Structure and Expression of the Human Globin Genes and Murine Histocompatability Antigen Genes

R.A. FLAVELL, F. GROSVELD, M. BUSSLINGER, E. DE BOER, D. KIOUSSIS, A.L. MELLOR, L. GOLDEN, E. WEISS, J. HURST, H. BUD, H. BULLMAN, E. SIMPSON,* R. JAMES,† A.R.M. TOWNSEND,‡ P.M. TAYLOR, ‡ W. SCHMIDT,§ J. FERLUGA,§ L. LEBEN,§ M. SANTAMARIA,‡ G. ATFIELD,§ AND H. FESTENSTEIN§

*Laboratory of Gene Structure and Expression, National Institute for Medical Research, Mill Hill, London NW7 1AA; *Clinical Research Centre, Northwick Park Hospital, Harrow, Middlesex HA1 3UJ; †Imperial Cancer Research Fund Tumour Immunology Unit, University College, London WC1E 6BT; ‡Division of Immunology, National Institute for Medical Research, Mill Hill,, London NW7 1AA; §Department of Immunology, The London Hospital Medical School, University of London, London E1 2AD*

The isolation of mammalian genes by molecular cloning and the development of the techniques of DNA-mediated transfer have made it possible to introduce genes into essentially any cell type and to study their expression at the sites of insertion. Globin genes have featured prominently in such experiments since they offer the potential for the study of those factors that govern tissue-specific gene expression and the modulation of gene expression during development; in addition, the globin system permits study of the general features of gene expression, such as the promotion of RNA synthesis.

We have studied the human β-globin system for the past five years. Structural studies (for review, see Maniatis et al. 1980) show that these β-related globin genes are found in a short segment of chromosome 11; there are five active globin genes present on the chromosome in the following 5′ to 3′ order: ε, expressed in early embryos; $^{G}\gamma$ and $^{A}\gamma$, expressed in the fetus; and δ and β, expressed in the adult (Fig. 1). There is a single pseudogene localized between the $^{A}\gamma$-globin and the δ-globin genes. We have cloned this locus as a series of overlapping cosmids spanning a total of about 120 kb of DNA (Fig. 1) (Grosveld et al. 1981; F. Grosveld and D. Kioussis, unpubl.).

Introduction of a given globin gene into mouse L cells results in constitutive gene expression (Mantei et al. 1979; Wold et al. 1979). We have developed a cosmid-cloning system in which cosmid vectors containing genes that are selectable in animal and bacterial cells are used to generate libraries of genomic DNA. Thus, vectors carrying the herpes simplex virus (HSV) thymidine kinase (*tk*) gene have been developed for use in tk$^-$ cell lines (Grosveld et al. 1982). Likewise, we have constructed and used vectors that carry the dominant selection markers aminoglycoside-3′-phosphotransferase and xanthine guanine phosphoribosyl transferase, which confer resistance to the antibiotic G418 (Jiminez and Davies 1980) and mycophenolic acid (Mulligan and Berg 1980), respectively (Grosveld et al. 1982).

In this way, genomic clones carrying the human $^{G}\gamma$-, $^{A}\gamma$-, δ-, and β-globin genes have been introduced into mouse L cells using the *tk*-selection system (H.H. Dahl et al., unpubl.). Under these conditions, the γ- and β-globin genes are expressed constitutively at levels up to 5000 copies per cell from a few copies of the cosmid DNA. Interestingly, the δ-globin-gene product is barely detectable under these conditions, which shows that the low level of expression characteristic of the δ-globin gene in vivo is also seen using transfected DNA. Need-

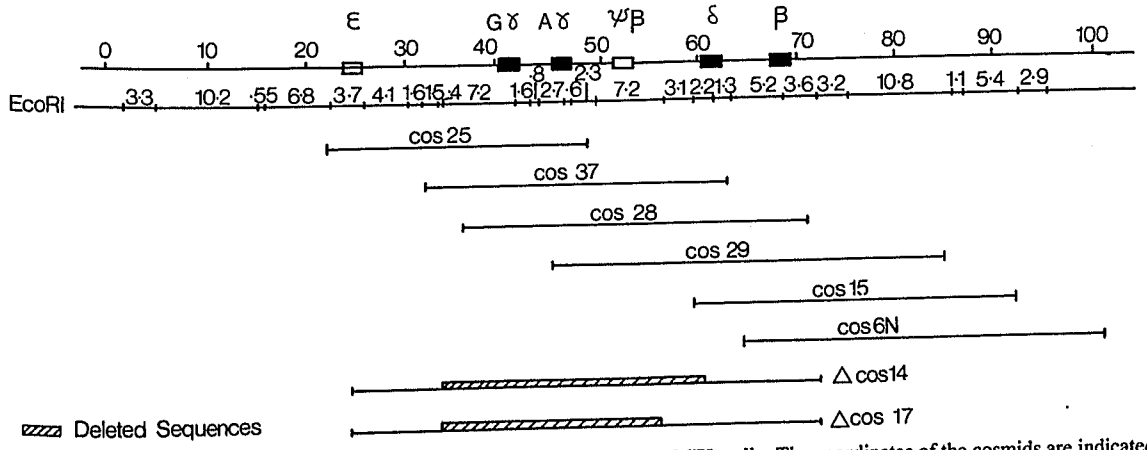

Figure 1. A map of human β-globin, showing cosmid clones used to transform MEL cells. The coordinates of the cosmids are indicated by the solid lines below the map. The $\psi\beta_2$ gene described previously (A. de Kleine and F.G. Grosveld; S. Shen and O. Smithies; both unpubl.) appears to be an artifact.

less to say, the endogenous globin genes of these non-erythroid cells are not expressed. It is therefore disturbing that the transfected genes are expressed in this non-physiological manner.

Specific Expression of the Human β-globin Gene in Murine Erythroleukemia Cells

The natural site of globin mRNA synthesis is the erythroid cell. These cells differentiate from the hematopoietic stem cell via a series of cell divisions; in the final stages of this process, globin mRNAs are made. The murine erythroleukemia (MEL) cell line ("Friend" cells) provide a useful model for the late stages of erythroid differentiation. Although these cells make little or no hemoglobin normally, upon induction with a variety of simple chemicals such as DMSO or hexamethylene-bisacetamide (HMBA), hemoglobin mRNA and protein synthesis are detected. We have examined the expression of transfected human globin genes in MEL cells before and after induction of erythroid differentiation. Willing et al. (1979) have shown previously that specific induction of the human β-globin gene is seen in MEL cells when it is introduced by the transfer of chromosome 11. Interestingly, γ-globin mRNA could not be found. We have introduced the cosmid DNAs shown in Figure 1 into a *tk*⁻ MEL cell line (Spandidos and Paul 1982) and selected *tk*⁺ colonies. These transformants were grown up, and RNA was extracted from induced and uninduced cells. The RNAs were examined for the presence of β- and γ-globin mRNA by nuclease-S1 mapping (Fig. 2). Although specific induction of β-globin mRNA was seen in about two thirds of the MEL clones, induction of γ-globin mRNA was not observed. Figure 2 shows two examples of RNA from cell lines transformed with cosHG28 (see Fig. 1). Although one of these lines shows an approximately tenfold induction of β-globin mRNA, the other shows a low constitutive level of expression; both show constitutive γ-globin mRNA synthesis.

Induction of β-globin mRNA is seen for transformants made with all the β-globin-gene-containing cosmids shown in Figure 1. It is likely that the induction phenomenon operates at the level of transcription rather than RNA stability. Constitutive levels of γ-globin mRNA are seen in most clones, and it is improbable that β- but not γ-globin mRNA would be degraded. Moreover, certain clones are constitutive for β-globin mRNA synthesis yet show similar behavior for both the γ-globin mRNA and the mouse α-globin mRNA used as an internal standard.

It appears, therefore, that the human β-globin gene responds to *trans*-acting signals in the MEL cell that suppress expression of β-globin mRNA prior to the induction process; this is the same result observed with the chromosome-transfer experiments. Expression of the γ-globin gene is in general not suppressed before induction (although in one exceptional case we have seen induction of γ-globin mRNA synthesis).

