Pyrosequencing Reveals Restricted Patterns of CD8⁺ T Cell Escape-Associated Compensatory Mutations in Simian Immunodeficiency Virus^{▽‡}

Benjamin J. Burwitz,^{1†} Jonah B. Sacha,^{2,3†} Jason S. Reed,^{2,3} Laura P. Newman,¹
Francesca A. Norante,¹ Benjamin N. Bimber,⁴ Nancy A. Wilson,¹
David I. Watkins,¹ and David H. O'Connor^{1,4*}

Department of Pathology, University of Wisconsin-Madison, Madison, Wisconsin 53706¹; Vaccine & Gene Therapy Institute² and Oregon National Primate Research Center,³ Oregon Health & Science University, Portland, Oregon 97006; and Wisconsin National Primate Research Center, Madison, Wisconsin 53706⁴

Received 12 July 2011/Accepted 10 October 2011

CD8⁺ T cells play a major role in the containment of human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) replication. CD8⁺ T cell-driven variations in conserved regions under functional constraints result in diminished viral replicative capacity. While compensatory mutations outside an epitope can restore replicative capacity, the kinetics with which they arise remains unknown. Additionally, certain patterns of linked mutations associated with CD8⁺ T cell epitope escape in these highly conserved regions may lead to variable levels of viral fitness. Here, we used pyrosequencing to investigate the kinetics and patterns of mutations surrounding the Mamu-A1*00101-bound Gag₁₈₁₋₁₈₉CM9 CD8⁺ T cell epitope. We obtained more than 400 reads for each sequencing time point, allowing us to confidently detect the emergence of viral variants bearing escape mutations with frequencies as low as 1% of the circulating virus. With this level of detail, we demonstrate that compensatory mutations generally arise concomitantly with Gag₁₈₁₋₁₈₉CM9 escape mutations. We observed distinct patterns of linked flanking mutations, most of which were found downstream of Gag₁₈₁₋₁₈₉CM9. Our data indicate that, whereas Gag₁₈₁₋₁₈₉CM9 escape is much more complex than previously appreciated, it occurs in a coordinated fashion, with very specific patterns of flanking mutations required for immune evasion. This is the first detailed report of the ontogeny of compensatory mutations that allow CD8⁺ T cell epitope escape in infected individuals.

Human immunodeficiency virus (HIV) strains, like those of other retroviruses, are genetically diverse within and between hosts. This diversity increases the fitness of HIV through evasion of host immune responses. Indeed, the evolution of HIV over time shows evidence of adaptation in response to CD8⁺ T cell responses restricted by HLA-A and HLA-B alleles (6, 9, 13, 14). Determination of an optimal set of HIV-derived CD8⁺ T cell antigens and deployment of these antigens in an appropriate vector are major goals of HIV vaccine research. Thus, understanding how host cellular immune responses drive viral sequence evolution is crucial for rational vaccine design.

Certain major histocompatibility complex (MHC) class I-bound epitopes found in individuals infected with HIV or simian immunodeficiency virus (SIV) show a limited capacity for variation, with escape mutations often only occurring late (if at all) in chronic infection (8, 15, 20). These epitopes are found in regions of the virus that are conserved due to structural or functional constraints (8, 15, 20). However, compensatory mutations found near a subset of these epitopes enable intraepitope variation and maintenance of viral replication (4,

7, 10, 21, 22, 26). The assortment of compensatory mutations found in these regions may affect viral variation within the epitope as well as viral fitness. A comprehensive analysis of the kinetics and patterns of epitope and compensatory mutations has yet to be undertaken.

Here, we utilized pyrosequencing to examine the ontogeny of compensatory mutations flanking the well-described, conserved Mamu-A1*00101-bound epitope Gag₁₈₁₋₁₈₉CM9. Previous studies identified an association between Gag₁₈₁₋₁₈₉CM9 escape and two compensatory mutations: Gag I161V and Gag I206V (7, 20). We expand on those studies by describing a diverse set of putative compensatory mutations not previously characterized and show that these novel mutations arise in an array of patterns and are associated with different amino acid substitutions at Gag T182, the second position of the Gag₁₈₁₋₁₈₉CM9 epitope. Furthermore, we observed strong clonal associations between Gag T182 mutants and flanking compensatory mutations, even when these mutations were rare within the viral quasispecies. Cumulatively, our data suggest that viral escape from Gag₁₈₁₋₁₈₉CM9-specific CD8⁺ T cells occurs only via discrete pathways involving coordinated amino acid substitutions both within and outside the viral epitope.

