Genome to Vaccinome:
Immunoinformatics & Vaccine design case studies

Urmila Kulkarni-Kale
Bioinformatics Centre
University of Pune
urmila@bioinfo.ernet.in
Outline

• Immunology basics
• What is reverse vaccinology?
• Immunoinformatics
  – Databases (Knowledgebases)
  – Algorithms (B- and T-cell epitope predictions)
  – Predictions
• Case studies
  – Mumps virus
  – Japanese encephalitis virus
The Immune System

- body's defense against infectious organisms
- The Innate immunity: first line of defense
  - rapid nonspecific responses
  - recognition of conserved structures present in many microorganisms
    - lipopolysaccharides in bacterial cell walls or proteins in flagella
- The adaptive immune response: second line of defense
  - tailored to an individual threat
  - specific to an infectious agent
  - memory cells persist that enable a more rapid and potent response on ‘re-infection’
The adaptive immune response

• Stimulated by receptor recognition of a specific small part of an antigen known as an epitope

• Two major arms:
  – The humoral immune response of antibody-secreting B lymphocytes (B cell epitopes)
  – The cellular immune response of T lymphocytes (T cell Th epitopes)
  – Response stimulated by receptor recognition
Antigen presentation and recognition: molecular and cellular processes.
Host-Pathogen interactions: Surface proteins

- In case of Viruses:
  - Capsid
  - Envelope
  - Membrane
Antigen-Antibody (Ag-Ab) complexes

- Non-obligatory heterocomplexes that are made and broken according to the environment
- Involve proteins (Ag & Ab) that must also exist independently
- Remarkable feature:
  - high affinity and strict specificity of antibodies for their antigens.
- Ab recognize the unique conformations and spatial locations on the surface of Ag
- Epitopes & paratopes are relational entities
Methods to identify epitopes

1. **Immunochemical methods**
   - ELISA: Enzyme linked immunosorbent assay
   - Immunofluorescence
   - Radioimmunoassay

2. **X-ray crystallography**: Ag-Ab complex is crystallized and the structure is scanned for contact residues between Ag and Ab. The contact residues on the Ag are considered as the epitope.

3. **Prediction methods**: Based on the X-ray crystal data available for Ag-Ab complexes, the propensity of an amino acid to lie in an epitope is calculated.
Antigen-Antibody complex

Number of Ab-binding sites on an antigen

Number of antibodies that could be raised against an antigen

A few antibodies may have overlapping binding sites on same antigen
Ab-binding sites:
Sequential & Conformational Epitopes!
Properties of Epitopes

• They occur on the surface of the protein and are more flexible than the rest of the protein.
• They have high degree of exposure to the solvent.
• The amino acids making the epitope are usually charged and hydrophilic.
B cell epitope prediction algorithms:

- Hopp and Woods – 1981
- Welling et al – 1985
- Parker & Hodges - 1986
- Kolaskar & Tongaonkar – 1990
- Haste et al., 2006

T cell epitope prediction algorithms:

- Margalit, Spouge et al - 1987
- Rothbard & Taylor – 1988
- Stille et al – 1987
- Tepitope - 1999
Hopp & Woods method

• Pioneering work
• Based on the fact that only the hydrophilic nature of amino acids is essential for an sequence to be an antigenic determinant
• Local hydrophilicity values are assigned to each amino acid by the method of repetitive averaging using a window of six
• Accuracy: 45-55%
Welling’s method

• Based on the % of each aa present in known epitopes compared with the % of aa in the avg. composition of a protein.

• assigns an antigenicity value for each amino acid from the relative occurrence of the amino acid in an antigenic determinant site.

• regions of 7 aa with relatively high antigenicity are extended to 11-13 aa depending on the antigenicity values of neighboring residues.
Parker & Hodges method

- Utilizes 3 parameters:
  - Hydrophilicity: HPLC
  - Accessibility: Janin’s scale
  - Flexibility: Karplus & Schultz

- Hydrophilicity parameter was calculated using HPLC from retention coefficients of model synthetic peptides.

- Surface profile was determined by summing the parameters for each residue of a seven-residue segment and assigning the sum to the fourth residue.

- One of the most useful prediction algorithms
Kolaskar & Tongaonkar’s method

• Semi-empirical method which uses physiological properties of amino acid residues

• frequencies of occurrence of amino acids in experimentally known epitopes.

