Although Divergent in Residues of the Peptide Binding Site, Conserved Chimpanzee Patr-AL and Polymorphic Human HLA-A*02 Have Overlapping Peptide-Binding Repertoires


*J Immunol* 2011; 186:1575-1588; Prepublished online 5 January 2011;
doi: 10.4049/jimmunol.1002990
http://www.jimmunol.org/content/186/3/1575

Supplementary Material
http://www.jimmunol.org/content/suppl/2011/01/05/jimmunol.1002990.DC1

References
This article cites 87 articles, 24 of which you can access for free at:
http://www.jimmunol.org/content/186/3/1575.full#ref-list-1

Subscription
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2011 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Although Divergent in Residues of the Peptide Binding Site, Conserved Chimpanzee Patr-AL and Polymorphic Human HLA-A*02 Have Overlapping Peptide-Binding Repertoires

Michael Gleimer,*‡ Angela R. Wahl,‡ Heather D. Hickman,§ Laurent Abi-Rached,‡ Paul J. Norman,‡ Lisbeth A. Guethlein,‡ John A. Hammond,‡ Monia Draghi,‡ Erin J. Adams,§ Sean Juo,§ Roxana Jalili,§ Baback Gharizadeh,§ Mostafa Ronaghi,§ K. Christopher Garcia,* William H. Hildebrand,‡ and Peter Parham‡

Patr-AL is an expressed, non-polymorphic MHC class I gene carried by ∼50% of chimpanzee MHC haplotypes. Comparing Patr-AL* and Patr-AL** haplotypes showed Patr-AL defines a unique 125-kb genomic block flanked by blocks containing classical Patr-A and pseudogene Patr-H. Orthologous to Patr-AL are polymorphic orangutan Popy-A and the S’ part of human pseudogene HLA-Y, carried by ∼10% of HLA haplotypes. Thus, the AL gene alternatively evolved in these closely related species to become classical, nonclassical, and nonfunctional. Although differing by 30 aa substitutions in the peptide-binding α2 bounds, Patr-AL and HLA-A*0201 bind overlapping repertoires of peptides; the overlap being comparable with that between the A*0201 and A*0207 subtypes differing by one substitution. Patr-AL thus has the A02 supertypic peptide-binding specificity. Patr-AL and HLA-A*0201 have similar three-dimensional structures, binding peptides in similar conformation. Although comparable in size and shape, the B and F specificity pockets of Patr-AL and HLA-A*0201 differ in both their constituent residues and contacts with peptide anchors. Uniquely shared by Patr-AL, HLA-A*0201, and other members of the A02 supertype are the absence of serine at position 9 in the B pocket and the presence of tyrosine at position 116 in the F pocket. Distinguishing Patr-AL from HLA-A*02 is an unusually electropositive upper face on the α2 helix. Stimulating PBMCs from Patr-AL chimpanzees with B cells expressing Patr-AL produced potent alloreactive CD8 T cells with specificity for Patr-AL and no cross-reactivity toward other MHC class I molecules, including HLA-A*02. In contrast, PBMCs from Patr-AL chimpanzees are tolerant of Patr-AL. The Journal of Immunology, 2011, 186: 1575–1588.
human ancestors 6–10 mya (23). Whereas the other human and chimpanzee MHC class I genes are present on all MHC haplotypes, Patr-AL is present only on ~50% of chimpanzee MHC haplotypes (23); an even distribution suggestive of a balancing selection that maintains MHC haplotypes with and without Patr-AL. Such selection is a general feature of MHC variation (25). As a consequence of this distribution, a majority of chimpanzees have Patr-AL, but, importantly, a significant minority does not. Indeed, the chimpanzee MHC haplotype sequenced by Anzai et al. (15) has Patr-A, -B, -C, -E, -F, and -G, but lacks Patr-AL. Accordingly, the first objective of our investigation was to define the location and environment of the Patr-AL gene in the chimpanzee MHC, thus defining the genes and genomic region that humans have lost.

In previous analysis, we showed that Patr-AL exhibits modest polymorphism and in this regard resembles the nonclassical HLA-E, -F, and -G genes (23). From a functional perspective, however, Patr-AL can be considered as a gene having two balanced alleles with dramatic functional difference: one makes a functional protein, the other (gene absence) does not. Although related to MHC-A, Patr-AL differs from Patr-A and HLA-A by >40 aa substitutions, including 30 in the α1 and α2 domains that form the peptide binding site. Thus, a second objective for this investigation was to determine if Patr-AL has peptide-binding function and with what specificity and structural nuance. Our third and final objective was to determine if Patr-AL has the potential to function as a histocompatibility Ag and be recognized by TCRs.

Materials and Methods

DNA sequencing and analysis

The CHORI-251 bacterial artificial chromosome (BAC) library (Children’s Hospital of Oakland Research Institute, Oakland, CA) was screened with a Patr-AL cDNA probe. Positive clones were end-sequence-determined to define their relative location in the chimpanzee MHC, and clone 639P10 was selected based on its coverage of the entire region containing Patr-AL. A shotgun library with insert sizes of 1 to 2 kb was made using the TOPO Shotgun kit (Invitrogen, Carlsbad, CA). Two thousand clones were sequenced by Sanger sequencing. Finishing was performed at the Stanford Genome Technology Center (Palo Alto, CA) using pyrosequencing on the 454 platform, as previously described (26). The sequence of the region was extended by primer-walking that generated an additional 8 kb of sequence from the overlapping BAC clone 243L17. Sequences were assembled using Staden 1.6.0 (27). The complete sequence has been deposited in GenBank under accession number HM629932. In addition, 5’ flanking sequences of Gogo-A*0401, Gogo-A*0501, and intron 3 of Gogo-A*0501 were amplified, cloned, and sequenced from the gorilla Radi; similarly, the 5’ flanking sequence of the human HLA-Y pseudogene was obtained from the B cell line WON-M. The sequences have been deposited in GenBank under accession numbers HM629928 (Gogo-A*0501 5’ flanking sequence), HM629929 (Gogo-A*0401 5’ flanking sequence), HM629930 (HLA-Y 5’ flanking sequence), and HM629931 (intron 3 of Gogo-A*0501). The primers A210_SENSE (5’-GGGCCATGATCCATCAGCTAGC-3’) and A210_ANTl (5’-GACGGCGATCCGCAAGGC-3’), annealing with the 5’ flanking region of A210, respectively, were used for amplification of the 5’ region, and the primers G5_X3_SENSE (5’-GGGAGTGGCTC CGCGATAGA-3’) and G5_X4_ANTl (5’-CCTCATGGTCAGAACAGAG TCTGG-3’), annealing to sites in exons 3 and 4, were used for the amplification of intron 3.

