AIDScience Vol. 1, No. 2, June 2001

Designing HIV-1 vaccines to reflect viral diversity and the global context of HIV/AIDS

A.S. De Groot[1-3*], H. Sbai[1], J. Frost[1], C. Saint-Aubin[1], N. Chinai[1], W. Martin[2], A. Bosma[1], G. Skowron[3], K. H. Mayer[1]

[1]TB/HIV Research Laboratory, Brown University, Providence, RI 02912, United States
[2]EpiVax, Inc., Providence, RI 02906, United States
[3]Brown University AIDS Program, Brown University, Providence, RI 02912, United States
[*]To whom correspondence should be addressed. E-mail: Anne_DeGroot@brown.edu

Abstract. One of the greatest barriers to developing an effective HIV-1 vaccine is the diversity of HIV-1 viral strains and clades. The number of unique HIV-1 sequences in public databases has been steadily increasing every year, with no end in sight. The most effective type of vaccine in the global context of the HIV epidemic would include immunogenic regions, or epitopes, of the HIV-1 genome that are highly conserved across clades and strains of HIV-1. Until recently, discovery of conserved epitopes in the HIV-1 genome has been hampered by a lack of effective tools that would enable researchers to mine large HIV-1 protein sequence databases for vaccine components. This article reviews the current status of bioinformatics tools for HIV-1 sequence database mining and concludes that the necessary tools for attempting to prepare "cross-clade" vaccines are at hand.

Introduction

HIV is much like a language that is a collection of dialects. Some words and phrases are conserved across the languages, others differ by a small amount, and others are completely lacking. Global communication with the dialects would be less than perfect for example, although it may be true that speaking a dialect of French is enough to be able to communicate in every francophone country, it is more likely that Haitian Creole would be poorly understood in the Ivory Coast, even though some standard French words are conserved. Likewise, although vaccinating individuals against one subtype of HIV-1 might be sufficient protection against other subtypes of HIV-1, it is more likely that vaccinating with a single HIV-1 strain may not be a successful means of protecting against challenge by strains belonging to other clades of HIV-1. Most of the HIV-1 vaccines exiting the developmental pipeline have been developed in the U.S. and Europe by research groups working with the most readily available strains of HIV-1, namely, those clustered under subtype or clade B (the subtype of HIV that predominates in the United States and Europe)(*1*)(Table 1). Concern about the "dialects" or strains of HIV-1 has stimulated a flurry of scientific debate about the efficacy of HIV-1 vaccines developed from Clade B HIV-1 strains.

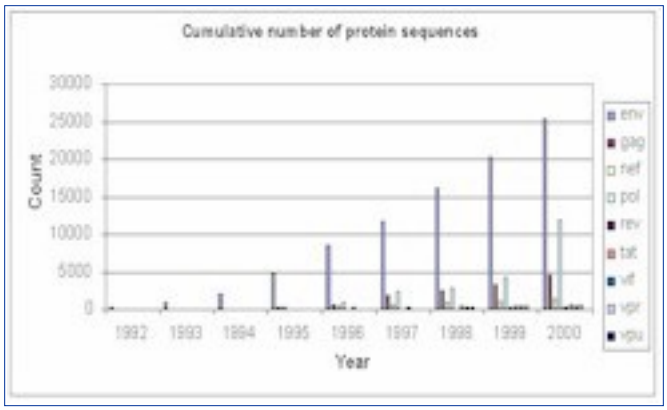
HIV-1 vaccines in development fall roughly into three categories: recombinant proteins, vectored vaccines, and "replicons" (highly engineered viral vectors). Of the 13 or so

recombinant envelope vaccines that have been studied in clinical trials to date, very few have incorporated non-clade B strain components. Vectored vaccines using vaccinia and canarypox have also focused on clade B strains (2). Concerns about cross-clade efficacy initially dampened enthusiasm for vaccine trials of these vaccines in developing countries. However, canarypox was shown to induce CTL against non-clade B subtypes (*in vitro*). Therefore the National Institutes of Health approved a phase I trial of CP205 (canarypox, clade B, env gag and pol), launched in Uganda in 1999. Clade B canarypox vectors may also enter phase I/II trials in Brazil, Haiti, and Trinidad.

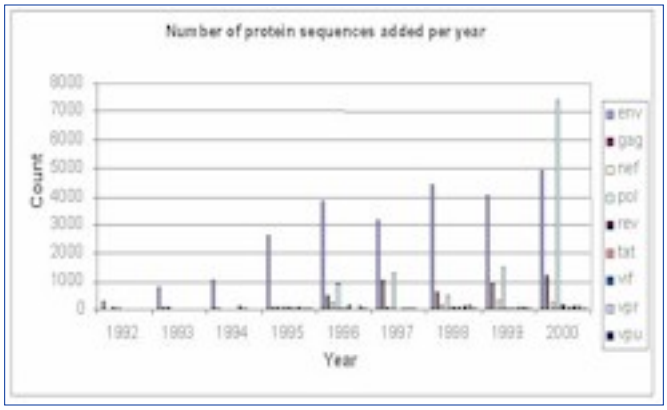
Research groups have also attempted to address the problem of HIV-1 variation by developing vaccines containing sequences from non-clade B HIV-1 strains. A canarypox product containing a clade A envelope (vCP1452-A) is likely to enter phase I trials later this year. A B/E recombinant envelope (gp120) vaccine (Vaxgen's AIDSVAX B/E) is now in phase III efficacy trials in Thailand. A replicon vaccine based on Venezuelan equine encephalitis virus (VEE)(*3*, *4*), containing a clade C HIV-1 sequence is in development by AlphaVax. The Oxford AIDS Vaccine Initiative (OXAVI) is also developing a clade C vaccine using DNA and viral vectors (5). Since each of these vaccines is expected to protect against HIV strains of one clade, successful protection against several clades may require immunization with more than one vaccine.

One solution to the problem of immunizing against many HIV clades may be to search for components of HIV-1 that are highly conserved, and to build a vaccine based on these components. Just as there may be certain words that are conserved in all dialects of French, there may be a set of epitopes (regions of HIV-1 that stimulate immune response) that are conserved across all strains and clades of HIV. These "cross-clade epitopes" may be the critical elements required for developing an effective HIV-1 vaccine or enhancing existing vaccines, given the extent of HIV-1 variability and the global context of HIV.

The variability of HIV-1 genomes can be measured, in part, by viewing the list of HIV-1 sequences archived at the Los Alamos National Laboratory HIV-1 Sequence Database (*6*) or at Genbank (*7*). The number of unique HIV-1 sequences in these public databases has been steadily increasing every year, with no end in sight (Figures 1a and 1b). Whereas only 16,000 HIV-1 protein sequences were listed in 1996, currently more than 44,000 HIV-1 protein sequences are listed. With the sequencing of more full-length genomes, the ability of HIV to mutate and recombine in chimeric forms (A/C, B/E) is beginning to be appreciated. Given the tendency of the HIV-1 retrovirus to evolve within individuals and populations, new variants will continue to be described.



A [View larger version of this image]



B [View larger version of this image]

Figure 1 (A and B). Exponential growth in the HIV-1 protein sequence databases is apparent from these two figures. A complete list of HIV-1 sequences can be obtained at the Los Alamos National Laboratory HIV-1 Sequence Database or at Genbank.

Discovery of conserved, immunologically relevant cross-clade regions of the HIV-1 genome has been hampered by a lack of powerful tools that would enable researchers to mine existing large HIV-1 sequence databases for vaccine components. This article will discuss the current status of bioinformatics tools for HIV-1 sequence database mining, review research on cross-clade responses to HIV-1 epitopes, and outline a new approach to the preparation of cross-clade vaccines.

HIV-1 Strain variation affects HIV-1 epitope processing and presentation

T cell epitopes

Successful vaccination leads to expansion of a set of T cells that are specific for T cell epitopes contained within the protein sequence of the vaccine and establishment of T cell memory to those epitopes (*8*). The hypothesis that expanded T cell responses to HIV-1 peptides following vaccination may be correlated with protection from HIV-1 remains to be proven in vaccine trials. However, both T helper (Th) and Cytotoxic T cells (CTL) responses play a role in protective responses against chronic HIV-1 infection (*9*, *10*), and appear to play a protective role against HIV-1 challenge in certain situations (*11*, *12*), therefore the extrapolation to protection against HIV-1 challenge seems reasonable.

