Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vaccine 21 (2003) 4486–4504
Vaccine
www.elsevier.com/locate/vaccine

Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach

Anne S. De Groot[a,b,*], Bill Jesdale[a], William Martin[b], Caitlin Saint Aubin[a], Hakima Sbai[a], Andrew Bosma[a], Judy Lieberman[c], Gail Skowron[d], Fadi Mansourati[e], Kenneth H. Mayer[e]

[a] *TB/HIV Research Lab, Brown University, Providence, RI 02912, USA*
[b] *EpiVax, Inc., Providence, RI, USA*
[c] *Center for Blood Research, Boston, MA, USA*
[d] *Roger Williams Hospital, Providence, RI, USA*
[e] *Brown University AIDS Program, Brown University, Providence, RI, USA*

Received 11 July 2002; received in revised form 11 March 2003; accepted 15 March 2003

Abstract

The genomic variability of HIV viruses circulating in different regions of the world has impeded the development of a globally relevant HIV vaccine. Broadly conserved HIV-1 cytotoxic T cell (CTL) epitopes were identified by screening protein sequences in the Los Alamos National Laboratory (LANL) HIV sequence database with a sequence parsing and matching algorithm (Conservatrix). Putative HIV-1 CTL epitopes were selected from this list using the epitope prediction tool EpiMatrix. *Methods*: One hundred peptides representing putative HLA A*0201, HLA A*1101, HLA A*0301, and HLA B*07 ligands conserved in many isolates of HIV-1 were synthesized. Seventy-five HLA A*0201, HLA A*1101 and HLA B*07 peptides were incubated with transport associated protein (TAP)-deficient T2 cells transfected with the gene for the corresponding human HLA molecule (HLA A*0201, HLA A*1101, and HLA B*07). Binding and stabilization of peptide–HLA complexes on the surface of the T2 cells was measured by FACS. T cell responses to the entire set of 100 peptides (HLA A*0201, HLA A*1101, HLA A*0301, and HLA B*07) were measured in ELIspot assays using PBMC from healthy HIV-1 infected subjects who possessed a matching HLA allele. *Results*: Fifty-seven (76%) of the 75 peptides tested in binding studies, including all (three of three) of the control (published) ligands bound to the T2 cells expressing the corresponding MHC molecule. Forty-three of the 100 peptides (43%) including all (four of four) of the control (published) epitopes tested in ELIspot assays stimulated γ-interferon release. Thirty-one of these 43 epitopes are novel, highly conserved HIV-1 epitopes. EpiMatrix predicted and assays confirmed MHC-restriction by more than one HLA allele for nine of the 43 novel epitopes; of these epitopes five were recognized in the context of MHC "supertypes" and four were promiscuous epitopes. *Conclusion*: Epitopes identified using this approach were conserved in a broad range of HIV-1 sequences derived from isolates obtained in Latin America, Africa, Asia, the Pacific Islands, Europe and the US. The successful identification of cross-clade epitopes by this bioinformatics approach may accelerate the development of a globally relevant HIV-1 vaccine.

Keywords: HIV-1 vaccine; Imminoformatics; Cytotoxic T cell; Epitope

1. Introduction

The vast majority of the 15,000 new HIV-1 infections that occur every day are acquired by individuals who live in developing countries, where the isolates of HIV that are transmitted are significantly different from the isolates selected for most of the HIV-1 vaccines currently under development [3]. HIV-1 subtypes or clades A, C, and D predominate in most of sub-Saharan Africa, clade E (A/E) is the most prevalent in Thailand [27], and new A/G chimeras are emerging in West Africa [4]. Recent data points to regional clusters within subtypes, such that the isolates within clade C that are circulating in South Africa are significantly different from the isolates within clade C that are circulating in India [49]. There is roughly 15–20% divergence between the nucleic acid sequences of different clades and approximately 7–12% variation within a clade.

* Corresponding author. Tel.: +1-401-277-3654; fax: +1-401-277-3656.
E-mail address: anne_degroot@brown.edu (A.S. De Groot).
URL: http://tbhiv.biomed.brown.edu.

0264-410X/$ – see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0264-410X(03)00390-6

Sequence modifications at the amino acid level may affect the immunogenicity of vaccines by modifying essential immunogenic regions, or T cell epitopes, in at least three ways: by affecting intracellular processing of the epitope, by interfering with binding (of the peptide epitope) to the MHC molecule, and by interfering with binding of the MHC–peptide complex to the T cell receptor (TCR). Changes in amino acid sequence associated with HIV-1 diversity may therefore prevent cross-clade protection against HIV-1 challenge by T cell clones raised against clade B vaccine constructs. This hypothesis is supported by numerous observations of viral escape from immune detection that have been linked to amino acid substitutions in HIV-1 T cell epitopes [19,24,36]. Thus, immunization with vaccines containing epitopes derived exclusively from clade B may not protect against challenge by HIV-1 isolates that are divergent, at the epitope level, from the vaccine strain. Current efforts to develop and test HIV vaccines based on clade B may not be relevant given the global context of the HIV epidemic [14].

One alternative to developing clade B vaccines for use in countries where non-clade B HIV-1 isolates are circulating has been to derive the vaccine from isolates representing other clades. Accordingly, HIV vaccines containing sequences from clade A [25] and clade C isolates [11] and the re-engineered vaccines containing non-clade B (E) genes or highly conserved genes such as gag (Vaxgen and Aventis Pasteur) [13] are in development. However, the same question that had been raised by critics of clade B vaccines is likely to be reiterated: will these single-clade vaccines be sufficient to protect against challenge by heterologous HIV-1 isolates?

Sequencing studies have now confirmed that isolates derived from many different clades circulate within each geographic area. For example, clade A and clade B HIV-1 isolates (and numerous chimeras) are circulating in Kenya. Both clades B and A/E isolates circulate in Thailand. HIV-1 vaccines that include highly conserved (cross-clade or multi-clade) T cell epitopes would, perhaps, be the most effective types of vaccine in the global context of the HIV epidemic. Many of these epitopes should be cytotoxic T cell (CTL) epitopes since CTL response to HIV-1 epitopes contributes to protection both prior to infection and after exposure [26,28,30,43,45]. Supertype epitopes (epitopes restricted by more than one HLA allele belonging to a family of HLA with similar restriction characteristics [48]) and promiscuous epitopes (epitopes restricted by HLA belonging to HLA families with disparate restriction characteristics [37,38]) are also important components of epitope-driven vaccines. Discovery of highly conserved sequences that are also immunogenic has been hampered by the lack of means to screen the large number of HIV-1 genome variants for epitopes. More than 65,000 HIV-1 protein sequences representing the eight clades of HIV-1 have been filed in public databases [1,5]. Direct evaluation of each overlapping peptide in this vast database of sequences would require the synthesis of millions of peptides and blood samples from thousands of volunteers.

A more rapid approach to identifying highly conserved HIV-1 epitopes for a novel HIV vaccine using bioinformatics tools and in vitro assays is described in this article. This approach may accelerate the development of a globally accessible, globally relevant cross-clade HIV-1 vaccine.

2. Methods

2.1. Epitope mapping

The entire list of HIV-1 sequences available in the 1997 version the Los Alamos National Laboratory (LANL) HIV sequence database site was obtained [5]. Conservatrix, a sequence matching and counting tool that compares the sequence of every 10-amino-acid-long-peptide in the sequence database for identity with every other 10-amino acid sequence, was used to identify broadly conserved (across clades) HIV-1 epitopes [16]. Conservatrix can be configured to allow amino acid substitution at non-anchor positions. However, for this study, the program was configured to search for peptides based on absolute conservation (no amino acid substitutions at any position). EpiMatrix, a matrix-based epitope search algorithm, was used to score the conserved ligands. The EpiMatrix method for selecting T cell epitopes has been described [17].

The criteria for selecting study peptides were as follows: (1) Maximum conservation and predicted immunogenicity—the 1000 most conserved peptides were derived from the 1997 LANL database using Conservatrix, evaluated for potential immunogenicity by EpiMatrix, and ranked by EpiMatrix score. (2) By proportional representation among major HIV proteins—the number of peptides selected was proportional to the entire protein length of each of the eight major proteins. (3) By choosing the highest scoring, 25 peptides for four MHC alleles (A2, A3, A11, and B7). The EpiMatrix matrices used corresponded to HLA subtypes HLA A*0201, HLA A*0301, HLA A*1101, and HLA B7 (all subtypes). EpiMatrix score was the primary criteria for selecting epitope candidates from the list of 1000 highly conserved peptides. Based on previous analyses, higher EpiMatrix scores suggest greater MHC binding potential [29]. In this manner, 100 of the most conserved potential A2-, A3-, A11- and B7-restricted ligands for each of eight HIV-1 proteins (env, gag, pol, rev, tat, vif, vpr, and vpu) were selected.

Conservatrix analysis was repeated on these proteins in May 2002, using a new HIV-1 sequence database derived from all Genbank HIV-1 entries as of April 2002. Information on the “country of origin” of the HIV-1 sequence was available for approximately 25% of the sequences in the Genbank database.

Twenty-four of the peptides selected for each allele were neither published CTL epitopes nor known MHC ligands at the time the peptides were selected (1997). A 25th peptide for each allele was selected based on the three study criteria listed above in addition to having been previously identified and published as a CTL epitope prior to this study [9]. This published epitope served as the positive control for the in vitro binding assays and CTL assays. The A2 published epitope selected for this study was SLYNTVATLY (a 10-mer variation of the published epitope) [30]; the A3 and the A11 published epitope was TVYYGVPVWK (a supertype epitope) [48]; and the B7 published epitope was GPGHKARVLA [24].

2.2. Peptide synthesis

Peptides corresponding to the epitope selections were prepared by 9-fluoronylmethoxy-carbonyl (Fmoc) synthesis on an automated Rainen Symphony/Protein Technologies synthesizer (Synpep, Dublin, CA, purity 90% by HPLC).

2.3. Binding assays

In vitro evaluation of MHC binding was performed by measuring the ability of exogenously added peptides to stabilize the class I MHC/beta 2 microglobulin structure on the surface of transport associated protein (TAP)-deficient T2 cell lines [34]. The T2-A2 cell line, expressing HLA A2 [47], was generously provided by Dr. C. Brander (Massachusetts General Hospital, Charlestown, Massachusetts). The T2-A11 cell line, expressing HLA A11 [52], was generously provided by Dr. J. McNicholl (Centers for Disease Control, Atlanta, Georgia). The T2-B7 cell line, expressing HLA B7 [47], was generously provided by Dr. R. Koup (NIH). Monoclonal antibody lines were also obtained: the BB7.2 cell line, secreting anti-HLA A2 [39], was obtained from ATCC; the A11.1M [22] and the Auf 5.13 cell lines [15], secreting anti-HLA A11 IgG, were obtained from Dr. J. McNicholl; and the ME-1 cell line, secreting anti-HLA B27, B7 and Bw22 IgG, was obtained from ATCC. Binding assays were not performed for the HLA3 peptides.

In vitro evaluation of MHC stabilization by the candidate peptide was performed as previously described by Ljunggren et al. [34], Nijman et al. [38], and Brander et al. [8]. Fluorescence of viable T2 cells (a marker of peptide binding) was measured at 488 nm on a FACScan flow cytometer (Becton-Dickinson, New Jersey) and binding results were interpreted as previously described [18]. Briefly, peptides qualified as binders if the mean linear fluorescence (MLF, averaged over 10,000 events) increased over background by 10% or more, with $P < 0.05$, for at least two of the three sets of wells containing different concentrations of peptide (10, 20 and/or 200 μg/ml). Each concentration of peptide was assayed in triplicate wells. Each binding assay was repeated four times; peptides were only considered to be binders if the above criteria were met in three out of the four repeated assays.

2.4. Subjects

HIV-1 infected subjects with viral loads below 10,000 copies/ml and absolute CD4 T cell counts above 200 cells per Cl were recruited from clinics located in Rhode Island (Roger Williams Hospital and Pawtucket Memorial Hospital) or in Massachusetts. HIV-seronegative controls were recruited from among laboratory personnel. All subjects were recruited in accordance with the Office for Human Research Protections (OHRP) guidelines, with Institutional Review Board (IRB) approval and signed informed consent. A total of 23 HIV-infected subjects' responses were evaluated in 34 ELIspot assays (Table 1). In four cases, subjects' PBMC were tested for responses to peptides restricted by more than one HLA allele.

2.5. HLA typing

HLA typing was performed using DNAzol (Gibco/Life Technologies) and HLA SSP ABC Typing Kits (One Lambda, Inc.) (Table 1). The kit that was available at the time the assays were performed did not permit subtyping. In some cases, the HLA type could not be resolved. These cases are designated either with multiple alleles (e.g. 14/8) where differentiation between alleles could not be determined with certainty or with "?" where no identifiable HLA type could be discerned.

