

IVAX Vaccine Design Toolkit

Vaccine Design for Beginners

Matthew Ardito

Prepared For:
Vaccine Renaissance IV - 2010




1

Overview

- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads



Overview




- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

Confidential 3

Gathering Raw Data

Popular Databases



- **GenBank (frequently used)**
 - A genetic sequence database, an annotated collection of all publicly available DNA sequences¹.
 - There are approximately 106,533,156,756 bases in 108,431,692 sequence records
- **SwissProt**
 - integral part of the UniProt Knowledgebase (UniProtKB).
 - a curated protein sequence database which provides a high level of annotation, minimal level of redundancy, and high level of integration with other databases
- **The Institute for Genomic Research (TIGR)**
 - leading center for microbial genome sequencing
 - now absorbed by the J Craig Venter Institute (JCVI)
- **Los Alamos National Laboratory (LANL)**
 - HIV/HCV genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials
- **PoxVirus**
 - a comprehensive Poxviridae informational and analytical resource

¹Nucleic Acids Research, 2008 Jan;36(Database issue):D25-30

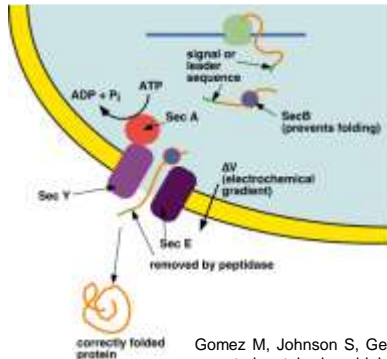
Confidential 4

Proteome Annotation/Protein Triage

Intracellular Pathogens - SECRETION



- Intracellular Pathogens take up residence in a host cell and secretes proteins that alter the chemistry of the host cell
 - These secreted proteins will eventually be degraded by the host cell's recycling machinery
- These proteins are presumably visible to the immune system.



Does the protein have a signal motif at its n terminus?

YES

Does the protein have additional transmembrane regions elsewhere?

NO

Does the protein contain a lipoprotein attachment site?

NO

Our protein is most likely secreted.

Gomez M, Johnson S, Gennaro ML. Identification of Mycobacterium tuberculosis secreted proteins by a bioinformatic approach. Infect Immunol 2000;68(4):2323-7.

Confidential

5

Proteome Annotation/Protein Triage

Intracellular Pathogens - SECRETION



- Signal P
 - <http://www.cbs.dtu.dk/services/SignalP>
 - predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms
- Phobius
 - <http://phobius.sbc.su.se/>
 - A combined transmembrane topology and signal peptide predictor
- LipoP
 - <http://www.cbs.dtu.dk/services/LipoP/>
 - produces predictions of lipoproteins and discriminates between lipoprotein signal peptides, other signal peptides and n-terminal membrane helices in Gram-negative bacteria

Confidential

6

Gathering Raw Data

FASTA format



- FASTA is a very simple file format, and one of the most common in the field.
- A sequence in FASTA format begins with a single-line description, followed by lines of sequence data.

```
>gi-168805691 hemagglutinin [Influenza A virus (A/Brisbane/59/2007 (H1N1))]
MKVKLLVLLCTFTATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLENSHNGKLCCLKGIAPLQLG
NCSVAGWILGNPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKESSW
PNHTVTGVSASCSHNGESSFYRNLLWLTGKNGLYPNLSKSYANNKEKEVLVWGVHHPNIGDQKALYHT
ENAYVSVVSSHYSRKFTPEIAKRPKVRDQEGRINYYWTLLEPGDTIIFEANGNLIAPRYAFALSRGFGSG
IINSNAPMDKCDKACQTPQGAINSSLPFQNVHPVTIGCEPKYVRSALRMVTGLRNIPSIQSRGLFGAIA
GFIEGGWTGMVDGWYGYHHQNEQSGSYAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERM
ENLNKKVDDGFI DIWTYNAELLVLENERLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNDE
CMESVKNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLLVSLGAI SFWMC SNGSLQ
CRICI
```

The description line starts with a greater than symbol (">").

The text after (">") is the sequence identifier (unique)

The text after the ID is the description (optional)

Next lines display the AA sequence in 80 character lines

Confidential

7

Gathering Raw Data

FASTA format



- A given FASTA file may contain one or more sequence entries.

```
Sequence 1
>gi-168805691 hemagglutinin [Influenza A virus (A/Brisbane/59/2007 (H1N1))]
MKVKLLVLLCTFTATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLENSHNGKLCCLKGIAPLQLG
NCSVAGWILGNPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSFERFEIFPKESSW
PNHTVTGVSASCSHNGESSFYRNLLWLTGKNGLYPNLSKSYANNKEKEVLVWGVHHPNIGDQKALYHT
ENAYVSVVSSHYSRKFTPEIAKRPKVRDQEGRINYYWTLLEPGDTIIFEANGNLIAPRYAFALSRGFGSG
IINSNAPMDKCDKACQTPQGAINSSLPFQNVHPVTIGCEPKYVRSALRMVTGLRNIPSIQSRGLFGAIA
GFIEGGWTGMVDGWYGYHHQNEQSGSYAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERM
ENLNKKVDDGFI DIWTYNAELLVLENERLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNDE
CMESVKNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLLVSLGAI SFWMC SNGSLQ
CRICI


Sequence 2
>gi-168805697 neuraminidase [Influenza A virus (A/Brisbane/59/2007 (H1N1))]
MNPNQKIITIGSISIAIGIISLMLQIGNIISIWASHSIQTGSQNTGICNQRIITYENSTWVNHTYVNI
NTNVVAGEDKTSVTLAGNSSLCSISGWAIIYTKDNSIRIGSKGDVFVIREPFISCSHLECRFFFLTQGALL
NDKHSNGTVKDRSPYRALMSCPLGEAPSPYNSKFESVAWSASACHDGMGWLTIIGISGPDNGAVAVLKYNG
IITGTIKSWKKQILRTQSECEVCMNGSCFTIMTDGSPNSKAASYKIFKIEKGKVTKSIELNAPNFHYEES
CYPDTGIVMVCVRDNWHGNSRNPWVSNQNLDYQIGYICSGVFGDNPRPEDGEGSCNPVTVDGANGVKGFS
YKYDNGVWIGRTKSNRLRKGFEIMIWDPNGWNTNTDSDFSVKQDVVAITDWSGYSGSFVQHPLELTGLDCIRP
CFWVELVRGLPRENTTIWTSGSSISFCGVNSDTANWSWPDGAELPFT
```

- FASTA format is the same whether your data is encoded as amino acids or nucleotides.

Confidential

8

Overview



- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

Confidential 9

Data Management

Uploading Raw Data



iVAX A safer, faster approach to vaccine design. [Home](#) | [Contact](#)

[About](#) [CCHI Collaborators](#) [iVAX Tool Kit](#) [News](#)

[Home](#) [Data Management](#) [Conservation Analysis](#) [Class I Analysis](#) [Class II Analysis](#) [BLAST Analysis](#) [Vaccine Design](#) [Ad Hoc Analysis](#)

Data Management

[File Manager](#)
Use this Link to Manage Uploaded Datasets

[Upload Proteins](#) ←
Use this Link to Upload Protein Data for Analysis


[Upload Clusters](#)
Use this Link to Upload T cell Epitope Clusters for Analysis

[Upload Archive](#)
Use this Link to Upload an Archived Analysis

Confidential 10

Uploading Proteins

Browse for Protein File or...



Upload a Protein File for Analysis

[Or Paste a Sequence for Analysis](#)

Choose a Protein File to Upload:	<input type="text"/> <input type="button" value="Browse..."/>
File Label (optional):	<input type="text"/>
Source of File:	GENBANK
Action:	<input checked="" type="radio"/> New <input type="radio"/> Overwrite
<input type="button" value="Upload"/>	

File format help:
Data files for upload must be named: *unique_file_name.pep*.

File format is FASTA. For example:

```
>Unique_sequence_name Optional sequence label
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES


>Unique_sequence_name Optional sequence label
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES

>Additional protein sequences...
```

Confidential 11

Uploading Proteins

Paste a Sequence



Paste a Sequence for Analysis

[Or Upload a Data File for Analysis](#)

File Name:	SAMPLE
File Label (optional):	<input type="text"/>
Source of File:	GENBANK
Action:	<input checked="" type="radio"/> New <input type="radio"/> Overwrite <input type="radio"/> Append


Enter an Amino Acid Sequence in FASTA format:
>Unique_sequence_name Optional sequence label
AMWQ_ACID_SEQUENCE_IN_80_CHARACTER_LINES


```
>PROTEIN-1
GKLPQWRNSTALCLGRSAVFNGLVYTIINRQKQVTHATKLVQSSSTGRICSPHAILRGRNCTLRAL
LGRFPCRSFQWQWHLFVYRGRAYSMCFPHVPTASLRSLVALSSGTLKFINRPFMTOVAQWQSSACK
RQSVNSFFSALNGLHRSKYKYPALNVTMFINRKYRKYIINQVSRPTRRQTSLYVAASGRVTYSTRSQ
QTVIPIHGGSPWYRGGSSKISLYWTVKSGRILLHSIDHLLAPRQVFKIRNGKSSDHSRAPIGTCSK
CITFHQIIPHRQFFQHVRLITYGACPRVYVQHTLKLATDQGVYRQTE
```

12

Data Management

Uploading Raw Data





A safer, faster approach to vaccine design.

[Home](#) | [Contact](#)

About
CCHI Collaborators
iVAX Tool Kit
News

Home
Data Management
Conservation Analysis
Class I Analysis
Class II Analysis
BLAST Analysis
Vaccine Design
Ad Hoc Analysis

Data Management

[File Manager](#) ←

- Use this Link to Manage Uploaded Datasets
- [Upload Proteins](#)
Use this Link to Upload Protein Data for Analysis
- [Upload Clusters](#)
Use this Link to Upload T cell Epitope Clusters for Analysis
- [Upload Archive](#)
Use this Link to Upload an Archived Analysis

Confidential
13

File Manager / Editor

Description



Home
Data Management
Protein Analysis
Cluster Analysis
Blast Analysis
Ad Hoc Analysis
Logout

File Manager

(Query By Example)

RowNum	File	File Label	Source	# Proteins	Date
1	SAMPLE		GENBANK	1	12-JUL-10 10:07:49

[Simple Search \(simple AND\)](#)

[Advanced Search \(SQL Where Clause\)](#)

File Editor

File:

File Label:


Source:

Proteins: [1 \(Edit\)](#)

UserName:

Date Time:

[Return to File List](#)



Sequence Manager / Editor

Description

Sequence Manager
(Query By Example)

RowNum	File	Sequence ID	Sequence Label	AA SEQUENCE	DataLen	Temp Score	Date
1	SAMPLE	PROTEIN-1	PROTEIN-1	QKLPGRHNSATLCLQHP	329	Empty	14-OCT-10

[Simple Search \(simple AND\)](#) [Execute Query](#) [Advanced Search \(SQL Where Clause\)](#)

Sequence Editor

File:	SAMPLE
Sequence ID:	PROTEIN-1
Sequence Label:	PROTEIN-1
Sequence:	QKLPGRHNSATLCLQHP
Data Length:	329
Temp Score:	
UserName:	MARDITO
Date Time:	14-OCT-10 03:10:03

[Update](#) [Delete](#) [Insert](#)


[Return to Sequence Manager](#)

Sequence Insert

File:	SAMPLE
Sequence ID:	
Sequence Label:	
Sequence:	
Data Length:	
Temp Score:	
UserName:	MARDITO
Date Time:	14-OCT-10 03:10:21

[Submit](#)

[Return to Sequence Manager](#)



Overview

- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

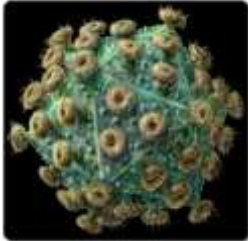
Confidential 16

Conservation Analysis

Variable Pathogens



- The genetic variability of some pathogens constitutes a significant challenge to efforts to design a vaccine driven by cellular immune response.



