A Dominant Role for CD8⁺-T-Lymphocyte Selection in Simian Immunodeficiency Virus Sequence Variation†

David H. O'Connor,¹ Adrian B. McDermott,¹ Kendall C. Krebs,¹ Elizabeth J. Dodds,¹ Jacqueline E. Miller,¹ Edna J. Gonzalez,¹ Timothy J. Jacoby,¹ Levi Yant,¹ Helen Piontkivska,² Ralph Pantophlet,³ Dennis R. Burton,³ William M. Rehrauer,¹ Nancy Wilson,¹ Austin L. Hughes,² and David I. Watkins^{1*}

Wisconsin Primate Research Center and Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, Wisconsin¹; Department of Biological Sciences, University of South Carolina, Columbia, South Carolina²; and The Scripps Research Institute, La Jolla, California³

Received 11 May 2004/Accepted 4 August 2004

CD8⁺ T lymphocytes (CD8-TL) select viral escape variants in both human immunodeficiency virus and simian immunodeficiency virus (SIV) infections. The frequency of CD8-TL viral escape as well as the contribution of escape to overall virus diversification has not been assessed. We quantified CD8-TL selection in SIV infections by sequencing viral genomes from 35 SIVmac239-infected animals at the time of euthanasia. Here we show that positive selection for sequences encoding 46 known CD8-TL epitopes is comparable to the positive selection observed for the variable loops of *env*. We also found that >60% of viral variation outside of the viral envelope occurs within recognized CD8-TL epitopes. Therefore, we conclude that CD8-TL selection is the dominant cause of SIV diversification outside of the envelope.

CD8⁺-T-lymphocyte (CD8-TL) responses against human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) select viral escape variants that are poorly recognized by the immune system. The number of CD8-TL responses that select escape variants and the contribution of CD8-TL selection to viral diversification are unknown. CD8-TL responses against HIV and SIV are broadly directed, as they target epitopes in all viral proteins (11, 26). If many of these responses select for viral variants, CD8-TL escape may be principally responsible for sequence changes that accumulate within infected individuals, an outcome recently hypothesized by Overbaugh and Bangham (40).

Until recently, the molecular and immunologic tools that are necessary to analyze CD8-TL selection at the genome level have been unavailable. Most studies of CD8-TL escape have focused on single epitopes or small groups of epitopes clustered in single SIV and HIV genes (4, 7, 16, 20, 22, 43, 44). While they were instrumental in demonstrating that CD8-TL selection is a cardinal feature of SIV and HIV infections, these studies could not comprehensively evaluate the frequency of CD8-TL escape throughout the viral genome.

High-throughput DNA sequencing technologies now enable the sequencing of entire SIV genomes from infected animals (38). Additionally, 46 CD8-TL epitopes bound by the high-frequency Indian rhesus macaque major histocompatibility complex class I (MHC class I) alleles Mamu-A*01 (3), -A*02 (28a), and -B*17 (34) are now known. By sequencing complete

viral genomes spanning these 46 CD8-TL epitopes, we can now directly measure the frequency of CD8-TL selection.

With this study, we found that CD8-TL escape is common, with sequences encoding 23 of 46 epitopes accumulating variations that are consistent with escape. Within a single animal, >64% of the nonsynonymous viral variants outside of the envelope were located within recognized CD8-TL epitopes, suggesting that CD8-TL escape is a primary selector of within-host diversity.

MATERIALS AND METHODS

Viruses and animals. The 35 animals used for the bulk of the analyses were infected with SIVmac239 between 1998 and 2001. These animals were challenged with SIV in trials of vaccine efficacy and pathogenesis (2, 21, 33, 39). Several of the animals received vaccines encoding SIV gene fragments, but we do not believe that these vaccines altered the natural course of infection or modulated the overall immune responses to the virus. Viral loads were monitored either by quantitative kinetic reverse transcription-PCR as previously described (29) or by branched DNA assays (Bayer Diagnostics, Emeryville, Calif.). CD4⁺ counts were determined as previously described (15). Animals were euthanized when they showed signs of physical distress or developed opportunistic infections typical of simian AIDS.

MHC genotyping. Allele-specific primers for Mamu-A*01, -A*02, and -B*17 were used to genotype all 35 animals as previously described (25).

Statistical analyses. The number of synonymous substitutions per synonymous site (d_s) and the number of nonsynonymous substitutions per nonsynonymous site (d_n) were estimated by the method of Nei and Gojobori (35) for the sequence obtained from each virus and for the inoculum sequence. Since the Vif₁₀₀₋₁₀₇ VI8 CD8-TL epitope was identified partly based on variation data from this study, it was excluded from statistical analyses. In preliminary analyses, more complex models of nucleotide substitution (28, 50) yielded identical results, as expected when (as in the present data) the number of substitutions per site is low (36). In the case of ambiguous nucleotides, we assumed equal occurrences of the possible nucleotides in the viral population in any given monkey. Because the virus from each monkey was compared independently with the inoculum, each comparison was phylogenetically (17) and thus statistically independent. The property of statistical independence enabled us to use statistical methods that are not dependent on assumptions regarding the substitution model and thus have more robustness than model-based methods that are typically applied for analyses of nucleotide substitution patterns. Undetermined bases, which likely

* Corresponding author. Mailing address: Department of Pathology, University of Wisconsin, 1300 University Ave., Madison, WI 53706. Phone: (608) 265-3380. Fax: (608) 265-8084. E-mail: watkins@primate.wisc.edu.

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

resulted from insertion-deletion replacements, were excluded from the analyses.

Variable loop regions. All variable loop regions were designated as described in another study (12).

Viral sequencing. Viruses were sequenced at the time of host death as described previously (38), with slight modifications. An additional set of eight PCR primer pairs was used to amplify problematic regions of the genome. Mixed populations of sequences with different lengths, particularly in *env*, could not be resolved by direct sequencing and are annotated with "X's" in Fig. S1 in the supplemental material. The sequences were run on either an ABI 377 or ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, Calif.). Sequences were edited with Sequencer 4.1 (Genecodes, Ann Arbor, Mich.), and mixed-base polymorphisms were identified automatically by Sequencer. Nucleotide sequences were aligned with the wild-type SIVmac239 sequence in MacVector 7.2, trial version (Accelrys, San Diego, Calif.). These nucleotide alignments were conceptually translated into amino acid alignments that distinguish mixed-base substitutions from complete substitutions by use of a script written by D. H. O'Connor in Lasso dynamic markup language (Blueworld, Bellevue, Wash.).

Enumerating SIV-specific CD8-TL responses by IFN- γ ELISPOT. An enzyme-linked immunospot (ELISPOT) assay for gamma interferon (IFN- γ) was performed as previously described (29). For investigations of CD8-TL responses spanning variable sites, 15-mer peptides spanning variable sites in multiple orientations (e.g., 15-mers with the variable site at the beginning, middle, or end) were tested for their reactivities. Experimental samples with >10 spots and a mean spot count of more than the following were considered positive: (mean spot-forming count [SFC] of wells containing no peptide) + 2 $\times$ (standard deviation of wells containing no peptide). All peptides were tested in duplicate.

Peptide pools containing 10 peptides each were tested for their reactivities with cells from SIVmac239-infected animals infected at the National Primate Research Center between 1999 and 2004. The same criteria for positivity were used for peptide pools as for individual peptides. Each peptide pool was tested for recognition with cells from at least 22 animals. Peptide pools were tested at least once per animal, and in many instances, the peptide pools were tested at several time points for a single animal. The frequency of recognition was determined by calculating the number of animals whose cells recognized a peptide pool and dividing this by the number of animals whose cells were tested with the peptide pool. The average ELISPOT reactivity was calculated by taking the arithmetic mean of the recognition frequencies for all peptide pools for a particular protein.

Nucleotide sequence accession numbers. The SIVmac virus sequences examined in this study are available in GenBank under accession numbers AY611486 to AY611495, AY576480 to AY576481, AY599198 to AY599201, AY607701 to AY607704, AY600249, AY587015, AY597209, AY603959, and AY588945 to AY588946.

