The MHC-E Locus Is the Most Well Conserved of All Known Primate Class I Histocompatibility Genes

Leslie A. Knapp, Luis F. Cadavid and David I. Watkins

*J Immunol* 1998; 160:189-196; ;
http://www.jimmunol.org/content/160/1/189
The MHC-E Locus Is the Most Well Conserved of All Known Primate Class I Histocompatibility Genes1,2

Leslie A. Knapp,* Luis F. Cadavid,* and David I. Watkins†

The HLA-E locus is characterized by limited polymorphism and low levels of cell surface expression. However, the function of the products of this nonclassical MHC class I gene remains unknown. To evaluate the conservation of the MHC-E locus throughout anthropoid primate evolution, we identified the homologue of the HLA-E locus in six different New World monkey species. Full-length sequencing of MHC-E cDNAs in four unrelated cotton-top tamarins (Saguinus oedipus) revealed no evidence for polymorphism. Using the PCR, denaturing gradient gel electrophoresis, and direct sequencing, we also identified MHC-E alleles in five other New World monkey species, representing all extant platyrrhine families. In contrast to all other classical and nonclassical MHC class I genes in primates, the rate of synonymous nucleotide substitution is much greater than the rate of nonsynonymous nucleotide substitution within exons 2 and 3 encoding the peptide binding region (PBR) in MHC-E genes. The PBR of the MHC-E molecule, therefore, has evolved under purifying selective pressures, and the very unusual evolutionary history of this ancient gene provides further evidence that the products of the HLA-E locus serve a critical immunological function. Given the remarkable conservation of the PBR during primate evolution, this critical immunologic function is probably related to the peptide binding ability of the MHC-E protein. The Journal of Immunology, 1998, 160: 189–196.

MHC class I molecules are cell surface glycoproteins that play a critical role in immune response by binding peptides and presenting them to CTL (1). In humans, the products of the classical MHC class I (or class Ia) HLA-A, -B, and -C loci are expressed on the surface of nearly all cell types. These three class Ia genes are further characterized by an extensive degree of allelic variation in the second and third exons, which is primarily maintained by positive selection pressure (2). Nucleotide sequence data from alleles of the HLA-A, -B, and -C loci reveal that the number of nonsynonymous substitutions resulting in amino acid replacements in the peptide binding regions (PBR)4 is significantly greater than that in non-PBR sites in the same molecules (2). Contrastingly, human nonclassical MHC class I (class Ib) loci such as HLA-F and -G are less polymorphic than the MHC class Ia genes and typically exhibit more restricted tissue distribution (3). Alleles from the human nonclassical HLA-E locus also have higher rates of synonymous nucleotide substitutions in the putative Ag recognition sites than their classical MHC class I counterparts (4), suggesting that these molecules have been conserved to serve an important biologic function that remains as yet undescribed.

Interestingly, HLA-E shares a number of characteristics with the mouse nonclassical MHC class Ib locus Qa-1. Both genes encode molecules that exhibit extensive tissue distribution (5, 6), low cell surface expression (5, 6), and limited polymorphism (5). HLA-E and Qa-1 also share unusual structural similarities within the putative peptide binding groove. Specifically, both loci are characterized by an alanine at position 67 and serines at positions 143 and 147 (5), replacements that are not found in human or mouse classical MHC class I molecules. Recent studies of peptide binding capacity have also shown that, like Qa-1 molecules (7), HLA-E can bind peptides derived from class I signal sequences (8). Taken together, the unusual and highly convergent features shared by the Qa-1 and HLA-E genes lend further support for the hypothesis that these nonclassical MHC class I molecules perform an important immunologic function that is perhaps regulated by the expression of other MHC class I molecules.

Among the nonhuman primates, both classical and nonclassical MHC class I genes have been identified. In apes and Old World monkeys, homologues of the classical HLA-A and -B loci are polymorphic, with a high rate of nonsynonymous substitution in the PBR (9–21). Homologues of the nonclassical HLA-E and -F loci have also been identified in orangutans (22) and macaques (23, 25–25). Recently, MHC-E orthologues were identified in gorillas, chimpanzees, and bonobos (26) and in vervet (green) monkeys (23). Like their human counterparts, the nonclassical MHC-E and -F genes in apes and Old World monkeys exhibit conservation of the codons encoding amino acids in the PBR, particularly with respect to the MHC-E locus (24). In the cotton-top tamarin, a New World primate, the classical MHC class I genes exhibit reduced polymorphism and limited sequence variation (27, 28). Furthermore, comparisons with human, chimpanzee, and macaque MHC-F cDNA sequences have revealed that the cotton-top tamarin MHC-F gene has accumulated more nonsynonymous than synonymous differences in the PBR (25).

