*Proc. Natl. Acad. Sci. USA*
Vol. 93, pp. 9096–9101, August 1996
Immunology

# Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex

(evolution/synteny)

MASANORI KASAHARA*†, MASARU HAYASHI*, KEIJI TANAKA‡, HIDETOSHI INOKO§, KIMIHIKO SUGAYA¶, TOSHIMICHI IKEMURA¶, AND TERUO ISHIBASHI*

*Department of Biochemistry, Hokkaido University School of Medicine, Sapporo 060, Japan; ‡Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan; §Division of Molecular Life Science, Tokai University School of Medicine, Isehara 259-11, Japan; and ¶Department of Evolutionary Genetics, National Institute of Genetics, and The Graduate University for Advanced Studies, Mishima 411, Japan

*Communicated by D. Bernard Amos, Duke University Medical Center, Durham, NC, May 22, 1996 (received for review March 20, 1996)*

**ABSTRACT** **Proteasomes are the multi-subunit protease thought to play a key role in the generation of peptides presented by major histocompatibility complex (MHC) class I molecules. When cells are stimulated with interferon γ, two MHC-encoded subunits, low molecular mass polypeptide (LMP) 2 and LMP7, and the MECL1 subunit encoded outside the MHC are incorporated into the proteasomal complex, presumably by displacing the housekeeping subunits designated Y, X, and Z, respectively. These changes in the subunit composition appear to facilitate class I-mediated antigen presentation, presumably by altering the cleavage specificities of the proteasome. Here we show that the mouse gene encoding the Z subunit (*Psmb7*) maps to the paracentromeric region of chromosome 2. Inspection of the mouse loci adjacent to the *Psmb7* locus provides evidence that the paracentromeric region of chromosome 2 and the MHC region on chromosome 17 most likely arose as a result of a duplication that took place at an early stage of vertebrate evolution. The traces of this duplication are also evident in the homologous human chromosome regions (6p21.3 and 9q33–q34). These observations have implications in understanding the genomic organization of the present-day MHC and offer insights into the origin of the MHC.**

Accumulated evidence indicates that the proteasome plays a key role in the generation of cytosolic peptides presented by major histocompatibility complex (MHC) class I molecules (reviewed in refs. 1–3). The 20S proteasome, which functions as the proteolytic core of the 26S protease complex, has a cylindrical structure with four layers of rings, each composed of seven subunits (4–6). The outer two rings are made up of α-type subunits, while the inner two rings are composed of β-type subunits. The catalytic sites of the 20S proteasome reside in the β-type subunits, with the hydroxyl group of the amino-terminal threonine acting as the nucleophile in peptide hydrolysis (7–9). Two of the β-type subunits, known as LMP2 (low molecular mass polypeptide 2) and LMP7, are encoded within the class II region of the human, mouse, and rat MHC (10, 11). Like the class I and TAP (transporter associated with antigen processing) molecules, the expression of these subunits is upregulated by interferon γ (IFN-γ). Recent studies (12–14) showed that, upon stimulation with IFN-γ, LMP2 and LMP7 are incorporated into the proteasomal complex, presumably by displacing the housekeeping β-type subunits designated Y (also known as subunit δ, 2, PSMB6) and X (also known as subunit ε, MB1, 10, PSMB5), respectively. These changes in the subunit composition appear to facilitate class I-mediated antigen presentation, presumably by altering the cleavage specificities of the proteasome (15–17).

Recently, Hisamatsu *et al.* (18) identified a third pair of structurally similar proteasome β-type subunits, designated MECL1 and Z, whose expression is regulated reciprocally by IFN-γ. When cells are stimulated with IFN-γ, the MECL1 subunit is thought to be incorporated into the proteasome by displacing the housekeeping Z subunit. Unlike the LMP2 and LMP7 subunits, the IFN-γ-inducible MECL1 subunit is encoded outside the MHC (19). Here we show that the mouse gene (*Psmb7*)‖ encoding the Z subunit maps to the paracentromeric region of chromosome 2. Inspection of the mouse loci adjacent to the *Psmb7* locus provides evidence that the paracentromeric region of chromosome 2 and the MHC region on chromosome 17 most likely arose as a result of an ancient duplication. We discuss the implications of this observation, with emphasis on the origin and evolution of the MHC.

## MATERIALS AND METHODS

**Isolation of Mouse *Psmb7* cDNA Clones.** The Uni-ZAP XR cDNA library made from C57BL/6J mouse epididymal mRNA (20) was screened according to the standard procedure (21) using the human Z subunit cDNA clone (18) as a probe. The nucleotide sequences of positive cDNA clones were determined by the chain-termination method (22).

**Mapping of the Mouse *Psmb7* Gene.** The mouse *Psmb7* gene was mapped by PCR using the interspecific backcross mouse DNA panel (23) obtained from The Jackson Laboratory. PCR was performed essentially as described (24) using ≈70 ng of genomic DNA as a template. The conditions of PCR were 40 cycles of 40 sec at 94°C, 2 min at 58°C, 2 min at 72°C, and a final extension of 10 min at 72°C. The sequences of the sense and antisense primers were 5′-GAATGCTGTCTTGGAAGCG-3′ and 5′-TCTAGCTGGAAGCTTGAGTCC-3′, respectively. The sequence of the sense primer, which corresponds to nucleotides 67–85 in Fig. 1, is encoded by exon 2 of the mouse *Psmb7* gene (M.H. and M.K., unpublished data). The antisense primer was designed based on the intron 2 sequence (M.H. and M.K., unpublished data). The coding region of the *Mus spretus*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ABC, ATP-binding cassette; IFN-γ, interferon γ; LMP, low molecular mass polypeptide; MHC, major histocompatibility complex; TAP, transporter associated with antigen processing.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. D83585).

