Journal of Immunogenetics (1990), **17,** 133–150.

INACTIVATION OF THE *H-2K1*^{k} GENE COULD INVOLVE THE SUBSTITUTIONS OF METHYLATED CpGs

B. C. SIM,* F. G. GROSVELD† AND K. M. HUI*

**Institute of Molecular and Cell Biology, National University of Singapore, Singapore, and*
†Gene Structure and Expression, National Institute for Medical Research, London, England

(*Received 1 November 1989; accepted 20 February 1990*)

SUMMARY

By the isolation of overlapping cosmid clones and 'chromosome walking' studies from the *H-2K*^{k} gene, we have obtained cosmid clones encoding the *H-2K1*^{k} gene from two separate cosmid libraries. The nucleotide sequence of one of the clones was determined. The cloned *H-2K1*^{k} gene could be transcribed *in vitro* to give a normal H-2 class I mRNA of 1·7 kb. However, the deletion of four nucleotides in exon 3 of the *H-2K1*^{k} gene results in a translation termination codon at the beginning of exon 4. In agreement with this, when expressed in human cells, the *H-2K1*^{k} gene gave a truncated, cytoplasmic polypeptide of M_r 36,000. Therefore, although the *H-2K1*^{k} gene is homologous to other class I MHC genes in its molecular organization and nucleotide sequence, it is a pseudogene. When compared to the nucleotide sequence of the *H-2K*^{k} gene, the *H-2K1*^{k} gene has undergone many substitutions of methylated CpG residues (^{me}CpG). This represents further evidence to suggest that this gene is inactive.

INTRODUCTION

The molecular organization of the mouse class I MHC genes has been studied extensively. There are 26 class I genes reported for the B10 mouse (Flavell *et al.*, 1985) and at least 32 class I genes are present in the BALB/c mouse (Hood *et al.*, 1983). These genes are distributed among four genetic regions, H-2K; H-2D, H-2 L; Qa and Tla. The H-2K region of the B10, BALB/c, AKR and C3Hf/HeN mice are reported to contain two class I genes (Flavell *et al.*, 1985; Hood *et al.*, 1983). In the B10 mouse, these two genes are about 15 kb apart and are arranged in a head-to-tail configuration (Flavell *et al.*, 1985). One of these two genes, *H-2K*, encodes the classical transplantation antigens. This gene has been extensively studied and the nucleotide sequences of many of its alleles have been determined (Weiss *et al.*, 1983; Mellor *et al.*, 1983; Arnold *et al.*, 1984; Reddy & Pan, 1985; Minamide *et al.*, 1988). The second gene, termed *H-2K1* for the b and k haplotypes (Flavell *et al.*, 1985) and *H-2K2* for the d haplotype

Correspondence: Kam M. Hui, Institute of Molecular and Cell Biology, National University of Singapore, Kent Ridge, Singapore 0511.

(Hood *et al.*, 1983), has not been well studied. The nucleotide sequence of the *H-2K1*^{k} gene of the C3H mouse has only recently been reported (Watts *et al.*, 1989b). However, the function and evolutionary implications of the *H-2K1*^{k} gene have not been studied.

The genetic organization of the H-2K region of the mouse is also unique from an evolutionary point of view. In most animals, including man, all class I loci are telomeric to the class II genes (Ploegh *et al.*, 1981). In the mouse, however, the H-2K region is centromeric to the class II genes (Klein, 1975). For this reason, it has been suggested that the H-2K locus arose by translocation to its present site during evolution (Bodmer, 1981). In support of this hypothesis, Flavell and colleagues have reported the similarity between the H-2K and Qa region. They reported that the H-2K and Qa region genes occur as gene pairs in a head-to-tail configuration (Flavell *et al.*, 1985) and that they have similar hybridization patterns when examined with 4 extragenic probes (Weiss *et al.*, 1984). From these observations, these authors suggested that the H-2K region was generated by the translocation of a 'Q6 and Q7-like' gene pair from the Qa region (Weiss *et al.*, 1984).

Previously, we have cloned and sequenced the *H-2K*^{k} gene from the C3Hf/HeN mouse (Minamide *et al.*, 1988). The H-2K molecules are the classical transplantation antigens, which serve as restriction molecules to permit cytotoxic T cells to recognize foreign antigens during cell-mediated immune responses (Zinkernagel & Doherty, 1979). The only reported role of the H-2K1 gene to date has been its implication as the donor gene to generate mutant K^{bm4} gene (Nathenson *et al.*, 1986). To study its potential immune function and genetic origin, we have cloned and studied the H-2K1 gene from the C3Hf/HeN mouse, so that a comparison at the nucleotide sequence level can be made with the *H-2K*^{k} and Qa region genes.

MATERIALS AND METHODS

Isolation of cosmid clones

Cosmid libraries were constructed in the vector pTCF (Grosveld *et al.*, 1982) using high molecular weight genomic DNA isolated either from C3Hf/HeN mouse livers or from the AKR thymoma K36.16 tumour cells (Hui *et al.*, 1984). Methods were as described by Minamide and colleagues (1988). The C3Hf/HeN mice were obtained from Harlan Spraque Dawley via the Animal Production Unit of the National Cancer Institute (Frederick, MD, U.S.A.) (Callahan *et al.*, 1981). This particular subline of C3Hf/HeN mouse has been subsequently reported by us (Minamide *et al.*, 1988) and confirmed by others (Vogel *et al.*, 1988), to be identical to the C3H/HeN mouse. Over one million recombinant bacterial colonies of each library were screened. Class I MHC cosmid clones were selected by hybridization to the pH2IIa, pH2III and the 640 bp *Bam*HI 5′-flanking DNA probes. The H-2 class I DNA probes and the conditions for hybridization employed in this study have been described (Hui *et al.*, 1986). The pH2IIa and pH2III probes were from M. Steinmetz. The 640 bp *Bam*HI 5′-flanking probe was from R. A. Flavell.

DNA sequence analysis

Nucleotide sequences were obtained by both the M13 dideoxy chain termination method (Sanger *et al.*, 1977) and the chemical degradation method (Maxam & Gilbert, 1977). The dideoxy-sequencing kit was purchased either from Amersham or United States Biochemical Corporation (Sequenase). Nucleotide sequences were subsequently processed and analysed by the DNASTAR microcomputer software package (DNASTAR Inc., Europe).

DNA-mediated gene transfer

Exogenous DNA (50 μg) was introduced into human cells by electroporation (Potter, 1988). For K562 cells this was achieved with a Progenetor (Hoeffer) and the delivery of two shocks at 300 V, 80 msecs in 279 mM sucrose and 1 mM $MgCl_2$. For HeLa cells, electroporation was performed with the Gene-Pulser (BioRad) by shocking the cells once at 240 V, 960 μFD in phosphate-buffered saline. The cell concentration employed was 1×10^7/ml for all conditions. Total mRNA was subsequently harvested after 2 days and used for Northern blots. The α-actin probe was obtained from M. Buckingham (Minty *et al.*, 1981).

Biochemical analysis of the H-2K1^k molecules

Two dimensional SDS-PAGE was performed as previously described (Jones, 1977) using 12·5% acrylamide gels. Immunoprecipitation of ^{35}S-labelled cell lysates was carried out with the monoclonal antibody (MoAb) 100-30 which recognizes a public antigenic determinant of the k haplotype class I MHC molecules (kindly provided by Dr B. A. Askonas). Estimates of molecular weight were obtained by plotting relative migration (Rf values) against log M_r using high molecular weight protein standards (cat. No. 6001LA, Bethesda Research Laboratories) as comparison.

RESULTS

Isolation and linkage of cosmid clones carrying the H-2K^k and H-2K1^k genes

Cosmid clones containing MHC class I sequences were selected from two cosmid libraries, derived separately from genomic DNA of the C3Hf/HeN mouse and from the K36.16 tumour cells. Clones were first selected by virtue of their hybridization to the 640 bp *Bam*HI 5′-flanking DNA probe derived from the *H-2K*^{b} gene (Flavell *et al.*, 1985). The selected cosmids were subsequently screened with both the pH2III and pH2IIa DNA probes (cDNA sequences derived from the 5′- and 3′-region respectively of an H-2 mRNA, Steinmetz and colleagues (1981a)). Cosmid clones which hybridized to all three H-2 probes were then mapped by restriction endonucleases. From these clones, we previously sequenced and characterized a cosmid clone, C3Hf-8, which contains the entire *H-2K*^{k} gene of the C3Hf/HeN mouse (Minamide *et al.*, 1988). To identify the *H-2K1*^{k} gene which maps 3′ to the *H-2K*^{k} gene (Fig. 1), we have isolated a 2·7 kb *Sal*I-*Eco*RI fragment which maps to the 3′ end of clone C3Hf-8 (Fig. 1). The cosmid clone C3Hf-8 contained a 5 kb 3′ *Eco*RI end-fragment which on further digestion with *Sal*I, in order to remove the vector pTCF sequence, generated a 2·7 kb fragment. This 2·7 kb fragment was then used as a probe to screen the pre-selected cosmid clones. Several positive clones were isolated. One of the clones isolated from the C3Hf/HeN cosmid library was designated C3Hf-19. Another clone, isolated from the K36·16 cosmid library, was designated 407-7. The relative positions of the *H-2K*^{k} (clone C3Hf-8) and the *H-2K1*^{k} (clones C3Hf-19 and 407-7) genes on the chromosome and their restriction endonuclease maps are summarized in Fig. 1.

