A cluster of transcribed sequences between the *Pb* and *Ob* genes of the murine major histocompatibility complex

(LMP antigens/class II genes/transport proteins)

Sungae Cho, Michelle Attaya, Michael G. Brown, and John J. Monaco

Department of Microbiology and Immunology, Medical College of Virginia–Virginia Commonwealth University, Richmond, VA 23298-0678

ABSTRACT   The region of the murine major histocompatibility complex (MHC) between the *Pb* ($A_{\beta 3}$) and *Ob* ($A_{\beta 2}$) genes controls the expression of an intracellular complex named the LMP (low molecular weight polypeptide) complex. DNA probes for at least seven different genes mapping to this region were isolated. These hybridize to a minimum of eight different transcripts ranging from approximately 1.3 to 3.7 kilobases (kb). The deduced amino acid sequences of the corresponding cDNAs indicate that three of these genes are new members of the MHC class II gene family. These genes are transcribed in a tissue-specific pattern similar to that of the traditional class II genes. Two of the remaining four genes, *HAM1* and *HAM2*, are homologous to one another and to a family of eukaryotic and prokaryotic transport proteins and may be involved in antigen processing. The tissue distribution of *HAM1* transcripts is consistent with its proposed role in class I-restricted antigen processing, whereas *HAM2* transcription appears more restricted and may be involved in antigen processing for class II-restricted T cells. The *HAM2* gene may produce two differentially spliced transcripts. The identity of the remaining two genes is not known. Analyses of transcript sizes, tissue distribution, sequence, and genetic mapping data suggest that none of these genes code for LMP antigens.

The murine major histocompatibility complex (MHC), *H-2*, contains a cluster of genes intimately involved in the immune response and encompasses ≈4000 kilobases (kb) (2 centimorgans) of murine chromosome 17 (1). MHC-encoded molecules have been classified into three groups based on functional and biochemical homology. Class I and II molecules bind processed peptide fragments of foreign antigens and present them to cytotoxic T cells and helper T cells, respectively. Class III molecules are components of the complement system. Approximately 32 class I antigen genes, 7 class II antigen genes (1, 2), and 4 class III antigen genes (3) have been identified in the *H-2* region. In addition to these classical MHC-encoded molecules, several other genes have been identified within the MHC, such as genes for tumor necrosis factor and lymphotoxin (4) and steroid 21-hydroxylase (5).

Genes for class II molecules, or I-region-associated (Ia) antigens, are clustered near the centromeric end of the MHC (Fig. 1). The two classical murine Ia antigens, I-A and I-E, are expressed on the cell surface as heterodimers composed of α and β chains. Three other β-chain genes have been identified by cross-hybridization (see Fig. 1). No protein products have yet been detected corresponding to *Ob* or *Eb2*. The *Pb* gene is a pseudogene (6), and the expression of I-E heterodimers is mouse haplotype-dependent (7). The distance between *Pb* and *Ea* is ≈270 kb (8).

Monaco and McDevitt (9) demonstrated that the region between the *Pb* and *Ob* genes controlled the expression of the

LMP (low molecular weight polypeptide) antigen complex. This complex has a native molecular weight of ≈580,000 and is composed of 16 noncovalently associated polypeptides with molecular weights between 15,000 and 30,000. The LMP antigens are distinct from previously defined MHC-encoded molecules with respect to tissue distribution, biochemistry, serology, and genetics (9–11). Evidence that LMP antigens are involved in immune function includes their genetics, regulation (20- to 25-fold induction) by interferon γ (IFN-γ), similar to class I, II, and III antigens (12–14), and their expression in macrophage and lymphoid cell lines (11).

Serological analysis, using intra-I-region recombinant mouse strains, mapped at least one LMP antigen gene to a region of ≈100 kb between the *yl* (≈20 kb proximal to *Ob*) and *bql* (≈50 kb distal to *Pb*) recombination points (11). As one approach to further characterize the LMP antigens, we attempted to isolate the corresponding genes by screening a set of overlapping genomic cosmid clones spanning this 100-kb segment for transcribed sequences. We show here that eight previously undetected transcripts originate from this region, corresponding to at least seven different genes, and describe their tissue specificity, sizes, and potential relationship to LMP antigens.