†To whom reprint requests should be addressed at: Department of Biochemistry, Hokkaido University School of Medicine, Kita 15, Nishi 7, Sapporo 060, Japan. e-mail: Kasahara@hucc.hokudai.ac.jp.

‖The officially approved gene symbol for the human Z subunit is *PSMB7*. Therefore, the corresponding mouse gene was designated *Psmb7* after consultation with The Mouse Genome Database, The Jackson Laboratory, Bar Harbor, ME.

*Psmb7* cDNA was amplified by reverse transcriptase–PCR using spleen cDNA as a template. The sequences of the sense and antisense primers used for this amplification were 5′-GCTGTGTCGGTGTTTCAGCCAC-3′ (nucleotides 11–32 in Fig. 1) and 5′-AGTCATCTAGAGTTATCCAACCA-3′ (nucleotides 852–874 in Fig. 1), respectively.

**Phylogenetic Tree Analysis.** The amino acid sequences of the proteasome subunits were aligned by using the CLUSTAL program of the LASERGENE software (DNAstar, Madison, WI). The *p*-distance was calculated for all pairs of sequences. The distance matrix thus obtained was used to construct a neighbor-joining tree (25). The reliability of the branching pattern was assessed by the bootstrap analysis. Phylogenetic analyses were performed by using the MEGA computer software package (26).

## RESULTS

**Primary Structure of the Mouse Proteasome Z Subunit.** The mouse epididymal cDNA library ($\approx 3 \times 10^5$ plaques) was screened using the human *PSMB7* cDNA clone (18) as a probe. Among the 15 positive clones, 8 clones thought to have full-length or near full-length inserts were chosen and partially sequenced. Only one type of cDNA sequence was identified. Fig. 1 shows the sequence of the cDNA clone with the longest insert. The longest open reading frame encodes a protein made up of 277 amino acids, of which the N-terminal 43 residues are thought to be removed by posttranslational processing (18). Thus, the mature form of the mouse Z subunit is predicted to contain 234 amino acids with a calculated molecular mass of 25,254 Da. Its calculated isoelectric point is 6.020. The mature forms of the mouse and human Z subunits show 97% amino acid sequence identity.

**Mapping of the Mouse *Psmb7* Gene by Interspecific Backcross Analysis.** The chromosomal localization of the *Psmb7* gene was determined by using progeny derived from the [(C57BL/6Ei × *M. spretus*)$F_1$ × *M. spretus*] interspecific backcross (23). To identify polymorphism informative for mapping, we amplified by PCR the coding region of the *M. spretus Psmb7* cDNA and determined its partial sequence. This showed that guanine at nucleotide position 85 (Fig. 1) is substituted to thymine in *M. spretus* (data not shown). We therefore designed the sense primer so that its 3′ terminus ends with guanine at position 85. The combination of this primer and the antisense primer located in intron 2 produced a single band of 175 bp in C57BL/6Ei mice but no bands in *M. spretus*. Analysis of 94 progeny showed that the mouse *Psmb7* gene maps to the paracentromeric region of chromosome 2 (Fig. 2). No recombination was observed between *Psmb7* and the DNA marker *D2Ucl21*, indicating that they lie within 3.8 cM at the 95% confidence level.

Inspection of the genes previously mapped to the paracentromeric region of mouse chromosome 2 (27) shows that this region and the MHC on chromosome 17 contain at least seven pairs of evolutionarily related genes thought to have originated by duplications (Table 1). This result indicates that the paracentromeric region of chromosome 2 and the MHC region on chromosome 17 most likely arose as a result of a chromosomal duplication. The MHC-encoded TAP molecule, which translocates peptides from the cytosol into the endoplasmic reticulum, is a member of the ABC (ATP-binding cassette) superfamily of transporters (31). Interestingly, the mouse gene encoding the transporter molecule of the ABC superfamily, designated *Abc2* or *D2H0S1474E*, maps to the paracentromeric region of chromosome 2 within 3.7 cM from the *Notch1* locus (28).