To prove that clones C3Hf-19 and 407-7 are adjacent to the *H-2K*^{k} gene, the 3′ *Eco*RI end-fragment was purified from the two *H-2K*^{k} cosmid clones C3Hf-8 (Minamide *et al.*, 1988) and c27·2 (Arnold *et al.*, 1984). These two purified 3′ *Eco*RI end-fragments (a 2·7 kb fragment for C3Hf-8 as described above, and a 3 kb fragment for c27·2) were used to compare with the 3 kb *Eco*RI 5′ end-fragments of the cosmid clones C3Hf-19 and 407-7. All four purified end-fragments were separated on agarose gels and multiple Southern blots were obtained from the

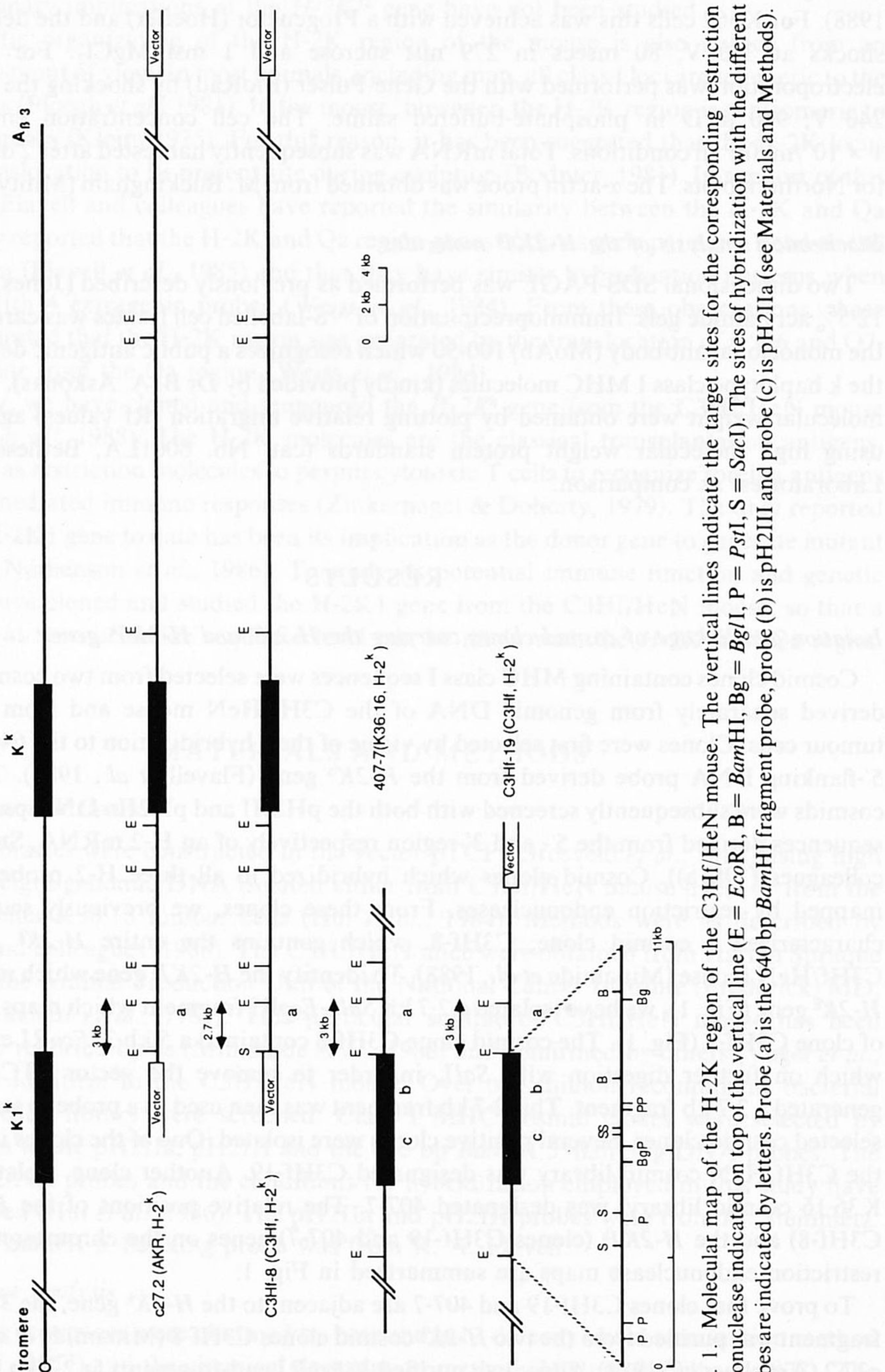


FIG. 1. Molecular map of the H-2K region of the C3Hf/HeN mouse. The vertical lines indicate the target sites for the corresponding restriction endonuclease indicated on top of the vertical line (E = *Eco*RI, B = *Bam*HI, Bg = *Bgl*I, P = *Pst*I, S = *Sac*I). The sites of hybridization with the different probes are indicated by letters. Probe (a) is the 640 bp *Bam*HI fragment probe, probe (b) is pH2III and probe (c) is pH2IIa (see Materials and Methods).

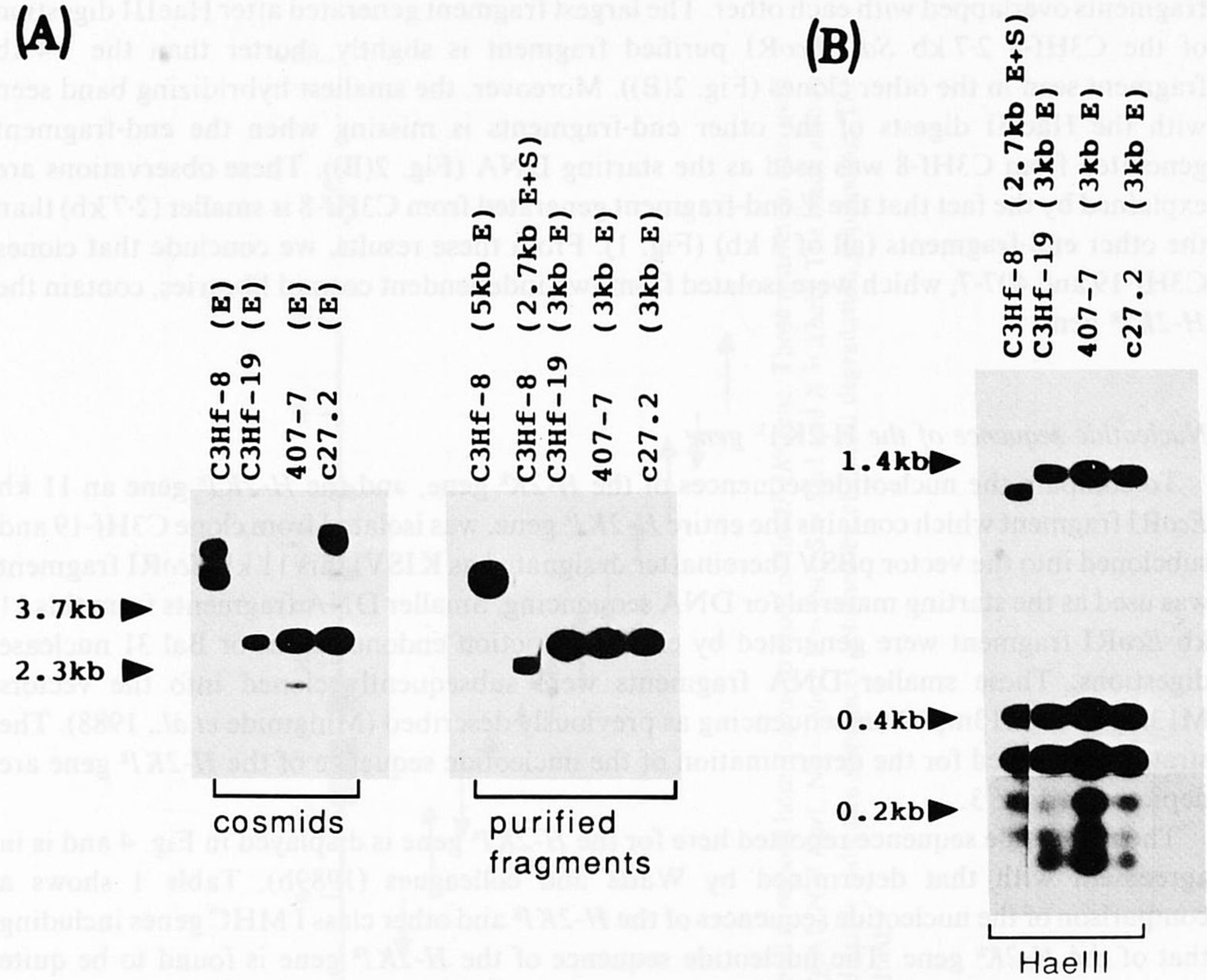


FIG. 2. Linking of *H-2K1^k* to *H-2K^k*. (A) Purified end-fragments of the different cosmid clones (see Materials and Methods) were compared with *Eco*RI digests of the entire clone of the different cosmids. Multiple blots were hybridized separately with each of the different end-fragments. An example autoradiogram, obtained after hybridization to the 3 kb *Eco*RI 5′ end-fragment of clone C3Hf-19, is shown (E = *Eco*RI and S = SalI). (B) The different purified end-fragments obtained were digested with HaeIII and separated on agarose gels. As in (A), multiple blots were hybridized separately with each of the different end-fragments. The autoradiogram obtained after hybridization to the 3 kb *Eco*RI 5′ end-fragment of clone C3Hf-19 is shown. The lower-half of the gel was over-exposed during printing so that all small fragments present can be seen clearly.

same gel. In addition, each of the four purified end-fragments was also used as a DNA probe for hybridization to different set of filters obtained from multiple blots. All patterns obtained after hybridization were identical. A representative pattern obtained with the 3 kb *Eco*RI 5′ end-fragment (isolated from C3Hf-19) as the DNA probe for hybridization is shown in Fig. 2(A). Extra hybridization fragments seen with *Eco*RI digests of clones C3Hf-8 and c27·2 (Fig. 2(A)) are possibly due to incomplete digestion by the restriction endonuclease. Further, these purified fragments were digested with the restriction endonuclease HaeIII. After Southern blotting and hybridization with each of the four purified end-fragments, the HaeIII fragment patterns obtained with the C3Hf-19 3 kb *Eco*RI 5′ end-fragment as the hybridization probe is again shown as an example (Fig. 2(B)). These results confirmed that all the isolated end-

fragments overlapped with each other. The largest fragment generated after HaeIII digestion of the C3Hf-8 2·7 kb *Sal*I-*Eco*RI purified fragment is slightly shorter than the 1·4 kb fragment seen in the other clones (Fig. 2(B)). Moreover, the smallest hybridizing band seen with the HaeIII digests of the other end-fragments is missing when the end-fragment generated from C3Hf-8 was used as the starting DNA (Fig. 2(B)). These observations are explained by the fact that the 3′ end-fragment generated from C3Hf-8 is smaller (2·7 kb) than the other end-fragments (all of 3 kb) (Fig. 1). From these results, we conclude that clones C3Hf-19 and 407-7, which were isolated from two independent cosmid libraries, contain the *H-2K1*^{k} gene.

