

ELF: An Analysis Tool for HIV-1 Peptides and HLA Types

Charles Calef¹, Rama Thakallapally², Richard A. Kaslow³,
Mark Mulligan⁴, Bette Korber¹

¹ *Theoretical Biology and Biophysics (T10), Los Alamos National Laboratory, Los Alamos, NM 87545*

² *Michigan Proteome Consortium, 4326 Med Sci I, University of Michigan, Ann Arbor, MI 48109*

³ *Program in Epidemiology of Infection and Immunity, University of Alabama at Birmingham, 220 Ryals Bldg., 1665 University Blvd., Birmingham, AL 35294-0022*

⁴ *Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, AL 35243*

SUMMARY

The Epitope Location Finder (ELF) site (http://hiv-web.lanl.gov/ALABAMA/epitope_analyzer.html) performs various analyses of HIV-1 peptide and protein sequences with the intention of facilitating the task of identifying optimal reactive CTL epitopes. The input requires some combination of: a reactive peptide that is known to stimulate CTL responses, the HLA type of the individual whose CTL response is under study, and the protein sequence that was used as a basis for peptide design. There are multiple output options, ranging from minimal to more complex and voluminous.

To summarize, ELF can:

- list the HLA genotypes associated with a submitted HLA serotypes, or vice versa, and create a table of related anchor residue motif patterns;
- scan a submitted HIV-1 amino acid string representing a peptide that is able to stimulate a CTL response for potential epitopes, based on anchor motifs for specified HLA molecules;
- find the HXB2 coordinates of a reactive peptide and extract an alignment of the peptide against the sequences in our master alignment sets;
- search our HIV epitope database for known epitopes within the boundaries of the submitted peptide;

- draw maps of HIV proteins showing location of epitopes, highlighted in red, that have HLA presenting molecules that agree with the submitted HLA;
- provide information about epitopes that might have been missed because the protein sequence used to design an overlapping peptide set differed in a known epitope.

INPUT SUBMISSION FORM

The input form page (Figure 1) has data input boxes, to be filled in by the user, and a list of possible outputs which will be computed when the output check boxes are checked. The output option called Summary tables is selected by default but may be turned off if desired.

Figure 1. The data input submission form for ELF. Input fields occupy the upper part of the screen, and output option check boxes are at the bottom. Each field label is a link to a file explaining its use.

ANALYSIS OPTIONS

The Patient ID and Peptide ID are optional inputs for the convenience of the user who may have several different data sets to analyze. They are simply printed, unchanged, in the output.

Associated HLAs

The Patient HLA input field can be filled in with one or more HLA types. Both serotypes (e.g., A2) and genotypes (e.g., B*1801) are acceptable. Multiple HLAs can be submitted if they are separated by a space character or comma, e.g., “B8 B60” or “B8, B60”. If patient HLA information is submitted, with the Summary tables output option selected, a simple analysis is run in which the B8 and B60 serotypes are “expanded” to include associated HLA genotypes and presented as a table (Table 1). The listing of known serotypes with their corresponding defined genotypes was primarily based on: “The HLA dictionary 1999: 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,” G. M. Th. Schreuder, C. K. Hurley, S. G. E. Marsh, M. Lau, M. Maiers, C. Kollman, H. Noreen. *Tissue Antigens*54:409–437 (1999).

Table 1 HLAs associated with users input

Submitted HLAs	Associated HLAs
B8 B60	B8, B*0801, B*0802, B*0803, B*0806, B60, Bw60, B*4001, B*40012

Potential Anchor Motifs

An Anchor Motifs table (Table 2) shows any anchor residue motifs associated with the HLAs submitted. In the motifs column a “.” character means any amino acid may occur at that position, while square brackets list the amino acids required at that position. For example in the B8 motifs below, a K or an R must occur at position 5. HLA anchor residue motifs used here are listed in *The HLA Fact Book* by S. Marsh, P. Parham, and L.Barber, published by the Academic Press, 2000, and *MHC Ligands and Peptide Motifs*, by H. G.

Rammensee, J. Bachmann, and S. Stevanovic, Chapman and Hall publishers, 1997. To find out more detailed information concerning HLA motifs, see the SYFPEITHI Web site at <http://syfpeithi.bmi-heidelberg.com>.

Similar output can be obtained using the program Motifscan; class II anchor motifs or proteins from other organisms can also be included in Motifscan searches, or when using some of the non-HIV specific tools in ELF.