**The Duplication Involving the MHC Region Is Evident Also in the Human Genome.** One of us (18) showed recently that the human gene encoding the Z subunit (*PSMB7*) maps to the region 9q34.11–q34.12, which is homologous to the paracentromeric region of mouse chromosome 2 (27). We therefore examined whether the human MHC and its adjacent region

```
          +         +         +         +         +         +         +         +
caagATGGCGGCTGTGTCGGTGTTTCAGCCACCGGTCGGAGGCTTTTCTTTTGATAATTGTCGCAGGAATGCTGTCTTGG        80
   M  A  A  V  S  V  F  Q  P  P  V  G  G  F  S  F  D  N  C  R  R  N  A  V  L  E      -18
                  (Y) (A)
                                                    +1
AAGCGGATTTCGCAAAAAAGGGGTTCAAGCTCCCGAAAGCTCGGAAAACTGGCACTACCATCGCGGGGGTGGTGTATAAG       160
  A  D  F  A  K  K  G  F  K  L  P  K  A  R  K  T  G  T  T  I  A  G  V  V  Y  K        9
           (R)   (Y)         (V)

GATGGCATAGTTCTTGGAGCAGACACGAGAGCAACTGAAGGGATGGTTGTTGCTGACAAAAACTGTTCAAAAATTCACTT       240
D  G  I  V  L  G  A  D  T  R  A  T  E  G  M  V  V  A  D  K  N  C  S  K  I  H  F       36

CATATCTCCTAATATTTATTGCTGTGGTGCTGGGACAGCTGCAGACACAGACATGACAACCCAGCTTATTTCTTCCAACT       320
 I  S  P  N  I  Y  C  C  G  A  G  T  A  A  D  T  D  M  T  T  Q  L  I  S  S  N  L      63

TGGAGCTCCACTCTCTGACCACTGGCCGCCTCCCGAGAGTTGTTACAGCTAATCGGATGCTGAAGCAGATGCTCTTCAGG       400
  E  L  H  S  L  T  T  G  R  L  P  R  V  V  T  A  N  R  M  L  K  Q  M  L  F  R       89
              (S)

TATCAAGGTTACATTGGTGCAGCCCTAGTTTTGGGGGGAGTAGATGTTACTGGACCTCATCTCTACAGCATCTATCCTCA       480
Y  Q  G  Y  I  G  A  A  L  V  L  G  G  V  D  V  T  G  P  H  L  Y  S  I  Y  P  H      116

TGGATCAACTGATAAATTGCCTTATGTCACCATGGGTTCTGGCTCCTTGGCAGCAATGGCTGTGTTTGAAGATAAGTTTA       560
 G  S  T  D  K  L  P  Y  V  T  M  G  S  G  S  L  A  A  M  A  V  F  E  D  K  F  R      143
                                                          (L)

GGCCAGATATGGAGGAGGAAGAAGCCAAGAAGCTAGTGAGTGAGGCTATTGCAGCTGGCATCTTTAATGACTTGGGGTCT       640
  P  D  M  E  E  E  E  E  A  K  K  K  L  V  S  E  A  I  A  A  G  I  F  N  D  L  G  S      169
                        (N)

GGAAGCAACATTGATCTGTGTGTTATCAGCAAGAGCAAGCTGGACTTTCTTCGTCCATTCTCAGTGCCCAACAAGAAAGG       720
G  S  N  I  D  L  C  V  I  S  K  S  K  L  D  F  L  R  P  F  S  V  P  N  K  K  G      196
                           (N)                        (Y) (T)

GACCAGGCTTGGCCGGTACAGATGTGAGAAAGGCACTACCGCTGTCCTCACCGAGAAAGTTACCCCTCTGGAGATTGAGG       800
 T  R  L  G  R  Y  R  C  E  K  G  T  T  A  V  L  T  E  K  V  T  P  L  E  I  E  V      223
                                                       (I)

TGCTAGAAGAGACTGTTCAGACAATGGATACTTCGTGAatggtgctgtgggtggttggataactctagatgactcggggt       880
  L  E  E  T  V  Q  T  M  D  T  S  Ter                                                 234

gcataccacccctaccccaaccccattctaccccaaccagaaaacatgcctttggcaattgctcaattaaaataaataaa       960
taaaaagac(A)n                                                                         969
```

FIG. 1. Nucleotide and deduced amino acid sequence of the mouse Z subunit cDNA clone. The N-terminal amino acids thought to be removed by posttranslational processing are indicated by negative numbers and italicized. Four overlapping putative polyadenylylation signals (AATAAA) are underlined and italicized. The amino acid sequence of the human Z subunit is shown in parentheses only when it differs from the mouse sequence. (A)n, poly(A) tail.

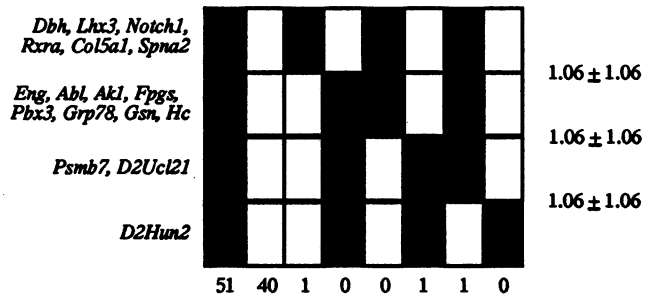


FIG. 2. Segregation patterns of *Psmb7* and neighboring loci in [(C57BL/6Ei × *M. spretus*)$F_1$ × *M. spretus*] (BSS type) interspecific backcross mice. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6Ei × *M. spretus*)$F_1$ parent. Black boxes indicate the inheritance of the C57BL/6Ei allele, and white boxes indicate the inheritance of the *M. spretus* allele. The number of offspring inheriting each type of chromosome is shown at the bottom of each column. Recombination frequencies between pairs of loci (percentage recombination with standard errors) are given on the right. For gene symbols, see Table 1. Gene names not listed in Table 1 can be obtained from the Mouse Genome Database maintained at The Jackson Laboratory.

(6p21.3) contain genes evolutionarily related to those mapped to the 9q34 region. The results summarized in Table 1 show that the regions 6p21.3 and 9q33–q34 contain at least nine sets of genes thought to have arisen by duplications. This result indicates that the duplication of the chromosomal segment involving the MHC took place before the divergence of the primate and rodent lineages.