Large-scale sequence alignment was performed and visualized using the VISTA program (http://www.-gsd.lbl.gov/vista/). Local alignments were performed using MAFFT (28). Phylogenetic analysis for the T, W, and K pseudogenes was performed using MEGA3 (29). Analysis of divergence time for the T, W, and K pseudogenes was performed using the MCMCtree program implemented in the Phylogenetic Analysis by Maximum Likelihood package [PAML (30)], calibrated by fossil-based speciation time estimates for Mucaca mulatta (23–33 mya), Gorilla gorilla (10.5–12 mya), and Pan troglodytes (7–9 mya) (31, 32).

We assumed that the ancestor of Patr-AL and HLA-Y was not fixed at the time of the human/chimpanzee divergence because neither is fixed in the modern species, and the deletions in the Patr-AL and HLA-Y haplotypes share the same breakpoints. Simulations tested if both Patr-AL and Patr-AL” haplotypes could be retained under neutrality (i.e., in the absence of selection to keep both of them in the population) from the time of the human/chimpanzee divergence until present. The simulations recorded the allele-frequency change per generation and stopped when one haplotype was lost. Forward-time population simulation was performed using simaPOP (33), assuming a generation time of 15 y, N, = 30,000 (34), random mating, and starting haplotype frequencies of 50%. The simulations were conservative because reduction in population size, generation time, or unequal starting frequencies would increase the probability of losing one haplotype, as would selection for one haplotype.

Phylogenetic analysis of MHC-A, MHC-H, and MHC-A–related gene sequences

MHC-A, MHC-H, and MHC-A–related gene sequences were aligned using MAFFT (28) and manual correction of the resulting alignments. The aligned sequences were then investigated for the presence of recombinant segments using a combination of domain-by-domain phylogenetic analyses and recombination detection methods, as implemented in the recombination detection program, RDP (35). Neighbor-joining (NJ) analyses were conducted with MEGA4 (29) using the Tamura–Nei method with 500 replicates, windows including one to eight segments (introns and/or exons) were used. To confirm the results of the recombination analysis, phylogenetic analyses were conducted on 10 data sets representing the full MHC class I gene sequence (with some overlap) with three methods: maximum likelihood (ML), NJ, and maximum parsimony (MP). NJ analyses were performed as indicated above. PAUP*4.0b10 (36), with the tree biscorrection branch swapping algorithm, was used for MP analyses with 500 replicates and a heuristic search. ML analyses were performed with RAxML7 (37) under the GTR+CAT model with 500 replicates (rapid bootstrapping).

For the full-genome analysis, the final set of sequences was obtained by excluding the recombinant sequences, the pseudogenes, as well as exon 4, encoding the α1 domain, and the noncoding regions in the 5’ and 3’ of the gene. To complement this analysis and extend the set of sequences that could be included, we analyzed a smaller gene segment beginning 300 bp upstream of the ATG start codon and ending in exon 2. Phylogenetic analyses were conducted with three methods, as indicated above.

Ancestral sequence reconstructions for the peptide-binding domain were performed with the CODEML program of the PAML package (30) using the marginal reconstruction approach and the M0 model. The tree topologies used for these reconstructions were obtained using the ML approach described above.

Characterization of Patr-AL and HLA-A peptide pools

Soluble Patr-AL, HLA-A*0201, and HLA-A*0207 were secreted by transfected 721.221 cells (hereafter referred to as 221 cells) grown for 60 d in a Unisyn CP2500 bioreactor (Biostev International, Minneapolis, MN) as described (38). Soluble class I protein was affinity-purified and the peptides acid-eluted. Peptide pools were subjected to Edman sequencing on a model 492A pulsed liquid phase protein sequencer (PerkinElmer, Wallingford, MA), with cysteine un-derivatized, to determine the peptide motifs. Alternatively, pools were fractionated by reverse-phase HPLC and fractions analyzed using a Q-Star QTOF mass spectrometer (PerSeptive Scieq, Foster City, CA). Sequence assignment was performed using MASCOT (Matrix Science, London, U.K.).

Combinatorial prediction of the overlap between the Patr-AL and HLA-A*02 peptide pools was performed assuming average 40% overlap between the Patr-AL pool and individual HLA-A*02 allele pools and 60% overlap between HLA-A*02 allele pools. The binominal expansion describing the percentage of peptides unique to the Patr-AL pool converges to 1 – (0.40/6) = 33.3%, indicating 66.6% overlap between the Patr-AL pool and the collective HLA-A*02 pool.

