I-B

Enhanced Motif Scan: A Tool to Scan for HLA Anchor Residues in Proteins

Karina Yusim[a], James J. Szinger[a], Isobella Honeyborne[b,c], Charles Calef[a], Philip J. R. Goulder[b,c], Bette T. M. Korber[a]

I-B-1 Introduction

The HIV Immunology Database at `http://www.hiv.lanl.gov/` is a repository of information about HIV T-cell and antibody epitopes, integrated with the sequence variability data from the HIV Sequence Database [Kuiken2003]. The immunology database includes tables, maps and alignments of HIV-specific cytotoxic T lymphocyte (CTL) epitopes that are updated through annual summaries of the literature. We also develop simple web-based tools with the goal of assisting immunologists in experimental design and interpretation of their results. While these tools emphasize ease of application for HIV studies, many are also useful for studies of other pathogens and immune responses.

Several tools in the immunology database that rely on HLA anchor motif assignments are available to help identify potential CTL epitopes in HIV proteins[Calef2001, Calef2002, Thakallapally2001]. HLA anchor motifs are the conserved elements in epitopes that allow binding with specific HLA class I or class II proteins for presentation on the cells surface and recognition by T-cells. The peptide binding groove is divided into six pockets (A, B, C, D, E and F). Only 2 to 4 of these are generally occupied, although most of the residues within a peptide contribute to some degree to MHC binding. There are usually two dominant binding anchors for an epitope bound to an MHC class I molecule. These are usually the second position residue and the carboxy-terminus (C-terminus), and these residues bind to the B and F pockets, respectively, of the MHC class I peptide-binding groove. Class I epitopes are typically 9 amino acids long, ranging between 8-12 amino acids long. The optimal length of class II presented epitopes is often higher, and the spacing of the anchor residues relative to each other within the epitope tends to be more variable.

Anchor motif patterns are relatively, although not perfectly, preserved in peptides that are presented by specific HLA molecules. Only a fraction of the peptides that retain an anchor motif will actually bind to the appropriate HLA molecule (for an example, see Altfeld2001), and only a small fraction of those peptides that bind to HLA molecules will actually be properly processed and presented and stimulate specific T-cells responses. So the presence of an anchor motif only suggests the possibility of an epitope that could be presented by a specific HLA molecule. Knowledge of the peptide-binding motif can be very useful, however, for example in fine mapping of novel epitopes when one has identified a CTL response against a longer peptide that contains an epitope.

We have updated our HLA binding motif scanner tool [Thakallapally2001] (Motif Scan) to include a more comprehensive list of known anchor motifs. The interface has been improved, and now includes the ability to track anchor motifs in alignments of proteins, not just single proteins; and to differentiate between C-terminal and interior anchor motif residues. Also, we are working on the feature that would allow to identify all available motifs within protein fragments of up to 100 amino acids in length. Some basic reference sets of HIV proteins can be automatically entered for searches. The general purpose of this tool is to help users to identify known anchor residue motifs for epitopes presented by class I and class II HLA molecules, and then use these motifs to highlight potential epitopes within a protein sequence. Although the automated search capabilities of Motif

[a]Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico.

[b]HIV Pathogenesis Programme, University of KwaZuluNatal, Durban, South Africa

[c]Department of Paediatrics, Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, Oxford OX1 3SY, UK

In *HIV Immunology and HIV/SIV Vaccine Databases 2003.* Bette T. M. Korber, Christian Brander, Barton F. Haynes, Richard Koup, John P. Moore, Bruce D. Walker, and David I. Watkins, editors. Publisher: Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR 04-8162. pp. 25–36.

Scan are based on HIV proteins, it can be applied to any protein. The tool could also be applied to look for general functionally important motifs of amino acid or nucleotides in sequences that have characteristic spacing. As the program is searching sequences of letters for motifs of interest, it would work equally well for identifying repeated patterns in proteins and in DNA. For example, if one put the restriction enzyme BamHI site `[G]-[G]-[A]-[T]-[C]-[C]` in as a custom motif and searched a DNA alignment, BamHi sites would be highlighted within the alignment.

I-B-2 HLA anchor motif sources

Primary anchor motif sources

The main data source for Motif Scan is a database of HLA binding motifs stored on our website. The HLA binding motif database has been recently updated to include two major motif libraries from *The HLA Facts Book* [Marsh2000] (`http://www.anthonynolan.com/HIG/`) and *MHC Ligands and Peptides Motifs* [Rammensee1997] (`http://syfpeithi.bmi-heidelberg.com`). We also searched the literature for the new motifs, not yet listed in these two major sources. What we found in the primary literature is presented as an additional source. Because motifs presented in different sources sometimes differ, in our tool the motifs presented are listed along with their sources, and the user can choose which ones to use for scanning the protein sequences for specific motif patterns. A number of HLA class I and class II alleles have not been characterized with respect to their peptide binding motif, and so we will continue to periodically update the database as new motifs are defined.

Table I-B-1: HLA motifs that are predicted based on conserved patterns found in a minimum of two optimal HIV epitopes in positions P2 or C-term (see Frahm2004, page 3 this volume, for lists of well defined epitopes presented by these HLA's).

