The EMBO Journal Vol.2 No.2 pp.245–254, 1983

Mouse histocompatibility genes: structure and organisation of a K^d gene

S. Kvist*, L. Roberts[1] and B. Dobberstein

European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

Communicated by B. Dobberstein
Received on 2 December 1982

The gene coding for the mouse H-2K^d antigen has been isolated by using a K-locus specific cDNA probe. The complete nucleotide sequence of the gene reveals eight exons separated by seven introns. Transcriptionally important DNA sequences (CCAAT and TATA) precede the first exon. Comparison with other H-2 genes shows extensive homology in exons as well as in introns. Two cDNA clones encoding K^d antigens have been analysed and provide evidence for at least two expressed K^d genes in the DBA/2 mouse. Comparison of the K^d antigen sequence to three other H-2 antigens indicates that gene conversion mechanism(s) act on H-2 genes. Analyses of exon donor and acceptor sites of different H-2 genes and cDNAs show that alternative splicing sites are used by different genes.

Key words: alternative splicing/gene conversion/K locus genes/nucleotide sequence

Introduction

The major histocompatibility complex (MHC) of the mouse, H-2, is a tightly linked cluster of genes located on chromosome 17. These genes play a key role in the immune response of the animal and encode three distinct classes of antigens: Class I, comprising the transplantation antigens K, D and L; Class II, the immune response-associated or Ia antigens; and Class III, the complement factors (Klein, 1975, 1979; Ploegh *et al.*, 1981).

Genes that code for lymphoid differentiation antigens, Qa and T1a (Flaherty, 1980) have also been found within the H-2 complex. The Qa and T1a antigens are structurally very similar to the Class I antigens, and we will hereafter consider the Qa and T1a antigens as belonging to the Class I molecules. Class I antigens are cell membrane glycoproteins with a mol. wt. of 43 000–47 000 daltons. They are non-covalently associated with β_2-microglobulin, mol. wt. 12 000 (Coligan *et al.*, 1981; Michaelson *et al.*, 1977). Structural analyses of the Class I antigens suggest that they can be divided into three extracellular domains – each consisting of ~90 amino acids, a trans-membrane segment of ~25 amino acids and a cytoplasmic domain of 30–40 amino acids (Coligan *et al.*, 1981).

A remarkable feature of the Class I antigens is the very high degree of polymorphism they display. About 50 different alleles have been identified at the K and D loci, respectively (Klein, 1979). No single allele seems to appear with high frequency but rather a large number of alleles occur with low frequency.

More detailed analyses of the genes within the H-2 complex have recently been made. Several laboratories have prepared cDNA from the mRNA coding for the H-2 antigens (Ploegh *et al.*, 1980; Kvist *et al.*, 1981; Sood *et al.*, 1981; Steinmetz *et al.*, 1981a). Southern blot analyses using these cDNA clones indicated that the genes encoding Class I molecules constitute a large multigene family (Steinmetz *et al.*, 1981a,1981b; Cami *et al.*, 1981). Large fragments of DNA cloned into cosmids were employed to show that the genome of the BALB/c mouse has ~35 class I H-2 genes (Steinmetz *et al.*, 1982). Two of these genes, one coming from the Qa region and the other one coding for the L^d antigen, have been sequenced (Steinmetz *et al.*, 1981b; Moore *et al.*, 1982). The analysis of these two sequences revealed a high degree of homology indicating that Qa genes (and most likely also those of T1a) have a very similar structure to the classical H-2 genes. To identify and define a gene or a subfamily of genes within the H-2 complex, it is essential to use DNA probes of greater specificity. In a previous report we identified a probe which is specific for a subfamily of H-2 genes, at least one of which encodes the H-2 K^d antigen (Xin *et al.*, 1982). In this communication we present the complete nucleotide sequence of the K^d gene, its comparison with three other known H-2 antigen sequences and a possible explanation for the generation of the H-2 gene polymorphism within the context of gene conversion.

Results and Discussion

Isolation of genomic clones by using a low copy H-2 cDNA probe

About 6 x 10^5 lambda 1059 clones of DBA/2 liver DNA were screened independently with two different H-2 cDNA probes. The first probe is specific for a subfamily of the H-2 genes and recognizes at least one gene mapped to the K locus (Xin *et al.*, 1982). This probe we refer to as pH-2^d-5b. The second probe is a subclone of pH-2^d-1 (Kvist *et al.*, 1981) containing a *Pst*I fragment consisting of the common part of the 3′ non-coding region in addition to the sequence encoding the cytoplasmic tail, the trans-membrane segment and the third external domain. This probe does not show any specificity for a particular gene but seems to recognize most, if not all, of the H-2 class I genes. We refer to this probe as pH-2^d-1a. Both probes have been described earlier (Xin *et al.*, 1982).

Using pH-2^d-1a, we isolated 30–40 clones, only four of which were recognized by pH-2^d-5b. DNAs from these four clones were prepared and restricted with the enzymes *Bgl*II and *Eco*RI. In a Southern blot analysis, one clone, 2.14, displayed a hybridizing pattern indistinguishable from total genomic DNA using the K locus-specific probe pH-2^d-5b, i.e., a 3.7-kb *Bgl*II and a 13-kb *Eco*RI fragment (Xin *et al.*, 1982). It was likely that 2.14 represents a clone encompassing the K^d gene and was therefore chosen for further analyses. To prove that clone 2.14 is a functional gene, we transfected tk^- mouse L cells with DNA containing 2.14 together with the thymidine kinase gene and were able to isolate tk^+ transformants which also express the gene product of 2.14 on their cell surface. Expression was identified through the use of a

[1]Present address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

*To whom reprint requests should be sent.