• Data of 169 epitopes from 34 different proteins was collected of which 156 which have less than 20 aa per determinant were used.

• Antigen: EMBOSS
CEP Server

• Predicts the conformational epitopes from X-ray crystals of Ag-Ab complexes.
• uses percent accessible surface area and distance as criteria
An algorithm to map sequential and conformational epitopes of protein antigens of known structure

- Percent accessible surface area (\%ASA) using Lee & Richards (1971) algorithm modified by Shrake & Rupley (1973) and implemented in InsightII is calculated.

- Residues having \%ASA $\geq$ 25 are termed as accessible residues.

- A contiguous stretch of more than three accessible residues is termed as an antigenic determinant (AD).

- Every antigenic determinant is extended to N- and C-termini, only if accessible amino acid(s) is present after an inaccessible amino acid residue.

- Antigenic determinants are listed with the details like chain ID, length of determinant, start and end position and the amino acid sequence.
The distance between every atom of residues from the $i^{th}$ determinant and every atom of residues from the $j^{th}$ determinants is calculated.

If the distance between any pair of atoms of at least two residues from sequentially distinct determinants is found to be $\leq$ predetermined cutoff distance, then $j^{th}$ determinant is termed to be part of a conformational epitope that consists of $i^{th}$ and $j^{th}$ determinants.

Such distance calculations are carried out for every sequential determinant ($j = 1, n$ and $j \neq i$) with $i^{th}$ determinant as a reference. The reference sequential determinant is then varied from 1 to $n$. The list of conformational epitopes is computed.

Individual accessible residues that are within the distance cutoff from the individual ADs making the CE are also included as a part of predicted CE.

List of CEs is printed in html format.
CE: Features

• The first algorithm for the prediction of conformational epitopes or antibody binding sites of protein antigens

• Maps both: sequential & conformational epitopes

• Prerequisite: 3D structure of an antigen
CEP: Conformational Epitope Prediction Server
http://bioinfo.ernet.in/cep.htm

Predicted AD

<table>
<thead>
<tr>
<th>AD No.</th>
<th>Antigenic Determinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_ Y</td>
<td>Y_1-KVFGRCE-7</td>
</tr>
<tr>
<td>2_ Y</td>
<td>Y_13-KRHGIDNyrRGyS-24</td>
</tr>
<tr>
<td>3_ Y</td>
<td>Y_37-NfNtQaTNRNMDG-49</td>
</tr>
<tr>
<td>4_ Y</td>
<td>Y_65-NDGRPGSRnLnNicPcSAILSSDITA-90</td>
</tr>
<tr>
<td>5_ Y</td>
<td>Y_100-SDGN-103</td>
</tr>
<tr>
<td>6_ Y</td>
<td>Y_112-RNRcKTGVQA-122</td>
</tr>
</tbody>
</table>

Predicted CE

<table>
<thead>
<tr>
<th>CE No</th>
<th>AD within 6A of Reference AD</th>
<th>Res within 6A of Ref AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y_1-KVFGRCE-7</td>
<td>Y_128: G</td>
</tr>
<tr>
<td></td>
<td>Y_37-NfNtQaTNRNMDG-49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_65-NDGRPGSRnLnNicPcSAILSSDITA-90</td>
<td>Y_129: L</td>
</tr>
<tr>
<td>2</td>
<td>Y_13-KRHGIDNyrRGyS-24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_112-RNRcKTGVQA-122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_65-NDGRPGSRnLnNicPcSAILSSDITA-90</td>
<td>Y_129: L</td>
</tr>
<tr>
<td></td>
<td>Y_100-SDGN-103</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Y_65-NDGRPGSRnLnNicPcSAILSSDITA-90</td>
<td>Y_97: K</td>
</tr>
<tr>
<td></td>
<td>Y_1-KVFGRCE-7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_37-NfNtQaTNRNMDG-49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_13-KRHGIDNyrRGyS-24</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Y_100-SDGN-103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_13-KRHGIDNyrRGyS-24</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Y_112-RNRcKTGVQA-122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y_13-KRHGIDNyrRGyS-24</td>
<td></td>
</tr>
</tbody>
</table>

