

The Structure and Expression of Mammalian Gene Clusters

R.A. Flavell, H. Bud, H. Bullman, M. Busslinger, E. deBoer, A. deKleine, L. Golden, J. Groffen, F.G. Grosveld, A.L. Mellor, N. Moschonas, and E. Weiss

INTRODUCTION

The past 5 years has seen a revolution in our understanding of the structure and organization of the genes of higher organisms. Of the models that have been prominent in this period, surely the haemoglobin gene system has been one of the most useful. This system lends itself to this role for a number of reasons: 1) Most of the protein products have been characterized in terms of primary and higher-order structure. 2) The primary DNA sequence of the genes has been elucidated. 3) The changes in the expression of the different genes during development are known. In the case of the human β -related genes, the ϵ gene is expressed in the early embryo, the γ genes in the foetus, and the δ and β genes in the adult. 4) The general arrangement of these genes has been elucidated (Fig. 1). They are present in a cluster on chromosome 11 in the same order as they are expressed during development.

Several defects in the functioning of the globin genes have been defined: 1) Abnormal globin protein can be produced, usually because of point mutations (eg, sickle cell anaemia) and less commonly because of deletions and fusions (eg, haemoglobin Lepore). 2) The amount of the globin proteins can be reduced to a lower level (β^+) or can even be absent (β^0) in thalassemias. A number of these are due to gross alterations in the structure of the β -globin locus (Fig. 1).

Structural analysis of cloned globin genes and indirect restriction mapping performed by genomic blotting experiments have permitted the detailed elu-

cidation of globin gene organization (see, for example, Maniatis et al [1980] for a recent review). A number of generalizations can be made from these observations.

1. *The related genes are clustered in a short region of the chromosome.* They are present in the order 5' $\epsilon^G\gamma^A\gamma\beta$ 3' for the β -related human globin gene locus and $\zeta\alpha_1\alpha_2$ for the human α locus. In these cases, the genes are present on the chromosome in the order in which they are expressed, but this does not appear to be so for the chicken globin genes, where the two adult α genes are flanked on either side by an embryonic α -globin gene. Though the α - and β -related genes are normally on separate chromosomes, this is clearly not obligatory, since in *Xenopus* the α - and β -globin genes are closely linked [Jeffreys et al, 1980]. Within a given cluster the globin genes appear to be transcribed from the same DNA strand.

2. *The distances between the globin genes by far exceed the size of the genes themselves.* Thus, the distances are ϵ –13.5 kb– γ^G –3.5 kb– γ^A –13.5 kb– δ –5.5 kb– β ; each globin gene (including transcribed introns) is about 1.5 kb in length.

3. *Repeated sequences constitute much of the intergenic DNA.* Numerous reiterated sequences are found at several sites within the β -related globin gene cluster. These have been mapped extensively in the case of the rabbit [Shen and Maniatis, 1980; Hoeijmakers-van Dommelen et al, 1980] and human [Fritsch et al, 1980; Coggins et al, 1980; Kaufman et al, 1980; Groffen J, Grosveld, F, deKleine A, and Flavell RA, in preparation] β -related globin genes. Both species contain several different repeated sequences, although the human gene cluster has many more different repeats than the rabbit cluster (about 11 for human versus 5 for rabbit). The arrangements of these reiterated

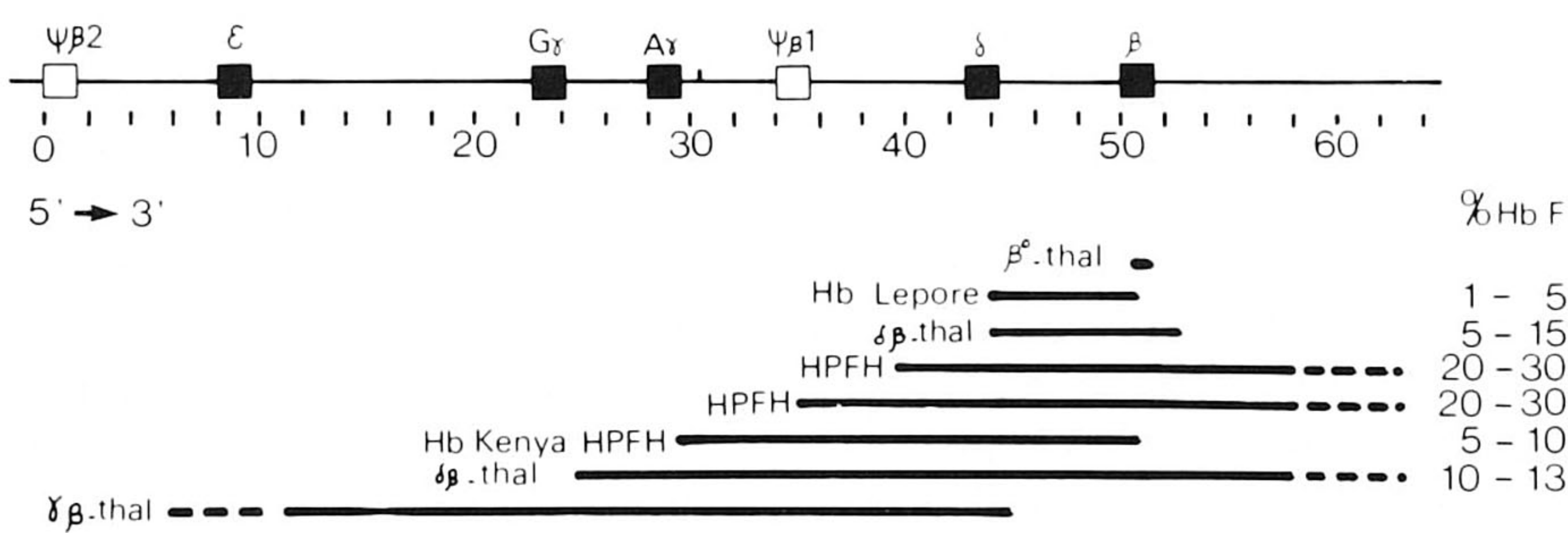


Fig. 1. The structure of the $\epsilon\gamma\delta\beta$ globin region in normal DNA and in a number of hereditary diseases. The DNA deleted in the hereditary diseases indicated is depicted by the horizontal line. A dashed line indicates that the precise location of the end point of the deletion in that area is not known. The deletion in Hb Kenya has not been established at the DNA level. See also Maniatis et al [1980] and Flavell et al [1981].

sequences is interesting, since in at least two locations they seem to be present as clusters of apparently contiguous repeats. Thus between the ϵ and $G\gamma$ -globin genes a long repeated structure is present that contains repeats numbered 1, 2, and 3 in Figure 2. A similar, but not identical, long repeated sequence (first described by Kaufman et al [1980]) is found to the 3' side of the β -globin gene. Here repeats 1 and 2 are present, but this time they are apparently contiguous with repeats 4, 10, and another (called Kpn B in Figure 2). This fluid nature of repeated sequences, where they are found intimately associated with other repeated sequences in the genome, is of great interest.