Nucleotide sequence of the H-2K1^k *gene*

To compare the nucleotide sequences of the *H-2K*^{k} gene, and the *H-2K1*^{k} gene an 11 kb *Eco*RI fragment which contains the entire *H-2K1*^{k} gene, was isolated from clone C3Hf-19 and subcloned into the vector pBSV (hereinafter designated as KISV) this 11 kb *Eco*RI fragment was used as the starting material for DNA sequencing. Smaller DNA fragments from this 11 kb *Eco*RI fragment were generated by either restriction endonucleases or Bal 31 nuclease digestions. These smaller DNA fragments were subsequently cloned into the vectors M13mp18 or M13mp19 for sequencing as previously described (Minamide *et al.*, 1988). The strategy employed for the determination of the nucleotide sequence of the *H-2K1*^{k} gene are depicted in Fig. 3.

The nucleotide sequence reported here for the *H-2K1*^{k} gene is displayed in Fig. 4 and is in agreement with that determined by Watts and colleagues (1989b). Table 1 shows a comparison of the nucleotide sequences of the *H-2K1*^{k} and other class I MHC genes including that of the *H-2K*^{k} gene. The nucleotide sequence of the *H-2K1*^{k} gene is found to be quite similar to that of the *H-2K*^{k} gene (Fig. 4 and Table 1). An insertion of 189 bp, however, was detected in the fifth intron of the *H-2K1*^{k} gene (Fig. 4). A search for homology between this sequence and other reported H-2 gene sequences in the DNA data bank revealed that it is also present in (1) the 5′-flanking region of clone 27·1 (Q7^d) (Steinmetz *et al.*, 1981b) from the Qa region; (2) the 3′-non-coding region of clones T3 and T13 (Pontarotti *et al.*, 1986); (3) the clone 17·3 (Fisher *et al.*, 1985) from the T1a region; (4) the 5′-flanking region of the *H-2K*^{b} gene (Weiss *et al.*, 1983) (Fig. 5(A)). On further analysis of this insertion, we detected a 39 bp DNA sequence which is identical to the reported mouse B1 repeat sequence (Kalb *et al.*, 1983; Propst & Vande Woude, 1984) (Fig. 5(A), bold-types). We subsequently cloned the 189 bp insertion and used it as a probe to screen *H-2*^{b} cosmid clones (obtained from A. Mellor). The insert sequence clearly hybridized to most of the Qa region genes (Fig. 5(B)). However, only three TL region genes (namely T3, T12 and T14) hybridized to this fragment (Fig. 5(B)). This suggests that the B1 repeat sequence is present more frequently in the Qa region than in the T1a region of the b haplotype mouse. This concurs with the observation by Ronne and colleagues (1985), who concluded from both sequence homology and similar insertion patterns that the H-2 genes are more related to the Qa genes than to the TL genes.

Other noticeable features of the nucleotide sequence of the *H-2K1*^{k} gene are the absence of a TATAA consensus sequence and the loss of multiple GpC dinucleotides in the 5′-flanking regions (Fig. 4). These observations suggest that the *H-2K1*^{k} gene may not be functional. However, the absence of a TATAA box has been reported in many constitutively expressed housekeeping and other genes (Bird, 1986). These genes generally contain CpG islands that are completely unmethylated both in the germ line and in all somatic cells (Bird, 1986). Point

