MOLECULAR AND CELLULAR BIOLOGY, Jan. 1990, p. 138–145
0270-7306/90/010138-08$02.00/0

Vol. 10, No. 1

# A Putative Transmembrane Protein with Histidine-Rich Charge Clusters Encoded in the *H-2K/t*$^{w5}$ Region of Mice

BENOIT ST.-JACQUES, TAE-HEE HAN, ARMAND MACMURRAY, AND HEE-SUP SHIN*

*Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142*

Received 26 July 1989/Accepted 3 October 1989

**The *H-2* complex of mice contains many genes in addition to the gene families involved in immune reactions. Some of them are believed to function in mouse development, as suggested by the findings that several embryonic lethal mutations map within or near the *H-2* complex. We have analyzed the *H-2K/t*$^{w5}$ region in an attempt to study non-*H-2* genes encoded in this region. Overlapping cosmid clones spanning about 170 kilobase pairs of DNA, including the *H-2K/t*$^{w5}$ region of the mouse, have been screened for genes expressed in embryonic carcinoma cells. A transcript of 2.8 kilobase pairs (K. Abe, J.-F. Wei, F.-S. Wei, Y.-C. Hsu, H. Uehara, K. Artzt, and D. Bennett, EMBO J. 7:3441-3449, 1988) encoded by the KE 4 gene flanking *H-2K* distally was identified. The transcript was abundantly expressed in embryonic carcinoma cells but was present at low levels in other tissues in adults. A cDNA for this transcript was isolated from the F9 embryonic carcinoma cell line and sequenced. It potentially encodes a protein of 436 amino acids with several interesting features. First, it contains two regions made of well-conserved repeats unusually rich in histidine residues. In the repeats, histidine alternates with other amino acids, notably glycine or serine. Second, the two histidine-rich regions are separated by three putative membrane-spanning domains. Third, the N-terminal part of the sequence shows characteristics of a signal peptide. The results indicate that the protein coded by the gene may be a transmembrane protein with histidine-rich charge clusters. A similar sequence motif found in other known genes allows speculation on the possible function of this gene.**

The major histocompatibility complex (MHC) includes many genes involved in immune functions. The MHC I and II gene families code for cell surface receptors involved in immune recognition, and the MHC III genes encode components of the complement system. Other genes, such as the genes for tumor necrosis factors α and β (40) and the gene for steroid 21 hydroxylase (2, 46), have also been mapped within the MHC. This chromosomal region is believed to include many other genes that are yet to be detected or characterized. For example, susceptibility to a number of diseases is associated with certain MHC alleles in humans (8, 35). Genes linked to the mouse MHC are implicated in affecting the rate of early mouse development (16). Some of the *H-2*-linked repeated sequences are transcribed during embryogenesis (9, 30). Several embryonic lethal mutations of the *t* complex map within *H-2* (4); primarily, the *t*$^{w5}$ mutation has never been separated from *H-2K* by recombination (3). *t*$^{w5}$ is one of the embryonic recessive lethal mutations in the *t* complex of the mouse. Embryos homozygous for *t*$^{w5}$ develop normally until the elongated egg cylinder stage. At this time, the endoderm becomes morphologically abnormal and the embryonic ectoderm shows signs of pyknosis. Death of the embryos usually occurs within the next 2 days (5).

Identification and characterization of non-MHC genes within the MHC may provide insights into the genetics and evolution of the complex and its relationship with many of the MHC-linked diseases. Furthermore, identification of the gene product of any of the *H-2*-linked genes that are expressed in embryogenesis is likely to contribute to our understanding of mammalian development.

The MHC region has been scrutinized for expressed genes both in humans and in mice. This region is an especially good target for such an approach since substantial portions of the *HLA* and the *H-2* complex have been molecularly cloned. For example, five novel transcripts have been detected in the vicinity of *HLA-B* (39), and a gene containing a reiteration of a dipeptide made of a basic residue next to an acidic residue has been detected within the MHC III region in humans and mice (28). In the *H-2K* region of the mouse, five different transcripts have been identified (1). The embryonically expressed genes in the *H-2K* region would be particularly interesting subjects for a detailed analysis, since they can be potential candidates for *t*$^{w5}$. Therefore, we have isolated and characterized the cDNA for one of the embryonic transcripts within the *H-2K/t*$^{w5}$ region. Analysis of the primary structure of the cDNA reveals that the gene may code for a transmembrane protein with charge clusters consisting of histidine-rich repeats. The possible function of this gene is discussed in light of the sequence motifs found in other known genes.

## MATERIALS AND METHODS

**DNA probes.** Cloned fragments or whole cosmids were labeled by nick translation or by random priming to a specific activity of at least $10^8$ cpm/μg, using [α-$^{32}$P]dCTP as described previously (13, 38). Unincorporated nucleotides were separated from the DNA by passing the mixture through a 0.8-ml spun column of Sepharose CL-6B.

**Probe prehybridization.** To block repetitive regions of radiolabeled probes, hybridization solution containing 7 × $10^6$ cpm of probe, 200 μg of sonicated mouse genomic DNA (average size, 300 base pairs [bp]), and 200 μg of sonicated salmon sperm DNA per ml was incubated at 42°C for 3 h. The hybridization solution was then added to the filter bag and incubated for 40 h at 52°C. Washing consisted of two 20-min 2× SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS) washes at room temperature, three 20-min 0.2× SSC–0.1% SDS washes, the first at room temperature and the second and third at 52°C,

* Corresponding author.