**Phylogenetic Tree of the Proteasome β-Type Subunits.** Isolation of the Z subunit cDNA most likely completes the cloning of mammalian β-type proteasome subunits. To gain insights into the evolution of the β-type subunits, we constructed a neighbor-joining tree using an α-type subunit HC3 as an outgroup (Fig. 3). The tree confirms that the Z subunit is most closely related to the MECL1 subunit and shows that the three pairs of β-type subunits regulated reciprocally by IFN-γ (LMP7/X, LMP2/Y, and MECL1/Z) emerged by gene duplication from their respective ancestral subunits. These three pairs of β-type subunits, and the yeast subunits, PRE2, PRE3, and PUP1, which are X-, Y-, and Z-like, respectively, have two sequential threonine residues thought to be important for catalytic activity at their N termini (18). The tree shows that the subunits with the two sequential threonine residues are more closely related to each other than to the remaining β-type subunits that do not have the two threonine residues and hence might be catalytically inactive.

## DISCUSSION

Michaelson (34) suggested more than 10 years ago that mouse chromosomes 2 and 17 might contain homologous regions. This suggestion was based on the fact that chromosomes 2 and 17 contain not only the genes encoding the C1-SET members of the immunoglobulin (Ig) superfamily ($\beta_2$-microglobulin versus MHC molecules), histocompatibility antigens (H3 versus H2), complement components (C5 versus C3), and enzymes with some structural similarity at the three-dimensional level (adenylate kinase-1 versus phosphoglycerate kinase-2), but also the genes affecting the tail length (*Sd* versus *T*) and the immune response (*Ir2* versus *Ir1*). However, other chromosomes also contain genes encoding the C1-SET members of the Ig superfamily, histocompatibility antigens, and complement components. Therefore, Michaelson's proposal has been viewed with skepticism (35). Our present study demonstrates that the region surrounding the *Psmb7* locus is indeed duplicated on mouse chromosome 17 (Table 1). However, the duplicated segment appears to be confined to the region of mouse chromosome 2 with synteny to the human 9q33–q34 region. Thus, neither the $\beta_2$-microglobulin (*$B_2m$*) locus on chromosome 2 nor the *C3* locus on chromosome 17 is likely to fall within the duplicated segment. Examination of the mouse chromosome 2 loci, which were proposed by Michaelson (34)

Table 1. The region flanking the *Psmb7* (*PSMB7*) locus is duplicated in the MHC

| | Mouse chromosome 2 | | Human chromosome 9 | | | Mouse chromosome 17 (*H2* complex) | | Human chromosome 6 (*HLA* complex) | |
|---|---|---|---|---|---|---|---|---|---|
| | Gene symbol | Distance, cM | Gene symbol | Location | | Gene symbol | Distance, cM | Gene symbol | Location (6p21.3) |
| Retinoid X receptor α | *Rxra* | 15.4 | *RXRA* | 9q34 | Retinoid X receptor β | *Rxrb* | 18.49 | *RXRB* | Class II |
| Procollagen, type V, α1 | *Col5a1* | 15.4 | *COL5A1* | 9q34.2–q34.3 | Procollagen, type XI, α2 | *Col11a2* | 18.51 | *COL11A2* | Class II |
| ATP-binding cassette 2 | *Abc2* | 12.2 | *ABC2* | 9q34 | Transporter associated with antigen processing | *Tap1* | 18.60 | *TAP1* | Class II |
| | | | | | | *Tap2* | 18.62 | *TAP2* | |
| Proteasome β-type subunit Z | *Psmb7* | 17.6 | *PSMB7* | 9q34.11–q34.12 | Proteasome β-type subunits LMP2, LMP7 | *Lmp2* | 18.59 | *LMP2* | Class II |
| | | | | | | *Lmp7* | 18.61 | *LMP7* | |
| *Drosophila* Notch gene homolog 1 | *Notch1* | 15.4 | *NOTCH1* | 9q34.3 | Mammary tumor virus integration site 3 | *Int3* | 18.73 | *INT3* | Class III |
| Pre-B cell leukemia transcription factor 3 | | | *PBX3* | 9q33–q34 | PBX2 homeobox | | | *PBX2* | Class III |
| Tenascin | | | *HXB* | 9q32–q34 | Tenascin-like | | | *TN-X* | Class III |
| Complement C5 | *Hc* | 16.5 | *C5* | 9q33 | Complement C4 | *C4* | 18.8 | *C4A/C4B* | Class III |
| Glucose regulated protein 78 kDa | *Grp78* | 16.5 | *GRP78* | 9q33–q34.1 | Heat shock protein, 70 kDa | *Hsp70-1* | 18.94 | *HSPA1* | Class III |
| | | | | | | *Hsp70-3* | | *HSPA1L* | |
| | | | | | | *Hsp70t* | | *HSP70-hom* | |
| Valyl-tRNA synthetase 1 | | | *VARS1* | 9 | Valyl-tRNA synthetase 2 | | | *VARS2* | Class III |

With the exception of *Abc2*, the map positions of the mouse chromosome 2 loci were estimated by using the same interspecific backcross panel. The mouse *Abc2* gene was mapped by Allikmets *et al.* (28). Its position in centimorgan (cM) from the centromere was obtained from the Mouse Genome Database maintained by The Jackson Laboratory. The map positions of mouse chromosome 17 loci were taken from Ref. 29. References to the map positions of the human loci can be obtained from the Genome Data Base, a computerized data base of human linkage information maintained by the William H. Welch Medical Library of The Johns Hopkins University (Baltimore, MD). The map position of *PSMB7* was described by Hisamatsu *et al.* (18). Human *INT3* is a *Notch* related gene described by Sugaya *et al.* (30).