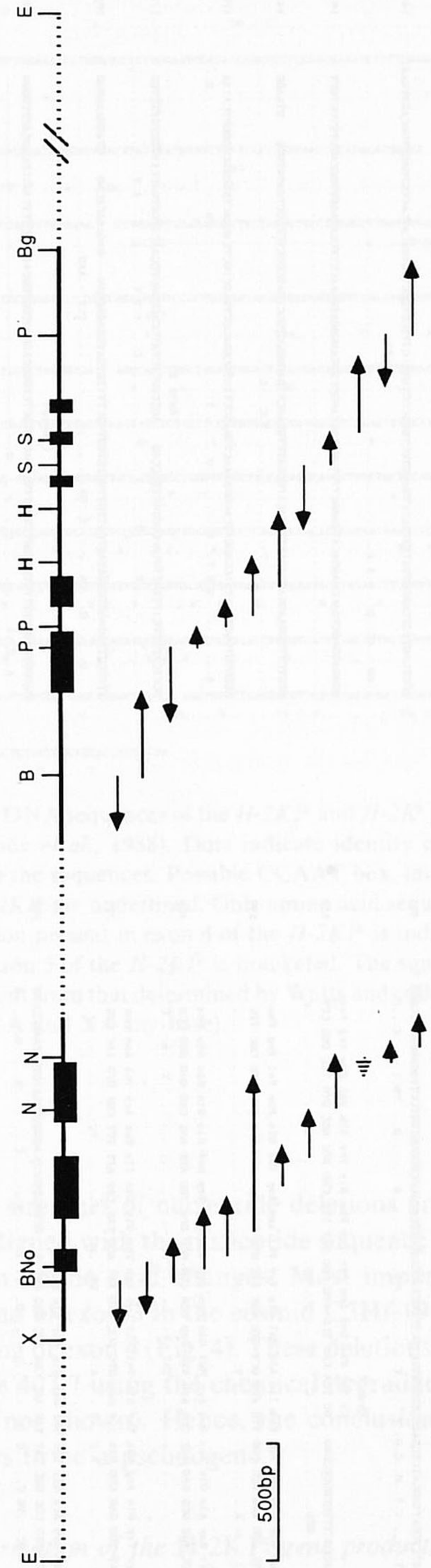


FIG. 3. Sequencing strategy for the *H-2K1^k* gene. Dark boxes correspond to regions homologous to exons of the *H-2K^k* gene. Those restriction sites used in sequencing are shown (B = *Bam*HI, Bg = *Bgl*I, E = *Eco*RI, H = *Hgi*AI, N = *Nar*I, No = *Not*I, P = *Pst*I, S = *Sac*I and X = *Xba*I). The black arrows indicate sequences obtained by dideoxy-sequencing and the hatched arrow indicates sequences obtained by chemical degradation sequencing. The dotted lines indicate DNA sequences which have not been determined.

```
H-2K^k (C3Hf/HeN)      GGTCAGGGCTGGGGAAGCCCAGGGCTGGGGATTCCCCATCTGGTCAGTTTCACTTCTGCTCCTAAC      66
H-2K^k (C3Hf/HeN)      ........G.............................CCA.....................

CTGGGTCAGGCGCTTCTGTCAGGACACTGATGACACGCCTACAGGTCTCACACCCAATGGGTGGCGCGTTCCTGGGGAACCAACCA      152
..........TC..C.....C.........T....G...AGT...C...T..C....T............A..ACCAA.......T..

                                                                          Exon 1
                                                                          Met Ala
                                                                            (
GCTTCGC---GGA----GGC---AAAGTCCACACAGCTAGCAGAACTCAGATGCCCCGGATCCCAGATAGGATTGG   ATG GCT      224
.TG....CGC...CGCT...TAT.........G....CC..G.........A.T.T..A...G.C..CC..TGC-.    ... ..A

Trp Arg Thr Val Leu Leu Leu Leu Ala Ala Ala Leu Ala Pro Thr Gln Thr Cys Ala G

TGG CGC ACA GTG CTC CTG CTG CTG GCG GCC GCC CTG GCC CCG ACT CAG ACC TGT GCG G   GTGAGTG      289
CCC T.. .TG C.. ... ... ... T.. ... ... ... ... ... ... ... ... ... C.C ... .   ......A

CGGGGTCGGGAGGGAAACAGCCTCTGTGGGGAGGGGCGGGGC[-]AACACAGGGGTAGCCTCGTCCCTGTGTGGCTTATCTGATCCTTC      375
                                           [C]
.C..AC............G.......A.............---   --...C....A....G.....TC.G..C..CC.C.G..C...C.

ACCCGATCTCCATCAGCATCTCCAGCCCCGCGTCCTGCTCCCCTCCCAGCCCGCGCACCCGCCCGTGGTCCTGGGAGGAAGTCGGG      461
G...CT......C.T..G..C.G..T......C............T..........G.......G....TG.T.....G...C..

                Exon 2
                Ly Ser His Ser Leu Arg Tyr Phe His Thr Ala Val Ser Arg Pro
GTCTCACCGCTTTCCGCCCCCAG   GC TCA CAC TCG CTG CGG TAT TTC CAC ACC GCC GTG TCC CGG CCT      528
..........GCG.........    .. C.. ..T ... ... A.. ... ... ... ... ... ... ... ... ..C

Gly Leu Gly Glu Pro Arg Phe Ile Ala Val Gly Tyr Val Asp Asp Thr Gln Phe Val Arg Phe
GGT CTC GGG GAG CCC CGG TTC ATC GCT GTC GGC TAC GTG GAC GAC ACG CAG TTC GTG CGC TTC      591
..C ... ... A.. ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ...

Asp Ser Asp Ala Glu Asn Pro Arg Met Glu Pro Arg Ala Arg Trp Met Glu Gln Glu Gly Pro
GAC AGC GAC GCG GAG AAT CCG AGG ATG GAG CCG CGG GCG CGG TGG ATG GAG CAG GAG GGG CCG      654
... ... ... ... ... ... ... ... ... TAT ... ... ... ... .T. ... ... ... ... ... .T. .A. ..C

Glu Tyr Trp Glu Arg Glu Thr Gln Val Ala Lys Asn Thr Glu Gln Ile Ser Arg Val Asn Leu
GAG TAT TGG GAG CGG GAG ACA CAG GTC GCC AAG AAC ACG GAG CAG ATT TCT CGA GTG AAC CTG      717
... ... ... ... ... A.C ..G ... A.. ... ... GG. .AT ... ... ... .TC ... ... ... ...

Arg Thr Leu Arg Gly Tyr Tyr Asn Gln Ser Glu Gly G
AGG ACC CTG CGC GGC TAT TAC AAC CAG AGC GAG GGC G   GTGGGTGACCCAGGGTCGGATGTCACAACCCCTC      788
... ... GC. .TG C.. ..C ... ... ... ... .C. ... .   ...A.......C........G.....G......A

CATGCCCGGACAAGGAGAC-CTGAGATGGGAAGTTCCCGGGCCCCAAGTCCGAGGTTCGGGAGCAGAACCACAACCCAGCACTGGT      873
..CTT..C....CA.G...G....CG--------.......T.....................A..A...GGACC.GGGA.CGGTT.

CCCTTTCAGTTTGTAGGTGTCCGCTAGTGGGCGG[-]TCCGGCAGAACACGGGAGGGGGGTCAGGGGGTGCTGACCGCGGGGTCCCGCA      959
                                  [G]
........C....G...A......GG.C......   ---...------.....C...T.AGC....---.....................

    Exon 3
    Ly Ser His Thr Tyr Gln Arg Met Ser Gly Cys Asp Val Gly Ser Asp Gly Arg Leu Leu Arg
G   GC TCT CAC ACT TAC CAG CGG ATG TCT GGC TGT GAT GTG GGG TCC GAC GGG CGC CTC CTC CGT      1022
.   .. ... ... ..G .T. ..A ... ... .AC ... ... ..G ... ... ..G ... T.. ... ... ... ..C
```

```
Gly Tyr Trp Gln Tyr Ala Tyr Asp Gly Gln Asp Tyr Ile Ala Leu Asn Glu Asp Leu Lys Thr Trp
GGG TAC TGG CAG TAT GCC TAC GAT GGC CAG GAT TAC ATC GCC CTG AAC GAA GAC CTG AAA ACG TGG      1088
... ... GA. ... ..C ..A ... ..C ... TGC ... ... ... ... ... ... ... ... ... ... ... ...

Thr Thr Ala Asp Ile Ala Ala Gln Ile Thr Arg Gln Lys Leu Glu Gln Ala Gly Val Ala Glu Arg
ACG ACC GCA GAC ATA GCA GCT CAG ATT ACC AGA CAA AAA CTG GAA CAG GCT GGT GTT GCA GAA AGA      1154
... G.G ..C ... ..G ..G ..G .T. ..C ... .A. ..C ..G TG. ..G ... ... ... .A. ... ..G ...

  Glu Gly Leu Pro Gly Gly Arg Val His Gly Val Ala Pro Gln Ile Pro Gly Ala Trp Glu Gly
GA--G GGC CTA CCT GGA GGG CGC GTG CAT GGA GTC GCT CCG CAG ATA CCT GGA GCT TGG GAA GGA      1217
..CC. ... ... ... ... ... .A. ... .G. ... ..G ... ... ... ... ... .C. ... C.. ... C.C

Asp Ser   Cys Ala Gln
GAC AGC T--GC GCA CAG   GTGCAGGCG  //  GACACATAGTTCACTGTGATCATTGGTTTAGCTGAGTCTTGTGTAGA      1288
... GCT GCC.. ... ...   .......GC       .................C.......A....A..........G.....

                                                    Exon 4
                                                    Ile Pro Gln Arg Pro Met TRH
TTTCAGTTTGTATTGTTAATTGTGGGATTTCTTAAATCTTCCACACAG  ATC CCC CAA AGG CCC ATG TGA   CCCATC      1363
..........C...........T.....................  ..T ... ... ... ... ... ...   ...G..

ACCCCAGTTCTCAAGGTGATGTCACCCTGAGGTGCTGGGCCCTGGG[C]TTCTACCCTGCTGACATCACCCTGACCTGGCAGTTGAAT      1449
                                               [T]
..AG...AC..G...A.A.A.....................  . ..................................................

GGGGAGGAGCTGACCCAGGACATGGAGCTTGTGGAGACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCATCTGTGGTGGT      1535
..................................................................................................

GCCTCTTGGGAAGGAGCAGAATTACACATGCCATGTGGAACATGAGGGGCTGCCTGAGCCCCTCACCCTGAGATGGG   GTAAGG      1618
.................T.................T.C...C.....................................   ......

AGGGTGTGGGTGCATAACCGGGATCAGGGAAAACTGGAGCCTTCTGCAGACCCTGAGCTGCTCAGGACTGAGAGCTGGGGTCATAA      1704
..A...........G.G.T...G.........G.......T.........................G.................G.

                                          (Exon 5)
CCCTCACCTTCATTTCCTGTACCGGTCCTTCCCAG   AGCCTCCTCTATCCACCGACTCCAACATGGTGACCATTGCTGTTCTGG      1787
.................T......T...........   .........C......T.T........C...A.T..............

TTGTCCTTGGAACTGTGATAGTCATTGTTGCTGTGGTGGCTTTTGTGATGAAGAGGAG---GAGAAACACAG   GTAGGAAAGGG      1867
...........G...CA.......C..GA.........................T...AAG...........   ...........

                                                (Insert
CAGGGTGTGCGTTTTCTCTCAGTCTCCTTTAGAAGTGTGCTCTGGCCAGGCGTGGTGGCACACACCTTTAATCCCAGCACTTGGGA      1953
...A..C..A............C.........G----------------------------------------------------

GGCAGAGACAGGAGGATTTCTGAGTTCAAGGCCAGCCTGGTCTACAAAGTGAGTTCCAGGACAGCCAGGGCTTCAGAGAAACCCTA      2039
--------------------------------------------------------------------------------------

TCTTGAAAAACAAAAAAGCAAACAAACAAACAAACAAACAAAAGAATTGTGC-TCTGCTCATCAATGGGGAACACAGCCACAC      2125
-------------------------------------------------.....G.....................G.....

[-]CG--CATTGCTACT--CTCTAACTGGGTCCTTTGTCAGTTCTGGGAACTCC--AGTGTCAAGATCTTCCTTGAACTCTCACATCTT      2205
[C]
   .CCA.........GT.............TGC.................T.CT.................G...........G...
```