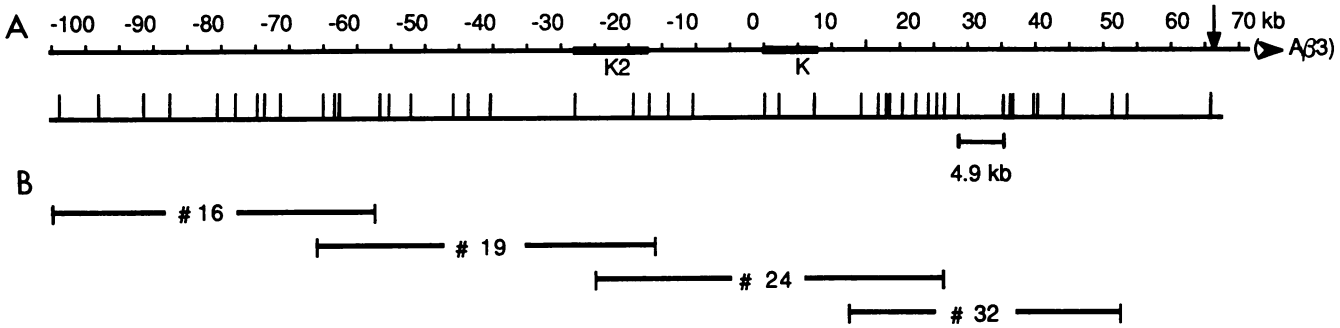


FIG. 1. (A) Molecular map of the *H-2K* region of the $t^{w5}$ haplotype. Locations of the *H-2K* and *K2* MHC I genes are shown as solid boxes at the top. The arrowhead at the right indicates that the MHC II gene, *Aβ3*, is located within the next 10 kb of DNA past the end of the map. Since the MHC is inverted in *t* haplotypes (36), the centromere is on the right side of the map and the telomere is on the left. The arrow at about +65 kb indicates the centromeric limit of the region where the $t^{w5}$ gene should reside (3). Vertical lines indicate sites for the restriction enzyme *Bam*HI. (B) Regions contained in cosmids 16, 19, 24, and 32. The 4.9-kb *Bam*HI fragment of cosmid 32, which was used to screen an EC cell cDNA library, is also indicated.

and one 30-min 0.1× SSC–0.1% SDS wash at 68°C. The washed filters were then exposed with X-ray film for 12 to 48 h at −70°C.

**RNA analysis and cDNA cloning.** Total RNA was isolated from tissues or F9 embryonal carcinoma (EC) cells by lysing in guanidine thiocyanate and centrifugation through 5.7 M cesium chloride by standard procedures (29, 37). Selection of poly(A)$^+$ RNA was done on oligo(dT)-cellulose columns as described by Maniatis et al. (29). For all Northern (RNA) blots, RNAs were electrophoresed in formaldehyde gels containing 1% agarose and transferred to nylon membranes (GeneScreen Plus; Dupont, NEN Research Products). Filters were hybridized, washed, and autoradiographed by standard procedures (29, 37). Double-stranded cDNA was prepared from the poly(A)$^+$ RNA of F9 cells by the method of Gubbler and Hoffman (17) and then cloned into a bacteriophage vector, λgt11 (21). The recombinant DNAs were packaged in vitro into phages (Gigapack; Stratagene), and the phages were plated and lifted onto nylon membranes (GeneScreen Plus) as described previously (38). The filters were autoclaved at 100°C for 1 min, hybridized with a radiolabeled probe, washed, and autoradiographed as described previously (38).

**Sequencing strategy.** The 2.7-kilobase-pair (kb) *Eco*RI insert of clone 10.2a was digested with *Bam*HI to yield four fragments of 1,121, 913, 525, and 241 bp, respectively. Each fragment was subcloned in the pBluescript-KS(+) vector (Stratagene) in both orientations. Deletion subclones were generated by exonuclease III and S1 nuclease digestions (Erase-A-Base system; Promega Biotec). Consecutive deletion subclones generally overlapped by 50 to 100 bp. Cesium chloride-isolated plasmid DNA was used as a template for sequencing by the dideoxy method (34), using $^{35}$S-labeled nucleotides and Sequenase (United States Biochemical Corp.). Every sequence was read from both strands. Reactions were run in 6% polyacrylamide gels, using the Sequigen nucleic acid sequencing cell (Bio-Rad Laboratories). For G+C-rich sequences in which band compression was a problem, sequencing reactions were performed by using dITP or 7-deaza dGTP instead of dGTP. The 2.7-kb *Eco*RI fragment cloned in pBR322 was used as a template to sequence over the *Bam*HI sites. Synthetic oligonucleotides (18-mers) were used as primers.

## RESULTS

**Isolation of cDNAs for embryonic transcripts encoded in the *H-2K*/$t^{w5}$ region.** An overlapping set of cosmid clones spanning about 170 kb of DNA including the *H-2K* region has been previously isolated and characterized (42). The attempt to identify the $t^{w5}$ gene among these cosmid DNAs by a strategy based on restriction fragment length polymorphism between DNA from mutant ($t^{w5}/t^{w5}$) EC cells and from revertant ($t^{w5G}/t^{w5G}$) mouse liver has not yet been informative. The revertant (a gift from K. Artzt) was spontaneously derived from the $t^{w5}$ haplotype and contains the full complement of a complete $t^{w5}$ haplotype except the lethal mutation, as judged by genetic, molecular, and biochemical markers (K. Artzt, personal communication). We used four cosmids spanning the region to probe a genomic Southern blot containing mutant and revertant DNA restricted with seven different enzymes. This probing did not reveal any difference between the mutant and revertant DNAs but ruled out the possibility of a gross insertion or deletion of a transposon between the two DNAs in this region. We have therefore chosen to identify embryonically active genes embedded in this region. Northern blots of poly(A)$^+$ RNA isolated from F9 EC cells were probed directly with whole cosmids, using sonicated mouse genomic DNA to block repetitive sequences in the probes (see Materials and Methods). Cosmids 16, 19, and 32 (Fig. 1) each identified a different transcript, whereas an MHC I probe hybridized to the same RNA did not reveal any transcript (Fig. 2). The transcript in cosmid 32 corresponded to the KE 4 gene of Abe et al. (1) and was unrelated to the transcripts detected by cosmids 16 and 19. At least part of the transcribed sequence in cosmid 32 was localized to a single-copy 4.9-kb *Bam*HI fragment. On a Northern blot, this probe detected two transcripts of about 2.8 and 1.5 kb. The 2.8-kb transcript was expressed abundantly in EC cells and faintly in adult tissues, whereas the 1.5-kb transcript was expressed abundantly in adult kidney cells but faintly in other tissues and EC cells (Fig. 3).