MATERIALS AND METHODS

Cell Lines and Mice. All cell lines were cultured in RPMI 1640 medium supplemented with penicillin (100 units/ml), streptomycin (100 μg/ml), glutamine (2 mM), and 10% fetal bovine serum. The IFN-γ-treated cells were grown in medium supplemented with IFN-γ (ICN; 15 units/ml) for 3–4 days. The effect of Con A supernatant was determined by incubation of WEHI-3 cells with 20% (by volume) of a culture supernatant derived from Con A (Sigma)-activated BALB/c mouse spleen cells (15). For the analysis of tissue-specific expression, total RNA was obtained from brain, heart, lung, thymus, kidney, liver, and spleen of BALB/c mice; B-cell line TA3 (16); T-cell lines RMA and RMA-S (17); fibroblast cell lines 3T3 and L-M(TK⁻); and myelomonocytic cell line WEHI-3.

Cosmids. Genomic cosmid clones (designated 4.24, 5.22, 5.10, and 5.9; Fig. 1) covering the region between the *bql* and *yl* crossover points were kindly provided by M. Steinmetz (8). *Eco*RI-digested cosmid fragments, designated in alphabetical order for each cosmid from the largest fragment, were separated by electrophoresis in a 1% agarose gel. Fragments used to probe Northern blots were radiolabeled with [α-³²P]dCTP by random priming (18).

Northern Blot Analysis. Cytoplasmic RNA was isolated from both IFN-γ-treated and untreated WEHI-3 cells by the method of Favaloro *et al.* (19) with some modification (ref. 20, pp. 4.1.2–4.1.6). Total RNA was isolated by a modification of the method of Glisin *et al.* (28). Poly(A)⁺ mRNA was

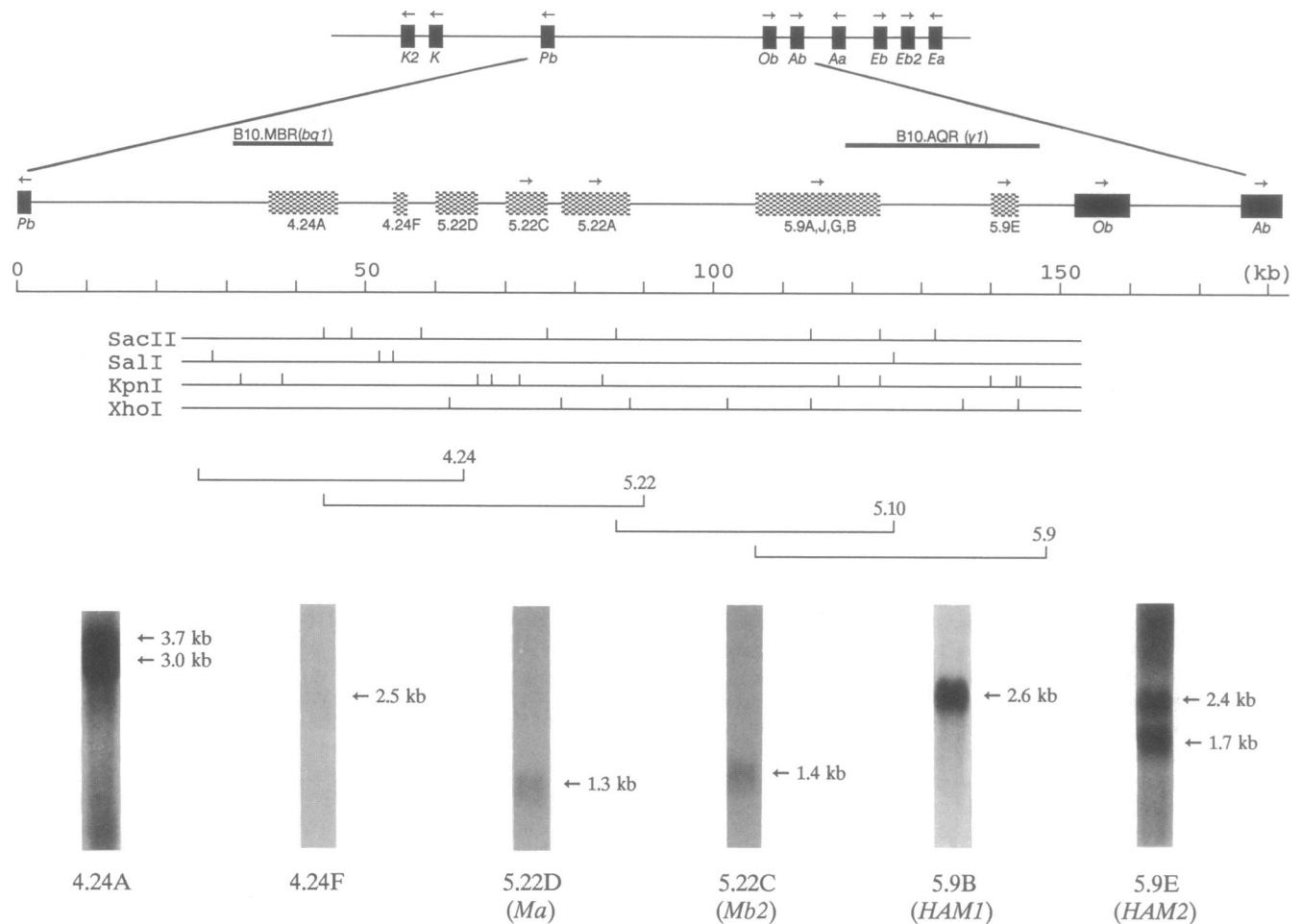Abbreviations: IFN-γ, interferon γ; MHC, major histocompatibility complex.