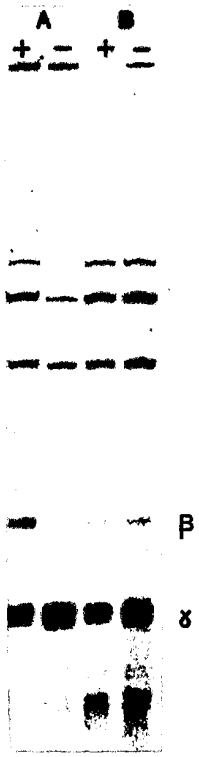

Figure 2. Nuclease-S1 mapping of RNA from MEL cells before (−) and after (+) induction of erythroid differentiation. 3′-labeled probes were used to measure simultaneously γ- and β-globin mRNA. The positions of the relevant bands are indicated; the bands seen in the upper part of the gel are input fragments. *A* and *B* are two independently isolated transformed cell lines.

DNA Methylation Suppresses Globin-gene Expression

It has been known for some time that there is an inverse correlation between the presence of 5-methylcytosine (5^mC) groups in eukaryotic DNA and gene expression. Waalwijk and Flavell (1978b) showed that DNA methylation in the rabbit β-globin gene varied in a tissue-specific manner, and subsequently several studies showed that active genes are hypomethylated (Mandel and Chambon 1979; McGee and Ginder 1979; van der Ploeg and Flavell 1980). Interestingly, in the human β-globin locus, when the adult globin genes are expressed (as in erythroblasts in the adult), the δ- and β-globin region is hypomethylated, whereas the γ-globin-gene region has a high level of methylation; the reciprocal result is obtained when the fetal globin genes are expressed early in development (van der Ploeg and Flavell 1980; T. de Lange and R.A. Flavell, unpubl.). This suggests that DNA methylation might play a role in the developmental regulation of globin-gene expression.

To establish that DNA methylation is the cause rather than the effect of gene expression, it is necessary to ask whether a DNA molecule, once methylated, can be expressed. To address this question, we have generated methylated α- and γ-globin genes and asked whether these are expressed in mouse L cells. The approach we have used is that of Stein et al. (1982) (see Fig. 3). A

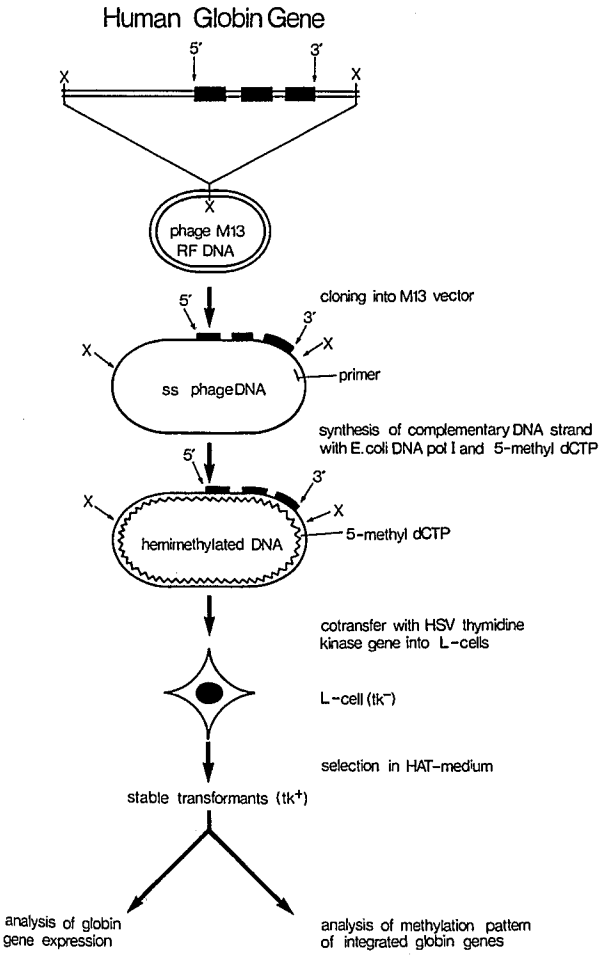

Figure 3. Strategy used to study the effect of DNA methylation on globin-gene expression in vivo. First, a human globin gene is cloned into the RF DNA of bacteriophage M13. Single-stranded (SS) DNA of the recombinant phage is used as a template for the synthesis of its complementary DNA strand by *E. coli* DNA polymerase I in the presence of 5-methyl dCTP instead of normal dCTP. The resulting hemimethylated DNA is introduced into mouse L cells by calcium phosphate coprecipitation with the HSV *tk* gene as a selective marker and is subsequently integrated into nuclear DNA. Transformed cell clones are selected by their ability to grow in HAT medium. Prior to replication in these cells, the hemimethylated DNA becomes methylated at residues that are recognized by cellular methylases, i.e., predominantly at CpG residues. This methylation pattern is faithfully inherited during subsequent replication rounds and can be analyzed in the DNA of stable transformants by the Southern blot technique, using methylation-sensitive restriction enzymes. On the other hand, RNA of transformed cells is analyzed for globin transcripts by nuclease-S1 mapping to determine whether or not the methylated globin genes are expressed in L cells. (■) The three exons of the globin gene; (×) the restriction enzyme used for the initial cloning experiment, i.e., *Pst*I for the α- and β-globin gene and *Hin*dIII for the $^{A}\gamma$-globin gene. The resulting recombinant phages are named $M\alpha_1$, $M\beta_1$, and $M\gamma_1$, respectively.

given globin gene is cloned into the replicative form (RF) of the single-stranded bacteriophage M13. The single-stranded form of this phage is then used as a template for the synthesis of a methylated minus strand by using an oligonucleotide primer and d^{m}CTP, dATP, dTTP, and dGTP. The hemimethylated DNA is then used to transform mouse Ltk^{-} cells. In this transformation, we include an unmethylated β-globin gene as an internal control; the HSV *tk* gene functions as a selective marker, and in most cell lines obtained, the two globin genes and the *tk* genes are taken up and stably maintained by the L cells.

Stein et al. (1982) and Wigler et al. (1981) have shown that the pattern of methylation is inherited at ^{m}CpG residues but that the remaining methylcytosine residues are not retained; this fits the idea that a methylase specific for CpG methylates the DNA at hemimethylated CpG sites but cannot methylate the remaining sites. When the hemimethylated human α-globin gene is transfected into L cells and these cells are analyzed after about 50 generations, no α-globin mRNA is detected; the control β-globin gene, however, is clearly transcribed, as is the unmethylated α-globin gene in separate control experiments (Fig. 4). We have examined the DNA in these cells by Southern blotting, using the restriction enzyme pair *Msp*I and *Hpa*II (Waalwijk and Flavell 1978a). This shows that the methylcytosine residues (C^{m}CGG) at all *Msp*I sites in the 1.5-kb *Pst*I DNA fragment cloned in M13 are still present since this fragment is totally resistant to cleavage by *Hpa*II, although it is cut by *Msp*I (Fig. 5). In addition to the results obtained with the α-globin gene, we have performed the same experiments with the methylated γ-globin gene (Fig. 6). Again, the methylated γ-globin gene is not transcribed in the L cells (clone 10), whereas the unmethylated γ- and β-globin genes (clone 9) are transcribed. It follows therefore that methylation of the CpG residues in the human α-globin or γ-globin genes is sufficient to prevent transcription in L cells.

We have started a study of the effect of methylation of specific regions of the γ-globin gene on transcription. To do this, a given DNA fragment is hybridized to the γ-globin gene cloned in M13 and used as a primer for DNA synthesis using d^{m}CTP and dATP, dTTP, and dGTP. This results in a duplex circle that is hemimethylated over the newly synthesized region but that is unmethylated on both strands in the region covered by the primer. We prepared DNA in this way by using as a primer the *Bgl*II/*Hin*dIII fragment that runs from position +100 (in the first exon of the γ-globin gene) to a position 380 nucleotides to the 3′ side of the $^{A}\gamma$-globin gene. This provides us with a mosaic γ-globin, methylated in the entire 5′ flanking region and the first 100 nucleotides of the gene but unmethylated in the body of the gene. This gene has been used to cotransform mouse L cells together with a control unmethylated β-globin gene and the HSV-I *tk* gene. In two clones (12 and 13) the partially methylated γ-globin gene is not transcribed, whereas the β-globin gene added as an internal control is transcribed (Fig. 6). In one clone (11), however, abundant transcripts of the γ-globin gene are present. Inspection of the methylation status of the promoter region of the interpreted $^{A}\gamma$-globin genes in these clones shows that in the cases where no γ-globin transcripts are seen, the ^{m}CpG-sensitive enzymes (*Hpa*II and *Hha*I) do not cut at their respective sites

