A Comparative Analysis of Viral Peptides
Presented by Contemporary Human and Chimpanzee MHC Class I Molecules

Hanneke W. M. van Deutekom, Ilka Hoof, Ronald E. Bontrop and Can Kesmir

*J Immunol* 2011; 187:5995-6001; Prepublished online 31 October 2011;
doi: 10.4049/jimmunol.1102236
http://www.jimmunol.org/content/187/11/5995

Supplementary Material
http://www.jimmunol.org/content/suppl/2011/11/01/jimmunol.1102236.DC1

References
This article cites 37 articles, 12 of which you can access for free at:
http://www.jimmunol.org/content/187/11/5995.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
A Comparative Analysis of Viral Peptides Presented by Contemporary Human and Chimpanzee MHC Class I Molecules

Hanneke W. M. van Deutekom,* Ilka Hoof,* Ronald E. Bontrop,*† and Can Keşmir*

Genetic factors such as the MHC influence the immunocompetence of an individual. MHC genes are the most polymorphic genes in primates, which is often interpreted as an adaptation to establish good T cell responses to a wide range of (evolving) pathogens. Chimpanzee MHC (Patr) genes are less polymorphic than human MHC (HLA) genes, which is surprising because chimpanzee is the older species of the two and is therefore expected to display more variation. To quantify the effect of the reduced polymorphism, we compared the peptide binding repertoire of human and chimpanzee MHC molecules. Using a peptide-MHC binding predictor and proteomes of >900 mammalian viruses, we show that, at the population level, the total peptide binding repertoire of Patr-A molecules is ~36% lower than that of their human counterparts, whereas the reduction of the peptide binding repertoire of the Patr-B locus is only 15%. In line with these results, different Patr-A molecules turn out to have largely overlapping peptide binding repertoires, whereas the Patr-B molecules are more distinct from each other. This difference is somewhat less apparent at the individual level, where we found that only 25% of the viruses are significantly better presented by “simulated” humans with heterozygous HLA-A and -B loci. Taken together, our results indicate that the Patr-B molecules recovered more after the selective sweep, whereas the Patr-A locus shows the most signs of the selective sweep with regard to its peptide binding repertoire. The Journal of Immunology, 2011, 187: 5995–6001.

Human MHC class I molecules, also known as HLAs, present peptides derived from degraded proteins to cytotoxic T cells. The three different loci of the HLA class I genes, A, B, and C, originate from ancient gene duplications (1). Of these, the HLA-B gene is the most polymorphic gene in the human genome, with >1900 alleles, followed by HLA-A (>1300) and HLA-C (>900) (2, 3) (http://www.ebi.ac.uk/imgt/hla/stats.html).

The high polymorphism of the HLA genes is thought to be a result of host/pathogen coevolution and is not specific to the human population (4). Indeed, orthologs of HLA genes have been identified in nonhuman primates. Especially, the organization and linkage of HLA genes and chimpanzee MHC genes (Patr genes) is similar because the Patr class I region also comprises an A, B, and C locus (5–7). Phylogenetic analysis of primate MHC-A alleles, based on exon sequences, suggest the existence of two ancestral lineages: the A2 lineage includes human and gorilla alleles, whereas the A3 lineage consists of human, chimpanzee, and bonobo alleles (8, 9). Indeed, all known Patr-A alleles are related to the HLA-A*01, -A*03, and -A*11 families that are part of the A3 lineage (5, 10, 6, 11). Because the common ancestor of gorilla and human predates the common ancestor of chimpanzee and human, it has been suggested that chimpanzees and bonobos lost their MHC molecules belonging to the A2 lineage (6). A characteristic of the chimpanzee MHC region is an additional A-like MHC class I gene, Patr-AL, that does not have a human counterpart (6). Patr-AL groups outside of both the A2 and A3 lineages based on the coding sequence and is present only on ~50% of the haplotypes. With a low level of polymorphism as well as low cell surface expression, Patr-AL shows the characteristics of nonclassical class I molecules (6). Constructing the phylogeny of the MHC-B alleles is challenging due to the frequent recombinations occurring at this locus (12, 13). de Groot et al. (14) analyzed intron 2 sequences of MHC class I molecules and concluded that the diversity of Patr-A alleles is even more reduced than the diversity of Patr-B alleles, which was subsequently supported by microsatellite data (15). The reduced variation of both Patr-A and Patr-B molecules is thought to represent the signature of a selective sweep, which took place 2–3 million years ago (14). A possible candidate as selective agent was suggested to be a HIV-1/SIVcpz–related retrovirus because humans and chimpanzees show differences in pathology after HIV-1/SIVcpz infection; chimpanzees rarely develop AIDS (16–18), whereas in the absence of therapy, the vast majority of HIV-1–infected humans do. Subsequently, the contemporary Patr molecules were found to be similar to the HLA molecules of human long-term nonprogressors (17, 19, 20).