The impact of HIV-1 strain variation on T cell responses has been explored by a number of

scientific groups (Table 2). In general, these studies involve mapping responses to HIV-1 at the level of the interaction between the T cell and the antigen presenting cell (APC) - more specifically, at the level of the peptide epitope, which links the APC's major histocompatability complex (MHC) molecule and the effector T cell's receptor (Figure 2a).

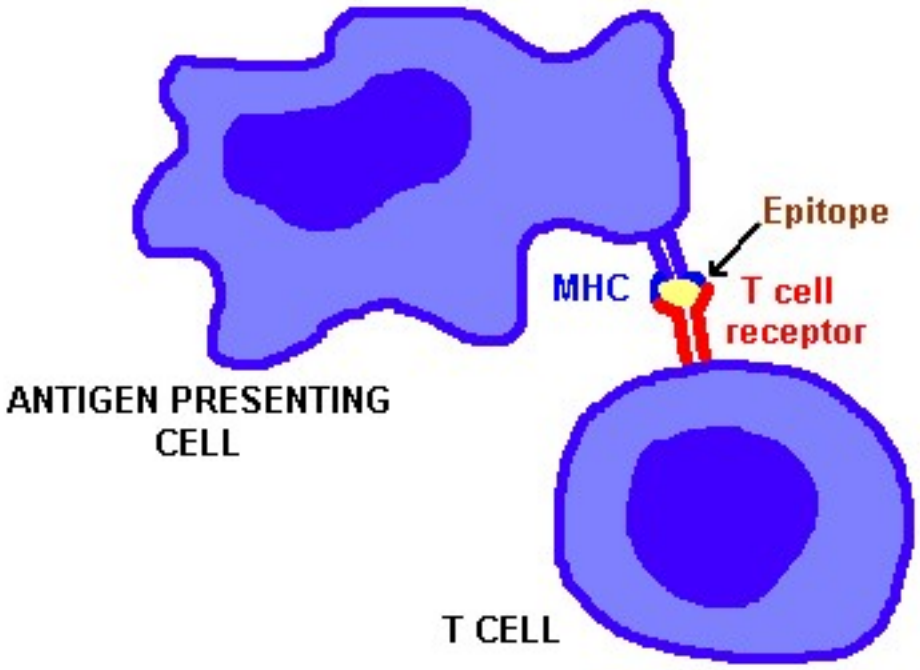


Figure 2a. Interaction between the MHC, peptide, and T cell.
Recognition of a foreign antigen by T cells requires that the antigen derived peptides be displayed within the context of an MHC molecule binding pocket. For a viral pathogen such as HIV, this peptide is processed internally within the cytoplasm and displayed within the cleft of MHC class I molecules. Each type of MHC class I molecule binds a unique set of peptides as a result of the polymorphism within a species and the diversity among individual members of a species, which are then presented to CD8+ T-cells. Each T cell clone recognizes one specific epitope-MHC complex.

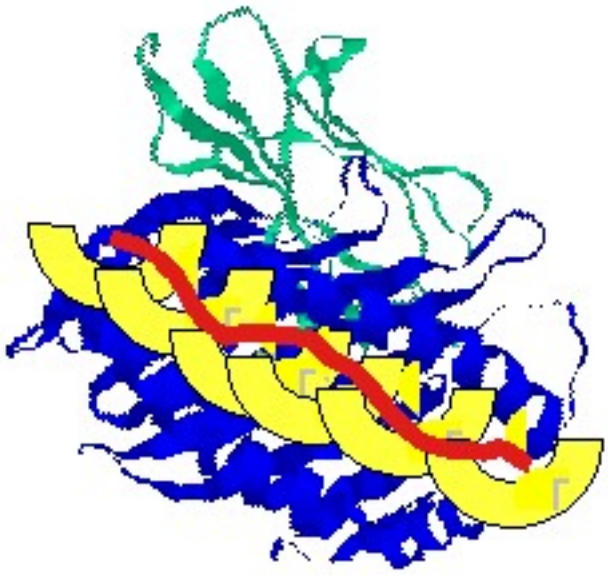


Figure 2b. Interaction between MHC ligand peptide and the MHC molecule.
A peptide (in red) derived from a pathogen such as HIV binds in the cleft of the MHC molecule. R group side chains from the peptide extend into pockets (in yellow) in MHC binding cleft. The configuration of the pockets is unique to each MHC, thus only selected peptides can bind in each MHC binding cleft.

The process of antigen processing is complex. Ultimately, short peptides (MHC ligands) derived from larger proteins bind in the binding groove of MHC Class I and Class II molecules and are transported to the APC's surface (Figure 2a), where they interact with the T cell receptor. For formation of class I peptides, the process involves cleavage in the APC's proteasome, binding to a trans-membrane transport protein called TAP, and further processing in the endoplasmic reticulum. Class II peptides are processed in the APC's proteolytic vesicles. On the surface of the APC, the MHC class I-peptide complex interacts with the T cell receptor of class I restricted T cells (usually CTL) and the MHC class II-peptide complex interacts with the T cell receptor of class II restricted T cells (usually Th). Each T cell receptor is specific for one MHC-peptide complex. Recognition of the MHC-peptide complex by the T cell receptor triggers a cascade of events that culminate in T cell response (secretion of cytokines or direct attack on the APC). MHC ligands that trigger T cell responses are called T cell epitopes.

The critical step in this process appears to be binding to the MHC. MHC ligands/T cell epitopes bind to the MHC through interactions between the peptide's R group side chains and pockets located on the floor of the MHC (Figure 2b). Different MHC molecules (defined by HLA alleles) have distinct types of MHC binding pocket-R group interactions, limiting the set of peptide ligands that can be presented in the context of any given MHC. It is now clear that MHC-defined structural constraints on peptides that bind in the MHC groove result in the "genetic restriction" of immune response, described by Zinkernagel and Doherty more than two decades ago (*13*).

Due to the combined effects of intracellular processing and MHC structure, only a subset of the entire set of peptides derived from any given HIV-1 protein sequence is likely to bind to an

MHC molecule. Changes in the sequence of HIV-1 peptides that conform to the binding specificities of a particular MHC can compromise binding of the modified peptide to that MHC (*14*). Therefore, the changes in amino acid sequence that are associated with HIV-1 diversity may interfere with HIV-1 ligand processing and binding in the MHC binding groove. Failure to bind may diminish cross-clade protection against HIV-1 challenge by T cell clones raised against peptides derived from a clade B vaccine construct.

Analyses of HIV-1 immunopathogenesis have provided ample evidence of the highly epitope-specific nature of immune response to HIV-1. Modifications of MHC ligands at the amino acid level have been associated with failure to bind, or failure to be recognized by the T cell, resulting in viral escape from immune response (*15*, *16*, *17*). Furthermore, sequence modifications may affect the intracellular processing of the epitopes prior to MHC binding, since these modifications could affect processing of the native protein into shorter peptides by the proteasome or transport of the protein through TAP into the endoplasmic reticulum (*18*, *19*). Alternatively, variant peptides that still bind to the MHC may fail to engage the TCR, acting as an "antagonist" to T cell response (*20*).