HIV



HCV



Influenza

- Vaccines designed to protect against a given strain or clade of a quickly mutating pathogen may be ineffective when faced with a heterologous challenge.
- One approach to solving this problem is to include conserved and functionally (or structurally) important epitopes in the vaccine.

Confidential

17

Conservation Analysis

Multiple Alignments (ClustalW)



- <http://www.ebi.ac.uk/Tools/clustalw2/index.html>
- Perform alignment of all available isolates and manually extract conserved segments.
- This might be a viable strategy assuming the protein you are interested in is
 - a) fairly short and b) you have a limited number of isolates to align.

The screenshot shows the ClustalW web interface with the following sections:


- YOUR SEIAL:** Input field for sequence ID.
- ALIGNMENT TITLE:** Input field for title, with 'Sequence' selected.
- RESULTS:** Includes 'interactive' (selected), 'SCORE TYPE' (percent), 'NO END GAPS' (yes), and 'GAP EXTENSION' (yes).
- ALIGNMENT:** Includes 'full' (selected), 'TORDIAG', 'PAREGAP', 'GAP DISTANCES' (def), and 'NUMBER' (1).
- OUTPUT:** Includes 'OUTPUT FORMAT' (tab writers), 'OUTPUT ORDER' (assigned), 'TREE TYPE' (none), 'PHYLOGENETIC TREE' (correct dist, off), 'SHRINK GAPS' (off), and 'CLUSTERING' (FLS).
- Input Field:** A large text area with the prompt 'Enter or paste a set of sequences in any supported format' and a 'Help' button.
- Buttons:** 'Submit' and 'Cancel' buttons at the bottom right.

Confidential

18

Conservation Analysis

Multiple Alignments (ClustalW)



```

Q1-46957562 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q2-46957540 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q3-46957540 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q4-46957562 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q5-46957566 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q6-46957544 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q7-46738225 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q8-46738215 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
.....

Q1-46957562 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q2-46957540 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q3-46957540 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q4-46957562 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q5-46957566 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q6-46957544 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q7-46738225 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 117
Q8-46738215 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 116
.....


Q1-46957562 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q2-46957540 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q3-46957540 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q4-46957562 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q5-46957566 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q6-46957544 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q7-46738225 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q8-46738215 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
.....

Q1-46957562 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q2-46957540 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q3-46957540 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q4-46957562 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q5-46957566 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q6-46957544 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q7-46738225 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q8-46738215 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 224
.....
    
```

19

Conservation Analysis

Consensus Sequences (Consensus)



- Once you have an alignment, you can create a consensus sequence
- <http://coot.embl.de/Alignment/consensus.html>

Consensus

```

Q1-46957562 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q2-46957540 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q3-46957540 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q4-46957562 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q5-46957566 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q6-46957544 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q7-46738225 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
Q8-46738215 MDVWQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 60
.....

Q1-46957562 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q2-46957540 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q3-46957540 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q4-46957562 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q5-46957566 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q6-46957544 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 120
Q7-46738225 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 117
Q8-46738215 AKAVDIEVINPMAATHACVFTDQHSQEVVLGHWTEHFNHSHHSHVQGHEDIISLWDQGLR 116
.....

Q1-46957562 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q2-46957540 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q3-46957540 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q4-46957562 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q5-46957566 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q6-46957544 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q7-46738225 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
Q8-46738215 DCVLLIELCVTLKCTHIVN-----SHITNRRSGDQ-IHNCSEFVETDIDRDTIR 167
.....

Q1-46957562 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q2-46957540 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q3-46957540 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q4-46957562 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q5-46957566 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q6-46957544 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q7-46738225 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 227
Q8-46738215 EYALFVHLDPVQIIRHYQHLHLINWNGIMLLGHLICHAEEIKLWVTVYGVFVWMAATTLFCADD 224
.....
    
```


Note: Frequency ONLY. Does not weigh for immunogenicity

Threshold: [10] [20] [30] [40] [50] [60] [70] [80] [90] [100]

[Submit] [Submit All] [Exit]

You could see the study the "study" of this score.

Overview




- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

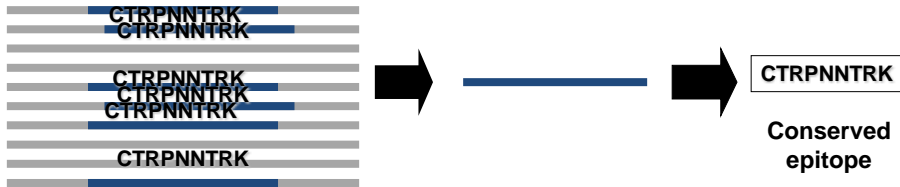
Confidential 21

Conservation Analysis

Conservation without MultiAlign



Conservatrix Finds Conserved 9mers



Identifying the most conserved 9mers will allow us to protect against more strains with fewer epitopes

You can set a cutoff for the acceptable level of conservation so that only the results you want are displayed.

Conservatrix will list the conserved peptide, and which proteins it appears in

Conservatrix does not care where a 9mer appears within a protein, only that it does appear

Confidential 22

Conservatrix Analysis

Execution





A safer, faster approach to vaccine design.

[Home](#) | [Contact](#)

[About](#) [CCHI Collaborators](#) [iVAX Tool Kit](#) [News](#)

[Home](#) [Data Management](#) [Conservation Analysis](#) [Class I Analysis](#) [Class II Analysis](#) [BLAST Analysis](#) [Vaccine Design](#) [Ad Hoc Analysis](#)

Conservation Analysis

[Conservatrix](#) ←
Use this Link to execute a conservation analysis w/ Conservatrix.


[Aggregatrix](#)
Use this Link to summarize coverage w/ Aggregatrix.

[EpiAssembler](#)
Use this Link to summarize coverage w/ EpiAssembler.

Confidential
23

Conservatrix Analysis

Execution




Execute Conservatrix to Identify Conserved Peptides

Choose an Accession:	<input type="text" value="FLU-HA"/>
New Sequence:	<input type="text" value="FLU-HA-CVX-DATA"/>
Parse:	<input checked="" type="radio"/> By Nine <input type="radio"/> By Ten
Show Sequence Name?	<input checked="" type="radio"/> True <input type="radio"/> False
Cutoff % of 100:	<input type="text" value=".5"/>
<input type="button" value="Execute"/>	

Confidential
24

Conservatrix Analysis

Frequency Table




Conservatrix Results

Current Accession: FLU-HA
Report Date: 24-SEP-2010 : 12:49

#	Accession	Sequence	Parse	Peptide	Count	Percent
1	FLU-HA	FLU-HA-CVX-DATA	NINE	ACPRVVKQN	4	1.00
SEQS: NOME-2050.BOSTON-2025.CHICAGO-2013.PROVIDENCE-2012						
2	FLU-HA	FLU-HA-CVX-DATA	NINE	CPRVVKQNT	4	1.00
SEQS: NOME-2050.BOSTON-2025.CHICAGO-2013.PROVIDENCE-2012						
3	FLU-HA	FLU-HA-CVX-DATA	NINE	GACPRVVKQ	4	1.00
SEQS: NOME-2050.BOSTON-2025.CHICAGO-2013.PROVIDENCE-2012						
...
...
...
99	FLU-HA	FLU-HA-CVX-DATA	NINE	CTLIDALLG	2	.50
SEQS: NOME-2050.BOSTON-2025						
100	FLU-HA	FLU-HA-CVX-DATA	NINE	CLGHHAVPN	2	.50
SEQS: CHICAGO-2013.PROVIDENCE-2012						

Confidential
25

Overview




- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

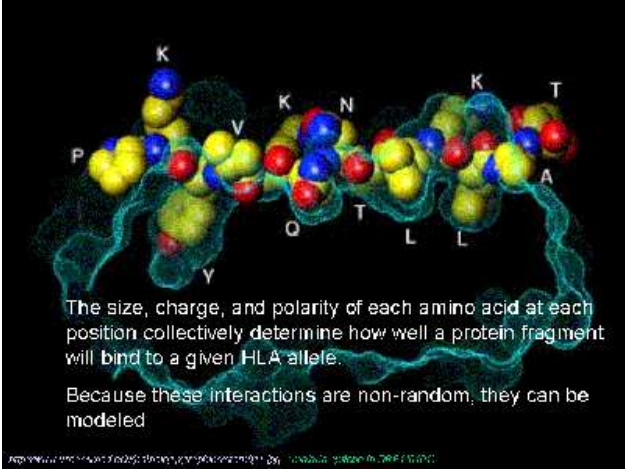
Confidential
26

Screening for T cell Epitopes

MHC/Ligand Interactions



- Binding is mediated by the interaction of the candidate peptides side chains with specific regions in the floor of the MHC binding groove.



The size, charge, and polarity of each amino acid at each position collectively determine how well a protein fragment will bind to a given HLA allele.


Because these interactions are non-random, they can be modeled

Representational and visualization of molecular models by Molecular Modeling in DRP/ICDC

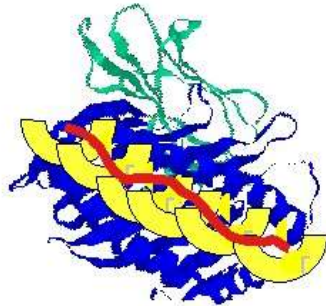
Confidential 27

Screening for T cell Epitopes

MHC/Ligand Interactions



- The binding groove of both Class I and Class II can be divided into 9 such regions

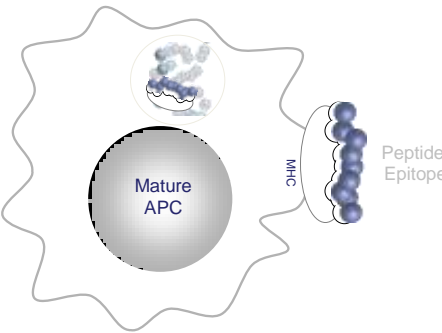


- In the case of Class I MHC, the binding groove is closed-ended and can accommodate peptides between 8 and about 11 amino acids in length, although 9-mer and 10-mer peptides are preferred.
- In the case of Class II MHC the binding groove is open ended. Class II MHC can accommodate longer peptides, typically 12-20 amino acids in length.

Confidential 28

Screening for T cell Epitopes

How does in silico mapping work?



The cell epitopes are linear and restricted by MHC (HLA).

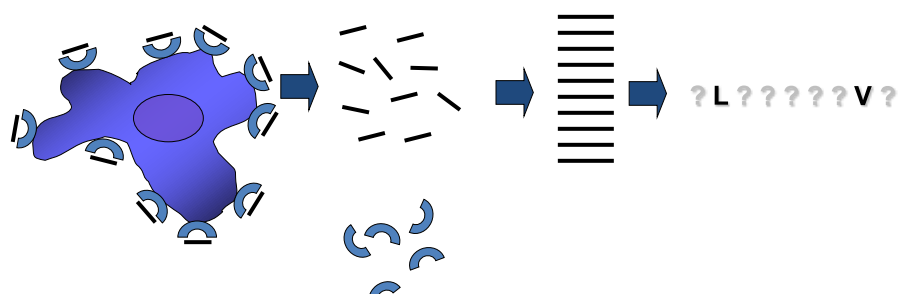
Because the pockets of the HLA are well known, interactions with peptides can be modeled.

The EpiMatrix algorithm scores all the 9-mers in a given sequence for binding affinity across a range of common HLA and reports both detailed and aggregated results.

29

Screening for T cell Epitopes

Constructing the matrix



- Early researchers eluted, sequenced, and aligned peptides bound to MHC
- They (Falk et al.) discovered that certain amino acids appeared in certain positions more often than others.
- This information was used to develop rudimentary binding “motifs”

Falk, K., Rötzschke, O., Stevanovic, S., Jung, G. and Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 1991, 351,290-296.