Allele	Anchor motif	Number of epitopes that share this pattern
A*2501	`xxxxxxxxx[W]`	2
A*3201	`x[I]xxxxxxx[W]`	2
A*6802	`x[TV]xxxxxx[VL]`	3
B*4002	`x[E]xxxxxx[IAVL]`	5
B*4201	`x[P]xxxxxx[L]`	3
B*8101	`x[P]xxxxxx[L]`	2

Additional proposed anchor motifs

We have added several additional possible motifs, listed in Table I-B-1, based on conserved patterns in known optimal epitopes in HIV proteins [Frahm2004]. These motifs are not well established, and we will update the suggested motifs as they become better characterized.

Predicting HLA-C motifs by genetic similarities in HLA molecules

Although many of the A and B alleles have had their motif described, very few of the anchor motifs for HLA-C alleles have been characterized to date. As an increasing number of HLA-C-restricted CTL responses are being characterized, we have sought to predict the peptide-binding motif of the currently uncharacterized HLA-C alleles by comparing them to pockets with similar residues. Table I-B-2 shows the positions of amino acids lining the pockets of MHC class I molecules. Predicting the motif is straightforward if there is an identical or similar molecule that has been previously described, because a pocket having the same amino acid side chains can be expected to bind similar residues. Previously described F pockets of Cw*0102 and Cw*0304, for example, are identical, except for a leucine to isoleucine change at amino acid 95 and their motifs have both been found to bind small hydrophobic residues. Characteristics such as size, polarity and hydrophobicity (Table I-B-3) of the amino acid side-chains that line the pocket directly influence the motif. For example, a pocket with an overall strong positive charge will be expected to bind residues with negatively charged side chains, whereas a strongly hydrophobic pocket, where the constituent residues have bulky aromatic side chains, may often bind a small hydrophobic residue in that pocket. Based on this rationale, we have made predictions regarding potential anchor motifs for HLA-C subtype proteins that do not otherwise have a defined anchor motif. For the purposes of prediction we have focused on the principal pockets, B and F, although occasionally alleles have dominant anchor residues in other pockets. Elucidation of Cw*0102 has shown, for example, that residue 3 is an important anchor in the D pocket, and the anchor motif for Cw*0102 includes a proline at residue 3.

Tables I-B-4 and I-B-5 compare the variation in key binding residues along the B and F pockets of HLA-C alleles respectively, and, on the basis of similarity and known anchor motifs, make predictions regarding additional motifs. Table I-B-6 provides the predicted HLA-C summary motifs for B and F pockets.

Here is an example of our approach for motif prediction. The motif for the molecule Cw*0304 has previously been described. In pocket B (Table I-B-4) the residues have canceling charges and several large hydrophobic aromatic

Table I-B-2: The positions of the amino acids lining the HLA-C class I peptide binding site.

Pocket	Constituent residues	Peptide position accomodated
A	5, 7, 59, 63, 66, 99, 159, 163, 167, 171	1
B	7, 9, 24, 25, 34, 45, 63, 66, 67, 70, 99	2
C	9, 70, 73, 74, 97	6
D	99, 113, 114, 155, 156, 159, 160	3
E	97, 114, 147, 152, 156	7
F	77, 80, 81, 84, 95, 116, 123, 143, 146, 147	Carboxy terminus

Table I-B-3: Amino acid categories.

Category of amino acid	Members in group
Non-polar, hydrophobic	A, V, I, L, M
Non-polar, hydrophobic, large aromatic ring structure	F, W, Y
Non-polar, hydrophobic, small	P
Non-polar, hydrophilic	G, S, T, C, N
Positively charged, hydrophilic	R, H, K
Negatively charged, hydrophilic	D, E

residues. This pocket has been found in Cw*0304 to prefer the small hydrophobic residue alanine, and based on the B pocket similarity, we predicted alanine to be the B pocket anchor residue for Cw*0302 and Cw*0303.

The F pocket (Table I-B-5) for these alleles is likely to differ, however, since at position 116 there is a change from tyrosine to serine in Cw*0302 compared to Cw*0304 and Cw*0303. This substantial change would be expected to result in a more spacious F pocket for Cw*0302 than for Cw*0303 and Cw*0304. Thus, since for Cw*0304, the published motif at the F pocket is a medium-sized hydrophobic residue (L or M), for Cw*0302 we would predict a larger hydrophobic residue such as F, W or Y.

We are currently testing the validity of these predictions. For example, we have identified six HIV-1 relevant HLA-Cw*18 restricted responses toward 18-mer peptides. Using the motifs predicted above we can now compare the predicted optimal epitopes with actual epitopes.

Table I-B-4: Anchor motif predictions for the HLA-C locus B pocket. For each allele, the amino acids lining the pocked are shown along with the published or predicted motif. Amino acids at positions 9, 45, 63 and 67, shown in bold, are believed to be particularly important in determining the resulting motif.

