Rational Peptide Selection To Detect Human Immunodeficiency Virus Type 1-Specific T-Cell Responses under Resource-Limited Conditions

C. B. Willberg, S. K. Pillai, E. R. Sharp, M. G. Rosenberg, J. D. Agudelo, J. D. Barbour, and D. F. Nixon

Division of Experimental Medicine, Department of Medicine, UCSF, San Francisco, California 94158; Veterans Affairs Medical Center, San Francisco, California 94121; Jacobi Medical Center, Albert Einstein College of Medicine, Bronx, New York 10461; and Positive Health Program, San Francisco General Hospital, UCSF, San Francisco, California 94143

Received 23 January 2007/Returned for modification 16 February 2007/Accepted 21 March 2007

Understanding human immunodeficiency virus type 1 (HIV-1)-specific cytotoxic T-lymphocyte responses is important for the development of vaccines and therapies. We describe a novel method for the rational selection of peptides that target stable regions of the HIV-1 genome, rich in epitopes specifically recognized by the study population. This method will be of particular use under resource/sample-limited conditions.

Understanding the fine specificity of cytotoxic T-lymphocyte (CTL) responses is important for vaccine monitoring and design. The most commonly used assay for the detection of antigen-specific CTLs is the enzyme-linked immunospot (ELISPOT) assay. This rapid assay can detect cytokines secreted from virus-specific T cells in response to antigens, usually in the form of synthetic peptides based on consensus or autologous virus sequences. The number of antigens that can be interrogated by this assay depends on the availability of both peripheral blood mononuclear cells (PBMC) and synthetic peptides. Responses to the entire human immunodeficiency virus type 1 (HIV-1) proteome may be mapped, as has been shown, but this requires a large number of both PBMC and peptides (1, 2). An alternative approach for reducing the number of PBMC needed is to use the use of pooled peptides. This has been an effective method of reducing the number of PBMC needed, but it remains dependent on a large number of peptides. Moreover, there are situations in which the number of viable PBMC are severely limited, such as in the pediatric setting, or the cost of manufacturing large sets of peptides is prohibitive. Under these circumstances, a more targeted approach is needed to provide information on responses using the minimum number of cells and peptides.

Here, we describe a quantitative method for selecting panels of individual peptides to maximize data on the breadth and magnitude of a response while minimizing the number of PBMC and peptides needed and taking into account the underlying HLA structure of the sampled population.

Our method selects a pool of peptides for interrogation that fall under epitopes of higher HIV-1 prevalence in the studied population, weighted (penalized) for cases of high entropy (sequence variation). In so doing, we focus on regions of the HIV-1 proteome that are most likely to elicit immunogenic responses from the greatest fraction of the population. Moreover, our weighting strategy leads to higher peptide scores for peptides of low entropy, even for those epitopes falling under low-prevalence HLA types, ensuring the study of immunogenic peptides even among that fraction of the sampled population that bears low-frequency HLA types.

We targeted regions of the HIV-1 proteome that were rich in CTL epitopes (3, 11, 12). These regions were then weighted according to the diversity of the HLA restrictions of the epitopes and with further biased selection toward HLA types of higher frequency within the study population. Finally, to reduce the chance of false-negative results, as described by Altfeld et al., we incorporated the entropy of each amino acid within the selected region (1). This method is described in greater detail below.

Identification of epitopes. We based our analysis on the usage of peptides of 15 amino acids in length. To determine the number of epitopes each peptide contained, we mapped major histocompatibility complex class I (MHC-I)-restricted epitopes, obtained from the Los Alamos database (http://hiv-web.lanl.gov/content/immunology/tables/ctl_summary.html), onto the HIV-1 HXB2 proteome. The HLA restrictions were standardized to the two-digit molecular HLA type nomenclature (8). The epitopes were mapped onto the HXB2 proteome as shown in the examples from HIV-1 Gag p17 (Fig. 1A).

Amino acid score. To maximize the recognition of the selected peptides, epitopes were weighted based on the HLA prevalence (as a percentage) of our North American study group. The HLA prevalence for North America was obtained from the dbMHC IIHWG Anthropology database (http://www.ncbi.nlm.nih.gov/mhc/MHC.fcgi?cmd=init). Each amino acid within the HIV-1 proteome received a score that equaled the total of the HLA prevalence for each epitope that covered it ($\Sigma_{\text{HLA}}$); an example from two 15-mer peptides from p17 is shown (Fig. 1B). When extended across all of HIV-1 Gag p17, the $\Sigma_{\text{HLA}}$ can be displayed as a heat map (Fig. 1B, part iii). The sum of each amino acid's $\Sigma_{\text{HLA}}$ contributed to the final peptide score (Fig. 1D).