To evaluate the level of conservation of the MHC-E locus throughout anthropoid primate evolution, we identified the homologue of the HLA-E locus in the cotton-top tamarin (Saguinus oedipus) and five other New World monkey species. Together, the species studied represent all extant platyrrhine clades (Cebidae, Atelidae, and Pitheciidae) (29). Our comparisons of platyrrhine...
and other primate MHC-E locus alleles demonstrate that this genetic locus is very ancient, and the majority of amino acids in the putative PBR have been conserved throughout anthropoid primate evolution.

Materials and Methods

Animals studied

A total of 10 New World monkeys, representing all extant platyrrhine clades (Cebidae, Atelidae, and Pitheciidae) (29), were studied. Four unrelated cotton-top tamarins (Calibatrachus jacchus), and one owl monkey (Aotus trivirgatus) were analyzed using B lymphoblastoid cell lines. One of the cotton-top tamarins (So 100-75) and the owl monkey (At-666) were originally housed at the New England Regional Primate Research Center (Southborough, MA). A blood sample from the white-faced saki monkey (Aotus trivirgatus) and one brown-headed spider monkey (Ateles belzebuth) and one brown-headed spider monkey (A. fusciceps) were obtained from the Zoologico Santa Cruz (Mesitas del Colegio, Research Center (Madison, WI). A blood sample from the Roger Williams Park Zoo (Providence, RI). Blood samples from one long-haired spider monkey (Ateles belzebuth) and one brown-headed spider monkey (A. fusciceps) were obtained from the Zoologico Santa Cruz (Mesitas del Colegio, Colombia).

RNA extraction, cDNA synthesis, and PCR

Total RNA was extracted from peripheral blood or lymphoblastoid cell lines from all animals using RNaseasy (Qiagen, Inc., Chatsworth, CA). One microgram of the RNA was then used to synthesize cDNA with 1 μg of random hexamers (Promega, Madison, WI) and 50 U of SuperScript II reverse transcriptase (Life Technologies, Gaithersburg, MD) in 20 μl of 1× PCR buffer (GeneAmp, Perkin-Elmer, Foster City, CA), 5 mM MgCl₂, 20 U of RNase inhibitor (Promega), and 1 mU of each of the four deoxyribonucleotide triphosphates (dGTP, dATP, dTTP, and dCTP, GeneAmp, Perkin-Elmer). The cDNA synthesis was conducted in a Perkin-Elmer 9600 thermal cycler at room temperature for 10 min, at 42°C for 15 min, at 99°C for 5 min, and at 5°C for 5 min. PCR amplification was then conducted using 25 pmol of the primers listed in Table I. The PCR mixture also included 1× PCR buffer, 2 mM MgCl₂, 2.5 U of Taq polymerase (Perkin-Elmer), and 20 μl of cDNA for a final volume of 100 μl. The reactions were heated to 94°C for 2 min and then to cycling conditions consisting of 30 rounds of 60-sec denaturation at 94°C, 60-sec annealing at 60°C, and 90-sec extension at 72°C. The reaction was concluded with a 10-min final extension at 72°C.

Full-length Saoe-E allele sequencing

In preparation for direct sequencing, PCR products from the 5′LPX/ SOER1 and SOEF2/SOENDR2 amplifications of cDNA from animal S. oedipus 100-75 were electrophoresed on a 1% agarose gel. The appropriately sized bands were excised and purified using the Qiaex Gel Extraction Kit (Qiagen, Inc., Chatsworth, CA). Fifty to one hundred nanograms of each purified PCR product was sequenced using the Taq DyeDeoxy Ter-
of three different paralogous classical MHC class I loci. The mean dI genes and functional non-MHC type I IFN genes, IFN-a regions and non-receptor binding regions of mammalian IFN-N were calculated by the method of Nei and Jin (38), and statistical significance was determined using a test.