RESULTS

Viral variation is consistent with escape from 23 SIV CD8-TL epitopes. We assessed the frequency of CD8-TL escape within individual hosts by examining 46 CD8-TL responses restricted by the high-frequency Indian rhesus macaque MHC class I alleles Mamu-A*01, -A*02, and -B*17 (3, 34). In addition to the 45 previously described epitopes, a Mamu-A*01-restricted CD8-TL response in Vif (Vif₁₀₀₋₁₀₇VI8) that was previously deemed nonreactive (3) was identified during this study. These epitopes bind to MHC class I molecules with high affinities and are recognized by CD8-TL isolated from SIV-infected animals, so this epitope set was not intentionally biased with respect to escape tendency, epitope localization, and response avidity.

The sequences encoding the 46 CD8-TL epitopes were evaluated for viruses isolated from 35 SIVmac239-infected animals at the time of euthanasia. Of the 35 animals, 21 were positive for Mamu-A*01, 6 were positive for Mamu-A*02, and 4 were positive for Mamu-B*17. The survivorship of the 35 animals ranged from 14 to 167 weeks, with viral loads at the time of sacrifice ranging from 9,500 to 410,000,000 viral RNA copies per ml of plasma (Table 1). These values are typical of animals

TABLE 1. Survival, clinical parameters at time of sacrifice, and PCR-SSP MHC class I typing of 35 SIVmac239-infected animals

Animal no.	Survival (wks)	No. of CD4 ⁺ cells/ μ l	Viral load (RNA copies/ml of plasma)	Presence of allele		
				Mamu-A*01	Mamu-A*02	Mamu-B*17
80025	14	314	4.1×10^6	+	—	—
90131	16	180	1.7×10^7	—	—	—
87108	17	476	4.1×10^6	+	—	—
96114	22	NA ^a	1.2×10^5	+	—	—
95112	22	1216	4.6×10^7	—	—	—
96081	26	NA ^a	4.1×10^8	—	—	—
97074	28	NA ^a	NA ^a	—	—	+
97009	28	158	NA ^a	—	—	—
81035	29	101	5.0×10^4	—	—	—
92077	31	485	5.2×10^6	+	—	—
87082	32	113	1.2×10^7	—	+	—
92050	38	238	1.6×10^7	—	+	—
96016	39	NA ^a	2.3×10^7	+	—	—
96135	40	2441	1.4×10^5	+	—	—
95086	44	672	2.8×10^6	+	—	—
93062	44	269	2.6×10^5	—	—	—
80035	44	50	9.7×10^4	+	—	—
96123	45	NA ^a	1.3×10^6	+	—	—
95045	54	644	9.7×10^6	+	—	—
96020	57	NA ^a	NA ^a	—	+	—
85013	58	203	2.5×10^5	+	—	—
96104	60	NA ^a	5.8×10^6	—	—	—
96093	64	NA ^a	3.7×10^6	—	—	—
95084	66	1241	6.1×10^5	+	+	—
96072	67	NA ^a	8.5×10^6	—	+	+
93057	68	NA ^a	2.6×10^5	+	—	—
2127	69	NA ^a	NA ^a	+	—	—
95003	70	1733	4.5×10^6	—	—	—
2065	75	NA ^a	9.5×10^3	+	—	+
95058	89	264	5.4×10^6	+	—	—
95115	90	294	2.0×10^5	+	—	—
1975	100	667	1.5×10^5	+	—	—
96118	109	NA ^a	NA ^a	+	—	—
96031	163	NA ^a	NA ^a	+	+	—
1937	167	46	3.6×10^5	+	—	+

^a NA, not performed at time of death.

infected with SIVmac239, for which the median time of survival is approximately 52 weeks (39).

The virus isolated at the time of sacrifice from each animal was sequenced to identify the dominant nucleotide(s) present at each site in the viral genome (38) (see Fig. S1 in the supplemental material). The sequences at the time of host death were conceptually translated and then aligned with SIVmac239 protein sequences.

Sequences encoding 57% (8 of 14) of the Mamu-A*01-restricted CD8-TL epitopes had amino acid changes in at least 1 of the 21 Mamu-A*01-positive animals (Fig. 1). Viruses from individual animals had amino acid replacements in one to seven Mamu-A*01-restricted CD8-TL epitopes (Fig. 1). The number of epitopes harboring variation correlated with survivorship: an average of 1.9 Mamu-A*01-restricted epitopes accumulated variations in animals that survived for less than 1 year, while animals that survived for more than 1 year had variations in 4.2 Mamu-A*01-restricted epitopes. This suggests that variation accumulates throughout an infection, with some variations arising only after a prolonged infection. Seven of the eight Mamu-A*01-positive animals with the longest times of survival and known viral loads at the time of sacrifice had viral

A. Mamu-A*01

	Gag ₁₈₁₋₁₈₉ CM9	Gag ₃₇₂₋₃₇₅ LF8	Pol ₁₄₇₋₁₅₆ LV10	Pol ₅₂₅₋₅₃₁ SV9	Vif ₁₀₀₋₁₀₇ V18	Tat ₂₈₋₃₅ SL8	Env ₂₃₃₋₂₄₁ CL9	Env ₅₂₀₋₅₂₈ TL9	Env ₇₂₆₋₇₃₅ ST9
	CTPYDINQM	LAPVPIPF	LGPHYTPKIV	STPPLVRLV	VTFPHYADI	STPESANL	CAPPQYALL	TVPWPNASL	SSPPSYFQQT
80025					..h...t	P.....P			
87108					..i...p				
96114					..i...L..				
92077									
96016						P...l..			
96135					S	P.....			
95086					..h...t	P.....P			
80035					t	P.....EF			
96123					S	P.....W			
95045						P.....			
85013	..A.....				..h...t	P.....			
95084	..i.....				..c...t	P.....Q			
93057	..I.....				..D...t	P...l..			
2127	..C.....				..D...t	KP			
2065					..c...X	P.....			
95058	..I.....				..Dc..X	P...g..			
95115						P.....			
1975	G.....	..A.....			..d...X	P.....			
96118	..A.....					P.....Q		N.....	
96031	f.....				..dC..V	P.....R			
1937	..C.....					P.....		..l...XNI	
90131									
95112									
96081									
97074						P.....			
97009									
81035									
87082									
92050									
93062									
96020									
96104									
96093									
96072									
95003									

(over-lapping Mamu-A*02 epitope Vif₁₀₀₋₁₀₇WY8 in Mamu-A*02+ 96020 and 96072)

B. Mamu-A*02

	Gag ₇₁₋₇₅ GY9	Vif ₉₇₋₁₀₄ WY8	Env ₇₈₇₋₇₉₄ RY8	Nef ₁₀₂₋₁₀₉ LY9	Nef ₁₁₀₋₁₁₉ TM10	Nef ₁₂₅₋₁₃₇ YY9	Nef ₂₂₁₋₂₂₉ YY9	Nef ₂₃₈₋₂₄₅ LM9
	GSENLKSLY	WTDVTPNY	RTIISLNKY	RTLLSRVY	LLRARGETY	TMSYKLAIDM	YTSQPGIRY	YTYEAYVRY
87082								
92050								
96020		D.						
95084	F.....							
96072	D.							
96031	s.d..r...	dC						
80025		h.						
90131								
87108								
96114					C			
95112								
96081								
97074								
97009								
81035								
92077								
96016								
96135			H					
95086		h.						
93062								
80035								
96123								
95045								
85013		h.						
96104								
96093								
93057		D.						
2127		D.						
95003								
2065		C						
95058		Dc						
95115								
1975		d.						
96118								
1937								

(over-lapping Mamu-A*01 epitope Vif₁₀₀₋₁₀₇WY8 in Mamu-A*01+)

(over-lapping Mamu-B*17 epitope Nef₁₂₅₋₁₃₇YW9 in Mamu-B*17+)