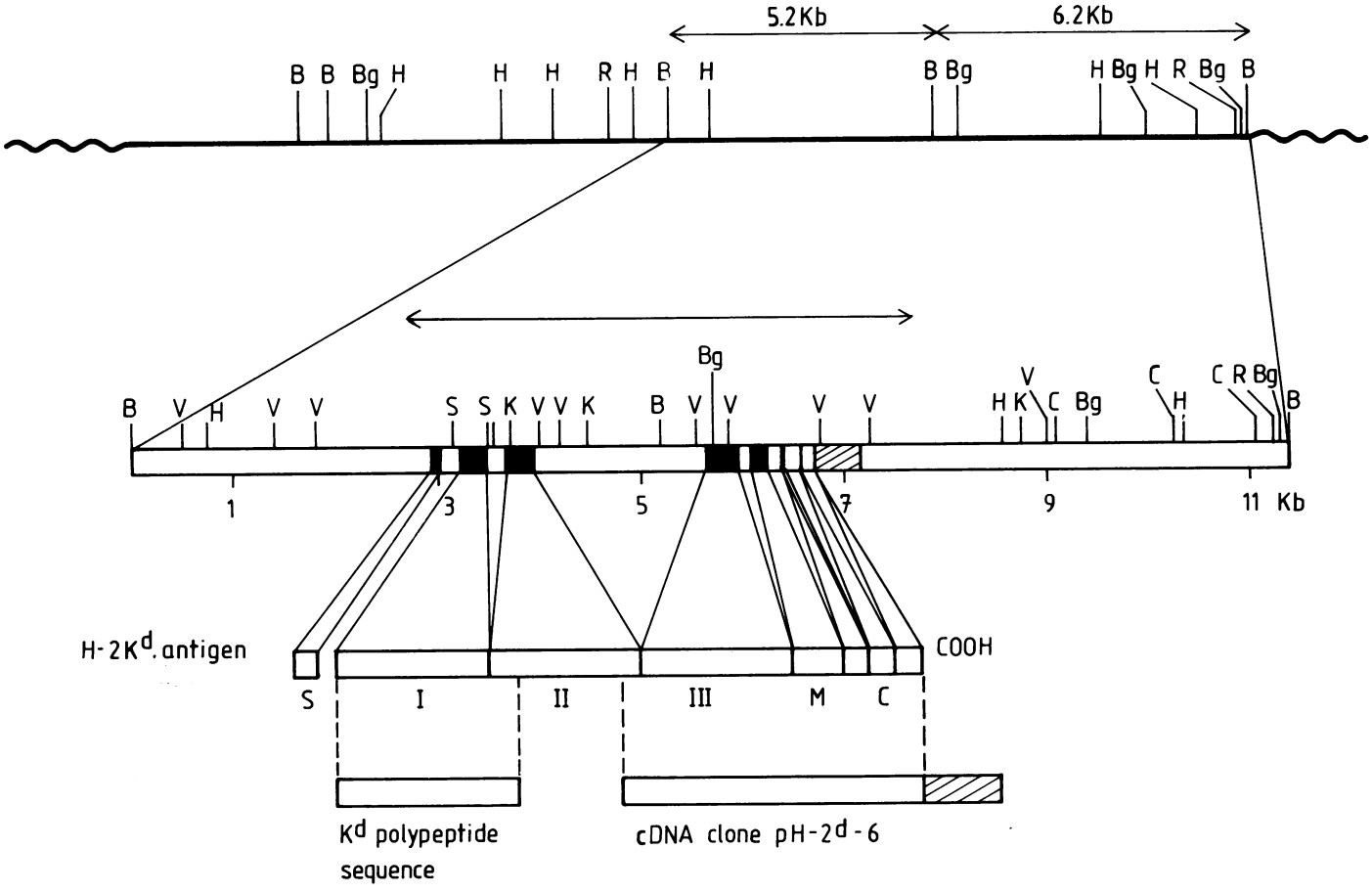

Fig. 1. Restriction map of lambda clone 2.14. The length of the insert in clone 2.14 is ~22 kb. The *Bam*HI site at the right end of the insert was constructed from the partial digest of DBA/2 mouse DNA with *Sau*3A1 and cloned into the *Bam*HI sites of lambda 1059. The 5.2-kb and 6.2-kb *Bam*HI fragments were recloned into pBR322 and subjected to the deletion subcloning procedure in order to be sequenced (Frischauf *et al.*, 1980). Restriction maps of these fragments were made according to Smith and Birnstiel (1976). The location of the exons are indicated by black boxes. The hatched boxes denote the 3′ non-coding region. The arrow indicates the sequenced region of the clone. Exon-intron boundaries were found by comparing: (i) the deduced amino acid sequence of clone 2.14 with the known protein sequence of the K^d antigens (exons 2–3) (Coligan *et al.*, 1981); and (ii) the DNA sequence of clone 2.14 with the cDNA clone pH-2^d-6 (exons 3–4, 4–5, 5–6, 6–7 and 7–8). The signal sequence was localized by typical criteria for such sequences described by Kreil (1981). S, signal sequence; I, II and III, the three external domains; M, membrane-spanning segment; C, cytoplasmic region. Restriction enzymes: B, *Bam*HI; Bg, *Bgl*II; C, *Hind*II; H, *Hind*III, K, *Kpn*I; R, *Eco*RI; S, *Sma*I; V, *Pvu*II.

monoclonal antibody directed against the K^d molecule. The details of these experiments will be given elsewhere (Burgert *et al.*, in preparation). We conclude from all of the data presented above that we have isolated the functional gene for the K^d antigen.

Restriction endonuclease mapping, nucleotide sequencing and identification of the K^d gene within clone 2.14

DNA of clone 2.14 was digested with several restriction enzymes and a partial map was constructed (Figure 1). The total insert has a length of ~22 kb. By Southern blot analyses with our cDNA clones for both coding and 3′ non-coding regions, the gene was mapped to two *Bam*HI fragments, of 5.2 kb and 6.2 kb, respectively. The 5′ end of the gene was found on the 5.2-kb fragment in the middle of the insert, whereas the 3′ non-coding region mapped to the 6.2-kb fragment on the right-hand side in the insert. These two fragments were recloned into the *Bam*HI site in the tetracycline-resistant gene of pBR322. Detailed restriction maps of the two *Bam*HI fragments are shown in Figure 1. To determine the nucleotide sequence, both fragments were subjected to deletion subcloning (Frischauf *et al.*, 1980) and overlapping clones were sequenced by the method of Maxam and Gilbert (1980). The complete nucleotide sequence of the gene is shown in Figure 2. The size of the gene is ~5000 nucleotides and consists of eight exons interrupted by seven introns. The exon-intron boundaries were identified by comparing the deduced amino acid sequence of the gene with the first 98 amino acids of the known protein sequence of the K^d antigen (Coligan *et al.*, 1981) and a cDNA clone, pH-2^d-6, constructed in our laboratory and believed to code for a K^d or a K^d-like antigen (see below).

Gene 2.14 contains a nucleotide sequence that is highly homologous to two other H-2 genes (Steinmetz *et al.*, 1981b; Moore *et al.*, 1982). The first exon of the gene encodes a polypeptide consisting of 21 amino acids. In the middle of the sequence it contains a stretch of hydrophobic amino acids typical of a signal sequence known to be required for translocating the molecule through the membrane of the rough endoplasmic reticulum (Blobel and Dobberstein, 1975; Dobberstein *et al.*, 1979; Kreil, 1981). An intron of ~200 nucleotides divides the signal sequence from the first external domain (exon 2). An intron of the same size separates the first domain from the second (exon 3). In contrast, a large intron of ~1700 bp divides the second domain from the third (exon 4). The membrane spanning segment consists of 39 amino

Fig. 2. The DNA sequence of the K^d gene in clone 2.14. The amino acid sequences encoded by the exons is shown above the DNA sequence. The TATA box and the CCAAT box are found 51 and 77 nucleotides, respectively, upstream from the initiation codon and are underlined. The *Bam*HI site at nucleotide 2460 is indicated by a star.

acids and is encoded by exon 5. Its middle part is devoid of charged residues and is thought to be located within the lipid membrane. It is flanked on either side by polar amino acids. Exons 6, 7 and 8 code for 11, 13 and 10 amino acids, respectively, and together constitute the hydrophilic C terminus of the antigen. The last exon also contains the complete 3′ non-coding region found in the cDNA clones pH-2^d-4 and pH-2^d-5 (see below).