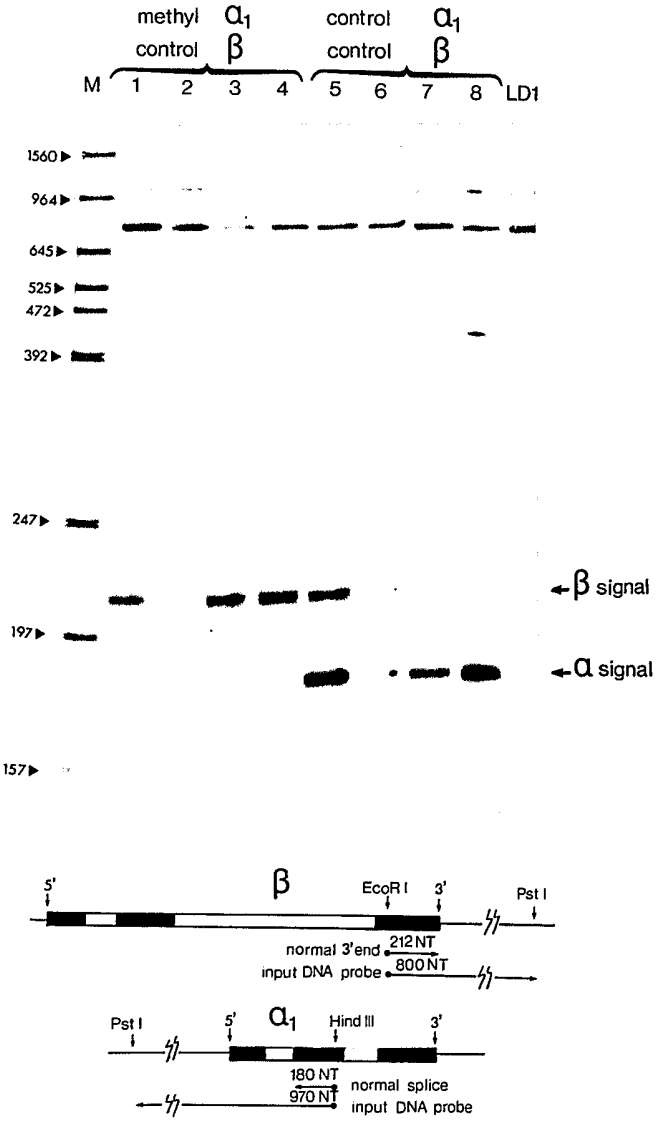

Figure 4. Effect of DNA methylation on human α_1-globin-gene expression. Cell clones 1 to 4 were transformed with in vitro hemimethylated Mα_1 DNA, and clones 5 to 8 were transformed with unmodified Mα_1 RF DNA. All these clones were furthermore transformed with the unmodified Mβ_1 RF DNA containing the human β-globin gene. The analysis of the methylation pattern of the α_1-globin gene is shown in Fig. 5 for clones 2–5. Clone 1 contains about 3–5 copies of fully methylated α_1-globin genes (data not shown). Cytoplasmic RNA isolated from all these cell clones was analyzed for the presence of globin transcripts by nuclease-S1 mapping as described by Busslinger et al. (1981). The β-globin-specific DNA probe was end-labeled at the *Eco*RI site of the β-globin gene by reverse transcriptase, and the α_1-specific DNA was labeled at the *Hin*dIII site of the α_1-globin gene by T4 polynucleotide kinase. Nuclease-S1-resistant DNA fragments were separated on an 8% polyacrylamide sequencing gel. RNA isolated from L cells was used as a control (lane *LD1*), and end-labeled ϕX174 DNA was digested with *Rsa*I as a size marker (lane *M*). The position of the input DNA probes, as well as the nuclease-S1-resistant DNA fragments, is shown (arrows) below the map of the respective globin gene.

(Fig. 7). In the case where γ-globin transcripts are found, we see that the methylcytosine residues in this region have been lost; we believe that this occurred by some repair process in vivo. These data are therefore all consistent with the hypothesis that methylation in the promoter region is sufficient to suppress transcription in vivo. This experiment further suggests that the promoter region may be the target for DNA methylation in vivo; however, to prove this, more work needs to be done. It is obviously possible that methylation of only part of the γ-globin gene at *any* site is sufficient to suppress transcription, and we are actively testing this idea.

Structure and Function of the Genes Encoding the Murine MHC

The major histocompatibility complex (MHC) plays an important role in the regulation of the immune response in vertebrates. In mammals there are two major classes of MHC proteins, the classic transplantation antigens (class I: in mouse, H-2K,D,L proteins; in man, HLA-A,B,C) and the immune-response-associated antigens (class II: in mouse, Ia antigens; in man, HLA-DR). Also associated with the complex are some of the genes involved in the complement reaction (class III). Class-I antigens appear to function in presenting foreign antigens (e.g., viruses) to cytotoxic T cells of the immune system. These proteins are found on virtually all cell types. Class-II antigens are believed to function in the presentation of foreign antigens to another type of T cell, the T-helper cell. These proteins are found on the surface of lymphocytes and macrophages. All these genes are closely linked on chromosome 17 of the mouse (Fig. 8). In addition, adjacent to the *H-2* locus in mouse is the *TL* complex, which contains several genes encoding lymphoid differentiation antigens, the Qa and T1a antigens.

Both the class-I antigens (K,D,L) and the Qa and T1a antigens are integral membrane proteins of 40,000–45,000 molecular weight found associated with a non-polymorphic light-chain β_2-microglobulin (m.w. = 12,000), which is not encoded in the MHC. Because of

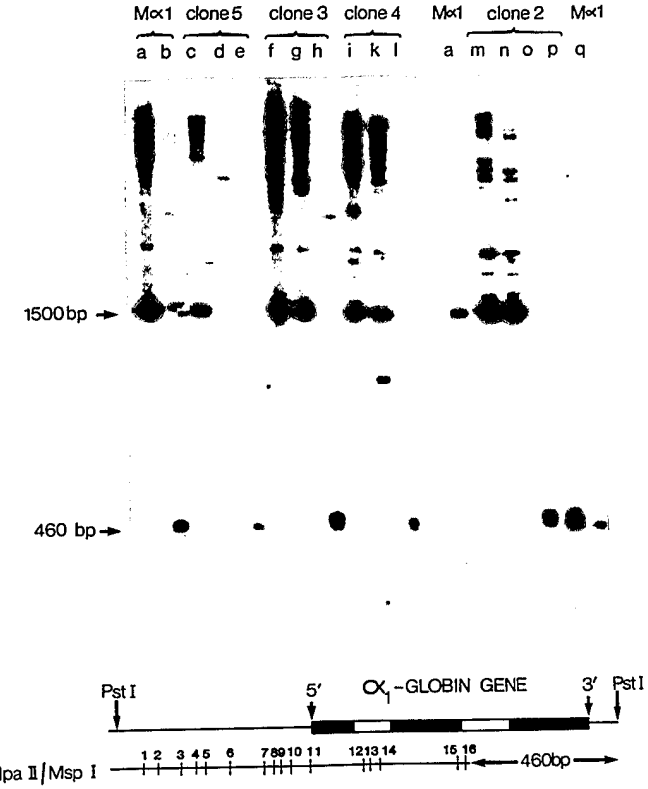

Figure 5. Inheritance of DNA methylation of the human α_1-globin gene. DNA (5 μg) of cell clones transformed either with hemimethylated Mα_1 DNA (clones 2–4) or with unmethylated Mα_1 RF DNA (clone 5) was digested with *Pst*I alone (lanes *c, f, i, m*), with *Pst*I and *Hpa*II (lanes *d,g,k,n*), or with *Pst*I and *Msp*I (lanes *e, h, l, o, p*). As marker DNA, Mα_1 RF DNA together with 5 μg of L-cell DNA was digested with *Pst*I (lane *a*) or with *Pst*I and *Msp*I (lanes *b, q*). DNA fragments were separated on a 1.5% agarose gel, transferred to nitrocellulose, and hybridized to the nick-translated *Pst*I DNA fragment containing the α_1-globin gene. The amount of Mα_1 RF DNA in lanes *a* and *b* corresponds to 5 copies per haploid mouse genome and that in lane *q* corresponds to 20 copies per haploid genome. Lane *p* corresponds to lane *o*, but mixed into the DNA of clone 2, it contains the same amount of *Msp*I-digested Mα_1 RF DNA as lane *q*. All lanes to the left (*a–l*) were autoradiographed for a longer period than those to the right (*m–q*). The restriction map for the enzyme *Hpa*II and its isoschizomer *Msp*I was determined by partial restriction mapping of the cloned *Pst*I insert of Mα_1 and is shown below the map of the human α_1-globin gene. *Hpa*II restriction sites are numbered from left to right.

their similar structure and close genetic linkage it is likely that all these class-I-like proteins are evolutionarily related and perhaps members of a dynamic multigene family (Michaelson et al. 1977; 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 serological analyses and protein sequencing. In fact, about 50 alleles at each of the *H-2K* and *H-2D* loci have been observed (for review, see Klein 1979). *H-2* haplotypes are defined serologically and consist of a particular set of alleles found at each *H-2* locus (e.g., the C57 black/10 [B10] mouse is *H-2*^{b} at each *H-2* locus, the AKR strain is *H-2*^{k}, DBA/2 is *H-2*^{d}, etc.). Some inbred mouse strains appear to be natural recombinants between other strains (e.g., the A mouse strain with the *H-2*^{a} haplotype is *H-2K*^{k} but *H-2D*^{d}). In addition, a variety of recombinant lines have been generated from many of the inbred strains.

