PART II: HIV-1 CTL EPITOPES

SUMMARY

Part II includes tables, maps, and alignments of HIV-specific CTL epitopes arranged sequentially according to the location of the proteins in the HIV-1 genome. We attempted to make this section as comprehensive as possible, requiring that the epitope be contained within a defined region of a maximum of 30 amino acids, but not that the precise boundaries be defined. Studies that were based on the analysis of whole proteins are described at the end of each protein section. The same epitope can have multiple entries, as each entry represents a single publication in this section of the database. For more recent updates and useful searching capabilities, please see our web site: http://hiv-web.lanl.gov/immunology. For a concise listing of the best defined CTL epitopes, see the summary by Christian Brander and Philip Goulder in part I. CTL protein reactions with no well-defined epitopes are listed at the end of each protein section.

A. CTL EPITOPE TABLES

Each CTL reference has a six part basic entry:

- **HXB2 Location:** The viral strain HXB2 (GenBank Accession Number K03455) is used as a reference strain throughout this publication. The position of the defined epitope location on the sequence of the HXB2 protein is indicated. Obviously HXB2 may not be identical to a given defined reactive sequence, so we are simply indicating the location of the aligned positions. The HXB2 numbering is used in the protein maps in this database and is the reference strain in the HIV Sequence Compendium. HXB2 was chosen as the reference clone because it is the most intensively studied strain in terms of immunogenecity, structure, and function.
- **Author Location:** The amino acid positions of the epitope boundaries and the reference sequence are listed as given in the primary publication. Frequently, these positions as published are imprecise, and do not truly correspond to the numbering of the sequence, but they provide a reasonable guide to the peptide's approximate location in the protein. Also, in many cases the reference sequence identification was not provided, and in such cases it is not possible to use these numbers to specify precise locations. If you are interested in finding the precise positions of epitopes you are studying relative to the HXB2 strain, please try using the interactive position locator at our web site: http://hiv-web.lanl.gov/NUM-HXB2/HXB2.MAIN.html.
- **Epitope Sequence:** The amino acid sequence of the epitope of interest as defined in the reference, based on the reference strain used in the study defining the epitope. On rare occasions, when only the epitope location and not the actual epitope sequence was specified in the original publication, and the sequences were numbered inaccurately by the primary authors, we may have misrepresented the epitope's amino acid sequence. Therefore epitopes that were not explicitly written out in the text in the primary publication, those that we determined by looking up the reference strain and the numbered location, are followed by a question mark in the table.
- **Immunogen:** The antigenic stimulus of the CTL response.
- **Species(HLA):** The species responding and HLA of MHC specificity of the epitope.
- **Reference:** The primary reference (sometimes two or more directly related studies are included).

Following the entry for a given CTL epitope are brief comments explaining the context in which the epitope was studied. If the same epitope was studied in several labs, each study is cited in its own bulleted entry.

B. HIV PROTEIN EPITOPE MAPS

Because of the increasing number of defined epitopes, only HIV CTL epitopes defined in primates and mapped to within a region of 21 amino acids or less, with a known HLA specificity, are indicated on the HIV protein epitope maps.

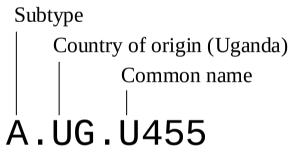
The location and HLA restriction elements of CTL epitopes are indicated on protein sequences of HXB2. These maps are meant to provide the relative location of epitopes on a given protein, but the HXB2 sequence may not actually carry the epitope of interest, as it may vary relative to the sequence for which the epitope was defined. Epitopes are numbered in bold on the maps; the map numbering corresponding to the numbering of the epitope sequence alignments.

C. REFERENCES AND NOTES

ALIGNMENTS

Space and binding limitations have forced us to print the alignments of CTL epitopes as a document separate from this compendium. You should have received this alignment Appendix as a shrink-wrapped, three-hole punched package in the same mailer as your book. As with the maps, only HIV CTL epitopes defined within a region of 21 amino acids or less, with a known HLA specificity, have corresponding alignments. For each numbered epitope in the epitope-protein maps, an alignment was generated from the protein sequence alignments in the HIV-1 genetic sequence database. All epitopes are aligned to the HXB2 sequence and the sequence used to define the epitope is indicated directly above it. In subtype consensus sequences an upper case letter indicates the amino acid was present in all sequences, a lower case letter indicates the amino acid was present in most sequences in a given position, and a question mark indicates two or more amino acids were represented with equal frequency. The master alignment files from which the epitope alignments were created are available at `http://hiv-web.lanl.gov/ALIGN_CURRENT/ALIGN-INDEX.html`.