To quantify the functional consequences of the selective sweep on chimpanzee MHC molecules, one needs to compare the peptide binding repertoires of Patr and HLA molecules in relationship to all viruses known to infect these hosts. Such an analysis has been performed on a small scale, that is, where the presentation of HIV-1/SIVcpz Gag epitopes was compared within a few Patr and HLA molecules (20). The results of this study are in line with the hypothesis that an HIV/SIVcpz-like virus could have been the agent causing the selective sweep. However, as the number of peptides binding to a MHC molecule is one of the many factors shaping an efficient T cell response, much more data are needed to determine the extent

*Theoretical Biology and Bioinformatics, Utrecht University, 3584 CH Utrecht, The Netherlands; and Biomedical Primate Research Centre, 2280 GH Rijswijk, The Netherlands.

Address correspondence and reprint requests to Hanneke W.M. van Deutekom, Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. E-mail address: H.W.M.vanDeutekom@uu.nl

The online version of this article contains supplemental material.

Copyright © 2011 by The American Association of Immunologists, Inc. 0022-1767/11/S16.00

www.jimmunol.org/cgi/doi/10.4049/jimmunol.1102236
of the selective sweep. Generating more data using biological material from chimpanzees is difficult because of the regulation to minimize invasive biomedical research using great ape species (21). Therefore, we instead employ an in silico approach to compare the peptide binding repertoire of contemporary Patr molecules with that of HLA molecules, using a large set of viral proteomes that is a nearly complete representation of the viral world both species are facing.

Materials and Methods

MHC-I binding predictions
NetMHCpan-2.0 (22) was used to predict binding affinities of peptides to MHC class I molecules, because this is the only predictor that can predict binding affinities of less characterized MHC molecules, such as Patr molecules. NetMHCpan is a neural network-based predictor that uses MHC class I binding groove polymorphisms to predict peptide-MHC binding affinities. The 2.0 version is based on a training set including nonhuman binding data. When using NetMHCpan to compare a wide array of MHC molecules, applying a fixed affinity threshold (e.g., 500 nM IC50) to define the peptide repertoires is not feasible, because the range of prediction scores varies considerably between alleles. Therefore, we chose to take a 1% cutoff to define the peptide repertoire of each MHC molecule, which makes a comparison between alleles more reliable. In other words, peptides are defined as binders by sorting them with respect to their binding affinities and selecting those that rank among the top 1% of peptides in the viral world, and thus we assume that all MHC molecules bind, in total, the same number of peptides. For individual viruses, however, this allele-specific threshold allows for a difference in the number of binding peptides between MHC molecules (e.g., the percentage of HTLV-1 peptides binding to HLA molecules ranges from 0.2 to 2.2%). To simulate the low expression level of Patr-AL, we defined the set of peptides binding to Patr-AL at the top 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, and 1% of peptides with the highest binding affinities, where, for example, the top 0.4% indicates that the number of peptides binding to Patr-AL is 40% of the peptides relative to other MHC molecules.

Selection of MHC molecules
HLA molecules with a frequency > 0.5% were included in this study (National Marrow Donor Program, http://bioinformatics.nmdp.org/HLA/ Haplotype Frequencies, May 2010). For all of these HLA molecules, the pseudosequence, defined as the amino acid residues in contact with the peptide (23), is used as input for NetMHCpan (22). This forced us to use only HMC molecules with unique pseudosequences, resulting in 37 HLA-A and 65 HLA-B molecules. Without taking frequency into account, there are 79 unique HLA-A pseudosequences and 175 unique HLA-B pseudosequences. HLA-C molecules are not included in this study, as the NetMHCpan predictions for HLA-C molecules are of lower quality due to the absence of HLA-C molecules from the training data (for comparison of peptide binding predictions to HLA-A, -B, and -C molecules, see Ref. 22).

All Patr molecules whose sequence was available were included in this analysis, resulting in 29 Patr-A and 40 Patr-B molecules with unique pseudosequences.

Subsets of MHC molecules were created by randomly selecting MHC molecules without replacement, within one locus and one species, for all different subset sizes.