Fɪɢ. 1.   Identification of probes for transcribed sequences. (*Upper*) Genomic organization (left to right = centromeric to telomeric) of the proximal portion of *H-2*. The region between *Pb* and *Ab* is expanded below, with the scale in kilobases. Probes used for Northern blots are marked as stippled boxes. Arrows indicate 5' → 3' orientation of the corresponding genes, where known. Sites for four restriction enzymes (*Sac* II, *Sal* I, *Kpn* I, and *Xho* I) were mapped for comparison with the original restriction maps of cosmids 4.24, 5.22, 5.10, and 5.9 (8). The sizes of genomic *Eco*RI fragments used as probes are as follows: 4.24A, 9.6 kb; 4.24F, 1.6 kb; 5.22D, 6.0 kb; 5.22C, 7.9 kb; 5.9B, 4.5 kb; 5.9E, 3.3 kb. (*Lower*) WEHI-3 cells were cultured with IFN-γ (15 units/ml) for 3–4 days. Poly(A)$^+$ RNA was isolated, electrophoresed, transferred to nylon, and probed with ^{32}P-radiolabeled *Eco*RI cosmid fragments.

isolated by oligo(dT) (Pharmacia) affinity chromatography. From 3 to 10 μg of poly(A)$^+$ mRNA or 30 μg of total RNA was run in a 2.2 M formaldehyde/1% agarose gel and transferred to nylon membrane (BioTrace RP, Gelman) in 20× standard saline citrate (SSC) overnight.

Southern Blot Analysis. Total genomic DNA was isolated from BALB/c mouse liver by the method of Ausubel *et al.* (ref. 20, pp. 2.2.1–2.2.3). Ten micrograms of total genomic DNA was digested with *Eco*RI and electrophoresed in a 1% agarose gel. The separated genomic DNA was transferred to nylon membrane with 0.4 M NaOH/1 M NaCl overnight. Cosmid fragments hybridizing to a single band on Southern blots (i.e., free of repetitive sequence) were subcloned for cDNA library screening.

Construction and Screening of cDNA Libraries. Two λgt10 WEHI-3 cDNA libraries were constructed from RNA extracted from IFN-γ-treated WEHI-3 cells; one library was made from poly(A)$^+$ RNA by using the Librarian X (Invitrogen, San Diego) cDNA library kit and the other from total RNA according to the protocol of Ausubel *et al.* (ref. 20, 5.5.1–5.8.5). Size-selected cDNAs between 0.7 and 5 kb were blunt-end ligated with *Eco*RI linkers and inserted into λgt10 arms. Libraries were plated on *Escherichia coli* C600 Hfl hosts, and plaques were transferred to nitrocellulose as described by Benton and Davis (21).

RESULTS

Eight hybridization bands were detected by Northern blot analysis of WEHI-3 (myelomonocytic cell line) RNA using radiolabeled genomic cosmid DNA fragments as probes (Fig. 1). The genomic probes and characterization of the corresponding transcripts and cDNAs are discussed individually below. The size of each transcript is indicated in parentheses.