MATERIALS AND METHODS

Animals and vaccinations. The Indian rhesus macaques used in this study were all previously described in other publications (Table 1). Some of the animals were vaccinated before infection with SIVmac239 as previously described. Animals were infected with the cloned virus SIVmac239. The Wisconsin National

* Corresponding author. Mailing address: Department of Pathology and Laboratory Medicine, University of Wisconsin—Madison, 555 Science Dr., Madison, WI 53711. Phone: (608) 265-3389. Fax: (608) 265-8084. E-mail: doconnor@primate.wisc.edu.

† B.J.B. and J.B.S. contributed equally to the work.

‡ Supplemental material for this article may be found at <http://jvi.asm.org/>.

[▽] Published ahead of print on 12 October 2011.

TABLE 1. Indian rhesus macaques used in this study

Animal	Gag CM9 escape mutation with highest frequency	Vaccination ^a	Reference
r00044	T182A	DNA/Ad5 (Gag/Tat/Rev/Nef)	25
r00060	T182A	DNA/Ad5 (Gag/Tat/Rev/Nef)	25
r95114	T182A		1
r97113	T182I	DNA/Ad5 (Gag/Tat/Rev/Nef)	25
r95058	T182I	DNA/MVA (Gag)	2
rh1937	T182I		17
r00014	T182S/C	DNA/Ad5 (Gag/Tat/Rev/Nef)	25
rh2127 (r31157)	T182S/C	DNA/rMVA (Tat/Rev/Nef/Gag CM9/Tat SL8)	24
r00045	T182S/L		25
r02120			23

^a Ad5, adenovirus type 5; MVA, modified vaccinia Ankara virus; rMVA, recombinant MVA.

Primate Research Center (University of Wisconsin, Madison, WI) maintained the SIV-infected macaques in accordance with experimental protocols approved by the University of Wisconsin Research Animal Resources Center review committee.

SIV sequencing. Cell-free plasma was obtained from EDTA-treated anticoagulated whole blood by using Ficoll-Paque Plus (GE Healthcare Bioscience) and density centrifugation. Plasma was frozen at -80°C until the time of viral RNA extraction. Viral RNA in the plasma was isolated using a QIAamp MinElute virus spin kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Viral RNA was reverse transcribed (RT) and amplified using a SuperScript III One-Step RT-PCR system with Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA) and MID-tagged primers (454; Life Sciences, Branford, CT) spanning Gag amino acid residues 155 to 253 (see Table S1 in the supplemental material). The RT-PCR conditions were as follows: 50°C for 15 min; 94°C for 2 min; 40 cycles of 94°C for 15 s, 58°C for 30 s, and 68°C for 50 s; and 68°C for 5 min. Following RT-PCR, DNA in each sample was purified using two rounds of Ampure XP magnetic bead separation (Agencourt, Bendigo, Victoria, Australia) (bead/PCR product ratio, 0.6:1). Samples were quantified using Quant-iT high-sensitivity (HS) reagents (Invitrogen, Carlsbad, CA). Amplicons were pooled at equimolar ratios, and the final library was diluted to 1.5×10^6 copies/ μL . Emulsion PCR (emPCR) was performed using 1.5 copies/bead and a GS Junior Titanium emPCR kit (Lib-A) per the instructions of the manufacturer (Roche, Indianapolis, IN). Enriched DNA beads (5×10^5) were sequenced using a Roche/454 GS Junior system.

SIV sequence analysis. Pyrosequencing flowgram data were processed using Roche Amplicon Variant Analyzer (AVA) software. Within Roche AVA, flowgram reads were aligned to the reference SIVmac239 sequence (GenBank accession no. M33262), and then identical read results were assembled into groups. Manual editing was performed using CodonCode alignment software (CodonCode Corp., Dedham, MA) only when homopolymeric regions differed by ± 1 nucleotide and resulted in frameshifts. Improved base-calling using Roche AVA software compensates for base addition artifacts in homopolymeric regions, and manual editing was kept to a minimum.

Pyrosequencing read groups generated by Roche AVA were analyzed using customized BLAST-based software (written by B. N. Kimber) to identify and translate positions of difference between the experimental data and the reference SIVmac239 genome (3). The source code for this pipeline can be obtained from a subversion repository (<https://hedgehog.flhrc.org/tor/stedi/trunk/server/customModules/SequenceAnalysis>). Instructions for accessing this repository, including a guest password, can be found at the following URL: <https://www.labkey.org/wiki/home/Documentation/page.view?name=syn>. The pipeline itself has been integrated into the LabKey software platform as the SequenceAnalysis module, which provides a graphical, Web-based platform to initiate analysis pipelines and view results. LabKey is a free, open-source software package available at www.labkey.org. For each individual time point, only mutations present in more than 1% of total reads were considered to represent data above the background.