30

Screening for T cell Epitopes

Other methods



- As additional training data became available, many statistically based techniques were used to model the data:
 - Frequency Analysis
 - Hidden Markov Models (HMM)
 - Support Vector Machines (SVM)
 - Stabilization Matrix Methods (SMM)
 - Random Forests
 - Naive Bayesian Analysis

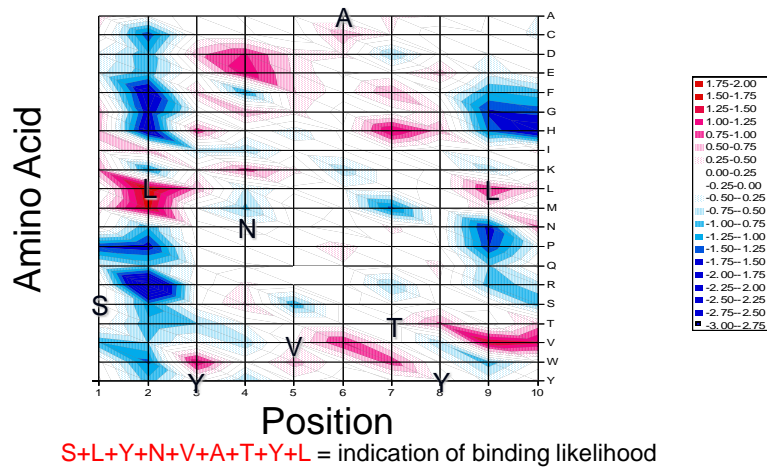
31

Screening for T cell Epitopes

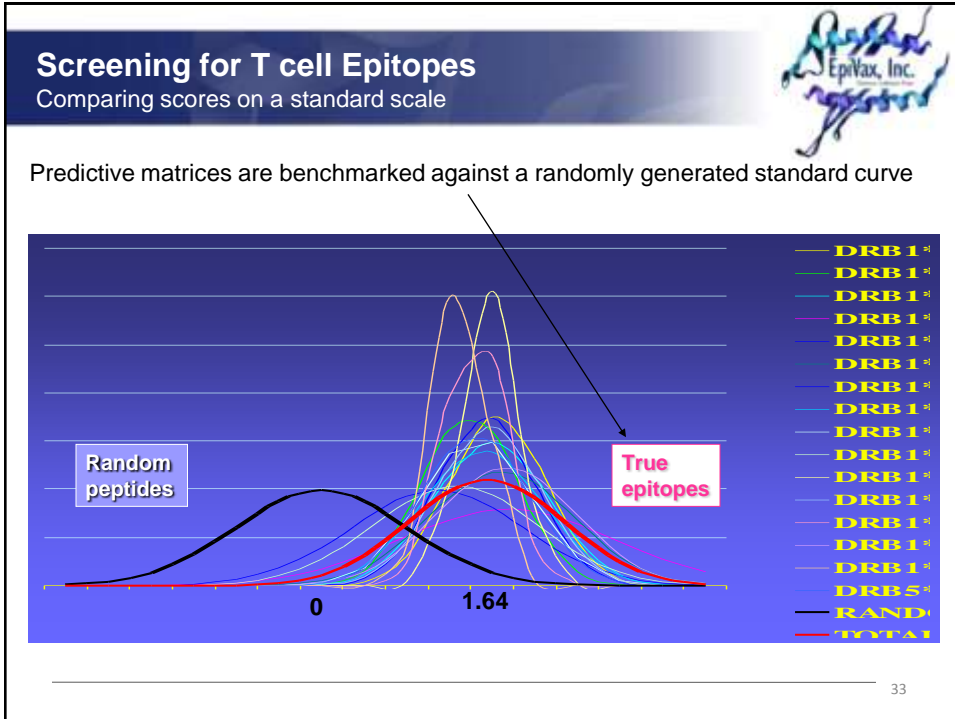
How EpiMatrix Scores



Graphical Representation of A*0201 Coefficient matrix




32



Screening for T cell Epitopes

EpiMatrix: Class I Analysis



iVAX A safer, faster approach to vaccine design. [Home](#) | [Contact](#)

[About](#) [CCH Collaborators](#) [iVAX Tool Kit](#) [News](#)

[Home](#) [Data Management](#) [Conservation Analysis](#) [Class I Analysis](#) [Class II Analysis](#) [BLAST Analysis](#) [Vaccine Design](#) [Ad Hoc Analysis](#) [Tutorial](#)

Class I EpiMatrix Analysis

[Execute a Standard EpiMatrix Class I Analysis](#) ←
 Use this Link to Execute a Standard EpiMatrix Class I Analysis

[Execute an Extended EpiMatrix Class I Analysis](#)
 Use this Link to Execute an Extended EpiMatrix Class I Analysis

[Class I Protein Summary Report](#)
 Use this Link to Display the Protein Summary Report

[Class I Protein Immunogenicity Scale](#)
 Use this Link to Display the Protein Immunogenicity Scale

[Class I Protein Immunogenicity Report](#)
 Use this Link to Display the Results of EpiMatrix Analysis.

Class I Cluster Analysis

[Find Class I T Cell Epitope Clusters](#)
 Use this Link to Screen Protein Sequences for Class I T Cell Epitope Clusters

[Interactive Class I Cluster Immunogenicity Report](#)
 Use this Link to Display the Results of an EpiMatrix Class I Cluster Analysis


[Printable Class I Cluster Immunogenicity Report](#)
 Use this Link to Print or Download an EpiMatrix Class I Cluster Analysis

Confidential

34

Screening for T cell Epitopes

EpiMatrix: Class I Standard Analysis



Score Standard ClassI File with EpiMatrix

Choose a File:

Choose a Protein or Null for All Proteins:

Execute

FLU-HA

BOSTON-2025
 CHICAGO-2013
 FLU-HA-CVX-DATA
 NOME-2050
 PROVIDENCE-2012

Confidential 35


EpiMatrix Results

Current File: SAMPLE
Current Sequence: SAMPLE_01

Top 10% of Z-Scores			Top 5% of Z-Scores				Top 1% of Z-Scores			
All Z-Scores In Top 5% are Considered "Hits"										
Matrix:			KB_A0101_09	KB_A0201_09	KB_A0301_09	KB_A2402_09	KB_B0702_09	KB_B4403_09	Hit	Average
AA Sequence	AA Start	GRAVY	Z-Score	Z-Score	Z-Score	Z-Score	Z-Score	Z-Score	Count	Z-Score
MGARASVLT	1	0.79	-0.34	0.24	-0.48	-1.15	-0.39	-0.14	0	-0.38
GARASVLTG	2	0.53	-0.85	-0.36	0.17	-1.28	0.16	0.32	0	-0.31
ARASVLTGS	3	0.49	-0.01	-0.19	-0.52	-0.91	-0.3	0.05	0	-0.31
RASVLTGSK	4	-0.14	0.75	-0.24	2.22	-0.73	0.54	-0.49	1	0.34
ASVLTGSKL	5	0.78	0.42	0.29	0.33	0.93	1.26	1.15	0	0.73
SVLTGSKLD	6	0.19	-0.01	0.5	0.73	-0.64	-0.6	-0.51	0	-0.09
VLTGSKLDA	7	0.48	0.03	1.3	0.28	-0.59	-0.22	-0.16	0	0.11
LTGSKLDAW	8	-0.09	0.68	-0.09	-0.18	-0.04	0.5	0.44	0	0.22
TGSKLDAWE	9	-0.9	-0.2	-0.84	-1.15	-0.18	-1.05	-1.49	0	-0.82
GSKLDAWEQ	10	-1.21	0.34	-1.13	-0.51	-1.49	-1.55	-0.9	0	-0.87
SKLDAWEQI	11	-0.67	-1.24	0.73	-1.07	0.94	0.68	0.57	0	0.1
KLDAWEQIR	12	-1.08	0.33	0.97	1.57	-0.91	-1.52	-0.92	0	-0.08
LDWEQIRL	13	-0.22	-0.01	0.91	-1.04	0.39	1.28	-0.1	0	0.24
DAWEQIRLK	14	-1.08	0.45	-0.45	1.06	-0.76	-1.07	-0.21	0	-0.16
AWEQIRLKP	15	-0.87	0	-1.56	-1.08	0.3	-0.46	-0.08	0	-0.48
WEQIRLKPG	16	-1.11	-2.36	-0.86	-1.43	-0.72	-0.5	2.66	1	-0.54
EQIRLKPGC	17	-0.73	-1.32	-0.54	-0.77	-0.91	-2.12	0.06	0	-0.93
QIRLKPGCCK	18	-0.78	0.04	-0.33	2.47	-0.95	-0.29	-1.72	1	-0.13
IRLKPGCCKK	19	-0.82	-0.16	-0.83	1.72	-0.5	-0.47	-1.07	1	-0.22
RLKPGCCKKK	20	-1.76	0.15	0.47	2.16	-0.22	-0.13	-0.77	1	0.38
LKPGCCKKKY	21	-1.4	2.06	-1.37	0.36	-0.15	-0.31	-0.01	1	0.1
KPGCKKKYR	22	-2.32	-1.1	-1.24	1.04	-1.41	1.34	-0.24	0	-0.27
PGCKKKYRL	23	-1.47	-1.13	-0.42	-1.14	0.61	0.02	-0.13	0	-0.37
GCKKKYRLK	24	-1.72	-0.42	-1.63	1.28	-1.17	-1.4	-1.17	0	-0.75
.
PFASLKSFL	488	0.88	0.16	-1.64	-0.31	2.02	0.34	0.59	1	0.19
FASLKSFLG	489	1.01	0.88	0.52	0.49	-0.79	-0.09	-0.22	0	0.13
ASLKSFLGT	490	0.62	0.39	0.76	0.59	-0.31	-0.67	0.81	0	0.26
SLKSFLGTD	491	0.03	-0.79	0.85	0.83	-0.22	-0.24	-0.64	0	-0.04
LKSFLGTDQ	492	-0.27	0.22	-0.79	-0.9	-0.63	-0.88	-1.44	0	-0.74
Maximum			3.67	3.1	2.94	2.84	4.11	3.31	2	--
Sum of Significant Z			25.98	49.57	50.35	25.54	55.06	34.08	--	--
Hit Count			11	23	22	11	33	25	125	--
Total Assessments:	2952	Total Significant Z:	240.58	Expected Z:	326.38	Deviation:	-75.8	Deviation per 1000:	-14.36	

Screening for T cell Epitopes

EpiMatrix: Class I Extended Analysis



Score Extended ClassI File with EpiMatrix

Choose a File:

Choose a Protein or Null for All Proteins:

Choose a Parse Number: By Nine By Ten

A Locus

<input checked="" type="checkbox"/> A0101 By 9	<input checked="" type="checkbox"/> A0201 By 9	<input checked="" type="checkbox"/> A0301 By 9	<input type="checkbox"/> A1101 By 9	<input checked="" type="checkbox"/> A2402 By 9	<input checked="" type="checkbox"/> A6801 By 9
<input type="checkbox"/> A0101 By 10	<input type="checkbox"/> A0201 By 10	<input type="checkbox"/> A0301 By 10	<input type="checkbox"/> A1101 By 10	<input type="checkbox"/> A2402 By 10	<input type="checkbox"/> A6801 By 10

B Locus

<input checked="" type="checkbox"/> B0702 By 09	<input checked="" type="checkbox"/> B0801 By 09	<input checked="" type="checkbox"/> B2705 By 09	<input checked="" type="checkbox"/> B3501 By 09	<input checked="" type="checkbox"/> B4403 By 09	Reserved
<input type="checkbox"/> B0702 By 10	<input type="checkbox"/> B0801 By 10	<input type="checkbox"/> B2705 By 10	<input type="checkbox"/> B3501 By 10	<input type="checkbox"/> B4403 By 10	Reserved

Confidential

37

Screening for T cell Epitopes

EpiMatrix: Class II Standard Analysis





A safer, faster approach to vaccine design.

[Home](#) | [Contact](#)

[Home](#)
[Data Management](#)
[Conservation Analysis](#)
[Class I Analysis](#)
[Class II Analysis](#)
[BLAST Analysis](#)
[Vaccine Design](#)
[Ad Hoc Analysis](#)
[Tutorial](#)

Class II EpiMatrix Analysis

[Execute a Standard EpiMatrix Class II Analysis](#) ←

Use this Link to Execute a Standard EpiMatrix Class II Analysis.

[Execute an Extended EpiMatrix Class II Analysis](#)

Use this Link to Execute an Extended EpiMatrix Class II Analysis.