2.6. ELIspot assays

Peripheral blood mononuclear cells (PBMC) were separated from heparinized peripheral blood samples using Lymphoprep (Nycomed Pharma) density centrifugation. The PBMC were pre-incubated with peptide (peptide stimulation), or PHA (PHA stimulation) or both (Peptide/PHA stimulation) for 5–10 days according to published protocols. In all cases, 20 U/ml IL2 (Sigma) were added 2 or 3 days after cultures were initiated and every 2 days thereafter. PBMCs were harvested after stimulation and plated at 10,000–100,000 cells per well in an ELIspot plate (Millipore, Inc.) which was pre-coated with Mouse anti-human IFN-γ monoclonal antibody (Pharmingen), 15 μg/ml. All ELIspot assays were performed using a single peptide per well. At the time of the final assay, target peptides were added at 10 μg/ml concentration to wells and incubated for 18–20 h. Autologous PBMC or T2 cells expressing the relevant MHC molecule were used as antigen presenting cells; these target cells were pulsed with peptides in the assay wells. Cells were also plated with PHA, 10 μg/ml, for the positive control wells, and with no peptide added for the negative control wells. Cells were discarded and the plate was washed with 0.05% Tween 20/PBS (Gibco/Life

Table 1
Study subject cohort

	Peptides tested	ELIspot responses	HLA A		HLA B		HLA C	
A2								
H0204R	A2	0	A2	?	B40	B57	Cw03	Cw07
H0007M	A2	0	A2	A29	B15	B44	Cw03	Cw16
902991	A2	4	A2	A30	B39	?	ND	
H0014M	A2	3	A2	A1	B8	?	Cw16	
517001	A2	2	A2	A3	B44	?	Cw05	Cw07
H0023M	A2	2	A2	A30	B35	B49	Cw04	Cw07
829001	A2 (B7)	1 (B7)	A2	A3	B7	B58	Cw07	
906002	A2 (A3)	1 (A2)	A2	A3	B8	B51	Cw01	Cw07
Average no. of A2 responses/responders		2						
A11								
H0018M	A11	0	A11	A2	B44	B51	Cw05	Cw15
H0201R	A11	0	A11	A68	B42	B45	Cw16	Cw17
523001	A11 (A3)	4 (A11)	A11	A3	B14	B51	Cw08	Cw13
606001	A11	5	A68	A1	B15	B40	Cw03	?
202001	A11	3	A68	A25	?	?	?	?
718001	A11	4	A11	A68	B42	B45	Cw16	Cw17
Average no. of A11 responses/responders		4						
B7								
228	B7	8	A3	A24	B7	B38	Cw07	Cw12/13
H0015M	B7	2	A1	A3	B7	B8	Cw03	Cw07
829001	(A2) B7	1	A2	A3	B7	B58	Cw07	?
411001	B7	4	A29	A30	B8	B44	Cw07	Cw16
Average no. of B7 responses/responders		4						
A3								
H0012	A3	0	A32	A3	B14	B40	Cw08	
H0013M	A3	0	A66	A3	B35	B41	Cw04	Cw17
H0032M	A3	0	A33	A3	B14	B15	Cw07	Cw08
419001	A3	11	A23	A3	B49	B57	Cw06/07/12/13	?
411002	A3	6	A24	A3	B27	B57	Cw13	Cw18
517001	A3	6	A2	A3	B44	?	Cw05	Cw07
906002	(A2) A3	5 (A3)	A2	A3	B8	B51	Cw01	Cw07
509002	A3	4	A26	A3	B8	B52	Cw07/12/13	?
523001	(A11) A3	3 (A3)	A11	A3	B14	B51	Cw08	Cw13
Average no. of A3 responses/responders		2						

Technologies). A secondary antibody, biotinylated mouse anti-human IFN-γ monoclonal antibody (Pharmingen) was added to the wells for 3–4 h at 1 μg/ml, then washed as before. Streptavidin-alkaline phosphatase (Pharmingen) was added for a 1 h incubation, with subsequent washes as before. Lastly, BCIP-NBT buffer (Sigma) was added for color development for 45 min. The plate was washed several times with deionized water and allowed to dry thoroughly.

ELIspot assays were performed in triplicate. Spots were counted using a dissecting microscope (Leica, Inc.). ELIspot wells that contained a number of spots that was at least twice the average number of background spots and also contained greater than 20 spots per 1 million cells (equivalent to a ratio of 1 responder per 50,000 PBMC, above background) were considered positive and included in the averaged responses for peptide, according to the criteria described by Schmechel et al. [44].

3. Results

3.1. Peptide selection

One hundred peptides were selected based on the criteria described in methods. The peptides scored in the same range as a set of known HLA A2, HLA A11, HLA A3, and HLA B7 binders archived in a publicly-accessible MHC binding peptide database [40]. EpiMatrix scores for a set of 10,000 random peptides [2], the set of Rammensee known ligands, and the study set of 25 HLA-restricted peptides are shown, for each of the alleles, in Fig. 1a–d.

A summary of the results for the 75 peptides tested in binding assays and 100 peptides tested in ELIspot assays is presented in Table 2. Fifty-seven (76%) of 75 peptides tested in binding studies bound to the T2-HLA cells expressing the corresponding MHC molecule, including all (three of three) of the control (published) ligands. Forty-three of

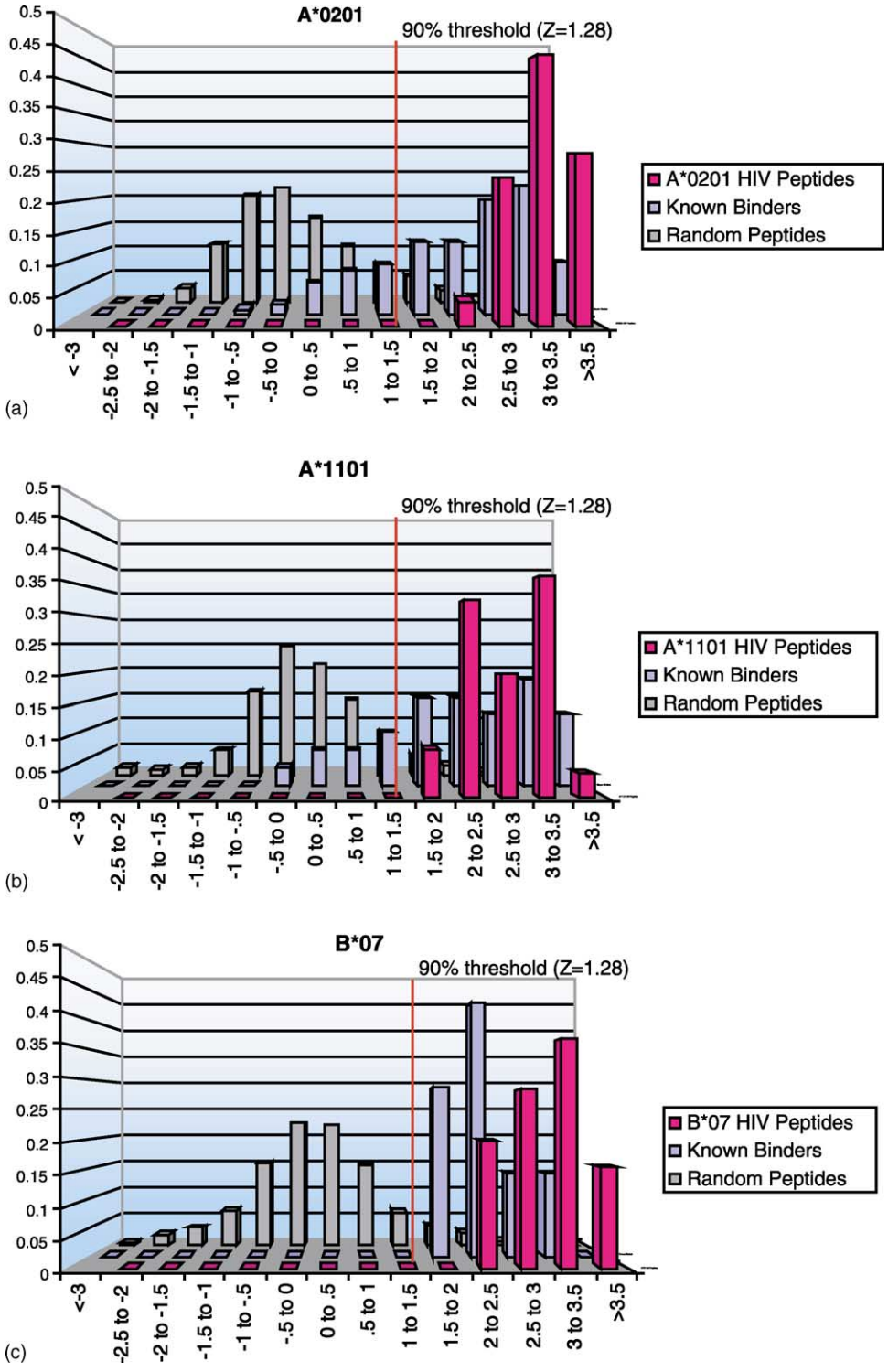


Fig. 1. Scores for study peptides contrasted with published HLA ligands and random peptides. EpiMatrix scores for a set of 10,000 random peptides of natural amino acid composition, the set of Rammensee's large database of known ligands searchable by HLA allele ligands [SYFPEITHI], and the study set of 25 HLA-restricted peptides are shown in (a–d). The 10,000 random peptides of natural amino acid composition were derived from the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (Randseq, http://www.expasy.ch/tools/randseq.html). (a) Scores for 25 A*0201-restricted epitopes selected for this study are contrasted with scores for 10,000 random peptides and 280 known A*0201 ligands. (b) Scores for 25 A11-restricted epitopes are contrasted with scores for 10,000 random peptides and 32 known A11 ligands. (c) Scores for 25 B7-restricted epitopes are contrasted with scores for 10,000 random peptides and seven known B7 ligands. (d) Scores for 25 A3-restricted epitopes are contrasted with scores for 10,000 random peptides and 44 known A3 ligands. As is shown in these figures, the set of peptides selected for this study scored either above or within the range of EpiMatrix scores received by known binders and epitopes, and scored well above the mean of scores for a set of 10,000 random peptides, for the corresponding HLA allele.

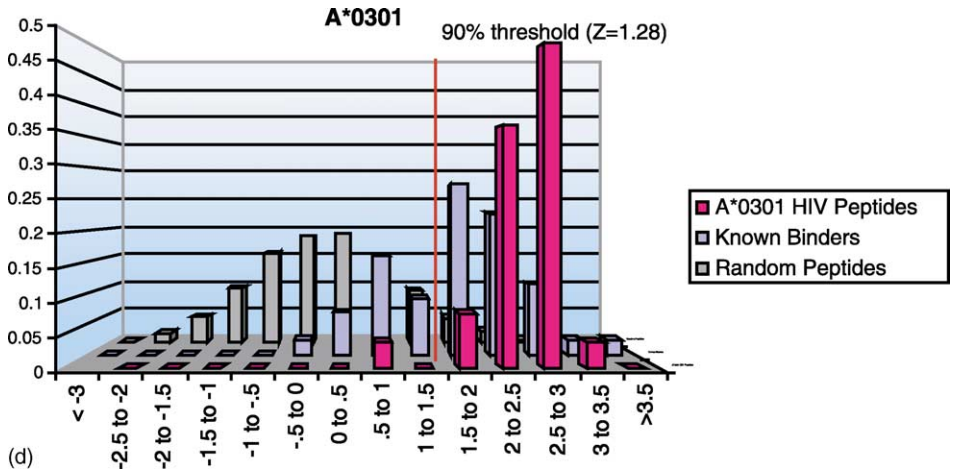


Fig. 1. (*Continued*).

100 peptides (43%) including all (four of four) of the control (published) epitopes tested in ELIspot assays stimulated γ-interferon release. The 31 new epitopes identified in the course of this study are listed in Table 3. Detailed MHC binding results and ELIspot results for each of the study peptides are provided in Table 4a–d.

EpiMatrix predicted and in vitro assays confirmed MHC-restriction by more than one HLA allele for eight of the novel epitopes; of these epitopes five were recognized in the context of MHC "supertypes" and three were promiscuous epitopes. Eighteen of the 43 confirmed epitopes in this study (and 12 of the 32 novel epitopes) were completely conserved in more than 1 in 10 (10%) HIV-1 protein sequences in the Genbank database (Table 3).

3.2. A2 peptide results

3.2.1. A2 peptide binding

Results for the 25 A2 peptides are shown in Table 4a. Thirteen of the 25 A2 HIV-1 peptides (52%) selected by Conservatrix and EpiMatrix for this study bound to T2 cells expressing HLA A2 (T2-A2), including the previously published A2 binder (SLYNTVATLY, a 10-mer variation on SLYNTVATL). Peptides 1257 (NMWQEVGKAM, env) and 1249 (ILKEPVHGVY, pol), two published A2-restricted epitopes, were not shown to bind to A2 cells by our criteria, although both did stimulate γ-interferon release by PBMC in ELIspot assays (see below).

Table 2
Summary of world clade binding and ELIspot assays

Allele	No. tested	No. of binders	%Binders	No. of ELIspot	%ELIspot
A2	25	13	52	6	24
A11	25	23	92	10	40
B7	25	21	84	11	44
A3	25	N.D.	N.D.	16	64
All peptides	100	57	76	43	43

In contrast, in negative control assays, none of eight non-A2-restricted peptides (B7 and A11 peptides synthesized for this study) stabilized the HLA A2 MHC molecule on T2-A2 cells (data not shown).