Phylogenetic analysis
The HLA and Patr sequences were downloaded from the IPI-MHC Database (http://www.ebi.ac.uk/ipd/mhc, November 2010). The HLA-A and Patr-A sequences were aligned with ClustalW (24), and a tree was generated with PhyML v2.4.5 using BioNJ. A bootstrap of 100 replicates was used to define the confidence in tree nodes (25).

Functional clustering of MHC class I molecules
For each MHC class I molecule, the binding affinity was predicted for a set of 100,000 random natural nonamer peptides, resulting in an affinity vector per molecule. These affinity vectors were then used to calculate pairwise distances (d) between the binding specificities of two molecules, defined as d = 1 − Pcorr, where Pcorr is Pearson’s correlation coefficient between the corresponding affinity vectors. Based on the resulting pairwise distance matrices, hierarchical clustering was performed using the unweighted pair group method with arithmetic mean (average linkage clustering) algorithm implemented in the PHYLP software package v3.68.

Bootstrapping was performed by choosing random sets of binding affinities with replacement. A majority consensus clustering was calculated and visualized using the software SplitTree4 (26).

Creating a “viral world”
All viral proteomes and genomes were downloaded from the European Bioinformatics Institute (http://www.ebi.ac.uk, October 2008). Because the host annotation did not always allow us to differentiate between primates and other mammals, we included all mammalian viruses in this study (n = 904). At least 55% of these mammalian viruses are known to infect pri-mates. All unique nonamers from these viruses were used to create the viral world, which resulted in 3.2 × 106 unique nonamers. Viral worlds with octamers and decamers were constructed in the same way.

Once nonredundant viruses were selected for the individual analysis, that is, the nonamers of the viruses were compared in a pairwise manner and in case of an overlap of >80%, one of the viruses was randomly selected. Performing the homology reduction at the proteome level hardly changes the set of nonredundant viruses.

Results

Characterization of “holes” in the peptide binding repertoire of Patr-A molecules
Several studies have addressed the MHC repertoire reduction in chimpanzee using genomic sequence analyses (5, 6, 14). Usually all exon sequences or intron 2 sequences of human and chimpanzee MHC molecules are used to construct a phylogenetic tree to look for clusters of MHC molecules that contain exclusively human sequences. Such a phylogenetic tree was reconstructed using the 37 most common HLA-A and all (n = 29) known Patr-A molecules (Fig. 1A). As expected, the A2 and A3 lineages form distinct clusters and the A2 lineage does not contain any Patr-A molecules.

Based on their peptide binding preferences, HLA-A molecules can be assigned to different supertypes (27–33). The molecules in the A2 lineage belong to three HLA-A supertypes: A01, A02, and A03. All three supertypes prefer small and aliphatic hydrophobic residues as anchor residue at position 2, yet they differ in their amino acid preference at the C terminus. Specifically, the A02 supertype is characterized by aliphatic hydrophobic residues at the C terminus, whereas the A01 supertype prefers aromatic and large hydrophobic residues and the A03 supertype is the only one that binds basic residues at the C terminus. The A02 supertype uniquely occurs in the A2 lineage, whereas the A01 and A03 supertypes are present in both the A2 and A3 lineages (Fig. 1A). The loss of the A2 lineage in the chimpanzee implies, first, the absence of the A02 supertype binding specificity.