We know nothing of the function of these repeated sequences. The sequences known as Alu repeats have been the subject of the most attention and they have an intriguing set of properties [Jelinek et al, 1980]: 1) They are homologous both to an abundant class of small nuclear RNAs and to the double-stranded regions of heterogeneous nuclear RNA. This shows that at least some of the repeats, though not necessarily all, are transcribed in vivo. 2) They serve as efficient transcription templates for RNA polymerase III in vitro. 3) They contain a sequence that is homologous to a sequence found near the replication origin of SV40, polyoma, and BK viral DNAs. From the last observation it is tempting to speculate that these sequences may represent cellular origins of replication. There is no evidence, however, to support this idea at present.

GENE CONVERSION IN THE GLOBIN GENES

Comparisons between the DNA sequences of a given gene in different but related organisms have shown that, in general, the intronic DNA sequences diverge more rapidly than the coding sequences. Surprising results have been obtained from a comparison of the structures of two given duplicated genes within the same organism. The two nonallelic γ -globin genes code for proteins that differ at only a single amino acid position; $G\gamma$ -globin has glycine and $A\gamma$ -globin has alanine at position 136. Normally one would expect the DNA sequence to show more differences, since the third position of most codons can be varied without altering the amino acid encoded, and indeed, it is generally true that a nucleic acid sequence is more divergent than the protein sequence. The sequence of both the $G\gamma$ - and $A\gamma$ -globin genes has recently been reported by Slightom et al [1980]; they sequenced the $G\gamma$ and $A\gamma$ genes from one chromosome, together with allelic $A\gamma$ gene from the second chromosome, all cloned from the DNA of a single individual. The sequence of all these genes was strongly conserved, and surprisingly this was also true for both introns. Amazingly, the sequences of the $G\gamma$ and $A\gamma$ genes on a single chromosome were more similar to one another than were

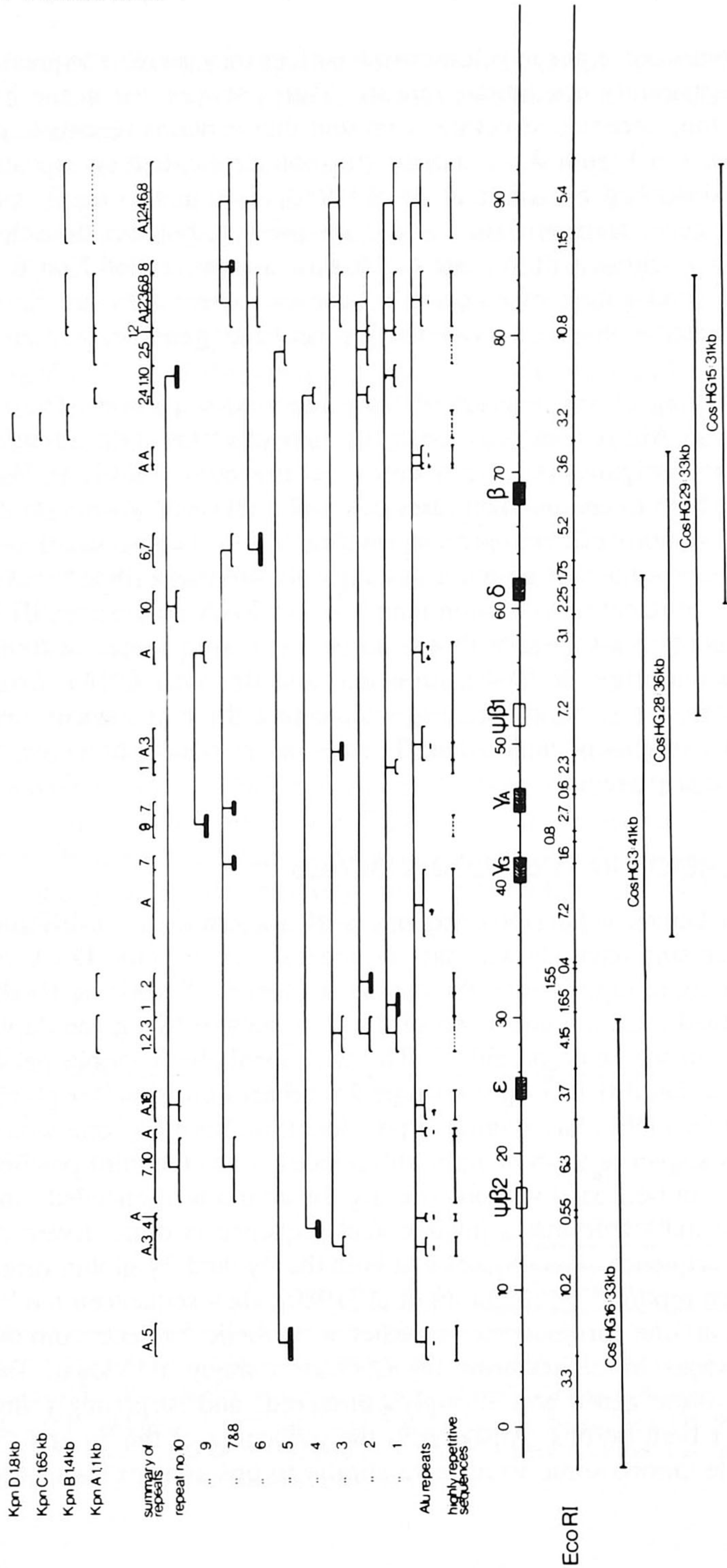


Fig. 2. Classification of intergenic repeat sequences in the human β -globin locus. Repeat sequences were classified in eleven types (Alu and 1 through 10) on the basis of one- and two-dimensional electrophoresis-hybridization experiments. The size of each fragment containing a repeat sequence is illustrated by a bracket as compared to the EcoRI map illustrated below the gene map (indicated by hatched boxes). Fragments containing homologous repeat sequences are connected by a straight line. The cosmid recombinants that were used for this analysis are shown at the bottom of the figure.

the sequences of the two allelic $A\gamma$ genes! A detailed study shows that the latter observation holds for the DNA sequences of the 5' two-thirds of the large intron (the sequences further 5' than this are almost identical in all three genes, so no comparison is possible). In the remainder of the large intron preconceptions apparently prevail, since here the allelic $A\gamma$ genes are more similar to one another than are the $G\gamma$ and $A\gamma$ genes. The DNA sequence at the boundary of these two dissimilar regions is a segment of simple-sequence DNA consisting of alternating TG or CG, and the authors have therefore suggested that this acts as a hot spot for recombination. They suggest that the sequences of the two nonallelic genes on a single chromosome are kept similar by recombination events, which serve to correct the sequence of the one gene against the other. In the model, one of the genes, say the $G\gamma$ gene of chromosome a pairs with the $A\gamma$ gene of chromosome b and after strand cleavage and DNA replication, a converted $A\gamma$ gene is present on the progeny of chromosome G.

There is some evidence that this phenomenon may well be widespread. The two human nonallelic α -globin genes are more similar than expected [Efstratiadis et al, 1980], despite the fact that the α -gene duplication is found in most primates. In addition, Goossens et al [1980] showed that certain humans have three α -globin genes on a single chromosome. This is what would be expected from an unequal crossing-over event and it is one of the specific predictions of the recombination mechanism discussed by Slightom et al [1980]. In addition, good evidence of gene conversion has been obtained for certain yeast genes (reviewed by Baltimore [1981]).