3.2.2. A2 ELIspot results

PBMC from eight subjects who possessed the A2 allele were obtained. Six of the 25 A2 peptides, including the control peptide, stimulated γ-interferon secretion from HIV-infected subjects PBMC in vitro (24%). Two study subjects, H0007M and H0204R (both HLA A2) responded to none of the selected peptides, including the control peptide SLYNTVATLY, but their cells did release γ-interferon in response to stimulation by PHA (data not shown). PBMC from six subjects responded to at least one A2 peptide. The average number of responses per subject, excluding subjects who did not respond to any peptides, was two (Table 1).

Subject 0902991 responded to four peptides (1256 (NLWVTVYYGV, env), 1257 (NMWQEVGKAM, env), 1249 (ILKEPVHGVY, pol), and 1237 (DLADQLIHLY, vif)). 1237 was the only peptide that appeared to bind in T2-A2 assays. Subject H0014M (A1, A2, B8, ?, Cw16, ?) responded to three peptides (1261 (SLYNTVATLY, gag), 1247 (HLKTAVQMAV, pol), and 1237). All three of these peptides were confirmed A2 binders.

Subject 0517001 and H0023M each responded to two HLA A2 study peptides. Both subjects' PBMC responded to peptide 1249. In addition, subject 0517001 responded to 1257, while subject H0023M responded to 1261 (control epitope). Subjects 0829001 and 0906002 each responded to one novel epitope, 1242 (EMMTACQGVG, gag, a published A2 epitope but not a T2-A2 binder) and 1261, (control epitope), respectively.

Three new A2-restricted epitopes were identified. Three published HLA A2-restricted epitopes were reconfirmed, one of which was published prior to the study (Table 4a). One subject (0829001, A2, A3, B7, B58, Cw07) responded to peptide (1242 (EMMTACQGVG, gag)). Since 1242 did

Table 3
New and published epitopes described in the world clade study

Allele	Peptide no.	Sequence	Protein	EBP (%)	Binder	CTL ELIspot 10E6	No. of responders	Supertype (ST)/promicuous (P)/immunodominant (ID)	Comments
New epitopes									
A2	1256	NLWVTVYYGV	env	84	N	388	1		A2
A2	1237	DLADQLIHLY	vif	54	Y	190	1		A2
A2	1247	HLKTAVQMAV	pol	82	Y	230	1		A2
A11	1284	VTENFNMWKN	env	17	Y	81, 50	2	ST	A11 binder, A11, A68
A11	1281	TTPDKKHQKE	pol	60	Y	131, 83	2	ST	A11 binder, A11, A68
A11	1264	AVFIHNFKRK	pol	53	Y	1478, 59, 50	3	ST, ID	A11 binder, A3, A11, A68
A11	1274	NTPVFAIKKK	pol	53	Y	2283, 1600	2	ST, ID	A11 binder, A11, A68
A11	1276	PGMDGPKVKQ	pol	52	Y	61	1	P	A11 binder, A11 or A68, Published B8
A11	1267	FTTPDKKHQK	pol	36	N*	71	1	ST	A11 binder, A11, A68
A11	1277	PTVLESGTKE	rev	16	Y	81	1		A68, not bound to A11
A11	1279	TACNNCYCKK	tat	74	Y	207	1		A68, not bound to A11
B7	1298	KPVVSTQLLL	env	46	Y	22, 172	2	P	Promiscuous (B8 and B7), B8 motif
B7	1309	TPQDLNMMLN	gag	31	Y	48	1		B7
B7	1302	QPDKSESELV	pol	36	N	44	1		B7
B7	1307	SPIETVPVKL	pol	12	Y	65, 31	2	P	A2, B7, B8, B61
B7	1294	IPAETGQETA	pol	8	Y	87	1		B7
B7	1296	KPPLPSVKKL	vif	23	Y	45	1		B7
B7	1300	LVILAIVALV	vpu	6	N	103	1		B7
A3	1325	HSFNCGGEF	env	76	N.D.	160	1		Pulsed on T2-A3 targets
A3	1329	VSFEPIPIH	env	58	N.D.	207, 150	2		Published (unknown HLA)
A3	1265 of A11	CTRPNNNTRK	env	51	N.D.	140	1	P	A2
A3	1328	SLWDQSLKP	env	50	N.D.	103, 1370	2	ID	Published (unknown HLA)
A3	1331	LARNCRAPRK	gag	35	N.D.	973, 1247, 703	3	ID	Only tested in A3
A3	1264 of A11	AVFIHNFKRK	pol	52	N.D.	67, 17	2		A3
A3	1335	QLDCTHLEGK	pol	61	N.D.	89, 73	2		A3
A3	1336	QIIEQLIKK	pol	48	N.D.	87	1		A3
A3	1340	KLVDFRELNK	pol	36	N.D.	453, 33	2		A3
A3	1337	GIPHPAGLK	pol	20	N.D.	60	1		A3
A3	1341	KILYQSNPY	rev	36	N.D.	20	1		A3
A3	1342 (1279 10-mer)	TACNNCYCK	tat	46	N.D.	33	1		A3
A3	1344	KLTEDRWNK	vif	54	N.D.	63	1		A3
Published epitopes—confirmed									
A2	1257	NMWQEVGKAM	env	50	N	333, 72	2	A2	A2, (other unknown HLA)
A2	1261	SLYNTVATLY	gag	78	Y	127, 259, 204	3	A2	A2
A2	1249	ILKEPVHGVY	pol	96	N	2380, 247, 117, 172	4	P	A2, Bw62
A11	1283	TVYYGVPVWK	env	18	Y	696, 96	2	P	A2, A3, A11, A6801, B18
A11	1271	IRLRPGGKKK	gag	43	Y	59	1	P	A3, B62, Bw62, A11 binder
B7	1305	RPVVSTQLLL	env	41	Y	31	1	ST	Published B7, responses by B8 subject
B7	1291	GPGHKARVLA	gag	28	Y	500, 241	2	B7	B7 (control in this assay)
B7	1306	SPAIFQSSMT	pol	13	Y	129	1	B7	B7
B7	1293	GPKVKQWPLT	pol	27	Y	78, 311	2	P	B7, B8
A3	1283 of A11 (KB)	TVYYGVPVWK	env	59	N.D.	80	1	P, ST	A2, A3, A301, A11, A6801, B18
A3	1332	RLRPGGKKK	gag	34	N.D.	1500, 1900, 770	3	P, ID	A3, A3.1, B62, Bw62, B42
A3	1339	AIFQSSMTK	pol	59	N.D.	17, 23, 289, 107, 200	5	ST	A3, A3.1, A11, A6801, A33
Published epitopes—not confirmed									
A2	1238	DLNTMLNTVG	gag	65	N			Not confirmed	B14, other
A2	1241	ELRSLYNTVA	gag	71	Y			Not confirmed	A2, other
A11	1272	KIRLRPGGKK	gag	36	Y			Not confirmed	A3, B27, B62, Bw62 (A11 binder)
B7	1295	KPCVKLTPLC	env	27	Y			Not confirmed	Published (unknown HLA)
B7	1308	TPQDLNTMLN	gag	31	Y			Not confirmed	B14, (unknown HLA)

New HIV-1 epitopes determined using the methods described in this report are listed in the table. The estimated binding probability (EBP), level of response (number of CTL ELIspots/10E6 cells) and the type of response is also provided, along with the type of response (immunodominant, supertype, or promiscuous, see text). Those epitopes that were predicted and have since been published by other laboratories (and peptides that have been published but were not confirmed, in this study) are also listed.

Table 4
Conservation of peptides

Peptide no.	Sequence	Protein	EBP (%)	Binder	2002 numerator (ABS)	2002 numerator (2 AA)	2002 denominator	2002 total (ABS) (%)	2002 total (2 AA) (%)
(a) HLA A2 peptides									
1261 (CTR)	SLYNTVATLY								
1256	NLWVTVYYGV	env	84	N	440	922	32684	1	3
1244	GIKQLQARVL	env	74	Y	1133	1571	32684	3	5
1259	QLQARVLAVE	env	65	N	837	1554	32684	3	5
1253	KLTPLCVTLN	env	51	N	1625	2480	32684	5	8
1257	NMWQEVGKAM	env	50	N	4839	8495	32684	15	26
1260	QMHEDIISLW	env	48	N	745	1135	32684	2	3
1239	DMRDNWRSEL	env	44	N	1945	2832	32684	6	9
1261	SLYNTVATLY	gag	78	Y	975	3360	6269	16	54
1241	ELRSLYNTVA	gag	71	Y	655	3121	6269	10	50
1238	DLNTMLNTVG	gag	65	N	1178	1845	6269	19	29
1242	EMMTACQGVG	gag	65	N	1244	1364	6269	20	22
1255	LLWKGEGAVV	pol	99	Y	482	492	21075	2	2
1249	ILKEPVHGVY	pol	96	N	986	1381	21075	5	7
1252	KLLWKGEGAV	pol	89	Y	466	491	21075	2	2
1247	HLKTAVQMAV	pol	82	Y	549	590	21075	3	3
1245	GLKKKKSVTV	pol	76	Y	8819	13621	21075	42	65
1240	ELHPDKWTVQ	pol	72	Y	5633	6172	21075	27	29
1250	KIEELRQHLL	pol	71	N	6446	12951	21075	31	61
1251	KLLRGTKALT	pol	62	N	1308	3349	21075	6	16
1258	QILVESPTVL	rev	69	N	57	193	5666	1	3
1243	FLNKGLGISY	tat	10	Y	66	252	1139	6	22
1237	DLADQLIHLY	vif	54	Y	103	660	905	11	73
1246	HIPLGDARLV	vif	46	Y	206	770	905	23	85
1254	LLEELKNEAV	vpr	89	Y	416	893	914	46	98
1248	ILAIVVWTIV	vpu	90	Y	60	533	728	8	73
(b) HLA A11 peptides									
1283 (CTR)	TVYYGVPVWK								
1282	TVQCTHGIKP	env	44	Y	926	3149	32684	3	10
1266	FAILKCNDKK	env	33	Y	462	1160	32684	1	4
1280	TITLPCRIKQ	env	23	Y	1935	8787	32684	6	27
1265	CTRPNNNTRK	env	18	Y	12642	17499	32684	39	54
1283	TVYYGVPVWK	env	18	Y	794	817	32684	2	2
1284	VTENFNMWKN	env	17	Y	793	1038	32684	2	3
1271	IRLRPGGKKK	gag	43	Y	748	1307	6269	12	21
1272	KIRLRPGGKK	gag	36	Y	991	1385	6269	16	22
1273	LVQNANPDCK	gag	18	Y	1524	1637	6269	24	26
1263	ARNCRAPRKK	gag	11	N*	535	1058	6269	9	17
1281	TTPDKKHQKE	pol	60	Y	7215	12395	21075	34	59
1264	AVFIHNFKRK	pol	53	Y	438	507	21075	2	2
1274	NTPVFAIKKK	pol	53	Y	11098	13953	21075	53	66
1276	PGMDGPKVKQ	pol	52	Y	9505	11461	21075	45	54
1278	QVRDQAEHLK	pol	49	Y	436	581	21075	2	3
1270	IPHPAGLKKK	pol	41	Y	8946	13362	21075	42	63
1268	GIPHPAGLKK	pol	37	Y	11040	13689	21075	52	65
1267	FTTPDKKHQK	pol	36	N*	6435	12223	21075	31	58
1275	NVTENFNMWK	pol	24	Y	34	45	21075	0	0
1277	PTVLESGTKE	rev	16	Y	54	126	5666	1	2
1279	TACNNCYCKK	tat	74	Y	36	780	1139	3	68
1269	IKPPLPSVKK	vif	48	Y	72	440	905	8	49
1262	ALTALITPKK	vif	32	Y	50	531	905	6	59
1286	WTLELLEELK	vpr	8	Y	652	734	914	71	80
1285	WTIVFIEYRK	vpu	16	Y	55	382	728	8	52
(c) HLA B7 peptides									
1291 (CTR)	GPGHKARVLA								
1298	KPVVSTQLLL	env	46	Y	1387	4612	32684	4	14
1305	RPVVSTQLLL	env	41	Y	2729	4588	32684	8	14
1295	KPCVKLTPLC	env	27	Y	1307	1484	32684	4	5
1304	RCSSNITGLL	env	20	N	3695	9184	32684	11	28
1288	APTKAKRRVV	env	18	Y	1050	2272	32684	3	7
1303	RAIEAQQHLL	env	17	Y	878	1641	32684	3	5
1290	GPCKNVSTVQ	env	15	Y	760	2570	32684	2	8
1309	TPQDLNMMLN	gag	31	Y	489	1802	6269	8	29

Table 4 (*Continued*)