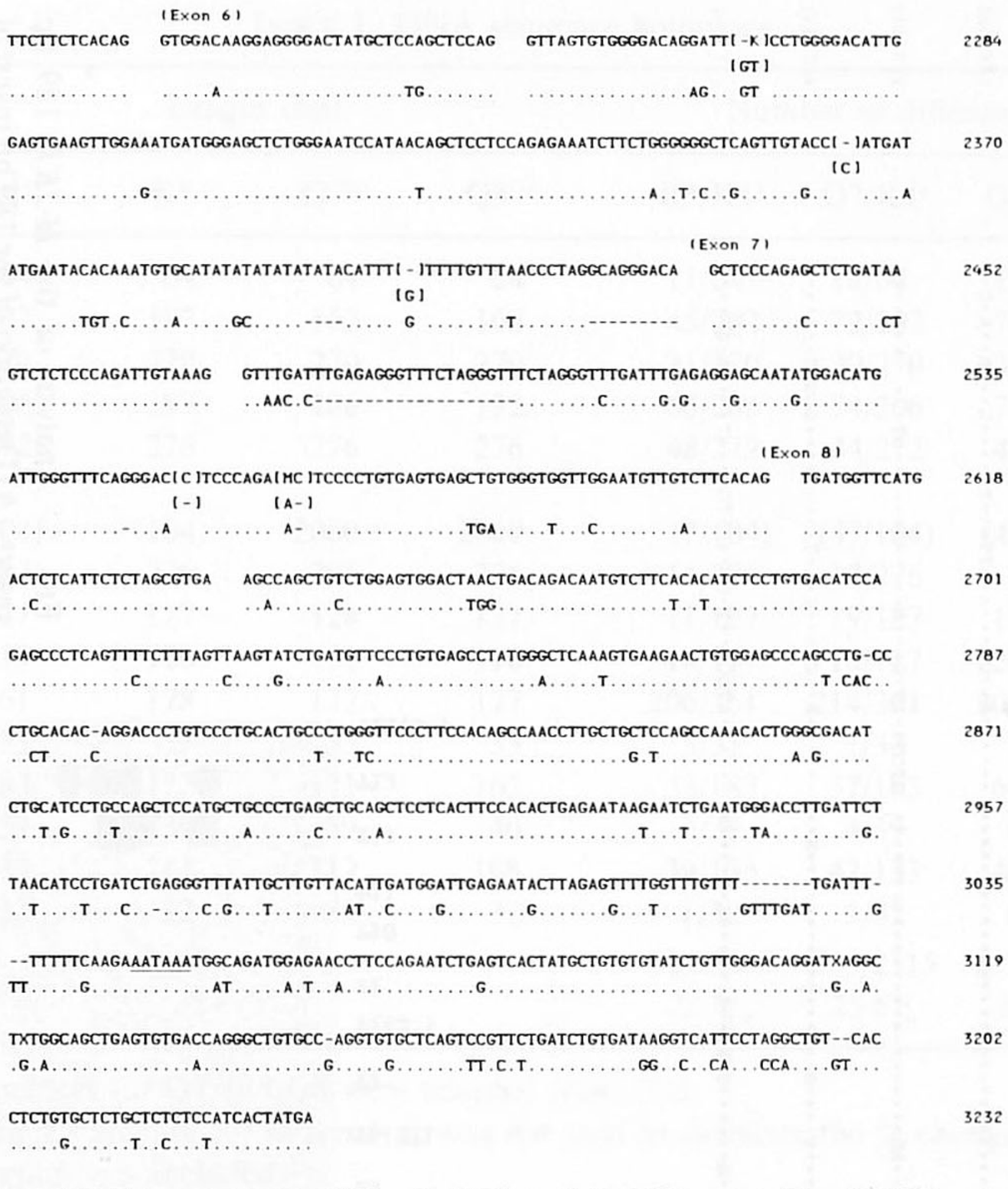


FIG. 4. Aligned DNA sequences of the *H-2K1*^{k} and *H-2K*^{k} genes. The *H-2K*^{k} sequence is from C3H-8 (Minamide *et al.*, 1988). Dots indicate identity of DNA sequence and dashes are inserted to align the sequences. Possible CCAAT box, initiation codon and polyadenylation signal of the *H-2K1*^{k} are underlined. Only amino acid sequence for the *H-2K1*^{k} is shown. The termination codon present in exon 4 of the *H-2K1*^{k} is indicated in bold-types and the insert sequences in intron 5 of the *H-2K1*^{k} is bracketed. The square brackets, [], enclose sequences which are different from that determined by Watts and colleagues (Watts *et al.*, 1989b). (K = T or G, M = C or A and X = any base).

mutations and short stretches of nucleotide deletions are also observed in the exons of the *H-2K1*^{k} gene when aligned with the nucleotide sequence of the *H-2K*^{k} gene (Fig. 4). In some cases, these result in amino acid changes. Most importantly, two deletions, each of two nucleotides, at the end of exon 3 in the cosmid C3Hf-19 results in a translation termination codon at the beginning of exon 4 (Fig. 4). These deletions are also confirmed as present in the *H-2K1*^{k} cosmid clone 407-7 using the chemical degradation sequencing method (Maxam & Gilbert, 1977; data not shown). Hence, the conclusion from sequencing data is that the *H-2K1*^{k} gene appears to be a pseudogene.

Biochemical characterization of the H-2K1^k *gene product*

To determine whether the cloned *H-2K1*^{k} gene is functional *in vitro*, DNA transfection and expression studies were performed with the construct K1SV which was subcloned from the

(A)

	(intron 5) B1 repeats	
H-2K1^k	CTTTAATCCCAGCAC-TTGGGAGGCAGAGACAGGAGGATTT--CTGAGTTCA **AGGCCAGCCTGGTCTACAAAGTGAGTTCCAGGACAGCCA** -GGGCTTCAGAGAGAAACCCTA-TCTTGAAA--AACAAA	
	(3' non-coding)	
Clone 17.3^c	·A·····T·····-·····G···CA·····G ••••••••••••••••••••••••••••••••••• -·····A··C·····TT·····G·····A···	(90.2%)
	(3' non-coding)	
T13^c	·A·····T·····A·····G···CA·····G ••••••••••••••••••••••••••••••••••• -·····A··C·····TT···T·····G··AA··A·C·	(86.5%)
	(5' flanking)	
Clone 27.1	·····-·····G···T··C··T·····G ••••••••••AG••••••••••••••••• GTCT··AT·C···A·····CT·····A···	(85.6%)
	(3' non-coding)	
T3^b	·A·····T·····A·····G···CA··G·····G ···G·····T···G·····A·· -·····A··C···C···T·····GGGGG··A···	(82.5%)
	(5' flanking)	
H-2K^b	···A·AC-A·····-··A···T···A··G···AT·T··C·CTG····· ···G·····A·····T·A···A··A···A·T····· -·····GTGC···T·G·-··TG···-T···CA·G····	(71.8%)

(B)

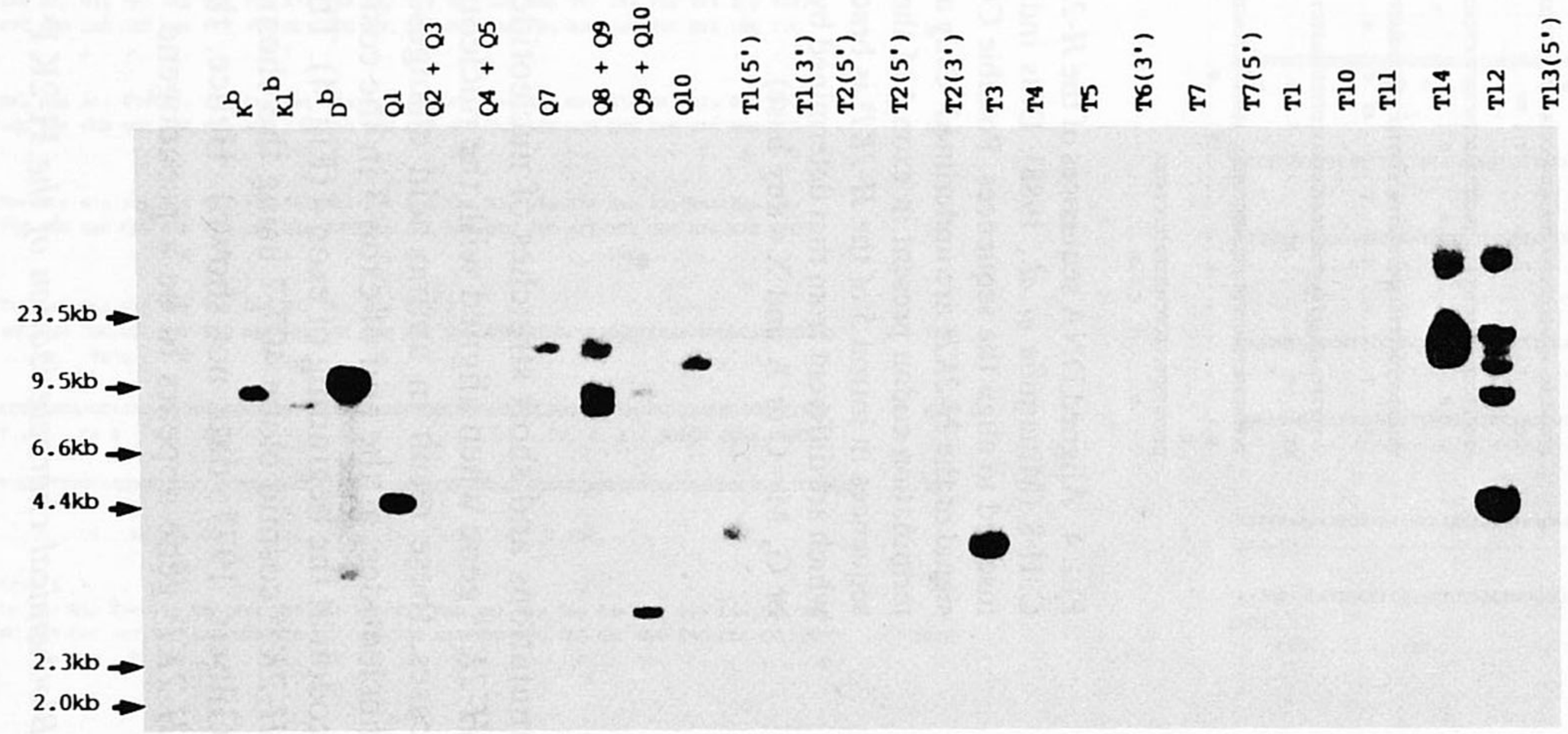


FIG. 5. Analysis of the *H-2K1*^{k} 189 bp insert. (A) Detection of the 189 bp insert of the *H-2K1*^{k} in other MHC class I genes. The letters in bold-type indicate the B1 repeat sequence within the 189 bp insert (B) Screening of cosmid clones containing class I MHC genes isolated from a H-2^b cosmid library with the 189 bp insert.

TABLE 1. DNA sequence homology

	Length (bp)				Number of different nucleotides			
	K1^k	K^k	Q7[(a)]	Q8[(a)]	K^k/K1^k	Q7/K1^k	Q8/K1^k	Q8/Q7*
exon 1	64	64	64	64	11/64	18/64	18/64	0/64
intron 1	202	197	163	162	45/202	72/202	74/202	1/163
exon 2	270	270	270	270	31/270	30/270	33/270	16/270
intron 2	206	187	186	192	63/206	74/206	71/206	10/186
exon 3	272	276	276	276	48/272	44/272	45/272	14/276
intron 3 (partial)†	(104)	(104)	2060	2060	(7/104)	(47/104)	(40/104)	11/2060
exon 4	287	276	285	276	14/276	17/276	9/276	11/285
intron 4	127	127	126	127	11/127	19/127	12/127	13/126
exon 5	117	120	117	116	16/117	16/117	20/117	14/117
intron 5‡	361	178	172	177	206/361	214/361	208/361	25/172
exon 6	33	33	33	33	3/33	1/33	2/33	3/33
intron 6	183	172	171	163	33/183	37/183	66/183	56/171
exon 7	39	39	39	40	3/39	4/39	4/39	7/39
intron 7	133	111	112	108	39/133	42/133	49/133	19/112
exon 8	32	32	32	32	1/32	5/32	6/32	6/32
changes					526/2315	593/2315	617/2315	206/2046
% changes					22·7%	25·6%	26·6%	10%

* Nucleotide sequences for Q7 and Q8 were adapted from (33).