GENETIC ANALYSIS OF GLOBIN GENE EXPRESSION

Structural analysis alone clearly cannot explain the subtleties of gene expression. To understand gene expression we must employ genetics. Two options are available at present. The approach most frequently followed is to analyse the phenotype conferred by mutant DNA molecules that have been generated in vitro—the process that has been called “reversed genetics.” This powerful method has been particularly useful for the study of eukaryotic promoters. The second, classical approach is to study natural mutants. Structural analysis of mutations that confer a specific and well-defined phenotype would alone be very informative. Since, however, detailed genetic analysis of mammals is limited, it is commonly not possible to map mutation precisely, nor to determine an exact phenotype in vivo. The best strategy available, therefore, is to combine classical genetics with reversed genetics. The approach in this case is to clone a gene from an individual with a given phenotype. The structure of the mutant gene is determined and the phenotype conferred by this mutant gene is analysed in detail, using reversed genetics

in a given assay system (see below). In this article, we will illustrate the use of this approach in the study of human genetic diseases.

THE β -THALASSAEMIAS AS A MODEL FOR A HUMAN GENETIC DEFECT

The thalassaemias constitute a heterogeneous collection of diseases in which the functioning of one or more of the β -related globin genes is disturbed. In a number of rarer types of thalassaemia, large deletions are evident. These deletions, as mapped by Southern blotting experiments, are shown in Figure 1 (see also Flavell et al [1979], Fritsch et al [1979], Bernards et al [1979], Ottolenghi et al [1979], Bernards and Flavell [1980], and van der Ploeg et al [1980]). These deletions have helped our understanding (at this time, primitive) of the way globin gene expression is regulated, but since these problems have been considered elsewhere, we will not discuss them in this article. (See Maniatis et al [1980] for a recent review.)

THE MOLECULAR DEFECT IN β^+ AND β^0 THALASSAEMIA

β -Thalassaemia is by far the commonest form of thalassaemia in the β -related globin gene family. β -Thalassaemia is generally divided into two classes. In the first, β^+ -thalassaemia, a low but detectable level of β -globin protein is found; this low level of β -globin protein reflects a low level of β -globin mRNA. It seems likely, a priori, that this disease would be the result of a defect in the transcription or processing of β -globin mRNA precursors. In β^0 -thalassaemia no β -globin protein is detected. In this case, therefore, a number of causes of the genetic lesion can be envisaged. In some cases, globin mRNA is readily detectable [Temple et al, 1977], whereas in others, β -globin mRNA could not be detected by cDNA hybridization techniques (eg, see Ottolenghi et al [1975]). It is likely, therefore, that there are several different molecular forms of β^0 -thalassaemia.

Unlike the diseases discussed above, in β -thalassaemia gross alterations of DNA structure are not usually observed [Flavell et al, 1979; Orkin et al, 1979]. In one case, however, difference can be seen between the normal and thalassaemic globin genes [Flavell et al, 1979]. In this exceptional form of β -thalassaemia the abnormal globin gene had a deletion of about 600 bp, which included the third coding segment of the β -globin gene [Flavell et al, 1979]. The detailed structure of this thalassaemic gene has been elucidated by cloning [Orkin et al, 1980].

A major form of β^0 -thalassaemia found in Southern Italy differs from the above type in that no gross deletion of DNA could be detected; it is characterized by levels of β -globin that are lower than the detection levels of

cDNA hybridization (eg, see Ottolenghi et al [1975] and Tolstoshev et al [1976]). We have analyzed the molecular defect in more detail in this type of β^0 -thalassaemia.

As a source of clinical material we were particularly fortunate in obtaining blood samples from the patient described in Ottolenghi et al [1975]. This patient is a double-heterozygote for $\delta\beta^0/\beta^0$ thalassaemia; on one chromosome, therefore, a large deletion has occurred that eliminates the entire β -globin gene. Studies of β -globin gene expression in this patient, therefore, only relate to the single β -globin gene present.

We have cloned the β -globin gene from this patient as a 7.5-kb HindIII fragment and have determined the entire DNA sequence of this gene. A comparison of this sequence with that of the normal β -globin gene shows that the only difference in the entire gene (including both intervening sequences) is a single base substitution that generates a stop codon at the position corresponding to amino acid residue 39 (Fig. 3). A similar result has been obtained by Trecartin et al [1981] for a case of β^0 -thalassaemia from Sardinia. This result is similar to that described previously in a Chinese β^0 -thalassaemic; in this case, a termination codon was present at amino acid 17 [Chang and Kan, 1979].

This result clearly explains why this β -globin gene is thalassaemic. Why, then, does this patient not produce detectable levels of β -globin mRNA? To reexamine this point, we have probed for α -, β -, and γ -globin mRNA using sensitive S_1 -nuclease mapping procedures (not shown). We have shown that β -globin mRNA is in fact detectable by these methods, but that this level is very low. A comparison of the level of β -globin mRNA with that of the α - or γ -globin mRNAs suggests that only about 2–5% of the normal level of β -globin mRNA was present.

A reasonable, but untested, explanation for the low level of β -globin mRNA in this patient is that the failure of this mRNA to be translated effectively causes it to be more susceptible to nucleases in the cytoplasm and therefore to be degraded rapidly after synthesis.

PHENOTYPIC ASSAY OF INHERITED DEFECTS IN HUMAN GENES

The structural analysis of the type of β^0 -thalassaemia discussed above obviously gives a clear answer. A stop codon present in the β -globin gene causes the β^0 -thalassaemic phenotype. Yet it is likely that molecular lesions will exist in which the molecular defect is not so easy to discern, either because the mutation detected by structural analysis is not clearly interpretable, or because a large number of polymorphic differences are present that mask the true mutation that causes the disease. To solve this problem, it is essential to correlate the structural data with phenotypic expression.

We have recently used the techniques of reversed genetics to study β -thalassaemia. The approach followed is shown in Figure 4 and derives from the work of Banerji et al [1981]. A globin gene is linked to a segment of SV40 that contains the DNA sequences around the origin cloned in a bacterial plasmid. This triple recombinant is then introduced into Hela cells by calcium phosphate coprecipitation. A high proportion of the cells take up the DNA and up to 30% of those cells express the SV40 early region as estimated by immunofluorescence. We have shown that under these conditions the rabbit β -globin gene is expressed specifically from its own promotor by two criteria. First, the expression of the β -globin gene is independent of the orientation of the β -globin gene with respect to the SV40 DNA sequences. Second, the 5' ends of the β -globin RNA map at the natural cap site of the β -globin gene.