Entropy. The variability of HIV-1 antigens is known to interfere with the accurate detection and measurement of HIV-1-specific T-cell responses (1). Differences in sequence between the consensus-based synthetic peptide antigens and an individual's autologous virus often result in false-negative re-
The hypothetical example is shown in blue (part i); the joined epitope constitutes the HLA-A2-restricted SLYNTVTAL epitope. (B) Generation of specific amino acid sequences; their HLA restriction is shown within. Overlapping epitopes were combined to make one continuous epitope region. Aligning of variable regions was performed with the Se-Al gap parameters and the "IUB" DNA weight matrix (10). Sequences were then realigned by using ClustalW with default sequence per individual to minimize the sampling bias. Alignments were restricted to one database (http://hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN_INDEX.html). Alignments were restricted to one protein to minimize the sum of the entropy from each protein were aligned with the HIV-1 HXB2 sequence (4), according to the formula

\[ \sum_{i=1}^{n} \frac{f(A_i)}{\ln(\text{Total length})} \] (6, 9). Sites within each protein were aligned with the HIV-1 HXB2 sequence (4) (Fig. 1C, parts i and ii). The sum of each amino acid’s entropy within a peptide was used in the final peptide score (Fig. 1D).

**Final peptide score.** In this example, the HIV-1 group M consensus (15-mer) peptide sequences, from the NIH ARRRP, were mapped onto the HIV-1 proteome. The final peptide scores were based on the sum of the amino acid’s entropy within a peptide from across the peptide divided by the sum of the entropy from each amino acid within the peptide (Fig. 1D). Peptides from each HIV-1 protein were then ranked; Table 1 shows an example for HIV-1 Gag p17 (Fig. 1D). Peptides from each HIV-1 protein were then ranked; Table 1 shows an example for HIV-1 Gag p17. To determine the breadth of responses, top-ranking peptides from each protein were selected for use in an ELISPOT assay.

**FIG. 1. Generation of peptide scores.** (A) Epitopes were mapped onto the HXB2 genome. Epitopes are shown as white boxes under their specific amino acid sequences; their HLA restriction is shown within. Overlapping epitopes were combined to make one continuous epitope region. A hypothetical example is shown in blue (part i); the joined epitope constitutes the HLA-A2-restricted SLYNTVTAL epitope. (B) Generation of peptide panels. Panel A contains peptides that received the highest scores for each protein; panel B contains the lowest-scoring peptides.

**TABLE 1. Peptide scores for the p17 protein**

| Protein | Peptide sequence | HLA  | Entropy | Peptide | Score |
|---------|------------------|------|---------|---------|-------|
| p17     | MGGARASLGGELDR  | 1.04 | 4.21    | 0.03    |
|         | ASVLSGGELDREWKEI| 1.06 | 4.52    | 0.23    |
|         | SGGLDREWKLRPLRP| 3.75 | 3.77    | 0.99    |
|         | LDRWKEIRLRPPGGK| 6.59 | 2.79    | 2.36    |
|         | EKIRLPQGGKKYKLY| 9.22 | 4.34    | 2.12    |
|         | LRPQGGKKYKLYKHH| 9.96 | 4.99    | 2.00    |
|         | GKKYKLKHYWASR  | 8.75 | 5.08    | 1.72    |
|         | YKLHYWASRELNG  | 6.74 | 3.06    | 2.20    |
|         | HIVHASRELFAVNN | 4.61 | 2.36    | 1.95    |
|         | ASRELFAFVAPPGL| 2.39 | 1.55    | 1.55    |
|         | LERFANPGLLETSE | 1.22 | 0.55    | 0.48    |
|         | AVNPGGLLETSECGRQ| 0.00 | 4.49    | 0.00    |
|         | GLLLETSCGROILGQ| 0.00 | 6.01    | 0.00    |
|         | EVKDTEKALEKIEE | 0.00 | 7.54    | 0.00    |
|         | CROIGLOPSLQOGT| 0.00 | 5.89    | 0.00    |
|         | LGOLOPSLQTGSEEL| 0.65 | 6.17    | 0.11    |
|         | QPSLQTGSEELRSY| 3.08 | 6.13    | 0.50    |
|         | QTGSEELRSLYNTVA| 6.31 | 5.76    | 1.10    |
|         | EELRSLYNTVTALC| 7.00 | 5.98    | 1.17    |
|         | SLNTYVLATLYCHROR | 7.15 | 6.39    | 1.12    |
|         | TVATLCHQRIEVK  | 7.07 | 7.38    | 0.96    |
|         | LYCHQRIEVDKTKE | 6.36 | 4.94    | 1.29    |
|         | HQRJVEKDTEKALEK| 6.34 | 1.87    | 3.39    |
|         | EVKDTEKALEKIEE | 5.04 | 8.46    | 5.83    |
|         | TKJEALKEIEEONKS| 4.36 | 4.03    | 10.23   |
|         | LKJEIEEONKSKKRA| 0.09 | 5.86    | 0.01    |
|         | EEEONKSKKAQQAOAA| 0.00 | 5.52    | 0.00    |
|         | NKKSSKAOAAADTNG | 0.14 | 7.79    | 0.02    |
|         | KKAQQAAADTNGSSQO| 0.85 | 10.28   | 0.08    |
|         | QAAADTNGSSQVSON| 1.76 | 9.44    | 0.19    |
|         | DTNGSSQVSONPYIP| 2.28 | 8.14    | 0.28    |
|         | SSQVSONYPVONOQ| 2.26 | 6.16    | 0.37    |
|         | SQPYPONVQGOMV  | 1.45 | 2.30    | 0.63    |