Table 2 Anchor residue motif patterns associated with user-submitted HLAs

HLA	Anchor Residue Motifs
B8	.. [K].[KR]... [L]
B8	.. [K].[KR].. [L]
B8	.. [K].[KR].... [L]
B*40012	.[E]..... [L]
B*40012	.[E]..... [L]
B*40012	.[E]..... [L]

Note: An updated listing of serotypes, genotypes and anchor motifs including several new references is underway, we hope to have it available within the Elf and Motifscan programs by fall 2002.

To summarize this combination of input and output options:

Input	Output	Results
HLA only	Summary Tables only	Expanded list of “associated HLA” and table of anchor residue motifs associated with those HLAs

Potential Epitopes in User-supplied Peptide

The Peptide Sequence data entry box on the input form (Figure 1) is the place to enter a peptide of interest for analysis, such as an immunologically reactive peptide. This sequence, in conjunction with HLA information, and the Summary Tables output selection produces not only the “Associated HLAs” and “Anchor residue motifs” (Tables 1 & 2) shown above, but also identifies the location of this peptide in the HIV genome, and creates a table of “Possible Epitopes” lurking in the peptide (Table 3). The possible epitopes table looks like this for the submitted peptide HQREIKDSTKEALDKIEE and HLA B40 combination.

Position in query peptide	AA sequence	HLA
(3-12)	REIKDTKEAL	B40

The table shows that at position 3-12 of the hypothetical submitted sequence "HQREIKDTKEALDKIEE" there occurs an amino acid string which matches one of the B*40012 anchor residue motifs in the Anchor Motifs Table above.

```

REIKDTKEALdkiee    submitted peptide
| | | | | | | |      matches
.E . . . . . L     a B*40012 anchor motif
    
```

Note, that possible epitopes are just possibilities; they may well have never been observed, and may not occur in our database. They are intended to suggest possibilities for optimal epitopes within a larger reactive peptide, and may help streamline the screening process.

To summarize this combination of input and output options:

Input	Output	Results
HLA + peptide sequence	Summary Tables only	The tables listed above + a list of possible epitopes within the peptide sequence that match the anchor motifs

Alignment

The submitted Peptide Sequence will be aligned to the set of complete HIV database alignments if you select the Align peptide sequence output option. A portion of this alignment is shown in Figure 2.

The epitope alignment was created from [HIV-1 sequence web alignments](#). Sequences are labeled by: [subtype.country.sequenceID](#)

[View/save](#) the Table format alignment.

Query:	HQREIKDTKEALDKIEE
Query Length:	17
HXB2 Location:	p17(89-106)
Alignment:	Gag, 185 sequences

"." = gap in sequence
 "-" = identity to query sequence
 "RED" = perfect identity to query sequence

```

QUERY                                HQR.EIKDTKEALDKIEE

CONSENSUS_A( 10)                      ---Idv-----
A.BY.97.97BL006                        ---I-----
A1.KE.93.Q23-17                         ---IDV-----
A1.SE.94.SE7253                          ---I-----
A.SE.94.SE7535                           ---INVT-----
A.SE.95.SE8538                           ---IDV-----L--
A.SE.95.SE8891                           --KIAV-----
A.SE.95.UGSE8131                         ---IDV-----
A1.UG.92.92UG037                         ---I-V-----
A1.UG.85.U455                            ---IDV-----N---

CONSENSUS_A2( 2)                       --???V-----
A2.CD.-.97CDKTB48                       --KI-V-----
    
```

Figure 2. Part of the alignment generated by ELF.

Known Epitopes

To determine whether any of the known epitopes listed in our database occur within the bounds of your submitted peptide sequence, then in the output selection table check the option called "Known epitopes in database". This will produce, in addition to the tables already described, a link to all records in our CTL database of known epitopes whose position in the HIV protein is

contained within the region covered by a submitted peptide. The HLA of these database epitopes may differ from the submitted HLA.

To summarize this combination of input and output options:

Input	Output	Results
HLA + peptide sequence	Summary Tables + Known epitopes in database	The tables listed above + link to database records

Maps

Another output option called Maps will draw maps that show the location of all known epitopes of length < 22 AA for each HIV protein. The epitopes whose HLA matches any of the HLAs in the input list will be highlighted in red. The example shown (Figure 3) below is part of the Gag p17 protein with all A2 and associated HLAs highlighted. The maps are in pdf format which can be viewed in Adobe Acrobat and printed. The Maps option is the most time consuming computation at this site; it takes about 10–15 seconds to complete.