Comparison of peptide-binding profiles

From the sequences of nonamer binding peptides, either deposited for 16 HLA-A alleles in the SYFPEITHI database (39) or obtained in this study for Patr-AL, A*0201, and A*0207, we compiled 19 allotype-specific data sets. Only peptides originating from unfractionated cells and alleles having five or more peptides defined were included in the analysis. For each data set, peptide scoring matrices (representing the peptide-binding profile) were generated using the PROPHET program of the EMBOS package (40) (Gribskov scoring scheme, gap open, and extension penalties of 500). The matrices were built using the complete nonamer sequences, thereby giving the unbiased estimation of the residue preference at each of the matrix positions. The 19 matrices were then used to score the peptides of the 19 peptide data sets using the PROPHET program of the EMBOS package (40). Each resulting matrix:peptide score gives a relative measure of the
ability of that particular peptide to bind to the MHC allele represented by that particular matrix. We reasoned that by comparing the CSDs of the scores obtained for each of the matrix-peptide combinations, we would obtain an indication of how the peptide-binding properties of the MHC alleles are related to each other. To compare how the 19 different scoring matrices responded to each individual peptide set, we tabulated the absolute difference (ID) between every pair of mean scores that was obtained using that peptide set. The resulting 19 tables were condensed to a single pairwise distance table by calculating the mean ID for each peptide-matrix combination. Using this pairwise table as input, an NJ tree was generated with the NEIGHBOR program of the PHYLIP package (41).

Expression, crystallization, and structure determination
Soluble Patr-AL was produced in Escherichia coli and refolded with β2-microglobulin and peptide ALDKATVLL (Anaspec, Fremont, CA), as described (42). Complexes were affinity-purified by nickel-nitrilotriacetic acid chromatography (Qiagen, Valencia, CA) using histidine-tagged HA chimi- S200 gel filtration (Amersham, Piscataway, NJ), and MonoQ ion exchange (Amersham). Protein was concentrated to 10 mg/ml and crystals grown by sitting-drop vapor diffusion at 22°C, with well solution containing 22% PEG-12000, 100 mM Tris pH 8.5, and 200 mM ammonium sulfate. X-ray diffraction data were collected at the Stanford Synchrotron Radiation Laboratory and processed using the HKL-3000 suite (43). The structure analysis and comparisons

Structure analysis and comparisons
For comparison of peptide conformation, residues 1–180 of Patr-AL were aligned using PyMOL to the following structures (Protein Data Bank ID indicated in parentheses): HLA-A*0201-FLWGPRALV (1QEW), HLA-A*0201 ALWGFFPVL (1B0G), HLA-A*0201 GILGFVFVL (1HHI), HLA-A*0201 TLTCNTCSV (1HHG), HLA-B*4401 (1SYV), HLA-B*4403 (1SYS), HLA-A*0101 (1W72), HLA-A*0201 (1HHI), HLA-B*5301 (1A1O), HLA-C*0401 (1QOD), HLA-B*2705 (1A83), HLA-B*2501 (2C1K), HLA-E*0101 (1MHE), HLA-B*1501 (1XR9), HLA-C*0301 (1EFX), HLA-A*0101 (2D31), HLA-B*0801 (1M05), HLA-A*1101 (2HNT), HLA-A*1101 (1X7Q), HLA-A*2402 (2BCK). Root mean square deviation (RMSD) between Cα carbons of peptides in different structures were measured in PyMOL. For comparison of the Patr-AL isolectric point with those of other MHC class I molecules, human HLA allotype sequences were extracted from the International ImMunoGeneTics Database (13) (http://www.ebi.ac.uk/imgt/hla/). Non-human MHC class I protein sequences were obtained from the following databases at the National Center for Biotechnology Information: mouse, NCBI (www.ncbi.nlm.nih.gov/); rat, NCBI (www.ncbi.nlm.nih.gov/); dog, NCBI (www.ncbi.nlm.nih.gov/); cattle, CAA63476; cat, NP_001041626; pig, NP_001090900; horse, NP_001090900. Isoelectric points were estimated using the EMBL calculator (http://www.embl-heidelberg.de/cgi/pi-wrapper.pl).

CTL lines
Chimpanzee PBMCs were isolated from whole blood by Ficoll gradient separation. Alloreactive CTL lines were generated by stimulating chim- panzee PBMCs with autologous feeder cells and gamma-irradiated 221 cells expressing Patr-AL, in culture for 14 d. CTLs produced during the separation. Alloreactive CTL lines were generated by stimulating chim-

Results
Patr-AL marks a unique genomic block absent from human MHC haplotypes
A Patr-AL cDNA probe was used to screen a BAC library made from genomic DNA of a Patr-AL” chimpanzee. By sequencing two overlapping BAC clones, we defined a 214-kb region that contained Patr-AL. In this region, the telomeric 22-kb segment and the centromeric 67-kb segment correspond with contiguous sequence in the Patr-AL” haplotype described by Anzai et al. (15). In contrast, the central 125-kb region, which contains the Patr-AL gene, has no counterpart in the Patr-AL” haplotype. Eight sequenced HLA haplotypes (45–47) also lack this 125-kb genomic block and have the same breakpoints with the Patr-AL” haplotype as the Patr-AL” haplotype (Fig. 1). The absence of MHC-AL from these human and chimpanzee MHC haplotypes almost certainly derives from one deletion event in an ancestral MHC-AL” haplotype, which occurred prior to separation of human and chimpan-

From forward simulations under neutral evolution, we estimate that the mean time required for either loss or fixation of Patr-AL in the chimpanzee population would have been 83,300 generations (1.24 million years), a time period much shorter than that elapsing since separation of human and chimpanzee ancestors >6 mya (31).

Patr-A, the orthologue of HLA-A, lies in the centromeric 67-kb segment of the 214-kb region we sequenced, at a distance of 70 kb from Patr-AL (Fig. 1A). Flanking Patr-AL are several MHC class I pseudogenes and gene fragments (Patr-K and Patr-U upstream, Patr-W and MIG-D downstream) in identical configuration to that observed in the Patr-AL” and HLA haplotypes (Fig. 1A). Patr-AL is also flanked by pseudogenes and gene fragments, which we term Patr-K-like (Patr-KL), Patr-W-like (Patr-WL), MIG-D-like (MIG-DL), and Patr-T-like (Patr-TL), according to their similarities with human pseudogenes (48). Importantly, the organization of flanking pseudogenes and gene fragments differs between Patr-AL and Patr-A/HLA-A. First, the W pseudogene is downstream from Patr-A and HLA-A, whereas WL is upstream of Patr-AL. Second, the TL pseudogene downstream from Patr-AL has no counterpart in the genomic block containing Patr-AL. These differences show that the blocks containing Patr-AL and Patr-A are not simply products from a single duplication of a common ancestral block. Upstream of the block containing Patr-AL lies the block containing the Patr-H pseudogene, and downstream of the block containing Patr-A is the Patr-J pseudogene. Like Patr-AL, the H pseudogene (49) is related to MHC-A, as is the J pseudogene, but to a lesser extent (50). That the TL pseudogene downstream of Patr-AL corresponds with the T pseudogene downstream of Patr-H and HLA-H raised the possibility that the blocks containing Patr-H and Patr-AL were the duplicated products of a common ancestor (Fig. 1B).

