

A shared MHC supertype motif emerges by convergent evolution in macaques and mice, but is totally absent in human MHC molecules

Alessandro Sette · John Sidney · Scott Southwood ·
Carrie Moore · Jessica Berry · Courtney Dow ·
Kate Bradley · Ilka Hoof · Mark G. Lewis ·
William H. Hildebrand · Curtis P. McMurtrey ·
Nancy A. Wilson · David I. Watkins · Bianca R. Mothé

Received: 7 November 2011 / Accepted: 25 December 2011 / Published online: 10 February 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract The SIV-infected rhesus macaque (*Macaca mulatta*) is the most established model of AIDS disease

Electronic supplementary material The online version of this article (doi:10.1007/s00251-011-0598-5) contains supplementary material, which is available to authorized users.

A. Sette · J. Sidney · S. Southwood · C. Moore · K. Bradley ·
B. R. Mothé
Department of Vaccine Discovery,
La Jolla Institute for Allergy and Immunology,
La Jolla, CA 92037, USA

J. Berry · C. Dow · B. R. Mothé (✉)
Department of Biological Sciences,
California State University-San Marcos,
San Marcos, CA 92096, USA
e-mail: bmothe@csusm.edu

I. Hoof
Theoretical Biology & Bioinformatics, Utrecht University,
3584 CH Utrecht, The Netherlands

M. G. Lewis
Bioqual, Inc.,
9600 Medical Center Drive,
Rockville, MD 20850, USA

W. H. Hildebrand · C. P. McMurtrey
Department of Microbiology and Immunology,
The University of Oklahoma,
Oklahoma City, OK 73104, USA

N. A. Wilson · D. I. Watkins
Department of Pathology and Laboratory Medicine,
University of Wisconsin-Madison,
Madison, WI 53706, USA

systems, providing insight into pathogenesis and a model system for testing novel vaccines. The understanding of cellular immune responses based on the identification and study of Major Histocompatibility Complex (MHC) molecules, including their MHC:peptide-binding motif, provides valuable information to decipher outcomes of infection and vaccine efficacy. Detailed characterization of Mamu-*B*039:01*, a common allele expressed in Chinese rhesus macaques, revealed a unique MHC:peptide-binding preference consisting of glycine at the second position. Peptides containing a glycine at the second position were shown to be antigenic from animals positive for Mamu-*B*039:01*. A similar motif was previously described for the D^d mouse MHC allele, but for none of the human HLA molecules for which a motif is known. Further investigation showed that one additional macaque allele, present in Indian rhesus macaques, Mamu-*B*052:01*, shares this same motif. These “G2” alleles were associated with the presence of specific residues in their B pocket. This pocket structure was found in 6% of macaque sequences but none of 950 human HLA class I alleles. Evolutionary studies using the “G2” alleles points to common ancestry for the macaque sequences, while convergent evolution is suggested when murine and macaque sequences are considered. This is the first detailed characterization of the pocket residues yielding this specific motif in nonhuman primates and mice, revealing a new supertype motif not present in humans.

Keywords HLA supertype · MHC · Peptide-binding motif · Rhesus macaque

Introduction

Rhesus macaques have been extensively used in infectious disease research. This model provided key insights into disease pathogenesis and allowed for the evaluation of novel vaccine concepts. Two populations of rhesus macaques, consisting of Indian-origin rhesus macaques and Chinese-origin rhesus macaques, have been utilized extensively in AIDS research and for other models of infectious diseases (Gardner and Luciw 2008; Giraldo-Vela et al. 2008; Miller et al. 1989; Smith et al. 1999). While these two sets of animals appear to be indistinguishable physiologically, genetic factors affecting immune responses appear to be quite distinct. This is particularly true in terms of the Major Histocompatibility Complex (MHC) class I molecules, which determine the repertoire of T-cell responses that an individual can develop against Simian Immunodeficiency Virus (SIV) and/or any other foreign pathogen (Parham 2005). Recent studies have shown that the MHC gene regions of Indian and Chinese macaques differ appreciably in terms of the degree of polymorphism, the specific allelic variant, and their functional characteristics (Solomon et al. 2010; Southwood et al. 2011; Wiseman et al. 2009).

In previous studies, we have investigated the MHC:peptide-binding repertoire of several MHC class I and II molecules in Indian-origin rhesus macaques, such as Mamu-A*01 (Allen et al. 1998), -B*17 (Mothe et al. 2002), -B*01 (Loffredo et al. 2005), -DRB1*0406, and -DRB*w201 (Dzuris et al. 2001) among others (Dzuris et al. 2000; Giraldo-Vela et al. 2008; Loffredo et al. 2008; Loffredo et al. 2007; Loffredo et al. 2004; Reed et al. 2011; Sette et al. 2005). While Chinese rhesus macaques are of value as animal models for AIDS vaccine development and for other pathogens, their full potential has not been realized due to missing functional MHC and genetic information.

We have recently started to define peptide-binding motifs for common Chinese macaque MHC class I molecules and to date have elucidated detailed quantitative motifs for three alleles: Mamu-A1*026:01, Mamu-B*083:01 (Southwood et al. 2011), and Mamu-A1*022:01 (Solomon et al. 2010). Interestingly, we found that all of these molecules are associated with motifs that are overlapping with well-known HLA supermotifs. Specifically, Mamu-A1*026:01 shares a high degree of cross-reactivity with the HLA-A2 supertype allele HLA-A*02:02, as well as -A*02:01 and -A*02:03, while Mamu-B*083:01 is highly cross-reactive with HLA-A3 supertype alleles HLA-A*31:01, -A*03:01, and -A*68:01 (Southwood et al. 2011). Additionally, Mamu-A*022:01 is highly cross-reactive with the HLA-B7 supertype alleles B*07:02 and B*35:01 (Solomon et al. 2010). The HLA-B7, -A3, and -A2 supertypes are the three most abundant supertypes in the human population, each present

with a phenotypic frequency of greater than 40%, averaged across various ethnic groups (Sette and Sidney 1998; Sette and Sidney 1999).

To fully realize the potential of Chinese-origin macaques in infectious disease research, further investigation and characterization of the most commonly expressed MHC alleles is necessary. This will facilitate accurate assessment of cellular immunity and immune correlates of protection. In the present set of experiments, we focused on a high-frequency allele in Chinese rhesus macaques, Mamu-B*039:01, which is present in 5.8% of Chinese rhesus macaques (Solomon et al. 2010). We sought to characterize Mamu-B*039:01 by determining its specific MHC:peptide-binding motif. We were surprised to find that its motif consisted of a preference for glycine (G) at position 2. Analysis of published literature revealed that only the murine class I molecule H-2 D^d has been reported to share this specificity. The analysis of known Mamu and HLA sequences resulted in the definition of structural characteristics associated with the preference for glycine in position 2 and led us to discover another macaque allele that we subsequently demonstrated to be associated with similar peptide-binding specificity. Structural and phylogenetic analysis suggested that this MHC motif is not found among HLA alleles. Hence, herein we propose a new supertype motif present in macaques and mice, but not in humans.

Materials and methods

Creation of stable Mamu-B*039:01 and Mamu-B*052:01 transfectant cell lines

To produce secreted Mamu molecules in the context of endogenous ligand identification, α -chain complementary DNAs (cDNAs) of Ch-Mamu-B*039:01 and In-Mamu-B*052:01 were modified at the 3' end by PCR mutagenesis to delete codons 5–7 encoding the transmembrane and cytoplasmic domains and to add a 30 bp tail encoding the ten amino acid rat very low-density lipoprotein receptor (VLDLr), SVVSTDDDLA, for purification purposes (Hickman et al. 2000). sMHC-VLDLr were cloned into the mammalian expression vector pcDNA3.1 (Invitrogen). The MHC class I-deficient EBV-transformed B-lymphoblastoid cell line 721.221 cells were transfected with sMHCs Mamu-B*39TVLDLr and Mamu-B*52TVLDLr by electroporation. After 48 h of incubation, the cells were plated in 96-well plates (Falcon) in RPMI 1640 containing the antibiotic Geneticin. Transfectants were tested for the production of sMHC molecules by a VLDLr-specific ELISA (Hawkins et al. 2008).