The variation in the *H-2K* and *H-2D* genes is greater than any other known gene, and the nature of this polymorphism is completely unknown. It is likely that the analysis of these genes by recombinant DNA technology will provide some clues to this variation and the mechanisms involved in its generation.

A number of groups have isolated cDNA clones that encode proteins whose sequences are homologous to known transplantation antigens (see, e.g., Kvist et al. 1981; Steinmetz et al. 1981).

Southern blotting analysis of genomic DNA probed with these cDNAs shows many cross-hybridizing DNA fragments, which 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) constructed from spleen DNA from the B10 mouse (A. Mellor et al., unpubl.). Analysis of these clones by standard restriction enzyme digestion techniques establishes that many of these clones overlap one another and define a number of gene clusters (Table 1); in total, about 17 different *H-2*-like genes have been isolated in these cosmids. Steinmetz et al. (1982) have cloned about 36 *H-2* genes from the *H-2*^{d} haplotype and arranged these in more than 9 gene clusters.

The number of independent *H-2*-like genes in these DNA clones exceeds that expected from the 4–5 known classic transplantation antigen (*H-2*) genes. This suggests that the cDNA probes used detect other *H-2*-related sequences present in the mouse genome. These could be multiple copies of the known *H-2* genes, nonfunctional pseudogenes, or other functional, evolutionarily related genes (e.g., the Qa and T1a antigens). Indeed, Steinmetz et al. (1982) have cloned and sequenced an *H-2*-related gene that is found in the *Qa2/3* region and that is part of a cluster of at least 7 *H-2* genes (Steinmetz et al. 1982).

A first step in identification of the *H-2* genes that we have cloned 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, e.g., B10 (*H-2*^{b}), AKR (*H-2*^{k}), and BALB/c (*H-2*^{d}) mice with a number of restriction enzymes and

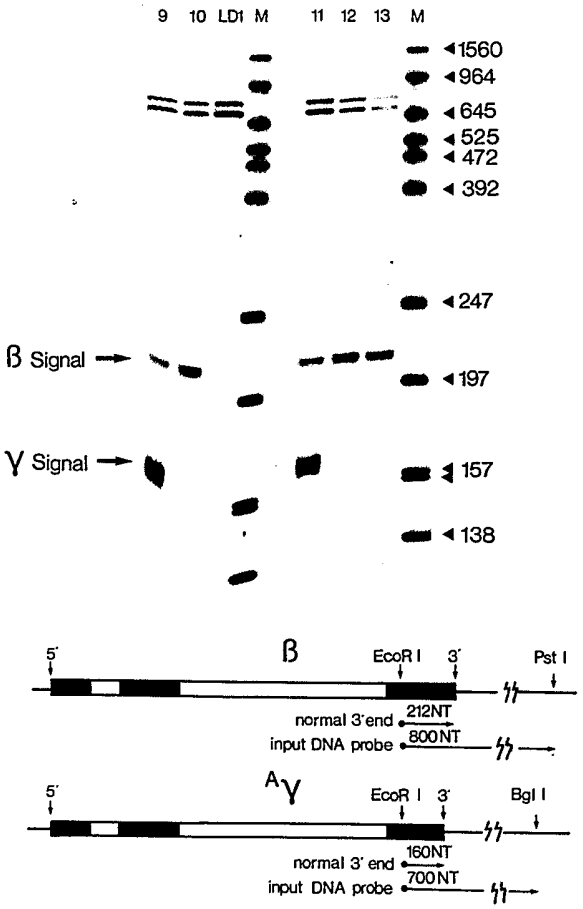

Figure 6. Effect of DNA methylation on human $^{A}\gamma$-globin-gene expression. Cytoplasmic RNA isolated from cell clones 9–13 was analyzed for the presence of globin transcripts by nuclease-S1 mapping using a β-globin-specific as well as an $^{A}\gamma$-globin-specific DNA probe. Both DNA probes were previously end-labeled by reverse transcriptase at the *Eco*RI site in the third exon of the respective gene. Nuclease-S1-resistant DNA fragments were separated on an 8% polyacrylamide sequencing gel. Cytoplasmic RNA isolated from LD1 cells was used as a control (lane *LD1*) and 5′-end-labeled φX174 DNA digested with *Rsa*I was used as a size marker (lane *M*). The positions of the input DNA probes as well as the nuclease-S1-resistant DNA fragments are shown (arrows) below the maps of the globin genes. The analysis of the methylation pattern of the $^{A}\gamma$-globin gene contained in clones 9, 10, 11, and 13 is shown in Fig. 7. The methylation pattern of the $^{A}\gamma$-globin gene of clone 12 corresponds to that of clone 13 (data not shown).

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 inbred strains. For example, using a probe from a clone in cluster 4 on *Pst*I digests of these mouse DNAs shows a 2.5-kb band in B10 DNA but a 3.8-kb band in AKR DNA. As mentioned above, recombinant mice have been produced that have undergone a recombination event within the *H-2* locus, resulting in a given section of the *H-2* map deriving from one haplotype (e.g., AKR *H-2*^{k}) and a second segment from a second mouse (e.g., C57 black/6 or B6 *H-2*^{b}). Where the recombination event has occurred within the complex has been shown by serological analysis of the recombinant mouse strains. Thus, in the recombinant mouse B6K2 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 *Pst*I-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 gene-cluster 4 is localized in the AKR segment of the B6K2 mouse, i.e., in the *Tla* region.

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 cross-hybridizing sequences or related genes 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 10^{6}–10^{7} bp of DNA. This gene family might then be two orders of magnitude larger than the β-globin cluster described in the previous section. In addition, one of the cosmids, H41, maps outside of the MHC. We have not determined where it is located on a separate chromosome, nor whether the gene on this cosmid is functional or expressed.

Structural studies of the *H-2* genes must surely be directed toward the elucidation of the mechanism whereby polymorphic differences between *H-2* genes are generated and maintained in the population. Any such explanation must take into account a number of basic facts about this system. First, the number of different alleles at a given locus is very high, commonly more than 50 at a given locus. Second, the protein sequences of two nonallelic gene products in a given haplotype are commonly more similar to one another (e.g., H-2K^{b} and H-2D^{b}) than are those of two alleles (such as H-2D^{b} and H-2D^{d}). This second fact is reminiscent of observations made by Slightom et al. (1980) in the human γ-globin system. They observed that the nonallelic $^{G}\gamma$- and $^{A}\gamma$-globin genes from a given human chromosome were more similar to each other in DNA sequence than were two allelic $^{A}\gamma$-globin genes from the same individual. They concluded that this similarity between nonallelic genes was the result of the exchange of genetic information between the two nonalleles by gene conversion. Clearly, such an observation is capable of explaining the similarity between the *H-2K*^{b} and *H-2D*^{b} genes, and DNA sequence analyses of these genes currently underway should test this hypothesis.