Analysis of a gorilla MHC haplotype (GenBank accession numbers CU104658 and CU104664) showed it aligns with the chimpanzee and human haplotypes and has the blocks containing MHC-H and MHC-J flanking a block containing AL (Fig. 1A, 1C). Analysis of repetitive elements, as well as MHC class I pseudogenes and gene fragments, demonstrated that the block containing MHC-A is absent from the gorilla haplotype, whereas a block corresponding to the Patr-AL block is present (Fig. 1A, 1C). In the same position as the Patr-AL gene is a gene corresponding with a previously characterized cDNA, called Gogo-Oko, which has a divergent recombinant structure that puts it apart from other cDNA sequences named in the Gogo-A series (51, 52). From domain-by-domain phylogenetic analysis, we now see that Gogo-Oko has segments in common with AL, H, and the A2 lineage of A (Fig. 2B). That only the S1’ end of Gogo-Oko remains orthologous
to *Patr-AL* shows the extent to which this putative gorilla equivalent of *Patr-AL* has been replaced by segments of *H* and *A*. That several gorilla *Gogo-A* are more related to *HLA-A* and *Patr-A* than to *Gogo-Oko* (51) raises the possibility that some gorilla MHC haplotypes have retained the block containing the *A* locus.

**Evolution of duplicated blocks and haplotypes containing *H*, *AL*, and *A* genes**

In higher primates, the MHC class I gene family has expanded and diversified through duplications, followed by deletions, of a building block of ~50 kb containing one MIC gene and one MHC class I gene (53). Evidence for these processes is seen in Fig. 1A. Our results are consistent with an evolutionary model in which the modern genomic blocks containing *H*, *AL*, and *A* evolved from a common ancestor by two successive duplications (Fig. 1B). The initial duplication produced one block containing the common ancestor of *AL* and *H* and a second block containing the ancestor of *A* and a fourth unidentified gene (*A* in Fig. 1B). The second duplication produced a haplotype with four blocks: three ancestral to the modern blocks containing *A*, *AL*, and *H*, whereas the fourth block lost the unidentified *A* class I gene and its associated *U* pseudogene as part of the deletion that gave rise to the structure of the *AL* haplotype, with its unique set of genes and gene fragments upstream of MHC-AL (Fig. 1B). From this
haplotype, the AL⁺ haplotype was formed by deletion of the 125-kb block containing AL, whereas the AL⁻ haplotype arose by deletion of the ~80-kb block containing A. Phylogenetic analysis of the hominoid T, W, and K pseudogenes is consistent with the model (Fig. 1D) and allowed us to estimate the time of the duplications using a Bayesian approach (30). The first duplication occurred 23.3 mya (95% confidence interval, 15.9–30.3 mya), consistent with the earlier estimate of 26 mya for divergence of A and AL (23); the second occurred 15.3 mya (95% confidence interval, 11.6–23.2 mya) (Fig. 1D).

Species-specific evolution of AL gene diversity and function

To search for counterparts of Patr-AL in other hominoid species, we performed phylogenetic analyses of MHC class I gene sequences (Fig. 2). Fig. 2A shows the tree obtained with the complete full-length gene sequences. The sequences were also divided into smaller segments that were separately subjected to phylogenetic analysis, and a summary of the relationships observed is given in Fig. 2B by colored shading of the segments.

Most closely related to Patr-AL is orangutan Popy-A (Fig. 2A), which is orthologous to Patr-AL throughout the gene, with the exception of an ~10-bp segment in exon 2 (encoding residues 65–67 of the α₁ domain) where Patr-AL is more related to H (Fig. 2B), a likely consequence of gene conversion. Unlike Patr-AL, Popy-A is a highly polymorphic gene and for this reason was previously considered to be orthologous to HLA-A and Patr-A (54). Since the time of the last common ancestor of chimpanzee and orangutan, the AL gene evolved differently in the two species. In orangutan it became a polymorphic classical MHC class I gene, whereas in chimpanzee it became a conserved nonclassical class I gene.

Although complete Patr-AL orthologues were not identified in gorilla and human, both species have MHC class I genes with segments related to AL. Exon 1 and intron 1 of gorilla Gogo-Oko and exons 1, 2 and introns 1, 2 of a human pseudogene, HLA-Y (45, 55, 56), are orthologous to Patr-AL. In contrast, the remaining exons and introns of HLA-Y appear orthologous to the A locus. Of two ancient lineages of HLA-A alleles (54, 57), HLA-Y is closer to the A2 family than the A3 family (Fig. 2B). As is the case for Patr-AL, HLA-Y is not fixed (45, 55, 56), and neither is it represented in the eight sequenced HLA haplotypes (45). Although the precise genomic location of HLA-Y remains unknown, linkage disequilibrium between HLA-Y and a subset of HLA-A demonstrates its presence in the MHC. HLA-Y has been detected on all haplotypes that have HLA-A*2901, *3001, *33, *3401, or *6802 and on most haplotypes that have HLA-A*0201, *0205, or *33J (55, 56, 58, 59). The HLA-A alleles associated with HLA-Y are more frequent in non-Caucasian populations, and we estimate that ~20% of the human population carries HLA-Y. Whereas AL sequences contribute to expressed functional genes in chimpanzee, orangutan, and gorilla, in humans the AL sequences appear only in the nonfunctional form of HLA-Y.