[Class II Protein Summary Report](#)

Use this Link to Display the Protein Summary Report.

[Class II Protein Immunogenicity Scale](#)

Use this Link to Display the Protein Immunogenicity Scale.

[Class II Protein Immunogenicity Report](#)

Use this Link to Display the Results of EpiMatrix Analysis.

Class II Cluster Analysis

[Find Class II T Cell Epitope Clusters](#)

Use this Link to Screen Protein Sequences for Class II T Cell Epitope Clusters.

[Interactive Class II Cluster Immunogenicity Report](#)

Use this Link to Display the Results of an EpiMatrix Class II Cluster Analysis.

[Printable Class II Cluster Immunogenicity Report](#)


Use this Link to Print or Download an EpiMatrix Class II Cluster Analysis.

Confidential

38

Screening for T cell Epitopes

EpiMatrix: Class II Standard Analysis



Score Standard ClassII File with EpiMatrix

Choose a File:

Choose a Protein or Null for All Proteins:

BOSTON-2025
 CHICAGO-2013
 FLU-HA-CVI-DATA
 NOME-2050
 PROVIDENCE-2012

Success! Standard Class II Scoring Complete

ClassII EpiMatrix Protein Summary Report

Choose a File:


Choose a Protein:

Confidential

39

Screening for T cell Epitopes

EpiMatrix: Class II Standard Analysis



Overview of ClassII Protein Selections

Current File: FLU-HA - Current Sequence: All Proteins
September 29, 2010 ;(Epx Ver. 1.2);

[Click to Print](#) [Click to Download](#)

File	Protein Sequence	Protein Label	Length	EpiMatrix Hits	EpiMatrix Score	Reg Adjusted Epx Score
FLU-HA	BOSTON-2025	hemagglutinin	329	99	-30.78	-30.78
FLU-HA	CHICAGO-2013	hemagglutinin	329	189	44.68	44.68
FLU-HA	NOME-2050	hemagglutinin	329	170	29.37	29.37
FLU-HA	PROVIDENCE-2012	hemagglutinin	329	247	89.24	89.24

↑

Information on the proteins analyzed as part of this analysis is provided above.
EpiMatrix Hits is the number of EpiMatrix Z scores above 1.64 found within the sequence.
The EpiMatrix Score is derived from the number and intensity of the EpiMatrix Hits normalized for the length of the protein.
In other words the Score is the excess or shortfall in predicted aggregate immunogenicity relative to a random peptide standard.


Confidential

40

ClassII EpiMatrix Report

File: FLU-HA - Sequence: PROVIDENCE-2012
September 29, 2010 (Epx Ver. 1.2)
[Click to Print](#) [Click to Download](#)

Frame Start	AA Sequence	Frame Stop	DRB1*0101 Z-Score	DRB1*0301 Z-Score	DRB1*0401 Z-Score	DRB1*0701 Z-Score	DRB1*0801 Z-Score	DRB1*1101 Z-Score	DRB1*1301 Z-Score	DRB1*1501 Z-Score	Hits
1	QKLPGNRRNS	9	0.58	0.25	1.06	-0.2	0.23	0.95	0.45	0.31	0
2	KLPGNRNST	10	0.41	0.48	-0.19	0.57	0.34	0.8	-0.75	0.84	0
3	LPGNRRNSTA	11	0.94	1.03	1.26	0.5	0.49	0.96	0.92	0.84	0
4	PGNRNSTAT	12	0.35	-1.66	-0.43	0.23	-0.21	-1.07	-0.57	0.18	0
5	GNRNSTATL	13	1.1	0.74	1.16	0.6	0.62	0.95	0.95	0.2	0
6	NRNSTATLCL	14	1.14	0.29	0.51	1.04	0.07	-0.39	-0.05	0.58	0
.
307	RYVKQNTLK	315	-0.34	0.32	-0.1	-0.51	0.58	-0.57	0.77	-0.23	0
308	YVKQNTLKL	316	3.06	2.28	3.18	2.81	2.43	2.81	3.11	2.55	8
309	VKQNTLKL	317	0.97	1.51	0.95	1.06	1.62	2.01	1.7	1.41	2
310	KQNTLKLAT	318	0.49	-0.1	0.22	0.54	1	0.89	0.86	1.34	0
311	QNTLKLATG	319	0.15	-0.22	0.23	-1.29	1.19	1.26	0.23	0.01	0
312	NLKLATGTM	320	0.24	0.63	-0.41	0.33	-0.02	-0.77	1.07	-0.44	0
313	TLKLATGMR	321	0.9	0.78	0.81	0.55	1.24	0.7	-0.09	0.46	0
314	LKLATGMRN	322	1.93	1.17	1.92	1.23	1.86	1.4	0.2	2.35	4
315	KLATGMRNV	323	-0.23	-1.07	-0.73	0.29	-0.96	-0.6	-0.69	-0.55	0
316	LATGMRNV	324	-0.57	0.68	-0.6	-0.24	-0.07	0.37	-0.55	0.01	0
317	ATGMRNVPK	325	1.18	0.31	1.32	0.09	0.6	0.74	0.15	0.6	0
318	TGMRNVPKK	326	-0.24	-1.37	-0.28	-0.98	0.52	0.23	-0.18	-1	0
319	GMRNVPKKQ	327	0.86	0.13	0.98	1.06	1.04	1.26	0.34	0.04	0
320	MRNVPKKQT	328	0.81	0.36	-0.03	0.85	1.13	1.58	0.77	1.41	0
321	RNVPKKQTR	329	-1.22	1.25	-1.36	-1.75	0.16	-0.74	0.7	0.34	0
Summarized Results			DRB1*0101	DRB1*0301	DRB1*0401	DRB1*0701	DRB1*0801	DRB1*1101	DRB1*1301	DRB1*1501	Total
Maximum Single Z score			3.06	2.69	3.18	3.02	2.77	3.29	3.11	2.55	--
Sum of Significant Z scores			65.04	58.45	52.95	64.72	51.65	81.61	62.54	72.96	509.92
Count of Significant Z Scores			30	29	25	30	25	38	31	36	244
Total Assessments Performed: 2568			Deviation from Expectation: 224.26				Deviation per 1000 AA: 89.24				
Adjusted for Regulatory Epitopes			Deviation from Expectation: 224.26				Deviation per 1000 AA: 89.24				



Screening for T cell Epitopes

EpiMatrix: Class II Standard Analysis

Overview of ClassII Protein Selections

Current File: FLU-HA - Current Sequence: All Proteins
September 29, 2010 (Epx Ver. 1.2);
[Click to Print](#) [Click to Download](#)

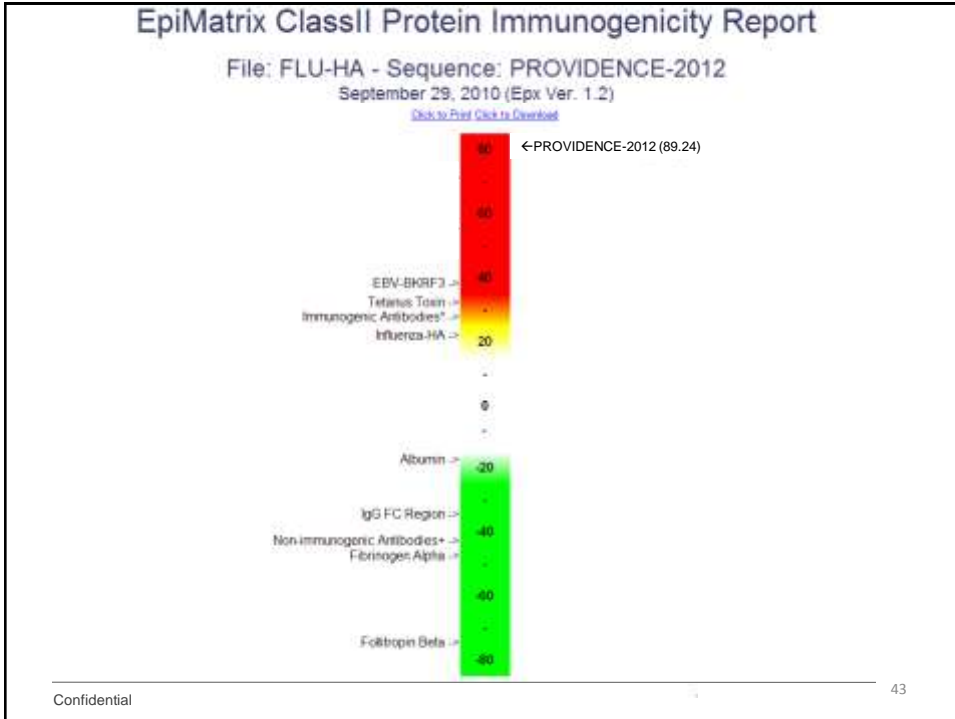
File	Protein Sequence	Protein Label	Length	EpiMatrix Hits	EpiMatrix Score	Reg Adjusted Epx Score
FLU-HA	BOSTON-2025	hemagglutinin	329	99	-30.78	-30.78
FLU-HA	CHICAGO-2013	hemagglutinin	329	189	44.68	44.68
FLU-HA	NOME-2050	hemagglutinin	329	170	29.37	29.37
FLU-HA	PROVIDENCE-2012	hemagglutinin	329	247	89.24	89.24

↑

Information on the proteins analyzed as part of this analysis is provided above.
EpiMatrix Hits is the number of EpiMatrix Z scores above 1.64 found within the sequence.
The EpiMatrix Score is derived from the number and intensity of the EpiMatrix Hits normalized for the length of the protein.
In other words the Score is the excess or shortfall in predicted aggregate immunogenicity relative to a random peptide standard.

42

Confidential



Screening for T cell Epitopes

EpiMatrix: Class II Extended Analysis

Score Extended ClassII File with EpiMatrix

Choose a File:

Choose a Protein or Null for All Proteins:

<input type="checkbox"/> DRB1*0101	<input type="checkbox"/> DRB1*0301	<input type="checkbox"/> DRB1*0401	<input type="checkbox"/> DRB1*0701	<input type="checkbox"/> DRB1*0901	<input type="checkbox"/> DRB1*1101	<input type="checkbox"/> DRB1*1301	<input type="checkbox"/> DRB1*1501
<input type="checkbox"/> DRB1*0102	<input type="checkbox"/> DRB1*0302	Reserved	<input type="checkbox"/> DRB1*0702	<input type="checkbox"/> DRB1*0902	<input type="checkbox"/> DRB1*1102	<input type="checkbox"/> DRB1*1302	<input type="checkbox"/> DRB1*1502
<input type="checkbox"/> DRB1*0104	<input type="checkbox"/> DRB1*0304	Reserved	Reserved	<input type="checkbox"/> DRB1*0804	<input type="checkbox"/> DRB1*1104	<input type="checkbox"/> DRB1*1304	<input type="checkbox"/> DRB1*1504
<input type="checkbox"/> DRB1*0105	<input type="checkbox"/> DRB1*0307	Reserved	Reserved	<input type="checkbox"/> DRB1*0908	<input type="checkbox"/> DRB1*1106	<input type="checkbox"/> DRB1*1305	<input type="checkbox"/> DRB1*1508
<input type="checkbox"/> Reserved	<input type="checkbox"/> DRB1*0308	Reserved	Reserved	<input type="checkbox"/> DRB1*0817	<input type="checkbox"/> DRB1*1107	<input type="checkbox"/> DRB1*1311	Reserved
<input type="checkbox"/> Reserved	<input type="checkbox"/> DRB1*0309	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1114	<input type="checkbox"/> DRB1*1321	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1116	<input type="checkbox"/> DRB1*1322	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1120	<input type="checkbox"/> DRB1*1323	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1121	<input type="checkbox"/> DRB1*1327	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1127	<input type="checkbox"/> DRB1*1328	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1128	<input type="checkbox"/> DRB1*1334	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1130	Reserved	Reserved
<input type="checkbox"/> Reserved	Reserved	Reserved	Reserved	Reserved	<input type="checkbox"/> DRB1*1133	Reserved	Reserved

Confidential 44

Screening for T cell Epitopes

EpiMatrix: Class II Standard Analysis



iVAX A safer, faster approach to vaccine design. [Home](#) | [Contact](#)

[About](#) [CCHI Collaborators](#) [iVAX Tool Kit](#) [News](#)

[Home](#) [Data Management](#) [Conservation Analysis](#) [Class I Analysis](#) [Class II Analysis](#) [BLAST Analysis](#) [Vaccine Design](#) [Ad Hoc Analysis](#) [Tutorial](#)

Class II EpiMatrix Analysis

[Execute a Standard EpiMatrix Class II Analysis](#)
Use this Link to Execute a Standard EpiMatrix Class II Analysis.