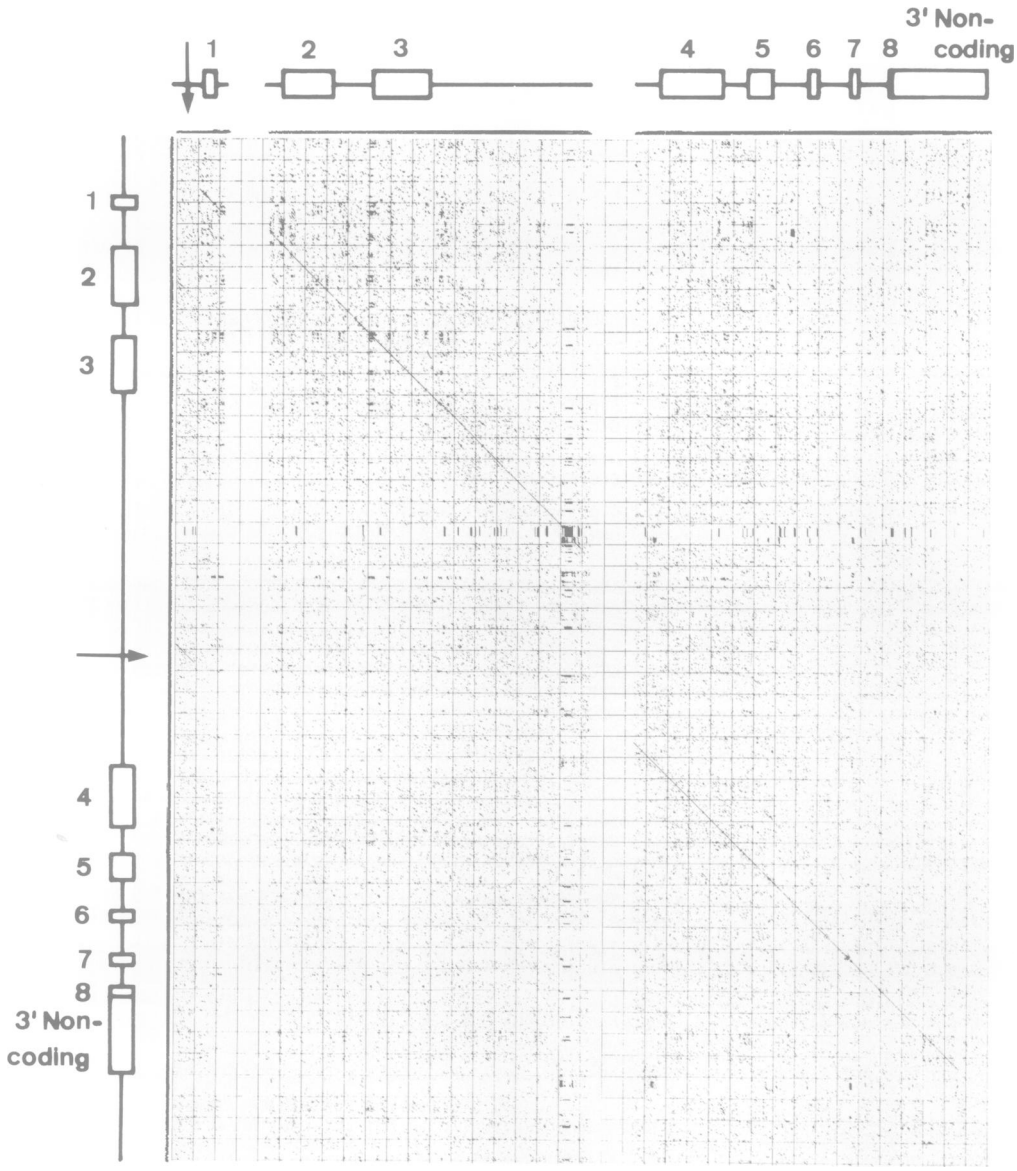

Fig. 3. Homology between the K^d gene and the L^d gene analysed with the dot matrix method. The gross structures of the genes are shown along the sides of the figures. The K^d gene along the vertical side and the L^d gene along the horizontal side. Exons are indicated with boxes and are numbered. At two positions the sequence of the L^d gene is missing and gives rise to breaks in the homology line. The arrows denote the first ~100 nucleotides of the L^d gene found in the third intron of the K^d gene. For details see Materials and methods.

It is interesting to note, that the sizes of exons 1–5 correlate well with protein sequences involved in a particular function (signal sequence and membrane spanning segment) or with biochemically and structurally determined domains (the three external domains). The cytoplasmically exposed C-terminal portion is encoded in three separate exons. No structural or functional subdivision of this segment has yet been revealed. Gene deletion or exchange experiments could be used to elucidate whether protein sequences encoded by exons 6–8 perform separate functions.

Two sequences known to be involved in regulation of DNA transcription precede exon 1 in clone 2.14. These sequences, CCAAT and TATA, are found 77 and 51 nucleotides, respectively, upstream of the first codon of exon 1 (Figure 2). Since, the TATA-box for all genes so far examined is located 26–34 nucleotides upstream from the mRNA start site (Breathnach and Chambon, 1981) the 5′ non-coding region of the mRNA for the gene in 2.14 would comprise only some 20 nucleotides. A similar length was found for the 5′ non-coding region of an immunoglobulin kappa light chain gene (Bodary and Mach, 1982). To demonstrate unequivocally the origin of transcription, S1 mapping analysis must be performed.

In the 3′ non-coding region of 2.14 we found two polyadenylation signals AATAAA (Breathnach and Chambon, 1981). It seems likely that the first signal is used in cDNA clone pH-2^d-5 which is polyadenylated 12 nucleotides downstream. Multiple polyadenylation sites are not unusual and have also been found for the mouse dihydrofolate reductase gene (Setzer *et al.*, 1982).

Comparison of the K^d gene with the L^d gene

Using a computer-generated dot matrix we compared the K^d gene with the L^d gene. This analysis compares every stretch of nine nucleotides from the K^d gene with every nine nucleotide sequences of the L^d gene and when homology is found in at least six nucleotides it places a dot in the two-