Allele	Binding pocket residues											Published	Predicted	Comments
	7	**9**	24	25	34	**45**	**63**	66	**67**	70	99	Motif	Motif	
Cw*0102	Y	**F**	S	V	V	**G**	**E**	K	**Y**	Y	C	A,L		
Cw*0103	Y	**F**	S	V	V	**G**	**E**	K	**Y**	Y	C		A,L	Pocket is identical to Cw*0103
Cw*0202	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0203	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0302	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0303	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0304	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y	A		Strong preference for small hydrophobic residues
Cw*0305	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0306	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0307	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0308	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0309	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0401	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F	Y,P		
Cw*0402	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F		Y,P	Pocket is identical to Cw*0401
Cw*0403	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	F		P	Pocket is similar to Cw*0401 but smaller S to Y change
Cw*0404	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F		Y,P	Pocket is identical to Cw*0401
Cw*0405	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F		Y,P	Pocket is identical to Cw*0401
Cw*0406	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	F		P	Pocket is similar to Cw*0401but smaller S to Y change
Cw*0501	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0502	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0602	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	Y		R,Q	Pocket is negatively charged predict positive residues
Cw*0603	Y	**Y**	S	V	V	**G**	**E**	K	**Y**	Q	Y		A,L,P	Pocket is similar to Cw*0602 but less negatively charged
Cw*0604	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	Y		R,Q	Pocket is identical to Cw*0602
Cw*0701	Y	**D**	S	V	V	**G**	**E**	N	**Y**	Q	Y		R,H,K	Pocket is similar to Cw*0602 but more negatively charged
Cw*0702	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	S	Y,P		
Cw*0703	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	S		Y,P	Pocket is identical to Cw*0702
Cw*0704	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	Y		R,Q	Pocket is identical to Cw*0602
Cw*0705	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	Y		R Q	Pocket is identical to Cw*0602
Cw*0706	Y	**D**	S	V	V	**G**	**E**	N	**Y**	Q	Y		R,H,K	Pocket is similar to Cw*0602 but more negatively charged
Cw*0707	Y	**D**	S	V	V	**G**	**E**	N	**Y**	Q	Y		R,H,K	Pocket is similar to Cw*0602 but more negatively charged
Cw*0708	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	F		R,Q	Pocket is almost identical to Cw*0602
Cw*0709	Y	**D**	S	V	V	**G**	**E**	N	**Y**	Q	Y		R,H,K	Pocket is similar to Cw*0602 but more negatively charged

Table I-B-4: Anchor motif predictions for the HLA-C locus B pocket (cont.)

Allele	Binding pocket residues											Published	Predicted	Comments
	7	**9**	24	25	34	**45**	**63**	66	**67**	70	99	Motif	Motif	
Cw*0710	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	S		Y,P	Pocket is identical to Cw*0702
Cw*0711	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	Y		R,Q	Pocket is identical to Cw*0602
Cw*0712	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	Y		R,Q	Pocket is identical to Cw*0602
Cw*0801	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0802	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0803	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0804	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0805	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*0806	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1202	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1203	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1204	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1205	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1206	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1402	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F		Y,P	Pocket is identical to Cw*0401
Cw*1403	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F		Y,P	Pocket is identical to Cw*0401
Cw*1404	Y	**S**	A	V	V	**G**	**E**	K	**Y**	Q	F		Y,P	Pocket is identical to Cw*0401
Cw*1502	Y	**Y**	A	V	V	**G**	**E**	N	**Y**	Q	Y		A	Similar to Cw*0304 but less positively charged K to N change
Cw*1503	Y	**Y**	A	V	V	**G**	**E**	N	**Y**	Q	Y		A	Similar to Cw*0304 but less positively charged K to N change
Cw*1504	Y	**Y**	A	V	V	**G**	**E**	N	**Y**	Q	Y		A	Similar to Cw*0304 but less positively charged K to N change
Cw*1505	Y	**Y**	A	V	V	**G**	**E**	N	**Y**	Q	Y		A	Similar to Cw*0304 but less positively charged K to N change
Cw*1506	Y	**Y**	A	V	V	**G**	**E**	N	**Y**	Q	Y		A	Similar to Cw*0304 but less positively charged K to N change
Cw*1507	Y	**Y**	A	V	V	**G**	**E**	N	**Y**	Q	Y		A	Similar to Cw*0304 but less positively charged K to N change
Cw*1601	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1602	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1604	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1701	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1702	Y	**Y**	A	V	V	**G**	**E**	K	**Y**	Q	Y		A	Pocket is identical to Cw*0304
Cw*1801	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	F		R,Q	Pocket is similar to Cw*0602 except Y to F change
Cw*1802	Y	**D**	S	V	V	**G**	**E**	K	**Y**	Q	F		R,Q	Pocket is similar to Cw*0602 except Y to F change

Table I-B-5: Anchor motif predictions for the HLA-C locus F pocket. For each allele, the amino acids lining the pocked are shown along with the published or predicted motif. Amino acids at positions 77, 80, 81, 95 and 116, shown in bold, are believed to be particularly important in determining the resulting motif.