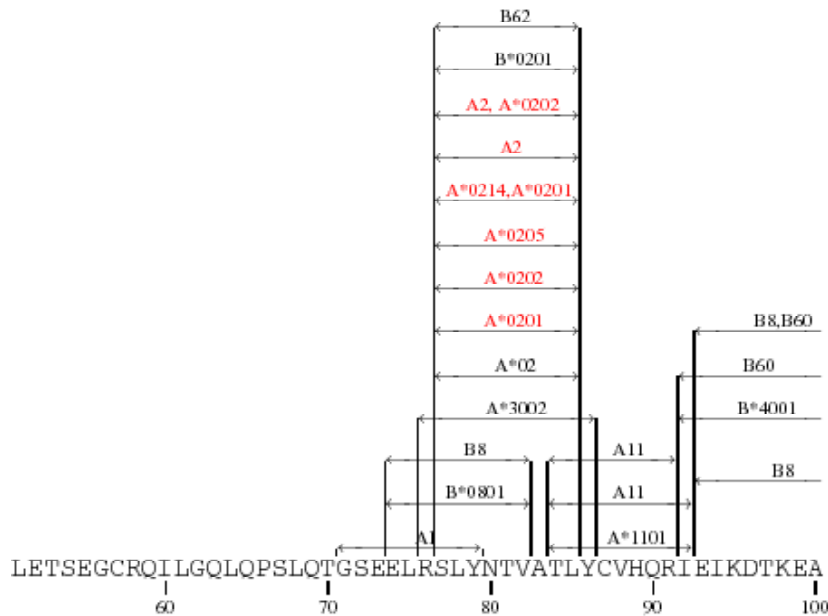


Figure 3. Part of the p17 protein with HLA A2 epitopes highlighted in red.

Finding “missed” epitopes

The last of the input options (Figure 1) called “Protein Sequence of User’s Viral Strain” is used in conjunction with the output option “Find missed epitopes.” This option allows the user to search an entire (or partial) HIV protein sequence to find sequence substitutions relative to known epitopes presented by a given HLA. Thus the HLA type of an individual is used to focus on previously defined epitopes with appropriate HLA presenting molecules; if the protein sequence used to generate peptides has substitutions relative to the reactive epitope, it is flagged as an epitope that might have been missed in the peptide screening for reactive peptides, due to natural HIV variation. This output option alerts the user to potential immunologically significant sequence differences between the user’s isolate sequence and known epitopes in the database. To illustrate, if HLA B35 was submitted along with the Nef protein from user’s isolate “x” (only part of which is shown here, with the region of interest underlined):

...TNAACAWLEAQEEEEVGFVPTPQVPLRPMTYKAAVDLSHFLKEKGGLEGL...

Table 4 would result. The second column of the table shows known epitopes from our database whose HLA agrees with the query HLA (B35), but whose amino acid sequence differs from the equivalent amino acid sequence in the user’s HIV isolate, shown in column 3. The user’s sequence has a “T” in position 4 of the epitope whereas the database epitope has an R at that position. Column 1 of the table contains a link to the database record which can be examined for further details about this epitope including the reference to the study that originally defined the epitope.

Table 4. Known Epitopes Potentially Missed in Users Isolate

Position	Epitope in database	Epitope in ref. strain	Epitope in CON-B	HLA
Nef(68–76)	FPVRPQVPL	FPVTPQVPL	fpvrpqvpl	B35
Nef(68–76)	FPVRPQVPL	FPVTPQVPL	fpvrpqvpl	B*3501
Nef(73–82)	SVPLRPMTYK	QVPLRPMTYK	qvplrPmtyk	B35 C4

To summarize this combination of input and output options:

Input	Output	Results
HLA + peptide sequence + sequence of HIV strain	Summary Tables + Known epitopes in database + Find missed epitopes	The tables listed above + table of known potentially missed epitopes and links to database

It is understandable that the reader may be confused by the many options discussed above in a static presentation on paper. A clearer understanding of the ELF site can best be had by connecting to it—http://hiv-web.lanl.gov/ALABAMA/epitope_analyzer.html—and running the “Sample Input”. Or you can run your own data with various input and output options. The Map output option is the slowest computation, so it may be preferable to leave it unchecked at least while getting a feel for how the site works. Please send suggestions for modifications or error reports to btcl@lanl.gov and cxc@lanl.gov.

ACKNOWLEDGMENTS

This suite of programs was developed at Los Alamos National Laboratory to support the experimental immunology being conducted at the University of Alabama under the NIH contract: HLA Typing and Epitope Mapping to Guide HIV Vaccine Design in the Adolescent Medicine HIV/AIDS Research Network (AMHARN).

The web interface was developed through the NIH-DOE interagency agreement that supports the Los Alamos HIV database with the hope these tools might be broadly helpful to T-cell immunologists.