The peptide-binding specificity and repertoire of Patr-AL is like that of HLA-A*02

Patr-AL was affinity-purified from the supernatant of 221 cells secreting soluble Patr-AL (38). Edman-sequencing showed that the peptides bound to Patr-AL were predominately nonameric peptides constrained by three anchor positions: P2 preference for leucine, smaller aliphatic residues, and glutamine; P3 preference for aspartate; and C-terminal preference for aliphatic residues. This motif resembles that common to several HLA-A*02 subtypes (Fig. 3A) and differs from the peptide-binding motifs of other HLA-A and Patr-A [SYFPEITHI database (39)] (Supplemental Fig. 1).
The pool of Patr-AL binding peptides was fractionated by reverse-phase HPLC, and individual peptides were sequenced by tandem mass spectrometry. Because initial analysis identified ALDKATVLL, a known HLA-A*0207 binding peptide (60), we systematically compared the peptides eluted from Patr-AL with those obtained from HLA-A*0207, and also HLA-A*0201, the prototypical HLA-A*02. Sequencing 126 abundant peptides (Supplemental Fig. 2) uncovered extensive overlap between the peptides bound by Patr-AL, HLA-A*0201, and HLA-A*0207. The peptides binding to each MHC class I are defined by the differently colored circles: Patr-AL, red; A*0201, blue; and A*0207, green. The four overlapping regions between the circles define the peptides bound by all three MHC class I and by the three combinations of two of them. On the left under “Sequenced peptides” is shown the analysis for the peptides for which the amino acid sequences are known. On the right under “Molecular ions” is shown a second, independent study in which peptides were defined by the weight of their molecular ions. C, NJ phylogenetic tree to compare the peptide-binding specificities of Patr-AL, A*0201, and A*0207 as defined in our analysis (green and boxed) with those previously defined for A*0201, A*0207, A*0214 (green) and 13 other HLA-A allotypes obtained from the SYFPEITHI database (39). The number of unique peptides in each data set is shown in parentheses after the allotype name.

The pool of Patr-AL binding peptides was fractionated by reverse-phase HPLC, and individual peptides were sequenced by tandem mass spectrometry. Because initial analysis identified ALDKATVLL, a known HLA-A*0207 binding peptide (60), we systematically compared the peptides eluted from Patr-AL with those obtained from HLA-A*0207, and also HLA-A*0201, the prototypical HLA-A*02. Sequencing 126 abundant peptides (Supplemental Fig. 2) uncovered extensive overlap between the peptides bound by Patr-AL, HLA-A*0201, and HLA-A*0207 (Fig. 3B, left). This was pursued by examining 849 molecular ions, each randomly selected from one of the peptide profiles and then assessed for its presence in the other two profiles (Fig. 3B, right).

In magnitude, the overlap between peptides bound by Patr-AL and either A*0201 or A*0207 was comparable with that between the two A*02 subtypes (Fig. 3B, right): 52% of Patr-AL–bound peptides also bound A*0201 or A*0207; 49% of A*0207–bound peptides also bound A*0201. We estimate that 66% of the peptides bound by Patr-AL also bind to an HLA-A*02 subtype and that 40% of the peptides bound by the entire family of A*02 subtypes can bind to Patr-AL. Conversely, review of the literature and the content of the SYFPEITHI database for MHC ligands and peptide motifs (39) showed that none of the peptides bound by Patr-AL has been found to bind to any MHC class I allotype other than HLA-A*02. Thus, the peptide-binding function of chimpanzee Patr-AL is like that expected of a novel HLA-A*02 subtype. Such similarity was totally unexpected because Patr-AL differs from A*0201 by 30 aa substitutions in the peptide-binding α1 and α2 domains (Supplemental Fig. 3) (23), 13 of which are predicted to contact peptide (61). By contrast, substitution of tyrosine for cysteine at position 99 is all that distinguishes A*0207 from A*0201.

Although the above analysis demonstrated considerable overlap between the repertoires of peptides bound by Patr-AL and HLA-A*02 and failed to find any commonality with other HLA class I molecules, we developed an independent method to give unbiased comparison of the peptide-binding specificity of Patr-AL with a broader range of HLA-A allotypes embracing both the A2 and A3 ancient families of HLA-A alleles. We created a peptide-binding scoring matrix for each of the 16 HLA-A allotypes for...
which sufficient information was available in the SYFPEITHI database (39). The peptides eluted in this study from the HLA-A*0201 and A*0207 allotypes were also used to generate independent scoring matrices. Each matrix was then used to score each set of eluted peptides and pairwise comparisons of the scores used to generate a distance-tree. In this tree the Patr-AL peptide-binding specificity is clearly seen to cluster within the group of HLA-A*02 subtypes and to be apart from the other HLA-A (Fig. 3C).

Although the Patr-AL and A*0202 specificities are clearly differentiated from those of other HLA-A, the peptides defined in our analysis group A*0201 and A*0207 more closely together than is seen from analysis of the A*0201 and A*0207 binding peptides defined in other studies (Fig. 3C). This likely reflects differences in the methods used to assay and define the binding peptides. All our data came from sequence analysis of peptides eluted from HLA class I secreted from 221 cells, whereas the sequences in SYFPEITHI derive from a variety of cellular and molecular methods (62, 63). For example, a study based on the binding of synthetic peptides found that the peptide-binding repertoire of A*0207 was largely limited to a subset of that bound by A*0201 (60), whereas in our analysis 30 of the 72 sequenced peptides eluted from A*0207 were not among the 49 peptides eluted from A*0201 (Fig. 3B). It is also likely that the small number of nonamer sequences in SYFPEITHI representing some allotypes led to imprecision in the values for their scoring matrices.