We also compared patterns of nucleotide substitution between MHC class I genes and functional non-MHC type I IFN genes, IFN- and IFN- genes. Using a method described by Nei and Gojobori (37), we estimated the proportion of synonymous nucleotide substitutions per synonymous site ( and the proportion of nonsynonymous nucleotide substitutions per nonsynonymous site ( in the PBR and non-PBR for pairwise comparisons of primate MHC-A, -B, -E, and -F orthologues. For comparative purposes, the same method was used to evaluate patterns of nucleotide substitution in the receptor binding regions and non-receptor binding regions of mammalian IFN- and IFN- alleles. As suggested by Hughes (39), the estimates of and were not corrected for multiple hits because the correction is not applicable to short domains when the nucleotide sequences are distantly related.

MHC-E nomenclature

Official locus and allele names were given in accordance with the rules designated by the nonhuman primate MHC nomenclature committee (40). New sequences have been deposited in GenBank (AF004918 through AF004925).

Results

The MHC-E locus is conserved, but not polymorphic, in cotton-top tamarins

To determine the degree of conservation of the MHC-E locus in primates, we designed primers to amplify MHC-E in cotton-top tamarins (Table IA). With the 5'LPXI/SOER1 and SOEF2/SOENDR2 primers, we PCR-amplified two overlapping cDNA fragments representing the full-length cotton-top tamarin MHC-E gene (Fig. IA). Direct sequencing of MHC-E cDNAs from four unrelated tamarins revealed that the sequence was identical in all individuals.

The putative PBR of MHC-E cDNAs are conserved in cotton-top tamarins

The predicted amino acid sequence of the cotton-top tamarin cDNA, Saoe-E*01, is shown in Figure 2. Interestingly, the Saoe-E*01 sequence shares unique characteristics with the HLA-E alleles. For example, like the MHC-E sequences described for humans and the orangutan, Saoe-E*01 contains a deletion in exon 7 that may result in a shortened cytoplasmic domain. Surprisingly, this same deletion has not been observed in the MHC-E alleles of macaques. However, both the rhesus macaque allele, Mafa-E*05, and the cotton-top tamarin, Saoe-E*01, have a three-nucleotide deletion in the transmembrane region that is absent in HLA-E, Popy-E, and Mafa-E.

Alignment of the predicted amino acid sequence of Saoe-E*01 and MHC-E alleles from Old World monkeys, apes, and humans also revealed that the MHC-E locus has been well conserved in anthropoid primates. A comparison between the PBR of MHC-E sequences in the orangutan and cotton-top tamarin revealed just five replacements within the 53 possible positions (Fig. 2). These five residues (positions 65, 73, 82, 149, and 170) form portions of the a-helixes and include one TCR contact residue (position 65) (41) and one C peptide-binding pocket residue (position 73) (42).

Consistent with previous analyses of patterns of synonymous substitutions within the putative PBR of MHC-E alleles in other primates, comparisons between Saoe-E*01 and HLA-E*01 demonstrate that the number of synonymous (d) nucleotide substitutions was significantly greater than the number of nonsynonymous (d) substutions (Table II). The d values, in changes per 100 synonymous sites, and d values, in changes per 100 nonsynonymous sites, between Saoe-E*01 and HLA-E*01 in the PBR are 33.0 ± 11.0 and 6.7 ± 2.3, respectively. In non-PBR sites the values for d are 2 to 5 times greater than the d values in the same region. Analysis of the PBR sites of classical MHC class I molecules encoded by the MHC-B or Saoe-G loci demonstrates that d is significantly greater than d (Table II). Thus, in contrast to the low d/d ratio reported for PBR vs non-PBR sites in homologues of the classical MHC class I loci, the d/d ratios for PBR sites in MHC-E alleles are extraordinarily high. The high d/d ratios reported for the PBR of MHC-E alleles, however, is not observed when the PBR of another nonclassical MHC class I locus is compared in humans, rhesus macaques, and cotton-top tamarins (HLA-F, Mafa-F, and Saoe-F, respectively) is analyzed. Instead, the d/d ratios for the comparison between human and rhesus macaque PBR and between human and cotton-top PBR are 1.1 and 0.38, respectively (Table II).