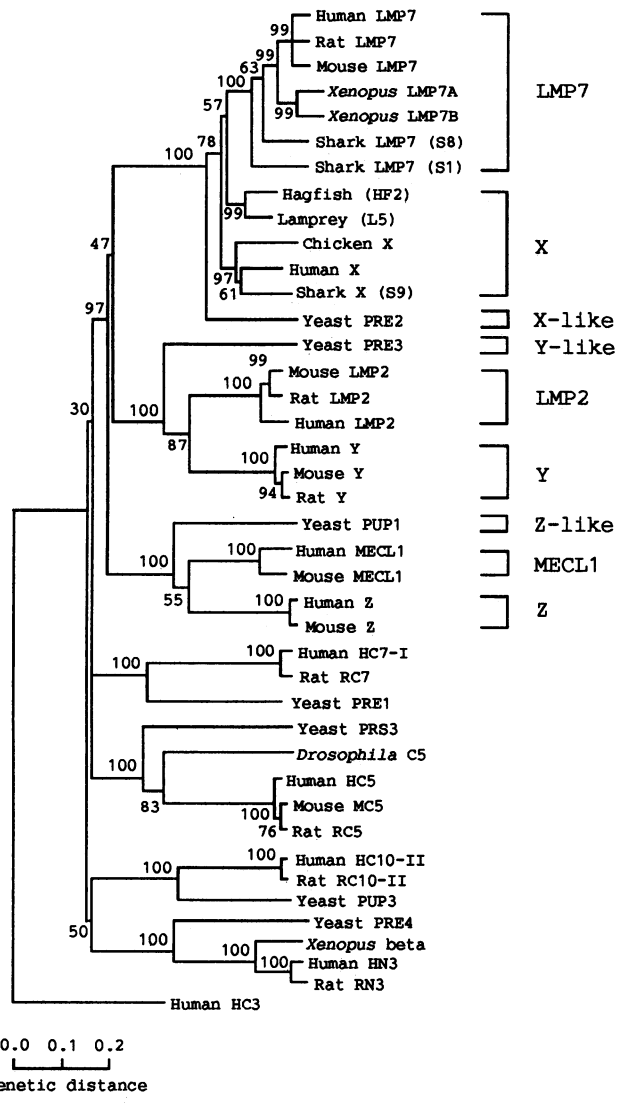


FIG. 3. Phylogenetic tree of proteasome β-type subunits. To assess the reliability of branching patterns, 500 bootstrap replications were performed. Numbers at the nodes indicate the bootstrap confidence level in percent. The sources of the sequences were as follows: shark LMP7 (clones S1 and S8), hagfish (clone HF2), lamprey (clone L5), and shark X (clone S9) (32); human Z (18); mouse Z (this study); human MECL1 (19); and mouse MECL1 (M.H., K.T., and M.K., unpublished data). References for the remaining sequences can be found in refs. 12 and 33.

to have their duplicated counterparts on chromosome 17, indicates that, with the exception of the adenylate kinase-1 locus, they all map outside the region with synteny to the human 9q33–q34 region (27). Hence, most, if not all, of the arguments that formed the basis of his proposal appear to be inappropriate. It is therefore truly surprising that Michaelson's proposal, though not in its original form, turned out to be partially correct.

The existence of the duplicated segment in both humans and mice (Table 1) indicates that the chromosomal duplication occurred before the divergence of the primate and rodent lineages. Two lines of evidence suggest that this duplication actually took place much earlier than mammalian radiation. First, the complement components, C5 and C4, encoded by the genes on mouse chromosomes 2 and 17, respectively, are structurally related proteins thought to have originated by gene duplication at or before the emergence of the bony fish (36). Jawless fishes represented by hagfishes and lampreys appear to lack a C5-like protein, with only the alternative pathway of complement activation (37). These findings suggest that the C4/C5 duplication, and hence the chromosomal duplication in question, took place at or before the emergence of the bony fish but after the emergence of jawless fishes. Second, the cartilaginous fish, thought to have emerged ≈400 million years ago, have both LMP7 and X subunits (ref. 32; Fig. 3), consistent with the fact that they have the MHC–peptide-based T-cell recognition system (38). In contrast, jawless fishes appear to have only the X-like subunits which cannot be unambiguously classified either as LMP7 or X (ref. 32; Fig. 3). These results suggest that the gene duplication that gave rise to LMP7 and X probably took place at or before the emergence of the cartilaginous fish but after the emergence of jawless fishes. The fact that the proteasome genes coding for the LMP and Z subunits are mapped to the different copies of the duplicated chromosomal segment indicates that the three pairs of β-type subunits arose simultaneously as a result of a single chromosomal duplication. Thus, the age of the chromosomal duplication estimated from the analysis of the proteasome β-type subunits is in good accord with that estimated from the analysis of the complement components. We conclude therefore that the chromosomal duplication involving the MHC region took place at an early stage of vertebrate evolution, probably at or before the emergence of the cartilaginous fish but after the emergence of jawless fishes. This conclusion enables us to predict that most, if not all, of the genes listed on the right half of Table 1 are probably still encoded within the MHC in various classes of vertebrates ranging from mammals to the cartilaginous fish. The fact that the *Hsp70* and *Lmp7* genes are linked to the *Xenopus* MHC (39, 40) is consistent with this prediction.