**Patr-AL and HLA-A*02 bind peptides with similar conformation**

The complex of Patr-AL bound to the ALDKATVLL peptide was crystallized and a three-dimensional structure determined at 2.7 Å resolution (Table I). Patr-AL has a typical MHC class I structure (Fig. 4A), in which the Cα traces of the H chain and β2-microglobulin superimpose with their counterparts in other HLA class I structures. Notably, the root mean square deviation between Cα carbons of the Patr-AL and A*0201 chains was 0.557 Å.

Despite the common structure and peptide-binding specificity, Patr-AL is distinguished from HLA-A*02 and all other forms of MHC class I by the unusually electropositive solvent-accessible surface of its Cα helix (64). The Cα helix of Patr-AL has one lysine (position 161) and five arginine (positions 141, 145, 151, 152, and 163) residues in addition to the three lysines (positions 144, 146, and 176) and six arginines (positions 108, 111, 131, 157, 169, and 170) present in A*0201 (Fig. 4B). This preponderance of positive charge is such that Patr-AL is the only known MHC class I isoform with a basic isoelectric point (pI = 8.5) (Fig. 4C), and it is poorly resolved by the conditions usually used to distinguish HLA-A and B variants (65). For other MHC class I molecules, this surface of the Cα helix binds to the Vα domain of TCRs (66).

The conformation of bound ALDKATVLL peptide and its interactions with the binding groove are well resolved in the Patr-AL structure (Supplemental Fig. 4). The N terminus is deeply buried in the groove, and the side chain of the P2 anchor points into the B pocket between the Cα helix and the β-sheet floor. The peptide backbone then arches up to overcome an obstruction in the peptide-binding groove caused by the bulky His70 and Tyr99 side-chains. The arch peaks at residue P4 and then slopes down into the groove, allowing the P9 anchor to engage the F pocket. This conformation is very similar to those observed for six different peptides bound to HLA-A*0201 (Fig. 4D) [Protein Data Bank IDs: 1QEW, 1B0G, 1HHG, 1HHI, 1HHJ, 1HHK (67, 68)].

We quantified the conformational differences between the peptides bound to Patr-AL and those bound to 21 other MHC class I molecules by calculating the pairwise RMSDs of the peptide Cα backbones. Comparison of Patr-AL with HLA-A*0201 gave RMSDs of 0.58–1.05 Å, well within the range defined by the six A*0201-binding peptides (0.44–1.56 Å). In contrast, comparison of Patr-AL with 15 other HLA class I isoforms gave RMSDs of 1.03–2.17 Å (Fig. 4E). The striking conformational similarity of peptides bound to Patr-AL and HLA-A*02 cannot be attributed solely to anchor residue preferences because the distantly related HLA-B*0801, E*0101, and G*0101 isoforms share the preference of Patr-AL for aliphatic anchors at P2 and P9, yet their bound peptides deviate by 2.07 Å, 1.88 Å, and 1.95 Å, respectively, from that of Patr-AL.

**The specificity-determining pockets of Patr-AL and HLA-A*0201 have similar architecture despite containing nonconservative substitutions**

Because the B and F specificity-determining pockets play a major role in determining which peptides bind to MHC class I (69), we compared their architecture in Patr-AL and HLA-A*0201 (Fig. 5). As a negative control we also examined HLA-B*0801, which has a non-overlapping peptide-binding repertoire with Patr-AL and A*0201.

The B pockets of Patr-AL (Fig. 5A, left) and A*0201 (Fig. 5A, center) are both deep and hydrophobic. Despite their similarities in size and shape, the two pockets differ by nonconservative substitutions at positions 66, 67, and 70. Substitution of lysine 66 in A*0201 for isoleucine in Patr-AL appears functionally neutral because the lysine side chain contributes four aliphatic carbons to the wall of the pocket while the charged e-amino group remains solvent-accessible at the top of the groove. Serine 67 at the bottom of the Patr-AL B pocket is similar in size to valine 67 in A*0201, but its hydroxyl group is available for hydrogen bonding, which

<table>
<thead>
<tr>
<th>Table I. Summary of crystallographic analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection and Processing</strong></td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Cell dimensions</td>
</tr>
<tr>
<td>α, β, γ (%)</td>
</tr>
<tr>
<td>Reflection σ cutoff</td>
</tr>
<tr>
<td>Reflections (total / test)</td>
</tr>
<tr>
<td>R/Rmerge factors</td>
</tr>
<tr>
<td>Number of atoms</td>
</tr>
<tr>
<td>RMSD bond length (Å)</td>
</tr>
<tr>
<td>Average B factor</td>
</tr>
<tr>
<td>MHC H chain (A, 2,256 atoms)</td>
</tr>
<tr>
<td>Light chain β2m (B, 833 atoms)</td>
</tr>
<tr>
<td>Peptide (C, 66 atoms)</td>
</tr>
<tr>
<td>Water (42 atoms)</td>
</tr>
<tr>
<td><strong>PROCHECK Statistics</strong></td>
</tr>
<tr>
<td>Residues in most favored regions (%)</td>
</tr>
<tr>
<td>Residues in additional allowed regions (%)</td>
</tr>
<tr>
<td>Residues in generously allowed regions (%)</td>
</tr>
<tr>
<td>Residues in disallowed regions (%)</td>
</tr>
</tbody>
</table>

*Values in parentheses are for highest-resolution shell.
β2m, β2-microglobulin.

Downloaded from http://www.jimmunol.org/ by guest on September 15, 2017
could explain the preference of Patr-AL for glutamine at P2 (Fig. 3A, left). Histidine 70 in A*0201 obstructs the groove’s floor, like tyrosine 70 in Patr-AL. The hydrophobic edge of the indole ring of histidine 70 faces the B pocket and preserves its hydrophobic nature. Contrasting with these similarities, the B pocket of B*0801 differs in size, shape, and composition from its Patr-AL and A*0201 B counterparts (Fig. 5A, right). Bulky phenylalanine 67 makes the pocket shallower, and substitution of aspartate for phenylalanine at position 9 allows tyrosine 99 to adopt a different rotamer, which disrupts the pocket wall causing the bound peptide to sink deeper into the groove.