A cDNA library was constructed in the vector λgt11, using mRNA isolated from F9 cells. The library was screened with the 4.9-kb *Bam*HI fragment from cosmid 32, yielding several positive clones. The sizes of the inserts ranged from 800 bp to 2.7 kb, with no clear clustering of sizes. Clone 10.2a contained the largest insert. Its size (2.7 kb) corresponded to that of the larger transcript detected on Northern blots, suggesting that it is a near-full-length cDNA. Three other clones, of 2.2 kb, 2.0 kb, and 860 bp, have been characterized by restriction enzyme digests and partial sequencing; they appear to be shorter cDNAs for the same transcript as 10.2a. Probing of a Northern blot with the 2.7-kb cDNA indicated that the smaller 1.5-kb kidney-specific transcript

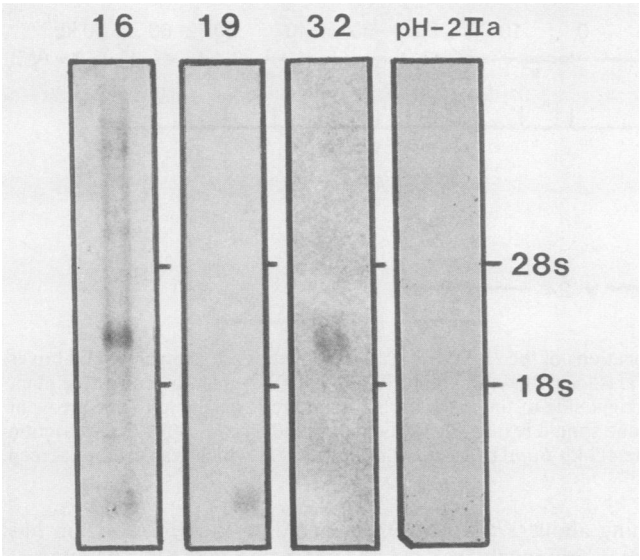


FIG. 2. Northern blots of poly(A)$^+$ RNA isolated from F9 EC cells and probed with whole cosmids from the *H-2K/t*$^{w5}$ region. The probe used in each case is indicated at the top. Cosmids 16, 19, and 32 are depicted in Fig. 1. pH-2IIa is an MHC I probe (38). Positions of 18S and 28S RNAs are indicated at the right.

does not share any homology with this cDNA. Another gene must therefore be at least partially encoded in the 4.9-kb *Bam*HI fragment of cosmid 32. This gene has not been characterized further.

Because embryonic genes in this region are potential candidates for *t*$^{w5}$, the 2.7-kb cDNA was also used as a probe on a blot containing poly(A)$^+$ RNA from wild-type and mutant (*t*$^{w5}$/*t*$^{w5}$) EC cells. No difference between the two cell types was found (data not shown). Finally, in an attempt to detect small differences in the DNA encoding the 2.7-kb transcript between the mutant and the revertant, the cDNA clone was used to probe denaturing gradient gels containing restricted genomic DNA from mutant (*t*$^{w5}$/*t*$^{w5}$) EC cells and from revertant (*t*$^{w5G}$/*t*$^{w5G}$) mouse liver. No difference was detected (data not shown).

**Predicted amino acid sequence suggests that the cDNA codes for a transmembrane protein.** The 2.7-kb insert of clone 10.2a was subcloned in the pBluescript vector and sequenced by the dideoxy method, using plasmid DNA as a template (34). The strategy of nested deletions was used to complete the sequencing (Erase-A-Base system; Promega Biotec). The sequence of clone 10.2a is shown in Fig. 4. Two ATG codons, at positions 1 to 3 and 7 to 9, define the start of an open reading frame (ORF) of about 1.3 kb. The sequence ACCATGG around the second ATG codon is identical to the optimal consensus sequence for translation initiation by eucaryotic ribosomes (26). The region upstream of the initiator codons contains a number of in-frame stop codons and has a high G+C content. The ORF ends with a TAG stop codon at position 1309. A canonical polyadenylation signal (47) is located 22 nucleotides upstream from an 18-nucleotide-long poly(A) tail. The presence of regions of internal homology is evident when the 10.2a sequence is compared with itself in a dot matrix analysis. These sequences that share homology make up almost 30% of the ORF. They code for the histidine-rich regions described below.

The predicted amino acid sequence determined from the 1.3-kb ORF encodes a protein of 436 amino acids with a molecular mass of 47 kilodaltons. The sequence is rich in histidine (14.7%) and glycine (13.5%). Comparison of codon usage of this ORF with that of other eucaryotic proteins (using the CODON FREQUENCY program) indicates that the codons used in clone 10.2a in general are among the frequently used ones. Three cysteines are present, at positions 276, 382, and 427. A single putative N-linked glycosylation site (Asn-Phe-Thr) is present starting at position 337.

The N-terminal part of the ORF possibly encodes a signal sequence. Although there exists no consensus for signal sequences (32), they are generally between 18 and 25 amino acids long and comprise three distinct regions: an N-terminal region (n) variable in length and composition but always carrying a positive net charge, a central hydrophobic region (h) also variable in length but enriched in the amino acids Phe, Ile, Leu, Met, and Val, with very few charged or polar residues, and a more polar C-terminal region (c) that consists of five or six residues and seems to define the cleavage site (45). The N-terminal region of the ORF of the 10.2a sequence fits well with this description. It can be divided into (i) a positively charged n region from Met-1 (or Met-3) to His-9, with a net charge of +2 (one Arg residue plus the amino group of the terminal Met residue), (ii) a more hydrophobic h region from Trp-10 to Ala-26 that contains three Val, five Leu, and no charged residues, and (iii) a more polar c region from Gly-27 to Gly-30 or Gly-32, depending on the location of the cleavage site. The exact cleavage site of this putative signal sequence cannot be clearly predicted, but considering the −3/−1 principle of Von Heijne (43, 44), it is likely to take place either after Gly-30 or after Gly-32.

