# HIV Database Workshop

## www.hiv.lanl.gov

## seq-info@lanl.gov


**Presenters**: Will Fischer, Elizabeth-Sharon Fung

**Database PIs**: Bette Korber, Thomas Leitner, Karina Yusim, Brian Foley

**Additional database staff:**
Werner Abfalterer, Kumkum Ganguly, Jennifer Macke, James Szinger, Elena Giorgi, Hyejin Yoon

**Contract Officer Representative**: Anjali Singh, NIAID, NIH

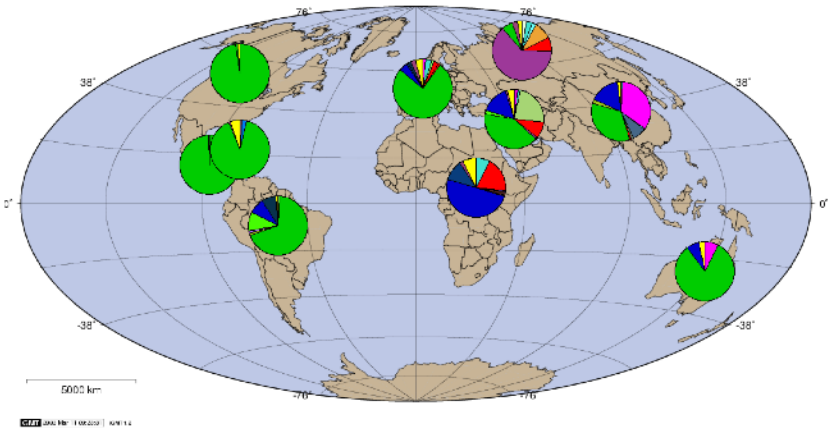
*Theoretical Biology and Biophysics, T-6*
*Los Alamos National Laboratory*

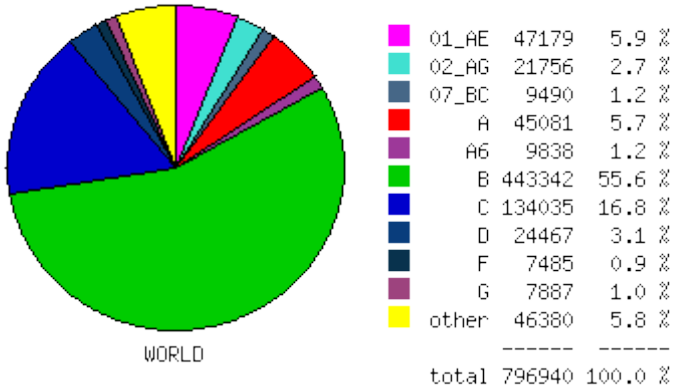

**HIV DB Workshop slides:**
https://tinyurl.com/HIVDB-2019-IEDB

# Los Alamos HIV Databases

- Database statistics, for the month of July 2019
  - (see: www.hiv.lanl.gov/tmp/awstats/hiv-stats.pdf)
  - 643,187 hits
  - 26,286 visits (i.e. uses)
  - The 2019 update includes all sequences through Dec 2018
- HIV Sequence Database: Over 922,972 searchable annotated HIV/SIV sequences total.
  - Stored metadata enables us to provide custom made alignments or pre-made 1-sequence-per-person alignments.
- HIV Immunology Database: Searchable annotated T cell epitopes and Antibody entries
  - 10,455 CD8+ epitope entries from 1,353 papers
  - 1,609 CD4+ epitope entries from 389 papers
  - 3,549 distinct monoclonal antibody entries
  - Neutralization data accessible through CATNAP
    - For 423 Abs, 40 antibody mixtures, and 20 polyclonal sera
    - 1054 pseudoviruses tested, including 819 with sequences
- >60 bioinformatics tools with simple web interfaces
- multiple search interfaces
  - Tools split ~ 50/50 between HIV-specific and general-use

Global Clade and CRF distribution






Los Alamos NATIONAL LABORATORY

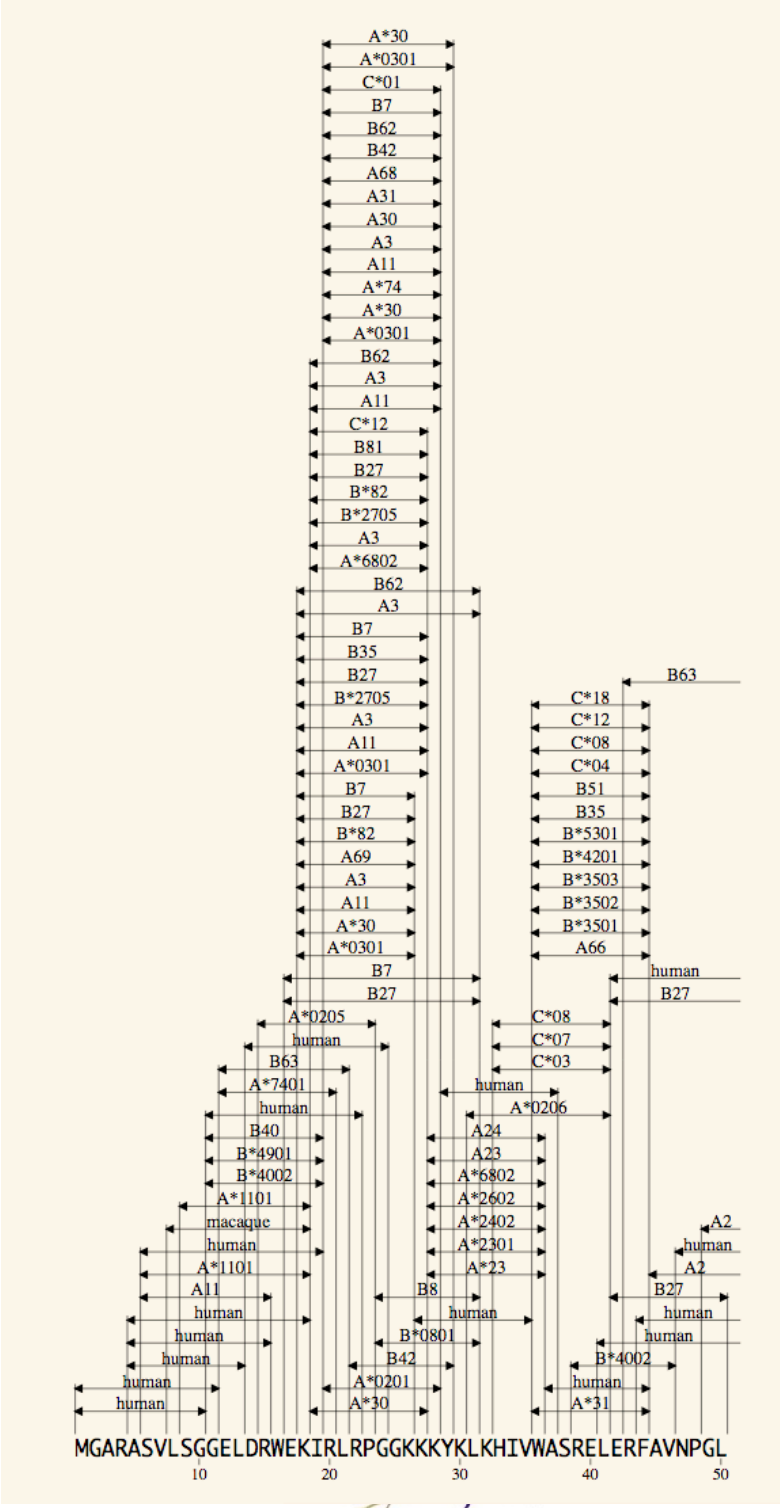
# Los Alamos HIV Databases

- Database statistics, for the month of July 2019
  - (see: www.hiv.lanl.gov/tmp/awstats/hiv-stats.pdf)
  - 643,187 hits
  - 26,286 visits (i.e. uses)
  - The 2019 update includes all sequences through Dec 2018

- HIV Sequence Database: Over 922,972 searchable annotated HIV/SIV sequences total.
  - Stored metadata enables us to provide custom made alignments or pre-made 1-sequence-per-person alignments.