With the exception of position 95, all the residues lining the F pocket are conserved in Patr-AL, HLA-A*0201, and HLA-B*0801 (Fig. 5B). At the bottom of the F pocket, A*0201 and B*0801 have valine 95, whereas Patr-AL has the larger isoleucine, which is accommodated by a different rotamer of leucine 81. Consequently, the F pocket of Patr-AL is wider and shallower, consistent with its increased capacity to bind peptides with C-terminal...
phenylalanine (Fig. 3A). From this analysis, we see that the specificity-determining pockets of A*02 and Patr-AL accomplish the same functional effect, but in different ways, using different amino acid residues and different molecular contacts.

Evolution of the peptide-binding specificity shared by Patr-AL and HLA-A*02

MHC class I allotypes can be clustered according to their peptide-binding specificity, as assessed by the combination of peptide anchor residues preferentially bound by the B and F pockets of the peptide binding site; such clusters being referred to as supertypes. HLA-A allotypes have been grouped into six supertypes (A01, A02, A03, A24, A01-A03, and A01-A24) (69). The A02 supertype principally consists of A*02 and the related A*69 and some A*68 subtypes. From examining the patterns of substitution in the B and F pockets, we discovered that A02 peptide specificity correlates simply with the amino acid residues at position 9 in the B pocket and at position 116 in the F pocket (Fig. 6). Only allotypes of the A02 and A24 supertypes can accommodate aliphatic residues in the F pocket, a feature correlating with presence of tyrosine at...
position 116. Distinguishing A24 from A02 is the capacity to bind aromatic residues in the B pocket, which is dependent on serine at position 9. Thus, the A02 supertype is uniquely defined by the lack of serine at position 9 and the presence of tyrosine at position 116. Consistent with this definition, Patr-AL has phenylalanine at position 9 and tyrosine at position 116, as do two of three Popy-A allotypes (Fig. 6). In contrast, none of the 30 Patr-A allotypes has this motif, as it is also true for five bonobo Papa-A allotypes, four Gogo-A allotypes and Gogo-Oko, and two gibbon Hyla-A allotypes (70). This correlation is only relevant in the context of the A-related genes, because many HLA-B and C allotypes, including HLA-A*02.01, do not have peptide-binding specificities similar to HLA-A*02. Thus, other residues that distinguish HLA-A from HLA-B and C (71) make important contributions to the A02 supertype. However, in the context of A-related genes, the residues at positions 9 and 116 provide simple evolutionary switches that can introduce or take away the A02 supertype specificity.

To track how the A02 specificity has evolved, we performed ancestral sequence reconstructions at positions 9 and 116. Our goal was to assess if the shared peptide-binding specificity of A*02 and Patr-AL had been maintained since the time of their common ancestor or if it has periodically been lost and regained. The phylogenetic tree in Fig. 7A (72) examines both positions 9 and 116, whereas that in Fig. 7B concentrates on position 9. The common ancestor of Patr-AL and HLA-A is predicted to have had phenylalanine 9 and tyrosine 116, the combination retained by Patr-AL. Thus, we predict this ancestor had the A*02/AL peptide binding specificity. In contrast, the last common ancestor of all HLA-A allotypes has the combination of tyrosine 116 without serine 9. Thus, the A02 supertype is uniquely defined by the lack of aromatic residues in the B pocket, which is dependent on serine at position 9 of the peptide-binding domain (highlighted purple).

Patr-AL is recognized by TCRs and influences the T cell repertoire

Because Patr-AL is not fixed in the chimpanzee genome (23), individual chimpanzees can either have or lack Patr-AL. To see if Patr-AL functions as a T cell alloantigen, PBMCs from three Patr-AL+ chimpanzees were stimulated with class I-deficient 221 cells transfected with Patr-AL. Vigorous cellular proliferation yielded CD8+ CTL lines that killed 221 cells transfected with Patr-AL, but not untransfected cells or cells transfected with Patr-A (Fig. 8A, left panels). As a control, PBMCs from three Patr-AL+ chimpanzees were similarly cultured with Patr-AL–expressing 221 cells. This stimulation gave less proliferation, and the T cells produced did not kill Patr-AL–expressing 221 cells (Fig. 8A, right panels). This result shows that Patr-AL is recognized functionally by TCRs and induces self-tolerance in Patr-AL+ individuals. It also implies that Patr-AL is expressed in the thymus where it participates in negative selection of the T cell repertoire.

The CTLs raised against Patr-AL were tested for their capacity to kill 221 transfected cells expressing a variety of Patr-A and HLA-A allotypes, including HLA-A*0201. The CTLs were exquisitely specific for Patr-AL, showing no reactivity with 221 cells expressing any other form of MHC class I. This was not so surprising a result given the extensive sequence divergence of Patr-AL in the upper faces of the α1 and α2 helices that interact with TCR (23), particularly the uniquely electropositive face of the Patr-AL α2 helix predicted to interact with the TCRα-chain (73).

Discussion

Coding-region sequences group HLA-A alleles into six families—roughly corresponding with the broad serological types—derived from two ancient lineages. The A3 lineage comprises the A9, A80, and A11/A3/A11 families, and the A2 lineage comprises the A2, A10, and A19 families (54, 57, 74). Patr-A, the chimpanzee orthologue of HLA-A, has only A13/A11 family alleles of the A3 lineage (75). This restriction reduces Patr-A diversity compared with HLA-A, whereas Patr-B and C are more diverse than human HLA-B and C. Because chimpanzee genomes are overall more diverse than human genomes (8, 76, 77), this unusual reversal for MHC-A suggested that pathogen-mediated selection has favored...
preservation of A2 lineage alleles on the human line and/or their extinction on the chimpanzee line (78).