The predicted amino acid sequence was used to generate a

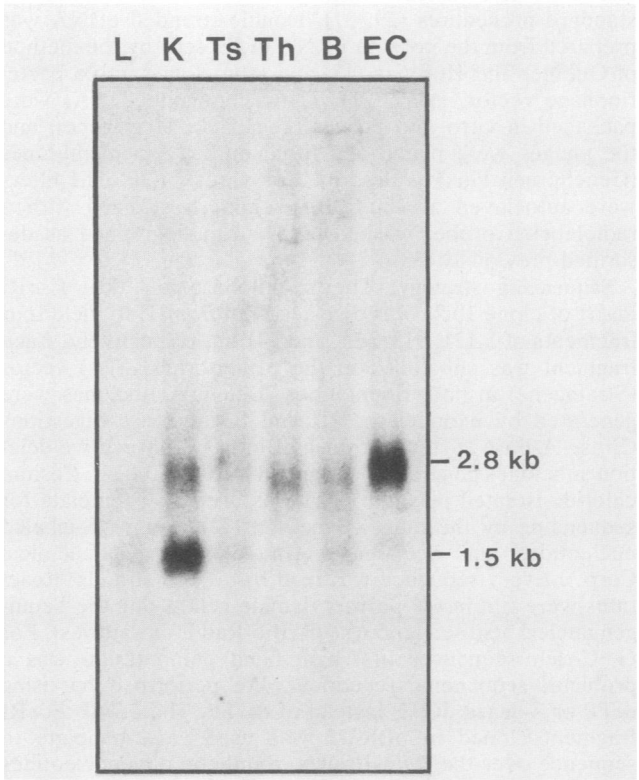


FIG. 3. Tissue distribution of the cosmid 32 transcripts. The 4.9-kb *Bam*HI fragment from cosmid 32 was used to probe poly(A)$^+$ RNA from the sources indicated at the top (L, adult lung; K, kidney; Ts, testis; Th, thymus; B, brain; EC, EC cells). The sizes of the two transcripts are indicated at the right.

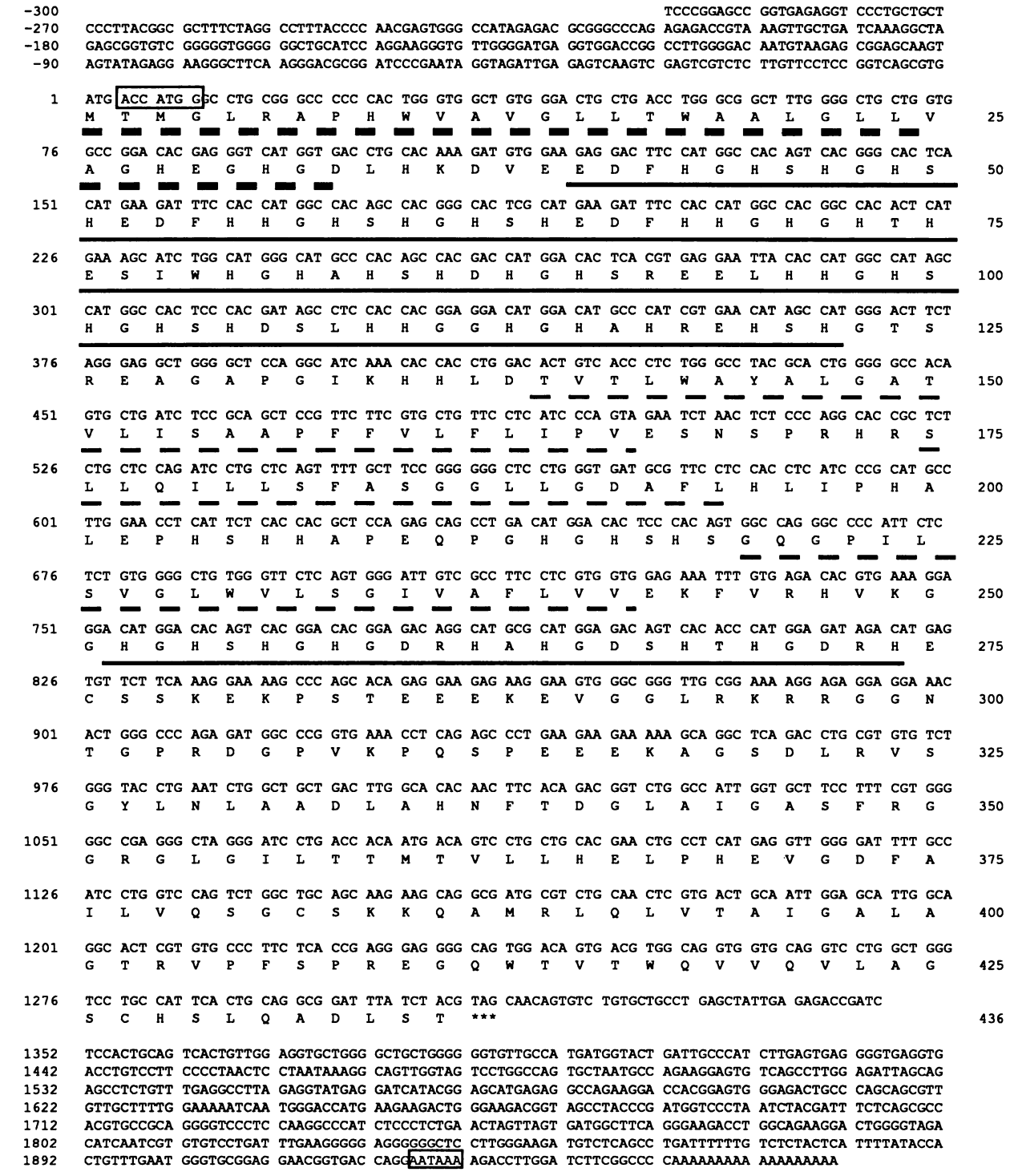


FIG. 4. DNA sequence and deduced amino acid sequence of the cDNA clone 10.2a. The amino acid sequence is shown, in the one-letter code, under the nucleotide sequence. Nucleotides are numbered at the left, and amino acids are numbered at the right. The sequence ACCATGG, identical to the optimal consensus sequence for initiation of translation (26), and the canonical polyadenylation signal AATAAA are boxed. The first in-frame stop codon is indicated by asterisks. Other features of the sequence are underlined: the putative signal sequence (bold dashed line), the two histidine-rich regions (continuous lines), and the three putative transmembrane domains (fine dashed lines).