FIG. 1. Amino acid replacements in sequences encoding 23 known CD8-TL epitopes. Nucleotide sequences spanning each CD8-TL epitope were conceptually translated. Dots indicate identity with the wild-type amino acid residue. Amino acid replacements that resulted from a mixed population of the wild-type sequence and a variant sequence are indicated with lowercase one-letter amino acid abbreviations of the variant amino acids. For example, a mixed base in the dominant sequence within the second codon of Gag₁₈₁₋₁₈₉CM9 in animal 95084 gave rise to two possible amino acid outcomes, the wild-type threonine (T) or a variant isoleucine (I). In the alignment, this change is denoted with "i." Codons for which more than two amino acids can be inferred from the direct sequence are annotated with an "X." Complete nucleotide substitutions that changed the resultant amino acid are indicated with uppercase one-letter amino acid abbreviations. (A) Amino acid replacements in eight previously identified Mamu-A*01-restricted CD8-TL epitopes, plus a ninth epitope (Vif₁₀₀₋₁₀₇V18) that was identified in this study. (B) Amino acid replacements in nine Mamu-A*02-restricted CD8-TL epitopes. (C) Amino acid replacements in five Mamu-B*17-restricted CD8-TL epitopes.

C. Mamu-B*17

	Vif ₈₅₋₇₃ SW9	Env ₈₁₆₋₈₂₈ LY10	Env ₈₃₀₋₈₃₈ FW9	Nef ₁₆₅₋₁₇₃ IW9	Nef ₁₉₅₋₂₀₃ MW9
Mamu-B*17 Positive					
97074	SHLEVGQYW	LRTELTYLQY	FHEAVQAVW	IRYPKTFGW	MHPAQTQW
96072		..I.....	..Y...F.	T.....	I.....
2065		a.	..X.....	T.....	X.....
1937	..q.....		..Y.....	T...H...	I.....
80025					
90131					
87108					
96114					
95112			..q.....		
96081					
97009					
81035					
92077					
87082		A.		..f.....	
92050		A.		..F.....	
96016					
96135					
95086					
93062			..Q.....		
80035					
96123					
95045					
96020			X.	..F.....	
85013					
96104			..q.....		
96093			g.		..V....
95084				..F.....	
93057					
2127					
95003	..Y.....				
95058					
95115			X.		
1975					
96118				T...i...	
96031				..F.....	..V....
(overlapping Mamu-A*02 epitope Nef ₁₅₉₋₁₆₇ YY9 in Mamu-A*02+)					
Mamu-B*17 Negative					

FIG. 1—Continued.

loads below 10⁶ copies/ml, which may reflect a replicative disadvantage to viruses containing mutations in multiple CD8-TL epitopes. Alternately, lower viral loads may simply delay the onset of AIDS (30), lengthening the time of survival and increasing the likelihood of developing CD8-TL epitope mutations. Viruses from 14 Mamu-A*01-negative animals were sequenced to exclude the possibility that this variation was due to epitope localization in mutational hot spots or to genetic drift. Sixty-seven Mamu-A*01-restricted epitopes accumulated variations in the 21 Mamu-A*01-positive macaques (mean, 3.2 per animal), and only nine sequences encoding Mamu-A*01-restricted epitopes contained variants in the 14 Mamu-A*01-negative animals (mean, 0.6 per animal) (Fig. 1; also see Fig. S2 in the supplemental material). Similar analyses were performed for the six Mamu-A*02-positive and four Mamu-B*17-positive animals. Forty-seven percent (9 of 19) of the Mamu-A*02-restricted epitopes had variations in Mamu-A*02-positive animals (Fig. 1). Each Mamu-A*02-positive animal had variations in at least two Mamu-A*02 epitopes, while the longest lived Mamu-A*02-positive animal, 96031, had variations in seven Mamu-A*02-restricted epitopes. Forty-two percent (5 of 12) of the Mamu-B*17-restricted epitopes accumulated substitutions in Mamu-B*17-positive animals, with an average of 3.8 variable epitopes per animal (Fig. 1).

Most epitopes had a characteristic pattern of amino acid variation, with variations accumulating at the same residues in multiple animals. For example, viruses from all six Mamu-A*02-positive animals contained a replacement of serine for

leucine at amino acid 5 of the Env₇₈₇₋₇₉₄RY8 epitope (Fig. 1). None of the sequences from Mamu-A*02-negative animals, in contrast, had variations in this epitope. This reproducible variation suggests that the sequence space available to CD8-TL epitopes is restricted, and thus most possible variants will have deleterious structural or functional consequences (18, 37, 42). The emergence of identical amino acid variants is a signature of CD8-TL selection, as previously suggested for an analysis of six SIV CD8-TL epitopes (6). Interestingly, variation does not occur in the same epitopes in all animals. The Mamu-A*01-restricted Env₆₂₀₋₆₂₈TL9 epitope, for example, accumulated variations in 52% of the Mamu-A*01-positive animals (11 of 21). The prevalent amino acid substitution in this epitope, however, was consistent, with a serine-for-asparagine replacement at position 8 in 8 of the 11 animals.

We also investigated whether animals expressing a combination of MHC class I alleles harbor viruses with escape epitopes bound by multiple alleles. Five of the animals expressed two of the MHC class I alleles considered in this study. The Mamu-A*01- and -A*02-positive macaques 95084 and 96031 had amino acid replacements in 10 and 13 CD8-TL epitopes, respectively (Fig. 1). Similarly, sequences from the Mamu-A*01- and -B*17-positive animals 2065 and 1937 had variations in an average of 7.5 epitopes. In both instances, the number of epitopes with variations was approximately double the number of epitopes that varied in animals with only one of these alleles. Viruses from Mamu-A*01-, -A*02-, and -B*17-negative animals, in contrast, averaged variation in only 1.9 epitopes. Therefore, the amount of variation in sequences encoding CD8-TL epitopes appears to be independent of the MHC class I allele and is additive.

The majority of SIV variation outside of the envelope occurs in CD8-TL epitopes. Since most rhesus macaques express more than six MHC class I alleles (8), Mamu-A*01-, -A*02-, and -B*17 epitope variations likely underestimate the contribution of CD8-TL pressure to SIV diversification. We tested this with animal 2065, which was selected because it expressed both Mamu-A*01 and Mamu-B*17 and had an intermediate time of survival (75 weeks). The virus recovered from animal 2065 had 50 nonsynonymous nucleotide substitutions at the host's time of death. Twenty-five of these nonsynonymous substitutions were in *env*. Env is a target for immune selection by both antibodies (41, 45, 48) and CD8-TL (7). Overlapping epitopes recognized by both humoral and cellular immune responses (26), as well as CD8-TL epitopes that contain residues responsible for altered cell tropism (46), confound the attribution of particular sequence changes to discrete selective pressures. Therefore, substitutions in *env* were excluded from our analysis. Three of the 25 non-*env* substitutions were suboptimal nucleotides in the SIVmac239 molecular clone (1). These suboptimal nucleotides were also excluded from further analyses.