On examining noncoding sequences, notably intron 2 that separates exons 2 and 3 encoding the MHC class I peptide binding site, de Groot et al. (79, 80) detected lower diversity in Patr-A, B, and C than in their human orthologues. This pointed to chimpanzees experiencing general reduction in MHC class I diversity during the selective sweep, which did not affect other gene systems. Another potential consequence of the sweep was fixation in chimpanzee of the deletion that recombined MICA with MICB to give the chimeric MICA/B gene (15, 81). The sweep was estimated to have occurred after separation of human and chimpanzee ancestors 6–9 mya but before chimpanzee subspeciation ~1.5 mya. de Groot et al. (80) speculated that the selective sweep was caused by a simian ancestor of HIV-1, which could explain why modern chimpanzees are more resistant to HIV-1 infection than are humans. Circumstantial evidence suggests HIV-1 has adapted to HLA-A*02 during the current epidemic and that T cells responding to viral Ags presented by A*02 are ineffective (82–84). Even with the application of modern medicine, this situation is expected to lead to reduced A*02 frequencies in human populations and the possibility of its extinction in some of them. de Groot et al. proposed that comparable adaptation of pathogen to A2 lineage alleles led to their extinction in ancestral chimpanzee populations.
A ratio of 2.5. Data in specific for Patr-AL, exhibiting no significant cross-reactivity with any of human and chimpanzee MHC-A allotypes. The CTLs are exquisitely individual chimpanzee and transfected 221 target cells expressing a range of cytotoxicity assays performed with Patr-AL–specific CTLs from one CTL lines. As shown for CTL produced from PBMCs of Patr-AL expressing transfected 221 cells led to the generation of Patr-AL–specific T cells. Although Patr-AL binds peptides like HLA-A*02, Patr-AL, no CTLs reactive with Patr-AL were produced (right panels).

The peptide-binding specificities of HLA-A allotypes have been grouped into six supertypes: A01, A02, A03, A24, A01-A03, and A01-A24 (69). Extending the analysis to chimpanzee MHC class I identified examples of the A01-A24 supertypes and absence from the B pocket of serine 9. Patr-AL and HLA-A*02 have very similar three-dimensional structures and conformations of bound peptide. However, the anchoring interactions of peptide residues 2 and 9 with the B and F pockets, respectively, differ significantly in detail because of nonconservative substitutions in the residues lining these pockets in Patr-AL and HLA-A*02.

We demonstrate that chimpanzee Patr-AL binds peptides like HLA-A*02 and is of the A02 supertype. Neither Patr-AL nor HLA-A*02 is fixed, and their gene frequencies are comparably high. HLA-A*02 and Patr-AL are both alloantigens that interact with the αβ receptors of CD8 T cells and will induce T cell tolerance when expressed as self-MHC class I. Thus, Patr-AL has the potential to contribute to chimpanzee immunity by presenting peptide Ags to CD8 T cells. Although Patr-AL binds many of the same peptides as HLA-A*02, Patr-AL–specific T cells do not recognize HLA-A*02 and other MHC class I allotypes, which we attribute to the numerous nonconservative substitutions that distinguish Patr-AL from other MHC class I in the upward face of the class I molecule that contacts TCRs. Further differentiating Patr-AL from HLA-A*02 and other MHC-A is its much reduced polymorphism, lower levels of gene and cell surface expression, and a more restricted tissue distribution (23). In aggregate, these differences argue for chimpanzee Patr-AL and human HLA-A*02 having qualitatively different functions.

The peptide-binding specificities of HLA-A allotypes have been grouped into six supertypes: A01, A02, A03, A24, A01-A03, and A01-A24 (69). Extending the analysis to chimpanzee MHC class I identified examples of the A01-A24 supertypes and absence from the B pocket of serine 9. Patr-AL and HLA-A*02 have very similar three-dimensional structures and conformations of bound peptide. However, the anchoring interactions of peptide residues 2 and 9 with the B and F pockets, respectively, differ significantly in detail because of nonconservative substitutions in the residues lining these pockets in Patr-AL and HLA-A*02.

Although losing the ancient A2 lineage of MHC-A alleles, the chimpanzee has not lost the peptide-binding specificity of the A02 supertype. In considering the overall reduced diversity of chimpanzee MHC class I compared with human MHC class I, Patr-AL stands out as a factor that chimpanzees have and humans lack. Clearly, Patr-AL survived the selective sweep postulated by de Groot et al. (79, 80), and it is conceivable that its function was beneficial and has been a target for positive selection, as is consistent with our demonstration that Patr-AL has been subject to balancing selection. In this scheme of things, absence of a human equivalent of Patr-AL could make humans more susceptible to HIV/AIDS.

HLA-A*02, Patr-AL, and orangutan Popy-A last shared a common ancestor 14–30 mya. We predict this ancestor had phenylalanine 9 and tyrosine 116, the combination retained by Patr-AL and one of three Popy-A allotypes (Popy-A*03) (70). Thus, it is likely that functional MHC class I of the A02 supertype was maintained throughout the evolution of Patr-AL on the chimpanzee lineage. That is not the case for the human line, because ancestral HLA-A had serine 9 and tyrosine 116, which lacks the A02 supertype. Acquisition of phenylalanine 9 by A*02 was an event specific to human evolution. Thus, the A02 supertype appears to have been eliminated at some point during human evolution and then regained much later. This provides potential precedent in human history for a pathogen-mediated selective sweep that eliminated MHC-A allotypes of the A02 supertype.
Since then, the A02 supertype evolved anew and was driven by selection to high frequency in the modern human population. Future epidemiological studies should determine if the current epidemic of HIV/AIDS is acting to reverse that trend and reduce the frequency of HLA-A*02 and the A02 supertype.

Acknowledgments
We thank the Yerkes Regional Primate Center for the samples of chimpanzee peripheral blood.

Disclosures
The authors have no financial conflicts of interest.

References