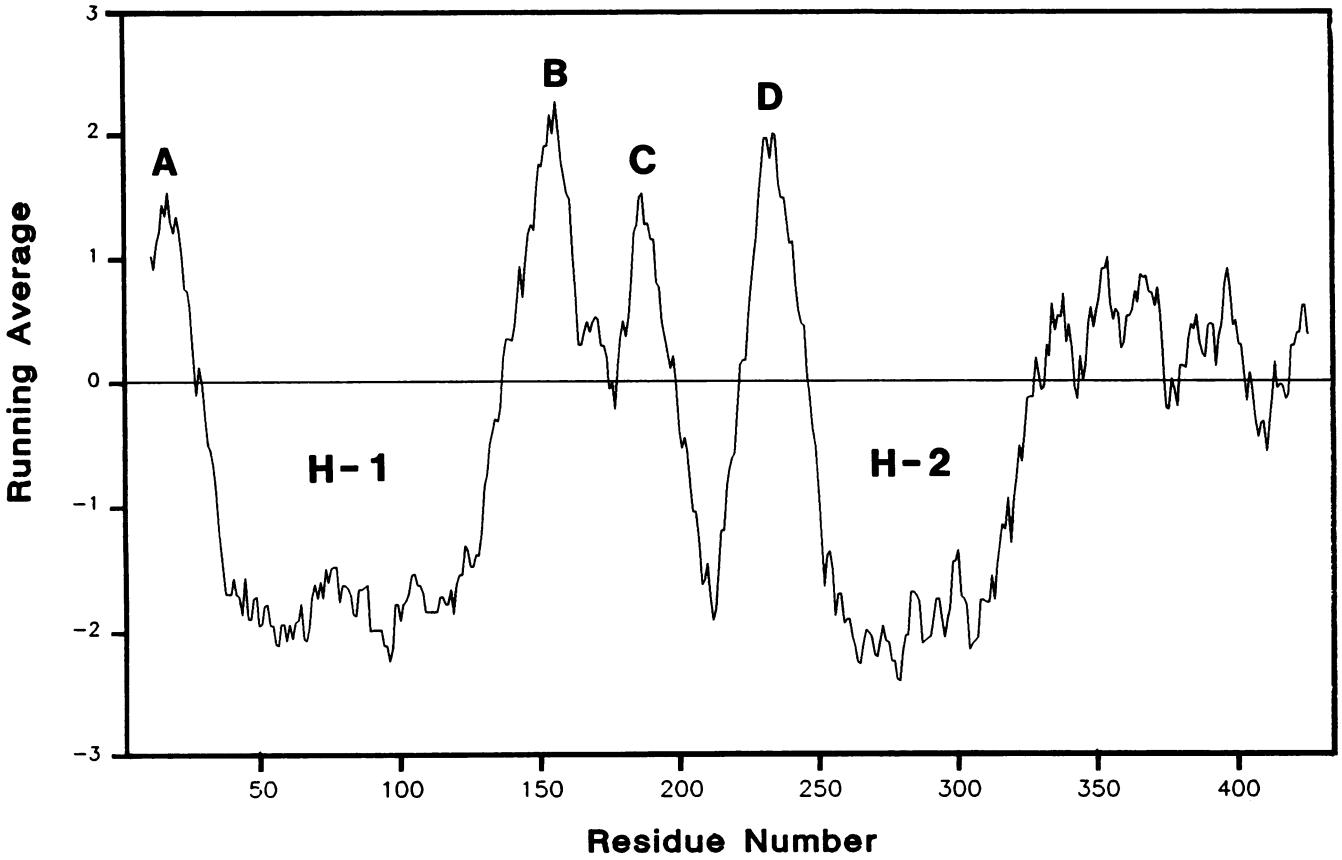


FIG. 5. Hydropathy plot of the predicted amino acid sequence. The hydrophobicity index of each residue, averaged over a window of 21 amino acids by the method of Kyte and Doolittle (27), is plotted with respect to the position along the sequence. Positive values indicate hydrophobic regions. A, Putative signal sequence; B, C, and D, three putative transmembrane domains predicted by the algorithm of Eisenberg (12); H-1 and H-2, two histidine-rich regions (see text).

Kyte-Doolittle (27) hydropathy plot (Fig. 5). With the algorithm of Eisenberg (12), three putative membrane-spanning domains were found in the center of the amino acid sequence. The two histidine-rich regions are located before and after, respectively, the three transmembrane domains.

**Charge clusters with histidine-rich repeats.** The most striking feature of the amino acid sequence is the presence of two charge clusters rich in histidine (Fig. 6). The first region from residues 40 through 123 is made up of six well-conserved repeats in tandem. Each repeat starts with one or two negatively charged residues, followed by a nonpolar residue and a stretch of 9 to 15 amino acids in which histidine alternates with mostly glycine or serine. This histidine-rich sequence constitutes a statistically significant cluster of charged residues as defined by Karlin et al. (22). The second region, from residues 252 to 274, consists of two almost identical repeats separated by a single alanine residue. Each repeat unit is made up almost exclusively of histidines alternating with glycines and serines.