Of the remaining 22 substitutions, 18% (4 of 22) (Fig. 2) were located within Mamu-A*01- and Mamu-B*17-restricted CD8-TL epitopes. Two of these epitopes (Pol₁₄₇₋₁₅₆LV10 and Nef₁₆₅₋₁₇₃IW9) were recognized at 70 weeks postinfection by CD8-TL from animal 2065 in an IFN-γ ELISPOT assay (data not shown). The Tat₂₈₋₃₅SL8 epitope was not recognized by CD8-TL at 70 weeks postinfection, but acute-phase CD8-TL against this epitope in animal 2065 have been described previously (39). Epitope variation in the fourth epitope,

Gag

SIYmac239 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170
 2065 MVTNVLVSGKADALEKTLRPNKGYMLKHVVWANEIDRFGALSLENKSGCKLISVLAPVPTTOSKLSIYTNVYINCIBAEKVKHTEAKQIVQRLVETGTTMPKSRPTAPSSRGNGYVQQIGNYVHLRSPRTLNAMVAKLIEKKFGAV
 SIYmac239 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340
 2065 VPOFQALSECCTFYDINQMLACVGBDQAAQMIIRDIINEEAADMDLQHPQAPQOGQLRPSGSDIAGTSSVDQICIMWYQNPFPVGNVYRWIQLGLQKCYVMYNTNILDVKGQKPEPQSVYDFYKSLRAEQTDAAVKNMWTQTLIQANPDCKLVLKGLGV
 SIYmac239 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510
 2065 NPTLEMLTACGGYGGQKARLMAEALKEALAPVFPFAAQQRQFKPIKCMCKGBGSAQRAPRQCKQCKMDKVMKAPDRQAGFLGQFKPKPFPMAQVHQGMPTAPDEPDVLDLLKNYMQLKQKQREKPEYKVEDLLHNSLFGDQ*

Pol

SIYmac239 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
 2065 VLELWERTGTCAMQSKFKTKMLNKKNGFCYGMFQGTGGFFRPSWGEKAPQFFHGSASGADANCSFGPSGSAKELHAYQAARKEKRAQOGDGFAPQPSIMRRPVTAHIEGQFVFLDTGADDSVTGIELAPFPYFKIVGGGFINTEKXNVEYLGKRI
 SIYmac239 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360
 2065 KGTIMTGDTPINIFGRNLLTALQMSLNFPIAKVPEVVKALFGDGPLKQWPLSKKIVALREICERKEDQGLEAPTPNPTFTAIKKDKNKWRMLIDFRELNRVTQDFTVEQLGIPHPAGLAKRRITVLDIGONFSLDEEFRTAFTLPSYNNAEFGKRIYKVLFPQQ
 SIYmac239 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540
 2065 WKGSFAPQITWREHVLEPRKAMPDVTLYQMDILLASDRTDLEHDAVILSKELLNSIGHSPEEFKQDPPQWGYELPTKWKILQELPORSTWVNDIQKLYGNLWAAQIYPQGIKTHCLRLISGKMTLLEVTQWTEMAEYENKILLSQEGCTYQEGKPLEATVKS
 SIYmac239 550 560 570 580 590 600 610 620 630 640 650 660 670 680 690 700 710 720
 2065 QDNQWKRIGQEDKILKVGKFAKINKTHINGVELLARVYQKIGKEALVINGQVFKFHPVETWQCAWIDYQVNTWIFEDFISTEPFLVLPVNDVDFEETIIDGSCNKSKEGAGYITDGRKDVKYLEQTINQQAELFAFLALITDSGPKANIIVDSQIVMGIITCPTES
 SIYmac239 730 740 750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900
 2065 ESRLYNQIIEEMIKKSEIYAVVPAHKIGIGNQIEIDHLSQIRQVLFLEKIEPAQEHKYNKVELFKGLPRIVARQIVDTCDCQKGEAIGHANSDLGTWMDCTHLSGKIITIAVHVASGFIEARVPOQTGROTALFLLLKLAGRWPITLHHTDNGANFASQEVKIVAWWA
 SIYmac239 910 920 930 940 950 960 970 980 990 1000 1010 1020 1030 1040 1050
 2065 GIEHTFGVFNPGQGVVFAAMNHLKNGIDIRIEQANSVETIVLMAVCMNFRGGIGDMTPAERLNMITTEQIQQQSKNSKFFVYRREGDQMKMGELMKWEGAVILKVTQDIKVPVREKAKIKIDYGGKEVDSHSHMEDTGEAREVA*

Vif

SIYmac239 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
 2065 MEERKLVNPTWPRLESMHSLKYLKTKLQKVCYVPHFVQKAWMTCSRVIPFLQEGSHLVQCYTHLTPTKTMHLSITAVRITWYSKNWNTDTPYVAVIILESTYPTCTAGEVRAIRGEQLLSCFRPANKVQVPSIQY
 SIYmac239 160 170 180 190 200 210 220 230 240 250 260
 2065 LALKVSDVRSQNGENTWQKWRDNRGLMAKQNSRGKQKQKPKTKNAFFPLAKVGLTA*

Vpx

SIYmac239 10 20 30 40 50 60 70 80 90 100 110
 2065 MSDPERIPGNSGEETIGEAPEWLNKTVSEINREAVNHLPELILFQWQSRWEYWDQMSPSYKRVLCLOKALPMCKKGCRLGEGHAGGWAPGPPPPPGFLA*

Vpr

SIYmac239 10 20 30 40 50 60 70 80 90 100
 2065 MEERPENEGQREPDWVYVLEELKEALKHPDPLLLTALGNHYNRHGDTLEGAGELIRLOALPMHFRGQTHSRICQPGGNGPLSALPPSSML*

Tat

SIYmac239 10 20 30 40 50 60 70 80 90 100 110 120 130
 2065 METPLREQNSLESSNERSISCISEADSTPESANIGHEILSOLYRLEACVNTCYCKCHQCFCKLGLGICVQSKEKRRTPKAKANTSSASNKPISNRTRECOPEKAKKETVEKAVATAPGLGR*

Rev

SIYmac239 10 20 30 40 50 60 70 80 90 100
 2065 MSNHERRELKRLRLHLHQTNPYTPGTANQRQKRWREWOQLALADRIYSPDPPTDPLDIAIQQLQMLAIESIPDPPTNTPALCDPTEDSGSPQ*

Nef

SIYmac239 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
 2065 MGALEHAKRERPSGDLRQKLLRARGETTYGRLGVEGDSQSGPDGLGSSISCEQKNGYNTFWNPAERKIAFRKQNMDDIDEEDDLQVGSVRFPVPLRNGYKLAIDMSFTKEKGLGEGYYSARRRILDIYEEER
 SIYmac239 160 170 180 190 200 210 220 230 240 250 260
 2065 GIIPWDQDTSPIHLVYKPTFGMLKVLVFNYSDEADEEYVHMPVACQTSQDDPQVGEVLAWEFDPLATVYAVYEPFGSKGLSEEVRRRLTARGILNMAQKKETR*

Non-Synonymous Variants:
 Suboptimal nucleotide sites:
 Recognized epitopes in variable regions:
 Recognized Mamu-A*01 epitopes without variation:
 Recognized Mamu-B*17 epitopes without variation:
 Known Mamu-A*01 epitopes not recognized in 2065:
 Known Mamu-B*17 epitopes not recognized in 2065:
 Known Mamu-A*01 epitopes not tested for recognition in 2065

FIG. 2. CD8-TL responses against variable regions in the virus from macaque 2065 at the time of sacrifice. The nucleotide sequence of the virus from animal 2065 was translated and aligned with the wild-type SIVmac239 sequence as described in the legend to Fig. 1. The 25 amino acid variants are boxed in black. Three of these 25 amino acid variants resulted from the selection of variants at suboptimal nucleotides in the SIVmac239 molecular clone, and these are shaded green. We first examined whether the remaining 22 variants were located within known Mamu-A*01- and Mamu-B*17-restricted CD8-TL epitopes recognized by animal 2065 (1). Recognized Mamu-A*01-restricted epitopes containing variations are boxed in magenta with yellow shading. Similarly, recognized Mamu-B*17-restricted epitopes containing variations are boxed in blue with yellow shading. Recognized Mamu-A*01 and Mamu-B*17 epitopes against invariant regions in the virus from animal 2065 virus are shaded magenta and blue, respectively. Known Mamu-A*01 and Mamu-B*17 epitopes that were not recognized by CD8-TL from animal 2065 are boxed in magenta and blue, respectively. Two known Mamu-A*01 epitopes were not tested and are indicated with red boxes with double lines. The Mamu-A*01-restricted Vif₁₀₀₋₁₀₇ V18 epitope that was discovered in this study is shown with a red dashed box. Yellow shading with no flanking box indicates CD8-TL-reactive regions outside of known Mamu-A*01 and Mamu-B*17 epitopes that were identified by IFN- γ ELISPOT.