Clearly this assay can be used to study the expression of a defective globin gene from a thalassaemic patient. To investigate this possibility, we have recently studied β^+ -thalassaemia. The DNA sequence of the β -globin gene has been determined for two cases of β^+ -thalassaemia [Spritz et al, 1981; Westaway and Williamson, 1981]. Interestingly, both genes show a single-base substitution when compared with the normal β -globin gene—TTGG \rightarrow TTAG at a site 21 nucleotides to the 5' side of the 3' splice junction of the small intervening sequence of the β -globin gene. Among the possible phenotypes that could be postulated to explain this result, the most interesting is that this generates a new splice acceptor site in the small intervening sequence. We have recently established this using the SV40-Hela cell system [Busslinger et al, 1981]. S₁ mapping of the β -globin RNA produced in Hela cells from a β^+ -thalassaemic β -globin gene shows that about 95% of the β -globin mRNA produced is spliced incorrectly, that is, to this abnormal splice site (Fig. 5). To establish the nature of the aberrant RNA produced, we have used a 5' labeled DNA fragment as a primer to synthesize, with reverse transcriptase, a cDNA complementary to this β^+ -globin RNA. This cDNA has been purified and its DNA sequence determined. As a control, cDNA synthesized from globin mRNA from a patient with sickle cell haemoglobin was also sequenced (Fig. 6). It can be seen that the abnormal splicing of the small intron generates a β -globin mRNA with an additional 19 nucleotides that derive from the small intervening sequence; since this is not a multiple of 3, this insertion results in the alteration of the translational reading frame and in fact generates an in-phase stop codon five amino acids after the splice.

From the above, it should be clear that a combination of structural and functional analyses will enable us in the future to elucidate the molecular nature of defects in a large number of human diseases. In the examples provided here, we show that defects at the level of protein synthesis and RNA processing can be elucidated by these systems. Since the globin mRNA

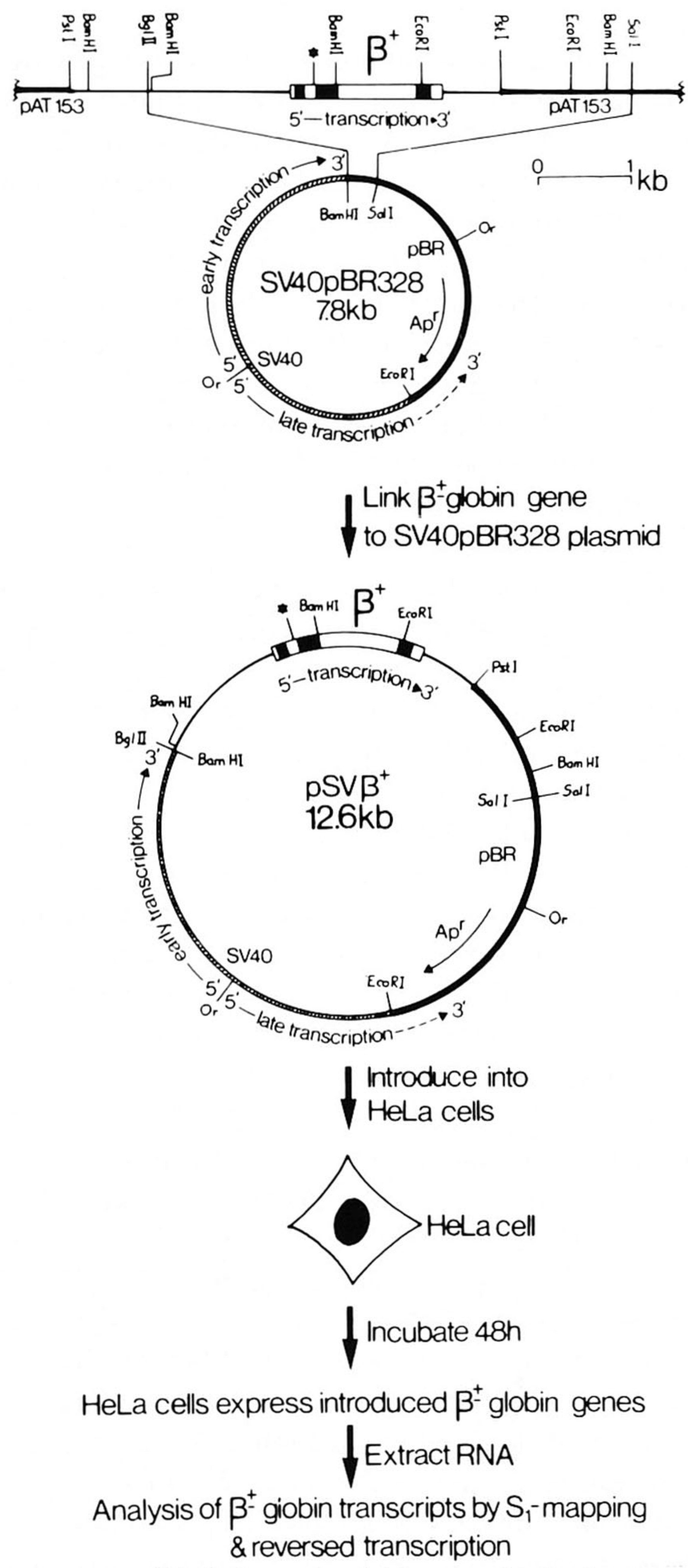


Fig. 4. Procedure to test a β^+ -thalassemic globin gene for its functional defect. The β^+ -globin gene was originally cloned into the plasmid pAT153 by Westaway and Williamson [1981] from the DNA of a Turkish-Cypriot patient diagnosed as doubly heterozygous for β^+ -thalassemia and Hb Lepore. The construction of the recombinant plasmid pSV β^+ is described in detail in Busslinger et al [1981]; pSV β^+ DNA was introduced into HeLa cells by the calcium phosphate coprecipitation method [Wigler et al, 1977]. From Busslinger et al [1981] with permission of MIT press. Hatched lines denote SV40 DNA; solid lines correspond to plasmid sequences and thin lines to human DNA. The β -globin coding sequences are represented by