An exhaustive homology search, using the FASTA program of Pearson and Lipman (31), has not shown extensive sequence identity to any known protein. The search, however, revealed several protein sequences rich in histidine and glycine. One group of such sequences includes the bovine and human high-molecular-weight kininogens (HMWK). HMWK are plasma glycoproteins that carry the vasoactive peptide bradykinin and also play a role as stoichiometric cofactors for the activation of factor XI and prekallikrein in the intrinsic pathway of blood coagulation (15). The identity between the clone 10.2a predicted amino acid sequence and the sequence of HMWK is about 35% in an 83-amino-acid overlap. It is confined to the first histidine-rich domain of clone 10.2a and fragment 1 (glycopeptide) plus fragment 2 (histidine-rich peptide) of the HMWK described by Han et al. (18, 19). The homology includes one Lys-His and seven His-Gly or Gly-His pairs. Another protein, the human histidine-rich glycoprotein, also contains a histidine-rich region that is 50% homologous to the histidine-rich region of bovine and human HMWK (25) and therefore also shares some conserved His-Gly pairs with clone 10.2a.

The pattern of histidine alternating with glycine or serine found in the KE 4 sequence has not been described before but is reminiscent of the pattern found in the paired repeat of *Drosophila melanogaster*, in which histidine alternates with proline in the sequence His-Pro-His-Pro-X-Pro-His-Pro-His-X-His-Pro-X-His-Pro-His-X-His-Pro-X-X-His-Pro-His (14). In addition to the paired repeat, the *Drosophila* gene deformed also contains the sequence His-Ser-His-Ser-His-Ser-His-Thr-His-Ser, which is very close to the motif found in KE 4 (33).

## DISCUSSION

Recently, Abe et al. (1) have searched for expressed genes in the *H-2K* region by using two strategies: identification of unique or low-copy-number sequences that are conserved

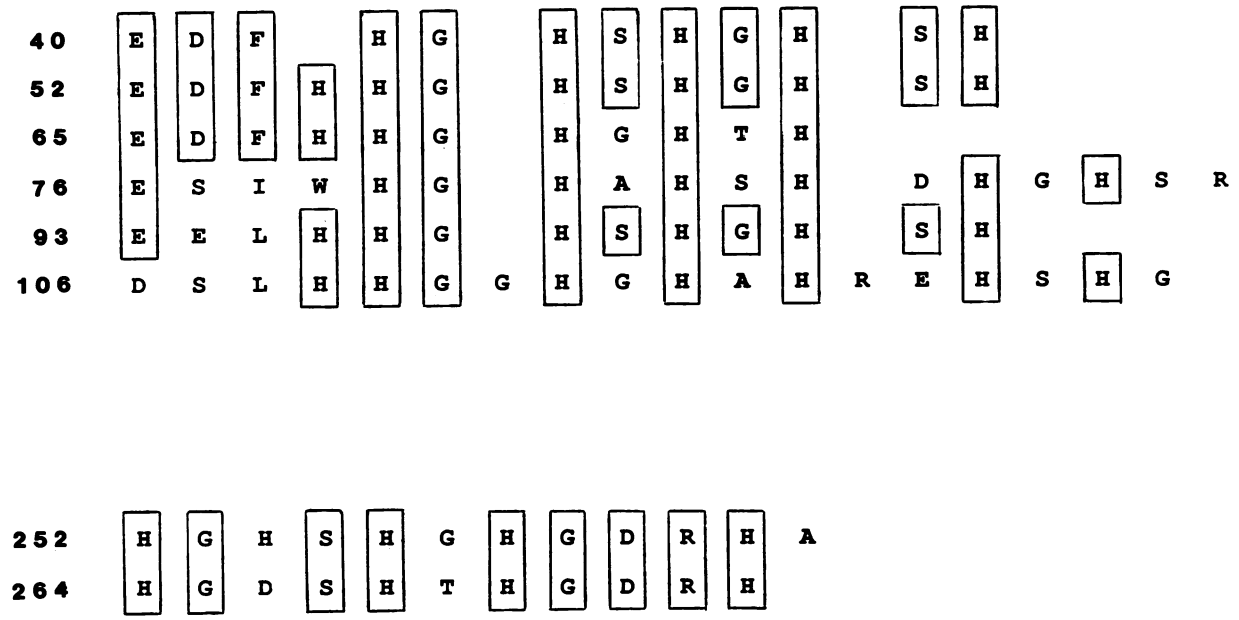


FIG. 6. Alignment of the histidine-rich repeats. The first region consists of six repeats (from residues 40 to 124); the second region consists of two repeats (from residues 252 to 274). Numbers on the left indicate the position of the first amino acid of each repeat. Conserved amino acids are boxed.

across species, and localization of methylation-free CpG islands. They identified five novel genes in the *K* region of the mouse MHC which they named KE 1 to 5. The gene present in cosmid 32, for which we have isolated a 2.7-kb cDNA, corresponds to the KE 4 gene of Abe et al. (1). This group reported that a nonmethylated CpG-rich island is present at the telomeric end of the genomic 4.9-kb *Bam*HI fragment that we used to screen the cDNA library and that sequences homologous to at least part of this gene are also present in the genomes of other mammalian species, including rat, Chinese hamster, mink, cat, and human. The pattern of tissue expression of genes KE 1 to 4 (abundant in EC cells) and their evolutionary conservation suggest that they may play some important role in embryogenesis. Furthermore, as discussed by Abe et al. (1), these genes must be considered as candidates for the $t^{w5}$ gene, particularly KE 4, which is expressed in 5.5-day-old embryos, a time when $t^{w5}$ is thought to act (5).