Mamu endogenous ligand determination

Separately, approximately 25 mg of Mamu-B*39TVLdLr and Mamu-B*52TVLdLr molecules from the 721.221 cell line were purified over an affinity column composed of anti-VLdLr antibody (ATCC clone CRL-2197) coupled to CNBr-activated Sepharose 4B (GE Healthcare, Piscataway, NJ). sMHC molecules were then eluted in 0.2 N acetic acid, brought up to 10% acetic acid, and heated to 76°C for 10 min. Peptides were separated from heavy and light chains by ultra-filtration in a stirred cell with a 3-kDa molecular weight cutoff cellulose membrane (Millipore, Bedford, MA). The peptide batch was flash frozen and lyophilized. The peptides were then reconstituted in 10% acetic acid.

Following isolation, 10% of the peptide pool was subjected to 14 rounds of N-terminal sequencing by Edman degradation. A motif was generated by calculating the fold increase of each amino acid over the prior round. A hierarchy was then determined based on the amino acid composition at each position (Falk et al. 1991).

Peptides were reverse-phase HPLC fractionated using a Jupiter Proteo C12 column (Phenomenex, Torrance, CA) on a Paradigm MG4 system (Michrom Bioresources, Auburn, CA). A standard CH₃CN gradient was employed to generate approximately 40 peptide-containing fractions. UV absorption was monitored at 215 nm. Peptide fractions were concentrated to dryness and reconstituted in 20 µl of nanospray buffer composed of 50% methanol, 50% H₂O, and 0.5% acetic acid. Nano-electrospray capillaries (Proxeon, Denmark) were loaded with 1 µl of each peptide fraction and infused at 1,100 V on a Q-Star Elite quadrupole mass spectrometer with a time of flight detector (Applied Biosystems, Foster City, CA). Ion maps were generated for each fraction in a mass range of 300–1,200 amu. Using independent data acquisition for selection, ions (putative peptides) were fragmented by tandem mass spectrometry (MS/MS). An amino acid sequence was assigned using the publicly available, web-based MASCOT (Matrix Science Ltd., London, UK) and/or de novo sequencing.

Positional scanning combinatorial library and peptide synthesis

Positional scanning combinatorial libraries (PSCLs) were synthesized as previously described (Pinilla et al. 1999). In the PCSL, each pool in the library contains randomized 9-mer peptides with one fixed residue at a single position. With each of the 20 naturally occurring residues represented at each position along a 9-mer backbone, the entire library consisted of 180 peptide mixtures.

Peptides utilized in screening studies were purchased as crude or purified material from Mimotopes (Minneapolis, MN, USA/Clayton, Victoria, Australia), Pepscan Systems

B. V. (Lelystad, Netherlands), A&A Labs (San Diego, CA, USA), Genescript Corporation (Piscataway, NJ, USA), or the Biotechnology Center at the University of Wisconsin-Madison (Madison, WI, USA). Peptides synthesized for use as radiolabeled ligands were synthesized by A&A Labs and purified to >95% homogeneity by reverse-phase HPLC. Peptide purity was determined with analytical reverse-phase HPLC and amino acid analysis, sequencing, and/or mass spectrometry. Peptides were radiolabeled utilizing the chloramine T method (Sidney et al. 2001). Lyophilized peptides were resuspended at 20 mg/ml in 100% DMSO, then diluted to required concentrations in PBS+0.05% (v/v) nonidet P40 (Fluka Biochemika, Buchs, Switzerland). SIV peptide sequences were derived from the SIVmac239 sequence, Genbank accession M33262 (Kestler et al. 1990).

MHC purification and peptide-binding assays

Mamu class I MHC purification was performed by affinity chromatography using the W6/32 and/or B123.2 class I antibodies, as previously described (Allen et al. 2001; Loffredo et al. 2009; Sidney et al. 2005). Protein purity, concentration, and depletion efficiency steps were monitored by SDS-PAGE.

Quantitative assays for peptide binding to detergent-solubilized MHC class I molecules were based on the inhibition of binding of a high affinity radiolabeled standard probe peptide and performed as detailed in prior studies (Loffredo et al. 2009; Schneidewind et al. 2008; Sidney et al. 2001; Sidney et al. 2005). Peptides were tested at six different concentrations covering a 100,000-fold dose range in three or more independent assays. Control wells to measure nonspecific (background) binding were also included. In each experiment, a titration of the unlabeled version of the radiolabeled probe was also tested as a positive control for inhibition.

The radiolabeled peptides utilized for the Mamu-B*039:01 and Mamu-B*052:01 assays were 3422.03 (sequence YGFSDPLTF) and 3289.0028 (VGNVYVKF), respectively, representing sequences identified by Edman degradation and mass spectrometry analysis (described above). For each peptide, the concentration of peptide yielding 50% inhibition of the binding of the radiolabeled probe peptide (IC₅₀) was calculated. Under the conditions used, where [radiolabeled probe] < [MHC] and IC₅₀ ≥ [MHC], the measured IC₅₀ values are reasonable approximations of the true K_d values (Gulukota et al. 1997; Sette et al. 1994).

PCR-SSP

Peripheral blood mononuclear cells (PBMCs) from two SIVmac239 challenged and infected Chinese rhesus macaques were supplied by Bioqual (Rockville, MD). The total

RNA was extracted from ten million PBMCs using the Qiagen Qias shredder tissue homogenizer and the AllPrep DNA/RNA Mini Kit catalog #80204 and the manufacturer's protocol (Qiagen, Valencia, CA, USA). First-strand cDNA was synthesized for each sample from 50 ng of RNA using the Super Script III First-Strand Synthesis for RT-PCR kit catalog #18080-51 (Invitrogen, Carlsbad, CA, USA). The cDNA was then used in MHC typing utilizing sequence-specific primers in polymerase chain reaction (PCR-SSP) as previously described (Karl et al. 2008). The presence of a Mamu-B*039:01 specific 500 bp in length product was visualized for both animals by agarose gel electrophoresis.

IFN- γ ELISPOT assay

We performed enzyme-linked immunosorbent spot (ELISPOT) assays to assess SIV-specific T-cell responses, as described previously (Allen et al. 2001). Briefly, 1×10^5 PBMCs were used per well in precoated ELISpot^{PLUS} kits (Mabtech, Mariemont, OH) according to the manufacturer's instructions for the detection of IFN- γ -secreting cells. All tests were performed in triplicate using individual peptides at 10 μ M. The positive control, Con A (Sigma-Aldrich), was used at a final concentration of 5 μ g/ml. The negative control wells were devoid of any stimulation. The 96-well plates were incubated for 20 h at 37°C in 5% CO₂. Wells were imaged and counted with Zeiss KS EliSpot reader and analyzed. Spots were counted by an automated system with set parameters for size, intensity, and gradient. Background (the mean of wells without peptide stimulation) levels were subtracted from each well on the plate. Responses were considered positive if the mean of the number of SFC was more than 50 spots per million cells and the $p \leq 0.05$. The level of statistical significance was determined with a Student's *t* test using the mean of triplicate values of the response against individual peptides versus the response against the negative control.

Bioinformatic analysis

Analysis of the PSCL data was performed as described previously (Sidney et al. 2008). Briefly, IC₅₀ nM values for each mixture were standardized as a ratio to the geometric mean IC₅₀ nM value of the entire set of 180 mixtures and then normalized at each position so that the value associated with optimal binding at each position corresponds to 1. For each position, an average (geometric) relative binding affinity (ARB) was calculated, and then, the ratio of the ARB for the entire library to the ARB for each position was derived. We have denominated this ratio, which describes the factor by which the normalized geometric average binding affinity associated with all 20 residues at

a specified position differs from that of the average affinity of the entire library, as the specificity factor (SF). As calculated, positions with the highest specificity will have the highest SF value. Primary anchor positions were then defined as those with an SF ≥ 2.4 . This criterion identifies positions where the majority of residues are associated with significant decreases in binding capacity.

Secondary anchor designations were based on the standard deviation (SD) of residue specific values at each position. Dominant secondary anchor positions were defined as those where the SD was >3 and the SF <2.4 , as well as positions associated with an SD >2 if the SF is between 1.5 and 2.4. Weak secondary anchors have been defined as positions associated with a SD in the 2.5–3 range with an SF <1.5 , or an SF in the 1.5–2.4 range with an SD <2 .