DGGE analysis of the MHC-E locus in New World monkeys: the MHC-E locus is conserved and polymorphic in some platyrhines

To identify MHC-E genes in other New World monkey species, a second set of PCR primers was designed to amplify the most polymorphic portions of exons 2, 3, and 4 of MHC-E genes (Fig. 1B and Table IB). Following reverse transcription-PCR amplification with the EA1MID and A3MID+GC-clamped primers, we analyzed our 510-bp PCR product using DGGE. Because each allele at a particular locus is the same length, individual alleles cannot be discriminated on agarose gels. Instead, alleles can be separated on a denaturing gradient gel according to differences in sequence composition. When DNA fragments are electrophoresed through an increasing gradient of denaturants composed of urea and formamide, the two DNA strands of each allele dissociate at different rates because of sequence-specific differences in denaturation. As a consequence, each allele migrates at a unique rate through the gel. To maximize the separation of alleles that differ by only one or two nucleotide substitutions, a G+C-rich sequence (GC-clamp) is attached to the PCR product during amplification (31). The optimum denaturing gradient for each PCR product can then be determined by analyzing DNA fragments with a perpendicularly denaturing gradient according to a method described by Myers and co-workers (30). Once the optimum gradient is identified, PCR products can be electrophoresed on an appropriate parallel denaturing gradient (30). Perpendicular DGGE analysis of MHC class I genes in cotton-top tamarins using perpendicular DGGE indicated that the optimal denaturing gradient for separation of our reverse transcription-PCR products was 50 to 65%.
FIGURE 2. Predicted amino acid sequence of New World monkey MHC-E molecules compared with that of HLA-E*0101. Identity with Saoe-E*01 is shown by dashes, while differences are given by letter substitutions. Periods denote gaps introduced to maximize sequence alignment. Asterisks denote amino acids that are located in the putative PBR. New World monkey alleles are abbreviated according to the rules designated by the nonhuman primate MHC nomenclature committee (40): Saoe, Saguinus oedipus; Caja, Callithrix jacchus; Aotr, Aotus trivirgatus; Pipi, Pithecia pithecia; Atbe, Ateles belzebuth; and Atfu, Ateles fusciceps. Previously described primate MHC-E sequences have been included for comparison: HLA-E*0101 (43), HLA-E*0102 (6), Popy-E*01 (22), Mamu-E*05 (21), Mafa-E*01 and -E*02 (24), Mamu-F*01 (25), Saoe-G*06 and -G*08, and Saoe-F*01 (28).
Parallel DGGE analysis of MHC-E locus cDNAs in New World monkeys on a 50 to 65% denaturing gradient revealed one or two MHC-E alleles in each species. Figure 3 shows the pattern of homoduplexes and, in two cases, heteroduplexes identified in our DGGE analysis of 10 New World monkeys. In the common marmoset and the brown-headed spider monkey, two different alleles were identified for each individual. For these samples, heteroduplex bands were also formed during the PCR when complementary strands from the two alleles hybridized. The heteroduplex bands exhibited less fluorescence and migrated at higher positions in the gradient gel. Only one MHC-E locus sequence was identified in our four unrelated cotton-top tamarins, the owl monkey, the long-haired spider monkey, and the white-faced saki monkey.

Following DGGE and direct sequencing, we compared MHC-E cDNAs from all six New World monkey species with HLA-E alleles. Analysis of predicted amino acid sequences from the middle of exon 2 through the middle of exon 4 revealed that the MHC-E locus is highly conserved in platyrrhines, particularly within the putative PBR (Fig. 2). In the Callitrichidae, both cotton-top tamarins and common marmosets possessed nearly identical MHC-E cDNAs within the region studied. Specifically, Saoe-E*01 and the Caja-E*01 and -E*02 alleles differed by one and two amino acids, respectively. Furthermore, all three cDNAs were identical in the 36 peptide binding sites examined in the DGGE analysis. The majority of residues in the PBR of the owl monkey (Aotr-E*01) and spider monkeys (Aitu-E*01, E*02, and Aite-E*01) were also conserved. However, it is noteworthy that these four sequences are distinguished by an alanine, rather than a serine, at position 147. The white-faced saki monkey cDNA, Piti-E*01, shared more common substitutions with the MHC-E sequences of humans, apes, and Old World monkeys than with the New World monkeys included in this study.