Nef₁₉₅₋₂₀₃MW9, was observed by 29 weeks postinfection, and it is possible that our failure to detect this response at 70 weeks postinfection was due to this early variation.

We synthesized 15-mer peptides spanning each of the remaining 18 variants and tested them for CD8-TL reactivity following peptide stimulation. CD8-TL responses that select escape variants are difficult to detect *ex vivo* during chronic infections (24), particularly if the escape occurs early during infection (7, 44). Therefore, we tested samples collected at 16, 64, and 75 weeks postinfection. Peptides spanning 11 of the 18 amino acid variants were recognized by CD8-TL (Fig. 2; also see Fig. S3 in the supplemental material). In total, 64% (14 of 22) of the amino acid substitutions in animal 2065 occurred within recognized CD8-TL epitopes, demonstrating that CD8-TL selection is the dominant cause of SIV variation outside of *env*.

Positive selection of known CD8-TL epitopes and in *nef*.

Envelope sequence changes are concentrated in five hypervariable variable loops (termed V1 to V5) that may protect HIV from antibody recognition (9, 49). These loops contribute extensively to the overall variation of the envelope gene, which is widely cited as the apotheosis of HIV sequence diversity (19, 40). Since MHC-restricted variation frequently accumulates in CD8-TL epitopes, we reasoned that the intensity of selection within the CD8-TL epitopes might be comparable to selection in the envelope hypervariable loops.

The mean number of nonsynonymous substitutions per nonsynonymous site ($d_N = 0.0113 \pm 0.0029$ [mean $\pm$ standard error of the mean]) was significantly larger (by a paired *t* test) than the mean number of synonymous substitutions per synonymous site ($d_S = 0.0062 \pm 0.0031$) for the variable loops of *env*, consistent with the positive selection of variants (Fig. 3A). In contrast, for the remainder of *env*, the mean d_N (0.0047 ± 0.0005) and mean d_S (0.0053 ± 0.0008) were not significantly different. Surprisingly, the mean d_N for the variable loops was not significantly different from that for the non-*env* CD8-TL epitopes (0.0165 ± 0.0016). Therefore, in our cohort, nonsynonymous nucleotide substitutions accumulated as rapidly in the regions of the virus encoding known CD8-TL epitopes as in the variable loops of the envelope. Unexpectedly, we also found that the mean d_N for the variable loops did not differ significantly from that for the *nef* gene (0.0089 ± 0.0011). While other selective pressures may contribute to the unusually high mean d_N for *nef* (23), these data suggest that selection by Nef-specific CD8-TL may be unusually prominent.

We next investigated whether the sequences encoding CD8-TL epitopes throughout the entire SIV genome evolved under positive selection. We tested for differences in the patterns of nucleotide substitution among nonpeptide regions, CD8-TL epitopes that are known to be restricted by the class I MHC of the monkey from which the virus was obtained (restricted epitopes), and the same CD8-TL epitopes in animals that did not express the restricting MHC class I allele (nonrestricted epitopes). The mean d_S values showed no significant differences for restricted epitopes, nonrestricted epitopes, and nonpeptide regions (Fig. 3B). In contrast, the mean d_N was significantly higher for restricted epitopes than for either nonrestricted epitopes or nonpeptide regions (Dunnett's test; $P < 0.001$). For nonpeptide regions, the mean d_S (0.0045 ± 0.0004) did not differ significantly from the mean d_N (0.042 ± 0.0002) (paired *t* test). Similarly, for nonrestricted epitopes,

FIG. 3. Sequences encoding CD8-TL epitopes evolve under positive selection. (A) Mean numbers of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N) for the *nef* gene, restricted CD8-TL epitopes (excluding the *env* gene), and variable regions of *env*. (B) Mean numbers of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N) for non-epitope regions, nonrestricted epitopes, and restricted epitopes. A factorial analysis of variance in d_N values by a general linear model approach showed significant effects for the epitopes ($P < 0.001$). Numbers in parentheses are the numbers of comparisons versus the inoculum on which each mean is based. *, comparisons for which the mean d_N was significantly larger than the mean d_S .

the mean d_S (0.0052 ± 0.0010) and mean d_N (0.0061 ± 0.0007) did not differ significantly (paired t test). However, for restricted epitopes, the mean d_N (0.0141 ± 0.0013) was significantly larger than the mean d_S (0.0053 ± 0.0010) (paired t test; $P < 0.001$).

d_N significantly exceeded d_S in these comparisons, suggesting that positive Darwinian selection is a factor in enhancing the rate of amino acid sequence evolution in SIV. Since this pattern was observed for restricted epitopes but not for either nonrestricted epitopes or non-epitope regions, these results implicate CD8-TL as the selective factor.

Positive selection in *nef*, *rev*, and *vif* correlates with CD8-TL responses against epitopes in these proteins. We next investigated the selection of individual SIV genes. For all genes in all monkeys, the mean d_S (0.0045 ± 0.0005) was very similar to the

mean d_N (0.0040 ± 0.0003). However, both the mean d_S and the mean d_N were strikingly different among genes (Fig. 4). The *vpx* gene showed the lowest mean d_N (0.0010 ± 0.0004); this value was significantly lower than those for *vif*, *vpr*, *tat*, *rev*, *env*, and *nef* (simultaneous $P < 0.05$; Dunnett's test). For *nef*, *rev*, and *vif*, the mean d_N values for the 35 monkeys were significantly larger than the mean d_S values (paired-sample t tests; for *vif*, $P < 0.001$; for *rev*, $P = 0.005$; for *nef*, $P < 0.001$).

Given our observation that most nonsynonymous variation occurred within recognized CD8-TL epitopes, we hypothesized that the elevated d_N values for *nef*, *rev*, and *vif* were indicative of CD8-TL responses against epitopes in these proteins. To test this hypothesis, we analyzed SIV-specific IFN- γ ELISPOT results in SIVmac239-infected animals. Peripheral blood mononuclear cells (PBMC) from these animals were stimulated with pools of approximately 10 peptides each, as described previously (33). On average, each peptide was tested with cells from 29 animals, and all peptides were tested with cells from at least 22 animals. We calculated the percentages of animals that recognized, on average, all of the peptide pools for each SIV protein. The most frequently recognized proteins in this data set were Nef, Rev, Gag, and Vif (Fig. 4B). The lack of nonsynonymous variation in *gag*, in spite of its frequent recognition by CD8-TL, may be due to structural and functional constraints that limit tolerated variation (18, 42).

DISCUSSION

CD8-TL play a dominant role in selecting amino acid variations in SIV. We showed here that 50% of known SIV CD8-TL epitopes accumulate amino acid replacements consistent with CD8-TL escape and that SIV amino acid replacements outside of Env accumulate primarily as a result of CD8-TL selection. Nonetheless, the actual effect of CD8-TL escape on virus variation may still be underestimated in this study. We did not examine compensatory substitutions that occur in regions flanking CD8-TL epitopes but that may mitigate the fitness cost of maintaining amino acid replacements (18, 22, 42) or may directly reduce CD8-TL recognition by affecting antigen processing (13, 51, 52). For example, virus particles containing a substitution of cysteine for tyrosine at position Nef₁₉₃ from animal 2065 might escape recognition by Nef₁₉₅₋₂₀₃MW9 CD8-TL.

If CD8-TL escape is so pervasive, why was the existence of this phenomenon initially so controversial (5)? Our results supply three possibilities. Firstly, while we showed that 50% of known CD8-TL responses select for escape variants, studies that utilized small epitope groups may have inadvertently focused on responses that rarely select for escape variants. Indeed, while most of the strongest CD8-TL responses select for epitope variants, epitopes such as Mamu-A*01-restricted Gag₃₇₂₋₃₇₉LF8 elicit persistent, strong CD8-TL responses but rarely select viral variants until very late in infection (data not shown).