[Execute an Extended EpiMatrix Class II Analysis](#)
Use this Link to Execute an Extended EpiMatrix Class II Analysis.

[Class II Protein Summary Report](#)
Use this Link to Display the Protein Summary Report.

[Class II Protein Immunogenicity Scale](#)
Use this Link to Display the Protein Immunogenicity Scale.

[Class II Protein Immunogenicity Report](#)
Use this Link to Display the Results of EpiMatrix Analysis.

Class II Cluster Analysis

[Find Class II T Cell Epitope Clusters](#) ←
Use this Link to Screen Protein Sequences for Class II T Cell Epitope Clusters.


[Interactive Class II Cluster Immunogenicity Report](#)
Use this Link to Display the Results of an EpiMatrix Class II Cluster Analysis.

[Printable Class II Cluster Immunogenicity Report](#)
Use this Link to Print or Download an EpiMatrix Class II Cluster Analysis.

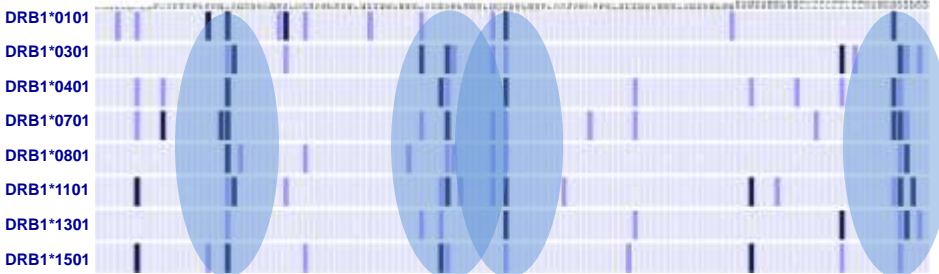
Confidential 45

Screening for T cell Epitopes

Epitopes Tend to Cluster



- T cell epitopes are not randomly distributed throughout protein sequences but instead tend to **cluster** in specific regions.
 - These clusters can be very powerful. One or more dominant T-cell epitope clusters can enable significant immune responses to even otherwise low scoring proteins.
- ClustiMer is used to identify T-cell epitope clusters. It identifies polypeptides predicted to bind to an unusually large number of HLA alleles.




- T-cell epitope clusters make excellent vaccine candidates:
 - compact; relatively easy to deliver as peptides; highly reactive in-vivo

46

Screening for T cell Epitopes

Finding Class II Clusters



Class II Cluster Analysis

[Find Class II T Cell Epitope Clusters](#) ←
Use this Link to Screen Protein Sequences for Class II T Cell Epitope Clusters

[Interactive Class II Cluster Immunogenicity Report](#)
Use this Link to Display the Results of an EpiMatrix Class II Cluster Analysis

[Printable Class II Cluster Immunogenicity Report](#)
Use this Link to Print or Download an EpiMatrix Class II Cluster Analysis

Success! 12 Clusters Found

Find ClassII T Cell Epitope Clusters


Choose a File:	FLU-HA
Choose a Protein or Null for All Proteins:	
<input type="button" value="Execute"/>	<div style="border: 1px solid gray; padding: 2px;"> BOSTON-2025 CHICAGO-2013 NOME-2050 PROVIDENCE-2012 </div>

Confidential

47

Screening for T cell Epitopes

Finding Class II Clusters



EpiMatrix ClassII Interactive Immunogenicity Report


Choose a File:	FLU-HA
Choose a Protein or Null for All Proteins:	PROVIDENCE-2012
Summarize EpiBars?	<input checked="" type="radio"/> True <input type="radio"/> False
<input type="button" value="Execute"/>	

Confidential

48

Screening for T cell Epitopes

Interactive Class II Cluster Report



Overview of Cluster Selections

Current File: FLU-HA - Current Sequence: PROVIDENCE-2012
September 29, 2010 (Epx Ver. 1.2)

[Click to Print](#) [Click to Download](#)


File	Input Sequence	Cluster Address (w/ FLANKS)	Cluster Sequence	Hydrophobicity	EpiMatrix HITS (w/o FLANKS)	EpiMatrix CLUSTER SCORE (w/o FLANKS)	EpiBars? (#)
FLU-HA	PROVIDENCE-2012	22-37	SDTAVKTTNRGLFTV	-0.30	8	14.02	Yes (1)
FLU-HA	PROVIDENCE-2012	31-54	IRGLVYVNAISLVGSSSTGRIQRI	-0.65	17	27.6	Yes (3)
FLU-HA	PROVIDENCE-2012	61-78	SENGTSLRALLQRPFCSE	-0.29	8	13.35	Yes (1)
FLU-HA	PROVIDENCE-2012	76-99	CRSFQNKKRLVVKRSKAYSNCY	-1.11	16	31.98	Yes (1)
FLU-HA	PROVIDENCE-2012	89-123	EIVYFRVLAISLISVASSDTLQVSE	-0.17	25	43.69	Yes (3)
FLU-HA	PROVIDENCE-2012	116-132	SDTLEP INARPNQVTAQ	-0.38	7	10.02	No
FLU-HA	PROVIDENCE-2012	141-164	RSQVNSFFARGLMILHRSKRSKIPAL	-0.69	19	30.03	Yes (3)
FLU-HA	PROVIDENCE-2012	185-207	RSTRNAGTSLTYRASQVYVYVSE	-0.91	12	17.06	Yes (1)
FLU-HA	PROVIDENCE-2012	219-234	SDPWRQQRISLITM	-0.78	12	24.87	Yes (2)
FLU-HA	PROVIDENCE-2012	231-256	IKYPTLYVSTFQELLINSIQRLAIVR	0.60	32	57.72	Yes (5)
FLU-HA	PROVIDENCE-2012	254-278	PRDPTLIRKGGSSIRRSAPLQDNR	-0.66	30	65.01	Yes (4)
FLU-HA	PROVIDENCE-2012	305-325	CRSFVYQHTLILALATGRPRKTFE	-0.73	14	28.24	Yes (2)

T

Confidential
49

Screening for T cell Epitopes

Class II Cluster Detail Report



EpiMatrix Cluster Detail Report


File: FLU-HA Sequence: PROVIDENCE-2012 Cluster: 76
September 29, 2010 (Epx Ver. 1.2)

[Click to Print](#) [Click to Download](#) [Back to Cluster Summary](#)

Frame Start	AA Sequence	Frame Stop	Hydrophobicity	DRB1*0101	DRB1*0301	DRB1*0401	DRB1*0701	DRB1*0801	DRB1*1101	DRB1*1301	DRB1*1501	Hits
				Z-Score	Z-Score	Z-Score	Z-Score	Z-Score	Z-Score	Z-Score		
76	CRSFQNKKW	84	-0.37				1.49					0
77	RSFQNKKWR	85	-0.54							1.79		1
78	SFQNKKWRL	86	-0.34					1.4				0
79	FQNKKWRLF	87	-1.2		1.48		2.39	1.89	1.84	1.44		3
80	QNKKWRLFV	88	-1.04									0
81	NKKWRLFVK	89	-1.09									0
82	KKWRLFVKK	90	-1.2									0
83	KWRLFVKRS	91	-0.86	1.39		1.39			1.44			0
84	WRLFVKRSK	92	-0.86	1.41		1.57		2.77	3.29	1.42	1.44	2
85	RLFVKRSKA	93	-0.56		1.69				1.4	1.77	1.52	2
86	LFVKRSKAY	94	-0.2		1.28					1.92		1
87	FVKRSKAYS	95	-0.71	1.46	1.45			2.73	2.72	2.7	1.4	3
88	VKRSKAYSN	96	-1.41	2.17	1.8	2.02	2.31	1.86	2.02	1.3	2.3	7
89	KRSKAYSNC	97	-0.34									0
90	RSKAYSNCY	98	-0.28		1.84						1.31	1
91	SKAYSNCYP	99	-0.21									0
Summarized Results				DRB1*0101	DRB1*0301	DRB1*0401	DRB1*0701	DRB1*0801	DRB1*1101	DRB1*1301	DRB1*1501	Total
Maximum Single Z score				2.17	1.84	2.02	2.39	2.77	3.29	2.7	2.3	--
Sum of Significant Z scores				2.17	5.34	2.02	4.7	9.25	9.88	8.18	2.3	43.84
Count of Significant Z Scores				1	3	1	2	4	4	4	1	20
Total Assessments Performed: 128				Hydrophobicity: -1.11			EpiMatrix Score: 30.66			EpiMatrix Score (w/o flanks): 31.98		

Screening for T cell Epitopes

Interactive Class II Cluster Report



Overview of Cluster Selections

Current File: FLU-HA - Current Sequence: PROVIDENCE-2012
September 29, 2010 (Epx Ver. 1.2)

[Click to Print](#) [Click to Download](#)


File	Input Sequence	Cluster Address (w/ FLANKS)	Cluster Sequence	Hydrophobicity	EpiMatrix HITS (w/ FLANKS)	EpiMatrix CLUSTER SCORE (w/ FLANKS)	EpiBars? (#)
FLU-HA	PROVIDENCE-2012	22 - 37	SDTAVKTTNRGLFVV	-0.30	8	14.02	Yes (1)
FLU-HA	PROVIDENCE-2012	31 - 54	IRGLVYVNAISLVGSSSTGRIQRI	-0.65	17	27.6	Yes (3)
FLU-HA	PROVIDENCE-2012	61 - 78	SENGTSLRALLGRPHQCE	-0.29	8	13.35	Yes (1)
FLU-HA	PROVIDENCE-2012	76 - 99	CESTFQKRRRLPVKRSKATANCFE	-1.11	16	31.98	Yes (1)
FLU-HA	PROVIDENCE-2012	89 - 123	EYVYPRASLRSIVASQTLACIHE	-0.17	25	43.69	Yes (3)
FLU-HA	PROVIDENCE-2012	116 - 132	SDTLPINRPNPTQTGQ	-0.38	7	10.02	No
FLU-HA	PROVIDENCE-2012	141 - 164	EGGVNSFFARGLMLHRSKRIKIPAL	-0.69	19	30.03	Yes (3)
FLU-HA	PROVIDENCE-2012	185 - 207	RSTRNGTSLTYASQRYTYSCE	-0.91	12	17.06	Yes (1)
FLU-HA	PROVIDENCE-2012	219 - 234	SDPWQGRISLITD	-0.78	12	24.87	Yes (2)
FLU-HA	PROVIDENCE-2012	231 - 256	ELYPTIVTSFRIILLINSIQRLAIFD	0.60	32	57.72	Yes (5)
FLU-HA	PROVIDENCE-2012	254 - 278	PRDPTFLKRRGSSIRSRAPLQDS	-0.66	30	65.01	Yes (4)
FLU-HA	PROVIDENCE-2012	305 - 325	CEETVYGHILLALZGRKVFEE	-0.73	14	28.24	Yes (2)

↑

Confidential 51

Screening for T cell Epitopes

Class II Cluster Scale



EpiMatrix Cluster Immunogenicity Report

File: FLU-HA
September 29, 2010 (Epx Ver. 1.2)


[Click to Print](#) [Click to Download](#)

Target	Score
Tetrax Toxin (825-850) ->	40
Hepatitis C NS3 (1248-1267) ->	30
Influenza HA (307-319) ->	20
Southwood-1998 (1188, 16) ->	10
Human CLIP ->	10
Southwood-1998 (1188, 32) ->	10
Minimum 20-Mer Peptide ->	-10

← PROVIDENCE-2012:76,(31-98)