Generation of SIV Gag 3D models. Five separate models of the N-terminal 283-residue fragment of the immature SIVmac239 Gag polyprotein were generated using the Geno3D automatic molecular modeling tool (<http://pbil.ibcp.fr/htm/index.php>) and model PDB 1L6N (homologous HIV sequence) as a template. Three-dimensional (3D) structures were visualized using PyMol freeware

(www.pymol.org). Biochemical bonds and distances between amino acids were separately investigated using PyMol for all five models.

RESULTS

*Mamu-A1*00101*-positive rhesus macaques infected with SIVmac239 consistently mount an immunodominant CD8⁺ T-cell response to the Gag₁₈₁₋₁₈₉CM9 epitope. However, immune escape within this epitope does not typically occur until late in chronic infection. Gag₁₈₁₋₁₈₉CM9 lies within a highly conserved portion of the Gag protein. Nevertheless, escape mutations eventually arise at position Gag T182 (position 2 of the Gag₁₈₁₋₁₈₉CM9 epitope) in many *Mamu-A1*00101*-positive animals, and escape has been associated with two flanking compensatory mutations: Gag I161V and Gag I206V (7, 20). Escape within the Gag₁₈₁₋₁₈₉CM9 epitope has also been observed to occur in the absence of either Gag I161V or Gag I206V, however, suggesting that other compensatory mutations may exist (7, 20).

In order to comprehensively characterize the kinetics and patterns by which compensatory mutations arise, we selected nine SIVmac239-infected *Mamu-A1*00101*-positive rhesus macaques and examined viral evolution occurring within and flanking the Gag₁₈₁₋₁₈₉CM9 epitope by Roche/454 pyrosequencing. These animals were selected based on the availability of longitudinal samples and the results of previous Sanger-based sequencing showing escape at position T182 of the Gag protein. Specifically, we stratified our animals into three groups based on T182 escape: (i) T182A; (ii) T182I; and (iii) T182S/C/L (Table 1). Six of the *Mamu-A1*00101*-positive animals had been vaccinated prior to SIVmac239 infection, although this did not appear to affect the timing or route of Gag₁₈₁₋₁₈₉CM9 escape. Finally, to verify the sensitivity of our approach, we analyzed the Gag sequence from one *Mamu-A1*00101*-negative rhesus macaque (r02120; Fig. 1A). We were unable to detect viral variations at above 0.6% of total reads at any position within Gag₁₈₁₋₁₈₉CM9 in the sequences obtained from this animal (Fig. 1A). We therefore set the limit of detection of viral variation in this study at 1% of total sequencing reads.

We longitudinally examined three rhesus macaques (r95114, r95058, and r00014) over the course of acute and chronic SIV infections. Time-of-death Sanger sequencing showed that each of these animals harbored one of the three major types of Gag T182 escape: (i) T182A; (ii) T182I; or (iii) T182S/C (Table 1). Strikingly, we observed an emergence of the T182A escape mutation that was coincident with an S242T mutation in r95114 in the absence of the previously described I161V or I206V mutations (Fig. 1B). We detected the I161V and I206V compensatory mutations, but they emerged only after the establishment of virus harboring the T182A and S242T mutations (Fig. 1B). Interestingly, we found a downstream S242T mutation at late time points in chronic infection in two additional rhesus macaques with the T182A escape mutation (r00044 and r00060; see Fig. S1 in the supplemental material and https://xnight.primate.wisc.edu:8443/labkey/project/WNPRC/WNPRC_Laboratories/oconnor/public/publications/begin.view?). This S242T mutation had been observed previously in SIVmac239-infected, *Mamu-A1*00101*-positive rhesus macaques (7), but its potential role in Gag₁₈₁₋₁₈₉CM9 escape is not clear. We detected sim-

FIG. 1. Consensus sequences of the Gag₁₈₁₋₁₈₉ CM9 region in rhesus macaques. The Gag₁₈₁₋₁₈₉ CM9 region from *Mamu-A1*00101*-positive rhesus macaques was pyrosequenced. (A) r02120; (B) r95114; (C) r95058; (D) rh2127. Each consensus sequence contains each mutation(s) present in 1% or more of total reads. Changes are indicated by the use of the single-letter amino acid code for substitutions where only one residue with a change was found at a given position with a frequency > 1%. All other changes are indicated with the letter X; the frequencies of mixed amino acids are listed below the corresponding positions. Nonsynonymous mutations are colored according to prevalence (indicated by the color key on the right). Positions where nonsynonymous mutations were observed are listed across the top. Yellow highlighting denotes the nine amino acids comprising the CD8⁺ T cell epitope Gag₁₈₁₋₁₈₉ CM9.

ilar frequencies of the T182A escape mutation and the S242T mutation at all time points of sequencing for r95114, a situation analogous to the similar frequencies of T182A and I206V observed in r00044 and r00060 (Fig. 1B; see also Fig. S1 in the supplemental material).