For single amino acid substitution (SAAS) panels, larger SD values are inherent comparable to PSCL, and therefore, more stringent criteria were employed for defining primary and secondary anchor positions. For SAAS, a primary anchor position is defined as one in which the SF ≥ 3.5 . Secondary anchor positions were defined as those where the SD was >12 with an SF <3.5 or at positions with an SD >10 and an SF in the 1.5–3.5 range.

To identify predicted binders, all possible 9-mer peptides in SIVmac239 sequences were scored using the matrix values derived from the PSCL analyses of Mamu-B*039:01 and Mamu-B*052:01 (Sidney et al. 2008). The final score for each peptide represents the product of the corresponding matrix values for each peptide residue–position pair. Peptides scoring among the top 3.0% ($n=100$) were selected for binding analysis.

Phylogenetic reconstruction

We obtained a multiple sequence alignment of 266 full-length Mamu-B nucleotide sequences from the IPD-MHC database release 1.6.0 (Robinson et al. 2010). The human allele HLA-B*07:02 was used as an outgroup. From the aligned MHC sequences, we deleted 51 codons encoding residues that form the binding pockets of the MHC molecules in order to avoid binding specificity-determining positions biasing the phylogeny. For phylogenetic reconstruction, we used the maximum-likelihood-based method Phyml (Guindon and Gascuel 2003) with the HKY model of nucleotide substitution (Hasegawa et al. 1985), a maximum-likelihood estimate of the proportion of invariable sites, four discrete gamma-distributed substitution rate categories, and 1,000 bootstrap runs. The HKY model was selected based on jModeltest (Posada 2008). The phylogenetic trees were visualized using SplitsTree (Huson and Bryant 2006) and Dendroscope (Huson et al. 2007), respectively.

Results

Preference for glycine in position 2 of peptides binding to Mamu-B*039:01

We were interested in defining the MHC:peptide-binding motif for Mamu-B*039:01 because, with a reported frequency of 5.8% in the captive-bred population (Solomon et al. 2010), it represents one of the most frequent Chinese rhesus macaque alleles whose binding specificity has yet to be characterized. Previous studies have demonstrated that the elution and characterization of naturally bound ligands is an effective method for determining the peptide-binding specificity of class I MHC molecules (Hickman-Miller et al. 2005; Kubo et al. 1994). Accordingly, we evaluated the natural ligands associated with soluble Mamu-B*039:01 by sequencing 21 different endogenously loaded Mamu-B*039:01 peptide ligands representing six different peptide fractions (Table 1). The majority of ligands (14 of 21; 67%) were nine residues in length. Ligands of ten residues ($n=4$) were also identified. Ligands of 8, 11, and 12 residues were only rarely found (one instance each). These observations suggest that nonamers represent the preferred size for peptide ligands bound to Mamu-B*039:01.

Table 1 Endogenous Mamu-B*039:01 ligands identified by MS/MS sequencing analysis

Fraction	Sequence	Length	Binding affinity (IC ₅₀ nM)
47	RGVEKPPHL	9	1.0
52	SGYEGRVPL	9	4.8
52	KGAGTGGLGL	10	1.1
52	RGAGTGGLGL	10	4.8
57	IGVSHPLSI	9	63
57	SGFPAKVTAHW	11	1.3
57	SGKLVPEW	9	19
57	KGfvrenvw	9	0.20
57	RGTKVILHL	9	7.1
60	AGVfvprsf	9	0.62
60	IGYPKALL	9	0.13
60	LGGVAVGI	8	8.0
60	KKYDVPLSW	9	1.8
68	QGFPTIKIF	9	2.5
68	KGYGLGSGISLF	12	0.45
68	VGLGPALEF	10	1.3
70	SGFWKALTF	9	0.63
70	AGIFPRLSF	9	0.49
70	AGFRALGLF	9	0.31
70	QGFLVfHSF	9	0.12
70	YGfSDPLTF	10	1.2

Next, we aligned the endogenously bound peptides and tabulated the frequency of various amino acid residues at each position (Fig. 1). The results indicated an overwhelming preference for glycine (G) in position 2, as demonstrated by its presence in 20 of the 21 natural ligands (95%). At the C-terminus, aromatic residues phenylalanine (F) and tryptophan (W) accounted for 62% of residues observed (13/21) and the aliphatic residues leucine (L) and isoleucine (I) accounted for the remaining eight (38%). This information identifies position 2 and the C-terminus as the likely main anchor positions for Mamu-B*039:01 binding.

To confirm that the sequenced peptides were indeed natural ligands of B*039:01, a binding assay was developed using purified MHC, as described in the “Materials and methods” section. Using this assay, we were able to confirm that the eluted ligands, the basis of the preliminary motif, do indeed bind Mamu-B*039:01 with high affinity. Specifically, 20 of the 21 endogenous peptides identified bound with IC₅₀s of 50 nM or better.

Derivation of a detailed quantitative motif

The Mamu-B*039:01 binding capacity of a nonamer PSCL was determined next. As shown in Fig. 2, the PSCL analysis confirmed the preliminary motif defined on the basis of the pool sequencing data. Specifically, using the criteria outlined in the “Materials and methods” section, position 2 and the C-terminus were identified as the primary binding anchors, with G being dominant at position 2 and aromatic residues F and W preferred at the C-terminus.

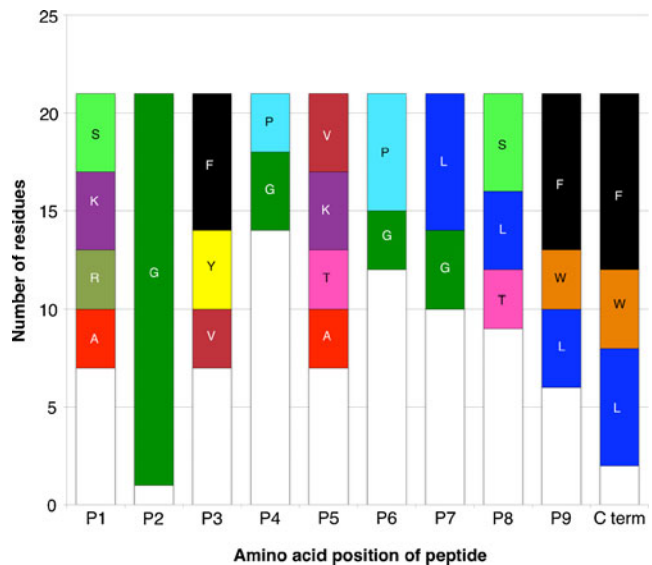


Fig. 1 Residue frequency summary of sequenced Mamu-B*039:01 ligands. Specific amino acids are indicated and color-coded if present at a given position in greater than 10% of ligands identified. White bars represent all other unspecified residues

Residue	Position								
	1	2	3	4	5	6	7	8	9
A	0.260	0.057	0.038	0.480	0.288	0.280	0.953	1.000	0.009
C	0.161	0.003	0.015	0.179	0.254	0.165	0.729	0.068	0.083
D	0.009	0.001	0.021	0.311	0.015	0.038	0.037	0.128	0.003
E	0.020	0.002	0.007	0.037	0.092	0.028	0.024	0.033	0.012
F	0.065	0.008	0.799	0.673	0.121	0.236	0.226	0.173	1.000
G	0.084	1.000	0.041	0.252	0.109	0.226	0.131	0.648	0.012
H	0.233	0.008	0.038	0.292	0.059	0.199	0.281	0.062	0.005
I	0.270	0.009	0.071	0.224	0.103	0.323	0.237	0.235	0.330
K	1.000	0.028	0.017	0.400	0.039	0.155	0.124	0.033	0.119
L	0.328	0.004	0.030	0.607	0.058	0.233	0.335	0.086	0.151
M	0.580	0.013	0.076	0.461	1.000	0.170	0.245	0.102	0.032
N	0.057	0.002	0.007	0.604	0.047	0.312	0.181	0.150	0.005
P	0.009	0.003	0.004	0.407	0.216	1.000	1.000	0.224	0.003
Q	0.114	0.027	0.007	0.548	0.143	0.208	0.213	0.204	0.003
R	0.645	0.013	0.040	0.310	0.138	0.138	0.645	0.187	0.007
S	0.154	0.013	0.022	0.316	0.109	0.339	0.296	0.412	0.003
T	0.106	0.006	0.047	0.337	0.160	0.312	0.188	0.352	0.003
V	0.202	0.003	0.096	0.277	0.142	0.268	0.293	0.274	0.012
W	0.127	0.002	0.049	1.000	0.150	0.196	0.050	0.166	0.895
Y	0.095	0.003	1.000	0.801	0.096	0.124	0.136	0.145	0.016
Geomean	0.12	0.010	0.040	0.36	0.11	0.20	0.21	0.16	0.020
SD	3.5	4.8	4.1	2.0	2.4	2.1	2.7	2.4	7.0
SF	0.67	10.8	2.33	0.23	0.73	0.43	0.39	0.51	4.11