Peptide no.	Sequence	Protein	EBP (%)	Binder	2002 numerator (ABS)	2002 numerator (2 AA)	2002 denominator	2002 total (ABS) (%)	2002 total (2 AA) (%)
1308	TPQDLNTMLN	gag	31	Y	1143	1806	6269	18	29
1291	GPGHKARVLA	gag	28	Y	677	1196	6269	11	19
1287	APRKKGCWKC	gag	23	Y	1155	1540	6269	18	25
1302	QPDKSESELV	pol	36	N	257	596	21075	1	3
1293	GPKVKQWPLT	pol	27	Y	9479	11594	21075	45	55
1306	SPAIFQSSMT	pol	13	Y	9096	13261	21075	43	63
1307	SPIETVPVKL	pol	12	Y	9941	11150	21075	47	53
1250 of A2	KIEELRQHLL	pol	11	Y	6446	11476	21075	31	54
1278 of A11	QVRDQAEHLK	pol	11	Y	436	583	21075	2	3
1294	IPAETGQETA	pol	8	Y	520	582	21075	2	3
1301	LVSQIIEQLI	pol	8	N	277	540	21075	1	3
1299	LPPLERLTLD	rev	47	Y	152	305	5666	3	5
1292	GPKESKKKVE	tat	9	Y	50	182	1139	4	16
1297	KPPLPSVTKL	vif	27	Y	141	849	905	16	94
1296	KPPLPSVKKL	vif	23	Y	166	849	905	18	94
1289	FPRIWLHSLG	vpr	44	Y	100	821	914	11	90
1300	LVILAIVALV	vpu	6	N	28	76	728	4	10
(d) HLA A3 peptides									
1283 (CTR)	TVYYGVPVWK								
1325	HSFNCGGEF	env	76	N.D.	7919	11307	32684	24	35
1326	TLFCASDAK	env	74	N.D.	940	1065	32684	3	3
1327	HSFNCRGEF	env	70	N.D.	911	11200	32684	3	34
1283 of A11 (KB)	TVYYGVPVWK	env	59	N.D.	794	1038	32684	2	3
1329	VSFEPIPIH	env	58	N.D.	1196	2409	32684	4	7
1265 of A11	CTRPNNNTRK	env	51	N.D.	12642	23911	32664	39	73
1328	SLWDQSLKP	env	50	N.D.	930	1201	32684	3	4
1330 (1261 10-mer of A2)	SLYNTVATL	gag	59	N.D.	1000	3511	6269	16	56
1331	LARNCRAPRK	gag	35	N.D.	313	1115	6269	5	18
1332	RLRPGGKKK	gag	34	N.D.	770	1043	6269	12	17
1333	ILDIRQGPK	gag	28	N.D.	1272	2024	6269	20	32
1335	QLDCTHLEGK	pol	61	N.D.	475	501	21075	2	2
1339	AIFQSSMTK	pol	59	N.D.	8132	13178	21075	39	63
1334	LVDFRELNK	pol	57	N.D.	12284	13799	21075	58	65
1264 of A11	AVFIHNFKRK	pol	52	N.D.	438	507	21075	2	2
1336	QIIEQLIKK	pol	48	N.D.	427	601	21075	2	3
1338	KVYLAWVPA	pol	42	N.D.	323	589	21075	2	3
1340	KLVDFRELNK	pol	36	N.D.	12242	13665	21075	58	65
1337	GIPHPAGLK	pol	20	N.D.	11785	13122	21075	56	62
1341	KILYQSNPY	rev	36	N.D.	148	203	5666	3	4
1342 (1279 10-mer of A11)	TACNNCYCK	tat	46	N.D.	100	359	1139	9	32
1344	KLTEDRWNK	vif	54	N.D.	581	823	905	64	91
1343 (1262 10-mer of A11)	ALTALITPK	vif	16	N.D.	66	442	905	7	49
1286 of A11	WTELLEELK	vpr	21	N.D.	652	746	914	71	82
1345	RLIDRIRER	vpu	46	N.D.	127	521	728	17	72

The peptide sequences of 96 novel World Clade peptides and four control peptides are shown in the table. Part (a) shows the peptides selected for the study that were predicted using the EpiMatrix A2 matrix; part (b) shows the peptides selected for the study using the A11 matrix; part (c) shows the peptides selected using the B7 matrix; and part (d) shows the peptides selected using the A3 matrix. The peptide sequences (including the control peptide), protein source, estimated binding probability (based on EpiMatrix score), and binding assay results for each of the peptide sequences, are shown in the first four columns. Binding assays were not performed (N.D.) for the A3 peptides. The number of HIV-1 sequences in Genbank (April 2001) database that were absolutely conserved (contained the exact same sequence) in 2002 is shown in the column marked "numerator (ABS)" and the number of sequences on file for each protein in 2002 is shown as the "denominator". In the column marked "2002 numerator (2 AA)", the number of peptides that varied by as many as two amino acids but were still predicted to be immunogenic based on EpiMatrix score is shown. The final two columns show the percent conservation (number conserved/2002 denominator) for 2002 (ABS) and 2002 (2 AA), respectively. EpiMatrix score was the main criteria for inclusion in the original selections. The most highly conserved peptides were not always the ones that scored highest on EpiMatrix. Therefore it is not surprising that only 45 of the peptides were absolutely conserved in more than 10% of their respective protein database. Allowing two amino acids to vary while maintaining EpiMatrix score as the main inclusion criteria (an adjustment that reflects the fact that binding to MHC, and not T cell epitope interaction, is the rate-limiting factor for immunogenicity) increased the relative conservation of the peptides. Using the data generated by this adjustment, 63 peptides were found to be conserved in more than 10% of their respective protein databases. The peptides for this study were originally selected in 1997, at a time when the collective databases (Los Alamos National Laboratory and Genbank) were biased towards clade B sequences. The 2002 databases include a greater variety of non-clade B sequences. Using the same selection criteria for choosing peptides on the 2002 databases of HIV-1 sequences, a different set of peptides would have been selected (data not shown).

not bind to A2 in vitro, and was previously published as B7, this peptide was not included in the list of six positive A2 responses.

3.2.3. A2 peptide conservation

Peptide 1247, a novel pol epitope, is conserved in 549 of 21,075 pol sequences in public databases (3%, Table 4a and Fig. 2b). Only 25% of pol sequences available in Genbank are linked to a country of origin. Peptide 1237 (a novel vif epitope) is also conserved (103 out of 660 sequences, or 11%, Table 4a and Fig. 2c) within the comparatively small dataset of HIV vif sequences available in the HIV-1 sequence databases. Peptide 1256, an env peptide, is less well conserved across HIV isolates (440 out of 32,684, or 1%, Table 4a). However, closer evaluation of the origin of isolates that contain the identical epitope (Fig. 2a) suggests that this epitope may still be very relevant for vaccine design.

3.3. A11 results

3.3.1. A11 peptide binding

Binding assay results for the peptides selected by Conservatrix and EpiMatrix are shown in Table 4b. Twenty-three of the 25 A11 HIV-1 peptides selected by Conservatrix and EpiMatrix for this study bound to T2 cells expressing the A11 allele (92%, Table 4b), including the published HLA A11-restricted control epitope, 1282 (TVQCTHGIKP, env). In contrast, none of six A2 and B7 HIV-1 peptides (selected from peptides synthesized for this study) bound to T2-A11 cells (data not shown). All but one of the peptides (1267) to which PBMC responded in ELIspot were binders in the T2-A11 binding assay.

3.3.2. A11 ELIspot results

PBMC from six subjects possessing the HLA A11 allele or an allele belonging to the HLA A3, A11, A68

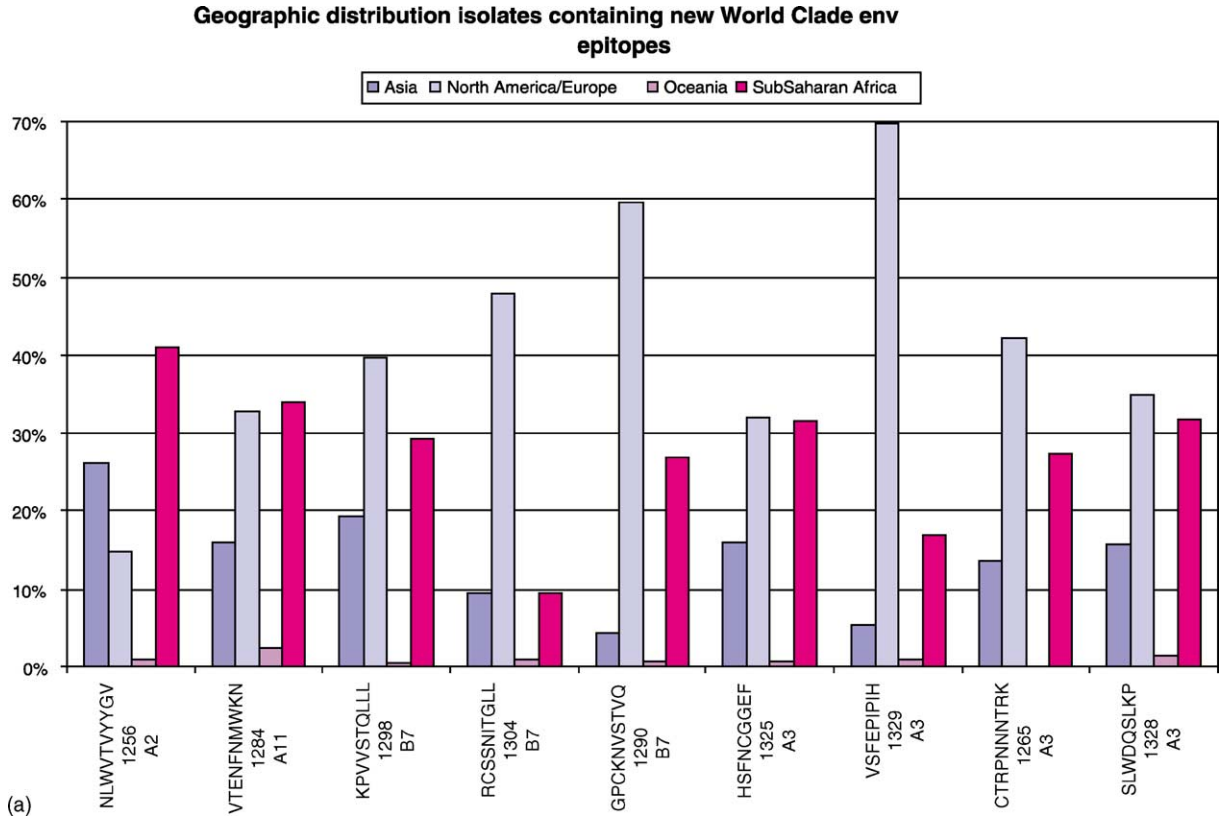
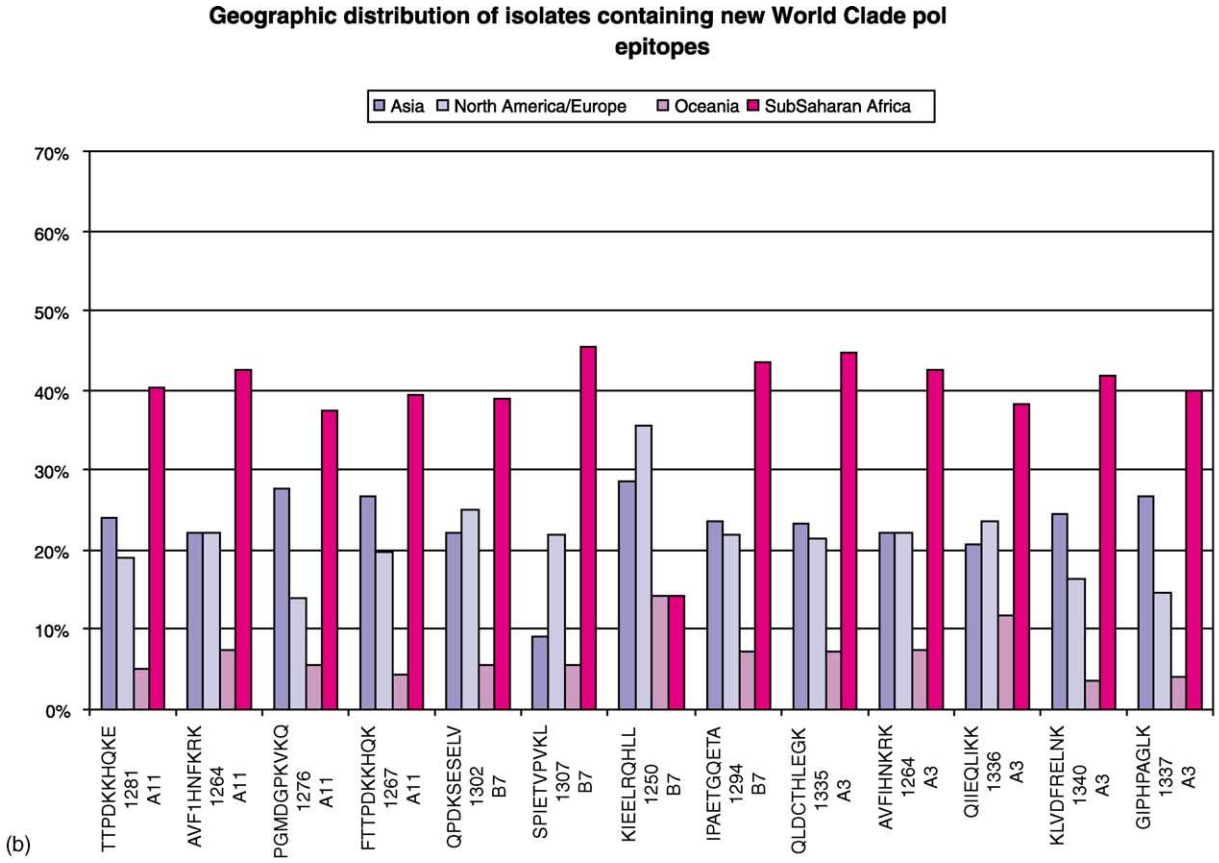
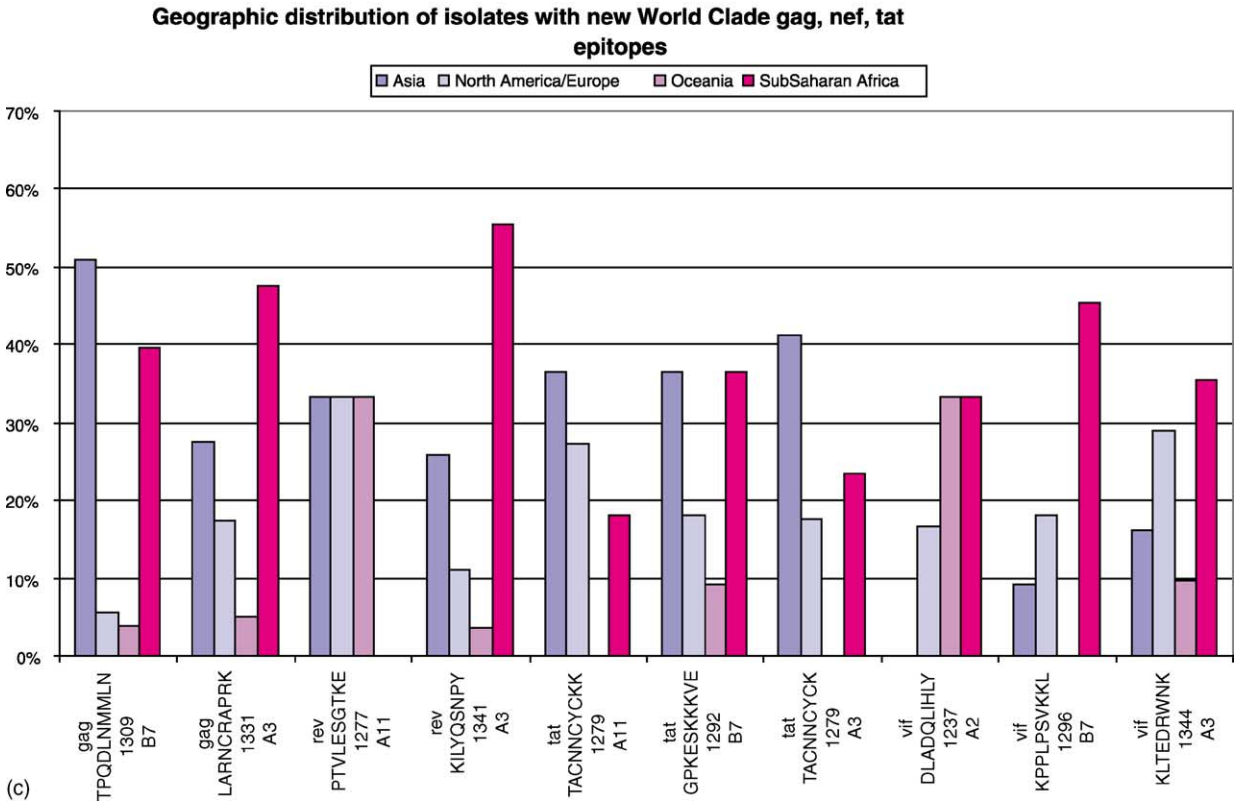