A gene tree of the full-length Saoe-E*01 cDNA and MHC class I alleles from Old World monkeys, apes, and humans was performed to evaluate the relationship between Saoe-E*01 and other primate MHC class I genes. Figure 4 demonstrates that Saoe-E*01 is orthologous to HLA-E. A second analysis of the new platyrrhine MHC-E alleles clusters all the new E locus alleles together based upon sequence from the middle of exon 2 through the middle of exon 4 (Fig. 5).

The evolutionary history of the MHC-E locus in primates is unlike the history of classical MHC class I genes

To examine the extent to which MHC-E genes have been subject to purifying selection, we compared patterns of nucleotide substitu-
mature protein. The patterns of nucleotide substitution were examined by estimating the proportion of synonymous nucleotide substitution per synonymous site and the proportion of nonsynonymous nucleotide substitution per nonsynonymous site, $p_s$ and $p_N$, respectively (37). In the case of MHC-A and MHC-B loci, mean $p_s$ exceeds mean $p_N$ in non-PBR sites. This pattern is seen in gene regions where purifying selection acts to eliminate a substantial portion of nonsynonymous mutations. On the other hand, the $p_s/p_N$ ratio is very low in the PBR sites of MHC-A and MHC-B alleles, where positive selection maintains diversity. The MHC-F locus in primates exhibits a very low $p_s/p_N$ ratio in the PBR and an extraordinarily high $p_s/p_N$ ratio in the non-PBR sites. In contrast to the pattern of classical MHC class I genes and MHC-F, a high $p_s/p_N$ ratio is found in both PBR and non-PBR sites of MHC-E alleles. Notably, a high $p_s/p_N$ ratio is also found in the binding sites of type I IFN-α and IFN-β genes, where strong purifying selection acts to conserve the receptor binding region (38).

**Discussion**

To evaluate the conservation of the MHC-E locus in anthropoid primates, we identified the homologue of the HLA-E locus in the cotton-top tamarin (Saguinus oedipus) and five other New World monkey species. Previously, MHC-E genes had been identified in Old World monkeys (21, 23, 24) and apes (22, 26). Sequence analysis revealed that Saoe-E*01 was very similar to HLA-E, particularly within the codons encoding the putative PBR. Comparisons of $d_S$ and $d_N$ of the nucleotides encoding the PBR and the nucleotides encoding non-PBR sites demonstrated a unique pattern of amino acid replacement relative to that described for other MHC class I genes. Our analyses revealed that the MHC-E locus has evolved under purifying selective pressures and suggests that the products of the HLA-E locus serve an important biologic function.

Recent studies of diverse human populations describe the presence of two HLA-E alleles that differ by a single base pair substitution in exon 3, resulting in an amino acid change from arginine (R) to glycine (G) at position 107 (26, 44), which is found in the loop between the β-pleated sheet and the α-helix segments of the peptide binding groove (45). In Hutterites, African-Americans, and Hispanics, these two HLA-E alleles are found in nearly equal frequency, and it has been suggested that functional differences between the alleles may exist (44). Among all nonhuman primates studied to date, only glycine is found at position 107. However, the number of nonhuman primate samples that have been analyzed remains too small to draw any conclusions regarding functional differences between alleles within the same species. Nevertheless, our demonstration that the PBR is well conserved among primates suggests that the MHC-E molecule serves a critical immunologic function that is probably related to peptide binding. This unusual conservation of the PBR of MHC-E combined with its limited polymorphism may be related to the role of MHC-E molecules in the regulation of NK cell activity.
Previously, molecular cloning and expression studies of MHC genes in the cotton-top tamarin have shown that the limited polymorphism and variability in the classical MHC class I genes of this species are due, at least in part, to rapid turnover of MHC class I loci. There is evidence for duplication of expressed MHC class I genes (27, 46) and recent inactivation of other MHC class I genes (47). Additionally, a small founder population may have contributed to fixation of newly arisen MHC class I genes (28). Even the HLA-F orthologue in the cotton-top tamarin has accumulated a large number of nonsynonymous substitutions, suggesting that this locus is not as well conserved in cotton-top tamarins as it is in Old World monkeys, apes, and humans. Despite the phylogenetic distance between humans and tamarins and the high rate of gene turnover in the callitrichids, the cotton-top tamarin MHC-E locus has remained remarkably intact and highly conserved. Thus, it is likely that purifying selective pressures have maintained the MHC-E locus throughout anthropoid primate evolution.