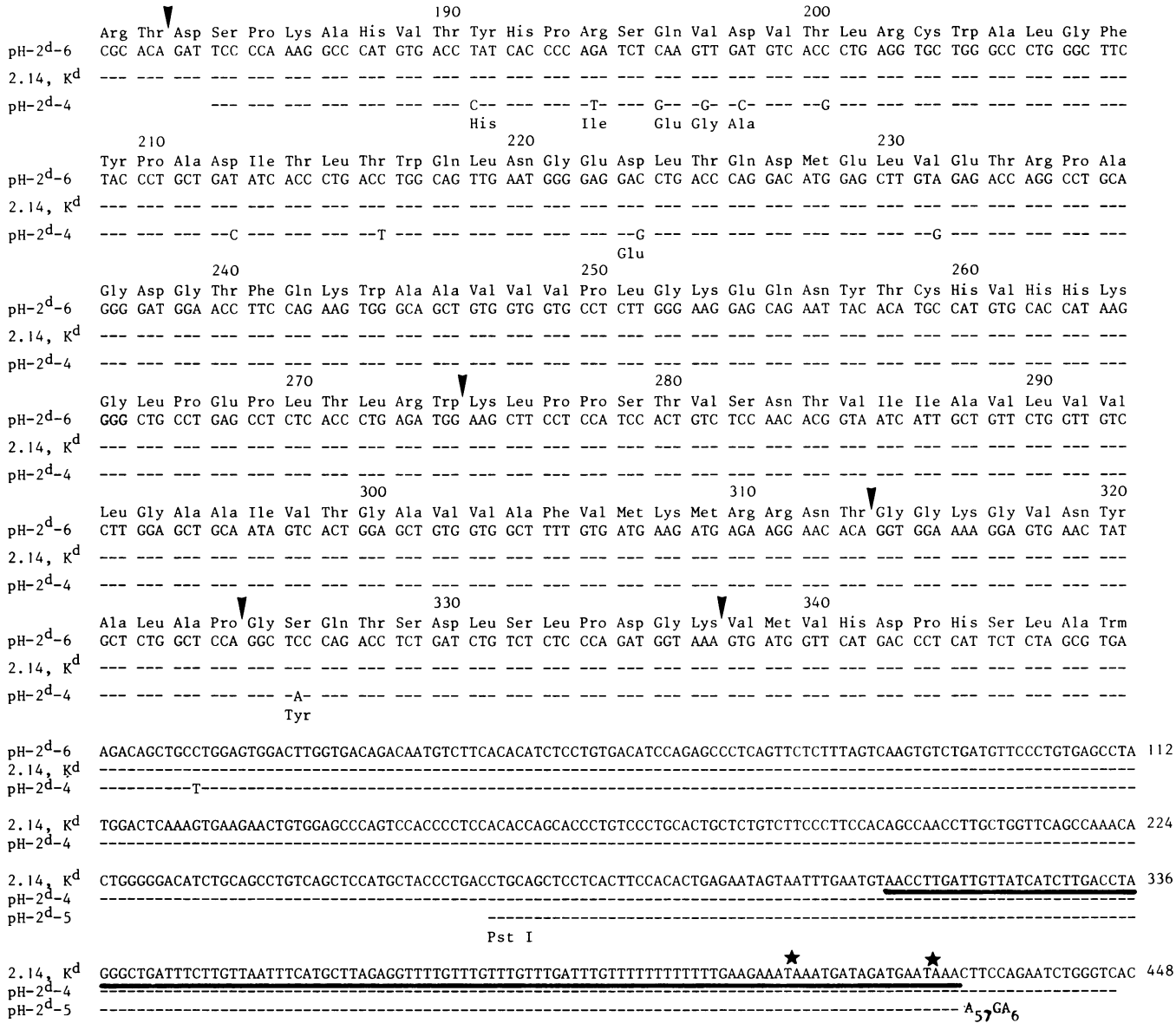

Fig. 4. Alignment of the DNA sequences of pH-2^d-6, the K^d gene in clone 2.14, pH-2^d-4 and pH-2^d-5. Amino acid sequences are shown above the DNA sequence for pH-2^d-6 and under the DNA sequence for pH-2^d-4. Homology is indicated by a dash. The arrows denote the splicing points as compared to the K^d gene in clone 2.14. The amino acids are numbered. The K-specific sequence in the 3′ non-coding region is underlined and the two poly(A) sites are indicated by stars. The numbering on the right hand side of the figure refers to nucleotides after the termination codon (untranslated region). Note that pH-2^d-6 stops at nucleotide 112 in the non-coding region and pH-2^d-5 starts first at nucleotide 267 in the same region.

dimensional matrix seen in Figure 3. With the exception of the extreme 3′ and 5′ ends, extensive homology was found over the whole length of the genes between exons as well as introns.

In the introns, differences between H-2 sequences resulted from nucleotide changes as well as deletions (ranging from 1 to 20 nucleotides), whereas in the exons such differences arose only from nucleotide changes. The percent homology in the introns was between 80 and 90%. When only nucleotide changes were counted the homology was >90%.

Surprisingly there seems to be no homology in the sequence preceding the first exon. This sequence in the L^d gene is found in the long intron separating exon 3 from 4 in the K^d gene (indicated in Figure 3 by the arrows) (Moore *et al.*, 1982). As fragments used in the sequencing of the K^d gene extensively overlapped and moreover the putative transcription consensus sequences were found, we feel confident that the fragments have been placed correctly. The differences in the case of the L^d gene might be due to a transposition such as reported earlier for a 3′ non-coding segment of the Qa gene 27.1 (Steinmetz *et al.*, 1981b), to a cloning artefact, or a misplacement of M13 clones used for the sequencing of gene 27.5 (Moore *et al.*, 1982).

Evidence for at least two K^d-like genes

O'Neill and Parish (1981) have shown that two K locus genes might exist in the k haplotype. We have indications that this might also be true for the d haplotype. Comparison of the K^d protein sequence and several cDNA clones support this.

A comparison of 114 amino acids deduced from clone 2.14 with the known K^d protein sequence (Pease *et al.*, 1982; Col-