- HIV Immunology Database: Searchable annotated T cell epitopes and Antibody entries
  - 10,455 CD8+ epitope entries from 1,353 papers
  - 1,609 CD4+ epitope entries from 389 papers
  - 3,549 distinct monoclonal antibody entries
  - Neutralization data accessible through CATNAP
    - For 423 Abs, 40 antibody mixtures, and 20 polyclonal sera
    - 1054 pseudoviruses tested, including 819 with sequences

- >60 bioinformatics tools with simple web interfaces
- multiple search interfaces
  - Tools split ~ 50/50 between HIV-specific and general-use




Los Alamos NATIONAL LABORATORY

# HIV Immunology Database Workshop

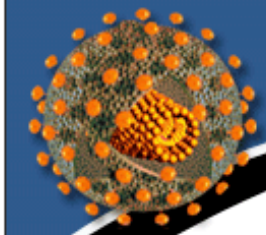
- Today
  - Overview of the HIV Immunology and HIV Sequence Databases
  - T cell epitope data and search interface
  - Peptide tools
- Tomorrow
  - Integration of Antibody and Sequence Data (a walk-through)
  - Neutralizing Antibody Resources
  - CATNAP
    - neutralization exploration
    - tailored for HIV but pathogen-agnostic
  - CombiNaber, applicable for any pathogen
  - Glycan shield
  - HIV Genome Browser
  - Vaccine design and evaluation tools

Los Alamos
NATIONAL LABORATORY

# HIV Immunology Database Workshop

- Today
  - Overview of the HIV Immunology and HIV Sequence Databases
  - T cell epitope data and search interface
  - Peptide tools
- Tomorrow
  - Integration of Antibody and Sequence Data (a walk-through)
  - Neutralizing Antibody Resources
  - CATNAP
    - neutralization exploration
    - tailored for HIV but pathogen-agnostic
  - CombiNaber, applicable for any pathogen
  - Glycan shield
  - HIV Genome Browser
  - Vaccine design and evaluation tools

Los Alamos
NATIONAL LABORATORY



# HIV DATABASES

## http://hiv.lanl.gov

The **HIV databases** contain comprehensive data on HIV genetic sequences and immunological epitopes. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000. Our content is reviewed by an Editorial Board.

SEQUENCE DATABASE ▸

IMMUNOLOGY DATABASE ▸

OTHER VIRUSES ▸

News:

Archived News ▸

CATNAP: two new features

CATNAP now provides an option to calculate geometric mean estimates including tests that were above threshold (setting a score of 100 ($IC_{50/80}$) or 20 ($ID_{50/80}$) for the purpose of the estimation). Also, we have introduced a "Trim-and-Re-calculate" feature in the analysis which enables users to select data from specified papers instead of using the full set in CATNAP collection. This could be useful to reduce data redundancy or to address inconsistencies between studies (for instance, changes in pipette tips used for serial dilution). *20 February 2019*

GenSig

*GenSig*, a signature analysis web interface, is now available online. It can identify genetic signatures in a DNA alignment with associated phenotypic data. Also we have integrated our signature code into CATNAP allowing signature analysis to be conducted on-the-fly as new bNAb data is entered into the database. *10 January 2019*

# Integration of HIV Sequence and Immunology databases

- Los Alamos HIV Database: the first pathogen-specific database
  - HIV Sequence Database – founded in 1986 by G. Myers
  - HIV Immunology Database - founded in 1994 by B. Korber

- Integration of HIV sequence and immunological data via multiple tools, for example:
  - CATNAP superimposes Ab neutralization data with the virus data, and links to structures, germline V/D/J genes, Ab sequences, Ab contact residues, Env alignments, positions associated with neutralization sensitivity …
  - AnalyzeAlign shows the diversity and HIV variability of epitopes
  - HIV Genome Browser provides an interactive detailed view of the HIV genome or proteome with HIV sequence variability, functional domains and antibody and T cell epitopes marked by genome position
  - Multiple tools tap into the Donors database, containing available donor HIV sequences, Ab sequences, monoclonal and polyclonal Ab data, HLAs, and T-cell epitopes

Los Alamos
NATIONAL LABORATORY

# Beyond HIV

- **About one dozen of our computational tools (20%) are strictly HIV-specific. The remaining 80% are partially or fully applicable to other organisms**

- A striking example of successful extension beyond HIV is Mosaic/Epigraph vaccine design:
  - Rabies in bats (Stading *et al*, Plos Negl Trop Dis, 2017)
  - Filoviruses (Theiler *et al*, Sci Rep. 2016, Fenimore, PLoS One, 2012)
  - Chlamydia trachomatis (Badamchi-Zadeh *et al*, Front Immunol, 2016)
  - Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in pigs (Cui *et al*, Vaccine reports, 2016)
  - Hepatitis C (Yusim *et al*, Clin Vaccine Immunol, 2013)
  - Foot-and-Mouth Disease in livestock (Devendra *et al*, in preparation)
  - Hepatitis B (Yusim *et al*, in preparation)
  - **... and the HIV-1 mosaic designs are moving into Phase III human trials**

- The database structure and tools are transferrable to other pathogens. We have created several pathogen databases prototyped on the HIV database, and translating multiple tools: (https://www.hiv.lanl.gov/content/otherviruses.html):
  - **HCV Sequence** (Kuiken et al, Nucleic Acid Res, 2008) and Immunology (Yusim et al, Appl Bioinformatics, 2005) Databases
  - **Hemorrhagic Fever Viruses (HFV) Sequence Database (**80 viral species, found in 10 different genera comprising five different families: arena-, bunya-, flavi-, filo- and togaviridae) (Kuiken et al, Nucleic Acid Res, 2012)
  - **Filovirus Sequence and Immunology Database** (Yusim et al, Database, 2016) (hfv.lanl.gov)
  - Because of lack of individual funding, only the sequence portions of these databases are automatically updated

Los Alamos NATIONAL LABORATORY

# Many HIV Database tools are broadly applicable

- Tools list is color-coded by range of use

# HIV Database Tools

- Tools specific for HIV/SIV
- General use tools with some HIV/SIV-specific features
- General use tools

## Analysis and Quality Control

- Entropy quantifies positional variation in an alignment using Shannon Entropy
- GenSig identifies genetic signatures. It can work on any phenotype file given in conjunction with a codon-aligned nucleotidealignment of a protein coding region
- Glycan Shield Mapping shows mapping absent hole-causing potential N-linked glycosylation sites (PNGS) on predicted glycan shields for an ENV sequence
- HIV BLAST finds sequences similar to yours in the HIV database
- Hypermut detects hypermutation
- jpHMM at GOBICS detects subtype recombination in HIV-1; hosted at GOBICS as a collaboration between the Department of Bioinformatics, University of Göttingen and the Los Alamos HIV Sequence Database
- N-Glycosite finds potential N-linked glycosylation sites
- PCOORD multidimensional analysis of sequence variation
- Quality Control runs several tools for quick troubleshooting of HIV-1 sequences; optional step prepares sequence submission for GenBank
- RAPR (Recombinant Analysis PRogram) uses the Wald-Wolfowitz Runs Test to check for recombination in every triplet in the alignment.
- RIP (Recombinant Identification Program) detects HIV-1 subtypes and recombination
- SNAP calculates synonymous/non-synonymous substitution rates
- SUDI Subtyping plots the distance of your sequence to established subtypes
- Variable Region Characteristics summarizes features of peptides from a protein alignment, reporting peptide characteristics such as length and N-linked glycosylation sites; particularly useful for characterizing variable regions of Env
- VESPA (Viral Epidemiology Signature Pattern Analysis) detects residues with different frequencies in two sequence sets

Los Alamos NATIONAL LABORATORY

# HIV Database Tools

- Tools specific for HIV/SIV
- General use tools with some HIV/SIV-specific features
- General use tools