Probes 4.24A (3.7 and 3.0 kb) and 4.24F (2.5 kb). cDNA clones corresponding to probe 4.24A have been isolated (C. Gilchrist and J.J.M., unpublished data) but have not been fully characterized, and no cDNA clones corresponding to 4.24F have been found, perhaps because of the relatively low level of expression inferred from the blots shown in Fig. 1. Thus, the transcriptional orientation of these two genes is not known. Preliminary data suggest that the two transcripts detected with the 4.24A probe are related to each other and are either generated by alternative splicing or have enough internal homology to cross-hybridize. The expression of 4.24A transcripts has been variable in different RNA preparations from WEHI-3 cells, for reasons that are not fully understood. Subtle differences in culture conditions such as serum components may cause induction or repression of this gene.

Probes 5.22A and 5.22C (Both 1.4 kb) and 5.22D (1.3 kb). With the 5.22C fragment as probe, a 1.2-kb cDNA clone was

isolated from the WEHI-3 cDNA library and sequenced. The deduced amino acid sequence has significant homology to class II β-chain sequences (unpublished data). Comparison of the restriction maps of the 5.22C cDNA clone and the genomic fragment identifies the orientation of the 5.22C gene as designated in Fig. 1. The entire 1.2-kb cDNA sequence is contained within the 8-kb 5.22C *Eco*RI fragment. The 5.22A fragment, which lies in close proximity to 5.22C (Fig. 1), also contains a copy of this gene. Partial sequence comparisons of the genomic fragments to the cDNA sequence have not revealed any differences in coding sequence between 5.22C and 5.22A, although multiple nucleotide substitutions are found in intron sequences (unpublished data). It is therefore unclear whether only one copy or both copies of this gene are transcribed and from which copy the cDNA sequence is derived. The orientation of both copies is the same as for other functional class II β-chain genes. The genes contained in 5.22A and 5.22C will be referred to as *Mb1* and *Mb2*, respectively.

A 1.1-kb cDNA clone corresponding to probe 5.22D was isolated from the WEHI-3 library. Preliminary sequence analysis has revealed significant homology with class II α chains (unpublished data). This gene will be referred to as *Ma*.

As shown in Fig. 2, the level of *Ma* and *Mb* transcripts is increased by treatment either with supernatants from Con A-stimulated T cells or with IFN-γ. Both Con A supernatant and IFN-γ are known to enhance the expression of class II genes (22, 23). Although it is difficult to compare directly the intensity of the signals on the two sets of blots (because both the blots and probes were prepared at different times, and the specific activity of the probe was not controlled), transcription of both *Ma* and *Mb* appears to be induced to a greater extent by Con A supernatant than by IFN-γ alone. It is possible that other components in Con A supernatant may synergize with IFN-γ in the induction of these genes. *Ea*

cDNA was used as a positive control probe with the same RNA preparations; increased levels of *Ea* transcripts were observed for both treatments.

Ma and *Mb* expression was observed in a B-cell line (TA3) and in spleen, and weakly in the WEHI-3 macrophage cell line and kidney, but not in T-cell lines (RMA, RMA-S) or other tissues (Fig. 3), consistent with the tissue specificity of conventional class II genes (24).

Probes 5.9B (2.6 kb) and 5.9E (1.7 kb and 2.4 kb). cDNAs of 2.4 and 0.9 kb were isolated with probes 5.9B and 5.9E, respectively. As inferred from the number of cDNA clones isolated, 5.9B and 5.9E are the most abundant among the newly found transcribed sequences, with a frequency of ≈0.005% of total mRNA. The complete nucleotide sequence of the 2.4-kb 5.9B cDNA was determined, and this gene has been designated *HAM1* (25). Fragments 5.9G, 5.9J, and 5.9A also detect a transcript of ≈2.6 kb, and subsequent studies show that these fragments also contain portions of the *HAM1* gene, which spans ≈11 kb in genomic DNA (M.A. and J.J.M., unpublished data). The order of these fragments and the orientation of *HAM1* as shown in Fig. 1 were determined by comparison of restriction maps and sequences of the cDNA and genomic clones. The level of *HAM1* transcripts in WEHI-3 cells is increased by treatment with Con A supernatant (data not shown).