† Due to incomplete sequence for intron 3, it was not used to calculate the % changes.

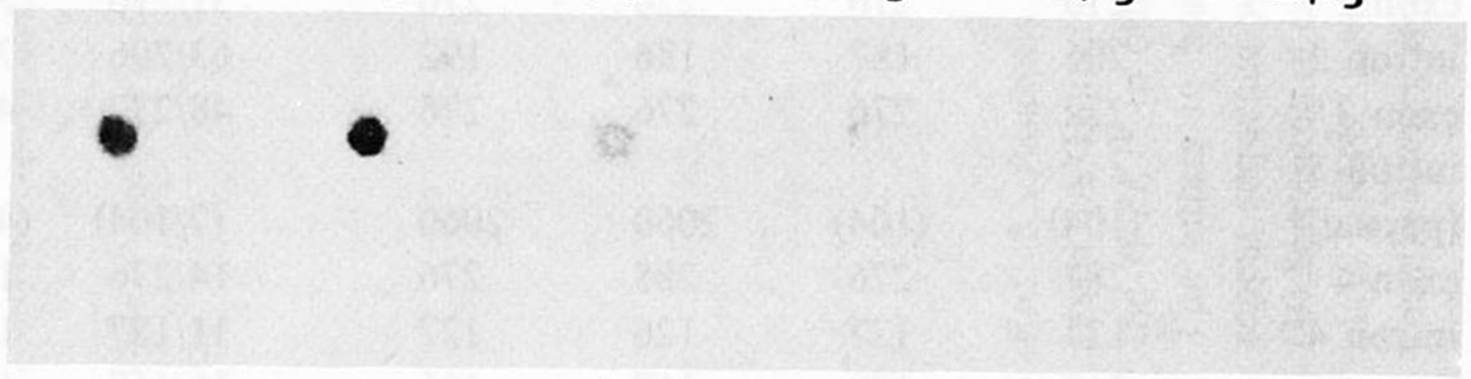
‡The 189bp insertion was included.

cosmid C3Hf-19. Cosmid DNA was introduced into human cells by electroporation as described above. Total cellular mRNA was isolated, blotted onto nitrocellulose membranes and hybridized to the pH2IIa cDNA probe. When compared with total mRNA isolated from untransfected HeLa cells, the mRNA isolated from K1SV-transfected HeLa cells produced a stronger signal, by a factor of at least 20-fold, after hybridization with the pH2IIa DNA probe (Fig. 6(A)). These results were further confirmed by Northern blots using mRNA of K1SV-transfected K562 cells and the pH2IIa probe. The K1SV-transfected K562 cells gave a normal 1·7 kb H-2 class I mRNA transcript which was identical to that obtained from mRNA isolated either from normal AKR mouse livers or *H-2K^k*-(clone c27·2-)transfected K562 cells (Fig. 6(B)). This 1·7 kb H-2 transcript was not present in K562 cells (Fig. 6(B)). Hence, it appears that the *H-2K1^k* could give a complete, full-length H-2 transcript *in vitro*.

The gene product encoded by the *H-2K1^k* gene was also studied by DNA transfection experiments with K1SV. When introduced into human K562 cells, K1SV consistently gave a product of M_r 36,000 on immunoprecipitation with the monoclonal antibody 100-30 (Fig. 7(A)(ii), arrow). This product is not present in untransfected K562 cells (Fig. 7(A)(i)) and is not present on the cell surface (possibly due to the lack of association with the β2-microglobulin) (Fig. 7(A)(iii)). More importantly, this M_r 36,000 product was not competed out by excess untransfected, unlabelled K562 human cell lysate (Fig. 7(A)(iv), arrow). Hence, as would have been predicted from the nucleotide sequence, the *H-2K1^k* gene gave a truncated, cytoplasmic polypeptide of M_r 36,000 due to the presence of a translation

(A) Dot blot (with pH2IIa DNA probe)

(i) Hela mRNA



(ii) KlSV - Hela mRNA

(B) Northern

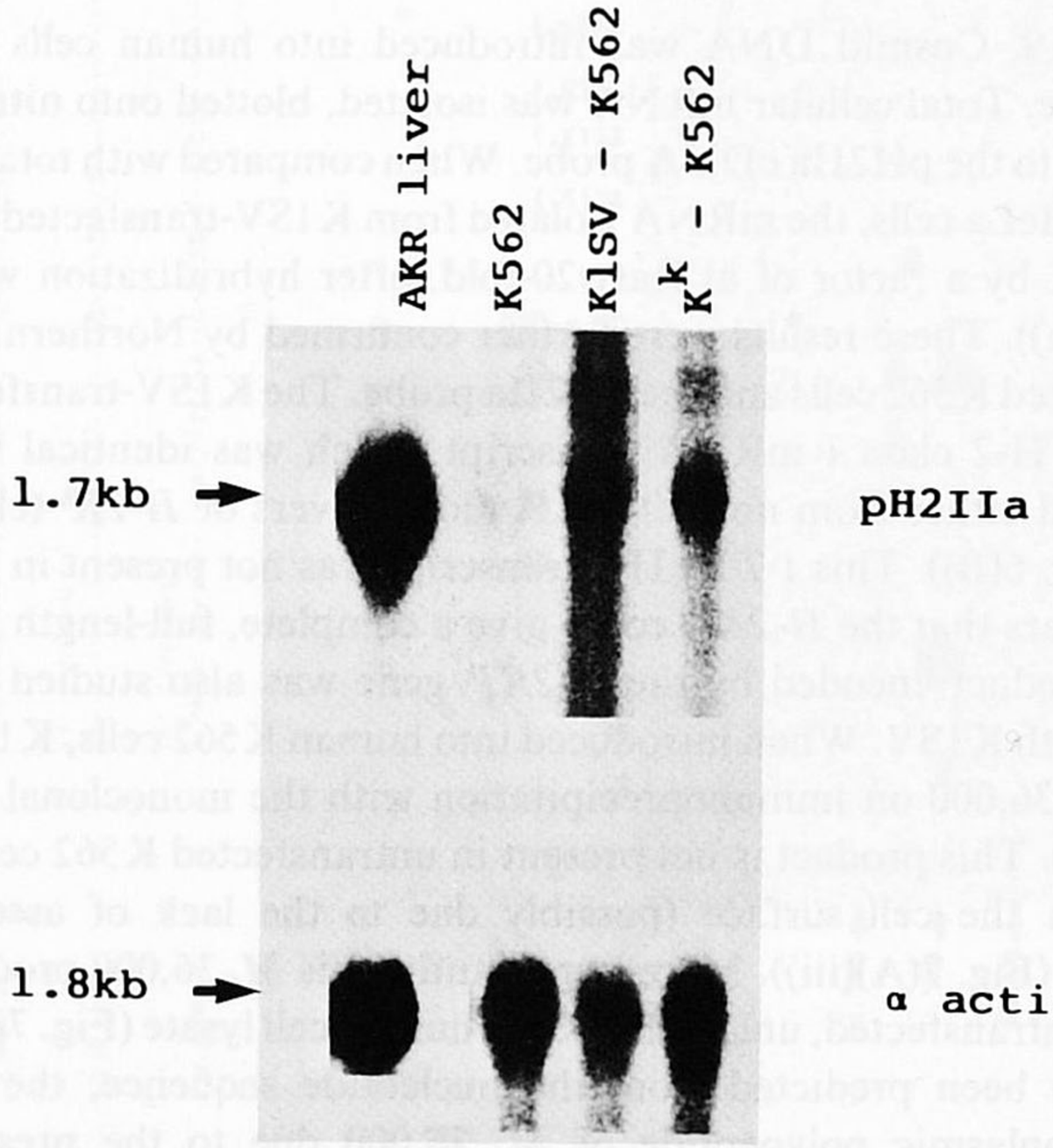


TABLE 2. Comparison of the number of CpG, TpG and CpA occurrences in the H-2K1^{k}and H-2K^k genes of the C3Hf/HeN mouse.

Dinucleotide	H-2K1^k	H-2K^k
CpG	87*	129
GpC	189	181
TpG	271	261
GpT	164	177
CpA	240	193
ApC	168	162

* The figure shown is the actual number of occurrences for each of the corresponding dinucleotides. For each dinucleotide sequence under examination, the group having the reverse DNA sequence is also counted and used as an internal control to ensure random distribution. The 189 bp insert in intron 5 of the H-2K1^k gene was not included in the comparison.

termination codon in exon 4 (Fig. 4). The presence of this polypeptide was also confirmed in K1SV-transfected Hela cells (Fig. 7(B)(ii), arrow).