Fig. 2 PSCL-derived matrix describing 9-mer binding to Mamu B*039:01. The PSCL was tested for binding, the data analyzed, and primary and secondary anchor positions defined, as described in the “Materials and methods”. Values shown represent the average relative binding (ARB) of the corresponding library relative to other pools with the same fixed position. Values have been normalized to the optimal residue at the corresponding position. *SD* indicates the standard deviation between the ARB of pools at the same position. *SF* is the

specificity factor, calculated as described in the “Materials and methods” representing the ratio of the average binding of the entire library to the average of pools at the indicated position. At the primary anchor positions (*SF* > 2.4; *blue shading*), the most preferred residues, associated with an ARB > 0.1 are highlighted by *bold yellow font*. *Green shading* highlights secondary anchor positions. The library average binding for Mamu-B*039:01 was 1,772 nM

Aliphatic, hydrophobic residues I and L were also tolerated at the C-terminus, as well as positively charged lysine (K). Additionally, positions 1 and 3 were identified as dominant secondary anchors. A pictorial summary representation of the Mamu-B*039:01 motif is shown in Fig. 3a.

Peptides with a glycine at the second position are antigenic in Mamu-B*039:01 animals

To validate the proposed motif at the functional level, we analyzed cellular responses from two Chinese rhesus macaques previously infected with SIVmac239 and positive for Mamu-B*039:01 by PCR-SSP, as detailed in the “Materials and methods” section. Cellular immune responses in these animals were evaluated using IFN- γ ELISPOT assay and SIVmac239-derived peptides. Specifically, the PCSL data were used to generate a predictive algorithm by standard methods (Southwood et al. 2011). SIVmac239 sequences scoring in the upper 3.0% range ($n=100$) were synthesized and tested for in vitro MHC:peptide-binding affinity, 40 of which bound Mamu-B*039:01 with an IC_{50} < 500 nM.

These peptides were tested in IFN- γ ELISPOT assays using cells from B*039:01 SIVmac239-infected animals to identify antigenic peptides (Table 2). Of the 40 SIV peptides tested, 28 contained a glycine at the second position.

In terms of responses, a total of 8 of the 40 peptides generated positive responses in at least one animal (see Table 2). Specifically, ChRh 1 had three peptides with average SFCs of >50. ChRh 2 had five peptides with an average SFC of >50. Out of those eight responses, three responses were induced by peptides which had a glycine at position 2, indicating that these peptides are antigenic in the setting of SIV infection.

Similarities of the B39 motif with the known H-2 D^d motif

Given the unusual preference of a glycine at the second position of the MHC:peptide-binding motif, we investigated whether other MHC molecules might also possess this particular characteristic. A thorough literature search revealed that the only other known MHC class I molecules with a reported preference of glycine at position 2 specificity were

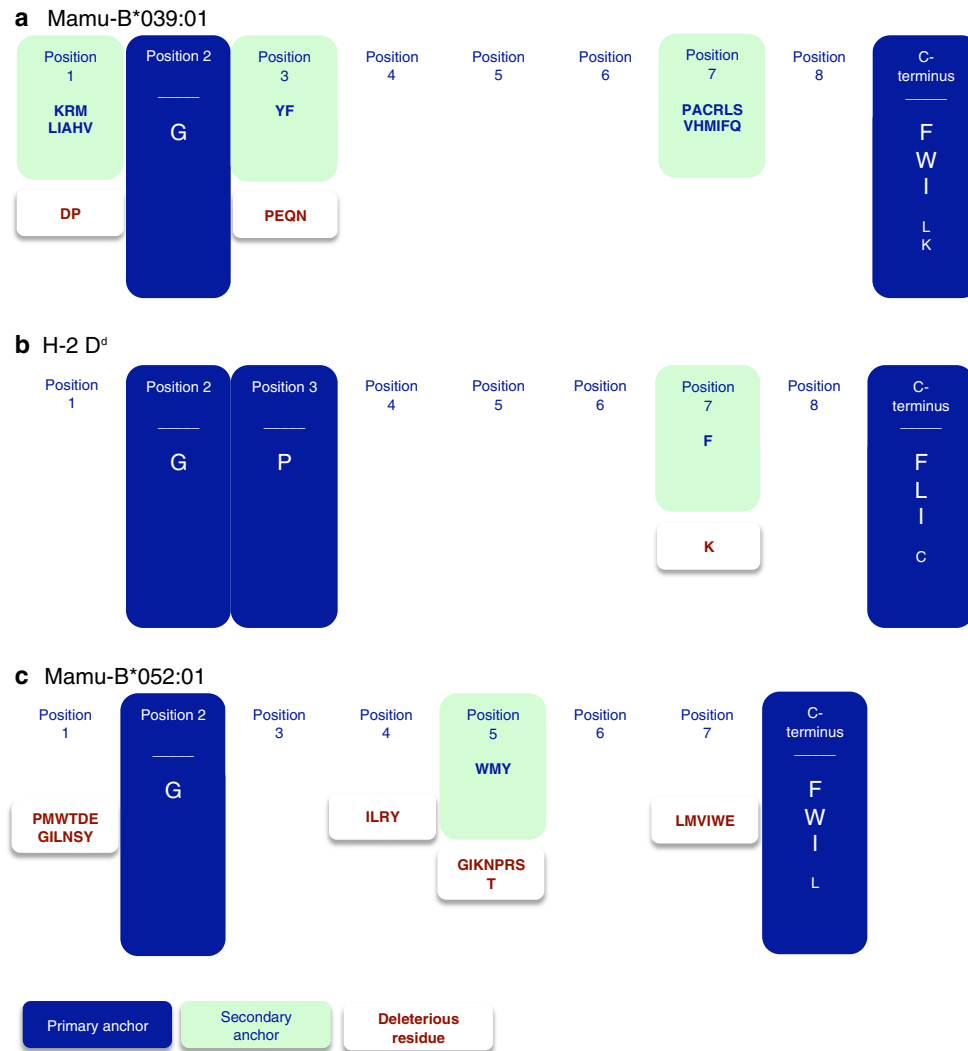


Fig. 3 Map of the **a** Mamu-B*039:01, **b** H-2 D^d, and **c** Mamu-B*052:01 motif. Pictorial summary of the PSCL matrices, indicating primary (in blue) and secondary (in green) anchor positions with

associated preferred and tolerated residues. Also represented are those residues that are deleterious to binding at indicated positions (brown font)

the murine D^d MHC molecule (Corr et al. 1993; Li et al. 1998) and the Mamu-B*004 molecule (Dzuris et al. 2000). Previously published data for H-2 D^d demonstrated a strong requirement for G in position 2, along with proline (P) in position 3 and hydrophobic residues at the C-terminus (Fig. 3b) (Corr et al. 1993). Data for Mamu-B*004 likewise showed a specificity for G in position 2, but displayed no strict dependence on a C-terminal anchor (Dzuris et al. 2000).

Similarities of the Mamu-B*039:01 motif with the Mamu-B*052:01 motif

Concurrently with these experiments, Mamu-B*052:01, expressed in 8.4% Indian rhesus macaques, was also being characterized. In a similar fashion to Mamu-B*039:01, we

eluted and sequenced peptides from this MHC molecule. Sixteen peptides, consisting of 8-mers, 9-mers, and 10-mers, were eluted and characterized. The most frequent size observed was 8-mers, representing half of the endogenous peptides identified. The other half was split equally between 9- and 10-mers at four peptides (25%) each. Similar to Mamu-B*039:01, a dominant preference for G at position 2 was noted, as all peptides possessed this residue (Table 3).