Despite the epitope-specificity of human response to HIV-1, T cells induced by recombinant canarypox vectors based on clade B HIV-1 strains have been able to kill cells infected with HIV-1 or cells transfected with HIV-1 genes from other clades (*21*). These successes can be attributed to one of several special factors: (1) In general, vaccines that have induced these cross-clade responses included proteins from HIV-1 that are more conserved proteins of HIV-1, such as gag and other internal proteins (*22*, *23*, *24*, *25*)(see Table 2 for a partial listing of cross-clade studies). (2) The majority of published studies were performed using recombinant vaccinia virus constructs expressing whole HIV-1 genes and mixed populations of T cells derived from HIV-1 infected subjects. Quite a few of the epitopes contained in genes presented by these vaccinia and whole gene constructs can be shown to be conserved across clades (see Table 3). These types of "whole gene" cross-clade studies may simply be confirming the existence of cross-clade epitopes (amino acid sequences that are conserved across clades of HIV-1) rather than proving that clade B vaccines are acceptable immunogens in the context of a non-clade B HIV-1 challenge. And finally, (3) several studies providing more exacting proof that specific HIV-1 epitopes derived from different strains of HIV-1 can stimulate cross-clade responses were performed using highly conserved peptides with no substitutions in the amino acids that are associated with binding to the MHC (the amino acid "anchors" in the peptide sequence)(*26*, *27*)(Table 2). The latter studies are consistent with recent discoveries about the "wobbly" nature of T cell receptor (TCR) interaction with the MHC-peptide complex (*28*, *28*). The wobbly TCR permits a broader range of possible MHC-peptide combinations per epitope-specific T cell receptor than previously estimated.

Of note, few "cross-clade" studies have been performed using peptides representing epitopes from regions of HIV-1 that dramatically *differ* (particularly in the anchor regions) across clades. Results of these studies are perhaps too predictable: CTL response would probably not be conserved against dissimilar epitopes. It is also much easier to confirm a positive "cross-clade" response (using more conserved epitopes) than to clearly define a negative, or missing, cross-clade response.

B cell epitopes

The adverse impact of HIV-1 strain variation on the efficacy of antibodies raised by clade B vaccines has also been profound. The envelope protein (composed of gp120 and gp41) against which humoral response is directed is extremely variable (more than 25,000 env sequences are currently available in public databases, Figures 1a and 1b).

A few cross-neutralizing antibody epitopes to gp120 of HIV-1 grown *in vitro* have been identified (*29*). Antibodies to these epitopes have shown to be relatively inefficient for the neutralization of HIV-1 isolates from infected patients. A single neutralizing epitope has been defined on HIV-1 gp41 (*29*). This epitope is relatively conserved across HIV-1 strains, but not very immunogenic. In general, antibodies to the handful of cross-clade conserved neutralizing epitopes that have been identified are rarely elicited in the humoral response after HIV-1 infection. Cross-neutralizing antibodies (able to inhibit infection by heterologous HIV-1 isolates) have also not been elicited in human trials of the two vaccines currently in phase II and III trials: the canarypox vaccine (*30*) and the recombinant gp120 vaccine (*31*). Antibodies raised against HIV-1 envelope have usually been specific for the clade B isolates in the vaccine formulations, and have only rarely neutralized primary isolates of HIV derived from patients (*32*, *33*). Even those antibodies elicited in improved vaccination (prime-boost) protocols have had a limited breadth of reactivity (*30*).

Research groups developing HIV-1 vaccines directed at stimulating the humoral arm of the immune system have addressed the problem of HIV-1 variability in a number of different ways. For example, some groups have devoted their attention to the development of oligomeric recombinant envelope protein, which more closely resembles the native configuration of the protein in HIV-1 virions (*34*). Other groups have experimented with modifiying delivery mechanisms to boost cross-clade immunogenicity. Recently, two clade B vaccine candidates, the Chiron DNA-prime, recombinant protein boost (both clade B) vaccine and the Vaxgen recombinant gp120 (clade B) vaccine, were reported to induce cross-clade neutralizing antibodies, measured *in vitro* (*35*, *36*). In addition, researchers at Maxgen have plans to use directed molecular evolution technologies to generate HIV-1 proteins that elicit broader, stronger antibody responses than wild type HIV-1 proteins (*37*).

These studies may lead to the development of improved HIV-1 vaccine candidates. However, HIV-1 envelope protein variation is protean. Bioinformatics tools to map cross-clade B cell epitopes do not yet exist. There appears to be no easy solution to the problem of envelope variability and antibody responses, raising concern about the ability of single-clade vaccines to induce effective antibody responses against HIV-1 challenge.

Designing a cross-clade, epitope-driven HIV-1 vaccine using bioinformatics

A number of researchers including the group at the TB/HIV Research Laboratory at Brown University have been promoting and pursuing the development of a novel HIV-1 vaccine that reflects the global diversity of HIV-1 strains and the HLA variability of human populations (*38*, *39*, *40*, *41*, *42*, *43*, *44*, *45*). The approach of these groups generally consists of searching for

conserved peptide sequences, determining whether these sequences are immunogenic, and then cobbling these epitopes together in a multi-epitope vaccine.

Sophisticated sequence tools to search protein sequences for T cell epitopes have been available for approximately 10 years. These informatics tools have only recently been applied to the problem of HIV-1 vaccines. The combined use of several informatics tools is now making it possible to analyze *all published variants* of the HIV-1 genome and prospectively identify both class I and class II-restricted T cell epitopes for use in epitope-driven HIV-1 cross-clade vaccines.

Tools for identifying cross-clade T cell epitopes

The discovery of MHC binding motifs prompted the development of new bioinformatics tools for vaccine design. The MHC-binding-motif based algorithms were first described by Rotzschke and Falk and Leighton *et al*. in 1991 (*46*, *47*), by Lipford *et al*. in 1993 (*48*), by Parker *et al*. in 1994 (*49*) and by the TB/HIV Research Laboratory in 1995 (*50*). Other bioinformatics approaches to predicting T cell epitopes include artificial neural networks (first described by Brusik *et al*. in 1994)(*51*) and structural approaches, as described by Delisi (*52*) and by Altuvia *et al*. (*53*). A variation on predicting peptides using MHC binding motif-based algorithms known as "extended MHC binding motifs" or "peptide side chain scanning" was described by Sette *et al*. (*54*) and Hammer *et al*. (*55*). These methods allowed for the construction of a matrix of all possible amino acid side chain effects for a single MHC-binding motif. A similar approach using synthetic undecapeptides, was developed by B. Fleckenstein *et al*. and reported in 1996 (*56*). Further variations on this theme, termed "matrix-based" T cell epitope selection algorithms, were developed by Davenport and Ho Shon for one class II allele in 1995 (*57*) and by Jesdale and De Groot for an array of class I and class II alleles in 1997 (EpiMatrix)(*58*). Most recently, the team of Hammer, Sturniolo, *et al*. made another important advance in epitope mapping: similarities in MHC binding constraints that are reflected in commonalities in the structure of MHC binding pockets allow new motifs to be developed by mixing and matching binding pocket characteristics (*59*). Bioinformatics developers at EpiVax have expanded this pocket profile approach to include a total of 64 class II motifs for use with EpiMatrix.

EpiMatrix

This algorithm features matrix motifs for 32 HLA class I alleles, several murine alleles, and the ninety matrix motifs for human class II alleles mentioned above. The algorithm ranks 9 to 10 amino acid long segments overlapping by 8 to 9 amino acids, respectively, derived from any protein sequence for match to the selected MHC matrix motif. An estimated binding probability (EBP) is derived by comparing the EpiMatrix score to those of known binders and presumed non-binders (*60*). EpiMatrix has been successfully applied to the retrospective analysis of published epitopes (*60*), to the prospective selection of HLA B7-restricted CTL epitopes from an HIV-infected patient's HIV quasispecies sequences (*61*), to the identification of novel HLA B7-restricted *Mycobacterium tuberculosis* epitopes (*62*), to the selection of novel HLA B35-restricted malaria epitopes (*63*), and to the discovery of HIV-1 clade E HLA A11-restricted

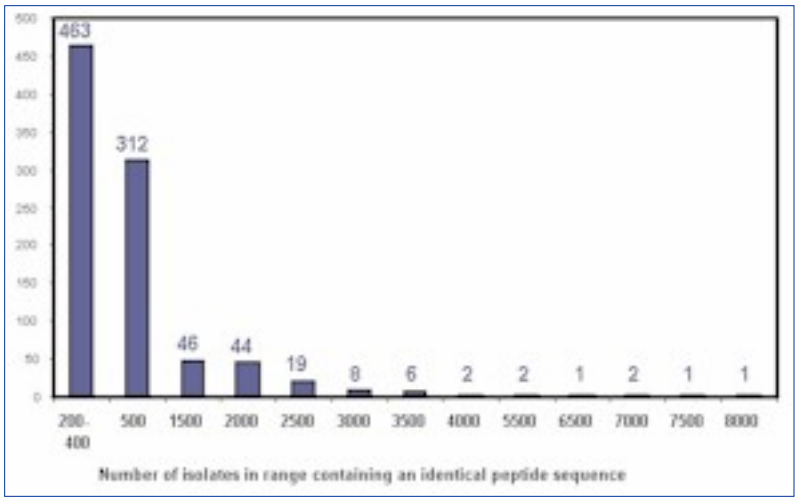
epitopes (*64*).