Secondly, and paradoxically, it is possible that CD8-TL escape has been difficult to detect because it is so common. d_N/d_S analyses, which are frequently used to test selection by CD8-TL, assume that sequences flanking known epitopes are devoid of other epitopes that simultaneously select for escape variants. In the animals used for this study, the mean d_N values

FIG. 4. Significant elevation of d_N in three SIV genes that are frequently recognized by CD8-TL. (A) Mean numbers of synonymous substitutions per synonymous site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N) for the nine SIV protein-coding genes. Means for both d_S and d_N were significantly different among genes (one-way analysis of variance; $P < 0.001$ in both cases). *, genes for which the mean d_N was significantly larger than the mean d_S . (B) Frequencies of IFN- γ ELISPOT recognition of SIV. Pools of 10 peptides were tested for IFN- γ ELISPOT reactivity with cells from at least 22 SIVmac239-infected animals, and the frequencies of recognition are shown with gray bars. The location of the gray bar within a protein corresponds with the proteomic location of the peptide pool (e.g., the left-most gray bar for Gag is the peptide pool containing the amino terminus of Gag). The mean frequency of recognition for the peptide pools in a protein is shown with a red bar to the right of each peptide pool series.

were significantly elevated for *vif*, *rev*, and *nef*, an observation most consistent with frequent, widespread CD8-TL selection. Therefore, d_N/d_S studies involving sequences encoding known CD8-TL epitopes may be confounded by the presence of additional epitopes that are also evolving under CD8-TL pressure.

Thirdly, CD8-TL responses against variable regions declined after the epitope that they recognized escaped. For example, responses against Pol₆₀₁₋₆₁₅, Vif₆₉₋₈₃, and Vif₁₉₇₋₂₁₁ that were present at 16 weeks postinfection became undetectable at later time points (see Fig. S2 in the supplemental material). If we had not known the composition of our input virus and monitored the immune responses during early infection, we would not have detected CD8-TL-mediated selection in sequences encoding these epitopes.

The interplay between CD8-TL selection and viral diversity is complex. Within an individual, the majority of variation

outside of the envelope is clustered within CD8-TL epitopes, yet not all CD8-TL responses select for escape variants. The factors that govern the tendency of responses to select escape variants have not been clearly delineated, but they may be influenced by immunological factors, including the CD8-TL avidity (14, 38) and response magnitude (47) as well as the genomic plasticity of epitope sequences (22).

Our results add a layer of intricacy to the already controversial relationship between CD8-TL selection and disease progression. Goulder and colleagues showed that late escape from an HLA-B*27-restricted cytotoxic T-lymphocyte (CTL) epitope correlated with the progression to AIDS (20) in an elegant experiment pointing to an important role for these CTL in maintaining viral control. Other studies have not been able to show consistent relationships between CD8-TL escape and disease progression among individuals expressing HLA-B*5701 (31), another allele that is strongly associated with

long-term nonprogression. Mamu-A*01- and -B*17-positive macaques resemble HLA-B*27- and HLA-B*5701-positive individuals in that many, but not all, animals exhibit an improved outcome and very low viral loads (39). Associations between CD8-TL selection and disease progression were investigated with two Mamu-A*01- and -B*17-positive animals, 2065 and 1937, in this study. Detectable associations should be most pronounced for these animals, as their viral control was likely mediated by strong CD8-TL responses. In both animals, CD8-TL escape occurred within multiple known Mamu-A*01- and Mamu-B*17-restricted epitopes, affecting six epitopes in animal 2065 and eight epitopes in animal 1937. Additional epitopes that were not restricted by Mamu-A*01 or Mamu-B*17 also selected for escape variants in animal 2065 and, presumably, animal 1937. Did any of these escape variants correlate with disease progression? It is impossible to answer this question with certainty, but it seems unlikely. Of the six Mamu-A*01- and Mamu-B*17-restricted epitopes that varied in animal 2065, variation first appeared within two epitopes during the first 16 weeks of infection, within one epitope between weeks 16 and 29, within one epitope between weeks 29 and 52, within one epitope between weeks 51 and 72, and within one epitope between week 72 and the animal's sacrifice at week 75 (data not shown).

The viral loads in the plasma of animal 2065 ranged from 3×10^5 to 3×10^6 copies per ml during the first 48 weeks of infection, while the same four epitopes also accumulated variations in animal 1937, which had viral loads between 7×10^2 and 5×10^4 between weeks 12 and 75 postinfection. One could claim that the control observed in animal 1937 was due to a reduced in vivo replicative fitness of viruses with escape mutations in these four Mamu-A*01 and Mamu-B*17 epitopes, but this argument is contradicted by the observation that the same epitopes accumulated variations in animal 2065. Although we could not correlate escape from any single epitope with a loss of viral control, it remains possible that the aggregate effects of escape from multiple responses on viral fitness and immune control exert a strong influence on disease progression. It is also possible that the extent of variation within CD8-TL epitopes influences recognition and disease progression, as previously hypothesized (39). Viruses from both animal 1937 and animal 2065 contained single amino acid substitutions in the Tat₂₈₋₃₅SL8 epitope, while viruses from 13 of the other 19 Mamu-A*01-positive animals contained multiple substitutions in this epitope. The differences in variation patterns were not as pronounced for other epitopes. We are currently investigating these effects by passaging the late-stage virus from animal 1937 into MHC-disparate hosts and measuring the stabilities of the epitope mutations in these new hosts.

One limitation of the present analysis compared to studies of CD8-TL selection and disease progression in HIV-infected humans is the lack of high-resolution MHC class I genotyping for Indian rhesus macaques. Most known SIV CD8-TL epitopes are bound by Mamu-A*01, -A*02, -A*11, and -B*17, and PCR-allele-specific genotyping tests are available for all of these alleles. These alleles, however, represent only a small fraction of the alleles that are present in Indian rhesus macaques. Meaningful relationships between CD8-TL escape and SIV disease survivorship might become apparent only when large numbers of animals are examined with complete MHC

class I genotyping. This limitation was evident when we compared our animal cohort, containing 35 animals with MHC genotypes for three alleles, to studies of HIV-infected humans. For example, Moore et al. found that the extent of HLA-associated selection in an individual is a better predictor of the pretreatment HIV viral load than HLA genotypes alone, but this required an analysis of 473 HLA-typed individuals who were monitored for >2,200 patient years (32).

A hallmark of the HIV pandemic is its enormous global sequence diversity. Conventional wisdom suggests that much of the diversity is found within the viral envelope, in which the genetic distance among viruses can exceed 35% (19). Within *env*, diversity is concentrated in hypervariable regions that are frequently targeted by antibodies (10, 27, 53). According to this study, CD8-TL escape is the dominant cause of virus evolution outside of *env*, and CD8-TL epitopes accumulate amino acid variations as rapidly as the *env* hypervariable loops. The ubiquity and frequency of CD8-TL escape provide compelling evidence that CD8-TL responses play a central role in shaping AIDS virus diversity.

ACKNOWLEDGMENTS

We thank the veterinary staff of the Wisconsin Primate Research Center for animal husbandry, infection, and clinical core support; Jacque Mitchen for coordinating the animal experiments and necropsies; and Thomas Friedrich for helpful discussions.

This work was supported by NIH grants RO1-AI-46366, RO1-AI-49120, and RO1-AI-52056 to D.I.W. and P51 RR000167 to the WPRC as well as by NIH grant GM043940 to A.L.H.