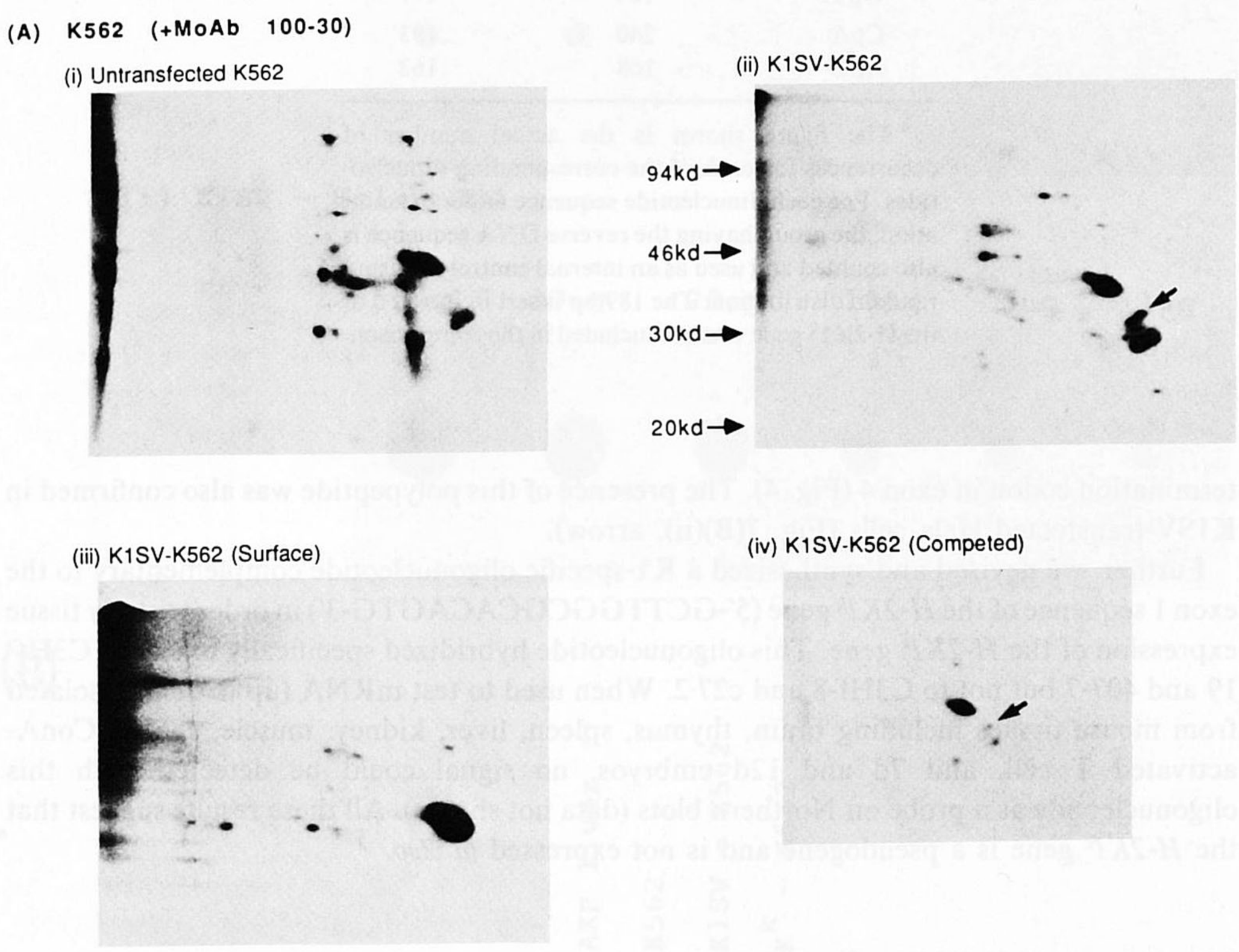
Further, we devised and synthesized a K1-specific oligonucleotide complementary to the exon 1 sequence of the *H-2K1^k* gene (5′-GCTTGGCGCACAGTG-3′) in order to study tissue expression of the *H-2K1^k* gene. This oligonucleotide hybridized specifically to clones C3Hf-19 and 407-7 but not to C3Hf-8 and c27·2. When used to test mRNA (up to 50 μg) isolated from mouse tissues including brain, thymus, spleen, liver, kidney, muscle, sperm, ConA-activated T cell, and 7d and 12d embryos, no signal could be detected with this oligonucleotide as a probe on Northern blots (data not shown). All these results suggest that the *H-2K1^k* gene is a pseudogene and is not expressed *in vivo*.

Mutations in the H-2K1^k *gene are not random*

A closer examination of the nucleotide sequence of the *H-2K1^k* gene of the C3Hf/HeN mouse and comparison with that of the *H-2K^k* gene, revealed that the mutations in the *H-2K1^k* gene do not appear to be random substitutions. Many of the CpG dinucleotides in the *H-2K^k* gene have been converted to TpG or CpA in the *H-2K1^k* gene. As a consequence, the *H-2K1^k* gene has relatively small number of detectable CpGs compared with the *H-2K^k* gene

FIG. 6. Transcription of the *H-2K1^k* gene *in vitro*. (A) Comparative dot blots of untransfected and K1SV-transfected human He1a cells. Different concentrations of purified mRNA were spotted onto nitrocellulose membranes and hybridized to the cDNA probe pH2IIa. (B) Northern blot comparison of mRNA isolated from K1SV-transfected and untransfected human K562 cells with mRNA of normal AKR liver and K^k-transfected K562 cells. The filter was hybridized first with pH2IIa, and subsequently with the α-actin probe.

(Table 2). This observation could not be explained solely by the differences in nucleotide sequences between the *H-2K1*^{k} and the *H-2K*^{k} genes. This is concluded from the fact that, when the number of occurrences of GpC, GpT and ApC are used as internal controls, their frequencies in the *H-2K1*^{k} and *H-2K*^{k} genes are fairly constant (Table 2). These results suggest that the CpGs in the *H-2K1*^{k} gene might have been heavily methylated and that these methylated CpGs were subsequently mutated during DNA replication as a result of inappropriate base pairing.



DISCUSSION

The H-2K region of the mouse is unique with regard to its position within the MHC. The mouse is the only species studied that has two class I genes centromeric to the class II region. The significance of this is unknown. It has been suggested that the H-2K region arose by a translocation to its present site (Bodmer, 1981). However, genetic evidence of this translocation was not reported. It has been reported that in the Qa region of the B10 mouse, alternating gene regions are more similar to one another than to adjacent gene regions (Weiss *et al.*, 1984; Devlin *et al.*, 1985). Similarly, on comparing the nucleotide sequences of the *H-2K1*^{k} gene and the *H-2K*^{k} gene (Fig. 4), we have shown that the overall homology between the different *H-2K*^{k} alleles is stronger (over 99%, Minamide *et al.*, 1988) than that between the *H-2K*^{k} and *H-2K1*^{k} genes (77·3%, Table 1). The fact that adjacent genes are less homologous, as reported for the Qa region genes (Flavell *et al.*, 1985), suggests that the two genes present in the H-2K region arose by the translocation of a gene pair, rather than the translocation of a single precursor gene followed by gene duplication. Devlin *et al.* (1985) suggested that *H-2K*^{b} genes are more similar to the Q6 and Q8 genes, whereas the K1^b gene is more similar to the Q7 and Q9 genes. On comparison, the Q7-Q8 gene pair is found to be more homologous to one another (10% changes) than in the K1^k-K^k gene pair (over 20% changes) (Table 1). Therefore, if the K1^k-K^k gene pair originated by translocation from a 'Qa-equivalent' precursor gene pair, the H-2K region could be considered as a recombination 'hot-spot' where subsequent deletion and substitution events occur more frequently than in the Qa region.

The evolutionary significance of the presence of a B1 repeat sequence in the fifth intron of the *H-2K1*^{k} gene is presently unknown (Fig. 4). Recently, these repeat sequences have also been found in the H-2K region of two t-haplotype mice (Uehara *et al.*, 1987). The similarity of the Qa and H-2K region genes is further demonstrated by the observation that when the 189 bp insert element from the *H-2K1*^{k} gene fifth intron was used as a probe, it hybridized to more Qa region genes (50%) than T1a region genes (20%) in the B10 mouse. These repeat sequences have been suggested as responsible for the generation of diversity in the Qa and T1a regions of the mouse genome (Kress *et al.*, 1984; Ronne *et al.*, 1985) and may therefore contribute to the genetic instability of the Qa region. This is supported by the observation that the absolute number of Qa region genes in different mouse strains varies (Eastman O'Neill *et al.*, 1986; Watts *et al.*, 1989a). The postulated function of B1 repeats in the generation of diversity may also be operative in the *H-2K1*^{k} gene.

FIG. 7. Two-dimensional SDS-PAGE analysis of the gene product encoded by *H-2K1*^{k}. (A) Human K562 cells. (i) ^{35}S-methionine labelled K562 cell lysate was prepared as described in Materials and Methods and allowed to react with the MoAb 100-30. The immunoprecipitate obtained was then analysed by two dimensional SDS-PAGE. The figure shows autoradiogram obtained. (ii) K562 cells were transfected with the plasmid K1SV containing the *H-2K1*^{k} gene as described in Materials and Methods. After two days, the transfected cells were labelled with ^{35}S-methionine and cell lysate was prepared and the immunoprecipitate analysed as in (i). The surface-binding experiment, (A) (iii), was performed with K1SV-transfected K562 cells. These transfected cells were labelled with ^{35}S-methionine before being reacted with the MoAb 100-30. After binding to the MoAb, the cells were washed, cell lysate was prepared and the immunoprecipitate analysed as previously described. For the competition experiment, (A) (iv), cell lysate was prepared after mixing both ^{35}S-labelled K1SV-transfected (8×10^6 cells) and unlabelled K562 cells (4×10^7 cells). The immunoprecipitate subsequently obtained was then analysed. (B) (i) and (ii): same as in (A) (i) and (ii), except the human He1a cell line was employed. The M_r 36,000 spot, if present, is indicated by an arrow.

Interactions between the Qa and the H-2K genes have also been well documented in the generation of variant K molecules (Nathenson *et al.*, 1986). Specifically, for the bm1 mouse, the donor gene suggested was Q10 (Mellor *et al.*, 1985): and for the bm6 mouse, the donor gene is homologous with sequences in the Q4 gene (Geliebter *et al.*, 1986). Such genetic evidence suggests that evolutionary relatedness or proximity of location of the K-region and Qa-region genes has allowed such intimate genetic interactions. It is conceivable, therefore, that the *H-2K1*^{k} and *H-2K*^{k} genes in the C3Hf/HeN mouse have arisen as the result of translocation of a gene pair from the Qa region.