Binding Site: 1FDL
- Select BS
- Select CE: CEB1, CEB2, CEB3, CEB4, CEB5 (CE<=xB5)
T-cell epitope prediction algorithms

- Considers amphipathic helix segments, tetramer and pentamer motifs (charged residues or glycine) followed by 2-3 hydrophobic residues and then a polar residue.
- Sequence motifs of immunodominant secondary structure capable of binding to MHC with high affinity.
- Virtual matrices are used for predicting MHC polymorphism and anchor residues.
MHC-Peptide complex
Epitome database

Epitome is a database of all known antigenic residues and the antibodies that interact with them, including a detailed description of residues involved in the interaction and their sequence/structure environments. Additionally, Interactions can be visualized using an interface into Jmol.

Publication

Complementary resources

- Dr. Andrew C.R. Martin's Group at UCL - general information about antigens and antibodies [http://www.bioinf.org.uk](http://www.bioinf.org.uk)
- The international ImMunoGeneTics information system [http://imgt.cines.fr/](http://imgt.cines.fr/)
- Darren Flower’s databases of quantitative functional peptide data for immunology:
  1. JenPep - [http://www.jenner.ac.uk/JenPep](http://www.jenner.ac.uk/JenPep)
  2. AntiJen - [http://www.jenner.ac.uk/antijen/](http://www.jenner.ac.uk/antijen/)
- BciPep - a database of B cell epitopes [http://www.imtech.res.in/raghava/bcipep](http://www.imtech.res.in/raghava/bcipep)
- The HIV Molecular Immunology Database [http://hiv-web.lanl.gov/content/immunology/index.html](http://hiv-web.lanl.gov/content/immunology/index.html)

© 2005 rostlab.org • columbia

• [http://cubic.bioc.columbia.edu/services/epitome/](http://cubic.bioc.columbia.edu/services/epitome/)
CED database

A brief introduction to CED

- What is CED?
- What is ... epitope?
- Why did you build CED?
- How did you construct CED?
- Why small in size at present?
- Browse, search, and view CED entries
- Are there any other epitope database?
- What will you do with CED in the future?
- CED logo, a little bit similar to KEGG logo?

- What is CED?

CED is short for Conformational Epitope Database.

- What is ... epitope?

http://web.kuicr.kyoto-u.ac.jp/~ced/intro.html
BciPep Database

Download BciPep Data

The data files of BciPep can be obtained by getting an username and password from the administrator. For getting username and password mail to raghava@imtech.res.in.

The data can be downloaded using NETSCAPE browser.

<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
<th>Get It!</th>
</tr>
</thead>
<tbody>
<tr>
<td>BciPep.dat</td>
<td>File contains the 2479 entries. Each entry is having information about peptide sequence, source protein, pathogen group, immunogenicity, neutralization, source protein, experimental method, model organism, database reference, antibody, publication reference and antigen structure.</td>
<td>Download</td>
</tr>
<tr>
<td>Non-redundant_bciPep.dat</td>
<td>The file has the unique entries of BciPep database.</td>
<td>Download</td>
</tr>
<tr>
<td>Antibody structure</td>
<td>The file contains atomic coordinate files (PDB files) of all antibodies available in BCiPep database.</td>
<td>Download</td>
</tr>
</tbody>
</table>
Welcome to the Antigen Antibody Interactions Database

Developed by Bioinformatics Centre, University of Pune, India. The database contains the data about the various types of interactions present in the Antigen and Antibody complex interaction sites, like Van der Waals, Hydrogen bonds, Salt Bridges etc. The primary data source is the Antigen-Antibody complexes present in the Protein Data Bank. In this version only those entries are considered where the antigens are proteins and...
Rational Vaccine design: Challenges & opportunities

Genomic Data of viruses

Variations/conservations

Antigen: 3D structure(s)

Epitope Prediction software

- Annotations
- Organisations
- Data mining

- Rules for predictions
- Accuracy related issues
- Experimental validations

- Relatively very few
- Modeling is only solution
Reverse Vaccinology workbench: list of parts

- The components are—
  - A curated genomic resource (VirGen). 2004
  - A server for prediction of epitopes (CEP) 1999; 2005
  - A knowledge-base to study Ag-Ab interactions (AgAbDb) 2007
  - A server for variability analyses (PVIS) 2009
  - A derived database of 3D structures of viral proteins
    - Compilation of experimental structures of viral proteins from PDB
    - Predicted structures using homology modeling approach 1999; 2007