Primitive eukaryotes like yeast have two major forms of structurally related HSP70 molecules known as HSP70 and GRP78 (41). Therefore, the duplication of the genes encoding HSP70 and GRP78 must have predated the emergence of vertebrates. We presume that, when the chromosomal duplication took place, the *HSP70* and *GRP78* genes were still linked to each other, but that the former was lost subsequently from the chromosomal segment that eventually became the part of chromosome 2. Likewise, we presume that the latter was lost from chromosome 17. A partial amino acid sequence comparison of the TAP and ABC2 transporter molecules suggests that their ancestors might have diverged before the emergence of eukaryotes (42). Therefore, a situation analogous to that postulated for HSP70 might apply to these transporter genes. Accumulating evidence indicates that higher vertebrates have a large number of ABC transporters (43). Therefore, we cannot rule out the possibility that the *Abc2* gene was inserted by chance into the vicinity of the *Psmb7* gene after the chromosomal duplication.

The *Col5a1* and *Col11a2* genes (Table 1) encode highly similar polypeptides that are both components of type V/XI collagen (44), consistent with our proposal that these genes arose as a result of the chromosomal duplication involving the MHC region. However, the *Col11a2* gene is more closely related to the chromosome 3-encoded *Col11a1* gene than it is to the *Col5a1* gene. This observation suggests that, after the chromosomal duplication, the ancestral *Col11a* gene was duplicated to give rise to *Col11a1* and *Col11a2*, and that *Col11a1* was subsequently translocated to the region that became a part of mouse chromosome 3.

The budding yeast *Saccharomyces cerevisiae* has X-, Y-, and Z-like proteasome subunits (Fig. 3). Taking into account this observation and the foregoing discussion, we propose a hypothetical model that explains how the present-day MHC might have been assembled (Fig. 4). This model assumes that, along with the ancestors of the other genes listed in Table 1, the genes encoding the three β-type proteasome subunits (X-, Y-, and Z-like), HSP70, GRP78, a precursor of the complement components C3, C4, and C5, and probably the transporters of

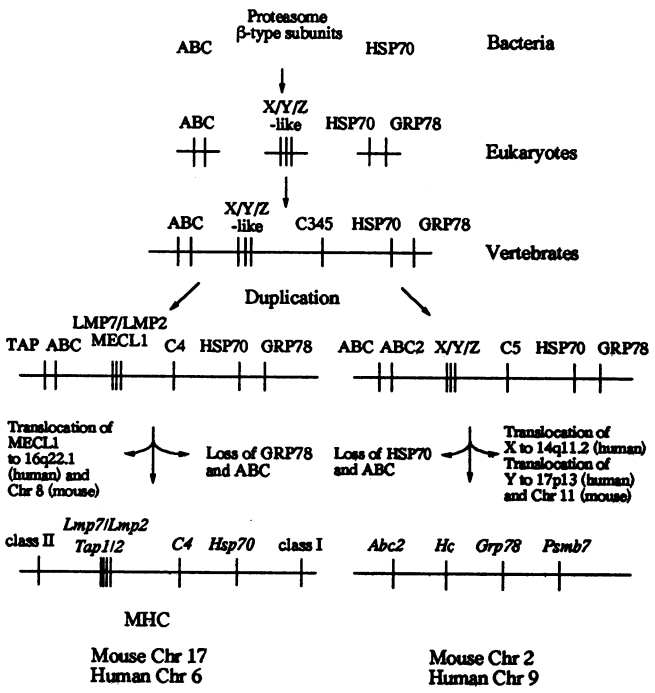


FIG. 4. A hypothetical model that explains how the present-day MHC might have been assembled. Transporters of the ABC superfamily (abbreviated as ABC), proteasome β-type subunits, and 70-kDa heat shock proteins (HSP70) are molecules of ancient origin that can be found even in bacteria. By the time eukaryotes emerged, an ancestral β-type subunit gene had undergone successive duplications to produce X-, Y-, and Z-like subunits. The *HSP70* gene was also duplicated to create two major forms of HSP70 molecules: HSP70 and GRP78. Before or at the emergence of vertebrates, the genes encoding the three β-type proteasome subunits (X-, Y-, and Z-like), HSP70, GRP78, a precursor of the complement components C3, C4, and C5 (abbreviated as C345), and probably the transporters of the ABC superfamily are thought to have been assembled into a single chromosomal region. This region was then duplicated and one of the duplicated region eventually became the present-day MHC. The other set of the duplicated region is now on chromosome 2 in mice and chromosome 9 in humans. Other events postulated to have taken place are indicated in the figure. The maps at the bottom are those of mouse chromosomes 17 and 2 and list only key loci. Like its human counterpart, the mouse gene encoding the MECL1 subunit is encoded outside the MHC and maps to chromosome 8 (M.H. and M.K., unpublished data).

the ABC superfamily were recruited to a single chromosomal region at or before the emergence of vertebrates. It further assumes that this region underwent duplication at an early stage of vertebrate evolution, with one of the duplicated region eventually becoming the present-day MHC. This model explains why the LMP2 and LMP7 subunits, which do not share an immediate common ancestor (Fig. 3), are encoded in tandem within the MHC (14) and why the mouse *Psmb7* gene is located on chromosome 2. However, not all of the mouse genes shown in Fig. 2 have their duplicated copies on chromosome 17, and the *H2* complex contains many genes with no apparent duplicated copies on chromosome 2. These observations suggest that the actual sequence of events was probably more complicated than that shown in Fig. 4.