Fig. 2. These histograms illustrate the distribution of isolates derived from different regions of the world (Europe and North America, Africa, Asia, and Oceania) that contain peptide sequences that are identical to the novel epitopes described in this manuscript. Not all HIV-1 sequences in our database could be classified, since only 25% of HIV-1 sequences in Genbank have attached country identifiers. In some cases where country identifiers could not be differentiated from the location of the individual or institution reporting the sequence, the sequence may be mis-classified. For example, some isolates described as being European or North America in these figures may actually be non-clade B isolates. All isolates identified as having been described or identified in Australia were included under “Oceania”. (a) Describes the distribution of identical peptides in env isolates. Note that conservation of these epitopes is very skewed, in the case of env isolates, reflecting the dominance of North American and European sequences in the early years of the epidemic (1997, when the original candidate epitopes were selected). (b) Illustrates the distribution of sequences that are exactly identical to the pol epitopes identified in this study. In this figure, the high relative conservation of the novel epitopes in HIV-1 sequences classified as having been African in origin is notable, perhaps reflecting a more balanced list of pol sequences in 1997, as well as the relatively well conserved nature of the pol protein. (c) Illustrates the distribution of sequences that are exactly identical to the gag, nef, and tat epitopes identified in this study. Again, African isolates predominate.



(b)



(c)

Fig. 2. (*Continued*).

supertype [48] were evaluated in ELIspot assays. Two subjects (H0018M: A2, A11, B44, B51, Cw05, Cw15 and H0201R: A11, A68, B42, B45, Cw16, Cw17) responded to none of the peptides, but did respond to PHA in vitro. As shown in Table 4b, 10 (40%) of the A11 peptides, including the control epitope 1283 (TVYYGVPVWK, env), stimulated ELIspot responses from PBMC obtained from the remaining four subjects. The average number of responses per subject (of the four subjects responding to peptides) to the HLA A11 peptides was four. Two published A11-restricted epitopes were reconfirmed in these assays; eight novel HLA A11-restricted epitopes were defined.

Only two of the four study subjects responding to A11 peptides possessed the A11 allele, which is much more common among Asian and Indian populations than the Providence population of HIV-1 infected subjects. Both of these subjects (523001 and 718001) also possessed another A3 supermotif alleles at the A locus (A11, A3 and A11, A68, respectively). PBMC from the two additional subjects were tested with the A11 peptides because the subjects possessed either the A3 allele or the A68 allele, members of the A3 supermotif family.

Subject 606001 (A1, A68, B15, B40, Cw03) responded to five HLA A11-selected peptides (1274 (NTPVFAIKKK, env), 1284 (VTENFNMWKN, env), 1281 (TTPDKKHQKE, pol), 1264 (AVFIHNFKRK, pol), and 1279 (TACNNCYCKK, tat)), all of which were shown in binding assays to bind to HLA A11. This subject did not respond to the published control A3 supertype restricted epitope (TVYYGVPVWK, env).

Subject 0523001 (A3, A11, B14, B51, Cw08, Cw13) responded to peptides 1283 (control epitope) and 1264 in addition to peptides 1284 and 1271 (IRLRPGGKKK, gag). Subject 0202001 (A25, A68, B18, B27) also responded to peptides 1274 and 1264 in addition to peptide 1283 (control epitope).

Subject 718001 (A11, A68, B42, B45, Cw16, Cw17) was the only subject whose PBMC responded to peptides 1276 (PGMDGPKVKQ, pol), 1267 (FTTPDKKHQK, pol, the sole peptide that did not bind), 1277 (PTVLESGTKE, rev), in addition to recognizing a peptide that was recognized by one other subject (1281; 606001 [A1, A68, B15, B40, Cw03]). The response to peptide 1267 was observed in the context of peptides pulsed on T2-A11 cells (the method used for this subject's PBMC) and two of the other peptides to which this subject responded did bind to T2-A11 cells (1277 and 1281). These data suggest that peptide 1267 is indeed A11-restricted. Also, this subject's PBMC did not respond to the control peptide pulsed on T2-A11 cells. It is, however, possible that this subject was responding to peptides in the context of HLA A68 (an allele shared with subject 606001). In contrast, for this cohort, none of the subjects' HLA B and HLA C alleles overlapped, making it unlikely that any of the peptides tested were presented in the context of other HLA alleles.

3.3.3. HLA A3, A11, A68 supertype and promiscuous epitopes

Seven of the peptides appear to have been restricted by more than one A3 supertype allele (A11–A3–A68) either because the peptide bound to HLA A11, and was shown to stimulate γ-interferon release in ELIspot assays performed with PBMC from study subjects possessing a A3 supertype allele (A3, A11, or A68), and/or was published as an A3-restricted epitope by another laboratory group (Table 4b). Two of these supertype epitopes have previously been published (the control epitope, 1283 (TVYYGVPVWK, env, and 1271 IRLRPGGKKK, gag)).

These data suggest that five additional novel supertype peptides are described here, including 1284 (VTENFNMWKN, env), 1272 (KIRLRPGGKK, gag), 1281 (TTPDKKHQKE, pol), 1264 (AVFIHNFKRK, pol), and 1274 (NTPVFAIKKK, pol), based on binding to HLA A11 in T2A11 binding assays and recognition of the epitope by PBMC from subjects possessing the HLA A68 allele. In addition, two promiscuous (non-supertype) epitopes are described; 1283 (TVYYGVPVWK, env) and 1276 (PGMDGPKVKQ, pol), peptides previously published for HLA A2 and HLA B8, respectively. Both were shown to bind to HLA A11 on the surface of T2-A11. Peptide 1276, a published B8 epitope, stimulated a response from PBMCs from a subject possessing HLA A1, A68, B15, B40, Cw03. Subject 0202001 (A25, A68, B18, B27) and subject 0523001 (A3, A11, B14, B51, Cw08, Cw13) responded to peptide 1283 (a T2-A11 binder and published A2, A11, A68, B18 epitope). Thus, peptides 1276 and 1283 can be considered "promiscuous" epitopes.

3.3.4. A11 peptide conservation

Eleven of these A11 peptides were conserved in more than 10% of the sequences obtained from public databases. Peptide 1274 (NTPVFAIKKK, pol), a supertype peptide that was also immunodominant (ID) (responses greater than 1000 spots per 10E6 SFC) is conserved in 11,098 of 13,689 pol sequences in public databases (53%, Table 4b). 1281 (TTPDKKHQKE, pol) is conserved in 7215 of 13,689 database entries (Table 4b). Peptide 1284 (VTENFNMWKN, env) is conserved in 793 env sequences (only 2% of the 32,684 sequences in the 2002 databases, Table 4b). However, isolates containing the exact peptide sequence have been isolated from Brazil [7], Burkina Faso [1], the Central African Republic [1], Cote d'Ivoire [1], Zaire [4], Zambia [2], Thailand [1], Jamaica [1], China [1], and Sweden [1] and 20 additional countries (Fig. 2a). Peptide epitope 1276 (PGMDGPKVKQ, pol) is conserved in 9505 of 21,065 (42%) pol sequences in public databases (Table 4b). Peptide 1279 (TACNNCYCKK, tat) is conserved in 36 out of 780 tat isolates (3%, Table 4b). See Fig. 2a–c for information on the regional conservation of the other confirmed HLA A11-restricted epitopes examined in this study.

3.4. B7 results

3.4.1. B7 peptide binding

Twenty-one of the 25 peptides selected for this study by Conservatrix and EpiMatrix stabilized B7 molecules in the HLA B7-transfected T2 cell (T2-B7) binding assay (84%), including the published B7 control peptide (1291 (GPGHKARVLA, gag)) (Table 4c). B7-restricted binding was also confirmed for two B7-restricted peptides that were published subsequent to the start of this study (1305 (RPVVSTQLLL, env) and 1306 (SPAIFQSSMT, pol)). The remaining 18 peptides are novel HLA B7 ligands. None of eight A2 and A11 peptides (selected from peptides synthesized for this study) stabilized B7 (data not shown).

3.4.2. B7 ELIspot results

PBMC from four study subjects were evaluated for γ-interferon secretion to the peptides in ELIspot assays (Table 4c). Three of the subjects possessed the HLA B7 allele. One possessed the HLA B8 allele, which has similar binding motif requirements (EpiVax, unpublished). Eleven of the 25 B7 peptides stimulated γ-interferon response (44%). Seven of these epitopes are novel HLA B7-restricted epitopes (1298, 1309, 1302, 1307, 1294, 1296, and 1300). In addition, peptide 1308 (TPQDLNTMLN, gag), while not recognized in ELIspot assays, did bind to HLA B7 in the T2-B7 assay (and was recognized in a separate study of Côte d'Ivorian subjects, see Section 4). Two of these epitopes did not bind to T2-B7 cells (1300, 1302). PBMC from all four subjects responded to the peptides. The number of responses per B7 subject ranged from one to eight; the average number of responses was four.