Screening for T cell Epitopes

Interactive Class II Cluster Report



Overview of Cluster Selections

Current File: FLU-HA - Current Sequence: PROVIDENCE-2012
September 29, 2010 (Epx Ver. 1.2)

[Click to Print](#) [Click to Download](#)

File	Input Sequence	Cluster Address (w/ FLANKS)	Cluster Sequence	Hydrophobicity	EpiMatrix HITS (w/ FLANKS)	EpiMatrix CLUSTER SCORE (w/ FLANKS)	EpiBars? (#)
FLU-HA	PROVIDENCE-2012	22 - 37	SDTAVKTTNRGLFVV	-0.30	8	14.02	Yes (1)
FLU-HA	PROVIDENCE-2012	31 - 54	IRGLVSTNAILVGLSSSTGRIQRI	-0.65	17	27.6	Yes (3)
FLU-HA	PROVIDENCE-2012	61 - 78	SENGTSLRALLGPHQCE	-0.29	8	13.35	Yes (1)
FLU-HA	PROVIDENCE-2012	76 - 99	CESTQNKRRLLVVRKRAYANCFE	-1.11	16	31.98	Yes (1)
FLU-HA	PROVIDENCE-2012	89 - 123	EYVYPRASLRSIVASSTLACVSE	-0.17	25	43.69	Yes (3)
FLU-HA	PROVIDENCE-2012	116 - 132	SDTLPINRPNPTVTAQ	-0.38	7	10.02	No
FLU-HA	PROVIDENCE-2012	141 - 164	EGVNSFFARGLMLHRSKRIKPAI	-0.69	19	30.03	Yes (3)
FLU-HA	PROVIDENCE-2012	185 - 207	RSTRNGTSLTYRASQVYTSYSE	-0.91	12	17.06	Yes (1)
FLU-HA	PROVIDENCE-2012	219 - 234	SDPWVQGRISITD	-0.78	12	24.87	Yes (2)
FLU-HA	PROVIDENCE-2012	231 - 256	ELVYFVSTFRIILLINSIQRLAFLD	0.60	32	57.72	Yes (5)
FLU-HA	PROVIDENCE-2012	254 - 278	PRDPTLIRKGGSSIRSRAPLQDS	-0.66	30	65.01	Yes (4)
FLU-HA	PROVIDENCE-2012	305 - 325	CEETVYGHILLALATGRKRTFE	-0.73	14	28.24	Yes (2)


↑

Confidential

53

Screening for T cell Epitopes

Class II Logo Report



EpiMatrix Peptide Logo Report


Individual Amino Acid Contribution to Binding Averaged Across Alleles and Frames
(font size/color is keyed to average score)

File: FLU-HA - Sequence: PROVIDENCE-2012 - Cluster: 76
September 29, 2010 (Epx Ver. 1.2)

[Click to Print](#) [Click to Download](#) [Back to Summary Report](#)

76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
C	R	S	F	Q	N	KK	W	R	L	F	V	K	R	S	K	A	Y	S	N	C	Y	P	
0	4.72	2.48	15.79	6.25	9.68	17.09	19.25	10.01	18.41	30.7	17.46	11.01	19.27	18.41	16.19	19.27	8.78	4.4	11.82	9.08	43	1.9	0

Overview



- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

Confidential 55

Searching for Homology

BLAST at NCBI





The screenshot shows the NCBI homepage with a search bar at the top. The left sidebar contains a 'Resources' menu with categories like 'All Resources (A-Z)', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Maps', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Small Molecules', 'Taxonomy', 'Training & Tutorials', and 'Variation'. The main content area features a 'Welcome to NCBI' message, a 'Genome' banner announcing 1000 prokaryotic genomes, and a 'How To...' section with links for determining conserved synteny, finding homologs, coten articles, and designing PCR primers. The right sidebar lists 'Popular Resources' including BLAST, Bookshelf, Gene, Genome, Nucleotide, OMIM, Protein, PubChem, Pubmed, and SDRP, along with 'NCBI News' updates.

Confidential 56

Searching for Homology

BlastiMer



Success! 12 Clusters Submitted and 12 Clusters Returned.

EpiMatrix Class II Cluster Blast

Choose a File:

Choose a Protein or Null for All Proteins:

Choose a Database to Search: Genbank Non-Redundant Proteins Genbank Patent Database

Search Human Sequences Only? Search Entire Database Limit Blast to Human Sequences


Overwrite Existing Blast Data

Confidential

59

Searching for Homology

Interactive BLAST Report



Interactive Class II Cluster BLAST Report

Choose a File:

Choose a Protein or Null for All Proteins:

Overview of Cluster Blast Results


Current File: FLU-HA - Current Sequence: PROVIDENCE-2012

[Click to Print](#), [October 15, 2010](#), [Click to Download](#)

File	Input Sequence	Cluster Address (w/ FLANKS)	Cluster Sequence	EpiMatrix CLUSTER SCORE (w/o FLANKS)	Link to Alignment Report	Link to Summary Report
FLU-HA	PROVIDENCE-2012	22 - 37	NGLLVKTTINRQIKVT	14.02	Alignment	Summary
FLU-HA	PROVIDENCE-2012	31 - 54	RSQIKVTNATHKLVQSSSTORICRS	27.8	Alignment	Summary
FLU-HA	PROVIDENCE-2012	61 - 78	GKNCETLIKALLGRPCRS	13.35	Alignment	Summary
FLU-HA	PROVIDENCE-2012	76 - 96	CRSPQNKWRLPVKRSKAYSNCYP	31.98	Alignment	Summary
FLU-HA	PROVIDENCE-2012	99 - 123	PTVFPRIASLSLIVASSGTLAFINK	43.69	Alignment	Summary
FLU-HA	PROVIDENCE-2012	115 - 132	SGTLKPFINKRPHWTOVAG	10.02	Alignment	Summary
FLU-HA	PROVIDENCE-2012	141 - 164	RGSVNSFFSRLLNLHNSKIKYFAL	30.03	Alignment	Summary
FLU-HA	PROVIDENCE-2012	185 - 207	PSTRRRQTSLYVRASORVTVETK	17.05	Alignment	Summary
FLU-HA	PROVIDENCE-2012	219 - 234	SRPMVROQSSRSISYH	24.97	Alignment	Summary
FLU-HA	PROVIDENCE-2012	231 - 256	SIYWTIVKPGRIILLINSIGNLIAPRG	57.72	Alignment	Summary
FLU-HA	PROVIDENCE-2012	254 - 278	PROYFKIHNOKSSIHNSRAPIQTCR	55.01	Alignment	Summary
FLU-HA	PROVIDENCE-2012	305 - 325	CFRTYKQETLALGSRVYFK	28.24	Alignment	Summary

Searching for Homology

BLAST Alignment Report



Blast Alignment Report for FLU-HA_PROVIDENCE-2012-115

Analysis of Homologous Sequences

October 15, 2010

[Click to Print](#) [Click to Download](#) [Back to Cluster Summary](#)


# BLAST HITS	SEQUENCE	AA IDENTITIES	FILE	DESCRIPTION	ORGANISM
--	SUTLQYLRKRFNHTQWQ	--	FLU-HA	--	n/a
1	-----SD-----	16/18	ABQ1244	hemagglutinin	[Influenza A virus (A/New York/757/1995)(H3N2)]
1	-----TQ-S-----	15/18	ACV42084	hemagglutinin	[Influenza A virus (A/swine/Kansas/015252/2009)(H3N2)]
1	-----SD-----T-	15/18	ABM9172	hemagglutinin	[Influenza A virus (A/swine/Italy/630/1987)(H3N2)]
22	-----E--SA-----	15/18	ABO10167	hemagglutinin	[Influenza A virus (A/England/121/1995)(H3N2)]
1	-----SD-----T-	15/18	CAC40046	haemagglutinin	[Influenza A virus (A/swine/Italy/636/87)(H3N2)]
1	-----E--SA-----	15/18	AAT64846	hemagglutinin	[Influenza A virus (A/Lyon/22686/93)(H3N2)]
5	-----E--SD-----	15/18	AAK67172	hemagglutinin	[Influenza A virus (A/Seoul/16/89)(H3N2)]
64	-----E--SD-----	15/18	ACU12494	hemagglutinin	[Influenza A virus (A/Siena/3/1991)(H3N2)]
4	-----E--ES-----T-	14/18	ACH95746	hemagglutinin	[Influenza A virus (A/Taiwan/VGHYM0904-08/1980)(H3N2)]
Total:					
					100

Confidential

61

Searching for Homology

BLAST Summary Report



Cluster Blast Summary Report for FLU-HA_PROVIDENCE-2012-115

Analysis of Homologous Sequences

October 15, 2010


[Click to Print](#) [Click to Download](#) [Back to Cluster Summary](#)

115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132
S	G	T	L	K	F	I	N	K	R	F	N	W	T	G	V	A	Q
100/100	100/100	100/100	100/100	96/100	100/100	99/100	95/100	99/100	86/100	100/100	100/100	100/100	100/100	100/100	100/100	94/100	100/100
				K 4/100		T 1/100	Q 3/100	K 1/100	N 22/100							T 6/100	
									G 5/100								
									S 5/100								
									A 1/100								

Confidential

62


Overview



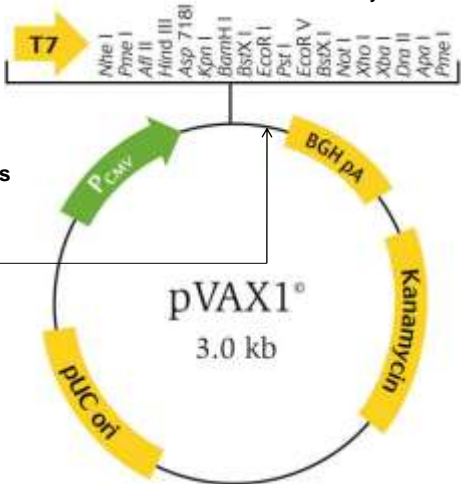
- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

Confidential 63

Vaccine Design Concepts



- Immunogenic peptides can be formulated and delivered in various ways
 - as peptides
 - as pseudo-proteins
 - in formulation with micro-spheres
 - linked to carrier proteins
 - embedded into viral genomes.
 - **Plasmid DNA based formulations**




Epitopes in a single contiguous sequence that is back translated to DNA is inserted into a carrier

Need to construct

Confidential 64

Overview




- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- **Vaccine Design**
 - Concepts
 - Immunogenic Consensus
 - String of beads

Confidential 65

Vaccine Design

Immunogenic Consensus - EpiAssembler




- EpiAssembler
- Start with a core, immunogenic 9mer conserved in multiple strains
- Identify overlapping immunogenic 9mers
- Generate an ICS using pieces of sequences from each of the strains
- ICS may be distinct from any one strain of the pathogen, but contains immunogenic elements from each strain, making it cross-reactive with many strains

