LA-UR- 09-00148

Approved for public release;
distribution is unlimited.

Title:	Expanded Breadth of the T-Cell Response to Mosaic HIV-1 Envelope DNA Vaccination.
Author(s):	B. Korber, Z# 108817, T-6/T-Division W. Fisher, Z# 103477, T-6/T-Division T. Wallstrom, Z# 110720, T-4/T-Division
Intended for:	Journal: Journal of Virology

Form 836 (7/06)

Expanded Breadth of the T-cell Response to Mosaic HIV-1 Envelope DNA Vaccination

Running Title: Increased T-cell epitope recognition for AIDS vaccines

Wing Pui Kong[1], Lan Wu[1], Timothy C. Wallstrom[2], Will Fischer[2], Zhi-Yong Yang[1], Sung-Youl Ko[1], Norman Letvin[3], Barton F. Haynes[4], Beatrice Hahn[5], Bette Korber[2,6] and Gary J. Nabel[1]*

[1]Vaccine Research Center
NIAID, National Institutes of Health
Bldg. 40, Room 4502, MSC-3005
40 Convent Drive
Bethesda, MD 20892-3005

[2]Los Alamos National Laboratory
MS K710, T-10
Los Alamos, NM 87545

[3]Division of Viral Pathogenesis
Department of Medicine
Beth Israel Deaconess Medical Center
Harvard Medical School
RE113, P.O. Box 15732
Boston, MA 02115

[4]Duke Human Vaccine Institute
Duke University Medical Center
MSRBII Building, Room 4085, P.O. Box 103020
Research Drive
Durham, NC 27710

[5]Department of Medicine
University of Alabama at Birmingham
Kaul 816
720 20th Street South
Birmingham, AL 35294

[6] Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501

*Corresponding author. Tel: 301-496-1852, Fax: 301-480-0274, E-mail: gnabel@nih.gov.

Word counts: Abstract: 222; Text: 7,290.

ABSTRACT

An effective AIDS vaccine must control highly diverse circulating strains of HIV-1. Among HIV-1 gene products, the envelope (Env) protein contains variable as well as conserved regions. In this report, an informatic approach to the design of T-cell vaccines directed to HIV-1 Env M group global sequences was tested. Synthetic Env antigens were designed to express mosaics that maximize the inclusion of common potential T-cell epitope (PTE) 9-mers and minimize the inclusion of rare epitopes likely to elicit strain-specific responses. DNA vaccines were evaluated using intracellular cytokine staining (ICS) in inbred mice with a standardized panel of highly conserved 15-mer PTE peptides. 1, 2 and 3 mosaic sets were developed that increased theoretical epitope coverage. The breadth and magnitude of T-cell immunity stimulated by these vaccines were compared to natural strain Env's; additional comparisons were performed on mutant Env's, including gp160 or gp145 with or without V regions and gp41 deletions. Among them, the 2 or 3 mosaic Env sets elicited the optimal CD4 and CD8 responses. These responses were most evident in CD8 T cells; the 3 mosaic set elicited responses to an average of 8 peptide pools compared to 2 pools for a set of 3 natural Env's. Synthetic mosaic HIV-1 antigens can therefore induce T-cell responses with expanded breadth and may facilitate the development of effective T-cell-based HIV-1 vaccines.

INTRODUCTION

The development of AIDS vaccines has been advanced recently by demonstrations of increased survival and decreased viral load following vaccination with T-cell vaccines in non-human primate models (12,19,23,26,31,37). Although such vaccine studies have implied that T cells may contribute to the control of viremia in the highly lethal SIVmac251 challenge model, the applicability of these results in human studies remains uncertain. The major concern regarding the efficacy of HIV vaccines in humans is the extraordinary genetic diversity of the virus. The sequence similarity of HIV-1 envelope from diverse isolates within a clade can diverge as much as 15%, and between alternative clades may approach 30% (10). In addition, the diversity of the viral Gag gene product can approach similar levels, particularly in p17 and p15, which are much more diverse than p24 (6), although Gag does not have the extreme localized diversity seen in the highly variable regions of Env (6,10). While the approach to viral diversity has been addressed in existing vaccines through the use of envelopes derived from representative viruses in the major clades, increasing knowledge about the genetic diversity of naturally occurring isolates has enabled alternative approaches that enhance population coverage of vaccine-elicited T-cell responses.

Approaches under consideration include the use of central gene sequences based on ancestral, consensus, or center of the tree genetic analyses (5,10,18,31,36). Such prototypes are derived by selection of the most common amino acids at each residue (10,16,17,21,25,36), identifying the most recent common ancestor of diverging viruses in a vaccine target population (5,10,18,36), or modeling the sequence at the center of the phylogenetic tree (29), respectively. Peptides based on any of these three centralized

protein strategies enhanced the detection of T-cell responses in natural infection relative to the use of peptides based on natural strains; however, all three strategies behaved equivalently (7).

The use of a single M group consensus/ancestral Env sequence has been shown to elicit T-cell responses with greater breadth of cross-reactivity than single natural strains in animal models (31,36). Such central sequences do not exist in nature, and even phylogenetic ancestral reconstructions are just an approximate model of an ancestral state of the virus (8). Thus, central sequence strategies have provided evidence that various informatically-derived gene products can elicit immune responses to T-cell epitopes found in diverse circulating strains, leading to the possibility of using computational strategies to design polyvalent vaccines which optimize T-cell coverage (6,24). In this study, we have evaluated for the first time the ability of non-natural mosaic Env immunogens (6) to elicit T-cell responses of increased cross-reactivity against epitopes represented in naturally circulating viruses in animals.

Mosaic HIV-1 envelope genes were derived using an informatic approach, whereby *in silico* generated recombinants of natural variants from the Los Alamos database M group Env alignment were created, scored, and selected in combination to optimize the coverage of 9-mers in the global database for a given vaccine cocktail size. While mosaic proteins are artificial constructs that do not occur in nature, they align well to natural proteins, and any short span found in mosaics will tend to be found repeatedly among natural strains (although some of the hypervariable loop regions of Env are so extremely variable that they are not repeated among circulating strains, and necessitate bridging these regions with segments found in a single strain). *In silico* recombination

breakpoints are constrained to create fusion points found in natural sequences. It is possible to provide increased breadth of coverage with a single mosaic, providing the maximum possible single-antigen diversity coverage for stretches of 9 amino acids. Alternatively, multiple mosaics can increase the breadth of representation but have the drawback of requiring the synthesis of additional vectors for clinical use. Mosaics also preserve a natural Env-like sequence to retain normal antigen processing. Here, we have compared single, double or triple mosaic envelope antigen sets to naturally circulating strains or other derivatives for their ability to elicit immune responses of increased breadth. The data suggest that mosaic HIV-1 envelope sequences provide an approach that may be useful in the development of HIV vaccines that respond to T-cell epitopes represented in naturally circulating strains.

MATERIALS AND METHODS

Immunogens

Plasmid DNAs containing five different modifications of the canonical Env gene were generated for each of the antigen sets used in this study: full length Env proteins ("gp160"), and four variants with short deletions. Deletion variants were as follows: 1) the full length Env protein with variable loops 1, 2, 4 and 5 deleted, Env gp160ΔVs; 2) the full length Env protein with deletion of the fusion domain, the cleavage domains, and shortened interspace between heptad 1 (H1) and heptad 2 (H2) (4), Env gp160ΔCFI; 3) the full length Env protein with deletion of the fusion domain, the cleavage domains, shortened interspace between heptad 1 (H1) and heptad 2 (H2), and also with variable loops 1, 2, 4 and 5 deleted, Env gp160ΔCFIΔVs; and 4) the Env protein without cytoplasmic domain, with deletion of the fusion domain, the cleavage domains, shortened interspace between heptad 1 (H1) and heptad 2 (H2), and also with variable loops 1, 2, 4 and 5 deleted (Env gp145ΔCFIΔVs F1, F2, F3, F4). All modified HIV Env genes were synthesized using human-preferred codons (GeneArt, Regensburg, Germany) (15) or by preparation of oligonucleotides of 75 bp overlapping by 25 or of 60 bp overlapping by 20 and assembled by *Pwo* (Boehringer Mannheim) and Turbo *Pfu* (Stratagene) as described previously (4,14). All deletions or other modifications were generated by site-directed mutagenesis using a QuickChange kit (Stratagene, La Jolla, CA). The cDNAs were cloned into a plasmid expression vector, pCMV/R, which mediates high level expression and immunogenicity *in vivo* (2,38).