PBMC from subject 228 (A3, A24, B7, B38, Cw07, Cw12/13) responded to eight peptides (1309 (TPQDLNMMLN, gag), 1291 (GPGHKARVLA, gag [control epitope]), 1302 (QPDKSESELV, pol), 1293 (GPKVKQWPLT, pol), 1306 (SPAIFQSSMT, pol), 1307 (SPIETVPVKL, pol), 1296 (KPPLPSVKKL, vif), 1300 (LVILAIVALV, vpu)). Of these eight epitopes, one pol peptide (1302 (QPDKSESELV, pol)) and one vpu peptide (1300 (LVILAIVALV, vpu)) were not shown to bind to HLA B7 in T2-B7 assays. As these peptides did not give a positive EpiMatrix score for any of the other HLA alleles possessed by this subject (data not shown) and the ELIspot responses were well above background (44 spots per million PBMC and 103 spots per million, respectively), these two peptides were assigned to the list of new HLA B7-restricted epitopes.

PBMC from subject H0015M-100700 (A1, A3, B7, B8, Cw03, Cw07) responded to two peptides: to 1291 (control epitope) and peptide 1298 (KPVVSTQLLL, env), a novel HLA B7-restricted epitope, both in the absence of HLA B7-restricted targets (peptide pulsed on autologous PBMC) and in the presence of HLA B7-transfected B7-restricted target cells (peptide pulsed T2-B7 targets). PBMC from subject 0829001 (A2, A3, B7, B58, Cw07) responded to a single peptide, 1294 (IPAETGQETA, pol), a T2-B7 binder.

The fourth subject possessed the HLA B8 allele (subject 0411001: A29, A30, B8, B44, Cw07, Cw16). The HLA B8 motif is similar to the B7 motif in the second to last position (L in positions 8 or 9). This study subject was evaluated in order to compare the subjects responses to those of subject H0015M-100700 (A1, A3, B7, B8, Cw03, Cw07). The latter subject responded to several peptides that corresponded to both the B7 and the B8 motifs. Subject 0411001 (B8) responded to four peptides in ELIspot assays (1298 (KPVVSTQLLL, env), 1305 (RPVVSTQLLL, env), 1293 (GPKVKQWPLT, pol), and 1307 SPIETVPVKL, pol), all T2-B7 binders. This subject shared a response with subject H0015M for only one peptide (1298). This subject did not respond to the control epitope, peptide 1291. However, peptide 1305 is a published HLA B7-restricted epitope (published subsequent to the initiation of this study). These responses suggest that B7 and B8 share motif similarities (either at the C-terminal anchor and, possibly, for these peptides, the P in position 2).

One allele-mismatched subject 0509001 (A1, A26, B27, B51) did not respond to any of the B7 epitopes but did respond to PHA. No HIV-seronegative subjects' responses to the B7 peptides were evaluated.

3.4.3. B7 published epitopes, supertype epitopes and promiscuous epitopes

Two published B7 supertype epitopes (1291, the control epitope, and peptide 1309) were reconfirmed. In one case, a novel epitope (1298) was recognized by more than one subject, who possessed diverse HLA B alleles (B8) in addition to binding to HLA B7 on T2-B7 cells. This epitope was classified as promiscuous (B7, B8). Peptide 1292, a published B8 epitope, was predicted and confirmed to be restricted by B7 by binding assays and ELIspot assays in these studies, and was therefore also classified as promiscuous. Similarity between the B7 and B8 MHC binding motifs has been noted above. And finally, peptides 1307 and 1250 are previously published epitopes restricted by other HLA alleles, A2 (1307) and Bw60 (1250), respectively, and were therefore classified as promiscuous epitopes, for a total of four promiscuous B7 epitopes.

3.4.4. B7 peptide conservation

Eleven of the B7 peptides were conserved in more than 10% of isolates filed in public databases. In particular, 1307 (SPIETVPVKL, pol), a promiscuous epitope, was conserved in 9941 of 21,075 pol sequences (47%, Table 4c and Fig. 2b). Similarly, peptide 1250 (KIEELRQHLL, pol), was conserved in 6446 of 21,075 pol sequences in public databases (31%). Peptide 1309 (TPQDLNMMLN gag) is conserved in 489 of 6269 gag sequences in public databases (8%, Table 4c and Fig. 2c). Peptide 1298 (KPVVSTQLLL, env) is conserved in 1387 of 32,684 env isolates in public databases (4%, Table 4c). See Fig. 2a–c for the regional distribution of sequences containing each of the epitopes.

3.5. A3 peptide results

3.5.1. A3 peptide binding

Functional monoclonal reagents for T2-A3 binding assays (that did not give a high background) could not be obtained for this study. Accordingly, T cell responses to the A3 peptides were measured but binding assays were not performed. (Table 4d).

3.5.2. A3 ELIspot results

Sixteen peptides stimulated γ-interferon release in ELIspot assays (64%), including the published A3-restricted control peptide (1283 (TVYYGVPVWK, env)). All six subjects responding to the A3 selected peptides possessed the A3 allele. Only one of the six possessed another allele (A11) that could confound the results. Three subjects (H0012M-101700 (A3, A32, B14, B40, Cw08), H0013M-103100 (A3, A66, B35, B41, Cw04, Cw17), H0032M-050801 (A3, A33, B14, B15, Cw07, Cw08)) did not respond to any A3 peptides, including the control peptide, although these subjects did respond to PHA. Three published A3 epitopes were reconfirmed in this study and 14 novel HLA A3-restricted epitopes were defined. The published epitopes were 1283 (TVYYGVPVWK, env), 1332 (RLRPGGKKK, gag), and 1339 (AIFQSSMTK, pol) which were recognized by one, three, and five subjects, respectively. The number of responses per subject, when non-responders were excluded, ranged from 11 to 3. The average number of responses per subject was six.

Subject 419001 (A3, A23, B49, B57, C?) responded to 11 peptides. These included 1329 (VSFEPIPIH, env), 1265 (CTRPNNNTRK, env, a published A2 epitope), 1331 (LARNCRAPRK, gag), a novel immunodominant A3 epitope, 1332 (RLRPGGKKK, gag), 1340 (KLVDFRELNK, pol), 1337 (GIPHPAGLK, pol), 1335 (QLDCTHLEGK, pol), 1264 (AVFIHNFKRK, pol), 1336 (QIIEQLIKK, pol), 1337 (GIPHPAGLK, pol), and 1342 (TACNNCYCK, tat).

Subject 411002 (A3, A24, B27, B57, Cw13, Cw18) responded to six epitopes including 1329, 1331, 1332, 1264, 1344 (KLTEDRWNK, vif), and 1343 (ALTALITPK, vif). Subject 912002 (A2, A3, B8, B51, Cw01, Cw07) also responded to six peptides, including 1325 (HSFNCGGEF, env), 1328 (SLWDQSLKP, env), 1330 (SLYNTVATL, gag), 1339 (AIFQSSMTK, pol), 1340 (KLVDFRELNK, pol), and 1341 (KILYQSNPY rev). As peptide 1330 is a published A2-restricted epitope and no binding assay results were available to confirm HLA A3-restricted binding, this response was not categorized as an A3-restricted response for this subject.

Subject 0906002 (A2, A3, B8, B51, Cw01, Cw07) responded to five peptides. These were 1339, 1325, 1340, and 1330. Subject 509002 (A3, A26, B8, B52, ?) responded to four peptides. These were 1283, the published A3 control epitope (TVYYGVPVWK, env), 1331, 1332, and 1339. Subject 523001 (A3, A11, B14, B5, Cw08) responded to three peptides including 1335 (QLDCTHLEGK, pol), 1339, and 1340.

3.5.3. HLA A3 supertype and promiscuous epitopes

Three previously published supertype epitopes were re-confirmed (1283 (TVYYGVPVWK env), 1332 (RLRPGGKKK gag), and 1339 (AIFQSSMTK, pol)). One novel supertype peptide was identified (1264 AVFIHNFKRK, pol), a peptide that stimulated ELIspot responses from three supertype subjects (see A11 data, HLA A locus alleles for those three subjects were A25, A68, A3, A11, and A1 A68) and two A3 subjects (A3, A23, A3, A24). One promiscuous epitope, peptide 1265, a previously published HLA A2-restricted epitope (CTRPNNNTRK, env) was identified (response by subject 419001: A3, A23, B49, B57, C?).

3.5.4. A3 peptide conservation

Perhaps the most conserved peptide of the list of A3 epitope candidates was 1265 (CTRPNNNTRK, env) which was conserved in 12,642 of 32,684 env sequences in public databases (39%, Table 4d). The fact that this peptide is so highly conserved, and presented by more than one HLA allele (A2 published and A3, predicted and confirmed in this study), makes it an excellent candidate for inclusion in an HIV vaccine. Another highly conserved env peptide that is newly described in this report is peptide 1325 (HSFNCGGEF, env) which is conserved in 11,307 of 32,684 env sequences listed in public databases (24%, Table 4d). See Fig. 2a for information on the regional conservation of these two env peptides. A pol peptide, 1340 (KLVDFRELNK) is conserved in 12,242 of 21,075 pol sequences (58%, Table 4d). Peptide 1337 (GIPHPAGLK, pol), is conserved in 11,785 of 21,075 pol sequences (56%, Table 4d) in public databases. See Fig. 2b for the regional conservation of these new pol epitopes. Peptide 1344 (KLTEDRWNK vif), was conserved in 581 of 905 vif sequences (64%, Table 4d) in public databases. Additional highly conserved A3 peptides are described in Table 4d, and the regional distribution of sequences containing these epitopes is shown in Fig. 2a–c.

3.6. Subject responses to epitopes restricted by more than one HLA allele

If the study subjects possessed more than one allele for which peptides had been selected, assays were performed for both sets of peptides (if sufficient samples were available). For example, subject 0523001 (A3, A11, B14, B51, Cw08, Cw13) responded to four A3 peptides (1283 (TVYYGVPVWK, env), 1264 (AVFIHNFKRK, pol), 1284 (VTENFNMWKN, env) and 1271 (IRLRPGGKKK, gag)), and three A11 peptides including 1335 (QLDCTHLEGK, pol), 1339 (AIFQSSMTK, pol), and 1340 (KLVDFRELNK, pol). (Table 4a and d). All three A11 peptides were shown to bind on T2-A11 cells. HLA A3 and A11 belong to the same supertype, therefore some of the cross-allele presen-

tation observed here may be due to the supertype binding and presentation.

Subject 0517001 (HLA A2, A3, B44, Cw05, Cw07) responded to A2 peptide 1249 (ILKEPVHGVY, pol) and to 1257 (NMWQEVGKAM, env), but not to the control epitope 1261, and also responded to A3 peptides 1341 (KILYQSNPY, rev) but not to the control peptide nor to the other dominant epitopes. (Table 4a and d).

Subject 0906002 (A2, A3, B8, B51, Cw01, Cw07) responded to five A3 peptides. These were 1339 (AIFQSSMTK, pol), 1325 (HSFNCGGEF, env), 1340 (KLVDFRELNK, pol), 1328 (SLWDQSLKP, env), and 1330 (SLYNTVATL, gag). This subject also responded to one A2 peptide; 1261 (SLYNTVATLY, gag). The latter peptide is a 10-mer version of 1330. Since this subject possessed both the A2 and the A3 allele, response to A2 peptide 1261 and A3 peptide 1330 may have been due to presentation of the peptide in HLA A2. This subject's response to A3 peptides was much broader (four epitopes) than to A2 epitopes (two epitopes). (Table 4a and d).

3.7. Dominant versus subdominant epitopes

Where at least one of the study subjects' responses exceeded 1000 spots per 10E6 PBMC, the epitope was labeled immunodominant. These epitopes are identified as ID in Table 3. The 1264 and 1274, both novel A11 epitopes derived from pol, stimulated γ-interferon release from more than 1000 cells per 10E6 (1 in 1000 PBMC), as did two A3 epitopes (1328 and 1331). Two published epitopes, 1249 (A2) and 1332 (A3) also stimulated responses at this level. The remaining 40 epitopes identified, however, stimulated responses in the range of 100–700 spots per 10E6 PBMC, and may be categorized as "subdominant" epitopes. Individual subjects demonstrated responses to both immunodominant epitopes (spots per 10E6 > 1000) and to subdominant epitopes. In one case, where information on response to the study peptides was available at several time points, a study subject appeared to respond to a greater number of subdominant epitopes at the second time point (data not shown).

4. Discussion

The purpose of these studies has been to confirm that a specific bioinformatics approach, that is, the combination of the Conservatrix and EpiMatrix algorithms, permits the selection of highly conserved HIV-1 T cell epitopes from among *tens of millions* of epitope candidates (more than 65,000 HIV-1 sequences × average 600 amino acids per sequence × 10mer overlapping frames), and, to begin evaluation of these epitopes for inclusion in an "epitope-driven vaccine". The only alternative to using algorithms like Conservatrix and EpiMatrix to define epitopes that are relevant across HIV-1 clades and isolates is to perform epitope mapping using thousands of peptides and thousands of study volunteers. This pilot study proves that a bioinformatics approach can be a much more efficient means of identifying epitope candidates for a truly cross-clade HIV-1 vaccine. Other epitope mapping algorithms may also be useful tools for discovering highly conserved HIV-1 epitopes. A review of these algorithms is published elsewhere [53].

Representative conserved peptides for eight major HIV-1 proteins were selected for this study (env, gag, pol, rev, tat, vif, vpr, vpu). This analysis was performed for all matrices available for the EpiMatrix tool [17]. For this pilot study, a list of 25 peptides for four HLA alleles HLA A2, A3, A11, and B7 were selected for in vitro studies (Table 1). HLA A2 and A3 are the highly prevalent HLA worldwide, while A11 is more common in Asian populations and B7 is more common among African and African-American populations.