66

Vaccine Design


EpiAssembler – Raw Data



STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L

Vaccine Design


EpiAssembler – Core Epitope



STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L

Vaccine Design

EpiAssembler – Flanking Epitopes




STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L

W P K V E Q F W A K H M W N F
W P K V E Q F W A

Vaccine Design

EpiAssembler – Flanking Epitopes



STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L

W P K V E Q F W A K H M W N F
Q A S W P K V E Q W P K V E Q F W A

Vaccine Design

EpiAssembler – Flanking Epitopes

STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L

	W	P	K	V	E	Q	F	W	A	M	W	N	F	I	S	G	I	Q		
Q	A	S	W	P	K	V	E	Q	F	W	A	M	W	N	F	I	S	G	I	Q

Vaccine Design


EpiAssembler – Flanking Epitopes

STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L

	W	P	K	V	E	Q	F	W	A	M	W	N	F	I	S	G	I	Q		
Q	A	S	W	P	K	V	E	Q	F	W	A	M	W	N	F	I	S	G	I	Q
	N	F	I	S	G	I	Q	Y	L											


Vaccine Design

EpiAssembler – Final Immunogenic Consensus Sequence



STRAIN 01	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	X	W	N	X	I	S	X	I	Q	Y	L													
STRAIN 02	Q	A	S	W	P	K	V	E	X	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L													
STRAIN 03	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X													
STRAIN 04	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	Q	Y	L													
STRAIN 05	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L													
STRAIN 06	Q	A	S	W	P	K	X	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X													
STRAIN 07	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	X	N	F	I	S	G	I	Q	Y	L													
STRAIN 08	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L													
STRAIN 09	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	F	X	S	X	I	X	Y	X													
STRAIN 10	Q	A	S	W	P	X	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L													
STRAIN 11	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L													
STRAIN 12	Q	A	S	W	X	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	X													
STRAIN 13	Q	A	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L													
STRAIN 14	Q	A	S	W	X	K	X	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L													
STRAIN 15	Q	A	S	W	P	K	V	E	X	F	W	X	K	H	M	W	N	F	I	S	G	I	Q	Y	L													
STRAIN 16	Q	X	S	W	P	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	X	G	I	Q	Y	L													
STRAIN 17	X	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	X													
STRAIN 18	Q	X	S	W	P	K	X	E	Q	F	W	A	K	H	M	W	N	X	I	S	G	I	Q	Y	L													
STRAIN 19	Q	A	S	W	X	K	V	E	Q	F	W	A	K	H	M	W	N	F	I	S	X	I	Q	Y	L													
STRAIN 20	Q	A	S	W	P	K	V	E	Q	F	W	A	X	H	M	W	N	F	I	S	G	I	Q	Y	L													
										F	W	A	K	H	M	W	N	F																				
										W	P	K	V	E	Q	F	W	A					M	W	N	F	I	S	G	I	Q							
										Q	A	S	W	P	K	V	E	Q					N	F	I	S	G	I	Q	Y	L							
										Q	A	S	W	P	K	V	E	Q					F	W	A	K	H	M	W	N	F	I	S	G	I	Q	Y	L

Overview



- Gathering raw data
 - Popular sources
 - Data Management
- Conservation Analysis
 - Multiple Alignments / Consensus Sequences
 - Conservatrix
- Screening for putative T cell epitopes (EpiMatrix)
 - MHC Class I and II
- Searching for homology
 - Human genome / Other pathogens (GenBank)
- Vaccine Design
 - Concepts
 - Immunogenic Consensus
 - String of beads

Confidential 74

VaxCAD

Overview



- VaxCAD – computer assisted vaccine design
- VaxCAD algorithm can be used to create synthetic antigens from T cell epitopes, epitope clusters, and ICSs

SYFPEITHI SLYNTVATL ILKEPVHGV



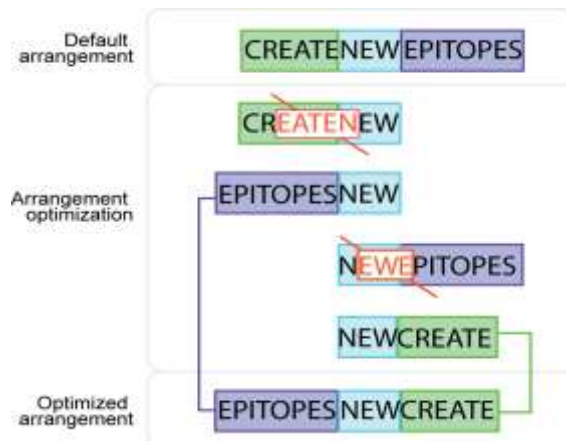
SYFPEITHISLYNTVATLILKEPVHGV
"String-of-Beads"

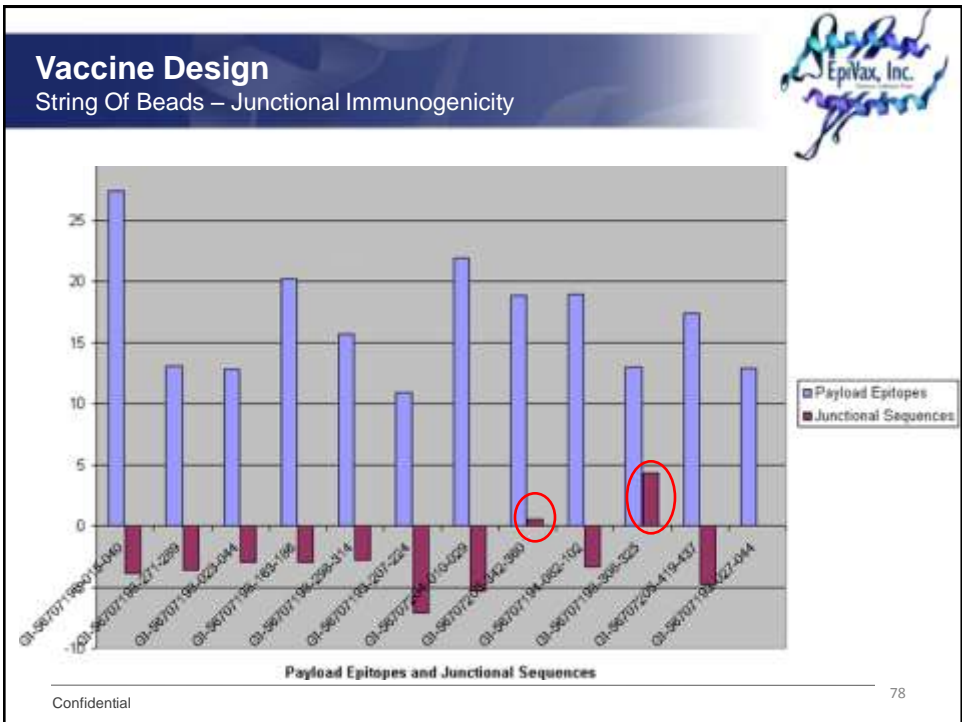
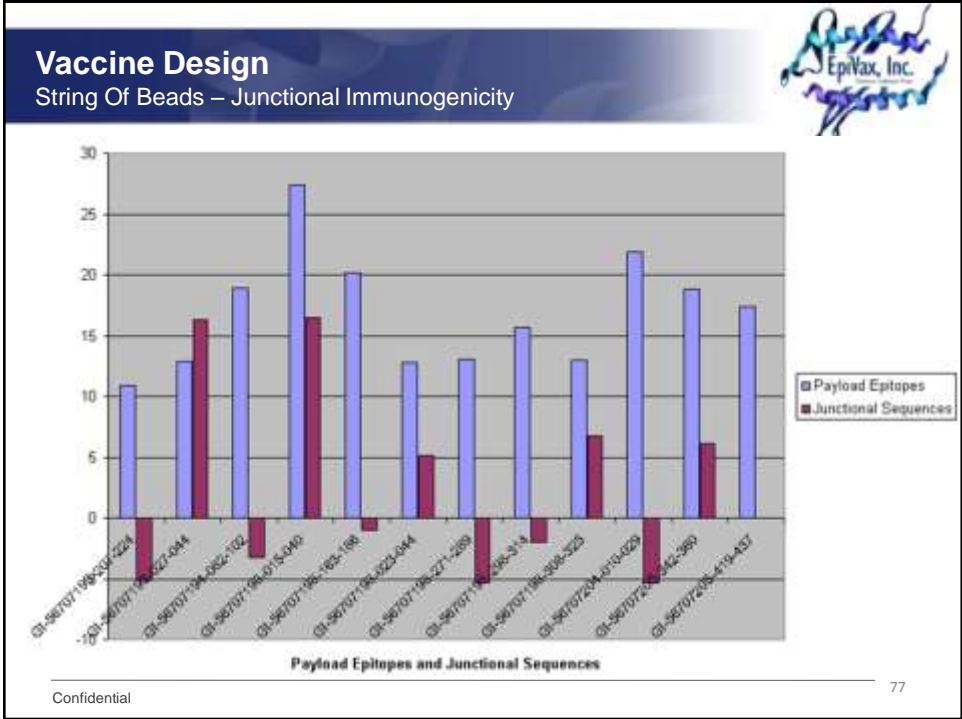
VaxCAD

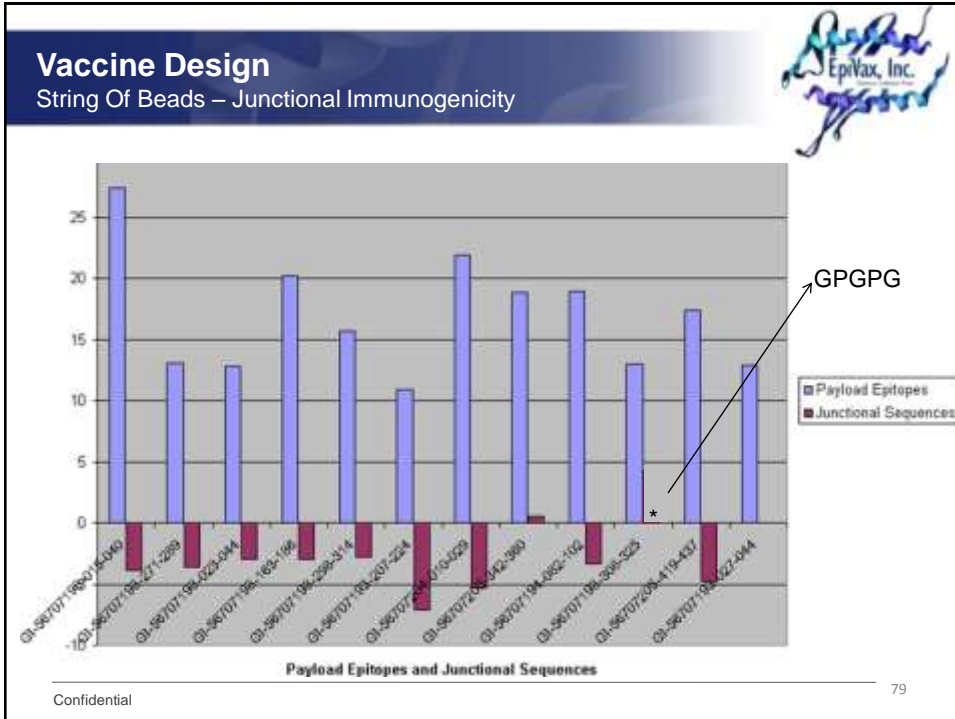
Junctional Epitopes



VaxCAD will identify junctional epitopes and rearrange chosen epitopes to reduce junctional epitope formation







Vaccine Design

VaxCad

iVAX A safer, faster approach to vaccine design.

Home | Contact

About CCHI Collaborators iVAX Tool Kit News

Home Data Management Conservation Analysis Class I Analysis Class II Analysis BLAST Analysis Vaccine Design Ad Hoc Analysis Tutorial


Vaccine Design

[VaxCad](#)
Use this Link to Design a DNA construct w/ VaxCad

Confidential 80

Vaccine Design

VaxCad



Execute Vaccine CAD to Reorder an Accession


Choose a File:	FLU-HA		
Choose a Protein:	PROVIDENCE-2012		
Research Program:	FLU-HA-VAX		
Junction Threshold:	0		-999 to Force Spacers
Class:	<input type="radio"/> Class1		<input checked="" type="radio"/> Class2
Start w/Spacer:	<input checked="" type="radio"/> NO		<input type="radio"/> YES
<input type="button" value="Execute"/>			