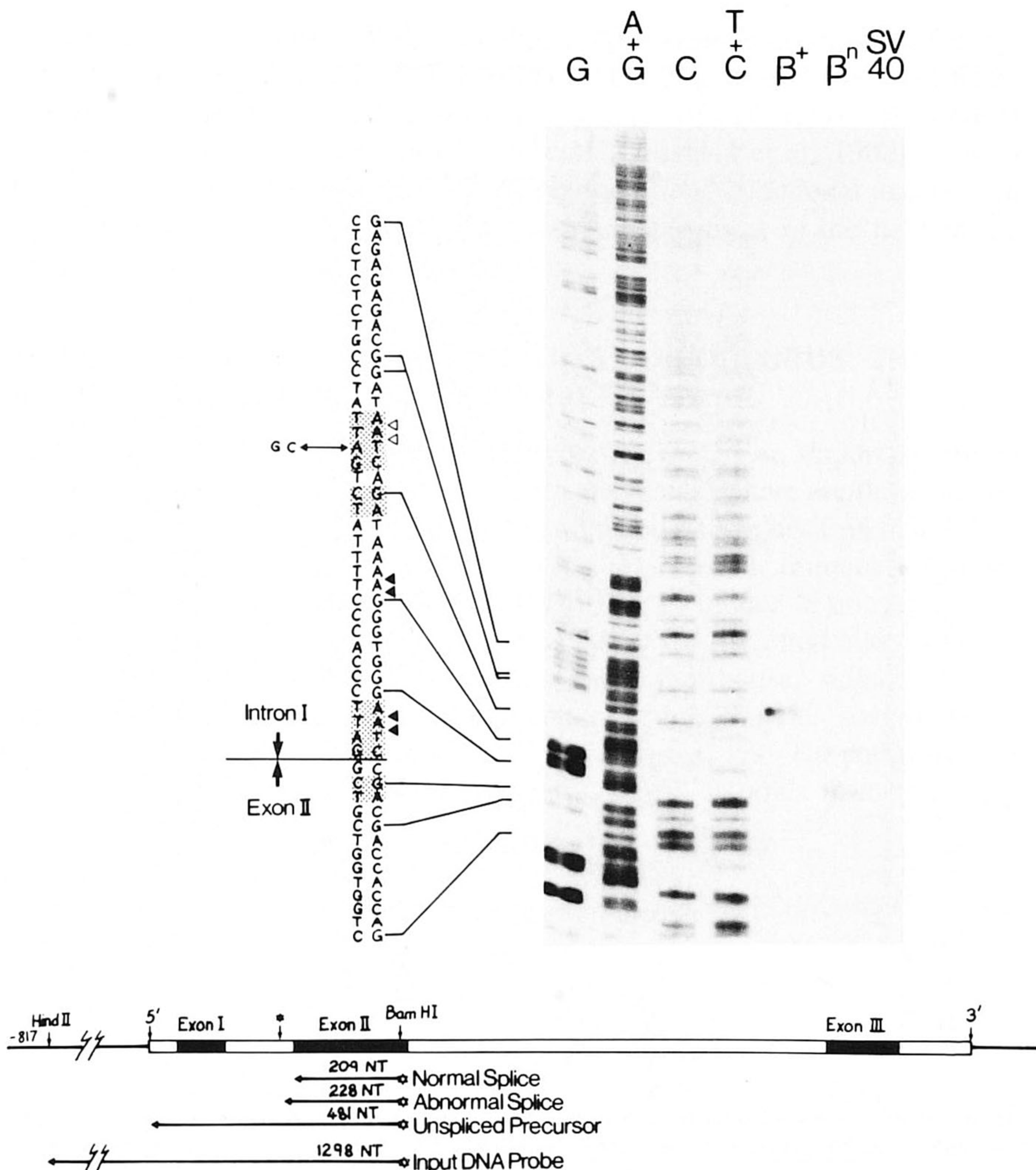


Fig. 5. S_1 mapping of the splice between the first intron and second exon of the β^+ -globin mRNA: S_1 nuclease digestion of hybrids at 20°C and 40°C. The HindII/BamHI DNA probe shown in the scheme was hybridized to RNA from HeLa cells transformed with either the β^+ -thalassemic (pSV β^+ DNA) or the normal β -globin gene (pSV β^n DNA). As a control, RNA from HeLa cells transformed with the SV40-pBR328 vector alone was used. Hybrids were digested with S_1 nuclease either at 40°C for 1 hr or at 20°C for 2 hr. The protected DNA fragments (lanes β^+ , β^n , SV40) were separated on an 8% polyacrylamide sequencing gel. IP denotes input DNA probe. 5' end-labeled pBR322 DNA \times HpaII was used as a size marker (M). The position of the input DNA probe as well as of the S_1 nuclease-resistant DNA fragments are depicted as arrows below the map of the β^+ -globin gene. From Busslinger et al [1981] with permission of MIT press.

filled blocks; introns and sequences coding for the untranslated regions of the β -globin mRNA are shown by open blocks. The star above the first intron indicates the position where the normal and the β^+ -thalassemic globin genes differ by a single base-change (GC \leftrightarrow AT; Westaway and Williamson [1981]).

produced in these cells is spliced, polyadenylated, and exported to the cytoplasm [Busslinger et al, 1981], it is likely that defects at the level of RNA transport can also be studied. Elsewhere, we have described how this system can be used to study transcriptional defects [Grosveld et al, 1982]. It is to be anticipated that the combination of structural and functional analyses in the next few years will bring some exciting discoveries in the field of the molecular biology of human diseases.

MOLECULAR ANALYSIS OF A COMPLEX GENE LOCUS: THE MAJOR HISTOCOMPATIBILITY COMPLEX

The major histocompatibility complex (MHC) plays an important role in the regulation of the immune response in vertebrates. There are three classes of MHC proteins: the classical transplantation antigens (class I: in man HLA-A, -B, and -C; in mouse the H-2 K, L, and D proteins), immune response-associated antigens (class II: in man HLA-DR; in mouse Ia antigens), and the complement products (class III). The genes for the three classes of antigens are closely linked on chromosome 17 of the mouse. Adjacent to the H-2 complex in mouse is the TL complex, which contains several genes encoding lymphoid differentiation antigens, the Qa, and Tla antigens. See Figure 7 for a schematic genetic map of the MHC complex in mouse.