Allele	Binding pocket residues										Published	Predicted	Comments
	77	**80**	**81**	84	**95**	**116**	123	143	146	147	Motif	Motif	
Cw*0102	**S**	**N**	**L**	Y	**L**	**Y**	Y	T	K	W	L		Strong preference for small hydrophobic residues
Cw*0103	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L	Pocket almost identical to Cw*0102 except Y to F change
Cw*0202	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*0203	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*0302	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket similar to Cw*0304 except Y to S change
Cw*0303	**Y**	**N**	**L**	Y	**I**	**Y**	Y	T	K	W		L,M	Pocket is identical to Cw*0304
Cw*0304	**S**	**N**	**L**	Y	**I**	**Y**	Y	T	K	W	L,M		Strong preference for hydrophobic residues
Cw*0305	**S**	**N**	**L**	Y	**I**	**Y**	Y	T	K	W		L,M	Pocket is identical to Cw*0304
Cw*0306	**S**	**N**	**L**	Y	**I**	**Y**	Y	T	K	W		L,M	Pocket is identical to Cw*0304
Cw*0307	**N**	**K**	**L**	Y	**I**	**Y**	Y	T	K	W		L,F	Pocket is similar to Cw*0401
Cw*0308	**S**	**N**	**L**	Y	**I**	**Y**	Y	T	K	W		L,M	Pocket is identical to Cw*0304
Cw*0309	**S**	**N**	**L**	Y	**I**	**Y**	Y	T	K	W		L,M	Pocket is identical to Cw*0304
Cw*0401	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W	L	F	Strong preference for hydrophobic residues
Cw*0402	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0403	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0404	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0405	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0406	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0501	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0502	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,F	Pocket is identical to Cw*0401
Cw*0602	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W	L		
Cw*0603	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*0604	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*0701	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	L		Y	Pocket is identical to Cw*0702
Cw*0702	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	L	Y		
Cw*0703	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		Y,L	Pocket is similar to Cw*0702 but smaller L to W change
Cw*0704	**S**	**N**	**L**	Y	**F**	**F**	Y	T	K	L		L,M	Pocket is similar to Cw*0304
Cw*0705	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	L		Y	Pocket is identical to Cw*0702
Cw*0706	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	L		Y	Pocket is identical to Cw*0702
Cw*0707	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	L		Y,L	Pocket is similar to Cw*0602 but larger
Cw*0708	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	L		Y,L	Pocket is identical to Cw*0702
Cw*0709	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	L		Y,L	Pocket is similar to Cw*0602 but larger

Table I-B-5: Anchor motif predictions for the HLA-C locus F pocket (cont.)

Allele	Binding pocket residues										Published	Predicted	Comments
	77	**80**	**81**	84	**95**	**116**	123	143	146	147	Motif	Motif	
Cw*0710	**S**	**N**	**L**	Y	**I**	**S**	Y	T	K	L		F,W,Y	Pocket is similar to Cw*0302
Cw*0711	**S**	**N**	**L**	Y	**F**	**F**	Y	T	K	L		L,M	Pocket is similar to Cw*0304 but smaller I to F change
Cw*0712	**S**	**N**	**L**	Y	**F**	**F**	Y	T	K	W		L,M	Pocket is similar to Cw*0304 but smaller I to F change
Cw*0801	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L,M	Pocket is almost identical to Cw*0304
Cw*0802	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L,M	Pocket is almost identical to Cw*0304
Cw*0803	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L,M	Pocket is almost identical to Cw*0304
Cw*0804	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L,M	Pocket is almost identical to Cw*0304
Cw*0805	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L,M	Pocket is almost identical to Cw*0304
Cw*0806	**S**	**N**	**L**	Y	**L**	**F**	Y	T	K	W		L,M	Pocket is almost identical to Cw*0304
Cw*1202	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1203	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1204	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*1205	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*1206	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1402	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1403	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1404	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1502	**N**	**K**	**L**	Y	**I**	**L**	Y	T	K	W		L,M,Y,F	Pocket is similar to Cw*0401 except L to I change and F to L change
Cw*1503	**N**	**K**	**L**	Y	**I**	**L**	Y	T	K	W		L,M,Y	Pocket is similar to Cw*0401 except L to I change and F to L change
Cw*1504	**N**	**K**	**L**	Y	**I**	**S**	Y	T	K	W		L	Pocket is almost identical Cw*0602 but smaller L to I change
Cw*1505	**N**	**K**	**L**	Y	**I**	**F**	Y	T	K	W		L	Pocket is similar to Cw*0202 but smaller S to F change
Cw*1506	**N**	**K**	**L**	Y	**I**	**Y**	Y	T	K	W		L,M	Pocket is similar to Cw*0602 but smaller S to Y change
Cw*1507	**S**	**N**	**L**	Y	**I**	**L**	Y	T	K	W		L,M,Y	Pocket is similar to Cw*0304 but larger Y to L change
Cw*1601	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		F,W,Y	Pocket is identical to Cw*0302
Cw*1602	**N**	**K**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0602
Cw*1604	**S**	**N**	**L**	Y	**L**	**S**	Y	T	K	W		L	Pocket is identical to Cw*0302
Cw*1701	**N**	**K**	**L**	Y	**I**	**F**	Y	S	K	L		L	Pocket similar to Cw*0602
Cw*1702	**N**	**K**	**L**	Y	**I**	**F**	Y	S	K	L		L	Pocket similar to Cw*0602
Cw*1801	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,Y	Pocket is identical to Cw*0401
Cw*1802	**N**	**K**	**L**	Y	**L**	**F**	Y	T	K	W		L,Y	Pocket is identical to Cw*0401

Table I-B-6: Anchor motif predictions for the HLA-C locus. The motifs for Cw*0102, Cw*0304, Cw*0401, Cw*0602 and Cw*0702 are previously published and are shown in bold.