Study of sequence→structure→function (antigenicity) to identify & prioritize vaccine candidates
VirGen home

Menu to browse viral families
Search using Keywords & Motifs
Genome analysis & Comparative genomics resources
Guided tour & Help

Navigation bar

http://bioinfo.ernet.in/virgen/virgen.htm
Sample genome record in VirGen

Retrieve sequence in FASTA format

Tabular display of genome annotation

<table>
<thead>
<tr>
<th>VirGen Annotation</th>
<th>Alternate Names</th>
<th>Length of Protein</th>
<th>Residue (start-end)</th>
<th>Base (start-end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyprotein</td>
<td>Polyprotein</td>
<td>3432</td>
<td>1..3432</td>
<td>96..10394</td>
</tr>
<tr>
<td>ancC</td>
<td>anchored capsid protein (core)</td>
<td>127</td>
<td>1..127</td>
<td>96..476</td>
</tr>
<tr>
<td>C</td>
<td>anchored capsid protein (C)</td>
<td>105</td>
<td>1..105</td>
<td>96..410</td>
</tr>
<tr>
<td>prM</td>
<td>putative anchored core protein (C)</td>
<td>167</td>
<td>128..294</td>
<td>477..977</td>
</tr>
<tr>
<td>V3</td>
<td>V3 (50 kd membrane-associated glycoprotein)</td>
<td>75</td>
<td>220..294</td>
<td>753..977</td>
</tr>
</tbody>
</table>

‘Alternate names’ of proteins
Graphical view of Genome Organization

Viral polyprotein along with the UTRs

Graphical view generated dynamically using Scalable Vector Graphics technology
### Multiple Sequence Alignment

#### MSA

<table>
<thead>
<tr>
<th>Genus</th>
<th>Protein</th>
<th>Cluster Alignment</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gagC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Link for batch retrieval of sequences

![Dendrogram](image)

[MSA_Genus](http://example.com)
Browsing the module of Whole Genome Phylogenetic trees

Most parsimonious tree of genus Flavivirus
Input data: Whole genome
Method: DNA parsimony
Bootstrapping: 1000
Welcome to the Antigen Antibody Interactions Database

Developed by Bioinformatics Centre, University of Pune, India. The database contains the data about the various types of interactions present in the Antigen and Antibody complex interaction sites, like Van der Waals, Hydrogen bonds, Salt Bridges etc. The primary data source is the Antigen-Antibody complexes present in the Protein Data Bank. In this version only those entries are considered where the antigens are proteins and...
AgAbDB: summary of interacting residues

General Statistics

Query to get quick summary about the interacting residues in Antigen and Antibody

**Antibody Information**

The Number of Interaction for Antibody residue

<table>
<thead>
<tr>
<th>Residue</th>
<th>CHAIN</th>
<th>KABAT NUMBER</th>
<th>PDB NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>L</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Tyr</td>
<td>L</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Tyr</td>
<td>L</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Tyr</td>
<td>H</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td>Tyr</td>
<td>L</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>

**Antigen Information**

The Number of Interaction for Antigen residue

<table>
<thead>
<tr>
<th>Residue</th>
<th>CHAIN</th>
<th>KABAT NUMBER</th>
<th>PDB NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
</tbody>
</table>

**Antibody**

<table>
<thead>
<tr>
<th>RESIDUE</th>
<th>CHAIN</th>
<th>KABAT NUMBER</th>
<th>PDB NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>H</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>His</td>
<td>L</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Tyr</td>
<td>L</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Leu</td>
<td>L</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>His</td>
<td>L</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>Tyr</td>
<td>H</td>
<td>52</td>
<td>50</td>
</tr>
</tbody>
</table>

**Antigen**

<table>
<thead>
<tr>
<th>RESIDUE</th>
<th>CHAIN</th>
<th>KABAT NUMBER</th>
<th>PDB NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>P</td>
<td>664</td>
<td></td>
</tr>
</tbody>
</table>
Interactions mapped on structure

Get Interactions

- Summary
- Residue Level
- Burried Residue
- Atomic Level
- View Interactions
- Statistics
- PPI Output

The Interaction View for 1TJI

Click on Buttons to display interacting residues with

- Whole Complex
- All Interactions
- van der Waal's
- Hydrogen Bonds
- Salt Bridges