To determine whether this is the case, we performed functional clustering of the human and chimpanzee MHC-A molecules based on their peptide binding specificities. To this end, the functional similarity between two MHC molecules was defined as the correlation between the predicted binding affinities among a large set of random natural peptides using the in silico predictor NetMHCpan (22) (see Materials and Methods). A high correlation of the binding specificities of two MHC molecules for the same set of peptides indicates similar binding preferences and thus a large functional similarity, whereas a low correlation indicates distinct binding preferences. These correlations were subsequently used to construct a functional clustering tree of the HLA-A and Patr-A molecules (Fig. 1B). The clustering based on peptide binding specificities is very different from the clustering based on exon sequences. As expected, the binding specificities of HLA-A molecules cluster into the four distinct supertypes described earlier. The peptide binding specificities of the contemporary Patr-A molecules are distributed rather evenly among the A01, A03, and A24 supertypes, whereas the A02 supertype binding specificity is absent in chimpanzees. Thus, the holes in the chimpanzee peptide binding repertoire roughly correspond to the A02 supertype.
binding specificity. Moreover, HLA-A and Patr-A molecules cluster differently within the A01, A24, and A03 supertypes: whereas HLA-A molecules frequently have long branches and create separate clusters, Patr-A molecules exhibit much shorter branches and tend to create single clusters with many molecules, indicating that Patr-A molecules have very similar peptide binding specificities. However, this clustering does not give an estimate of the size of the peptide binding repertoire of Patr-A molecules relative to that of the HLA-A molecules. To quantify the effect of the absence of the A2 lineage on the Patr-A peptide binding repertoire, a large data set of mammalian viral proteomes \((n = 904)\) was generated to represent the viral world the two species are facing (see Materials and Methods). This data set contains more than 3 million unique peptides of nine amino acids in length. The binding affinities of these peptides to common HLA-A molecules and all known Patr-A molecules were predicted using NetMHCpan. Peptides that potentially bind to a particular MHC molecule are defined as those peptides with a binding affinity among the top 1% highest affinities of all peptides within the data set, resulting in the same number of predicted binders by only 0.9% (10.6–11.5%; Fig. 2A). Moreover, doubling the number of HLA-A molecules in this analysis (from 37 to 79) increases the predicted level of saturation, however, depends on the prediction method used and might be underestimated by NetMHCpan (data not shown). Next, the peptide binding repertoire of HLA-A molecules was compared with that of the Patr-A molecules (Fig. 2A). Because the numbers of HLA-A and Patr-A molecules in our analysis differ, we based our comparison on heterogeneous subsets of different numbers of MHC molecules, selected randomly \((n = 2, 3, 5, 10, 15, 20, 25)\), where a large subset of MHC molecules mimics the total peptide binding repertoire of a population. For all subsets, the fraction of binding peptides is significantly smaller for Patr-A molecules than for HLA-A molecules (Fig. 2A). Specifically, upon approaching the population level \((n = 25)\), the peptide binding repertoire of Patr-A molecules is 36% smaller than the peptide binding repertoire of HLA-A molecules.

Next, we quantified the holes in the chimpanzee peptide binding repertoire as the following. All the viral peptides in our data set that are predicted to bind to a HLA-A and/or Patr-A molecule \((\sim12\%)\) of all viral peptides) were divided into three groups—peptides that bind to HLA-A only \((47%)\), peptides that bind to Patr-A only \((13\%)\), and peptides that bind at least one HLA-A and one Patr-A molecule \((40\%)\)—suggesting that HLA-A molecules have an almost 4-fold larger “species specific” peptide binding repertoire. We subsequently calculated, for each HLA-A molecule, the fraction of binders that are not predicted to bind to any Patr-A molecule, that is, that are unique to HLA-A (Table I). As expected, the HLA-A molecules belonging to the A02 supertype have very large percentages of peptides \((80–90\%)\) that are predicted not to bind to any Patr-A molecule. Similarly, the human A2 lineage molecules belonging to other supertypes than the A02 supertype, for example, HLA-A*2601/02, A*6601/02, A*3401, and A*2501, have a large fraction of binders that do not bind to any Patr-A molecule \((59–67\%; Table I)\). In contrast, some other human A2 lineage molecules, e.g., HLA-A*7401, share 95% of their peptide binding repertoire with Patr-A molecules (results not shown). Whether these binding specificities have been recovered in the chimpanzee population after the loss of the A2 lineage or have always been there remains unknown.

Until now, our analysis suggests that the largest difference in the peptide binding repertoire of HLA-A and Patr-A molecules lies in the absence of the A02 supertype binding specificities in chim-

**FIGURE 1.** A phylogenetic tree based on all exon sequences of MHC-A molecules \((A)\) and the clustering of the same molecules based on their binding specificity \((B)\). The A3 lineage HLA-A molecules are depicted in blue, A2 lineage HLA-A molecules in black, Patr-A molecules in red, and Patr-AL in green.
panzees. Therefore, we next looked at the coverage of the viral world by HLA-A molecules after removing the A02 supertype. As expected, the remaining HLA-A molecules bind a smaller percentage of peptides (8.4% for $n = 25$); however, they still bind significantly more viral peptides than do Patr-A molecules (Fig. 2A, blue line). When the complete A2 lineage is instead removed from our analysis, the size of the peptide binding repertoire of the human A3 lineage molecules is almost the same as the size of the Patr-A peptide binding repertoire (Fig. 2A, green line). Taken together, these results suggest that the loss of the A2 lineage in Patr-A molecules affects several supertype binding specificities, with the A02 supertype being most affected.

Patr-B molecules recovered more than Patr-A molecules

A selective sweep could, for instance due to linkage disequilibrium, easily affect other MHC class I loci in addition to the A locus. Detecting the loss of the A2 lineage among contemporary Patr-A molecules has been straightforward using phylogenetic analysis (10, 14) (Fig. 1A). This is not true for the B locus, for which the phylogenetic tree has little substructure and different ancestral B lineages cannot be defined (10). However, based on intron 2 sequence analysis, the effect of the selective sweep was posited to be most pronounced in the Patr-B locus (14). This hypothesis was further supported by microsatellite data analysis and functional studies (15, 20).