The association between escape at position 2 of the Gag₁₈₁₋₁₈₉CM9 epitope (T182A) and the S242T mutation indicated that a more complex system of compensatory mutations may exist in SIVmac239-infected, *Mamu-A1*00101*-positive animals than had been previously appreciated. Furthermore, Gag₁₈₁₋₁₈₉CM9 position 2 (T182) escape mutations are predominantly alanine, isoleucine, or cysteine substitutions, as shown by historical Sanger sequencing (7, 16). We therefore hypothesized that differing amino acid substitutions present at Gag₁₈₁₋₁₈₉CM9 position 2 (T182) would be associated with different patterns of flanking mutations.

We observed a pattern of mutations associated with the T182I escape mutation in r95058 that was different from the patterns seen with viruses harboring the T182A escape mutation in r95114 (Fig. 1C). The coincident I206V and S242T mutations that occurred with the emergence of the T182A escape mutation were found in fewer than 18% of the total sequence reads during the emergence of the T182I escape mutation in r95058. Instead, we found a distinct pattern of flanking mutations, with V243A and D244E mutations found at frequencies similar to the T182I escape frequency (Fig. 1C). However, upon further inspection of two additional rhesus macaques with the T182I escape mutation, we again discovered the downstream S242T mutation (see Fig. S2 in the supplemental material and https://xnlight.primat.wisc.edu:8443/labkey/project/WNPRC/WNPRC_Laboratories/oconnor/public/publications/begin.view?). These data indicate the existence of multiple pathways that allow Gag₁₈₁₋₁₈₉CM9 escape.

Finally, we assessed the flanking mutations associated with the T182S/C escape mutation found in animal rh2127. We again found a pattern of flanking mutations that was unique compared to those assessed in animals r95114 and r95058. We found similar frequencies of the T182S/C escape mutation and the S242T mutation, as seen with the pattern observed in r95114 (Fig. 1B and D). However, we discovered an absence of the I206V and V243A mutations, common in animals with T182A or T182I escape mutations, respectively, in the sequences at all time points (Fig. 1D). To confirm our results, we sequenced virus from an additional animal with the T182S/C escape mutation (see Fig. S3 in the supplemental material and https://xnlight.primat.wisc.edu:8443/labkey/project/WNPRC/WNPRC_Laboratories/oconnor/public/publications/begin.view?). We found a similar pattern of Gag₁₈₁₋₁₈₉CM9 flanking mutations, with low frequencies of I206V and an absence of V243A.

We observed a particularly complex pattern of flanking mutations accompanying Gag₁₈₁₋₁₈₉CM9 escape in animal r00045 (Fig. 2). We found multiple Gag₁₈₁₋₁₈₉CM9 escape variants that began with a predominating T182S escape mutation and shifted over time to a T182L escape mutation (Fig. 2). Intriguingly, this shift within the viral quasispecies corresponded to changes in the frequencies of mutations P221Q and A222T (Fig. 2).

Our alignment of consensus sequences in investigations of the Gag₁₈₁₋₁₈₉CM9 region indicated that novel compensatory mutations may exist and that particular patterns of mutations

in the capsid protein may be required to allow viral escape. However, to interpret the importance of these mutations for Gag₁₈₁₋₁₈₉CM9 escape, we needed to better understand their clonal associations within the viral quasispecies. To this end, we measured the frequency of each of several unique amino acid sequences within our virus population (Fig. 3). We found that greater than 80% of viral genomes in r95114 with a T182A escape mutation also harbored the S242T mutation (Fig. 3A). We observed similar viral variants in rh2127, where the T182S and T182C escape mutations were also clonally associated with the S242T mutation (Fig. 3C). These data support our conclusion that the S242T mutation allows Gag₁₈₁₋₁₈₉CM9 escape.

The emergence of the T182I escape variant in r95058 occurred at between 47 and 67 weeks postinfection (wpi), and we did not have viral RNA samples from this time interval to examine (Fig. 3B). Therefore, no flanking mutations associated with the emergence of T182I escape could be identified. However, we did observe mutations on viral genomes that were rarely linked to the T182I escape mutation (Fig. 3B). The I206V mutation was rarely linked to either the S242T mutation or the M250T mutation, again suggesting that specific patterns of mutations downstream of the Gag₁₈₁₋₁₈₉CM9 escape sequence are required for viral replication.

Given the complexity of the T182S/C/L escape mutations in r00045, we were particularly eager to assess clonal associations between the Gag₁₈₁₋₁₈₉CM9 escape sequence and flanking mutations. The T182A escape mutation was clonally associated with the I206V and I161V mutations at 60 wpi, supporting our previous finding observed in animals with the T182A escape mutation (Fig. 1B and 4; see also Fig. S1 in the supplemental material). Additionally, the T182S escape mutation was clonally associated with the unique A222T mutation at this same time point (Fig. 4). We were surprised to find that none of the viruses with the T182S escape mutation contained the I206V mutation (Fig. 2). Interestingly, we discovered a clonal association between the emerging T182L escape mutation and the P221Q mutation that began at 119 wpi (Fig. 4). These two mutations continued to rise in frequency over time within the viral quasispecies and were linked at all subsequent time points.