A more direct test of the idea would be to examine the DNA sequence of a given *H-2* gene before and after a putative gene conversion has occurred. A potential test system like this has been provided by Nathenson and his colleagues (Nairn et al. 1980) in the study of natural mutants in the *H-2* system. These workers isolated mice with mutant *H-2*^{b} haplotypes by using the graft-rejection assay. A number of mutants were painstakingly isolated, and the partial protein sequence of a number of these mutants was determined. Interestingly, some of the mutants showed multiple amino acid changes; e.g., in the mutant *bm1*, which has an altered H-2K^{b} molecule, two residues, Arg_{155}-Leu_{156}, have changed to Tyr-Tyr. Multiple changes suggest a novel genetic mechanism, and gene conversion by copying a stretch of a gene

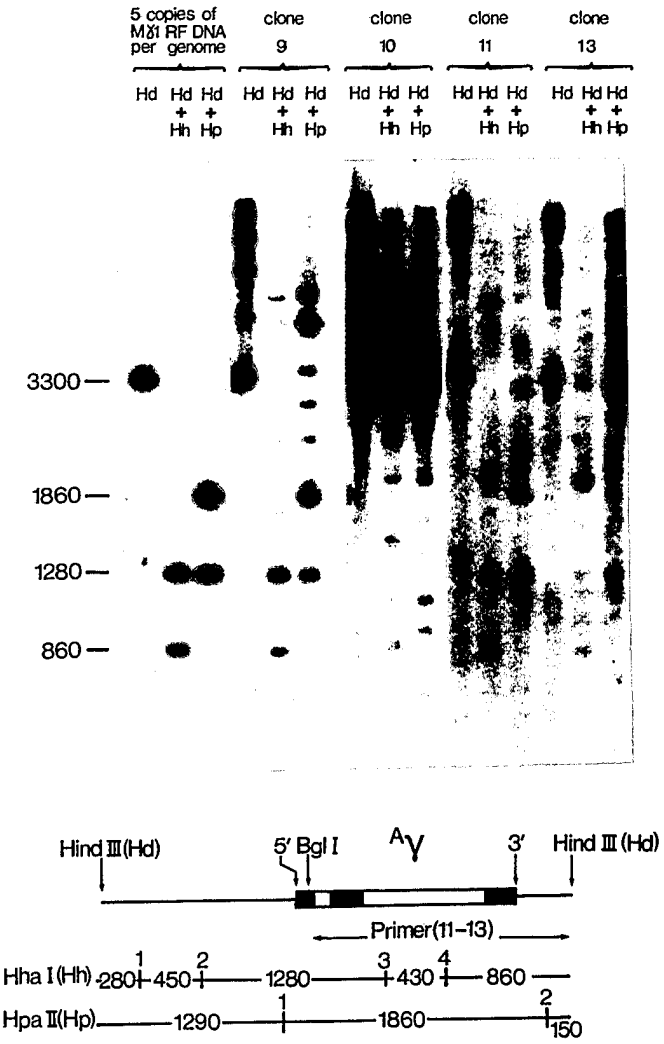

Figure 7. Inheritance of DNA methylation of the human $^{A}\gamma$-globin gene. DNA (10 μg) of LD1 cells transformed either with unmethylated Mγ_1 RF DNA (clone 9) or with totally (clone 10) or partially (clones 11–13) hemimethylated Mγ_1 DNA was subject to *Hin*dIII digestion alone or to double digestion with *Hin*dIII and *Hha*I or with *Hin*dIII and *Hpa*II. As marker DNA, Mγ_1 RF DNA together with 10 μg of L-cell DNA was digested with the same restriction enzymes. DNA fragments were separated on a 1% agarose gel, transferred to nitrocellulose, and hybridized to the nick-translated *Hin*dIII DNA fragment containing the $^{A}\gamma$-globin gene. The restriction maps for the enzymes *Hha*I and *Hpa*II are shown below the map of the human $^{A}\gamma$-globin gene as is the position of the *Bgl*I-*Hin*dIII DNA primer used for the synthesis of the partially hemimethylated Mγ_1 DNA that gave rise to clones 11–13. *Hha*I and *Hpa*II restriction sites are numbered from left to right, and the lengths of the restriction fragments are given in numbers of base pairs.

having the appropriate Tyr-Tyr residues into the *H-2K*^{b} gene could explain the results.

Accordingly, we have cloned the *H-2K* region from the mutant *bm1* mouse as a series of overlapping cosmids and compared the structure of this gene with the structure of the normal *H-2K*^{b} gene. Figure 9A shows that the general sequence arrangement of the normal and mutant *H-2K* regions are identical. No gross rearrangements of DNA sequences have occurred, and the regions encoding the third domain and the remaining carboxyterminal portion of the *H-2* molecule seem to be the same for both genes. Careful inspection of the fine restriction map of the coding regions of the two genes, however, reveals few differences (Fig. 9B). As expected, the *bm1* gene lacks a *Hin*fI site (present at the sequence encoding the normal amino acid residues 155 and 156), but it has also gained a *Pst*I site within the exon encoding the second domain of the *H-2* molecule. We have sequenced 450 bp (comprising the entire second domain and part of each of the two flanking introns) of both the *H-2K*^{b} and *H-2K*^{bm1} genes. The two se-

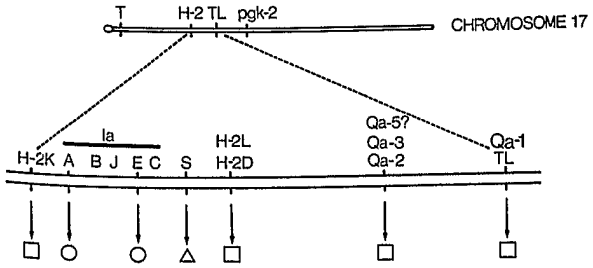

Figure 8. Schematic genetic map of the *H-2* and *TL* complexes of chromosome 17 of the mouse. The relative order of the loci listed vertically is not known. The distance from *H-2K* to *H-2D* is about 0.3 cM and from *H-2D* to *T1a* is about 1 cM. (□) Class-I antigens; (○) class-II antigens; (△) class-III antigens.

Table 1. Analysis of B10 Mouse Genomic Clones

Cluster	Cosmids	kb	Genes	Location
1	7	95	2	*H-2K*
2	2	40	1	*H-2D*
3	18	120	5	*Qa2,3*
3b*	5	?	?	*Qa2,3*
4	39	90	5	*T1a*
5	10	?	5	*Qa2,3* or *T1a*
6	1	40	1,5	*Qa2,3*