Nothing is yet known of the function of these genes. Nevertheless, the predicted amino acid sequence for the KE 4 gene allows educated speculations on its functions. KE 4 shows a 35% sequence identity with bovine and human HMWK over an 83-amino-acid overlap restricted to the histidine-rich region. This similarity in amino acid composition might be indicative of a conserved function for this peptide. It is close to the 25% identity over 100 amino acids that is generally taken to indicate a genuine relatedness of two peptides (11). The function of HMWK as a nonenzymatic cofactor in contact activation of the intrinsic pathway of blood coagulation is well characterized. It has been shown that HMWK is necessary for normal binding of factor XI and prekallikrein to negatively charged surfaces in the milieu of the plasma (48). Furthermore, the light chain of HMWK, which contains the histidine-rich sequence and binds firmly to negatively charged resins, is essential for the cofactor activity (41). It is also noteworthy that a loss of procoagulant activity of bovine HMWK results from a release of the histidine-rich fragment by bovine plasma kallikrein (24). Thus, it appears that HMWK binds to negatively charged surfaces through the histidine-rich region and acts as a surface-bound receptor for factor XI and prekallikrein (24). The other histidine-rich plasma protein, histidine-rich glycoprotein, has been shown to bind heparin with high affinity in a way that probably involves the histidine residues (10).

The fact that the KE 4 protein may be a transmembrane protein with a histidine-rich domain like those of HMWK and histidine-rich glycoprotein suggests the possibility that part of the KE 4 protein is involved in cellular binding to extracellular matrix, collagens, or basement membrane molecules. Assuming that the single potential N-glycosylation site (position 337) is located on the extracytoplasmic side of the membrane, the most likely configuration for this protein would place the second histidine-rich region on the extracytoplasmic side as well, a position in which it could be involved in extracellular binding. The first histidine-rich region, on the other hand, can be located on either side, depending on how many membrane-spanning domains are actually functional between the two histidine-rich regions. Some binding specificity might be provided by sequences outside the histidine-rich region or by another molecule associated with this protein. This possibility is particularly interesting since the gene products of some *t* lethal mutations have long been suggested to be associated with cell surface molecules that play a role in cell-cell interactions in early embryos (6).

The characteristic pattern found in the histidine-rich region of KE 4 has not been described before, but another sequence, the paired repeat, is also made up of histidines alternating with other residues, in this case mostly proline. This sequence, first described in the paired gene of *D. melanogaster*, defines (by hybridization experiments) a set of at least 11 other genes containing related His-Pro-rich sequences (14). The role of these conserved histidine-rich sequences is not known, but paired (14), bicoid (7), and deformed (33), which contain such sequences, are all three involved in early development and also contain homeobox

domains. Other genes of this set are suspected of having a developmental function on the basis of their temporal expression patterns (14).

Recently, Karlin et al. (22) have developed a statistical approach for the detection of "significant" clusters of charged residues in the primary structure of a polypeptide. They showed that such clusters are lacking for the bulk of eucaryotic cytoplasmic enzymes and housekeeping proteins as well as most procaryotic proteins but are abundant in (i) cellular proteins involved in regulatory functions, including nuclear transcription factors, steroid and thyroid hormone receptors, nuclear proto-oncogenes, developmental control proteins, high-molecular-weight heat shock proteins, and transactivators of eucaryotic DNA viruses, and (ii) transmembrane proteins such as voltage-gated ion channels, growth factors and neurotransmitter receptors, and opsins (22, 23; V. Brendel and S. Karlin, unpublished data). The presence of two significant charged clusters (mixed charges) in the amino acid sequence described in this report also suggests that this gene might belong to one of these functional categories of proteins.

To learn more about the protein encoded by KE 4, antibodies specific for the protein will be used to examine the spatial localization of the protein at both cell and tissue level. The temporal regulation will also be analyzed by examining protein expression in embryos at various stages of development.

## ACKNOWLEDGMENTS

We thank K. Artzt, D. Bennett, M. Steinmetz for critical comments, and G. Connelly for technical help.

This work was supported by grants from the American Cancer Society (CD-321) and the Markey Foundation.