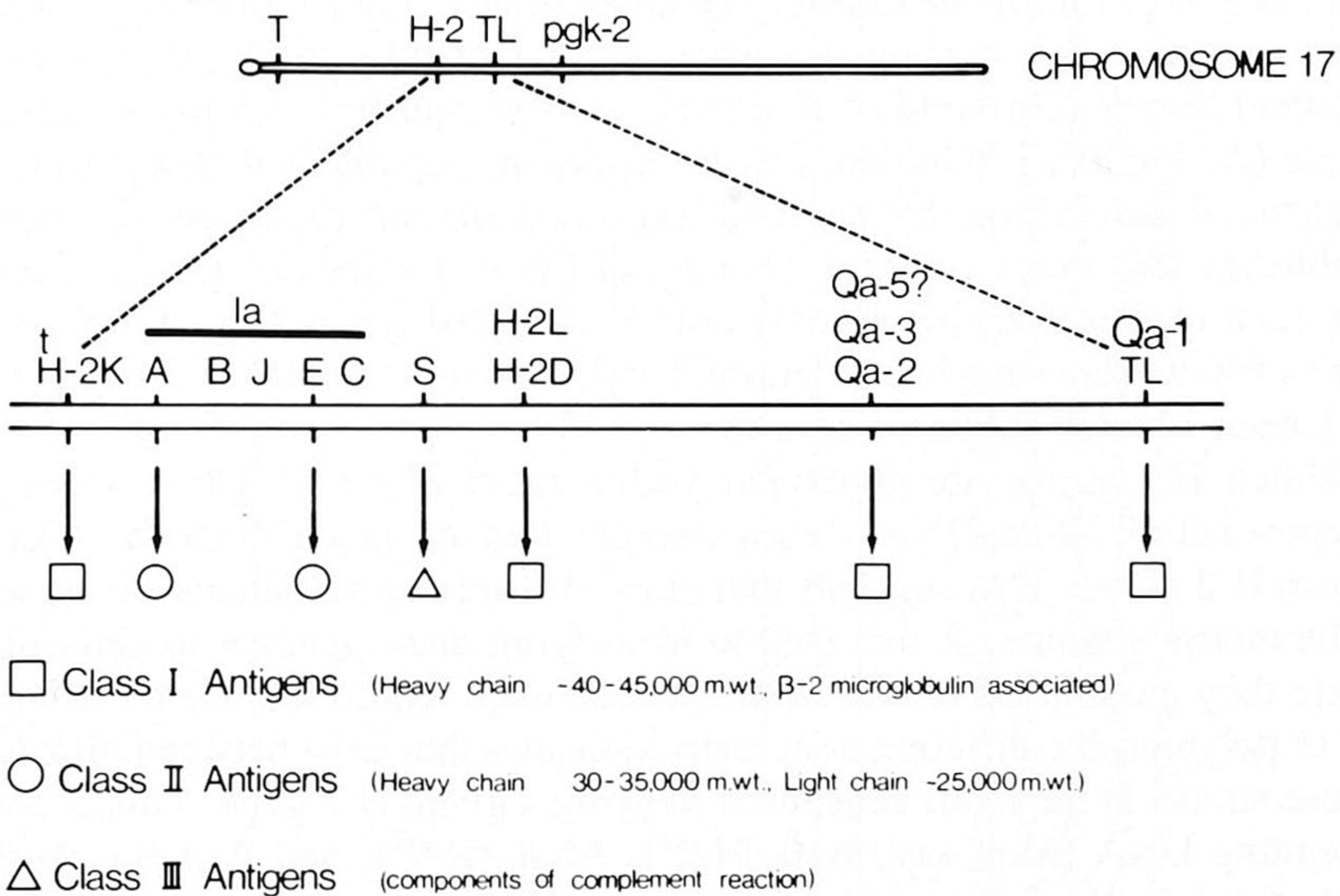


Fig. 7. Schematic genetic map of the H-2 and TL complexes of chromosome 17 of the mouse. The relative order of the loci listed vertically is not known. The distance from H-2K to H-2D is about 0.3 cM and from H-2D to Tla is about 1 cM.

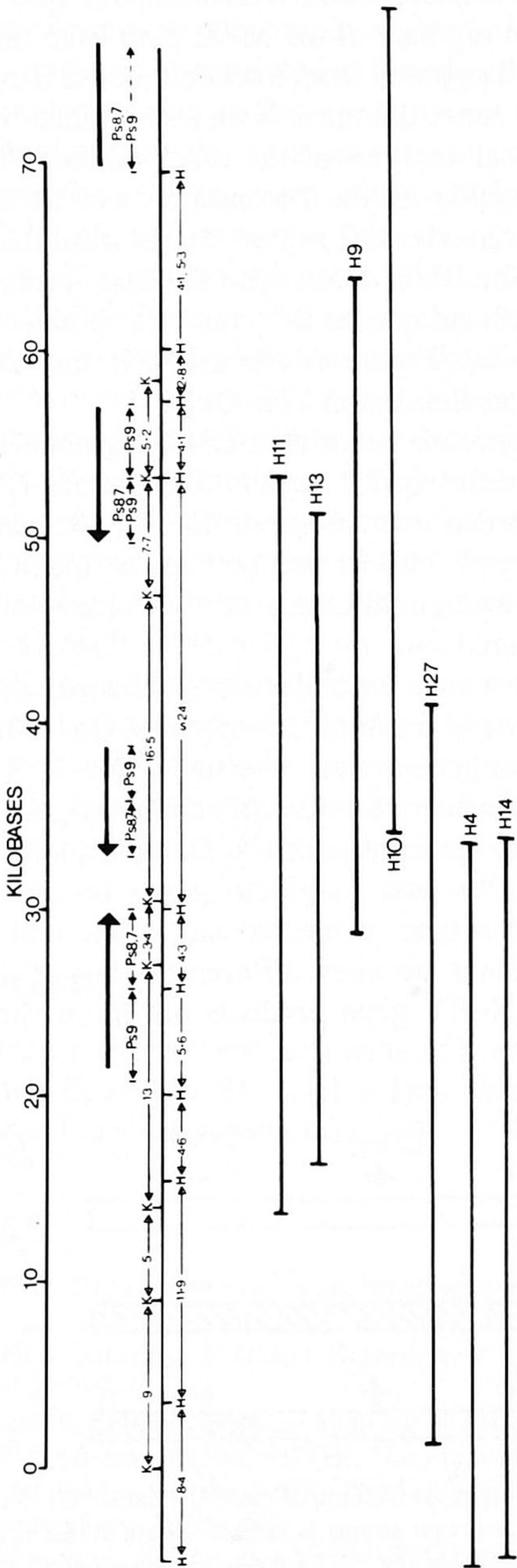
Both the H-2 antigens (K, D, L, and R) and the Qa and Tla are integral membrane proteins of 40,000–45,000 molecular weight found associated with β_2 -microglobulin. It is therefore possible that these antigens are evolutionarily related (see, for example, Michaelson et al [1977] and Stanton and Hood [1980]).

One of the most intriguing aspects of the H-2 antigens is their extreme polymorphism. Allelic products from different inbred mouse strains differ in most cases as shown by both serologic analyses and protein sequencing. In fact, about 50 alleles at each of the H-2K and H-2D loci have been observed (for a review see Klein [1979]). H-2 haplotypes are defined serologically and consist of a particular set of alleles found at each H-2 locus (eg, the C57 black/10, or B10, mouse is H-2^b at each H-2 locus, the AKR strain H-2^k, DBA/2 H-2^d and so on). Some inbred mouse strains appear to be natural recombinants between other strains (eg, the A mouse strain with the H-2^a haplotype is H-2K^k but H-2D^d). It is likely that the analysis of the gene clusters by recombinant DNA technology will provide some clues to the nature of the polymorphic differences.

A number of groups have isolated cDNA clones that encode proteins whose sequences are homologous to known transplantation antigens (see, for example, Kvist et al [1981] and Steinmetz et al [1981]). Southern blotting analysis of genomic DNA probed with these cDNAs shows many cross-hybridizing DNA fragments, a finding that reinforces the notion that the MHC is a large multigene family. We have isolated a large number of mouse genomic clones that contain sequences complementary to H-2 cDNA from a cosmid library [Grosveld et al, 1981] by using spleen DNA from the B10 mouse (A. Mellor, L. Golden, and H. Bullman, unpublished observations). Analysis of these clones by standard restriction enzyme digestion techniques establishes that many of these clones fall into a number of gene clusters. One such cluster contains at least four H-2-related genes (Fig. 8). Several other such clusters have been identified and about 15 different H-2-like genes have been isolated in these cosmids.

Which H-2 genes are contained within these clusters? The number of independent H-2-like DNA clones exceeds that expected from the 4 or 5 known H-2 genes. This suggests that other H-2-related sequences are present in the mouse genome. A first step to identifying these genes is to determine where they are located on the MHC genetic map. To do this we have made use of polymorphic differences at restriction sites that exist between different mouse strains in the DNA sequences flanking a given H-2 gene. This is done by cutting DNA from, say, B10 (H-2^b), AKR (H-2^k), and BALB/c (H-2^d) mice with a number of restriction enzymes and blotting this onto nitrocellulose filters. The filters are then probed with a fragment from a given cosmid in order to detect polymorphic restriction site differences between the parental inbred strains. For example, using a probe from a clone in the cluster depicted

H11 COSMID FAMILY



in Figure 8 in PstI digests of these mouse DNAs shows a 2.5-kb band in B10 DNA, a 3.8-kb band in AKR DNA. Recombinant mice have been produced that contain a given segment of the MHC map from one haplotype (say AKR; H-2^k) and a second segment from a second mouse (say C57 black/6 or B6; h-2^b). The origin of genes that have been recombined in these mice has been shown by serological analysis of the recombinants. Thus, in the recombinant mouse B6K2 (see Fig. 9) the Tla and Qa1 loci are derived from the AKR (H-2^k) mouse, whereas the H-2 region comes from the B6 mouse. Blotting PstI-cut DNA from the B6K2 DNA with the same probe shows that the AKR pattern (3.8-kb band) and not the B6 pattern is obtained. It follows, therefore, that the gene cluster of Figure 8 is localized in the AKR segment of the B6K2 mouse, that is, in the region Tla, Qa1.