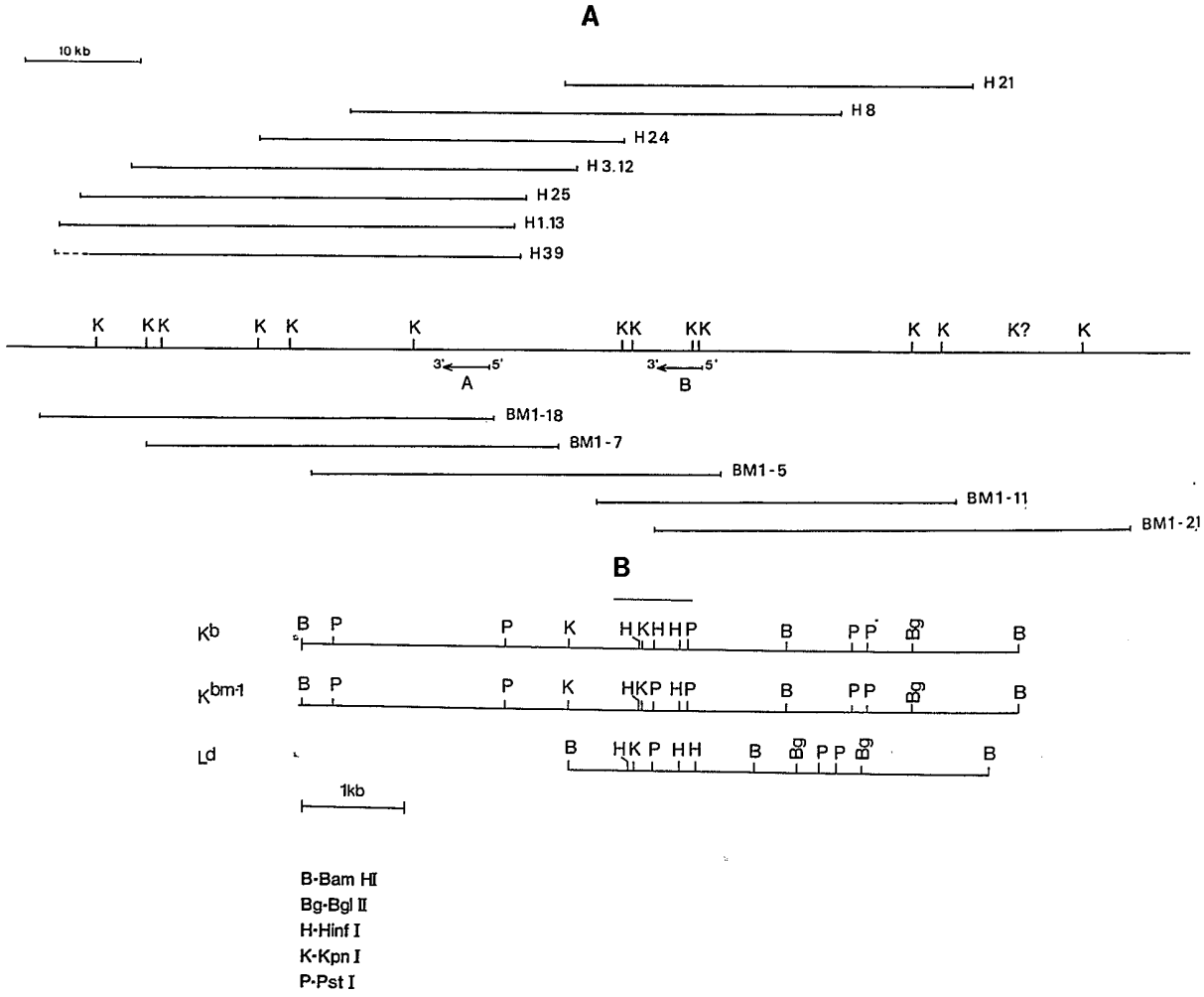

Figure 9. Restriction maps of the *H-2K* region in the B10 or mutant B6*bm1* mouse. (*A*) Overlapping cosmid clones from the B10 (H) and B6*bm1* libraries are indicated. (K) *Kpn*I. (*B*) Fine map of K^b, K^{bm1}, and L^d (Evans et al. 1982) gene regions oriented with 5′ ends on the left. The line above the maps indicates the extent of the region sequenced in the K^b and K^{bm1} genes. *Hin*fI sites in this region alone are shown.

quences are identical except for the seven nucleotide changes clustered in the sequence-encoding residues 152 to 156 of the H-2K protein molecule (Fig. 10). These differences cause three amino acid changes—Glu_{152} (GAA)→Ala_{152} (GCT), Arg_{155} (AGA)→Tyr_{155} (TAT), and Leu_{156} (CTC)→Tyr_{156} (TAC). The latter two changes correspond to those already found in the partial amino acid sequencing; the former change presumably went undetected in the protein work (Nairn et al. 1980).

Since gene conversion is the likely mechanism for this donation of new DNA sequences, we would like to determine which *H-2* (or other) gene is the donor. A hint in this direction comes from the data of Moore et al. (1981) and Evans et al. (1982), who cloned and sequenced the $H\text{-}2L^d$ gene from the BALB/c mouse. Evans et al. (1982) noted that the sequence of the H-2L^d protein, deduced from the gene sequence was the same as that of the $H\text{-}2K^{bm1}$ mutant gene and suggested that gene conversion from an $H\text{-}2L^d$-like gene in the $H\text{-}2^b$ haplotype might be the mechanism of mutagenesis. In Figure 10 we compare the DNA sequences of $H\text{-}2K^b$, $H\text{-}2L^d$, and $H\text{-}2K^{bm1}$. It can be seen that all seven abnornal nucleotides seen in $H\text{-}2K^{bm1}$ are found in the $H\text{-}2L^d$

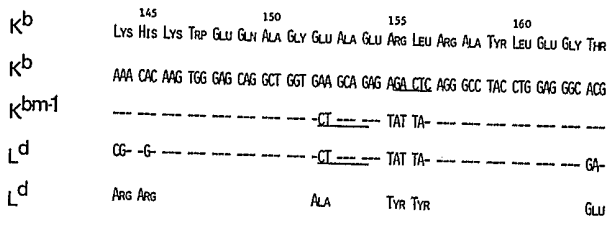

Figure 10. Comparison of the DNA sequences of the $H\text{-}2K^b$, $H\text{-}2L^d$ (Moore et al. 1981; Evans et al. 1982), and $H\text{-}2K^{bm1}$ genes in the second domain of the *H-2* molecule, between amino acids 144-163 (Nairn et al. 1980). Changed bases only are indicated in the latter two sequences. Restriction sites for *Pst*I (CTGCAG) and *Hin*fI (GANTC) are underlined.

gene. This suggests that these changes result from a gene-conversion event between the *H-2K*^{b} gene and the *H-2L*^{d} gene or an *H-2L*^{d}-like gene in the *H-2*^{b} haplotype; since the *bm1* mutant was isolated from a C57BL/6 (=B6≅B10)×BALB/c heterozygote (*H-2*^{b}/*H-2*^{d}), both explanations are possible. The *H-2L*^{d} gene differs from the *H-2K*^{bm1} gene at codons 144, 145, and 163. These differences limit the maximum extent of the gene-conversion event to the DNA sequences between these sites, given that the donor gene was *H-2L*^{d}. If the donor gene derived from the *H-2*^{b} chromosome, then the latter argument would, of course, not hold. Although there is no unequivocal evidence for the existence of an *H-2L*^{b} gene, an *H-2L*-like gene or pseudogene may exist in the B6 mouse, and we are currently screening our libraries for such genes.

It is likely that this type of novel genetic mechanism is responsible for the generation of other mutant H-2 molecules. Both the *H-2K*^{bm6} and *H-2K*^{bm9} mutant genes have at least two amino acid differences. Moreover, as we have previously discussed (Flavell et al. 1981), we believe that gene conversion between functional *H-2* genes and other *H-2*-related genes, most of which are concentrated in the *Qa,T1a* region, may be the most important mechanism for the generation of polymorphic differences.

Expression of the *H-2K*^{b} and *H-2D*^{b} Genes in Mouse L Cells

Although cosmid clones can be mapped to a given locus by the mapping procedures described above, to establish the identity of any given gene, a more detailed analysis is required. We have therefore introduced *H-2* cosmids into mouse L cells to ask if these cells containing a given cosmid express a known *H-2* gene. We have transformed L cells that are mutant for thymidine kinase (*tk*$^-$) with cosmid DNA as a calcium phosphate coprecipitate and then selected for stable *tk*$^+$ transformants. Since the cosmid vector used for most of our experiments, pOPF1 (Grosveld et al. 1982), contains the HSV *tk* gene, transformants are readily obtained. Transformants were screened for the binding of monoclonal antibodies that react with the H-2K^b or H-2D^b molecules. L cells transformed with the cosmid H8 bind the H-2K^b monoclonal antibody specifically, whereas none of this group binds the H-2D^b monoclonal antibody (not shown). Cosmid clones H8, H24, and H25 are all found in cluster 1 and map at the *H-2K* locus. This cluster (Fig. 9A) contains two genes; since both H24 and H25 contain gene *A* and do not express H-2K^b, it is clear that gene *B* is the *H-2K*^{b} gene, and the sequence data reported above confirm this conclusion. L cells containing the *B* gene of cluster 1 (called gene *IB* here) bind a variety of other alloantisera directed against the H-2K^b molecule and immunoprecipitations with anti-H-2K^b antibodies reveal the expected 45K protein. We are therefore confident that the cells express the H-2K^b protein.

In addition, we have determined which of the cosmid clusters contains the *H-2D*^{b} gene. When L cells containing cosmids from cluster 2 (Fig. 11) are examined for their ability to bind monoclonal antisera, they are found to react positively to anti-H-2D^b antisera. In this case, the cosmid cluster in question contains only one gene. Thus, this gene maps to the *H-2D* locus and expresses the H-2D^b protein as assayed by monoclonal antibody binding, as well as other assays to be described below.

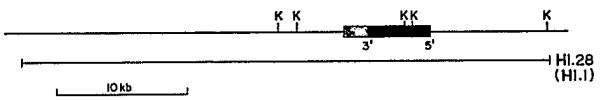

Figure 11. Restriction map of the *H-2D*^{b} gene region. (K) *Kpn*I.

Cytotoxic-T-cell-mediated Lysis of Transformed Cells

The presence of the H-2K^b molecule on a cell renders these cells as targets for allogeneic cytotoxic-T-cell (T_c-cell)-mediated killing by T_c cells raised against H-2^b antigens (Doherty and Zinkernagel 1975).