An additional feature of EpiMatrix is that it can measure the MHC binding potential of each 10 amino acid long snapshot to a number of human HLA, and therefore can be used to identify regions of MHC binding potential clustering. Other laboratories have confirmed cross-presentation of peptides within HLA "superfamilies" (such as the A3 superfamily: A11, A3, A31, A33 and A68) described by Walker, Sette, *et al.* (*54*). Presumably, vaccines containing such "clustered" or promiscuous epitopes will have an advantage over vaccines composed of epitopes that are not promiscuous. EpiMatrix has been used to discover many such promiscuous HIV-1 epitopes, some of which are illustrated in Table 4. The EpiMatrix algorithm has been made available on the internet for use by HIV researchers at the TB/HIV Research Laboratory Web site.

EpiMatrix is just one of a number of tools available on the internet that can be used for T cell epitope searches (*65*). Several epitope mapping tools are available to researchers, including the tool available at the SYFPEITHI Web site (*66*) and the HLA binding prediction tool available on the site authored by Ken Parker at the National Institutes of Health (BIMAS)(*67*). Neither of these sites returns exactly the same predictions as EpiMatrix. Algorithms such as EpiMatrix and Conservatrix (see next section) may be useful for evaluating exisiting vaccines for cross-clade immunogenicity as well as for developing vaccines.

Conservatrix

The key to identifying conserved regions across clades is to search for conserved amino acid "strings" of the same length. The Conservatrix algorithm, another bioinformatics tool developed by the TB/HIV Research Laboratory, accomplishes this by parsing every sequence in a given database (the current HIV-1 sequence database is compiled from the Los Alamos National Laboratory HIV-1 Sequence Database and Genbank) into 9 to 10 amino acid long text strings. The algorithm then performs a simple string-of-text-based search, similar to the approach used by the "find" function in word-processing programs. Each of these text strings is then ranked by number of times it occurs in the set of text strings. (Figure 3) Highly conserved peptide text strings are then input into EpiMatrix and ranked for immunogenicity by EBP. This tool has been applied to the analysis of HIV-1, hepatitis C, and human papilloma virus.



A [View larger version of this image]

Figure 3. A total of 907 HIV-1 envelope peptides are conserved in more than 250 envelope protein sequences; many of these have proven to be T helper or CTL epitopes when tested *in vitro*.

Designing a cross-clade, epitope-driven HIV-1 vaccine using bioinformatics

The approach to developing a "World Clade" vaccine at the TB/HIV Research Laboratory (in collaboration with EpiVax) has been to apply both Conservatrix and EpiMatrix to select highly conserved HIV-1 T cell epitopes (*83*). These epitopes are confirmed *in vitro* and a vaccine is then constructed from these epitopes by cloning their (DNA) coding sequences into a vector plasmid or viral vaccine vector in a tandem string. This vaccine construct, if effective, would be developed as an adjunct to another vaccine that effectively induces cross-clade humoral responses.

The World Clade vaccine approach is illustrated in the next few paragraphs. The first step to developing this vaccine has been to gather sequences from existing public databases (the Los Alamos National Laboratory Web site, supplemented with Genbank) for each of nine distinct HIV proteins. The next step was to use Conservatrix to search for conserved 9 and 10 amino acid long peptides in the HIV-1 protein database. The criteria established for selecting peptides is extremely conservative and similar to the approach suggested by Ferrari *et al*.: *"no more than one single [amino acid] substitution outside the anchor residues in no more than 50% of the isolates analyzed for each clade* (*21*)*."* In Figure 3, the results of a Conservatrix analysis of 18,313 env sequences are shown. 1000 env peptides are categorized by level of conservation. Information on peptides conserved in less than 250 sequences is not shown. Those peptides that are extremely well conserved (in more than 250 HIV-1 sequences for env, for example) were selected for further evaluation by EpiMatrix. This list was scanned for sequences that matched any of EpiMatrix's current list of 30 class I and 74 class II MHC binding motifs.

Selected peptides were synthesized, and both MHC binding capability and T cell responses to the peptides evaluated *in vitro*. MHC binding was evaluated using the T2 cell binding assay (*68*). T cell responses to the peptides were measured in standard gamma-interferon release ELISpot assays (*69*). Table 4 illustrates results obtained for a set of peptides selected for conservation and restriction by the HLA A3 allele. 10 of the original 25 HIV-1 peptides

selected for this experiment were determined to be novel, highly conserved, HLA A3-restricted CTL epitopes, and three previously published A3-restricted peptides were re-confirmed.

Several highly conserved, promiscuous epitopes are shown in the table. For example, CTRPNNTRK is conserved in 7,562 of more than 18,000 env sequences available in public HIV-1 sequence databases. This peptide is a published A2 epitope. EpiMatrix predicted that the epitope would be promiscuous (restricted by more than one HLA allele), and experimental results confirmed that the peptide is also recognized in the context of HLA A3. EpiMatrix also selected epitopes from minor HIV-1 proteins that have not previously been extensively mapped. One such epitope, vif KLTEDRWNK, is extensively conserved across published HIV-1 vif sequences (411 out of 507, or 81%).

These pilot studies provided confirmation that a bioinformatics analysis of a large database of HIV-1 sequences can prospectively identify highly conserved HIV-1 epitopes. EpiVax and the TB/HIV Research Laboratory are now the process of completing T cell assays and assembling a number of HIV-1 epitopes selected by Conservatrix and EpiMatrix in a DNA vaccine vector.

Develop new "Cross-Clade" vaccines or modify existing ones?

The epitope-driven vaccine concept is an attractive one that is being successfully pursued in a number of laboratories (*70*, *71*, *72*). Complex vaccines containing T helper and B cell epitopes alongside CTL epitopes derived from a variety of pathogens (such as 5 viruses and 1 bacterium) have already been constructed and tested (*71*). A typical epitope-based vaccine construct contains a single start codon with coding sequences for epitopes inserted consecutively in the construct, with or without intervening spacer amino acids. *In vitro* studies of several such constructs have confirmed that the epitopes are expressed, stimulate protective immune response, and do not to interfere with one another. Another epitope-driven vaccine approach is to mix several plasmids together, each of which contains genes for different proteins or different minigene epitopes. These discoveries suggest that epitope-based vaccines containing a mixture of HIV-1 epitopes that are highly conserved across clades and strains of HIV-1 are feasible.

Proof that epitope-driven minigene vaccination can stimulate protective immune responses has been obtained by researchers carrying out minigene vaccination studies in a range of animal models (*73*, *74*, *75*, *76*, *77*, *78*). Including a mixture of CTL and Th epitopes may be critical to protection (*79*). For example, immunization with a cocktail of peptides consisting of a B-cell epitope, a T-helper epitope, and a CTL epitope linked to a fusion peptide resulted in a 190-fold reduction in the titer of virus infection in a murine model of a human viral disease (*80*). These results highlight the ability of epitope-based DNA immunization to induce protective immune responses to well-defined epitopes and indicate the potential of this approach for the development of vaccines against HIV. Composing a vaccine of highly conserved sequences from HIV-1 may stimulate a broader range of T cell responses than a "whole gene" vaccine, since no single HIV-1 gene could include all possible conserved epitopes.

Conclusion

Concerns about the global context of the HIV/AIDS epidemic are being addressed by

researchers developing HIV vaccines containing clade A genes, clade C genes, and re-engineering existing vaccines by inserting non-clade B (E) genes or highly conserved genes such as gag. Despite the number of research studies and publications that demonstrate cross-clade responses to Clade B vaccines and epitopes, few HIV-1 vaccine researchers would refute the utility of an HIV-1 vaccine composed of components that are highly conserved across all, or most, HIV-1 clades.