We have used the same approach to localize DNA segments in the H-2K region, the H-2D region, and the Qa2,3 region. Thus, class I, H-2 genes, or pseudogenes are found spread over chromosome 17 between the H-2K and TL genetic loci; a region of 1.3 cM and perhaps as much as 10⁷ base pairs of DNA. This gene cluster might then be two orders of magnitude larger than the β -globin cluster.

There is little information yet as to the molecular mechanism that generates the polymorphism characteristic of this multigene family. One clue, however, might come from the β -globin gene system. As stated above, the evolution of the β -related globin gene has been strongly influenced by gene conversion which has resulted in the homogenization of the DNA sequence of, in one case for example, nonallelic ^G γ - and ^A γ -globin genes on the same chromosome. We consider that this type of mechanism is one that is likely to generate polymorphic differences between different alleles of a given H-2 gene. Thus, the H-2K^b and H-2D^b gene products are more similar to one

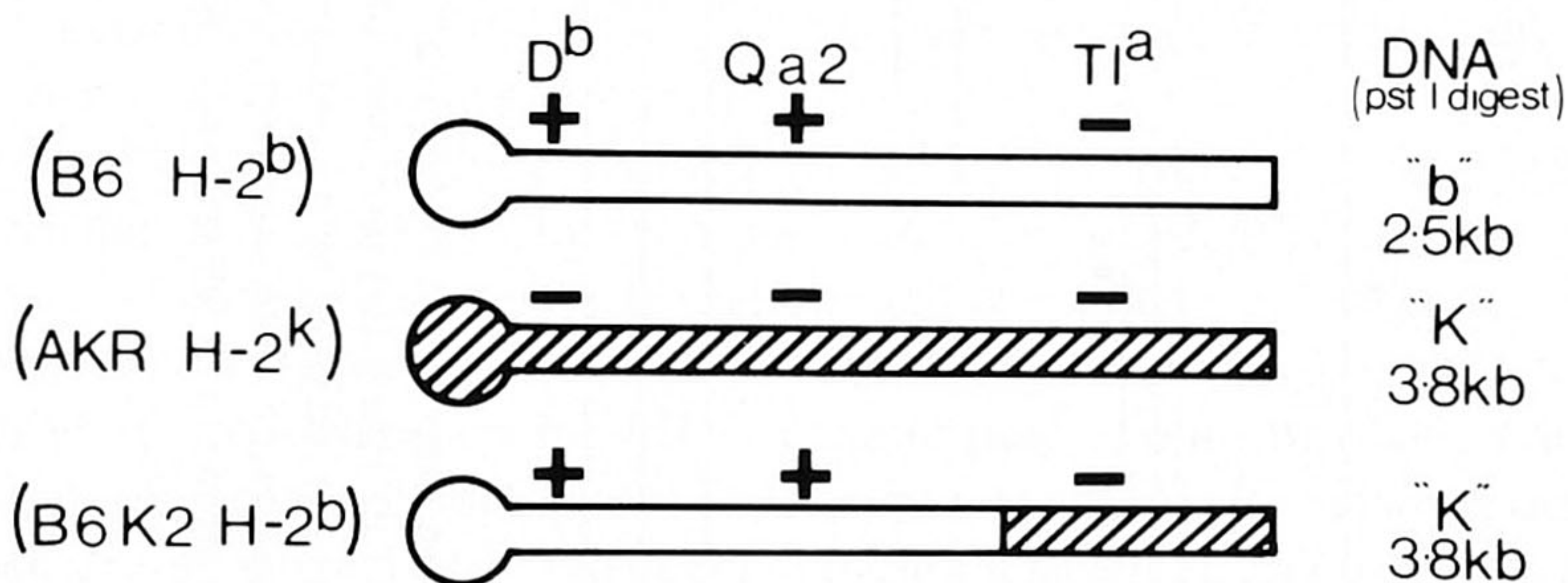


Fig. 9. Scheme for the determination of the location of the H-2-related "H11" cosmid family. The chromosomal makeup of three relevant strains is shown. Strain B6K2 is a recombinant that has Qa1 and Tla region derived from the AKR mouse. Southern blots of B6K2 DNA probed with a labeled DNA fragment show the AKR (3.8-kb fragment) and not the B6 (2.5-kb) pattern. The H11 gene cluster must therefore lie in the Qa1 Tla region.

another than are the H-2K^b and H-2K^k gene products; Baltimore [1981] has suggested that this type of similarity might result from gene conversion events between the two genes in question in a given haplotype. Clearly such gene conversion events are not restricted to functional genes alone and, to us, one particularly attractive idea is that pseudogenes are used to donate new sequences to functional genes by a conversion event. This can occur by the copying of a nondefective segment of a given pseudogene into the functional H-2 gene. There are two attractive features of such a mechanism to create diversity. First, if many pseudogenes exist, the number of possible new combinations that will generate a new gene are enormous, since different segments, say 5' or 3', of different pseudogenes or genes can be used in each conversion event. Second, the pseudogene sequences will accumulate mutations that in turn can be brought into the active gene by gene conversion; in this way the effective mutation rate for the functional gene will be increased. Such a model for the generation of diversity is directly testable because gene conversion events are not restricted to the coding segments of genes; in the case of the human γ -globin genes, it will be recalled that the introns had also been corrected by this mechanism. If gene conversion has played a role, we will expect to find a similar level of homology between the nonfunctional gene segments (introns 3' untranslated regions) and exons of the nonalleles that have taken part in the conversion event; and we will expect to find that the sequence differences between the alleles (say H-2K^b and H-2K^d) will be similar for the exons and the nonfunctional segments of the gene.

ACKNOWLEDGMENTS

The following persons were recipients of postdoctoral fellowships: M.B. (Swiss National Funds), L.G. (Helen Hay Whitney), N.M. (EMBO), and E.W. (Deutsche Forschungsgemeinschaft).

REFERENCES

- Baltimore D (1981): Gene conversion: Some implications for immunoglobulin genes. *Cell* 24:592-594.
- Banerji J, Rusconi S, Schaffner W (1981): Expression of a genomic segment of simian virus 40 DNA. *Cell* 27:299-304.
- Bernards R, Flavell RA (1980): Physical mapping of the globin gene deletion in hereditary persistence of foetal haemoglobin (HPFH). *Nucleic Acids Res* 8:1521-1534.
- Bernards R, Kooter JM, Flavell RA (1979): Physical mapping of the globin gene deletion in ($\delta\beta$)⁰thalassaemia. *Gene* 6:265-280.
- Busslinger M, Moschonas N, Flavell RA (1981): β^+ thalassaemia: Aberrant splicing results from a single point mutation in an intron. *Cell* 27:289-298.
- Chang JC, Kan YW (1979): β^0 thalassemia, a nonsense mutation in man. *Proc Natl Acad Sci USA* 76:2886-2889.