To determine the effect of the selective sweep on the peptide binding repertoire of Patr-B molecules, we repeated our analysis for the B locus. Similar to HLA-A molecules, common HLA-B molecules ($n = 65$) bind 12% of the viral world peptides. Increasing the number of HLA-B molecules ($n = 175$) increases the peptide binding repertoire by only 2%, indicating that the HLA-B peptide binding repertoire reaches saturation rather quickly (Fig. 2B). Surprisingly, despite their higher polymorphism, HLA-B molecules together bind a similar fraction of viral peptides as do HLA-A molecules. As in the case of Patr-A molecules, Patr-B molecules bind significantly less diverse viral peptides than does their human counterpart: the peptide binding repertoires of random heterogeneous subsets of HLA-B molecules is 15% larger than that of Patr-B molecules when compared at population level ($n = 25$). However, the difference at the B locus is significantly smaller than the 36% difference at the A locus ($p < 0.0001$, Mann–Whitney $U$ test). This could mean either that the selective sweep only slightly affected the Patr-B locus, or the peptide binding repertoire of the Patr-B locus may have recovered more than the Patr-A peptide binding repertoire following the selective sweep. The latter is in line with the fact that the A locus evolves by point mutation, a slower process than evolving by recombination as the B locus (12, 13).

The peptide binding repertoires of Patr-A molecules strongly overlap

Careful inspection of the binding specificity tree (Fig. 1B) reveals that, within the A03 supertype, the branch lengths within Patr-A molecules are shorter than those of HLA-A molecules. This
suggests that the peptide binding repertoires among Patr-A molecules might be more similar. Indeed, within the A03 supertype, the Patr-A molecules are significantly more related to each other than are the HLA-A molecules \((p = 0.001\), Mann–Whitney \(U\) test). The same tendency is observed for the A01 supertype, although it is not significant \((p = 0.08)\). To quantify this further, the fraction of peptides that uniquely binds to a specific MHC molecule was calculated. To this end, we used 1000 heterogeneous random subsets of 25 MHC molecules and calculated for each MHC molecule the fraction of binders that do not bind to the other 24 MHC molecules. In this instance, a low fraction of uniquely binding peptides implies a large overlap with the binding repertoire of other MHC molecules. Using this measure, we found that the fraction of peptides that uniquely binds to Patr-A molecules is lower than that binding to HLA-A, HLA-B, and Patr-B molecules \((p < 0.0001\), Mann–Whitney \(U\) test; Fig. 3A), whereas the binding motifs of Patr-B molecules were found to be about as unique as those of HLA-A and HLA-B molecules \((p > 0.15\); Fig. 3A).

As an alternative approach to quantify the overlap in peptide binding repertoires, we estimated the degree of overlap among MHC molecules. If the peptide binding repertoires of MHC molecules have a large overlap, a peptide that binds a specific MHC molecule could also bind to other MHC molecules. For all the predicted binders, we calculated the mean number of MHC molecules that they can bind. The largest overlap was found for the peptide binding repertoires of Patr-A molecules, where on average 7.3 Patr-A molecules could bind the same peptide, followed by Patr-B (6.0), HLA-B (4.8), and HLA-A (3.9) molecules (Fig. 3B; \(p < 0.0002\), Mann–Whitney \(U\) test, for all comparisons). Taken together, these results again suggest that Patr-B molecules recovered much more rapidly after the selective sweep via the generation of molecules with distinct binding motifs, whereas the functional repertoire of Patr-A molecules still carries clear signs of the selective sweep.

The HLA and Patr peptide binding repertoires are similar at the individual level

The results presented above suggest that the signs of the selective sweep are still visible when comparing the peptide binding repertoires of Patr with those of HLA molecules at the population level. However, infectious agents impose selection pressure at the individual level because the fitness of an individual is (partially) determined by how well it recovers from an infectious disease to survive long enough to give rise to offspring. Therefore, to determine the effect of the reduced peptide binding repertoire of Patr molecules at the individual level and per virus, we simulated MHC heterozygous individuals by randomly selecting two distinct MHC-A and two distinct MHC-B molecules. The peptide binding repertoires of individuals were then determined for all the nonredundant viruses in our data set (see Materials and Methods). For each virus, the number of binding peptides of a simulated human individual, having two HLA-A and two HLA-B molecules, was compared with the number of binding peptides of a simulated chimpanzee individual, having two Patr-A and two Patr-B molecules.