Mosaic proteins were designed using the methods described in Fischer et al. (6); a web-based suite of tools is now available that enables generation of candidate mosaic

sequences for any set of variable pathogen proteins, and epitope length sequence coverage comparison of different vaccine antigen candidates (34). Mosaics are optimized as a set for a particular size of cocktail, so were designed separately for the 1, 2, and 3 antigen combinations (*i.e.*, the single mosaic is not found in the 2 or 3 mosaic set). The input data was an unaligned version of the full Env M group alignment from the Los Alamos National Laboratory HIV database, as of July 2006 (restricted to include a single sequence per person). Sequences were generated as recombinants of that set, and optimized for 9-mer coverage of that set. Unnatural breakpoints were excluded. We also selected the 3 natural sequences that in combination provided the optimal 9-mer coverage of that same data set, either with or after exclusion of the V-loops, using the same software suite (9,34) (http://www.hiv.lanl.gov/content/sequence/MOSAIC). The length of 9 amino acids was selected for the optimization criteria because it is the most common length of optimal CD8 epitopes; nearby lengths (8, 10, 11, 12...) also get greatly enhanced coverage through the process of optimizing on 9-mers (data not shown). The full length Env protein amino acid sequences of the three sets of mosaics, gp160 and various mutants, are shown (Fig. S1).

Vaccine antigen comparisons

The basic antigen design strategies included the following sets: 3 natural strains that have been previously studied as a polyvalent vaccine in the modified form gp145ΔCFI, each from different clades (Env ABC); 1, 2, or 3 gp160 mosaics (mos.1, mos.2, mos.3); 3 natural strains selected to in combination provide optimal M group coverage of gp160 9-mers (nat.3) (CRF01AE, FIN92168 AF219267; clade B, QH0908 AF277072; and clade C, 93IN101 AB023804; listed as clade, sequence name and accession number), or to provide optimal M group coverage if the V regions were excluded (natΔV.3): clade C, 99BW46424 AF443084; clade B, QH0908 AF277072; clade A, KNH1088 AF457063. These baseline sets were further modified to enable

direct comparisons of T-cell responses of the full gp160 proteins to previously studied envelope modifications; thus, gp160 responses for a given antigen set were compared to gp145ΔCFI and gp160ΔCFI modifications. A negative control (negc) consisting only of the CMV/R vector was included.

Splenocytes from immunized mice were analyzed by intracellular cytokine staining (ICS) for TNF-α and IFN-γ T-cell responses against the approximately 100 different peptide pools described below. Responses from $CD4^+$ and $CD8^+$ T lymphocytes were measured separately. The data were analyzed for the magnitude of overall response (strength) and the number of positive responses (breadth). Both TNF-α and IFN-γ responses were measured and analyzed. IL-2 responses were also measured, but there was very little signal and the measurements were dominated by noise, and so these were not included in further analysis. Between 7 and 10 *vac*cine antigen/protein *mod*ifications (vac/mod) configurations were tested on 12 separate days. The 12 sets of experiments were grouped into six pairs; in each pair, the same set of antigen configurations was tested. The magnitude of the overall responses varied by a factor of up to about six on different days, and this effect was corrected through statistical methods, as described below. Not all configurations of vaccine plus modification were tested, and the number of times a particular configuration was repeat tested ranged from 2 to 12. In Table S1, we indicate the number of microtiter plates measured for each vector.

Vaccination

Six- to eight-week-old B6D2F1/J (H2 Haplotype b/d) female mice (Jackson Laboratory, Bar Harbor, ME) were used for these immunogenicity studies. Mice (ten per group) were immunized with a total of 15 μg of DNA (100 μL in PBS), 4x at two-week

intervals. Immunizations were administered bilaterally into the muscle of the hind leg using needle and syringe. The groups included pCMV/R with no insert (Control); vaccines containing 15 μg of 1 mosaic plasmid DNA; 7.5 μg of each plasmid in the two-plasmid groups, e.g. a combination of 2 mosaics; or 5 μg of each plasmid in the three-plasmid groups, e.g. 3 mosaics, Natural-Strains(Set1), Natural-Strains(Set2), and Trivalent. All animal experiments were reviewed and approved by the Animal Care and Use Committee, Vaccine Research Center (VRC), National Institute of Allergy and Infectious Diseases (http://www.niaid.nih.gov/vrc) and performed in accordance with all relevant federal and National Institutes of Health guidelines and regulations.

Peptides for ICS stimulation

In this study, we used 492 Env peptides for intracellular cytokine staining (ICS) stimulation. For this ICS analysis, 15 mer PTE peptides (20) were used to evaluate the vaccines as the common standardized panel of HIV-1 peptides for T-cell-based vaccines. The 492 Env peptide sequence set was designed to permit expression of the potential T-cell epitopes (PTE) found most frequently in the sequences of circulating worldwide HIV-1 strains, based on 549 full-length HIV-1 genome sequences obtained from the Los Alamos National Laboratory (LANL) HIV sequence database as of February 2005. All synthesized peptides (New England Peptide, Gardner, MA) are 15 a.a. in length with naturally occurring 9 a.a. sequences that are potential T-cell determinants captured in an unbiased manner (20,22). Briefly, frequencies were computed for all 9-amino-acid subsequences in the dataset, and then 15-mer peptides were selected in order of the summed frequencies of the previously unincluded 9-mers they contained, using a forward stepwise algorithm. This algorithm selects first for the highly conserved PTEs, then the

less conserved. 492 PTE peptides were generated (for a coverage threshold of 15% (20)), and grouped into 78 pools of 6 to 12 PTE 15-mer peptides such that the peptides that carried the highest frequency 9-mers were grouped in the first pool, continuing so that the peptides with the rarest 9-mers were in the 78^{th} pool. All but the pools representing the rarest potential epitopes contained 6 peptides each; the 3 pools of rarest potential epitopes contained 10-12 peptides. We refer to these sets as PTE pools. Four pools with larger numbers of peptides were also tested, with 114 for the first 3 large pools, 148 for the 4th large pool; we refer to these sets as PTE superpools. Pooled sets of peptides, 15-mers overlapping by 11, corresponding to each of the three Envelopes included in the Env ABC polyvalent vaccine were also used as previously described (2-4,6,14,32).

Cellular Immune Analysis

Two weeks after the last immunization, spleens from 3 mice in each group were harvested aseptically, gently homogenized to form a single-cell suspension, washed, and resuspended to a final concentration of approximately 10^7 cells/ml. All groups of harvested spleen cells (maximum of 10^6 cells/peptide pool) were stimulated for 5 h in the presence of 2 μg of anti-CD28 and anti-CD49d mAbs/ml (BD PharMingen, San Diego, CA), and also with 10 μg/ml brefeldin A (Sigma, St. Louis, MO). Cells were stimulated for 5 hours with a) 15-mers of 6-peptide pools or 12-peptide pools of PTE, as the target testing stimulating agents, b) no stimulation for background control, c) Ebola GP protein as the negative control, and d) phorbol myristate acetate (PMA) with ionomycin as the positive control. Env A, Env B, and Env C pools derived from three candidate genes previously described (4) were included as additional controls. Cells were then washed and stained with Vivid dye (Invitrogen, Carlsbad, CA) to determine their viability. FC

block monoclonal antibodies were added to the cells followed by staining with surface antigens (rat anti-mouse cell surface antigens CD3-PerCP-Cy5.5, CD4-AlexaFluor700 and CD8-APC-Cy7, BD PharMingen, San Diego, CA). The cells were washed again, permeabilized, fixed with Cytofix/Cytoperm and stained with monoclonal antibodies (rat anti-mouse cell surface antigens CD3-PerCP-Cy5.5, CD4-AlexaFluor700 and CD8-APC-Cy7, and rat anti-mouse cytokines IFN-γ-APC, IL-2-PE, and TNF-α-PE-Cy7, BD PharMingen, San Diego, CA) followed by multi-parametric flow cytometry analysis to detect the IFN-γ, IL-2, or TNF-α positive cells in the $CD4^+$ or $CD8^+$ T-cell population. Another 3 mice in each group were subjected to the same analysis two days after the initial test to repeat the analysis.

Flow Cytometry

Stained cells were assayed on the BD™ LSR-II Flow Cytometer using FACSDiva software (BD Biosciences, San Jose, CA). The data were analyzed with FlowJo 8.6.1 software (Tree Star, Ashland, OR).

Statistical Methods

The objective of the analysis is to compare the strength and breadth of different vaccine strategies and Envelope modifications. A statistical model was used that enabled us to control for the variability between assays done on different days (we will call this the "date effect"), and so to assess the contribution of the vaccine strategy (the "vaccine effect") to the outcome. These effects should be independent, since the date effect will depend on the measurement process and the variation between mice, and the vaccine effect will depend on the vaccine that was given. The usual procedure for dealing with such independent effects is to adopt a balanced experimental design, so that the

measurement of a particular vaccine is randomized over the different dates. However, the adoption of such a design was inconsistent with the exploratory manner in which the data was acquired, and in any case, the significance of the date effect was not fully recognized in advance. It turns out that the strength of the different vaccines varies substantially, by a factor of about six, but the date effect is roughly comparable in magnitude, complicating the assessment of both strength and breadth.