Allele	Anchor motif
Cw*0102	x[**AL**]xxxxxx[**L**]
Cw*0103	x[AL]xxxxxx[L]
Cw*0202	x[A]xxxxxx[L]
Cw*0203	x[A]xxxxxx[L]
Cw*0302	x[A]xxxxxx[FWY]
Cw*0303	x[A]xxxxxx[LM]
Cw*0304	x[**A**]xxxxxx[**LM**]
Cw*0305	x[A]xxxxxx[LM]
Cw*0306	x[A]xxxxxx[LM]
Cw*0307	x[A]xxxxxx[LF]
Cw*0308	x[A]xxxxxx[LM]
Cw*0309	x[A]xxxxxx[LM]
Cw*0401	x[**YP**]xxxxxx[**LF**]
Cw*0402	x[YP]xxxxxx[LF]
Cw*0403	x[P]xxxxxx[LF]
Cw*0404	x[YP]xxxxxx[LF]
Cw*0405	x[YP]xxxxxx[LF]
Cw*0406	x[P]xxxxxx[LF]
Cw*0501	x[A]xxxxxx[LF]
Cw*0502	x[A]xxxxxx[LF]

Allele	Anchor motif
Cw*0602[1]	x[**RQ**]xxxxxx[**L**]
Cw*0603	x[ALP]xxxxxx[L]
Cw*0604	x[RQ]xxxxxx[L]
Cw*0701	x[RHK]xxxxxx[Y]
Cw*0702	x[**YP**]xxxxxx[**Y**]
Cw*0703	x[YP]xxxxxx[YL]
Cw*0704	x[RQ]xxxxxx[LM]
Cw*0705	x[RQ]xxxxxx[Y]
Cw*0706	x[RHK]xxxxxx[Y]
Cw*0707	x[RHK]xxxxxx[YL]
Cw*0708	x[RQ]xxxxxx[YL]
Cw*0709	x[RHK]xxxxxx[YL]
Cw*0710	x[YP]xxxxxx[FWY]
Cw*0711	x[R]xxxxxx[LM]
Cw*0712	x[R]xxxxxx[LM]
Cw*0801	x[A]xxxxxx[LM]
Cw*0802	x[A]xxxxxx[LM]
Cw*0803	x[A]xxxxxx[LM]
Cw*0804	x[A]xxxxxx[LM]
Cw*0805	x[A]xxxxxx[LM]
Cw*0806	x[A]xxxxxx[LM]

Allele	Anchor motif
Cw*1202	x[A]xxxxxx[FWY]
Cw*1203	x[A]xxxxxx[FWY]
Cw*1204	x[A]xxxxxx[L]
Cw*1205	x[A]xxxxxx[L]
Cw*1206	x[A]xxxxxx[FWY]
Cw*1402	x[YP]xxxxxx[FWY]
Cw*1403	x[YP]xxxxxx[FWY]
Cw*1404	x[YP]xxxxxx[FWY]
Cw*1502	x[A]xxxxxx[LMYF]
Cw*1503	x[A]xxxxxx[LMYF]
Cw*1504	x[A]xxxxxx[L]
Cw*1505	x[A]xxxxxx[L]
Cw*1506	x[A]xxxxxx[LM]
Cw*1507	x[A]xxxxxx[LMY]
Cw*1601	x[A]xxxxxx[FWY]
Cw*1602	x[A]xxxxxx[L]
Cw*1604	x[A]xxxxxx[L]
Cw*1701	x[A]xxxxxx[L]
Cw*1702	x[A]xxxxxx[L]
Cw*1801	x[RQ]xxxxxx[LY]
Cw*1802	x[RQ]xxxxxx[LY]

[1] The published motif for Cw*0602 is xxxxxxxx[L].

I-B-3 Using Motif Scan

Data dictionaries

The HLA genotype/serotype classification used in the tool was based on the *The HLA Facts Book* [Marsh2000] and checked against more recent sources [Marsh2002, Schreuder2001]. Motif Scan can also be used as a quick web-based reference to look up associated HLA genotype/serotype nomenclature. Also, we added class I supertype classification and binding supermotifs from work of Sette and Sidney [Sette1999].

Search fields

The main web page of Motif Scan contains links to the downloadable lists of HLA genotypes, serotypes and supertypes, choice for motif length, window for the custom motif and motif sources. Here are detailed explanations for these search fields

HLAs The user can select multiple HLA class I or class II alleles to search the database for known or predicted motifs. (To find motifs for more than one allele, use the mouse to click on the first choice, and then hold down the control key while clicking on additional alleles for the search). HLA alleles may be specified by genotype, serotype or supertype.

Genotype HLA genotypes are in general specified by four digit number, for example A*0201.

Serotype If a user selects a serotype as a search field, anchor residues that have been defined for all related genotypes will be returned. For example, A2 will return motifs for the set of A2 related genotypes, including: A*0201, A*0202, etc. If a user knew only the serotype of the individual, it might be useful to have the anchor residues for all related genotypes displayed; if the specific HLA genotype was defined, then the specific anchor residues would be of greater interest. Because of the way HLA nomenclature has evolved, occasionally a two-digit HLA specification will be related to a genotype in a non-intuitive manner. For example, the genotype B*1513 might be specified by the serotype B15 or B77, where B15 is inclusive of many B*15 genotypes, and B77 is specifically B*1511. The serotype-genotype dictionary link on the main page provides a quick reminder of the naming conventions and relationships.