Included in the epitope alignments are only those sequences which completely span the protein. Short fragments of sequences are excluded. The subtype designation and the country of isolation are indicated along with the common name of the sequence. A key to the two-letter country codes can be found on our website at `http://hiv-web.lanl.gov/HTML/databasecountrycode.html`, or in *Human Retroviruses and AIDS*, our annual HIV sequence compendium.



The alignments were modified in some cases to optimize the alignment relative to the defined epitope and to minimize insertions and deletions. A dash indicates identity to the consensus sequence, and a period indicates an insertion made to maintain the alignment. Stop codons are indicated with a $, and frameshifts by a #; they are inserted to maintain the alignments.

Appendix: Alignments of CTL Epitopes

Space and binding limitations have forced us to print the alignments of CTL epitopes as an appendix separate from the bound compendium. Only epitopes <22 amino acids long, with known HLA specificity are shown. These alignments correspond to the sets of protein alignments available at:

`http://hiv-web.lanl.gov/ALIGN_CURRENT/ALIGN-INDEX.html`

Each alignment on the pages that follow has a common layout that begins with a headline announcing the protein and the epitope numbers represented by that alignment. The epitope numbers correspond to the numbers in the epitope maps in section II-B. Duplicate epitopes—those with the same sequence and HLA—are shown only once. The epitopes have been grouped into sets of overlapping epitopes such as

```
KIRLRPGGKKKYKL
KIRLRPGGKKK
  RLRPGGKK
    RPGGKKYK
```

Below the aligned list of epitopes that comprise the set is the equivalent amino acid sequence from the isolate B.FR.HXB2. This B.FR.HXB2 sequence is the "model" against which all other sequences in the alignment have been compared. When amino acids and their positions in the subject sequence agree with those in the HXB2 sequence a "-" character is written; an "x" means an unknown amino acid, a "$" means a stop codon, and a # means a frame shift. For example

```
HXB2  VARELHP
seqx  VARQLHP
```

would appear as

```
HXB2  VARELHP
seqx  ---Q---
```

These epitope alignments differ from standard alignments, however, in that gaps placed to bring sequences into alignment have been squeezed out and the alignment shifted rightwards (toward the C-term end). For example, using "." as the gap character and "-" to indicate identity, the alignment

```
HXB2  VARELHP
seqx  VAR.LHP
```

would be printed as

```
HXB2  VARELHP
seqx  QVAR---
```

Q is the amino acid one position to the left of the V. As a result of squeezing gaps and shifting characters rightward, alignments in gappy regions, such as the V3 loop of gp160 will look "bad." Here is an actual example from gp160. The standard alignment with gaps, printed below, contains the region RPVVSTQLLLNGSLAEEEVV

```
HXB2  RPVVSTQLLL.NGSLAEEE.VVIRSVNFT
PVPI  RPVISTQLLL.NGSLAEEKDVQIRSENIT
```

The HXB2 sequence contains two gaps compared with only one in PVPI. When the gaps are squeezed, the HXB2 sequence moves two locations right while the PVPI sequence moves only one place right. As a result the alignment of the epitope now looks like:

```
HXB2        RPVVSTQLLLNGSLAEEEVV
A.RW.PVPI   PVISTQL--NGSLAE-KD-Q
```

which looks wrong.

The one exception to these rules involves epitope 50 in gp160 RIQRGPGRAFVTIGK, the tip of the V3 loop. HXB2 (the model sequence to which all others are compared) has an unusual QR insertion before the GPGR tip. The alignment of this epitope contains two gaps corresponding to those inserted amino acids.

Consensus sequences were generated for all HIV subtypes and CRFs that contained five or more loci representative of that subtype. In these consensus sequences a capital letter means complete conservation of that amino acid at that position, a lowercase letter means >30% conservation (not the usual 50%) and a ? character means no amino acid occurs with 30% frequency. Gaps have been squeezed from the consensuses but their original letters have been retained.