While some vaccines were repeated many times, other vaccines were only measured on a few days, often only two. If, for example, the overall response to the vaccine was low, it was not clear whether this was due to the vaccine or to the day on which it was measured. The date effect also complicated the assessment of breadth. If we had used a fixed threshold to assess positivity, as is customary, we would have missed positive responses on days when the overall response was low, and interpreted random noise as a positive response on days when the overall response was high. We tested this approach, and found great variation in the number of positives for the same vac/mod on different days. A routine analysis, not correcting for the date effect, would have led to greatly increased noise in the breadth assessments and would have prevented us from making meaningful comparisons.

To deal with this problem, we adopted a statistical model that enabled us to correct for the date effect. We call the corrected data the "date-corrected" data. Using the date-corrected data, we can compare vaccine strengths directly, and use a common threshold for assessing positivity. Because the date effect is uncertain, the date-corrected data acquires some additional uncertainty, but the results are nevertheless highly significant. Intuitively, what makes this approach possible is that some of the vaccines

were tested on most or all of the dates, and the difference in their responses provides the necessary information about the date effect.

In order to account for the date effect, we modeled the logarithm of the vaccine strength, rather than the strength itself; this converts the multiplicative variation into an additive variation that can then be estimated using a linear model. Accordingly, the following "two-way layout" was adopted:

$$l_{ij} = v_i + d_j + \varepsilon_{ij}, \qquad (1)$$

where l_{ij} is the (natural) logarithm of the strength of the responses for the vac/mod i on day j, v_i and d_j are quantifications of the vaccine and date effects, respectively, in this model, and the ε_{ij} are identical and independently distributed Gaussian random errors, to account for natural mouse-to-mouse variation and other stochastic effects. We describe how we determined l_{ij} below. Note that the log-response is additive in v_i and d_j, which reflects the independence of the date and vaccine effects.

We use the data l_{ij} to make estimates, $\hat{v}_i$ and $\hat{d}_j$, for the vaccine and date effects. The interpretation of these numbers, roughly speaking, is the following: If vaccines 1 and 2 are measured on the same day, then we expect the response to vaccine 1 to be $\exp(\hat{v}_1)/\exp(\hat{v}_2)$ times larger than the response to vaccine 2. Similarly, if the same vaccine is measured on day 1 and day 2, then we expect the response on day 1 to be $\exp(\hat{d}_1)/\exp(\hat{d}_2)$ times larger than the response on day 2. The analysis only gives ratios of the strengths (or differences in the log strengths). Thus, we measure all vaccine strengths relative to the negative control, and all date effects relative to an arbitrarily chosen fiducial date.

The date corrected log-strength is

$$\tilde{l}_{ij} = l_{ij} + \hat{d}_0 - \hat{d}_j ; \qquad (2)$$

this is the log-strength that would have been expected had the data been measured on the fiducial date. The expected difference in the date-corrected log-strengths, for two different vaccines, depends only on the vaccine, not on the day:

$E(\tilde{l}_{ij} - \tilde{l}_{i'j}) = \hat{v}_i - \hat{v}_{i'}$.

The date-corrected responses to individual peptide pools are obtained by multiplying the data on day j by the factor $\exp(\hat{d}_0)/\exp(\hat{d}_j)$, where $\hat{d}_0$ is the fiducial date effect. It is the data we would expect had the data been measured on the fiducial date. (Note that this is a slight approximation, in that the factor should strictly be the expectation of $\exp(d_0)/\exp(d_j)$, but the difference should be small.)

To assess uncertainties in vaccine strength, we calculate the variance of $v_i - v_0$, where v_0 is the vaccine effect for the negative control. These uncertainties are easily determined from the linear model, Eq. (1), using standard methods.

The date effects also depended somewhat on the T-cell type and cytokine, so separate models were fitted to all combinations of CD4 and CD8, and of IFN and TNF. However, for the same T-cell type, cytokine, and date, different parts of the data gave very similar estimates for d_j.

To assess functionality, we computed a matrix whose rows denote particular experiments and whose columns denote small peptide pools. For each element of the matrix, we assigned the number 0, 1, or 2, depending on the number of positive responses for TNF-α and IFN-γ observed for the corresponding experiment and peptide pool. Some

experiments were also performed testing IL-2 responses, but the results were weak and sporadic and thus excluded from further analyses. We then used a standard agglomerative clustering algorithm (35), using Euclidean distances, to cluster the experiments (row vectors) and the peptide pools (column vectors) (http://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap.html, based on the R package heatmap.2). These cluster patterns are shown on the margins of the heatmaps (Fig. 6), which were generated by color-coding responses to indicate those that generated no response (pale yellow), 1 response to either TNF-α or IFN-γ (orange) or responses to both (red). Statistical support for the various clusters is indicated on the dendrogram branch points, based on the approximately unbiased test of multistep-multiscale bootstraps (33).

The data in these experiments came from a total of 352 microtiter plates, each of which measured IFN or TNF responses to $CD4^+$ or $CD8^+$ T cells for a particular vaccine modality (vac/mod) on a particular day. By vac/mod we refer to the DNA vaccine antigen cocktails (including 1, 2 or 3 mosaics; 3 natural strains selected to provide in combination optimal 9-mer coverage; three natural strains, one each from clades A, B and C…) and the Env modifications (including gp160, gp145ΔCFI, gp160ΔCFI, and gp145ΔCFIΔV and gp145ΔCFIΔV, where ΔV refers to removal of the hypervariable loops, and ΔCFI refers to deletions of the cleavage site, fusogenic domain, and spacing of heptad repeats 1 and 2 (4). In some cases, all or part of the data from a given plate was clearly affected by systematic error, as indicated by trends or consistently elevated responses from pools in contiguous regions of the plate. Such plates, of which 17 involved CD8 and two CD4, were left out of the analysis. Thus, a total of 333 plates

were used. Among these plates, there was also a very small fraction of small peptide pools (0.3%) for which data was unavailable. We did not try to estimate the missing data.

RESULTS

Vaccine insert design

Polyvalent mosaic vaccines were designed using a genetic algorithm (6) to assemble *in silico* recombinants of natural Env proteins with breakpoints that do not disrupt the protein and which optimize potential epitope coverage of a diverse population. All stretches of 9 amino acids (or 9-mers) were considered potential epitopes, and the presence of rare 9-mers was minimized. The M group sequence alignment from the 2006 HIV database (www.hiv.lanl.gov) was used as a baseline; M designates the "main" group of HIV-1 sequences that includes all of the standard clades (A-K) and their recombinants, the over-arching set of diverse HIV sequences responsible for the global HIV pandemic.

We performed separate optimizations for 1, 2 or 3 mosaic gp160 DNA vaccine antigen combinations. Each set is comprised of plasmid DNAs encoding distinct proteins, which in combination yield optimal coverage for a given number of antigens. These designs were subsequently modified to parallel Env modifications that have been previously explored as vaccine antigens, including ΔCFI deletions, where ΔCFI refers to deletions of the cleavage site, fusogenic domain, and spacing of heptad repeats 1 and 2 (4), and gp145, to compare the impact of Env modifications relative to intact gp160 on T-cell responses to the various antigens (Fig. 1A). In addition, the three natural strains were selected that in combination optimize coverage of the M group, either including or excluding the hypervariable V1, V2, V4 and V5 regions (ΔVs). The ΔVs were optimized independently, and are distinct sets of proteins. We also did several exploratory immunizations using mosaic constructs that had the hypervariable loops removed; no particular benefit was conferred (data not shown), and as these constructs were not as

extensively tested as the others, we did not include these results in further analysis. The number of mice that received each vaccine and a description of the peptide pools used in the analysis is shown (Table S2).

The rationale for the deletion of the hypervariable regions is that they are often unique, hence would be strain-specific, and responses to these regions might divert the vaccine-induced immune response away from more conserved and potentially cross-reactive regions of Env. The code for both designing mosaics and selecting optimal natural sequences is available on-line: http://www.hiv.lanl.gov/content/sequence/MOSAIC/ (34). A comparison of the basic gp160 vaccine designs to the M group sequences from the 2006 HIV-1 database Env alignment (www.hiv.lanl.gov) is shown, and the impact on numbers of potential epitopes lost by deleting parts of the protein in gp145ΔCFI constructs is indicated (Fig. 1B). We included for comparison and as a positive control a polyvalent vaccine that included one A, one B and one C clade Env gene (Env ABC). These strains were not optimized for 9-mer coverage, and in combination have previously been shown to elicit immune responses (2-4,6,14,32).

Induction of T-cell responses with different gp160 mosaics and mosaic deletion mutants

To evaluate the T-cell responses to different gp160 mosaics, B6D2F1/J mice were immunized with plasmid DNA vaccines encoding gp160 mosaics encoded by 1, 2 or three plasmids, one natural strain or a natural strain with deleted V regions, and compared to a mixture of clade A, B, C gp145s with deletions in the cleavage, furin and interhelical domains (ΔCFI) described previously (4,14). Mice (n=10 per group) were immunized

with a total of 15 μg of DNA four times at two-week intervals. Two weeks after the last immunization, the splenocytes from different groups of 3 immunized mice were isolated, pooled and stimulated by different stimulants as in all experiments in this study: unstimulated, 492 Env peptides PTE in 78 pools (p1 to p78). Additional controls included Env ABC peptide pools, an irrelevant peptide pool control, and non-specific stimulants (data not shown). The CD4 and CD8 cell responses were measured using ICS for IFN-γ and TNF-α (Fig. 2, left and right, solid and open bars respectively). The minimal threshold response indicated by horizontal dashed lines was defined as 2 times the negative control. The CD8 response to the 1 plasmid (mos 1) gp160 mosaic was lower than the response to the mos 2 or mos 3 mosaics (Fig.2, right panel 2 vs. 3,4), with 4 compared to 7 positive pools above background. Some peptide pools were shared in common among the three different mosaic combinations while others were unique.