Supertype Supermotifs will be provided for a given supertype, as defined by Sette and Sidney [Sette1999].

Motif Source Multiple sources (see above) may be selected to search for the variants of the suggested binding motif.

Motif Syntax The anchor residues are shown in the square brackets. The preferred but not dominant amino acids in the anchor positions are shown in parentheses. For example, motif for A*2602 taken from SYFPEITHY library is `x-[VTILF]-x-x-x-x-x-x-[YF(ML)]`. This means that second and C-terminal positions are anchor positions. The dominant amino acids at the second position are `V`, `T`, `I`, `L`, and `F`. At the C-terminal anchor position the dominant amino acids are `Y` and `F`, while `M` and `L` are also found, but not as commonly, among A*2602 epitopes. Note that as a default, Motif Scan will search on all possible anchor position amino acids, both dominant and preferred. It is possible to restrict a search to the dominant amino acids only, by composing a custom motif excluding the amino acids in parentheses (see below).

Supermotif Syntax For the supermotifs, residues within brackets are residues predicted to be tolerated and sometimes cross-presented by multiple Class I molecules within the putative supertype.

Motif Length Motifs are stored in the database with a length of 9 amino acids and the motifs for other lengths (8, 10 or 11 amino acids) are computed on-the-fly by adding or removing amino acids in front of the C terminus. Lengths are adjusted only for motifs from HLA class I genotypes, serotypes and supertypes. Anchor motifs for class II HLAs and custom motifs are not adjusted in length, and often for class II epitopes the last anchor residue of the motif will be embedded within the epitope, and not be the C-terminal amino acid.

Custom Motif A helpful feature of Motif Scan is the ability for users to define custom motifs. The syntax for the custom motif is the same as for the database motifs. Positions where several amino acids may be possible include all amino acids listed within square brackets [], and an `x` denotes arbitrary residues that specify the spacing in a motif. One can optionally use a dash (`-`) to separate the residues. For example, `x[LM]xxx[K]xx[V]` or `x-[LM]-x-x-x-[K]-x-x-[V]` are equivalent. With regard to defining epitopes, this feature, for example, allows the user to restrict a search to dominant anchor residues, or to add in auxiliary residues that are available in the listings of Marsh2000 and Rammensee1997. Auxiliary residues in epitopes do not bind directly in the pockets of HLA proteins that hold the anchor residues, but still influence the ability of a peptide to bind; auxiliary residues are not included in the Motif Scan database. The custom motif feature is useful also for the purposes other than scanning sequences for the HLA binding motifs. One can scan query sequences for any pattern of interest, for example, studying motifs with characteristic functions found in functional domains.

Sequences and sequence alignments

To scan for a binding motif, a user can select from our predefined HIV protein sequences or upload or copy and paste their own sequences.

Predefined Sequences The predefined sequences include either all proteins from the HIV-1 HXB2 reference strain, or alignments of consensus and ancestral protein sequences for M group (subtypes A-D, F-H, J, K and circulating recombinant forms CRF01, CRF02) and O Group of HIV-1, from Los Alamos HIV Sequence Database, 2002 [Kuiken2003]. The alignments can give a rapid assessment of how well the search motif is preserved among the diverse forms in the epidemic. This will be updated periodically, as consensus and ancestral sequences are updated.

User's Own Sequences The user can upload or cut and paste either an individual protein sequence, or a set of individual sequences, or an alignment of sequences. User sequences should be input as simple text files in FASTA or TABLE format. Examples of sequence formatting can be obtained by clicking on the links provided on the web page. Individual sequences are stripped of gaps before processing. Gaps inserted to maintain the alignment are specified with a dash (-). If the sequences are aligned and the user wants to preserve the alignment, the user should check "Yes" in the box "Are the input sequences aligned"? In this case the gaps will be maintained and the alignment will be preserved, but they will not be counted as characters in terms of motif spacing.

Output

The results of the program are presented in several stages. First, the motifs corresponding to the input HLA types are presented. Then, the user chooses which motifs or set of motifs to scan with, chooses motif length (between 8-11 amino acids, one can select all possible lengths to be comprehensive), uploads query sequences or chooses predefined sequences, and initiates the scanning of these sequences for the respective motifs.

The final output is organized by search pattern, and all motifs with identical search patterns are grouped together. The matching motifs are presented on the input sequences in two colors: C-terminal anchor amino acids are shown in magenta and anchor amino acids in the other positions are shown in cyan. If a given amino acid is matched by more than one motif, then it is highlighted as a C-terminal anchor amino acid if any of the motifs are matched at the C-terminal anchor. All motif amino acids are shown in uppercase and non-anchors are lowercase. Following the sequences is a list of potential epitopes showing their positions in the input sequences.

The output can be viewed and downloaded in a format convenient for further coding and analysis: sequences can be downloaded in the FASTA format where the anchor amino acids are presented in uppercase and all the remaining ones in lowercase, but the colors are omitted. Alternatively, the color-highlighted motifs can be retained and copied and pasted from the web page directly into a word processor file. The list of potential epitopes and their positions in the protein can be downloaded in CSV (comma-separated value) format, which can be read into a spreadsheet.