A key question that remains to be answered is whether the duplication predated or postdated the emergence of MHC class I and II genes. The model depicted in Fig. 4 is compatible with both possibilities. If the latter is the case, class I and II genes must have been lost from one of the duplicated regions, presumably because of the constraint on the number of functionally important MHC genes (45, 46). On the other hand, if the former is the case, class I and II genes preassembled elsewhere might have invaded the region already equipped with the ancestors of the genes involved in antigen presentation, or the class I and II genes might have been assembled newly in one of the duplicated regions. These scenarios, which we think more likely, imply that jawless fishes, which presumably emerged before the chromosomal duplication, have neither class I nor class II genes. This is not inconsistent with the fact that attempts to isolate Ig, T-cell receptor, MHC class I, MHC class II, or recombination activating genes from jawless fishes have been invariably unsuccessful (reviewed in ref. 47).

It is interesting to note that the ancestral syntenic group identified in our study (Table 1) contains the genes coding for HSP70, proposed to have donated its peptide-binding domain (PBD) to MHC molecules (47–51). Although the recent nuclear magnetic resonance study (52) appears to argue against the possibility that the entire PBD of HSP70 molecules was donated *en bloc* to the nascent MHC molecule, the reported secondary structure is not inconsistent with the possibility that half of the PBD domain of HSP70 was donated to MHC molecules. Therefore, regardless of whether the duplication predated or postdated the emergence of class I and II genes, there is some attraction in the idea that the MHC class I and II molecules might have been assembled using the domains of the molecules encoded by the ancestral syntenic group. The *RAGE* (receptor for advanced glycosylation end products of proteins) gene encoded in the class III region of the human MHC (30) is a member of the Ig superfamily with one V- and two C-like domains (53). If this gene is also a member of the ancestral syntenic group, its C-like domain is a potential precursor of the membrane-proximal domain of MHC class I and II molecules.

Ohno (54) proposed that the genome underwent a few rounds of duplications early in vertebrate evolution. Closely linked pairs of duplicated genes, known as paralogous genes, are thought to represent the remnants of these duplications (reviewed in refs. 55 and 56). We suggest that the MHC region might have been created as a part of such genome-wide duplications.

It has been assumed generally that some MHC genes, in particular the class I, class II, *Lmp2*, *Lmp7*, and *Tap* genes, are linked to maintain certain allelic combinations that confer upon the organism the selective advantage (57). It is, however, difficult to envisage similar selective forces acting on the homologous genes located on mouse chromosome 2 or human chromosome 9. This reasoning leads us to the conclusion that the linkage of most, if not all, of the genes encoded within the MHC has been maintained without apparent selective pressures and possibly by chance for over ≈400 million years.

We thank Ms. Keiko Kohda for her technical assistance, Ms. Noriko Namerikawa for her secretarial help, and one anonymous reviewer for valuable comments. This work was supported by Grants-in-Aid for Scientific Research from The Ministry of Education, Science, Sports and Culture of Japan (to M.K., K.T., H.I. and T. Ikemura).

1. Goldberg, A. L. & Rock, K. L. (1992) *Nature (London)* **357,** 375–379.
2. Goldberg, A. L. (1995) *Science* **268,** 522–523.
3. Heemels, M.-T. & Ploegh, H. (1995) *Annu. Rev. Biochem.* **64,** 463–491.
4. Tanahashi, N., Tsurumi, C., Tamura, T. & Tanaka, K. (1993) *Enzyme Protein* **47,** 241–251.
5. Peters, J.-M. (1994) *Trends Biochem. Sci.* **19,** 377–382.
6. Rubin, D. M. & Finley, D. (1995) *Curr. Biol.* **5,** 854–858.
7. Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995) *Science* **268,** 533–539.
8. Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R. & Baumeister, W. (1995) *Science* **268,** 579–582.
9. Fenteany, G., Standaert, R. F., Lane, W. S., Choi, S., Corey, E. J. & Schreiber, S. L. (1995) *Science* **268,** 726–731.
10. Monaco, J. J. (1993) *Curr. Opin. Immunol.* **5,** 17–20.