The responses elicited by the natural strain plasmids, with or without V region deletions, or gp145ΔCFI trivalent plasmids were decreased relative to the mosaics using the PTE peptide pools. For example, the 3 mosaic set induced a CD8 response above 0.25% in 7 peptide pools compared to only 2 pools with 3 natural Env's (Fig. 2, right panel 4 vs. 5), and these responses were higher in magnitude. Similar results were seen in CD4 responses (Fig. 2, left panel), though the wild type natural strain plasmid induced a larger number of CD4 epitope responses. The additional animals in each group were sacrificed and analyzed in the same way, and statistical analysis applied to the responses measured from the entire group (below). From the initial observation, it was clear that the gp160 mosaic plasmids elicited detectable T-cell responses against PTE peptide pools after immunization.

Immunogenicity of gp160ΔCFI, gp160ΔV, gp160ΔVΔCFI, and gp145ΔVΔCFI mosaics

We next tested alternative mosaic plasmids, with or without the variable region, ΔCFI, or both deletions. These vectors were in turn compared to the multiclade Env gp145ΔCFI vectors shown previously to improve the breadth of the Env T cell response compared to single strain Env immunogens (14,32). These comparisons were done as described above (Figs. S2-5). The resultant immune responses for the complete groups of ten animals were then analyzed to determine their comparative magnitude and breadth of response. Breadth was defined here as the number of responses to different PTE peptide pools. Comparing the number of responses to PTE peptides, it was evident that the responses in these groups immunized with deleted mosaics was lower than with the wild type gp160 mosaic, particularly for CD8 responses, although this result may in part reflect mismatches between the immunogen and assay peptides. Nonetheless, we chose to focus on the more immunogenic inserts, 2 and 3 set gp160 mosaics, according to this criteria.

Strength of vaccine-induced responses to Env A, B, C, PTE and PTE grouped peptide pools

The analysis of immune responses of diverse immunogen combinations to many peptides were limited technically by the number of animals that could be analyzed on the same day. In order to compare groups analyzed with different immunogens on different days, internal references were used, and we evaluated this data further with a second approach using formal statistical analysis to compare relative vaccine potency. Specifically, we looked at the breadth (number of epitopes) and magnitude of cellular

immune responses, and attempted to control for inter assay variability on samples analyzed on different days. To characterize the ability of different immunogens to elicit a vaccine response, we determined the vaccine effect, the magnitude and breadth of CD4 and CD8 responses, by comparing the IFN-γ and TNF-α ICS responses to the different peptide pools: Env ABC peptide pools, the PTE pools, and the PTE superpools (Figs. 3A, B). Six immunogens were compared; sets of mos 1, 2, or 3 mosaics, and the pool of three natural strains, with or without V regions, in addition to a negative control and a positive control Env A, B, C vaccine. Reassuringly, the Env ABC vaccine elicited strong responses to the Env ABC peptide pools. Thus the overall vaccine response tested using autologous peptides suggested a robust response to the matched vaccine. The response of CD4 cells to the vaccine was similar in both the small size and superpools, and it was highly significant for both IFN-γ and TNF-α responses ($P < 10^{-4}$). For CD8 responses to the Env ABC immunogens, the IFN-γ and TNF-α ICS responses were highest against B clade pooled peptides. Among the individual A, B, and C pools, the responsiveness of CD8 T-cells was reflected in a rank order, B > A ~ C (paired Wilcoxon B > A, p = 0.0005; B > C, p = 0.003). CD4 T-cell responses were more consistent across all three clade pools, with B ~ A > C, paired Wilcoxon B > C p = 0.009, A > C p = 0.052 (data not shown).

Magnitude of vaccine-induced immunity to PTE pools and superpools

The strength of responses to different vaccines was also assessed by measuring the magnitude of T-cell responses to subgroups of the PTE peptide pools, each containing 10 peptides. Because peptide pools were used, a strong response might reflect either an intense response to a single peptide in the set, or a sum of moderate or low responses to

several distinct peptides in the set. For CD8 cells, the strength of the observed responses to the mosaics was much greater than that of the naturals or the positive control (Fig. 3). Among the mosaics, the 2 and 3 mosaic set responses were stronger than the single mosaic. The overall strength of the response to PTE pools was much greater than the response to the Env A, B, and C pools. The gp145ΔCFI immunogens induced a consistently weak response, which may have resulted from a loss of epitopes due to the deletions in the open reading frame (see Fig. 1B) which then caused mismatch between the immunogens with regard to the number of peptides made on the basis of the wild type sequence. The response profile for CD4 is quite different from that of CD8 (Fig. 3). Here, the strength of the responses to mosaics and the naturals was roughly comparable, and there was less variation in the response of the different vaccines than in CD8 cells. The data suggests that the ΔCFI modification may reduce the response, possibly because of less matching to the ICS assay peptides, but the 145ΔCFI modification, which suppressed the CD8 response, does not do the same for the CD4 response. Finally, vaccine-induced immunity was determined by ICS with PTE superpools. Response strength to the different vaccines was also assessed by measuring T-cell responses to the 4 peptide pools including large numbers of peptides (Fig. 3). The strength estimates for the large peptide pools are very similar to those obtained with the small peptide pools, despite the fact that they were obtained from separate experiments.

Breadth of Responses

Vaccine breadth was assessed using date-corrected (normalized) responses to the 78 small peptide PTE pools. Data for identical vaccines or vaccine variants on particular dates were pooled, and responses were deemed positive if the median response exceeded

a fixed threshold as described in the statistical methods. The ICS IFN-γ responses in CD8 cells were compared among six different vaccines.

The breadth of the mosaics responses was dramatically greater than that of the naturals, including both Env ABC and the three naturals selected to optimize epitope coverage (Fig. 4A). The mosaics all show clear spikes for numerous peptide pools. The number of such pools increases from 4 to 10 in the monovalent vs. the trivalent mosaics. The best natural set of Envs shows only two weak spikes, at pools 7 and 13. The results for the TNF-α ICS responses in CD8 cells were very similar, both in magnitude of the response and in the specific peptide pools that test positive (Fig. 4B).

Analogous plots for CD4 IFN-γ and TNF-α ICS revealed similar patterns of responsiveness (Fig. 5A, B). The mosaics still generated more positive responses than the naturals, although the difference was less striking. For example, the 2 mosaic and 3 mosaic sets generated a vigorous IFN-γ response to 15 and 14 pools, respectively, but three natural proteins still do quite well (10 pools each for Env ABC, Natural ΔV, and Natural (N1/N2/N3).

Although the correction for inter-assay variation improved the experimental consistency between experiments, the identification of positive responses was not sensitive to uncertainty in the strength correction. In fact, the greater breadth of CD8 response in the mosaics was clear, even if no strength correction is used at all. With regard to the threshold, changes in the threshold do affect the number of positive responses, and these changes can affect different vaccines differently, depending on how many positives are near the threshold. The CD8 breadth comparisons are very robust to changes in the threshold; the CD4 comparisons are not as robust. Inasmuch as the

breadth estimate depends on a choice of threshold, and there is uncertainty as to the correct value, the evidence for increased CD4 breadth in the mosaics should be viewed as suggestive but not conclusive.

Concordant vs. discordant synthesis of IFN-γ and TNF-α

To evaluate the qualitative nature of the T-cell response, we examined whether these cells synthesized either IFN-γ or TNF-α alone, or both, in response to vaccination, the latter a surrogate for a multifunctional cytokine production. For this analysis, the responses were evaluated using a heatmap representation of functional CD8 responses to IFN-γ and TNF-α, with a threshold of 0.1 (Fig. 6A). No response, a response to either IFN-γ or TNF-α, or a response to both IFN-γ and TNF-α were represented by increasing "heat" (pale yellow, orange, red represent negative, monofunctional, and bifunctional responses, respectively). Vaccines that had similar behavior in terms of responses to particular peptide pools are clustered together by row, and peptides pools that had similar patterns of eliciting response are clustered together in columns. All gp160 2 and 3 set mosaics are contained in a single large cluster; i.e., they displayed consistent patterns of frequent bifunctional responses to particular peptide pools. The gp160 mos 1 mosaic responses occupied a single cluster with intermediate reactivity. A cluster for the natural gp160ΔCFI set was also evident. All of the mosaics elicited many bifunctional responses; the pattern is less apparent in the naturals, although this may be due to a generally lower overall response. The consistency of the responses to closely related vaccines is apparent. Analysis of the CD4 responses with this heatmap revealed that the mosaics, as well as the naturals, are broadly bifunctional (Fig. 6B). Indeed, the number of bifunctional responses exceeds the number of monofunctional responses. The CD4

responses of the mosaics and the naturals were quite similar, not only in magnitude and breadth, but in the specific peptide pools generating a positive response.