## Phylogenetics

- AnnotateTree creates a colored and weighted phylogenetic tree
- Branchlength calculates branch lengths between internal and end nodes; now included in the TreeRate tool
- FindModel finds which evolutionary model best fits your sequences
- IQ-TREE is a fast and effective stochastic algorithm for finding Maximum Likelihood trees, including site-specific rates of evolution at each alignment position
- PhyloPlace reports phylogenetic relatedness of an HIV-1 sequence with reference sequences
- PhyML generates much better trees than our simple TreeMaker tool
- Poisson-Fitter estimates time since MRCA and star-phylogeny. For use with acute (low diversity) samples
- Rainbow Tree Color code phylogenetic tree branches according to labels in the sequence names
- TreeMaker generates a Neighbor Joining phylogenetic tree
- TreeRate finds the phylogenetic root of a tree and calculates branch lengths and evolutionary rate

# HIV Database Tools

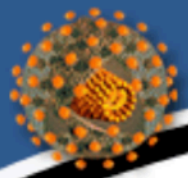
- Tools specific for HIV/SIV
- General use tools with some HIV/SIV-spe
- General use tools

## Format and display

- Format Converter converts between alignment formats
- GenBank Entry Generation produces GenBank Sequin files for HIV-1, HIV-2, and SIV sequences, plus associated metadata
- Genome Browser uses jBrowse to display diverse data about the HIV-1 genome and proteome
- Highlighter highlights mismatches, matches, transitions and transversion mutations and silent and non-silent mutations in an alignment of nucleotide sequences
- Protein Feature Accent provides an interactive 3-D graphic of HIV proteins; can map a sequence feature (a short functional domain, epitope, or amino acid) and see it spatially
- Recombinant HIV-1 Drawing Tool creates a graphical representation of your HIV-1 intersubtype recombinant
- SeqPublish makes publication-ready alignments

## Alignment and sequence manipulation

- Align Multi-tool manipulates sequence alignments, including sorting, pruning, and renaming
- Alignment Slicer cuts vertical slices from sequence alignments
- Analyze Align shows weblogos, calculates frequency by position, and finds variants in an alignment
- Codon Alignment takes a nucleotide alignment and returns a codon alignment and translation
- Consensus Maker computes a customizable consensus
- ElimDupes compares the sequences within an alignment and eliminates any duplicates
- Gap Strip/Squeeze removes columns with more than a given % of gaps
- Gene Cutter clips genes from a nucleotide alignment, codon-aligns, and translates
- HIValign uses our HMM alignment models to align your sequences
- PepMap can be used to map epitopes, functional domains, or any protein region of interest
- Pixel generates a PNG image of an alignment using 1 or more colored pixel(s) for each residue
- QuickAlign (formerly Epilign and Primalign) aligns short nucleotide or protein sequences (e.g., primers, epitopes) to our prebuilt genome or protein alignments, or to a user alignment
- Sequence Locator finds the standard numbering of your HIV or SIV nucleotide or protein sequence
- SynchAlign aligns overlapping alignments to one another
- Translate nucleotide sequences to 1-letter amino acids





# HIV Database Tools

- Tools specific for HIV/SIV
- General use tools with some HIV/SIV-specific features
- General use tools

## Database search interfaces

- Advanced Search creates a custom search interface
- Antibodies search for HIV antibodies by protein, immunogen, AB type, isotype, author, keywords
- CTL/CD8+ Search searches for CD8+ epitopes by protein, immunogen, HLA, author, keywords
- Geography shows the geographic distribution of sequences in the database
- Reference Sequence Coordinate Search retrieves HXB2 and Mac239 features and coordinates
- Intra-patient Search retrieves intra-patient sequence sets
- Neutralizing Antibody Contacts and Features retrieves neutralizing antibody contact sites and other HIV-1 Env features
- Sequence Search searches for sequences based on numerous criteria
- T-Helper/CD4+ Search search for CD4+ epitopes by protein, immunogen, HLA, author, keywords
- Vaccine Trials Database finds past vaccine trials and their results

## Immunology

- CATNAP (Compile, Analyze, Tally NAb Panels) provides meta-analysis of published neutralization panel data
- CombiNAber predicts and analyzes combination antibody neutralization scores using $IC_{50}$ and/or $IC_{80}$ for individual antibodies
- ELF (Epitope Location Finder) identifies known and potential epitopes within peptides
- Epigraph Tool Suite uses input of diverse sequences to generate Epigraph sequences for use in vaccine or reagent design
- Epilign (QuickAlign) aligns a protein sequence (e.g., epitope) to the appropriate protein alignment
- Heatmap displays a table of numbers by using colors to represent the numerical values
- Hepitope identifies potential epitopes based on HLA frequencies
- Neutralization Index computes a tier-like score for sera (using ID50 titers) or antibodies (using IC50 titers)
- Mosaic Vaccine Tool Suite designs and assesses polyvalent protein sequences for T-cell vaccines
- Motif Scan finds HLA anchor motifs in protein sequences for specified HLA serotypes, genotypes or supertypes
- PeptGen generates overlapping peptides from a protein sequence

# HIV Immunology Database Entries and Annotation

- HIV T cell epitopes and Antibody data organization
  - T Cells (CTL and Helper epitopes)
    - One reference per entry, epitope/HLA combinations are often repeated
    - CTL and T-helper database organization is identical
  - B Cells (Antibodies)
    - One entry for each monoclonal antibody
    - Many references per entry (> 800 for some well studied mAbs)

- Descriptions of HIV T cell epitopes and Antibodies with associated data are harvested from regular periodic literature searches:
  - Epitope sequence, location, immunogen, vaccine details, subject details…
  - Epitope Variants (escape, reduced binding, etc.)
  - Host HLA or MHC, binding region, germline genes, etc
  - Neutralizing Antibody Resources, contact residues, positions related to neutralization sensitivity or resistance, etc.
  - Notes summarizing main findings

- Multiple search interfaces and database products:
  - 5 search interfaces for T cell epitopes, epitope variants and antibodies
  - Computational tools for immunologists
  - Epitope maps and summary tables that can also serve as search interfaces
  - HLA typing and epitope mapping data sets
  - Neutralizing antibody resources:
    - Neutralization, germline and antibody sequence data through CATNAP
    - Links to Germline Antibody Reconstruction tools
    - Search interface and a table for Ab contact residues, positions related to neutralization sensitivity or resistance, etc.
    - Assay protocols and neutralization serotype discovery data

Los Alamos
NATIONAL LABORATORY

# HIV DATABASES

The **HIV databases** contain comprehensive data on HIV genetic sequences and immunological epitopes. The website also gives access to a large number of tools that can be used to analyze and visualize these data. This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000. Our content is reviewed by an Editorial Board.

SEQUENCE DATABASE ▸ IMMUNOLOGY DATABASE ▸

OTHER VIRUSES ▸

News: Archived News ▸

2018 Alignments

The 2018 *Web*, *Filtered Web*, and *Super Filtered Web Alignments* are now available online. These curated alignments contain a full range of sequences available through the end of 2018. *08 October 2019*

**https://hiv.lanl.gov**

Los Alamos NATIONAL LABORATORY

# HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

## Search Interfaces

- CTL/CD8+ search
- T Helper/CD4+ search
- Antibody search
- CTL variant search
- T Helper variant search