A number of indirect arguments suggest that the presence on the cell membrane of a given H-2 polypeptide, such as the H-2K^b molecule, is sufficient to generate a target for T_c cells. For example, this process can be blocked by monoclonal antibodies to a single H-2 polypeptide (Weyland et al. 1981). To test this directly and also to characterize further the H-2K^b polypeptide detected by serological assays on the surface of L cells transformed by cosmid H8, we generated T_c cells directed against H-2^b or H-2^k antigens and tested their ability to kill various target cells (not shown). Whereas anti-H-2^k T_c cells show specific killing of all L-cell lines tested, anti-H-2^b T_c cells show specific killing of only control B10 target cells and two L-cell clones (LH8-1 and LH8-2) transformed by cosmid H8. Untransformed L cells (Ltk$^-$) and clones transformed by cosmids H24 and H39 (data not shown) are not seen as targets for killing by anti-H-2^b T_c cells. These data confirm that the H-2K^b molecule detected by serological assays in cells transformed by cosmid H8 can act as a target for anti-H-2^b T_c-cell killing.

We performed a similar experiment using B10.D2 (*H-2*^{d}) T_c cells generated by immunization with target spleen cells from the recombinant mouse B10.A (5R) (*H-2K*^{b}, *H-2D*^{d}). Such T_c cells kill cells expressing the H-2K^b but not the H-2D^b molecule. These T_c cells also kill L cells transformed with cosmid H8 but not L cells transformed with cosmid H39 (not shown).

Cytotoxic-T-cell-mediated Lysis of Cells Infected with Influenza Virus

T_c-cell-mediated lysis of cells expressing viral antigens on their surface can only take place when the viral antigen is presented to T_c cells in association with an appropriate H-2 antigen; in this way H-2 antigens act as a restriction element for the cytotoxic-T-cell response

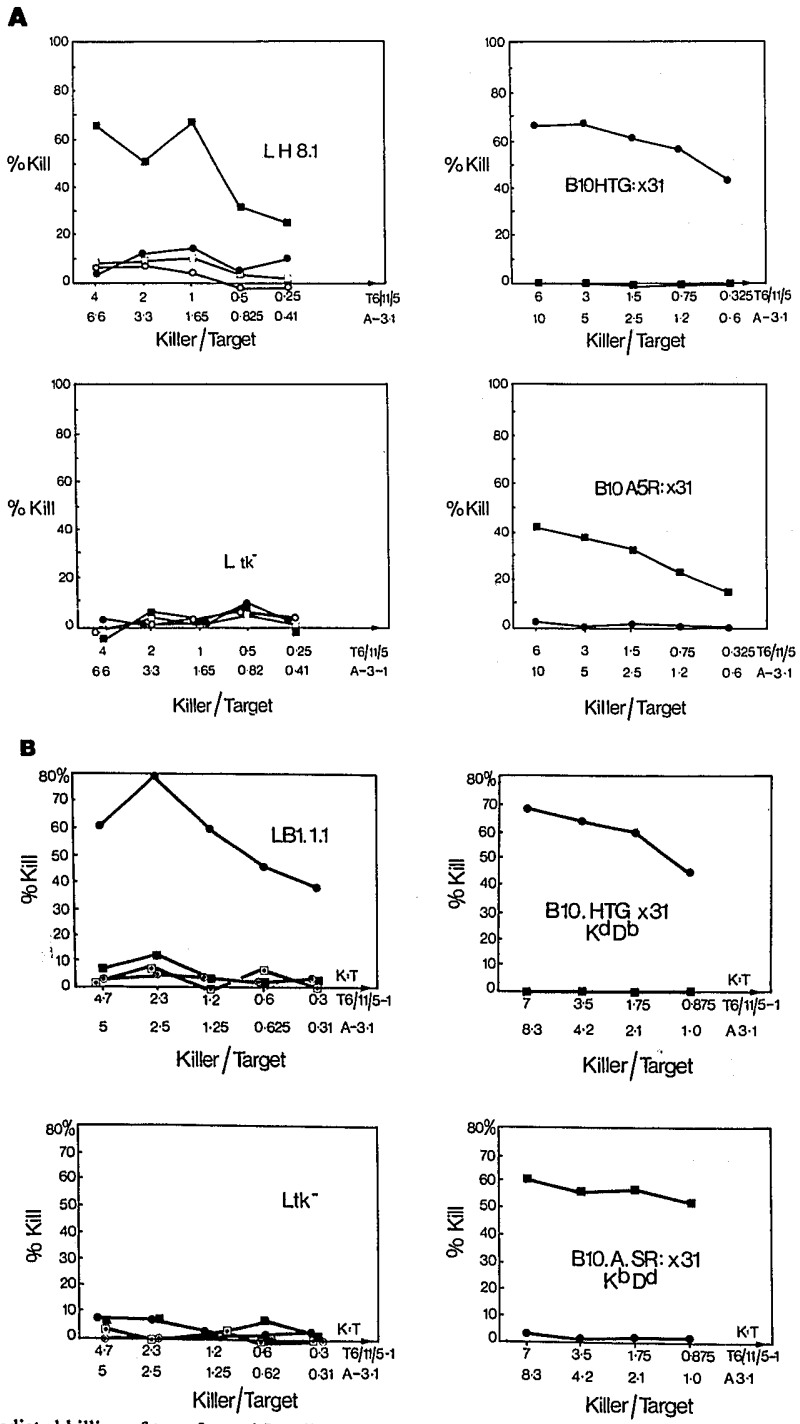

Figure 12. T_C-cell-mediated killing of transformed L cells infected with influenza virus. (*A*) Two influenza-specific T_C clones—T6/11/5 (*H-2K*^{b}-restricted) (□, ■) or A3.1 (*H-2D*^{b}-restricted) (○, ●) were incubated with LH8-1 or Ltk$^-$ (L-cell control) target cells that were either infected (■, ●) or not infected (□, ○) with influenza virus strain X31. The percent specific lysis estimated by ^{51}chromium release (29) at each ratio of killer (T_C cell)-to-target (LH8.1 or Ltk$^-$) cells was calculated using the formula $\frac{(E-C)}{(M-C)} \times 100\%$, where E=cpm from target cells incubated with killer cells, C=cpm from target cells with medium alone, and M=cpm from target cells lysed with 5% Triton. Controls using influenza-virus-infected target lymphocytes from B10HTG mice (*H-2K*^{d}, *D*^{b}) or B10A5R (*HP2K*^{b}, *D*^{d}) demonstrate the *H-2K*^{b}- and *H-2D*^{b}-restricted killing of T_C clones T6/11/5 and A3.1, respectively. The isolation and characterization of these T_C clones is described in detail elsewhere (A.R.M. Townsend and P. Taylor, in prep.). (*B*) The same experiment performed with L cells transformed with the *H-2D*^{b} gene. Symbols are as in *A*.

(Doherty and Zinkernagel 1975; Zinkernagel and Doherty 1979).

Thus, we tested the ability of the H-2K^b and the H-2D^b molecules expressed on the surface of L cells transformed by cosmid H8 to act as a restriction element for T_C-cell-mediated killing of cells infected by influenza virus.

Influenza-specific T_C clones restricted to one *H-2* re-

gion can be selected and grown in the presence of T-cell growth factor (Lin and Askonas 1981). L cells ($H\text{-}2^k$) and LH8-1 were infected with type-A influenza virus (A/X31) and used as target cells for two influenza-specific $H\text{-}2^b$ T_C clones, one restricted to H-2K^b (T6/11/5) and the other (A/3.1) to H-2D^b molecules (A. Townsend and P.M. Taylor, unpubl.). Figure 12A shows that X31-infected, untransformed L cells ($H\text{-}2^k$) are unable to act as targets for killing by either the $H\text{-}2K^b$- or $H\text{-}2D^b$-restricted T_C-cell clones. LH8-1 cells infected with X31, however, act as targets for killing by the $H\text{-}2K^b$-restricted T_C-cell clone but not the $H\text{-}2D^b$-restricted clone. This shows that the H-2K^b molecule expressed on the surface of LH8-1 cells is able to act as a restriction element for T_C-cell killing of X31-virus-infected cells. We have also carried out the same experiments with L cells transformed with several of the cosmids from cluster 2. Again, using an $H\text{-}2D^b$-restricted, influenza-virus-specific T_C-cell clone, we have obtained specific lysis (Fig. 12B).