DISCUSSION

In this report, we have analyzed the ability of HIV-1 mosaic genes encoded in single, double or triple plasmids to confer increased reactivity against T-cell epitopes derived from commonly circulating viruses. The peptide designs, either those in the mosaics or in the PTE peptides for immune assays, were based on 9 mers in Env sequences and were independent of MHC, either human or mouse. Initial analyses could therefore be performed in inbred mice where immune function can be assessed with greater consistency. For this analysis, we immunized mice with these plasmids or with previously-defined mutants derived from natural isolates and compared their ability to respond to T-cell epitopes that have been identified in a high proportion of naturally circulating strains. The study was performed in inbred mice where the effects of these vectors could be assessed on the background of restricted MHC haplotypes, allowing comparisons on a constant genetic background. These data may underestimate the ability of such vectors to elicit responses to a wide array of genes, given the increased MHC complexity in humans and non-human primates. At the same time, the possibility remains that Env epitopes have undergone negative selection to escape human MHC recognition, which would act oppositely and possibly reduce immunogenicity in humans.

The data indicate that the mosaic vaccine antigens are able to expand the breadth of T-cell responses to Env PTE. The responses appear to be greatest in animals receiving combinations of two- or three-mosaic inserts that give greater predicted coverage of T-cell epitopes based on informatic analyses. Interestingly, little difference was noted between two or three mosaic Env immunization, possibly because of the restricted MHC

haplotype in mice. Whether this would hold true in nonhuman primates or humans will require additional testing.

The reason for the improved immunogenicity of the Env mosaics compared to the optimal natural Env glycoproteins in terms of breadth and magnitude of responses may relate in part to the peptides used for the ICS analysis. The number of responses to three mosaic proteins was surprisingly enriched over the three natural strains selected to give optimal M group coverage. The best naturals and the mosaics were selected relative to the M group collection of HIV sequences in the HIV database. Comparing the gp160 trivalent antigen designs, the set of M group Envs from the database had 44% of its 9-mers perfectly matched in the 3 mosaic set, while 34% were matched by the 3 optimal natural strains (Fig. 1B). This relatively modest advantage, however, seems unlikely to account for the profound experimental advantage seen for mosaics over natural proteins (Figs. 5 and 6). One possibility is that PTE peptide pools were selected in particular to cover 9-mers that were most common, as were the mosaics, and the relative advantage of the mosaics to the optimal naturals is enhanced when tracking 9-mer identities found among the PTE set: 3 mosaics matched 67% of PTE 9-mers, and the best naturals matched only 44%. The number of identical 9-mers (potential CD8+ T-cell epitopes) shared between a vaccine (including all antigens in a polyvalent vaccination) and a given peptide pool can provide a rough estimate of the potential number of CD8+ T-cell responses to that vaccine that could be detected by ICS using that peptide pool. The 2- and 3-mosaic antigens match more 9-mers than the natural-sequence polyvalent antigens (Fig. S6). In particular, even the 2 mosaic vaccine has a higher number of 9-mer matches (match counts) with PTE peptides than any of the natural-sequence candidates, although

the latter have 3 sequences each. The 3 mosaic vaccine has higher match counts than the natural-sequence vaccines in 70 of 78 individual pools (Fig. S6), and a large overall advantage (Table S3).

A second factor is that, by design, mosaics minimize the inclusion of rare and unique 9-mers. In contrast, natural strains inevitably contain many unusual, type-specific 9-mers. T-cell reactions against rare variant epitopes found in natural strains would go undetected in PTE-based assays, and in fact would be of little use in a vaccinated population. If immunodominant, they could limit the potential to stimulate responses with more useful epitopes with the potential for greater cross-reactivity. For mosaic vaccine antigen designs based on more conserved proteins like Gag (6) we were able to require that every single 9-mer included in a given mosaic antigen set was found at least 3 times in the population. For the Env-based mosaics used here, however, to span hypervariable regions and weave together intact proteins, we were forced to require inclusion of a small number of rare 9-mers, and 3 mosaic gp160 proteins contained 128 9-mers that were found <3 times in the M group alignment used in Fig. 1B. In contrast, the mixture of three natural strains contained 406 9-mers found <3 times in the population. Similarly, the PTE peptides are designed to emphasize the inclusion of the most common potential epitopes, and so have more peptides in common with the mosaic proteins. If any of the vaccines elicited responses specific for epitopes that are not included in the PTE sets, they would be missed. The natural three gp160 cocktail contains 1,072 9-mers that are not found among any of the PTE peptides, while the mos 3 mosaic vaccine has many fewer, 684, so it is possible that the natural vaccine has elicited

a higher frequency of responses that were missed by the assay system, because they are specific for the vaccine strain,

While the approach to diversification of envelope immunogens appears promising, it is clear that this approach will need to be combined ultimately with an informatic approach to optimize the response to other viral gene products, including Gag and possibly Nef. Whether the addition of a relatively conserved gene product such as polymerase can improve this response further will also require testing. It may be argued that the polymerase is a less effective immunogen for protection against HIV-1 or SIV challenge because this gene product is conserved, present in low abundance, and is not likely to represent a viral gene product that is highly selected by immune pressure. At the same time, the variability of the envelope gene and the identification of T-cell escape mutants in nonhuman primates and humans suggest that this gene product, as well as the Gag protein, is recognized and under strong selective pressure by the immune system. The viral Env is under further selection by antibodies, which appear to evolve during the course of infection and drive the generation of new mutants (13,27,28,30).

The data here suggest that the informatic approach to enhancement of viral diversity coverage increases the immune recognition of diverse viral epitopes. Whether this increased recognition can give rise to improved protection remains unknown. There are two ways that a polyvalent mosaic might improve a vaccine: the first is in cross-reactive protection against circulating strains, the second is the potential to block fit immune escape routes. The present study suggests that the mosaic concept should be tested further in nonhuman primate models. In particular, the SIV sooty mangabey virus pool may provide an opportunity to evaluate the efficacy of mosaic immunization in

protecting against infectious challenge. A variety of viruses from diverse clades have been identified and in many cases adapted to growth in Indian rhesus macaques (1,11). By assembling additional viral sequences and generating SIV_{smm} mosaic sequences and relevant challenge strains, it should be possible to test whether this approach can show efficacy in the context of viral infection and merit further evaluation in human efficacy studies.

The recent results of the STEP trial have suggested that immunization with recombinant adenovirus (rAd) 5 vectors shows no efficacy in humans and may have the potential to increase HIV acquisition in Ad5 seropositive individuals. While the basis for this lack of efficacy is unknown, it is clear that the induction of a T-cell immune response remains a desirable goal for HIV vaccines. Even in the event that broadly neutralizing antibody immunogens could be elicited, the likelihood that a number of circulating viruses would remain resistant to neutralization or the evolution of variants that would escape antibody neutralization would suggest that stimulatory T-cell responses could help to control infection and/or contain viremia. It is therefore important to recognize that the approach to the development of a T-cell-based vaccine that would address diverse relevant strains of circulating virus remains distinct from the approach to the elicitation of neutralizing antibodies. It is unlikely that the informatic approach derived here to address T-cell diversity would resolve the issues related to broadly neutralizing antibody immunogens but a combination of this approach and rationally designed broadly neutralizing antibody immunogens may improve the likelihood of containment or prevention of natural HIV infection.

ACKNOWLEDGMENTS

We thank Ati Tislerics for assistance with manuscript preparation, Brenda Hartman for figure preparation, and members of the Nabel lab for helpful discussions and advice. This work was supported in part by the Intramural Research Program of the National Institutes of Health, Vaccine Research Center, National Institute of Allergy and Infectious Disease and Los Alamos National Laboratory directed research funding.

REFERENCES

1. **Apetrei, C., A. Kaur, N. W. Lerche, M. Metzger, I. Pandrea, J. Hardcastle, S. Falkenstein, R. Bohm, J. Koehler, V. Traina-Dorge, T. Williams, S. Staprans, G. Plauche, R. S. Veazey, H. McClure, A. A. Lackner, B. Gormus, D. L. Robertson, and P. A. Marx**. 2005. Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. J Virol. **79**:8991-9005.

2. **Barouch, D. H., Z. Y. Yang, W. P. Kong, B. Korioth-Schmitz, S. M. Sumida, D. M. Truitt, M. G. Kishko, J. C. Arthur, A. Miura, J. R. Mascola, N. L. Letvin, and G. J. Nabel**. 2005. A human T-cell leukemia virus type 1 regulatory element enhances the immunogenicity of human immunodeficiency virus type 1 DNA vaccines in mice and nonhuman primates. J. Virol. **79**:8828-8834.