The motif was further investigated by assessing the binding capacity using a PSCL (Fig. 4), as utilized in Mamu-B*039:01 characterization. Analyzing the data as described in the “Materials and methods” section, position 2 was identified as a secondary anchor position, where G was identified as a preferential, albeit not dominant residue. At this point, it was noted that 8-mer peptides were the dominant species for Mamu-B*052:01, comprising half of

Table 2 SIVmac239 peptides in Mamu-B*039:01-positive Chinese rhesus macaques

Sequence	Protein	Position	Mamu-B*039:01 IC ₅₀ nM	SFC's 50 or greater
KAVATAPGL	Tat	120	0.17	
AGFLGLGPW	Gag	432	0.25	
LGIPHPAGL	Pol	300	0.89	
RGETYGRLL	Nef open	24	1.0	
RGFAAPQFS	Pol	105	1.3	
QGQYMNTPW	Nef open	62	1.4	
VAVHVASGF	Pol	842	2.8	
KGIGGNQEI	Pol	747	4.4	
QGWKGSPI	Pol	359	6.7	
KGLGVNPTL	Gag	336	8.0	ChRh 2
STYAVRITW	Vif	83	8.8	ChRh 2
QGIRQVLF	Pol	761	9.1	
KRGVFLVGF	Env	524	15	ChRh 1
SGFIEAEVI	Pol	848	15	
RGVFLVGLF	Env	525	19	
KARLMAEAL	Gag	360	36	
KGWLSTYAV	Vif	79	46	
TVFYGVPAW	Env	27	46	
SGFMPKCSK	Env	249	47	
QGYWHLTPE	Vif	70	56	ChRh 2
GGGNPLSAI	Vpr	86	65	
KGSPAIFQY	Pol	362	66	ChRh 1
GGGNSSWPW	Env	756	93	
RGGRWILAI	Env	858	99	
GGFFRPWSM	Pol	40	173	
AGRWPITHL	Pol	872	181	
FGAEVVPGF	Gag	166	190	
LGMSLNFP	Pol	202	231	
KGANFPGLA	Vif	199	249	
TGRQTALFL	Pol	860	307	
SLYNTVCVI	Gag	77	314	
IGFSTPEEK	Pol	420	346	
KHYWDAIRF	Env	222	391	
LGFLATAGS	Env	533	415	
KIVGGIGGF	Pol	154	419	ChRh 2
PGIRYPKTF	Nef open	163	444	
LKNYMQLGK	Gag	470	446	ChRh 2
RQLIRLLTW	Env	774	474	ChRh 1
KEALAPVPI	Gag	369	486	
FGFNGTRAE	Env	275	495	

IFN- γ ELISPOT assays were used to evaluate CD8+ T-cell responses in two Mamu-B*039:01-positive animals using Mamu-B*039:01 high-affinity peptide binders. Peptides derived from SIVmac239 were tested for in vitro MHC:peptide-binding affinity against Mamu-B*039:01. SIVmac239 peptides with binding affinity <500 nM are shown. Peptide sequences in bold font indicate a positive response with a glycine at the second position. Responses were considered positive if the mean of the number of SFC was more than 50 spots per million cells and the $p \leq 0.05$

endogenous peptides sequenced. The PSCL, however, was based on 9-mers. To ensure thorough characterization of the motif, and since an 8-mer PSCL was not available, we tested a set of SAAS of an 8-mer ligand eluted from Mamu-B*052:01 (sequence VGNVYVKF; IC₅₀ 7.4 nM). This method identified position 2 as a primary anchor with a dominant preference for G (Fig. 5), with every other

residue associated with a greater than two orders of magnitude reduction in binding capacity. Similarly to the case of Mamu-B*039:01, the C-terminus was also identified as a primary anchor position, with aromatic and aliphatic residues being preferred. Position 5 was identified as a dominant secondary anchor position by definition. The summary motif for Mamu-B*052:01 is shown in Fig. 3c.

Table 3 Endogenous Mamu-B*052:01 ligands identified by MS/MS sequencing analysis

Fraction	Sequence	Length	Binding affinity (IC ₅₀ nM)
55	KGFSEEHNTW	10	0.77
54	AGAAALRPFA	10	11
59	TGYVATRW	8	41
59	KGFTDADNTW	10	1.5
59	SGGPKYEYRW	10	1.7
62	VGNVYVKF	8	7.4
62	LGRPDAEYW	9	5.2
62	RGPDAFRF	8	134
62	LGRPDEEYW	9	12
62	NGFQEETW	8	3.2
64	AGVWSQDKW	9	23
64	IGHRYIEVF	9	3.4
67	YGPKVDIW	8	12
72	HGVEFDI	8	0.31
72	VGPRWMEW	8	1.3
72	EGPQYWEW	8	4.8

A unique B pocket structural motif is shared between Mamu-B*039:01, H-2 D^d, and Mamu-B*052:01

Based on the crystal structures of MHC class I molecules (Achour et al. 1998; Li et al. 1998; Madden 1995; Saper et al. 1991), the MHC B pocket is hypothesized to interact with the residue in position 2 of peptide ligands. Previously (Sette et al. 2003), we had investigated similarities in position 2 peptide-binding specificity between human and mouse class I molecules and found that the similarities were associated with very different B pocket structures, reflecting convergent evolution. Given the similar and unusual preferences of Mamu-B*052:01, Mamu-B*039:01, and murine H-2 D^d for G in position 2, we next investigated whether a common structural motif could be identified. We further sought to determine if the shared specificity would be reflective of convergent evolution or the result of common ancestry.

Accordingly, we analyzed the residues in positions 7, 9, 24, 34, 45, 63, 66, 67, 70, and 99 which compose the B pocket of MHC class I molecules (Achour et al. 1998; Li et al. 1998; Madden 1995; Saper et al. 1991) of Mamu-B*039:01, H-2 D^d, and Mamu-B*052:01. Surprisingly, this analysis revealed that all three alleles had very similar B pocket structures (Table 4). Specifically, Mamu-B*039:01, H-2 D^d, and Mamu-B*052:01 shared identical residues in positions 7, 24, 34, 63, 66, and 67, where tyrosine (Y), glutamic acid (E), valine (V), E, arginine (R), and alanine (A) were found, respectively. With the exception of an A/Y dichotomy in position 99, in which alanine is a small

nonpolar residue and tyrosine is a large polar aromatic, variations at other positions were more conservative in nature.

To determine if this structural motif was found in other primate MHC class I alleles, we next compiled the corresponding B pocket residues for all other available full Mamu and HLA sequences. This analysis revealed that the residues present in positions 24 (E) and 66 (R) of Mamu-B*039:01, H-2-D^d, and Mamu-B*052:01 were relatively frequent in other macaque class I sequences (Table 4). In macaques, R was present in position 66 in 13% and E in position 24 in 15% of the 650 Mamu alleles analyzed; as a tandem, they were found in an appreciable 5.7% of the macaque sequences. As a whole, six Mamu alleles (Mamu-B*004:01, Mamu-B*029:01:01, Mamu-B*029:01:02, Mamu-B*029:02, Mamu-B*029:03, and Mamu-B*032:01) matched the entire consensus sequence shown in Table 4. By contrast, these residues, as a pair or individually, were not found in any of the over 945 HLA alleles analyzed (A in position 67, fairly common in macaques, was also not found in any HLA sequence). These differences in frequency are highly significant ($p < 0.0001$ by the Fisher's exact test). These observations demonstrate that the structural similarities associated with the specificity for G in position 2 are not uncommon in macaques while this type of peptide motif is not found in humans.

Taken together, the data in this section have defined a common B pocket structural motif for alleles with specificity for G in position 2. This structural motif was associated with the presence of E in position 24 and R in position 66. The presence of these residues in the corresponding positions is found in appreciable frequencies in macaques, at both the phenotypic and genotypic levels, and absent in humans.