Courtesy: Jmol
Study of variations at different levels of Biocomplexity

- Strains/isolates of a virus
- Serotypes of a virus
- Viruses that belong to same genus
- Viruses that belong to same family

Implications of variations in designing vaccines
Protein Variablility Index Server (PVIS)
Beta test version

- PVIS takes MSA as an input and calculates variability of amino acids using Wu-Kabat’s coefficient at each position of the consensus sequence

- Features:
  - Interactive, GUI based alignment output format
  - No limit on input length of MSA
  - At each position of alignment, user can view consensus residue and its corresponding variability
  - Generates CSV (Comma Separated File) of Variability values against their positions in consensus sequence
Various output formats

\[ VI = \frac{\text{number of different amino acids at } \text{i}^{\text{th}} \text{ position}}{\text{frequency } (F) \text{ of the most common amino acid at } \text{i}^{\text{th}} \text{ position}} \]

where

\[ F = \frac{\text{Number of occurrences of most common amino acid at } \text{i}^{\text{th}} \text{ position}}{\text{total number of sequences considered for multiple alignment}} \]
Antigenic diversity of mumps virus: an insight from predicted 3D structure of HN protein
Mumps Virus: at a glance

Source: VirGen database

Order: Mononegavirales
Family: Paramyxoviridae
Subfamily: Paramyxovirinae
Genus: Rubulavirus
Species: Mumps virus

Genome: -ve sense ssRNA
Genotypes: 10: A → J (SH gene)
Known antigenic proteins: F & HN
SBL-1 HN:

Fold: β propeller
Monomer: 6 bladed propeller with 4-stranded β sheet & 4 helices

Helices: Red  Strands: yellow  Turns: blue, Coils: green
A new site for neutralisation: mapping antigenicity using parts list approach

<table>
<thead>
<tr>
<th>Vaccine Strain</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lzagreb</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>L-zagmas</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>SIPAR02</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>Urabev</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>SkBv</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>Miya</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>Hoshino</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>Rubv</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>Rub1v</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>j12</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>j12m</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>j12m1</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>j15</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>j15ma</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
<tr>
<td>j15ma1</td>
<td>QRSTSWWPYELLYEISFTFTN8GQSVPNMSWIPIYSFTRPG9GNC8GENVCPTACVSAY 480</td>
</tr>
</tbody>
</table>

Total variations: 47

Hypervariable region of HN identified using MSA of Vaccine strains (Majority marked with yellow screen)

Residues 462, 464, 468, 470, 473, 474 present on surface; Known escape mutants are in proximity
Mapping mutations on 3D structure of Mumps virus: a case study

Colour: according to majority
• Case study: Design & development of peptide vaccine against Japanese encephalitis virus
We Have Chosen JE Virus, Because

- JE virus is endemic in South-east Asia including India.

- JE virus causes encephalitis in children between 5-15 years of age with fatality rates between 21-44%.

- Man is a "DEAD END" host.
We Have Chosen JE Virus, Because

- Killed virus vaccine purified from mouse brain is used presently which requires storage at specific temperatures and hence not cost effective in tropical countries.
- Protective prophylactic immunity is induced only after administration of 2-3 doses.
- Cost of vaccination, storage and transportation is high.
Predicted structure of JEVS
Mutations: JEVN/JEVS
CE of JEVN Egp
Species and Strain specific properties: TBEV/ JEVN/JEVS

• Loop1 in TBEV: LA EEH QGGT
• Loop1 in JEVN: HN EKR ADSS
• Loop1 in JEVS: HN KKR ADSS

Antibodies recognising TBEV and JEVN would require exactly opposite pattern of charges in their CDR regions.