We complemented our analysis of Gag₁₈₁₋₁₈₉CM9 escape by generating three-dimensional models of viral variants that we detected in the viral quasispecies of different animals. The region of the capsid protein encompassing the Gag₁₈₁₋₁₈₉CM9 epitope has been shown to play a role in multiple events, including capsid multimerization, capsid maturation, and uncoating of the capsid following cell entry, within the SIV life cycle (11, 12, 19). Therefore, compensatory mutations allowing viral escape presumably restore the structure and function of the capsid protein. We were surprised to find that S242, which mutated in many animals showing Gag₁₈₁₋₁₈₉CM9 escape, was spatially closer to T182 than the previously described I206V compensatory mutation (Fig. 5). However, there were no direct bonds found between T182 and the common S242, V243, and D244 amino acids associated with escape (Fig. 5). This is not an entirely surprising result, given that the previously described I206V compensatory mutation, which has been shown to restore viral fitness *in vitro*, also shares no direct bonds with T182. Additionally, we were unable to detect biochemical as-

FIG. 2. Consensus sequence of the Gag₁₈₁₋₁₈₉CM9 region in rhesus macaque r00045. The Gag₁₈₁₋₁₈₉CM9 region from Mamu-A1*001-positive rhesus macaque r00045 was pyrosequenced. Each consensus sequence contains each mutation(s) present in 1% or more of total reads. Changes are indicated by the use of the single-letter amino acid code for substitutions where only one residue with a change was found at a given position with a frequency > 1%. All other changes are indicated with the letter X; the frequencies of mixed amino acids are listed below the corresponding positions. Nonsynonymous mutations are colored according to prevalence (indicated by the color key on the right). Positions where nonsynonymous mutations were observed are listed across the top. Yellow highlighting denotes the nine amino acids comprising the CD8⁺ T cell epitope Gag₁₈₁₋₁₈₉CM9.

sociations between T182 and any of the flanking mutations we observed in this study (data not shown).

DISCUSSION

Gag₁₈₁₋₁₈₉CM9 lies within the capsid protein of SIVmac239 (20). Specifically, the epitope is found in a stretch of 20 amino acids that is highly conserved in all retroviruses, indicating the structural importance of this region. Therefore, the common T182 escape mutation within this epitope is hindered by structural and functional constraints (21). Two mutations flanking Gag₁₈₁₋₁₈₉CM9, I161V and I206V, allow epitope escape by restoring viral fitness (7). We show here that Gag₁₈₁₋₁₈₉CM9 escape is associated with much more complex patterns of flanking mutations (Fig. 6). Specifically, we discovered distinct patterns of mutations associated with different types of

Gag₁₈₁₋₁₈₉CM9 escape. The T182A escape mutation was accompanied by high-frequency I161V and I206V mutations (Fig. 6A). In contrast, viruses with the T182I or T182S/C/L escape mutation lacked high-frequency I161V and I206V mutations (Fig. 6B and C). Viruses with the T182I escape mutation were associated with both V243A and D244E, whereas viruses with T182S/C/L were associated with D244E only (Fig. 6B and C). Interestingly, in both cases where the predominant virus population exhibited Gag₁₈₁₋₁₈₉CM9 escape but neither I206V nor S242T, we discovered a pair of mutations at positions D244 and M250. Therefore, it is possible that the combination of these two mutations may substitute for a single compensatory mutation, such as I206V. These flanking mutations often arise contemporaneously on the same viral genomes as Gag₁₈₁₋₁₈₉CM9 escape variants and represent putative novel compensatory mutations. Indeed, the

FIG. 3. Linkage of Gag₁₈₁₋₁₈₉CM9 escape and flanking mutations in rhesus macaques. (A) r95114; (B) r95058; (C) rh2127. Each table shows the amino acid substitution within Gag₁₈₁₋₁₈₉CM9 (left, bold). Positions where nonsynonymous mutations were observed are listed (top left). Time points postinfection are listed (top right). The frequency of each pattern of linked mutations within the viral quasiespecies is listed (bottom right).

importance of the I206V compensatory mutation to SIVmac239 viral fitness has been previously demonstrated (7), and we found clonal associations between the T182A escape mutation and the I206V compensatory mutation in our study. Similar clonal associations between the T182 escape mutation and the downstream S242T mutation strongly suggest that this is a novel compensatory mutation. Animals with the T182 escape mutation and without either the I206V or S242T mutation (r95058 and r00045) exhibited differing patterns of flanking mutations, indicating that the system of compensatory mutations associated with Gag₁₈₁₋₁₈₉CM9 escape is highly complex. Additionally, we observed a single instance where the T182A escape mutation was not accompanied by flanking mu-

tations (r97113 at 98 wpi). Therefore, it is likely that additional Gag₁₈₁₋₁₈₉CM9-associated compensatory mutations exist outside the region of Gag we examined.