- Search help
- Variant search help

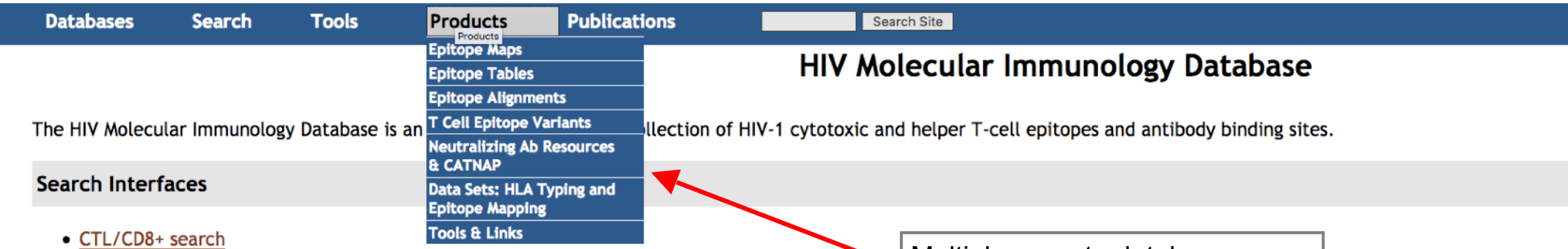
## Database Products

- All Database products and publications
- Epitope maps
- Epitope tables
- Epitope alignments
- Epitope density plots
- T cell epitope variants and escape mutations
- Neutralizing antibody resources & CATNAP
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology Database
- How to cite this database
- Frequently-asked Questions (FAQ)

## Tools and Data Sets

- Tools & Links for immunologists
- SIV Epitopes (PDF) review article summarizing known SIV epitopes
- Identifying HLA-Associated Polymorphisms in HIV-1 (PDF) review article summarizing HIV polymorphism associated with escape mutations. Also a table of polymorphisms.
- HLATEM HLA Typing and Epitope Mapping Data Sets
- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

lamos ABORATORY




# HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an ...llection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

## Search Interfaces

- CTL/CD8+ search
- T Helper/CD4+ search
- Antibody search
- CTL variant search
- T Helper variant search

- Search help
- Variant search help

Antibody Search

Multiple ways to database products and tools

## Database Products

- All Database products and publications
- Epitope maps
- Epitope tables
- Epitope alignments
- T cell epitope variants and escape mutations
- Neutralizing antibody resources & CATNAP
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology Database
- How to cite this database
- Frequently-asked Questions (FAQ)

T cell epitope variants and escape mutations

Neutralizing Antibody Resources

## Tools and Data Sets

- Tools & Links for immunologists
- SIV Epitopes (PDF) review article summarizing known SIV epitopes
- Identifying HLA-Associated Polymorphisms in HIV-1 (PDF) review article summarizing HIV polymorphism associated with escape mutations. Also a table of polymorphisms.
- HLATEM HLA Typing and Epitope Mapping Data Sets
- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

*https://www.hiv.lanl.gov/content/immunology/index.html*

Los Alamos NATIONAL LABORATORY

Databases Search Tools Products Publications Search Site

# HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

## Search Interfaces

- CTL/CD8+ search
- T Helper/CD4+ search
- Antibody search
- CTL variant search
- T Helper variant search

- Search help
- Variant search help

## Database Products

- All Database products and publications
- Epitope maps
- Epitope tables
- Epitope alignments
- T cell epitope variants and escape mutations
- Neutralizing antibody resources & CATNAP
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology Database
- How to cite this database
- Frequently-asked Questions (FAQ)

## Epitope Tables

These tables summarize the epitopes from our database. HIV-1 epitope data may also be obtained in the form of downloadable maps or alignments.

- CTL epitopes
- Best-defined ("A-list") CTL epitopes
- CTL epitope variants and escape mutations
- T-helper epitopes
- T Helper epitope variants and escape mutations

- Antibody epitopes
- Best Neutralizing Antibodies
- Antibody-Dependent Cell-Mediated Cytotoxicity (ADCC)
- Antibody index by name
- Antibody index by binding type

- SIV epitopes

- Neutralizing antibody resources

Epitope alignments: epitopes aligned to HIV subtype Reference sequences in Fasta format

## Tools and Data Sets

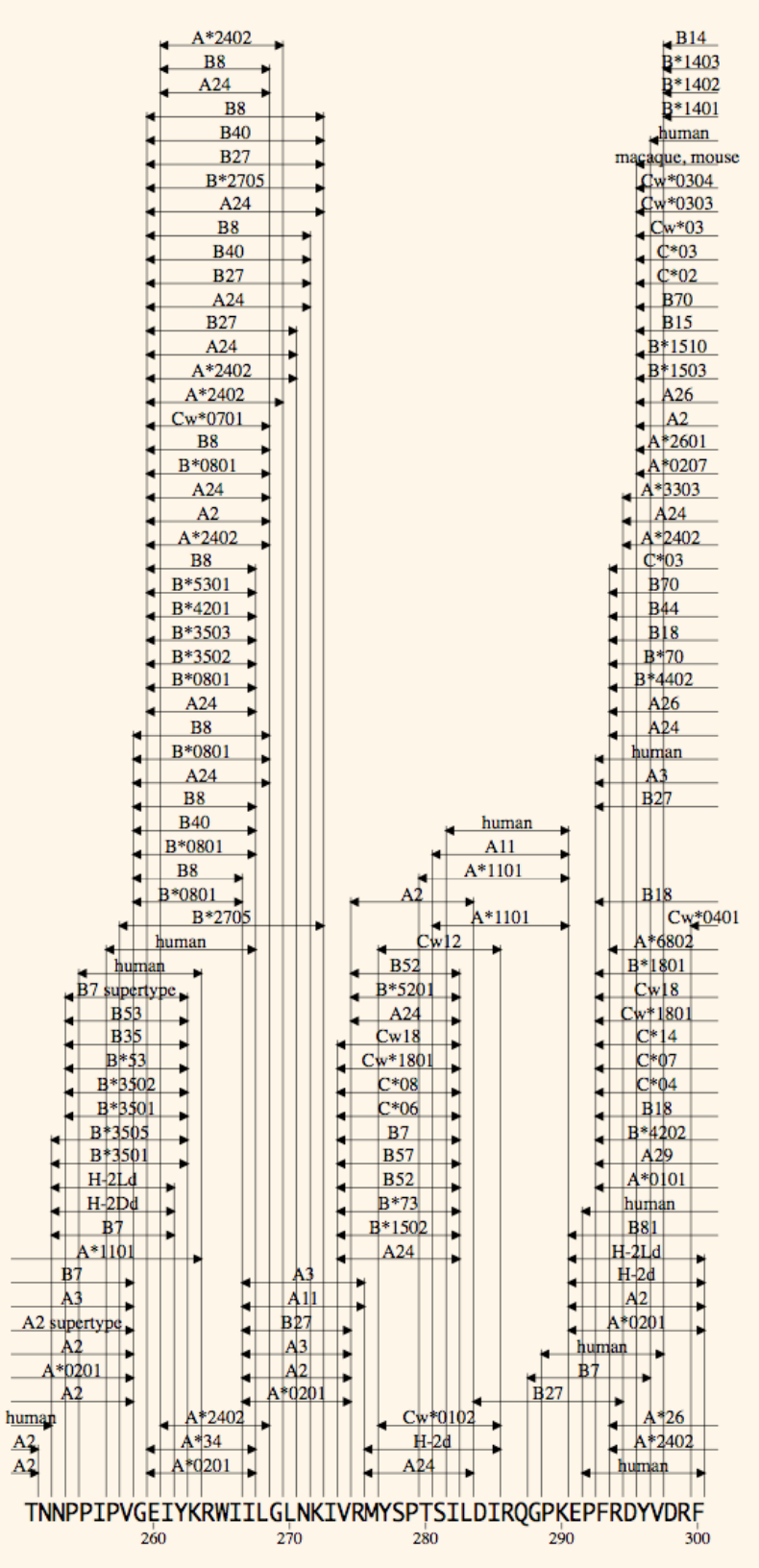
- Tools & Links for immunologists
- SIV Epitopes (PDF) review article summarizing known SIV epitopes
- Identifying HLA-Associated Polymorphisms in HIV-1 (PDF) review article summarizing HIV polymorphism associated with escape mutations. Also a table of polymorphisms.
- HLATEM HLA Typing and Epitope Mapping Data Sets
- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

Reactive peptide maps and tables (with HLA and other subject data) from several large-scale studies scanning HIV proteins.