Evolutionary origin of the pocket

We next examined whether the G2 structural motif in macaques might be the result of convergent evolution of Mamu-B*039:01 and Mamu-B*052:01 or a reflection of common ancestry. For this analysis, we reconstructed a phylogenetic tree for all available full-length Mamu allele sequences, five common mouse alleles, and a set of common HLA alleles (online resource 1). As expected, the mouse alleles are only distantly related to the primate MHC alleles, confirming that the emergence of the G2 motif in macaques and mice is the result of convergent evolution. Mamu-B*039:01 and Mamu-B*052:01, however, cluster together, suggesting that the shared structural motif is due to a common ancestry. In order to reveal the phylogenetic relationship between Mamu-B alleles in more detail and independently of the binding motif determining residues, we omitted 51 codons encoding the residues that form the pockets in the peptide-

Residue	Position								
	1	2	3	4	5	6	7	8	9
A	0.170	0.111	0.230	0.137	0.262	0.152	0.371	0.146	0.011
C	0.171	0.023	0.072	0.128	0.195	0.104	0.158	0.064	0.010
D	0.102	0.114	0.346	0.145	0.201	0.160	0.414	0.099	0.008
E	0.081	0.172	0.253	0.028	0.126	0.044	0.091	0.015	0.024
F	0.113	0.073	1.000	0.412	0.667	1.000	1.000	1.000	0.538
G	0.092	1.000	0.587	0.172	0.144	0.084	0.239	0.148	0.017
H	0.125	0.080	0.333	0.191	0.256	0.127	0.570	0.056	0.006
I	0.158	0.155	0.479	0.144	0.543	0.176	0.227	0.052	0.075
K	0.274	0.039	0.049	0.157	0.053	0.067	0.039	0.065	0.004
L	0.144	0.038	0.209	1.000	0.315	0.066	0.321	0.235	0.032
M	1.000	0.093	0.495	0.359	0.200	0.136	0.268	0.160	0.014
N	0.079	0.125	0.289	0.217	0.031	0.100	0.190	0.076	0.012
P	0.043	0.056	0.316	0.203	0.222	0.144	0.262	0.132	0.002
Q	0.089	0.125	0.210	0.159	0.348	0.126	0.160	0.156	0.008
R	0.604	0.063	0.122	0.229	0.145	0.107	0.078	0.103	0.009
S	0.254	0.080	0.387	0.181	1.000	0.132	0.370	0.090	0.006
T	0.137	0.079	0.311	0.133	0.396	0.184	0.168	0.130	0.007
V	0.152	0.034	0.287	0.064	0.248	0.185	0.217	0.145	0.011
W	0.135	0.060	0.454	0.314	0.199	0.919	0.340	0.218	1.000
Y	0.146	0.137	0.449	0.652	0.328	0.568	0.376	0.305	0.321
Geomean	0.15	0.090	0.28	0.19	0.23	0.15	0.23	0.12	0.020
SD	2.0	2.2	2.0	2.1	2.2	2.2	2.1	2.3	5.2
SF	0.88	1.53	0.47	0.70	0.59	0.86	0.57	1.13	6.93

Fig. 4 PSCL-derived matrix describing 9-mer binding to Mamu-B*052:01. The PSCL was tested for binding, the data analyzed, and primary and secondary anchor positions defined, as described in the “Materials and methods”. Values shown represent the average relative binding (ARB) of the corresponding library relative to other pools with the same fixed position. Values have been normalized to the optimal residue at the corresponding position. *SD* indicates the standard deviation between the ARB of pools at the same position. *SF* is the

specificity factor, calculated as described in the “Materials and methods” representing the ratio of the average binding of the entire library to the average of pools at the indicated position. At the primary anchor positions ($SF > 2.4$; *blue shading*), the most preferred residues, associated with an ARB > 0.1 are highlighted by *bold yellow font*. *Green shading* highlights secondary anchor positions. The library average binding for Mamu-B*052:01 was 1,422 nM

binding groove. In the resulting phylogenetic tree, Mamu-B*039:01 and Mamu-B*052:01 cluster together (Online resource 2). Hence, within the same species, the shared binding specificity of these alleles is the result of a common ancestry, while between species, it is the result of convergent evolution.

Discussion

In this study, we have described the peptide-binding specificity of the macaque MHC class I molecules, Mamu-B*039:01, and Mamu-B*052:01. Both molecules were found to preferentially bind peptides with glycine at the second residue. Previously, Mamu-B*004 binding of a dominant T-cell epitope was found to be dependent upon G in position 2 (Dzuris et al. 2000). While more detailed analyses of the specificity of this molecule were not undertaken due to the unavailability of reagents, pocket analyses suggest that this allele may also be a member of the supertype delineated in the present study. Also, previous studies

revealed that the mouse MHC molecule H-2 D^d also has a preference for glycine at the second position (Corr et al. 1993; Li et al. 1998). Strikingly, our analysis demonstrated that this commonality hinges on convergent evolution independently generating the same pocket structure in distantly related MHC genes. Even more strikingly, this unique specificity is encoded in about 6% of macaque class I genes but is totally absent in humans.

An analysis of all the residues comprising the MHC B pocket of these macaque (Mamu) molecules, as well as the mouse molecule D^d, revealed that they share very similar B pocket structures. Together, the three molecules have identical residues in six of the ten positions hypothesized to form the B pocket and largely chemically conserved residues at the remaining positions. Most striking, however, is the shared presence of E in position 24 and R in position 66. These two residues, in these positions, are not found in any of the almost 1,000 HLA alleles sequenced thus far.

Crystal structures of peptides bound to H-2 D^d (Achour et al. 1998; Li et al. 1998) have suggested the presence of R in position 66 likely drives the requirement for glycine in

Residue	Position							
	1	2	3	4	5	6	7	8
A	0.040	0.007	0.246	0.045	0.175	0.587	0.955	0.032
C	1.000	0.003	0.067	0.032	0.015	0.602	0.018	0.020
D	0.004	0.003	0.553	0.037	0.010	0.147	0.011	0.001
E	0.004	0.003	0.280	0.023	0.020	0.713	0.007	0.001
F	0.011	0.002	0.013	0.039	0.084	0.131	0.012	1.000
G	0.005	1.000	0.113	0.021	0.005	0.403	1.000	0.001
H	0.013	0.009	0.609	0.026	0.017	0.419	0.037	0.000
I	0.009	0.000	0.100	0.005	0.002	0.147	0.002	0.358
K	0.052	0.000	0.033	0.022	0.000	0.041	0.080	0.005
L	0.008	0.000	0.121	0.009	0.010	0.307	0.001	0.114
M	0.001	0.001	1.000	0.195	0.471	0.201	0.001	0.069
N	0.009	0.001	0.978	0.023	0.008	0.524	0.876	0.001
P	0.001	0.001	0.780	0.011	0.000	0.775	0.011	0.000
Q	0.065	0.002	0.294	0.036	0.027	0.586	1.000	0.000
R	0.010	0.000	0.044	0.005	0.000	0.551	1.000	0.000
S	0.009	0.002	0.795	0.081	0.008	0.510	0.015	0.001
T	0.003	0.001	0.620	1.000	0.004	0.316	0.028	0.001
V	0.080	0.001	0.094	0.083	0.180	0.983	0.001	0.006
W	0.002	0.001	0.088	0.017	1.000	0.887	0.008	0.350
Y	0.007	0.001	0.098	0.009	0.324	1.000	0.457	0.084
Geomean	0.010	<0.001	0.19	0.030	0.010	0.39	0.030	0.010
SD	5.4	6.0	3.5	3.5	12.8	2.3	11.8	14.3
SF	2.4	13.8	0.13	0.83	1.7	0.060	0.74	3.6

Fig. 5 SAAS-derived matrix describing 8-mer binding to Mamu-B*052:01. A panel of single amino acid substitution analogs of the Mamu-B*052:01 binding 7.4 (sequence VGNVYVKF) was tested for binding, and the data analyzed, as described for the PSCL in the “Materials and methods”. Because of the larger SD in the panels of SAAS, compared to PSCL, more stringent criteria to define primary

and secondary anchor positions have been utilized, as also described in the “Materials and methods”. At the primary anchor positions (SF > 3.5; blue shading), the most preferred residues, associated with an ARB > 0.1 are highlighted by bold yellow font. Green shading highlights secondary anchor positions. The library average binding for Mamu-B*052:01 was 18.0 nM

position 2 of the corresponding peptide ligands. In both reported structures, the arginine at position 66 is shown to point down into the binding groove, where it forms salt bridges with E24 and E63. Contact between R66

and Y45 is also noted. Together, these contacts result in a narrowing of the B pocket, such that the presence of any residue except G in position 2 would likely result in steric clashes with R66.