**TABLE 2. Peptide panels**

| Protein | Panel A | Panel B |
|---------|---------|---------|
| p17     | 7896    | 7897    |
|         | 7895    | 7890    |
| p24     | 7916    | 7941    |
|         | 7957    | 7962    |
| p2 p7 p2 p6 | 7977 | 7978 |
|         | 7964    | 7989    |
| Protease | 5498    | 5491    |
|         | 5490    | 5490    |
| RT      | 5526    | 5596    |
|         | 5552    | 5534    |
| Integrase | 5682 | 5684    |
|         | 5668    | 5668    |
| Tat     | 5117    | 5135    |
|         | 5118    | 5125    |
| Rev     | 6008    | 6002    |
|         | 6014    | 6001    |
| Vif     | 6058    | 6045    |
|         | 6057    | 6040    |
| Vpr     | 6072    | 6064    |
|         | 6071    | 6081    |
| Vpu     | 6426    | 6428    |
|         | 6427    | 6422    |
| Env     | 8771    | 8832    |
|         | 8772    | 8862    |
| Nef     | 5155    | 5177    |
|         | 5156    | 5139    |

*Panel A contains peptides that received the highest scores for each protein; panel B contains the lowest-scoring peptides.*

Sponses. In order to minimize this bias and systematically focus on conserved regions of HIV-1, we computed the amino acid entropy (variability) of sites across the HIV-1 proteome and incorporated this information into our peptide scoring scheme.

In this example, the entropy of each amino acid was determined with full-length subtype B protein sequence alignments of all HIV-1 proteins, downloaded from the Los Alamos database (http://hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN_INDEX.html). Alignments were restricted to one sequence per individual to minimize the sampling bias. Sequences were then realigned by using ClustalW with default gap parameters and the “IUB” DNA weight matrix (10). Manual aligning of variable regions was performed with the Se-Al sequence alignment editor, version 2.0 (http://evolve.zoo.ox.ac.uk). The HyPhy software package was implemented to compute site-specific Shannon entropy scores for each coding region from the frequency, \( f \), of amino acid \( A \) at position \( i \), according to the formula

\[ \sum_{i=1}^{n} \frac{f(A_i)}{\ln(\text{Total length})} \] (6, 9). Sites within each protein were aligned with the HIV-1 HXB2 sequence (4) (Fig. 1C, parts i and ii). The sum of each amino acid’s entropy within a peptide was used in the final peptide score (Fig. 1D).

**Final peptide score.** In this example, the HIV-1 group M consensus (15-mer) peptide sequences, from the NIH ARRRP, were mapped onto the HIV-1 proteome. The final peptide scores were based on the sum of the amino acid’s entropy within a peptide divided by the sum of the entropy from each amino acid within the peptide (Fig. 1D). Peptides from each HIV-1 protein were then ranked; Table 1 shows an example for HIV-1 Gag p17. To determine the breadth of responses, top-ranking peptides from each protein were selected for use in an ELISPOT assay.
Panel B (subjects revealed that panel A was significantly better recognized than responses from nine (panel A) and five (panel B) HIV-1-infected panel B (7.4 peptides induced responses in panel A, compared to 0 in responses was directed toward panel A (Fig. 2). An average of approximately 4 million PBMC were used to cover each peptide.

Proof of principle. To test this method, we screened gamma interferon (IFN-γ) responses from 10 HIV-1-infected subjects. The HLA types of the individuals were unknown. The North American HLA prevalence was applied to generate two panels. The HLA types of the individuals were unknown. The North American HLA prevalence was applied to generate two panels. The IFN-γ ELISPOT assay was conducted by the standard protocol, as previously described (S, 7). In this example, approximately 4 million PBMC were used to cover each peptide panel and controls. As predicted, the highest number of responses was directed toward panel A (Fig. 2). An average of 7.4 peptides induced responses in panel A, compared to 0 in panel B (P = 0.0053) (Fig. 2).