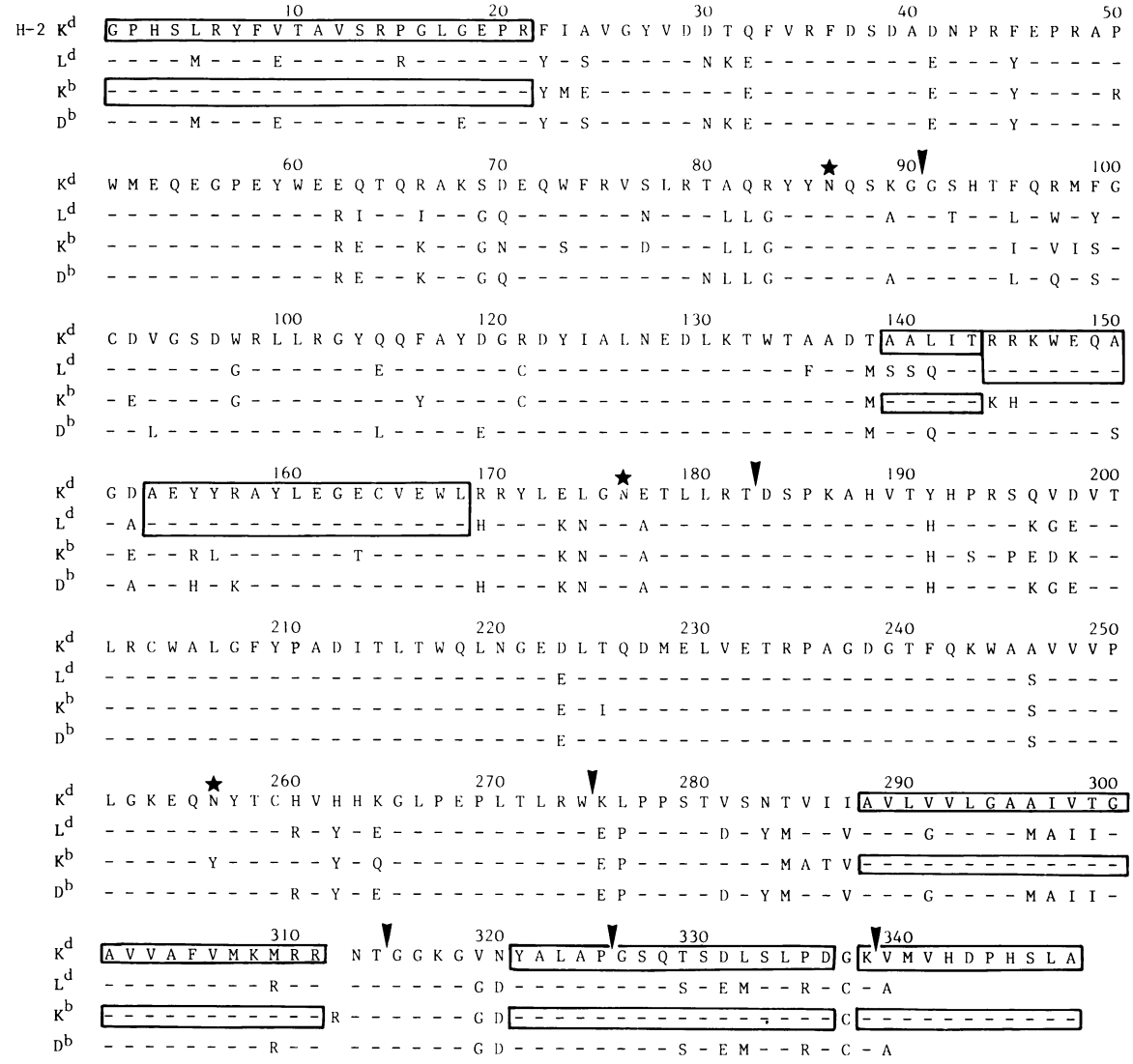

Fig. 5. Amino acid comparison of four H-2 antigens. A dash indicates homology with the K^d antigen on top. Homology over long regions between two antigens are drawn in boxes. The amino acids are numbered above the sequences. The arrows indicate the putative splicing points in corresponding genes or cDNAs. Potential glycosylation sites are denoted with stars. The single letter code for the amino acids is used where A is Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu, M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. The L^d sequence is taken from Moore *et al.* (1982); the K^b sequence is from Coligan *et al.* (1981) and Reyes *et al.* (1982); and the D^b sequence from Reyes *et al.* (1982b) and Maloy and Coligan (1982).

igan *et al.*, 1981) showed identity in 109 positions. The five differences found could be explained by the presence of two H-2K genes in the d haplotype or by sequencing errors.

We have isolated three cDNA clones which seem to encode H-2K^d or K^d-like molecules (Figure 4). The gene present in clone 2.14 shows a complete identity with pH-2^d-6 (Figure 4). pH-2^d-6 enabled the localization of the exon-intron boundaries between exon 3 – 4, 4 – 5, 5 – 6, 6 – 7 and 7 – 8.

The clones pH-2^d-4 (Lalanne *et al.*, 1982) and pH-2^d-5 (Xin *et al.*, 1982) carry the K locus-specific fragment pH-2^d-5b at their 3′ non-coding end which we used to isolate clone 2.14. Unfortunately, the non-coding region of pH-2^d-6 is very short. Its 3′ end stops before running into the sequence of pH-2^d-5 (i.e., it ends at nucleotide 112 after the termination codon). Because pH-2^d-5 contains only the extreme 3′ non-coding sequence, we cannot exclude that pH-2^d-5 and pH-2^d-6 indeed represent two cDNA clones belonging to the same species of mRNA and coding for the same K^d antigen. However, clone pH-2^d-4 displays a sequence that differs by 12 nucleotides (11 in the coding region and one in the 3′ non-coding region) from pH-2^d-6 and clone 2.14. Of these 12 nucleotides, seven give rise to amino acid replacements. Five of these seven amino acid differences are found within the hypervariable cluster located between amino acids 190 and 200 (Figures 4 and 5). The clone pH-2^d-4 also encodes 10 amino acids in exon 8, thereby displaying the typical H-2K antigen C terminus (see below). We conclude that pH-2^d-4 is encoded by a gene different from the gene in clone 2.14. Most likely this gene is also located in the K region.

Comparison of different H-2 antigens

To localize hypervariable as well as more constant regions, we compared four H-2 antigens (Coligan *et al.*, 1981; Reyes *et al.*, 1982a, 1982b; Maloy and Coligan, 1982) (Figure 5). The first and the second domains are much more variable than the third. Stretches of up to 24 amino acids in a row are found to be identical in the third domain. Differences are spread throughout the total sequence making it hard to identify true

Table I. Amino acid sequence homology of H-2 antigens

Haplotypes compared	% Homology in different regions of the molecules					
	I	II	III	M	C	Total
$K^d - K^b$	81	80	87	82	91	84
$K^d - D^b$	78	83	90	69	68	81
$K^d - L^d$	77	82	90	69	68	80
$K^b - D^b$	86	78	90	72	80	83
$K^b - L^d$	83	79	90	72	80	82
$L^d - D^b$	92	85	100	100	100	94

The regions compared are: I, II and III, the three external domains; M, the membrane-spanning segment (exon 5); C, the cytoplasmic domain. The sequences are taken from the same references as in the legend to Figure 5.

hypervariable clusters. Regions with the most pronounced differences are found between amino acids 62 and 80, 95 and 99, 193 and 198 (Figure 5). The part of the membrane-spanning segment exposed on the cell surface (amino acids 275 – 287) also varies considerably whereas the part facing the cytoplasm is well conserved (Dobberstein *et al.*, 1983). Throughout the sequence the D^b and L^d antigens are very similar (Table I). With respect to the three external domains K^b is more similar to D^b and L^d than to K^d. However, from residue 278 onwards the K^d molecule shows many homologies to the K^b, apart from exon 7, where it differs by two residues. These two amino acids are the same in K^b, L^d and D^b. The K^d and K^b molecules are, in addition, nine amino acids longer at their C termini. This seems to be a unique feature of the antigens near the K end of the H-2 complex. It has been shown before that D^b and L^d antigens are more similar to one another than D^b is to D^d. Therefore, it has been suggested that D^b and L^d represent true alleles (Maloy and Coligan, 1982). These two molecules do not differ in a single residue from amino acid number 158 onwards and show 94% homology over all. Such is not the case between K^d and K^b. They differ extensively in the three external domains and the high percentage of homology, 83%, is due to the great similarity in exons 5 – 8 (Figure 5 and Table I).

The K^d, L^d and D^b molecules each have three potential glycosylation sites for asparagine-linked carbohydrates located at residues 86, 176 and 256. In the K^b molecule the asparagine residue at position 256 is replaced by a tyrosine resulting in only two potential carbohydrate binding sites in this molecule. The K^d molecule has two fewer cysteine residues than the K^b molecule (amino acids 121 and 337). The arginine at positon 121 in the K^d molecule is also found in D^b. At position 337, however, the K^d antigen is unique, having a glycine instead of a cysteine.

Does gene conversion act on H-2 genes?

Several mechanisms have been proposed to explain the polymorphism of the H-2 genes. Among these mechanisms are control of gene expression and gene duplication with generation of diversity through point mutations (Bodmer, 1973; Silver and Hood, 1976). Control of gene expression can be ruled out, as it has clearly been shown by Southern blot analysis that H-2 genes of different haplotypes show true polymorphism (Cami *et al.*, 1981; Steinmetz *et al.*, 1982). The question remains whether point mutations alone are responsible for the observed high degree of polymorphism.