Ultimately, an effective HIV-1 vaccine will need to stimulate both humoral and cellular arms of the immune response, and to induce antibodies and T cells that are capable of recognizing HIV-1 in its protean manifestations. A cross-clade immunogen, such as the one to be developed by Maxygen with International AIDS Vaccine Initiative funding may succeed in stimulating broadly neutralizing antibodies. Training the immune system to recognize many different HIV-1 epitopes, selected for their conservation across clades as illustrated by the TB/HIV Research Laboratory's "World Clade" project, is another approach to developing an effective HIV-1 vaccine. Until now, discovery of highly conserved sequences that are also immuno-stimulatory has been hampered by the lack of bioinformatics tools. Researchers who apply these tools are likely to be successful in their search for a truly polyglot HIV-1 vaccine that addresses the diversity of HIV-1 in the global context of HIV/AIDS.

References and notes

1. S. Bende, M.I. Johnston. *AIDS Read.* **10**, 526 (2000). Available online
2. M. Clements-Mann *et al*. *J. Infect. Dis.*, **177**, 1230 (1998). PubMed
3. N.L. Davis *et al*. *J. Virol.* **74**, 3430 (2000). Free full text article available
4. N.L. Davis, K.W. Brown, R.E. Johnston. *J. Virol.* **70**, 3781 (1996). PubMed
5. T. Hanke, A.J. McMichael. *Nat. Med.* **6**, 951 (2000). PubMed
6. HIV sequence database, Los Alamos HIV database, 1999, B. Korber, G. Meyers eds., Los Alamos National Laboratories, New Mexico 1999. Available online
7. S.F. Altschul *et al*., *Nucleic Acids Res.* **25**, 3389 (1997). PubMed
8. A. Seth *et al*., Proc. Natl. Acad. Sci. U.S.A. 95, 10112 (1998). Free full text article available
9. B.D. Wallker, E.S. Rosenberg, C.M. Hay, N. Basgoz, O.O. Yang. *Adv. Exp. Med. Biol.* **452**, 159 (1998). PubMed
10. M. Altfeld, *et al*. *J. Exp. Med.* **193**, 169 (2001). PubMed
11. S.L. Rowland-Jones, *et al*. *J. Clin. Invest.* **102**, 1758 (1998). PubMed
12. S.L. Rowland-Jones, *et al*. *Lancet* **341**, 860 (1993). PubMed
13. R.M. Zinkernagel, P.C. Doherty. *Immunol. Today* **18**, 14 (1997).
14. Y. Zevering, C. Khamboonruang, M.F. Good. *Eur. J. Immunol.* **24**, 1418 (1994).

15. A.J. McMichael, R.E. Phillips. *Annu. Rev. Immunol.* **15**, 271 (1997).

16. P.J. Goulder, *et al*. *Nat. Med.* **3**, 212 (1997).

17. D.T. Evans, *et al*. *Nat. Med.* **5**, 1270 (1999).

18. Y. Altuvia, H. Margalit. *J. Mol. Biol.* **295**, 8790 (2000).

19. A. Paradela, *et al*. *J. Immunol.* **164**, 329 (2000).

20. P. Klenerman. *Nature* **369**, 403 (1994).

21. G. Ferrari, *et al*. *Proc. Natl. Acad. Sci. USA* **94**, 1396 (1997).

22. G. Ferrari, *et al*. *AIDS Res. Hum. Retroviruses* **16**, 1433 (2000).

23. J.A. Lynch, *et al*. *J. Infect. Dis.* **178**, 1040 (1998).

24. F. Buseyne, *et al*. *Virology* **250**, 316 (1998).

25. S.E. Wilson, *et al*. *AIDS Res. Hum. Retroviruses* **14**, 925 (1998).

26. S.C. Threlkeld, *et al*. *J. Immunology* **159**, 1648 (1997).

27. S. Rowland-Jones, *et al*. *Dev. Biol. Stand.* **92**, 209 (1998).

28. B. Hemmer, M. Vergelli, C. Pinilla, R. Houghten, R. Martin. *Immunology Today* **19**, 163 (1998).

29. Los Alamos HIV database 1999 (Review Articles), Editors: B. Korber *et al*., Los Alamos National Laboratories, New Mexico 1999. (A.F. Labrjin, P.W.H.I. Parren review). Available online

30. R.B. Belshe, *et al*. *AIDS* **12**, 2407 (1998).

31. M.J. McElrath, *et al*. *AIDS Res. Hum. Retroviruses* **16**, 907 (2000).

32. R. Mascola, *et al*. *J. Infect. Dis.* **173**, 340 (1996).

33. D.R. Burton. *Proc. Natl. Acad. Sci. USA* **94**, 10018 (1997).

34. R.W. Sanders, *et al*. *J Virol.* **74**, 5091 (2000).

35. S.W. Barnett, presentation 27 at the 8th conference on Retroviruses and Opportunistic infections, Chicago, IL, 4-8 February, 2001. Available online

36. R. Shibata, *et al*., poster 190 presented at the 8th conference on Retroviruses and Opportunistic infections, Chicago, IL, 4-8 February, 2001. Available online

37. R.G. Whalen, R. Kaiwar, N.W. Soong, J. Punnonen. *Curr. Opin. Mol. Ther.* **3**, 31 (2001).

38. HIV Molecular Immunology Database 1995, Editors: B. Korber *et al*., Los Alamos National Laboratories, New Mexico, IV, 17-28 (1995). Available online

39. J.R. Schafer, poster and abstract presented at Molecular Approaches to the Control of Infectious Diseases, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, September 1996.

40. J.R. Schafer, B.M.Jesdale, J.A. George, N.M. Kouttab, A.S. De Groot. *Vaccine* **16**, 1880 (1998).

41. K. Cease, J. Berzofsky. *Annu. Rev. Immunol.* **12**, 923 (1994).

42. B. Haynes. *Lancet* **348**, 933 (1996).

43. A.V. Hill, M.P. Davenport. *Mol. Med. Today* **2**, 38 (1996).

44. Los Alamos HIV database 1995 (Review Articles), Editors: B. Korber *et al*., Los Alamos National Laboratories, New Mexico 1995 (review by F.E. Ward, S.T. Tuan, B. Haynes). Available online

45. C.C. Wilson, *et al*. *J. Virol.* **75**, 4195 (2001).

46. O. Rotzschke, *et al*. *Eur. J. Immunol.* **21**, 2891 (1991).

47. J. Leighton, *et al*. *J. Immunol.* **147**, 198 (1991).

48. G.B. Lipford, M. Hoffman, H. Wagner, K. Heeg. *J. Immunol.* **150**, 1212 (1993).

49. K.C. Parker, M.A. Bednarek, J.E. Coligan. *J. Immunol.* **152**, 163 (1994).

50. G.E. Meister, C.G.P. Roberts, J.A. Berzofsky, A.S. De Groot. *Vaccine* **13**, 581 (1995).

51. V. Brusic, G. Rudy, L.C. Harrison. In: R.J. Stonier, X.S. Yu, eds., Complex systems, mechanisms of adaption. Amsterdam; IOS Press 253 (1994).

52. R. Rosenfeld, Q. Zheng, S. Vajda, C. Delisi. *Genet. Anal.* **12**, 1 (1995).

53. Y. Altuvia, O. Schueler, H. Margalit. *J. Mol. Biol.* **249**, 244 (1995).

54. A. Sette, *et al*. *J. Immunol.* **151, 3163 (1993).**

55. J. Hammer, *et al*. *J. Exp. Med.* **180**, 2353 (1994).

56. B. Fleckenstein, *et al*. *Eur. J. Biochem.* **240**, 71 (1996).

57. M.P. Davenport, I.A. Ho Shon, A.V. Hill. *Immunogenetics* **42**, 392 (1995).

58. B.M. Jesdale, *et al*. Vaccines '97, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1997.

59. T. Sturniolo, *et al*. *Nat. Biotechnol.* **17**, 555 (1999).

60. A.S. De Groot. *AIDS Res. and Hum. Retroviruses* **7**, 139 (1997).

61. X. Jin, *et al*. *AIDS Res. Hum. Retroviruses* **16**, 67 (2000).

62. Y. Bushkin, poster and abstract, Cold Spring Harbor Molecular Immunology Conference, Cold Spring Harbor, NY, September 1999.