In summary, we have described a novel method for the rational selection of peptides for screening HIV-1-specific CTL responses in resource-limited settings. This protocol targets stable regions of the HIV-1 genome that are rich in epitopes that are specifically recognized by the study population where underlying HLA prevalences are known. In the example given, only 4 million PBMC, which can be isolated from as little as 5 ml of blood, were used and only 26 peptides were needed. If these peptides had been pooled, cell numbers could have been reduced further, although this would reduce the information on the breadth of responses.

This method could easily be adapted to other viruses, such as hepatitis C virus and human cytomegalovirus, and/or expanded to incorporate CD4+ T-cell epitopes. Due to the low number of resources needed, this approach will be of particular benefit in field therapeutic vaccine studies, as well as studies conducted under resource-limited conditions.

This work was funded in part by NIH grant AI060379, with support from the UCSF AIDS Research Institute (ARI). The study received IRB approval.

We thank J. E. Snyder-Cappione, K. E. Garrison, and L. C. Ndhlovu for critical reading of the manuscript.

REFERENCES
1. Altfeld, M., M. M. Addo, R. Shankarappa, P. K. Lee, T. M. Allen, X. G. Yu, A. Rathod, J. Harlow, K. O’Sullivan, M. N. Johnston, P. J. Goulder, J. I. Mullins, E. S. Rosenberg, C. Brander, B. Korber, and B. D. Walker. 2003. Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences. J. Virol. 77:7330–7340.
2. Betts, M. R., D. R. Ambrozak, D. C. Douek, S. Bonhoeffer, J. M. Brenchley, J. P. Casazza, R. A. Koup, and L. J. Picker, 2001. Analysis of total human immunodeficiency virus (HIV)-specific CD4+ and CD8+ T-cell responses: relationship to viral load in untreated HIV infection. J. Virol. 75:11983–11991.
3. Kaufmann, D. E., P. M. Bailey, J. Sidney, B. Wagner, P. J. Norris, M. N. Johnston, L. A. Cosimi, M. M. Addo, M. Lichterfeld, M. Altfeld, N. Frahm, C. Brander, A. Sette, B. D. Walker, and E. S. Rosenberg. 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4+ responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. J. Virol. 78:4463–4477.
4. Korber, B. T., B. T. Foley, C. L. Kuiken, S. K. Pillai, and J. G. Sodroski. 1998. Numbering positions in HIV relative to HXB2. CG p. 102–111. In B. Korber et al., Human retroviruses and AIDS 1998: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Physics Group, Los Alamos National Laboratory, Los Alamos, NM.
5. Larsson, M., X. Jin, B. Ramratnam, G. S. Ogg, J. Engelmaier, M. A. Demotête, A. J. McMichael, W. I. Cox, R. M. Steinman, D. Nixon, and N. Bhadrwaj. 1999. A recombinant vaccinia virus based ELISPOT assay detects high frequencies of Pol-specific CD8 T cells in HIV-1-positive individuals. AIDS 13:767–777.
6. Pond, S. L., S. D. Frost, and S. V. Muse. 2005. PhyHy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.
7. Sandberg, J. K., N. M. Fast, K. A. Jordan, S. N. Furlan, J. D. Barbour, G. Fennell, J. Dobroszycki, H. M. Spiegel, A. Weiner, M. G. Rosenberg, and D. F. Nixon. 2003. HIV-specific CD8+ T cell function in children with vertically acquired HIV-1 infection is critically influenced by age and the state of the CD4+ T cell compartment. J. Immunol. 170:4403–4410.
8. Schreuder, G. M., C. K. Hurley, S. G. Marsh, M. Lau, M. A. Fernandez-Vina, H. J. Noreen, M. Setterholm, and M. Maiers. 2005. HLA dictionary 2004: summary of HLA-A, -B, -C, -DRB1/3/4/5, -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, -DQ antigens. Hum. Immunol. 66:170–210.
9. Shannon, C. E. 1997. The mathematical theory of communication, 1963. M.D. Comput. 14:306–317.
10. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
11. Yu, X. G., M. M. Addo, B. A. Perkins, F. Wei, A. Rathod, S. C. Geer, M. Parta, D. Cohen, D. R. Stone, C. J. Russell, G. Tanzi, S. Mei, A. G. Wurcel, N. Frahm, M. Lichterfeld, L. Heath, J. I. Mullins, F. Marincola, P. J. Goulder, C. Brander, T. Allen, Y. Cao, B. D. Walker, and M. Altfeld. 2004. Differences in the expressed HLA class I alleles effect the differential clustering of HIV type 1-specific T cell responses in infected Chinese and Caucasians. AIDS Res. Hum. Retroviruses 20:557–564.
12. Yusim, K., C. Kesmir, B. Gaschen, M. M. Addo, M. Altfeld, S. Brunak, A. Chigaev, V. Detsou, 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.