*https://www.hiv.lanl.gov/content/immunology/index.html*

# p17 CTL/CD8+ Epitope Map

- Epitopes up to 14 aa long are mapped on HXB2
- HXB2 sequence may differ
- Epitopes with identical boundaries and HLA fields are included in the maps only once
- The epitope maps are interactive!
  - Clicking on an epitope leads to the epitope entry




*www.hiv.lanl.gov/content/immunology/maps/maps.html*

# CTL/CD8+ Epitope Summary (B-list)

- A comprehensive list of all unique epitopes in the database (including with unknown HLA, boundaries not fully defined…)
- Similar lists for Helper epitopes and linear Ab binding sites
- Unlike epitope maps that show epitope locations, each epitope sequence is shown

| Epitope | Protein | HXB2 Location | Subtype | Species | HLA |
|---|---|---|---|---|---|
| MGARASVLSG | p17 | 1-10 | CRF01_AE | human | |
| ASVLSGGEL | p17 | 5-13 | B | human | |
| ASILRGGKLDK | p17 | 5-15 | C | human | |
| SVLSGGQLDR | p17 | 6-15 | B | human | A11 |
| LSGGELDRWEK | p17 | 8-18 | | macaque | |
| GELDRWEKI | p17 | 11-19 | B | human | B*4002, B40 |
| GQLDRWEKI | p17 | 11-19 | B | human | |
| GKLDSWEKIRLR | p17 | 11-22 | A, CRF01_AE, CRF02_AG | human | |

*www.hiv.lanl.gov/content/immunology/tables/ctl_summary.html*

# Best-defined CTL/CD8+ Epitope Summary (A-list)

- Experimentally validated optimal epitopes with known HLA presenting molecules
- Defined/curated by Christian Brander and colleagues

| Epitope | Protein | HXB2 Location | Subtype | Species | HLA |
|---|---|---|---|---|---|
| GELDRWEKI | p17 | 11-19 | | human | B*4002 |
| KIRLRPGGK | p17 | 18-26 | | human | A*0301 |
| IRLRPGGKK | p17 | 19-27 | B | human | B*2705 |
| RLRPGGKKK | p17 | 20-28 | | human | A*0301 |
| RLRPGGKKKY | p17 | 20-29 | B | human | A*0301 |
| GGKKKYKLK | p17 | 24-32 | B | human | B*0801 |
| KYKLKHIVW | p17 | 28-36 | B | human | A*2402 |
| HLVWASREL | p17 | 33-41 | | human | Cw*0804 |

*www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html*

## Epitope variants and escape mutations

- Experimental epitope variants from the literature
  - Search interfaces
  - Summary tables (~3500 CTL epitope variants)
- HLA associated HIV polymorphisms (Zabrina Brumme, Bruce Walker)
  - Database review and a table

*www.hiv.lanl.gov/content/immunology/pdf/2010/escape_article_supplement.html*

Los Alamos NATIONAL LABORATORY

# CTL/CD8+ Search (www.hiv.lanl.gov/content/immunology/ctl_search)

| | | | |
|---|---|---|---|
| | Proteins with defined epitopes | Proteins with undefined epitopes | |
| HIV protein | – ALL –<br>p17<br>p17-p24<br>p24<br>p24-p2p7p1p6 | – ALL –<br>Gag<br>Gag/Pol<br>Pol<br>Vif | |
| HXB2 location | — | | Results overlap with query location |
| Epitope | ISPRTLNAW | | Results contain query sequence |
| Epitope name | | | |
| Record number | | | |
| Subtype | – ALL – | | |
| Immunogen | – ALL –<br>computer prediction<br>HIV-1 and GBV-C co-infection<br>HIV-1 and HCV co-infection<br>HIV-1 exposed seronegative<br>HIV-1 infected monocyte-derived<br>HIV-1 infection | | |
| Vaccine details if Immunogen is Vaccine | Vaccine type | – ALL – | |
| | Vaccine strain | – ALL – | |
| | Vaccine component | – ALL – | |
| | Adjuvant | – ALL – | |
| Species | – ALL – | | |
| MHC/HLA | – ALL –<br>A*01<br>A*0101<br>A*02<br>A*0201<br>A*02.01<br>A*020101 | | |
| Author | Pillay | | ☑ First ☐ Last |
| Country | – ALL – | | |
| Keywords | – ALL –<br>acute/early infection<br>adjuvant comparison<br>antagonism<br>antibody binding site definition and exposure<br>assay development, comparison, standardization, improvement<br>autologous responses | | |
| Note | | | |

Search Reset Click for Search Help

- Search by HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords
- Search on epitope location and find fuzzy matches, overlaps and embedded epitopes
- Search examples:
  - *Example:*
    - SLYNTVATL – 285 entries
    - Narrow the search with keyword “escape” – 35 entries

**Search for ISPRTLNAW**
**With the first author Pillay**

Los Alamos NATIONAL LABORATORY

## Search CTL/CD8+ T-Cell Epitope Database

Found 1 matching record:

### Displaying record number 53832

Immunological, virological, and epidemiological contexts:

Link to Epitope Maps → p24 Epitope Map

Link to Epitope Alignment → Epitope Alignment

Variant details with annotator's notes → Show epitope variants

| | |
|---|---|
| HXB2 Location | p24(15-23) |
| Author Location | Gag(147-155) |
| Epitope | ISPRTLNAW |
| Subtype | C |
| Species (MHC/HLA) | human(B57) |
| Immunogen | HIV-1 infection |
| Donor MHC/HLA | A*3001, A*66, B*4201, B*5802, Cw*0602, Cw*1701; A*66, A*68, B*57, B*5802, Cw*0602, Cw*0701 |
| Country | South Africa |
| Experimental methods | CD8 T-cell Elispot - IFNy |
| Keywords | epitope processing, responses in children, mother-to-infant transmission, escape, acute/early infection |

### Notes

- HIV-specific CTLs in infants were shown to be able to select for viral escape variants early in life, despite a lack of escape with the same CTL specificity in the mother. Infant CTL responses may be compromised by transmission of escape variants that arose in the mother and also those that arose in the father, if the father was the source of the mother's infection.
- ISPRTLNAW is the C consensus form of the epitope and was the autologous form in the mother, and was transmitted to her infant. By 33 weeks a new dominant form of the epitope had emerged in the infant, mSPRTLNAW, and two additional variants had arisen, one with a substitution proximal to the epitope, plSPRTLNAW, and lSPRTLNAW.

### References

**Pillay2005** Thillagavathie Pillay, Hua-Tang Zhang, Jan W. Drijfhout, Nicola Robinson, Helen Brown, Munira Khan, Jagadesa Moodley, Miriam Adhikari, Katja Pfafferott, Margaret E. Feeney, Anne St. John, Edward C. Holmes, Hoosen M. Coovadia, Paul Klenerman, Philip J. R. Goulder, and Rodney E. Phillips. Unique Acquisition of Cytotoxic T-Lymphocyte Escape Mutants in Infant Human Immunodeficiency Virus Type 1 Infection. *J. Virol.*, 79(18):12100-12105, Sep 2005. PubMed ID: 16140787. Show all entries for this paper.