The great reduction in the number of CpG residues in the nucleotide sequence of the *H-2K1*^{k} gene (Table 2) is an interesting observation. The class I MHC genes are regarded as examples of housekeeping genes (Bird, 1986), and they are expressed in many different tissues (Hood *et al.*, 1983). The CpG residues within housekeeping genes occur predominantly in functionally crucial positions in DNA sequences and therefore exert a regulatory role (Bird, 1986; Doerfler, 1983). Hence, for the MHC genes, more frequent clusters of CpGs can be found close to the promoter/leader and/or 5′ regions of these gene (Tykocinski & Max, 1984). The presence of un-methylated CpGs is essential for normal gene activity (Bird, 1986; Doerfler, 1983). Over a period of time, 5-methylycytosine tends to deaminate to thymidine, resulting in the depletion of CpGs by conversion of ^{me} CpGs to either TpG or CpA (depending on which DNA strand underwent the mutation). A great reduction in the number of CpGs in the *H-2K1*^{k} gene (Table 2) suggests that internal CpGs might have been methylated leading to gene inactivation. This explanation is also supported by the presence of an excess of TpGs and CpAs in the *H-2K1*^{k} nucleotide sequence (Table 2) and by the detection of methylated cytosine residues in exon 3 of clones C3Hf-19 and 407–7 as shown by chemical sequencing (data not shown). Table 2 shows that a mutation to CpAs is preferred. However, at present we do not understand why the complementary (non-coding) strand should be preferably deaminated.

ACKNOWLEDGMENTS

We thank Drs Andrew Mellor, Elizabeth Weiss and Michael Steinmetz for providing DNA probes and the *H-2*^{b} cosmid clones. We also thank Drs Andrew Mellor and Jeff Bidwell for their suggestions and comments on the manuscript and Mr Kian Jin Jek for his help with the computer search. This research is supported by the Institute of Molecular and Cell Biology, National University of Singapore and by the Medical Research Council, United Kingdom.

REFERENCES

ARNOLD, B., BURGERT, H.-G., ARCHIBALD, A.L. & KVIST, S. (1984) Complete nucleotide sequence of the murine H-2K^k gene. Comparison of three H-2K locus alleles. *Nucleic Acids Research,* **12,** 9473.

BIRD, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature,* **321,** 209.

BODMER, W.F. (1981) HLA structure and function: a contemporary views. *Tissue Antigens,* **17,** 9.

CALLAHAN, G.N., WALKER, L.E. & MARTIN, W.J. (1981) Biochemical comparison of the H-2K antigen isolated from C3HfB/HeN and C3H/HeN mice. *Immunogenetics,* **12,** 561.

DEVLIN, J.J., WEISS, E.H., PAULSON, M. & FLAVELL, R.A. (1985) Duplicated gene pairs and alleles of class I genes in the Qa2 region of the murine major histocompatability complex: a comparison. *EMBO Journal,* **4,** 3203.

DOERFLER, W. (1983) DNA methylation and gene activity. *Annual Review of Biochemistry*, **52,** 93.

EASTMAN O'NEILL, A., REID, K., GARBERI, J.C., KARL, M. & FLAHERTY, L. (1986) Extensive deletions in the Q region of the mouse major histocompatibility complex. *Immunogenetics*, **24,** 368.

FISCHER, D.A., HUNT III, S.W. & HOOD, L. (1985) Structure of a gene encoding a murine thymus leukemia antigen, and organization of Tla genes in the BALB/c mouse. *Journal of Experimental Medicine*, **162,** 528.

FLAVELL, R., ALLEN, H., HUBER, B., WAKE, C. & WIDERA, G. (1985) Organization and expression of the MHC of the C57BL/10 mouse. *Immunology Review*, **84,** 29.

GELIEBTER, J., ZEFF, R.A., SCHULZE, D.H. *et al.* (1986) Interaction between K^b and Q4 gene sequences generates the K^{bm6} mutation. *Molecular and Cellular Biology* **6,** 645.
147, 136.

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

HOOD, L., STEINMETZ, J. & MALISSEN, B. (1983) Genes of the major histocompatibility complex of the mouse. *Annual Review of Immunology*, **1,** 529.

HUI, K., GROSVELD, F. & FESTENSTEIN, H. (1984) Rejection of transplantable AKR leukamia cells following MHC DNA-mediated cells transformation. *Nature*, **311,** 750.

HUI, K., MINAMIDE, L., PRANDONI, N., FESTENSTEIN, H. & GROSVELD, F.G. (1986) Structural variations in the H-2 genes of AKR lymphomas. *Journal of Immunogenetics*, **13,** 117.

JONES, P.P. (1977) Analysis of H-2 and Ia molecules by two-dimensional gel electrophoresis. *Journal of Experimental Medicine*, **146,** 126.

KALB, V.F., GLASSER, S., KING, D. & LINGREL, J.B. (1983) A cluster of repetitive elements within a 700 base pair region in the mouse genome. *Nucleic Acid Research*, **11,** 2177.

KLEIN, J. (1975) *Biology of the Mouse Histocompatibility complex*. Springer-Verlag, Berlin.

KRESS, M., BARRA, Y., SEIDMAN, J.G., KHOURY, G. & JAY, G. (1984) Functional insertion of an Alu Type 2 (B2 S1WE) repetitive sequence in murine class I genes. *Science*, **226,** 974.

MAXAM, A.M. & GILBERT, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences U.S.A.*, **74,** 560.

MELLOR, A.L., WEISS, E.H., RAMACHANDRAN, K. & FLAVELL, R.A. (1983) A potential donor gene for the bm1 gene conversion event in the C57BL mouse. *Nature*, **306,** 792.

MINAMIDE, L.S., CALLAHAN, G.N., GROSVELD, F.G. & HUI, K.M. (1988) The nucleotide sequence of the H-2K gene of C3Hf/HeN mice. *Immunogenetics*, **27,** 148.

MINTY, A., CARAVATTI, M., ROBERT, B. *et al.* (1981) Mouse actin messenger RNAs. Construction and characterization of a recombinant plasmid molecule containing a complementary DNA transcript of Mouse alpha-actin mRNA. *Journal of Biological Chemistry*, **256,** 1008.

NATHESON, S.G., GELIEBTER, J., PFAFFENBACH, G.M. & ZEFF, R.A. (1986) Murine major histocompatibility complex class I mutants: Molecular analysis and structure-function implications. *Annual Review of Immunology*, **4,** 471.

PLOEGH, H.L., ORR, H.T. & STROMINGER, J.L. (1981) Major histocompatibility antigens the human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. *Cell*, **24,** 287.

PONTAROTTI, P.A., MASHIMO, H., ZEFF, R.A. *et al.* (1986) Conservation and diversity in the class I genes of the major histocompatibility complex: sequence analysis of a Tla^b gene and comparison with a Tla^c gene. *Proceedings of the National Academy of Sciences*, **83,** 1782.

POTTER, H. (1988) Electroporation in Biology: methods, applications, and instrumentation. *Analytical Biochemistry*, **174,** 361.

PROPST, F. & VANDE WOUDE, G.F. (1984) A novel transposon-like repeat interrupted by an LTR element occurs in a cluster of B1 repeats in the mouse c-mos locus. *Nucleic Acids Research*, **12,** 8381.

REDDY, E.S.P. & PAN, J.Y. (1985) Molecular cloning and sequencing of H-2K^k cDNA: comparison with other H-2 genes and evidence for alternative splicing. *Gene*, **38,** 239.

RONNE, H., WIDMARK, E., RASK, L. & PETERSON, P.A. (1985) Intron sequences reveal evolutionary relationships among major histocompatibility complex class I genes. *Proceedings of the National Academy of Sciences U.S.A.*, **82,** 586000.

SANGER, F., NICKLEN, S. & COULSON, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences U.S.A.*, **74,** 5463.

STEINMETZ, M., FRELINGER, J.G., FISHER, D. *et al.* (1981a) Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell*, **24,** 125.

STEINMETZ, M., MOORE, K.W., FRELINGER, J.G. *et al.* (1981b) A pseudogene homologous to mouse transplantation antigens: transplantation antigens are encoded by eight exons that correlate with protein domains. *Cell*, **25,** 683.

TYKOCINSKI, M. & MAX, E. (1984) CG dinucleotide clusters in MHC genes and in 5′ demethylated genes. *Nucleic Acids Research*, **12,** 4385.

UEHARA, H., ABE, K., PARK, C.-H.T., SHIN, H.-S., BENNETT, D. & ARTZT, K. (1987) The molecular organization of the H-2K region of two t-haplotypes: implications for the evolution of genetic diversity. *EMBO Journal*, **6,** 83.

VOGEL, J., DAVIS, A., MCKINNEY, D., MCMILLAN, M., MARTIN, J. & GOODENOW, R. (1988) Molecular characterization of the C3HfB/HeN H-2K^{km2} mutation. *Journal of Experimental Medicine*, **168,** 1781.

WATTS, S., DAVIS, A., GRANT, B., WHEELER, C., HILL, L. & GOODENOW, R. (1989a) Organization and structure of the Qa genes of the major histocompatibility complex of the C3H mouse: implications for Qa function and class I evolution. *EMBO Journal*, **3,** 1749.

WATTS, S., DAVIS, A. & GOODENOW, R. (1989b) Sequence analysis of the C3H H-2K1^{k} gene: relationship to the H-2 loci. *Immunogenetics* **29,** 355.

WEISS, E.H., GOLDEN, L. & ZAKUT, R. (1983) The DNA sequence of the H-2K^{b} gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *EMBO Journal*, **2,** 453.

WEISS, E.H., GOLDEN, L., FAHRNER, K. *et al.* (1984) Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse. *Nature*, **310,** 650.

ZINKERNAGEL, R.M. & DOHERTY, P.C. (1979) MHC-restricted cytotoxic T cells: studies on the role of polymorphic major transplantation antigens determining T-cell restriction, specificity, function, and responsiveness. *Advanced Immunology*, **27,** 51.