The 0.9-kb 5.9E partial cDNA clone contains a polyadenylylation signal sequence, AATAAA, indicating that this clone corresponds to the 3' end of the gene. This cDNA sequence is homologous to the corresponding portion of the *HAM1* gene, and the gene corresponding to 5.9E has been designated *HAM2* (25). It is unclear whether the 2.4-kb transcript detected with the 5.9E genomic fragment represents weak cross-hybridization to the homologous 2.6-kb *HAM1* transcript (the size difference between 2.4 and 2.6 kb may be due to variation in measurement) or whether the two transcripts arise from alternative splicing, although the bulk
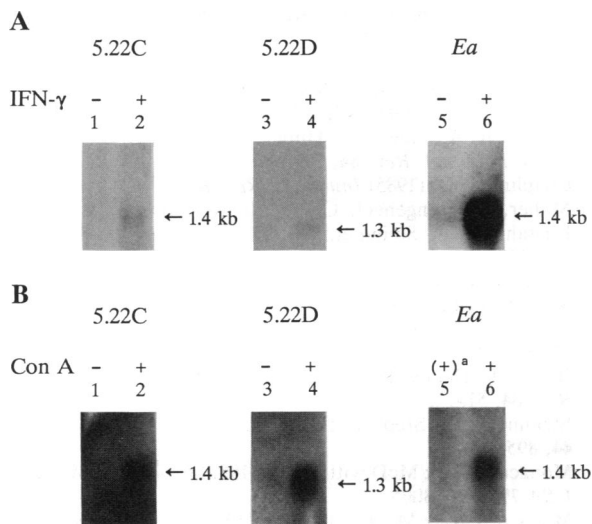

FIG. 2. Effect of IFN-γ and Con A supernatant on transcription of class II genes. (*A*) Three micrograms of poly(A)⁺ RNA obtained from untreated (lanes 1, 3, and 5) or IFN-γ (15 units/ml)-treated (lanes 2, 4, and 6) WEHI-3 cells was run in a formaldehyde/1% agarose gel and transferred onto nylon. RNA blots were hybridized with ³²P-radiolabeled 5.22C (lanes 1 and 2) and 5.22D (lanes 3 and 4) genomic probes. *Ea* cDNA was used as control (lanes 5 and 6). (*B*) Thirty micrograms of total RNA extracted from WEHI-3 cells incubated with (lanes 2, 4, 5, and 6) or without (lanes 1 and 3) Con A supernatant was transferred to nylon and probed as described above. Lane 5 marked as (+)ᵃ corresponds to RNA extracted from cells treated with 10% Con A supernatant and lanes 2, 4, and 6 contain RNA from cells incubated with 20% Con A supernatant.
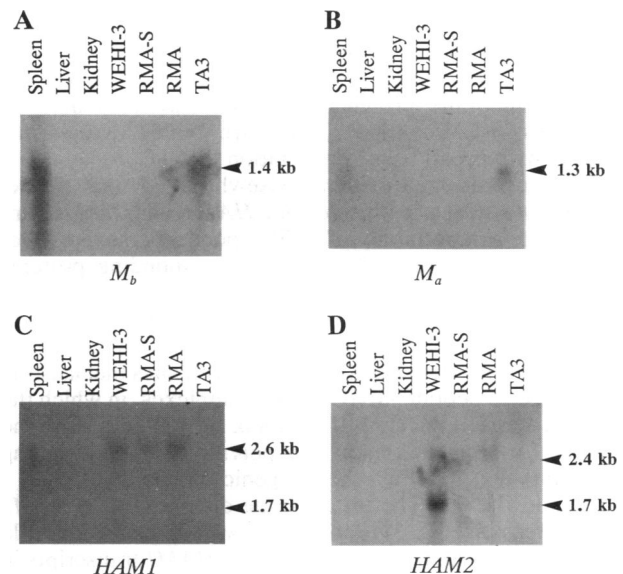

FIG. 3. Tissue distribution of class II and transport protein transcripts. Thirty micrograms of total RNA was purified from the tissues and cell lines indicated, electrophoresed in a formaldehyde/1% agarose gel, transferred to nylon membrane, and hybridized to 1.3 × 10⁷ cpm of the following probes: CIII-12, a 1.2-kb cDNA containing the complete *Mb* coding sequence (see text) (*A*); DII-1, a 1.1-kb cDNA containing the complete *Ma* coding sequence (*B*); B5, a 1.4-kb cDNA containing the 3' end of the *HAM1* coding sequence (25) (*C*); EII-4, a 0.95-kb cDNA containing the 3' end of the *HAM2* coding sequence (*D*). All blots were exposed to film at −70°C for 96 hr with a DuPont Cronex Lightning Plus intensifying screen, except for *B*, which is a 24-hr exposure.

of the current evidence favors the latter. The relative amounts of the two transcripts differ in different preparations of WEHI-3 RNA (Fig. 1 vs. Fig. 3).