Recently, a novel means for exchange of genetic information among reiterated gene families has attracted considerable

Table II. Occurrence of nucleotide substitutions in different regions of H-2 antigens

Haplotypes compared	Replacement					Silent				
	I	II	III	M	C	I	II	III	M	C
$K^d - K^b$	17	18	12	7	3	ND	12	6	1	0
$K^d - D^b$	20	16	9	13	8	ND	9	4	9	0
$K^d - L^d$	21	17	9	13	8	14	7	4	9	0
$K^b - D^b$	13	21	9	13	5	ND	14	3	7	0
$K^b - L^d$	14	21	9	13	5	ND	14	3	7	0
$L^d - D^b$	6	15	0	0	0	ND	9	0	0	0

The regions are listed in the legend to Table I. The numbers given are actual nucleotide substitutions. Silent mutations are counted as substitutions not changing the amino acid. For multiple events we counted two substitutions as one silent and one replacement and for three substitutions, two replacements and one silent. The reason for this is the random expectation for a silent mutation being 24%. ND, not determined. The DNA sequences were taken from: K^b, Reyes *et al.* (1982a); L^d, Moore *et al.* (1982); D^b, Reyes *et al.* (1982b).

attention. This phenomenon, called gene conversion, occurs readily in yeast (Klein and Petes, 1981; Jackson and Fink, 1981) and probably also within the globin (Slightom *et al.*, 1980) and the immunoglobulin gene families (Egel, 1981; Schreier *et al.*, 1981; Dildrop *et al.*, 1982; Ollo and Rougeon, 1982; Gough, 1982). Gene conversion differs from unequal crossing-over in that the latter often increases or decreases the number of genes in a family of closely related genes, whereas the former does not change the number of genes. Gene conversion seems to affect a local region of DNA, and in most cases only one of the two genes involved undergoes conversion (Egel, 1981). In addition, it appears that only genes very homologous to each other can exchange DNA this way. Since the histocompatibility gene family with its ~40 genes (at least in the d haplotype; Steinmetz *et al.*, 1982) displays a very high degree of homology between independent genes; gene conversion could have played a role in generating the polymorphism (Pease *et al.*, 1982).

By comparing the sequences of the four different antigens, we observed regions where two or three sequences show identity (Figure 5). Such segments of identity within a highly variable surrounding are difficult to explain by point mutations alone. They are, however, fully consistent with gene conversion or double crossing-over. For the creation of polymorphism in H-2 genes, mutations, gene conversion and crossing-over all seem to have been involved. Deciding the frequency of each, and thereby the most dominating event, has to await the comparison of many more sequences from the H-2 gene family.

Replacement substitutions versus silent mutations

To be able to understand the evolution of related genes, like those of MHC, it is important to compare changes in the nucleotide sequence. These changes can either lead to a replacement of a particular amino acid by another residue, or they can be 'silent' and not lead to replacements. We have compared the codons of the H-2 genes in Table II. Independently of which combinations were tested, the silent substitutions in the three exons of the external domains account for 25 – 41% of the total number of substitutions of these genes. Interestingly, no silent mutations are found in the exons coding for the cytoplasmic C terminus. However, as many as eight substitutions in this part of the molecule lead to replace-

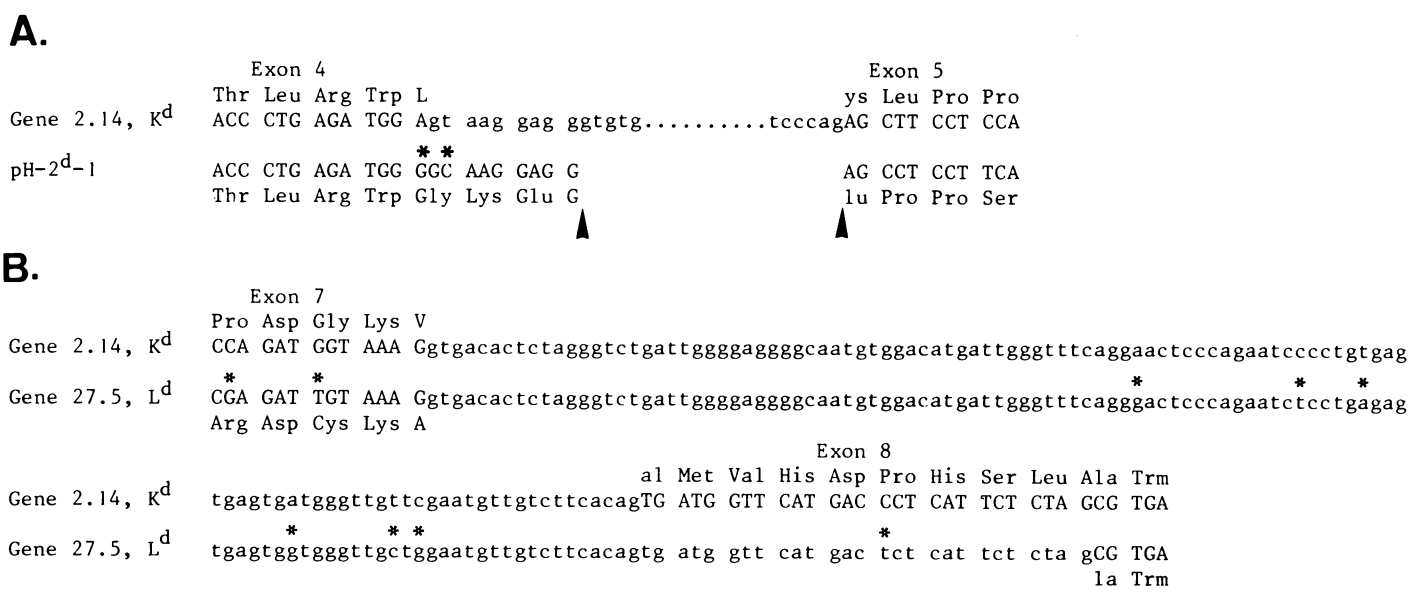

Fig. 6. DNA sequences surrounding two different exon-intron boundaries in H-2 genes. **(A)** The donor site of exon 4 and the acceptor site of exon 5 in the K^d gene are compared to corresponding DNA sequences in pH-2^d-1. The arrows indicate the putative splicing point in pH-2^d-1. **(B)** Comparison of the donor site of exon 7, intron 7 and the acceptor site of exon 8 in the K^d and L^d genes. For both **(A)** and **(B)**: DNA sequences of exons are shown in capital letters and of introns in small letters. Nucleotide changes are indicated by stars. Amino acids are shown above the DNA sequences for the K^d gene and below the DNA sequences for pH-2^d-1 and the L^d gene. The L^d sequence is taken from Moore *et al.* (1982).

ment of amino acid residues. At present we do not know the explanation for this phenomenon.

Alternative splicing sites are used by different H-2 genes

We have previously observed that the length of exon 4 seems to vary between different H-2 genes. The donor site of exon 4 in the K^d gene reveals an adenosine as the last nucleotide (Figure 6A). Together with the acceptor site of exon 5 (sequence AG), this sequence (AAG) codes for a lysine residue. Interestingly, a cDNA clone from our laboratory (pH-2^d-1), believed to code for a D^d antigen (Kvist *et al.*, 1981), has three additional amino acids in the corresponding sequence (Figure 6A). The DNA sequence encoding the additional residues constitutes the first nine nucleotides of the following intron of the K^d gene. The adenosine nucleotide at the donor site in the K^d gene is changed for a guanosine in pH-2^d-1. However, adenosine and guanosine can both be used as donor sites even if guanosine is preferred in 90% of all cases (Lewin, 1980). More important is the fact that the sequence GT at the start of the intron in the K^d gene is changed to GC in pH-2^d-1. This destroys the consensus sequence at the start of the intron in the gene corresponding to pH-2^d-1. The second splicing site, nine nucleotides downstream is then used.