Additional information provided in the entry:

- Location, Donor MHC/HLA, experimental methods, Notes
- Link to all entries for a reference
- PubMed links to papers
- Link to Epitope Maps
- Link to Epitope Alignment (aligned to large set of seq.)
- Epitope variants if studied in the paper

Los Alamos NATIONAL LABORATORY

# Variant details

**Displaying record number 53832**

| | | |
|---|---|---|
| HXB2 Location | p24(15-23) | p24 Epitope Map |
| Epitope | ISPRTLNAW | Epitope Alignment |
| | mSPRTLNAW | escape documented in this paper |
| Variants | lSPRTLNAW | diminished response |
| | p\|lSPRTLNAW | not determined |
| Species (MHC/HLA) | human(B57) | |

Link back to epitope entry

**Variant Details**

Showing all 3 variants.

| | |
|---|---|
| **Variant ID.** | 1413 |
| **Epitope Seq.** | ISPRTLNAW |
| **Variant Seq.** | mSPRTLNAW |
| **Mutations** | I/M |
| **Epitope Location** | I1M |
| **HXB2 Location** | I15M |
| **Mutation Type** | E: escape documented in this paper |
| **Method** | CD8 T-cell Elispot - IFNy, Sequence |
| **Note** | This is de novo variant seen in infant by week 33 of age. The index peptide was recognized, but not the variant. |

| | |
|---|---|
| **Variant ID.** | 1414 |
| **Epitope Seq.** | ISPRTLNAW |
| **Variant Seq.** | lSPRTLNAW |
| **Mutations** | I/L |
| **Epitope Location** | I1L |
| **HXB2 Location** | I15L |
| **Mutation Type** | DR: diminished response |
| **Method** | CD8 T-cell Elispot - IFNy, Sequence |

Mutation type

Note describing why the variant was designated as a particular mutation type

Mutation type examples:

- E escape
- IE inferred escape
- DR diminished response
- SF susceptible form
- etc…

Los Alamos NATIONAL LABORATORY

# QuickAlign

https://www.hiv.lanl.gov/content/sequence/QUICK_ALIGNv2/QuickAlign.html

- Aligns query sequence to an alignment, creates WebLogos, calculates frequency by position, tallies variants in an alignment
- Can be used to align epitopes, functional domains, or any protein or any region of interest
- Shows results by groupings (subtypes for example) and all groups together

| | |
|---|---|
| Query: | SLYNTVATL |
| Query Length: | 9 |
| HXB2 Location: | Gag 77-85 = p17 77-85 |
| Alignment: | GAG, 458 sequences |

Summarize

```
Query                    SLYNTVATL
A1.KE.86.ML170           --F------
A1.KE.94.Q23             --F------
A1.SE.94.SE7253          --F----V-
A1.SE.94.SE7535          ---------
A1.SE.95.SE8538          ---------
A1.SE.95.SE8891          ---------
A1.SE.95.UGSE8131        ---------
A1.TZ.97.97TZ03          --F----V-
```

Summary for subtype A

| Variant | Count | Percent |
|---|---|---|
| SLYNTVATL | | |
| --F------ | 11 | 47.83 |
| --------- | 7 | 30.43 |
| --F--I-V- | 1 | 4.35 |
| --F----V- | 1 | 4.35 |
| -------V- | 1 | 4.35 |
| ----L---- | 1 | 4.35 |
| --F-A--V- | 1 | 4.35 |

Total sequences = 23

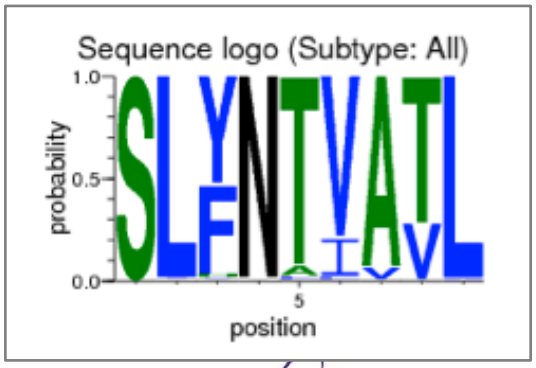
Number of variants = 7

Variant frequency summary

Frequency by position — Go to top

See full raw counts — cutoff: 95%

| Position | Percentage and raw count of non-gap | Non-gap/total (percentage) |
|---|---|---|
| 1 | S: 99.90% (3113) other: 0.10% (3) | 3116/3119 (100.00%) |
| 2 | L: 98.90% (3068) other: 1.10% (34) | 3102/3119 (99.55%) |
| 3 | Y: 52.71% (1633) F: 43.77% (1356) other: 3.52% (109) | 3098/3119 (99.42%) |
| 4 | N: 99.68% (3104) other: 0.32% (10) | 3114/3119 (99.94%) |
| 5 | T: 92.86% (2887) A: 5.05% (157) other: 2.09% (65) | 3109/3119 (99.78%) |
| 6 | V: 79.35% (2448) I: 18.15% (560) other: 2.50% (77) | 3085/3119 (99.01%) |
| 7 | A: 92.95% (2889) V: 6.53% (203) other: 0.51% (16) | 3108/3119 (99.74%) |
| 8 | T: 72.52% (2254) V: 27.06% (841) other: 0.42% (13) | 3108/3119 (99.74%) |
| 9 | L: 99.00% (3078) other: 1.00% (31) | 3109/3119 (99.78%) |




Los Alamos NATIONAL LABORATORY

# AnalyzeAlign Output

## (Discontinuous sites are permitted)

Groups
[Download combined logs PDF EPS]
B A1 A2 C D F1 F2 G 01_AE 02_AG

Group B

Go to top

Logo


Download: PNG PDF EPS

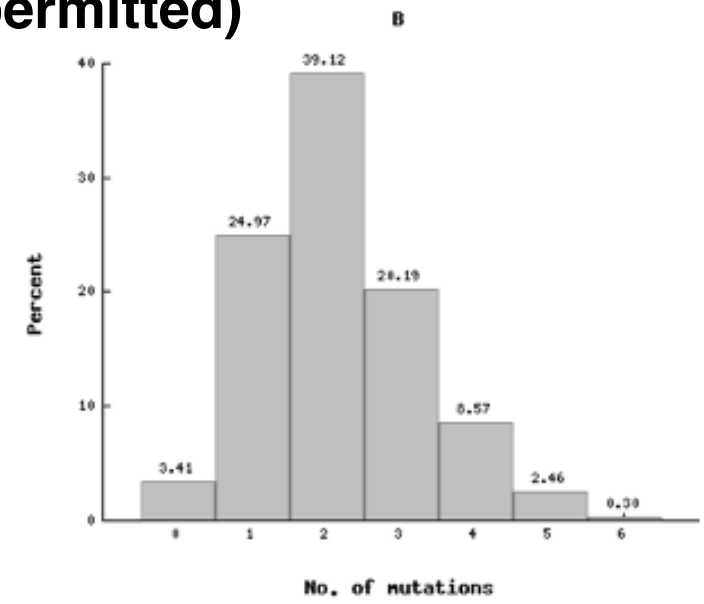
Frequency by position

See full raw counts

cutoff: 95%

| | Percentage and raw count of non-gap | Non-gap/total (percentage) | Gap/total (percentage) |
|---|---|---|---|
| 324 | G: 99.43% (1924) other: 0.57% (11) | 1935/1937 (99.90%) | 2/1937 (0.10%) |
| 325 | D: 80.84% (1565) N: 16.63% (322) other: 2.53% (49) | 1936/1937 (99.95%) | 1/1937 (0.05%) |
| 326 | I: 98.24% (1900) other: 1.76% (34) | 1934/1937 (99.85%) | 3/1937 (0.15%) |
| 327 | R: 98.76% (1913) other: 1.24% (24) | 1937/1937 (100.00%) | 0/1937 (0.00%) |
| 331 | C: 99.90% (1935) other: 0.10% (2) | 1937/1937 (100.00%) | 0/1937 (0.00%) |
| 332 | O: 85.23% (1651) T: 7.80% (151) N: 2.89% (56) other: 4.08% (79) | 1937/1937 (100.00%) | 0/1937 (0.00%) |
| 333 | I: 62.26% (1206) L: 31.80% (616) V: 5.73% (111) other: 0.21% (4) | 1937/1937 (100.00%) | 0/1937 (0.00%) |
| 334 | S: 85.54% (1657) O: 10.22% (198) other: 4.23% (82) | 1937/1937 (100.00%) | 0/1937 (0.00%) |
| 335 | R: 42.72% (827) K: 16.48% (319) G: 15.34% (297) E: 8.68% (168) S: 6.30% (122) A: 3.05% (59) I: 1.91% (37) T: 1.03% (20) other: 4.49% (87) | 1936/1937 (99.95%) | 1/1937 (0.05%) |
| 336 | T: 32.27% (625) A: 31.23% (605) E: 11.82% (229) K: 8.21% (159) G: 3.67% (71) V: 3.30% (64) S: 2.74% (53) Q: 1.91% (37) other: 4.85% (94) | 1937/1937 (100.00%) | 0/1937 (0.00%) |