Forty-three percent of epitopes selected using these tools stimulated ELIspot responses in vitro. Some of the epitopes were confirmed both in this study and in other published reports [23]. Confirmation of the epitopes by other laboratories (Table 3) provides external validation of the method and additional proof of principle for this bioinformatics approach to defining immunogenic epitopes.

The diverse, sometimes subdominant nature of the response to HIV is evident in this study (as many as 11 responses restricted by a single HLA molecule were observed). Individuals possessing the same HLA allele often responded to different epitopes, and responded to epitopes restricted by more than one allele, and no single epitope appeared to be immunodominant. These data are in complete agreement with several other studies [32,33]. Betts et al. observed that "an individual's overall CD8 T cell response to HIV is not adequately represented by the response to a single epitope and that individual's major histocompatibility complex class I alleles do not predict an immunodominant response restricted by that allele. Accurate quantification of total HIV-specific CD8 T cell responses will require assessment of the response to all possible epitopes" [6]. The experiments reported here, which were performed on PBMC from a relatively small study subject sample, over an 18-month period, and restricted to four HLA alleles, may significantly underestimate the number of epitopes that could be recognized by healthy HIV-1 infected subjects. The goal of the study was not to map HIV-1 infected subjects' responses to HIV in great detail. However, these data suggest that more extensive epitope mapping may need to be performed in order to quantify the breadth of human immune response to HIV-1.

Furthermore, many of the subjects belonged to a cohort that demonstrated excellent adherence to anti-retroviral therapy and low viral loads, which may have diminished our ability to detect HIV-specific CTL responses (GS, personal communication). Fewer than expected HLA A2 epitopes were discovered. Similar paucity of response to A2 epitopes has been noted by other investigators, who suggested that the HIV-1 genome has evolved to escape presentation by this highly prevalent HLA allele [31,46].

4.1. Discordant binding

For four of the six A2 peptides stimulating ELIspot response, no binding to HLA A2 could be detected. The line of T2-A2 cells used in this study has subsequently been shown to have a partially-functional TAP molecule [51] leading to the presentation of endogenously MHC-bound peptides on the surface of these cells. As a result, binding of exogenous A2-restricted peptides might not have been detected due to high background fluorescence associated with TAP mediated processing and presentation of endogenous peptides. In support of this hypothesis, two of the peptides, 1257 (NMWQEVGKAM, env) and 1249 (ILKEPVHGVY, gag), have been confirmed as A2-restricted epitopes by other research groups. Novel A2 epitopes 1256 (NLWVTVYYGV, env, subject 0902991: A2, A30, B39, ?), 1247 (HLKTAVQMAV, pol, subject H0014M: A1, A2, B8, ?, Cw16), and 1237 (DLADQLIHLY, vif, recognized by subjects 902991 and H0014M) may be restricted by A2, and not bind, or (unlikely in the cases where more than one subjects' PBMC responded) the peptide may be restricted by other alleles possessed by this subject.

Peptide 1267 (FTTPDKKHQK, pol) stimulated an ELIspot response but did not bind to T2-A11 in vitro. The response was observed for subject (718001, A11, A68, B42, B45, Cw16, Cw17) who possessed both the A11 and A68 alleles. Thus, it would be difficult to determine which allele was involved in presenting this peptide and whether the peptide might be presented in the context of another allele, including A68. This peptide was therefore classified as an A3 supertype epitope, restricted by A11 or A68.

In two cases, B7 peptides that stimulated ELIspot response did not bind to HLA B7-transfected T2 cells (1302 (QPDKSESELV, pol) and 1300 (LVILAIVALV, vpu)). In both cases, the peptide stimulated response from subject 228. These peptides therefore may be presented in the context of another allele possessed by this subject, such as A3, A24, B38, Cw07, or Cw12/13, however, EBP for these alleles, with the exception of the A24 and the B38 motif, was low. Of note, none of seven B7 peptides bound to HLA A2 or A11 when tested in T2-A2 and T2-A11 binding assays.

4.2. Supertype and promiscuous epitopes

Five epitopes were recognized in the context of MHC "supertypes" and three were promiscuous epitopes. One peptide, CTRPNNTRK, was predicted to be restricted by A3 (and A11), a prediction that was confirmed in these studies. This peptide has been published as an A2-restricted epitope. Thus, as predicted by our tools, this peptide is both promiscuous (A3, A11, A2) and highly conserved in 10,143 out of 24,599 env sequences in public databases (41%). Even though the peptides were selected first for conservation and, second, for probable presentation by one HLA allele, and the promiscuity function of EpiMatrix (now called "ClustiMer") [16] was not applied to this study, this study still serves to illustrate the use of Conservatrix and EpiMatrix as tools for discovery of supertype and/or promiscuous epitopes. Future HIV genome analyses will apply Conservatrix and EpiMatrix (with the clustering function ClustiMer) to select peptides for supertype and/or promiscuous characteristics.

Epitopes identified using this method are highly conserved in isolates derived from a wide range of countries. Since the Conservatrix analysis was performed independently of any starting sequence (whether clade B, or other subtype), the novel epitopes discovered in this study appear to be conserved in isolates from many different countries, across continents, and appear to be independent of clade or subtype grouping. Conservation of these novel epitopes across four regions of the world (Asia, North America/Europe, South Central America, and sub-Saharan Africa) is illustrated in Fig. 2a–c.

It is entirely possible that this analysis of HIV sequences has uncovered regions of HIV-1 that are essential, in some way, to the survival of the virus (the Achilles' heel, or heels, of the HIV-1 genome). For example, these regions may be relevant for binding to cellular receptors, to the function of certain proteins, or may be related to the three-dimensional configuration of one of the virus' proteins. However, it should also be noted that only 45 of the peptides were conserved in more than 10% of their respective protein database, after selection by EpiMatrix for high-scoring peptides. The degree of conservation for each of the epitopes would expand markedly if the amino acids were allowed to vary at "non-anchor" positions. A reanalysis of the peptides, allowing two amino acids to vary at non-anchor positions, is provided in the last three columns of Table 4. Most of the peptides appear to be highly conserved (63 would be conserved in greater than 10% of the sequences). A realistic estimate of the degree of conservation across clades may lie somewhere between these two estimates (absolute conservation and allowing two amino acids to vary).

Proteasomal cleavage sites adjacent to these conserved epitopes may affect the processing of the peptides, and may not be as well conserved as the epitopes themselves. Variation of the proteasomal cleavage sites might abrogate the natural processing of the epitopes in the context of infection by variant HIV-1 strains. Additional confirmation of the epitopes in the context of natural infections in other regions of the world is planned, as it is currently difficult, using available proteasomal cleavage site algorithms, to accurately predict proteasomal cleavage (Bill Martin, EpiVax, unpublished internal evaluation of proteasomal cleavage algorithms).

Several aspects of this study deserve additional scrutiny. For example, all of the study subjects were recruited in the Northeastern United States (Rhode Island and Massachusetts). As their HIV-1 isolates were not sequenced, it would be impossible to determine whether the subjects were infected with clade B or non-clade B isolates of HIV. Furthermore, whether or not the sequences of the HIV virus species infecting the subjects corresponded exactly to the epitopes selected for this study is unknown. However, the

data presented here are consistent with the hypothesis that the responses observed are due to T memory cells responding to epitopes that are processed and presented in the course of natural HIV infection.

In this study, we have not evaluated responses to the peptides using blood samples obtained from individuals known to be infected with other clades of HIV-1. However, there is (1) evidence from our conservation analysis that these epitopes are conserved in a wide range of isolates from Africa, Asia, South America, and the Pacific and (2) additional proof of cross-clade recognition from as-yet unpublished studies conducted by Dr. Janet McNicholl and her colleagues using PBMC from clade A/C subjects in Côte d'Ivoire. Seven selections from our list of peptides were confirmed in small cohort of 16 Ivoirian subjects (Ann Charoudi, Dr. Janet McNicholl, unpublished data). In those studies, peptide 1261 (SLYNTVATLY, gag) was recognized; both variations of our B7 gag epitope (1308 TPQDLNMMLN and 1309 TPQDLNTMLN) were recognized, and several subjects also responded to peptide 1295 (KPCVKLTPLC, gag). Three A3 peptides 1283 (TVYYGVPVWK, env, A3 control epitope), 1331 (LARNCRAPRK, gag, novel A3 epitope), and 1332 (RLRPGGKKK, gag, published epitope) were also recognized. These observations confirm our rationale for selecting cross-clade epitopes.

In addition, in studies performed by Bond et al. in Thailand, CTL responses to five A11 peptides that were identical to peptides selected for the present study were evaluated in nine clade E-infected Thai subjects [7]. CTL responses were observed in this study to four cross-clade epitopes that did not stimulate T cell responses in Bond's cohort of nine Thai subjects. Conversely, Thai study subjects responded to one epitope (1268 GIPHPAGLKK, two of nine subjects tested) to which subjects recruited in Providence did not respond. Whether these differences are due to epitope processing or to individual subject differences as described by Lieberman et al. [32] and Betts et al. [6], is unknown. Bond and McNicholl's group also found CTL responses to a 11-mer variant of our peptide 1283 TVYYGVPVWK (they evaluated peptide V̲TVYYGVPVWR).

CD8+/CD4+ depletion was not performed prior to ELIspot assays, thus some of the responses observed could be due to class II restricted responses. However, the HLA restriction for most of these epitopes was confirmed in binding studies using T2 cells expressing a single MHC molecule, and, in general, these peptides did not bind to T2 cells expressing MHC class I molecules for which they were predicted not to bind. Furthermore, where more than one subject responded to a peptide, the subjects were only matched for the HLA A or B allele corresponding to the peptide selections. Since it is extremely unlikely the responding cells were matched at any of their class II alleles, by chance, all of the in vitro responses observed here are likely be due to CD8+ restricted responses.

In general, ELIspot responses to these peptides provide additional confirmatory evidence that cross-clade CTL epitopes can be identified. Cross-clade epitopes have previously been described by Walker and Weinhold and others [10,21,35,48,50]. The studies described here appear to be consistent with studies published by Weinhold and co-workers [20] and Ferrari et al. [20,21]. The latter studies used recombinant vaccinia virus constructs expressing whole genes and mixed populations of T cells derived from HIV-1 infected subjects [10,35,50]. Thus, the genes present in those constructs probably included some of the epitopes identified using the bioinformatics approach used in the present study.

In other studies, Rowland-Jones et al. have evaluated highly conserved *peptide* epitopes (instead of whole genes cloned into vaccinia) for cross-clade T cell recognition (using cells from subjects infected with various clades of HIV) [42,43,48]. These studies provide exciting proof that non-identical sequences representing non-clade B versions of an epitope can stimulate a single T cell clone. Thus, the promiscuous nature of TCR interaction with MHC–epitope complexes will further expand the potential for "cross-clade" recognition of conserved epitopes. Accordingly our estimates of cross-clade conservation (Table 3), which are based on absolute conservation of every amino acid in the sequence, may significantly underestimate the number of isolates, with which T cells recognizing these epitopes may be able to cross-react.

For the final version of this manuscript, we reanalyzed the conservation our original 1997 peptides using an updated and inclusive 2002 database of HIV-1 sequences (Table 4a–d). In performing the reanalysis, we found that less than 50% of the peptides we originally identified in 1997 would have been selected using the same criteria on the 2002 database, reflecting the fact that a greater variety of HIV-1 strains have been sequenced in 2002, with less clade B bias. Furthermore, we found that HIV-1 peptides are generally poorly conserved. For example, less than 250 env peptides are conserved in more than 5% of the 2002 env sequence database (data not shown)—a fact that reflects the well known diversity of the HIV-1 genome and raises concern for HIV-1 vaccine development in general. While some of the new peptides we would select from the 2002 database were better conserved across clades than our 1997 selections, fewer of these conserved peptides were immunogenic based on EpiMatrix score (data not shown).

This study provides an illustration of the bioinformatics approach to selecting broadly conserved T cell epitopes for HIV-1 vaccine development. Studies of acutely infected subjects and long-term non-progressors (LTNP) have now convincingly shown that broad CTL and Th responses to both dominant and subdominant epitopes, restricted by multiple HLA alleles, are associated with better control of HIV-1 infection [41], accentuating the importance of broadly conserved CTL epitopes. The epitopes described in this report are to be included in a DNA-vectored vaccine that is similar in concept but different in approach from other epitope-based HIV vaccines [12,20,25].

5. Disclosure

Two of the contributing authors, William Martin and Anne S. De Groot, are senior officers and majority shareholders at EpiVax, a privately-owned vaccine design company located in Providence, RI. EpiVax holds the license to EpiMatrix and associated epitope mapping tools developed at Brown University. These authors acknowledge that there is a potential conflict of interest related to their relationship with EpiVax and attest that the work contained in this research report is free of any bias that might be associated with the commercial goals of the company.