Comparing 100 human and 100 chimpanzee individuals generated this way, we found that for most of the viruses, the average number of binding peptides in simulated humans was higher (76%) than that in simulated chimpanzees (Supplemental Fig. 1), of which 25.2% were significant \((p < 0.01\), Mann–Whitney \(U\) test, corrected for multiple testing using Bonferroni). Only two viruses, T cell lymphotrophic virus type I (human and simian variant) and the human rhinovirus (four distinct types), had significantly more binding peptides in simulated chimpanzees than in simulated humans \((p < 0.01\); see Supplemental Table I). Of all the nonredundant viruses, we investigated the primate retroviruses in more detail because the selective sweep in chimpanzees was suggested to be caused by HIV-1/SIVcpz or a related ancestral retrovirus (20, 19) \((n = 20\); Supplemental Table I). Except for the T cell lymphotrophic virus type I, none of the retroviruses in our data set has convincingly more binding peptides in chimpanzees than in humans. However, the difference on the average number of binding peptides is rather small; for example, simian T cell lymphotrophic virus type I has 12% more binding peptides in chimpanzees, which corresponds to an additional nine peptides on average (74 HLA binders versus 83 Patr binders). Whether presenting a few more peptides would make a quantitative difference in the generation of T cell responses is rather unclear, and more
functional assays are necessary to support the claim that any of the viruses listed high up in Supplemental Table I could have caused the selective sweep.

A characteristic of the chimpanzee MHC region is the additional nonclassical MHC class I gene, Patr-AL. Having this additional Patr-AL molecule might have a significant contribution in generating T cell responses and could diminish the differences shown above. Recently, Gleimer et al. (34) showed that eluted peptides from the Patr-AL molecule are largely overlapping with those of HLA-A*02, suggesting that the functionality of the A02 supertype might have been maintained in the chimpanzee population. To test whether Patr-AL can close the “hole” in the peptide binding repertoire of chimpanzees, Patr-AL was included in our analysis. In line with the findings of Gleimer et al. (34), predicted Patr-AL binders are most similar to the predicted peptides of HLA molecules belonging to the A02 supertype (Fig. 1B). Nevertheless, Patr-AL shows the highest overlap of peptide binding repertoire with HLA-A*3201, a molecule not included in the study of Gleimer et al. (34) (38% overlap with HLA-A*3201 compared with 28% with HLA-A*0201; data not shown). To estimate the effect of Patr-AL as an extra MHC molecule in an individual, we simulated chimpanzees having an additional Patr-AL molecule with different peptide presentation efficiency levels (Supplemental Table II). Assuming an efficiency level similar to other MHC molecules, we found that chimpanzees reach a 12% larger peptide repertoire than humans. However, the Patr-AL molecule is known to have a lower expression level (6). Our results show that when the Patr-AL molecule binds as much as 40% of the peptides bound to a MHC molecule, the average number of binding peptides in simulated humans and chimpanzees becomes equal (Supplemental Fig. 1). This suggests that when the Patr-AL molecule had a rather efficient Ag presentation, the difference between the peptide binding repertoire of human individuals and that of chimpanzee individuals would disappear.

Discussion

Several previous studies applied phylogenetic analysis using genome sequences to suggest that chimpanzees experienced an ancient selective sweep affecting the MHC class I repertoire (5, 10, 6, 14). The loss of the A2 lineage in chimpanzees was straightforward to infer from genomic data, however de Groot et al. (14, 15) proposed that the selective sweep has been most pronounced in the Patr-B locus. To quantify the effect of such a selective sweep on the entire Patr peptide binding repertoire requires elution of peptides from every single contemporary Patr molecule and those of their human counterparts, which is extremely labor intensive. As an alternative, in this study, we used an in silico approach to compare the peptide binding repertoire of HLA and Patr molecules derived from a large set of mammalian viral proteomes ($n = 904$).

The primary outcome of this comparative approach is that on the population level the reduction in the peptide repertoire of Patr-B molecules is much less prominent compared with that of Patr-A molecules (Fig. 2). The contemporary Patr-A peptide binding repertoire contains clearly defined holes. For example, none of the contemporary Patr-A molecules shows a binding specificity similar to the A02 supertype (Fig. 1). Moreover, the loss of binding specificities was also observed, although to a lesser extent, for HLA-A*6601, A*6602, A*2501, A*2601, A*2602, and A*3401, which do not belong to the A02 supertype. Consistent with this finding, the functional diversity of Patr-A molecules is much smaller than that of the Patr-B molecules, reflected by the larger number of HMC molecules that bind the same peptide (Fig. 3B).