We also examined the kinetics of compensatory mutations in relation to Gag₁₈₁₋₁₈₉CM9 escape in multiple *Mamu-A1*00101*-positive rhesus macaques. It has been hypothesized that delays in the emergence of Gag₁₈₁₋₁₈₉CM9 variants are due to a requirement for a secondary compensatory mutation, making the combination of events highly unlikely to occur simultaneously. Indeed, the probability of two amino acid substitutions occurring simultaneously within the SIV genome, assuming that nucleotide mutations at each position are equally likely and a conservative reverse transcriptase error

r00045

								Week 60	Week 65	Week 104	Week 112	Week 119	Week 123	Week 128	Week 132	Week 157
								228 Reads	1,355 Reads	6,756 Reads	1,752 Reads	1,735 Reads	1,830 Reads	1,597 Reads	1,698 Reads	44 Reads
I161	P172	CM9	I206	P221	A222	D244	M250									
.	.	CSPYDINQ	.	.	.	E	.	3.51%	0.96%	0.03%	0.11%	0.00%	0.11%	0.00%	0.00%	0.00%
.	.	CSPYDINQ	.	.	T	E	.	95.18%	98.82%	21.05%	11.07%	6.40%	3.50%	0.56%	0.12%	0.00%
.	S	CSPYDINQ	.	.	.	E	.	0.00%	0.00%	0.21%	0.17%	0.35%	0.16%	0.44%	0.12%	2.27%
.	S	CSPYDINQ	.	.	T	E	.	0.00%	0.00%	77.52%	79.85%	63.46%	56.01%	37.07%	17.31%	0.00%
.	S	CSPYDINQ	.	Q	.	E	.	0.00%	0.00%	0.00%	6.28%	28.70%	37.71%	60.36%	80.91%	93.18%
.	S	CSPYDINQ	.	Q	.	E	A	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.27%
Other (with T182S mutation)								1.31%	0.22%	1.09%	2.52%	1.09%	2.51%	1.57%	1.54%	2.28%

								Week 119	Week 123	Week 128	Week 132	Week 157	Week 198
								39 Reads	115 Reads	215 Reads	349 Reads	1963 Reads	5424 Reads
I161	P172	CM9	I206	P221	A222	D244	M250						
.	.	CLPYDINQ	.	Q	.	E	.	0.00%	1.74%	0.47%	0.29%	0.36%	0.00%
.	S	CLPYDINQ	.	.	T	E	.	15.38%	8.70%	6.05%	2.29%	0.00%	0.00%
.	S	CLPYDINQ	.	Q	.	E	.	84.62%	88.70%	93.02%	95.99%	91.75%	2.29%
.	S	CLPYDINQ	.	Q	.	E	A	0.00%	0.00%	0.00%	0.00%	6.57%	95.06%
Other (with T182L mutation)								-	0.86%	1.46%	1.43%	1.32%	2.65%

								Week 60	Week 157
								38 Reads	26 Reads
I161	P172	CM9	I206	P221	A222	D244	M250		
.	.	CAPYDINQ	.	.	.	E	.	5.26%	0.00%
.	.	CAPYDINQ	.	.	T	E	.	44.74%	0.00%
.	.	CAPYDINQ	.	Q	.	E	.	0.00%	3.85%
.	.	CAPYDINQ	V	.	.	E	.	26.32%	0.00%
V	.	CAPYDINQ	V	.	.	E	.	23.68%	0.00%
.	S	CAPYDINQ	.	Q	.	E	.	0.00%	92.31%
.	S	CAPYDINQ	.	Q	.	E	A	0.00%	3.85%
Other (with T182A mutation)								-	-

								Week 119	Week 123	Week 157
								19 Reads	20 Reads	75 Reads
I161	P172	CM9	I206	P221	A222	D244	M250			
.	.	CPYDINQ	.	.	T	E	.	21.05%	5.00%	0.00%
.	.	CPYDINQ	.	Q	.	E	.	0.00%	0.00%	1.33%
.	S	CPYDINQ	.	.	T	E	.	52.63%	40.00%	0.00%
.	S	CPYDINQ	.	Q	.	E	.	21.05%	50.00%	86.67%
.	S	CPYDINQ	.	Q	.	E	A	0.00%	0.00%	8.00%
Other (with T182P mutation)								5.27%	5.00%	4.00%