EpiVax and the TB/HIV Research Lab declared the "GAIA" vaccine project described in this manuscript to be a not-for-profit project in August 2000. Anne S. De Groot is founder and scientific advisor to the GAIA Vaccine Foundation, a 501c3 organization that supports the development of a globally applicable AIDS vaccine and plans to distribute the vaccine worldwide based on need rather than profit.

Acknowledgements

This work was supported by Rhode Island Center for Cellular Medicine Phase I grant to EpiVax and the TB/HIV Research Lab, by the National Institutes of Health, NIAID, Division of AIDS (R01 AI 40888, HIV Evolution and Cellular Immune Response, R43 AI 46212 Epitope-driven HIV vaccine targeting dendritic cells and R21 AI 45416, A cross-clade approach to HIV Vaccine Development) and by the Lifespan CFAR (Lifespan CFAR Pilot Project Validation of candidate HIV vaccine T cell epitopes). We are grateful for the support and assistance of many individuals, without whom the completion of this portion of the "GAIA" (highly conserved HIV-1 vaccine) pilot project would have been impossible. First of all, we wish to thank the HIV-infected individuals who generously donated their blood to this study. We also thank our clinical collaborators and nurse clinicians (Diane Dean, Grace Accetta) who recruited the study subjects and provided us with important feedback on our work. In addition, we wish to thank the following individuals for their tireless enthusiasm and encouragement: either presently or formerly at the Division of AIDS, Neal Nathanson, Margaret (Peggy) Johnston, Jim Bradac, Michael Pensiero, and Fred Vogel; at Emory University, Janet McNicholl, and at Brown University, Fran McCrossan, Frank Rothman, Michael Lysaght, Sandy Deitch, and Regina Berenbach. Brown University graduates Natasha Chinai, Jean David Barnea and James Robert Schafer also contributed to the success of this project. Craig Grein, Aminata De Groot, and Zeno Trocchi also provided moral support, patience, and encouragement.

This work is dedicated to the memory of Bill Beckler, Shorty Boy Torres, Darlest Chenevert, and our friends and colleagues who are still living with HIV: for the cure. In a further step designed to accelerate the process of HIV vaccine development, the GAIA Vaccine Foundation is providing HIV researchers with free access to EpiMatrix by Internet (http://www.GAIAvaccine.org).

References

[1] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
[2] Bairoch A. Amino acid scale: amino acid composition (%) in the SWISS-PROT protein sequence data bank. SWISS-PROT release 38; 1999 [online].
[3] Bende S, Johnston M. Immunization update: search for an AIDS vaccine. AIDS Read 2000;10:526.
[4] Korber B, Meyers G. Los Alamos HIV sequence database. Los Alamos (NM): Los Alamos National Laboratories; 2002 [online].
[5] Korber B, Meyers G. Los Alamos HIV sequence database. Los Alamos (NM): Los Alamos National Laboratories; 1997 [online].
[6] Betts MR, Casazza JP, Patterson BA, Waldrop S, Trigona W, Fu TM, et al. Putative immunodominant human immunodeficiency virus-specific CD8(+) T-cell responses cannot be predicted by major histocompatibility complex class I haplotype. J Virol 2000;74: 9144–51.
[7] Bond K, Pau CP, Malegam JY, De Groot AS, Szu E, Hodge TW, et al. HLA A11 cytotoxic T lymphocyte epitopes in HIV-1 env subtype E. AIDS Res Hum Retroviruses 2001;17:703–18.
[8] Brander C, Pichler WJ, Corradin G. Identification of HIV protein-derived cytotoxic T lymphocyte (CTL) epitopes for their possible use as synthetic vaccine. Clin Exp Immunol 1995;101: 107–13.
[9] Brander C, Walker BD. The HLA class I restricted CTL response in HIV-1 infection: systematic identification of optimal epitopes, vol. LAUR 97-1150. Los Alamos (NM): Los Alamos National Laboratory; 1997.
[10] Buseyne F, Chaix ML, Fleury B, Manigard O, Burgard M, Blanche S, et al. Riviere Y. Cross-clade-specific cytotoxic T lymphocytes in HIV-1-infected children. Virology 1998;250:316–24.
[11] Caley IJ, Betts MR, Irlbeck DM, Davis NL, Swanstrom R, Frelinger JA, et al. Humoral, mucosal, and cellular immunity in response to a human immunodeficiency virus type 1 immunogen expressed by a Venezuelan equine encephalitis virus vaccine vector. J Virol 1997;71:3031–8.
[12] Cease K, Berzofsky J. Toward a vaccine for AIDS: the emergence of immunobiology based vaccine development. Ann Rev Immunol 1994;12:923–89.
[13] Clements-Mann ML, Weinhold K, Matthews TJ, Graham BS, Gorse GJ, Keefer MC, et al. Immune responses to human immunodeficiency virus (HIV) type 1 induced by canarypox expressing HIV-1MN gp120, HIV-1SF2 recombinant gp120, or both vaccines in seronegative adults. J Infect Dis 1998;177:1230–46.
[14] Cohen J. AIDS research. Debate begins over new vaccine trials. Science 2001;Sep. 14:293.
[15] Colombani J, Lepage V, Degos L, Kalil J, Fellous M. HLA monoclonal antibody registry: a proposal. Tissue Antigens 1982;20: 161–71.
[16] De Groot AS, Bosma A, Chinai N, Frost J, Jesdale BM, Gonzalez MA, et al. From genome to vaccine: in silico predictions, ex vivo verification. Vaccine 2001;19(31):4385–95.
[17] De Groot AS, Jesdale BM, Szu E, Schafer JR, Chicz R, Deocampo G. An interactive web site providing MHC ligand predictions: application to HIV research. AIDS Res Hum Retroviruses 1997;7: 139–42.
[18] De Groot AS, Saint Aubin C, Bosma A, Sbai H, Rayner J, Martin W. Rapid determination of HLA B*07 ligands from the West Nile virus NY99 genome. Emerg Infect Dis 2001;7:706–13.

[19] Evans DT, O'Connor DH, Jing P, Dzuris JL, Sidney J, das Silva J, et al. Virus-specific CTL responses selected for amino-acid variation in simian immunodeficiency virus env and nef. Nat Med 1999;5: 1270–6.
[20] Ferrari G, Kostyu DD, Cox J, Dawson DV, Flores J, Weinhold KJ, et al. Identification of highly conserved and broadly cross-reactive HIV type 1 cytotoxic T lymphocyte epitopes as candidate immunogens for inclusion in Mycobacterium bovis BCG-vectored HIV vaccines. AIDS Res Hum Retroviruses 2000;16:1433–43.
[21] Ferrari G, Humphrey W, McElrath MJ, Excler JL, Duliege AM, Clements ML, et al. Clade B-based HIV-1 vaccines elicit cross-clade cytotoxic T lymphocyte responses in uninfected volunteers. Proc Natl Acad Sci USA 1997;94:1396–401.
[22] Foung SK, Taidi B, Ness D, Grumet FC. A monoclonal antibody against HLA A11 and A24. Hum Immunol 1986;15:316–9.
[23] Gianfrani C, Oseroff C, Sidney J, Chestnut RW, Sette A. Human memory CTL response specific for influenza A virus is broad and multispecific. Hum Immunol 2000;61:438–52.
[24] Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 1997;3:212–7.
[25] Hanke T, McMichael AJ. Design and construction of an experimental HIV-1 vaccine for a year-2000 clinical trial in Kenya. Nat Med 2000;6:951–5.
[26] Haynes BF. HIV vaccines: where we are and where we are going. Lancet 1996;348:933–7.
[27] Hu DJ, Dondero TJ, Rayfield MA, George JR, Schochetmen EA. The emerging genetic diversity of HIV. JAMA 1996;275:210–6.
[28] Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J. Dramatic rise in plasma viremia after CD8 (+) T cell depletion in simian immunodeficiency virus-infected macaques. J Exp Med 1999;189:991–8.
[29] Jin X, Nixon DF, Roberts CGP, Nixon DF, Safrit J, Zhang LQ, et al. Identification of subdominant cytotoxic T lymphocyte epitopes encoded by autologous HIV type 1 sequences, using dendritic cell stimulation and computer-driven algorithm. AIDS Res Hum Retroviruses 2000;16:67–76.
[30] Johnson RP, Trocha A, Yang L, Mazarra GP, Panicali DL, Buchanan TM, et al. HIV-1 gag-specific cytotoxic T lymphocytes recognize multiple highly conserved epitopes: fine specificity of the gag-specific response defined by using unstimulated peripheral blood mononuclear cells and cloned effector cells. J Immunol 1991;147:1512.
[31] Lieberman J. The cytotoxic T cell response to HIV-1. In: Huber ME, Wahn P, editors, Proceedings of the Symposium in Immunology V. Berlin: Springer-Verlag; 1996.
[32] Lieberman J, Fabry JA, Fong DM, Parkerson GRR. Recognition of a small number of diverse epitopes dominates the cytotoxic T lymphocyte response to HIV type 1 in an infected individual. AIDS Res Hum Retroviruses 1997;13:383–92.
[33] Lieberman J, Fabry JA, Kuo MC, Earl P, Moss B, Skolnik PR. Cytotoxic T lymphocytes from HIV-1 seropositive individuals recognize immunodominant epitopes in Gp160 and reverse transcriptase. J Immunol 1992;148:2738–47.
[34] Ljunggren HG, Stam NJ, Ohlen C, Neefjes JJ, Hoglund P, Heemels MT, et al. Empty MHC class I molecules come out in the cold. Nature 1990;346:476–80.
[35] Lynch JA, deSouza M, Robb MD, Markowitz L, Nitayaphan S, Sapan CV, et al. Cross-clade cytotoxic T cell response to human immunodeficiency virus type 1 proteins among HLA disparate North Americans and Thais. J Infect Dis 1998;178:1040–6.
[36] McMichael AJ, Phillips RE. Escape of HIV from immune control. Ann Rev Immunol 1997;15:271–96.
[37] Meister GE, Roberts CGP, Berzofsky JA, De Groot AS. Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences. Vaccine 1995;13(6): 581–91.
[38] Nijman HW, Houbiers JGA, Vierboom MPM, van der Burg SH, Drijfhout JW, D'Amaro J, et al. Identification of peptide sequences that potentially trigger HLA A2.1-restricted cytotoxic T lymphocytes. Eur J Immunol 1993;23:1215–9.
[39] Parham P, Brodsky FM. Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA A2 and a variant of HLA A28. Hum Immunol 1981;3: 277–99.
[40] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Imunogenetics 1999;50(3–4):213–9.
[41] Rosenberg ES, Altfeld M, Poon SH, Phillips MN, Wilkes BM, Eldridge RL, et al. Immune control of HIV-1 after early treatment of acute infection. Nature 2000;407:523–6.
[42] Rowland-Jones S, Dong T, Krausa P, Sutton J, Newell H, Ariyoshi K, et al. The role of cytotoxic T-cells in HIV infection. Dev Biol Standard 1998;92:209–14.
[43] Rowland-Jones SL, Dong T, Fowke KR, Kimani J, Krausa P, Newell H, et al. Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. J Clin Invest 1998;102:1758–65.
[44] Schmechel SC, Russell N, Hladik F, Lang J, Wilson A, Ha R, et al. Immune defense against HIV-1 infection in HIV-1-exposed seronegative persons. Immunol Lett 2001;79:21–7.
[45] Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, et al. Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes. Science 1999;283:857–60.
[46] Shankar P, Fabry JA, Fong DM, Lieberman J. Three regions of HIV-1 gp160 contain clusters of immunodominant CTL epitopes, Immunol Lett 1996;52:23–30.
[47] Storkus WJ, Alexander J, Payne JA, Cresswell P, Dawson JR. The alpha 1/alpha 2 domains of class I HLA molecules confer resistance to natural killing. Immunology 1989;143:3853–7.
[48] Threlkeld SC, Wentworth PA, Kalams SA, Wilkes BM, Ruhl DJ, Keogh E, et al. Degenerate and promiscuous recognition by CTL of peptides presented by the MHC class I A3-like superfamily. J Immunol 1997;159:1648–57.
[49] WHO-UNAIDS, VAC. Approaches to the development of broadly protective HIV vaccines: challenges posed by the genetic, biological and antigenic variability of HIV-1. AIDS 2001;15:W1–25.
[50] Wilson SE, Pedersen SL, Kunich JC, Wilkins VL, Mann DL, Mazzara GP, et al. Cross-clade envelope glycoprotein 160-specific CD8+ cytotoxic T lymphocyte responses in early HIV type 1 clade B infection. AIDS Res Hum Retroviruses 1998;14:925–37.
[51] Wolfel C, Drexler I, Van Pel A, Thres T, Leister N, Herr W, et al. Transporter (TAP)- and proteasome-independent presentation of a melanoma-associated tyrosinase epitope. Int J Cancer 2000;88: 432–8.
[52] Zhang QJ, Gavioli R, Klein G, Masucci MG. An HLA A11-specific motif in nonamer peptides derived from viral and cellular proteins. Proc Natl Acad Sci USA 1993;90:2217–21.
[53] De Groot AS, Sbai H, Saint Aubin C, McMurry JA, Martin W. Immuno-informatics: mining genomes for vaccine components. Immunol Cell Biol 2002;80(3):255–69.