Sequence variants



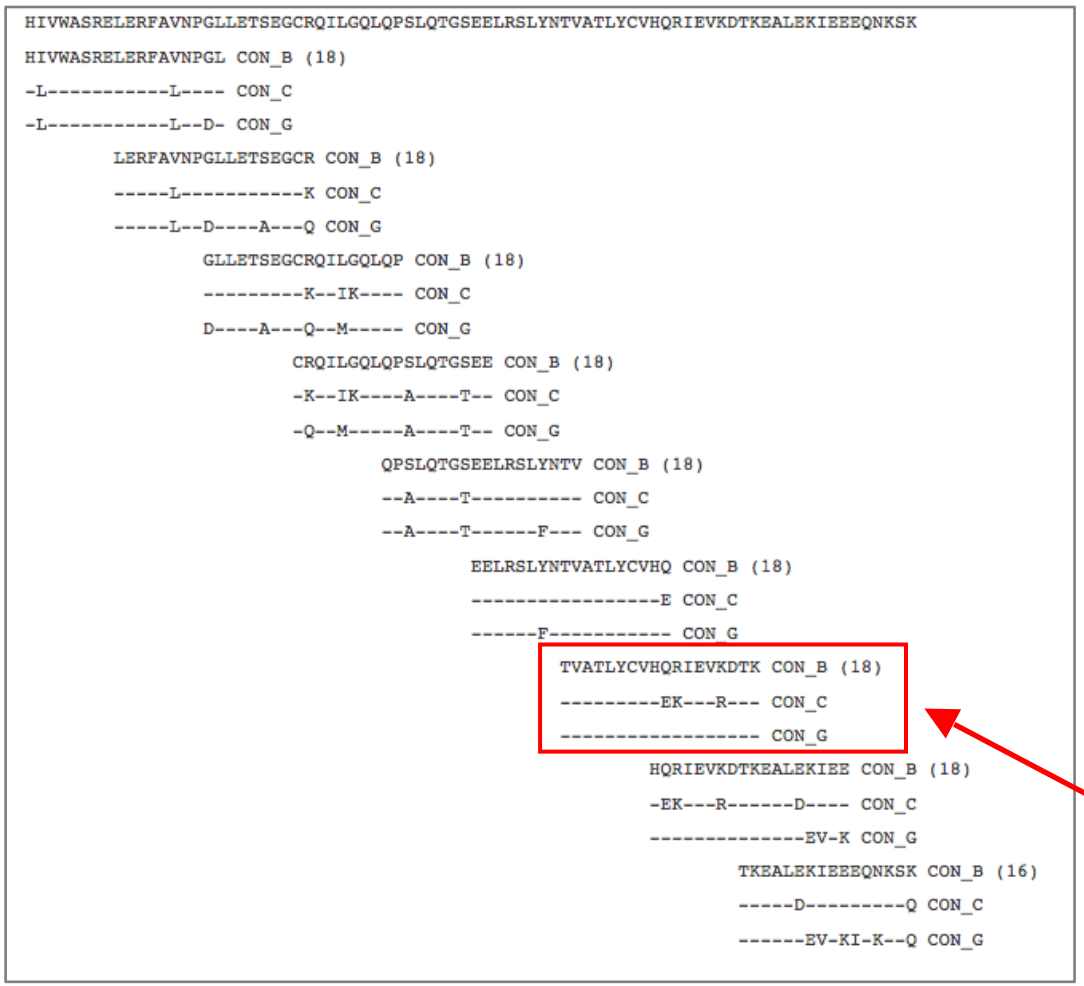

| Variant | Count | Pct. | No. of mutations |
|---|---|---|---|
| GDIRCOISRT | | | |
| ---------A | 134 | 6.92 | 1 |
| ------L--A | 89 | 4.59 | 2 |
| ---------- | 66 | 3.41 | 0 |
| ---------E | 65 | 3.36 | 1 |
| --------KA | 57 | 2.94 | 2 |
| ------L--- | 51 | 2.63 | 1 |
| --------G- | 47 | 2.43 | 1 |
| ------L-S- | 33 | 1.7 | 2 |
| ---------K | 31 | 1.6 | 1 |
| --------EA | 31 | 1.6 | 2 |
| -N--------A | 28 | 1.45 | 2 |
| --------GA | 27 | 1.39 | 2 |
| --------GE | 27 | 1.39 | 2 |
| --------K- | 26 | 1.34 | 1 |
| --------KE | 24 | 1.24 | 2 |
| ------L--E | 20 | 1.03 | 2 |
| -----TLOG- | 19 | 0.98 | 4 |

# PeptGen

https://www.hiv.lanl.gov/content/sequence/PEPTGEN/peptgen.html

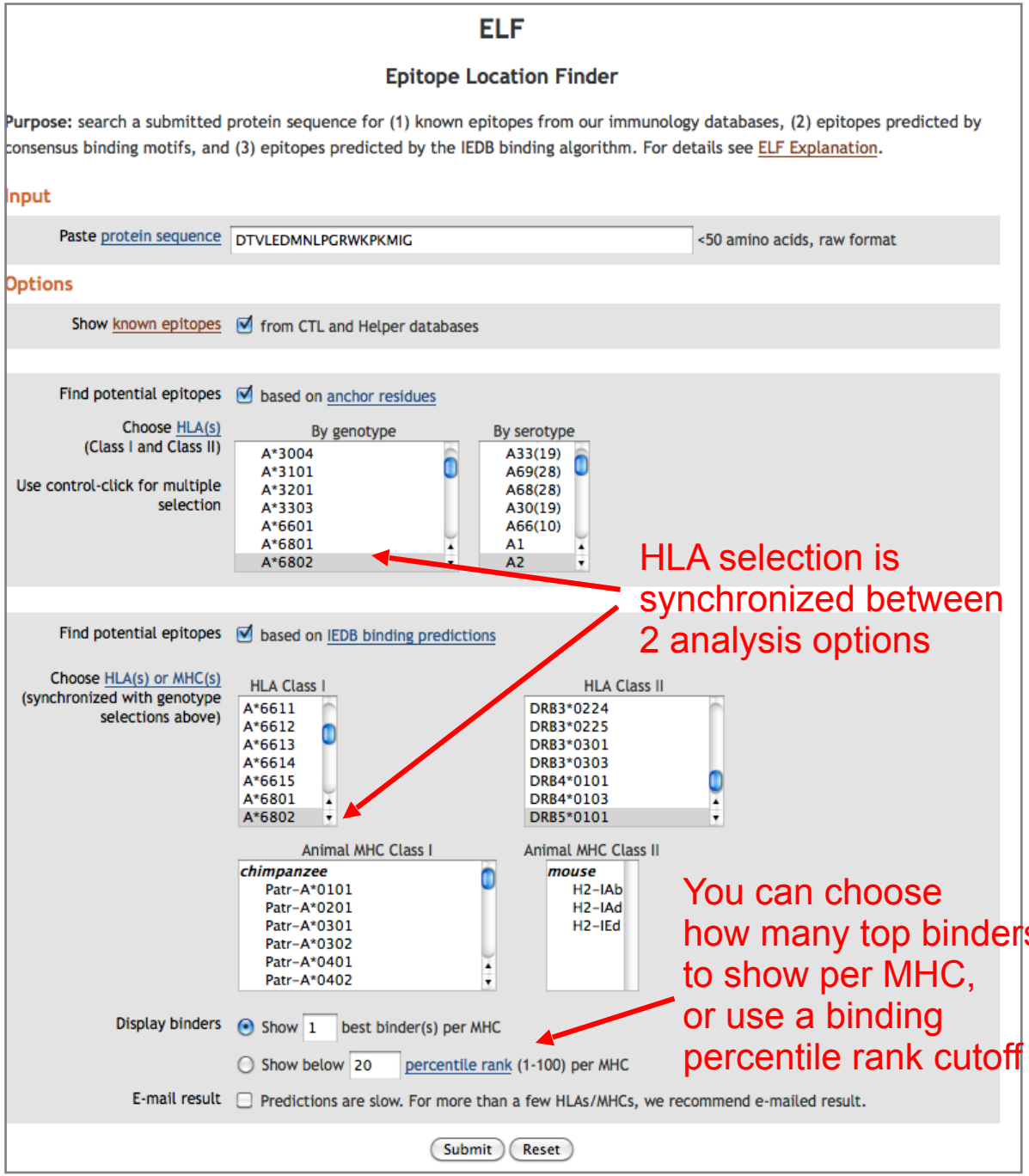
- Generates overlapping peptides for any protein sequence
- Takes alignment as an input and removes duplicate peptides

```
Seq1        HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK
Seq2        HLVWASRELERFALNPGLLETSEGCKQIIKQLQPALQTGTEELRSLYNTVATLYCVHEKIEVRDTKEALDKIEEEQNKSQ
Seq3        HLVWASRELERFALNPDLLETAEGCQQIMGQLQPALQTGTEELRSLFNTVATLYCVHQRIEVKDTKEALEEVEKIQKKSQ
```




```
1 HIVWASRELERFAVNPGL 1 s1 1  s1 - -
2 HLVWASRELERFALNPGL 1 s2 1  - s2 -
3 HLVWASRELERFALNPDL 1 s3 1  - - s3

4 LERFAVNPGLLETSEGCR 2 s1 1  s1 - -
5 LERFALNPGLLETSEGCK 2 s2 1  - s2 -
6 LERFALNPDLLETAEGCQ 2 s3 1  - - s3

7 GLLETSEGCRQILGQLQP 3 s1 1  s1 - -
8 GLLETSEGCKQIIKQLQP 3 s2 1  - s2 -
9 DLLETAEGCQQIMGQLQP 3 s3 1  - - s3

10 CRQILGQLQPSLQTGSEE 4 s1 1  s1 - -
11 CKQIIKQLQPALQTGTEE 4 s2 1  - s2 -
12 CQQIMGQLQPALQTGTEE 4 s3 1  - - s3

13 QPSLQTGSEELRSLYNTV 5 s1 1  s1 - -
14 QPALQTGTEELRSLYNTV 5 s2 1  - s2 -
15 QPALQTGTEELRSLFNTV 5 s3 1  - - s3

16 EELRSLYNTVATLYCVHQ 6 s1 1  s1 - -
17 EELRSLYNTVATLYCVHE 6 s2 1  - s2 -
18 EELRSLFNTVATLYCVHQ 6 s3 1  - - s3

19 TVATLYCVHQRIEVKDTK 7 s1&s3 2  s1 - s3
20 TVATLYCVHEKIEVRDTK 7 s2 1  - s2 -

21 HQRIEVKDTKEALEKIEE 8 s1 1  s1 - -
22 HEKIEVRDTKEALDKIEE 8 s2 1  - s2 -
23 HQRIEVKDTKEALEEVEK 8 s3 1  - - s3
```