REFERENCES

- Alexander, L., L. Denekamp, S. Czajak, and R. C. Desrosiers. 2001. Suboptimal nucleotides in the infectious, pathogenic simian immunodeficiency virus clone SIVmac239. *J. Virol.* 75:4019–4022.
- Allen, T. M., P. Jing, B. Calore, H. Horton, D. H. O'Connor, T. Hanke, M. Piekarczyk, R. Ruddersdorf, B. R. Mothe, C. Emerson, N. Wilson, J. D. Lifson, I. M. Belyakov, J. A. Berzofsky, C. Wang, D. B. Allison, D. C. Montefiori, R. C. Desrosiers, S. Wolinsky, K. J. Kunstman, J. D. Altman, A. Sette, A. J. McMichael, and D. I. Watkins. 2002. Effects of cytotoxic T lymphocytes (CTL) directed against a single simian immunodeficiency virus (SIV) Gag CTL epitope on the course of SIVmac239 infection. *J. Virol.* 76:10507–10511.
- Allen, T. M., B. R. Mothe, J. Sidney, P. Jing, J. L. Dzuris, M. E. Liebl, T. U. Vogel, D. H. O'Connor, X. Wang, M. C. Wussow, J. A. Thomson, J. D. Altman, D. I. Watkins, and A. Sette. 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, T. M., D. H. O'Connor, P. Jing, J. L. Dzuris, B. R. Mothe, T. U. Vogel, E. Dunphy, M. E. Liebl, C. Emerson, N. Wilson, K. J. Kunstman, X. Wang, D. B. Allison, A. L. Hughes, R. C. Desrosiers, J. D. Altman, S. M. Wolinsky, A. Sette, and D. I. Watkins. 2000. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* 407:386–390.
- Balter, M. 1998. Modest Briton stirs up storm with views on role of CTLs. *Science* 280:1860–1861.
- Barouch, D. H., J. Kunstman, J. Glowczwskie, K. J. Kunstman, M. A. Egan, F. W. Peyerl, S. Santra, M. J. Kuroda, J. E. Schmitz, K. Beaudry, G. R. Krivulka, M. A. Lifton, D. A. Gorgone, S. M. Wolinsky, and N. L. Letvin. 2003. Viral escape from dominant simian immunodeficiency virus epitope-specific cytotoxic T lymphocytes in DNA-vaccinated rhesus monkeys. *J. Virol.* 77:7367–7375.
- Borrow, P., H. Lewicki, X. Wei, M. S. Horwitz, N. Pfeffer, H. Meyers, J. A. Nelson, J. E. Gairin, B. H. Hahn, M. B. Oldstone, and G. M. Shaw. 1997. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat. Med.* 3:205–211.
- Boyson, J. E., C. Shuffebotham, L. F. Cadavid, J. A. Urvater, L. A. Knapp, A. L. Hughes, and D. I. Watkins. 1996. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *J. Immunol.* 156:4656–4665.
- Burns, D. P., and R. C. Desrosiers. 1991. Selection of genetic variants of