63. M.R. Klein, *et al*. *J. Inf. Dis.* **183**, 928 (2001).

64. K.B. Bond, *et al*. *AIDS Res. Hu. Retroviruses* **17** (2001).

65. EpiMatrix HIV Web site located at http://www.tbhiv.biomed.brown.edu/

66. H.G. Rammensee. *Immunogenetics* **50**, 213 (1999). See Web site

67. K.C. Parker. *J. Immunol.* **152**, 163 (1994). See Web site

68. H.G. Ljunggren, *et al*. *Nature* **346**, 476 (1990).

69. J. Lieberman, *et al*. *AIDS Res. Hum. Retroviruses* **13**, 383 (1997).

70. T. Hanke, J. Schneider, S.C. Gilbert, A.V.S. Hill, A. McMichael. *Vaccine* **16**, 426 (1998).

71. L.L. An, J.L. Whitton. *J. Virol.* **71**, 2292 (1997).

72. E.H. Nardin, *et al*. *J. Immunol.* **166**, 481 (2001).

73. C.P. Simmons, *et al*. *J. Immunol.* **166**, 1106 (2001).

74. E.D. Franke, *et al*. *Infect Immun.* **68**, 3403 (2000).

75. E.B. Schadeck, *et al*. *Virus Res.* **65**, 75 (1999).

76. K.S. Rosenthal, H. Mao, W.I. Horne, C. Wright, D. Zimmerman. *Vaccine* **17**, 535 (1999).

77. M. Tsuji, *et al*. *J. Virol.* **72**, 6907 (1998).

78. A.D. Hislop, *et al*. *Nat Med.* **4**, 1193 (1998).

79. A. Fomsgaard, *et al*. *Vaccine* **18**, 681 (1999).

80. S.C. Hsu, D. Chargelegue, O.E. Obeid, M.W. Steward. *J. Gen. Virol.* **80**, 1401 (1999).

81. The authors are extremely grateful to Dr. Frank Rothman of Brown University for his helpful suggestions and careful review of this manuscript. The World Clade Vaccine project is a not-for profit collaboration between EpiVax, Inc. and the TB/HIV Research Laboratory. A new non-profit initiative, GAIA Vaccine Foundation, has been established to support the development of a cross-clade vaccine against HIV-1.

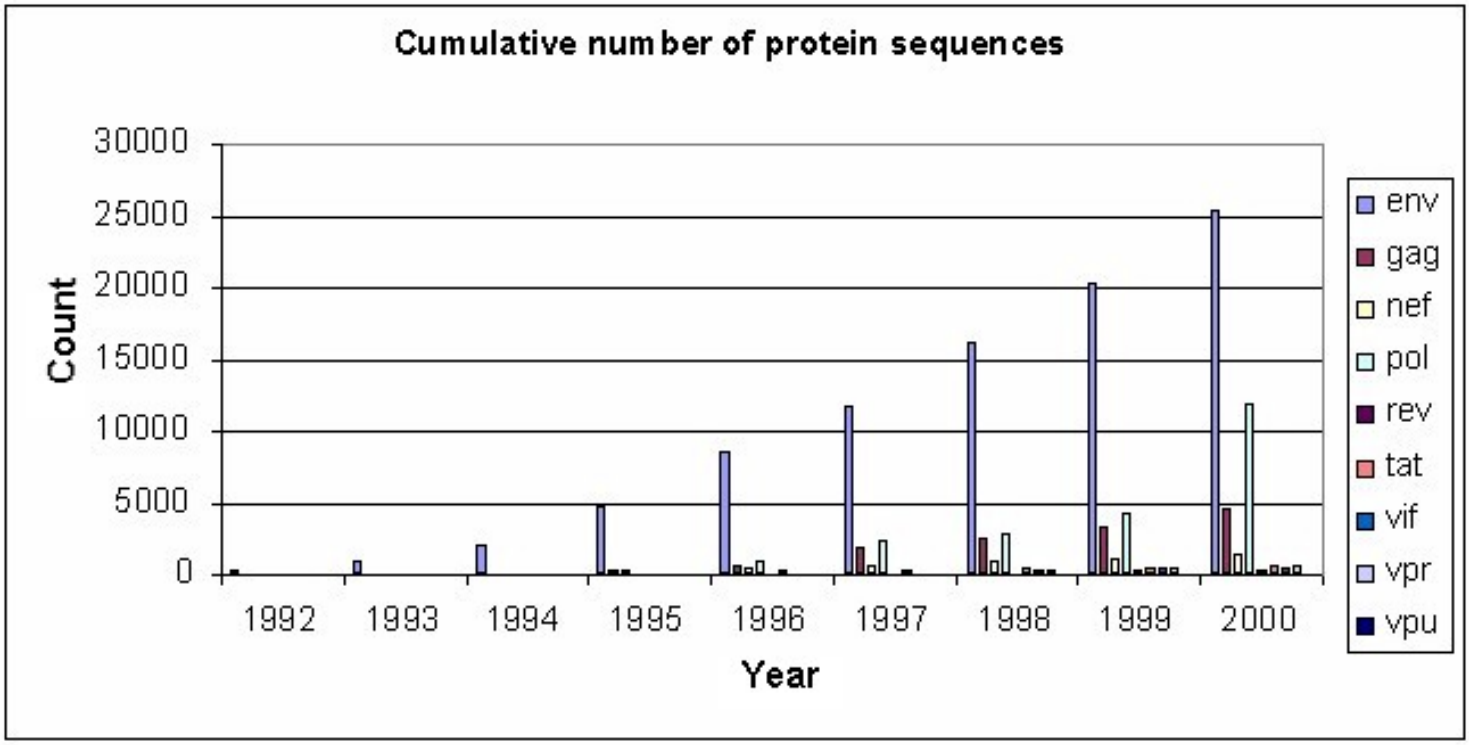
82. This work was funded by a developmental award from the Lifespan Center for AIDS Research and by an R21 award from the National Institutes of Health, Division of AIDS, to A.S. De Groot, and by an SBIR Phase I award to EpiVax, Inc.

83. The World Clade Vaccine project was declared a not-for-profit collaboration by the TB/HIV Research Laboratory and its partner in development, EpiVax Inc. on August 16, 2000. See the GAIA Vaccine Foundation site for more information.

84. For more information on HIV-1 clinical trials see: NIAID Division of AIDS, AIDS Vaccine Advocacy Coalition, International AIDS Vaccine Initiative, UNAIDS, GAIA, and AIDScience's Global Projects.

Table 1. HIV-1 clade composition of HIV-1 vaccine candidates in clinical trials.