								Week 132
								20 Reads
I161	P172	CM9	I206	P221	A222	D244	M250	
.	S	CYPYDINQ	.	Q	.	E	.	75.00%
.	S	CYPYDINQ	.	.	T	E	.	25.00%
Other (with T182Y mutation)								-

FIG. 4. Linkage of Gag₁₈₁₋₁₈₉CM9 escape and flanking mutations in rhesus macaque r00045. Each table shows the amino acid substitution within Gag₁₈₁₋₁₈₉CM9 (left, bold). Positions where nonsynonymous mutations were observed are listed (top left). Time points postinfection are listed (top right). The frequency of each pattern of linked mutations within the viral quasiespecies is listed (bottom right).

rate of 1×10^{-4} , is 9×10^{-8} (18). However, there are additional conditions that must be met for viruses with Gag₁₈₁₋₁₈₉CM9 escape to amass within the host. Our results support previous findings showing that the number of nonsyn-

onymous mutations allowed at T182 of Gag₁₈₁₋₁₈₉CM9 is limited, making the simultaneous occurrence of Gag₁₈₁₋₁₈₉CM9 escape and a compensatory mutation even more unlikely. Additionally, both necessary mutations must occur in a virus that

FIG. 5. Three-dimensional representation of the crystal structure of SIVmac239 capsid. This SIVmac239 capsid model was generated using Geno3D and HIV-1 protein databank file 1L6N as a template. (A) Distances between the carboxyl group of amino acid T182 and the carboxyl groups of amino acids I206 and S242. Yellow dotted lines represent distances in angstroms (Å). (B) No hydrogen bonds exist between amino acid T182 and amino acid S242, V243, or D244. Green dashed lines represent hydrogen bonds.

FIG. 6. Summary of nonsynonymous flanking mutations associated with Gag₁₈₁₋₁₈₉CM9 escape. Blue boxes indicate amino acid positions where the majority (>50%) of viral genomes carried a nonsynonymous mutation at the final time point of sequencing. Yellow highlighting denotes the nine amino acids comprising the CD8⁺ T cell epitope Gag₁₈₁₋₁₈₉CM9.

is replication competent, avoids host immunity, and produces infectious particles. Given these conditions, the accumulation of viruses harboring a Gag₁₈₁₋₁₈₉CM9 escape mutation is highly rare.

The diverse patterns of flanking compensatory mutations observed in this study, along with the distinct clonal associations we detected in animals with Gag₁₈₁₋₁₈₉CM9 escape, indicate that a second level of constraint may impede viral escape in this epitope. It is possible not only that reverse transcriptase must make errors within two specific codons in a relatively short period of time but also that these errors must generate specific amino acid substitutions which together are capable of retaining viral fitness. This idea is supported by recently published random matrix theory analyses of HIV-1 clade B Gag sequences, which concluded that groups of specific amino acid positions within Gag mutate in a highly interdependent manner (5), presumably to maintain viral fitness while avoiding host immune responses. That innovative approach to sequence analysis has revealed a higher order of conservation within Gag. Interestingly, Gag₁₈₁₋₁₈₉CM9 and large sections of the flanking sequences we examined lie within the SIVmac239 p27 protein in a region homologous to what was defined by Dahirel et al. as “Sector 3,” the most highly conserved portion of Gag, according to random matrix theory analysis (5). However, a more thorough investigation into structural changes caused by combinations of mutations within the capsid protein is required to elucidate the limitations of Gag₁₈₁₋₁₈₉CM9 escape, and further work studying the replicative capacity of engineered SIVmac239 viruses would be required to confirm our putative novel compensatory mutations.

Given that viral fitness is the consequence of a delicate interplay between sequence evolution and host immune pressure, understanding the delayed escape within Gag₁₈₁₋₁₈₉CM9 may also require sequencing of full SIV genomes. Since sequence divergence within multiple epitopes can compound

losses in viral fitness, viral variation may be restrained to a limited number of epitopes. Thus, epitope escape that occurs at a cost to the virus in the form of limiting its fitness must take place early in infection, making escape within structurally or functionally constrained epitopes more difficult. Indeed, longitudinal genome-wide studies may begin to reveal the restrictions placed on sequence evolution in HIV and SIV.

ACKNOWLEDGMENTS

We thank Austin Hughes for invaluable advice in determining Gag₁₈₁₋₁₈₉CM9 escape probabilities. We also thank Roger Wiseman, Julie Karl, and Simon Lank for maintaining the Roche/454 GS Junior sequencing instrument, as well as for troubleshooting advice.

This publication was made possible by grant P51 RR000167 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), to the Wisconsin National Primate Research Center, University of Wisconsin-Madison; the research was conducted at a facility constructed with support from Research Facilities Improvement Program (grant RR020141-01). Additional funding was provided by grants 1R01AI084787, R01AI077376-04A1, and R33AI082880 from the National Institutes of Health (NIH).