A second alternative splice site was found at the end of exon 8. A comparison of the K^d gene with the L^d gene (Moore *et al.*, 1982) shows that exon 8 in the former gene codes for 10 amino acids, whereas the latter gene only has one residue encoded by exon 8 (Figure 6B). However, the nucleotides encoding the additional nine amino acids in the K^d gene are preserved within the intron preceding exon 8 in the L^d gene. That is, the K^d gene uses an acceptor site 27 nucleotides upstream from the acceptor site used by the L^d gene. The homology in the intron between the 7th and 8th exon is extensive. Only six nucleotides are different in the K^d gene intron. None of these changes are in the immediate vicinity of the splicing consensus sequence of the intron. It is thus conceivable that pH-2^d-6 is not the only transcript from gene 2.14. A transcript might exist in which the splice site equivalent to the L^d gene has been used. Another possibility could be that the specificity of the splicing enzymes might require determinants located distant from the splice site and the six nucleotides changed in the intron might be sufficient to enable use of a different splice site. Also, as cDNA clone pH-2^d-6 is not of full length there is still the possibility that it is not coded by gene 2.14 but by a closely related gene, displaying sequence differences at the splice site of exons 7 and 8. Clearly, to demonstrate alternative splicing, the corresponding cDNAs must be characterized. In a human HLA-gene still another splicing pattern was observed (Malissen *et al.*, 1982). This gene lacks the intron between exon 7 and 8.

Conclusions and Implications

We have shown that the extensive variability and polymorphism of H-2 antigens extend into the cytoplasmic C terminus. It is generally believed that the polymorphism of the histocompatibilty antigens is a reflection of their function. The question arises as to whether the variabilty in the cytoplasmic segment and the fact that this part of the molecule is encoded by three separate exons have a functional significance. Zinkernagel and Doherty (1979) proposed that histocompatibility antigens function as restricting elements in T-cell killing of virus-infected cells. A physical association between the adenovirus-coded protein and histocompatibilty antigens has been shown (Kvist *et al.*, 1978). This interaction might require the presence of the cytoplasmic portion of the histocompatibility antigen (Signas *et al.*, 1982). The C termini of the H-2 antigens could, for example, function in the interaction with viral proteins, and their variability in amino acid sequence could increase the binding repertoire. If an association between histocompatibility antigen and viral protein is required for T-lymphocyte killing of infected cells, the degree of diversity at the C terminus would directly influence the T-killer cell response. Genetic engineering of the C-terminal exons and expression of the modified genes will be powerful tools in elucidating more precisely the functions of the C-terminal segment.

Materials and methods

Restriction enzymes, Klenow DNA polymerase and polynucleotide kinase were purchased from Boehringer (Mannheim, FRG) or from Biolabs (Bishop's Stortford, UK). [α-^{32}P]dNTPs (10 mCi/ml, 3000 Ci/mmol) and [γ-^{32}P]ATP (10 mCi/ml, 5000 Ci/mmol) were from Amersham, UK. Nitrocellulose filters were obtained from Schleicher and Schull (Dassel, FRG). Oligo(dT)-cellulose (T3), oligonucleotide linkers and bacterial alkaline phosphatase were from Collaborative Research (Waltham, MA). Reverse transcriptase, terminal transferase and low melting agarose were from Bethesda Research Labortories (BRL).

Construction of a cDNA library and isolation of H-2 cDNA clones

RNA was extracted from SL2 lymphoma cells grown in DBA/2 (H-2^d haplotype) mice as described previously (Kvist *et al.*, 1981). Poly(A)$^+$ RNA (mRNA) was isolated by using oligo(dT)-cellulose (Aviv and Leder, 1972). A cDNA library was constructed from the mRNA as described by Maniatis *et al.* (1982). The vector was pBR322. Colonies were selected for their ability to hybridize to clone pH-2^d-1 (Kvist *et al.*, 1981) and relevant plasmid DNAs were sequenced.

Isolation of genomic DNA and blot hybridization

DNA was isolated from a single mouse liver essentially as described by Blin and Stafford (1976). A lambda 1059 library (Karn *et al.*, 1980) containing liver DNA from a DBA/2 mouse was obtained from H. Lehrach and D.P. Leader and will be described elsewhere (H. Lehrach *et al.*, in preparation). Phage clones containing genomic DNA were identified by plaque hybridization (Benton and Davis, 1977). Phage DNA was prepared according to Karn *et al.* (1980). Southern blot analyses of genomic DNA or phage DNA were carried out by separating the restricted DNA on 0.5–1% agarose gels followed by transferring the DNA to nitrocellulose filters (Southern, 1975). DNA probes were labelled by nick-translation to a specific activity of $1-5 \times 10^8$ c.p.m./μg DNA (Rigby *et al.*, 1977). Hybridizations and washings of the filters were essentially as described by Gergen *et al.* (1979). Filters were hybridized for 20 h at 65°C in 10 x Denhardt (0.2% each of bovine serum albumin, Ficoll type 400 and polyvinylpyrrolidone PVP-360, Sigma, St. Louis), 4 x SET (0.6 M NaCl, 0.12 M Tris-HCl, pH 8.0 and 1 mM EDTA), 0.1% SDS and 0.1% sodium pyrophosphate. The hybridization solution also contained 50 μg/ml of poly(A) and poly(C), from Boehringer Mannheim. For phage hybridization, 50 μg/ml of sonicated *Escherichia coli* DNA was added. Washings were made in 4 x SET for 15 min twice, 2 x SET for 15 min twice and once in 0.1 SET for 15 min. When pH-2^d-5b was used as the probe the hybridization was made at 59°C instead of 65°C and the final wash was in 0.2 x SET.

DNA restriction mapping and sequence analysis

Phage DNA from clone 2.14 was digested with the restriction enzymes listed in the legends to Figure 1 and a partial restriction map was constructed. The 5.2- and 6.2-kb *Bam*HI fragments were cloned into the *Bam*HI site in the tetracycline-resistant gene of pBR322. Detailed restriction maps of both fragments were constructed by the method of Smith and Birnstiel (1976). For DNA sequence analysis of both the cDNA clone pH-2^d-6 and the two *Bam*HI fragments we used the subcloning procedure described by Frischauf *et al.* (1980) using *Cla*I linker. Overlapping clones were selected and DNAs from these were cut with *Cla*I and labelled at their 3′ end with [^{32}P]dCTP by using the Klenow fragment of DNA polymerase I. To label the 5′ end, fragments were cut with *Cla*I, the 5′ phosphate group removed with bacterial alkaline phosphatase and then labelled with [γ-^{32}P]ATP by using polynucleotide kinase. DNA fragments were then cut with *Eco*RI. Labelled DNA fragments were sequenced by the method of Maxam and Gilbert (1980). The following five reactions were used: G, G+A, A+C, T+C and C. To certify the sequence connecting the two *Bam*HI fragments, the 400-bp overlapping *Ava*II fragment was isolated, recloned by using *Bam*HI linkers into pUC8 and sequenced as described above.