Vaccine type	Company/ Group	Prime HIV-1 Strain	Boost HIV-1 strain
Recombinant protein (subunit)	Vaxgen	B/B	-
Recombinant protein (subunit)	Vaxgen	B/E	-
DNA/re-combinant protein	Chiron	B	B
Canarypox (Recombinant protein boost)	Aventis Pasteur	B	B
DNA	Merck	B	B
DNA/MVA	Oxford AIDS Vaccine Initative	A	A
Replicon	AlphaVax	C	C


Cumulative number of protein sequences
30000
25000
20000
15000
10000
5000
0
Count
1992
1993
1994
1995
1996
1997
1998
1999
2000
Year
env
gag
nef
pol
rev
tat
vif
vpr
vpu

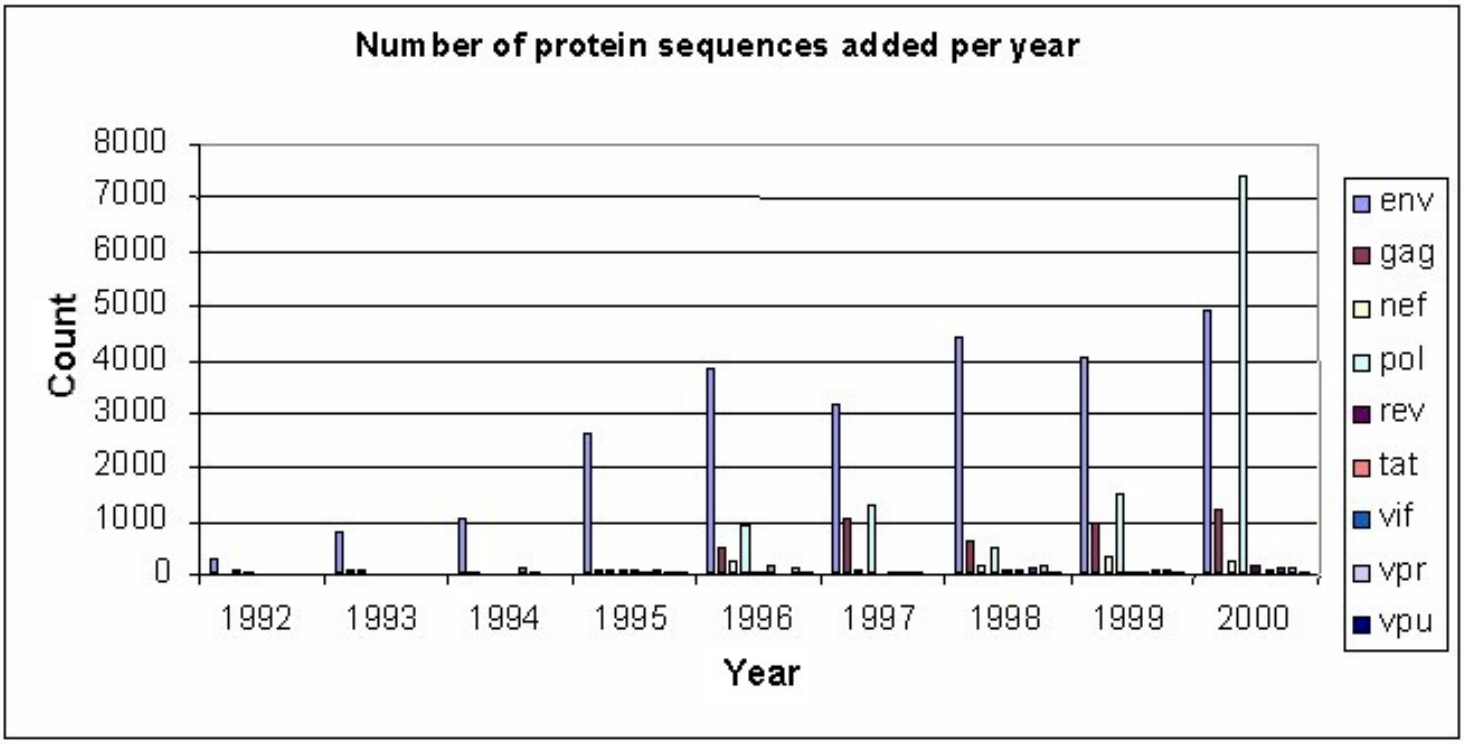

Number of protein sequences added per year
Count
8000
7000
6000
5000
4000
3000
2000
1000
0
1992
1993
1994
1995
1996
1997
1998
1999
2000
Year
env
gag
nef
pol
rev
tat
vif
vpr
vpu

Table 2. Recent studies of cross clade T cell responses.

Type of Report	Patients	Immunogen	Target HIV	Gene	Assay	Results	Authors	Journal	Year
Summary	HIV infected	Clade B	Other Clades	gag pol	Recomb'nt Vaccinia	Gag and Pol most important and conserved	Ferrari et al. (Wienholdt)	AIDS Res Hum Retroviruses 2000 Sep 20;16(14):1433-43	2000
Peptide / ELIspot	Hi-Exp Uninfected	Clade C, HIV2			Peptide/ Elispot	Cross reactive R seen against CONSERVED	Rowland-Jones SL et al. (McMichael)	Immunol Lett 1999 Mar;66(1-3):9-14	1999
Peptide / ELIspot	Hi Exposed HIV negative	Gambian, Kenyan	HIV-1, HIV-2		Peptide/ Elispot	epitopes highly conserved between B and the infecting strain	Rowland Jones (McMichael)	Dev Biol Stand 1998;92:209-14	1998
Peptide / ELIspot	Hi-Exp Uninfected	Clades A, C, D	Clade B		Peptide/ Elispot	50%had cross-clade responses to epitopes that were conserved in B Stronger responses were seen to A and C versions of B epitopes Estimates using a modified interferon-gamma Elispot assay indicate a circulating frequency of CTL to individual epitopes of between frequency of CTL to individual epitopes of between 1:3,200 and 1:50,000.	Rowland Jones et al. (McMichael)	J Clin Invest 1998 Nov 1;102(9):1758-65	1998
Whole Gene	HIV + North Am	Clade B	Clade E	gag Pol-RT Env Nef	Recomb'nt Vaccinia	most frequent to gag	Lynch et al. (Birx)	J Infect Dis 1998 Oct;178(4):1040-6	1998
	HIV+ Thais	Clade E	Clade B	gag Pol-RT Env Nef	Recomb'nt Vaccinia	most frequent to gag Pol-RT less frequent need population specific CTL data			
Whole Gene	HIV infected children	Clade B	Clade A	pol nef gag env	Recomb'nt Vaccinia	"proves single clade vaccine is ok"	Buseyne F, et al.	11: Virology 1998 Oct 25;250(2):316-24	1998
Whole Gene	HIV infected	Clade B	B(LAI) versus B(MN))	Env gp160-	Recomb'nt Vaccinia	2 had no cross reactivity against MN	Wilson SE, et al. (Sheppard)	AIDS Res Hum Retroviruses 1998 Jul 20;14(11):925-37	1998
	HIV infected	Clade B	A	gag, pol, nef	Recomb'nt Vaccinia	9 of 19 B inf also had A resp			
				gag,		12 of 19 B inf had C			

	HIV infected	Clade B	C	pol, nef	Recomb'nt Vaccinia	13 of 15 B Inf had C resp			
Whole Gene	HIV infected (A or C) HIV infected (B)	Ugandan A or C UK	A C D HIV2	p55 (gag)	Recomb'nt Vaccinia	had cross reactive responses	McAdam S. et al (Gotch)	AIDS 1998 Apr 16;12(6):571-9	1998
Whole Gene	HIV infected (African) SUBTYPED (only study so far)	A, A Env/G Gag recombinant, b	A, B, G	gag, nef	Recomb'nt Vaccinia		Durali et al.	J Virol 1998 May;72(5):3547-53	1998
Whole Gene	HIV infected Clade C	Zambian Clade B	HIV-1 Clade B	HIV gag-pol-env derived from B-clade HIV-1 (IIIB).	Recomb'nt Vaccinia	HIV clade C-infected had cross-reactive B responses against vaccinia targets	Betts, (Frelinger)	J Virol 1997 Nov;71(11):8908-11	1997
Whole Gene	HIV infected HIV infected	Clade B (US) A, C, G	A, B, C, G		Recomb'nt Vaccinia	all cross reactive with majority of vaccinia targets	Cao H, (Walker)	J Virol 1997 Nov;71(11):8615-23	1997
Vaccination	Chimpanzees	Clade B vaccine	Clade E env	env rev gag pol	Recomb'nt Vaccinia	crossreactive env responses	Boyer JD, et al. (Weiner)	Dev Biol Stand 1998;95:147-53	1998
Whole Virus	ALVAC/HIV1 gp160	Clade B (MN)	LAI HIV-1 infected	whole virus	autologous CD4+ lymphoblasts (EBV transformed B cells)	both a broad pattern of cytolysis in which viruses from all clades tested and a NARROW pattern were observed	Ferrari et al. (Weinhold)	Proc Natl Acad Sci U S A 1997 Feb 18;94(4):1396-401	1997

Table 3. Partial list of published epitopes conserved across HIV-1 strains and clades.