REFERENCES

1. Allen, T. M., et al. 2000. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* **407**:386–390.
2. Allen, T. M., et al. 2000. Induction of AIDS virus-specific CTL activity in fresh, unstimulated peripheral blood lymphocytes from rhesus macaques vaccinated with a DNA prime/modified vaccinia virus Ankara boost regimen. *J. Immunol.* **164**:4968–4978.
3. Bimber, B. N., et al. 2009. Ultradeep pyrosequencing detects complex patterns of CD8⁺ T-lymphocyte escape in simian immunodeficiency virus-infected macaques. *J. Virol.* **83**:8247–8253.
4. Crawford, H., et al. 2007. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *J. Virol.* **81**:8346–8351.
5. Dahirel, V., et al. 2011. Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. *Proc. Natl. Acad. Sci. U. S. A.* **108**:11530–11535.
6. Dilernia, D. A., et al. 2008. HLA-driven convergence of HIV-1 viral subtypes B and F toward the adaptation to immune responses in human populations. *PLoS One* **3**:e3429.

7. Friedrich, T. C., et al. 2004. Extraepitopic compensatory substitutions partially restore fitness to simian immunodeficiency virus variants that escape from an immunodominant cytotoxic-T-lymphocyte response. *J. Virol.* **78**: 2581–2585.
8. Honeyborne, I., et al. 2010. HLA-Cw*03-restricted CD8+ T-cell responses targeting the HIV-1 gag major homology region drive virus immune escape and fitness constraints compensated for by intracodon variation. *J. Virol.* **84**:11279–11288.
9. Kawashima, Y., et al. 2009. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* **458**:641–645.
10. Kelleher, A. D., et al. 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J. Exp. Med.* **193**:375–386.
11. Lanman, J., et al. 2003. Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. *J. Mol. Biol.* **325**: 759–772.
12. Li, S., C. P. Hill, W. I. Sundquist, and J. T. Finch. 2000. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**:409–413.
13. Matthews, P. C., et al. 2009. HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering. *J. Virol.* **83**:4605–4615.
14. Ngandu, N. K., C. Seoighe, and K. Scheffler. 2009. Evidence of HIV-1 adaptation to host HLA alleles following chimp-to-human transmission. *Virol. J.* **6**:164.
15. Nietfield, W., et al. 1995. Sequence constraints and recognition by CTL of an HLA-B27-restricted HIV-1 gag epitope. *J. Immunol.* **154**:2189–2197.
16. O'Connor, D. H., et al. 2004. A dominant role for CD8+ T-lymphocyte selection in simian immunodeficiency virus sequence variation. *J. Virol.* **78**:14012–14022.
17. O'Connor, D. H., et al. 2003. Major histocompatibility complex class I alleles associated with slow simian immunodeficiency virus disease progression bind epitopes recognized by dominant acute-phase cytotoxic-T-lymphocyte responses. *J. Virol.* **77**:9029–9040.
18. Overbaugh, J., and C. R. Bangham. 2001. Selection forces and constraints on retroviral sequence variation. *Science* **292**:1106–1109.
19. Owens, C. M., et al. 2004. Binding and susceptibility to postentry restriction factors in monkey cells are specified by distinct regions of the human immunodeficiency virus type 1 capsid. *J. Virol.* **78**:5423–5437.
20. Peyerl, F. W., et al. 2003. Simian-human immunodeficiency virus escape from cytotoxic T-lymphocyte recognition at a structurally constrained epitope. *J. Virol.* **77**:12572–12578.
21. Peyerl, F. W., et al. 2004. Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *J. Virol.* **78**:13901–13910.
22. Schneidewind, A., et al. 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.
23. Valentine, L. E., et al. 2009. Infection with “escaped” virus variants impairs control of simian immunodeficiency virus SIVmac239 replication in Mamu-B*08-positive macaques. *J. Virol.* **83**:11514–11527.
24. Vogel, T. U., et al. 2003. Multispecific vaccine-induced mucosal cytotoxic T lymphocytes reduce acute-phase viral replication but fail in long-term control of simian immunodeficiency virus SIVmac239. *J. Virol.* **77**:13348–13360.
25. Wilson, N. A., et al. 2006. Vaccine-induced cellular immune responses reduce plasma viral concentrations after repeated low-dose challenge with pathogenic simian immunodeficiency virus SIVmac239. *J. Virol.* **80**:5875–5885.
26. Yeh, W. W., et al. 2006. Compensatory substitutions restore normal core assembly in simian immunodeficiency virus isolates with Gag epitope cytotoxic T-lymphocyte escape mutations. *J. Virol.* **80**:8168–8177.