Interestingly, the nature and intensity of selective forces acting on the codons encoding the PBR are not identical in all MHC class I genes. While the accumulation of nonsynonymous nucleotide substitutions and the high degree of polymorphism in the codons encoding the PBR of classical MHC class I genes appear to have been driven by positive or diversifying selection (2), the high rate of synonymous nucleotide substitutions relative to the low rate of nonsynonymous nucleotide substitutions in the PBR of primate MHC-E alleles appears to be the consequence of purifying selection. Thus, the high rate of nonsynonymous nucleotide substitution in the PBR that is characteristic of most MHC class I molecules is absent in primate MHC-E genes. Instead, most of the variability reported for human, ape, and monkey MHC-E locus alleles is found in regions outside the PBR.

Although unusual with respect to the pattern of primate MHC class I gene evolution, the evolutionary history of MHC-E genes is remarkably similar to the evolution of non-MHC genes that remain highly conserved due to important functional constraints. Purifying selection has eliminated disadvantageous nonsynonymous substitutions in the β-strand-encoding region of primate β2m (48). The highly conserved reactive center region of α1-proteinase inhibitors, which control neutrophil elastase activity within the circulation, has also been maintained through purifying selection in primates (49). Additionally, a parallel can be drawn between the primate MHC-E molecules and the IFN-α and IFN-β genes. Molecular studies of mammalian type I IFN have demonstrated that IFN-α and IFN-β genes have been subject to purifying selection, particularly within the putative receptor binding domains. Strong functional constraints associated with IFN receptor binding, and hence powerful selective pressure to eliminate nonsynonymous mutations in regions involved in receptor binding, have dramatically shaped the evolution of the IFN-α and -β genes in mammals. Contrastingly, the pattern of nonsynonymous nucleotide substitutions observed within the IFN-ω genes suggests that positive Darwinian selection has acted to diversify mammalian IFN-ω (39). The pattern of diversification, common among IFN-ω genes in mammals and among classical MHC class I genes in primates, is virtually absent in mammalian IFN-α and -β genes and in primate MHC-E genes, where strong purifying selective pressures have conserved the critical receptor and peptide binding regions.

Although the precise function of HLA-E molecules remains undescribed, the extraordinary conservation of the codons encoding the PBR of this locus throughout anthropoid primate evolution suggests that the products of the MHC-E locus play an important role in the primate immune response. Moreover, the remarkable similarities between HLA-E and the murine nonclassical MHC class Ib gene, Qa-1, offer clues to the function of the products of the HLA-E locus. Both Qa-1 and HLA-E loci exhibit reduced polymorphism, shared residues in highly invariant positions within the PBR (5), similar patterns of gene expression (5, 6), and the ability to bind signal sequence-derived peptides (5, 8). In the case of Qa-1, the molecules are known to present leader peptides from murine H-2D class I proteins to alloreactive T cells (7). Importantly, MHC-E transcripts have also been found in placental tissue in humans (50, 51) and macaques (24). Taken together, these findings suggest that further insight into the function of HLA-E genes may come from studies of primate placental and extraplacental membranes.

It appears that evolutionary selective pressures have conserved amino acids within the PBR of primate MHC-E molecules. Despite the seemingly numerous recombinations, inactivations, and duplications that have occurred at adjacent MHC class I loci throughout primate evolution, primate MHC-E genes have remained remarkably
preserved over approximately 55 million yr (52). Our MHC-E sequence data from New World monkeys and comparisons with other primate MHC genes suggest that the MHC-E locus is the most well conserved of all known primate histocompatibility genes and, the remarkable evolutionary history of the codons encoding the PBR suggests an important immunologic role for HLA-E.

Acknowledgments

We thank Haydy Monsalve, Francisco Ruiz, and Anne Savage for help obtaining blood samples.

References