The 2.6-kb *HAM1* transcript is expressed in spleen, 3T3 fibroblast (data not shown), macrophage (WEHI-3), and T-cell lines (RMA, RMA-S) (Fig. 3). Long exposures of the blot in Fig. 3 also demonstrate low-level expression in the B-cell line TA3. In a separate experiment, *HAM1* expression was also observed in one preparation of liver RNA. Expression of the 1.7-kb *HAM2* transcript was detected only in the macrophage cell line; the 2.4-kb *HAM2* transcript was present at low levels in both macrophage and T-cell lines (Fig. 3). Interestingly, faint bands at ≈1.7 kb are visible in some preparations of RNA when probed with *HAM1* probes (Fig. 3), but not in others (Fig. 1). It is unlikely that this signal is due to cross-hybridization with *HAM2* transcripts, because of the restricted distribution of the latter. This 1.7-kb species may result from alternate splicing of the *HAM1* gene or from partial degradation in some of the RNA preparations.

DISCUSSION

The serological polymorphism in LMP antigens is controlled by genes mapping in the I-A region of the MHC (9–11). Only 2 of the 16 LMP subunits are polymorphic as determined by electrophoretic analysis; both of these map to the MHC, and at least one maps between *Pb* (previously A$_{\beta3}$) and *Ob* (previously A$_{\beta2}$) (11). The genes for the remaining 14 subunits have not been mapped.

Although no sequence information is yet available for the 4.24A or 4.24F genes, these are unlikely to belong to either the class I, the class II, or the LMP gene family, based on their transcript sizes. Recent genetic mapping data (C. K. Smith and J.J.M., unpublished) using several new intra-I-region recombinants (26), also excludes the possibility that these genes encode either of the polymorphic LMP subunits. The *Ma* and *Mb* genes (corresponding to probes 5.22D, 5.22C, and 5.22A) are members of the MHC class II gene family, based both on sequence homology and on tissue distribution; their gene products are thus expected to be integral membrane proteins and hence are also unlikely candidates for LMP antigens, which are soluble, cytoplasmic proteins with broad tissue distribution (9, 11).

We have presented evidence elsewhere (25) that the remaining two genes in this region, *HAM1* and *HAM2*, are members of a superfamily of ATP-dependent transport protein genes and suggested that the corresponding proteins function in the transport of antigen, or peptide fragments of processed antigens, into subcellular compartments containing newly synthesized MHC molecules. This hypothesis is based on evidence derived from mutant cell lines that lack the ability to form peptide–MHC class I complexes, in which the mutation maps to the class II region of the MHC, and on the mapping to the same region of a polymorphic, trans-acting gene (*cim*) that modifies the antigenic character of class I antigens in the rat. The broad tissue distribution of *HAM1* gene expression (Fig. 3) is consistent with its postulated role in class I expression. The expression of *HAM1* transcripts in class I-deficient RMA-S cells is not inconsistent with this hypothesis, since the RMA-S defect is presumed to be a point mutation (17). However, the more restricted tissue distribution of *HAM2* transcripts argues against such a role. Mutants defective in class II-restricted antigen presentation have been isolated from a cell line hemizygous in the *HLA-D* region (27), suggesting that the gene responsible maps within this region. It is possible that the defect in these cells may be in the human *HAM2* gene.

The lack of easily detectable transcripts from either *HAM1* or *HAM2* in TA3 cells, which express functional class I and class II antigens, is somewhat surprising. Either exception-

ally low levels of transcription are sufficient for full function or alternative proteins or pathways may be utilized to replace *HAM* gene function in these cells.

Two of the subunits of the LMP complex have been mapped to the MHC (11). The *HAM1* and *HAM2* genes map to a region of the MHC which would be consistent with the possibility that they encode these two LMP subunits. Even though the predicted *HAM1* and *HAM2* gene products are larger than any of the LMP subunits, it is possible that these two gene products are posttranslationally processed by proteolytic cleavage into two halves of approximately the correct molecular weight, roughly corresponding to the transmembrane and cytoplasmic domains of the predicted proteins (25). Indeed, for several of the prokaryotic members of this family of transport proteins, the membrane and cytoplasmic domains of the molecule reside on separate polypeptide chains of approximately equal size (25). Thus, we cannot definitely rule out the possibility that isolated cytoplasmic domains of these two molecules (*HAM1* and *HAM2*) might correspond to the two polymorphic MHC-linked LMP subunits. Again, however, by analogy to other family members, both domains of the transporter would be expected to remain tightly membrane-associated, in contrast to the soluble cytoplasmic LMP complex. Furthermore, the differing tissue distribution of *HAM1* and *HAM2* transcripts argues that they cannot *both* be LMP subunits. It therefore seems likely that there is at least one and possibly two more genes in this region yet to be discovered.