Further, modification in CDR is required to recognise strain-specific region of JEVS.
Multiple alignment of Predicted T\textsubscript{H}-cell epitope in the JE_Egp with corresponding epitopes in Egps of other Flaviviruses

\begin{align*}
\text{JE} & \quad \text{DFGSIGGVFNSIKKAVHQVF} \text{GAFRTLF} \text{FGGMS} \\
\text{MVE} & \quad \text{DFGSVGGVFNSIKKAVHQVF} \text{GAFRTLF} \text{FGGMS} \\
\text{WNE} & \quad \text{DFGSVGGVFTSVGKAIHQVF} \text{GAFRSLF} \text{FGGMS} \\
\text{KUN} & \quad \text{DFGSVGGVFTSVGKAVHQVF} \text{GAFRSLF} \text{FGGMS} \\
\text{SLE} & \quad \text{DFGSIGGVFNSIKKAVHQVF} \text{GAFRTLF} \text{FGGMS} \\
\text{DEN2} & \quad \text{DFGSLGGVFTSIGKALHQVF} \text{GAIYGAFFSGVS} \\
\text{YF} & \quad \text{DFFSSAGFFTSVGBKIGHTVF} \text{GSAFQGLFGGLN} \\
\text{TBE} & \quad \text{DFGSAGGFLSSIKKAVHTVLG} \text{GAFNSIFGGVG} \\
\text{COMM} & \quad \text{DFSGG} \quad \text{S} \quad \text{GK} \quad \text{H} \quad \text{V} \quad \text{G} \quad \text{F} \quad \text{G}
\end{align*}

Multiple alignment of JE_Egp with Egps of other Flaviviruses in the YSAQVGASQ region.

\begin{align*}
\text{JE} & \quad \text{SENHGNYSAQVGASQAAKFTITPNAPSITLKLG} \\
\text{MVE} & \quad \text{STSHGNYSTQIGANQAVRFTISPNAPIATAKMG} \\
\text{WNE} & \quad \text{VESHG----KIGATQAGRFSITPSAPSYTLKLG} \\
\text{KUN} & \quad \text{VESHGNYFTQTGAQQAGRFSITPAAPSYTLKLG} \\
\text{SLE} & \quad \text{STSHGNYSEQIQKNQQARFTISPQAPSFTANMG} \\
\text{DEN2} & \quad \text{HAVGNDTG------KHGKEIKITPQSSTTEAELT} \\
\text{YF} & \quad \text{QENWN--------TDIKTLKFDALSGSQEVEFI} \\
\text{TBE} & \quad \text{VAANETHS--------GRKTASFTIS--SEKTIILTNG}
\end{align*}
Peptide Modeling

Initial random conformation
Force field: Amber
Distance dependent dielectric constant $4r_{ij}$
Geometry optimization: Steepest descents & Conjugate gradients
Molecular dynamics at 400 K for 1ns
Peptides are:

`SENHGNYSAQVGASQ`
`NHGNYSAQVGASQ`
`YSAQVGASQ`
`YSAQVGASQAAKFT`
`NHGNYSAQVGASQAAKFT`
`SENHGNYSAQVGASQAAKFT`
Peptide Conformation
Region GASQ is shown in white.

YSAQVGASQ
NHGNYSQAQVGASQ
SENHNYSQAQVGASQ

YSAQVGASQAAKFT
NHGNYSQAQVGASQAAKFT
SENHNYSQAQVGASQAAKFT
Peptide Conformation
Region GASQ is shown in white.
Publications


Acknowledgements

• Prof. A. S. Kolaskar
• Ms. G. Sunitha Manjari, Bhakti Bhawat, Surabhi Agrawal & Shriram Bhosle
• M.Sc. / ADB Students@bioinfo
• Ms. Sangeeta Sawant, & Dr. M. M. Gore
• Ms. Janaki Oza, Prof. Deepti Deobagkar, Dr. Mallya, Dr. Dhere & Dr. Kapre

• Financial support:
  − Center of excellence (CoE) by both MCIT & DBT, Govt. of India
  − M.Sc. Bioinformatics programme from DBT, Govt. of India
  − Molecular modeling facility at Bioinformatics centre, University of Pune
  − Serum Institute of India

Thank you all!
Bioinformatics Centre

@ University of Pune

http://bioinfo.ernet.in
HRD Activities
In Bioinformatics and Biotechnology

Short Term Courses
Long Term Courses
Long Term Courses

M.Sc. Bioinformatics
Advanced Diploma in Bioinformatics (On hold)
CRCDM (PPP model)
Credit exchange program:
M.Sc. Zoology & Biotechnology
Contributory teaching:
M.Sc./M.Tech. Biotechnology (Integrated)
M.B.A. Biotechnology
M. Sc. Bioinformatics