11. Trowsdale, J. (1995) *Immunogenetics* **41,** 1–17.
12. Akiyama, K., Yokota, K., Kagawa, S., Shimbara, N., Tamura, T., Akioka, H., Nothwang, H. G., Noda, C., Tanaka, K. & Ichihara, A. (1994) *Science* **265,** 1231–1234.
13. Früh, K., Gossen, M., Wang, K., Bujard, H., Peterson, P. A. & Yang, Y. (1994) *EMBO J.* **13,** 3236–3244.
14. Belich, M., Glynne, R. J., Senger, G., Sheer, D. & Trowsdale, J. (1994) *Curr. Biol.* **4,** 769–776.
15. Driscoll, J., Brown, M. G., Finley, D. & Monaco, J. J. (1993) *Nature (London)* **365,** 262–264.
16. Gaczynska, M., Rock, K. L. & Goldberg, A. L. (1993) *Nature (London)* **365,** 264–267.
17. Gaczynska, M., Rock, K. L., Spies, T. & Goldberg, A. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9213–9217.
18. Hisamatsu, H., Shimbara, N., Saito, Y., Kristensen, P., Hendil, K. B., Fujiwara, T., Takahashi, E., Tanahashi, N., Tamura, T., Ichihara, A. & Tanaka, K. (1996) *J. Exp. Med.* **183,** 1807–1816.
19. Larsen, F., Solheim, J., Kristensen, T., Kolstø, A.-B. & Prydz, H. (1993) *Hum. Mol. Genet.* **2,** 1589–1595.
20. Mizuki, N. & Kasahara, M. (1992) *Mol. Cell. Endocrinol.* **89,** 25–32.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
23. Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J. & Birkenmeier, E. H. (1994) *Mamm. Genome* **5,** 253–274.
24. Kandil, E., Noguchi, M., Ishibashi, T. & Kasahara, M. (1995) *J. Immunol.* **154,** 5907–5918.
25. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
26. Kumar, S., Tamura, K. & Nei, M. (1993) *MEGA: Molecular Evolutionary Genetics Analysis* (Pennsylvania State Univ., University Park, PA).
27. Siracusa, L. D. & Abbott, C. M. (1994) *Mamm. Genome* **5** (Suppl.), S22–S39.
28. Allikmets, R., Gerrard, B., Glavac, D., Ravnik-Glavac, M., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., Modi, W. & Dean, M. (1995) *Mamm. Genome* **6,** 114–117.
29. Forejt, J., Artzt, K., Barlow, D. P., Hamvas, R. M. J., Lindahl, K. F., Lyon, M. F., Klein, J. & Silver, L. M. (1994) *Mamm. Genome* **5,** S238–S258.
30. Sugaya, K., Fukagawa, T., Matsumoto, K., Mita, K., Takahashi, E., Ando, A., Inoko, H. & Ikemura, T. (1994) *Genomics* **23,** 408–419.
31. Higgins, C. F. (1992) *Annu. Rev. Cell Biol.* **8,** 67–113.
32. Kandil, E., Namikawa, C., Nonaka, M., Greenberg, A. S., Flajnik, M. F., Ishibashi, T. & Kasahara, M. (1996) *J. Immunol.* **156,** 4245–4253.
33. Heinemeyer, W., Tröndle, N., Albrecht, G. & Wolf, D. H. (1994) *Biochemistry* **33,** 12229–12237.
34. Michaelson, J. (1983) *Immunogenetics* **17,** 219–260.
35. Klein, J. (1986) *Natural History of the Major Histocompatibility Complex* (Wiley, New York).
36. Nonaka, M. & Takahashi, M. (1992) *J. Immunol.* **148,** 3290–3295.
37. Nonaka, M., Fujii, T., Kaidoh, T., Natsuume-Sakai, S., Yamaguchi, N. & Takahashi, M. (1984) *J. Immunol.* **133,** 3242–3249.
38. Kasahara, M., Vazquez, M., Sato, K., McKinney, E. C. & Flajnik, M. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6688–6692.
39. Salter-Cid, L., Kasahara, M. & Flajnik, M. F. (1994) *Immunogenetics* **39,** 1–7.
40. Namikawa, C., Salter-Cid, L., Flajnik, M. F., Kato, Y., Nonaka, M. & Sasaki, M. (1995) *J. Immunol.* **155,** 1964–1971.
41. Boorstein, W. R., Ziegelhoffer, T. & Craig, E. A. (1994) *J. Mol. Evol.* **38,** 1–17.
42. Luciani, M. F., Denizot, F., Savary, S., Mattei, M. G. & Chimini, G. (1994) *Genomics* **21,** 150–159.
43. Dean, M. & Allikmets, R. (1995) *Curr. Opin. Genet. Dev.* **5,** 779–785.
44. Prockop, D. J. & Kivirikko, K. I. (1995) *Annu. Rev. Biochem.* **64,** 403–434.
45. Kobel, H. R. & Du Pasquier, L. (1986) *Trends Genet.* **2,** 310–315.
46. Sato, K., Flajnik, M. F., Du Pasquier, L., Katagiri, M. & Kasahara, M. (1993) *J. Immunol.* **150,** 2831–2843.
47. Kasahara, M., Flajnik, M. F., Ishibashi, T. & Natori, T. (1995) *Transplant Immunol.* **3,** 1–20.
48. Flajnik, M. F., Canel, C., Kramer, J. & Kasahara, M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 537–541.
49. Flajnik, M. F., Canel, C., Kramer, J. & Kasahara, M. (1991) *Immunogenetics* **33,** 295–300.
50. Salter-Cid, L. & Flajnik, M. F. (1995) *Crit. Rev. Immunol.* **15,** 31–75.
51. Rippmann, F., Taylor, W. R., Rothbard, J. B. & Green, N. M. (1991) *EMBO J.* **10,** 1053–1059.
52. Morshauser, R. C., Wang, H., Flynn, G. C. & Zuiderweg, E. R. P. (1995) *Biochemistry* **34,** 6261–6266.
53. Neeper, M., Schmidt, A. M., Brett, J., Yan, S. D., Wang, F., Pan, Y.-C. E., Elliston, K., Stern, D. & Shaw, A. (1992) *J. Biol. Chem.* **267,** 14998–15004.
54. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, New York).
55. Nadeau, J. H. & Kosowsky, M. (1991) *Mamm. Genome* **1,** S433–S460.
56. Lundin, L. G. (1993) *Genomics* **16,** 1–19.
57. Parham, P. (1990) *Nature (London)* **348,** 674–675.