- simian immunodeficiency virus in persistently infected rhesus monkeys. *J. Virol.* **65**:1843–1854.
10. Choi, W. S., C. Collignon, C. Thiriart, D. P. Burns, E. J. Stott, K. A. Kent, and R. C. Desrosiers. 1994. Effects of natural sequence variation on recognition by monoclonal antibodies neutralize simian immunodeficiency virus infectivity. *J. Virol.* **68**:5395–5402.
 11. Chouquet, C., B. Autran, E. Gomard, J. M. Bouley, V. Calvez, C. Katlama, D. Costagliola, and Y. Riviere. 2002. Correlation between breadth of memory HIV-specific cytotoxic T cells, viral load and disease progression in HIV infection. *AIDS* **16**:2399–2407.
 12. Cole, K. S., J. D. Steckbeck, J. L. Rowles, R. C. Desrosiers, and R. C. Montelaro. 2004. Removal of N-linked glycosylation sites in the V1 region of simian immunodeficiency virus gp120 results in redirection of B-cell responses to V3. *J. Virol.* **78**:1525–1539.
 13. Draenert, R., S. Le Gall, K. J. Pfaffert, A. J. Leslie, P. Chetty, C. Brander, E. C. Holmes, S. C. Chang, M. E. Feeney, M. M. Addo, L. Ruiz, D. Ramduth, P. Jeena, M. Altfeld, S. Thomas, Y. Tang, C. L. Verrill, C. Dixon, J. G. Prado, P. Kiepiela, J. Martinez-Picado, B. D. Walker, and P. J. Goulder. 2004. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J. Exp. Med.* **199**:905–915.
 14. Draenert, R., C. L. Verrill, Y. Tang, T. M. Allen, A. G. Wurcel, M. Bocznowski, A. Lechner, A. Y. Kim, T. Suscovich, N. V. Brown, M. M. Addo, and B. D. Walker. 2004. Persistent recognition of autologous virus by high-avidity CD8 T cells in chronic, progressive human immunodeficiency virus type 1 infection. *J. Virol.* **78**:630–641.
 15. Dykhuizen, M., J. Ceman, J. Mitchen, M. Zayas, A. MacDougall, J. Helge-land, E. Rakasz, and C. D. Pauza. 2000. Importance of the CD3 marker for evaluating changes in rhesus macaque CD4/CD8 T-cell ratios. *Cytometry* **40**:69–75.
 16. Evans, D. T., D. H. O'Connor, P. Jing, J. L. Dzuris, J. Sidney, J. da Silva, T. M. Allen, H. Horton, J. E. Venham, R. A. Rudersdorf, T. Vogel, C. D. Pauza, R. E. Bontrop, R. DeMars, A. Sette, A. L. Hughes, and D. I. Watkins. 1999. Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nat. Med.* **5**:1270–1276.
 17. Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* **125**:1–15.
 18. Friedrich, T. C., C. A. Frye, L. J. Yant, D. H. O'Connor, N. A. Kriewaldt, M. Benson, L. Vojnov, E. J. Dodds, C. Cullen, R. Rudersdorf, A. L. Hughes, N. Wilson, and D. I. Watkins. 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.
 19. Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B. H. Hahn, T. Bhattacharya, and B. Korber. 2002. Diversity considerations in HIV-1 vaccine selection. *Science* **296**:2354–2360.
 20. Goulder, P. J., R. E. Phillips, R. A. Colbert, S. McAdam, G. Ogg, M. A. Nowak, P. Giangrande, G. Luzzi, B. Morgan, A. Edwards, A. J. McMichael, and S. Rowland-Jones. 1997. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3**:212–217.
 21. Horton, H., T. U. Vogel, D. K. Carter, K. Vielhuber, D. H. Fuller, T. Shipley, J. T. Fuller, K. J. Kunstman, G. Sutter, D. C. Montefiori, V. Erle, R. C. Desrosiers, N. Wilson, L. J. Picker, S. M. Wolinsky, C. Wang, D. B. Allison, and D. I. Watkins. 2002. Immunization of rhesus macaques with a DNA prime/modified vaccinia virus Ankara boost regimen induces broad simian immunodeficiency virus (SIV)-specific T-cell responses and reduces initial viral replication but does not prevent disease progression following challenge with pathogenic SIVmac239. *J. Virol.* **76**:7187–7202.
 22. Kelleher, A. D., C. Long, E. C. Holmes, R. L. Allen, J. Wilson, C. Conlon, C. Workman, S. Shaunak, K. Olson, P. Goulder, C. Brander, G. Ogg, J. S. Sullivan, W. Dyer, I. Jones, A. J. McMichael, S. Rowland-Jones, and R. E. Phillips. 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.
 23. Khan, I. H., E. T. Sawai, E. Antonio, C. J. Weber, C. P. Mandell, P. Montbriand, and P. A. Luciw. 1998. Role of the SH3-ligand domain of simian immunodeficiency virus Nef in interaction with Nef-associated kinase and simian AIDS in rhesus macaques. *J. Virol.* **72**:5820–5830.
 24. Klennerman, P., U. C. Meier, R. E. Phillips, and A. J. McMichael. 1995. The effects of natural altered peptide ligands on the whole blood cytotoxic T lymphocyte response to human immunodeficiency virus. *Eur. J. Immunol.* **25**:1927–1931.
 25. Knapp, L. A., E. Lehmann, M. S. Piekarczyk, J. A. Urvater, and D. I. Watkins. 1997. A high frequency of Mamu-A*01 in the rhesus macaque detected by polymerase chain reaction with sequence-specific primers and direct sequencing. *Tissue Antigens* **50**:657–661.
 26. Korber, B., C. Brander, B. Haynes, R. Koup, C. Kuiken, J. Moore, B. D. Walker, and D. I. Watkins. 2002. *In* HIV molecular immunology compendium. Los Alamos National Laboratory, Los Alamos, N.Mex.
 27. Lee, C. N., J. Robinson, G. Mazzara, Y. L. Cheng, M. Essex, and T. H. Lee. 1995. Contribution of hypervariable domains to the conformation of a broadly neutralizing glycoprotein 120 epitope. *AIDS Res. Hum. Retrovir.* **11**:777–781.
 28. Li, W. H. 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitution. *J. Mol. Evol.* **36**:96–99.
 - 28a. Lofredo, J. T., J. Sidney, C. Wojewoda, E. Dodds, M. R. Reynolds, G. Napoe, B. R. Mothe, D. H. O'Connor, N. A. Wilson, D. I. Watkins, and A. Sette. 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.
 29. McDermott, A. B., J. Mitchen, S. Piaskowski, I. De Souza, L. J. Yant, J. Stephany, J. Furlott, and D. I. Watkins. 2004. Repeated low-dose mucosal simian immunodeficiency virus SIVmac239 challenge results in the same viral and immunological kinetics as high-dose challenge: a model for the evaluation of vaccine efficacy in nonhuman primates. *J. Virol.* **78**:3140–3144.
 30. Mellors, J. W., C. R. Rinaldo, Jr., P. Gupta, R. M. White, J. A. Todd, and L. A. Kingsley. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**:1167–1170.
 31. Migueles, S. A., A. C. Laborico, H. Imamichi, W. L. Shupert, C. Royce, M. McLaughlin, L. Ehler, J. Metcalf, S. Liu, C. W. Hallahan, and M. Connors. 2003. The differential ability of HLA B*5701⁺ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences. *J. Virol.* **77**:6889–6898.
 32. Moore, C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. A. Mallal. 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**:1439–1443.
 33. Mothe, B. R., H. Horton, D. K. Carter, T. M. Allen, M. E. Liebl, P. Skinner, T. U. Vogel, S. Fuenger, K. Vielhuber, W. Rehauer, N. Wilson, G. Franchini, J. D. Altman, A. Haase, L. J. Picker, D. B. Allison, and D. I. Watkins. 2002. Dominance of CD8 responses specific for epitopes bound by a single major histocompatibility complex class I molecule during the acute phase of viral infection. *J. Virol.* **76**:875–884.
 34. Mothe, B. R., J. Sidney, J. L. Dzuris, M. E. Liebl, S. Fuenger, D. I. Watkins, and A. Sette. 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.
 35. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
 36. Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, N.Y.
 37. Nietfield, W., M. Bauer, M. Fevrier, R. Maier, B. Holzwarth, R. Frank, B. Maier, Y. Riviere, and A. Meyerhans. 1995. Sequence constraints and recognition by CTL of an HLA-B27-restricted HIV-1 gag epitope. *J. Immunol.* **154**:2189–2197.
 38. O'Connor, D. H., T. M. Allen, T. U. Vogel, P. Jing, I. P. DeSouza, E. Dodds, E. J. Dunphy, C. Melsaether, B. Mothe, H. Yamamoto, H. Horton, N. Wilson, A. L. Hughes, and D. I. Watkins. 2002. Acute phase cytotoxic T lymphocyte escape is a hallmark of simian immunodeficiency virus infection. *Nat. Med.* **8**:493–499.
 39. O'Connor, D. H., B. R. Mothe, J. T. Weinfurter, S. Fuenger, W. M. Rehauer, P. Jing, R. R. Rudersdorf, M. E. Liebl, K. Krebs, J. Vasquez, E. Dodds, J. Lofredo, S. Martin, A. B. McDermott, T. M. Allen, C. Wang, G. G. Dosiadis, D. C. Montefiori, A. Hughes, D. R. Burton, D. B. Allison, S. M. Wolinsky, R. Bontrop, L. J. Picker, and D. I. Watkins. 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.
 40. Overbaugh, J., and C. R. Bangham. 2001. Selection forces and constraints on retroviral sequence variation. *Science* **292**:1106–1109.
 41. Petry, H., K. Pekrun, G. Hunsmann, E. Jurkiewicz, and W. Luke. 2000. Naturally occurring V1-Env region variants mediate simian immunodeficiency virus SIVmac escape from high-titer neutralizing antibodies induced by a protective subunit vaccine. *J. Virol.* **74**:11145–11152.
 42. Peyerl, F. W., D. H. Barouch, W. W. Yeh, H. S. Bazick, J. Kunstman, K. J. Kunstman, S. M. Wolinsky, and N. L. Letvin. 2003. Simian-human immunodeficiency virus escape from cytotoxic T-lymphocyte recognition at a structurally constrained epitope. *J. Virol.* **77**:12572–12578.
 43. Phillips, R. E., S. Rowland-Jones, D. F. Nixon, F. M. Gotch, J. P. Edwards, A. O. Ogunlesi, J. G. Elvin, J. A. Rothbard, C. R. Bangham, C. R. Rizza, et al. 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**:453–459.
 44. Price, D. A., P. J. Goulder, P. Klennerman, A. K. Sewell, P. J. Easterbrook, M. Troop, C. R. Bangham, and R. E. Phillips. 1997. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc. Natl. Acad. Sci. USA* **94**:1890–1895.
 45. Quinones-Kochs, M. I., L. Buonocore, and J. K. Rose. 2002. Role of N-linked glycans in a human immunodeficiency virus envelope glycoprotein: effects on protein function and the neutralizing antibody response. *J. Virol.* **76**:4199–4211.
 46. Safrit, J. T., A. Y. Lee, C. A. Andrews, and R. A. Koup. 1994. A region of the

- third variable loop of HIV-1 gp120 is recognized by HLA-B7-restricted CTLs from two acute seroconversion patients. *J. Immunol.* **153**:3822–3830.
47. Vogel, T. U., T. C. Friedrich, D. H. O'Connor, W. Rehrauer, E. J. Dodds, H. Hickman, W. Hildebrand, J. Sidney, A. Sette, A. Hughes, H. Horton, K. Vielhuber, R. Rudersdorf, I. P. De Souza, M. R. Reynolds, T. M. Allen, N. Wilson, and D. I. Watkins. 2002. Escape in one of two cytotoxic T-lymphocyte epitopes bound by a high-frequency major histocompatibility complex class I molecule, Mamu-A*02: a paradigm for virus evolution and persistence? *J. Virol.* **76**: 11623–11636.
 48. Wei, X., J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong, and G. M. Shaw. 2003. Antibody neutralization and escape by HIV-1. *Nature* **422**:307–312.
 49. Wyatt, R., P. D. Kwong, E. Desjardins, R. W. Sweet, J. Robinson, W. A. Hendrickson, and J. G. Sodroski. 1998. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**:705–711.
 50. Yang, Z., and R. Nielsen. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**: 32–43.
 51. Yokomaku, Y., H. Miura, H. Tomiyama, A. Kawana-Tachikawa, M. Takiguchi, A. Kojima, Y. Nagai, A. Iwamoto, Z. Matsuda, and K. Ariyoshi. 2004. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. *J. Virol.* **78**:1324–1332.
 52. Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detours, and B. T. Korber. 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol.* **76**:8757–8768.
 53. Zwick, M. B., R. Kelleher, R. Jensen, A. F. Labrijn, M. Wang, G. V. Quinnan, Jr., P. W. Parren, and D. R. Burton. 2003. A novel human antibody against human immunodeficiency virus type 1 gp120 is V1, V2, and V3 loop dependent and helps delimit the epitope of the broadly neutralizing antibody immunoglobulin G1 b12. *J. Virol.* **77**:6965–6978.