- Started in 2002
- Masters level
- 2 years (4 Semesters), full time
- 25 credits/semester + Project (16 credits)
- Intake thru entrance test
- No. of students: 30+1+2
### Integrated Course Management System (ICMS)

#### Bioinformatics Centre, University of Pune

**Monday, 28th December 2009**

---

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30</td>
<td><strong>BIM303(T)</strong> SHK</td>
<td><strong>BIM303(P)</strong> CGP</td>
<td><strong>BIM307(T)</strong> TTDM</td>
<td><strong>BIM307(T)</strong> TKJ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LH1</td>
<td>CL2</td>
<td>VS</td>
<td>LH1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:30</td>
<td><strong>BIM301(P)</strong> UKK</td>
<td><strong>BIM302(P)</strong> CL2</td>
<td><strong>BIM307(T)</strong> TKJ</td>
<td><strong>BIM307(T)</strong> TKJ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TAP</td>
<td>CL2</td>
<td>SHK</td>
<td>VS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td><strong>BIM301(P)</strong> UKK</td>
<td><strong>BIM302(P)</strong> CL2</td>
<td><strong>BIM303(T)</strong> CGP</td>
<td><strong>BIM303(T)</strong> CGP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TAP</td>
<td>CL2</td>
<td>UKK</td>
<td>LH1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:30</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:30</td>
<td><strong>BIM302(P)</strong> UKK</td>
<td></td>
<td><strong>BIM305(T)</strong> OORD</td>
<td><strong>BIM303(P)</strong> CGP</td>
<td><strong>BIM303(T)</strong> CGP</td>
<td><strong>BIM303(T)</strong> CGP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GDV</td>
<td>CL2</td>
<td>AK</td>
<td>LH1</td>
<td>WNG</td>
<td>WNG</td>
<td>CL2</td>
</tr>
<tr>
<td>15:30</td>
<td><strong>BIM302(P)</strong> UKK</td>
<td></td>
<td><strong>BIM305(T)</strong> OORD</td>
<td><strong>BIM303(P)</strong> CGP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GDV</td>
<td>CL2</td>
<td>AK</td>
<td>LH1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:30</td>
<td><strong>BIM302(P)</strong> UKK</td>
<td></td>
<td><strong>BIM305(T)</strong> OORD</td>
<td><strong>BIM303(P)</strong> CGP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GDV</td>
<td>CL2</td>
<td>AK</td>
<td>LH1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BINC: Bioinformatics National Certification examination

Instituted By
Department of Biotechnology
Government of India
New Delhi

BioInformatics National Certification Examination

Monday, December 28, 2009 | 5:09:06 AM

Registration starts: 01 December 2009 | Registration closes: 16 January 2010 | Examination: 20-21 February 2010

Attention: Registration for BINC Examination 2010 starts on 1 December 2010.

BioInformatics National Certification (BINC) Examination will be conducted on 20-21 February 2010. The Department of Biotechnology (DBT), Government of India has instituted this examination with an objective of certifying Bioinformatics professionals and to facilitate industries and potential employers for recruitment. University of Pune has been identified as a nodal agency by the Department of Biotechnology, Govt. of India to coordinate this examination along with six centres namely, Jawaharlal Nehru University, New Delhi; Anna University, Chennai; West Bengal University of Technology, Kolkata; Institute of Bioinformatics & Applied Biotechnology, Bangalore; North-Eastern Hill University, Shillong and University of Hyderabad, Hyderabad.

In the BINC 2009 examination, 519 candidates appeared out of which 30 candidates were certified.

The DBT awards DBT-BINC-Junior Research Fellowships (DBT-BINC-JRF) to top 15 BINC qualified Indian nationals in the order of merit to pursue Ph.D. in Indian Institutes/Universities. Note that the candidate must possess a postgraduate degree & meet the criteria of the institutes/universities in order to avail the DBT-BINC-Junior Research Fellowships (DBT-BINC-JRF). In addition, a cash prize of Rs. 10,000/- will be awarded to the top 10 BINC qualified Indian nationals.

“It is a privilege and an honor to be the coordinator of the BINC examination. It is just not a responsibility but a commitment to make a difference in the media and the students alike to publicize the BINC exam so that both, the bioinformatics students and the industries can benefit from this certification at the national level.”

The Coordinator BINC

~850 Registrations

13700 HITS
Thank You