In summary, we have identified seven genes in the *Pb–Ob* interval of the murine MHC. None of these genes appears likely to encode an LMP antigen. However, we have recently confirmed and refined the original LMP mapping data (C. K. Smith and J.J.M., unpublished work), which indicated that this region contains both of the polymorphic LMP genes. We therefore suggest that there are probably at least nine genes in the region between *Pb* and *Ob*.

1. Klein, J. (1986) *Natural History of the Major Histocompatibility Complex* (Wiley, New York), pp. 132–135.
2. Flavell, R. A., Allen, H., Huber, B., Wake, C. & Widera, G. (1985) *Immunol. Rev.* **84**, 29–50.
3. Chaplin, D. D. (1985) *Immunol. Rev.* **87**, 61–80.
4. Müller, U., Jongeneel, C. V., Nedospasov, S. A., Fischer-Lindahl, K. & Steinmetz, M. (1987) *Nature (London)* **325**, 265–267.
5. White, P. C., Chaplin, D. D., Weiss, J. H., Dupont, B., New, M. I. & Seidman, J. G. (1984) *Nature (London)* **312**, 465–467.
6. Widera, G. & Flavell, R. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5500–5504.
7. Bell, J. I., Denny, D. W. & McDevitt, H. O. (1985) *Immunol. Rev.* **84**, 51–71.
8. Steinmetz, M., Stephan, D. & Fischer-Lindahl, K. (1986) *Cell* **44**, 895–904.
9. Monaco, J. J. & McDevitt, H. O. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3001–3005.
10. Monaco, J. J. & McDevitt, H. O. (1984) *Nature (London)* **309**, 797–799.
11. Monaco, J. J. & McDevitt, H. O. (1986) *Hum. Immunol.* **15**, 416–426.
12. Wong, G. H. W., Clark-Lewis, I., Harris, A. W. & Schrader, J. W. (1984) *Eur. J. Immunol.* **14**, 52–56.
13. Rosa, F., Hatat, D., Abadie, A. & Fellous, M. (1985) *Ann. Inst. Pasteur/Immunol.* **136**, 103–119.
14. Strunk, R. C., Cole, F. S., Perlmutter, D. H. & Colten, H. R. (1985) *J. Biol. Chem.* **260**, 15280–15285.
15. Steeg, P. S., Moore, R. N. & Oppenheim, J. J. (1980) *J. Exp. Med.* **152**, 1734–1744.
16. Glimcher, L. H., Hamano, T., Asofsky, R., Sachs, D. H., Pierres, M., Samelson, L. E., Sharrow, S. O. & Paul, W. E. (1983) *J. Immunol.* **130**, 2287–2294.

17. Kärre, K., Ljunggren, H. G., Piontek, G. & Kiessling, R. (1986) *Nature (London)* **319**, 675–678.
18. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
19. Favaloro, J., Treisman, R. & Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749.
20. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
21. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
22. Mathis, D. J., Benoist, C. O., Williams, V. E., II, Kanter, M. R. & McDevitt, H. O. (1983) *Cell* **32**, 745–754.
23. Nakamura, M., Manser, T., Pearson, G. D. N., Daley, M. J. &

24. Le Meur, M., Gerlinger, P., Benoist, C. & Mathis, D. (1985) *Nature (London)* **316**, 38–42.
25. Monaco, J. J., Cho, S. & Attaya, M. (1990) *Science* **250**, 1723–1726.
26. Uematsu, Y., Fischer-Lindahl, K. & Steinmetz, M. (1988) *Immunogenetics (NY)* **27**, 96–101.
27. Mellins, E., Smith, L., Arp, B., Cotner, T., Celis, E. & Pious, D. (1990) *Nature (London)* **343**, 71–74.
28. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 7.19–7.22.

Gefter, M. L. (1984) *Nature (London)* **307**, 381–382.