3. **Catanzaro, A. T., M. Roederer, R. A. Koup, R. T. Bailer, M. E. Enama, M. C. Nason, J. E. Martin, S. Rucker, C. A. Andrews, P. L. Gomez, J. R. Mascola, G. J. Nabel, and B. S. Graham**. 2007. Phase I clinical evaluation of a six-plasmid multiclade HIV-1 DNA candidate vaccine. Vaccine **25**:4085-4092.

4. **Chakrabarti, B. K., W. P. Kong, B.-Y. Wu, Z.-Y. Yang, J. Friborg, Jr., X. Ling, S. R. King, D. C. Montefiori, and G. J. Nabel**. 2002. Modifications of the Human Immunodeficiency Virus envelope glycoprotein enhance immunogenicity for genetic immunization. J. Virol. **76**:5357-5368.

5. **Doria-Rose, N. A., G. H. Learn, A. G. Rodrigo, D. C. Nickle, F. Li, M. Mahalanabis, M. T. Hensel, S. McLaughlin, P. F. Edmonson, D. Montefiori, S. W. Barnett, N. L. Haigwood, and J. I. Mullins**. 2005. Human immunodeficiency virus type 1 subtype B ancestral envelope protein is functional and elicits neutralizing antibodies in rabbits similar to those elicited by a circulating subtype B envelope. J Virol. **79**:11214-11224.

6. **Fischer, W., S. Perkins, J. Theiler, T. Bhattacharya, K. Yusim, R. Funkhouser, C. Kuiken, B. Haynes, N. L. Letvin, B. D. Walker, B. H. Hahn, and B. T. Korber**. 2007. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. Nat. Med. **13**:100-106.

7. **Frahm, N., D. C. Nickle, C. H. Linde, D. E. Cohen, R. Zuniga, A. Lucchetti, T. Roach, B. D. Walker, T. M. Allen, B. T. Korber, J. I. Mullins, and C. Brander**. 2008. Increased detection of HIV-specific T cell responses by combination of central sequences with comparable immunogenicity. AIDS **22**:447-456.

8. **Gao, F., T. Bhattacharya, B. Gaschen, J. Taylor, J. P. Moore, V. Novitsky, K. Yusim, D. Lang, B. Foley, S. Beddows, M. Alam, B. Haynes, B. H. Hahn, and B. Korber**. 2003. Author response to Nickle *et al.* Science **299**:1517-1518.

9. **Gao, F., E. A. Weaver, Z. Lu, Y. Li, H. X. Liao, B. Ma, S. M. Alam, R. M. Scearce, L. L. Sutherland, J. S. Yu, J. M. Decker, G. M. Shaw, D. C. Montefiori, B. T. Korber, B. H. Hahn, and B. F. Haynes**. 2005. Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m consensus envelope glycoprotein. J Virol. **79**:1154-1163.

10. **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.

11. **Gautam, R., A. C. Carter, N. Katz, I. F. Butler, M. Barnes, A. Hasegawa, M. Ratterree, G. Silvestri, P. A. Marx, V. M. Hirsch, I. Pandrea, and C. Apetrei**. 2007. In vitro characterization of primary SIVsmm isolates belonging to different lineages. In vitro growth on rhesus macaque cells is not predictive for in vivo replication in rhesus macaques. Virology **362**:257-270.

12. **Kawada, M., T. Tsukamoto, H. Yamamoto, N. Iwamoto, K. Kurihara, A. Takeda, C. Moriya, H. Takeuchi, H. Akari, and T. Matano**. Gag-specific cytotoxic T lymphocyte-based control of primary simian immunodeficiency virus replication in a vaccine trial. J Virol., in press.

13. **Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw**. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A **105**:7552-7557.

14. **Kong, W. P., Y. Huang, Z. Y. Yang, B. K. Chakrabarti, Z. Moodie, and G. J. Nabel**. 2003. Immunogenicity of multiple gene and clade human immunodeficiency virus type 1 DNA vaccines. J. Virol. **77**:12764-12772.

15. **Kong, W.-P., C. Hood, Z.-Y. Yang, C. J. Wei, L. Xu, A. Garcia-Sastre, T. M. Tumpey, and G. J. Nabel**. 2006. Protective immunity to lethal challenge of the 1918 pandemic influenza virus by vaccination. Proc. Natl. Acad. Sci. USA **103**:15987-15991.

16. **Korber, B., B. Gaschen, K. Yusim, R. Thakallapally, C. Kesmir, and V. Detours**. 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. Br. Med. Bull. **58**:19-42.

17. **Kothe, D. L., J. M. Decker, Y. Li, Z. Weng, F. Bibollet-Ruche, K. P. Zammit, M. G. Salazar, Y. Chen, J. F. Salazar-Gonzalez, Z. Moldoveanu, J. Mestecky, F. Gao, B. F. Haynes, G. M. Shaw, M. Muldoon, B. T. Korber, and B. H. Hahn**. 2007. Antigenicity and immunogenicity of HIV-1 consensus subtype B envelope glycoproteins. Virology **360**:218-234.

18. **Kothe, D. L., Y. Li, J. M. Decker, F. Bibollet-Ruche, K. P. Zammit, M. G. Salazar, Y. Chen, Z. Weng, E. A. Weaver, F. Gao, B. F. Haynes, G. M. Shaw, B. T. Korber, and B. H. Hahn**. 2006. Ancestral and consensus envelope immunogens for HIV-1 subtype C. Virology **352**:438-449.

19. **Letvin, N. L., J. R. Mascola, Y. Sun, D. A. Gorgone, A. P. Buzby, L. Xu, Z. Y. Yang, B. Chakrabarti, S. S. Rao, J. E. Schmitz, D. C. Montefiori, B. R. Barker,**

F. L. Bookstein, and G. J. Nabel. 2006. Preserved CD4+ central memory T cells and survival in vaccinated SIV-challenged monkeys. Science **312**:1530-1533.

20. **Li, F., U. Malhotra, P. B. Gilbert, N. R. Hawkins, A. C. Duerr, J. M. McElrath, L. Corey, and S. G. Self**. 2006. Peptide selection for human immunodeficiency virus type 1 CTL-based vaccine evaluation. Vaccine **24**:6893-6904.

21. **Liao, H. X., L. L. Sutherland, S. M. Xia, M. E. Brock, R. M. Scearce, S. Vanleeuwen, S. M. Alam, M. McAdams, E. A. Weaver, Z. Camacho, B. J. Ma, Y. Li, J. M. Decker, G. J. Nabel, D. C. Montefiori, B. H. Hahn, B. T. Korber, F. Gao, and B. F. Haynes**. 2006. A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses. Virology **353**:268-282.

22. **Malhotra, U., F. Li, J. Nolin, M. Allison, H. Zhao, J. I. Mullins, S. Self, and M. J. McElrath**. 2007. Enhanced detection of human immunodeficiency virus type 1 (HIV-1) Nef-specific T cells recognizing multiple variants in early HIV-1 infection. J Virol. **81**:5225-5237.

23. **Matano, T., M. Kobayashi, H. Igarashi, A. Takeda, H. Nakamura, M. Kano, C. Sugimoto, K. Mori, A. Iida, T. Hirata, M. Hasegawa, T. Yuasa, M. Miyazawa, Y. Takahashi, M. Yasunami, A. Kimura, D. H. O'Connor, D. I. Watkins, and Y. Nagai**. 2004. Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial. J Exp. Med. **199**:1709-1718.

24. **Nickle, D. C., M. Rolland, M. A. Jensen, S. L. Pond, W. Deng, M. Seligman, D. Heckerman, J. I. Mullins, and N. Jojic**. 2007. Coping with viral diversity in HIV vaccine design. PLoS. Comput. Biol. **3**:e75.

25. **Novitsky, V., U. R. Smith, P. Gilbert, M. F. McLane, P. Chigwedere, C. Williamson, T. Ndung'u, I. Klein, S. Y. Chang, T. Peter, I. Thior, B. T. Foley, S. Gaolekwe, N. Rybak, S. Gaseitsiwe, F. Vannberg, R. Marlink, T. H. Lee, and M. Essex**. 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J Virol. **76**:5435-5451.

26. **Reynolds, M. R., A. M. Weiler, K. L. Weisgrau, S. M. Piaskowski, J. R. Furlott, J. T. Weinfurter, M. Kaizu, T. Soma, E. J. Leon, C. Macnair, D. P. Leaman, M. B. Zwick, E. Gostick, S. K. Musani, D. A. Price, T. C. Friedrich, E. G. Rakasz, N. A. Wilson, A. B. McDermott, R. Boyle, D. B. Allison, D. R. Burton, W. C. Koff, and D. I. Watkins**. 2008. Macaques vaccinated with live-attenuated SIV control replication of heterologous virus. J Exp. Med.