Computer programs

Comparison of the K^d gene with the L^d gene (Figure 3) was carried out by a computer program developed by H. Lehrach. In principle, this dot matrix program compares every sequence of nine nucleotides in the first DNA sequence to every stretch of nine nucleotides in the second DNA sequence. When homology is found for six or more of the nucleotides compared, the plotter inserts a dot. Two completely homologous sequences will then give rise to a diagonal line on the outprint.

Acknowledgements

We thank Drs. A.M. Frischauf and H. Lehrach for valuable advice and discussions throughout the work. We also appreciate their generous attitude of providing us with the genomic DNA library and cloning vehicles. We are grateful to Drs. D. Meyer and H. Garoff for critical reading of the manuscript and the comments they had. We also thank U. Hellert for excellent technical assistance and A. Steiner and W. Moses for typing the manuscript. L.R. was a recipient of a fellowship from the European Molecular Biology Organisation. This work was supported by the Deutsche Forschungsgesellschaft DO199/4-1.

References

Aviv,H. and Leader,P. (1972) *Proc. Natl. Acad. Sci.*, **69**, 1408-1412.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Blin,N. and Stattford,D.W. (1976) *Nucleic Acids Res.*, **3**, 2303-2308.
Blobel,G. and Dobberstein,B. (1975) *J. Cell Biol.*, **67**, 835-851.
Bodary,S. and Mach,B. (1982) *EMBO J.*, **1**, 719-724.
Bodmer,W.F. (1973) *Transplant. Proc.*, **4**, 1471-1475.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Cami,B., Bregegere,F., Abastado,J.P. and Kourilsky,P. (1981) *Nature*, **291**, 673-675.
Coligan,J.E., Kindt,T.J., Uehara,H., Martinko,J. and Nathenson,S.G. (1981) *Nature*, **291**, 35-39.
Dildrop,R., Bruggemann,M., Radbruch,A., Rajwesky,K. and Beyreuther,K. (1982) *EMBO J.*, **1**, 635-640.
Dobberstein,B., Garoff,H., Warren,G. and Robinson,P.J. (1979) *Cell*, **17** 759-769.
Dobberstein,B., Kvist,S. and Roberts,L. (1983) *Proc. R. Soc., London*, in press.
Egel,R. (1981) *Nature*, **290**, 191-192.
Flaherty,L. (1980) in Dorf,M.E. (ed.), *The Role of the Major Histocompatibility Complex in Immunology*, Garland STEM Press, NY, pp. 33-57.
Frischauf,A.-M., Garoff,H. and Lehrach,H. (1980) *Nucleic Acids Res.*, **8**, 5541-5549.
Gergen,J.P., Stern,R.H. and Wensink,P.C. (1979) *Nucleic Acids Res.*, **7**, 2115-2136.
Gough,N. (1982) *Trends Biochem. Sci.*, **7**, 307.
Jackson,J.A. and Fink,G.R. (1981) *Nature*, **292**, 306-311.
Karn,J., Brenner,S., Barnett,L. and Cesareni,G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5172-5176.
Klein,J. (1975), *Biology of the Mouse Histocompatibility Complex*, published by Springer Verlag, Berlin.
Klein,J. (1979) *Science (Wash.)*, **203**, 516-521.
Klein,H.L. and Petes,T.D. (1981) *Nature*, **289**, 144-148.
Kreil,G. (1981) *Annu. Rev. Biochem.*, **50**, 317-348.
Kvist,S., Ostberg,L., Persson,H., Philipson,L. and Peterson,P.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 5674-5678.
Kvist,S., Bregegere,F., Rask,L., Cami,B., Garoff,H., Daniel,F., Wiman,K., Larhammar,D., Abastado,J.P., Gachelin,G., Peterson,P.A., Dobberstein,B. and Kourilsky,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2722-2776.
Lalanne,J.L., Bregegere,F., Delarbre,C., Abastado,J.P., Gachelin,G. and Kourilsky,P. (1982) *Nucleic Acids Res.*, **10**, 1039-1049.
Lewin,B. (1980) *Cell*, **22**, 324-326.
Malissen,M., Malissen,B. and Jordan,B.R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 893-897.
Maloy,W.L. and Coligan,J.E. (1982) *Immunogenetics*, **16**, 11-22.
Maniatis,T., Fritsch,E.F., Sambrook,J. (1982) *Molecular Cloning, a Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Michaelson,J., Flaherty,L., Vitetta,E. and Poulik,M.D. (1977) *J. Exp. Med.*, **145**, 1066-1070.
Moore,K.W., Sher,B.T., Sun,Y.H., Eakle,K.A. and Hood,L. (1982) *Science (Wash.)*, **215**, 679-682.
Ollo,R. and Rougeon,F. (1982) *Nature*, **296**, 761-763.
O'Neill,H.C. and Parish,C.R. (1981) *Mol. Immunol.*, **18**, 713-722.
Pease,L.R., Schulze,D.H., Pfaffenbach,G.M. and Nathenson,S.G. (1982) *Proc. Natl. Acad. Sci. USA*, in press.
Ploegh,H.L., Orr,H.T. and Strominger,J.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6081-6085.
Ploegh,H.L., Orr,H.T. and Strominger,J.L. (1981) *Cell*, **24**, 287-299.
Reyes,A.A., Schold,M. and Wallace,R.B. (1982a) *Immunogenetics*, **16**, 1-9.
Reyes,A.A., Schold,M., Itakura,K. and Wallace,R.B. (1982b) *Proc. Natl. Acad. Sci. USA*, **79**, 3270-3274.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.
Schreir,P.H., Bothwell,A.L.M., Muller-Hill,B. and Baltimore,D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4495-4499.
Setzer,D.R., McGrogan,M. and Schimke,R.T. (1982) *J. Biol. Chem.*, **257**, 5143-5147.
Signas,C., Katze,M.G., Persson,H. and Philipson,L. (1982) *Nature*, **299**, 175-178.

Silver,J. and Hood,L. (1976) *Proc. Natl. Acad. Sci. USA,* **73**, 599-603.
Slightom,J.L., Blechl,A.E. and Smithies,O. (1980) *Cell,* **21**, 627-638.
Smith,H.O. and Birnstiel,M.L. (1976) *Nucleic Acids Res.,* **3**, 2387-2399.
Sood,A.K., Pereira,D. and Weissman,S.M. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 616-620.
Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.
Steinmetz,M., Frelinger,J.G., Fischer,D., Hunkapiller,T., Pereira,D., Weissman,S.M., Uehara,H., Nathenson,S. and Hood,L. (1981a) *Cell,* **24**, 125-134.
Steinmetz,M., Moore,K.W., Frelinger,J.G., Sher,B.T., Shen,F.W., Boyse, E.A. and Hood,L. (1981b) *Cell,* **25**, 683-692.
Steinmetz,M., Winoto,A., Minard,K. and Hood,L. (1982) *Cell,* **28**, 489-498.
Xin,J.-H., Kvist,S. and Dobberstein,B. (1982) *EMBO J.,* **1**, 467-471.
Zinkernagel,R.M. and Doherty,P.C. (1979) *Adv. Immunol.,* **27**, 51-141.