Confidential

81

Vaccine Design

VaxCad



Original Default Sequence

Peptide 1 ID	Peptide 1	Peptide 1 Score	Peptide 2 ID	Peptide 2	Peptide 2 Score	Junction Score
PROVIDENCE-2012	SIYWTIVLPGRILLNSIGNLIAPRG	83.43179701	PROVIDENCE-2012	PRGYFKIINGKSSIMRSRAPIGTCS	48.84200133	-7.11938
PROVIDENCE-2012	PRGVFKKINGKSSIMRSRAPIGTCS	48.94200133	PROVIDENCE-2012	PYKVPRYASLRSLVASSGTLKFINK	37.62859752	-7.11938
PROVIDENCE-2012	PYRVPRYASLRSLVASSGTLKFINK	37.62859752	PROVIDENCE-2012	CRSPQNKWRLPVKRSKAYSNCYP	29.60581088	12.73755
PROVIDENCE-2012	CRSPQNKWRLPVKRSKAYSNCYP	29.60581088	PROVIDENCE-2012	RGSVNSPFSRLNWLHRSKYKYPAL	29.38884758	-5.09290289
PROVIDENCE-2012	RGSVNSPFSRLNWLHRSKYKYPAL	29.38884758	PROVIDENCE-2012	CPRYVKQNTLKLATGMRNVPK	22.44223478	-1.73965406
PROVIDENCE-2012	CPRYVKQNTLKLATGMRNVPK	22.44223478	PROVIDENCE-2012	NRQKVTNATKLVQSSSTGRICRS	19.98416923	1.37201629
PROVIDENCE-2012	NRQKVTNATKLVQSSSTGRICRS	19.98416923	PROVIDENCE-2012	SKPWVKGQSSRSIYW	19.38889395	-5.44032434
PROVIDENCE-2012	SKPWVKGQSSRSIYW	19.38889395	PROVIDENCE-2012	PSTRRRQTSLYVKAAGRVTYSTK	8.47348192	-2.54883926
PROVIDENCE-2012	PSTRRRQTSLYVKAAGRVTYSTK	8.47348192	PROVIDENCE-2012	GKNCILIRALLGRPHCRS	7.74972749	-7.11938
PROVIDENCE-2012	GKNCILIRALLGRPHCRS	7.74972749	PROVIDENCE-2012	NGTLVKITINRQIKVT	6.90337832	-5.47324114
PROVIDENCE-2012	NGTLVKITINRQIKVT	6.90337832	PROVIDENCE-2012	SGTLKFINRNFNWTEGVAQ	4.4145833	-5.07349621
PROVIDENCE-2012	SGTLKFINRNFNWTEGVAQ	4.4145833	PROVIDENCE-2012	SIYWTIVLPGRILLNSIGNLIAPRG	83.43179701	-5.33664214
					TOTAL:	-35.01941155

Confidential

82

Vaccine Design

VaxCad

Reordered Sequence

Peptide 1 ID	Peptide 1	Peptide 1 Score	Peptide 2 ID	Peptide 2	Peptide 2 Score	Junction Score
PROVIDENCE-2012	SRPWVYRGQSSRSITW	19.39669995	PROVIDENCE-2012	PRGYFKRNGKSSIMSRAPIGTCS	48.94200133	-7.11938
PROVIDENCE-2012	PRGYFKRNGKSSIMSRAPIGTCS	48.94200133	PROVIDENCE-2012	SGTLKFKRNFNWTVGAQ	4.4145633	-7.11938
PROVIDENCE-2012	SGTLKFKRNFNWTVGAQ	4.4145633	PROVIDENCE-2012	CPRVYKQNTLKLATGMKNVFK	22.44223478	-7.11938
PROVIDENCE-2012	CPRVYKQNTLKLATGMKNVFK	22.44223478	PROVIDENCE-2012	RGSVNSFFSRNLNWLHKSKEYPAL	29.38084756	-2.83745573
PROVIDENCE-2012	RGSVNSFFSRNLNWLHKSKEYPAL	29.38084756	PROVIDENCE-2012	CRSFQNKRWLFVRSKAYSNCYP	29.60561086	-3.11041027
PROVIDENCE-2012	CRSFQNKRWLFVRSKAYSNCYP	29.60561086	PROVIDENCE-2012	PYRVPRYASRLVASSGTLKFINK	37.62859752	-4.78406996
PROVIDENCE-2012	PYRVPRYASRLVASSGTLKFINK	37.62859752	PROVIDENCE-2012	SIYWTIVKPGRELLNSIGNLIAPRG	63.43179701	6.09766441 ←
PROVIDENCE-2012	SIYWTIVKPGRELLNSIGNLIAPRG	63.43179701	PROVIDENCE-2012	PSTRRRQTSLYRASGRVTVSTK	9.47340192	-5.42113169
PROVIDENCE-2012	PSTRRRQTSLYRASGRVTVSTK	9.47340192	PROVIDENCE-2012	GKNCTLRALLGRPHCRS	7.74072740	-7.11938
PROVIDENCE-2012	GKNCTLRALLGRPHCRS	7.74072740	PROVIDENCE-2012	NGTLVKTTINRQIKVT	6.90337832	-5.47324114
PROVIDENCE-2012	NGTLVKTTINRQIKVT	6.90337832	PROVIDENCE-2012	NRQKIVTNATLKVSSSTGRICRS	19.96416923	-5.2805373
PROVIDENCE-2012	NRQKIVTNATLKVSSSTGRICRS	19.96416923	PROVIDENCE-2012	SRPWVYRGQSSRSITW	19.39669995	-5.44032434
					TOTAL:	-51.52694602

Confidential
83

Vaccine Design

VaxCad


Reordered Sequence with Spacers

Peptide 1 ID	Peptide 1	Peptide 1 Score	Spacer	Peptide 2 ID	Peptide 2	Peptide 2 Score	Junctional Score
PROVIDENCE-2012	SRPWVYRGQSSRSITW	19.39669995		PROVIDENCE-2012	PRGYFKRNGKSSIMSRAPIGTCS	48.94200133	-7.11938
PROVIDENCE-2012	PRGYFKRNGKSSIMSRAPIGTCS	48.94200133		PROVIDENCE-2012	SGTLKFKRNFNWTVGAQ	4.4145633	-7.11938
PROVIDENCE-2012	SGTLKFKRNFNWTVGAQ	4.4145633		PROVIDENCE-2012	CPRVYKQNTLKLATGMKNVFK	22.44223478	-7.11938
PROVIDENCE-2012	CPRVYKQNTLKLATGMKNVFK	22.44223478		PROVIDENCE-2012	RGSVNSFFSRNLNWLHKSKEYPAL	29.38084756	-2.83745573
PROVIDENCE-2012	RGSVNSFFSRNLNWLHKSKEYPAL	29.38084756		PROVIDENCE-2012	CRSFQNKRWLFVRSKAYSNCYP	29.60561086	-3.11041027
PROVIDENCE-2012	CRSFQNKRWLFVRSKAYSNCYP	29.60561086		PROVIDENCE-2012	PYRVPRYASRLVASSGTLKFINK	37.62859752	-4.78406996
PROVIDENCE-2012	PYRVPRYASRLVASSGTLKFINK	37.62859752	GPGPG	PROVIDENCE-2012	SIYWTIVKPGRELLNSIGNLIAPRG	63.43179701	164.12317 ←
PROVIDENCE-2012	SIYWTIVKPGRELLNSIGNLIAPRG	63.43179701		PROVIDENCE-2012	PSTRRRQTSLYRASGRVTVSTK	9.47340192	-5.42113169
PROVIDENCE-2012	PSTRRRQTSLYRASGRVTVSTK	9.47340192		PROVIDENCE-2012	GKNCTLRALLGRPHCRS	7.74072740	-7.11938
PROVIDENCE-2012	GKNCTLRALLGRPHCRS	7.74072740		PROVIDENCE-2012	NGTLVKTTINRQIKVT	6.90337832	-5.47324114
PROVIDENCE-2012	NGTLVKTTINRQIKVT	6.90337832		PROVIDENCE-2012	NRQKIVTNATLKVSSSTGRICRS	19.96416923	-5.2805373
PROVIDENCE-2012	NRQKIVTNATLKVSSSTGRICRS	19.96416923		PROVIDENCE-2012	SRPWVYRGQSSRSITW	19.39669995	-5.44032434
					TOTAL:	-60.86848726	

Confidential
84

Vaccine Design

VaxCad– Resulting Construct



Optimized w/Spacer Sequence

```

>FLU-HA-VAX_FLU-HA
SRPWVRGQSSRISIWPRGYFKIRNGKSSIMRSRAPIGTCSSGTLKFINKRFNWTGVA
QCPRYVKQNTLKLATGMRNVPKRGSVNSFFSRLNWLHKSXYKYPALCRSFQNKWRLE
VKRSKAYSNCYPPYRVPRYASLRSLVASSGTLKFINKGPGPGSIYWTIVKPGRILLIN
SIGNLIAPRGPSTRRRQTSLYVRASGRVTVSTKGKNCTLIRALLGRPHCRSNGTLVKT
ITNRQIKVTNRQIKVTNATKLVQSSSTGRICRS


```

- Once you have completed the design of your string of beads it is a good idea to reanalyze it with EpiMatrix just to confirm that the potential for junctional immunogenicity has been adequately addressed.
- Run through the BlastMer to look for newly created homologies, particularly to the human genome.



Confidential
85

Vaccine Design

Reverse Translation and Optimization



- There are many products available to accomplish the task of reverse translation and codon optimization.

- Most gene synthesis companies have methods of codon optimization that they feel enhance the quality of their products

IMPORTANT

- Codon optimization
 - Be sure to optimize for the system you will be testing in (human, mouse, etc)
- Restriction Sites
 - Be sure to map the restriction sites in your construct.
 - Make sure your construct does not include any copies of the restriction sites you will be using to digest your plasmid

http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/frequency_of_restriction_sites.asp

Confidential
86

Vaccine Design
Enhancing in-vivo performance

Protein
EpiVax, Inc.
supported

Add a Kozak sequence to enhance initiation of translation

ACC ATG G

Consensus

-6 -5 -4 -3 -2 -1 +1 +2 +3 +4

Acc AUGG

Each letter is written in proportion to its biased use in a sequence logo. If a base were used in all ~25,000 human mRNAs it would score 2 and be written two units tall. The most significantly biased bases are -3 (A) and +4 (G). The Initiation codon AUG is not drawn to scale (it would score 2 bits) of information on this scale.

Confidential 87

Vaccine Design
Enhancing in-vivo performance


Protein
EpiVax, Inc.
supported

Target CD8+ T cell epitopes to the Class I degradation pathway with Ubiquitin

Ubiquitin monomer

Confidential 88

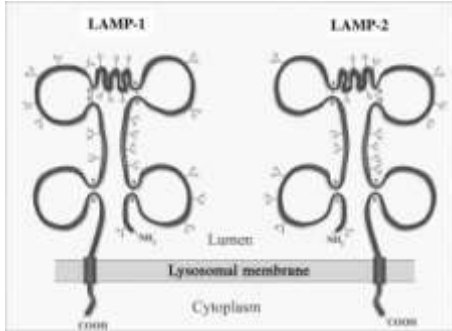
Vaccine Design
Enhancing in-vivo performance



Target CD4+ T cell epitopes the Class II secretory pathway

Secretion targeting sequence: (preferred strategy)
Tissue Plasminogen Activator Signal Sequence: MQMSPALTCLVLGLALVFGECSA


MIIC compartment targeting: LAMP (Lysosome-associated membrane protein-1)



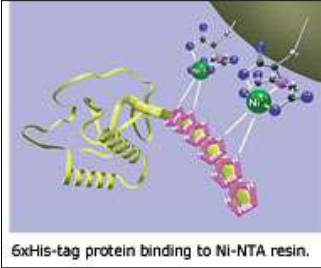
Confidential

89

Vaccine Design
Enhancing in-vivo performance



To enable detection of multi-epitope concatamer expression, engineer an expression tag at the end of the coding sequence immediately before the translation termination signal. In general, these tags should follow directly after the multi-epitope string



6xHis-tag protein binding to Ni-NTA resin.

6X histidine
HHHHHH
(preferred choice)

c-myc (myelocytomatosis cellular gene):
EQKLISEEDL

HA (Anti-Hemagglutinin Tag):
YPYDVPDYA


FLAG-tag:
DYKDDDDK

Confidential


90

Vaccine Design

Enhancing in-vivo performance



Enhance the detection of translation termination with a stop codon.



Confidential 91

iVAX Development Team



- William Martin, Chief Information Officer
 - Tel: (401) 272-2123 (ext. 102)
 - Email: martinb@epivax.com
- Frances Terry, Immunoinformatics Program Manager
 - Tel: (401) 272-2123 (ext. 125)
 - Email: fterry@epivax.com
- Gary Kurczy, iVAX Lead Programmer
 - Tel: (401) 272-2123 (ext. 128)
 - Email: gkurczy@epivax.com
- Matthew Ardito, Immunoinformatics Programmer/Analyst
 - Tel: (401) 272-2123 (ext. 128)
 - Email: mardito@epivax.com

Confidential 92