Table 4 Alignment of B pocket residues of alleles with a preference for G2 reveals a unique structural motif

Alignment of B pocket residues												
System	Region	Allele	B pocket position/residue									
			7	9	24	34	45	63	66	67	70	99
H-2	–	Dd	Y	V	E	V	Y	E	R	A	N	A
Mamu	Chinese	B*039:01	Y	T	E	V	L	E	R	A	T	Y
Mamu	Indian	B*052:01	Y	S	E	V	M	E	R	A	N	Y
Consensus sequence			Y	VTS	E	V	YLM	E	R	A	NT	YA
Frequency of consensus residue(s)												
System	No. of alleles	Full pocket match	B pocket position ^a									
			7	9	24	34	45	63	66	67	70	99
Mamu	650	8	1.0	0.32	0.15	0.88	0.65	0.39	0.13	0.47	0.27	0.74
HLA	945	0	1.0	0.11	0	1.0	0.46	0.58	0	0	0.50	0.91

Bold font highlights positions where the consensus sequence is found in less than one third of the corresponding MHC alleles

Since B*039:01 and B*052:01 share the E24-E63-R66 structural motif with H-2 D^d, it is reasonable to assume that a similar narrowing of the B pocket is present and drives the preference for G in position 2 of their ligands. The observation that both Mamu-B*039:01 and -B*052:01 can bind ligands with residues other than G in position 2, albeit with lower avidity, does suggest that there may be correspondingly less hindrance in their B pockets than is the case for H-2 D^d. This observation may be due to the fact that these two Mamu molecules have, respectively, L and methionine (M) in position 45, located on the floor of the B pocket. These residues occupy less volume than the corresponding Y present in H-2 D^d and accordingly result in a B pocket with more space available to accommodate larger residues.

Unlike Mamu-B*039:01 and -B*052:01, H-2 D^d has a very strong preference for proline in position 3 (Corr et al. 1993). This specificity is hypothesized to be driven by the presence of a narrow hydrophobic ridge, formed by a pair of tryptophans in positions 97 and 114, that splits the D pocket and reduces its depth and volume (Achour et al. 1998; Li et al. 1998). Besides providing stabilizing interactions with the resulting small pocket, the proline may allow the peptide conformation necessary to ascend the wall formed by the tryptophan pair. Neither Mamu molecule studied here contains the tryptophan pair present in the murine D^d molecule, and therefore, the corresponding hydrophobic ridge is likely not present, thus liberating Mamu-B*039:01 and B*052:01 from a similar constraint in position 3.

Given the absence of these particular amino acid preferences in HLA sequences, we were interested in how non-human primate sequences and a mouse sequence have evolved. We analyzed macaque and mouse sequences from a phylogenetic perspective. These analyses revealed that within the same species, the evolution of these sequences is the result of common ancestry, while between species, it is the result of convergent evolution. It is therefore interesting to speculate why these sequences are absent in humans. It is possible that HLA sequences have not evolved to develop this particular motif because the evolutionary pressure leading to the origination of this motif was absent or that the expression of such specificity might be for some reason detrimental in humans. These binding specificities may be associated with disease susceptibility or autoimmunity. A specific event, such as a selective sweep (de Groot et al. 2008; de Groot et al. 2002) or bottleneck effect can also be considered as a possible explanation.

We cannot speculate on the biological significance of the peptide-binding motif elucidated herein for mice and macaques. Besides not being found in humans, the hallmark E24-R66 structural motif has also not been reported to date in most other nonhuman primates, including gorillas, chimpanzees, orangutans, olive baboons, pigtail macaques, and

crab-eating macaques, as well as rats, sheep, and canines (Robinson et al. 2010).

On the other hand, it is found in a few cases in African green monkeys and swine and in several instances in cattle. The disparate phylogenetic relationships between the species that do show the motif strengthen the idea that it is the result of convergent evolution. At this time, however, we cannot speculate as to the specific mechanism(s) driving generation and maintenance of this motif, although the MHC-I richness, as seen in rhesus macaques, African green monkeys, rodents, and swine, might play a role in maintaining some “specialized” peptide-binding specificity.

In summary, herein we describe a new supertype motif that encompasses MHC molecules that preferentially bind peptides containing the glycine residue at the second position. These findings suggest that novel evolutionary mechanisms result in this particular MHC peptide-binding motif originating independently in nonhuman primates and murine species, but being totally absent in human MHC molecules.

Acknowledgments This research is supported by the National Institutes of Health: R01 AI070902-01A2 (AS and BRM), R24 RR015371 (DIW), and R15 AI064175-01 (BRM).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Achour A, Persson K, Harris RA, Sundback J, Sentman CL, Lindqvist Y, Schneider G, Karre K (1998) The crystal structure of H-2Dd MHC class I complexed with the HIV-1-derived peptide P18-I10 at 2.4 Å resolution: implications for T cell and NK cell recognition. *Immunity* 9:199–208
- Allen TM, Mothe BR, Sidney J, Jing P, Dzuris JL, Liebl ME, Vogel TU, O'Connor DH, Wang X, Wussow MC, Thomson JA, Altman JD, Watkins DI, Sette A (2001) CD8(+) lymphocytes from simian immunodeficiency virus-infected rhesus macaques recognize 14 different epitopes bound by the major histocompatibility complex class I molecule mamu-A*01: implications for vaccine design and testing. *J Virol* 75:738–749
- Allen TM, Sidney J, del Guercio MF, Glickman RL, Lensmeyer GL, Wiebe DA, DeMars R, Pauza CD, Johnson RP, Sette A, Watkins DI (1998) Characterization of the peptide binding motif of a rhesus MHC class I molecule (Mamu-A*01) that binds an immunodominant CTL epitope from simian immunodeficiency virus. *J Immunol* 160:6062–6071
- Corr M, Boyd LF, Padlan EA, Margulies DH (1993) H-2Dd exploits a four residue peptide binding motif. *J Exp Med* 178:1877–1892
- de Groot NG, Heijmans CM, de Groot N, Otting N, de Vos-Rouweller AJ, Remarque EJ, Bonhomme M, Doxiadis GG, Crouau-Roy B, Bontrop RE (2008) Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. *Mol Ecol* 17:2074–2088
- de Groot NG, Otting N, Doxiadis GG, Balla-Jhaghihoorsingh SS, Heeny JL, van Rood JJ, Gagneux P, Bontrop RE (2002) Evidence for an

- ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proc Natl Acad Sci U S A* 99:11748–11753
- Dzuris JL, Sidney J, Appella E, Chesnut RW, Watkins DI, Sette A (2000) Conserved MHC class I peptide binding motif between humans and rhesus macaques. *J Immunol* 164:283–291
- Dzuris JL, Sidney J, Horton H, Correa R, Carter D, Chesnut RW, Watkins DI, Sette A (2001) Molecular determinants of peptide binding to two common rhesus macaque major histocompatibility complex class II molecules. *J Virol* 75:10958–10968
- Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351:290–296
- Gardner MB, Luciw PA (2008) Macaque models of human infectious disease. *ILAR J* 49:220–255
- Giraldo-Vela JP, Rudersdorf R, Chung C, Qi Y, Wallace LT, Bimber B, Borchardt GJ, Fisk DL, Glidden CE, Loffredo JT, Piaskowski SM, Furlott JR, Morales-Martinez JP, Wilson NA, Rehrauer WM, Lifson JD, Carrington M, Watkins DI (2008) The major histocompatibility complex class II alleles Mamu-DRB1*1003 and -DRB1*0306 are enriched in a cohort of simian immunodeficiency virus-infected rhesus macaque elite controllers. *J Virol* 82:859–870
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
- Gulukota K, Sidney J, Sette A, DeLisi C (1997) Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J Mol Biol* 267:1258–1267
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160–174
- Hawkins OE, Vangundy RS, Eckerd AM, Bardet W, Buchli R, Weidanz JA, Hildebrand WH (2008) Identification of breast cancer peptide epitopes presented by HLA-A*0201. *J Proteome Res* 7:1445–1457
- Hickman HD, Batson CL, Prilliman KR, Crawford DL, Jackson KL, Hildebrand WH (2000) C-terminal epitope tagging facilitates comparative ligand mapping from MHC class I positive cells. *Hum Immunol* 61:1339–1346
- Hickman-Miller HD, Bardet W, Gilb A, Luis AD, Jackson KW, Watkins DI, Hildebrand WH (2005) Rhesus macaque MHC class I molecules present HLA-B-like peptides. *J Immunol* 175:367–375
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267
- Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R (2007) Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460
- Karl JA, Wiseman RW, Campbell KJ, Blasky AJ, Hughes AL, Ferguson B, Read DS, O'Connor DH (2008) Identification of MHC class I sequences in Chinese-origin rhesus macaques. *Immunogenetics* 60:37–46
- Kestler H, Kodama T, Ringler D, Marthas M, Pedersen N, Lackner A, Regier D, Sehgal P, Daniel M, King N et al (1990) Induction of AIDS in rhesus monkeys by molecularly cloned simian immunodeficiency virus. *Science* 248:1109–1112
- Kubo RT, Sette A, Grey HM, Appella E, Sakaguchi K, Zhu NZ, Arnott D, Sherman N, Shabanowitz J, Michel H et al (1994) Definition of specific peptide motifs for four major HLA-A alleles. *J Immunol* 152:3913–3924
- Li H, Natarajan K, Malchiodi EL, Margulies DH, Mariuzza RA (1998) Three-dimensional structure of H-2Dd complexed with an immunodominant peptide from human immunodeficiency virus envelope glycoprotein 120. *J Mol Biol* 283:179–191
- Loffredo JT, Bean AT, Beal DR, Leon EJ, May GE, Piaskowski SM, Furlott JR, Reed J, Musani SK, Rakasz EG, Friedrich TC, Wilson NA, Allison DB, Watkins DI (2008) Patterns of CD8+ immunodominance may influence the ability of Mamu-B*08-positive macaques to naturally control simian immunodeficiency virus SIVmac239 replication. *J Virol* 82:1723–1738
- Loffredo JT, Maxwell J, Qi Y, Glidden CE, Borchardt GJ, Soma T, Bean AT, Beal DR, Wilson NA, Rehrauer WM, Lifson JD, Carrington M, Watkins DI (2007) Mamu-B*08-positive macaques control simian immunodeficiency virus replication. *J Virol* 81:8827–8832
- Loffredo JT, Sidney J, Bean AT, Beal DR, Bardet W, Wahl A, Hawkins OE, Piaskowski S, Wilson NA, Hildebrand WH, Watkins DI, Sette A (2009) Two MHC class I molecules associated with elite control of immunodeficiency virus replication, Mamu-B*08 and HLA-B*2705, bind peptides with sequence similarity. *J Immunol* 182:7763–7775
- Loffredo JT, Sidney J, Piaskowski S, Szymanski A, Furlott J, Rudersdorf R, Reed J, Peters B, Hickman-Miller HD, Bardet W, Rehrauer WM, O'Connor DH, Wilson NA, Hildebrand WH, Sette A, Watkins DI (2005) The high frequency Indian rhesus macaque MHC class I molecule, Mamu-B*01, does not appear to be involved in CD8+ T lymphocyte responses to SIVmac239. *J Immunol* 175:5986–5997
- Loffredo JT, Sidney J, Wojewoda C, Dodds E, Reynolds MR, Napoe G, Mothe BR, O'Connor DH, Wilson NA, Watkins DI, Sette A (2004) Identification of seventeen new simian immunodeficiency virus-derived CD8+ T cell epitopes restricted by the high frequency molecule, Mamu-A*02, and potential escape from CTL recognition. *J Immunol* 173:5064–5076
- Madden DR (1995) The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* 13:587–622
- Miller CJ, Alexander NJ, Sutjipto S, Lackner AA, Gettie A, Hendrickx AG, Lowenstine LJ, Jennings M, Marx PA (1989) Genital mucosal transmission of simian immunodeficiency virus: animal model for heterosexual transmission of human immunodeficiency virus. *J Virol* 63:4277–4284
- Mothe BR, Sidney J, Dzuris JL, Liebl ME, Fuenger S, Watkins DI, Sette A (2002) Characterization of the peptide-binding specificity of Mamu-B*17 and identification of Mamu-B*17-restricted epitopes derived from simian immunodeficiency virus proteins. *J Immunol* 169:210–219
- Parham P (2005) MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol* 5:201–214
- Pinilla C, Martin R, Gran B, Appel JR, Boggiano C, Wilson DB, Houghten RA (1999) Exploring immunological specificity using synthetic peptide combinatorial libraries. *Curr Opin Immunol* 11:193–202
- Posada D (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253–1256
- Reed JS, Sidney J, Piaskowski SM, Glidden CE, Leon EJ, Burwitz BJ, Kolar HL, Eernisse CM, Furlott JR, Maness NJ, Walsh AD, Rudersdorf RA, Bardet W, McMurtrey CP, O'Connor DH, Hildebrand WH, Sette A, Watkins DI, Wilson NA (2011) The role of MHC class I allele Mamu-A*07 during SIV(mac)239 infection. *Immunogenetics* 63:789–807
- Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SG (2010) IPD—the Immuno Polymorphism Database. *Nucleic Acids Res* 38:D863–D869
- Saper MA, Bjorkman PJ, Wiley DC (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J Mol Biol* 219:277–319
- Schneidewind A, Brockman MA, Sidney J, Wang YE, Chen H, Suscovich TJ, Li B, Adam RI, Allgaier RL, Mothe BR, Kuntzen T, Oniangue-Ndza C, Trocha A, Yu XG, Brander C, Sette A, Walker BD, Allen TM (2008) Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. *J Virol* 82:5594–5605
- Sette A, Sidney J (1998) HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Curr Opin Immunol* 10:478–482

- Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50:201–212
- Sette A, Sidney J, Bui HH, Del Guercio MF, Alexander J, Loffredo J, Watkins DI, Mothe BR (2005) Characterization of the peptide-binding specificity of Mamu-A*11 results in the identification of SIV-derived epitopes and interspecies cross-reactivity. *Immunogenetics* 57(1–2):53–68
- Sette A, Sidney J, del Guercio MF, Southwood S, Ruppert J, Dahlberg C, Grey HM, Kubo RT (1994) Peptide binding to the most frequent HLA-A class I alleles measured by quantitative molecular binding assays. *Mol Immunol* 31:813–822
- Sette A, Sidney J, Livingston B, Dzuris J, Crimi C, Walker CM, Southwood S, Collins EJ, Hughes A (2003) Class I molecules with similar peptide binding specificities are the result of both common ancestry and convergent evolution. *Immunogenetics* 54:830–841
- Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B (2008) Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res* 4:2
- Sidney J, Southwood S, Mann DL, Fernandez-Vina MA, Newman MJ, Sette A (2001) Majority of peptides binding HLA-A*0201 with high affinity crossreact with other A2-supertype molecules. *Hum Immunol* 62:1200–1216
- Sidney J, Southwood S, Sette A (2005) Classification of A1- and A24-supertype molecules by analysis of their MHC-peptide binding repertoires. *Immunogenetics* 57:393–408
- Smith SM, Holland B, Russo C, Dailey PJ, Marx PA, Connor RI (1999) Retrospective analysis of viral load and SIV antibody responses in rhesus macaques infected with pathogenic SIV: predictive value for disease progression. *AIDS Res Hum Retroviruses* 15:1691–1701
- Solomon C, Southwood S, Hoof I, Rudersdorf R, Peters B, Sidney J, Pinilla C, Marcondes MC, Ling B, Marx P, Sette A, Mothe BR (2010) The most common Chinese rhesus macaque MHC class I molecule shares peptide binding repertoire with the HLA-B7 supertype. *Immunogenetics* 62:451–464
- Southwood S, Solomon C, Hoof I, Rudersdorf R, Sidney J, Peters B, Wahl A, Hawkins O, Hildebrand W, Mothe BR, Sette A (2011) Functional analysis of frequently expressed Chinese rhesus macaque MHC class I molecules Mamu-A1*02601 and Mamu-B*08301 reveals HLA-A2 and HLA-A3 supertypic specificities. *Immunogenetics* 63:275–290
- Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres E Jr, Wright C, Harkins T, O'Connor DH (2009) Major histocompatibility complex genotyping with massively parallel pyrosequencing. *Nat Med* 15:1322–1326