Example

The motif search on all motif libraries for HLA A*0214 reveals the following table of motifs:

Genotype	Serotype	Motif	Source	Scan?
A*0214	A2	x-[QV]-x-x-x-[K]-x-x-[VL]	Luscher2001	☑
A*0214	A2	x-[VQ(L)]-x-x-x-x-x-x-[L]	Marsh2000	☑
A*0214	A2	x-[VQL(A)]-x-x-x-x-x-x-[L(VM)]	SYFPEITHI	☑

(x can be any amino acid). Here the three sources listed give slightly different information. In Luscher2001, either Q or V is favored in the second position, K in the sixth position, and either V or L is favored in the C-terminal position. In the Marsh2000 and SYFPEITHI sources, only the second and C-terminal positions are listed as anchor positions. In Marsh2000, the second position favors either V or Q, or less frequently L, and the C-terminal position favors L. In SYFPEITHI, the second position favors either V or Q or L, or less frequently A, and the C-terminal position favors L, or less frequently V or M. We do not address this different sources motif discrepancy in Motif Scan, rather, we display all existing views and let the user decide which one (or more) to use for scanning the protein sequences. Additionally, the user can compose a custom motif that would combine or summarize the information we present.

Once the motif is chosen, the user can either use our predefined consensus and ancestral sequence alignments, or copy-paste or upload their own sequences. For example, we scanned the following two aligned test sequences to search for the Marsh2000 A*0214 motif x-[VQ(L)]-x-x-x-x-x-x-[L]:

```
>Test.1
KTIIFKVSSQGDPLIVLHSQN--LEFLYCNLTKLFNSTW
>Test.2
KTI--KKSSQGDPEIVLHSQNCGGEFLHCNSTQFFNSTW
```

The resulting output contains query sequences in lowercase letters, where the anchor residues are uppercase and colored: C-terminal anchor is magenta and the other anchor is cyan[2], and the list of potential epitopes based on these anchor motifs:

```
Test.1    ktiifkVssQ gdpLivLhsq n--LefLycn LtkLfnstw
Test.2    kti--kkssQ gdpeivLhsQ ncggefLhcn stqffnstw
```

Protein	Seq. Pos.	Aln. Pos.	Sequence	Anchors
Test.1	6-14	6-14	KVSSQGDPL	.V......L
Test.1	9-17	9-17	SQGDPLIVL	.Q......L
Test.1	21-29	21-31	NLEFLYCNL	.L......L
Test.1	24-32	26-34	FLYCNLTKL	.L......L
Test.2	7-15	9-17	SQGDPEIVL	.Q......L
Test.2	17-25	19-27	SQNCGGEFL	.Q......L

I-B-4 Summary and future directions

We envision several applications for Motif Scan. It is particularly useful for the situations when a CTL response is partially characterized from an individual with a known HLA type, and already localized to a protein or protein region. The presence of HLA appropriate anchor residues could help focus the search for potential epitopes in known reactive protein regions. These anchor residues have to be considered with caution however, as anchor residues are frequently present in regions with no reactive epitopes, and there are many true epitopes that do not contain the anchor residue motifs. This simple tool is useful, however, as an initial indication for potential epitopes. Many external tools for HLA binding predictions are also available. These include methods based on "extended" motifs including secondary anchors and disfavored residues and statistical matrices representing the weight of every amino acid in every position [De Groot1997, Parker1995, Rammensee1997] and relatively new methods based on artificial neural networks (ANNs) [Buus2003, Milik1998, Schönbach2002], which, in contrast to motif based methods, are well suited to recognize complex nonlinear sequence-dependent correlated effects.

We are working on adding several more features to Motif Scan in the near future. One useful feature would be to identify all possible HLA motifs listed in our database for a peptide. Currently it is possible to do so by scanning the protein sequence for all HLA motifs. However the current output then contains results for both positive matches (those HLAs that have anchor residues in the query peptide) and negative ones (those HLAs that do not), and is a little cumbersome to navigate. We plan in the future to have a special field for this kind of search and have output showing only positive matches in a convenient form.

Motif Scan is useful not only for HIV sequences, but for any protein or DNA sequence; the only feature of the tool that is specifically tailored for HIV is the ready availability of predefined HIV sequences. We are in the process of adapting this tool for the hepatitis C database website (`http://hcv.lanl.gov/`), where we will link Motif Scan with the hepatitis C consensus sequences as the predefined sequences.

I-B-5 References

[Altfeld2001] M. A. Altfeld, B. Livingston, N. Reshamwala, P. T. Nguyen, M. M. Addo, A. Shea, M. Newman, J. Fikes, J. Sidney, P. Wentworth, R. Chesnut, R. L. Eldridge, E. S. Rosenberg, G. K. Robbins, C. Brander, P. E. Sax, S. Boswell, T. Flynn, S. Buchbinder, P. J. Goulder, B. D. Walker, A. Sette, & S. A. Kalams. Identification of novel HLA-A2-restricted human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte epitopes predicted by the HLA-A2 supertype peptide-binding motif. *J Virol* **75**(3):1301–1311, 2001. On p. 25

[Buus2003] S. Buus, S. L. Lauemøller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, & S. Brunak. Sensitive quantitative predictions of peptide-MHC binding by a 'query by committee' artificial neural network approach. *Tissue Antigens* **62**(5):378–384, 2003. On p. 35