In summary, we have isolated most, if not all, of the class-I-related genes in the B10 genome and have located most of them to a region of the genetic map of the MHC locus. In particular, we have determined which of the cosmid clusters contain the $H\text{-}2K^b$ and the $H\text{-}2D^b$ genes, both by location and by immunological assays. In addition, we can conclude that the presence of an $H\text{-}2^b$ gene in mouse cells is sufficient to evoke essentially all the immunological properties that have been associated with these molecules by a number of more conventional approaches. The fact that these processes can be studied by DNA-mediated gene transfer will make it possible to carry out a detailed molecular analysis of the determinants recognized by antibodies and by the T-cell receptor. Specifically, it should now be possible to transform L cells with hybrid *H-2* genes, constructed in vitro, that contain DNA segments from, e.g., the $H\text{-}2K^b$ and $H\text{-}2D^b$ genes and to ask whether T_C-cell clones, either allospecific or *H-2*-restricted, can kill such target cells. In this way, we hope to map the determinants recognized by specific monoclonal antibodies and by the T-cell receptor.

An analysis of the structure and expression of the other genes isolated in our study may also elucidate the mechanisms involved in the generation of diversity in the *H-2* genes.

ACKNOWLEDGMENTS

M.B., L.G., and E.W. were the recipients of fellowships from the Swiss National Funds, the Helen Hay Whitney Foundation, and the Deutsche Forschungsgemeinschaft, respectively. Work in the laboratories of R.A.F, E.S., and A.R.M.T. was supported by the British Medical Research Council; work in the laboratory of H.F. was supported by the Cancer Research Campaign; and work by R.J. was supported by the Imperial Cancer Research Fund. Madlyn Nathanson is gratefully thanked for preparing this manuscript.

REFERENCES

BUSSLINGER, M., N. MOSCHONAS, and R.A. FLAVELL. 1981. β^+ thalassemia: Aberrant splicing results from a single point mutation in an intron. *Cell* **27:** 289.

DOHERTY, P.C. and R.M. ZINKERNAGEL. 1975. *H-2* compatibility is required for T-cell-mediated lysis of target cells infected with lymphocytic choriomeningitis virus. *J. Exp. Med.* **141:** 502.

EVANS, G.A., D.H. MARGULIES, R.D. CAMERINI-OTERO, K. OZATO, and J.G. SEIDMAN. 1982. Structure and expression of a mouse major histocompatibility antigen gene. $H\text{-}2L^d$. *Proc. Natl. Acad. Sci.* **79:** 1994.

FLAVELL, R.A., H. BUD, H. BULLMAN, M. BUSSLINGER, E. DE BOER, A. DE KLEINE, L. GOLDEN, J. GROFFEN, F.G. GROSVELD, A.L. MELLOR, N. MOSCHONAS, and E. WEISS. 1981. *The structure and expression of mammalian gene clusters.* A.R. Liss, New York. (In press.)

GROSVELD, F.G., H.H.M. DAHL, E. DE BOER, and R.A. FLAVELL. 1981. Isolation of β-globin-related genes from a human cosmid library. *Gene* **13:** 227.

GROSVELD, F.G., T. LUND, E.J. MURRAY, A.L. MELLOR, H.H.M. DAHL, and R.A. FLAVELL. 1982. The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Res.* **10:** 6715.

JIMENEZ, A. and J. DAVIES. 1980. Expression of a transposable antibiotic resistance element in *Saccharomyces. Nature* **287:** 869.

KLEIN, J. 1979. The major histocompatibility complex of the mouse. *Science* **203:** 516.

KVIST, S., F. BREGERERE, L. RASK, B. CAMI, H. GAROFF, F. DANIEL, K. WIMAN, D. LARHAMMAR, J.P. ABASTADO, G. GACHELIN, P.A. PETERSON, B. DOBBERSTEIN, and P. KOURILSKY. 1981. cDNA clone coding for part of a mouse $H\text{-}2^d$ major histocompatibility antigen. *Proc. Natl. Acad. Sci.* **78:** 2772.

LIN, Y.-L. and B.A. ASKONAS. 1981. Biological properties of an influenza A virus–specific killer T-cell clone. *J. Exp. Med.* **154:** 225.

MANDEL, J.L. and P. CHAMBON. 1979. DNA methylation: Organ-specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucleic Acids Res.* **7:** 2081.

MANIATIS, T., E.F. FRITSCH, J. LAUER, and R.M. LAWN. 1980. The molecular genetics of human hemoglobins. *Annu. Rev. Genet.* **14:** 145.

MANTEI, N., W. BOLL, and C. WEISSMANN. 1979. Rabbit β-globin mRNA production in mouse L-cells transformed with cloned rabbit β-globin chromosomal DNA. *Nature* **281:** 40.

MCGEE, J.D. and G.D. GINDER. 1979. Specific DNA methylation sites in the vicinity of the chicken β-globin genes. *Nature* **280:** 419.

MICHAELSON, J., L. FLAHERTY, E. VITETTA, and M. POULIK. 1977. Molecular similarities between the Qa-2 alloantigen and other gene products of the 17th chromosome of the mouse. *J. Exp. Med.* **145:** 1066.

MOORE, K.W., B.T. SHER, Y.H. SUN, K.A. EAKLE, and L. HOOD. 1981. DNA sequence of a gene encoding a Balb/c mouse L^d transplantation antigen. *Science* **215:** 679.

MULLIGAN, R. and P. BERG. 1980. Expression of a bacterial gene and mammalian cells. *Science* **209:** 1422.

NAIRN, R., K. YAMAGA, and S.G. NATHENSON. 1980. Biochemistry of the gene products from murine MHC mutants. *Annu. Rev. Genet.* **14:** 241.

SLIGHTOM, J.L., A.E. BLECHL, and O. SMITHIES. 1980. Human fetal $^G\gamma$ and $^A\gamma$ globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21:** 627.

SPANDIDOS, D. and J. PAUL. 1982. Transfer of human globin genes to erythroleukemic mouse cells. *EMBO J.* **1:** 15.

STANTON, T.H. and L. HOOD. 1980. Biochemical identification of the Qa-1 alloantigen. *Immunogenetics* **11:** 309.

STEIN, R., Y. GRUENBAUM, Y. POLLACK, A. RAZIN, and H. CEDAR. 1982. Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc. Natl. Acad. Sci.* **79:** 61.

STEINMETZ, M., A. WINOTO, K. MINARD, and L. HOOD. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell* **28:** 489.

STEINMETZ, M., J.G. FRELINGER, D. FISHER, T. HUNKAPILLER, D. PEREIRA, S.M. WEISSMAN, H. UEHARA, S. NATHENSON, and L. HOOD. 1981. Three cDNA clones encoding mouse transplantation antigens: Homology to immunoglobulin genes. *Cell* **24:** 125.

VAN DER PLOEG, L.H.T. and R.A. FLAVELL. 1980. DNA methylation in the human $\gamma\delta\beta$-globin locus in erythroid and non-erythroid tissues. *Cell* **19:** 947.

WAALWIJK, C. and R.A. FLAVELL. 1978a. *Msp*I, an isoschizomer of *Hpa*II which cleaves both unmethylated and methylated *Hpa*II sites. *Nucleic Acids Res.* **5:** 3231.

——. 1978b. DNA methylation at a CCGG sequence in the large intron of the rabbit β-globin gene: Tissue-specific variations. *Nucleic Acids Res.* **5:** 4631.

WEYLAND, C., G.H. HAMMERLING, and J. GORONZY. 1981. Recognition of *H-2* domains by cytotoxic T-lymphocytes. *Nature* **292:** 627.

WIGLER, M., D. LEVY, and M. PERUCHO. 1981. The somatic replication of DNA methylation. *Cell* **24:** 33.

WILLING, M.C., A.W. NIENHUIS, and W.F. ANDERSON. 1979. Selective activator of human β- but not γ-globin gene in human fibroblast × mouse erythroleukemia cell lines. *Nature* **277:** 534.

WOLD, B., M. WIGLER, E. LACY, T. MANIATIS, S. SILVERSTEIN, and R. AXEL. 1979. Expression of an adult rabbit β-globin gene stably inserted into the genome of mouse L-cells. *Proc. Natl. Acad. Sci.* **76:** 5684.

ZINKERNAGEL, R.M. and P.C. DOHERTY. 1979. MHC-restricted cytotoxic T-cells: Studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function and responsiveness. *Adv. Immunol.* **27:** 51.