27. **Richman, D. D., S. C. Morton, T. Wrin, N. Hellmann, S. Berry, M. F. Shapiro, and S. A. Bozzette**. 2004. The prevalence of antiretroviral drug resistance in the United States. AIDS **18**:1393-1401.

28. **Richman, D. D., T. Wrin, S. J. Little, and C. J. Petropoulos**. 2003. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc. Natl. Acad. Sci. U. S. A **100**:4144-4149.

29. **Rolland, M., M. A. Jensen, D. C. Nickle, J. Yan, G. H. Learn, L. Heath, D. Weiner, and J. I. Mullins**. 2007. Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins. J Virol. **81**:8507-8514.

30. **Rong, R., S. Gnanakaran, J. M. Decker, F. Bibollet-Ruche, J. Taylor, J. N. Sfakianos, J. L. Mokili, M. Muldoon, J. Mulenga, S. Allen, B. H. Hahn, G. M. Shaw, J. L. Blackwell, B. T. Korber, E. Hunter, and C. A. Derdeyn**. 2007. Unique mutational patterns in the envelope alpha 2 amphipathic helix and acquisition of length in gp120 hypervariable domains are associated with resistance to autologous neutralization of subtype C human immunodeficiency virus type 1. J Virol. **81**:5658-5668.

31. **Santra, S., B. T. Korber, M. Muldoon, D. H. Barouch, G. J. Nabel, F. Gao, B. H. Hahn, B. F. Haynes, and N. L. Letvin**. 2008. A centralized gene-based HIV-1 vaccine elicits broad cross-clade cellular immune responses in rhesus monkeys. Proc. Natl. Acad. Sci. U. S. A **105**:10489-10494.

32. **Seaman, M. S., L. Xu, K. Beaudry, K. L. Martin, M. H. Beddall, A. Miura, A. Sambor, B. K. Chakrabarti, Y. Huang, R. Bailer, R. A. Koup, J. R. Mascola, G. J. Nabel, and N. L. Letvin**. 2005. Multiclade human immunodeficiency virus type 1 envelope immunogens elicit broad cellular and humoral immunity in rhesus monkeys. J Virol. **79**:2956-2963.

33. **Shimodaira, H.** 2004. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. Ann. Statist. **32**:2616-2641.

34. **Thurmond, J., H. Yoon, C. Kuiken, K. Yusim, S. Perkins, J. Theiler, T. Bhattacharya, B. Korber, and W. Fischer**. 2008. Web-based design and evaluation of T-cell vaccine candidates. Bioinformatics. **24**:1639-1640.

35. **Ward, J. H.** 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. **58**:236-244.

36. **Weaver, E. A., Z. Lu, Z. T. Camacho, F. Moukdar, H. X. Liao, B. J. Ma, M. Muldoon, J. Theiler, G. J. Nabel, N. L. Letvin, B. T. Korber, B. H. Hahn, B. F. Haynes, and F. Gao**. 2006. Cross-subtype T-cell immune responses induced by a human immunodeficiency virus type 1 group m consensus env immunogen. J Virol. **80**:6745-6756.

37. **Wilson, N. A., J. Reed, G. S. Napoe, S. Piaskowski, A. Szymanski, J. Furlott, E. J. Gonzalez, L. J. Yant, N. J. Maness, G. E. May, T. Soma, M. R. Reynolds, E. Rakasz, R. Rudersdorf, A. B. McDermott, D. H. O'Connor, T. C. Friedrich, D. B. Allison, A. Patki, L. J. Picker, D. R. Burton, J. Lin, L. Huang, D. Patel, G. Heindecker, J. Fan, M. Citron, M. Horton, F. Wang, X. Liang, J. W. Shiver, D. R. Casimiro, and D. I. Watkins**. 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.

38. **Yang, Z.-Y., C. J. Wei, W.-P. Kong, L. Wu, L. Xu, D. F. Smith, and G. J. Nabel**. 2007. Immunization by avian H5 influenza hemagglutinin mutants with altered receptor binding specificity. Science **317**:825-828.

FIGURE LEGENDS

Figure 1. Schematic representation of modified HIV Env genes and comparison of the potential epitope coverage (9-mer matches) between different trivalent Env cocktails and the Los Alamos HIV Database of M group Env sequences.

(A) The major structural features of the Env proteins used in the present study (gp160, gp160ΔVs, gp160ΔCFI, gp160ΔCFIΔVs, and gp145ΔCFIΔVs (F1, F2, F3, F4)) are shown. LS represents the leader sequence. V_1, V_2, V_3, V_4, and V_5 indicate the respective variable regions. The transmembrane domain (TM) and other major structural domains are shown as indicated. Various domains of Env were deleted and the modified Env genes were inserted into the mammalian expression plasmid pCMV/R (2,38) for immunization.

(B) The Env alignment used for these comparisons is from the 2007 Los Alamos database, and includes one sequence from each of 1531 people, collected from all over the globe and aligned, including the spectrum of diversity found among known M group clades and recombinants. Red indicates the fraction of all 9-mers in the database set that are found perfectly matched in a given antigen cocktail, orange shows those that are matched in 8/9 positions, and yellow in 7/9 positions – some of these may be cross-reactive. The gp145ΔCFI-modified versions of these proteins, shown on the right, do not have as many matching potential epitopes, because sections of Env are deleted. The nat.3 set was optimized for gp160, and is not particularly well optimized for gp145ΔCFI, and gives comparable coverage to Env ABC in the retained regions. When the subset of sequences from clades A (82 sequences), B (454) or C (464) were each compared

separately to these M group-designed antigens, each clade behaved roughly comparably (within a few percent) to the full dataset results illustrated here (data not shown).

Figure 2. CD4 and CD8 T-cell responses in mice immunized with different gp160 mosaics, natural sequences and gp145ΔCFI. The indicated variants of gp160 mosaics and natural strain, with or without ΔVs (Fig. 1), were compared to gp145ΔCFI. Percent positive $CD4^+$ (left) and $CD8^+$ (right) cells of the total T cells were measured using ICS for IFN-γ (solid bars) and TNF-α (open bars). Immunization and ICS was performed as described in Materials and Methods. The vaccination groups included pCMV/R with no insert (Control) at 15 μg per animal, 15 μg per animal of the gp160 mosaic mos 1.0; 7.5 μg of each plasmid in the two-plasmid groups - gp160 mos 2.1 and 2.2; or 5 μg of each of gp160-mosaics mos 3.1, 3.2 and 3.3, three gp160 natural strains, gp160ΔVs natural strains, and gp145ΔCFI clades A, B and C. Percent positive $CD4^+$ (left) and $CD8^+$ (right) cells of the total T cells was measured using ICS for IFN-γ (solid bars) and TNF-α (open bars). The minimal threshold response indicated by red horizontal dashed lines was defined as 2 times the response from each group's unstimulated samples (un). Red bars indicate responses above the background, while black bars show the level of response to epitope pools below the background level of detection.

Figure 3. Estimates of the magnitude of the vaccine-induced immunity based on ICS responses to different peptide pools: Env ABC, PTE pools and grouped PTE peptides. Immune responses to the indicated immunogens by (A) IFN-γ and (B) TNF-α ICS. The figures plot the mean and uncertainty in the "vaccine effect" v_i, which is the

natural logarithm of the ratio of the vaccine strength to that of the negative control. Thus, vaccine strength goes as $\exp(v_i)$; cf. Statistical Methods for details. The negative control is always shown on the far left, in black, and has mean value zero, by definition. The vaccine effect of other vaccines is measured relative to that of the negative control. The green "X" plotted second from the left is a positive control, where an autologous response to the natural strains unselected A, B and C clade strains is measured against peptides derived from the vaccine protein sequences. Error bars denote $\pm 2\sigma$ uncertainties. The open circle is a negative control; the rectangle denotes gp160; the filled circle, gp145ΔCFI; and "X", gp160ΔCFI.

Figure 4. IFN-γ and TNF-α ICS responses of CD8+ T cells to 78 PTE peptide pools for selected vaccine antigen designs. Immune responses to the indicated immunogens by (A) IFN-γ and (B) TNF-α ICS. Data from all dates and modifications of basic vaccines are plotted. Different dates are indicated by different colors. Positive PTE pools, in which the median response of replicate experiments exceeds 0.1, are indicated by Xs. The median response for each pool is indicated by an open square. Negative control vaccines were uniformly negative, and so not shown.

Figure 5. IFN-γ and TNF-α ICS responses of CD4+ T cells to 78 PTE peptide superpools for selected vaccine antigen designs. Immune responses to the indicated immunogens by (A) IFN-γ and (B) TNF-α ICS. Cutoff is taken at 0.05, and the vertical

axis extends only to 1.0, rather than 2.0 as before. The date and vaccine effects for all models were highly significant ($P < 2\times10^{-7}$).