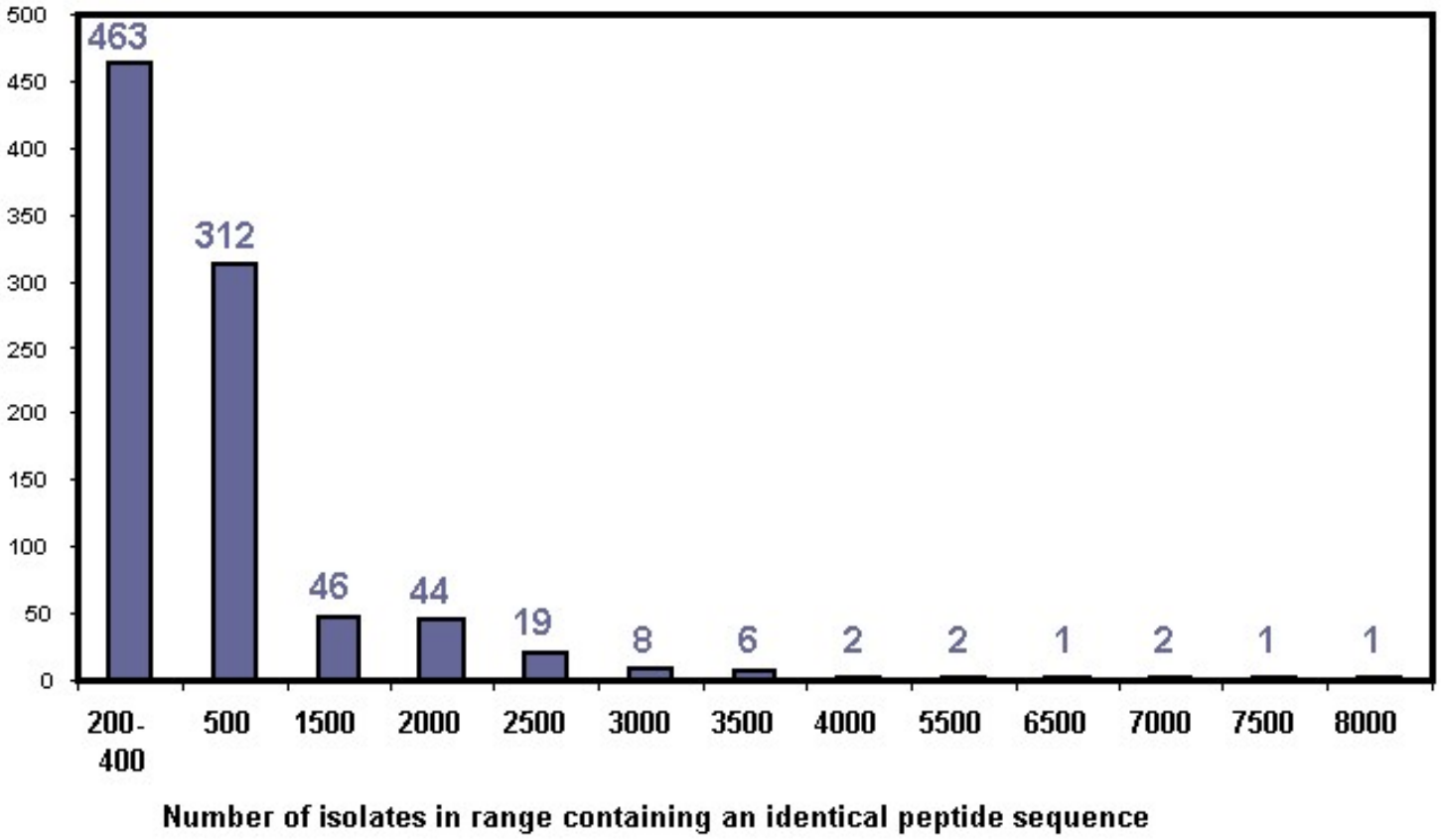
sequence	protein source	Published epitope, HLA type	Conserved in	Out of	percent CONSERVATION
NMWQEVGKAM	**env**	**A2**	2150	18482	**12%**
CTRPNNNTRK	**env**	**A2**	7576	18482	**41%**
APTKAKRRVV	**env**	**A2**	634	18482	**3%**
CTRPNNNTRK	env	A2	7526	18322	41%
TVYYGVPVWK	**env**	**A2,A3,A11,A68,B18**	422	18482	**2%**
TVYYGVPVWK	env	A2,A3,A11,A68,B18	421	18322	2%
RPVVSTQLLL	**env**	**B7**	1500	18482	**8%**
SLYNTVATLY	**gag**	**A2**	678	3249	**21%**
ELRSLYNTVA	**gag**	**A2**	502	3249	**15%**
KIRLRPGGKK	**gag**	**A3, B27, B62**	521	3249	**16%**
IRLRPGGKKK	**gag**	**A3, B62**	357	3249	**11%**
RLRPGGKKK	gag	A3,B62,B42	370	3249	11%
GPGHKARVLA	**gag**	**B7**	284	3249	**9%**
SPIETVPVKL	**pol**	**A2, B61**	195	4368	**4%**
ILKEPVHGVY	**pol**	**A2,B62**	141	4368	**3%**
AIFQSSMTK	pol	A3,A11,A68,A33	728	4368	17%
SPAIFQSSMT	**pol**	**B7**	765	4368	**18%**
QVRDQAEHLK	**pol**	**B7**	728	4368	**17%**
GPKVKQWPLT	**pol**	**B8**	654	4368	**15%**

Peptides shown in this table are published CTL epitopes that are highly conserved across clades and strains of HIV-1. They were identified by evaluating HIV-1 sequences for conserved regions, and then comparing these regions to EpiMatrix MHC binding motifs. This list of highly conserved potential MHC binders was then compared to the list of published HIV-1 sequences on the Los Alamos National Laboratory HIV-1 Sequence Database. This is only a partial list of the published epitopes containing conserved HIV-1 sequences.

Table 4. Cross-clade HLA A3-restricted epitopes selected using a bioinformatics approach.

A3 EpiMatrix Selection, Sequence	HIV-1 Prot.	EBP %	ELIspot, # patients with + Reponse	Published epitope? If so, HLA type	Conserved in	Out of	percent CONSERVATION
AVFIHNFKRK	pol	52	1	novel	173	4368	4
CTRPNNNTRK	env	51	1	A2	7526	18322	41
TVYYGVPVWK**	env	59	1	A2,A3,A11,A68,B18	421	18322	2
SLWDQSLKP	env	50	1	Human unk	514	18322	3
VSFEPIPIH	env	58	2	Human Unk	834	18322	5
LARNCRAPRK	gag	35	3	novel	125	3249	4
RLRPGGKKK	gag	34	3	A3,B62,B42	370	3249	11
QLDCTHLEGK	pol	61	2	novel	175	4368	4
QIIEQLIKK	pol	48	1	novel	86	4368	2
GIPHPAGLK	pol	20	1	novel	1125	4368	26
AIFQSSMTK**	pol	59	5	A3,A11,A68,A33	728	4368	17
KLVDFRELNK	pol	36	1	novel	1237	4368	28
KLTEDRWNK	vif	54	1	novel	411	507	81

Peptides shown in this table were identified by first searching the HIV-1 genome for conserved regions using the Conservatrix algorithm and then evaluating these regions for potential MHC binding capability, using the EpiMatrix algorithm. These peptides were selected for match to the A3 matrix motif by EpiMatrix (their scores are listed as EBP, or Estimated Binding Probabilities, in the table above). These peptides were then synthesized and studied in ELIspot assays. Of the 25 tested (not all shown), 13 (shown here) stimulated gamma interferon responses from patient PBMC in vitro. Of these 13, three were previously published A3 epitopes, and 10 are novel epitopes (not published).


500
450
400
350
300
250
200
150
100
50
0
463
312
46
44
19
8
6
2
2
1
2
1
1
200-400
500
1500
2000
2500
3000
3500
4000
5500
6500
7000
7500
8000
Number of isolates in range containing an identical peptide sequence