978). Nature 276, 292 - 294.
23. Flavell, R.A., Sabo, D.L., Bandle, E.F. and Weissman, C. (1974). J. Mol. Biol. 89, 255 - 272.