The effect of holes in the MHC peptide binding repertoire, created by the selective sweep, was also found at the individual level. For most viruses, simulated humans have more predicted binders compared with simulated chimpanzees (for 25% the difference is significant). However, with inclusion of Patr-AL, at an expression level of 40%, most of the difference between simulated humans and chimpanzees disappears. This suggests that if Patr-AL expression and functionality are efficient, the difference in the number of binding peptides between human and chimpanzee at the individual level can diminish. Recently, Gleimer et al. (34) demonstrated the functionality of Patr-AL in presenting peptides and in raising CD8+ T cell responses. Note, however, that in this study a strong promoter was used to induce high-level expression of Patr-AL molecules in transfected cells. In its natural genetic context, cell surface expression levels of Patr-AL have previously been shown to be low (6), which may affect its potential to present peptides. If so, this supports our conclusion that despite this additional MHC molecule, chimpanzees on average might have a reduced capacity to present a virus-specific peptide repertoire as compared with humans.

Our results are based on a computational method that predicts the binding affinities of peptides to MHC class I molecules and are, as such, dependent on the performance of the prediction method used. NetMHCpan shows a very high prediction performance for HLA-A and HLA-B molecules, whereas for chimpanzee MHC class I molecules the prediction performance is lower (22). Testing the performance of NetMHCpan on peptide sets eluted from several Patr molecules (20) revealed that only in one out of four cases NetMHCpan fails to predict the correct binding motif, whereas for the others the prediction accuracy is comparable to the HLA-A and HLA-B molecules (results not shown). Clustering analysis (as presented in Fig. 1B) suggests that low performance predictions overestimate the overlap in the peptide binding repertoire of MHC molecules (results not shown). This has two main implications for our results. First, we may underestimate especially the size of the Patr-B peptide binding repertoire, suggesting that the recovery of the Patr-B locus could even be larger than what we reported in this study. Second, we may overestimate the holes for both Patr-A and -B peptide binding repertoires. Note that NetMHCpan predicts the binding affinity of peptides to MHC class I molecules and does not provide information about how the actual CTL response is triggered by the pMHC complex. Because there are at the moment no prediction methods available for the interaction between CTLs and pMHC complexes, our results are limited to comparison of peptides binding to MHC molecules, but not of T cell responses between human and chimpanzee.

Previously, the selective sweep in chimpanzees was suggested to be caused by HIV-1/SIVcpz or a related ancestral retrovirus. The evidence used to support this hypothesis comes from the finding that both contemporary Patr molecules and HLA molecules associated with low viral load present similar regions in HIV-1/SIVcpz Gag (20, 19). Using a much less specific approach, that is, by performing a whole proteome analysis, we identified two extra viruses that have significantly more peptides that bind to Patr molecules (Supplemental Table I). However, for these viruses much more data on the effect of immunodominant CTL responses on disease progression and the pathogenicity in chimpanzees are needed to suggest them as possible agents causing the selective sweep. Additionally, the selective sweep and/or subsequent selection processes were suggested to have been more effective in West African chimpanzees than in other chimpanzee populations (35–37). Therefore, Patr class I repertoires could have been shaped differently depending on the chimpanzee subpopulations studied. Subsequently, a more specific analysis zooming in on Patr
molecules common in different chimpanzee populations may help to reveal the identity of the agent causing the selective sweep.

Acknowledgments

We thank Linda McPhee, Joost Beltman, and Lidija Berke for carefully reading the manuscript.

Disclosures

The authors have no financial conflicts of interest.

References

Figure S1: The ratio of the average number of binding peptides in 'simulated' chimpanzees (2 Patr-A + 2 Patr-B molecules) versus that of 'simulated' humans (2 HLA-A + 2 HLA-B molecules). The percentage of viruses, per interval of ratios, is given on the y-axis. The black line represents the ratio of the average number of uniquely binding peptides in chimpanzees to that in humans. Because humans bind more viral peptides in the majority of the viruses, the mean (dotted line) is at 0.95. Adding the Patr-AL molecule, capable of binding 40% of the peptides relative to other MHC molecules, to each chimpanzee shifts the mean to 1 (grey, dashed).