- Coggins LW, Grindley GJ, Voss JK, Slater A-A, Montagu P, Stinson MA, Paul J (1980): Repetitive DNA sequences near three human β -type globin genes. *Nucleic Acids Res* 8:3319-3333.
- Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeRiel JK, Forget B, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot, NJ (1980): The structure and evolution of the human β -globin gene family. *Cell* 21:653-668.
- Flavell RA, Bernards R, Kooter JM, deBoer E, Little PFR, Annison G, Williamson R (1979): The structure of the human β -globin gene in thalassaemia. *Nucleic Acids Res* 6:2749-2760.
- Flavell RA, Bud H, Bullman H, Dahl H, deBoer E, deLange T, Groffen J, Grosveld F, Grosveld G, Kioussis D, Moschonas N, Shewmaker C (1981): Globin gene expression *in vivo* and *in vitro*. In Stamatoyannopoulos G, Nienhuis AW (eds): "Organization and Expression of Globin Genes." New York: Alan R. Liss, p 119-126.
- Fritsch EF, Lawn RM, Maniatis T (1979): Characterization of deletions which affect the expression of fetal globin genes in man. *Nature* 279:598-603.
- Fritsch EF, Lawn RM, Maniatis T (1980): Molecular cloning and characterization of the human β -like globin gene clusters. *Cell* 19:959-972.
- Goossens M, Dozy AM, Embury SH, Zachariades Z, Jadjiminas MG, Stamatoyannopoulos G, Kan YW (1980): Triplicated α -globin loci in humans. *Proc Natl Acad Sci USA* 77:518-521.
- Grosveld FG, Dahl HHM, DeBoer E, Flavell RA (1981): Isolation of β -globin-related genes from a human cosmid library. *Gene* 13:227-237.
- Grosveld GC, DeBoer E, Shewmaker CK, Flavell RA (1982): DNA sequences necessary for the transcription of the rabbit β -globin gene *in vivo*. *Nature* 295:120-126.
- Hoeijmakers-Van Dommelen HAM, Grosveld GC, DeBoer E, Flavell RA, Varley JM, Jeffreys AJ (1980): The localization of repetitive and unique DNA sequences neighboring the rabbit β -globin gene. *J Mol Biol* 140:531-547.
- Jeffreys AJ, Wilson V, Wood D, Simons JP, Kay RM, Williams JG (1980): Linkage of adult α - and β -globin genes in *X. laevis* and gene-duplication by tetraploidization. *Cell* 21:555-564.
- Jelinek WR, Toomey TR, Leinwand L, Duncan CH, Biro PA, Choudary PN, Weissman SM, Rubin CM, Houck CM, Deininger PL, Schmid CW (1980): Ubiquitous, interspersed repeated sequences in mammalian genomes. *Proc Natl Acad Sci USA* 77:1398-1402.
- Kaufman RE, Kretschmer PJ, Adams JW, Coon HC, Anderson WF, Nienhuis AW (1980): Cloning and characterization of DNA sequences surrounding the human γ , δ and β -globin genes. *Proc Natl Acad Sci USA* 77:4229-4233.
- Klein J (1979): The major histocompatibility complex of the mouse. *Science* 203:516-521.
- Kvist S, Bregegere F, Rask L, Cami B, Garoff H, Daniel F, Wiman K, Larhammar D, Abastado JP, Gachelin G, Peterson PA, Dobberstein B, Kourilsky P (1981): cDNA clone coding for part of a mouse H-2^d major histocompatibility antigen. *Proc Natl Acad Sci* 78:2772-2776.
- Maniatis T, Fritsch EF, Lauer J, Lawn RM (1980): The molecular genetics of human haemoglobins. *Ann Rev Genet* 14:145-178.
- Maxam AM, Gilbert W (1980): Sequencing end-labelled DNA with base-specific chemical cleavages. In Grossman L, Moldave K (eds): "Methods in Enzymology." New York: Academic Press, p 499-560.
- Michaelson J, Flaherty L, Vitetta E, Poulik M (1977): Molecular similarities between the Qa-2 alloantigen and other gene products of the 17th chromosome of the mouse. *J Exp Med* 145:1066.
- Orkin SH, Old JM, Weatherall DJ, Nathan DG (1979): Partial deletion of β -globin gene DNA in certain patients with β -thalassemia. *Proc Natl Acad Sci USA* 76:2400-2404.
- Orkin SH, Kolodner R, Michelson A, Husson R (1980): Cloning and direct examination of a structurally abnormal human β^0 -thalassemia globin gene. *Proc Natl Acad Sci USA* 77:3558-3562.

- Ottolenghi S, Lanyon WG, Williamson R, Weatherall DJ, Clegg JB, Pitcher CS (1975): Human globin gene analysis for a patient with $\beta^0/\delta\beta^0$ -thalassemia. *Proc Natl Acad Sci USA* 72:2294–2299.
- Ottolenghi S, Giglioni B, Comi P, Gianni AM, Polli E, Acquaye CTA, Oldham JH, Masera G (1979): Globin gene deletion in HPFH, $\delta\beta$ thalassaemia and Hb Lepore disease. *Nature* 278:654–657.
- Shen C-KJ, Maniatis T (1980): The organization of repetitive sequences in a cluster of rabbit β -like globin genes. *Cell* 19:379–392.
- Slightom J, Blechl AE, Smithies O (1980): Human fetal G γ and A γ globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these two genes. *Cell* 21:627–638.
- Spritz RA, Jagadeeswaran P, Choudary PV, Biro PA, Elder JT, deRiel JK, Manley JL, Geffer ML, Forget BG, Weissman SM (1981): Base substitution in an intervening sequence of a β^+ -thalassemic human globin gene. *Proc Natl Acad Sci USA* 78:2455–2459.
- Stanton TH, Hood L (1980): Biochemical identification of the Qa-1 alloantigen. *Immunogenetics* 11:309–314.
- Steinmetz M, Frelinger JG, Fisher D, Hunkapiller T, Pereira D, Weissman SM, Uehara H, Nathenson S, Hood L (1981): Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell* 24:125–134.
- Temple GF, Chang JC, Kan YW (1977): Authentic β -globin mRNA sequences in homozygous β^0 thalassemia. *Proc Natl Acad Sci USA* 74:3047–3051.
- Tolstoshev P, Mitchell J, Lanyon G, Williamson R, Ottolenghi S, Comi P, Giglioni B, Basera G, Modell B, Weatherall DJ, Clegg JB (1976): Presence of gene for β -globin in homozygous β^0 thalassaemia. *Nature* 259:95–98.
- Trecartin RF, Liebhaber SA, Chang JC, Lee KY, Kan YW, Furbetta M, Angius A, Cao A (1981): β^0 thalassemia in Sardinia is caused by a nonsense mutation. *J Clin Invest* 68:1012–1017.
- Van der Ploeg LHT, Konings A, Oort M, Roos D, Bernini LF, Flavell RA (1980): γ - β -thalassaemia: deletion of the γ - and δ -genes influences β -globin gene expression in man. *Nature* 283:637.
- Westaway D, Williamson R (1981): An intron nucleotide sequence variant in a cloned β^0 -thalassaemia globin gene. *Nucleic Acids Res* 9:1777–1788.
- Wigler M, Sweet R, Sim GK, Wold B, Pellicer A, Lacy E, Maniatis T, Silverstein S, Axel R (1979): Transformation of mammalian cells with genes from procaryotes and eucaryotes. *Cell* 16:777–785.