## LITERATURE CITED

1. **Abe, K., J.-F. Wei, F.-S. Wei, Y.-C. Hsu, H. Uehara, K. Artzt, and D. Bennett.** 1988. Searching for coding sequences in the mammalian genome: the H-2K region of the mouse MHC is replete with genes expressed in embryos. EMBO. J. **7**:3441–3449.
2. **Amor, M., M. Tosi, C. Duponchel, M. Steinmetz, and T. Meo.** 1985. Liver mRNA probes disclose two cytochrome P-450 genes duplicated in tandem with the complement C4 loci of the mouse H-2S region. Proc. Natl. Acad. Sci. USA **82**:4453–4457.
3. **Artzt, K., K. Abe, H. Uehara, and D. Bennett.** 1988. Intra-H-2 recombination in the t haplotypes shows a hot spot and close linkage of $T^{w5}$ to H-2K. Immunogenetics **28**:30–37.
4. **Artzt, K., H.-S. Shin, and D. Bennett.** 1982. Gene mapping within the T/t complex of the mouse. II. Anomalous position of the H-2 complex in t-haplotypes. Cell **28**:471–476.
5. **Bennett, D.** 1975. The T-locus of the mouse: a review. Cell **6**:441–454.
6. **Bennett, D.** 1981. T/t locus, its role in embryogenesis and its relation to classical histocompatibility systems (a review). Prog. Allergy **29**:35–53.
7. **Berleth, T., M. Burri, G. Thoma, D. Bopp, S. Richtein, G. Frigerio, M. Noll, and C. Nusslein-Volhard.** 1988. The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. EMBO J. **7**:1749–1756.
8. **Brewerton, D. A., M. Caffrey, F. D. Hart, D. C. O. James, A. Nicholls, and R. D. Sturrock.** 1973. Ankylosing spondylitis and HL-A 27. Lancet **i**:904–907.
9. **Brickell, P. M., D. S. Latchman, D. Murphy, K. Willison, and P. W. J. Rigby.** 1983. Activation of a Qa/TIa class 1 major histocompatibility antigen gene is a general feature of oncogenesis in the mouse. Nature (London) **306**:756–760.
10. **Burch, M. K., M. N. Blackburn, and W. T. Morgan.** 1987. Further characterization of the interaction of histidine-rich glycoprotein with heparin: evidence for the binding of two molecules of histidine-rich glycoprotein by high molecular weight heparin and for the involvement of histidine residues in heparin binding. Biochemistry **26**:7477–7482.
11. **Doolittle, R. F.** 1986. Of URF's and ORF's: a primer on how to analyze derived amino acid sequences. University Science Books, Mill Valley, Calif.
12. **Eisenberg, D.** 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. Annu. Rev. Biochem. **53**:595–623.
13. **Feinberg, A. P., and B. Vogeistein.** 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
14. **Frigerio, G., M. Burri, D. Bopp, S. Baumgartner, and M. Noll.** 1986. Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. Cell **47**:735–746.
15. **Furie, B., and B. C. Furie.** 1988. The molecular basis of blood coagulation. Cell **53**:505–518.
16. **Goldbard, S. B., K. M. Verbanac, and C. M. Warner.** 1982. Genetic analysis of H-2 linked gene(s) affecting early mouse embryo development. J. Immunogenet. **9**:77–82.
17. **Gubbler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
18. **Han, Y. N., H. Kato, S. Iwanaga, and T. Suzuki.** 1976. Primary structure of bovine plasma high-molecular weight kininogen. J. Biochem. **79**:1201–1222.
19. **Han, Y. N., M. Komiya, S. Iwanaga, and T. Suzuki.** 1975. Studies on the primary structure of the bovine high-molecular weight kininogen. J. Biochem. **77**:55–68.
20. **Hood, L., M. Steinmetz, and B. Malissen.** 1983. Genes of the major histocompatibility complex of the mouse. Annu. Rev. Immunol. **1**:529–563.
21. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA Cloning, vol. 1. IRL Press, Oxford.
22. **Karlin, S., B. E. Blaisdell, E. S. Mocarski, and V. Brendel.** 1989. A method to identify distinctive charge configurations in protein sequences with application to human herpesvirus polypeptides. J. Mol. Biol. **205**:165–177.
23. **Karlin, S., and V. Brendel.** 1988. Charge configurations in viral proteins. Proc. Natl. Acad. Sci. USA **85**:9396–9400.
24. **Kerbiriou, D. M., and J. H. Griffin.** 1979. Human high molecular weight kininogen: studies of structure-function relationships and of proteolysis of the molecule occuring during contact activation of plasma. J. Mol. Chem. **254**:12020–12027.
25. **Koide, T., D. Foster, S. Yoshitake, and E. W. Davie.** 1986. Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA. Biochemistry **25**:2220–2225.
26. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
27. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
28. **Levi-Strauss, M., M. C. Carroll, M. Steinmetz, and T. Meo.** 1988. A previously undetected MHC gene with an unusual periodic structure. Science **240**:201–204.
29. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Murphy, D., P. M. Brickell, D. S. Latchman, K. Willison, and P. W. J. Rigby.** 1983. Transcripts regulated during normal embryonic development and oncogenic transformation share a repetitive element. Cell **35**:865–871.
31. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
32. **Randall, L. L., and S. J. S. Hardy.** 1989. Unity in function in the absence of consensus in sequence: role of leader peptides in export. Science **243**:1156–1159.
33. **Regulski, M., N. McGinnis, R. Chadwick, and W. McGinnis.**

1987. Developmental and molecular analysis of deformed; a homeotic gene controlling *Drosophila* head development. EMBO J. **6**:767–777.

34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schlosstein, L., P. I. Terasaki, R. Bluestone, and C. M. Pearson.** 1973. High association of an HL-A antigen, W27 with ankylosing spondylitis. N. Engl. J. Med. **288**:704–706.
36. **Shin, H.-S.** 1989. The T/t complex and the genetic control of mouse development, p. 443–471 *In* S. D. Litwin (ed.), Human immunogenetics: basic principles and clinical relevance. Marcel Dekker, Inc., New York.
37. **Shin, H.-S., T. A. Bargiello, B. T. Clark, F. R. Jackson, and M. W. Young.** 1985. An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates. Nature (London) **317**:445–448.
38. **Shin, H.-S., L. Flaherty, K. Artzt, D. Bennett, and J. Ravetch.** 1983. Inversion in the H-2 complex of t-haplotypes in mice. Nature (London) **306**:380–383.
39. **Spies, T., G. Blanck, M. Bresnahan, J. Sands, and J. L. Strominger.** 1989. A new cluster of genes within the human major histocompatibility complex. Science **243**:214–217.
40. **Spies, T., C. C. Morton, S. A. Nedospasov, W. Fiers, D. Pious, and J. L. Stominger.** 1986. Genes for the tumor necrosis factors α and β are linked to the human major histocompatibility complex. Proc. Natl. Acad. Sci. USA **83**:8699–8702.
41. **Thompson, R. E., R. Mandle, Jr., and A. P. Kaplan.** 1979. Studies of binding of prekallikrein and factor XI to high molecular weight kininogen and its light chain. Proc. Natl. Acad. Sci. USA **76**:4862–4866.
42. **Uehara, H., K. Abe, C.-H. T. Park, H.-S. Shin, D. Bennett, and K. Artzt.** 1987. The molecular organization of the H-2K region of two t-haplotypes: implications for the evolution of genetic diversity. EMBO J. **6**:83–90.
43. **Von Heijne, G.** 1983. Signal-sequence cleavage sites. Eur. J. Biochem. **133**:17–21.
44. **Von Heijne, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173**:243–251.
45. **Von Heijne, G.** 1985. Signal sequences, the limits of variation. J. Mol. Biol. **184**:99–105.
46. **White, P. C., D. D. Chaplin, J. H. Weis, B. DuPont, M. I. New, and J. G. Seidman.** 1984. Two steroid 21-hydroxylase genes are located in the murine 5 region. Science **312**:465–467.
47. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent mRNA 3′ end formation. Science. **226**:1045–1051.
48. **Wiggins, R. C., B. N. Bouma, C. G. Cochrane, and J. H. Griffin.** 1977. Role of high-molecular weight kininogen in surface-binding and activation of coagulation factor XI and prekallikrein. Proc. Natl. Acad. Sci. USA **74**:4636–4640.