[Calef2001] C. Calef, R. Thakallapally, D. Lang, C. Brander, P. Goulder, O. Yang, & B. Korber. PeptGen: Designing peptides for immunological studies and application to HIV consensus sequences. In B. Korber, C. Brander, B. Haynes, R. Koup, C. Kuiken, J. P. Moore, B. D. Walker, & D. I. Watkins, eds., *HIV Molecular Immunology 2000*, pp. I-63–I-100. Los Alamos National Laboratory, Theoretical Biology & Biophysics, Los Alamos, New Mexico, 2001. LA-UR 01-2430. On p. 25

[Calef2002] C. Calef, R. Thakallapally, R. Kaslow, M. Mulligan, & B. Korber. ELF: An analysis tool for HIV-1 peptides and HLA types. In B. Korber, C. Brander, B. Haynes, R. Koup, C. Kuiken, J. P. Moore, B. D. Walker, & D. I. Watkins, eds., *HIV Molecular Immunology 2001*, pp. I-21–I-25. Los Alamos National Laboratory, Theoretical Biology & Biophysics, Los Alamos, New Mexico, 2002. LA-UR 02-4663. On p. 25

[2] For better visibility in print, the example shows the C-terminal anchors in a black box (X) and shows the other anchors in a gray box (X).

[De Groot1997] A. S. De Groot, B. M. Jesdale, E. Szu, J. R. Schafer, R. M. Chicz, & G. Deocampo. An interactive web site providing major histocompatibility ligand predictions: Application to HIV research. *AIDS Res Hum Retroviruses* **13**(7):529–351, 1997. On p. 35

[Frahm2004] N. Frahm, P. J. Goulder, & C. Brander. Broad HIV-1 specific CTL responses reveal extensive HLA class I binding promiscuity of HIV-derived, optimally defined CTL epitopes. In B. Korber, C. Brander, B. Haynes, R. Koup, J. P. Moore, B. D. Walker, & D. I. Watkins, eds., *HIV Molecular Immunology 2002*, p. ???? Los Alamos National Laboratory, Theoretical Biology & Biophysics, Los Alamos, New Mexico, 2004. LA-UR 03-5816. On p. 26

[Kuiken2003] C. L. Kuiken, B. Foley, E. Freed, B. Hahn, P. A. Marx, F. McCutchan, J. W. Mellors, S. Wolinksy, & B. Korber, eds. *HIV Sequence Compendium 2002*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.M., 2003. LA-UR 03-3564. On p. 25, 34

[Marsh2000] S. G. E. Marsh, P. Parham, & L. D. Barber. *The HLA FactsBook*. Academic Press, San Diego, 2000. On p. 26, 33

[Marsh2002] S. G. E. Marsh, E. D. Albert, W. F. Bodmer, R. E. Bontrop, B. Dupont, H. A. Erlich, D. E. Geraghty, J. A. Hansen, B. Mach, W. R. Mayr, P. Parham, E. W. Petersdorf, T. Sasazuki, G. M. T. Schreuder, J. L. Strominger, A. Svejgaard, & P. I. Terasaki. Nomenclature for factors of the HLA system, 2002. *Tissue Antigens* **60**(5):407–464, 2002. On p. 33

[Milik1998] M. Milik, D. Sauer, A. P. Brunmark, L. Yuan, A. Vitiello, M. R. Jackson, P. A. Peterson, J. Skolnick, & C. A. Glass. Application of an artificial neural network to predict specific class I mhc binding peptide sequences. *Nat Biotechnol* **16**(8):753–756, 1998. On p. 35

[Parker1995] K. C. Parker, M. Shields, M. DiBrino, A. Brooks, & J. E. Coligan. Peptide binding to MHC class I molecules: Implications for antigenic peptide prediction. *Immunol Res* **14**(1):34–57, 1995. On p. 35

[Rammensee1997] H.-G. Rammensee, J. Bachmann, & S. Stevanović. *MHC Ligands and Peptide Motifs*. Landes Bioscience, Georgetown, Texas, 1997. On p. 26, 33, 35

[Schönbach2002] C. Schönbach, Y. Kun, & V. Brusic. Large-scale computational identification of HIV T-cell epitopes. *Immunol Cell Biol* **80**(3):300–306, 2002. On p. 35

[Schreuder2001] G. M. T. Schreuder, C. K. Hurley, S. G. E. Marsh, M. Lau, M. Maiers, C. Kollman, & H. J. Noreen. The HLA Dictionary 2001: A summary of HLA-A, -B, -C, -DRB1/3/4/5, -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR and -DQ antigens. *Tissue Antigens* **58**(2):109–140, 2001. On p. 33

[Sette1999] A. Sette & J. Sidney. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* **50**(3-4):201–212, 1999. On p. 33

[Thakallapally2001] R. Thakallapally, W. Kibbe, D. Lang, & B. Korber. Motifscan: A web-based tool to find HLA anchor residues in proteins or peptides. In B. Korber, C. Brander, B. Haynes, R. Koup, C. Kuiken, J. P. Moore, B. D. Walker, & D. I. Watkins, eds., *HIV Molecular Immunology 2000*, pp. I-101–I-102. Los Alamos National Laboratoy, Theoretical Biology & Biophysics, Los Alamos, New Mexico, 2001. LA-UR 01-2430. On p. 25