Figure 6. Hierarchical clustering and heatmap for CD8 and CD4 bifunctional (IFN-γ or TNF-α) T-cell responses to the 78 PTE peptide pools. (A) For CD8 cells, no response (pale yellow), a response to either IFN-γ or TNF–α (orange), or a response to both IFN-γ and TNF–α (red) are indicated. White squares indicate data that is not available. Rows are particular experiments, columns particular peptide pools. (B) For CD4 cells, the same descriptions apply, but in this case only the 4 most conserved peptide pools have very low reactivity and so are marked with arrows. The key at the bottom indicates the vac/mod used; each duplicated symbol indicates a repeated experiment on a different date. The number of antigens in the polyvalent vaccine is indicated at the left on the key (*e.g.*, a dark blue rectangle is mos.3, gp160, which was tested in 11 different repeat experiments). The color key/histogram shows the total number of each class of response, 0, 1, or 2 responses. The 6 peptide pools tested that contained the 36 peptides that bear the most frequent and highly conserved 9-mers are indicated by arrows underneath the appropriate columns, and tend to have sparse responses. Statistically robust clades in the dendrograms (33) are indicated by a number at the branch point.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Amino acid sequences of gp160 and mutant mosaics and best natural sequences. HIV-1 Env gp160 mosaic proteins were designed according to strategies previously described (6) and additionally described in Materials and Methods. The informatic mosaics were optimized as sets of one, two or three mosaics of cocktails. The antigen combinations were designed separately as: 1 mosaic containing mos 1; 2 mosaics containing mos 2.1 and mos 2.2; and 3 mosaics containing mos 3.1, mos 3.2 and mos 3.3. The sequence of the three natural sequences (N1, N2, N3) are also indicated. All the full-length gp160 sequences are indicated in (A). In addition, other Env variants with ΔCFI and gp160ΔV, gp160ΔCFI, gp160ΔVΔCFI, gp145ΔCFI and gp145ΔVΔCFI mutations are also indicated (B), (C), (D), (E) respectively.

Figure S2. gp160ΔCFI mos 2 and mos 3 mosaics elicited more breadth of $CD8^+$ T-cell responses than gp160ΔCFI single mosaic, the two gp160ΔCFI natural strains sets and gp145ΔCFI clades A, B and C. Different versions of gp160ΔCFI mosaics and natural strain (Fig. 1) were compared to gp160 mosaics and gp145ΔCFI clades A, B and C. Immunization and ICS were performed as described in Materials and Methods. The vaccination groups included pCMV/R with no insert (Control) at 15 μg per animal; 15 μg per animal of the gp160ΔCFI mos 1; 7.5 μg of each plasmid in the two-plasmid groups gp160-mos 2.1, 2.2 and gp160ΔCFI-mos 2.1, 2.2; or 5 μg of each of gp160-mos 3.1, 3.2 and 3.3, gp160ΔCFI-mos 3.1, 3.2 and 3.3; three gp160 natural strains, gp160ΔCFI natural strains; and gp145ΔCFI clades A, B and C. Percent positive $CD4^+$ (left) and $CD8^+$ (right) cells of the total T cells was measured using ICS for IFN-γ (solid bars) and TNF-α (open

bars). The minimal threshold response indicated by red horizontal dashed lines was defined as 2 times the response from each group's unstimulated samples (un). Red bars indicate responses above the background, while black bars show the level of response to epitope pools below the background level of detection.

Figure S3. gp160ΔVΔCFI mos 3 mosaics elicited similar breadth of $CD8^+$ T-cell responses as gp160ΔCFI mos 3 mosaics but less than gp160 mos 3 mosaics. gp160ΔCFIΔV mos 3 mosaics and gp160ΔVΔCFI of natural strains were compared to gp160 mos 2 mosaics, gp160 mos 3 mosaics, gp160ΔCFI mos 3 mosaics, gp160ΔCFI of natural strains and gp145ΔCFI clades A, B and C. Immunization and ICS were performed as described in Materials and Methods. Percent positive $CD4^+$ (left) and $CD8^+$ (right) cells of the total T cells were measured using ICS for IFN-γ (solid bars) and TNF-α (open bars).

Figure S4. gp145ΔCFI mosaics showed minimal $CD8^+$ T-cell responses to PTE peptide pool. CD4 and CD8 ICS responses of gp145ΔCFI mos 1 mosaics, gp145ΔCFI mos 2 mosaics and gp145ΔCFI mos 3 mosaics were compared with gp145ΔCFI and gp145ΔCFIΔVs of natural strains, gp160 mos 2 mosaics, gp160 mos 3 mosaics, gp160ΔCFI mos 3 mosaics and gp145ΔCFI clades A, B and C. Immunization and ICS was performed as in Figure S2. The gp145ΔCFI mos 1 mosaics, gp145ΔCFI mos 2 mosaics, and gp145ΔCFI mos 3 mosaics demonstrated minimal CD8 responses to PTE peptide pools.

Figure S5. gp160ΔV mosaics elicit fewer $CD8^+$ T-cell responses than gp160 mosaics. CD4 and CD8 responses of gp160ΔV mos 2 mosaics and gp160ΔV mos 3 mosaics were

compared with gp160 syn 1 mosaics, gp160 mos 2 mosaics, and gp160 mos 3 mosaics. Immunization, ICS, and labeling follows Figures S2, S3 and S4. The gp160ΔV mos 2 mosaics and gp160ΔV mos 3 mosaics elicited fewer CD8 responses to PTE peptides than gp160 mos 2 or mos 3 mosaics.

Figure S6. Matched 9-mers by pool for vaccine candidates. The number of 9-mers that exactly match 9-mers in each vaccine candidate is plotted for each of the 78 small (6-12) peptide pools; vaccine sequences are the full-length (gp160) unmodified Env sequences. Top panel: peptide pools in numerical order. Bottom panel: peptide pools ranked by number of matching counts.

SUPPLEMENTAL TABLES

	none	ΔCFI	gp145ΔCFI	Total
mos 1	22	8	8	38
mos 2	47	8	8	63
mos 3	47	24	8	79
N1-3	12	16	8	36
N1-3ΔV	13	16	8	37
Env ABC	-	–	36	36
Control	44	–	-	44
Total	185	72	76	333

Table S1. The number of plates for each combination of vaccine (row) and modification (column). IFN-γ or TNF-α responses of CD4+ or CD8+ T cells were tested for each vac/mod; thus, the number of plates is generally a multiple of four. When it is not, it indicates that some datasets needed to be discarded.

	Experiment												
Vaccine	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
Control	3	3	3	3	3	3	3	3	3	3	3	3	**36**
gp160 mos 1	3	3	3	3							3	3	**18**
gp160 mos 2.1/2.2	3	3	3	3	3	3	3	3	3	3	3	3	**36**
gp160 mos 3.1/3.2 /3.3	3	3	3	3	3	3	3	3	3	3	3	3	**36**
gp160 N1/N2/N3	3	3	3	3									**12**
gp160ΔV mos 2.1/2.2											3	3	**6**
gp160ΔV mos 3.1/3.2/3.3											3	3	**6**
gp160ΔV N1/N2/N3	3	3	3	3									**12**
gp160ΔCFI mos 1					3	3							**6**
gp160ΔCFI mos 2.1/2.2					3	3							**6**
gp160ΔCFI mos					3	3	3	3	3	3			**18**
gp160ΔCFI N1/N2/N3					3	3	3	3					**12**
gp160ΔVΔCFI mos							3	3					**6**
gp160ΔVΔCFI N1/N2/N3					3	3	3	3					**12**
gp145ΔCFI mos 1									3	3			**6**
gp145ΔCFI mos 2.1/2.2									3	3			**6**
gp145ΔCFI mos									3	3			**6**
gp145ΔCFI N1/N2/N3									3	3			**6**
gp145ΔVΔCFI N1/N2/N3									3	3			**6**
gp145ΔCFI EnvABC	3	3	3	3	3	3	3	3	3	3			**30**

Table S2. Experimental design of different antigens analyzed in this study.

Numbers of mice immunized with different antigens in each of the 12 experiments were. All the mice in each experiment were subjected to ICS after stimulation with the same set of peptide pools: 78 pools (6 to 12 peptides each) of the MHC-independent 492 potential T-cell epitopes (PTE) found most frequently in the sequences of circulating worldwide HIV-1 strains ; 4 superpools (114 or 148 peptides) of the 492 PTE peptides; EnvA(154 peptides), Env B(158 peptides), Env C(154 peptides) from three HIV genes described previously (4); Ebola GP protein (167 peptides) as the negative control, no stimulation for background control, and phorbol myristate acetate (PMA) with ionomycin as the positive control. Total numbers of mice immunized with the same vaccine analyzed by ICS assay in different experiments are summarized in the last column.

Vaccine Candidate	mean	median	minimum
1 mosaic (mos 1)	13.41	9.0	1
2 mosaics (mos 2.1,2.2)	22.47	21.0	3
3 mosaics (mos 3.1,3.2,3.3)	29.26	30.0	9
3 optimal natural (N1,N2,N3)	19.49	17.0	4
3 opt. natural (natΔV N1,N2, N3)	19.38	15.5	3

Table S3. Match counts between 9-mers found in vaccines (gp160 unmodified) and PTE peptide pools.