# ELF (Epitope Location Finder)




- **ELF** helps identify potential T cell epitopes in a reactive peptide from a person with known HLA type by
  - Highlighting appropriate HLA anchor motifs in the peptide
  - Aligning all known epitopes embedded in the peptide from the database to your query sequence, with links to epitope entries
  - Finding potential epitopes based on Immune Epitope Database (IEDB) binding predictions http://www.immuneepitope.org/
- We also have **MotifScan** tool that shows HLA binding and custom motifs on the sequence alignment

*https://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer.html*

Los Alamos NATIONAL LABORATORY

# ELF (reported epitopes in HIV database)

## Epitopes from our CTL database aligned to your query sequence

Bold red letters indicate residues that differ from the query sequence. The symbol ◀ means the HLA of the epitope matches one of your submitted HLAs. Click on the epitope to see full database entry. Click on "align" to align the epitope to the sequence database via QuickAlign.

Epitopes shown here are completely within the bounds of your query. Epitopes that overlap the ends of your query are included in the "View database records" links above.

Download this alignment in format table

DTVLEDMNLPGRWKPKMIG

DTVLEEMNL A*6802 align ◀ Epitopes
DTVLEDINL A*6802 align ◀ matching
DTVLEEWNL A*6802 align ◀ requested HLAs

DTVLEEMNL A68 align
DTVLEEMNL A28 align
DTVLEDMNL align
EEMNLPGRW B44 align
EEINLPGKW B44 align
EEMNLPGRW B*4402 align
EEMNLPGRW B*4403 align
EEMNLPGRW B18,B40,B44 align
EDMNLPGRW align
EEMNLPGRW B*44 align
EEINLPGKW B*4403 align
EEMNLPGRW align
LPGRWKPKMI Cw3 align
LPGRWKPKMI B7 align

Clicking on an epitope takes you to respective CTL or Helper epitope Database entries

Clicking on the “align” button takes you to “QuickAlign” for that epitope

# ELF (predicted MHC binding)

## Potential epitopes based on anchor residues

These peptides have C-terminal anchor residues, highlighted in **blue**, and internal anchors highlighted in **magenta**. These anchor residues match one or more motifs associated with the submitted HLA.

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG
DTVLEDMNL (A*0205 ........[L])
DTVLEDMNL (A*6802 .[TV].....[VL])
 TVLEDMNLP (A*0206 .[VQ].......)
   LEDMNLPGR (DRB5*0101,DRB5*0101 [FYLM]..[QVIM]....[RK])
```

## Motifscan

*https://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan*

## Potential epitopes based on IEDB binding predictions

Top binders for each MHC are highlighted in **blue**.
Prediction method: IEDB recommended
Low percentile = good binders
Show up to 1 binder(s) per MHC

**Class I**

Selected allele(s): A*6802, B*1501

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
     DMNLPGRW      B*1501 (26)
      MNLPGRWK     A*6802 (3.0)
```

**Class II**

Selected allele(s): DRB5*0101

Download this alignment in format table

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
 TVLEDMNLPGRWKPK    DRB5*0101 (17.17)
```

## IEDB binding predictions

Clicking on MHC links to the full list of IEDB predictions for that MHC (see next table)

Los Alamos NATIONAL LABORATORY

# Potential epitopes based on IEDB database
# MHC binding predictions

## IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

## MHC-I binding predictions - Prediction Results

**Input Sequences**

| # | Name | Sequence |
|---|---|---|
| 1 | sequence 1 | DTVLEDMNLPGRWKPKMIG |

**Prediction method: IEDB recommended | Low percentile = good binders**

Check to expanded the result: ☐

| Allele | # | Start | End | Peptide Length | Sequence | Method used | Percentile Rank |
|---|---|---|---|---|---|---|---|
| HLA-B*15:01 | 1 | 6 | 13 | 8 | DMNLPGRW | *NetMHCpan* | 26 |
| HLA-B*15:01 | 1 | 3 | 13 | 11 | VLEDMNLPGRW | *NetMHCpan* | 27 |
| HLA-B*15:01 | 1 | 3 | 11 | 9 | VLEDMNLPG | Consensus (*ANN,SMM,CombLib_Sidney2008*) | 27.60 |
| HLA-B*15:01 | 1 | 8 | 17 | 10 | NLPGRWKPKM | *NetMHCpan* | 31 |
| HLA-B*15:01 | 1 | 7 | 17 | 11 | MNLPGRWKPKM | *NetMHCpan* | 35 |
| HLA-B*15:01 | 1 | 2 | 9 | 8 | TVLEDMNL | *NetMHCpan* | 36 |
| HLA-B*15:01 | 1 | 2 | 11 | 10 | TVLEDMNLPG | *NetMHCpan* | 47 |
| HLA-B*15:01 | 1 | 4 | 11 | 8 | LEDMNLPG | *NetMHCpan* | 48 |

# Thank you for attending!

## Please send us comments, questions, and suggestions!

**Your comments will help us provide future training and better tools.**

**Slides available at https://tinyurl.com/HIVDB-2019-IEDB**

## Contact us: seq-info@lanl.gov or immuno@lanl.gov

Los Alamos
NATIONAL LABORATORY