Table SI: List of primate retroviruses where we explicitly looked at the difference between the number of binding peptides in 'simulated' chimpanzees and humans. Patr/HLA defines the ratio of the average number of binding peptides in chimpanzees versus that of humans. p-values are calculated using the Mann-Whitney U test with Bonferroni correction, nonsignificant p-values (p>0.05) are indicated with ns. The Rhinoviruses that also showed a significant difference have accession numbers: DQ473491, DQ473512 and L24917.

Table SII: Ratio of the average number of binding peptides in 'simulated' chimpanzees versus that of humans, where chimpanzees have an additional Patr-AL molecule. The expression of Patr-AL is the peptide presentation efficiency relative to other Patr/HLA molecules.
This figure illustrates the percentage of viruses with Patr/HLA values as shown in Table SI. The x-axis represents the Patr/HLA values ranging from 0.7 to 1.2, while the y-axis represents the percentage of viruses. The solid line indicates the distribution of viruses with Patr/HLA values, and the dashed line represents another related distribution. The peak at 1.0 suggests a concentration of viruses with this specific Patr/HLA value.
<table>
<thead>
<tr>
<th>Retrovirus</th>
<th>Accession number</th>
<th>Patr/HLA p-value</th>
<th>p-value with Patr-AL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human T-cell lymphotropic virus type 1</td>
<td>U19949</td>
<td>1.14</td>
<td>0.006</td>
</tr>
<tr>
<td>Simian T-cell lymphotropic virus type 1</td>
<td>AF074966</td>
<td>1.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Human T-cell lymphotropic virus type 2</td>
<td>Y13051</td>
<td>1.05</td>
<td>ns</td>
</tr>
<tr>
<td>Mason-Pfizer monkey virus</td>
<td>AF033815</td>
<td>1.05</td>
<td>ns</td>
</tr>
<tr>
<td>Simian T-cell lymphotropic virus type 2</td>
<td>Y14570</td>
<td>1.04</td>
<td>ns</td>
</tr>
<tr>
<td>Simian type D retrovirus</td>
<td>M11841</td>
<td>1.03</td>
<td>ns</td>
</tr>
<tr>
<td>Human T-lymphotropic virus 3</td>
<td>DQ093792</td>
<td>1.02</td>
<td>ns</td>
</tr>
<tr>
<td>Human immunodeficiency virus type 1</td>
<td>AJ291718</td>
<td>1.00</td>
<td>ns</td>
</tr>
<tr>
<td>Simian retrovirus 2</td>
<td>AF126467</td>
<td>0.99</td>
<td>ns</td>
</tr>
<tr>
<td>Simian immunodeficiency virus (SIVcpz)</td>
<td>AF382828</td>
<td>0.96</td>
<td>ns</td>
</tr>
<tr>
<td>Primate T-lymphotropic virus 3</td>
<td>AF391797</td>
<td>0.96</td>
<td>ns</td>
</tr>
<tr>
<td>Simian foamy virus (SFVcpz)</td>
<td>U04327</td>
<td>0.96</td>
<td>ns</td>
</tr>
<tr>
<td>Spider monkey foamy virus</td>
<td>EU010385</td>
<td>0.96</td>
<td>ns</td>
</tr>
<tr>
<td>Simian immunodeficiency virus (SIVagm)</td>
<td>M29975</td>
<td>0.95</td>
<td>ns</td>
</tr>
<tr>
<td>Human spumaretrovirus</td>
<td>U21247</td>
<td>0.95</td>
<td>ns</td>
</tr>
<tr>
<td>Human immunodeficiency virus type 2</td>
<td>U27200</td>
<td>0.95</td>
<td>ns</td>
</tr>
<tr>
<td>Simian immunodeficiency virus (SIVmnd)</td>
<td>AF328295</td>
<td>0.91</td>
<td>ns</td>
</tr>
<tr>
<td>Squirrel monkey retrovirus</td>
<td>M23385</td>
<td>0.90</td>
<td>0.0455</td>
</tr>
</tbody>
</table>

1 http://www.ebi.ac.uk
<table>
<thead>
<tr>
<th>Virus</th>
<th>Accession Number</th>
<th>Score</th>
<th>E-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baboon endogenous virus</td>
<td>D10032</td>
<td>0.88</td>
<td>0.0015</td>
<td>ns</td>
</tr>
<tr>
<td>Gibbon leukemia virus</td>
<td>U60065</td>
<td>0.86</td>
<td>0.0012</td>
<td>ns</td>
</tr>
<tr>
<td>Human Rhinovirus</td>
<td>DQ473490</td>
<td>1.10</td>
<td>0.005</td>
<td>4.9e-8</td>
</tr>
<tr>
<td>Expression AL (%)</td>
<td>Ratio Patr/HLA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>1.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>